



HAL
open science

La rentabilité de la fidélisation du consommateur : 3 essais complémentaires.

Thierry Vallaud

► **To cite this version:**

Thierry Vallaud. La rentabilité de la fidélisation du consommateur : 3 essais complémentaires.. Gestion et management. Conservatoire national des arts et métiers - CNAM, 2013. Français. NNT : 2013CNAM0873 . tel-00877355

HAL Id: tel-00877355

<https://theses.hal.science/tel-00877355v1>

Submitted on 28 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECOLE DOCTORALE ABBE GREGOIRE

Laboratoire Interdisciplinaire de Recherche en Science de l'Action

THÈSE présentée par :

Thierry Vallaud

soutenue le : 19 juin 2013

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Discipline/Spécialité : Sciences de gestion

La rentabilité de la fidélisation du consommateur
3 essais complémentaires

THÈSE dirigée par :

M. AIMETTI Jean-Paul
M. MERUNKA Dwight

Professeur, *Titulaire de la Chaire Marketing, CNAM*
Professeur, Aix-Marseille Université et Euromed Management

RAPPORTEURS :

M. DESMET Pierre
M. VALETTE-FLORENCE Pierre

Professeur, Université Paris-Dauphine
Professeur, Université de Grenoble 2

JURY :

M. ATTALI François
M. SAPORTA Gilbert

Professeur : Directeur Marketing Stratégique, Recherche et Innovation SODIAAL
Professeur Titulaire de la Chaire Statistique appliquée, *CNAM*

Remerciements :

En tout premier lieu nous tenons à exprimer notre gratitude envers le professeur Jean Paul Aimetti titulaire de la chaire Marketing du CNAM notre directeur de Doctorat et professeur Dwight Merumka d'Euromed notre directeur de Phd.

Nous remercions également les membres du jury.

Nos pensées vont aussi à Mr Christian Delom et Mr Jacques Paget qui ont encouragé ce travail de recherche ainsi qu'à mes collègues François Rosset et Abdelatif Oualouane qui m'ont donné de nombreux conseils statistiques.

Une pensée est souvent impactée par des rencontres, Sylvain Forestier, Daniel Larose, Peter Horn qui ont inspiré et guidé la plupart de nos travaux précédents. Nos collègues et clients ont également souvent participé à l'amélioration de notre réflexion et nous ont donné la matière pour réaliser la plupart des applications concrètes.

Ce travail ne serait rien sans ses relectrices assidues tant sur la forme que sur le fond : Catherine Roz et Sarah Fulcons.

La recherche est un travail personnel souvent difficile à supporter pour l'entourage, nous remercions à nouveau Catherine, mes enfants, Lucas et Felix et ma mère.

Résumé :

Dans cette thèse sur travaux, l'auteur part de deux travaux précédents sur la rentabilité de la fidélisation et la détermination du potentiel client pour faire un constat : une partie de la rentabilité de la fidélisation et du potentiel pour une marque est basée sur la part captable du chiffre d'affaire fait à la concurrence ; le taux de captation.

Dans ce nouveau travail il s'agit de montrer que le taux de captation est basé sur l'élasticité du taux de nourriture. A partir d'une analyse de la littérature et de plusieurs modélisations sur des données de panel scannées, l'auteur démontre, sur plusieurs marchés, que l'élasticité du taux de nourriture est contrainte et prévisible.

C'est donc en tenant compte de cet écart limité qu'une marque peut estimer le taux de captation et donc la rentabilité de la fidélisation ainsi que le potentiel client.

Mots-clés : fidélisation, taux de nourriture, élasticité, potentiel client, LTV, taux de captation, méthode des clones, prévision, panel de home scanning, ROI, valeur client, gros volume de données, modèle désagrégé

Résumé en anglais :

In this thesis based on works the author goes from two previous studies on the profitability of loyalty and customer potential determination to make a statement : part of the profitability of loyalty and of the potential for a brand is based on the reachable share of turnover done by the competition ; the catch rate.

In this new work it is shown that the catch rate is based on the elasticity of the share of category requirement. From a review of the literature and several modeling on scanning panel data the author demonstrates on several markets that elasticity of the share of category requirement is limited and predictable.

Then it's in taking into account this small difference that a brand can estimate the "catchable" rate and therefore the profitability of loyalty and potential of a customer.

Mots-clés : Loyalty, share of requirement, elasticity, customer's potential, LTV, catch rate, clones method forecasting, home scanning panel home, ROI, customer value, large volume of data (big data), disaggregated model

Table des matières :

Remerciements :	2
Résumé :	3
Résumé en anglais :	4
Table des matières :	5
Liste des tableaux :	10
Liste des figures :	13
Liste des graphiques :	15
Liste des annexes :	16
1. Introduction :	17
Introduction générale :	18
2. Synthèse de nos travaux :	18
2.1 LA FIDELISATION RENTABLE : PROPOSITION D'UN MODELE COMPOSITE ENTRE LA VALEUR D'UN CLIENT A L'INSTANT T ET SA VALEUR A VIE :	18
2.2 ESTIMATION DU POTENTIEL INDIVIDUEL DE CHIFFRE D'AFFAIRES EN UTILISANT LES DONNEES ISSUES D'UNE BASE DE DONNEES CLIENTS :	21
2.3 L'ELASTICITE DU TAUX DE NOURRITURE EST-ELLE CONTRAINTE ?	23
3. L'analyse de la littérature :	25
3.1 Définitions de quelques concepts :	25
3.1.1 La fidélisation :	25
3.1.2 La fidélité :	25
3.1.3 Le potentiel client :	25
3.1.4 Le taux de captation :	25
3.1.5 Le taux de nourriture :	26
3.2 L'importance de la fidélité pour les marques :	27
3.3 Ensemble de considération/ensemble évoqué/ensemble déterminant :	28
3.4 L'intensité des préférences :	28
3.5 Les comportements de remplacements (« <i>switching behaviour</i> ») :	29
3.6 La recherche de variété :	29
3.7 Les effets de satiété :	29
3.8 Les effets promotionnels :	29
3.9 Les actions marketing de toutes les marques d'un marché :	29

3.10	Le « <i>double jeopardy</i> » d'Erenberg et Uncles :.....	30
3.11	La prise en compte de la dimension temporelle :.....	31
3.12	La prise en compte des facteurs socio démographiques :	31
1.	<i>Recherche 1</i> : L'élasticité du taux de nourriture est-elle contrainte ?	32
	Introduction	33
1.1	Les questions de recherche :.....	34
1.2	La méthodologie :.....	34
1.2.1	Les données :.....	34
1.2.2	Les données de panel :	34
1.3	L'analyse préalable et les modélisations utilisées :.....	35
1.3.1	L'analyse descriptive des données :	35
1.3.2	Les modèles ARIMA :	35
1.3.3	Le perceptron multicouche :.....	36
1.3.4	Les techniques d'analyse du panier de la ménagère :	36
1.3.5	Séquence :	37
1.3.6	La modélisation de données de panel :.....	37
1.4	Le processus de validation des modèles.....	37
1.5	Résultats de la recherche :.....	37
1.5.1	L'analyse descriptive de données issues d'un panel de consommateurs :	37
1.6	Les principaux BPM (Brand Performance Measurement) étudiés :	37
1.7	Les évolutions par année :.....	38
1.8	L'évolution par année sur différentes périodes inter achats.....	39
1.9	Prévoir le taux de nourriture.....	42
1.10	Deux approches agrégées directes :	42
1.11	Une approche désagrégée indirecte :.....	43
1.11.1	Les approches agrégées :.....	43
1.11.2	Les approches désagrégées :	57
1.11.1	La cooccurrence d'achat :	57
1.11.2	La technique d'analyse de panel :	59
1.12	Le taux de nourriture selon le niveau de revenu	72
1.13	Discussion et conclusions.....	77
1.13.1	Réponse aux questions de recherche.....	77
1.13.2	Peut-on prévoir le taux de nourriture sur différentes durées de consommation ?.....	78

1.13.3	L'élasticité du taux de nourriture varie t'elle d'une catégorie de produit à l'autre :.....	78
1.13.4	L'élasticité bornée du taux de nourriture est-elle socialement explicable ?.....	79
1.14	Contribution et limites de notre recherche :	79
1.14.1	Les contributions :.....	79
1.14.2	Les limites :	79
1.15	Conclusions :.....	80
2.	<i>Recherche 2</i> : La fidélisation rentable : proposition d'un modèle composite entre la valeur d'un client à l'instant T et sa valeur à vie.....	81
	Avant-propos.....	82
	Introduction:	83
2.1	Contexte managérial et problématique de recherche	88
2.1.1	Importance de la fidélisation et du CRM	88
2.1.2	Les présupposés de la rentabilité de la fidélisation	88
2.1.3	Le client et ses différentes valeurs :	88
2.2	La durée de vie du client	89
2.3	Les modèles composites.....	95
2.4	Définition de nos questions de recherche.....	103
2.5	Modélisation du roi de la fidélisation à court terme (Objectif fidélisation®).....	103
2.6	L'analyse structurelle du marché	104
2.7	La part de marché de la Marque X (en valeur).....	105
2.8	La notoriété de la Marque X	105
2.9	Les cibles à fidéliser.....	106
2.9.1	Le nombre d'actes d'achat par acheteur.....	106
2.9.2	Les quantités achetées de produits d'Hygiène Beauté par acheteur.....	106
2.9.3	Les sommes dépensées pour les produits d'hygiène beauté par acheteur.....	106
2.10	Le taux de nourriture	107
2.11	Le nombre d'acheteurs	107
2.12	La quantité achetée par acte d'achat.....	107
2.13	Les marques « clef d'entrée »	109
2.14	Taille de la cible	110
2.15	Tableaux de simulation	112
2.16	ROI.....	119
2.17	Tableau de pilotage	119
2.18	Conclusion.....	119

2.19	Modélisation du roi de la fidélisation à long terme (la détermination de la LTV)	120
2.20	Durée de vie du consommateur	122
2.21	Captation à la marque et total marché :	126
2.22	Calcul des élasticités	126
2.23	Conclusion.....	128
2.24	Convergence et complémentarité des modèles	128
2.24.1	Mise en évidence des convergences	128
2.24.2	La valorisation du client :	129
2.24.3	Calcul de la VTC :.....	131
2.24.4	Une cible très identifiable :	136
2.24.5	Rappel sur le taux de couverture de cible d'Objectif Fidélisation [®] :.....	139
2.24.6	Elasticité et interprétation :	141
2.24.7	Part de la marge allouable au CRM :	143
2.24.8	Mise en perspective des coûts et des gains :	145
2.24.9	Choix définitif des cibles à potentiel :.....	145
2.24.10	Conclusion :.....	145
2.25	Réponses nos questions de recherche :	145
2.25.1	Tests et validation des questions :.....	145
2.26	Mise en évidence des lois sous-jacentes :	147
2.27	Interprétation des aspects complémentaires des approches sur un plan managérial:.....	147
2.28	Conclusion.....	148
3.	<i>Recherche 3</i> : Estimation du potentiel individuel de chiffre d'affaires en utilisant des données issues d'une base de données clients.....	150
3.1	Le contexte	151
3.2	Notre sujet de thèse	153
3.3	L'application précise du modèle	153
3.4	Les questions de recherche.....	154
3.5	Le processus de data mining utilisé.....	154
3.6	La compréhension des données	154
3.7	La préparation des données	155
3.8	Les modèles de classification et la détermination du potentiel client	158
3.8.1	Méthode des réseaux de Kohonen.....	158
3.9	Le développement du modèle	160

3.9.1	Les objectifs et la méthodologie.....	160
3.9.2	La fiabilité de la méthode de Kohonen	162
3.9.3	LA ROBUSTESSE DE LA CLASSIFICATION DE KOHONEN	164
3.9.4	Calcul des potentiels.....	174
3.9.5	Principaux résultats	177
3.9.6	Résumé des résultats	182
3.10	Les procédures de validation de nos modèles	183
3.11	Conclusions	184
3.11.1	Discussion des résultats de l'étude.....	184
3.11.2	Les limites et la contribution de notre étude	184
3.11.3	Pistes de recherche	184
4.	Conclusion générale	185
	ANNEXES	195
	Annexe 1 : Question sur la mesure de la valeur client par les entreprises françaises.....	196
	Annexe 2 : Glossaire CRM.	199
	Annexe 3 : Tableaux, graphiques et figures.....	210

Liste des tableaux

Tableau 1 : Les principaux BMP sur les catégories étudiées sur l'année 1	39
Tableau 2 : Les principaux BMP sur les catégories étudiées sur l'année 2	39
Tableau 3 : Evolution des principaux BMP sur les catégories étudiées par année	40
Tableau 4 : Indicateurs de dispersion du TN par marque.....	41
Tableau 5 : Récapitulatif des taux de nourriture de la marque 1 sur différentes périodes	41
Tableau 6 : Récapitulatif des taux de nourriture de la marque 2 sur différentes périodes	42
Tableau 7 : Résultat du modèle ARIMA, marque 1, données hebdomadaires	47
Tableau 8 : Poids et effets des variables dans le modèle ARIMA, marque 1, données hebdomadaires	48
Tableau 9 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 1, données hebdomadaires.....	49
Tableau 10 : Qualité de l'ajustement du modèle ARIMA , marque 1, données hebdomadaires	49
Tableau 11 : Résultat du modèle ARIMA, marque 1 sur 4 semaines (mois).....	50
Tableau 12 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 1 sur 4 semaines (mois).....	50
Tableau 13 : Qualité de l'ajustement du modèle ARIMA , marque 1 sur 4 semaines (mois).....	50
Tableau 14 : Résultat du modèle ARIMA, marque 14, données hebdomadaires	51
Tableau 15 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 14, données hebdomadaires.....	51
Tableau 16 : Qualité de l'ajustement du modèle ARIMA de la marque 14, données hebdomadaires	52
Tableau 17 : Résultat du modèle ARIMA, marque 14 sur 4 semaines (mois).....	52
Tableau 18 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 14 sur 4 semaines (mois).....	52
Tableau 19 : Qualité de l'ajustement du modèle ARIMA de la marque 14 sur 4 semaines (mois)...	53
Tableau 20 : Qualité de l'ajustement du Perceptron, marque 1, données hebdomadaires.....	54
Tableau 21 : Qualité de la prévision du Perceptron sur l'échantillon de validation, marque 1, données hebdomadaires.....	55
Tableau 22 : Qualité de l'ajustement du Perceptron, marque 1, données hebdomadaires	55
Tableau 23 : Qualité de l'ajustement du Perceptron de la marque 14 sur 4 semaines (mois)	56
Tableau 24 : Qualité de la prévision du Perceptron sur l'échantillon de validation de la marque 14 sur 4 semaines (mois).....	56
Tableau 25 : Qualité de l'ajustement du Perceptron au mois de la marque 14 sur 4 semaines (mois)	57
Tableau 26 : R ² des différents modèles agrégés	57
Tableau 27 : Taux de nourriture réel moyen des différents modèles agrégés entre le réel et le prédit	57
Tableau 28 : Cooccurrences des autres marques avec la marque 1 comme conséquence	58
Tableau 29 : Les plus fortes cooccurrences avec la marque 1 comme conséquence	58
Tableau 30 : Les plus fortes cooccurrences avec la marque 14 comme conséquence	59
Tableau 31 : Les indicateurs clefs en actes d'achats de la marque 1 à la semaine	60
Tableau 32 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 1 à la semaine.....	60
Tableau 33 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 1 à la semaine.....	60
Tableau 34 : Les résultats du modèle désagrégé de la marque 1 à la semaine.....	62
Tableau 35 : Résultats du modèle désagrégé pour la marque 1 à la semaine	63

Tableau 36 : Résultats du modèle que sur les effets intra pour la marque 1 à la semaine	64
Tableau 37 : Validation des résultats entre le prédit et le réel pour de la marque 1 à la semaine.....	65
Tableau 38 : Qualité de l'ajustement pour de la marque 1 à la semaine	65
Tableau 39 : Résultats du modèle pour la marque 1 sur 4 semaines (mois)	66
Tableau 40 : Validation des résultats entre le prédit et le réel de la marque 1 sur 4 semaines (mois)	66
Tableau 41 : Qualité de l'ajustement du modèle désagrégé pour la marque 1 sur 4 semaines (mois)	67
Tableau 42 : Les indicateurs clefs en actes d'achats de la marque 14 à la semaine.....	67
Tableau 43 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 14 à la semaine.....	68
Tableau 44 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 14 à la semaine.....	68
Tableau 45 : Résultats du modèle désagrégé de la marque 14 à la semaine	68
Tableau 46 : Validation des résultats entre le prédit et le réel du modèle désagrégé de la marque 14 à la semaine	69
Tableau 47 : Résultats du modèle désagrégé de la marque 14 sur 4 semaines (mois).....	70
Tableau 48 : Validation des résultats entre le prédit et le réel du modèle désagrégé de la marque 14 sur 4 semaines (mois).....	70
Tableau 49 : Qualité de l'ajustement du modèle désagrégé de la marque 14 sur 4 semaines (mois).	71
Tableau 50 : Synthèse des résultats des modèles désagrégés.....	72
Tableau 51 : Les groupes de la typologie.....	73
Tableau 52 : Tris croisés entre les groupes de la typologie et les classes TN et prix	73
Tableau 53 : Prix par segment.....	74
Tableau 54 : Croisement des segments de prix par PMG	74
Tableau 55 : Taux de nourriture par segment de prix et PMG.....	74
Tableau 56 : Typologie croisant le prix et le taux de nourriture	75
Tableau 57 : Types croisés par PMG	76
Tableau 58 : Types croisés par niveaux de vie en indice	76
Tableau 59 : Types croisés par nombre de personnes au foyer en indice	77
Tableau 60 : Elasticité des taux de nourriture et écarts types des élasticités par année et par catégorie de produits	77
Tableau 61 : Ecart types du taux de nourriture sur différentes périodes par année	78
Tableau 62 : R ² des différents modèles testés	78
Tableau 63 : Taux de nourriture moyen entre le prédit et le réel quels que soient les modèles à la semaine et au mois	78
Tableau 64 : Types croisés par nombre de personnes au foyer en indice	79
Tableau 65 : L'évolution du concept de marketing direct par année	85
Tableau 66 : L'importance du taux de nourriture	95
Tableau 67 : Evolution du taux d'activabilité.....	99
Tableau 68 : Elasticité et taux de captation.....	100
Tableau 69 : La notion de produits clefs d'entrée et de produits supports	101
Tableau 70 : La pénétration de la Marque X.....	105
Tableau 71 : Parts de marché	105
Tableau 72 : Le nombre d'actes d'achat par acheteur.....	106
Tableau 73 : Nombre de produits.....	106
Tableau 74 : Sommes dépensées	106
Tableau 75 : Le taux de nourriture de la Marque X par segment PMG.....	107
Tableau 76 : Le nombre d'acheteurs.....	107
Tableau 77 : La quantité achetée par acte d'achat	107

Tableau 78 : Les taux d'acheteurs, d'actes et de CA par segment PMG	108
Tableau 79 : Tableau récapitulatif.....	108
Tableau 80 : Elasticité de 20%	109
Tableau 81 : Rappel des pénétrations.....	109
Tableau 82 : Tableau de simulation : partie coût	112
Tableau 83 : Tableau de simulation : partie marché	112
Tableau 84 : Evolution de la marge brute en fonction des quantités	113
Tableau 85 : Simulation sur les gros consommateurs	113
Tableau 86 : Zoom sur les contributions par PMG : gros consommateurs	114
Tableau 87 : Zoom sur les contributions par PMG : Moyens consommateurs	115
Tableau 88 : Zoom sur les contributions par PMG : Petits consommateurs	117
Tableau 89 : Simulation de montée en puissance du programme : au global	118
Tableau 90 : Simulation de montée en puissance du programme : par année	118
Tableau 91 : ROI avec marge de 9,45€	119
Tableau 92 : Tableau de pilotage	119
Tableau 93 : Durée de vie du consommateur	123
Tableau 94 : Les variables en entrée du modèle	156
Tableau 95 : Le CA discrétisé en classe et ses indicateurs de dispersion	157
Tableau 96 : Les variables finales dans le modèle	161
Tableau 97 : Les classes de clones	165
Tableau 98 : Le tableau d'importance des variables dans le réseau bayésien.....	167
Tableau 99 : Le tableau d'importance des variables dans le réseau bayésien.....	168
Tableau 100 : Le poids des variables dans la réaffectation des clones	169
Tableau 101 : Taux de réaffectation correcte par les réseaux Bayésiens	171
Tableau 102 : Le poids des classes de clones entre l'ensemble d'apprentissage et l'ensemble de test	173
Tableau 103 : Taux d'évolution de CA par classe de clone	176
Tableau 104 : Tableaux de correction des taux d'évolution de CA aberrant par classe de clone ...	176
Tableau 105 : CA total, CA potentiel et taux d'augmentation	177
Tableau 106 : Croisement des segments Bronze, Argent, Or par PMG en CA en moyen.....	178
Tableau 107 : Croisement des segments Bronze, Argent, Or par PMG en CA potentiel moyen ...	179
Tableau 108 : Croisement des segments Bronze, Argent, Or par RFM en CA moyen.....	180
Tableau 109 : PMG en montant moyen de CA par classes de potentiel regroupées.....	182
Tableau 110 : PMG en CA potentiel moyen par classes de potentiel regroupées	183

Liste des figures :

Figure 1 : La détermination du taux de captation (en trois graphiques).....	26
Figure 2 : Citation du concept de taux de nourriture sur le web par année.....	27
Figure 3 : Principe de fonctionnement d'un réseau neurone de type Perceptron multicouches	36
Figure 4 : Liste des variables en entrée du modèle ARIMA	44
Figure 5 : Liste des variables en entrée du perceptron	54
Figure 6 : Liste des variables en entrée du modèle désagréé	61
Figure 7 : Les variables en entrée de la typologie.....	73
Figure 8 : La part du hors médias, l'enjeu.....	85
Figure 9 : La pyramide de la valeur client	89
Figure 10 : Durée de vie du client et évolution de sa valeur.....	89
Figure 11 : Modèle de détermination de la valeur client.....	90
Figure 12 : Modèle de détermination de la valeur client de Wayland et Cole d'après Venturi et Lefébure	91
Figure 13 : Modèle de détermination de la valeur client de Venturi et Lefébure	92
Figure 14 : Modèle de valeur client et probabilité d'achat	92
Figure 15 : Détermination des segments clients les plus rentables	94
Figure 16 : Simulation de couverture de cible	97
Figure 17 : Durée de vie du client et évolution de sa valeur	97
Figure 18 : Modèle de détermination d'accroissement de la valeur du portefeuille client.....	102
Figure 19 : Les effets de la fidélisation	102
Figure 20 : Les utilisations de Home Scan.....	104
Figure 21 : Taille de la cible.....	110
Figure 22 : Structure de la base de données	111
Figure 23 : L'étude SIMM.....	121
Figure 24 : Une représentation de la valeur client	125
Figure 25 : LTV, delta marché et captation.....	126
Figure 26 : Détermination des potentiels	127
Figure 27 : Pression concurrentielle pour Banque Populaire et Barclay's.....	127
Figure 28 : Méthodologie générale	128
Figure 29 : Calcul de la VTC	131
Figure 30 : Zoom sur les taux de fidélité de l'Etude Y	132
Figure 31 : Calcul de la VTC (cas particulier des pays avec des pneus d'hiver).....	133
Figure 32 : Durée de vie du client et évolution de sa valeur dans l'automobile	133
Figure 33 : Valeur Totale Client Renouvellement	134
Figure 34 : Segmentation PMG.....	135
Figure 35 : Zoom sur les bornes PMG (France et Allemagne)	135
Figure 36 : Zoom sur les bornes PMG (RU, Italie et Espagne)	136
Figure 37 : Bornes PMG VTC XXX (France et Allemagne).....	137
Figure 38 : Bornes PMG VTC XXX (RU et Italie)	137
Figure 39 : Bornes PMG VTC XXX (Espagne)	138
Figure 40 : Bornes PMG VTC _{TCRE} XXX (France et Allemagne).....	138
Figure 41 : Bornes PMG VTC _{TCRE} (RU et Italie).....	139
Figure 42 : Bornes PMG VTC _{TCRE} (Espagne)	139
Figure 43 : Taille de la cible (France et Allemagne).....	140
Figure 44 : Taille de la cible (RU, Italie et Espagne).....	140
Figure 45 : Captation de la cible sur un programme sortant (France et Allemagne)	141
Figure 46 : Captation de la cible sur un programme sortant (UK, Italie et Espagne)	141
Figure 47 : Marge captable et TN (France et Allemagne)	142

Figure 48 : Marge captable et TN (RU et Italie)	143
Figure 49 : Marge captable et TN (Espagne)	143
Figure 50 : Investissement CRM (France et Allemagne).....	144
Figure 51 : Investissement CRM (RU et Italie)	144
Figure 52 : Investissement CRM (Espagne)	145
Figure 53 : Convergence des modèles.....	146
Figure 54 : CRISP DM.....	154
Figure 55 : Le fonctionnement d'un réseau de Kohonen	159
Figure 56 : Le césurage du CA par groupe de clones	160
Figure 57 : le modèle sous Modeler	162
Figure 58 : Affichage de la PMG pas classes de clones.....	163
Figure 59 : Affichage de la RFM par classe de clones.....	163
Figure 60 : Le réseau bayésien de réaffectation	166
Figure 61 : Le dendrogramme des supers classes	173
Figure 62 : Césurage détaillé des classes de clones	175
Figure 63 : Effectif, CA et CA potentiel par classe de clones regroupés en Bronze, Argent, Or	177
Figure 64 : Croisement des segments Bronze, Argent, Or par PMG en effectif.....	178
Figure 65 : Croisement des segments Bronze, Argent, Or par PMG en CA.....	178
Figure 66 : Croisement des segments Bronze, Argent, Or par PMG en CA potentiel.....	179
Figure 67 : Croisement des segments Bronze, Argent, Or par RFM en effectif.....	179
Figure 68 : Croisement des segments Bronze, Argent, Or par RFM en CA	180
Figure 69 : Effectifs et CA par classes de potentiel regroupées.....	181
Figure 70 : PMG en effectif par classes de potentiel regroupées.....	181
Figure 71 : PMG en CA annuel par classes de potentiel regroupées	181
Figure 72 : PMG en CA potentiel par classes de potentiel regroupées.....	182

Nota :

- Objectif Fidélisation[®] est aujourd'hui une marque d'Arvato D Services France
- Home*Scan est une marque de Nielsen
- SIMM est une marque de Kantar Media

Liste des graphiques :

Graphique 1 : La série brute, le TN de la marque 1, la variable à prédire, donnée hebdomadaire	44
Graphique 2 : La série différenciée, le TN de la marque 1, la variable à prédire	45
Graphique 3 : La fonction d'autocorrélation de la série brute du modèle ARIMA, marque 1, données hebdomadaires.....	46
Graphique 4 : La fonction d'autocorrélation partielle sur la série différenciée du modèle ARIMA, marque 1, données hebdomadaires.....	47
Graphique 5 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision du TN, marque 1, données hebdomadaires.....	49
Graphique 6 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision, marque 1 sur 4 semaines (mois).....	50
Graphique 7 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision du TN, marque 14, données hebdomadaires.....	52
Graphique 8 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision du TN, marque 14 sur 4 semaines (mois).....	53
Graphique 9 : Qualité de l'ajustement du Perceptron entre le réel et la prévision du TN, marque 1, données hebdomadaires.....	55
Graphique 10 : Qualité de l'ajustement du Perceptron entre le réel et la prévision du TN de la marque 14 sur 4 semaines (mois).....	56
Graphique 11 : Qualité de l'ajustement de la marque 1 à la semaine.....	65
Graphique 12 : Qualité de l'ajustement du modèle désagrégé pour la marque 1 sur 4 semaines (mois)	67
Graphique 13 : Qualité de l'ajustement du modèle désagrégé de la marque 14 à la semaine	69
Graphique 14 : Qualité de l'ajustement du modèle désagrégé de la marque 14 sur 4 semaines (mois)	71
Graphique 15 : Répartition des types	76
Graphique 16 : Structure PMG sur un exemple du secteur cosmétique vendu en GMS	93
Graphique 17 : La pénétration de la Marque X.....	105
Graphique 18 : Parts de marché	105
Graphique 19 : Répartition des taux d'acheteurs, d'actes et de CA par segment PMG.....	108
Graphique 20 : Gros consommateurs : nombre d'individus en fonction du coût unitaire	114
Graphique 21 : Gros consommateurs : profit unitaire en fonction du nombre d'individus	115
Graphique 22 : Gros consommateurs : coût unitaire et profit unitaire en fonction du nombre d'individus.....	115
Graphique 23 : Moyens consommateurs : profit unitaire en fonction du nombre d'individus	116
Graphique 24 : Moyens consommateurs : profit unitaire en fonction du nombre d'individus	116
Graphique 25 : Moyens consommateurs : coût unitaire et profit unitaire en fonction du nombre d'individus.....	116
Graphique 26 : Petits consommateurs : profit unitaire en fonction du nombre d'individus	117
Graphique 27 : Petits consommateurs : profit unitaire en fonction du nombre d'individus	117
Graphique 28 : Petits consommateurs : coût unitaire et profit unitaire en fonction du nombre d'individus.....	118
Graphique 29 : "Valeur" de clients	124
Graphique 30 : Le cycle de vie d'un client bancaire.....	125
Graphique 31 : Le taux de captation	152
Graphique 32 : Les outliers	157

Liste des annexes :

ANNEXE 1 : QUESTION SUR LA MESURE DE LA VALEUR CLIENT PAR LES ENTREPRISES FRANÇAISES	191
Annexe 2 : Glossaire CRM.	194
Annexe 3 : Tableaux, graphiques et figures.	205

Introduction

Introduction générale

Notre recherche est une thèse sur travaux, dont les premiers ont commencé en 2000 lorsque nous avons mis au point pour Groupe D, une agence de marketing opérationnel, un modèle de rentabilité de la fidélisation inspiré de travaux de Robert Blattberg dans les années 1980/90 (Blattberg, 1986, 2001).

Notre thèse est basée sur trois travaux complémentaires qui portent sur la rentabilité de la fidélisation clients :

- 1° La valeur du client à un instant T vs sa valeur à vie (Vallaud, 2002, 2012)
- 2° La détermination d'un potentiel client (Vallaud, 2012)
- 3° L'élasticité du taux de nourriture du client, ce dernier travail est notre dernière recherche

Étalés sur 20 ans ces trois travaux sont complémentaires en ce sens que chacun apporte une information complémentaire au précédent en précisant à chaque fois ce que la marque peut gagner par la fidélisation dans le temps : la notion de taux de captation qui est la part qu'une marque peut prendre à la concurrence par une stratégie de fidélisation sur le taux de nourriture actuel du consommateur.

Nos travaux suivent une suite logique de modèles macros pour tendre vers un sujet « micro » : l'écart type du taux de nourriture du consommateur sur différentes durées de consommation qui serait la limite maximale du taux de captation.

Depuis notre premier travail nous évoquions le taux de captation comme contraint sans le démontrer. Dans notre dernier travail, nous démontrons que l'augmentation de valeur qu'une marque peut gagner dans la durée de vie du client est basée sur la bonne quantification de cet écart.

Nos deux premiers travaux portent sur des exemples à la fois en B to B et en B to C, le dernier travail ne porte que sur des exemples B to C. Nous pensons néanmoins que le raisonnement peut s'appliquer aux deux types de marché sous réserve d'adaptation.

1. Synthèse de nos travaux :

Nous résumons dans cette introduction générale une synthèse de nos différents travaux. Les concepts théoriques avec leurs références bibliographiques sont détaillés dans les travaux eux-mêmes.

1.1 LA FIDELISATION RENTABLE : PROPOSITION D'UN MODELE COMPOSITE ENTRE LA VALEUR D'UN CLIENT A L'INSTANT T ET SA VALEUR A VIE

Recruter ou fidéliser ? Privilégier les médias de masse ou le hors média ? Ces dernières années, la part des investissements dans le hors média n'a cessé d'augmenter, jusqu'à atteindre 63% des dépenses en 2009.

Pour justifier cette augmentation, les entreprises doivent être en mesure d'évaluer leurs rentabilités dans le mix marketing et cela passe par l'estimation de la valeur client.

Notre première étude portait sur l'évaluation de la rentabilité de programmes de fidélisation et de CRM. Le but étant de définir une nouvelle modélisation de cette rentabilité, basée sur la complémentarité de deux approches : un modèle qui estime la valeur d'un client à un instant t. C'est ce modèle que nous avons appelé Objectif Fidélisation^{®1} et le calcul de la LTV : Life Time Value (Dwyer et al 1989, Jackson, 1989, Jambu, 1998, Magson, 1998, Kim et al, 1999) , qui estime la valeur client sur un continuum temporel : une durée de vie.

Après avoir défini les concepts théoriques sur lesquels se basent les approches que nous venons de citer, nous avons appliqué empiriquement ces deux modèles, en utilisant des données portant sur de « vrais » produits.

¹ Objectif Fidélisation est une marque déposée qui appartient à Arvato

L'étape ultime a consisté à créer un modèle composite, également testé sur un cas concret avec des estimations de la valeur client à court terme sur plusieurs périodes et un modèle global de LTV.

Le concept de « *Life Time Value (LTV)* », qui se traduit par « la valeur du client à vie » et qui correspond à la somme des profits actualisés attendus sur la durée de vie du client, avait déjà été largement publié sous la formes de modèles plus ou moins proches pour déterminer la valeur client (Wayland et Cole 1997, Venturi et Lefebure 2000). Ces modèles utilisent les paramètres suivants : les quantités achetées sur la période, la marge par unité d'achat, la durée de la relation, les coûts d'acquisition et de rétention du client et la valeur capitalistique de l'investissement.

En fait, pour être rentable, la fidélisation doit valider différentes contraintes, qui seront les grandeurs clés des modèles décrits dans cette étude (Hallberg, 1995) :

- **Atteindre la couverture d'un cœur de cible déterminé** : le taux de couverture de cible est la part de la cible que l'on peut toucher via un programme de fidélisation. Par adhésion volontaire, on peut capter 30 % de la cible des clients d'une marque dans un programme de fidélisation classique, 50% pour la fidélisation d'une enseigne.

- **Recruter au programme les clients les plus contributifs** : le « Gros client » en CA est en général le plus contributif, c'est un gros consommateur de la marque (il dépense beaucoup en achats à la marque) mais son taux de nourriture est souvent faible. Il dépense aussi beaucoup ailleurs : 80% des achats annuels des Gros clients d'une marque sont faits à la concurrence (Hallberg, 1995). C'est le client qu'il ne faut pas perdre et qui a aussi le plus chance de passer à l'acte à la suite à une action de fidélisation.

- **Atteindre un taux d'activation** : cela correspond au taux de clients ayant réagi une seconde fois à une stimulation du programme de fidélisation. L'atteinte du taux d'activation souhaité s'obtient en faisant au consommateur des offres transactionnelles et des offres relationnelles, une fois son adhésion volontaire obtenue. Il faut que 50% des clients recrutés réagissent au moins une seconde fois au programme pour avoir une activité significative en termes d'achats.

- **Accroître le taux de captation** : Accroissement du taux de nourriture par une prise de part des achats faits à la concurrence. En pratique, on ne peut capter par la fidélisation, sur un grand nombre de clients, « que » 20 % en moyenne des quantités faites à la concurrence. Ce taux de « captation » de 20% n'est possible que si le taux de nourriture est inférieur ou égal à 50%. Au-delà de 20 % de captation, l'acheteur commencerait à percevoir un effort (une modification de ses habitudes de consommation). Cette affirmation partait à l'époque sur l'hypothèse non démontrée que le taux de nourriture d'un consommateur est contraint.

- **Obtenir une baisse puis un maintien des coûts unitaires de fidélisation** : le coût unitaire de fidélisation est déterminé une fois l'objectif de couverture de cible et la taille de base atteints. Ce coût doit s'amortir sur la marge des achats supplémentaires atteints par la captation sur le taux de nourriture. Il est donc stratégique de déterminer ce taux de part captable à priori : le taux de captation.

- L'atteinte aux 5 objectifs ci-dessus, doit se faire dans une dimension temporelle également déterminée à priori.

L'étape suivante de notre recherche avait consisté à appliquer à un cas de figure réel une approche empirique de la modélisation du ROI (Return On Investment/Retour Sur Investissement) de la fidélisation à court terme : application du modèle Objectif Fidélisation[®] pour utiliser et vérifier les concepts ci-dessus.

Par l'utilisation de ce modèle, on est capable de calculer les ROI du programme de fidélisation compte tenu des coûts et des gains et selon les hypothèses retenues par année : on identifie donc en quelle année le ROI est atteint.

Cette modélisation permet également de suivre les 5 grands objectifs de pilotage du programme par année en fonction des objectifs initiaux définis par le modèle.

La difficulté « marketing » est d'atteindre ces objectifs de façon simultanée afin d'obtenir la rentabilité du programme.

Si le modèle montre que la marque ne peut pas les atteindre, il n'est pas recommandé de se lancer dans un programme de fidélisation/CRM qui passerait par du marketing direct sortant. Il est donc primordial de bien modéliser ces objectifs en amont et d'en mesurer leurs réalisations sur le long terme.

Le second modèle étudié est le modèle de LTV. Il utilise, dans ses paramètres, les mêmes indicateurs qu'Objectif Fidélisation[®] et ces derniers s'appliquent de la même manière, mais sur des périodes beaucoup plus longues.

Le concept de VTC (Valeur Totale Client) est défini : il correspond à la marge totale dégagée sur toute la vie d'un client. On le calcule de la manière suivante : $VTC = VP \text{ (Valeur du produit)} \times TR \text{ (taux de renouvellement)} \times PDM_{\text{valeur}}$. La PDM_{valeur} (Part De Marché valeur) est la somme de la marge réalisée par la marque étudiée vs la somme de la marge totale de tous les individus. Sur le delta fait à la concurrence on applique alors un taux de captation. La somme de la valeur actuelle plus celle de la valeur captée sur la durée de vie, donne alors une VTC précise.

Ce modèle permet de créer un tableau de synthèse (comme c'était le cas pour Objectif Fidélisation[®]) qui croise les coûts inhérents au programme avec les gains espérés sur la durée de vie des clients par segment PMG (Division des clients en Petits, Moyen et Gros en fonction de leur chiffre d'affaires).

Le taux de couverture de cible et la réactivité ne sont pas présents de manière explicite mais ces deux notions sont implicitement incluses dans la LTV par la durée de vie des clients, l'attrition et la réaction aux actions marketing.

En créant en dernier lieu un modèle composite des deux approches, la recherche additionne une somme de modèles courts-termes montrant la cohérence avec l'approche par le LTV (Blattberg 1986, 1998, Carguillo, 1999, Bolton et al, 2000). On perçoit donc bien la complémentarité méthodologique des deux démarches et l'intérêt managérial :

- une prévision de la rentabilité de la fidélisation sur une période assez courte va être déterminante de la LTV
- la LTV est la mesure ultime des actions de fidélisation.

En pratique, ce qui a été fait dans cette étude s'applique à des cas concrets : en entreprise, avant de se lancer dans un programme de fidélisation, il est primordial d'en estimer la rentabilité. Les trois études de cas développées dans notre recherche démontrent que c'est applicable pour différents secteurs et que les données pour le faire sont disponibles.

En fonction du cas de figure (du produit, du marché, etc.), le modèle à appliquer pour évaluer la rentabilité du programme pourra être Objectif Fidélisation[®] pour un calcul à un instant t (bien adapté pour tous les produits de grande consommation) ou le modèle de la LTV pour un calcul sur toute la durée de vie du client.

Finalement, on retiendra de cette recherche les points suivants :

1. La rentabilité d'un programme de fidélisation doit être toujours évaluée avant d'être lancé,
2. Il y a plusieurs modèles utilisables pour l'évaluer et on privilégiera le modèle sur un instant t car il est prédictif du modèle sur la durée de vie. Le modèle court terme est plus facile à mettre en place.
3. Ces modèles se basent sur des indicateurs communs qui serviront à piloter le programme (couverture de cibles, structure PMG (Petit, Moyen, Gros clients en CA), réactivité, captation, marge allouable, coût unitaire).

Depuis la publication de la recherche, le modèle a été utilisé par un grand nombre de sociétés et on retrouve plus ou moins l'approche composite dans certains travaux (Blattberg, 1986, 1998, Allard, 2002).

1.2 ESTIMATION DU POTENTIEL INDIVIDUEL DE CHIFFRE D'AFFAIRES EN UTILISANT LES DONNEES ISSUES D'UNE BASE DE DONNEES CLIENTS :

Dans cette seconde recherche nous tendions à déterminer « un potentiel client ». Le potentiel client dans notre acception correspond à évaluer le chiffre d'affaires incrémental qu'une entreprise donnée pourrait atteindre avec ses clients actuels. C'est un indicateur stratégique pour toute entreprise, quel que soit son secteur d'activités.

Notre recherche concevait et mettait en application une méthode de détermination de ce potentiel client, en se basant uniquement sur des données classiquement présentes dans des bases de données de quasi-totalité des entreprises (variables sociodémographiques ou descriptives d'entreprises et enregistrements de leurs transactions). Ce point est très important pour la généralisation de la méthode.

Afin de calculer ce potentiel, les clients « qui se ressemblent » (des « clones ») vont être regroupés par utilisation d'une technique de classification automatique : les réseaux de Kohonen (Kohonen, 1998). L'hypothèse étant que les individus considérés comme des clones sont supposés avoir les mêmes caractéristiques et donc le même potentiel.

Les SOM de Kohonen sont des techniques de classification très puissantes qui permettent d'avoir un regroupement sur de grandes populations en minimisant le plus les différences.

Dans notre recherche, nous avons validé la classification de Kohonen par plusieurs techniques de classification confirmatoire et nous avons césuré le chiffre d'affaires de chaque groupe de clones pour nous permettre d'obtenir un potentiel « atteignable » pour chaque client.

Les données utilisées dans le test empirique décrit dans ce document, correspondent à la base de données de 5 millions de clients d'un des principaux distributeurs alimentaire français.

Les étapes de la méthodologie sont les suivantes : application de la méthode de classification par les réseaux de Kohonen et validation de la technique puis calcul des niveaux d'évolution dans les segments et enfin calcul du potentiel individuel. Nous allons à présent décrire succinctement ces trois étapes, en commençant par la classification par les réseaux de Kohonen.

Ce type de réseaux est le plus utilisé parmi les réseaux à apprentissage non supervisé. Il permet d'obtenir des groupes très homogènes et très stables avec de multiples individus et variables, avec la faculté de bien prendre en compte les valeurs manquantes, et autorise des relations non linéaires complexes entre beaucoup de variables et beaucoup d'individus. En tant que réseau de neurones, il se décompose en couches de nœuds connectées entre elles. Il convertit un signal d'entrée avec de nombreuses dimensions en une simple carte discrète, avec peu de dimensions : cela en fait un outil bien adapté à notre problématique.

En pratique, on obtient une cartographie qui permet d'identifier la similarité entre les groupes (deux groupes proches sur le graphique ont des caractéristiques proches).

Après avoir défini des classes par application de la technique décrite, plusieurs méthodes ont été testées afin de valider la précision et la robustesse de l'approche. Dans un premier temps et afin de déterminer les convergences entre les classes de Kohonen, la population a été divisée en sous populations pour vérifier la cohérence d'allocation des classes de clones : les classes de Kohonen se retrouvent bien aux mêmes poids entre deux échantillons.

Les classes de clones ont ensuite été croisées avec la PMG et la RFM (classement des clients selon la Récence du dernier achat, la Fréquence d'achat et le Montant d'achat) et globalement, les clients appartenant

aux mêmes segments PMG/RFM se retrouvent majoritairement dans les mêmes classes de Kohonen (c'est très clair visuellement sur la PMG).

La convergence de la méthode avec d'autres a également été vérifiée : deux types de modèles supervisés ont été testés (C5.0, Réseaux Bayésiens (Larose et Vallaud 2005, 2012)) et la réaffectation des groupes par ces modèles montre que les classes de Kohonen peuvent être reproduites : cela valide la robustesse de notre classification.

Enfin, la connectivité des super classes a été vérifiée (c'est un critère de qualité des cartes de Kohonen) et la forme de la classification obtenue satisfait aux critères de stabilité et de reproductibilité attendus.

L'étape suivante correspond au calcul des niveaux « d'évolution ». On définit l'évolution de la manière suivante : c'est le « petit saut » de chiffre d'affaires qu'un consommateur doit atteindre pour être regroupé avec les clients qui lui ressemblent le plus, sur toutes les variables sélectionnées par le modèle, mais qui ont un chiffre d'affaires plus élevé que lui. Le « saut » à franchir ne doit pas être trop important afin que l'estimation du potentiel reste la plus réaliste possible. En pratique, cela consiste à diviser le chiffre d'affaires annuel en déciles et à calculer la médiane de chaque décile.

Une valeur du CA potentiel (déduite du calcul des taux de potentiel d'accroissement entre les médianes et les déciles) est allouée à chacun des groupes de la classification de Kohonen. A chaque client est donc attribué un taux d'évolution, qui multiplié par le chiffre d'affaires du client, nous permet d'estimer son chiffre d'affaires potentiel. Ce césurage des potentiels par groupe est une des originalités de la méthode, le delta (le saut) correspondant à l'élasticité du taux de nourriture, la partie de potentiel captable.

Finalement, l'application de la méthode précédemment décrite sur les données étudiées, a permis de dégager les observations suivantes : les 5 373 026 clients génèrent en pratique un chiffre d'affaires annuel de 7.46 milliards d'euros mais représentent en fait un chiffre d'affaires potentiel de 953.2 millions d'euros. Le distributeur peut donc gagner presque 12.77% de chiffre d'affaires sur les clients déjà présents dans sa base de données.

On regroupe les clients en classes de potentiel. La base de ce distributeur se compose de 41.9% de clients « Bronze », de 14.1% clients « Argent » et de 44% de clients « Or ». Les meilleurs clients sont ceux avec les plus fortes valeurs absolues de potentiel.

76.6% sont des consommateurs « Or » et P et ils génèrent 29.9% du chiffre d'affaires annuel et 38,4% du chiffre d'affaires potentiel. Les clients qui ont le plus gros potentiel d'évolution en termes de chiffre d'affaires sont ceux qui actuellement dépensent le moins.

Finalement, l'association des deux approches que nous venons de résumer, à savoir l'utilisation de la classification de Kohonen pour obtenir des groupes de clones, puis le césurage en déciles du chiffre d'affaires des groupes de clones lui confère une grande originalité et une grande précision.

Le fait qu'elle ait été empiriquement testée sur de « vraies » données de distributeurs, confirme son applicabilité et en fait un outil directement utilisable par les entreprises désireuses de déterminer le potentiel de leurs clients.

A l'époque de notre recherche il y avait peu de travaux directement applicables à des cas réels sur la détermination de potentiel client. Nous avons appliqué la méthode sur de nombreux clients BtoB et BtoC et elle fonctionne très bien empiriquement.

Ces deux travaux ont en commun le postulat qu'il y a un potentiel de captation possible pour accroître le CA des clients étudiés que la marque peut prendre à ses concurrents. La détermination de ce potentiel, basé sur la bonne compréhension de l'élasticité du taux de nourriture, a été la motivation de cette nouvelle recherche

1.3 L'ELASTICITE DU TAUX DE NOURRITURE EST-ELLE CONTRAINTE ?

Dans notre dernière recherche, nous étudions l'élasticité du taux de nourriture.

Le taux de nourriture correspond à la part de marché d'une marque chez les consommateurs de cette marque (Sharp et Sharp, 1992) : en cela, il représente donc un des principaux indicateurs de la fidélité à la marque. L'objet de cette nouvelle étude consiste à montrer que l'élasticité de cet indicateur est une donnée « bornée » indépendamment de la catégorie de produits et que le taux de nourriture, en lui-même, est un indicateur prévisible sur la durée de consommation d'un individu.

Pour cela, nous avons considéré l'élasticité du taux de nourriture par l'écart type comme étant un indicateur capable de revenir à son état initial, ce qui revient à étudier déjà l'écart du taux de nourriture sur lui-même. Cette amplitude a été modélisée au niveau agrégé (marque, semaine, mois, année) et désagrégé (individu, marque, semaine, mois).

Les données utilisées en entrée sont des données de panel de home scanning de Nielsen sur un historique long (24 mois) et sur 6 marchés à cycles d'achats/rachats courts à longs (eau plate, café, plats surgelés, dentifrice, déodorisant et brosse à dent). Cet ensemble représente un très gros volume de données.

Pour chaque marché, ont été calculées les élasticités du taux de nourriture sur des dimensions individuelles et temporelles différentes. Les variables sociodémographiques, de segmentations clients (PMG, typologie) ainsi que quelques variables relatives au produit (prix, place sur son marché, etc) ont été utilisées en tant que variables explicatives.

Différentes techniques statistiques ont été mises en œuvre afin de modéliser l'élasticité du taux de nourriture. Dans un premier temps, une simple analyse descriptive des données a été menée en débutant par le calcul des indicateurs de base permettant de suivre les différents marchés étudiés (le nombre d'acheteurs, le nombre d'achats, la quantité vendue, le chiffre d'affaires généré, le taux de nourriture et les parts de marché sur plusieurs périodes). Cette analyse descriptive montre que le nombre de marques achetées et l'élasticité du taux de nourriture sont contraints dans des amplitudes « faibles » (sur 12 mois entre 24,4% et 29,2%): dans l'ensemble des cas étudiés, l'écart type du taux de nourriture est donc réellement « borné ».

Ensuite des méthodes agrégées ont été développées afin de tenter de prédire le taux de nourriture : un modèle ARIMA et un perceptron multicouche. Le modèle ARIMA (Auto Regressive Integrated Moving Average) est un modèle de prévision supervisé permettant de traiter des séries non stationnaires. Nous l'avons utilisé sur un historique de 104 semaines (90 d'apprentissage et 14 de test) afin de prédire le taux de nourriture de deux marques étudiées à la semaine. Nous avons obtenu une bonne prédiction du taux de nourriture pour ces deux marques (peu de différences entre le réel et le prédit par ARIMA). Le perceptron multicouches a été testé dans les mêmes conditions que l'ARIMA (mêmes marques, même historique) et a fourni des résultats équivalents en terme de prévision : le modèle est robuste et stable et prédit bien le taux de nourriture et ses écarts.

Des méthodes désagrégées d'analyse du panier de la ménagère ont ensuite été mises en œuvre. Il s'agit des techniques « Apriori » et « Sequence », qui utilisent des règles d'association dans le but d'identifier les cooccurrences d'achats. L'algorithme « Apriori » sert à reconnaître des propriétés qui reviennent fréquemment dans un ensemble de données et sait en déduire une hiérarchie d'occurrence. La technique « Séquence » est similaire mais prend en compte la notion d'ordre dans le temps. Cette dernière a finalement fourni beaucoup plus de règles significatives que l'algorithme « Apriori ». Il est donc possible de prédire par des modèles de cooccurrence d'achats les enième achats d'une marque, et ce de manière très fiable pour les règles les plus concurrentes. Ces algorithmes ont permis de démontrer que les consommateurs rachètent principalement la même marque, validant l'idée d'entropie du rachat et donc de stabilité du taux de nourriture.

La dernière méthode désagrégée utilisée est l'analyse de données de panel (Casin, 1999), elle permet de tenir compte à la fois de la variance des indicateurs dans le temps (dimension temporelle) et des panélistes entre

eux (dimension individuelle). Les résultats obtenus par l'application de cette technique ont été satisfaisant : bonne cohérence entre le réel et le prédit. Le fait que les variables socio démographiques jouent un rôle négligeable dans la prévision a renforcé l'idée que l'élasticité devait être prédite sur les variables endogènes au produit. Là aussi ce bon niveau de prévision même à un niveau individuel valide l'écart type « contraint » du taux de nourriture.

Finalement, les modèles désagrégés prédisent très bien le taux de nourriture : la prévision est fiable et le taux de nourriture est prévisible à un niveau individuel,

Enfin, et afin de répondre à notre question de recherche concernant l'aspect socialement explicable de l'élasticité du taux de nourriture, nous avons observé l'évolution de cet indicateur en fonction du niveau social des foyers (après création d'une typologie par la méthode de Kmeans). Les consommateurs qui achètent peu de fois la même marque ont un faible écart type du taux de nourriture. Les consommateurs les plus modestes changent constamment de marque afin de rechercher la marque la moins chère à l'instant t. L'élasticité du taux de nourriture permet aussi de faire ressortir ce type de phénomène.

En définitive, notre analyse nous a permis de répondre aux questions de recherche suivantes :

- l'élasticité du taux de nourriture est limitée quelle que soit la période temporelle utilisée : cela est dû principalement à l'inertie du consommateur qui répète ses achats aux marques qu'il consomme habituellement, dans les mêmes quantités,
- il est possible de prédire correctement le taux de nourriture sur des données agrégées sur des périodes différentes et ce via l'utilisation de différentes techniques : le taux de nourriture est donc prévisible,
- l'élasticité du taux de nourriture varie peu d'une catégorie de produit à l'autre : l'amplitude de variation reste en effet assez limitée,
- l'élasticité est marquée socialement : elle se réduit, en effet, chez les catégories sociales les plus défavorisées. L'évolution de l'élasticité du taux de nourriture est donc révélateur des difficultés financières des ménages ce qui en fait un outil exploitable dans l'aide à la détermination des prix.

Ces conclusions ont de nombreuses applications pratiques, en particulier au niveau managérial : la détermination des variations intra et inter-individuelles du taux de nourriture pourrait permettre d'estimer la part de marge « captable » d'un individu : le taux de captation et ainsi prévoir la rentabilité d'actions marketing. L'identification des clients capables d'augmenter leur consommation dans les limites de l'écart type de leur taux de nourriture permet de calculer le taux de captation (part que la marque peut gagner grâce à la fidélisation).

Fidéliser ses clients permet, en effet, non seulement à la marque de capter une part des achats de la concurrence mais également de maintenir les quantités achetées sur ses bons clients afin qu'ils ne perdent pas de leur valeur. Cette étude a donc permis de réaffirmer l'importance de la fidélisation dans la démarche marketing, dans un climat où les efforts ont plus tendance à se tourner vers des alternatives plus court-termistes : politique de prix bas et promotion.

Dans toutes nos publications précédentes, nous avons toujours à la fois essayé d'avoir un enjeu managérial mais aussi une perspective académique et scientifique. Nos publications précédentes sont des recherches « appliquées » car pour avoir « un temps nécessaire à la recherche » il faut justifier en entreprise d'une « certaine rentabilité ».

C'est ce grand écart qui est difficile à gérer cas sans application pratique, il est difficile en travaillant de faire de la recherche et à l'inverse toute recherche appliquée manque parfois d'un certain académisme nécessaire à une thèse. Nous en sommes conscients.

Nos travaux sont présentés dans ce document du plus récent aux plus anciens. Les travaux les plus anciens ont été entièrement remis à jour.

2. L'analyse de la littérature :

Comme notre recherche est basée sur nos travaux précédents nous allons rédiger un contexte de recherche lié à une analyse de la littérature en se limitant le plus possible aux concepts clefs de notre dernier sujet.

2.1 Définitions de quelques concepts :

Dans cette première partie nous définissons quelques concepts précédemment utilisés dans les travaux précédents.

2.1.1 La fidélisation :

La fidélisation (Cunningham, 1956, Meyen-Waarden, 2004) est la démarche marketing qui tente à favoriser la fidélité des clients. La fidélisation consiste à rechercher un état de fidélité voulu par les clients. Dans la fidélisation les clients sont « acteurs ».

La fidélisation est une démarche globale qui peut recourir à de nombreuses actions et à la mise en place de multiples dispositifs (Crié 1996, Trinquecost, 1996, Shepherdson 2000). Elle ne se résume pas à la mise en place d'un programme de fidélisation entendu au sens étroit du terme (carte + avantages).

2.1.2 La fidélité

On dissocie en général deux types de fidélité : la fidélité attitudinale vs la fidélité comportementale.

Jean Marc Lehu (Encyclopédie du Marketing, 2012) définit la fidélité comme l'attachement conscient ou non du consommateur à un produit, une marque, une entreprise ou un mode de distribution. Elle peut être mesurée par le taux de rachat, dès lors qu'après avoir essayé un produit, quatre comportements s'offrent au consommateur : l'abandon, l'essai à nouveau, la fidélité ou la consommation alternée (en revenant parfois aux produits consommés auparavant). C'est la fidélité comportementale.

Elle doit également prendre de plus en plus en considération la confiance portée par le consommateur au nom de l'entreprise, au partage de ses valeurs, au cautionnement de ses choix et de ses actes. C'est la fidélité attitudinale.

Dans notre présente recherche nous nous concentrons sur la fidélité « comportementale » (Berry, 1995).

2.1.3 Le potentiel client :

Nous l'avons défini dans un travail précédent (Vallaud, 2011) comme étant le chiffre d'affaires supplémentaire qu'une société donnée peut atteindre avec ses clients actuels.

2.1.4 Le taux de captation :

Nous l'avons aussi défini dans des travaux précédents (Vallaud, 2007, 2011, 2012), comme la part de quantité achetée qu'une marque peut « capter » sur ses concurrents. Nous l'avons aussi appelé le « potentiel captable ».

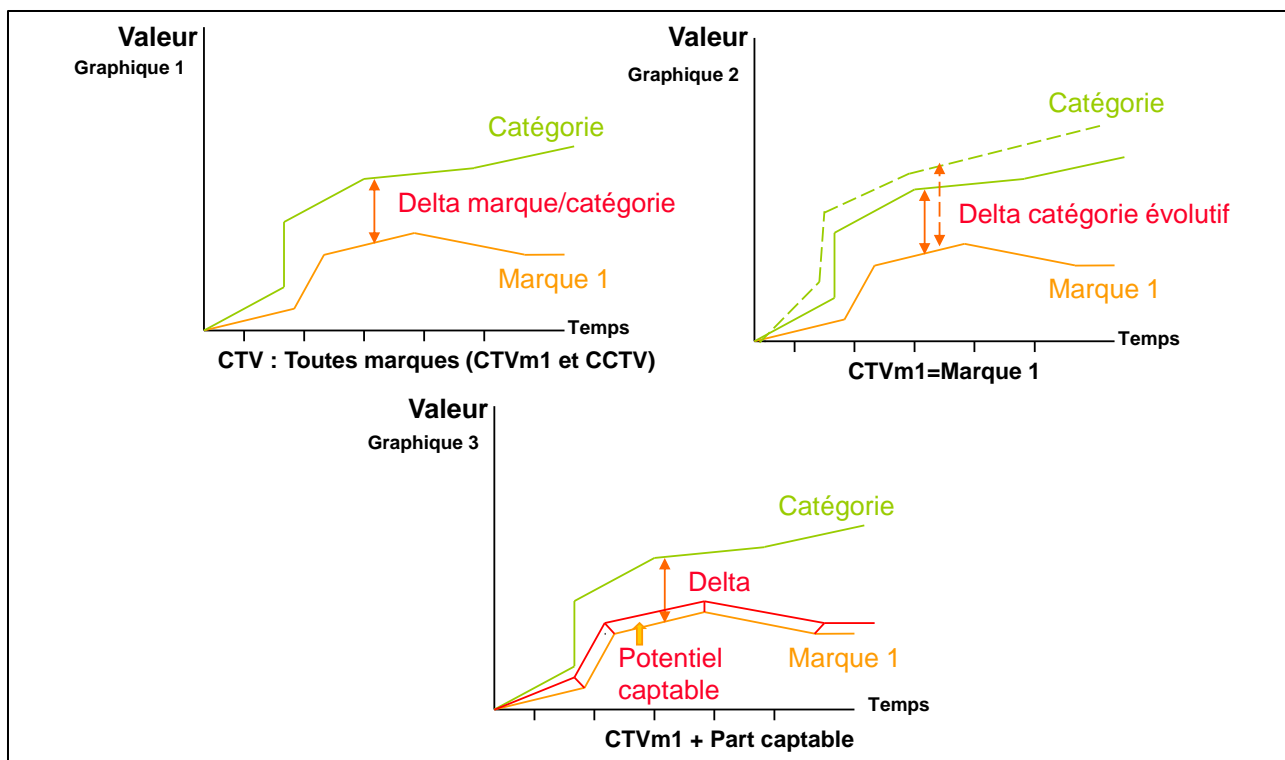


Figure 1 : La détermination du taux de captation (en trois graphiques)

La courbe orange symbolise la part de consommation en la marque étudiée en volume dans le temps par un consommateur, la courbe verte le total marché, le delta est donc ce que le consommateur consomme à la concurrence. La courbe verte en pointillés symbolise simplement que la consommation globale du consommateur sur le marché de la marque peut croître. La courbe rouge avec un léger écart par rapport à la courbe orange symbolise le delta que la marque peut prendre à la concurrence au fil du temps.

Nous retrouvons la notion de « captation »² dans la littérature « socio-économique » mais avec un sens un peu différent : capter le client dans un processus d'achat qui aurait pour but de l'enfermer dans la relation. Finalement capter une part d'esprit pour capter une part d'achat incrémentale, les concepts ne sont pas si éloignés.

2.1.5 Le taux de nourriture :

Pour définir le taux de nourriture nous reprenons la définition la plus courante, celle du portail e-marketing.fr : Taux de nourriture : Calcul en pourcentage exprimant la part de marché de la marque chez les consommateurs de la marque. Il constitue une mesure de la fidélité à la marque étudiée.

Il se définit aussi parfois par « *Share of Category Requirements (SCR)* » de la marque (Verhoef, 2003, Gupta, 2006, Sharp et Sharp, 1997, 1999, 2008, Rungie, Uncles et Laurent, 2009.).

² Pascale Trompette : Une économie de la captation, Revue Française de Sociologie, 2005

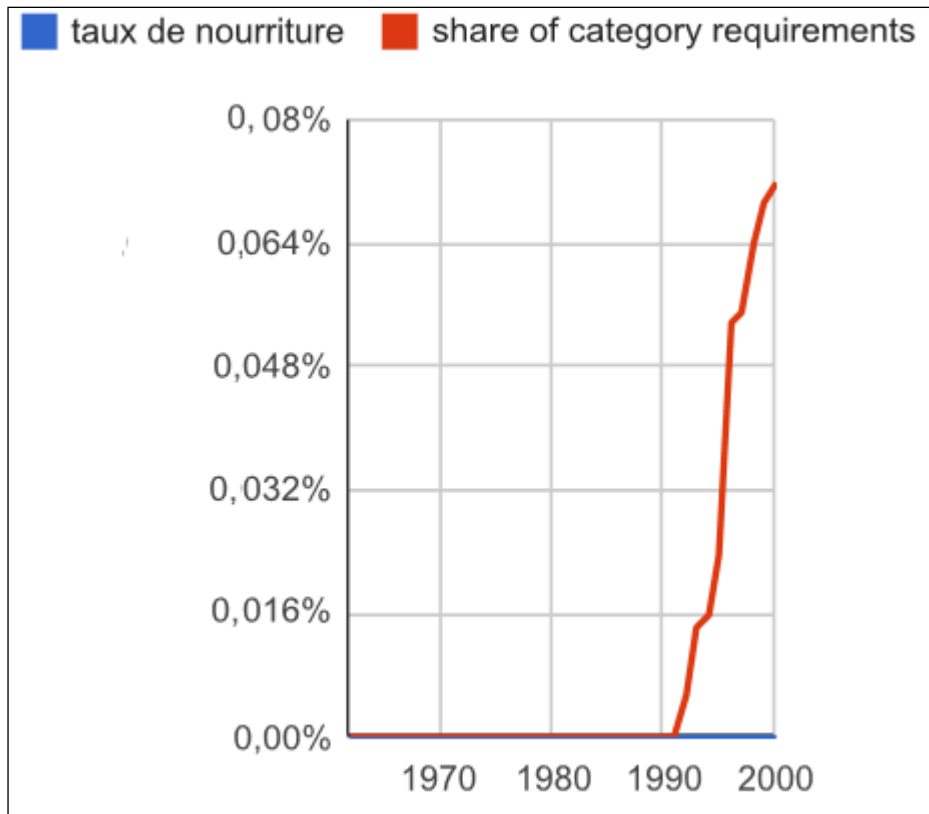


Figure 2 : Citation du concept de taux de nourriture sur le web par année

Evolution des deux concepts dans la littérature : Source Google Ngram. Notons une évolution importante et récente de « Share of Category Requirements ».

Nous reprenons la formulation mathématique suivante (Bhattacharya et al, 1996) :

$$SCR_{hi}(T) = \frac{i \in T Q_{hi}(t)}{j \in T Q_{hi}(t)}$$

$SCR_{hi}(T)$ = le taux de nourriture du foyer h pour la marque i sur la période T (qui peut être la semaine, le mois, le trimestre ou l'année).

$Q_{hi}(t)$ = la quantité de la marque i achetée par le foyer h sur l'occasion d'achat t (ou t est un indice de toutes les occasions d'achat durant la période T).

j = un indice de toutes les marques de la catégorie.

Les auteurs montrent que le SCR est le reflet de la fidélité, même si cette relation est faible, et que l'évolution du taux est de faible amplitude.

2.2 L'importance de la fidélité pour les marques :

Plusieurs ouvrages et articles ces 20 dernières années (Chestnut et Jacoby, 1978, Manuel et Xardel, 1980, Curry et Stora, 1993, Hallberg, 1995, Benavent et Crié, 1998, Volle, 1999, Lefebure et Venturi, 2000, Desmet, 2001, 2002) ont montré qu'il était moins coûteux de garder un consommateur que d'en recruter un nouveau : la fidélisation est donc apparue comme un enjeu majeur en période de concurrence exacerbée entre

les marques et de nombreux programmes de fidélisation (Wayland et Cole, 1997) qui se sont montés un peu partout dans la monde.

Nous avons remarqué que beaucoup de recherches (Bhattacharya 1997, Bhattacharya, Fader et Lodish 1996) essaient d'expliquer cette fidélité et de comprendre l'impact des variables marketing sur cette fidélité, mais sans se poser la question des limites à sa variation, au niveau individuel ou agrégé.

Les résultats de ces modèles (Dekimpe et al, 1997) semblent pourtant indiquer que de telles limites existent : ensemble de considération, ensemble déterminant, intensités des préférences, switching behaviors, recherche de variété, effets de satiété, effets promotionnels, actions marketing de toutes les marques du marché, effets de double jeopardy.

2.3 Ensemble de considération/ensemble évoqué/ensemble déterminant :

L'ensemble de considération, appelé aussi parfois « ensemble de notoriété » (Howard, 1989; Howard et Sheth, 1969, Jolivot et Chandon 2002), est l'ensemble de marques, produits ou services venant à l'esprit du consommateur lorsqu'il envisage un achat. Le consommateur effectue ensuite un traitement de ces informations ou des recherches pour effectuer son choix. Les campagnes de notoriété visent notamment une présence de la marque dans l'ensemble de considération des consommateurs.

L'ensemble évoqué (Guichard et Vanhems, 2004) est un sous ensemble de marque que le consommateur arrive à citer comme potentiellement achetable par lui-même.

L'ensemble déterminant (Hauser et Wernerfelt, 1980) est le sous ensemble des marques réellement achetées par le consommateur. Il peut avoir un ensemble de considération en général plus large que son ensemble déterminant. Les panels d'achat permettent d'isoler cet ensemble déterminant.

Ces auteurs montrent que ces différents sous-ensembles sont liés. Même si nous ne nous attachons pas à en comprendre le lien, il est intéressant d'émettre des hypothèses sur les raisons de ce lien et sur son impact avec le taux de nourriture du consommateur.

Selon cette théorie, certains auteurs admettent qu'il est difficile de modifier l'ensemble de considération d'un consommateur : celui-ci faisant de la résistance au changement. Donc au-delà d'un certain seuil de consommation d'un produit, le consommateur percevrait de manière plus ou moins consciente que son champ de choix se modifie et y résisterait. Il y aurait donc selon nous un seuil de tolérance où le consommateur accepte de consommer davantage une marque ; au-delà on modifierait de façon trop importante son champ de choix. Evidemment plus le produit serait d'un usage courant avec une fréquence d'achat élevée, plus la résistance au changement serait forte.

Dans la plupart des modèles de choix désagrégés, les auteurs incluent des critères qui agissent comme des poids d'entropie (Herniter et Howard 1964 ; Cooper et Nakanishi, 1989, Guadani et Little, 2008) : la préférence du contenant (nombre d'actes d'achat où un consommateur achète le même contenant sur le total de ses actes : un pack d'eau de 6 litres sur le total des packs d'eau achetés par exemple) ou le nombre d'achats antérieurs à la marque.

Cette entropie (Dall'Olmo et al, 1997) serait donc une des raisons qui ferait que le consommateur consomme toujours la même quantité à la marque avec un écart type limité.

2.4 L'intensité des préférences :

Le consommateur hiérarchise plus ou moins inconsciemment les raisons d'une préférence pour une marque vs une autre (Cook et Kress, 1985) avec dans un ensemble de marques limité des classements parfois

différents mais avec souvent une marque de prédilection qui représente la part la plus importante de ses choix.

2.5 Les comportements de remplacements (« *switching behaviour* ») :

Le « *brand switching* » (Neslin et al, 2003) est le phénomène par lequel les consommateurs passent d'une marque à une autre pour un même produit.

Les phénomènes de « *brand switching* » sont surtout consécutifs aux opérations de promotions de ventes mais peuvent également parfois être la conséquence de campagnes publicitaires, voire parfois d'un problème d'image affectant la marque quittée.

Le « *brand switching* » est une des composantes des ventes incrémentales liée à ces opérations.

Il est souvent ponctuel, le consommateur retournant en général à sa marque initiale.

2.6 La recherche de variété :

Le comportement de recherche de variété (Kahn et Ratner, 2005 ; Chintagunta, 1991, 1993, 1998) se définit comme « un changement de marque, sans insatisfaction à l'égard de la marque abandonnée ». Le consommateur va dans un ensemble « limité » de marques choisir tour à tour l'une de celles-ci, sans pour autant remettre en question la marque qu'il n'achètera pas, à un moment donné. Il arbitre donc ses achats entre différentes marques, pour ne pas avoir l'impression d'avoir une consommation routinière.

2.7 Les effets de satiété :

Il est démontré (Tanner et al, 2008) que pour certains produits de consommation courante le consommateur arrive vite à un niveau de satisfaction de son besoin alors que pour d'autres, les produits de luxe par exemple, ce niveau est plus long à atteindre. Selon les auteurs l'effet de satiété sous-tendrait donc une consommation par construction limitée.

2.8 Les effets promotionnels :

La promotion aurait pour effet de faire « *switcher* »³ les consommateurs d'une marque à l'autre sur le court terme, ces derniers étant attirés par la promotion (Fader 1993, Bell et al, 1999, Ailawadi et al, 2001), le consommateur revenant ensuite à son état initial.

C'est à peu près le même effet que le « *brand switching* ».

2.9 Les actions marketing de toutes les marques d'un marché :

Sur des marchés très concurrentiels à forte fréquence d'achat, les effets marketing de toutes les marques finissent par se compenser (c'est la loi d'équilibre des marchés) (Naik et Rammon ; 2003). L'information n'étant ni pure, ni parfaite parfois, certaines marques prennent l'avantage sur leurs concurrents pendant une durée parfois limitée.

Toutes ces théories plus ou moins « anthropiques » (Neslin et al, 1985, Jung Sang et al, Jain Dipak et al, 1991, Kumar et al 2001) nous amènent à justifier l'idée d'un taux de nourriture limité pour une marque dont l'élasticité serait contrainte :

³ Changer de marque pour une autre.

- Le consommateur n'est pas par essence limité dans sa consommation de produits
- Le consommateur a par nature, besoin de choix
- Mais ce choix est limité à quelques marques
- Le consommateur résiste, si on cherche à modifier ce choix
- La promotion peut l'amener à préférer une marque à une autre pour un moment donné puis à revenir à ses choix initiaux.

2.10 Le « double jeopardy » d'Ehrenberg et Uncles :

Une marque, avec une pénétration forte, a plus d'acheteurs et d'actes d'achats (avec donc plus de présence à l'esprit⁴, dans le champ de choix) car pénétration et fréquence d'achat sont liées. Cette marque a donc de plus gros consommateurs qui ont plus de chance de l'acheter plus souvent et en plus grosse quantité. En d'autres termes une marque à forte pénétration fidélise mieux ses gros clients. C'est le concept de Double Jeopardy (Ehrenberg et al, 1997, 1994, 1995) détaillé ci-dessous :

Au niveau de l'achat d'une marque donnée :

- Les marques ont des nombres d'acheteurs très différents en lien avec leurs parts de marché
- La fréquence d'achat moyenne est assez proche entre les marques
- Les petites marques ont moins d'acheteurs, qui en plus en achètent moins souvent
- Le montant moyen acheté, par occasion d'achat, varie peut d'une marque à l'autre
- La majorité des acheteurs d'une marque donnée l'achète assez irrégulièrement (une fois par an), seul un petit noyau de consommateur l'achète régulièrement.

Au niveau de l'achat de marques concurrentes :

- La plupart des consommateurs achètent plus d'une seule marque sur une période donnée
- Peu de consommateurs d'une marque sont fidèles à celle-ci à 100%, sur une longue période de temps. Ceci est plus vrai pour les petites marques où les consommateurs sont plus infidèles et moins vrai pour les « grosses » marques où les consommateurs sont à l'inverse plus fidèles.
- Les consommateurs qui achètent 100% de leurs achats à une marque donnée sont en général des petits consommateurs
- Un consommateur d'une marque donnée achète plus à d'autres marques sur une longue période qu'à la marque elle-même
- Les autres marques achetées par le consommateur d'une marque donnée sont souvent les mêmes d'une marque à l'autre
- Le facteur dominant pour la déduplication entre les marques est la pénétration des marques dans une proportion constante
- La plupart des marchés ne sont pas segmentés pour les marques plus ou moins substituables entre elles
- La part des marques est à peu près la même pour les petits et les gros clients

Chaque consommateur individuel achète à partir d'un petit répertoire de marques avec une probabilité d'achat stable.

Les huit concepts théoriques ci-dessus montrent plus ou moins une inélasticité, ou un « bornage » de la consommation, en fournissant des explications. Il y a donc des théories fortes qui soutiennent l'idée de taux de nourriture « borné ».

⁴ <http://skyisnolimit.wordpress.com/2008/09/14/double-jeopardy-les-implications-de-cette-theorie-en-marketing/>

2.11 La prise en compte de la dimension temporelle :

Certains auteurs (Loddish et Melac ; 2006) ont montré que la longueur de la période d'observation avait une influence sur la perception des résultats de l'analyse. Nous essaierons de travailler sur la période totale des 24 mois. Bien souvent les analyses de ce type sont réalisées sur des périodes courtes, 6/8 semaines.

2.12 La prise en compte des facteurs socio démographiques :

Sans analyser très profondément cette dimension socio démographique qui n'est pas l'objet principal de notre recherche mais aussi peu ou pas bien renseignée dans nos données, nous pensons qu'il est important de montrer la dimension sociale que reflète l'élasticité ou l'inélasticité du taux de nourriture.

La consommation alimentaire est de plus en plus une préoccupation forte de l'optimisation du budget des ménages et plusieurs recherches (Berger et al, 2007, Laisney et Soyeux, 2009) ont montré que la crise impactait fortement cette consommation.

Le taux de nourriture vs le prix est un moyen de mieux prendre en compte les difficultés des ménages (Dreze, Hoch et Purk (1994)) et cela peut avoir un impact sur le positionnement prix/produits.

Première Partie

Recherche 1 : L'élasticité du taux de nourriture est-elle contrainte ?

Introduction

Ce troisième travail de recherche est donc dans la continuité des deux précédents, en zoomant plus précisément sur l'élasticité du taux de nourriture, un concept lié au taux de captation.

A partir de données de panel de home scanning de Nielsen (ménagères qui scannent les achats rentrant dans son foyer) nous allons essayer d'observer sur un historique assez long (24 mois), comment le taux de nourriture d'une marque évolue client par client. Nous allons le faire sur 6 marchés (Gordon 1994) en prenant des marchés à cycles d'achats/rachats de cycles courts à cycles longs.

Ce travail sur de gros corpus de données se situe dans la poursuite de travaux similaires (Zyfriden et al, 1980, Fader 1990) mais assez peu réalisés en France.

Notre but principal est de montrer que l'élasticité du taux de nourriture d'un consommateur est une donnée « bornée » dont les écarts à la moyenne varient dans une amplitude donnée, sans rarement sortir de cette amplitude.

Dans notre étude, nous prenons une définition classique de l'élasticité⁵ qui est l'habitude d'un indicateur à revenir à son état initial, nous ne travaillons pas sur la notion classique en économie d'élasticité prix⁶. Là, nous nous concentrons uniquement sur l'écart du taux de nourriture sur lui-même.

Nous modéliserons cette amplitude à la fois à un niveau agrégé (marque, semaine, mois, année) et à un niveau désagrégé (individu, marque, semaine, mois) en utilisant plusieurs techniques statistiques.

Nous testons l'ensemble des modèles sur des périodes temporelles différentes et sur des marques avec des PDM différentes pour valider la stabilité empirique de nos résultats.

Bien que le but de notre recherche soit de démontrer empiriquement et mathématiquement le côté « borné » de l'élasticité du taux de nourriture, nous essayons préalablement de justifier par la littérature, pourquoi le taux de nourriture est borné et nous cherchons si des variables exogènes n'expliquent pas une partie de cette inélasticité telle que le prix, la sensibilité au prix, la part de MDD, le revenu du consommateur ou le champ de choix.

Si notre recherche est marketing, avec une part importante de modélisation statistique nous considérons la consommation comme un fait social. Nous essaierons donc de faire une typologie des causes de l'inélasticité du taux de nourriture en fonction de différents segments de clients en prenant en compte la dimension socio-économique du client.

Notre recherche est une avancée pour différentes raisons :

- Peu d'analyses ont démontré que le taux de nourriture est contraint
- Peu d'analyses ont justifié cette élasticité/inélasticité par des explications endogènes ou exogènes au consommateur.
- Peu d'analyses ont considéré l'élasticité du taux de nourriture comme une quantification des moyens allouables à la stratégie marketing : promotion, fidélisation, politique de prix
- Peu de ces analyses ont été réalisées sur un aussi gros corpus de données individuelles.

Nous pensons que notre recherche peut faire avancer à la fois la compréhension du comportement de l'acheteur et aussi avoir des implications dans l'allocation des moyens à la fidélisation ou à la promotion hors médias ou médias des marques. Nous pensons que notre recherche peut avoir également des

⁵ Définition du Larousse <http://www.larousse.fr/dictionnaires/francais/%C3%A9lasticit%C3%A9/28160>

⁶ [http://fr.wikipedia.org/wiki/%C3%89lasticit%C3%A9_\(%C3%A9conomie\)](http://fr.wikipedia.org/wiki/%C3%89lasticit%C3%A9_(%C3%A9conomie))

implications dans le positionnement du prix des marques, pour notamment les rendre plus accessibles aux consommateurs les plus modestes.

2.13 Les questions de recherche :

A l'issue de cette revue de littérature et de la synthèse de nos précédents travaux, nous émettons quatre questions de recherche :

- L'élasticité du taux de nourriture (le taux de captation) d'une marque (ce que la marque peut prendre à ses concurrents en termes de consommations sur ses acheteurs actuels) est-il borné ?
- S'il est borné peut-on le prévoir sur la durée de consommation d'un consommateur ?
- Varie-t-il d'une catégorie de produits à l'autre ?
- Son élasticité est-elle socialement explicable ?

Nous allons essayer de répondre à ces questions, bien que cela ne soit pas le cœur de notre sujet nous essaierons de donner des explications exogènes à chacune de nos réponses : le pourquoi. L'enjeu est au cœur des problématiques d'efficacité du marketing mix.

2.14 La méthodologie :

Nous travaillons sur un jeu de données à un niveau individuel désagrégé :

- Un panel consommateur exhaustif de tous les acheteurs et les achats sur 24 mois d'historique à la semaine sur plusieurs marchés

2.14.1 Les données :

2.14.2 Les données de panel :

La collecte des données utilisées pour la partie empirique de cette étude, viennent de Home Scan de Nielsen France.

Home Scan : Panel de 15 000 foyers qui scanne tous les achats de produits de grande consommation qui rentrent dans le foyer. Les foyers en scannant les codes barre des produits permettent à Nielsen de savoir ce que les foyers ont acheté, tel jour, à tel prix et en telle quantité.

Nous travaillons sur six marchés.

Nous nous intéressons depuis quelques années au taux de nourriture et à son élasticité dans les panels de Home Scanning.

Les données que nous utilisons là, portent sur les périodes 2010/2011.

Pour des raisons de confidentialité nous ne travaillons pas sur un historique récent et ne citerons pas les marques.

Les six marchés :

- EAU PLATE : Délais inter achat court
- CAFE : Délai inter achat moyen
- PLAT SURGELE : Délai inter achat moyen
- DENTIFRICE : Délais inter achat moyen
- DESODORISANT : Délais inter achat moyen/long
- BROSSE A DENT : Délais inter achat long

Nota : Nous avons arbitrairement discrétisé la durée inter achat en semaines pour les personnes ayant fait plus de deux achats dans une même catégorie de produits en cycle court, moyen, moyen/long et long.

Nous allons calculer systématiquement les élasticités du taux de nourriture pour chaque marché sur des dimensions individuelles et temporelles différentes.

Nous utiliserons soit l'écart à la moyenne, soit l'écart type sur la totalité des 24 mois d'historique pour chaque consommateur par marque et par marché.

Les dimensions : PMG (Petit, Moyen, Gros client en CA), sexe, âge, taille du foyer, nombre d'enfants, consommateurs de grandes marques vs les MDD, prix, typologie client, place du produit sur son marché, différentes périodes...seront des variables soit de contrôle, soit explicatives.

Zoom produits et période d'analyse :

Nous commençons notre analyse en travaillant à la catégorie de produits pour les 6 catégories retenues puis nous zoomons sur la catégorie Café, enfin sur ce marché nous zoomerons sur deux marques, la principale marque du marché en PDM (la marque 1) et une marque plus petite en PDM (la marque 14) également pour vérifier si les effets observés sont les mêmes.

Par ailleurs à chaque fois que cela est pertinent nous analysons le taux de nourriture à la semaine, au mois et à l'année pour vérifier que nos conclusions soient valides.

Dans la partie modélisation, l'année ne présentant que 2 périodes n'est évidemment pas pertinente, nous ne prenons que la semaine ou 4 semaines (le mois).

2.15 L'analyse préalable et les modélisations utilisées :

Dans notre recherche nous cherchons à mettre en compétition des techniques statistiques différentes pour modéliser l'élasticité du taux de nourriture et valider ou invalider nos hypothèses de recherches :

- L'analyse descriptive des données
- Des modèles de prévision de deux types : un modèle ARIMA et un perceptron multicouche
- Des techniques d'analyse du panier de la ménagère : Apriori, Séquence
- La modélisation de données de panel
- Une méthode de classification automatique pour l'analyse de la dimension socioéconomique.

Nous présentons brièvement les techniques statistiques citées ci-dessus :

2.15.1 L'analyse descriptive des données :

L'objectif est de décrire, résumer ou représenter, par des indicateurs statistiques simples (indicateurs de dispersion, corrélation...), les nombreuses données disponibles (Bouroche et Saporta, 1990).

Nous présentons dans ce document que quelques résultats pertinents de cette analyse.

2.15.2 Les modèles ARIMA :

Il s'agit d'un modèle de prévision avec une variable à prédire Y et des variables prédictives

Créé dans les années 70, les modèles ARIMA (Auto Regressive Integrated Moving Average) permettent de traiter les séries non stationnaires après avoir déterminé le niveau d'intégration (le nombre de fois qu'il faut différencier la série avant de la rendre stationnaire (Bourbonnais et Usunier, 2000, David et al, 1990). Ce modèle utilise souvent des valeurs retardées (d'où le terme Auto Regressive) et de chocs aléatoires qui sont en général de moyennes nulles, de variances constantes et non autos corrélés (bruit blanc) ; quand la variable qui représente ces chocs est retardée, on parle de moyenne mobile (Moving Average).

2.15.3 Le perceptron multicouche :

Il s'agit d'un modèle de prévision avec une variable à prédire Y et des variables prédictives

Créé dans les années 70 il s'inspire de la théorie cognitive d'où son appellation de réseaux neurones (Larose et Vallaud, 2005, Wheelwright et McGee, 2003).

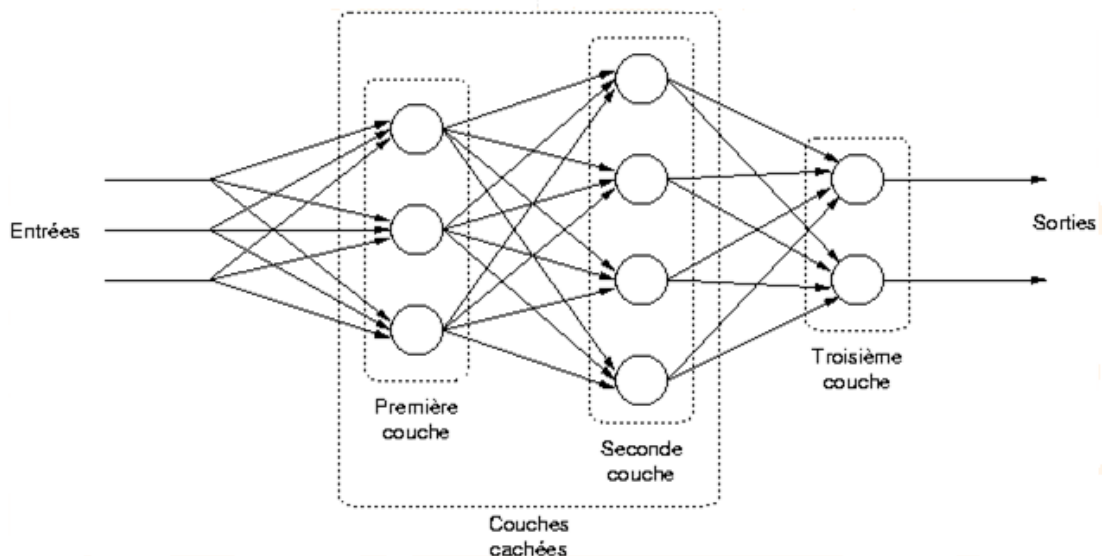


Figure 3 : Principe de fonctionnement d'un réseau neurone de type Perceptron multicouches

C'est un «classifieur» linéaire de type réseau neuronal formel organisé en plusieurs couches au sein desquelles une information circule de la couche d'entrée vers la couche de sortie uniquement. Chaque couche est constituée d'un nombre variable de neurones, les neurones de la couche de sortie correspondant toujours aux sorties du système.

L'architecture d'un réseau de neurones consiste à avoir une couche d'Input (variables X) et ensuite une succession de couches cachées qui peuvent être liées, puis d'une ou plusieurs sorties. Chaque couche intermédiaire prend ses entrées sur les couches précédentes sur lesquelles on applique une fonction dite de transfert. A chaque lien est associé un poids.

Le modèle cherche à ajuster sa prévision de Y en fonction des variables prédictives dont les neurones ajustent le poids pour obtenir la prévision la meilleure.

Nous avons déjà utilisé des réseaux dans notre modélisation du potentiel.

2.15.4 Les techniques d'analyse du panier de la ménagère :

Aussi appelée analyse d'affinité, l'analyse du panier de la ménagère (Tufféry, 2010) est l'utilisation de règles d'association pour déterminer quels produits sont achetés ensemble.

Nous utilisons deux techniques d'analyse de cooccurrence dans notre recherche.

2.15.4.1 Apriori :

Créé au début des années 90, l'algorithme Apriori (Larose et Vallaud, 2005) est un algorithme d'exploration de données conçu pour apprendre à formuler des règles d'associations. Il sert à reconnaître des propriétés qui reviennent fréquemment dans un ensemble de données et d'en déduire une catégorisation.

L'algorithme Apriori s'exécute en deux étapes :

- Génération de tous les items qui cooccurrent ensembles des plus fréquents au moins fréquents
- Génération de toutes les règles d'associations à partir des fréquences d'occurrences sur un seuil de support de la règle (nombre de cooccurrence sur la somme de toutes les concurrences) et un seuil de confiance donné.

2.15.5 Séquence :

Il s'agit à peu près de la même approche que dans Apriori mais ici le modèle tient compte du temps et donc de l'ordre des cooccurrences dans le temps : après tel achat et tel achat il se produira alors celui-ci.

2.15.6 La modélisation de données de panel :

L'analyse de données de panel (Casin, 1999) est une technique statistique particulière qui permet de tenir compte dans un modèle de deux dimensions à la fois de la variance dans le temps des indicateurs (la dimension temporelle) et les panélistes entre eux (la dimension individuelle).

Elle permet de tenir compte de l'influence de caractéristiques non observables des individus sur leur comportement :

- contrôle de l'hétérogénéité observée
- contrôle de l'hétérogénéité inobservée (effets spécifiques)

2.16 Le processus de validation des modèles

Pour valider les résultats de nos modèles nous utilisons des procédures classiques de validation en modélisation (Larose et Vallaud, 2005)

La mise en compétition de techniques très différentes sur le même ensemble de données pour mesurer la convergence des résultats est une démarche classique de validation.

Nous rajoutons une comparaison entre le taux de nourriture prévu et le taux réel sur un ensemble de validation où celui-ci est connu.

2.17 Résultats de la recherche :

2.17.1 L'analyse descriptive de données issues d'un panel de consommateurs :

Nous avons obtenu un fichier de données portant sur l'ensemble des achats des panélistes sur 24 mois d'historique au niveau individuel pour les différents marchés étudiés.

2.18 Les principaux BPM (Brand Performance Measurement) étudiés :

Nous avons reconstruit les indicateurs de base pour suivre nos différents marchés :

NA	Nombre d'acheteurs (pour calcul de la pénétration)		
AA	Nombre d'actes d'achat		
QA	Quantité	Brosse à dent	en nb d'unité
		Désodorisant	en nb d'unité
		Dentifrice	en nb de tube
		Plat Surgelé	en kg
		Café	en kg
		Eau Plate	en nb de litre
CA	Chiffre d'affaires en €		
TN	Taux de nourriture en %		
	TN = Nombre d'achats par marque / Total d'achat		

Ce tableau présente les principaux BPM et précise le type de quantité pour chaque catégories de produits étudiées, kilogrammes, litres, unités.

Tableau 1 : Les principaux BPM

Nous avons utilisé par ailleurs, parfois calculé, les PDM (Part De Marché) volumes et valeurs :

- PDM volumes : % des QA de la marque étudiée sur les QA marché
- PDM valeur : % du CA de la marque étudiée sur le CA marché

2.19 Les évolutions par année :

Pour chaque marché nous avons calculé les indicateurs résumés ci-dessous :

- Le nombre de marques par marché : plus il est élevé plus le choix du consommateur est important
- Le nombre de marques moyen achetées : ce nombre est assez stable d'un marché à l'autre montrant que le « *relevant set* » des consommateurs est finalement assez limité : pour un marché à 16 marques il peut être de 3,1 pour le café et pour un marché 170 marques il est plus faible à 2,2 (les déodorants). Le nombre de marques ne crée pas forcément la recherche de diversité chez le consommateur.
- Les quantités achetées par les consommateurs : exprimées en différentes unités selon les marchés montrent une grande volatilité.
- Le TN de nourriture en % il varie de 29,7 pour l'eau plate à 66,9 pour les brosses à dents, certains marchés présentent plus de risque de « *switches* » inter marques que d'autres.
- L'écart type du taux de nourriture est lui assez contraint entre 29,4% et 33,6. Le consommateur présente une élasticité assez limitée quels que soient les marchés.
- Le délai inter achat qui est assez variable selon les marchés.
- Le prix moyen d'un produit qui est lui-même assez variable selon les marchés.

Seul le nombre de marques achetées et l'élasticité du taux de nourriture semblent donc contraints dans des amplitudes « faibles ».

Année 1 :

Sur 12 mois		Nb de marque sur le marché	Nb de marque moyen achetées	QA/NA	TN en %	Ecart-type en %	Délai inter achats en semaines	Prix moyen en €
Brosse à dent	Long	28	1,7	2.9 unités	66,9	32,1	21,4	2,30 €
Désodorisant	Moyen Long	170	2,2	5.9 unités	52,4	33,6	11,6	2,50 €
Dentifrice	Moyen	47	2,2	7.7 tubes	50,3	32,4	10,2	1,70 €
Plat Surgelé	Moyen	150	2,1	5.6 kg	52,4	35,4	8,8	5,11 €
Café	Moyen	16	3,1	6 kg	36,7	31,5	3,7	9,40 €
Eau Plate	Court	115	3,9	166.4 litres	29,7	29,4	3,1	0,23 €
Moyenne		NS	2,5	NS	48,1	32,4	NS	3,54 €

Tableau 1 : Les principaux BMP sur les catégories étudiées sur l'année 1

Année 2 :

Sur 12 mois		Nb de marque sur le marché	Nb de marque moyen achetées	QA/NA	TN en %	Ecart-type en %	Délai inter achats en semaines	Prix moyen en €
Brosse à dent	Long	28	1,6	2.8 unités	68,2	31,9	22,7	2,40 €
Désodorisant	Moyen Long	170	2,3	5.6 unités	54,7	34	12,1	2,55 €
Dentifrice	Moyen	47	2,2	7.3 tubes	51,6	32,8	10,7	1,74 €
Plat Surgelé	Moyen	150	2	5.3 kg	55,3	35,9	9,5	5,10 €
Café	Moyen	16	3	6 kg	37,1	31,7	3,8	9,64 €
Eau Plate	Court	115	3,8	148 litres	30,7	29,8	3,4	0,23 €
Moyenne		NS	2,5	NS	49,6	32,7	NS	3,61 €

Les deux tableaux ci-dessus présentent sur deux années consécutives pour chaque catégorie de produits étudiée la longueur du cycle discrétisé en classe, le nombre de marques dans la catégorie, le nombre de marques moyen achetées par acheteur, les quantités moyennes par acheteur, le taux de nourriture par acheteur, l'écart type du taux de nourriture en %, les délais entre deux achats pour un acheteur de la catégorie en semaine et le prix moyen d'un produit de la catégorie acheté en euro.

Tableau 2 : Les principaux BMP sur les catégories étudiées sur l'année 2

Les écarts type des taux de nourriture sont assez similaires d'une année sur l'autre.

Le nombre de marques limitées et le faible écart type du taux de nourriture dans une moyenne de 32,7 montrent que le consommateur est grégaire, petit ensemble déterminant, peu d'écart de consommation ou du moins dans une amplitude limitée.

2.20 L'évolution par année sur différentes périodes inter achats

Nous concentrons notre analyse sur le marché du café et l'analysons sur les deux périodes toutes marques confondues.

Le taux de nourritures global à la semaine													
		Période 1				Période 2				Période Totale			
		P	M	G	Total	P	M	G	Total	P	M	G	Total
Moyenne		94,1%	89,2%	83,1%	87,5%	94,2%	89,4%	83,3%	87,7%	93,7%	89,3%	83,5%	87,6%
Min		14,3%	7,7%	4,3%	4,3%	11,1%	5,3%	6,3%	5,3%	9,1%	5,3%	4,3%	4,3%
Max		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Écart type		16,7%	21,7%	25,9%	23,2%	16,6%	21,6%	25,7%	23,0%	17,3%	21,6%	25,6%	23,1%
Intervalle de confiance	Min	61,3%	46,7%	32,4%	32,4%	61,7%	47,1%	32,9%	32,9%	59,8%	47,0%	33,3%	33,3%
	Max	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Nb moyen marque		2,4	3,6	4,0	3,1	2,3	3,5	4,0	3,0	3,0	4,6	5,1	3,9

Le taux de nourritures sur période intersite (4 semaines)

Le taux de nourritures sur période intersite (4 semaines)													
		Période 1				Période 2				Période Totale			
		P	M	G	Total	P	M	G	Total	P	M	G	Total
Moyenne		87,3%	76,7%	68,5%	75,2%	87,4%	76,9%	68,8%	75,5%	86,6%	77,0%	69,1%	75,4%
Min		9,1%	4,0%	3,3%	3,3%	9,1%	5,3%	4,0%	4,0%	5,9%	4,0%	3,3%	3,3%
Max		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Écart type		22,9%	28,0%	30,3%	29,0%	23,0%	27,9%	30,3%	28,9%	23,5%	28,0%	30,2%	29,0%
Intervalle de confiance	Min	42,3%	21,8%	9,2%	9,2%	42,4%	22,1%	9,5%	9,5%	40,5%	22,2%	9,8%	9,8%
	Max	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Nb moyen marque		2,4	3,6	4,0	3,1	2,3	3,5	4,0	3,0	3,0	4,6	5,1	3,9

Le taux de nourritures global sur 12 mois

Le taux de nourritures global sur 12 mois													
		Période 1				Période 2				Période Totale			
		P	M	G	Total	P	M	G	Total	P	M	G	Total
Moyenne		47,8%	32,2%	28,2%	36,7%	48,2%	32,5%	28,5%	37,1%	46,4%	32,6%	28,5%	36,0%
Min		7,1%	3,3%	0,5%	0,5%	7,1%	3,3%	0,5%	0,5%	3,7%	1,8%	0,5%	0,5%
Max		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Écart type		32,2%	29,0%	29,7%	31,5%	32,2%	29,0%	30,1%	31,7%	31,7%	28,7%	29,6%	30,9%
Intervalle de confiance	Min	-15,3%	-24,5%	-30,0%	-30,0%	-14,9%	-24,4%	-30,4%	-30,4%	-15,6%	-23,7%	-29,6%	-29,6%
	Max	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Nb moyen marque		2,4	3,6	4,0	3,1	2,3	3,5	4,0	3,0	3,0	4,6	5,1	3,9

Tableau portant uniquement sur le taux de nourriture et ses indicateurs de dispersion sur les deux années (Période 1, Période 2, la période totale étant la somme des deux périodes) par segment PMG. En vert les moyennes du taux de nourriture, en rouge l'écart type constaté sur période.

Tableau 3 : Evolution des principaux BMP sur les catégories étudiées par année

Si les taux de nourriture chutent fortement entre la semaine, 4 semaines (le mois) et l'année au global et par segment PMG, on constate que l'écart type oscille entre 23% et 31,5%

Si on fait une analyse similaire par marque on voit que le taux de nourriture à la semaine est évidemment élevé, compris entre 89% pour le max et 77% pour le min avec une moyenne à 84,9%. L'écart type est compris entre 28,3 et 19,5.

Dans le reste de notre étude nous allons choisir la marque 1, grande marque, avec une PDM de 16,3% et une petite marque, la marque 14 avec une PDM de 1,3%.

MARQUE	TN Moyen	TN Min	TN Max	TN Écart type
Marque 16	88,8%	6,3%	100,0%	22,1%
Marque 1	85,8%	10,0%	100,0%	26,9%
Marque 2	86,5%	11,1%	100,0%	24,2%
Marque 3	86,9%	5,3%	100,0%	23,6%
Marque 4	84,8%	10,0%	100,0%	25,0%
Marque 5	82,6%	9,1%	100,0%	26,4%
Marque 6	77,3%	9,1%	100,0%	28,3%
Marque 7	87,4%	9,1%	100,0%	22,7%
Marque 8	87,2%	12,5%	100,0%	23,3%
Marque 9	79,2%	11,1%	100,0%	27,9%
Marque 10	86,0%	7,7%	100,0%	24,4%
Marque 11	91,7%	7,7%	100,0%	19,5%
Marque 12	81,4%	9,1%	100,0%	26,8%
Marque 13	87,0%	8,3%	100,0%	22,2%
Marque 14	82,2%	8,3%	100,0%	26,7%
Marque 15	89,0%	10,0%	100,0%	21,9%
Moyenne	84,9%	9,0%	100,0%	24,5%

Le taux nourriture moyen et ses indicateurs de dispersion pour chacune des marques sur le marché du café. En rouge les deux marques que nous retenons pour une analyse plus détaillée.

Tableau 4 : Indicateurs de dispersion du TN par marque

A la semaine sur un achat de café donné le taux de nourriture d'une marque est évidemment élevé car la plupart des achats sur une période si courte sont souvent mono marque.

La marque 1 :

Unité temporelle	TN Moyen	TN Min	TN Max	TN Ecart type
Semaine	85,8%	10,0%	100,0%	26,9%
4 semaines (mois)	74,7%	1,2%	100,0%	31,3%
12 mois	24,8%	0,8%	100,0%	29,2%
			Moyenne	29,2%

Le taux de nourriture et ses indicateurs de dispersion sur 3 périodes temporelles différentes

Tableau 5 : Récapitulatif des taux de nourriture de la marque 1 sur différentes périodes

A la semaine le taux de nourriture de la marque 1 est de 85,8%.

Sur des périodes plus longues le taux de nourriture chute car le consommateur achète plus de marques mais l'écart type de celui-ci est compris, quelle que soit la période, entre 26,9 et 31,3 avec une moyenne de l'écart type à 29,2%.

La marque 14 :

Unité temporelle	TN Moyen	TN Min	TN Max	TN Ecart type
Semaine	89,0%	10,0%	100,0%	21,9%
4 semaines (mois)	71,0%	2,9%	100,0%	29,5%
12 mois	19,0%	0,9%	100,0%	22,0%
			Moyenne	24,4%

Tableau 6 : Récapitulatif des taux de nourriture de la marque 2 sur différentes périodes

A la semaine le taux de nourriture de la marque 14 est de 89%.

Sur des périodes plus longues le taux de nourriture chute car le consommateur achète plus de marques mais l'écart type de celui-ci est compris quelle que soit la période entre 21,9 et 29, (avec une moyenne de l'écart type à 24,4%).

Dans l'ensemble des cas ci-dessus (2 marques X 3 périodes) l'écart type du taux de nourriture est contraint.

2.21 Prévoir le taux de nourriture

Bien que notre démarche soit très empirique en maximisant les observations individuelles simples sur de gros volumes de données, sur de longues périodes historiques, nous procédons à des modélisations qui n'ont que des rôles confirmatoires (Herniter et Howard, 1964, Naert et Leeflang, 1978).

Nous cherchons à modéliser la stabilité d'un comportement de consommateur : l'écart type de son taux de consommation à un produit dans le temps. A priori les variables exogènes sont simplement explicatives mais pourraient ne pas rentrer dans le modèle : on ne cherche pas à faire un modèle causal, on cherche uniquement à « modéliser une entropie ou non » basée sur un comportement passé.

On utilise que les seules interactions, et le poids relatif des autres marques dans le set acheté par le consommateur pour prédire la part que va s'allouer la marque étudiée dans ses actes d'achat.

Ce qui va donc « peser » dans le modèle c'est surtout dans les achats antérieurs, la fréquence et la part des marques achetées auparavant par ce consommateur.

Pour prédire ce type de comportement, on trouve dans les données de panels, plusieurs types de modèles dont nous allons tester trois approches :

2.22 Deux approches agrégées directes :

On essaie de prévoir l'évolution du taux de nourriture de la marque sur le set des autres marques achetées sur une période.

C'est une approche agrégée (Wheelwright et McGee, 1983). On prédit le taux de nourriture et l'écart moyen de ce taux sur la période pour la marque 1 qui est la plus forte marque du marché en PDM (16,3% de PDM) puis pour la marque 14 qui est une marque plus challenger (1,3 pts de PDM)

On teste à chaque fois deux types de modèles :

- Une approche par modèle ARIMA (Bourbonnais et Usunier, 2000, David et al, 1990)
- Une approche par réseau Neurone (Larose et Vallaud, 2005, Wheelwright et McGee, 2003)

2.23 Une approche désagrégée indirecte :

On essaie de prévoir au niveau individuel, les achats totaux et les achats des marques 1 et 14 pour en déduire leur taux de nourriture prévisible.

2.23.1 Les approches agrégées :

2.23.1.1 Le modèle ARIMA :

Nous avons testé plusieurs modèles (Gosh, Neslin et Shoemaker, 1984) mais après plusieurs itérations nous ne gardons qu'un des meilleurs résultats obtenus.

On est sur un historique de 104 semaines, le taux de nourriture est calculé par semaine et par marque : $\text{taux de nourriture} = \frac{\text{la quantité achetée par marque}}{\text{la quantité totale achetée}}$.

On utilise 90 semaines pour l'apprentissage et l'estimation du modèle et les 14 dernières semaines pour le test et la validation du modèle.

Avec ce modèle on cherche à prédire le taux de nourriture des deux marques étudiées à la semaine.

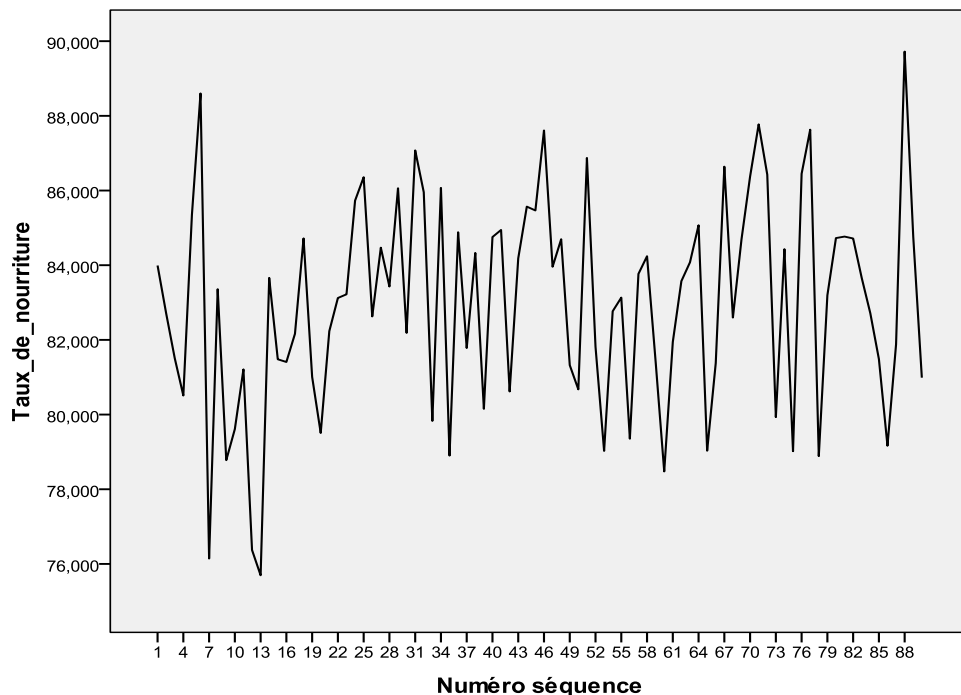
Champ	Type	Valeurs	Manquantes	Vérifier	Direction
Num Semaine	Intervalle	[1,104]		Aucun	In
Taux de nourriture	Intervalle	[75.69540613085766,90.894561...		Aucun	Out
Prix CARTE NOIR	Intervalle	[3.655611285266458,4.4771999...		Aucun	In
Prix moyen -Tous cafés	Intervalle	[2.73031758293833,3.40765815...		Aucun	In
PRIX Ecart type -Tous cafés	Intervalle	[0.9150933164514412,2.067429...		Aucun	In
Nb de marques présentes par semaine	Intervalle	[56,84]		Aucun	In
Nb de marques_AGGLOM	Intervalle	[1,3]		Aucun	In
Nb de marques_ATOMISE	Intervalle	[4,7]		Aucun	In
Nb de marques_GRAINS	Intervalle	[5,18]		Aucun	In
Nb de marques_LIQUIDE	Intervalle	[1,2]		Aucun	In
Nb de marques_LYOPHIL	Intervalle	[5,12]		Aucun	In
Nb de marques_MOULU	Intervalle	[47,71]		Aucun	In
Nb de marques_NORMAL	Intervalle	[4,8]		Aucun	In
Nb de marques_SOLUBLE	Intervalle	[1,2]		Aucun	In
Nb de marques_ARABICA	Intervalle	[41,63]		Aucun	In
Nb de marques_CAFE CHIC LAIT	Intervalle	[2,3]		Aucun	In
Nb de marques_CAFE CHIC +ING	Intervalle	[0,1]		Aucun	In
Nb de marques_CAFE CHICOREE	Intervalle	[4,8]		Aucun	In
Nb de marques_CAFE SOLUBLE	Intervalle	[6,13]		Aucun	In
Nb de marques_CHICOREE PURE	Intervalle	[1,6]		Aucun	In
Nb de marques_CHICOREE+INGRED	Intervalle	[0,2]		Aucun	In
Nb de marques_CHICOREE+LAIT	Intervalle	[0,1]		Aucun	In
Nb de marques_MELANGE	Intervalle	[14,26]		Aucun	In
Nb de marques_ROBUSTA	Intervalle	[4,13]		Aucun	In

Les données en entrée du modèle, à droite « Champs » sont les variables en clair, le type « échelle » veut dire que toutes les variables retenues au départ sont dans un format numérique et continu, « Valeurs » ce sont les valeurs des variables, « Manquante » le fait qu'il y a un remplacement de données manquantes, « Vérifier » la vérification de certaines valeurs et « Direction » précise le type de variable pour le modèle, la cible rouge, c'est une donnée à prédire, « in » ce sont les données explicatives.

Figure 4 : Liste des variables en entrée du modèle ARIMA

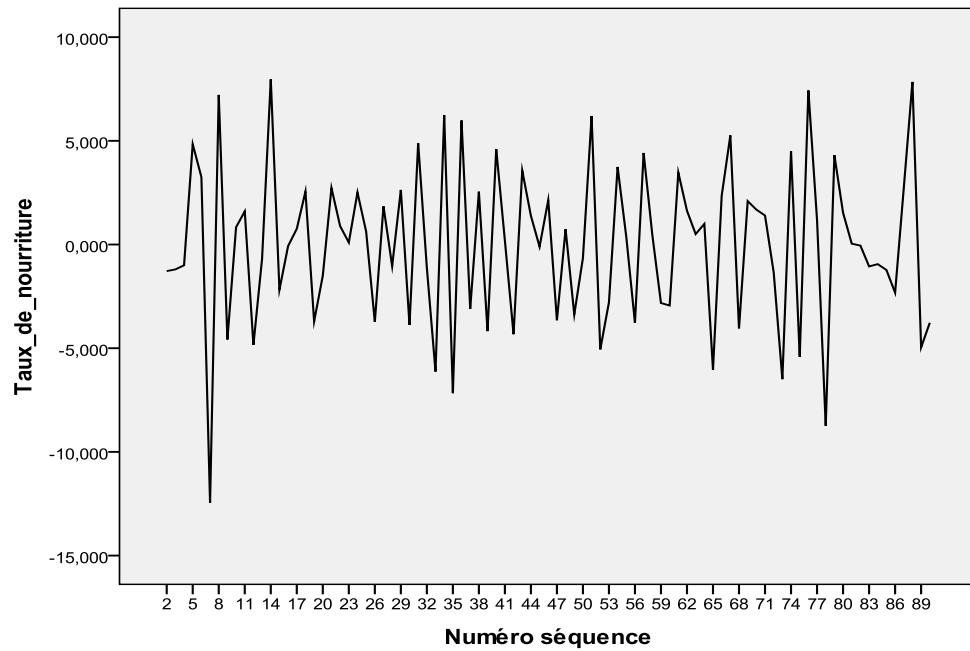
La marque 1 :

A la semaine :



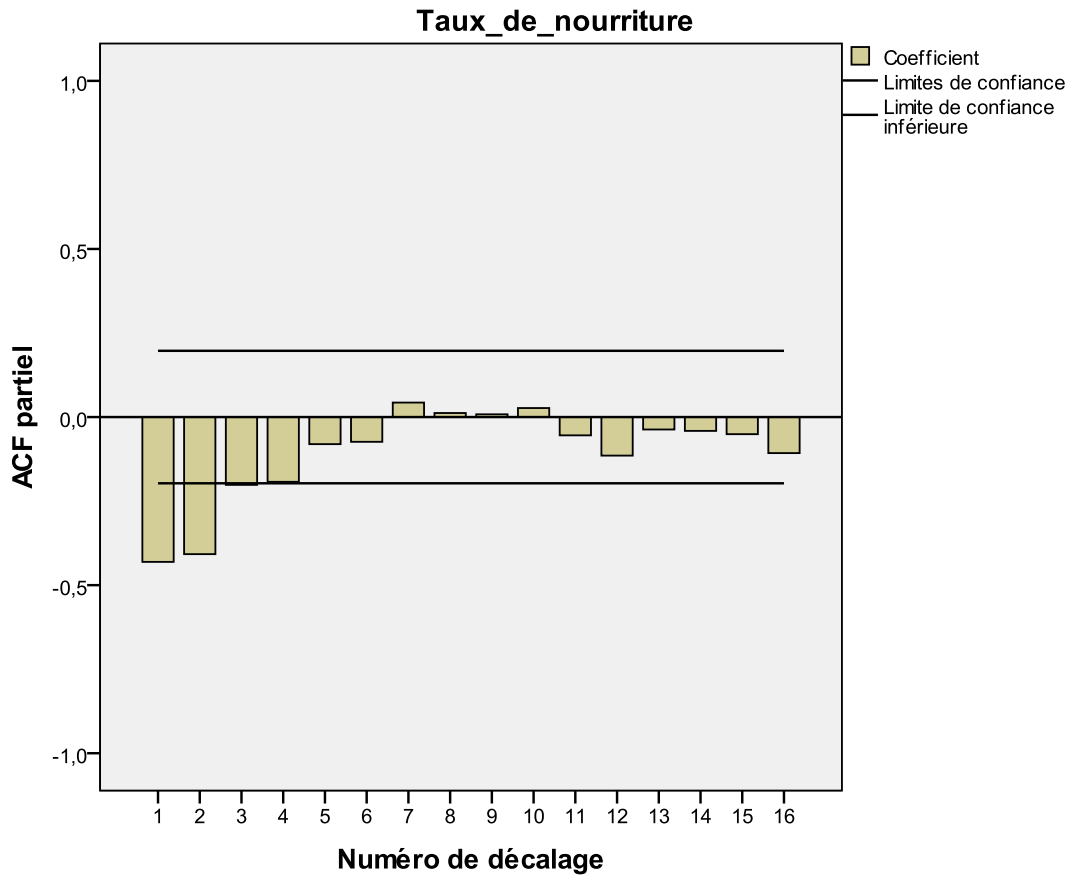
Graphique 1 : La série brute, le TN de la marque 1, la variable à prédire, donnée hebdomadaire

On a une tendance à la hausse, pour supprimer la tendance on différencie la série. Les données sont à la semaine, donc on n'a pas de saisonnalité.

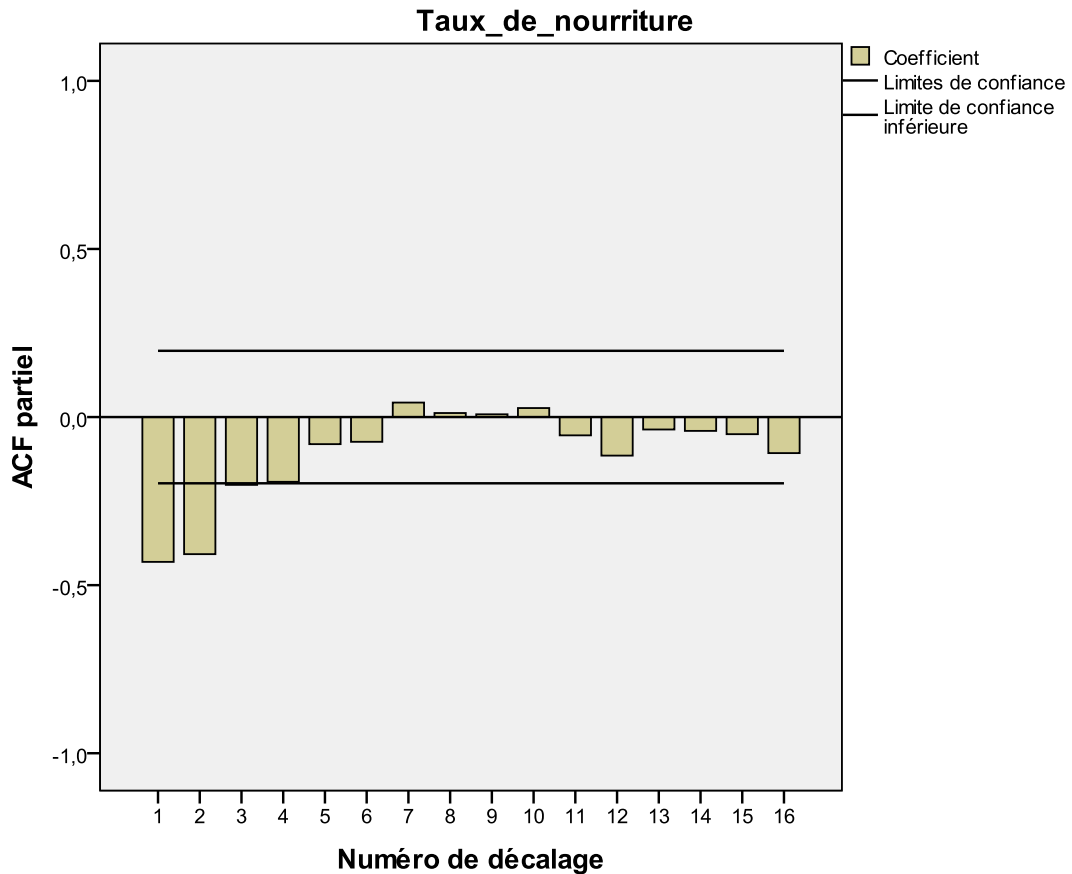


Transforme : différence(1)

Graphique 2 : La série différenciée, le TN de la marque 1, la variable à prédire



Graphique 3 : La fonction d'autocorrélation de la série brute du modèle ARIMA, marque 1, données hebdomadaires



Graphique 4 : La fonction d'autocorrélation partielle sur la série différenciée du modèle ARIMA, marque 1, données hebdomadaires

Après plusieurs essais et tests des paramètres, nous retenons un modèle : ARIMA(2, 1, 0) avec une fonction de transfert sur les variables explicatives (liste ci-dessous) :

Description du modèle

Description du modèle			Type de modèle
ID de modèle	Taux_de_nourriture	Modèle_1	ARIMA(2,1,0)

Estimation des paramètres

Modèle	Nombre de variables indépendantes	Statistiques de qualité d'ajustement du modèle				Ljung-Box Q(18)			Nombre de valeurs éloignées
		R-deux stationnaire	R-deux	RMSE	BIC standardisé	Statistiques	DDL	Sig.	
Taux_de_nourriture-Modèle_1	14	,833	,728	2,069	3,373	21,501	16	,160	0

R² Stationnaire : R² de la série prédite stationnaire (sans tendance et sans saisonnalité), R² : Ajustement de la série, RMSE : Racine moyenne des erreurs au carré, BIC standardisé : Critère d'information de Bayes plus il est grand plus le modèle est satisfaisant, Statistiques , DDL : Degré de liberté, Sig : Si Sig > à 0,05 les erreurs (résidus) sont aléatoires

Tableau 7 : Résultat du modèle ARIMA, marque 1, données hebdomadaires

Nous obtenons un bon R2 (0.83).

La statistique de Ljung-Box indique que les erreurs résiduelles des 2 modèles sont aléatoires (bruit blanc), une valeur de signification supérieure à 0,05 indique que les erreurs résiduelles sont aléatoires.

Paramètres du modèle ARIMA					Estimation
Taux_de_nourriture- Modèle_1	Taux_de_nourriture	Aucune transformation	AR	Lag 1	- ,708
				Lag 2	- ,777
Prix_CARTE_NOIR	Aucune transformation	Aucune transformation	Différence		1
			Numérateur	Lag 0	5,795
PRIX_Ecart_type_Tous_ca fés	Aucune transformation	Aucune transformation	Différence		1
			Numérateur	Lag 0	2,556
			Différence		1
Nb_de_marques_présente s_par_semaine	Aucune transformation	Aucune transformation	Dénominateur	Lag 1	,895
				Lag 2	- ,754
			Numérateur	Lag 0	- ,464
Nb_de_marques_LIQUIDE	Aucune transformation	Aucune transformation	Différence		1
			Dénominateur	Lag 1	- ,203
				Lag 2	,470
Nb_de_marques_LYOPHIL	Aucune transformation	Aucune transformation	Numérateur	Lag 0	,889
			Différence		1
			Dénominateur	Lag 1	-1,445
Nb_de_marques_MOULU	Aucune transformation	Aucune transformation		Lag 2	- ,948
			Numérateur	Lag 0	,119
			Différence		1
Nb_de_marques_ARABIC A	Aucune transformation	Aucune transformation	Dénominateur	Lag 1	,971
				Lag 2	- ,587
			Numérateur	Lag 0	,276
Nb_de_marques_CAFE_C HIC_LAIT	Aucune transformation	Aucune transformation	Différence		1
			Dénominateur	Lag 1	,226
				Lag 2	- ,022
Nb_de_marques_CAFE_C HIC_ING	Aucune transformation	Aucune transformation	Numérateur	Lag 0	,364
			Différence		1
			Dénominateur	Lag 1	- ,987
Nb_de_marques_CAFE_C HICOREE	Aucune transformation	Aucune transformation		Lag 2	- ,412
			Numérateur	Lag 0	-3,115
			Différence		1
Nb_de_marques_CAFE_C HICOREE	Aucune transformation	Aucune transformation	Dénominateur	Lag 1	,445
				Lag 2	,050
			Numérateur	Lag 0	1,211
Nb_de_marques_CAFE_C HICOREE	Aucune transformation	Aucune transformation	Différence		1
			Dénominateur	Lag 1	- ,080
				Lag 2	,736
Nb_de_marques_CAFE_C HICOREE	Aucune transformation	Aucune transformation	Numérateur	Lag 0	,873
			Différence		1
			Dénominateur	Lag 1	- ,235
				Lag 2	,133

Paramètres du modèle ARIMA					Estimation
Taux_de_nourriture- Modèle_1	Nb_de_marques_CAFE_SO LUBLE	Aucune transformation	Numérateur	Lag 0	,225
			Différence		1
Nb_de_marques_CHICOREE _PURE	Aucune transformation	Aucune transformation	Dénominateur	Lag 1	-1,256
				Lag 2	- ,949
			Numérateur	Lag 0	,086
Nb_de_marques_MELANGE	Aucune transformation	Aucune transformation	Différence		1
			Dénominateur	Lag 1	-1,416
				Lag 2	- ,985
Nb_de_marques_ROBUSTA	Aucune transformation	Aucune transformation	Numérateur	Lag 0	- ,051
			Différence		1
			Dénominateur	Lag 1	1,136
Nb_de_marques_ROBUSTA	Aucune transformation	Aucune transformation		Lag 2	- ,942
			Numérateur	Lag 0	,049
			Différence		1
Nb_de_marques_ROBUSTA	Aucune transformation	Aucune transformation	Dénominateur	Lag 1	- ,225
				Lag 2	-1,000
			Numérateur	Lag 0	

Ce tableau présente les poids du modèle associé au variable en fonction de différents effets retards (lag) pour trouver le « meilleur » modèle

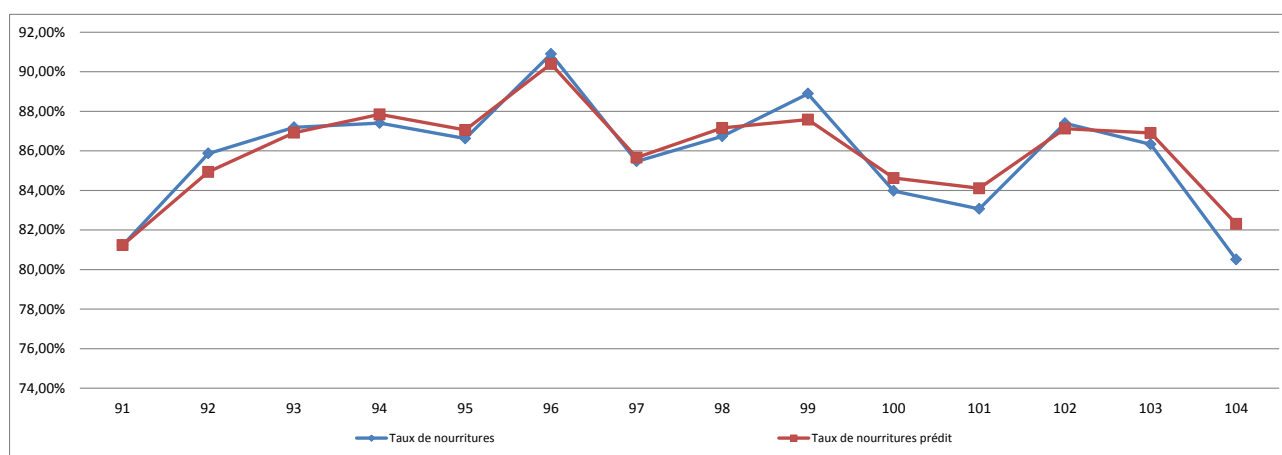
Tableau 8 : Poids et effets des variables dans le modèle ARIMA, marque 1, données hebdomadaires

On applique la prévision du modèle sur les 14 semaines non utilisées dans la phase apprentissage et estimation du modèle.

Semaines	Taux de nourritures	Taux de nourritures prédit	Erreur absolue
91	81,23%	81,24%	0,01%
92	85,86%	84,94%	0,93%
93	87,19%	86,92%	0,27%
94	87,40%	87,85%	0,44%
95	86,62%	87,05%	0,43%
96	90,90%	90,40%	0,50%
97	85,47%	85,66%	0,19%
98	86,73%	87,16%	0,43%
99	88,88%	87,58%	1,30%
100	83,97%	84,63%	0,65%
101	83,07%	84,11%	1,05%
102	87,40%	87,13%	0,27%
103	86,33%	86,90%	0,57%
104	80,50%	82,31%	1,81%
Erreur absolue moyenne			0,63%

Tableau 9 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 1, données hebdomadaires

L'erreur absolue moyenne est de 0.63% sur les 14 semaines utilisées pour le test. Les deux séries sont très bien ajustées.



Graphique 5 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision du TN, marque 1, données hebdomadaires

A la semaine les taux de nourriture entre la période observée et la période prédite sont :

TN moyen réel		TN moyen prédit	
	85,82%		85,99%

Tableau 10 : Qualité de l'ajustement du modèle ARIMA , marque 1, données hebdomadaires

Sur 4 semaines (mois)

Estimation des paramètres

Modèle	Nombre de variables indépendantes	Statistiques de qualité d'ajustement du modèle				Ljung-Box Q(18)			Nombre de valeurs éloignées
		R-deux stationnaire	R-deux	RMSE	BIC standardisé	Statistiques	DDL	Sig.	
Taux_de_nourriture-Modèle_1	17	,878	,847	5,818	6,567	15,771	16	,469	0

Tableau 11 : Résultat du modèle ARIMA, marque 1 sur 4 semaines (mois)

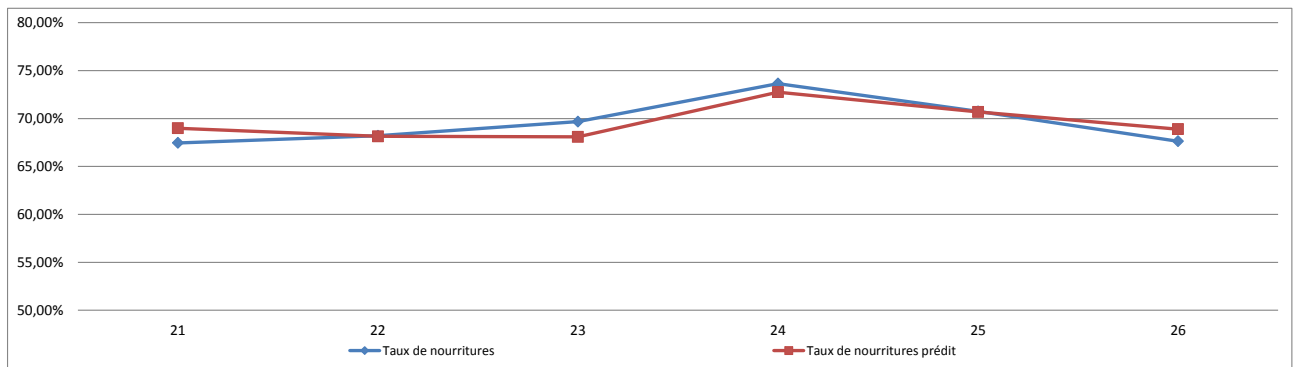
Un bon R2 (0.88)

La statistique de Ljung-Box indique que les erreurs résiduelles des 2 modèles sont aléatoire (bruit blanc), une valeur de signification supérieure à 0,05 indique que les erreurs résiduelles sont aléatoires.

4 semaines	Taux de nourritures	Taux de nourritures prédit	Erreur absolue
21	67,46%	68,99%	1,53%
22	68,20%	68,15%	0,06%
23	69,68%	68,09%	1,58%
24	73,65%	72,76%	0,90%
25	70,75%	70,69%	0,06%
26	67,63%	68,90%	1,27%
Erreur absolue moyenne			0,90%

Tableau 12 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 1 sur 4 semaines (mois)

L'erreur absolue moyenne est de 0.9% sur les 6 mois utilisés pour le test.



Graphique 6 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision, marque 1 sur 4 semaines (mois)

Sur quatre semaines les taux de nourriture entre la période observée et la période prédite sont :

TN moyen réel		TN moyen prédit	
	69,56%		69,60%

Tableau 13 : Qualité de l'ajustement du modèle ARIMA , marque 1 sur 4 semaines (mois)

Un très faible écart, une très bonne prévision.

La marque 14 :

A la semaine

Estimation des paramètres

Modèle	Nombre de variables indépendantes	Statistiques de qualité d'ajustement du modèle			Ljung-Box Q(18)			Nombre de valeurs éloignées
		R-deux stationnaire	R-deux	BIC standardisé	Statistiques	DDL	Sig.	
Taux_de_nourriture-Modèle_1	18	,823	,697	5,433	16,046	15	,379	0

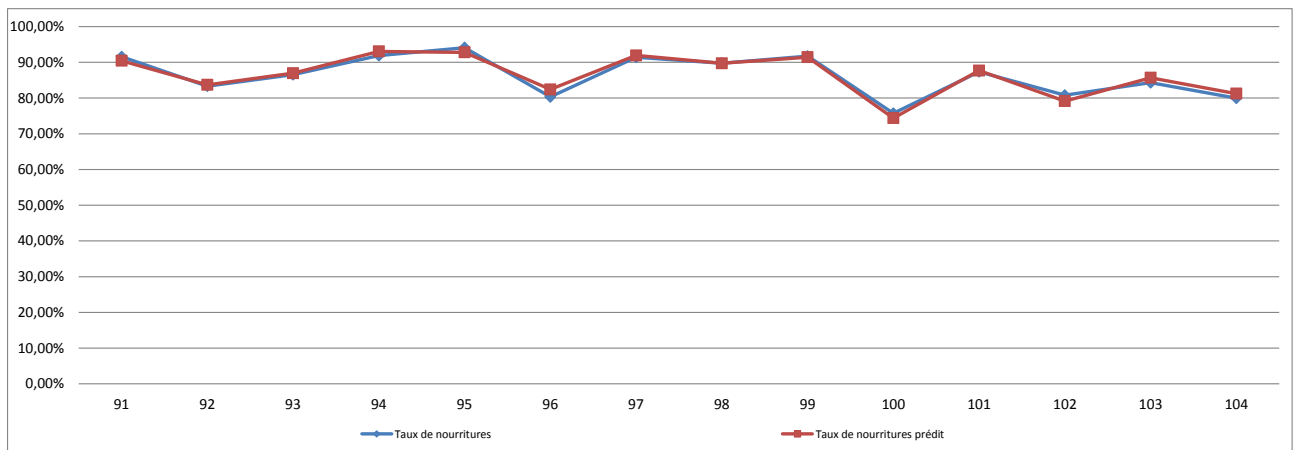
Tableau 14 : Résultat du modèle ARIMA, marque 14, données hebdomadaires

Nous obtenons un bon R2 (0.82).

Semaines	Taux de nourritures	Taux de nourritures prédit	Erreur absolue
91	91,56%	90,44%	1,12%
92	83,33%	83,68%	0,35%
93	86,53%	86,97%	0,44%
94	91,87%	93,05%	1,18%
95	94,10%	92,79%	1,32%
96	80,33%	82,40%	2,08%
97	91,40%	91,93%	0,52%
98	89,74%	89,72%	0,01%
99	91,72%	91,43%	0,29%
100	75,71%	74,38%	1,33%
101	87,35%	87,69%	0,34%
102	80,79%	79,14%	1,65%
103	84,30%	85,69%	1,39%
104	79,95%	81,24%	1,29%
Erreur absolue moyenne			0,95%

Tableau 15 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 14, données hebdomadaires

L'erreur absolue est faible, le R² élevé, la prévision est de bonne qualité.



Graphique 7 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision du TN, marque 14, données hebdomadaires

A la semaine les taux de nourriture entre la période observée et la période prédite sont :

TN moyen réel		TN moyen prédit	
	88,33%		86,47%

Tableau 16 : Qualité de l'ajustement du modèle ARIMA de la marque 14, données hebdomadaires

La prévision est d'assez bonne qualité.

Sur 4 semaines (mois)

Estimation des paramètres								
Modèle	Nombre de variables indépendantes	Statistiques de qualité d'ajustement du modèle			Ljung-Box Q(18)			Nombre de valeurs éloignées
		R-deux stationnaire	R-deux	BIC standardisé	Statistiques	DDL	Sig.	
Taux_de_nourriture-Modèle_1	15	,930	,845	6,440	5,870	15	,982	0

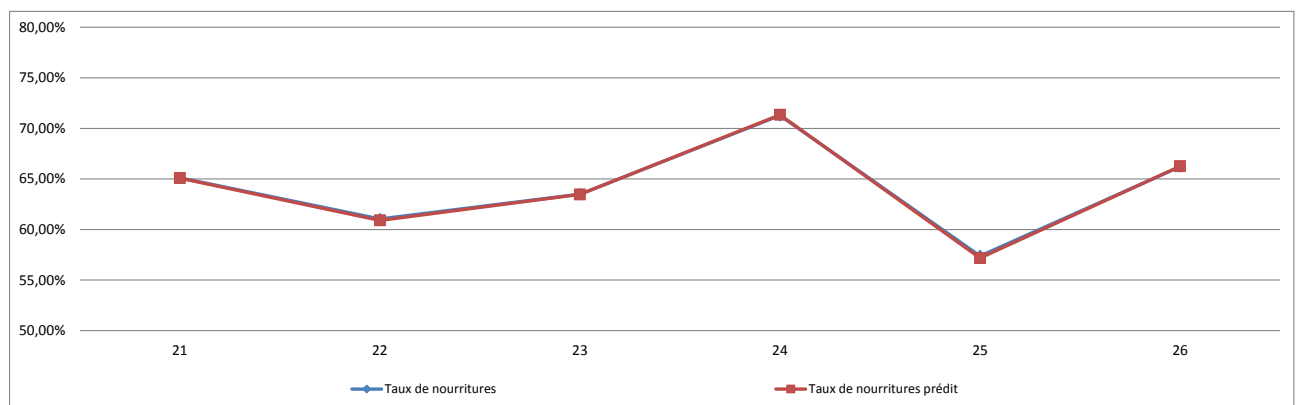
Tableau 17 : Résultat du modèle ARIMA, marque 14 sur 4 semaines (mois)

Un R² de 93%

Mois	Taux de nourritures	Taux de nourritures prédit	Erreur absolue
21	65,11%	65,08%	0,03%
22	61,04%	60,90%	0,14%
23	63,49%	63,47%	0,01%
24	71,28%	71,34%	0,06%
25	57,39%	57,17%	0,21%
26	66,21%	66,26%	0,05%
Erreur absolue moyenne			0,09%

Tableau 18 : Qualité de la prévision sur l'échantillon de validation du modèle ARIMA, marque 14 sur 4 semaines (mois)

L'erreur absolue moyenne est de 0.09% sur les 6 mois utilisés pour le test.



Graphique 8 : Qualité de l'ajustement du modèle ARIMA entre le réel et la prévision du TN, marque 14 sur 4 semaines (mois)

Sur 4 semaines les taux de nourriture entre la période observée et la période prédite sont :

TN moyen réel		TN moyen prédit	
64,09%		64,04%	

Tableau 19 : Qualité de l'ajustement du modèle ARIMA de la marque 14 sur 4 semaines (mois)

2.23.1.2 Les réseaux neurones : le perceptron multicouche :

On utilise 90 semaines pour l'apprentissage et l'estimation du modèle et les 14 dernières semaines pour le test et la validation du modèle.

On utilise les variables ci-dessous :

Champ	Type	Valeurs	Manquantes	Vérifier	Direction
Partition	Intervalle	[1,2]		Aucun	Partition
Num_Semaine	Intervalle	[1,104]		Aucun	In
Taux_de_nourriture	Intervalle	[75.69540613,90.89456129]		Aucun	Out
Taux_de_nourriture (t-1)	Intervalle	[75.69540613,90.89456129]		Aucun	In
Taux_de_nourriture (t-2)	Intervalle	[75.69540613,90.89456129]		Aucun	In
Taux_de_nourriture (t-3)	Intervalle	[75.69540613,90.89456129]		Aucun	In
Prix_CARTE_NOIR	Intervalle	[3.655611285,4.4772]		Aucun	In
Prix_CARTE_NOIR (t-1)	Intervalle	[3.655611285,4.4772]		Aucun	In
Prix_CARTE_NOIR (t-2)	Intervalle	[3.655611285,4.4772]		Aucun	In
Prix_moyen_Tous_cafés	Intervalle	[2.730317583,3.407658155]		Aucun	In
Prix_moyen_Tous_cafés (t-1)	Intervalle	[2.730317583,3.407658155]		Aucun	In
Prix_moyen_Tous_cafés (t-2)	Intervalle	[2.730317583,3.407658155]		Aucun	In
PRIX_Ecart_type_Tous_cafés	Intervalle	[0.915093316,2.067429708]		Aucun	In
PRIX_Ecart_type_Tous_cafés (t-1)	Intervalle	[0.915093316,2.067429708]		Aucun	In
PRIX_Ecart_type_Tous_cafés (t-2)	Intervalle	[0.915093316,2.067429708]		Aucun	In
Nb_de_marques_présentes_par_semaine	Intervalle	[56,84]		Aucun	In
Nb_de_marques_présentes_par_semaine (t-1)	Intervalle	[56,84]		Aucun	In
Nb_de_marques_présentes_par_semaine (t-2)	Intervalle	[56,84]		Aucun	In
Nb_de_marques_AGGLOM	Intervalle	[1,3]		Aucun	In
Nb_de_marques_AGGLOM (t-1)	Intervalle	[1,3]		Aucun	In
Nb_de_marques_AGGLOM (t-2)	Intervalle	[1,3]		Aucun	In
Nb_de_marques_ATOMISE	Intervalle	[4,7]		Aucun	In
Nb_de_marques_ATOMISE (t-1)	Intervalle	[4,7]		Aucun	In
Nb_de_marques_ATOMISE (t-2)	Intervalle	[4,7]		Aucun	In
Nb_de_marques_GRAINS	Intervalle	[5,18]		Aucun	In
Nb_de_marques_GRAINS (t-1)	Intervalle	[5,18]		Aucun	In
Nb_de_marques_GRAINS (t-2)	Intervalle	[5,18]		Aucun	In
Nb_de_marques_LIQUIDE	Intervalle	[1,2]		Aucun	In
Nb_de_marques_LIQUIDE (t-1)	Intervalle	[1,2]		Aucun	In
Nb_de_marques_LIQUIDE (t-2)	Intervalle	[1,2]		Aucun	In
Nb_de_marques_LYOPHIL	Intervalle	[5,12]		Aucun	In
Nb_de_marques_LYOPHIL (t-1)	Intervalle	[5,12]		Aucun	In
Nb_de_marques_LYOPHIL (t-2)	Intervalle	[5,12]		Aucun	In
Nb_de_marques_MOULU	Intervalle	[47,71]		Aucun	In
Nb_de_marques_MOULU (t-1)	Intervalle	[47,71]		Aucun	In
Nb_de_marques_MOULU (t-2)	Intervalle	[47,71]		Aucun	In
Nb_de_marques_NORMAL	Intervalle	[4,8]		Aucun	In
Nb_de_marques_NORMAL (t-1)	Intervalle	[4,8]		Aucun	In
Nb_de_marques_NORMAL (t-2)	Intervalle	[4,8]		Aucun	In
Nb_de_marques_SOLUBLE	Intervalle	[1,2]		Aucun	In
Nb_de_marques_SOLUBLE (t-1)	Intervalle	[1,2]		Aucun	In
Nb_de_marques_SOLUBLE (t-2)	Intervalle	[1,2]		Aucun	In
Nb_de_marques_ARABICA	Intervalle	[41,63]		Aucun	In
Nb_de_marques_ARABICA (t-1)	Intervalle	[41,63]		Aucun	In
Nb_de_marques_ARABICA (t-2)	Intervalle	[41,63]		Aucun	In

Champ	Type	Valeurs	Manquantes	Vérifier	Direction
Nb_de_marques_CAFE_CHIC_LAIT	Intervalle	[2,3]		Aucun	In
Nb_de_marques_CAFE_CHIC_LAIT (t-1)	Intervalle	[2,3]		Aucun	In
Nb_de_marques_CAFE_CHIC_LAIT (t-2)	Intervalle	[2,3]		Aucun	In
Nb_de_marques_CAFE_CHIC_ING	Intervalle	[0,1]		Aucun	In
Nb_de_marques_CAFE_CHIC_ING (t-1)	Intervalle	[0,1]		Aucun	In
Nb_de_marques_CAFE_CHIC_ING (t-2)	Intervalle	[0,1]		Aucun	In
Nb_de_marques_CAFE_CHICOREE	Intervalle	[4,8]		Aucun	In
Nb_de_marques_CAFE_CHICOREE (t-1)	Intervalle	[4,8]		Aucun	In
Nb_de_marques_CAFE_CHICOREE (t-2)	Intervalle	[4,8]		Aucun	In
Nb_de_marques_CAFE_SOLUBLE	Intervalle	[6,13]		Aucun	In
Nb_de_marques_CAFE_SOLUBLE (t-1)	Intervalle	[6,13]		Aucun	In
Nb_de_marques_CAFE_SOLUBLE (t-2)	Intervalle	[6,13]		Aucun	In
Nb_de_marques_CHICOREE_PURE	Intervalle	[1,6]		Aucun	In
Nb_de_marques_CHICOREE_PURE (t-1)	Intervalle	[1,6]		Aucun	In
Nb_de_marques_CHICOREE_PURE (t-2)	Intervalle	[1,6]		Aucun	In
Nb_de_marques_CHICOREE_INGRED	Intervalle	[0,2]		Aucun	In
Nb_de_marques_CHICOREE_INGRED (t-1)	Intervalle	[0,2]		Aucun	In
Nb_de_marques_CHICOREE_INGRED (t-2)	Intervalle	[0,2]		Aucun	In
Nb_de_marques_CHICOREE_LAIT	Intervalle	[0,1]		Aucun	In
Nb_de_marques_CHICOREE_LAIT (t-1)	Intervalle	[0,1]		Aucun	In
Nb_de_marques_CHICOREE_LAIT (t-2)	Intervalle	[0,1]		Aucun	In
Nb_de_marques_MELANGE	Intervalle	[14,26]		Aucun	In
Nb_de_marques_MELANGE (t-1)	Intervalle	[14,26]		Aucun	In
Nb_de_marques_MELANGE (t-2)	Intervalle	[14,26]		Aucun	In
Nb_de_marques_ROBUSTA	Intervalle	[4,13]		Aucun	In
Nb_de_marques_ROBUSTA (t-1)	Intervalle	[4,13]		Aucun	In
Nb_de_marques_ROBUSTA (t-2)	Intervalle	[4,13]		Aucun	In

Figure 5 : Liste des variables en entrée du perceptron

On a calculé pour toutes les variables y compris la cible, les valeurs à l'instant t-1 et t-2, exemples :

Taux_de_nourriture_Marque 1 (t-1) : Le taux de nourriture à l'instant t-1

Taux_de_nourriture_Marque 1 (t-2) : Le taux de nourriture à l'instant t-2

Prix_moyen : le prix moyen des marques à l'instant t

Prix_moyen (t-1) : le prix moyen des marques à l'instant t-1

Prix_moyen (t-2) : le prix moyen des marques à l'instant t-2

.....

On cherche comme dans le modèle précédent à modéliser le taux de nourriture des deux marques, la marque 1 et la marque 14.

La marque 1

A la semaine

On ne présente que le modèle qui a produit le meilleur résultat.

Net. name	Training perf.	Test perf.	Training error	Test error	Training algorithm	Error function	Hidden activation	Output activation
MLP 69-15-1	0.998817	0.768723	0.000051	0.011279	BFGS 36	SOS	Exponential	Logistic

Tableau 20 : Qualité de l'ajustement du Perceptron, marque 1, données hebdomadaires

Le modèle trouvé est un perceptron multicouche : la couche d'entrée contient 69 entrées (les variables d'entrées), la couche cachée contient 15 neurones et la couche de sortie donne le résultat du modèle.

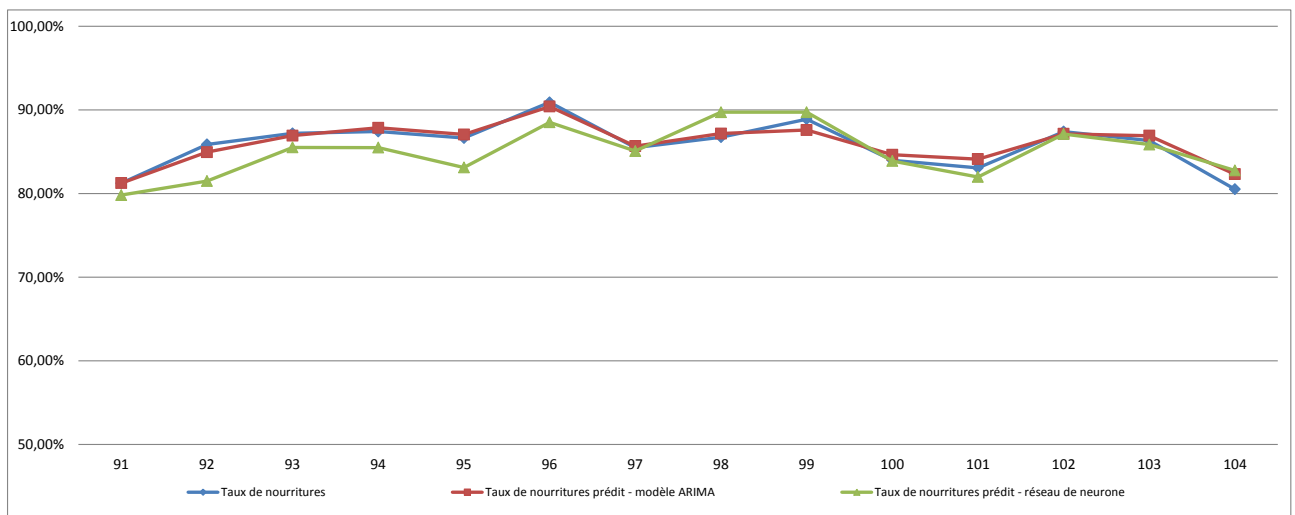
La corrélation entre le taux de nourriture et le taux de nourriture prédit à partir du modèle est d'environ 0.99 pour l'échantillon d'apprentissage du modèle et 0.77 pour l'échantillon test du modèle. Le modèle est robuste et stable et il prédit très bien le taux de nourriture.

Semaines	Taux de nourritures	Taux de nourritures prédit - modèle ARIMA	Taux de nourritures prédit - réseau de neurone	Erreur absolue
91	81,23%	81,24%	79,81%	1,42%
92	85,86%	84,94%	81,48%	4,38%
93	87,19%	86,92%	85,51%	1,68%
94	87,40%	87,85%	85,48%	1,92%
95	86,62%	87,05%	83,11%	3,52%
96	90,90%	90,40%	88,51%	2,39%
97	85,47%	85,66%	85,07%	0,40%
98	86,73%	87,16%	89,72%	2,99%
99	88,88%	87,58%	89,72%	0,84%
100	83,97%	84,63%	83,88%	0,09%
101	83,07%	84,11%	81,98%	1,09%
102	87,40%	87,13%	87,11%	0,29%
103	86,33%	86,90%	85,85%	0,48%
104	80,50%	82,31%	82,75%	2,25%
Erreur absolue moyenne				1,69%

Tableau 21 : Qualité de la prévision du Perceptron sur l'échantillon de validation, marque 1, données hebdomadaires

Ce modèle de réseau de neurone donne en moyenne des prévisions un peu moins bonnes que le modèle ARIMA, l'erreur absolue moyenne est de 1.69% sur les 14 semaines utilisées pour le test contre 0.63% pour le modèle ARIMA.

Les courbes s'ajustent très bien sur les semaines « prévues ».



Graphique 9 : Qualité de l'ajustement du Perceptron entre le réel et la prévision du TN, marque 1, données hebdomadaires

Une très bonne prévision du taux de nourriture hebdomadaire également avec le réseau de neurones.

TN moyen réel	TN moyen prédit
85,99%	85,00%

Tableau 22 : Qualité de l'ajustement du Perceptron, marque 1, données hebdomadaires

Sur 4 semaines (mois)

Net. name	Training perf.	Test perf.	Training error	Test error	Training algorithm	Error function	Hidden activation	Output activation
MLP 37-16-1	0.999993	0.995996	0.000001	0.020958	BFGS 55	SOS	Exponential	Tanh

Nom du modèle, Performance de l'apprentissage, Performance du test, Erreur de l'apprentissage, Algorithme utilisé pour l'apprentissage, Fonction d'erreur, Activation cachée, Sortie de l'activation

Tableau 23 : Qualité de l'ajustement du Perceptron de la marque 14 sur 4 semaines (mois)

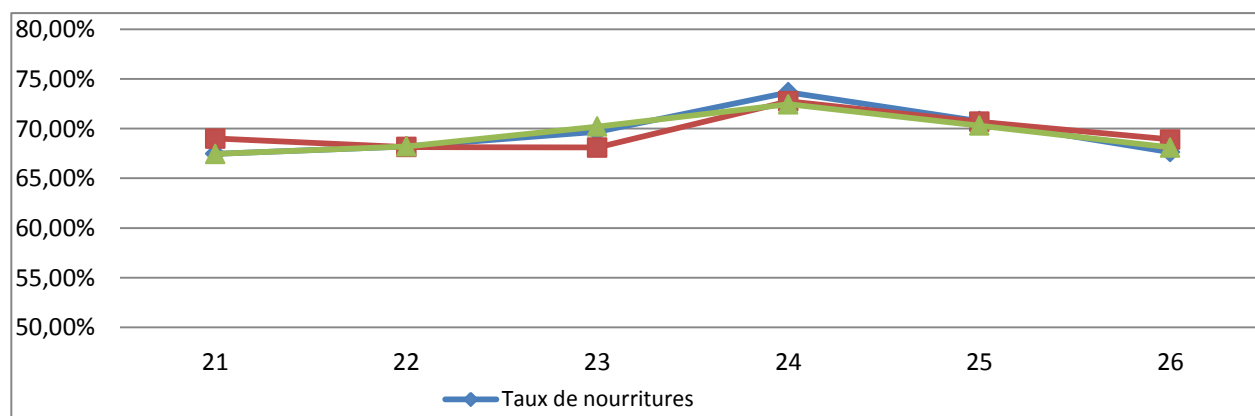
Le modèle trouvé est un perceptron multicouche (MLP), la couche d'entrée contient 37 entrées (les variables d'entrées), la couche cachée contient 16 neurones et la couche de sortie donne le résultat du modèle.

La corrélation entre le taux de nourriture et le taux de nourriture prédit à partir du modèle est d'environ 0.99 pour l'échantillon d'apprentissage du modèle et 0.99 pour l'échantillon test du modèle, le modèle est robuste et stable et il prédit très bien le taux de nourriture.

Mois	Taux de nourritures	Taux de nourritures prédit - modèle ARIMA	Taux de nourritures prédit - réseau de neurone	Erreur absolue
21	67,46%	68,99%	67,46%	0,01%
22	68,20%	68,15%	68,21%	0,00%
23	69,68%	68,09%	70,22%	0,54%
24	73,65%	72,76%	72,45%	1,20%
25	70,75%	70,69%	70,31%	0,44%
26	67,63%	68,90%	68,10%	0,47%
Erreur absolue moyenne				0,45%

Tableau 24 : Qualité de la prévision du Perceptron sur l'échantillon de validation de la marque 14 sur 4 semaines (mois)

Le modèle de réseau de neurone donne en moyenne des prévisions meilleures que le modèle ARIMA, l'erreur absolue moyenne est de 0.45% sur les 6 mois utilisés pour le test contre 0.90% pour le modèle ARIMA.



Graphique 10 : Qualité de l'ajustement du Perceptron entre le réel et la prévision du TN de la marque 14 sur 4 semaines (mois)

Une très bonne prévision du taux de nourriture au mois également avec le réseau de neurone.

TN moyen réel		TN moyen prédit	
	69,56%		69,60%

Tableau 25 : Qualité de l'ajustement du Perceptron au mois de la marque 14 sur 4 semaines (mois)

Les résultats du réseau neurone sur la marque 14 ne sont pas indiqués car ils sont très similaires au modèle ARIMA sur les deux périodes.

2.23.1.3 Conclusion sur les approches agrégées :

Les modèles agrégés prédisent très bien le taux de nourriture.

Les coefficients d'ajustement sont tous proches ou supérieurs à 0,75 tendant pour certains vers 1.

	ARIMA		Perceptron
	Marque 1	Marque 2	Marque 1
Semaine	0,83	0,82	0,77
Mois	0,88	0,93	0,99

Tableau 26 : R² des différents modèles agrégés

<u>Taux moyen</u>					
Semaine					
	TN moyen réel		TN moyen prédit		
		86,71%		85,82%	
Mois					
	TN moyen réel		TN moyen prédit		
		67,74%		67,75%	

Le tableau ci-dessus présente les moyennes des taux réels et prédit des deux modèles ARIMA et Perceptron.

Tableau 27 : Taux de nourriture réel moyen des différents modèles agrégés entre le réel et le prédit

Le taux de nourriture est évidemment inférieur sur 4 semaines (mois) qu'à la semaine mais il se prédit très bien dans les deux cas par les modèles agrégés. Son écart est donc prévisible et contraint.

2.23.2 Les approches désagrégées :

Dans cette partie nous repartons sur une désagrégation totale, un individu, une semaine, un achat.

Nous testons plusieurs approches de modélisation désagrégée :

- Une analyse de panier de la ménagère par deux méthodes Apriori et Séquence.
- Un modèle de panel avec variable temporelle

2.23.1 La cooccurrence d'achat :

On étudie d'abord la marque 1 puis la marque 14. Ici l'analyse ne peut se faire qu'à l'acte, donc la prise en compte de périodes d'achat différentes (semaine, mois, année) n'est pas pertinente.

2.23.1.1 Résultats de l'algorithme "A priori"

La marque 1 :

On ne sélectionne que les règles qui ont comme conséquence un achat du café "Marque 1".

Antécédent	Conséquence	% de prise en charge	% de confiance
MARQUE 2 = T and MAISON DU CAFE = T	MARQUE 1 = T	7,0%	83,7%
MARQUE 2 = T and MAISON DU CAFE = T and MDD AUTRES = T	MARQUE 1 = T	5,8%	83,0%
MARQUE 2 = T	MARQUE 1 = T	10,4%	79,4%
MARQUE 2 = T and MDD AUTRES = T	MARQUE 1 = T	8,1%	78,5%
MARQUE 3 = T and JACQUES VABRE = T	MARQUE 1 = T	5,8%	67,0%
MARQUE 3 = T and GRAND MERE = T	MARQUE 1 = T	5,4%	66,9%
MARQUE 4 = T and MDD AUTRES = T	MARQUE 1 = T	5,2%	66,7%
MARQUE 3 = T and JACQUES VABRE = T and MDD AUTRES = T	MARQUE 1 = T	5,1%	66,6%
MARQUE (= T and NESCAFE = T and MAISON DU CAFE = T	MARQUE 1 = T	6,4%	66,2%
MARQUE 4 = T	MARQUE 1 = T	6,7%	65,3%
MARQUE1 = T and NESCAFE = T and MAISON DU CAFE = T and MDD AUTRES = T	MARQUE 1 = T	5,5%	65,0%
MARQUE 3 = T and MAISON DU CAFE = T and MDD AUTRES = T	MARQUE 1 = T	9,2%	64,1%
MARQUE 3 = T and MAISON DU CAFE = T	MARQUE 1 = T	10,9%	64,0%
MARQUE 6 = T and JACQUES VABRE = T and MAISON DU CAFE = T and MDD AUTRES = T	MARQUE 1 = T	5,2%	63,6%

Tableau 28 : Cooccurrences des autres marques avec la marque 1 comme conséquence

Explications :

« L'antécédent » représente ce que l'on a acheté précédemment. Par exemple quand on achète les cafés "MARQUE 1" et "MAISON DU CAFE" alors 7% des clients achètent à nouveau le café "MARQUE 1" avec une probabilité de 83.7%. Le seuil de confiance du réel est très élevé.

La conséquence, c'est donc ce que l'on va acheter lors de l'achat suivant.

En rouge, à noter, les règles les plus significatives (pourcentages de confiance supérieurs à 70 %).

On trouve dans les règles les plus significatives, un rachat le plus probable à la marque 1 dont une grégarité du comportement.

La marque 2 :

Avec une plus faible part de marché en pénétration la marque ne permet pas d'avoir des règles significatives. Nous ne pouvons pas appliquer Apriori.

2.23.1.2 Résultats de l'algorithme "Séquence"

La marque 1 :

On ne sélectionne que les règles qui nous donnent un achat du café à la marque 1.

L'algorithme Séquence nous donne beaucoup plus de règles significatives que l'algorithme "Apriori". Quand on achète le café "MARQUE 1" 9 fois (acte d'achat) alors 10.5% des clients achètent à nouveau le café "MARQUE 1" avec une probabilité de 91%.

Antécédent	Conséquence	% de prise en charge	% de confiance
MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1	MARQUE 1	10,5%	91,0%
MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1	MARQUE 1	13,1%	89,7%
MARQUE 1 and MARQUE 1 and MARQUE 1 and MDD AUTRES and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1	MARQUE 1	5,8%	89,2%
MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1	MARQUE 1	11,8%	89,2%
MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1	MARQUE 1	14,8%	88,6%
MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MDD AUTRES and MARQUE 1 and MARQUE 1	MARQUE 1	5,7%	87,5%
MARQUE 1 and MDD AUTRES and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1	MARQUE 1	6,0%	87,1%
MARQUE 1 and MARQUE 1 and MARQUE 1 and MARQUE 1 and MDD AUTRES and MARQUE 1 and MARQUE 1	MARQUE 1	6,7%	87,0%

Tableau 29 : Les plus fortes cooccurrences avec la marque 1 comme conséquence

Vue partielle se limitant aux règles ayant plus de 87% de confiance.
L'acheteur de la Marque 1 a une forte probabilité de racheter la Marque 1 dans ces n ièmes achats.

La marque 14 :

On ne sélectionne que les règles qui nous donnent un achat du café à la marque 14

Quand on achète le café "Marque 14" 4 fois (acte d'achat) alors 1.9% des clients achètent à nouveau le café "Marque 14" avec une probabilité de 76.4%.

En rouge, à noter, les règles les plus significatives (pourcentages de confiance supérieurs à 70 %)

Antécédent	Conséquence	% de prise en charge	% de confiance
Marque 14 > Marque 14 > Marque 14 > Marque 14	Marque 14	1,9%	76,4%
Marque 14 > NESCAFE > Marque 14 > Marque 14	Marque 14	0,7%	75,7%
CARTE NOIRE > Marque 14 > Marque 14 > Marque 14	Marque 14	1,7%	71,8%
Marque 14 > Marque 14 > Marque 14	Marque 14	2,9%	67,9%
Marque 14 > CARTE NOIRE > Marque 14 > Marque 14	Marque 14	1,7%	67,5%
Marque 14 > Marque 14 > CARTE NOIRE > Marque 14	Marque 14	1,8%	67,2%
NESCAFE > Marque 14 > Marque 14 > Marque 14	Marque 14	0,8%	66,7%
Marque 14 > Marque 14 > MDD AUTRES > Marque 14	Marque 14	1,4%	66,5%
MDD AUTRES > Marque 14 > Marque 14 > Marque 14	Marque 14	1,5%	65,7%
NESCAFE > Marque 14 > Marque 14	Marque 14	1,2%	65,1%

Tableau 30 : Les plus fortes cooccurrences avec la marque 14 comme conséquence

L'acheteur de la Marque 14 a une forte probabilité de racheter la Marque 14 dans ses n ièmes achats avec une marque qui a pourtant une faible pénétration et une faible PDM.

On peut donc prédire par les modèles de cooccurrence d'achat les n ième achats d'une même marque à des niveaux de confiance très élevés, supérieurs à 70% pour les règles les plus occurrentes qui montrent que les consommateurs achètent principalement la même marque à la fois pour la marque 1 et la marque 14.

Cette approche valide l'idée d'entropie du rachat et donc la stabilité du taux de nourriture.

2.23.2 La technique d'analyse de panel :

A nouveau systématiquement :

- sur la partie temporelle nous travaillons à la semaine et au mois, nous ne retenons pas l'année sur la donnée individuelle car le modèle n'aurait pas de sens sur 24 occurrences.
- au niveau des marques nous testons deux marques, la marque 1 leader du marché et une plus petite marque, la marque 15.

La marque 1

A la semaine :

Le fichier : 11 148 Panélistes ayant acheté au moins une fois la Marque 1 sur 2 ans (104 semaines).

Dispersion du nombre de semaines :

Moyenne	50.395
Min	1
Max	104
Ecart-type	29.836
Médiane	50
Mode	4

Tableau 31 : Les indicateurs clefs en actes d'achats de la marque 1 à la semaine

On sélectionne les semaines où le panéliste a acheté la Marque 1. La base est alors de 5 051 panélistes.

Variable à expliquer :

Taux de Nourriture de la Marque 1 (QA Marque 1 / QA totale cafés) : voici ses caractéristiques en base individus sur 95 semaines maximum :

Effectif	5 051
Moyenne	0.854
Min	0.032
Max	1
Ecart-type	0.212
Médiane	0.975
Mode	1

Tableau 32 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 1 à la semaine

On constate qu'il y a une majorité de mono acheteurs.

Voici la distribution en base Panélistes * semaines d'achat :

Effectif	34 765
Moyenne	0.883
Min	0.016
Max	1
Ecart-type	0.24
Médiane	1
Mode	1

Tableau 33 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 1 à la semaine

Elle montre les mêmes tendances qu'en base individus.

Variables explicatives :

Champ	Graphique échan...	Mesure	Min	Max	Moyenne	Ecart-type	Asymétrie	Unique	Valide
⚙ Prix_moyen_Tous_Cafés		Continu	2.730	3.408	2.956	0.124	0.910	--	34765
⚙ Prix_EcartType_Tous_Cafés		Continu	0.915	2.067	1.132	0.171	2.001	--	34765
⚙ Nb_Marques_Achetées		Continu	1	7	1.263	0.521	2.095	--	34765
⚙ Nb_Marques_Présentes		Continu	56	84	68.482	5.312	0.394	--	34765
⚙ CA par semaine_CARTE NOIRE		Continu	1.300	79.800	6.123	4.271	3.110	--	34765
⚙ CA par semaine_Tous cafés		Continu	1.320	79.800	7.120	4.844	2.898	--	34765
⚙ Age du chef de famille		Continu	30	70	52.890	12.043	0.055	--	34765
⚙ Nombre de personnes du foyer		Continu	1	5	2.560	1.135	0.623	--	34765
⚙ Colocataires de 25 ans et plus		Continu	0	1	0.035	0.185	5.026	--	34765
⚙ Couple chef de fam 65 et + ss enfant		Continu	0	1	0.155	0.362	1.909	--	34765
⚙ Fam maternelle enfant 0 à 5 ans		Continu	0	1	0.054	0.227	3.930	--	34765
⚙ Fam secondaire enfant 12 à 17 ans		Continu	0	1	0.089	0.285	2.887	--	34765
⚙ Famille tertiaire enfant 18 à 24 ans		Continu	0	1	0.135	0.342	2.139	--	34765
⚙ Monofoyer de 35 à 50 ans		Continu	0	1	0.028	0.165	5.737	--	34765
⚙ Monofoyer de 50 à 64 ans		Continu	0	1	0.050	0.218	4.127	--	34765
⚙ Monofoyer de 65 ans et plus		Continu	0	1	0.061	0.239	3.672	--	34765
⚙ Unité urbaine de moins de 20000		Continu	0	1	0.170	0.375	1.760	--	34765
⚙ Niveau de vie_Aisé		Continu	0	1	0.186	0.389	1.618	--	34765
⚙ Niveau de vie_Modeste		Continu	0	1	0.097	0.296	2.723	--	34765
⚙ Niveau de vie_Moyen inférieur		Continu	0	1	0.277	0.448	0.996	--	34765
⚙ Niveau de vie_Moyen supérieur		Continu	0	1	0.316	0.465	0.790	--	34765
⚙ Présence denfants_Aucun enfant		Continu	0	1	0.576	0.494	-0.310	--	34765
⚙ bébé < 30 mois		Continu	0	1	0.054	0.226	3.952	--	34765
⚙ Enfants_de 0 à 17 ans		Continu	0	1	0.209	0.407	1.430	--	34765

Figure 6 : Liste des variables en entrée du modèle désagrégé

Modélisation :

Tous les modèles testés ont une performance similaire, nous retenons le meilleur modèle :

Random-effects GLS regression	Number of obs	=	34765
Group variable: id_pan	Number of groups	=	5051
R-sq: within = 0.7683	Obs per group: min	=	1
between = 0.8249	avg	=	6.9
overall = 0.7966	max	=	95
	Wald chi2(25)	=	122304.77
corr(u_i, X) = 0 (assumed)	Prob > chi2	=	0.0000
----- theta -----			
min	5%	median	95% max
0.1541	0.1541	0.3246	0.7080 0.8394

tn_ Marque 1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
prix_ Marque 1	-.0050078	.0031343	-1.60	0.110	-.011151	.0011354
prix_moyen_tous_cafes	.0117808	.0069172	1.70	0.089	-.0017767	.0253382
prix_ecarttype_tous_cafes	-.0030253	.0041922	-0.72	0.471	-.0112419	.0051912
nb_marques_achetees	-.3015761	.0018928	-159.33	0.000	-.3052859	-.2978663
nb_marques presentes	.0000682	.0001244	0.55	0.583	-.0001756	.0003121
ca_par_semaine Marque 1	.0293056	.0004328	67.71	0.000	.0284574	.0301539
ca_par_semaine_tous_cafes	-.0243167	.0004009	-60.65	0.000	-.0251026	-.0235309
age_du_chef_de_famille	.0001677	.0001871	0.90	0.370	-.000199	.0005344
nombre_de_personnes_du_foyer	.0044107	.0024709	1.79	0.074	-.0004321	.0092535
colocataires_de_25_ans_et_plus	-.0002002	.0080145	-0.02	0.980	-.0159083	.0155079
couple_chef_de_fam_65_et_ss_enf	.0069891	.0054111	1.29	0.196	-.0036165	.0175946
fam_maternelle_enfant_0_5_ans	.0161335	.0072916	2.21	0.027	.0018423	.0304248
fam_secondaire_enfant_12_17_an	.0060094	.0052011	1.16	0.248	-.0041846	.0162034
famille_tertiaire_enfant_18_24_	-.0012545	.0049014	-0.26	0.798	-.0108611	.0083521
monofoyer_de_35_50_ans	.0249164	.0076244	3.27	0.001	.0099728	.03986
monofoyer_de_50_64_ans	.0161035	.006228	2.59	0.010	.0038968	.0283102
monofoyer_de_65_ans_et_plus	.0167353	.0067978	2.46	0.014	.0034119	.0300587
unite_urbaine_de_moins_de_20000	.0020944	.0031487	0.67	0.506	-.004077	.0082657
niveau_de_vie_aise	-.0182784	.009298	-1.97	0.049	-.0365021	-.0000547
niveau_de_vie_modeste	-.025766	.0098299	-2.62	0.009	-.0450322	-.0064998
niveau_de_vie_moyen_inferieur	-.0169565	.0092804	-1.83	0.068	-.0351457	.0012327
niveau_de_vie_moyen_superieur	-.0144954	.0091087	-1.59	0.112	-.0323482	.0033574
presence_denfants_aucun_enfant	.0130194	.0088137	1.48	0.140	-.004255	.0302939
bebe_30_mois	-.0014872	.0081977	-0.18	0.856	-.0175544	.0145799
enfants_de_0_17_ans	.0058617	.0073709	0.80	0.426	-.008585	.0203084
_cons	1.220668	.0267949	45.56	0.000	1.168151	1.273185
sigma_u	.06001425					
sigma_e	.09519733					
rho	.28439997	(fraction of variance due to u_i)				

GLS : General list square ; within : intra ; between : inter ; overall : au global ; corr : correlation ; coef : coefficient ; std err : erreur standardisée (attachée à chaque coeff il faut que l'ordre de grandeur soit comparable) ; z : significativité du coefficient ; P> [z] : probabilité que le coeff soit dû au hasard ; [95% Conf Interval] : Intervalle de confiance du coeff (ne doit pas contenir de 0) ; cons : constante

Tableau 34 : Les résultats du modèle désagrégé de la marque 1 à la semaine

```

Between regression (regression on group means) Number of obs = 34765
Group variable: id_pan Number of groups = 5051

R-sq: within = 0.7667 Obs per group: min = 1
      between = 0.8259 avg = 6.9
      overall = 0.7953 max = 95

F(25,5025) = 953.34
sd(u_i + avg(e_i.)) = .0885596 Prob > F = 0.0000

```

tn_ Maroue 1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
prix_ Maroue 1	.014578	.0104927	1.39	0.165	-.0059922	.0351483
prix_moyen_tous_cafes	.0386663	.0236021	1.64	0.101	-.007604	.0849367
prix_ecarttype_tous_cafes	-.0167006	.0146096	-1.14	0.253	-.0453418	.0119406
nb_marques_achetees	-.3062848	.0050767	-60.33	0.000	-.3162374	-.2963322
nb_marques_presentes	.0008709	.0004237	2.06	0.040	.0000403	.0017016
ca_par_semaine_ Maroue 1	.0329303	.0011791	27.93	0.000	.0306188	.0352419
ca_par_semaine_tous_cafes	-.0262406	.0010659	-24.62	0.000	-.0283303	-.0241509
age_du_chef_de_famille	.0001797	.0002064	0.87	0.384	-.000225	.0005843
nombre_de_personnes_du_foyer	.0056389	.0027371	2.06	0.039	.0002729	.0110049
colocataires_de_25_ans_et_plus	.0002197	.0089998	0.02	0.981	-.0174239	.0178633
couple_chef_de_fam_65_et_ss_enf	.0103606	.0060181	1.72	0.085	-.0014374	.0221587
fam_maternelle_enfant_0_5_ans	.0171055	.0080924	2.11	0.035	.0012408	.0329702
fam_secondaire_enfant_12_17_an	.0045552	.0058164	0.78	0.434	-.0068475	.0159579
famille_tertiaire_enfant_18_24	.0005493	.0055192	0.10	0.921	-.0102708	.0113694
monofoyer_de_35_50_ans	.0331311	.0082276	4.03	0.000	.0170015	.0492608
monofoyer_de_50_64_ans	.0206331	.006828	3.02	0.003	.0072472	.0340189
monofoyer_de_65_ans_et_plus	.0205611	.0075315	2.73	0.006	.0057961	.0353261
unite_urbaine_de_moins_de_20000	.0010024	.0035092	0.29	0.775	-.0058771	.0078819
niveau_de_vie_aise	-.0210837	.0104507	-2.02	0.044	-.0415716	-.0005959
niveau_de_vie_modeste	-.0264222	.0110587	-2.39	0.017	-.0481022	-.0047423
niveau_de_vie_moyen_inferieur	-.0173833	.0104283	-1.67	0.096	-.0378273	.0030607
niveau_de_vie_moyen_superieur	-.0150853	.0102376	-1.47	0.141	-.0351554	.0049849
presence_denfants_aucun_enfant	.0112944	.0099694	1.13	0.257	-.00825	.0308387
bebe_30_mois	-.0015861	.0091683	-0.17	0.863	-.01956	.0163878
enfants_de_0_17_ans	.0052561	.0083526	0.63	0.529	-.0111186	.0216309
_cons	1.018127	.085911	11.85	0.000	.8497038	1.18655

Tableau 35 : Résultats du modèle désagrégé pour la marque 1 à la semaine

Le nombre de marques (de café) achetées ainsi que le CA dépensé, pour la marque 1 et, au global cafés ont une importance prépondérante dans la qualité du modèle (82% de variance expliquée).

Ces deux variables peuvent à priori être obtenues sur le consommateur, via une BDD et un panel consommateurs, et nous l'avons démontré, elles sont assez stables dans le temps.

Ce modèle n'est exploitable que si l'on dispose de ces variables. Sans elles, son pouvoir explicatif est négligeable, ou faut-il alors disposer d'autres descripteurs de l'acte d'achat.

Enfin, la socio-démographie joue un rôle négligeable : on le prouve en ne s'intéressant qu'aux effets "within", c'est-à-dire aux variables qui se répartissent entre les panélistes.


```

Fixed-effects (within) regression      Number of obs   =   34765
Group variable: id_pan                Number of groups =   5051

R-sq:  within = 0.7683                Obs per group: min =    1
      between = 0.8234                    avg =    6.9
      overall = 0.7960                    max =    95

                                         F(7,29707)     =  14071.67
corr(u_i, Xb) = 0.0772                 Prob > F       =   0.0000

```

tn_ Marque 1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
prix_ Marque 1	-.0065201	.0032748	-1.99	0.046	-.0129388	-.0001015
prix_moyen_tous_cafes	.010338	.0072097	1.43	0.152	-.0037933	.0244693
prix_ecarttype_tous_cafes	-.002205	.0043606	-0.51	0.613	-.010752	.0063421
nb_marques_achetees	-.2987587	.0020386	-146.55	0.000	-.3027545	-.2947629
nb_marques_presentes	.0000302	.0001297	0.23	0.816	-.0002241	.0002845
ca_par_semaine_ <small>Marque 1</small>	.0290009	.0004651	62.36	0.000	.0280893	.0299125
ca_par_semaine_tous_cafes	-.0242918	.0004319	-56.24	0.000	-.0251384	-.0234452
age_du_chef_de_famille	0	(omitted)				
nombre_de_personnes_du_foyer	0	(omitted)				
colocataires_de_25_ans_et_plus	0	(omitted)				
couple_chef_de_fam_65_et_ss_enf	0	(omitted)				
fam_maternelle_enfant_0_5_ans	0	(omitted)				
fam_secondaire_enfant_12_17_an	0	(omitted)				
famille_tertiaire_enfant_18_24_	0	(omitted)				
monofoyer_de_35_50_ans	0	(omitted)				
monofoyer_de_50_64_ans	0	(omitted)				
monofoyer_de_65_ans_et_plus	0	(omitted)				
unite_urbaine_de_moins_de_20000	0	(omitted)				
niveau_de_vie_aise	0	(omitted)				
niveau_de_vie_modeste	0	(omitted)				
niveau_de_vie_moyen_inferieur	0	(omitted)				
niveau_de_vie_moyen_superieur	0	(omitted)				
presence_denfants_aucun_enfant	0	(omitted)				
bebe_30_mois	0	(omitted)				
enfants_de_0_17_ans	0	(omitted)				
_cons	1.251699	.0257516	48.61	0.000	1.201225	1.302173
sigma_u	.0894066					
sigma_e	.09519733					
rho	.46866245	(fraction of variance due to u_i)				

F test that all u_i=0: F(5050, 29707) = 3.02 Prob > F = 0.0000

Tableau 36 : Résultats du modèle que sur les effets intra pour la marque 1 à la semaine

Le fait que les variables socio démographiques jouent peu dans la prévision (et non dans l'interprétation) de l'inélasticité relative du taux de nourriture est pour la généralisation de la méthode un point essentiel. L'élasticité devant être le plus possible prédite sur les variables endogènes au produit.

Voici les adéquations du modèle sur les ensembles d'apprentissage, de test et au global :

Apprentissage		
Coefficients de détermination (R ²) :	TN Marque 1	TN Marque 1 prédit
TN Marque 1	1	0.794
TN Marque 1 prédit	0.794	1

Test		
Coefficients de détermination (R ²) :	TN Marque 1	TN Marque 1 prédit
TN Marque 1	1	0.803
TN Marque 1 prédit	0.803	1

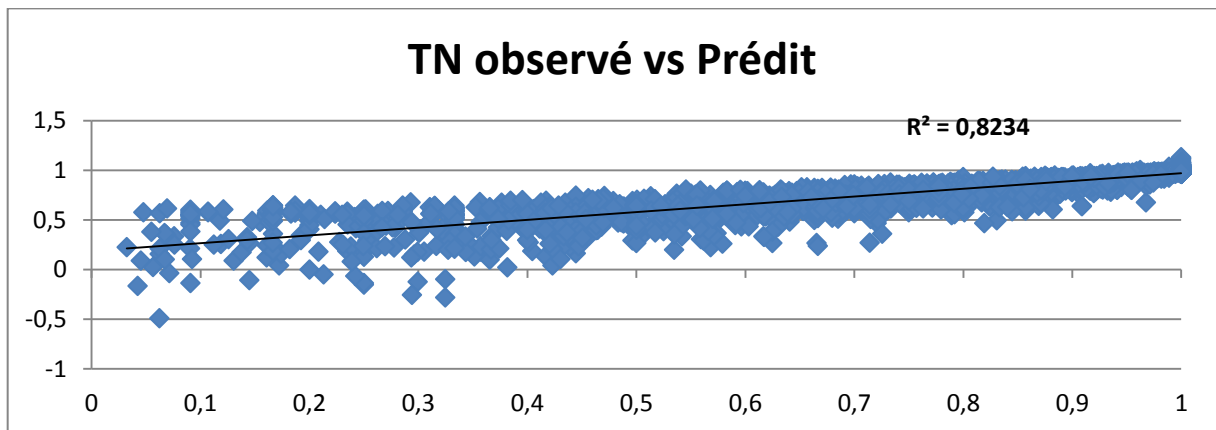
TOTAL		
Coefficients de détermination (R ²) :	TN Marque 1	TN Marque 1 prédit
TN Marque 1	1	0.796
TN Marque 1 prédit	0.796	1

Statistiques descriptives :	Observations	Minimum	Maximum	Moyenne	Ecart-type
TN Marque 1	34 765	0.016	1.000	0.883	0.240
TN Marque 1 prédit	34 765	-1.323	1.330	0.883	0.209

Tableau 37 : Validation des résultats entre le prédit et le réel pour de la marque 1 à la semaine

On retrouve une parfaite cohérence entre le réel et le prédit, l'écart type est compris entre 0,20 et 0,24.

Le graphique ci-dessous montre une correspondance entre le taux de nourriture réel (en abscisse) et le taux de nourriture prédit (en ordonnée).



Graphique 11 : Qualité de l'ajustement de la marque 1 à la semaine

On voit que l'estimation en est d'autant meilleure que l'on se rapproche de la moyenne de Y (à savoir 88%), donc les petits taux de nourriture sont logiquement un peu moins bien prédits.

TN moyen réel		TN moyen prédit	
	88,30%		88,30%

Tableau 38 : Qualité de l'ajustement pour de la marque 1 à la semaine

La prévision en taux moyen est « parfaite » même si le R² est de 82%.

Sur 4 semaines (mois) :

On agrège les semaines en mois. Les variables explicatives sont recalculées sur cette base.

```

Fixed-effects (within) regression           Number of obs   =   26192
Group variable: id_pan                     Number of groups =   5051

R-sq:  within = 0.5918                    Obs per group:  min =    1
        between = 0.7427                    avg   =    5.2
        overall = 0.6419                    max   =    26

                                           F(7,21134)     =   4377.99
corr(u_i, Xb) = 0.1198                    Prob > F       =    0.0000
    
```

Variable	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Marque 1 re	.0443568	.0049509	8.96	0.000	.0346527	.0540609
p1: Marque 1 re	.1109078	.0081757	13.57	0.000	.0948829	.1269327
prix_moyen_touscafs	-.0306618	.0145388	-2.11	0.035	-.0591591	-.0021646
prix_ecarttype_touscafs	-.2669644	.0016638	-160.46	0.000	-.2702256	-.2637033
nb_marquesachetes	.0046037	.0000354	130.09	0.000	.0045344	.0046731
nb_marquesprntes	-.0829741	.0037303	-22.24	0.000	-.0902857	-.0756625
qakgparmois_touscafs	.0071848	.0003493	20.57	0.000	.0065001	.0078696
caparmois_touscafs	.3989793	.0240893	16.56	0.000	.3517624	.4461961
_cons						
sigma_u	.11833413					
sigma_e	.1203566					
rho	.49152743	(fraction of variance due to u_i)				

F test that all u_i=0: F(5050, 21134) = 2.57 Prob > F = 0.0000

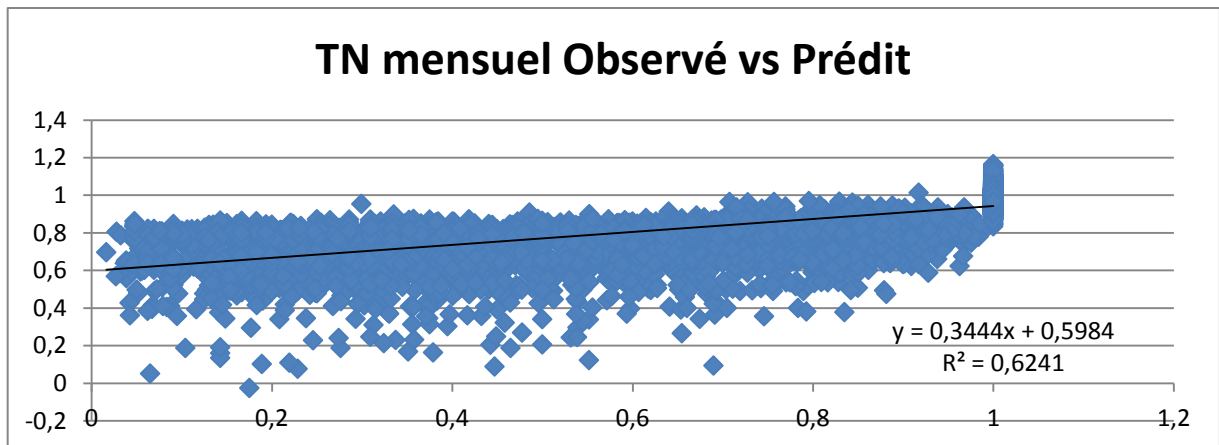
Tableau 39 : Résultats du modèle pour la marque 1 sur 4 semaines (mois)

Au mois, la qualité du modèle diminue : R² de 0,64. Perte de 8% de qualité d'ajustement.

Au niveau mensuel, l'élasticité du prix de la Marque 1 est supérieure à 0 ; même si l'effet est minime, le coefficient est significatif et la valeur nulle n'est pas comprise dans l'intervalle de confiance.

Statistiques descriptives :	Minimum	Maximum	Médiane	Moyenne	Ecart-type (n)
TN_Marque 1	0.016	1.000	1.000	0.878	0.234
TN_Marque 1 prédit	-0.025	1.165	0.937	0.901	0.102

Tableau 40 : Validation des résultats entre le prédit et le réel de la marque 1 sur 4 semaines (mois)



Graphique 12 : Qualité de l'ajustement du modèle désagrégé pour la marque 1 sur 4 semaines (mois)

Le taux de nourriture est élevé mais néanmoins assez bien prédit.

TN moyen réel		TN moyen prédit	
	87,80%		90,10%

Tableau 41 : Qualité de l'ajustement du modèle désagrégé pour la marque 1 sur 4 semaines (mois)

Les écarts types entre l'élasticité hebdomadaire et mensuelle sont similaires même si les taux de nourriture sont différents (cf. page 29 et 30). Ceci explique la qualité de la prévision.

La marque 14 :

Le fichier : 1 164 Panélistes ayant acheté au moins une fois la marque 14 sur 2 ans (104 semaines).

La dispersion du nombre de semaines :

Moyenne	47.365
Min	1
Max	104
Ecart-type	29.574
Médiane	47
Mode	13

Tableau 42 : Les indicateurs clés en actes d'achats de la marque 14 à la semaine

Variable à expliquer :

Taux de Nourriture de la marque 14 (QA Marque 14 / QA total cafés) : voici ses caractéristiques en base individus sur 95 semaines maximum :

Effectif	1 164
Moyenne	0.874
Min	0.099
Max	1
Ecart-type	0.203
Médiane	1

Mode	1
------	---

Tableau 43 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 14 à la semaine

On constate qu'il y a une majorité de mono acheteurs.

Voici la distribution en base Panélistes * semaines d'achat :

Effectif	3 938
Moyenne	0.902
Min	0.077
Max	1
Ecart-type	0.196
Médiane	1
Mode	1

Tableau 44 : Les indicateurs de dispersion du taux de nourriture sur les panélistes de la marque 14 à la semaine

Elle montre les mêmes tendances qu'en base individus.

A la semaine :

```

Fixed-effects (within) regression          Number of obs   =   3938
Group variable: id_pan                    Number of groups =   1164

R-sq:  within = 0.7465                    Obs per group:  min =    1
        between = 0.7544                  avg =    3.4
        overall = 0.7719                  max =    82

                                           F(7,2767)      =   1163.86
corr(u_i, Xb) = 0.1338                    Prob > F       =    0.0000

```

tn_ Marque 14	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
prix_ Marque 14	-.0019564	.0083171	-0.24	0.814	-.0182647	.0143519
prix_moyen_tous_cafs	-.0002662	.019373	-0.01	0.989	-.0382531	.0377207
prix_ecarttype_tous_cafs	-.0017241	.0119239	-0.14	0.885	-.0251046	.0216565
nb_marques_achetes	-.3091009	.0037987	-81.37	0.000	-.3165494	-.3016524
nb_marques_prsentes	-.0002763	.0003572	-0.77	0.439	-.0009766	.000424
qakgparsemaine_touscafs	-.1641165	.0113638	-14.44	0.000	-.186399	-.1418341
caparsemaine_touscafs	.0163208	.0012387	13.18	0.000	.0138919	.0187497
_cons	1.333394	.073651	18.10	0.000	1.188978	1.477811
sigma_u	.10184903					
sigma_e	.07925013					
rho	.62287403	(fraction of variance due to u_i)				

F test that all u_i=0: F(1163, 2767) = 2.23 Prob > F = 0.0000

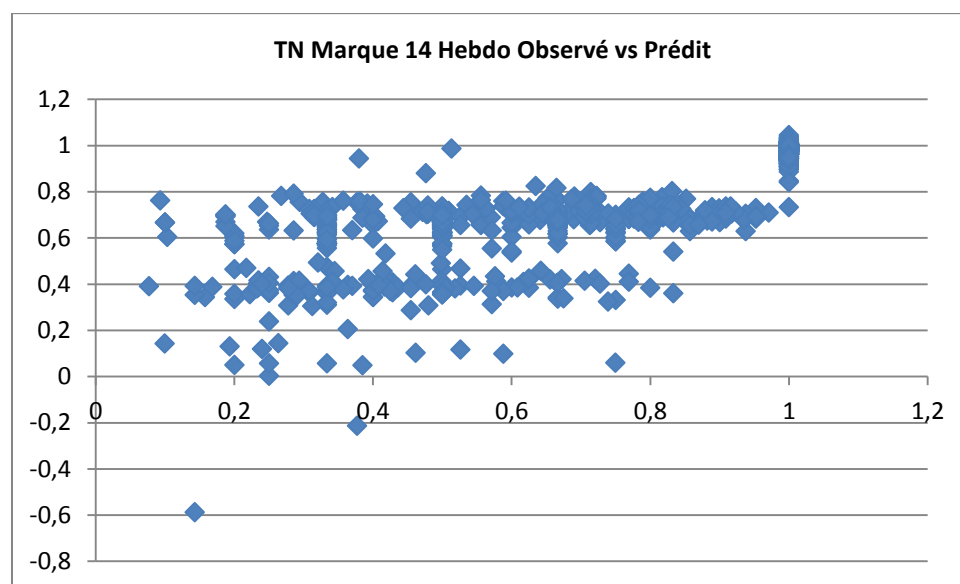
Tableau 45 : Résultats du modèle désagrégé de la marque 14 à la semaine

Le modèle est significatif globalement, mais la plupart des coefficients sont associés à des I.C. contenant la valeur 0. La constante et le nombre de marques achetées sont les deux seuls déterminants significatifs du taux de nourriture de la marque 14 hebdomadaire.

Résultat agrégé :

Echantillon	Minimum	Maximum	Médiane	Moyenne	Ecart-type (n)
TN_Marque 14	0.077	1.000	1.000	0.902	0.196
TN_Marque 14_Prédit	-0.588	1.044	0.984	0.902	0.163

Tableau 46 : Validation des résultats entre le prédit et le réel du modèle désagrégé de la marque 14 à la semaine



Graphique 13 : Qualité de l'ajustement du modèle désagrégé de la marque 14 à la semaine

Sur 4 semaines (mois) :

```

Fixed-effects (within) regression           Number of obs   =   3126
Group variable: id_pan                     Number of groups =   1164

R-sq:  within = 0.6196                    Obs per group:  min =    1
        between = 0.7183                    avg   =    2.7
        overall = 0.6767                    max   =    26

corr(u_i, Xb) = 0.2229                    F(7,1955)      =   454.97
                                                Prob > F       =    0.0000

```

t	Marque 14	ir	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	pr	Marque 14 ir	.0406266	.012398	3.28	0.001	.0163119 .0649412
	prix_moyen_touscafs		.1047559	.0208434	5.03	0.000	.0638783 .1456335
	prix_ecarttype_touscafs		-.0206429	.0397623	-0.52	0.604	-.0986239 .057338
	nb_marquesachetes		-.2376762	.0046513	-51.10	0.000	-.2467981 -.2285543
	nb_marquesprntes		.0041089	.0000959	42.84	0.000	.0039208 .0042971
	qakqparmois_touscafs		-.1545847	.0137492	-11.24	0.000	-.1815494 -.12762
	caparmois_touscafs		.015721	.0014924	10.53	0.000	.0127941 .0186479
	_cons		.4531405	.0696725	6.50	0.000	.3165003 .5897807
	sigma_u		.1170225				
	sigma_e		.09690083				
	rho		.59323533				(fraction of variance due to u_i)

F test that all u_i=0: F(1163, 1955) = 1.92 Prob > F = 0.0000

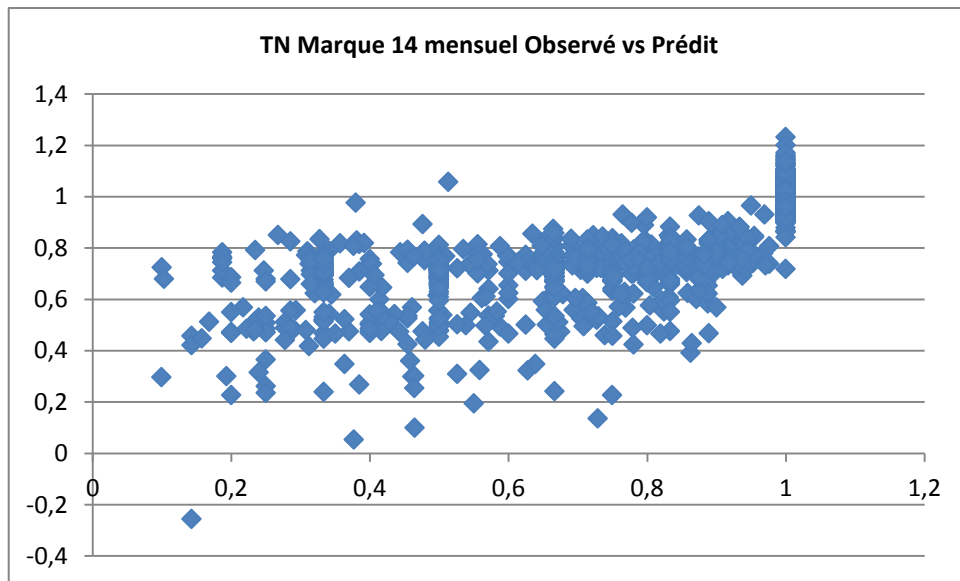
Tableau 47 : Résultats du modèle désagrégé de la marque 14 sur 4 semaines (mois)

Le Fit global est un peu moins bon qu'à la semaine, mais le modèle est mieux déterminé. Seul l'écart-type des prix (qui a peu d'influence) voit son coefficient compris entre les bornes de l'IC.

Résultat agrégé :

Echantillon	Minimum	Maximum	Médiane	Moyenne	Ecart-type (n)
TN_Marque 14	0.099	1.000	1.000	0.893	0.199
TN_Marque 14_prédit	-0.256	1.233	0.955	0.893	0.145

Tableau 48 : Validation des résultats entre le prédit et le réel du modèle désagrégé de la marque 14 sur 4 semaines (mois)



Graphique 14 : Qualité de l'ajustement du modèle désagrégé de la marque 14 sur 4 semaines (mois)

TN moyen réel		TN moyen prédit	
	89,30%		89,30%

Tableau 49 : Qualité de l'ajustement du modèle désagrégé de la marque 14 sur 4 semaines (mois)

Le taux de nourriture prédit moyen est le même que le taux de nourriture réel
Globalement, des résultats comparables à ceux obtenus sur la marque 1.

2.23.2.1 Conclusion sur les approches désagrégées :

Les modèles désagrégés prédisent très bien le taux de nourriture, entre 87,8% à 90,1% de « bien prédit » entre les modèles.

Les modèles à la semaine ont de meilleurs R^2 mais de plus faible ajustement entre le réel et le prédit. Les modèles au mois ont logiquement de plus faibles R^2 mais très acceptables pour des modèles désagrégés individuels.

	Marque 1	Marque 2			
Semaine	0,79	0,77			
Mois	0,64	0,67			
	Marque 1				
Semaine					
	TN moyen réel		TN moyen prédit		
		88,30%		88,30%	
Mois					
	TN moyen réel		TN moyen prédit		
		87,80%		90,10%	
	Marque 2				
Semaine					
	TN moyen réel		TN moyen prédit		
		90,20%		90,20%	
Mois					
	TN moyen réel		TN moyen prédit		
		89,30%		89,30%	

Tableau 50 : Synthèse des résultats des modèles désagrégés

La prévision est fiable et le taux de nourriture prévisible à un niveau individuel.

2.24 Le taux de nourriture selon le niveau de revenu

Nous émettons l'hypothèse que le taux de nourriture est socialement marqué, mais que l'amplitude de sa variation, son écart type est également contraint dans une certaine amplitude, quel que soit le niveau social des foyers mesuré par le revenu.

Nous réalisons d'abord une typologie par une méthode de classification des K moyennes

Les variables en entrée :

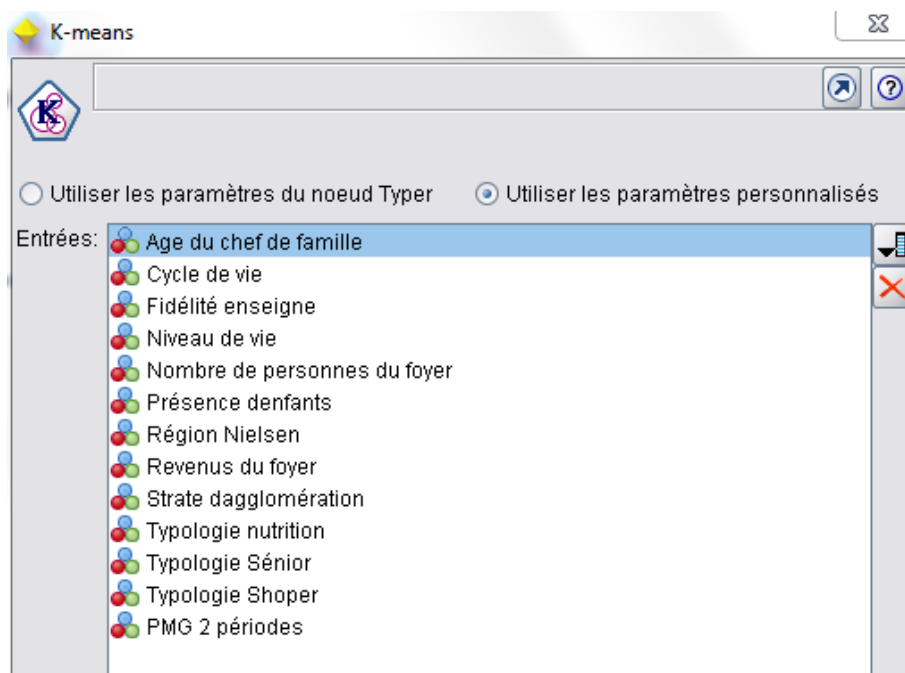


Figure 7 : Les variables en entrée de la typologie

Nous obtenons 5 groupes assez homogènes :

Typo	Comptage	%	Taux de nourriture	
			moyenne	Ecart type
Groupe 1	2 362	21,2%	55,5%	27,4%
Groupe 2	1 986	17,8%	50,5%	25,3%
Groupe 3	2 920	26,2%	43,9%	22,8%
Groupe 4	2 245	20,1%	42,7%	22,4%
Groupe 5	1 635	14,7%	42,3%	22,5%

Tableau 51 : Les groupes de la typologie

Nous calculons le taux de nourriture de ces groupes à la marque 1 sur la totalité de la période des 24 mois, nous nous apercevons que les taux de nourriture sont sensibles au prix, la variable « revenu » va donc jouer un rôle discriminant entre les acheteurs.

La variable revenu :

Typo	En effectif				Total	En indice			
	Taux de nourriture élevé & prix élevé	Taux de nourriture élevé & prix faible	Taux de nourriture faible & prix faible	Taux de nourriture faible & prix élevé		Taux de nourriture élevé & prix élevé	Taux de nourriture élevé & prix faible	Taux de nourriture faible & prix faible	Taux de nourriture faible & prix élevé
Groupe 1	824	1 188	199	151	2 362	115	99	69	94
Groupe 2	678	1 002	185	121	1 986	112	100	76	89
Groupe 3	925	1 390	378	227	2 920	104	94	106	114
Groupe 4	615	1 148	321	161	2 245	90	101	117	105
Groupe 5	344	914	275	102	1 635	69	110	138	91
Total	3 386	5 642	1 358	762	11 148	100	100	100	100

Tableau 52 : Tris croisés entre les groupes de la typologie et les classes TN et prix

Les foyers à plus faibles revenus ont des taux de nourriture plus faibles que les hauts revenus, ceux-ci pouvant acheter des produits à prix élevés (+/-13pts). En revanche l'écart type oscille entre 22,5% et 27,4%.

A la lumière de ces résultats, nous décidons de diviser les cafés par tranche de prix.

Les cafés sont classés en fonction du prix. On divise les acheteurs en trois strates : ceux qui achètent des cafés :

- A prix élevé
- A prix moyen
- A prix faible

Le prix en 3 tranches							
On divise le prix en 3 tranche : Prix élevé, Prix moyen et Prix faible							
Prix élevé : Prix \geq Prix moyen + écart type							
Prix moyen : Prix moyen - écart type < Prix < Prix moyen + écart type							
Prix faible : Prix \geq Prix moyen - écart type							
				Nb de clients		Prix moyen	Taux de nourriture moyen
				Eff.	%		
Prix faible				1 046	9,4%	4,1 €	54,7%
Prix moyen				8 825	79,2%	10,4 €	44,6%
Prix élevé				1 277	11,5%	25,5 €	58,1%

Tableau 53 : Prix par segment

Les quantités achetées par consommateur croisées par les types de prix :

Croisement de la PMG et les tranche de prix						
En nb de clients						
	En effectif			En %		
	P	M	G	P	M	G
Prix faible	661	266	119	12,0%	7,9%	5,2%
Prix moyen	4 028	2 836	1 961	73,0%	84,4%	86,5%
Prix élevé	830	259	188	15,0%	7,7%	8,3%

Tableau 54 : Croisement des segments de prix par PMG

Les taux de nourriture par PMG clients croisés par les types de prix

On calcule le taux de nourriture d'une marque moyenne sur la catégorie par consommateur.

En taux de nourriture									
	P			M			G		
	Taux de nourriture moyen	Ecart type du taux de nourriture	Etendue du taux de nourriture	Taux de nourriture moyen	Ecart type du taux de nourriture	Etendue du taux de nourriture	Taux de nourriture moyen	Ecart type du taux de nourriture	Etendue du taux de nourriture
Prix faible	59,4%	25,9%	85,2%	47,0%	22,8%	82,9%	45,3%	22,5%	87,0%
Prix moyen	55,4%	25,2%	87,5%	37,1%	17,8%	88,3%	33,2%	17,1%	89,0%
Prix élevé	66,5%	26,9%	81,0%	44,7%	20,9%	82,5%	39,0%	18,4%	83,6%

Tableau 55 : Taux de nourriture par segment de prix et PMG

On agrège les consommateurs sur le type de prix et le taux de nourriture :

Graphique 15 : Répartition des types

Sur la 1ère période de 12 mois					
En effectif					En part
P	M	G	Total		
Groupe1 : taux de nourriture élevé et prix élevé	1 816	786	647	3 249	29,9%
Groupe 2 : taux de nourriture élevé et prix faible	2 724	1 705	1 099	5 528	50,8%
Groupe 3 : taux de nourriture faible et prix faible	510	535	305	1 350	12,4%
Groupe 4 : taux de nourriture faible et prix élevé	234	293	220	747	6,9%
Total	5 284	3 319	2 271	10 874	100,0%

Sur la 2ème période de 12 mois					
En effectif					En part
P	M	G	Total		
Groupe1 : taux de nourriture élevé et prix élevé	1 656	735	567	2 958	29,6%
Groupe 2 : taux de nourriture élevé et prix faible	2 490	1 571	1 017	5 078	50,7%
Groupe 3 : taux de nourriture faible et prix faible	479	517	280	1 276	12,7%
Groupe 4 : taux de nourriture faible et prix élevé	254	263	181	698	7,0%
Total	4 879	3 086	2 045	10 010	100,0%

Tableau 57 : Types croisés par PMG

Sur une période courte de 24 mois, on voit que la part des consommateurs avec un taux de nourriture faible et qui achètent des cafés à prix faibles s'accroît. Ces consommateurs sont donc de moins en moins fidèles à des marques car ils ne peuvent se les offrir, et recherchent donc toujours les prix les plus bas.

On remarque aussi que les consommateurs qui achètent les marques les plus chères les achètent moins souvent.

Analyse par niveau de vie :

<u>Les 4 groupes par niveaux de vie</u>							
				En indice			
				Aisé	Moyen supérieur	Moyen inférieur	Modeste
Groupe1 : taux de nourriture élevé et prix élevé				121	99	83	78
Groupe 2 : taux de nourriture élevé et prix faible				91	99	106	106
Groupe 3 : taux de nourriture faible et prix faible				81	104	121	134
Groupe 4 : taux de nourriture faible et prix élevé				107	106	97	92
Total				100	100	100	100

Tableau 58 : Types croisés par niveaux de vie en indice

Les consommateurs du groupe 3 sont sur-représentés dans les niveaux de vie les plus « modestes ».

Les 4 groupes par Nombre de personnes du foyer :

	En indice				
	1 pers	2 pers	3 pers	4 pers	5 et +
Groupe1 : taux de nourriture élevé et prix élevé	113	82	98	105	93
Groupe 2 : taux de nourriture élevé et prix faible	100	106	97	94	95
Groupe 3 : taux de nourriture faible et prix faible	75	125	110	103	129
Groupe 4 : taux de nourriture faible et prix élevé	90	94	112	119	117
Total	100	100	100	100	100

Tableau 59 : Types croisés par nombre de personnes au foyer en indice

On voit que le taux de nourriture (la fidélité aux marques) vs le prix est un marqueur des difficultés sociales. Les consommateurs les plus modestes changent constamment de marque afin de rechercher la marque la moins chère à l'instant t. Ce phénomène s'accroît entre les deux périodes.

A date dans nos travaux précédents sur le taux de nourriture nous n'avions jamais imaginé utiliser le taux de nourriture des produits de grande consommation de cette façon, c'est aussi un des usages possible de travailler sur son élasticité.

2.25 Discussion et conclusions

Nous concluons notre recherche en

- Répondant aux questions de recherche
- Montrant les avancées et les limites de notre démarche
- Proposant des pistes pour des recherches complémentaires

2.25.1 Réponse aux questions de recherche

Dans cette partie, nous apportons des réponses aux questions de recherche que nous avons posées.

2.25.1.1.1 L'élasticité du taux de nourriture est-elle limitée :

En analysant les différents jeux de données nous avons montré par une analyse empirique que le taux de nourriture est limité avec une élasticité limitée et ce quelles que soient les périodes étudiées :

		Tx de nourriture		Ecart type du TN en %	
Sur 12 mois		Année 1	Année 2	Année 1	Année 2
Brosse à dent	Long	66,9	68,2	32,1	31,9
Désodorisant	Moyen Long	52,4	54,7	33,6	34
Dentifrice	Moyen	50,3	51,6	32,4	32,8
Plat surgelé	Moyen	52,4	55,3	35,4	35,9
Café	Moyen	36,7	37,1	31,5	31,7
Eau Plate	Court	29,7	30,7	29,4	29,8
Moyenne		48,1	49,6	32,4	32,7

Tableau 60 : Elasticité des taux de nourriture et écarts types des élasticité par année et par catégorie de produits

Si on calcule les écarts types moyen du taux de nourriture sur le marché du café, tout acheteur et toutes marques confondues, on voit que les écarts sont limités quelle que soit la période temporelle utilisée (semaine, mois, années).

	Année 1	Année 2	Total
A la semaine	23,2%	23,0%	23,1%
Au mois	29,0%	28,9%	29,0%
A l'année	31,9%	31,7%	30,9%

Tableau 61 : Ecarts types du taux de nourriture sur différentes périodes par année

2.25.2 Peut-on prévoir le taux de nourriture sur différentes durées de consommation ?

On a démontré que l'on peut prévoir avec différentes méthodes sur des données plus ou moins agrégées et sur des périodes différentes, le taux de nourriture de façon assez précise.

	ARIMA		Perceptron	Modèle de panel	
	Marque 1	Marque 14	Marque 1	Marque 1	Marque 2
Semaine	0,83	0,82	0,77	0,79	0,77
Mois	0,88	0,93	0,99	0,64	0,67

Tableau 62 : R² des différents modèles testés

Semaine					
	TN moyen réel		TN moyen prédit		
		88,46%		88,01%	
Mois					
	TN moyen réel		TN moyen prédit		
		78,52%		78,52%	

Tableau 63 : Taux de nourriture moyen entre le prédit et le réel quels que soient les modèles à la semaine et au mois

Au-delà des méthodes et sur des données avec des agrégats différents, on retrouve une stabilité de la prévision. Cette stabilité prouve indirectement des écarts limités car si les écarts étaient erratiques la prévision serait par définition « mauvaise ».

La donnée hebdomadaire étant la plus « bruitée » toutes les autres périodes donnent des prévisions encore plus stables.

2.25.3 L'élasticité du taux de nourriture varie t'elle d'une catégorie de produit à l'autre :

Sur 12 mois		Nb de marque sur le marché	Nb de marque moyen achetées	QA/NA	TN en %	Ecart-type en %	Délai inter achats en semaines	Prix moyen en €
Brosse à dent	Long	28	1,7	2.9 unités	66,9	32,1	21,4	2,30 €
Désodorisant	Moyen Long	170	2,2	5.9 unités	52,4	33,6	11,6	2,50 €
Dentifrice	Moyen	47	2,2	7.7 tubes	50,3	32,4	10,2	1,70 €
Plat Surgelé	Moyen	150	2,1	5.6 kg	52,4	35,4	8,8	5,11 €
Café	Moyen	16	3,1	6 kg	36,7	31,5	3,7	9,40 €
Eau Plate	Court	115	3,9	166.4 litres	29,7	29,4	3,1	0,23 €
	Moyenne	NS	2,5	NS	48,1	32,4	NS	3,54 €

Tableau 64 : Types croisés par nombre de personnes au foyer en indice

Elle varie mais dans une amplitude assez limitée même si les catégories sont différentes et que les taux de nourriture sont différents. En moyenne, un écart type du taux de nourriture de 32,4 avec un maximum à 35,4 et un min à 29,4.

On trouve l'idée d'une élasticité limitée par des caractéristiques endogènes au consommateur.

2.25.4 L'élasticité bornée du taux de nourriture est-elle socialement explicable ?

Nous avons montré qu'en regroupant les consommateurs sur le taux de nourriture et la sensibilité au prix des marques, on peut comprendre des dynamiques de « dé-consommation » (ou de moindre consommation de marques) qui sont le reflet de difficultés sociales. Nous avons vu que sur notre période d'analyse, cette dé-consommation s'est accrue avec une élasticité du taux de nourriture qui se réduit chez les catégories sociales plus défavorisées.

L'évolution de l'élasticité du taux de nourriture est donc révélateur des difficultés financières des ménages et peut donc être un outil d'aide à la détermination des prix.

2.26 Contribution et limites de notre recherche :

2.26.1 Les contributions :

Depuis notre recherche s'étalant sur 20 ans, nous essayons de déterminer la rentabilité de la fidélisation des consommateurs et leurs potentiels. Même si il y a sûrement un petit côté « obsessionnel » dans une telle focalisation sur un sujet « si précis », cette dernière recherche sur la part captable que peut s'allouer la marque par l'étude de l'élasticité du taux de nourriture est une avancée selon nous pour les points suivants :

- Elle travaille sur un grand corpus de données sur des données individuelles de home scanning sur une longue période et sur 6 marchés, on est presque dans du big data et on utilise un panel avec la lecture sur une longue période. On fait des comparaisons pour tirer des généralisations.
- Elle croise des modélisations différentes ARIMA, réseaux neurones, analyse de données de panels, modèle de cooccurrence d'achat pour essayer de montrer la convergence des résultats. En général ces techniques assez sophistiquées ne sont pas mises en compétition dans un même travail.
- Elle montre que le taux de nourriture et l'évolution de son élasticité est aussi un moyen de détecter la dé-consommation des ménages les plus modestes, pour pouvoir en tenir compte dans la politique de prix par exemple.

Les applications managériales sont nombreuses car via la détermination des variations intra-individuelles et inter-individuelles du taux de nourriture, on peut estimer la part de marge « captable » au niveau individuel et donc la rentabilité des actions marketing de manière prévisionnelle.

A ce stade le but marketing de notre recherche qui était de quantifier et de modéliser empiriquement l'élasticité du taux de nourriture pour en déduire une marge allouable à la fidélisation est atteint. Notre travail est suffisamment précis pour être à minima reproduit par d'autres chercheurs.

2.26.2 Les limites :

- Les données de panel individuels et magasins ne portent que sur des produits de grande consommation.
- Nous avons choisi des modèles précis ARIMA, perceptron, Apriori, Séquence, des modèles d'analyse de données de panel basés sur la régression, nous aurions pu choisir d'autres types de modèles.

- Elle ne porte que sur 6 marchés.
- Nous aurions pu analyser des sous marchés homogènes par marché ou des sous-ensembles de marques avec des caractéristiques similaires, nous avons limité notre choix de marques.
- Nous avons travaillé que sur un panel nous aurions pu, comme dans nos autres recherches, le faire aussi sur des bases de distributeurs.
- La période étudiée de 104 semaines n'est peut-être pas assez longue.
- Elle ne fait pas suffisamment de liens avec la LTV, recherche qui serait à mener.

2.27 Conclusions :

Cette thèse sur travaux m'a poussé à réactualiser deux de mes travaux passés sur la fidélisation et à réfléchir sur la détermination du taux de captation que j'avais déjà évoqué dans ces précédents travaux.

Il apparaît que le taux de nourriture et son élasticité ou son inélasticité sont des facteurs clefs pour déterminer les clients potentiellement rentables dans un programme de fidélisation, car pouvant accroître leurs consommations sans subir cet accroissement vs ceux qui ne le peuvent pas : c'est la notion même du taux de captation, la part que la marque peut capter grâce à la fidélisation.

L'explication de cette élasticité « limitée » du taux de nourriture est due principalement à l'inertie du consommateur qui répète ses achats aux marques qu'il consomme d'habitude, avec un niveau de besoin en quantité à peu près identique. Le consommateur est finalement très stable dans ses achats des produits étudiés. Ce n'est probablement pas une inertie due à un manque d'implication, ni un manque de satisfaction c'est une habitude non subie, positive.

Cette recherche est donc une continuité de travaux qui m'a amené à préciser un sous élément de la rentabilité de la fidélisation, l'élasticité du taux de nourriture pour en conclure à une élasticité « bornée », « limitée ». C'est uniquement dans la cadre de cette élasticité limitée que la marque peut déterminer la part captable des achats fait à la concurrence pour rentabiliser ses actions de fidélisation.

L'inertie du taux montre qu'il est très difficile d'aller au-delà.

Il est donc très important de déterminer par client quels sont les clients présentant la plus grande élasticité au-delà de la plus grande probabilité de ré-rachat. Cela devrait changer les segmentations clients en leur rajoutant systématiquement une part captable : l'élasticité.

Nous l'avons testé empiriquement cela fonctionne. Cette démarche remettra en question certaines conclusions sur la non rentabilité de la fidélisation. Par ailleurs, elle montre que ne pas encourager les bons clients élastiques à le rester, fragilise à moyen terme la rentabilité de la marque ou de l'enseigne. Le programme de fidélisation ayant pour but non seulement d'obtenir une part des achats à la concurrence, mais de garder l'élasticité du consommateur contributif. Si cette élasticité se réduit à terme, c'est un déclin de la valeur du consommateur.

A méditer quand on voit les efforts marketing se détourner de la fidélisation des meilleurs clients pour se tourner vers des réductions de prix.

La comparaison sur le long terme des deux stratégies, sur le taux de nourriture du consommateur sera l'objet d'autres recherches...

Le Taux de Nourriture (TN) est mal utilisé car très invariant. Espérons que cette recherche aura montré que sa variation limitée est une preuve de son utilité pour le décideur.

Deuxième Partie

***Recherche 2* : La fidélisation rentable : proposition d'un modèle composite entre la valeur d'un client à l'instant T et sa valeur à vie**

Avant-propos

La présente étude est une nouvelle version 2012 de précédents travaux publiés en 2002/2003.

Elle s'est nourrie de toutes les recherches que nous avons effectuées chez SOCIO Logiciels depuis notre modèle initial Objectifs Fidélisation mis au point à D Interactive et le calcul de LTV (Valeur à Vie du Client sur sa Durée de Vie)⁷, sans pour autant en modifier fondamentalement la forme originale.

Comme dans la première version, nous mettons en évidence à chaque étape, les points communs et les complémentarités de deux méthodes : une estimation de la valeur du client à l'instant T et une méthode de l'estimation de valeur du client sur sa durée totale de consommation. Nous voulons faire un travail qui explique simplement comment les deux méthodes se calculent et se complètent afin d'être accessibles au plus grand nombre. Sur ce point, cette nouvelle version s'est surtout améliorée dans sa précision méthodologique.

Ce travail de recherche porte sur la détermination de la rentabilité des programmes de fidélisation et de CRM (Customer Relationship Management : Gestion de la Relation Clients). Il a pour but de définir une nouvelle modélisation de cette rentabilité en tenant compte des modèles existants. Après avoir passé en revue les concepts théoriques sous tendus :

- couverture de cibles
- structure PMG (segmentation en Petits, Moyens, Gros clients basée sur le CA)
- activation
- taux de captation
- marge allouable
- maintien des coûts unitaires
- axe temps

Nous comparons deux modélisations à partir de deux exemples concrets :

- Objectif Fidélisation[®] est un modèle d'analyse de la valeur du client qui prend en compte le temps sur l'instant T
- Un modèle de LTV (Life Time Value : Valeur du Client à Vie) est un modèle d'analyse de la valeur du client qui considère le temps sur un continuum et donc cette valeur du client est estimée sur ce continuum.

Ces deux modèles sont construits en utilisant l'ensemble des concepts théoriques et définis et en vérifiant leur degré d'applicabilité. Une convergence des deux modèles est alors proposée sous la forme d'un modèle composite réunissant dans une même formulation les deux modèles précédents : la valeur du client à l'instant T vs sa valeur sur la totalité de sa durée de vie.

L'apport de la recherche étant de montrer la complémentarité des deux modèles.

Les avantages et les limites de l'approche sont alors mis en exergue.

⁷ Life time value est le terme anglais utilisé pour désigner la valeur vie client. La life time value est donc la somme des profits actualisés attendus sur la durée de vie d'un client. Le principe de la "life time value" est central dans le domaine des politiques d'acquisition et de fidélisation de la clientèle.

Introduction:

En cette période de crise et de concurrence exacerbée, les entreprises tentent de conquérir de nouveaux clients mais également de se concentrer et de développer leurs clients les plus rentables. Par ailleurs, la part des investissements en communication hors médias n'a cessé de croître pour atteindre 63,3% des dépenses hors média vs 36,7% pour les mass médias.

Le hors médias et plus précisément le marketing direct sont, par tradition, un poste de dépenses important pour les annonceurs où la rentabilité doit absolument être mesurée. Pourtant peu de travaux ont fait à notre connaissance, un point à date sur les différentes techniques d'estimation de la valeur du client, même si les sujets sur la fidélisation, le CRM ont été abondamment traités en 2000 et 2001.

Ce travail de recherche porte sur la détermination de la rentabilité des programmes de fidélisation et de CRM (Customer Relationship Management : Gestion de la Relation Clients).

Il a pour but de définir une nouvelle modélisation de cette rentabilité en tenant compte des modèles existants : modèle qui estime la rentabilité de la fidélisation à l'instant T, modèle qui estime la rentabilité de la fidélisation sur la durée de vie totale du client.

Après avoir passé en revue les concepts théoriques sous tendus : -couverture de cibles -structure PMG (segmentation des clients en Petits, Moyens, Gros clients basée sur le CA) -activation -taux de captation - marge allouable -maintien des coûts unitaires -axe temps, notre recherche compare alors deux modélisations à partir de deux exemples concrets :

- Objectif Fidélisation[®] est un modèle d'analyse de la valeur du client qui prend en compte le temps sur l'instant T, donc uniquement une partie de la durée de consommation du client
- Un modèle de LTV (Life Time Value : Valeur du Client à Vie) est comme son nom l'indique un modèle qui cumule la valeur du client sur l'ensemble de sa durée de consommation.

Le modèle court termiste de la durée de vie, Objectif Fidélisation est donc un modèle qui est par construction inclus dans le modèle de LTV. C'est en cela que l'approche est « composite ». Peu de travaux démontrent la complémentarité des deux approches, et comme elles sont prédictives l'une de l'autre.

Une convergence des deux modèles est alors proposée sous la forme d'un modèle composite réunissant dans une même formulation les deux modèles précédents. L'apport de la recherche étant de montrer la complémentarité des deux modèles. Les avantages et les limites de l'approche sont alors mis en exergue.

Nous définirons les termes clefs de notre recherche.

L'objet de la présente recherche est à la fois de faire une synthèse des techniques de mesure de la valeur du client quel que soit le type de « relations » entreprises avec lui pour justement essayer sans a priori, de définir un modèle d'estimation de cette valeur suffisamment robuste et simple pour être compris et appliqué par le plus grand nombre de décideurs.

Le but de nos travaux étant de fournir à la fois une technique d'estimation mais aussi des règles d'estimation de la valeur du client permettant de servir de base à l'établissement d'une stratégie marketing devant prioritairement déboucher sur des applications très opérationnelles.

Le décideur doit pouvoir choisir quel type de CRM, de marketing relationnel il va développer avec ses clients. Nous situons notre travail au-delà des modes, actuellement le marketing direct s'est renommé le CRM⁸ en incluant certes un plus vaste domaine mais comment s'appellera-t-il demain ?

⁸ La **gestion de la relation client** consiste à savoir cibler, attirer et conserver les bons clients et représente un facteur déterminant du succès de l'entreprise. Construire et développer des relations avec ses clients est un défi, particulièrement lorsque l'entreprise

Notre approche essaie de déterminer une approche « immuable » de la valeur du client pour estimer cette valeur dans le temps et perdurer même si, sur cette période, les techniques et les dénominations changent.

Notre modélisation est basée sur des présupposés méthodologiques que nous validons par une revue des théories qui les supportent.

Nous proposons à la fois un modèle conceptuel sur l'évolution de la valeur d'un client dans sa durée de vie et de la part attribuable à une marque : la part de la valeur que la marque peut utiliser pour à la fois couvrir les coûts du programme et dégager un profit.

Nous définissons ensuite une modélisation mathématique « composite » permettant de déterminer la valeur allouable par cette marque sur ce client pour le fidéliser et lui faire accroître sa consommation sur cette durée de vie.

L'aspect pratique de notre recherche ne domine pas sur l'aspect théorique puisque c'est une nouvelle modélisation dynamique que nous proposons au lecteur.

Par ailleurs, cette recherche est basée sur les travaux que nous avons réalisés pour nos clients au sein de D Interactive. Basée sur 25 modélisations à date dans des secteurs très divers, elle nécessite d'être adaptée à chaque secteur, chaque client. Néanmoins, le fondement théorique de la réflexion reste le même et se dégage, alors, une sorte de méta modèle : le modèle conceptuel.

Cette recherche vise trois publics :

D'abord, le décideur qui découvrira une publication lui permettant d'estimer la valeur d'un client, synthétisant un nombre important de travaux précédents.

Le chercheur qui pourra valider les enjeux théoriques en tenant compte de l'avis de praticiens/experts du domaine.

L'étudiant qui pourra lire un support pédagogique et exhaustif sur les moyens d'estimation de la valeur d'un client sur sa durée de vie. Beaucoup de lecteurs ont entendu parler de LTV (Life Time Value ou Valeur du Client sur sa Durée de Vie). Combien l'ont vu réellement calculée ?

possède des milliers (voire des millions) de clients qui communiquent avec celle-ci de multiples manières. Pour arriver à un résultat satisfaisant, les systèmes de gestion des relations clients (*customer relationship management* - CRM en anglais) doivent permettre aux responsables d'entreprise de mieux comprendre leurs clients pour adapter et personnaliser leurs produits ou leurs services. Il doit aussi permettre aux différents départements de l'entreprise de collaborer à travers le partage d'information concernant toute interaction avec le client.

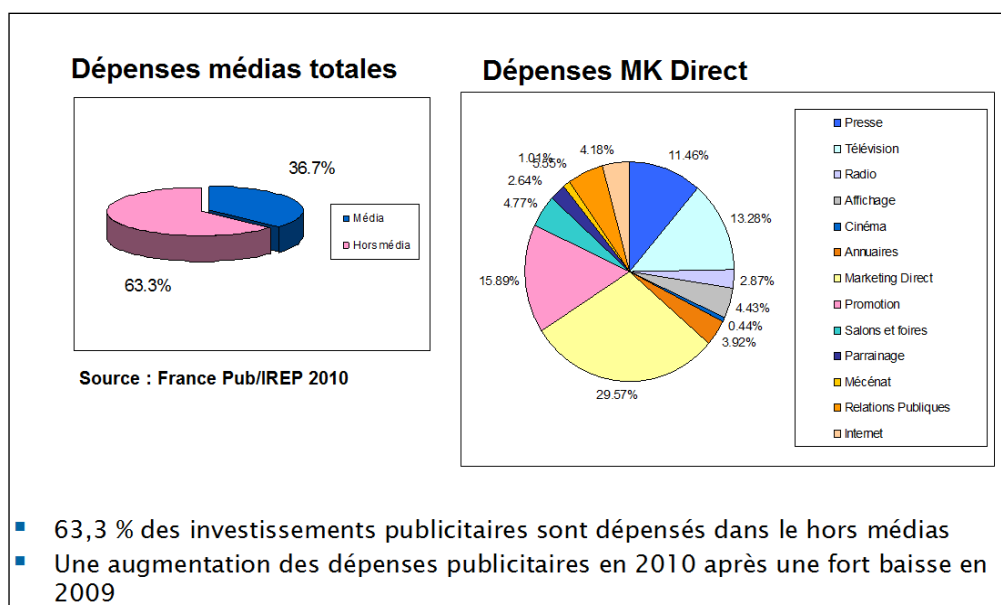


Figure 8 : La part du hors médias, l'enjeu

Cette figure remise à jour par rapport à la première édition, montre la part prégnante du Hors Média dans les dépenses des annonceurs avec 63% des dépenses et du Marketing Direct avec 29,6% en 2009. On voit la part très importante du Hors Médias et donc l'enjeu de la mesure de sa rentabilité dans le mix marketing.

	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	Cumul	% vert.
Marketing direct	23	14	59	32	42	37	13	29	18	39	35	34	43	24	20	5	467	70%
Base de données	5	2	0	14	2	2	0	0	3	0	19	11	8	8	10	0	84	13%
Relation client	0	0	0	0	0	0	0	0	0	0	0	0	11	21	19	0	51	8%
Marketing relationnel	0	0	0	0	0	0	0	0	0	0	4	0	0	6	11	5	26	4%
Gestion de la relation client	0	0	0	0	0	0	0	0	0	0	0	0	0	3	9	0	12	2%
CRM	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	0	10	2%
Data mining	0	0	0	0	0	0	0	0	0	0	0	1	6	1	2	0	10	2%
One to One	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	3	0%
Marketing de base de données	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
TOTAL	28	16	59	46	44	39	13	29	21	39	58	49	68	64	80	10	663	100%

Tableau 65 : L'évolution du concept de marketing direct par année

Etude de dénombrement des mots réalisée à partir du corpus complet de Marketing Magazine, Marketing Vente, Marketing Direct de 1988 à 2001.

Ce tableau est issu d'un dénombrement de mots choisis dans une page systématique du magazine Marketing Direct, il montre l'émergence soudaine et forte du mot CRM en 2000 vs d'autres mots. Le CRM va devenir le « mot marketing à la mode » des années 2000. L'évolution de « CRM » a entraîné un besoin croissant de mesure de son ROI.

Quelques définitions :

Rappelons ici quelques définitions clefs des concepts les plus importants de notre recherche.

CRM :

Selon Desmet (2001) : Mise en œuvre d'une réorganisation de l'information par les entreprises en vue d'intégrer toutes les sources d'information client détenues par les services (commercial, marketing, clients) dans une application unique de gestion de la relation. On perçoit dans la définition ci-dessus que la détention

d'information sur le client est un préalable à la détermination de sa valeur. Nous doutons malheureusement de la pérennité du concept.

Fidélisation :

Lorsque l'on définit la fidélisation, on détermine en général la notion de programme de fidélisation :
Selon Bénavent et Crié (1998) :

« Par programme de fidélisation on entend un ensemble d'actions organisées de telle manière que les clients les plus intéressants et les plus fidèles soient stimulés, entretenus, afin que l'attrition, c'est à dire le taux de clients perdus, soit minimisée et/ou que les volumes achetés soient augmentés ».

Selon Trinquecoste (1996) :

Un client fidèle est donc un client qui, de façon répétée voire systématiquement (c'est à dire exclusivement), fait le choix d'un même produit ou d'une même marque préférée à celui ou celle d'une ou plusieurs offres considérés comme substituables et provenant d'entreprises concurrentes.

La stratégie de croissance par développement des consommations de la clientèle actuelle.

Nous rappelons ce concept car nous opposerons deux situations possibles pour une marque face à la fidélisation :

- offensive : la marque peut accroître ses achats par « captation » d'unité que le client achète à la concurrence. Pour ce concept Trinquecoste parle de « culture intensive ».
- défensive : la marque essaie de maintenir les quantités actuelles de ses meilleurs clients car son taux de nourriture est très élevé sur cette population. Ce qui fait qu'elle ne peut pas gagner des achats incrémentaux sur la concurrence mais ne peut que perdre des unités.

Fidélité vs rétention :

Selon Crié (1996) :

« La **fidélité** peut être constatée comme la base du ré-rachat, par une attitude favorable ou encore le rachat, si l'on se trouve dans le cadre de biens durables. La fidélité se rapporte au niveau individuel de mesure. Elle indique la propension des individus à répéter leur choix de marque ou d'enseigne. Fidéliser, c'est inciter les clients actuels à choisir de nouveau notre enseigne à la prochaine occasion d'achat ».

Il oppose cette définition à la notion de rétention :

« La **rétention** représente simplement le fait de maintenir en activité une clientèle existante. Elle se mesure essentiellement à l'aide du taux d'attrition (proportion de clients perdus en général sur 12 mois). C'est une mesure faite à un niveau agrégé, qui peut masquer l'hétérogénéité d'une population de clients. Pratiquement, on peut considérer que la rétention est la conséquence d'une mauvaise fidélisation ».

Les deux notions sont intimement liées et il n'est pas de notre propos d'en affiner l'une ou l'autre. La fidélité est donc un acte proactif de la part de la marque, l'attrition serait un constat passif, un résultat.

De l'observation du taux d'attrition, le responsable marketing pourrait décider de mener une politique de fidélisation.

Même si le concept va au-delà de la fidélisation du client, il nous semble important de rappeler sa définition.

Valeur client

Le concept de Life Time Value (LTV) que nous traduisons par « valeur du client sur sa durée de vie » est défini par Don Peppers et Martha Roger dans un article récent paru dans DM Review (1998) : « La LTV est une quantité qui peut être précisément calculée pour chaque consommateur.

La Life Time Value est donc la somme des profits actualisés attendus sur la durée de vie d'un client.

On en a beaucoup parlé dans le monde des services (notamment de la banque) et on s'est souvent trompé en se perdant dans la conjecture de l'analyse des coûts rendant les modèles pour estimer la valeur des clients parfois trop complexes pour être utilisés opérationnellement.

Bien que l'analyse des coûts soit très enrichissante, par ailleurs, on peut affirmer que dans la plupart des cas, les plus gros clients en CA (mouvement du compte, nombre de produits différents souscrits et montant total des avoirs) sont les plus rentables. L'analyse des coûts venant simplement affiner et complexifier cette analyse.

2.28 Contexte managérial et problématique de recherche

2.28.1 Importance de la fidélisation et du CRM

Une tension économique de plus en plus forte :

- des marques très matures et hyper segmentées,
- une possibilité de différenciation réduite,
- des rivalités concurrentielles exacerbées.

« 72% des annonceurs déclarent « segmenter leurs clients », 15% « modéliser leur valeur », 6% « calculer la LTV ». Ces chiffres montrent un intérêt pour une politique client différenciée mais aussi une relative diversité des pratiques. (Source : « Etudes sur la mesure des clients par les entreprises françaises »⁹).

La croissance passe surtout par des gains de part de marché et la compétitivité par des réductions des économies d'échelle. La fidélisation des clients devenant un moyen moteur indispensable au marketing (JF Trinqucost, 1996).

La fidélisation et la relation client sont des attentes marketing importantes des entreprises.

2.28.2 Les présupposés de la rentabilité de la fidélisation

Selon Hallberg (1995) les programmes de fidélisation ne peuvent être rentables que si cinq contraintes sont atteintes simultanément :

- La couverture d'un cœur de cible préalablement identifié,
- La structure PMG (répartition du CA généré en Petits, Moyen, Gros clients) de la base de données,
- Le taux d'activabilité, (le % de personnes qui vont réagir au moins une seconde fois à un stimuli marketing après leur adhésion au programme de fidélisation)
- La baisse puis le maintien des coûts unitaires de fidélisation (le coût que représente une personne dans le programme de fidélisation)
- Le temps pour parvenir à atteindre les quatre objectifs ci-dessus simultanément.

Le temps est souvent un axe induit. Nous allons définir précisément chacune des conditions ci-dessus qui seront ensuite les paramètres de nos modèles.

2.28.3 Le client et ses différentes valeurs :

La LTV exprime les flux des futurs profits espérés, diminués des coûts des transactions avec le consommateur, diminués du taux du capital pour obtenir une valeur actuelle nette.

Nous retiendrons la formule de Don Peppers et Martha Roger car elle a le mérite de résumer le processus de calcul de la LTV.

Dans leur ouvrage *Le Client Capital de l'Entreprise* Curry et Stora (1993) définissent la pyramide de la valeur client. Ils en reprennent à leur tour l'explication de la loi de la contribution et montrent de façon très empirique la contribution marginale des meilleurs clients :

- La perte de 1% des clients les plus rentables peut entraîner une perte de 20% de la marge.
- Les clients les plus rentables restent aussi les plus élastiques à une offre de consommation incrémentale. Ces trois éléments sont fondamentaux de la « théorie » de la valeur client.

⁹ Etude annuelle réalisée par Marketing Magazine sur l'analyse de leurs propres clients par les entreprises françaises

Le problème de la démonstration de Curry et Stora est qu'ils ne citent dans leur ouvrage aucune théorie pour étayer leur démonstration. Ils basent leur ouvrage sur du pur empirisme.

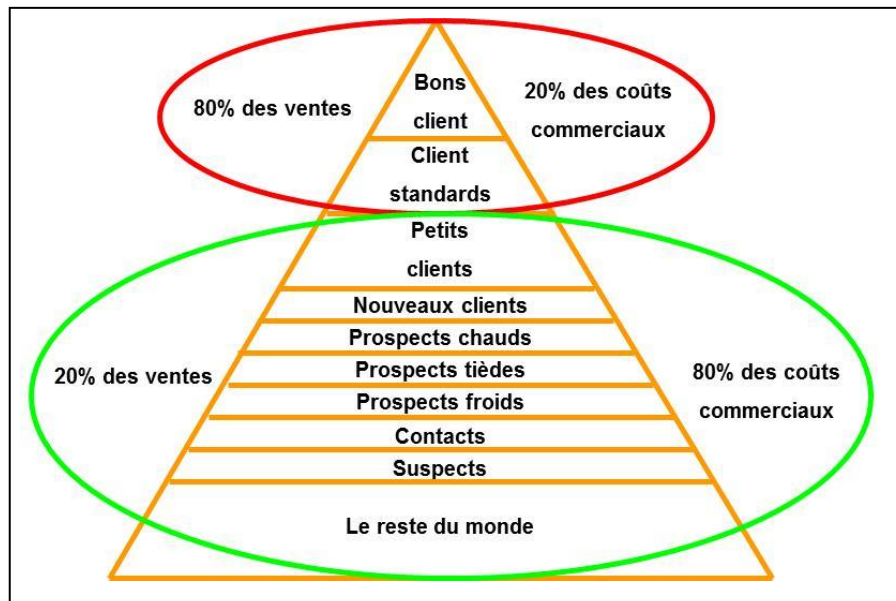


Figure 9 : La pyramide de la valeur client

La pyramide ci-dessus montre l'accroissement de la valeur du client pour l'entreprise au fur et à mesure que la relation se prolonge.

2.29 La durée de vie du client

Il convient de visualiser la LTV (Life Time Value) valeur du client sur sa durée de vie :

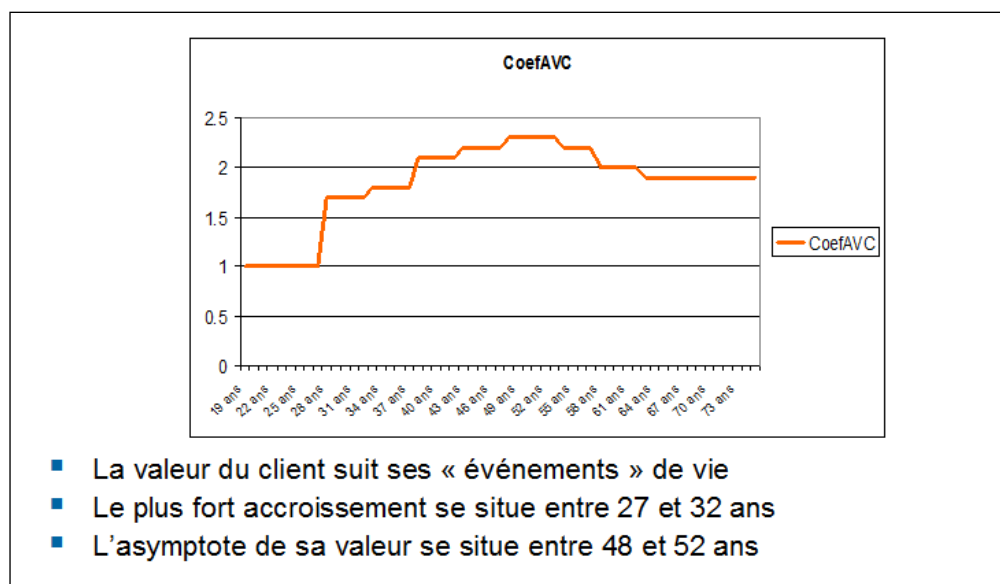


Figure 10 : Durée de vie du client et évolution de sa valeur

La courbe symbolise l'évolution de la valeur d'un client sur sa durée de vie pour un bien d'équipement dans le secteur de l'automobile. On voit que l'asymptote de la valeur « automobile » du foyer d'un individu se situe lorsqu'il a entre 48 et 52 ans. Données issue de l'étude SIMM de Kantar Média.

On peut tracer cette courbe pour tous les produits et services, que ce soit en B to C ou en B to B.

L'approche par la LTV et les coûts :

Nous faisons une comparaison de la formule originale de Wayland et Cole (1997) dans leur ouvrage Customer Connection et « leur » formule reprise par Venturi et Lefébure (2000) dans leur ouvrage Gestion de la relation client. Nous montrons que par cette comparaison les deux formules sont assez proches, seule la définition des paramètres est sensiblement différente.

Wayland et Cole :

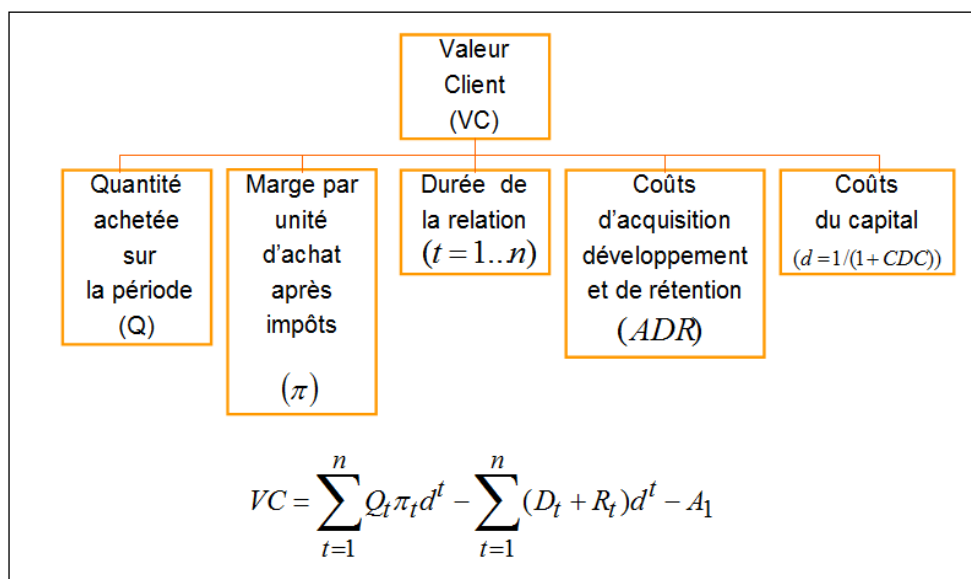


Figure 11 : Modèle de détermination de la valeur client

Ce type de modèle dont les différentes versions sont des dérivés du modèle initial tient compte des paramètres principaux de l'estimation de la LTV, les quantités vendues, la marge, la durée de la relation, la valeur du capital.

Ce modèle a été initialement défini par Dweyer (Dweyer 1989) et ensuite de nombreux auteurs l'ont complété, c'est le cas de Wayland et Cole.

$$VC_i = \sum_{i=1}^n Q_i \pi_i d^i - \sum_{i=1}^n (D_i + R_i) d_i - A_i$$

VC_i : valeur du client i
 Q : volume d'achat
 D : coût de développement
 R : coûts de rétention
 A : coûts d'acquisition
 d : pourcentage de remise avec $d = 1/(1+COC)$
 COC : coûts par rapport au capital investi
 $t = \sum_{i=0}^n t_i$
 t : durée de vie du client (en mois, années...)

Figure 12 : Modèle de détermination de la valeur client de Wayland et Cole d'après Venturi et Lefébure

Il nous semble que Venturi et Lefébure omettent le paramètre t symbolisant le temps pour Wayland et Cole avec le i du client i. Soit ici i symbolise le temps mais ce n'est pas ce que disent les deux auteurs. Venturi et Lefébure calculent le facteur t mais ne s'en servent pas réellement dans l'écriture de leur modèle général : il n'apparaît pas explicitement. De plus, sur le pourcentage de remise, d, (qui est en fait leur traduction de «taux d'escompte » pour coût du capital), ils ne l'appliquent pas en facteur comme le font Wayland et Cole.

Dans leur ouvrage, Venturi et Lefébure insistent sur le fait que « la valeur client doit distinguer la valeur potentielle des clients (espérance) de la valeur passée (constatée). La comparaison potentielle/constatée permet d'identifier les écarts et d'imaginer des stratégies de gestion de la valeur dans une perspective temporelle.... »

... « La détermination d'une valeur potentielle s'enrichit avec les éléments suivants :

- des coefficients d'actualisation pour mettre en avant qu'un euro d'aujourd'hui vaut plus qu'un euro dans un an.
- des coefficients d'incertitude, qui varient selon le secteur économique.
- la détermination d'un écart par rapport à une norme de comportement identifiée au même niveau du marché global.

Ces éléments visent généralement à relativiser le calcul de la fonction de valeur à long terme par rapport aux gains espérés à court terme.

L'intégration des incertitudes permet d'enrichir la valeur client par une notion plus explicite d'espérance de profitabilité client (EPC). Ce dernier terme est moins chargé de contresens que le terme traditionnel de valeur client et met en avant l'incertitude de l'apparition de ce profit ».

Ce sont les propos de Venturi et Lefébure dans La gestion de la Relation client (2000).qui proposent ensuite la formule suivante :

$$EPC = R + VAN \left(\sum_{i=1}^n (M_i - C_i - PM_i - PR) - F \right) + \Delta M$$

R : coûts de recrutement d'un nouveau client
 VAN : valeur actualisée des profits nets clients
 n : nombre d'années de relation
 M : marge brute générée par la relation sur la période ti
 C : coûts de gestion de la relation sur cette même période
 PM : coûts des opérations marketing effectuées sur le client
 PP : coûts des opérations promotionnelles effectuées sur le client
 F : coût de rupture de la relation
 ΔM : variation des marges sur les années estimées

Figure 13 : Modèle de détermination de la valeur client de Venturi et Lefébure

Cependant à la différence de Wayland et Cole cette formule ne tient compte ni de l'incertitude, ni d'un écart marché afin de déterminer une élasticité.

La formule ci-dessous de ces deux auteurs tient-elle réellement compte de ce phénomène ? Dans un article paru dans Journal of Database Marketing intitulé Calculating lifetime values of business customers for a telecommunications company, Kim, Min et Shin (1999) utilisent la part de marché « valeur » et une sorte de prévision pour déterminer l'incertitude marché.

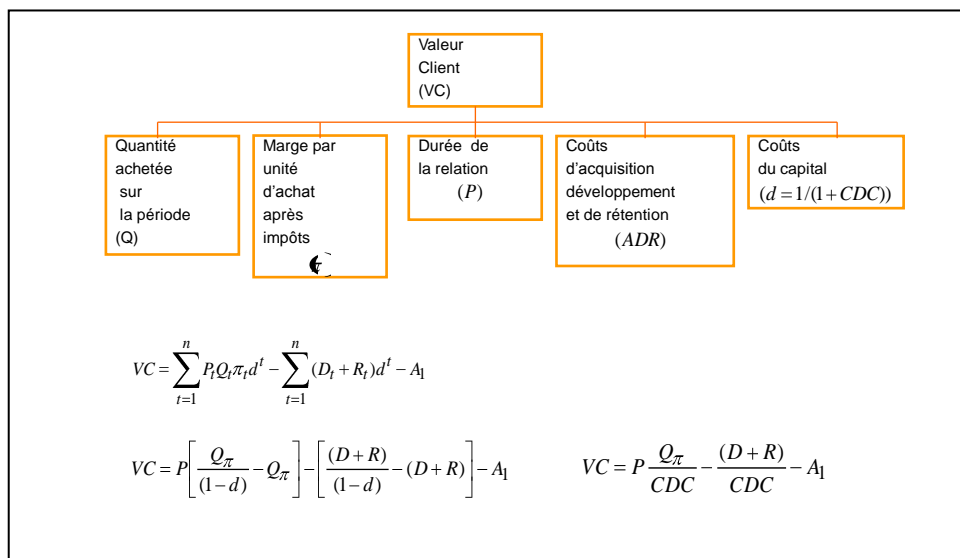
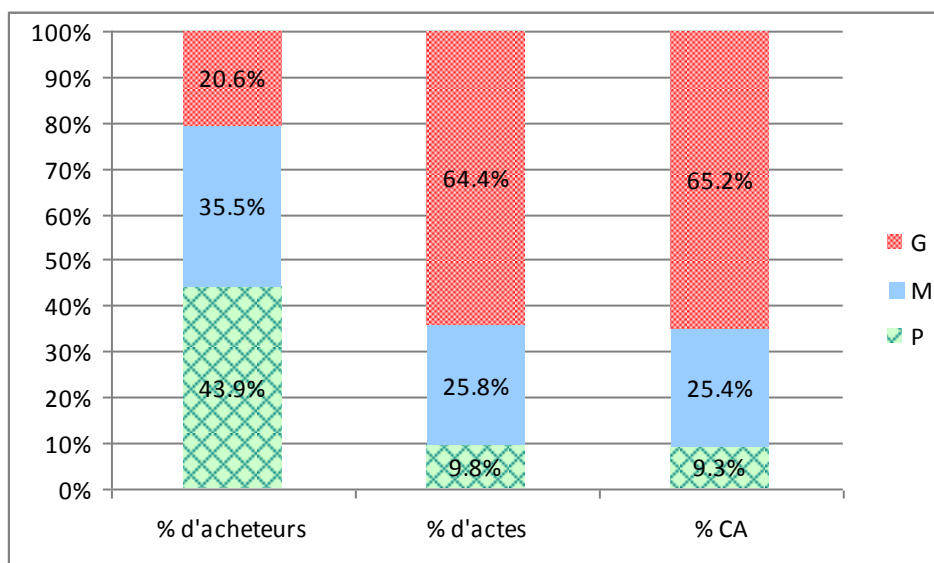


Figure 14 : Modèle de valeur client et probabilité d'achat

Nous préconisons comme ces auteurs d'utiliser la part de marché « valeur » de la marque pour estimer la part que celle-ci va pouvoir s'attribuer sur la consommation future du client. La part de marché « valeur » est l'agrégat au niveau marché des taux de nourriture individuels exprimés en « valeur ». L'objectif recherché est la maximisation de la valeur client par la fidélité dans la durée et la prise de valeur à la concurrence.

Qui fidéliser ?

Structure PMG sur un exemple du secteur cosmétique vendu en GMS :



Graphique 16 : Structure PMG sur un exemple du secteur cosmétique vendu en GMS

Selon Desmet (2001) les clients les plus rentables sont les gros clients, Pour simplifier ceux qui représentent un chiffre d'affaires en général un écart type au-dessus de la moyenne.

Comme le rappelle également Babette Leforestier de Sécodip, 36% des clients d'un hypermarché génèrent 78% du CA¹⁰. Le raisonnement semble donc pouvoir également s'appliquer à la grande distribution.

Ces clients sont fidèles, possèdent plus de produits que la moyenne. Ce sont des « G » ou CPP (Clients les Plus Précieux). Ce raisonnement est démontré dans plusieurs ouvrages pour les produits de grande consommation, pour les services banques/assurances et pour la distribution par Brian Woolf dans son ouvrage Le marketing de la différence.

Cet auteur souligne les points suivants :

Les coûts fixes des petites transactions sont supérieurs à ceux des gros achats.

Il démontre une corrélation entre la valeur du panier moyen et la marge de celui-ci. Il montre aussi que les gros clients ont, par définition, les plus gros paniers moyens donc de plus petits coûts de gestion que des clients plus petits. Ils sont marginalement plus rentables à fidéliser.

Les coûts unitaires de fidélisation sont donc marginalement mieux amortis sur les clients les plus contributifs que sur les autres. L'auteur parle aussi de la corrélation entre le PVC (Prix de Vente Client) des produits toujours plus élevés en moyenne chez les clients les plus contributifs.

Il rajoute « pourtant la plupart des distributeurs du secteur alimentaire font exactement l'inverse et donnent la priorité au client de proximité qui peut entrer n'importe quand dans un magasin, acheter l'offre spéciale du jour et payer à la caisse rapide... ». Nous avons nous même constaté la difficulté d'expliquer (notamment auprès de pétroliers) que le petit client (même de proximité) n'a pas de rentabilité en tant que telle et qu'il est pratiquement captif à l'enseigne de par sa proximité. Dans tous les cas, il ne dégage pas une marge suffisante pour être fidélisé comme les consommateurs les plus contributifs. Il faut viser le « gros » client.

¹⁰ Référence 2003 : étude annuelle réalisée par Kantar sur la GSA

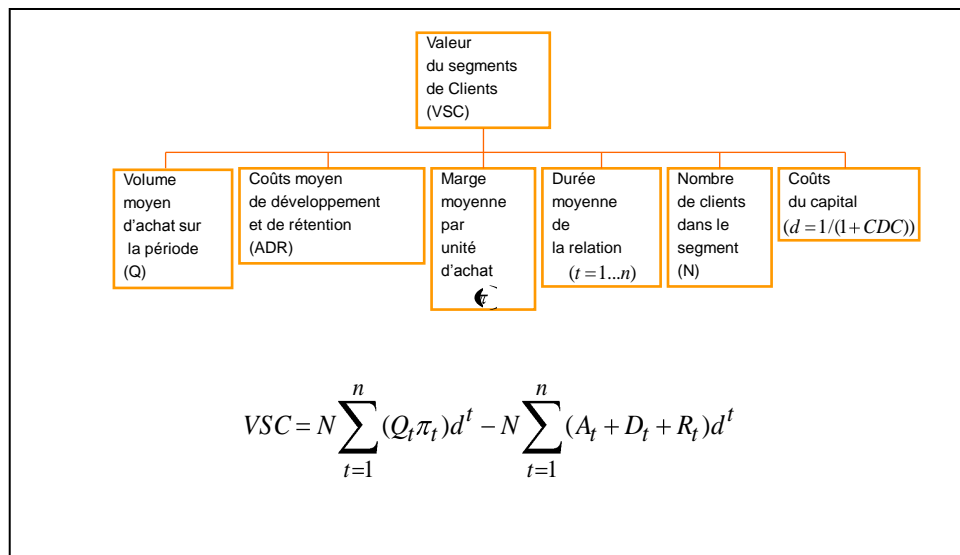


Figure 15 : Détermination des segments clients les plus rentables

Une approche complémentaire est de définir sur quel type de segmentation client se trouve le marché de la marque à fidéliser. Un certain nombre de grands faits s'opposent pour déterminer la valeur d'un ou plusieurs segments de clients :

- l'offre est un produit ou un service,
- on maîtrise ou non la distribution,
- la marge est forte ou faible en valeur relative et surtout en valeur absolue,
- le cycle d'achat ré-achat est court ou long (de quelques semaines à des années),
- le produit est impliquant pour le consommateur ou non
- la consommation du produit ne dure qu'une période (couche, médicament...).

Ces différents éléments vont avoir un effet direct et très important sur la prise en compte des segments P, M, G de clients.

Certains auteurs parlent de CVM, Customer Value Management, sorte d'acronyme tentant de remplacer le CRM. Le CVM consiste simplement à piloter la valeur des clients en fonction de leur rentabilité future. Il prône l'allocation segmentée des moyens moteurs en fonction de la propension des clients à consommer (John Distifano, DM News, 2000, Jan Mrazek, DM News, 2001).

Les clients les plus rentables sont aussi les plus élastiques à l'achat :

Garth Hallberg montre dans son livre « All consumer are not created equal », devenu l'ouvrage de référence sur le marketing de la différenciation des clients par la valeur, que selon l'étude MRCA¹¹ le taux de nourriture moyen des gros consommateurs d'une marque n'excède pas 20%, ceux-ci réalisant 80% de leur achat à la concurrence. Il cite une amplitude selon les secteurs étudiés allant de 48% à 5% de taux de nourriture pour les G sur une marque donnée.

Nous avons aussi vérifié dans beaucoup de cas ce phénomène : les meilleurs clients d'une marque (les plus gros en volume et en CA) achètent en moyenne plus de 50% de leurs achats à la concurrence.

On entrevoit donc bien l'importance du taux de nourriture de la marque sur les différents segments de clients pour la fidélisation.

¹¹ MCRA (Market Research Corporation of America) est une société qui gérait des panels consommateurs dans les années 1990. Elle a été absorbée depuis par IRI.

L'importance du taux de nourriture :

	P	M	G
NA : Nombre d'acheteurs	6 553 818	5 299 785	3 075 368
QA : Quantité Achetée à la Marque X par acheteur	1,4	4,1	11,9
QA autres marques par acheteur	35,2	41,8	54,5
Nombre d'actes	1,3	3,4	8,5
Somme dépensée (en euros)	5,55 €	15,12 €	38,78 €
Taux de nourriture	3,90%	9,00%	17,90%
Quantité achetée par acte d'achat	1,1	1,2	1,4

Tableau 66 : L'importance du taux de nourriture

Ce tableau est un exemple de données issues d'un panel scannérisé pour une marque de produit de grande consommation.

Dans la littérature récente on retrouve une confusion régulière sur les effets de la fidélisation sur le taux de nourriture :

- la fidélisation devrait avoir un impact positif sur le taux de nourriture. Dans les faits et dans ces articles c'est l'inverse en général qui est observé et donc cette hypothèse est systématiquement invalidée. La fidélisation n'accroît pas le taux de nourriture à la marque, elle accroît les quantités à la marque mais aussi éduque à la catégorie qui croît globalement, le taux de nourriture décroît donc sensiblement.
- En fait on observe en général un accroissement concomitant des QA des produits concurrents sur la catégorie au QA de la marque qui cherche à fidéliser. Bien que les QA augmentent et le CA aussi, le taux de nourriture reste souvent étalé ou diminué.

La fidélisation doit avoir un impact positif sur la part de marché : en fait là aussi comme il y a finalement un lien entre taux de nourriture et part de marché, on constate le même effet : les QA sur NA augmentent mais les quantités globales de la catégorie augmentent également par consommateur.

2.30 Les modèles composites

Nous entendons par « modèle composite » un modèle dynamique capable de déterminer la valeur du client à un instant t en prenant en compte les paramètres qui sont définis ci-dessous :

- couverture de cible
- structure PMG
- taux d'activabilité et taux de captation
- maintien des coûts unitaires

Le modèle doit porter sur un instant t mais doit aussi pouvoir tenir compte de l'effet sur la durée de vie globale du client si nécessaire.

Nota : on emploiera tour à tour activation et « activabilité » qui est un anglicisme impropre en français mais nous le préférons parfois, car il porte plus en lui-même la notion de proactivité attendue de la part du consommateur devant agir aux stimulations du programme de fidélisation. Le concept marketing le plus proche étant « activation » : le fait de faire acheter le client.

Rappel des principaux indicateurs marketing utiles à notre modèle : les concepts de notre modèle

Le taux de couverture de cible :

Le taux de couverture de cible est la part de la cible que l'on peut toucher via un programme de fidélisation.

Par adhésion volontaire, on peut capter 30 % de la cible des clients d'une marque dans un programme de fidélisation. Ce taux peut atteindre 50% si l'annonceur possède son propre réseau de distribution.

On observe ce taux de 30% empiriquement dans tous les programmes sur lesquels nous avons été amenés à travailler en France. A l'étranger, ce taux pourrait être différent. Au-delà, de ce taux, l'augmentation des moyens de recrutement risque d'impacter fortement les coûts. En plus, il y a un risque de recrutement hors cible. A ce moment-là, les coûts unitaires ne sont plus assez bas pour que les achats incrémentaux des fidélisés amortissent l'investissement du programme. La fidélisation devient alors plus chère et moins efficace.

Pourquoi ce taux ?

Compte tenu des moyens de recrutement existants pour la marque à un coût unitaire décent en maintenant une adhésion volontaire, il est très difficile d'aller au-delà, sans impacter gravement les coûts de recrutement.

Au-delà de 30%, le 1% suivant sort de la cible. En élargissant les filets, on recrute davantage de personnes hors cible. Il y a une notion de seuil maximal.

30% c'est aussi le nombre de personnes qui, dans les marchés de masse¹² permettent d'avoir un effet suffisant sur les ventes incrémentales afin d'amortir le programme. A 30% de la cible en base, l'investissement présentera un taux d'amortissement plus rapide à d'autres alternatives à la fidélisation dans le mix marketing de la marque. Ce taux de 30 % est nécessaire pour avoir un effet profitable sur la PDM (Part De Marché) à moyen terme.

En effet, au-dessous de ce taux de couverture de 30 %, le programme n'a pas d'effet volume, on couvre trop peu la cible pour qu'il y ait un effet visible.

Simulation de couverture de cible :

Exemple de visualisation de ciblage qui permet d'isoler dans le marché les sous cibles clefs.

¹² Pour un programme fidélisation compris entre 500 000 et plus de 1 M de personnes en base en fonction de ceux qui vont réagir en achetant d'avantage de produits de la marque fidélisante : les réactifs,

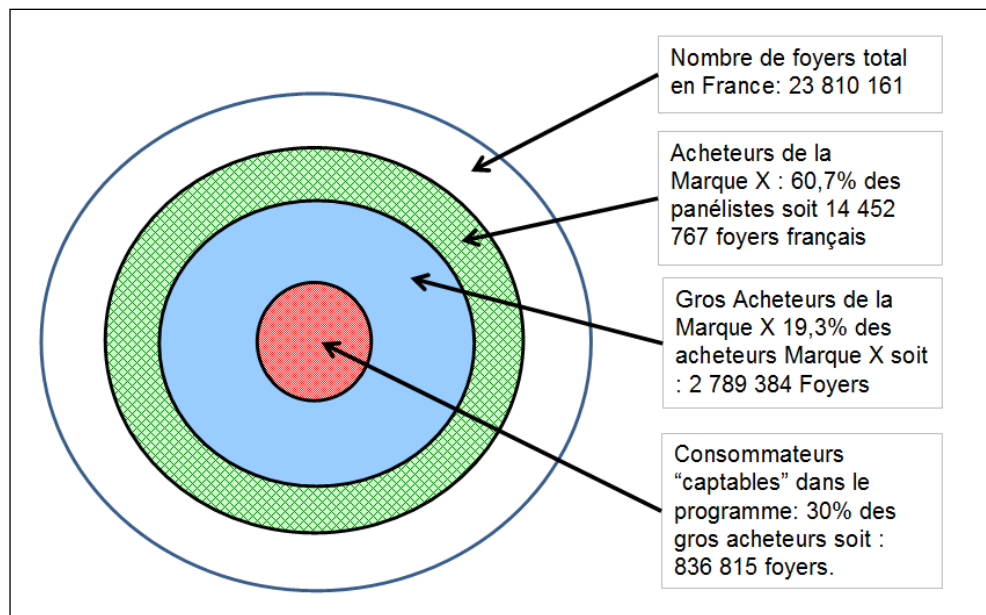


Figure 16 : Simulation de couverture de cible

- Nombre de foyers total en France: 23 810 161 Source INSEE 2002
- Acheteurs de la Marque X : 60,7% des panélistes, soit 14 452 767 foyers français
- Gros Acheteurs de la Marque X 19,3% des acheteurs Marque X, soit 2 789 384 Foyers
- Consommateurs "captables" dans le programme : 30% des gros acheteurs, soit 836 815 foyers.

Structure PMG

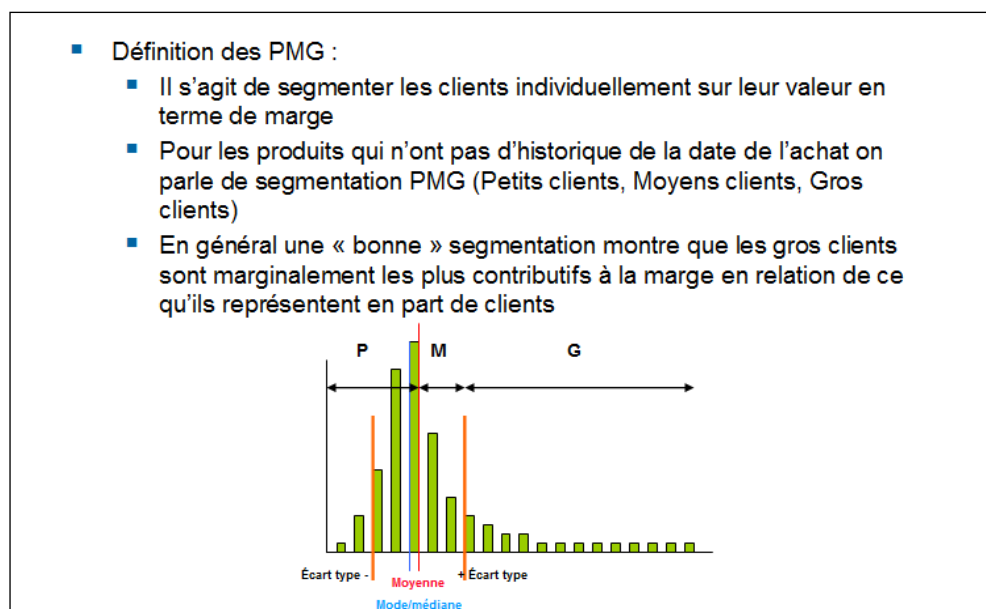


Figure 17 : Durée de vie du client et évolution de sa valeur

Le G est le client le plus élastique en général, il est gros consommateur de la marque mais aussi de la catégorie. Le taux de nourriture de la marque pour laquelle il est déjà un gros consommateur est souvent

faible, ce qui lui garantit une bonne élasticité. Il a un ensemble de marques qu'il achète régulièrement ; ce qui lui confère une forte probabilité de reporter sa décision d'achat vers les produits de la marque (Promotion et choix du point de vente, Volle, p 222).

En général, la cible d'un programme de fidélisation porte sur les gros clients de la marque quelle cherche à fidéliser. Les Gros clients sont en général les plus rentables (Kotler et Dubois) et aussi les plus élastiques. La perte d'un Gros client a un fort impact sur la marge de la marque.

L'exclusif à la marque est le plus souvent un P (petit client) marché présentant moins de potentiel de passage à l'acte.

Il faut que la base de données soit composée de 50% de ces Gros clients cible pour avoir une rentabilité.

En général, la base de données recrute le marché, pour obtenir 50% de Gros clients, il faut appliquer des règles de gestion et procéder au cleaning des consommateurs non G.

On peut optimiser à la marge le recrutement d'individus dans la cible en se concentrant sur les sources qui ramènent le plus de G. Malgré tout, cette optimisation reste difficile et on recrute finalement le marché.

Il est impossible d'avoir au recrutement 100% d'individus dans la cible.

Il est difficile de supprimer du programme (opération de nettoyage de la base de données) dès le départ, les clients hors cible qui ont manifesté par adhésion volontaire un intérêt pour le programme. Les supprimer, dès l'adhésion, suscite chez eux une frustration forte.

En-deçà de ce minimum de 50% d'individus dans la cible (des G), le programme n'aura pas suffisamment d'actifs (personnes achetant des produits en plus).

Évidemment, avec le temps, on peut se concentrer sur les sources qui apportent le plus de G au programme mais on a du mal à obtenir mieux que la structure du marché.

C'est donc par le cleaning des inactifs en général non G que l'on atteindra le taux des 50% de G.

Le taux de G étant directement corrélé à l'activité (l'achat), il est très important de se le fixer comme objectif et de l'atteindre.

Des sources trop intrusives et des moyens de recrutement essentiellement basés sur des mécaniques promotionnelles font baisser ce taux, elles sont donc à éviter. Sans suffisamment de G dans le programme de fidélisation (mauvais ciblage) celui-ci n'aura jamais de ROI positif.

Après étude des données « marché » via un panel de consommateurs, on peut connaître les caractéristiques qui déterminent un gros consommateur. En général, ce sont deux variables comportementales : la fréquence d'achat, le nombre de produits de la marque déclarée achetée et une variable socio démographique : la taille du foyer (Promotion et choix du point de vente, Volle, p 222).

Pour l'automobile, les variables sont assez différentes mais correspondent au même concept, deux variables comportementales : la marque et le modèle de véhicule possédé, le kilométrage parcouru par an et le nombre de véhicules dans le foyer.

Dans la distribution alimentaire c'est la fréquence de visite et la taille du foyer qui déterminent le plus la propension à se fidéliser de façon rentable pour l'enseigne.

Prises seules, les variables socio démographiques sans variable comportementale sont assez peu prédictives de la propension à consommer.

Le taux d'activabilité

Rappelons la définition du réactif d'ordre 2 : un réactif d'ordre 2 a réagi au moins une fois après son adhésion.

Selon Manuel et Xardel, à minima 50 % de l'ensemble des individus dans la base devront réagir une seconde fois après être rentrés dans la base de données. Ce seront les réactifs d'ordre 2. 50% est le minimum nécessaire pour assurer suffisamment de réactivité de rang supérieur : 50 % de réactifs d'ordre 2 assurent 60 % de réactifs d'ordre 3 (parmi les réactifs d'ordre 2), 70 % des réactifs d'ordre 3 seront réactifs d'ordre 4, et ainsi de suite. On retrouve ce raisonnement dans le cas du recrutement d'abonnés à des magazines, ou dans les comportements vépécistes. Manuel et Xardel définissent cela comme une échelle d'attrition. En fait, ce qui nous intéresse ici, c'est la rétention de clients actifs (réactifs).

Evolution du taux d'activabilité :

	Effectif des réactifs	Taux de réactifs	Taux de réactifs sur la base de départ
Réaction d'ordre 1 (adhésion)	10 000	100%	100%
Réaction d'ordre 2	5000	50%	50%
Réaction d'ordre 3	3000	60%	30%
Réaction d'ordre 4	2100	70%	21%
Réaction d'ordre 5	1680	80%	17%
Réaction d'ordre 6	1512	90%	15%
Réaction d'ordre 7	1361	90%	14%
Réaction d'ordre 8	1 225	90%	12%
Réaction d'ordre 9	1102	90%	11%
Réaction d'ordre 10	992	90%	10%

Ce tableau indique le taux de réactif à un programme de fidélisation au fil des réactions dans le temps. Il provient de l'analyse des réactifs en base de données d'une marque de grande consommation.

Tableau 67 : Evolution du taux d'activabilité

L'objectif d'activabilité ne peut être atteint que si les répondants ont adhéré volontairement au programme.

Les G sont en général sur représentés dans les réactifs d'ordre 2.

L'atteinte du taux d'activation souhaité s'obtient en faisant au consommateur des offres transactionnelles (BR : Bons de Réduction, boutique...) et des offres relationnelles (jeux, parrainage...) une fois son adhésion volontaire obtenue. Le mix des offres doit être particulièrement bien géré pour maintenir le consommateur actif le plus longtemps possible sans entamer le capital de marque : le rendre sensible au prix par de trop fortes offres promotionnelles par exemple.

Taux de captation d'achat à la marque

La notion de « culture intensive » : sur les consommateurs actuels de la marque, accroissement des quantités en valeur absolue et en part de marché dans le set des marques consommées.

Dans la théorie du comportement de l'acheteur :

Les acheteurs ont un set de marques difficilement modifiable, ceci est démontré dans la théorie de l'ensemble de considération d'Howard.

Selon Howard en moyenne (et bien que cela soit très différent en fonction des catégories de produit) un consommateur a 7 marques dans son ensemble de considération.

Le but des actions de fidélisation est de rendre la marque plus présente dans l'ensemble de considération au détriment des autres. En complément de l'ensemble de considération, on trouve la notion d'ensemble déterminant. L'ensemble déterminant est un sous-ensemble de l'ensemble de considération qui ne comporte que les marques achetées, le consommateur n'achetant pas toutes les marques de son ensemble de considération.

L'action de fidéliser à la marque correspond donc à rendre les marques plus présentes positivement dans l'ensemble de considération, ce qui augmenterait leur probabilité d'être également plus présentes dans l'ensemble déterminant.

En recrutant via un coupon porté sur ou dans l'emballage du produit, distribué en magasin ou directement au dos du bons de réduction portant sur le produit lui-même, on est sûr que la marque est dans l'ensemble déterminant du client.

Howard distingue les marques d'achat fréquent, d'achat routinier et d'achat ponctuel, la recherche d'information n'étant pas de même intensité.

Les consommateurs ont des comportements d'achat inertiels, ils ne modifient pas facilement leur set de marques (ils achètent presque toujours les mêmes marques) et n'accroissent pas ou ne diminuent pas leurs quantités de façon forte (ils achètent presque toujours les mêmes quantités). On retrouve cette inertie dans les travaux de Pierre Volle (Promotion et choix du point de vente, Volle, p 221) vis-à-vis des enseignes qui montrent que les efforts promotionnels ont du mal à affecter cette probabilité de choix du point de vente.

Pour cette raison d'inertie, on ne peut capter par la fidélisation sur un grand nombre de clients « que » 20 % en moyenne des quantités faites à la concurrence.

Ce taux de 20% de captation n'est possible que si le taux de nourriture est inférieur ou égal à 50%.

Au-delà de 20 % de captation, l'acheteur commencerait à percevoir un effort (une modification de ses habitudes de consommation). Cette modification devient alors très improbable, le client ayant du mal à modifier son ensemble déterminant.

Elasticité et taux de captation :

	P	M	G
Produits achetés en plus/élasticité 20%	7,0	8,4	10,9
Nouvelle QA Marque X achetée par acheteur	8,4	12,5	22,8
Nouvelle QA autre marques achetée par acheteur	28,2	33,4	43,6
Nouveau taux de nourriture	23,1%	27,1%	34,3%

Tableau 68 : Elasticité et taux de captation

Ce tableau est un exemple de données issues d'un panel scannérisé pour une marque de produit de grande consommation.

Maintien du coût unitaire de fidélisation

Le coût unitaire de fidélisation est déterminé une fois l'objectif de couverture de cible et la taille de base atteinte. Ce coût doit s'amortir sur la marge des achats supplémentaires atteints par la captation.

Le coût unitaire d'abord élevé dans un premier temps, doit arriver à l'objectif déterminé par la base en going. Ensuite, ce coût doit être impérativement maintenu. Une accélération ou une baisse de performance des autres objectifs peut faire remonter ce coût de façon symptomatique.

Axe temps :

L'atteinte des 4 objectifs ci-dessus doit se faire dans une dimension temporelle « descente » de retour sur investissement pour l'entreprise. En général 3/4 ans pour un programme de fidélisation grand public sur un produit de grande consommation. Pendant cette période et après une phase de tests, les coûts unitaires doivent tendre vers ceux fixés une fois la couverture de cible et la taille de base atteintes.

Cette dimension est une dimension induite dans Objectif Fidélisation[®] (notre premier modèle décrit dans la partie 4 du présent document). Cette dimension temporelle est prise en compte dans la projection de la durée du programme.

On comprend bien l'influence d'études de mesures fiables pour déterminer les variables ci-dessus :

- des études portant sur de grandes cohortes de consommateurs (méthode de l'individu « moyen » à un âge « moyen »).
- des études ayant une composante temporelle directe, la mesure est issue d'un panel où la mesure porte sur des individus d'âges différents permettant de retracer une courbe de vie théorique.

Pour la LTV, le second modèle présenté, la dimension temporelle est omniprésente.

La complémentarité des deux approches passe par une démonstration que les deux notions se complètent.

La notion de produits clefs d'entrée et de produits supports :

	Pénétration	PVC	Indice pénétration	Indice PVC	Indice composite
Produit 1	22,0%	3,4 €	253,7	64,6	159,1
Produit 2	18,4%	2,5 €	212,1	44,2	128,2
Produit 3	7,6%	8,6 €	87,6	150,6	119,1
Produit 4	5,5%	9,1 €	63,4	158,9	111,1
Produit 5	10,2%	5,4 €	117,6	94,3	105,9
Produit 6	9,6%	4,7 €	110,7	82,1	96,4
Produit 7	3,5%	7,7 €	40,4	134,8	87,6
Produit 8	4,4%	4,4 €	51,1	76,9	64,0
Produit 9	2,7%	5,6 €	31,1	98,7	64,9
Produit 10	2,8%	5,4 €	32,3	94,9	63,6
Moyenne	8,7%	5,7 €	100,0	100,0	100,0

Ce tableau représente les données issues d'un panel scanné pour une marque de produit de grande consommation.

Tableau 69 : La notion de produits clefs d'entrée et de produits supports

Les produits clefs d'entrée sont les produits qui présentent à la fois une bonne pénétration (forte probabilité d'achat fréquent) et une bonne marge (forte contribution marginale au programme), ils servent alors de support au recrutement. Les autres produits sont des produits présentant moins simultanément ces deux caractéristiques. Ce sont alors des produits « invités » dans le programme.

Cette notion de « produits fidélisant » est plus marginale dans la modélisation, peu d'auteurs y font référence, pourtant son impact est fort sur le ROI d'un programme de fidélisation.

Un modèle de la valeur client pour l'entreprise :

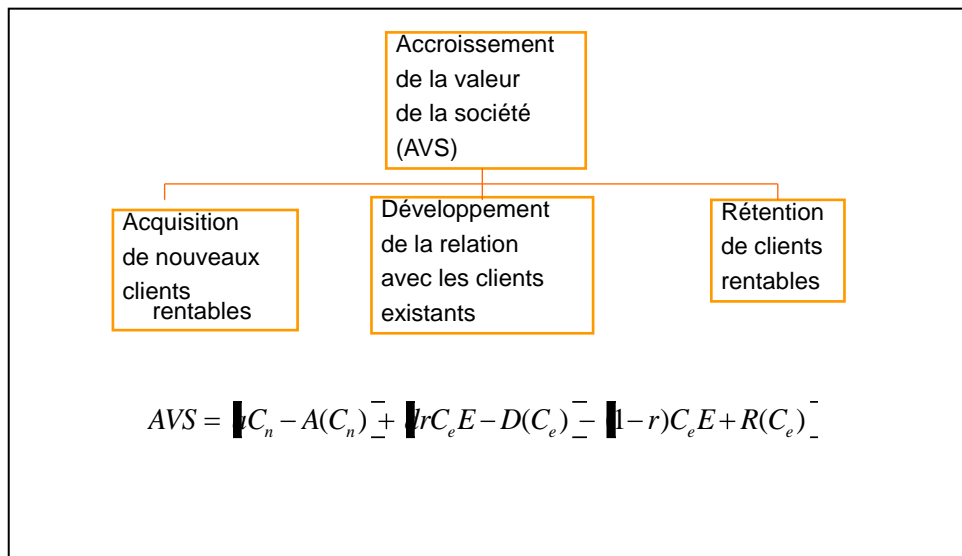


Figure 18 : Modèle de détermination d'accroissement de la valeur du portefeuille client

On retrouve dans la formule ci-dessus les trois dimensions du développement de la propension à consommer. Il porte sur une fidélité à une enseigne mais s'applique facilement à une marque et à un instant t. Les trois dimensions sont les « effets » attendus du mix marketing.

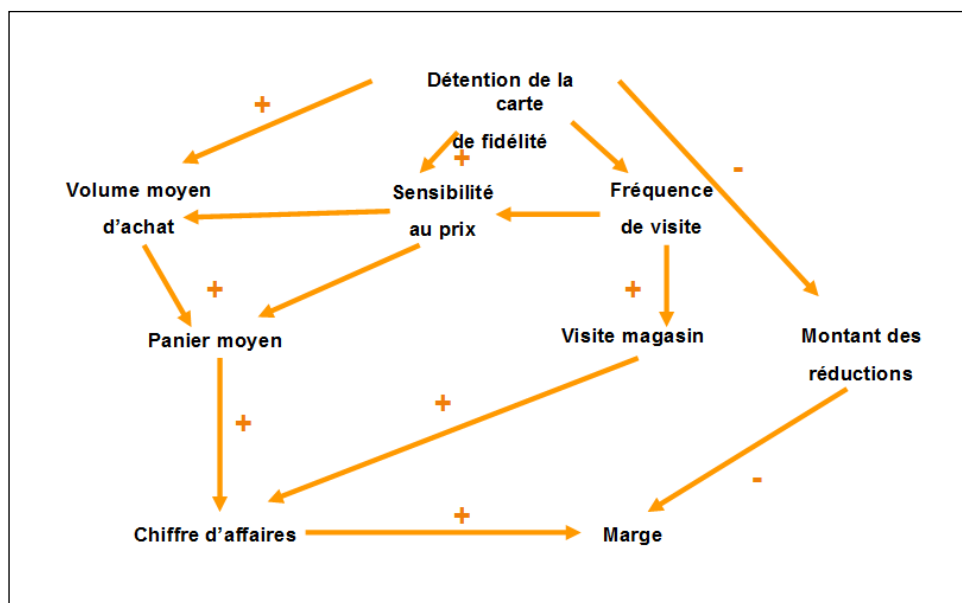


Figure 19 : Les effets de la fidélisation

Ce modèle comportemental explicite cette évolution de consommation. Il a été repris par Bénavent (Bénavent et Crié, 1998). On retrouve le schéma « classique » lié à la fidélisation. On entrevoit dans chaque phase où se situent les différentes composantes présentes précédemment dans les figures 10 et 11.

Après avoir défini les concepts de notre recherche et les différents modèles conceptuels et mathématiques associés, nous allons préciser notre problématique.

2.31 Définition de nos questions de recherche

La question fondamentale de notre recherche est : peut-on trouver un lien entre Objectif Fidélisation[®] et une approche par la LTV en définissant un modèle composite ?

Au-delà de ce lien, les deux approches sont-elles-dissociables ? Sous quelles conditions ?

Sont-elles obligatoirement complémentaires ou voire induites l'une en l'autre ?

Peut-on véritablement créer un modèle composite de ces deux approches ?

Nous émettons quatre questions de recherche qu'il nous faut formuler en hypothèses que nous essaierons de valider.

Pour nous, un modèle « composite » est bien l'étroite imbrication de modèles liés dont l'un peut découler de l'autre. Nous émettons des questions de recherche sous la forme des quatre hypothèses suivantes :

- **Q1 : On retrouve les mêmes composants (hors le temps) entre Objectif Fidélisation[®] et une approche par la LTV**

Objectif Fidélisation[®] utilise les mêmes variables que les modèles de LTV mais calcule la valeur du client sur une période définie et courte (12,24, 36 mois). Par définition un modèle de LTV calcule la valeur du client sur sa durée de vie totale, qui est une somme de courtes périodes de consommation. De par ses origines basées sur des données de panels de Home Scanning (ou les études de marché sur de grands échantillons mais sur une période donnée comme SIMM de Kantar) il est normal que Objectif Fidélisation[®] travaille sur des données « saturées » dans le temps.

- **Q2 : Seule la prise en compte du temps dans la durée de vie différencie Objectif Fidélisation[®] d'une approche par la LTV**

La LTV utilise la durée de vie totale du client, la durée de vie totale est divisée en périodes de courtes durées, c'est la somme de ces périodes.

- **Q3 : A un instant t, la modélisation de la LTV sur année moyenne donnée est égale au résultat d'Objectif Fidélisation[®]**

Le tout (la LTV) est bien la somme des parties (la valeur sur une courte période)

- **Q4 : On ne peut pas faire un calcul de ROI du programme de fidélisation sans prendre en compte la durée de la relation.**

Dans les approches basées sur des panels de Home Scanning et/ou études de marché, il faut donc bien tenir compte de la durée de vie du consommateur sur sa durée totale pour estimer la valeur client.

Nous allons présenter les deux modèles séparément Objectif Fidélisation[®] dans un premier temps, puis l'approche par la LTV à partir de deux exemples concrets. Ensuite, nous présenterons un modèle composite sur cas réels en faisant ressortir les éléments de validation de nos quatre hypothèses de recherche que nous testerons.

2.32 Modélisation du roi de la fidélisation à court terme (Objectif fidélisation[®])

Dans la partie du document qui suit nous présentons une première approche empirique de la modélisation du ROI de la fidélisation à court terme.

Pour la modélisation suivante, nous utilisons des données issues de Home*Scan de Nielsen. Ces données portent sur un cas réel « anonyme ».

Rappel sur Home*Scan de Nielsen

Le panel Home*Scan de Nielsen :

8 000 foyers représentatifs de la population française qui scannent tous les achats (codes-barres) des produits qu'ils achètent en GMS. Ce système garanti :

- Une donnée de ventes hebdomadaire non déclarative
- Une représentativité de la distribution
- Un historique très important

Le questionnaire média du panel Home*Scan de Nielsen :

Sur ce panel, Nielsen pose deux fois par an un questionnaire « médias » sur l'exposition de ses foyers aux mass médias (TV, Radio, Presse Mag, PQR, PQN, Affichage, Cinéma) construits dans les standards de questions des études spécifiques à ces médias.

- Plus de 70% des interrogés répondent de façon exploitable à ce questionnaire.
- Les résultats de ce questionnaire sont utilisés dans plusieurs outils de référence du marché : Scan Planning d'EMAP, Home*Scan Média de Carat, France Télévision.

Home*Scan est une forme de single source permettant d'avoir à la fois l'exposition aux médias, les achats des produits de grande consommation.

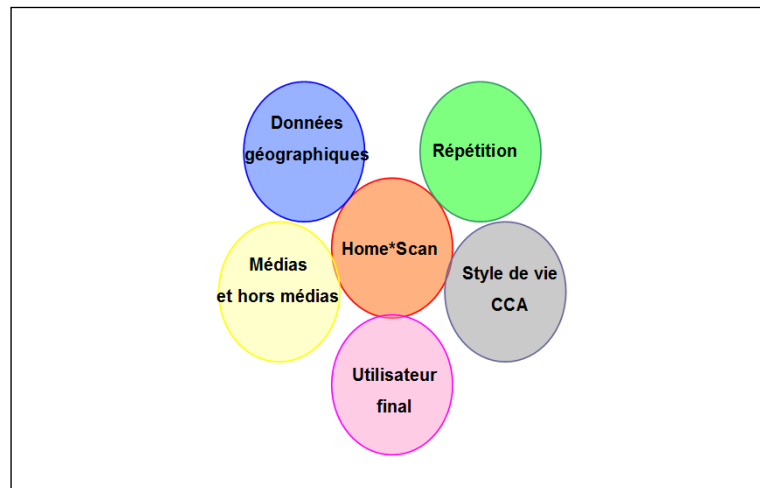


Figure 20 : Les utilisations de Home Scan

Source des informations : Nielsen

Le graphique ci-dessus montre ce qu'est une « single source ». Sur un même individu dans un panel on va collecter à la fois, sa socio démographie, son style de vie, sa consommation média, ses données d'achat, qui a consommé le produit dans le foyer, les données promo, etc.....

2.33 L'analyse structurelle du marché

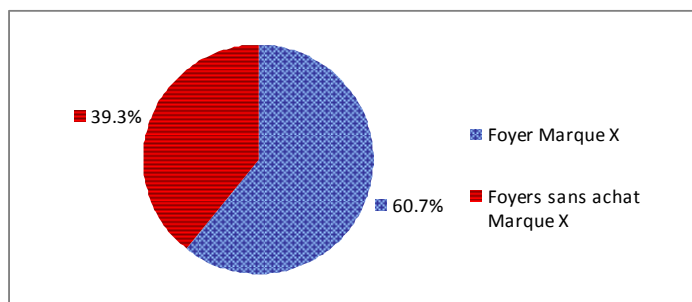
La pénétration de la Marque X

La marque X est une marque leader dans le secteur des produits d'hygiène/beauté.

La population des clients de la marque X est de 14 452 768 foyers actifs

	Pénétration Concurrents	Pénétration Marque X	NA Marque X extrapolés
Marché Hygiène Beauté	39,3%	60,7%	14 452 768

Tableau 70 : La pénétration de la Marque X



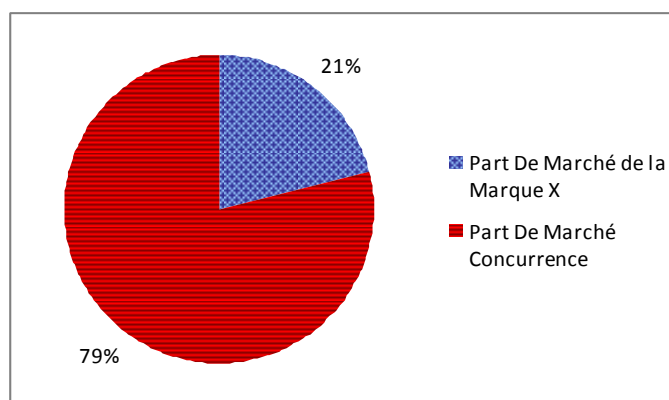
Graphique 17 : La pénétration de la Marque X

La pénétration de la Marque X sur le marché des produits d'Hygiène Beauté est de 60,7%. C'est donc une véritable marque de masse présente dans les 2/3 des foyers français. En pénétration, elle dépasse son principal concurrent de 6,4 points.

2.34 La part de marché de la Marque X (en valeur)

Part De Marché de la Marque X	Part De Marché Concurrence
21,4%	79,6%

Tableau 71 : Parts de marché



Graphique 18 : Parts de marché

La part de marché de la Marque X sur le marché des produits d'Hygiène Beauté est de 21,4%.

2.35 La notoriété de la Marque X

Néanmoins la Marque X a une faible notoriété spontanée (29%), ce qui peut être handicapant pour la fidélisation.

2.36 Les cibles à fidéliser

Un des objectifs de cette étude est de déterminer la ou les cible(s) d'individus à fidéliser. Les cibles à fidéliser doivent avoir les caractéristiques suivantes :

1. Un nombre d'achat et un panier moyen suffisamment élevé pour qu'ils ne perçoivent pas négativement l'effort des achats supplémentaires à la marque (attrition des achats à la concurrence)
2. Une élasticité théorique importante
3. Une forte probabilité d'acheter davantage de produits par acte d'achat.

2.36.1 Le nombre d'actes d'achat par acheteur

	TOTAL Marque X		
	P	M	G
TOTAL Marque X	1,6	4,2	9,8
Moyenne	5,2		

Tableau 72 : Le nombre d'actes d'achat par acheteur

Exemple de lecture : Les Gros consommateurs ont un nombre d'actes d'achat de 9,8 pour la marque Marque X.

2.36.2 Les quantités achetées de produits d'Hygiène Beauté par acheteur

	TOTAL HYGIENE BEAUTE			
	P	M	G	Moyenne
TOTAL Marque X	1,8	5,1	13,7	6,9
AUTRES Marques	36,4	44,5	57,1	46,0
TOTAL MARCHE	38,2	49,6	70,8	

Tableau 73 : Nombre de produits

Exemple de lecture : les Gros consommateurs achètent 13,7 produits de la Marque X contre 57 produits d'autres marques. Ils achètent 7,6 fois plus de produits de la Marque X que les Petits consommateurs et 2,7 fois plus que les Moyens.

2.36.3 Les sommes dépensées pour les produits d'hygiène beauté par acheteur

	TOTAL Marque X		
	P	M	G
TOTAL Marque X	7,65 €	20,52 €	49,71 €
Moyenne pondérée	19,33 €		

Tableau 74 : Sommes dépensées

Exemple de lecture : Le Gros consommateur dépense en moyenne 49,71 € en produits de la Marque X. Ils ont un nombre d'actes et une somme dépensée beaucoup plus importants que les Petits et les Moyens (6,1

fois plus d'actes que les Petits et 2,3 fois plus que les Moyens ; une somme dépensée 6,5 fois élevée à celle des Petits et 2,4 fois plus élevée que les Moyens). L'effort perçu d'actes en plus est donc moindre pour ces Gros consommateurs que pour les Petits et les Moyens.

2.37 Le taux de nourriture

	TOTAL Marque X		
	P	M	G
TOTAL Marque X	4,6%	10,3%	19,4%
TOTAL Concurrence	95,4%	89,7%	80,6%
TOTAL	100,0%	100,0%	100,0%

Tableau 75 : Le taux de nourriture de la Marque X par segment PMG

Les Gros consommateurs de la Marque X sont relativement élastiques avec un taux de nourriture relativement faible de 19,4% (80,6 % à la concurrence). Ils peuvent donc encore augmenter leur consommation en produits de la Marque X de façon significative.

Les Petits et les Moyens ont de faibles taux de nourriture de la Marque X, associés à sommes dépensées beaucoup plus faibles (cibles de conquête)

2.38 Le nombre d'acheteurs

Extrapolation à la population française (base : 23 810 161 foyers)

	TOTAL Marque X		
	P	M	G
TOTAL Marque X	7 659 967	4 003 417	2 789 384

Tableau 76 : Le nombre d'acheteurs

Il y a 2 789 384 foyers « Gros consommateurs » de la Marque X soit 19,3% des consommateurs de la Marque X.

2.39 La quantité achetée par acte d'achat

	Nbre d'actes d'achat Marque X		
	P	M	G
TOTAL Marque X	1,1	1,2	1,4

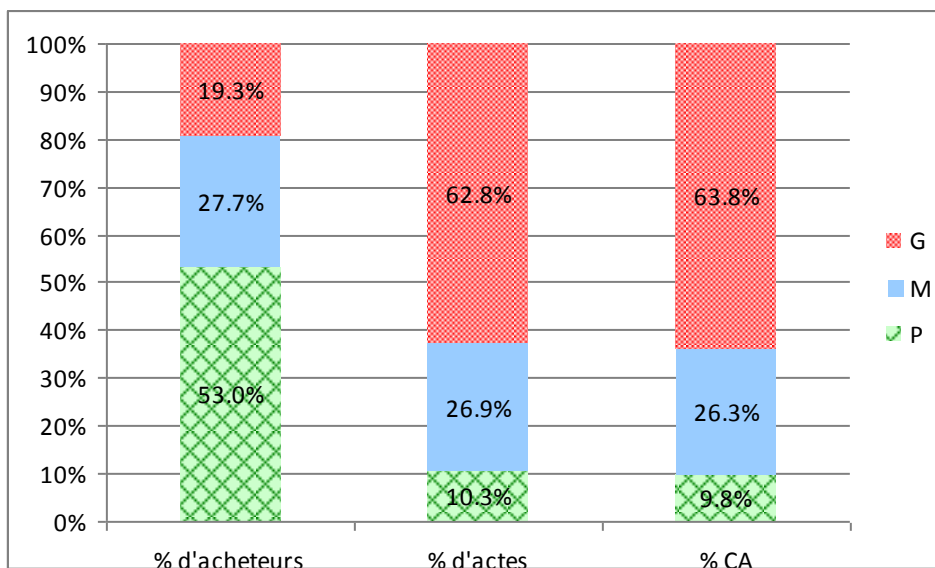
Tableau 77 : La quantité achetée par acte d'achat

Un Gros consommateur de la Marque X achète en moyenne 1,3 fois plus de produits de la Marque X par acte d'achat qu'un Petit consommateur et 1,2 fois plus de produits de la Marque X par acte d'achat qu'un consommateur Moyen.

	% d'acheteurs	% d'actes	% CA
P	53,0%	10,3%	9,8%
M	27,7%	26,9%	26,3%
G	19,3%	62,8%	63,8%

Tableau 78 : Les taux d'acheteurs, d'actes et de CA par segment PMG

Les Gros consommateurs représentent 19,3% des acheteurs de produits de la Marque X, 62,8% des actes d'achat en produits de la Marque X, 63,8% du chiffre d'affaires et probablement plus de 70 % de la marge.



Graphique 19 : Répartition des taux d'acheteurs, d'actes et de CA par segment PMG

Le Gros consommateur est le cœur de cible d'un programme de fidélisation.

	P	M	G
Nombre d'acheteurs	7 659 967	4 003 417	2 789 384
QA Marque X par acheteur	1,8	5,1	13,7
QA autres marques par acheteur	36,4	44,5	57,1
Nombre d'actes	1,6	4,2	9,8
Somme dépensée	7,65 €	20,52 €	49,71 €
Taux de nourriture	4,6%	10,3%	19,4%
Quantité achetée par acte d'achat	1,1	1,2	1,4

Tableau 79 : Tableau récapitulatif

Les Gros consommateurs de la Marque X présentent les caractéristiques énoncées plus haut :

- un nombre de produits achetés de la Marque X de 13,7
- supérieur à celui des Moyens (5,1) et des Petits (1,8).
- un taux de nourriture plus élastique (TN : 19,4%) que les Moyens (TN : 10,3%) et les Petits (TN : 4,6%).

- une propension à acheter plus de la Marque X par acte d'achat (1,4) que les Moyens (1,2) et les Petits (1,1). produits

Le Gros consommateur de la Marque X est plus apte à passer à l'acte que le Petit ou le Moyen consommateur, c'est à dire qu'il a plus de chance d'acheter des produits de la Marque X supplémentaires. Le panier moyen d'un Gros consommateur est élevé, ce qui lui permet de relativiser l'effort à fournir pour consommer davantage.

Les Petits et Moyens consommateurs ne présentent pas ces caractéristiques et ne seront pas intégrés systématiquement au programme de fidélisation.

En revanche, les Petits et Moyens consommateurs peuvent néanmoins être des cibles de conquête pour la Marque X (achat incrémental par le switch vers la Marque X). Ils seront présents dans la base de données et "cleanés" en fonction des règles de gestion préétablies.

Le Gros consommateur est le cœur de cible d'un programme de fidélisation.

	P	M	G
Produits achetés en plus/élasticité 20%	7,3	8,9	11,4
Nouvelle QA Marque X achetée par acheteur	9,1	14	25,1
Nouvelle QA autres marques achetée par acheteur	29,1	35,6	45,7
Nouveau taux de nourriture	23,8%	28,2%	35,5%

Tableau 80 : Elasticité de 20%

Avec une élasticité moyenne observée de 20%, le programme de fidélisation permet d'obtenir les résultats suivants : les Gros consommateurs de la Marque X achètent 25 produits, contre 14 pour les Moyens et 9 pour les Petits.

2.40 Les marques « clef d'entrée »

Cette partie étudie au sein des marchés porteurs, les marques contribuant le plus à l'efficacité économique du programme.

	Pénétration	PVC	Indice pénétration	Indice PVC	Indice composite
Sous Marque A	22,0%	3,69 €	253,1	64,6	158,9
Sous Marque B	18,4%	2,33 €	211,7	40,9	126,3
Sous Marque C	7,6%	8,60 €	87,4	150,6	119,0
Sous Marque D	5,5%	9,07 €	63,3	158,9	111,1
Sous Marque E	10,2%	5,38 €	117,3	94,3	105,8
Sous Marque F	9,6%	4,10 €	110,4	71,9	91,2
Sous Marque G	3,5%	7,68 €	40,3	134,8	87,5
Sous Marque H	4,4%	4,39 €	51,0	76,9	64,0
Sous Marque I	2,9%	5,63 €	33,4	98,7	66,0
Sous Marque J	2,8%	5,41 €	32,2	94,9	63,6
Moyenne	8,7%	3,96 €	100,0	100,0	100,0

Tableau 81 : Rappel des pénétrations

Dans le tableau ci-dessus : La pénétration, le PVC (Prix de Vente Consommateur) sont transformés en indices par rapport à la moyenne. L'indice composite est la moyenne des deux indices.

Les marques intéressantes comme support de recrutement pour le programme de fidélisation doivent avoir les caractéristiques suivantes :

- Une pénétration sur le marché supérieure à la moyenne (8,7%)
- Un PVC supérieur à la moyenne 3,96€

Par ailleurs le produit doit pouvoir porter un questionnaire de recrutement. En tenant compte de ces caractéristiques, les marques les plus intéressantes pour le programme de fidélisation, sous réserve qu'elles puissent comporter un questionnaire, sont :

- Sous Marque A
- Sous Marque B
- Sous Marque C
- Sous Marque D
- Sous Marque E (dans une moindre mesure)

Ces marques sont les marques supports de recrutement et seront systématiquement mises en avant.

Les autres marques ne présentent pas ces deux caractéristiques simultanément, elles seront des marques invitées dans le programme pour maximiser les taux de remontées des BR.

2.41 Taille de la cible

Le nombre de foyers français est de 23 810 161 foyers d'après le recensement de 1999.

Les acheteurs de produits de la Marque X représentent 60,7% de la population totale soit 14 452 767 foyers.

Sur le marché de l'Hygiène Beauté, les Gros consommateurs de la Marque X représentent 19,3 % des acheteurs de la Marque X soit 2 789 384 de foyers.

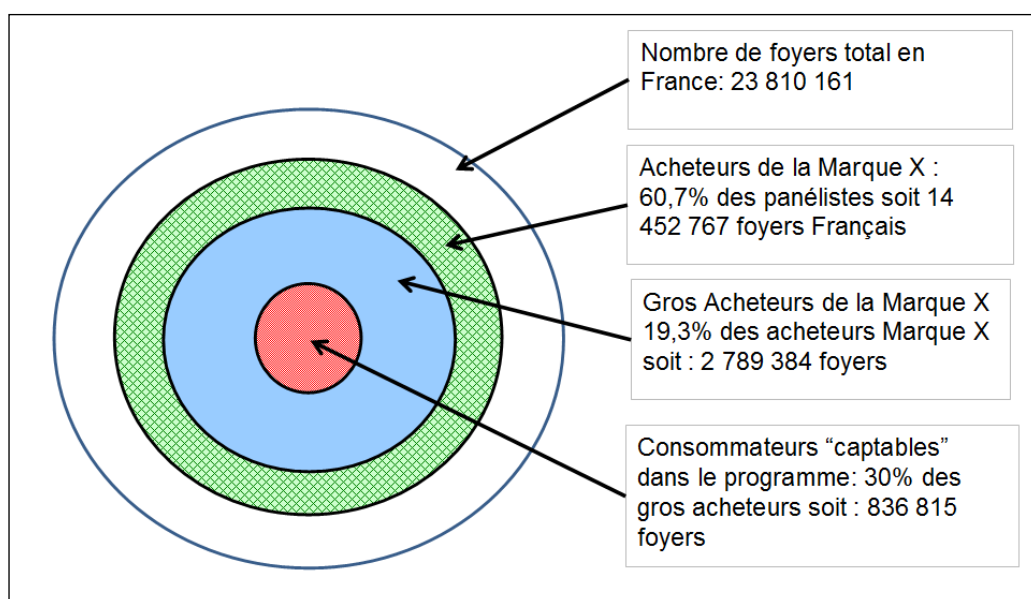


Figure 21 : Taille de la cible

Nombre de foyers total en France: 23 810 161

Acheteurs de la Marque X : 60,7% des panélistes soit 14 452 767 foyers français

Gros Acheteurs de la Marque X : 19,3% des acheteurs Marque X, soit : 2 789 384 foyers

Consommateurs "captables" dans le programme: 30% des gros acheteurs, soit 836 815 foyers.

La cible à fidéliser

30 % des G sont potentiellement captables dans le programme soit 836 815 foyers.

La structure de la base de données :

Sur ces 2 789 384 foyers, 30% sont "captables", soit 836 815 foyers Gros consommateurs. Ces Gros consommateurs représentent 70% de la base des foyers captés. La taille du programme optimisé est donc de 1.200 000 foyers.

En-deçà de ce taux de 30% de recrutement de notre cœur de cible, le programme n'a pas d'effet volume.

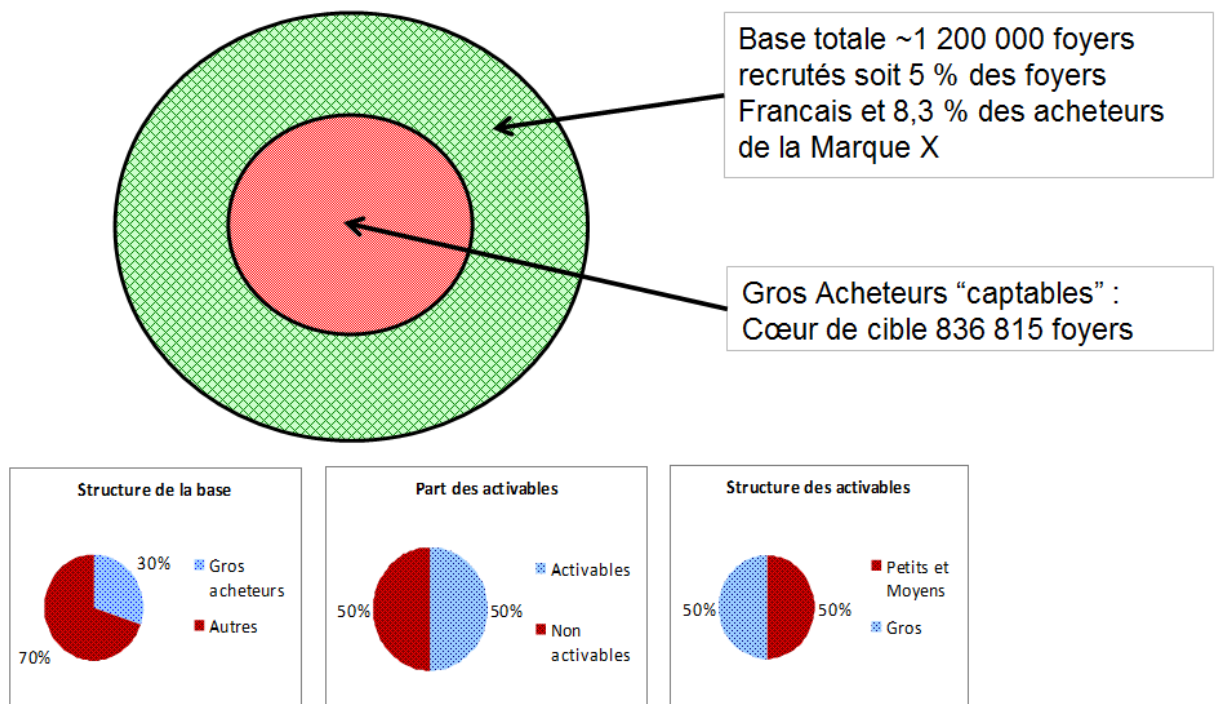


Figure 22 : Structure de la base de données

Nota : On ne peut pas améliorer de façon significative le recrutement en amont dans un programme de fidélisation, on recrute le marché :

- programme sur des produits de grande consommation : le questionnaire auto administré ramène des PMG dans les mêmes proportions que la structure du marché, cas ci-dessus, 30% de G vs 70% de P+M (autres).

- programme de grand distributeur : on pourrait être segmentant à la caisse mais pour des raisons logistiques et un effet de taille nécessaire à un feedback positif du réseau cela est rarement fait. Le distributeur recrute son marché.
- programme sur des services, banque, téléphonie : là les annonceurs pourraient être segmentant, mais en réalité la médiatisation de ces programmes les ouvre en général au large public, client de ces enseignes. C'est donc par le nettoyage de la base de données au fil de l'eau que le programme pourra viser sa cible.

2.42 Tableaux de simulation

PARTIE COÛT

	Année 1	Année 2	Année 3	Année 4	Année 5	Année 6	Année 7	Année 8
RECRUTEMENT	200 000	200 000	200 000	200 000	200 000	200 000	200 000	200 000
CUMUL BASE	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066
BUDGET RECRUTEMENT	548 816 €	548 816 €	548 816 €	548 816 €	548 816 €	548 816 €	548 816 €	548 816 €
COÛT UNITAIRE RECRUTEMENT	2,74 €	1,44 €	1,01 €	0,80 €	0,67 €	0,59 €	0,53 €	0,48 €
FIDELISATION								
CREATION + FRAIS TECHNIQUES 1 MESSAGE	38 112 €	38 112 €	38 112 €	38 112 €	38 112 €	38 112 €	38 112 €	38 112 €
PRODUCTION 1 MESSAGE	105 190 €	199 861 €	285 064 €	314 563 €	374 579 €	428 588 €	429 479 €	468 853 €
COMPENSATION BR (15% à 0,93 €)	27 898 €	53 007 €	75 604 €	95 942 €	114 246 €	130 719 €	145 546 €	158 889 €
AFFRANCHISSEMENT	82 322 €	156 413 €	223 094 €	283 110 €	337 119 €	381 156 €	429 479 €	468 853 €
COÛT DU MESSAGE	253 523 €	447 392 €	621 875 €	731 727 €	864 056 €	978 576 €	1 042 616 €	1 134 708 €
COÛT DU PROGRAMME n MESSAGES	3	760 568 €	1 342 176 €	1 865 624 €	2 195 182 €	2 592 169 €	2 935 727 €	3 127 847 €
COÛT UNITAIRE DE FIDELISATION	3,80 €	3,53 €	3,44 €	3,19 €	3,16 €	3,13 €	3,00 €	2,99 €
GESTION DE LA BASE	54 882 €	73 176 €	106 714 €	121 959 €	137 204 €	152 449 €	152 449 €	152 449 €
ETUDES	38 112 €	38 112 €	38 112 €	38 112 €	38 112 €	38 112 €	38 112 €	5 810 €
PLATEFORME TELEPHONIQUE	0 €	0 €	0 €	0 €	0 €	0 €	0 €	0 €
HONORAIRES AGENCES	27 889 €	27 889 €	27 889 €	34 861 €	34 861 €	34 861 €	39 509 €	39 509 €
FRAIS FIXES DU PROGRAMME	120 883 €	139 177 €	172 715 €	194 932 €	210 177 €	225 422 €	230 070 €	197 768 €
FRAIS FIXES UNITAIRES	0,60 €	0,37 €	0,32 €	0,28 €	0,26 €	0,24 €	0,22 €	0,17 €
COÛT UNITAIRE ANNUEL	4,41 €	3,90 €	3,76 €	3,48 €	3,42 €	3,37 €	3,22 €	3,16 €

Tableau 82 : Tableau de simulation : partie coût

Le tableau ci-dessus est un tableau de coût en fonction des volumes de clients fidélisés par année. Il dissocie les coûts par grands postes. Il multiplie les coûts des messages par le nombre de messages prévus. Ensuite il calcule le coût unitaire de fidélisation par poste.

PARTIE MARCHÉ

Marque X	Petits	Moyens	Gros	TOTAL / Moy
% DES ACHETEURS	53	27,7	19,3	100
% DES ACHATS Marque X EN VOLUME	18,7	28,3	53	100
BUDGET ANNUEL CONSACRE A LA CATE. DE PROD.	162,36 €	199,56 €	256,88 €	206,26 €
NB DE PRODUITS ACHETES / AN	38,2	49,6	70,8	52,87
BUDGET ANNUEL CONSACRE A Marque X	7,62 €	20,58 €	49,70 €	19,36 €
NB DE PRODUITS Marque X ACHETES / AN	1,8	5,1	13,7	7
MARGE BRUTE	3,02 €	8,55 €	22,97 €	11,59 €

Part d'activables 50%

Tableau 83 : Tableau de simulation : partie marché

Le tableau ci-dessus permet d'analyser la rentabilité actuelle des clients par PMG, il alloue une part de la marge sur les produits que le client pourrait consommer en plus via l'élasticité de son taux de nourriture.

EVOLUTION DE LA MARGE BRUTE EN FONCTION DES QUANTITES

Gros consommateurs	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066
	15,05 €	17,22 €	17,91 €	18,42 €	18,65 €	18,79 €	19,02 €	19,11 €
Moyens consommateurs	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066
	0,63 €	2,80 €	3,39 €	4,00 €	4,22 €	4,37 €	4,60 €	4,69 €
Petits consommateurs	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066
	-4,91 €	-2,73 €	-2,04 €	-1,54 €	-1,31 €	-1,16 €	-0,94 €	-0,85 €

Tableau 84 : Evolution de la marge brute en fonction des quantités

Le tableau ci-dessus croise les coûts par les gains en fonction de l'évolution par année de la taille totale de la base du programme

SIMULATION SUR LES GROS CONSOMMATEURS

BASE	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066
ACTIVABLES	100 000	190 000	271 000	343 900	409 510	468 559	521 703	569 533
BUDGET TOTAL	1 585 317 €	2 185 219 €	2 742 206 €	3 132 734 €	3 544 966 €	3 917 499 €	4 126 388 €	4 402 666 €
BUDGET REEL	1 585 317 €	2 185 219 €	2 742 206 €	3 132 734 €	3 544 966 €	3 917 499 €	4 126 388 €	4 402 666 €
COÛT UNITAIRE SUR LES ACTIVABLES	15,85 €	11,43 €	10,06 €	9,15 €	8,69 €	8,38 €	7,93 €	7,77 €
NOMBRE DE PRODUITS EN PLUS	9	7	6	5	5	5	5	5

Tableau 85 : Simulation sur les gros consommateurs

Le tableau ci-dessus est le tableau des résultats finaux qui indique qu'en fonction des coûts de fidélisation, des données marchés et hypothèses, quel est le nombre de produits à vendre en plus pour que le programme soit rentable.

Sur la base d'une marge théorique de 1,7 euros pour un produit en plus (moyenne des marges), il faut de 5 à 9 produits en plus pour amortir unitairement le programme sur 50% de foyers réactifs. La rentabilité s'amplifie avec les amortissements liés au temps : 5 produits en going.

Le nombre de messages est de 3 en moyenne :

- 1 à 2 pour les consommateurs peu ou pas intéressants (50% de la cible retenue)
- 3 pour les consommateurs Cœur de cible (50% de la cible retenue)
- Certains consommateurs Cœur de cible pourront recevoir 2 messages, 4 pour d'autres
- Ponctuellement, en fonction des opportunités promotionnelles des marques, des messages promotionnels mono-marque (hors programme) pourront être envoyés sur la base, notamment les Petits consommateurs et les Moyens consommateurs.

Le coût sur les activables est de 7,6 euros pour un gain théorique de 19,2 euros, soit 11,5 euros de bénéfice.

Notre objectif de captation sur les Gros est de 5 achats sur 57,1 achats faits à la concurrence.

La rentabilité du programme semble facilement atteignable.

On est rentable dès 200 000 individus en année 1 avec 13,2 € de marge allouable pour 7,9 € de coûts sur les actifs.

L'objectif de 30% des G reste néanmoins important et s'obtient à environ 1,2 millions de personnes.

Par nettoyage systématique les G représentent 50% de la base. Pour identifier les G on utilise une carte typologique : elle pose des questions clefs (golden questions) : taille du foyer, nombre de produits achetés, fréquence d'achat.

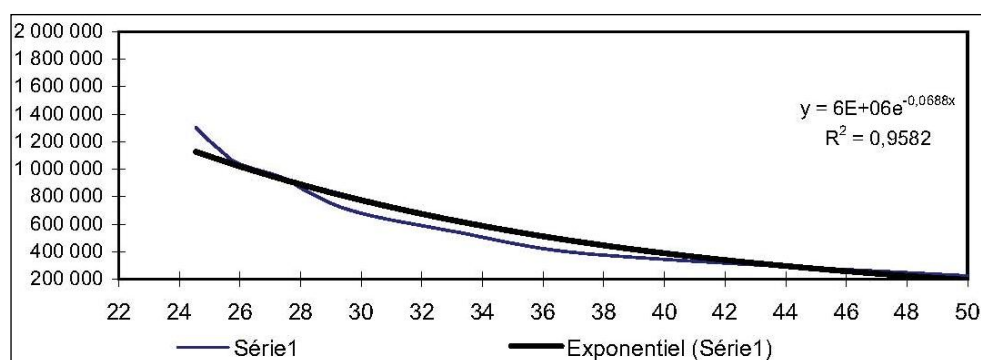
Analyse détaillées des contributions par PMG

GROS CONSOMMATEURS :

	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066	1 225 159	1 302 643
Cout Unitaire	7,93 €	5,75 €	5,06 €	4,56 €	4,33 €	4,18 €	3,95 €	3,86 €	3,80 €	3,74 €
Profit Unitaire	5,27 €	6,03 €	6,27 €	6,45 €	6,53 €	6,58 €	6,66 €	6,69 €	6,71 €	6,73 €

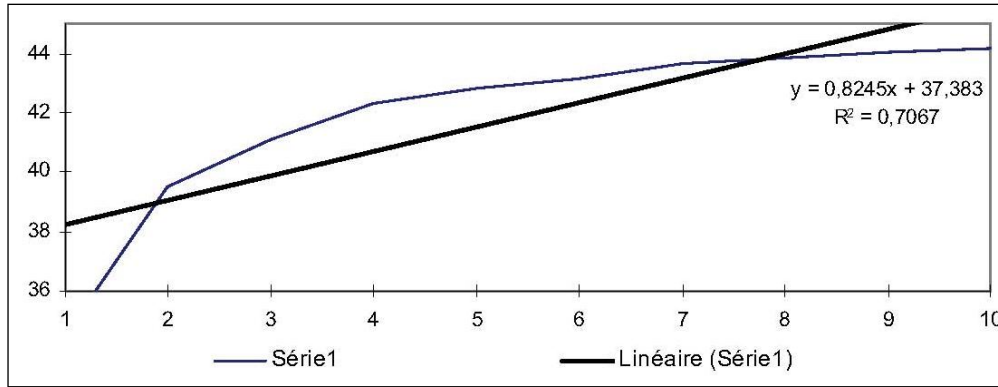
Tableau 86 : Zoom sur les contributions par PMG : gros consommateurs

Le tableau ci-dessus montre l'évolution des coûts et profits unitaires sur les clients G en fonction de la taille du programme.



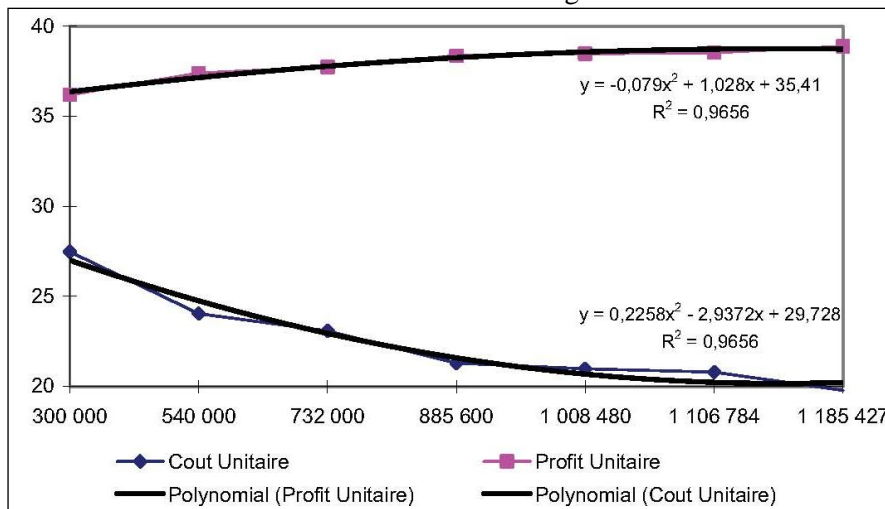
Graphique 20 : Gros consommateurs : nombre d'individus en fonction du coût unitaire

Le graphique ci-dessus montre sur les clients G la décroissance de coûts et la tendance de cette décroissance.



Graphique 21 : Gros consommateurs : profit unitaire en fonction du nombre d'individus

Le graphique ci-dessus montre sur les clients G la croissance de gains et la tendance de cette croissance.



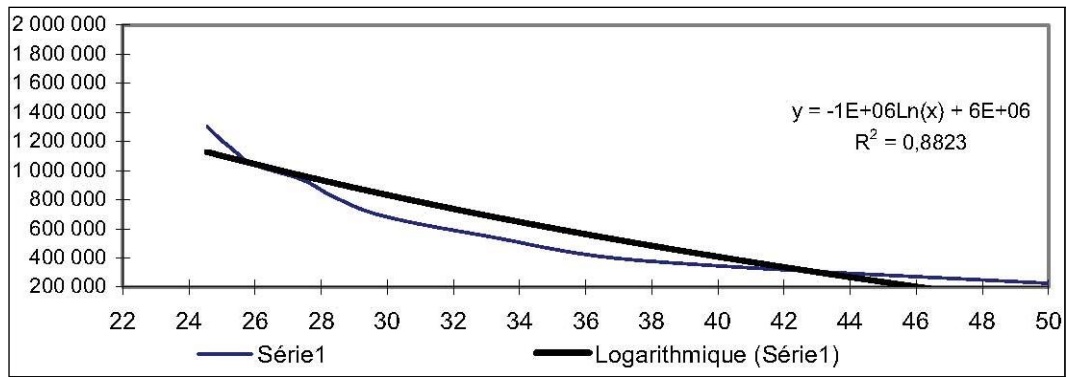
Graphique 22 : Gros consommateurs : coût unitaire et profit unitaire en fonction du nombre d'individus

Le graphique ci-dessus montre sur les clients G l'évolution des deux courbes, on voit qu'elles ne se croisent pas. Sur les clients G la rentabilité du programme est immédiate. Son gain unitaire couvre largement le coût unitaire de sa fidélisation. Ce gain croit avec le temps, les deux courbes restent tendanciellemment éloignées l'une de l'autre.

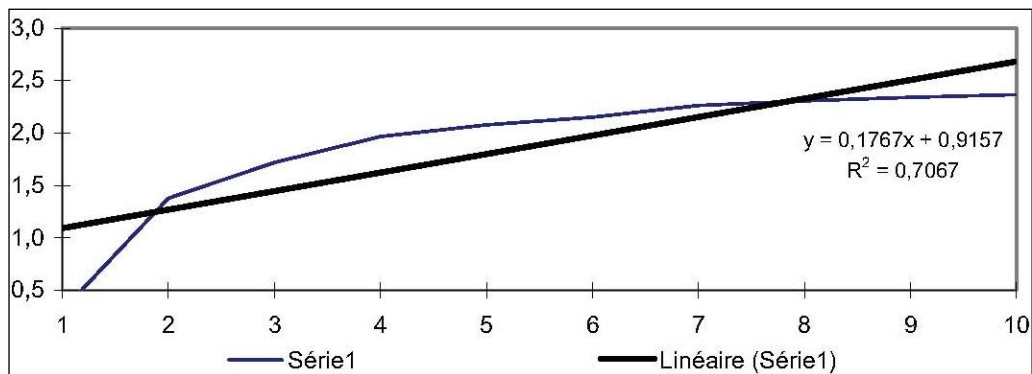
MOYENS CONSOMMATEURS :

	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066	1 225 159	1 302 643
Coût Unitaire	7,93 €	5,75 €	5,06 €	4,56 €	4,33 €	4,18 €	3,95 €	3,86 €	3,80 €	3,74 €
Profit Unitaire	0,05 €	0,21 €	0,26 €	0,30 €	0,32 €	0,33 €	0,34 €	0,35 €	0,36 €	0,36 €

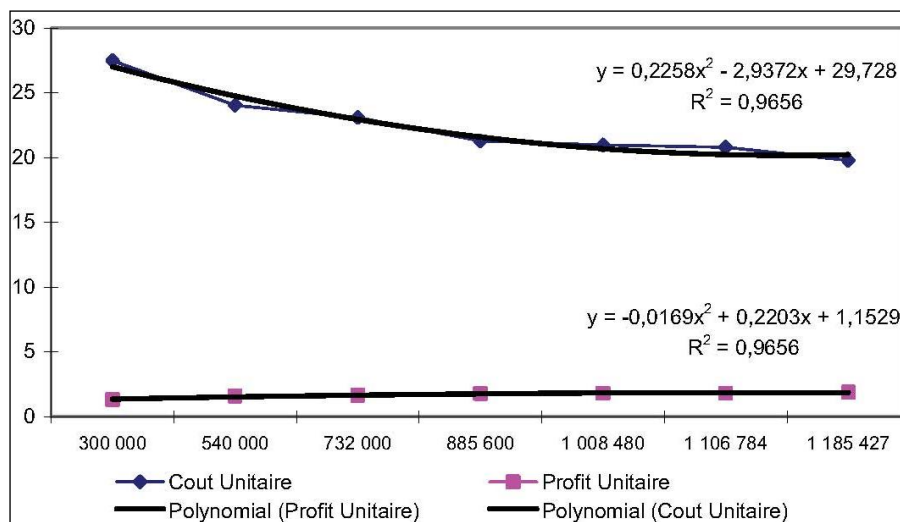
Tableau 87 : Zoom sur les contributions par PMG : Moyens consommateurs



Graphique 23 : Moyens consommateurs : profit unitaire en fonction du nombre d'individus



Graphique 24 : Moyens consommateurs : profit unitaire en fonction du nombre d'individus



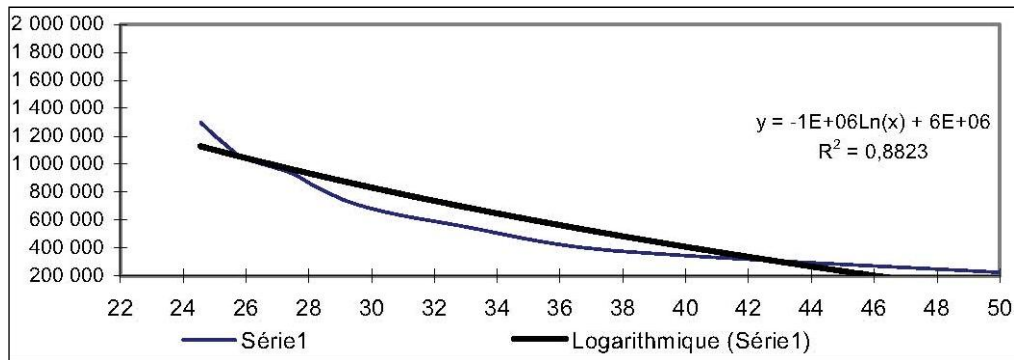
Graphique 25 : Moyens consommateurs : coût unitaire et profit unitaire en fonction du nombre d'individus

Le graphique ci-dessus montre que sur le client M, les deux courbes pourraient si elles se prolongeaient se croiser à un volume de clients fidélisés très important (hors du graphique), elles se rapprochent l'une de l'autre mais restent éloignées. La rentabilité est très lointaine sur le client Moyen. Les volumes de clients à fidéliser pour l'atteindre sont irréalisables.

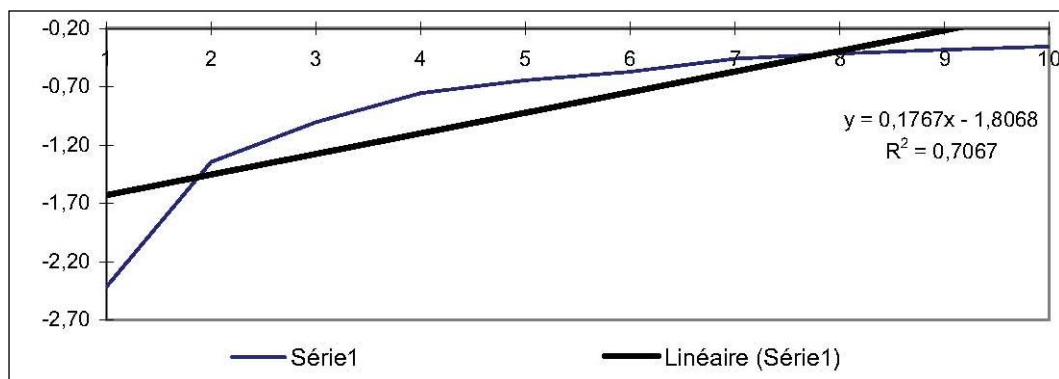
PETITS CONSOMMATEURS :

	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066	1 225 159	1 302 643
Cout Unitaire	7,93 €	5,75 €	5,06 €	4,56 €	4,33 €	4,18 €	3,95 €	3,86 €	3,80 €	3,74 €
Profit Unitaire	-0,37 €	-0,20 €	-0,15 €	-0,12 €	-0,10 €	-0,09 €	-0,07 €	-0,06 €	-0,06 €	-0,05 €

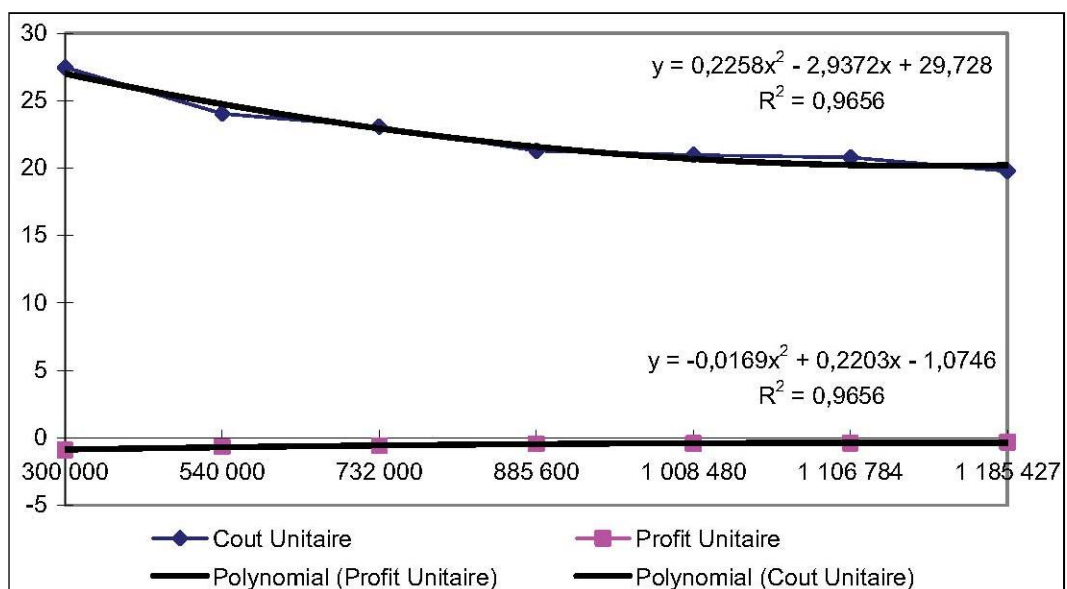
Tableau 88 : Zoom sur les contributions par PMG : Petits consommateurs



Graphique 26 : Petits consommateurs : profit unitaire en fonction du nombre d'individus



Graphique 27 : Petits consommateurs : profit unitaire en fonction du nombre d'individus



Graphique 28 : Petits consommateurs : coût unitaire et profit unitaire en fonction du nombre d'individus

Le graphique ci-dessus montre que sur le client P, les deux courbes ne se croiseraient pratiquement jamais. Fidéliser le client P vs son coût n'a pas de sens.

Simulation de montée en puissance du programme

En global

Possibilité de recrutement annuel	200 000
Possibilité de recrutement mensuel	20 000
Taux de recrutement mensuel	10,00%
Accroissement du programme par mois	20 000
Taille optimale du programme	1 200 000
Base de départ	50 000
Durée en mois	58
Taille du marché potentiel	23 810 161
Pénétration moyenne de la marque	60,70%
Clientèle de la marque	14 452 768
Part de la clientèle à fidéliser	8,30%

Tableau 89 : Simulation de montée en puissance du programme : au global

Le tableau ci-dessus est un récapitulatif des grands agrégats de la simulation.

Par année

Année 1		Année 2		Année 3		Année 4		Année 5	
Mois 1	70 000	Mois 13	310 000	Mois 25	550 000	Mois 37	790 000	Mois 49	1 030 000
Mois 2	90 000	Mois 14	330 000	Mois 26	570 000	Mois 38	810 000	Mois 50	1 050 000
Mois 3	110 000	Mois 15	350 000	Mois 27	590 000	Mois 39	830 000	Mois 51	1 070 000
Mois 4	130 000	Mois 16	370 000	Mois 28	610 000	Mois 40	850 000	Mois 52	1 090 000
Mois 5	150 000	Mois 17	390 000	Mois 29	630 000	Mois 41	870 000	Mois 53	1 110 000
Mois 6	170 000	Mois 18	410 000	Mois 30	650 000	Mois 42	890 000	Mois 54	1 130 000
Mois 7	190 000	Mois 19	430 000	Mois 31	670 000	Mois 43	910 000	Mois 55	1 150 000
Mois 8	210 000	Mois 20	450 000	Mois 32	690 000	Mois 44	930 000	Mois 56	1 170 000
Mois 9	230 000	Mois 21	470 000	Mois 33	710 000	Mois 45	950 000	Mois 57	1 190 000
Mois 10	250 000	Mois 22	490 000	Mois 34	730 000	Mois 46	970 000	Mois 58	1 210 000
Mois 11	270 000	Mois 23	510 000	Mois 35	750 000	Mois 47	990 000	Mois 59	1 230 000
Mois 12	290 000	Mois 24	530 000	Mois 36	770 000	Mois 48	1 010 000	Mois 60	1 250 000

Tableau 90 : Simulation de montée en puissance du programme : par année

Le tableau ci-dessus indique à quelle échéance le programme atteint sa taille critique pour être rentable compte tenu du recrutement mensuel et de l'attrition.

En résumé le programme parvient à maturité (30% de la cible des Gros soit 836 815 Gros sur une base de 1 200 000) en 58 mois. Il est rentable dès les 200 000 premiers individus en base, à 100% de marge allouable sur les produits supplémentaires consommés.

* en général une marque comme la Marque X avec 60,7% de pénétration nationale peut recruter 20 000 individus/mois.

2.43 ROI

Marge de 9,45 Euros

	Année 1	Année 2	Année 3	Année 4	Année 5	Année 6	Année 7	Année 8	Année 9	Année 10
Base	200 000	380 000	542 000	687 800	819 020	937 118	1 043 406	1 139 066	1 225 159	1 302 643
Activables	100 000	190 000	271 000	343 900	409 510	468 559	521 703	569 533	612 580	651 322
Marge générée par la fidélisation	945 184 €	1 795 849 €	2 561 448 €	3 250 487 €	3 870 623 €	4 428 744 €	4 931 054 €	5 383 132 €	5 790 003 €	6 156 187 €
Budget réel	1 585 317 €	2 185 219 €	2 742 206 €	3 132 734 €	3 544 966 €	3 917 499 €	4 126 388 €	4 402 666 €	4 651 316 €	4 875 102 €
Marge brute résiduelle	-640 133 €	-389 370 €	-180 757 €	117 753 €	325 657 €	511 246 €	804 666 €	980 466 €	1 138 687 €	1 281 085 €
% de marge brute résiduelle	-40,40%	-17,80%	-6,60%	3,80%	9,20%	13,10%	19,50%	22,30%	24,50%	26,30%

Tableau 91 : ROI avec marge de 9,45€

Le tableau ci-dessus calcule le ROI du programme compte tenu des coûts et des gains selon les hypothèses retenues par années. Il montre en quelle année le ROI est atteint.

En utilisant 4,72€ de marge (50 % de la marge disponible qui est de 9,15€ en année 5), la rentabilité intervient en année 4 mais s'accroît de façon significative rapidement.

2.44 Tableau de pilotage

Evaluation des objectifs définis en 2002 par Objectif Fidélisation®

	Phase test	Année 1	Année 2	Commentaires
Couverture de cible	30% de la cible soit 836 815 foyers clients			
Structure PMG Clients	70% de G soit 836 815 Gros consommateurs			
Réactivité d'ordre 2 membres	50%			
Coût unitaire	3,80 €			
	7,5 € sur les réactifs			

Tableau 92 : Tableau de pilotage

Le tableau ci-dessus récapitule les 4 grands agrégats de pilotage du programme à suivre par année en fonction des objectifs initiaux définis dans la phase test.

2.45 Conclusion

On retrouve tous les composants définis dans la première partie :

- couverture de cibles
- structure PMG à atteindre : activabilité
- maintien des coûts unitaires

Le fait de retrouver ces composants dans Objectif Fidélisation® est évident puisque le modèle a été conçu en les prenant pour paramètres. Evidemment, la difficulté est à la fois d'avoir une marque qui se prête à chacun d'entre eux et de les atteindre ensuite simultanément : modèle sous contraintes. Sans l'atteinte de ces objectifs, il n'y a pas de rentabilité escomptable de la fidélisation.

Une marque qui ne peut à priori pas les atteindre ne doit pas se lancer dans une démarche de fidélisation proactive de sa clientèle par un marketing direct sortant. On voit donc bien l'importance de définir ces objectifs au départ puis de les piloter dans la durée.

Le temps qui rapproche Objectif Fidélisation[®] de la LTV est induit dans Objectif Fidélisation[®]. «Induit», car on considère une année n moyenne, mais on simule la montée en puissance dans le temps, recrutement et accroissement des achats par captation à la concurrence. Le composant temporel est présent mais non modélisé sur la partie des achats incrémentaux. En effet, il est difficile de prévoir le comportement de l'acheteur dans le temps. Par contre le temps est bien pris en compte dans l'accroissement de la couverture de cible : la montée en puissance du programme. L'année n se perpétue « n » fois sur la durée du programme du client, les coûts remboursés décroissant jusqu'à une asymptote.

On entrevoit, par le composant temporel, le lien vers un modèle de LTV que nous allons définir maintenant, second modèle de notre recherche. Nous allons voir comment les composants présents dans Objectif Fidélisation[®] se retrouvent dans la LTV.

2.46 Modélisation du roi de la fidélisation à long terme (la détermination de la LTV)

A l'approche court terme du ROI de la fidélisation on peut proposer une modélisation tout autre sur la durée de vie du client.

Pour l'illustrer nous avons choisi dans cette partie un exemple dans le secteur bancaire à partir de l'étude SIMM de Kantar. Nous avons analysé sur la totalité de l'échantillon étudié de la vague 2000 toutes les données concernant la banque et les données sociodémographiques des individus répondants.

Rappel sur l'étude SIMM de Kantar :

Le principe de SIMM :

- Connaître les comportements de consommation et les expositions Média de la population française (étude Média/Produits).
- Étudier des habitudes de consommation individuelles – déclaratif
- Étudier des habitudes Média individuelles - déclaratif
- Fournir une base de données comportementales via la donnée individuelle, facile à traiter.
- Créer des études de ciblage et de profils tant marketing que media.
- SIMM est devenu l'étude de référence sur ces applications
- Avoir une approche comportementale et prospective.

Au départ SIMM a été importante en France dans les années 70 par Interdéco via une inspiration américaine, l'étude Média/Produits de Simmons. Depuis, elle a subi des évolutions notoires mais existe sur un historique important. Les données permettent facilement de remonter de 4 à 5 vagues en arrière, au-delà c'est plus fastidieux, les données étant sur des supports informatiques hétérogènes.

La méthode de SIMM :

- **Echantillon présélectionné, 10 000 individus** représentatifs des Français de 15 ans et plus

- **Questionnaire auto-administré** adressé par voie postale, taux de retour : 70%
 - 5 versions de questionnaires adaptés au statut de l'individu : Ménagères, Autres femmes, Chefs de foyer, Hommes seuls, Autres hommes
- **Terrain de l'enquête : mars-avril-mai**
- **Single source** : Toutes les questions recueillies simultanément dans un même questionnaire auprès du même individu
- **Calage des données presse** sur les études d'audience de référence C'est le plus gros échantillon en France sur ce type d'étude. On peut relativiser un peu le principe de single source car il s'agit de déclaratif. Les validations faites sur SIMM montrent que les individus sont néanmoins très cohérents dans leur déclaration.

Le contenu de SIMM :

Tous les secteurs de la consommation sont étudiés, ce qui rend cette étude utilisable tant pour les produits de grande consommation que pour les biens d'équipement et les services.



Figure 23 : L'étude SIMM

Ci-dessous une description de l'étude SIMM de Kantar :

- 31 secteurs étudiés -790 phrases d'attitudes, de styles de vie et d'opinions
- De nombreux thèmes abordés en termes de critères de choix
- Interrogation sur les processus d'achat, les attentes
- Les marques et les enseignes
- Déclinaison des marques ombrelles jusqu'aux références produits pour une analyse produit fine
- Présence des marques sélectives, de pharmacie, de vente directe
- Présence des marques de distributeurs (alimentation)
- Présence des marques atypiques (tendances bio, nature...)

Le niveau de détail permet relativement de descendre sur des marques ayant de faibles pénétrations (moins de 1% de pénétration).

Les usages marketing de SIMM :

Les usages des annonceurs

- Analyse de marché : services (banque, assurance), biens durables...
- Profil du consommateur du produit (versus acheteur au foyer)
- Vue « panoramique » de la consommation
- Identification d'une cible de consommation
- Segmentation de clientèle : styles de vie, typologie sectorielle
- Possibilité de travailler en évolution
- Les usages des médias ou des agences d'achat d'espace
- Ciblage
- Echelles de couverture vs coûts
- Média planning

Source des informations : Kantar Média

Nous avons vu que dans Objectif Fidélisation[®] l'axe temps est induit et que les composants définis pour déterminer la fidélisation sont présents. Nous allons décrire maintenant l'approche de la LTV et voir comment s'y appliquent certains des composants du modèle Objectif Fidélisation[®].

2.47 Durée de vie du consommateur

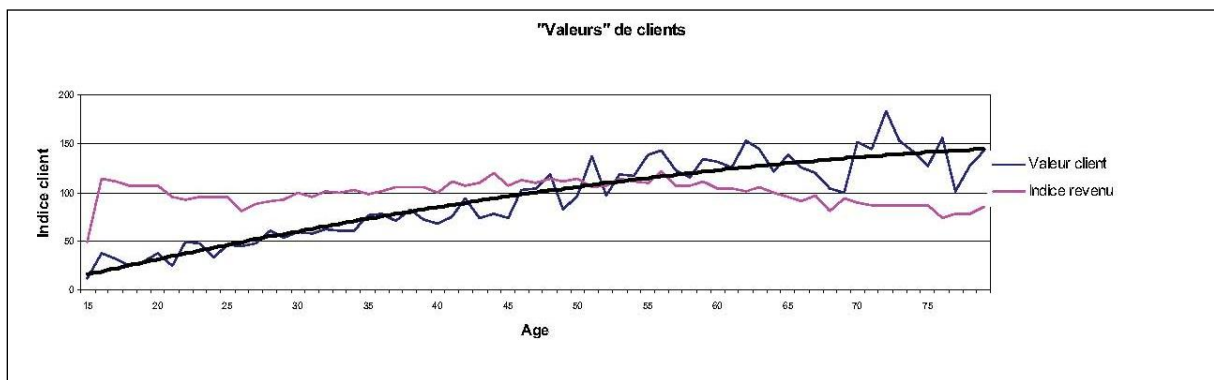
Age	Total	Somme moyenne remboursée par mois	Indice-prêt	Montant moyen des placements	Indice placement	Flux financier moyen	Valeur client	Revenu mensuel net du foyer	Indice revenu
15	2	0 €	0	6 860 €	14	6 860 €	12	1 067 €	49
16	200	788 €	96	12 685 €	26	22 137 €	38	2 454 €	113
17	130	1 063 €	130	6 301 €	13	19 061 €	32	2 400 €	111
18	140	762 €	93	5 519 €	11	14 666 €	25	2 315 €	107
19	134	489 €	60	11 331 €	23	17 196 €	29	2 291 €	106
20	134	791 €	97	12 311 €	25	21 802 €	37	2 285 €	106
21	112	446 €	55	9 504 €	19	14 855 €	25	2 066 €	95
22	134	1 000 €	122	16 986 €	35	28 981 €	49	1 990 €	92
23	136	876 €	107	17 087 €	35	27 596 €	47	2 055 €	95
24	149	665 €	81	11 176 €	23	19 150 €	33	2 062 €	95
25	132	694 €	85	18 681 €	38	27 006 €	46	2 052 €	95
26	115	887 €	108	15 714 €	32	26 359 €	45	1 739 €	80
27	130	865 €	106	17 797 €	36	28 182 €	48	1 908 €	88
28	163	740 €	90	26 191 €	53	35 072 €	60	1 956 €	90
29	163	793 €	97	21 461 €	44	30 982 €	53	1 993 €	92
30	165	826 €	101	24 687 €	50	34 602 €	59	2 171 €	100
31	170	947 €	116	22 597 €	46	33 965 €	58	2 058 €	95
32	175	810 €	99	26 784 €	55	36 500 €	62	2 181 €	101
33	185	875 €	107	25 106 €	51	35 610 €	60	2 163 €	100
34	164	845 €	103	25 963 €	53	36 099 €	61	2 203 €	102
35	185	855 €	105	34 371 €	70	44 631 €	76	2 119 €	98
36	254	882 €	108	35 260 €	72	45 845 €	78	2 193 €	101
37	242	839 €	103	31 227 €	64	41 297 €	70	2 265 €	105
38	193	851 €	104	38 093 €	78	48 305 €	82	2 264 €	105
39	202	858 €	105	32 148 €	65	42 445 €	72	2 262 €	105
40	180	863 €	106	29 615 €	60	39 975 €	68	2 171 €	100
41	195	863 €	106	33 749 €	69	44 107 €	75	2 412 €	111
42	204	906 €	111	43 723 €	89	54 596 €	93	2 303 €	106
43	149	828 €	101	33 276 €	68	43 208 €	73	2 387 €	110
44	138	863 €	106	35 120 €	72	45 475 €	77	2 585 €	119
45	135	856 €	105	32 760 €	67	43 035 €	73	2 290 €	106
46	120	882 €	108	49 713 €	101	60 297 €	102	2 427 €	112
47	127	794 €	97	50 914 €	104	60 439 €	103	2 365 €	19
48	144	857 €	105	59 149 €	120	69 433 €	118	2 466 €	114
49	107	935 €	114	37 205 €	76	48 427 €	82	2 404 €	111
50	130	825 €	101	47 183 €	96	57 080 €	97	2 439 €	113
51	146	815 €	100	70 707 €	144	80 490 €	137	2 282 €	105
52	172	785 €	96	47 079 €	96	56 494 €	96	2 310 €	107
53	193	828 €	101	59 569 €	121	69 500 €	118	2 441 €	113
54	167	808 €	99	58 688 €	120	68 381 €	116	2 400 €	111
55	171	881 €	108	70 432 €	143	81 001 €	138	2 387 €	110
56	107	867 €	106	73 553 €	150	83 962 €	143	2 625 €	121
57	147	695 €	85	64 339 €	131	72 683 €	123	2 297 €	106
58	127	627 €	77	60 018 €	122	67 547 €	115	2 309 €	107
59	130	713 €	87	70 295 €	143	78 855 €	134	2 412 €	111
60	122	784 €	96	67 727 €	138	77 140 €	131	2 226 €	103
61	137	793 €	97	64 163 €	131	73 673 €	125	2 231 €	103
62	159	669 €	82	81 825 €	167	89 852 €	153	2 193 €	101
63	168	734 €	90	75 731 €	154	84 535 €	144	2 282 €	105
64	150	644 €	79	63 535 €	129	71 267 €	121	2 150 €	99
65	162	622 €	76	73 860 €	150	81 319 €	138	2 066 €	95
66	208	645 €	79	65 734 €	134	73 468 €	125	1 978 €	91
67	202	738 €	90	61 568 €	125	70 424 €	120	2 085 €	96
68	194	496 €	61	55 362 €	113	61 319 €	104	1 753 €	81
69	188	545 €	67	51 734 €	105	58 276 €	99	2 034 €	94
70	192	767 €	94	79 620 €	162	88 819 €	151	1 932 €	89
71	167	693 €	85	76 293 €	155	84 613 €	144	1 876 €	87
72	155	868 €	106	97 555 €	199	107 972 €	183	1 869 €	86
73	149	588 €	72	83 053 €	169	90 110 €	153	1 893 €	87
74	138	495 €	61	77 358 €	158	83 304 €	141	1 864 €	86
75	139	464 €	57	69 359 €	141	74 926 €	127	1 863 €	86
76	131	535 €	65	84 724 €	173	91 141 €	155	1 589 €	73
77	113	659 €	81	51 831 €	106	59 738 €	101	1 690 €	78
78	108	449 €	55	69 608 €	142	74 992 €	127	1 669 €	77
79	82	415 €	51	78 678 €	160	83 654 €	142	1 834 €	85
Total	9 962	818 €	100	49 095 €	100	58 908 €	100	2 164 €	100

Tableau 93 : Durée de vie du consommateur

- **Total** : nombre de répondants par âge
- **Somme moyenne remboursée par mois** : montant moyen déclaré par âge de crédit remboursé mensuellement
- **Indice prêt** : montant des remboursements mensuels en indice par âge par rapport à la moyenne sur toutes les classes d'âges
- **Montant moyen des placements** : sommes moyennes déclarées placées par âge
- **Indice placement** : montant des placements par âge par rapport à la moyenne sur toutes les classes d'âges
- **Flux financier moyen** : montant moyen des flux financier déclarés par âge
- **Valeur client** : montant de la valeur client en indice par âge par rapport à la moyenne sur toutes les classes d'âges
- **Revenu mensuel du foyer** : revenu du foyer déclaré par mois
- **Indice de revenu** : montant du revenu mensuel du foyer en indice par âge par rapport à la moyenne sur toutes les classes d'âges

La « valeur client » du tableau ci-dessus est en fait le composant Q du modèle de Wayland et Cole. Aucun coût n'a pour l'instant été déduit. On calcule la valeur client (VC) sur une durée de vie moyenne.

Nous remarquons que la valeur du client est maximale pour un âge de 73 ans et minimale pour un âge de 16 ans. Elle a tendance à être croissante dans le temps.



Graphique 29 : "Valeur" de clients

Pour les 10 000 individus nous obtenons un âge renseigné et une consommation de produits bancaires. A un âge donné, nous pouvons calculer la valeur moyenne d'un client. Nous pouvons le faire pour tous les âges et tracer la courbe des valeurs moyennes constatées : la courbe du composant Q. On cherche à maximiser la valeur d'un client donné pour la banque étudiée en accroissant sa consommation.

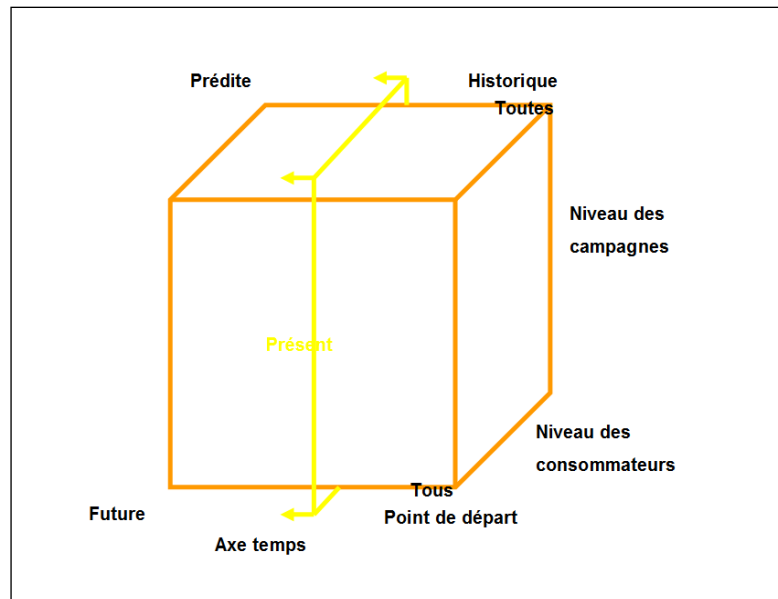


Figure 24 : Une représentation de la valeur client

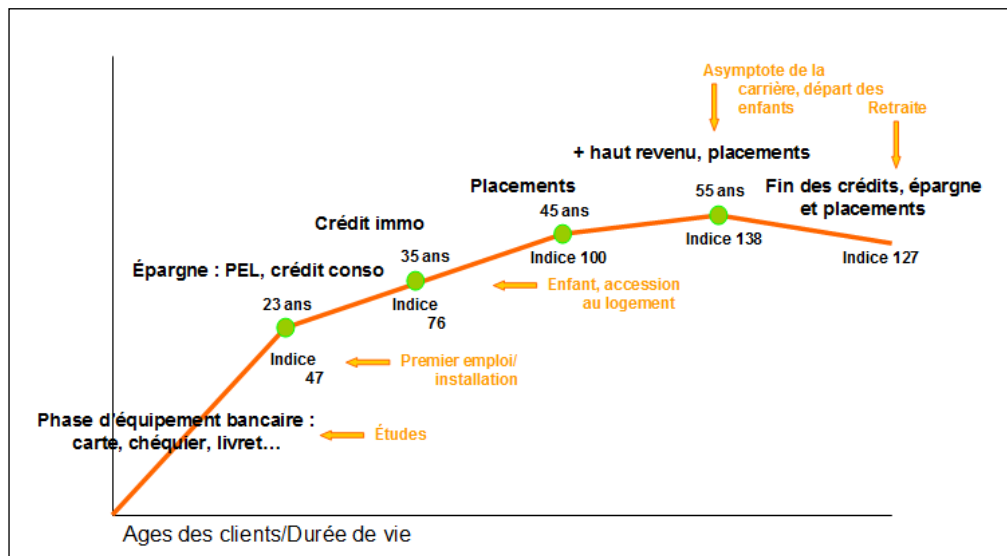
Nous pouvons ensuite estimer la droite de régression et la fonction. Cette fonction de régression simple est l'évolution de la valeur d'un client moyen pour une vie moyenne issue des courbes ci-dessous :

$$y = 2.0072x + 26.408$$

$$R^2 = 0.8431$$

L'âge moyen sur la durée de vie est le composant P (durée) du modèle de Wayland et Cole.

Lorsque l'on réfléchit sur cette courbe et qu'on la croise avec d'autres critères de consommation, on arrive à tracer un modèle explicatif assez cohérent des moments de vie d'un client.



Graphique 30 : Le cycle de vie d'un client bancaire

Ce graphique montre l'évolution de la valeur bancaire d'un client en fonction des grandes étapes de la vie liées à l'âge.

Une étude synchronique comme SIMM nous permet donc une diachronie c'est à dire une estimation moyenne de la durée de vie d'un client. La richesse de SIMM permet de le faire en :

- global marché,
- par produit bancaire,
- par enseigne,
- par segment PMG de client.

2.48 Captation à la marque et total marché :

Tous ces critères peuvent évidemment être croisés entre eux, on a donc une excellente source d'estimation de la valeur empirique marché de ces clients. On peut la comparer à d'autres études pour validation ou à des données issues de la base de données de la banque étudiée (par exemple pour déterminer le différentiel de valeur entre la banque étudiée et celle d'un client moyen marché). On appréhende bien la notion de taux de nourriture. Le taux de nourriture est la division de la consommation à la banque A sur le marché total soit : consommation totale de la banque A du client/consommation bancaire totale du client.

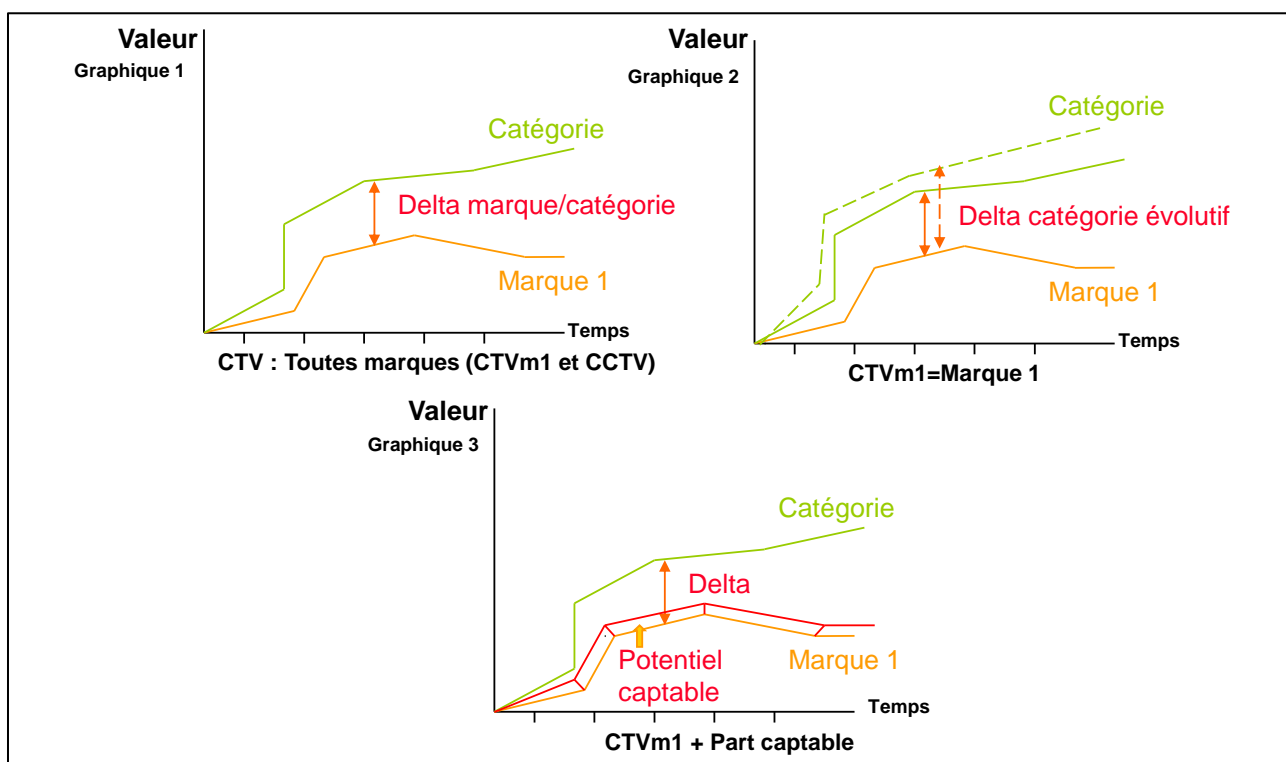


Figure 25 : LTV, delta marché et captation

Ce delta est important sur les gros clients car ceux-ci, on l'a vu dans la première partie, ont à la fois plus de chance de passer à l'acte et sont en général marginalement plus rentables. L'importance des G est le composant N de la version du modèle de LTV qui détermine les clients les plus rentables.

Evidemment, les Gros clients sont, dans ce cas, des clients très sollicités, souvent plus aguerris et donc plus difficiles à capter.

2.49 Calcul des élasticités

La notion de multibanquarité nous permet d'estimer une valeur marché total du client et sa valeur dans la banque étudiée, nous avons donc le taux de nourriture du client de la banque A.

On détermine alors de la même façon son potentiel de captation mais au lieu de le faire uniquement à un instant comme pour les produits de grande consommation, on le fait sur une durée de vie du client par segment P, M ou G.

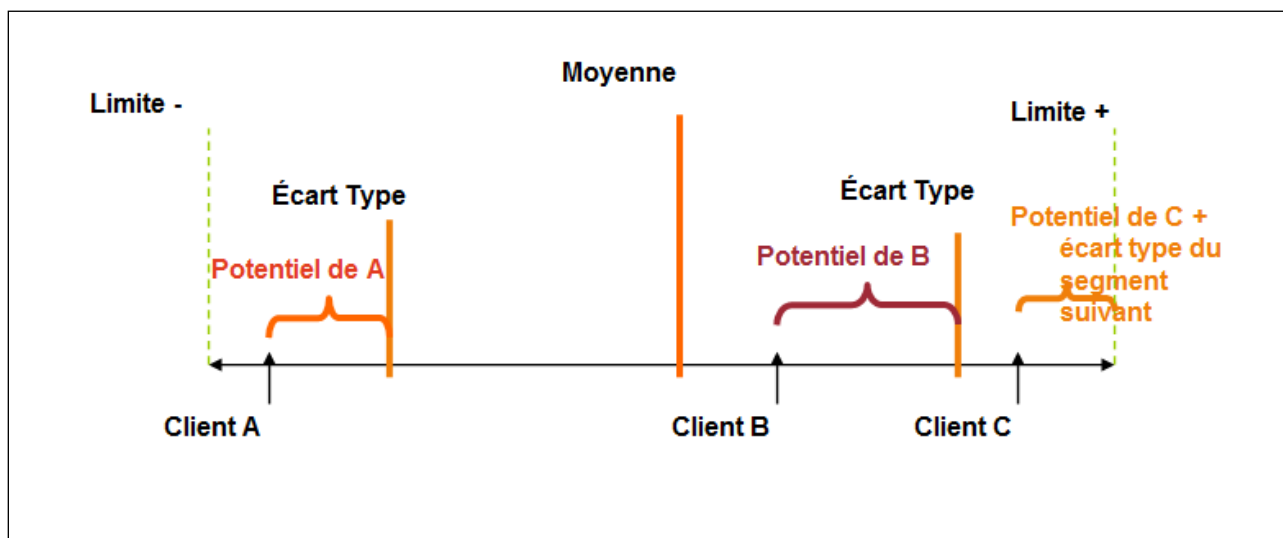


Figure 26 : Détermination des potentiels

On va donc appliquer le même taux de 20% par défaut sur le delta consommé à la concurrence. On retrouve le composant probabiliste P de la version « concurrentielle » du modèle de LTV :

$$VC = \sum_{t=1}^T P_t Q_t d_t$$

P étant en fait une probabilité que la relation se perpétue, associée à un accroissement des quantités.

Dans notre approche comme dans celle de Kim, Min et Shin nous utilisons, en général, un delta de 20% sur le delta de la PDM observée et de la PDM Valeur.

Dans cet exemple, sous SIMM comme il est facile de calculer le delta entre le marché et la banque étudiée sur l'ensemble de la durée, il est donc facile d'appliquer au moins linéairement pour l'instant le taux de captation de 20%. L'important c'est le delta entre la marque étudiée et le marché.

SIMM permet à la fois de le voir à un instant en fonction de l'âge mais aussi comme l'étude est annuelle elle permet d'analyser comment la concurrence évolue chaque année. Evidemment pour déterminer la captation, une bonne mesure de la pression concurrentielle est essentielle.

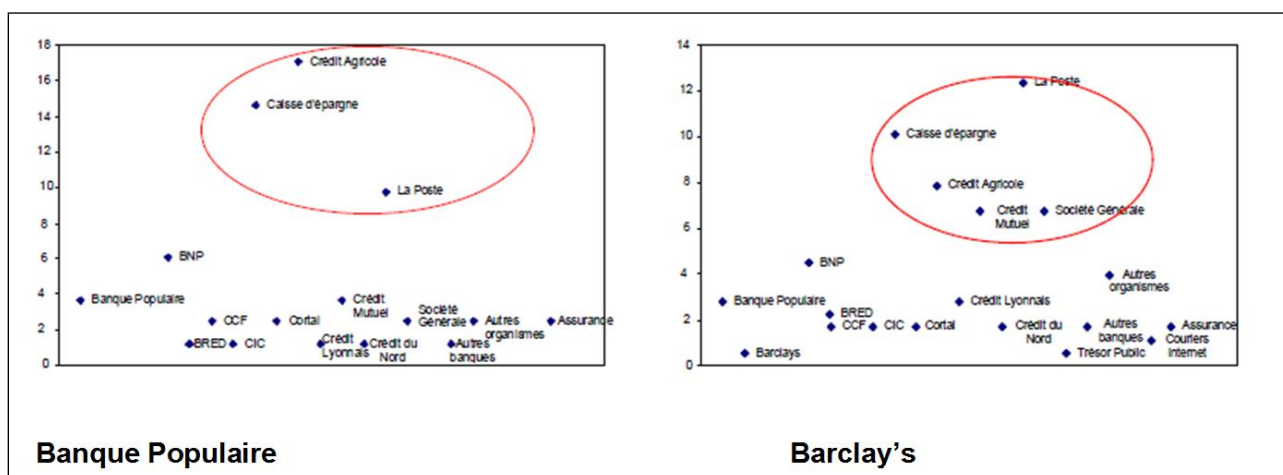


Figure 27 : Pression concurrentielle pour Banque Populaire et Barclay's

Nous avons, dans un modèle de LTV classique sur une donnée moyenne de consommation liée à un âge moyen, pu appliquer les composants P, M, G, captation d'Objectif Fidélisation[®]. Les autres composants : couverture de cible et réactivité sont des composants induits.

2.50 Conclusion

Le modèle de LTV utilise dans ses paramètres les mêmes indicateurs qu'Objectif Fidélisation[®]. Les difficultés rencontrées sont plus liées aux données qu'au modèle lui-même. SIMM nous apporte pratiquement toutes les données nécessaires pour le faire. La différence majeure est que pour les produits bancaires contrairement aux produits fmcg (fast moving consumer goods : produits de grande consommation), il faut obligatoirement le faire sur une durée de vie.

Les composants s'appliquent tous et pratiquement de la même façon. Seuls le taux de couverture de cible et la réactivité d'ordre 2 ne sont pas explicités, mais ils sont induits dans la LTV. Il faut évidemment un certain nombre de clients pour qu'une action de fidélisation entreprise à partir d'une modélisation de ROI sur la LTV ait une chance d'être positive.

On perçoit donc bien la complémentarité méthodologique des deux démarches, entre celle d'Objectif Fidélisation[®] exposée au départ et celle de la LTV. Nous allons maintenant faire un modèle test entièrement composite intégrant complètement les deux approches.

2.51 Convergence et complémentarité des modèles

Les données de l'exemple ci-dessous portent sur le marché du pneu mais ont été anonymées et transformées.

2.51.1 Mise en évidence des convergences

Il s'agit de modéliser le processus sur la LTV complète du produit étudié, on voit qu'il est finalement assez proche de celui d'Objectif Fidélisation[®] dont il s'inspire largement.

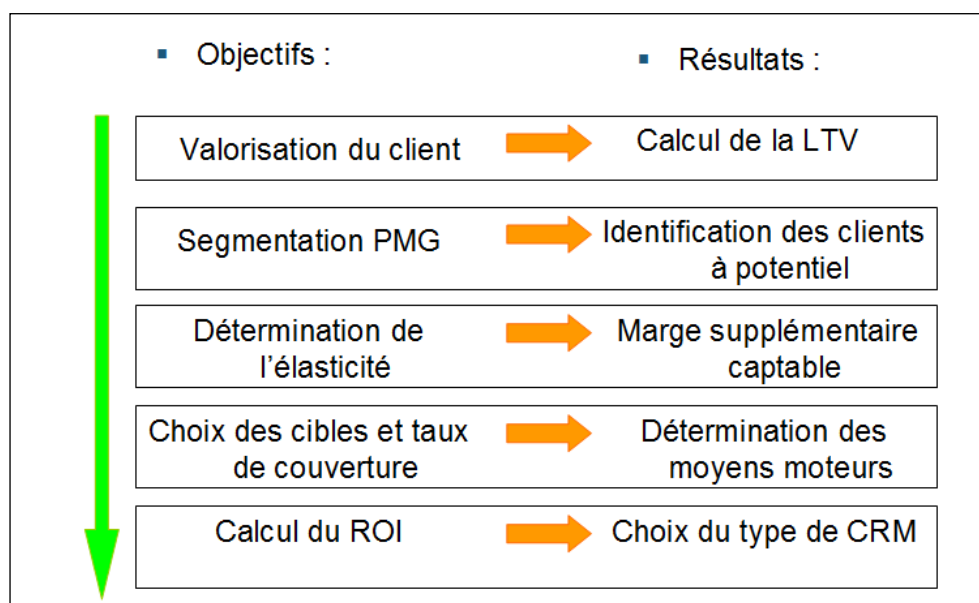


Figure 28 : Méthodologie générale

Le schéma ci-dessus récapitule les bénéfices apportés par chaque analyse en les hiérarchisant par ordre de réalisation.

Il s'agit pourtant de calculer une rentabilité de la fidélisation sur la durée de vie de segment client.

L'origine des données :

Pour mettre au point notre modèle « composite », nous utilisons sur les biens d'équipement automobile à cycle long une étude à un instant t assez proche dans sa forme de SIMM de Kantar Média.

Description méthodologique des données

L'Étude Y¹³ disponible sur les 5 principaux pays d'Europe :

- France, Allemagne, Royaume Uni, Espagne, Italie

Taille de l'échantillon : 19 454 individus sur les pays concernés

Période de l'étude : 1999 et 2000

L'Étude Y est notre source d'information principale pour Objectif Fidélisation, toutes choses égales par ailleurs lorsque ponctuellement d'autres sources sont utilisées.

Rappel des biais de l'Étude Y :

- Une mesure individuelle et non par foyer
- Surreprésentation des consommateurs de la marque étudiée
- Vision statique
- Absence de données identiques sur tous les individus

2.51.2 La valorisation du client :

Nous définissons le concept de valorisation :

- Il s'agit de déterminer par client de la marque étudiée de l'étude Y la valeur totale et la valeur de la marque étudiée sur sa durée de consommation.

Nous définissons également la terminologie employée pour faire un lien entre le modèle de LTV et notre approche « composite » :

- Valeur = Marge (source : la direction financière de la marque étudiée)
- Client = un client de la marque étudiée
- Valeur Totale du Client de la marque étudiée (VTC) = Marge totale dégagée sur toute la vie d'un client (achats de la marque étudiée + autres)

La valeur produit initiale est la marge du pneu déclaré dans l'étude Y.

Tous les calculs sauf précision sont réalisés pays par pays, l'Europe n'est pas une somme des pays mais un recalcul individuel.

¹³ Etudes internes réalisées par les constructeurs

Nous proposons le concept de VTC (Valeur Total Client) qui inclut celle de la marque VC (Valeur Client) plus la consommation faite à des concurrents de la marque étudiée. C'est un problème de précision des appellations.

Nous tenons compte d'un certain nombre de paramètres liés à la nature spécifique du produit étudié :

Les dimensions temporelles :

La période de référence de l'Étude Y : une année (12 mois glissants). La durée de consommation : 57 ans (de 18 à 75 ans).

Les événements clients majeurs dans sa vie de consommateur :

Montée en gamme, (évolution du type de véhicule),

Équipement par foyer, (évolution du nombre de véhicules par foyer).

Prise en compte des pneus hiver :

Différence d'influence en fonction des pays (pour la marque étudiée seule l'Allemagne, l'Italie, la France sont concernés par la vente de pneus hiver).

La règle retenue : un conducteur roule un tiers de son kilométrage annuel déclaré en pneus hiver (marge) et deux tiers en pneus été (marge).

Le renouvellement :

Pour pondérer la valeur du client sur sa durée de consommation (un client de la marque étudiée n'est pas fidèle à la marque sur sa durée de vie) nous appliquons à la VTC la PDM valeur de la marque étudiée. La PDM valeur est la part dépensée par tous les clients de la marque vs les dépenses totales du marché. La PDM étant la somme de toutes les consommations individuelles, elle est en moyenne assez proche d'un taux de nourriture individuel. On peut donc l'utiliser pour estimer la valeur à la marque d'un client quand on connaît la valeur totale de celui-ci sur sa durée de vie.

L'achat de véhicules neufs :

En fonction des pays on applique une hypothèse de part de véhicule acheté neuf vs les véhicules totaux (neufs plus occasions) qu'un client va pouvoir acheter dans sa vie. Plusieurs méthodes sont disponibles mais on peut regarder dans les courants commerciaux de la marque étudiée, la part de première monte (pneus vendus au constructeur), de pneus mis sur le véhicule à la construction.

C'est une approximation que l'on peut vérifier dans d'autres sources automobiles (NCBS, European Survey...¹⁴).

La plupart de ces règles sont contingentes au produit étudié mais on retrouve régulièrement ce genre d'exception dans les modélisations sur la durée de vie.

Chacune de ces règles est à prendre en compte. Il faut passer un certain temps au moment de la mise au point du modèle pour les isoler tout d'abord, puis pour les normaliser à partir des données.

Nous utilisons la PDM Valeur comme le font d'autres auteurs (Kim, Min et Shin, Journal of Data-base marketing, 1999).

¹⁴ Ces deux études sont des études de référence dans l'automobile pour décrire les acheteurs et leur comportement d'achat sur de longues périodes.

2.51.3 Calcul de la VTC :

$$VTC_i = VP \times TR \times 4 \times (PDM_{\text{valeur}})$$

VP = Valeur du Pneu déclaré dans l'Étude Y en marge

TR = Taux de Renouvellement (kilométrage parcouru déclaré dans l'étude Y/taux d'usure du pneu considéré)

Le kilométrage parcouru déclaré sous la forme d'un intervalle est moyenné. Le taux d'usure du pneu associé à sa marge est une source interne de la marque étudiée.

4 : On considère que les 4 pneus s'usent simultanément.

PDM valeur : En l'absence d'autres éléments, PDM valeur sur la marge nous permet d'obtenir le renouvellement à la marque étudiée par pays sur une année (chaque renouvellement n'étant pas systématiquement favorable à la marque étudiée). PDM valeur n'étant pas un « vrai » taux de fidélité, il reste modifiable à tout moment dans le modèle.

La PDM valeur est calculée sur la marge issue de l'étude Y. C'est la somme de la marge réalisée par la marque étudiée vs la somme de la marge totale de tous les individus.

Les taux de fidélité de l'étude Y sont très élevés (France : 79,5 entre le dernier achat et l'avant dernier, 84,2 entre le dernier et l'antépénultième). Ces taux étant basés sur du déclaratif et anormalement élevés, on a volontairement choisi la PDM valeur.

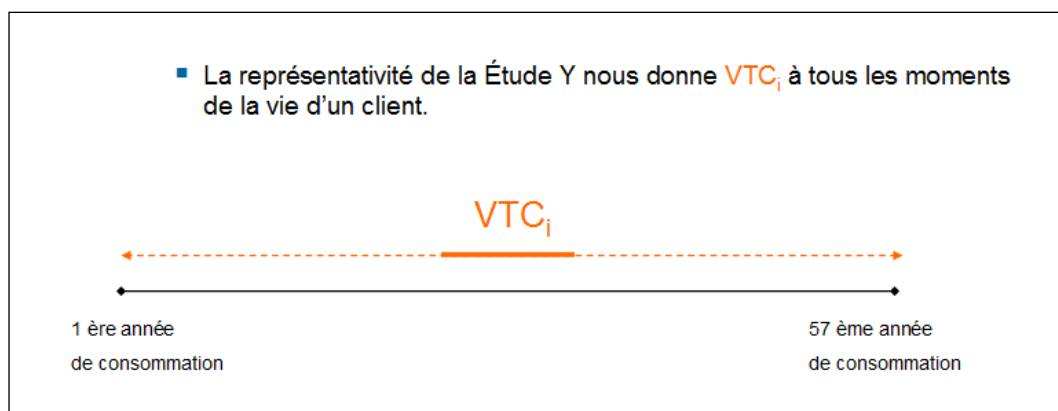


Figure 29 : Calcul de la VTC

On retrouve la visualisation de la courbe sur la durée de vie. Là, le modèle est saturé, nous estimons que la consommation a une fin entre la 1ère année de consommation (l'âge de conduire à minima : 18 ans en France) et la fin de la conduite active (18+57=75 ans). La courbe noire c'est la valeur totale du client sur sa durée de vie de conducteur, VTC, la courbe orange c'est ce que peut s'allouer une marque sur cette durée de vie.

	T0/T1		T1/T2		T0/T2	
	T1	T0	T2	T1	T2	T0
France	1 201	955	513	435	513	383
Allemagne	554	407	165	129	165	119
RU	469	262	162	135	162	107
Espagne	1 239	949	535	461	535	380
Italie	783	559	275	209	275	173

	T0/T1	T1/T2	T0/T2
France	79,52%	84,80%	74,66%
Allemagne	73,47%	78,18%	72,12%
RU	55,86%	83,33%	66,05%
Espagne	76,59%	86,17%	71,03%
Italie	71,39%	76,00%	62,91%

■ Les taux de fidélité de la Étude Y sont trop élevés pour être utilisé dans le calcul de la LTV

Figure 30 : Zoom sur les taux de fidélité de l'Étude Y

Le taux de fidélité est la division des clients qui déclarent dans l'étude Y avoir racheté la marque vs le total de ceux qui l'ont achetée. Ce taux est sur déclaré par rapport à la réalité connue.

Il s'agit d'une validation méthodologique, nous aurions pu utiliser le taux de fidélité déclaré à la marque pour calculer le delta restant à la concurrence sur la durée de vie mais ces taux semblent sur déclarés. Nous gardons donc la notion de PDM valeur pour estimer ce delta à la concurrence.

Cas particuliers des pays avec des pneus hiver :

Sur certains pays, le climat nécessite de changer de pneus en hiver, donc d'avoir systématiquement 2 jeux de 4 pneus, cela a un impact non négligeable sur la valeur, surtout que dans ces pays, les véhicules sont plutôt haut de gamme et nécessitent de forts diamétricités.

$$VTC_i = (2/3V_{Pe} + 1/3V_{Ph}) \times TR \times 4 \times (PDM_{valeur})$$

V_{Pe} = Valeur du Pneu été

V_{Ph} = Valeur du Pneu hiver

L'indice du pneu hiver est inférieur à 1 indice au pneu été.

L'hypothèse de répartition (2/3, 1/3) : le client utilise ses pneus hiver sur une période annuelle de 4 mois.

Les individus auxquels est appliquée cette formule sont tirés aléatoirement dans l'Étude Y dans les proportions suivantes :

- Allemagne : 28%
- Italie : 8%
- France : 6%

Ces chiffres applicables sur l'Étude Y ont été communiqués et validés par d'autres études internes.

Il s'agit d'un ajustement du modèle pour tenir compte de la dimension saisonnière de consommation du produit dans certains pays.

A partir de là, nous pouvons estimer la valeur pour la marque étudiée comme une part de la valeur totale des clients.

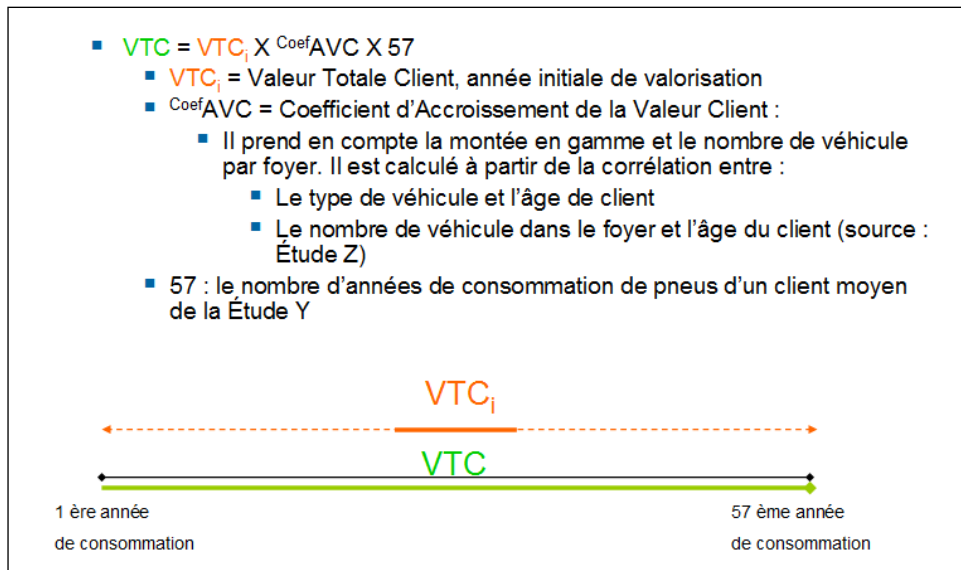


Figure 31 : Calcul de la VTC (cas particulier des pays avec des pneus d'hiver)

En observant le véhicule moyen à chaque âge nous pouvons observer avec le temps une montée en gamme du client. Nous retrouvons là un des concepts clefs de l'accroissement de la valeur client : avec le temps, la régularité d'achat et une augmentation de son pouvoir d'achat, le client monte en gamme dans ses achats. En calculant le différentiel de valeur des véhicules entre les âges, on obtient un coefficient de montée en gamme. Par âge, on peut aussi calculer dans le temps un nombre de véhicule probable dans le foyer, qui croît logiquement avec la maturité de la famille puis décroît. On peut déduire de cette observation un coefficient d'accroissement de la valeur client. La somme de ces deux coefficients ajustés donne CoefAVC .

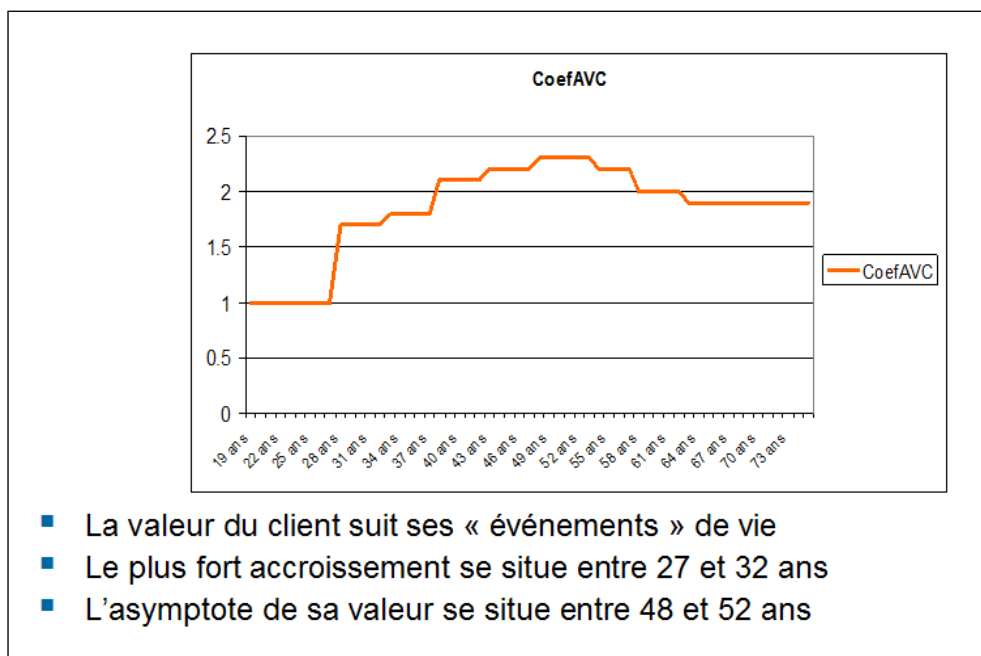


Figure 32 : Durée de vie du client et évolution de sa valeur dans l'automobile

L'achat de véhicules neufs :

Un client ne choisit pas ses pneus lorsqu'il achète un véhicule neuf, il faut donc retirer de sa valeur potentiellement captable cette part où la marque ne peut influencer sur son choix.

Selon l'étude Y un client moyen achète dans sa vie des véhicules neufs. En général, sur ces véhicules, il ne choisit pas ses pneus.

Par contre, pour les véhicules d'occasion, le cycle de renouvellement reste le même. Il faut donc enlever de la VTC la marge associée aux pneus du véhicule neuf.

En l'absence d'éléments plus précis, nous partons sur le raisonnement suivant au niveau individuel : Un foyer français change de véhicules tous les 4 - 5 ans (source : Syndicat des Constructeurs Automobiles). Sur 57 ans : 12.6 véhicules PVNp = Part de Véhicules Neufs par pays sur 57 ans : France 29%, Allemagne 33%, RU 21%, Italie 44%, Espagne 51% soit $12,6 \times PVNp = n$ actes d'achat de véhicules neufs en moyenne.

On retire donc de la VTC : valeur du pneu initiale dans l'étude Y \times PVNp \times 1.87 \times 4.

Il s'agit là aussi d'un ajustement du modèle lié au cas spécifique du produit. On arrive à formuler le modèle suivant qui regroupe tous les composants de Wayland et Cole.

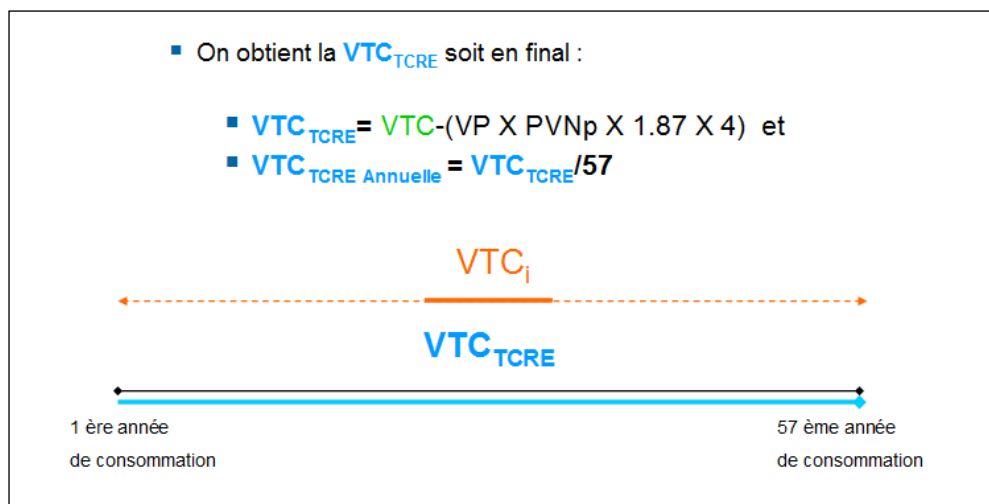


Figure 33 : Valeur Totale Client Renouvellement

La valeur du client pour la marque i, VTC_i est un sous ensemble de la valeur totale du client sur sa durée de vie. Cela est aussi vrai pour le client qui renouvelle lui-même ses pneus vs la première monte où il ne les choisit pas mais a déjà pourtant une valeur « pneus »

- Définition des PMG :
 - Il s'agit de segmenter les clients individuellement sur leur valeur en terme de marge
 - Pour les produits qui n'ont pas d'historique de la date de l'achat on parle de segmentation PMG (Petits clients, Moyens clients, Gros clients)
 - En général une « bonne » segmentation montre que les gros clients sont marginalement les plus contributifs à la marge en relation de ce qu'ils représentent en part de clients

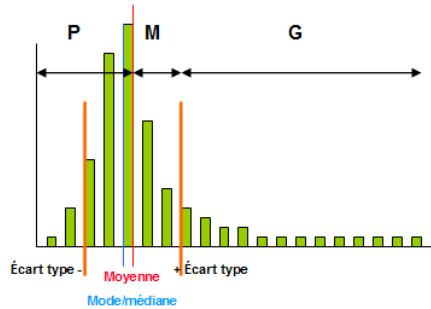


Figure 34 : Segmentation PMG

On retrouve la notion de PMG liée au type de véhicule et au kilométrage parcouru. Cette notion est celle d'Objectif Fidélisation[®], sa représentation est la même.

- Une distribution de la valeur client très concentrée sur peu de clients
- Une hyper concentration de la rentabilité
- Pas de modification significative du rapport NA/CA avec les nouveaux calculs
- Il y a un effet produit certain.
- Une distribution linéaire des gros (pas de groupe homogène de gros/très gros)
 - France :
 - Allemagne :

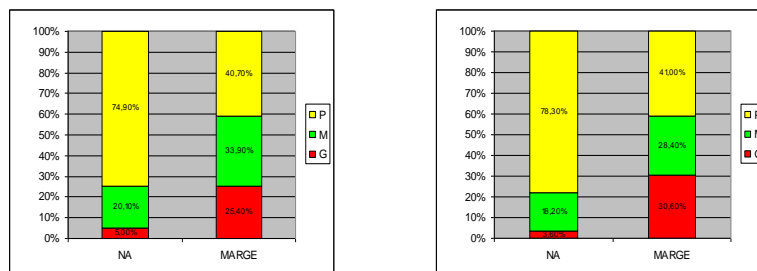


Figure 35 : Zoom sur les bornes PMG (France et Allemagne)

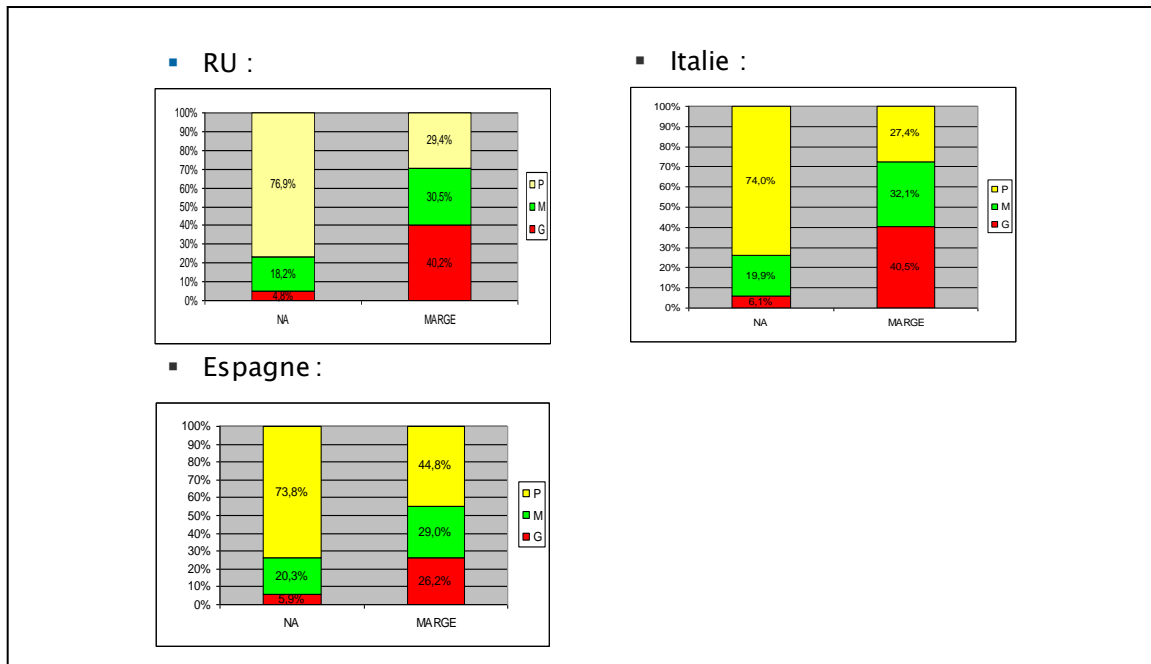


Figure 36 : Zoom sur les bornes PMG (RU, Italie et Espagne)

Dans les différents cas, on voit que les Gros clients, peu nombreux, représentent une part très importante des achats. Ils sont le cœur de cible. La cible la plus contributive au CA.

2.51.4 Une cible très identifiable :

Selon l'étude Y de la marque étudiée le client G à potentiel est :

- Un homme
- CSP plus
- Des revenus élevés (large executive)
- Grosse cylindrée puissante comme véhicule principal

Plus il est jeune avec ces critères, plus il est rentable.

Il y a peu de critères descriptifs dans l'Étude Y, néanmoins les critères ci-dessus sont faciles à utiliser pour cibler tactiquement.

Cette cible est valable pour tous les pays étudiés.

SIMM permet de faire des profils de cibles plus fins grâce à la richesse de ses questions. L'étude que nous avons utilisée dans le cas présent n'avait pas pour but de faire des profils extrêmement détaillés de la clientèle.

On définit plus précisément les bornes PMG qui vont nous permettre de définir la captation par segment :

- France :

	XXX			Marché
	Borne Inférieure	Moyenne	Borne supérieure	Moyenne
P	1,3 €	3,3 €	3,2 €	2,9 €
M	5,5 €	10,4 €	10,0 €	9,7 €
G	11,1 €	31,2 €	164,4 €	41,4 €

- Allemagne :

	XXX			Marché
	Borne Inférieure	Moyenne	Borne supérieure	Moyenne
P	1,2 €	3,1 €	4,2 €	3,5 €
M	4,6 €	9,6 €	12,6 €	10,3 €
G	13,7 €	47,3 €	144,3 €	43,8 €

Figure 37 : Bornes PMG VTC XXX (France et Allemagne)

- RU :

	XXX			Marché
	Borne Inférieure	Moyenne	Borne supérieure	Moyenne
P	1,6 €	2,2 €	5,5 €	1,8 €
M	7,5 €	9,5 €	15,8 €	7,4 €
G	18,3 €	47,2 €	144,3 €	41,9 €

- Italie :

	XXX			Marché
	Borne Inférieure	Moyenne	Borne supérieure	Moyenne
P	1 €	3 €	3 €	3 €
M	5 €	11 €	10 €	11 €
G	11 €	47 €	164 €	48 €

Figure 38 : Bornes PMG VTC XXX (RU et Italie)

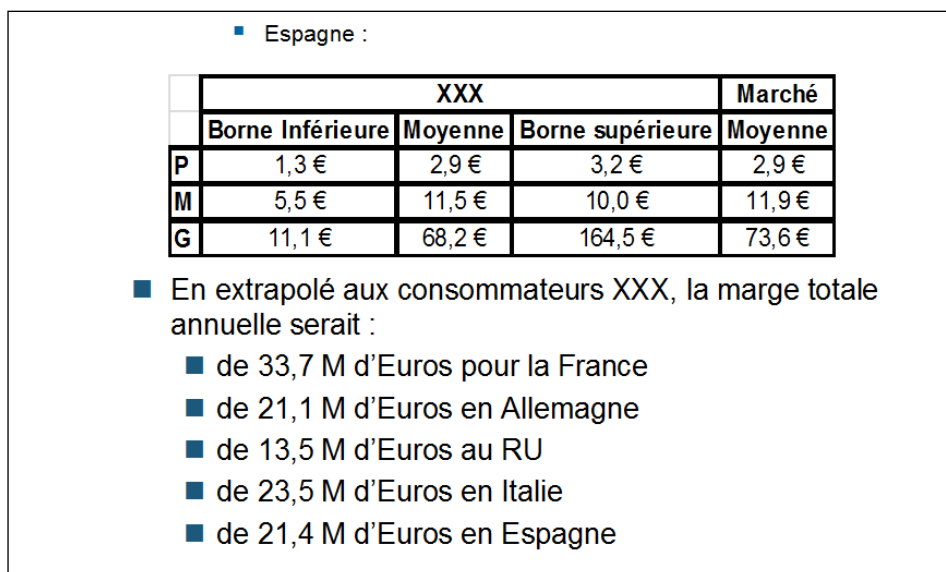


Figure 39 : Bornes PMG VTC XXX (Espagne)

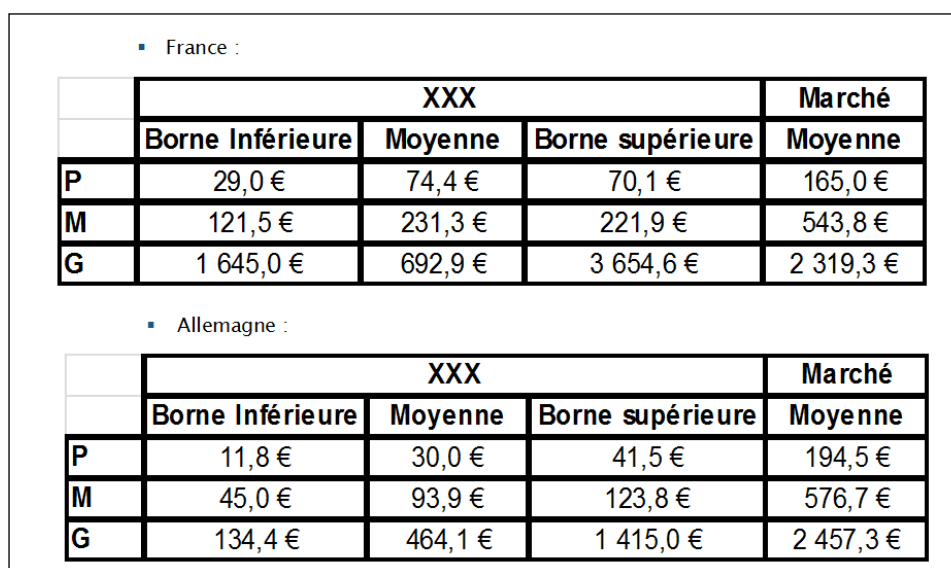


Figure 40 : Bornes PMG VTCTCRE XXX (France et Allemagne)

▪ RU :				
	XXX			Marché
	Borne Inférieure	Moyenne	Borne supérieure	Moyenne
P	13,0 €	18,1 €	45,7 €	103,2 €
M	62,1 €	79,3 €	131,3 €	415,2 €
G	152,4 €	393,1 €	1 201,1 €	2 360,2 €

▪ Italie :				
	XXX			Marché
	Borne Inférieure	Moyenne	Borne supérieure	Moyenne
P	19,4 €	38,9 €	46,7 €	144,3 €
M	81,0 €	170,0 €	148,0 €	596,5 €
G	164,7 €	701,1 €	2 437,5 €	2 650,9 €

Figure 41 : Bornes PMG VTCTCRE (RU et Italie)

▪ Espagne				
	xxx			Marché
	Borne Inférieure	Moyenne	Borne supérieure	Moyenne
P	19,4 €	42,9 €	46,7 €	159,3 €
M	81,0 €	170,9 €	148,0 €	662,3 €
G	164,7 €	1 010,4 €	2 437,5 €	4 096,6 €

▪ La valeur marché est le même calcul que la valeur XXX sans la PDM Volume, avec un valeur moyenne du pneu tenant compte de la concurrence dans la catégorie

Figure 42 : Bornes PMG VTCTCRE (Espagne)

2.51.5 Rappel sur le taux de couverture de cible d'Objectif Fidélisation[®] :

Dans tous les programmes de fidélisation sur des cibles de masse que nous avons gérés en CRM sortant par adhésion volontaire, nous n'avons pu capter que « 30% » de la cible des Gros avec les moyens moteurs du marché à notre disposition.

Dans la distribution, on peut aller jusqu'à 50%.

Il est très difficile d'aller au-delà.

On retrouve donc le concept de taux de couverture de cible explicité précédemment avec le même raisonnement sur ce produit. Cela diffère de l'approche de la LTV qui n'explique pas cette étape mais qui y est cependant induite. Par contre, cette détermination de la cible captable est une des étapes clés d'Objectif Fidélisation[®] : le taux de couverture de cible. Nous gardons dans le cas d'un programme de fidélisation sortant notre taux de couverture cible de 30%.

Que ce soit pour un modèle de fidélisation à court terme ou à long terme basé sur une détermination de la valeur client, la part de cible désirée que l'on arrive à faire entrer dans le programme est un élément clef de la rentabilité de celui-ci.

- On a pondéré les taux de pénétration par pays pour tenir compte de la surreprésentation XXX dans l'Étude Y.
 - Dans un programme de CRM « sortant » on peut et on doit capter 30% de la cible choisie.
 - En dessous on couvre trop peu la cible pour qu'il y ait un effet visible
 - Au delà des 30 premiers %, la captation devient plus chère et moins efficace, on recrute de plus en plus d'individus hors cible.
 - Règle présente dans la littérature et toujours empiriquement vérifiée.
- | | |
|---|---|
| <ul style="list-style-type: none"> ■ France : ■ Nombre total de foyer français : 24 100 000 ■ 78,9% ont au moins une voiture ■ Taux de pénétration corrigé de l'Étude Y par la PDM : 28,5 ■ Nombre de foyers acheteurs XXX : 5 413 732 ■ Acheteurs XXX G : 5% soit 270 687 ■ Acheteurs captable dans le programme : 30% soit 81 206 | <ul style="list-style-type: none"> ■ Allemagne : ■ Nombre total de foyer allemands : 36 413 000 ■ 74% ont au moins une voiture ■ Taux de pénétration corrigé de l'Étude Y par la PDM : 8,5% ■ Nombre de foyers acheteurs XXX : 2 297 384 ■ Acheteurs XXX G : 5% soit 114 678 ■ Acheteurs captable dans le programme : 30% soit 34 403 |
|---|---|

Figure 43 : Taille de la cible (France et Allemagne)

- | | |
|---|--|
| <ul style="list-style-type: none"> ■ RU : ■ Nombre total de foyer anglais : 29 300 000 ■ 72 % ont au moins une voiture ■ Taux de pénétration corrigé de l'Étude Y par la PDM : 11% ■ Nombre de foyers acheteurs XXX : 2 348 305 ■ Acheteurs XXX G : 14% soit 328 763 ■ Acheteurs captable dans le programme : 30% soit 98 629 | <ul style="list-style-type: none"> ■ Italie : ■ Nombre total de foyer allemands : 20 360 000 ■ 78 % ont au moins une voiture ■ Taux de pénétration corrigé de l'Étude Y par la PDM : 20.6 % ■ Nombre de foyers acheteurs XXX : 3 282 858 ■ Acheteurs XXX G : 7.6 % soit 248 097 ■ Acheteurs captable dans le programme : 30% soit 74 429 |
|---|--|
-
- | |
|--|
| <ul style="list-style-type: none"> ■ Espagne : ■ Nombre total de foyer allemands : 12 112 000 ■ 68 % ont au moins une voiture ■ Taux de pénétration corrigé de l'Étude Y par la PDM : 33,5 % ■ Nombre de foyers acheteurs XXX : 2 789 690 ■ Acheteurs XXX G : 4,7% soit 131 451 ■ Acheteurs captable dans le programme : 30% soit 39 435 |
|--|

Figure 44 : Taille de la cible (RU, Italie et Espagne)

- Même avec le meilleur des ciblages, dès l'adhésion dans un programme de CRM sortant une personne sur deux est hors cible
- Le cleaning des individus hors cible se fait à posteriori

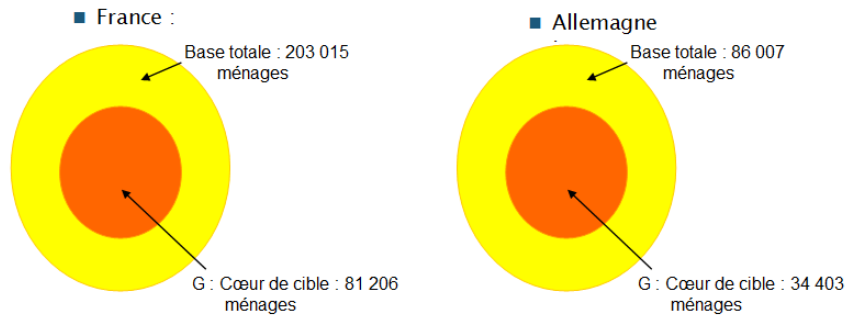


Figure 45 : Captation de la cible sur un programme sortant (France et Allemagne)

Dans un recrutement à un programme de fidélisation par adhésion volontaire « gratuite » un client sur deux environ ne fait pas partie des clients les plus contributifs. Il est difficile de sélectionner à l'entrée.

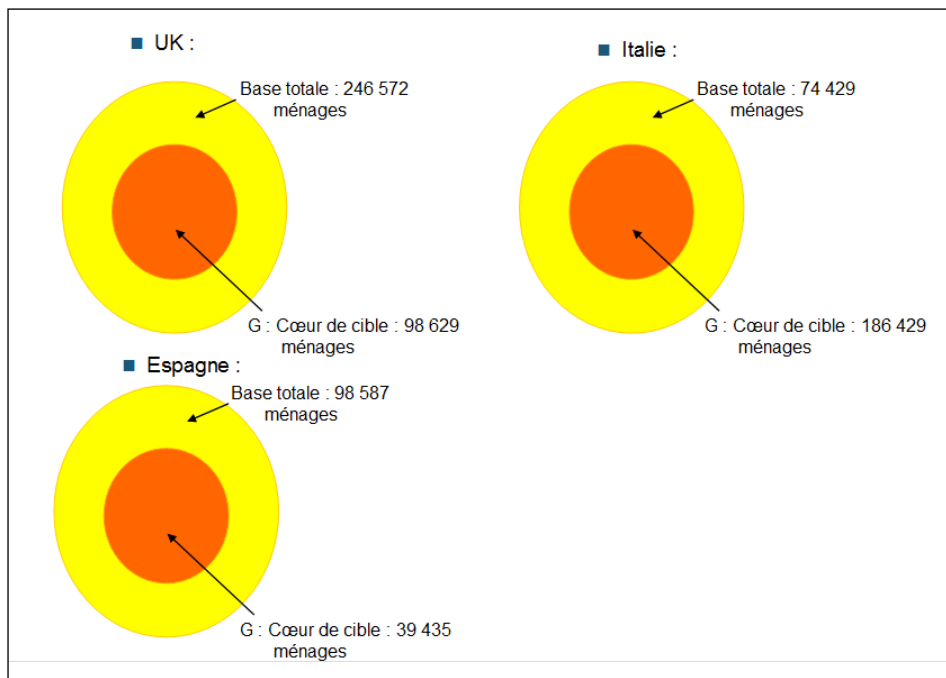


Figure 46 : Captation de la cible sur un programme sortant (UK, Italie et Espagne)

2.51.6 Élasticité et interprétation :

L'élasticité est une marge incrémentale :

$$\underline{\text{Élasticité}} = \text{VTC TCRE} \times \text{Coef Élasticité}$$

L'élasticité s'applique sur la part des achats actuels de XXX.

Le coefficient peut être inférieur ou égal à 20% des achats actuels. Nous utilisons les mêmes valeurs quels que soient les pays.

Notre simulation se fonde sur une élasticité en valeur absolue du client. L'élasticité liée à la « fidélisation » se faisant essentiellement par cannibalisation des achats à la concurrence.

On calcule le taux de nourriture $TN = VTC\ TCRE / \text{Valeur Marché}$.

Si ce taux de nourriture est inférieur ou égal à 50%, il y a une bonne chance de captation.

Le G est évidemment le consommateur le plus « probable » à cette captation. Chez le gros client (G) le taux de captation présente une somme « suffisante » en valeur absolue pour financer tel ou tel type de CRM.

On retrouve dans cette phase les concepts de LTV (VTC TCRE) et la notion d'élasticité appliquée sur le delta marché des PMG à la fois contenus dans le modèle de LTV et Objectif Fidélisation[®].

Rappel sur les 20% d'élasticité :

Dans la théorie du comportement de l'acheteur :

- L'acheteur a un set de marques difficilement modifiable,
- Son comportement d'achat est inertiel.

On est sur un marché de prescription avec une sensibilité au prix directe (le consommateur) et indirecte (le prescripteur). On est sur un marché globalement à faible implication. En général, on peut capter par la fidélisation 20% d'achats en plus faits à la marque. Au-delà de 20%, l'acheteur commence à percevoir un effort : il modifie son set de marques. Là aussi, on applique la notion de taux de captation car on ne peut prendre tout le différentiel consommé à la concurrence mais juste une part de ce différentiel qui est le taux de captation.

■ France :				
	Marge moyenne	Marge captable	TN	Nouveau TN
P	87,5 €	2,6 €	42,76%	6,61%
M	271,9 €	8,2 €	40,64%	6,29%
G	814,6 €	162,9 €	28,80%	5,19%

■ Allemagne :				
	Marge moyenne	Marge captable	TN	Nouveau TN
P	208,7	41,7	11,71%	14,05%
M	652,8	130,6	11,77%	14,13%
G	3 226,4	645,3	15,22%	18,27%

Figure 47 : Marge captable et TN (France et Allemagne)

■ RU :				
	Marge moyenne	Marge captable	TN	Nouveau TN
P	17,9 €	0,5 €	14,29%	2,21%
M	78,2 €	2,3 €	15,90%	2,46%
G	385,0 €	11,6 €	14,02%	2,17%

■ Italie :				
	Marge moyenne	Marge captable	TN	Nouveau TN
P	39,1 €	1,2 €	22,62%	3,50%
M	158,2 €	4,8 €	22,15%	3,43%
G	628,8 €	18,9 €	19,58%	3,03%

Figure 48 : Marge captable et TN (RU et Italie)

■ Espagne :				
	Marge moyenne	Marge captable	TN	Nouveau TN
P	42,8 €	1,3 €	23,93%	3,70%
M	171,6 €	5,2 €	22,95%	3,55%
G	1 013,8 €	30,4 €	21,39%	3,31%

Figure 49 : Marge captable et TN (Espagne)

2.51.7 Part de la marge allouable au CRM :

Sur la marge additionnelle, on peut décider d'allouer tout ou partie à la fidélisation.

Le réalisme de l'usage veut que l'on alloue une moyenne de 40% à la fidélisation. Le delta de marge servant à financer des éléments fixes qui s'affectent même sur les produits incrémentaux. Par ailleurs, la notion de réactif d'ordre 2 n'étant pas applicable en l'état sur la LTV, les 40% sont un moyen de l'appliquer car cela équivaldrait à ce que moins de la moitié des individus réagissent à nos offres.

En fonction de la somme investie en valeur absolue on peut déterminer le type de programme de fidélisation :

- Supérieur à 4,3 € par an : une fidélisation sortante ou mixte : on va pouvoir envoyer pro activement des éléments aux clients pour le fidéliser.
- Entre 1,4 et 4,3 € par an : une fidélisation entrante (marketing relationnel) : on va susciter que le client nous appelle, donc principalement à ses frais car les moyens moteurs sont insuffisants pour le contacter.
- Moins de 1,4€ : la fidélisation est un investissement pur (politique défensive)

Ces sommes sont très différentes ensuite en fonction de la nature (transactionnelle vs relationnelle) et de la qualité de la relation souhaitée (type de supports, fréquence...).

La somme diminue sensiblement avec la taille de la cible :

- Accroissement marginal du rendement dans le cadre d'un CRM entrant
- Limitation aux 30% de captation pour un CRM sortant.

A partir de là, on peut déterminer précisément quelle part investir à la fidélisation ou au CRM. On connaît la marge du client sur sa durée de vie et on sait quelle part allouer à la fidélisation en % pour le type de fidélisation donc on connaît maintenant les coûts unitaires. On obtient donc la somme allouable à la fidélisation au global sur la durée de vie du client et par an.

■ France :			
	Marge captable	Investissement	Investissement /an
P	114,8 €	6,9 €	0,1 €
M	356,7 €	21,4 €	0,4 €
G	1 068,7 €	64,2 €	1,1 €

■ Allemagne :			
	Marge captable	Investissement	Investissement /an
P	41,1 €	2,5 €	0,0 €
M	128,6 €	7,7 €	0,1 €
G	635,5 €	38,2 €	0,7 €

Figure 50 : Investissement CRM (France et Allemagne)

■ RU :			
	Marge captable	Investissement	Investissement /an
P	23,4 €	1,4 €	0,0 €
M	102,6 €	6,2 €	0,1 €
G	505,1 €	30,3 €	0,5 €

■ Italie :			
	Marge captable	Investissement	Investissement /an
P	41,1 €	2,5 €	0,0 €
M	128,6 €	51,4 €	0,9 €
G	635,5 €	254,2 €	4,5 €

Figure 51 : Investissement CRM (RU et Italie)

■ Espagne :

	Marge captable	Investissement	Investissement / an
P	56,17 €	3,37 €	0,06 €
M	225,10 €	13,52 €	0,24 €
G	1 329,97 €	79,87 €	1,40 €

Figure 52 : Investissement CRM (Espagne)

2.51.8 Mise en perspective des coûts et des gains :

On retrouve le tableau de synthèse d'Objectif Fidélisation[®] qui croise les coûts inhérents au programme de fidélisation ou de CRM avec les gains escomptables sur la durée de vie des clients par segment PMG.

A partir de ces résultats, le modèle est terminé et le décideur peut faire ses choix marketing associés à la fidélisation la mieux adaptée pour sa marque et présentant les meilleurs « chances » de rentabilité.

2.51.9 Choix définitif des cibles à potentiel :

- De tailles limitées
- Facilement identifiables et adressables
- Marge générée supérieure à la moyenne
- Élasticité satisfaisante
- Fréquence d'achat supérieure à la moyenne
- Marge allouable suffisante pour permettre tel ou tel CRM
- Choix définitif du type des moyens moteurs et du type de CRM
- Croisement des coûts et des gains sur des périodes temporelles différentes : à 1an, 3 ans, 5 ans, à 10 ans, à vie.

A partir de ces résultats, le modèle peut servir aussi d'indicateurs de pilotage pour s'assurer que le programme va dans le bon sens, on retrouve alors le tableau de suivi de Objectif Fidélisation[®].

2.51.10 Conclusion :

L'approche composite utilise tous les concepts d'Objectif Fidélisation[®] : -couverture de cible -structure PMG -réactivité -captation -marge allouable -coût unitaire, plus les concepts liés à LTV essentiellement la linéarité du temps. On peut formuler un modèle des deux modèles comme dans l'exemple précité.

Le modèle de durée de vie à court terme Objectif Fidélisation[®] se somme bien pour donner un modèle de durée de vie à long terme de type LTV. A l'inverse si on décompose la LTV par année calendaire on démontre que l'on peut arriver à un modèle de prévision de la valeur comme Objectif Fidélisation[®]. C'est l'inclusion de ces deux approches qui rend notre méthode « composite » des deux types de modèles.

2.52 Réponses nos questions de recherche :

2.52.1 Tests et validation des questions :

Il convient maintenant de rappeler nos questions de recherche et de montrer si on peut y répondre ou non.

Q1 : On retrouve les mêmes composants (hors le temps) entre Objectif Fidélisation[®] et une approche par la LTV ?

	Objectif Fidélisation	LTV	Modèle composite
Couverture de cible	Oui	Induit	Oui
Structure PNC	Oui	Oui	Oui
Réactivité d'ordre 2	Oui	Induit	Oui
Captation	Oui	Oui	Oui
Marge allouable	Oui	Oui	Oui
Maintient des coûts unitaires	Oui	Induit	Oui
Axe temps	Induit	Oui	Oui

Figure 53 : Convergence des modèles

Nous validons cette question.

Q2 : La prise en compte du temps dans la durée de vie différencie Objectif Fidélisation[®] d'une approche par la LTV.

En fait, la durée de vie est moins bien prise en compte dans Objectif Fidélisation[®], on pourrait estimer la valeur client sur l'ensemble de sa durée de vie, même sur les produits à cycle d'achat/rachat court, cela est possible. Objectif Fidélisation[®] travaille à un instant t.

Nous validons cette hypothèse : la durée de vie dans Objectif Fidélisation[®] est bien une dimension induite.

- dans Objectif Fidélisation[®] ≤ à 12 mois glissants
- dans LTV = +∞

Evidemment le composant t d'Objectif Fidélisation[®] est inclus dans t de la LTV → Objectif Fidélisation[®] inclus la LTV

Q3 : A un instant t, la modélisation de la LTV sur année moyenne donnée est égale au résultat d'Objectif Fidélisation[®]

Nous validons cette hypothèse : $LTV_{t=1} = VC_{\text{Objectif Fidélisation}^{\text{®}}}$

La division de la LTV à une année moyenne sur un produit à cycle court égale le résultat d'Objectif Fidélisation[®]. Les paramètres de l'un peuvent et doivent s'appliquer à l'une ou l'autre des deux approches.

Q4 : On ne peut pas faire un calcul de ROI du programme de fidélisation sans prendre en compte la durée de la relation :

C'est la durée de la relation qui accroît la valeur du client, c'est la logique des moments clients.

Dans la plupart des marchés plus un client consomme longtemps avec une fréquence plus ou moins régulière une marque, plus il contribue à de la valeur pour la marque. Il y a une relation entre le temps et la probabilité de rachat. Celle-ci est plus déterminante de l'accroissement de la valeur client que les quantités achetées par actes d'achats.

- Pour un produit alimentaire, c'est vers 35 ans que les foyers consomment le plus (2 adultes, 2 enfants entre 6 et 14 ans en moyenne, Source Conso-Scan de Sécodip).
- Un foyer atteint l'asymptote de sa consommation automobile entre 48 et 52 ans.
- Dans la banque, la valeur client est à son maximum vers 65 ans environ.

L'hypothèse est invalidée, il faut tenir compte de la durée de la relation ne serait-ce que pour estimer les coûts au-delà des gains. Objectif Fidélisation[®], lui, n'a pas besoin de la durée, même si celle-ci est induite. Nous rejetons donc cette hypothèse, la durée est omni présente mais elle n'est pas utile en tant que telle quels que soient les produits.

Nous validons trois hypothèses sur les quatre que nous avons émises. On peut à partir des deux modèles créer un modèle composite regroupant l'ensemble des composants.

2.53 Mise en évidence des lois sous-jacentes :

Trois hypothèses sont validées, une rejetée. Il y a bien une différence entre les deux modèles, mais ils sont fortement inclus l'un dans l'autre et leurs composants sont omni présents dans chacun d'entre eux. On peut parler néanmoins de modèle composite :

- la captation n'est pas exactement la même,
- le temps est pris en compte différemment.

Ces deux notions doivent être cependant investiguées davantage.

On peut faire un modèle composite d'une approche comme Objectif Fidélisation[®] qui porte sur un instant t et un modèle de LTV qui, lui, porte sur la durée de vie estimée ou totale du client.

Le raisonnement sur ce qui est atteignable et escomptable est le même, les indicateurs pratiquement similaires. Il y a quand même quelques différences.

2.54 Interprétation des aspects complémentaires des approches sur un plan managérial:

Sur un plan managérial, on démontre plusieurs choses :

- il faut estimer la rentabilité escomptable d'un programme de fidélisation à priori et cela est faisable. Les trois cas sur des secteurs différents, montrent que les données marché pour le faire existent,
- il faut ensuite se servir de ces indicateurs non seulement pour le go/no go mais également pour piloter le programme,
- on ne peut pas modéliser le ROI de la fidélisation sans prendre en compte tous les paramètres,
- prendre tous les paramètres ne veut pas dire non plus tomber dans un excès de comptabilité analytique, les faits sont en général suffisamment évidents pour que les grosses masses de rentabilité suffisent. La modélisation doit produire des résultats opérationnels pour rester « pilotables »,

- selon le produit ou le service il faut travailler le modèle sur la durée de vie du consommateur ou à un instant t .

Evidemment, le modèle reste perfectible et certains aspects nous ont peut-être échappés, mais nous avons essayé de faire une synthèse d'un ensemble d'existants théoriques récents et d'un certain nombre de cas empiriques.

2.55 Conclusion

Nous avons démontré que l'on peut calculer la rentabilité d'un programme de fidélisation de deux manières :

- à un instant t et cela s'applique pour tous les produits de vente à cycle court et dont la consommation peut s'étaler tout au long d'une vie (modèle dit non saturé). La périodicité inhérente aux données pousse à calculer une rentabilité de la fidélisation à court terme répond à un objectif mais indirectement la somme de cette valeur sur n périodes représente la valeur totale d'un client sur sa durée de vie si toutes les périodes sont bien prises en compte. En fait les modèles qui calculent la valeur à un instant comme Objectif Fidélisation[®] répondent à une détermination de la rentabilité des actions de fidélisation à court termes.

-sur la durée de vie par l'utilisation de modèle de LTV. Dans ce cas la période d'analyse est toute la durée de vie du consommateur, contrairement aux panels de Home Scanning et aux études de marché qui n'apportent qu'une vision partielle de cette durée.

Notre apport est surtout d'avoir montré la complémentarité de ces deux modèles et que les concepts de l'un s'applique très bien à l'autre. On est donc en mesure d'avoir une approche dite « composite » de la lecture de la rentabilité des actions de fidélisation : à court et à long termes.

L'axe temps est omniprésent dans les deux approches. Les concepts de couverture de cible, structure PMG, réactivité, captation, marge allouable, maintien des coûts s'appliquent bien aussi aux deux types de modèles dont à une approche de modélisation composite : les deux approches sont liées et se complètent.

Notre recherche présente quelques limites. Nous n'avons pas pu valider dans la littérature toutes les composantes de notre modèle composite.

Notamment le seuil de captation qui est évident mais difficilement démontrable car il emprunte des notions à des champs d'études assez différents, comportement de l'acheteur, modèle de satiété. Nous le ferons dans de prochains travaux. La notion de potentiel client est une notion induite par la captation.

Nous n'avons montré que quelques exemples, il en aurait fallu davantage notamment sur la fidélité dans la distribution.

Nous partons du principe qu'une marque ne peut capter qu'un delta de ce que le consommateur achète à court terme et tout au long de sa durée de vie. Nous ne démontrons pas bien si ce delta est effectivement « borné ». Déterminé avec précision, ce delta, aura un gros impact sur les allocations des moyens à la fidélisation.

Néanmoins notre recherche apporte aux lecteurs des avancées, pour la première fois à notre connaissance en France le calcul de la LTV est détaillé pas à pas de façon simple.

Nos exemples très précis et détaillés permettent au lecteur de refaire ses propres modèles. Le but est bien celui-ci : donner une vision simple et claire de ces modèles pour permettre leur recalcul.

Notre travail permet de creuser des pistes de recherche comme la détermination du potentiel client sous-jacent à la captation ou la détermination de l'élasticité du taux de nourriture d'une façon plus scientifique. Il permet de fournir les éléments théoriques et pratiques pour aller plus loin.

Notre travail implique également que l'on n'imagine pas qu'un programme de fidélisation client, de CRM, puisse se lancer sans un calcul préalable de sa rentabilité. A partir de ce calcul préalable, le décideur peut se servir en plus des résultats du modèle pour piloter le going de son programme.

Même si elle utilise une approche très proche de celle des grands cabinets de conseil, la modélisation présentée dans ce document est simple dans son application. On peut imaginer qu'elle permette à des sociétés de petite taille de la réaliser. Elle sort d'une approche plus « contrôle de gestion » qui souvent se focalise trop sur une prise en compte de tous les coûts à un niveau de détail très précis pour une approche plus macro mais aussi celle perçue par le client final, une approche « plus marketing » de ces calculs.

Il reste néanmoins à l'issue de ce travail de nombreuses pistes d'investigations complémentaires :
Il faut creuser la littérature sur les concepts attenants, notamment la captation, et en travaillant sur la détermination des potentiels clients et l'élasticité individuelle du taux de nourriture.

Il faut arriver à graduer l'évolution des quantités incrémentales sur l'axe temps. Mieux maîtriser l'effet du temps notamment dans Objectif Fidélisation[®], mais les travaux récents sur l'efficacité sur les ventes montrent que l'on est confronté à la fois à des problèmes théoriques et à des problèmes de mesure.

Il faut faire tourner le modèle sur des cas portant sur d'autres secteurs.

Nous constatons que le concept de CRM est moins utilisé ces dernières années, car comme nous l'avons démontré au début, il a été victime de sa sur médiatisation et de sa surenchère technologique, le marketing direct se reformulant souvent sous de nouvelles appellations sans pour autant remettre en question ses fondamentaux : aujourd'hui on parle de marketing comportementale, de vue à 360° du client... Si le nom change la problématique du ROI est plus que jamais d'actualité.

On entend actuellement que certains programmes de fidélisation ne sont pas rentables, ils ne le sont pas pour six principales raisons :

- ils ne touchent pas le bon cœur de cible,
- ils ne la touchent pas en nombre suffisant,
- ils ne la touchent pas par le bon canal,
- trop peu de gens du programme réagissent,
- les coûts vs les gains escomptables sont mal définis voire mal pilotés,
- la durée de captation et de ROI est sous-estimée.

Nous pensons que le présent document permet de mieux prendre en compte ces éléments clefs et nous espérons que ce travail va inspirer d'autres praticiens et chercheurs.

Troisième Partie

Recherche 3 : Estimation du potentiel individuel de chiffre d'affaires en utilisant des données issues d'une base de données clients

2.56 Le contexte

La plupart des sociétés, quel que soit leur secteur d'activité en BtoB ou en BtoC aimeraient connaître le potentiel de leurs clients en termes de chiffres d'affaires. Déterminer « un potentiel client » signifie identifier le chiffre d'affaires incrémental qu'une entreprise donnée pourrait générer sur ses clients actuels.

Les modèles de potentiel clients existent et sont principalement basés sur la détermination de la valeur client pendant la durée de vie de ce dernier : LTV (LTV = Life Time Value) (*Bénavent et Crié; Berger et Nasr 1998; Dwyer 1997; Venkaten, Rajkumar et Kumar 2004*).

Au-delà de ces modèles, d'autres approches existent qui estiment la part de dépense du client (*Cooil et al. 2007; Yuxing Du et al. ; Keimingham et al. 2007*).

Enfin d'autres modèles économétriques de potentiel ont été créés mais ils sont basés sur des données externes à la base de données clients (*Plastria 2001, Huff 2003, Reilly 1931*).

La consommation du client (sa valeur totale) représente la consommation de tous les achats d'un produit donné, sur sa durée de vie, par un consommateur donné. Elle se nomme la Valeur Totale du Client (VTC). Par exemple, au cours de sa vie la valeur totale d'un client pour un distributeur, est la somme de tous les achats qu'il aura fait dans le magasin de ce distributeur durant sa vie.

Il est possible d'estimer la consommation d'un consommateur sur un marché pour une marque et un produit donnés dans une catégorie de produits. Au cours de sa vie, le consommateur va consommer n marques m_1, m_2, \dots, m_n . Le rapport entre la consommation totale d'une de ces marques m_n et la consommation totale des marques, est le taux de nourriture de la marque $TnVTC_{mn}$ sur la durée de vie du client (graphique 1).

$$VTC_m = VTC_{m_1} + VTC_{m_2} + \dots + VTC_{m_n} \quad \text{Taux de nourriture de } TnVTC_{mn} = \frac{VTC_{mn}}{VTC}$$

La différence, notée delta dans les graphiques, entre la consommation totale du consommateur sur la catégorie de produit de la marque étudiée VTC et les consommations totales de la marque m_1 , VTC_{m_1} correspond à la consommation totale aux marques concurrentes $VTCC$ (Graphique 2).

$$VTCC = VTC - VTC_{m_1} \text{ donc } VTCC = VTC_{m_2} + \dots + VTC_{m_n}$$

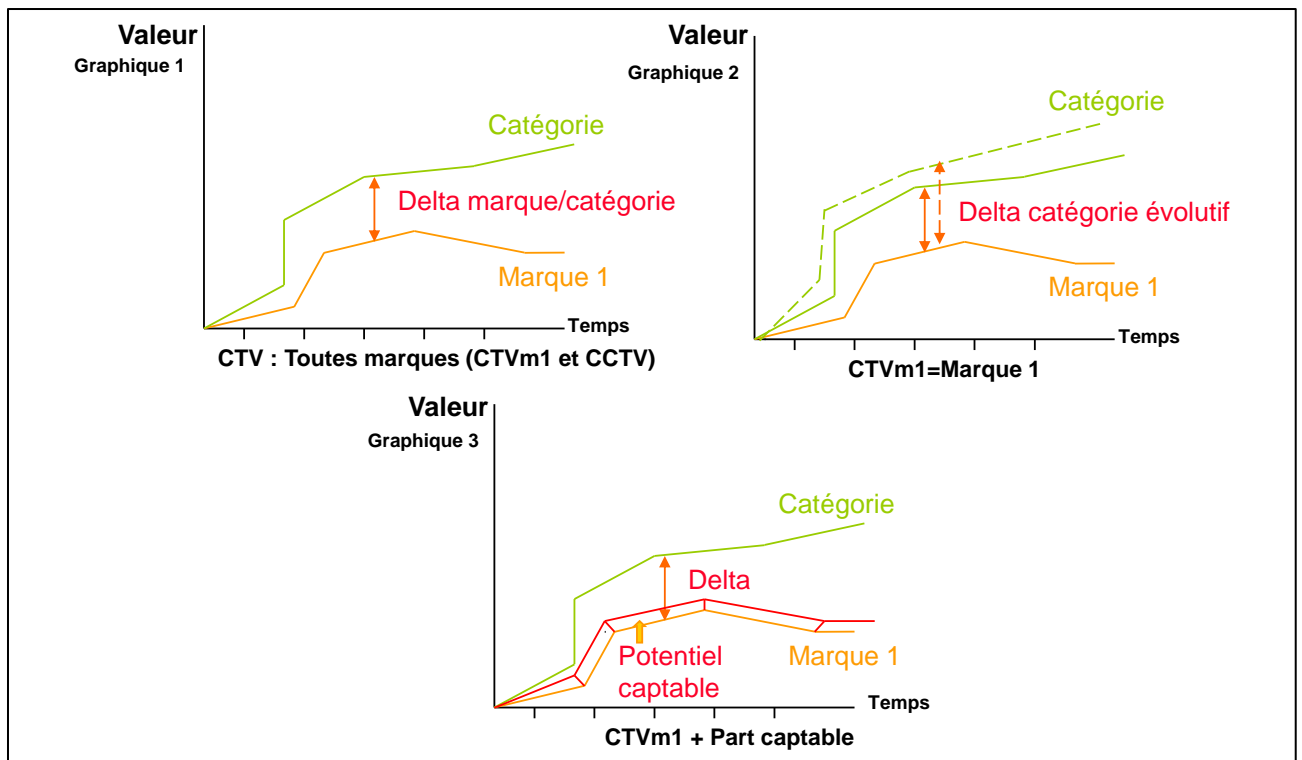
Selon les stimuli marketing de la marque m_1 , le client va prendre une part $TauxC_{m_1}$ de ce delta à la concurrence et/ou accroître sa consommation sur le total du marché :

L'accroissement de sa consommation totale marché

Soit : VA_{m_1} = La valeur actuelle pour la marque 1

$$VTC_{m_1} = VA_{m_1} + TauxC_{m_1} \times VTCC + \text{Taux d'accroissement de la consommation}$$

Les clients d'un distributeur vont consommer dans certains magasins concurrents et peuvent accroître leur consommation totale aux différents distributeurs.



Graphique 31 : Le taux de captation

Donc, le potentiel théorique d'un client est la consommation totale sur sa durée de vie CTV_{m1} qui peut être atteinte par ce client ; il peut être estimé par les moyens du modèle économétrique ci-dessous.

$$VTC_{m1} = VA_{m1} + TauxC_{m1} \times VTCC + \text{Taux d'accroissement de la consommation}$$

Où

$TauxC_{m1}$ = Part de la consommation prise aux concurrents (Graphique 3). C'est la part de marché « captable » sur les concurrents

Taux d'accroissement de la consommation = L'accroissement de sa consommation totale

Le potentiel captable du client correspond à ce que la marque a déjà atteint, augmenté de ce que le client pourra consommer en plus et/ou ce qui sera pris à concurrence. Ce potentiel atteignable peut être estimé de deux façons : en utilisant un modèle économétrique, ce qui nécessite des données exogènes à la base de données ou d'une autre façon en utilisant uniquement des données de la base de données clients, dans notre exemple la base des porteurs de la carte de fidélité de l'enseigne : la méthode des "clones".

Une marque donnée ne peut capter qu'un pourcentage donné du potentiel théorique ; selon plusieurs travaux de recherche dont ceux de Berend Wierenga et Gerrit en 2000. Le potentiel théorique est la consommation totale du consommateur.

Certains chercheurs en marketing ont montré qu'une marque peut accroître son taux de nourriture actuel à une marque de 30%, l'écart type moyen constaté des taux de nourriture. Au-dessus de ce taux, le consommateur perçoit un changement de son comportement de consommation et essaie alors d'y résister.

Au-dessus de 30% d'accroissement il y a trop de modifications de l'ensemble de choix du client¹⁵ (Bremer et Joyce, 1988). Ce sujet a déjà été traité dans une de nos précédentes recherches (Vallaud, 2003).

Les approches les plus avancées de détermination des potentiels essaient de déterminer la proportion de chiffre d'affaires incrémental qui peut être atteinte par une société, en se basant uniquement sur les seules données issues de la base de données clients de la société. Ces approches calculent un potentiel client par client mais évidemment doivent être en phase avec les macros données agrégées du marché.

2.57 Notre sujet de thèse

L'objectif est de travailler avec des algorithmes de classification¹⁶ (*Lerman 1970, Dorofeyuk 1971, Borko et al. Bernick 1963, Two Steps (Tan et al. 1997), K means (Hartigan et al. 1979, Fang et al. 1982), SOM (Teuvo Kohonen 1988, Vesanto 1997, Kaski 1997), etc.*), sur de bases de données de plusieurs millions d'enregistrements de sociétés commerciales (grands distributeurs, opérateurs téléphoniques, fournisseurs d'accès à Internet, sociétés de marketing direct, etc...).

Nous utilisons ces modèles dans le but de déterminer un potentiel client en utilisant une méthode que nous appelons « méthodes des clones » dans laquelle les clients qui se ressemblent le plus, sont considérés comme des clones, et sont supposés avoir le même potentiel.

Nous avons accès à diverses bases de données pour notre processus méthodologique. Dans ce document, nous allons faire un test empirique de notre méthode sur des données clients provenant uniquement de la base de données d'un grand distributeur français.

Au-delà de notre brève présentation du contexte, nous avons à aborder deux sujets majeurs :

- Le calcul du potentiel sur la valeur client en marketing et ces différents indicateurs associés : LTV, taux de nourriture, part de marché captable, etc...
- Les modèles mathématiques qui permettent à des individus similaires d'être regroupés dans des groupes homogènes : les techniques de classification.

Le champ d'investigation est multidisciplinaire avec, en mineur, les études marketing et, en majeur, la discipline des statistiques, du data mining et de la classification.

2.58 L'application précise du modèle

Une part importante des objectifs de notre recherche, est de tester séparément différentes techniques, et, si possible, conjointement, pour s'assurer que les groupes formés sont des groupes de clones homogènes. A delà du choix des modèles, une part importante de notre étude implique de définir les variables les plus informatives et donc une topologie des données en entrée du modèle, qui soit bien adaptée à celui-ci. Le but ici, est d'obtenir, des résultats les plus convergents et les plus pertinents possibles.

Une autre partie importante de notre étude est de sélectionner les méthodes mentionnées ci-dessus et de valider ces choix. L'objectif est de trouver une méthode qui converge suffisamment et qui puisse être validée par l'ensemble des approches ci-dessus. La modélisation devient alors une association de plusieurs modèles.

- La réduction des dimensions pour choisir les variables avec un très grand nombre de groupes et des écarts de valeurs important
- La projection des variables actives et passives¹⁷ dans les groupes,
- La réallocation des groupes par un algorithme supervisé,

¹⁵ L'ensemble de choix est un ensemble fini pour une catégorie de produits donnée qu'un client a l'esprit avant de faire un achat.

¹⁶ Les SOM (Self Organizing Maps, cartes auto organisatrices) appartiennent aux méthodes de classification, souvent appelées "typologies". Ces méthodes sont dites automatiques non supervisées.

Non supervisée veut dire que l'on ne définit pas à priori une variable cible à prédire.

¹⁷ Les variables actives sont utilisées pour construire les groupes eux-mêmes ; les variables passives sont des variables descriptives pour expliquer les groupes.

- La validation de la connectivité de « super classes » par d'autres méthodes de classification non supervisées,
- Une vérification empirique via un panel externe comme celui géré par Nielsen ou TNS Sofres¹⁸ qui représente la « réalité marché » du potentiel.

Le modèle définitif est créé en utilisant une plateforme logicielle standard du marché : la version française de PASW Modeler de SPSS/IBM (ex Clementine).

La contribution scientifique est :

- L'apport méthodologique pour sélectionner les modèles de classification et valider ces choix
- Une application sur des données réelles, validée par un business cas réel : calculer un vrai potentiel captable.

2.59 Les questions de recherche

- Pouvons-nous utiliser une technique de classification pour déterminer les consommateurs qui sont similaires entre eux et donc définir un potentiel réaliste en termes de chiffre d'affaires pour ces clients ?
- Pouvons-nous développer une méthode pérenne, que l'on puisse répliquer ?
- Comment pouvons-nous valider la stabilité des classes obtenues ?

2.60 Le processus de data mining utilisé

Nous utilisons le processus projet standard Cross-Industry Standard Process pour le Data Mining (CRISP)¹⁹ qui nous permet de guider notre approche d'analyse des données. Le processus standard CRISP est constitué des étapes suivantes :

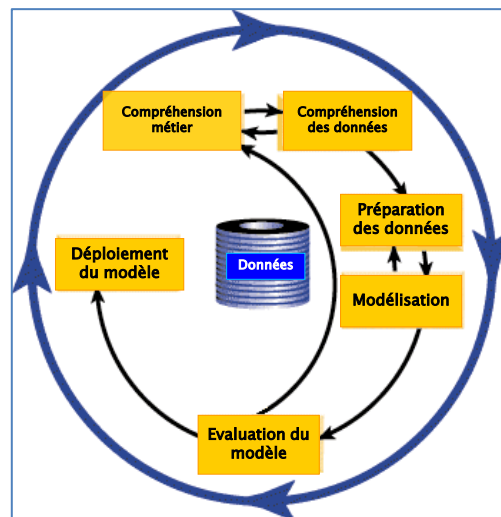


Figure 54 : CRISP DM

2.61 La compréhension des données

¹⁸ Nielsen et TNS sont des sociétés d'études de marché qui fournissent des panels dans lesquels leurs membres scannent leurs achats. Ces panels peuvent être croisés avec des bases de données clients pour mesurer les effets du marketing mix.

¹⁹ <http://www.crisp-dm.org/>

Nous allons travailler sur 5 373 026 individus issus de la base de données d'un grand distributeur français. Nous avons le détail de tous les tickets de caisses sur une période de 12 mois, de Janvier 2006 à Décembre 2006.

Pour le processus de validation externe, nous utilisons aussi une étude de marché disponible sur le marché Français : *Référence 2006* de TNS Sofres²⁰. Cette étude nous donne le taux de nourriture des principaux distributeurs Français²¹.

2.62 La préparation des données

Cette étape consiste à nous familiariser avec les données de la base des membres du programme de fidélité de l'enseigne, dans la but de bien préciser la structure de la base en se basant sur le dessin d'enregistrement, le niveau de complétude des champs contenus dans les fichiers, et aussi l'origine et la nature des données. Chaque champ est vérifié pour s'assurer qu'il n'affecte pas la stabilité du modèle. Nous avons fait un audit et une AED (Analyse Exploratoire des Données) en deux fois, seule la seconde AED est présentée dans ce document.

Sont vérifiés :

- La structure de la base de données
- L'origine et la nature des données (sociodémographiques / consommation)
- La possibilité de réaliser des analyses croisées des données (par marque/ rayon/famille de produits, etc....)
- La périodicité des données
- L'historique des données
- La complétude des données

Les principaux processus appliqués sur les données de la base incluent les étapes suivantes :

- Contrôler et valider le format des variables
- Recoder et corriger certaines variables appelées « variables aberrantes », parce que contenant des valeurs inattendues ou extrêmes par rapport à la moyenne (plus de 3 écarts types)
- Créer des agrégats spécifiques et utiles pour notre détermination des potentiels (chiffre d'affaire total par famille de produits, fréquence annuelle de visites, panier moyen d'achat...)
- Analyser les corrélations entre la variable cible (le chiffre d'affaire) et d'autres variables (des critères sociodémographiques, la fréquence d'achat...) dans le but de vérifier s'il existe des relations de dépendance entre ces variables.
- « Géocoder », étape utile pour enrichir les profils de certaines variables socio démographiques de l'INSEE²² via les IRIS²³ (unités françaises géographiques spécifiques)
- Calculer les distances entre le client et le point de vente (zone de chalandise)

L'analyse a été réalisée sur un chiffre d'affaire sur 12 mois glissants sur l'ensemble de l'historique, pour s'assurer que le modèle soit le plus fiable possible. Dans ce type d'analyse basée sur la compréhension des comportements d'achat des clients, plus l'historique des données est important et homogène, plus le modèle est stable et prédictif.

²⁰ Référence 2006 est une étude de marché monographique qui est réalisée sur la marche de la distribution alimentaire en France depuis 10 ans par TNS Sofres intégré en 2009 au groupe Kantar.

²¹ http://www.secodip.fr/worldpanel/htm/dossier_presse/tns-plusloin.asp/

²² L'INSEE en France (Institut National de la Statistique et des Études Économiques). Il collecte et publie des informations sur l'économie et la société française en menant régulièrement un recensement national. Situé depuis peu à Strasbourg il est la branche française de Euro Stat, le système des statistiques européen. L'INSEE a été créée en 1946.

²³ L'IRIS est une unité géographique liée aux données du recensement

Dans ce document, nous n'avons inséré que quelques exemples de la préparation des données, car notre démonstration est centrée sur le modèle et les résultats. Des détails de la seconde AED sont présentés en annexe 2 (p.39).

Les variables en entrée sont présentées dans le tableau ci-dessous :

Variables	Description
Identifiant Client	Pour lier les tables entre elles sur un même client
Ratio Autre	"
Ratio Bazar	"
Ratio BOF/APF	"
Ratio Charcuterie LS	"
Ratio Clients Animaux	"
Ratio Client Bébés	"
Ratio Clients Boucherie	"
Ratio Boulangerie	"
Ratio Charcuterie	"
Ratio Diététiques Bio	"
Ratio Fromage	"
Ratio Fruits et Légumes	"
Ratio Poisson	"
Ratio Surgelés	"
Ration Vins	"
Ratio DPH	"
Ratio Epicerie	"
Ratio Liquide	"
Ratio Textile	"
Ratio Ultra Frais	"
Ratio Volaille	"
Ratio Premier Prix	"
Ratio Marque Distributeur 1	"
Ratio Marque Distributeur 2	"
Nombre d'enfants au foyer	
CA Filtré 12 mois	Chiffre d'affaires diminué de quelques dépenses hors mag
CA Total	Chiffre d'affaires total
CA Promo Annuel	Chiffre d'affaires annuel réalisé sous promotion
Nb de points transformés sur 12 mois	Nombre de points "promotionnels" utilisés sur une année
CM transformé sur 12 mois	Part du chiffre d'affaires où le client a utilisé des points promotionnels
Cumul nb BA pris	Cumul des bons d'achat que le client a utilisé
PMG 12 mois ²⁴	Segmentation
RFM 3 mois ²⁵	Segmentation

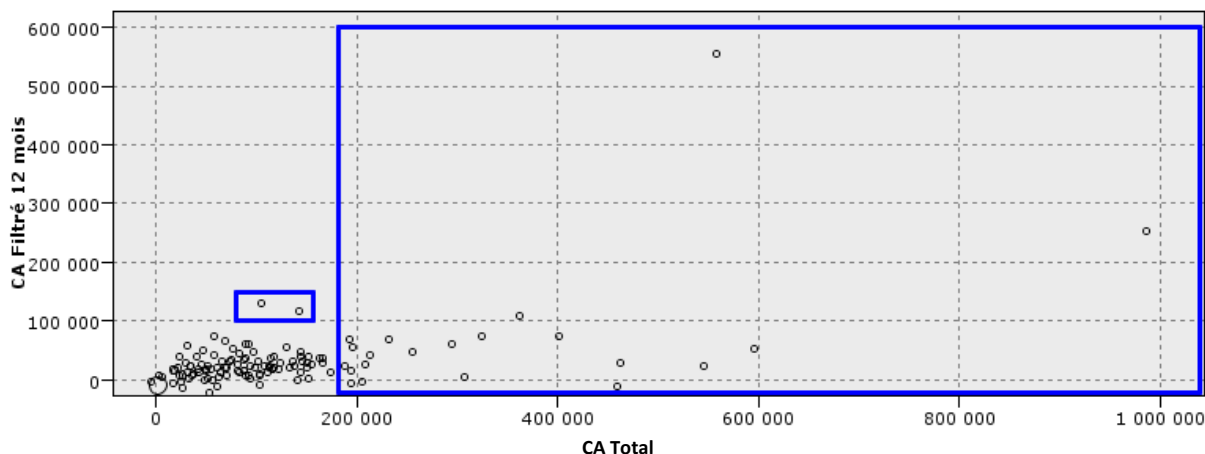
Tableau 94 : Les variables en entrée du modèle

²⁴ Segmentation PMG (Petit, Moyen, Gros) divise les consommateurs en fonction de leurs chiffres d'affaires

²⁵ Segmentation RFM (Récence, Fréquence, Montant) est une segmentation classique en marketing

Identification des clients aberrants :

Nous avons identifié et éliminé de notre étude quelques clients avec des comportements anormaux sur deux variables portant sur le chiffre d'affaire : le « CA filtré 12 mois²⁶ » et le « CA Total²⁷ ».



Graphique 32 : Les outliers

Nous utilisons uniquement ces deux variables pour la détection des clients aberrants, parce qu'elles sont très constitutives du potentiel lui-même.

Discrétisation:

Nous avons discrétisé quelques variables importantes et étudié leur dispersion. Ci-dessous un exemple en quintiles sur la variable CA Total.

	Effectif	%
0	3 985	4,0%
>0 < 280,52€	16 014	16,0%
>=280,52 < 1 538,50€	19 999	20,0%
>=1 538,50€ < 4 4648,2€	19 999	20,0%
>= 4 4648,2€ < 12 155,82€	19 999	20,0%
>= 12 155,82€ <=805 990,68€	20 000	20,0%
NR	4	0,004%
Total	100 000	100%

CA total en cumul					
Moyen	Minimum	Maximum	Ecart type	Médiane	Somme
7 031,88 €	0 €	806 991 €	10 377,17 €	2 757,18 €	703 187 673 €

Tableau 95 : Le CA discrétisé en classe et ses indicateurs de dispersion

²⁶ Le chiffre d'affaires filtré (CA filtré) c'est le chiffre d'affaires cumulé sur les 12 derniers mois moins les réductions de chiffre d'affaires faites par les achats sous promotion.

²⁷ Le chiffre d'affaires total c'est le chiffre d'affaires cumulé sur les 12 mois sans enlever les réductions de chiffre d'affaires faites par les achats sous promotion.

Nous fournissons une AED complète dans l'annexe 2 (p.39) avec une analyse descriptive avec des tableaux et des graphiques, estimation des corrélations, etc....

2.63 Les modèles de classification et la détermination du potentiel client

Le processus de modélisation est divisé entre trois grandes phases :

(1), La méthode de classification elle-même (2), le calcul des niveaux d'évolution et (3) le calcul du potentiel individuel :

1° La méthode de classification : puisque ce modèle doit être appliqué sur de très grandes bases de données avec un grand nombre de variables et d'enregistrements, les SOM (Self Organizing Map ou « carte auto organisatrice ») semblent particulièrement bien adaptés (Kohonen, 1988) :

- Les réseaux de Kohonen permettent d'obtenir des groupes très homogènes et très stables avec de multiples individus et variables,
- Les réseaux de Kohonen autorisent des relations non linéaires complexes entre beaucoup de variables sur beaucoup d'individus,
- Les réseaux de Kohonen prennent bien en compte les valeurs manquantes

2.63.1 Méthode des réseaux de Kohonen

Les réseaux de Kohonen sont une forme de réseaux de neurones qui utilisent un type de carte auto organisatrice qui elle-même représente une catégorie particulière de réseaux de neurones.

La méthode d'analyse est une méthode de classification. Son principal avantage est de convertir un signal d'entrée avec de nombreuses dimensions en une simple carte discrète, avec peu de dimensions. Les réseaux de Kohonen sont une méthode non supervisée, aucune variable cible ne doit être définie.

Les réseaux de Kohonen utilisent la méthode d'apprentissage de Kohonen. Soit un ensemble de m variables pour le n ième enregistrement qui servent d'entrée à un vecteur $x_n = x_{n1}, x_{n2}, \dots, x_{nm}$ dans l'ensemble des poids en cours m pour un nœud sortie donnée j devant être un vecteur de poids $w_j = w_{1j}, w_{2j}, \dots, w_{mj}$. Dans l'apprentissage de Kohonen, les nœuds dans le voisinage des nœuds gagnant ajustent leur poids en utilisant une combinaison linéaire des vecteurs d'entrée dans le vecteur des poids en cours (current) :

$$w_{ij} = w_{ijcurrent} + \eta(x_{ni} - w_{ijcurrent})$$

où η , $0 < \eta < 1$, représente le taux d'apprentissage. Kohonen précise que le taux d'apprentissage doit être une fonction décroissante des cycles d'apprentissage (mené sur l'ensemble des données).

Après chaque itération, il vérifie la fiabilité de ses regroupements antérieurs.

- C'est évidemment un processus assez long quand on l'applique sur un grand nombre d'individus avec beaucoup de variables ; il faut en tenir compte dans le choix de la puissance de calcul de l'ordinateur et éviter de relancer plusieurs fois le calcul.
- Chaque client est alloué à un groupe déterminé dont il est le plus proche sur toutes les variables qui le caractérisent.

Les réseaux Kohonen montrent trois processus caractéristiques :

1° Compétition : Les nœuds en sortie sont en concurrence entre eux pour produire la meilleure valeur pour une fonction de score donnée, en général la minimisation des distances euclidiennes.

2° Coopération : Les nœuds gagnants deviennent alors les centres des voisinages des neurones “excités”

3° Adaptation : Les nœuds dans le voisinage du nœud gagnant, participent à l’adaptation, c’est à dire l’apprentissage. Les poids de ces nœuds sont ajustés dans le but d’améliorer la fonction de score.

L’architecture du réseau :

Chaque neurone de la carte est lié à tous les autres neurones. Chacun d’eux reçoit une copie complète d’un vecteur en entrée.

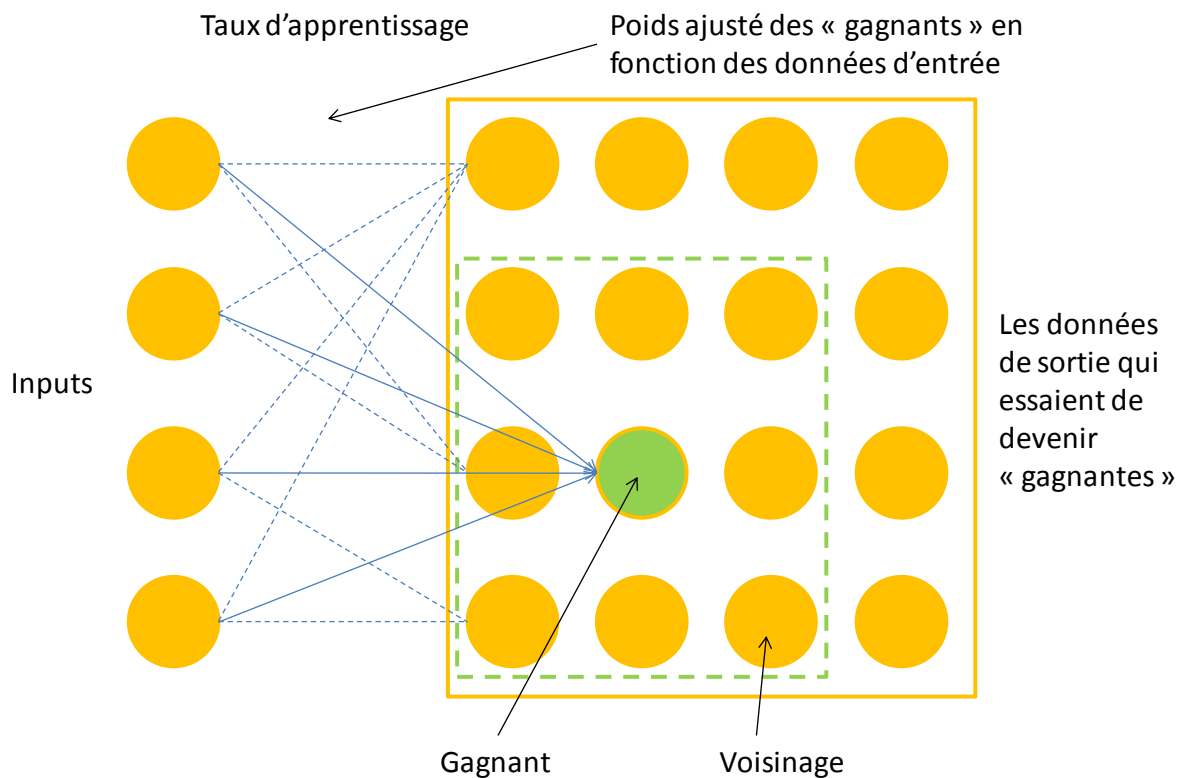


Figure 55 : Le fonctionnement d’un réseau de Kohonen

En représentant cartographiquement l’analyse nous pouvons évaluer la similarité entre les groupes. Deux groupes qui sont proches sur le graphique ont des caractéristiques similaires.

Le but est de trouver une méthode :

- Qui représente le meilleur compromis entre beaucoup de groupes, mais qui s’assure aussi que les groupes sont les plus précis possible avec des clients bien homogènes à l’intérieur, et que les groupes sont les plus différents possibles entre eux.
- Qui nous permet d’obtenir un potentiel client réaliste avec des groupes qui sont stables intrinsèquement.

2° Le calcul des taux « d’évolution » : « L’évolution » est le petit saut de chiffre d’affaire qu’un consommateur doit atteindre pour être regroupé avec les clients qui lui ressemblent le plus, sur toutes les

variables sélectionnées par le modèle mais qui ont un chiffre d'affaire plus élevé que lui. Cela nécessite une méthode de calcul basée sur la division de chaque décile d'une classe de « clones » par la médiane du décile.

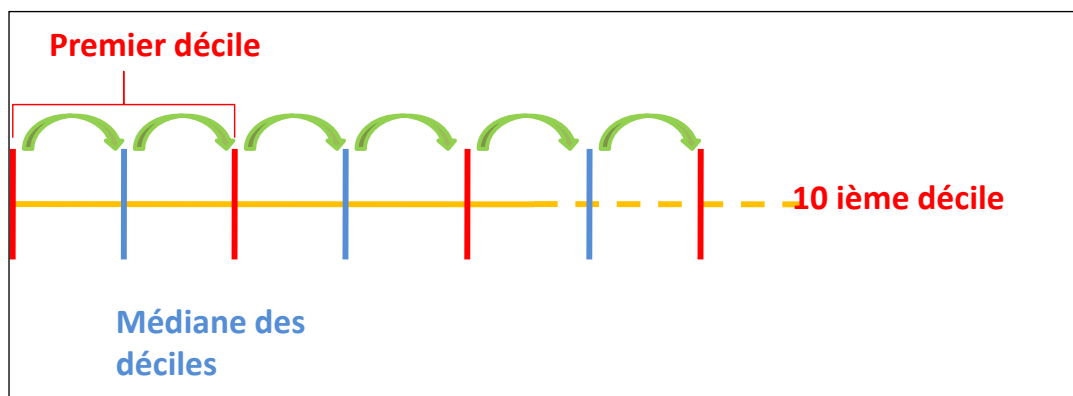


Figure 56 : Le césurage du CA par groupe de clones

Nous retenons la méthode des déciles qui permet de prendre en compte des variations assez significatives de chiffres d'affaires.

Les individus dans un groupe ne doivent pas avoir un écart important à franchir dans le but d'obtenir une estimation réaliste du potentiel²⁸: l'accroissement du chiffre d'affaire potentiel qui pourrait être réalisé après la mise en place des actions marketing appropriées. Nous allons essayer de justifier ce calcul par des moyens méthodologiques. Cette étape va nous donner les taux d'évolution dans les classes de Kohonen.

3° Le calcul du potentiel client individuel : une fois que les taux sont calculés, nous allons calculer, pour chaque client, un potentiel client individuel à atteindre. Ce calcul nécessite des ajustements spécifiques : pour tous les clients avec un taux d'évolution de potentiel supérieur à 100%, leur potentiel représente plus du double de leur CA actuel. Ils se voient allouer le taux d'évolution de potentiel moyen de tous les groupes, sans inclure évidemment ceux à plus de 100%.

2.64 Le développement du modèle

2.64.1 Les objectifs et la méthodologie

Pour compléter les segmentations basées sur le chiffre d'affaire futur, la segmentation PMG (*Brusset 2005*) et la segmentation RFM (*McCartya et Hastak 2007, Chen et al., 2008*), nous calculons un chiffre d'affaire potentiel pour chaque client dans la base de données du programme de fidélisation du distributeur.

Ce score est basé sur une approche itérative nous permettant de prédire la propension à consommer des clients, dans le but de déterminer le chiffre d'affaire potentiel futur.

L'approche consiste à regrouper ensemble les clients qui se ressemblent le plus entre eux, selon les variables présentes en base : variable socio démographiques variables de consommation...

Pour le calcul, nous allons utiliser les données de consommation enregistrées sur une période de 12 mois (de janvier 2006 à décembre 2006).

²⁸ Exemple: Le client A a un chiffre d'affaires actuel de 1 000€. Le client A appartient au premier décile pour un groupe dans lequel tous les clients se ressemblent le plus. Le chiffre d'affaire maximum du client de la limite supérieure de ce décile est de 1 200€. Donc le potentiel est la différence entre les 1 200€ du consommateur max et les 1 000€ du consommateur A : 200€ soit 20%

Les variables utilisées dans le modèle sont celles que nous avons décidé de conserver à l'issue de l'étape de préparation.

Statut familial et données de consommation

Nombre d'enfants au foyer
CA Filtré 12 mois
CA Total
CA Promo annuel
Nb de points transformés sur 12 mois
CM transformé sur 12 mois
Nb de bons de réduction utilisés
PMG 12 mois
RFM 3 mois
Situation familiale

Part de chiffres d'affaires selon les familles de produits

Ratio Clients Autre
Ratio Clients Bazar
Ratio Clients BOF/APF
Ratio Clients Charcuterie LS
Ratio Clients Animaux
Ratio Clients Bébés
Ratio Clients Boucherie
Ratio Clients Boulangerie
Ratio Clients Charcuterie
Ratio Clients Diététique Bio
Ratio Clients Fromage
Ratio Clients Fruits et Légumes
Ratio Clients Poisson
Ratio Clients Surgelés
Ratio Clients Vin AOC
Ratio Clients DPH
Ratio Clients Epicerie
Ratio Clients Liquide
Ratio Clients Textile
Ratio Clients Ultra Frais
Ratio Clients Volaille
Ratio Clients Premier Prix
Ratio Clients Marque Distributeur 1
Ratio Clients Marque Distributeur 2

Tableau 96 : Les variables finales dans le modèle

Les variables non retenues sont éliminées après l'analyse des corrélations pour les variables quantitatives (le chiffre d'affaires et le nombre d'actes d'achat par exemple) ou par des matrices de proximité pour les variables qualitatives. Nous n'utilisons pas d'ACP (Analyse en Composante Principale), de réduction des variables quantitatives, parce que nous voulons garder les informations au niveau le plus désagrégé possible, des variables originales de la base de données.

Les clients inactifs (les clients sans aucune transaction sur la période) sont éliminés.

Le flux PASW Modeler :

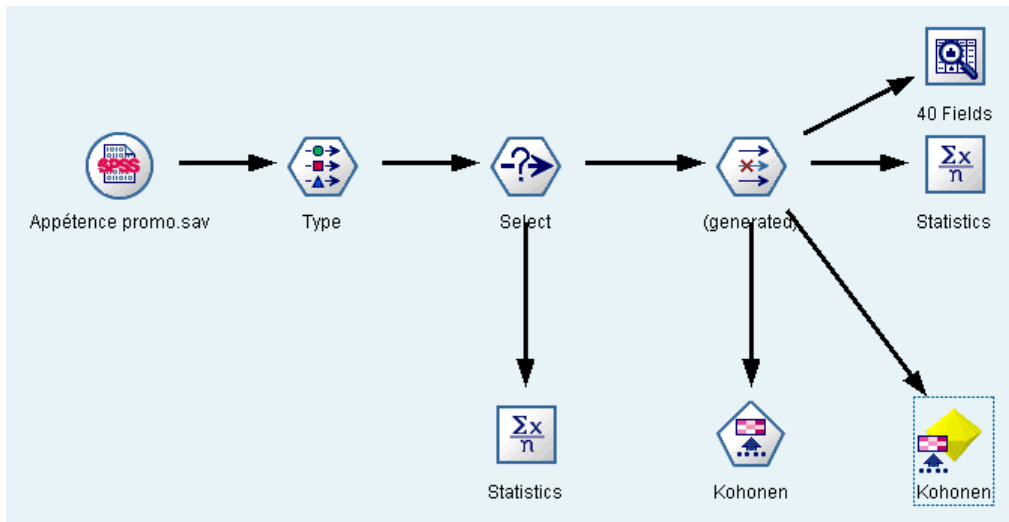


Figure 57 : le modèle sous Modeler

Cette illustration est ici pour montrer la façon dont un modèle est construit sur PASW Modeler de SPSS/IBM (ex Clementine). C'est un logiciel statistique qui utilise un langage objet pour construire les modèles, dans ce logiciel les objets sont appelés des « nœuds ». Le nœud de départ sur la gauche est noté SPSS car ce sont les données en entrée, les autres nœuds suivants les flèches sont des opérations sur les données. Le nœud noté « Kohonen » contient le modèle statistique, le diamant jaune en sortie en bas à droite contient les résultats du modèle.

Nous utilisons une méthode de classification pour créer les groupes de “clones” qui les rend fortement homogènes intrinsèquement mais différents entre eux.

Une fois que les familles de clones ont été obtenues et que les valeurs du potentiel ont été calculées, les principales familles de potentiel sont déterminées :

- « Or » : taux d'évolution supérieur à 20%
- « Argent » : taux d'évolution supérieur ou égale à 15% et inférieur à 20%
- « Bronze » : taux d'évolution inférieur à 15%

Il doit être souligné ici que le potentiel porte sur le potentiel absolu sur douze mois consécutifs.

Ce potentiel est exprimé sous la forme d'un taux. Pour des objectifs opérationnels, les valeurs de potentiels doivent être reclassées en valeur absolue :

- P1: Gros potentiel
- P2: Moyen potentiel
- P3: Petit potentiel

2.64.2 La fiabilité de la méthode de Kohonen

Nous avons testé plusieurs méthodes pour déterminer les convergences entre les classes de Kohonen.

2.1 La visualisation des convergences

Nous avons obtenu 40 classes, numérotées de 00 à 93 (les classes ne sont pas numérotées de manière continue).

Nous voulions obtenir un nombre assez important de classes pour minimiser au maximum l'écart type intra classes.

La cartographie : 00 est la classe de coordonnées 0 sur l'axe des X et 0 sur l'axe des Y et 93 est la classes de coordonnées 9 sur l'axe des X et 3 sur l'axe des Y.

Les classes de Kohonen x la segmentation PMG (12 mois)

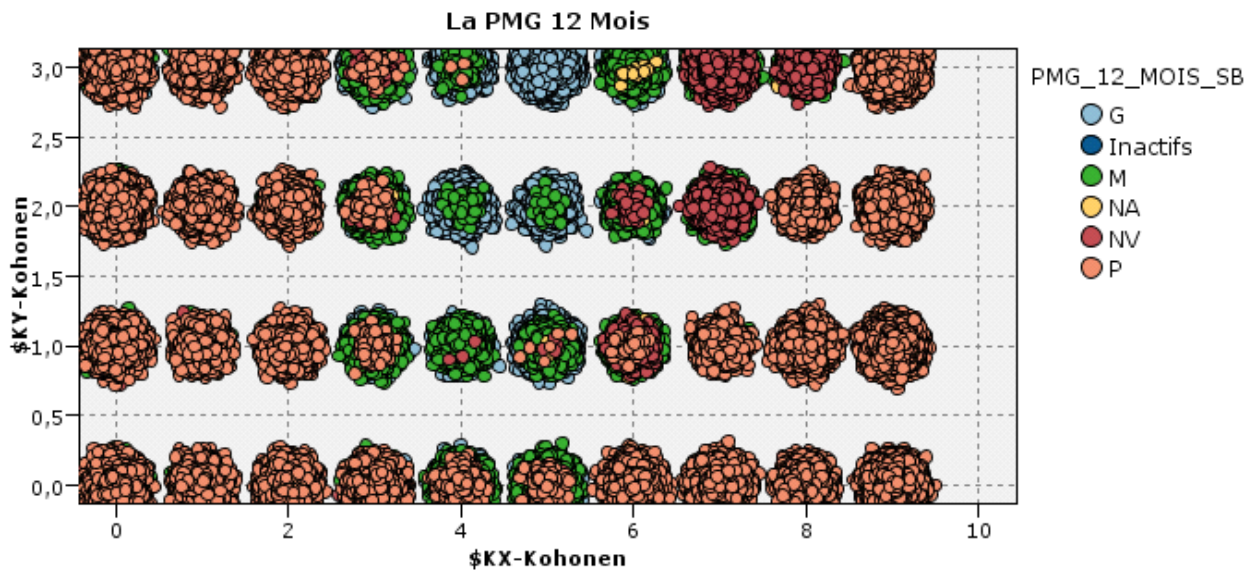


Figure 58 : Affichage de la PMG pas classes de clones

Les couleurs représentent les segments PMG, Nouveaux (NV), Inactifs et Non Affectés à un segment (NA)

Les couleurs qui représentent les segments PMG de clients sont en général bien regroupées, les clients appartenant aux mêmes segments PMG étant ensemble.

Visuellement, la position des PMG dans les classes montre de la stabilité.

Les classes de Kohonen x la segmentation RFM (3 mois)

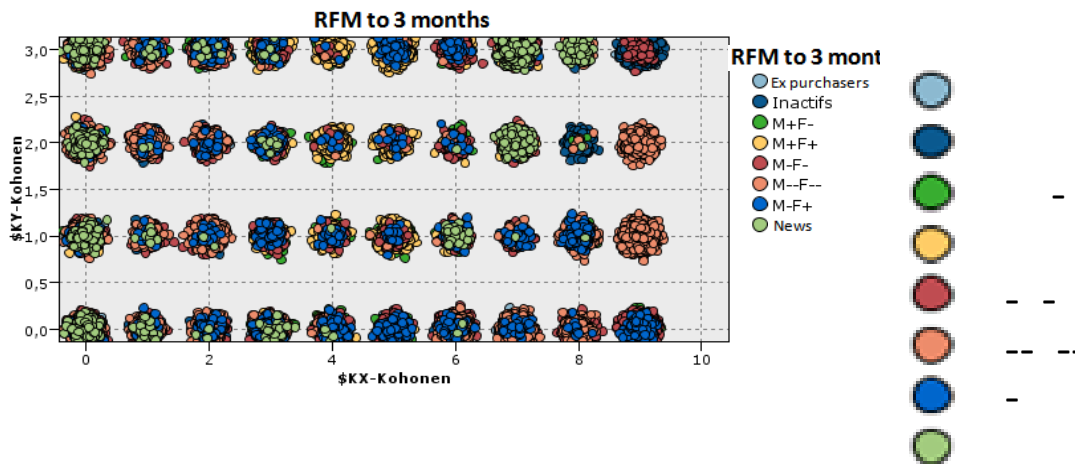


Figure 59 : Affichage de la RFM par classe de clones

Les couleurs représentent les segments RFM, les Anciens clients sans achat depuis plus de 48 mois, les Inactifs sans achat depuis plus de 24 mois, les Nouveaux apparus dans les 6 derniers mois, les M—F--, les M-F-, les M+F-, les M-F+, les M+F+ qui sont les segments basés sur le montant et le fréquence du moins actif au plus actif.

Les couleurs sont généralement assez bien regroupées, les clients appartenant aux mêmes segments RFM (les segments vont de M+F+ à M--F--) se retrouvant majoritairement dans les mêmes classes de Kohonen.

Dans la RFM le mélange de couleurs est plus important à l'intérieur de chaque classe, mais les consommateurs appartenant aux mêmes segments RFM se retrouvent principalement dans les mêmes classes de Kohonen ; l'homogénéité des classes est néanmoins moins évidente que dans le mapping de la PMG.

2.64.3 LA ROBUSTESSE DE LA CLASSIFICATION DE KOHONEN

La distribution de la population est-elle stable ? Nous répondons à cette question de quatre façons différentes :

A/ Y a-t-il une convergence entre le poids des classes entre l'échantillon des observations actives et les observations passives ?

B/ Le regroupement peut-il être reproduit par un réseau Bayésien (Pourret et al, Jensen, Stephenson 2000) ?

C/ Peut t'on reproduire la classification par une segmentation telle que C5.0²⁹ (Quinlan 1993, 1996, 2004) ?

D/ Y a-t-il connectivité des super classes³⁰ ?

A/ La convergence de la méthode

Nous pouvons vérifier l'allocation des clients sur deux échantillons aléatoires de notre corpus de données. Les classes de Kohonen se retrouvent aux mêmes poids entre les deux échantillons.

²⁹ C5.0 est un algorithme qui permet de faire des scores discriminant entre des groupes d'individus. Ici nous l'utilisons pour confirmer la classification de Kohonen. C5.0 devant parvenir avec les variables les plus discriminantes de chaque type à réaffecter les individus dans les mêmes classes que celles où les a affecté l'algorithme de Kohonen.

³⁰ Connectivité des super classes : si on regroupe les classes de Kohonen (en groupe plus important : super classes). Les regroupements sont-ils pertinents, les classes rapprochées sont-elles bien les plus proches ?.

Clones	Total		Echantillon Apprentissage		Echantillon Test	
	Effectif	%	Effectif	%	Effectif	%
KH01	271 944	5,06%	10 949	5,13%	260 995	5,06%
KH02	171 396	3,19%	6 983	3,27%	164 413	3,19%
KH03	261 136	4,86%	10 498	4,92%	250 638	4,86%
KH04	289 912	5,40%	11 508	5,39%	278 404	5,40%
KH05	80 239	1,49%	3 214	1,50%	77 025	1,49%
KH06	40 698	0,76%	1 596	0,75%	39 102	0,76%
KH07	64 515	1,20%	2 550	1,19%	61 965	1,20%
KH08	93 685	1,74%	3 768	1,76%	89 917	1,74%
KH09	95 415	1,78%	3 757	1,76%	91 658	1,78%
KH10	91 169	1,70%	3 681	1,72%	87 488	1,70%
KH11	57 384	1,07%	2 235	1,05%	55 149	1,07%
KH12	181 691	3,38%	7 224	3,38%	174 467	3,38%
KH13	142 728	2,66%	5 624	2,63%	137 104	2,66%
KH14	83 298	1,55%	3 260	1,53%	80 038	1,55%
KH15	65 365	1,22%	2 597	1,22%	62 768	1,22%
KH16	152 665	2,84%	6 153	2,88%	146 512	2,84%
KH17	119 559	2,23%	4 797	2,25%	114 762	2,22%
KH18	45 360	0,84%	1 794	0,84%	43 566	0,84%
KH19	73 151	1,36%	2 783	1,30%	70 368	1,36%
KH20	35 914	0,67%	1 378	0,65%	34 536	0,67%
KH21	120 165	2,24%	4 688	2,20%	115 477	2,24%
KH22	137 752	2,56%	5 462	2,56%	132 290	2,56%
KH23	36 215	0,67%	1 417	0,66%	34 798	0,67%
KH24	267 939	4,99%	10 739	5,03%	257 200	4,99%
KH25	193 624	3,60%	7 581	3,55%	186 043	3,61%
KH26	50 454	0,94%	2 019	0,95%	48 435	0,94%
KH27	26 271	0,49%	1 036	0,49%	25 235	0,49%
KH28	76 724	1,43%	3 082	1,44%	73 642	1,43%
KH29	199 372	3,71%	7 810	3,66%	191 562	3,71%
KH30	28 913	0,54%	1 102	0,52%	27 811	0,54%
KH31	124 878	2,32%	4 922	2,30%	119 956	2,32%
KH32	347 565	6,47%	13 963	6,54%	333 602	6,47%
KH33	75 304	1,40%	2 998	1,40%	72 306	1,40%
KH34	103 656	1,93%	4 107	1,92%	99 549	1,93%
KH35	24 658	0,46%	989	0,46%	23 669	0,46%
KH36	31 206	0,58%	1 272	0,60%	29 934	0,58%
KH37	301 456	5,61%	11 863	5,55%	289 593	5,61%
KH38	252 820	4,71%	10 042	4,70%	242 778	4,71%
KH39	130 904	2,44%	5 193	2,43%	125 711	2,44%
KH40	425 926	7,93%	16 942	7,93%	408 984	7,93%
Total	5 373 026	100,00%	213 576	100,00%	5 159 450	100,00%

Tableau 97 : Les classes de clones

B/ La réallocation en utilisant un réseau Bayésien

Le tableau ci-dessus confirme que l'algorithme est capable de reproduire la distribution sur un ensemble de données plus important (Ensemble d'apprentissage vs Ensemble de test).

Quoi qu'il en soit, c'est en utilisant un autre algorithme que nous pouvons déterminer si oui ou non, la classification peut être reproduite et si elle est stable ou non.

Une fois encore, l'échantillon d'apprentissage est divisé en deux sous échantillons indépendants : l'échantillon d'apprentissage qui inclut 70% des observations, l'échantillon test qui inclut les 30% restant.

Nous utilisons un réseau Bayésien car, pour réallouer 40 groupes, une analyse discriminante n'est pas bien adaptée pour réallouer un si grand nombre de groupe avec autant de variables.

Le réseau Bayésien permet une approche pas à pas ; nous pouvons également fixer le niveau de probabilité des liens que nous retenons entre les variables. Si nous fixons une probabilité de 0,9, les résultats sont présentés sous la forme du graphique ci-dessous.

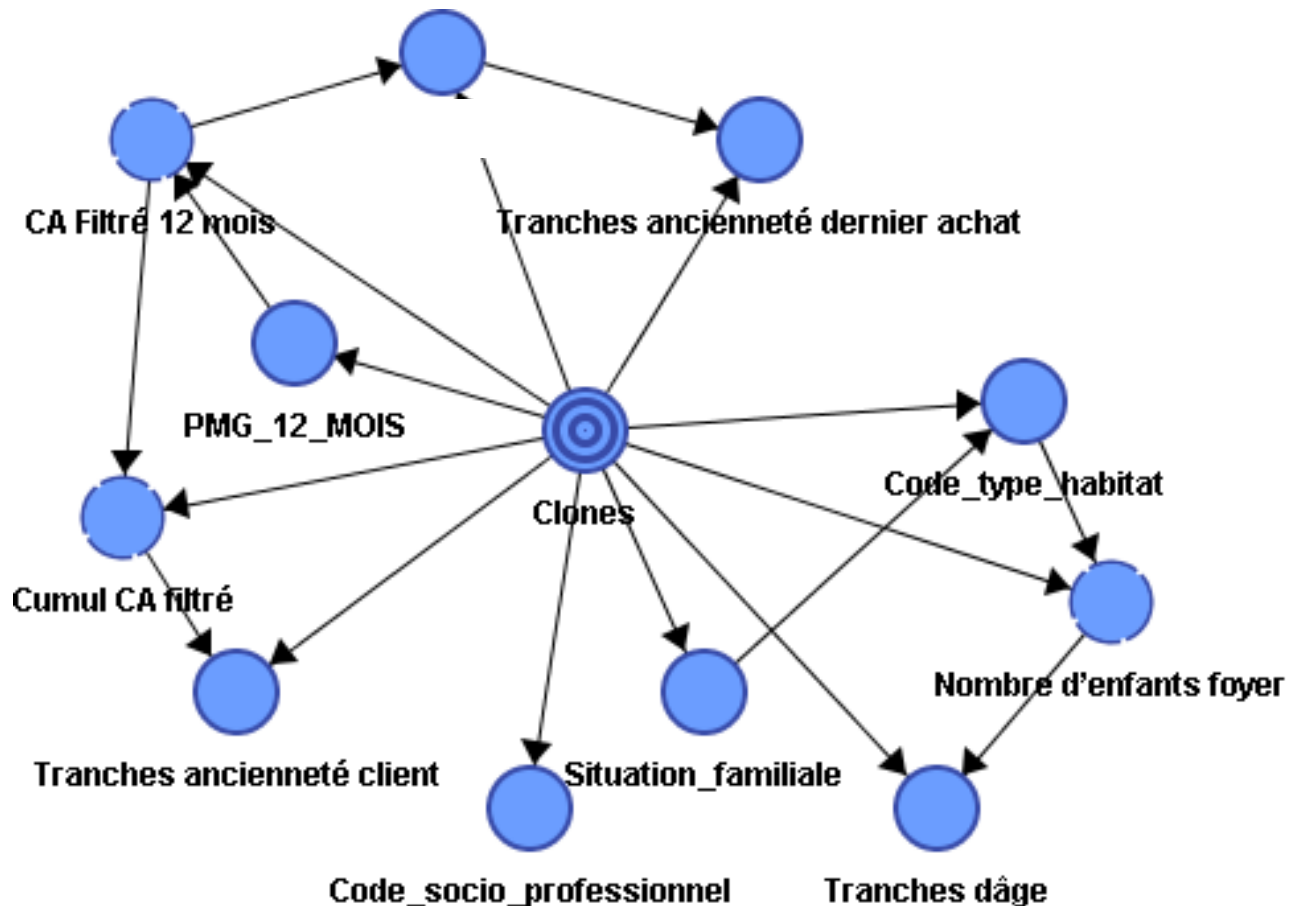


Figure 60 : Le réseau bayésien de réaffectation

Le réseau utilise 11 variables : le chiffre d'affaires et aussi des variables socio démographiques. On peut remarquer que la PMG et la RFM sont retenues par le modèle. Ces résultats valident la représentativité des densités. Ci-dessous le poids des variables dans le modèle.

Importance des noeuds en termes d'apport d'information sur la connaissance de Clones			
Noeud	Information mutuelle	% d'information mutuelle (Part de l'information contenue dans le réseau qui provient de la variable)	Importance relative
Situation familiale	1,0177	20,53%	1
Code socio professionnel	0,9956	20,08%	0,9782
LA RFM Champion	0,9797	19,76%	0,9626
PMG 12 MOIS	0,9674	19,51%	0,9506
CA Filtré 12 mois	0,9297	18,75%	0,9135
Code type habitat	0,9229	18,61%	0,9068
Tranches d'âge	0,8771	17,69%	0,8618
Cumul CA filtré	0,6254	12,61%	0,6145
Tranches ancienneté client	0,5086	10,26%	0,4997
Nombre d'enfants foyer	0,4609	9,30%	0,4528
Tranches ancienneté dernier achat	0,4113	8,30%	0,4041

Tableau 98 : Le tableau d'importance des variables dans le réseau bayésien

Importance des noeuds en termes d'apport d'information sur la connaissance de Clones			
Noeud	Information mutuelle	% d'information mutuelle (Part de l'information contenue dans le réseau qui provient de la variable)	Importance relative
Situation familiale	1,02	20,53%	1
Code socio professionnel	1,00	20,08%	0,98
LA RFM	0,98	19,76%	0,96
PMG 12 MOIS	0,97	19,51%	0,95
CA Filtré 12 mois	0,93	18,75%	0,91
Code type habitat	0,92	18,61%	0,91
Tranches d'âge	0,88	17,69%	0,86
Cumul CA filtré	0,63	12,61%	0,61
Tranches ancienneté client	0,51	10,26%	0,50
Nombre d'enfants foyer	0,46	9,30%	0,45
Tranches ancienneté dernier achat	0,41	8,30%	0,40

Tableau 99 : Le tableau d'importance des variables dans le réseau bayésien

La mesure de Kullback-Leibler http://www.it-innovations.ae/iit005/proceedings/articles/E_6_IIT05_Khalid.pdf vient de la théorie de l'information. C'est une mesure de la convergence entre deux séries après qu'elles aient été recodées dans un format binaire (0 ou 1). Plus la valeur est élevée, plus grande est la probabilité que ces valeurs aient une distribution jointe.

Analyse des relations				
Parent	Enfant	Distance de Kullback-Leibler	Poids relatif	Contribution globale
Clones	Situation familiale	1,02	1	11,44%
Clones	Code socio professionnel	1,00	0,98	11,20%
Clones	PMG 12 MOIS	0,97	0,95	10,88%
Clones	Code type habitat	0,87	0,86	9,79%
Clones	Tranches d'âge	0,75	0,73	8,39%
CA Filtré 12 mois	Cumul CA filtré	0,58	0,57	6,55%
PMG 12 MOIS	CA Filtré 12 mois	0,55	0,54	6,15%
Clones	LA RFM	0,54	0,53	6,07%
Clones	Nombre d'enfants foyer	0,46	0,46	5,22%
Clones	Tranches ancienneté client	0,46	0,45	5,20%
Cumul CA filtré	Tranches ancienneté client	0,44	0,43	4,95%
CA Filtré 12 mois	LA RFM	0,35	0,34	3,89%
LA RFM	Tranches ancienneté dernier achat	0,31	0,30	3,43%
Clones	Cumul CA filtré	0,19	0,18	2,09%
Clones	Tranches ancienneté dernier achat	0,14	0,14	1,63%
Clones	CA Filtré 12 mois	0,11	0,10	1,19%
Code type habitat	Nombre d'enfants foyer	0,06	0,06	0,72%
Situation familiale	Code type habitat	0,06	0,05	0,62%
Nombre d'enfants foyer	Tranches d'âge	0,05	0,05	0,59%

Tableau 100 : Le poids des variables dans la réaffectation des clones

Le résultat du scoring au niveau individuel : sur l'ensemble d'apprentissage, 90,7% des individus sont correctement réaffectés.

Sur l'échantillon test, le pourcentage est de 90.1%

Ci-dessous on trouve les taux en % d'individus correctement réaffectés par les Réseaux Bayésiens pour chacune des 40 classes.

Les classes de Kohonen peuvent être reproduites.

	APPRENTISSAGE			TEST		
	EFFECTIF	PRECISION	FIABILITE	EFFECTIF	PRECISION	FIABILITE
Total	149 241	90,75%		64 335	90,10%	
KH01 prédit / observé	7 120	93,23%	95,81%	3 069	92,66%	94,58%
KH02 prédit / observé	4 637	95,39%	89,86%	2 013	94,86%	88,72%
KH03 prédit / observé	7 265	98,75%	90,55%	3 090	98,38%	90,38%
KH04 prédit / observé	7 639	94,29%	97,51%	3 202	94,01%	97,27%
KH05 prédit / observé	1 845	80,36%	75,83%	727	79,19%	74,11%
KH06 prédit / observé	875	78,41%	86,38%	354	73,75%	78,84%
KH07 prédit / observé	1 413	80,60%	84,97%	619	77,67%	85,14%
KH08 prédit / observé	2 348	90,62%	83,62%	1 057	89,80%	82,84%
KH09 prédit / observé	2 170	82,64%	92,30%	906	80,11%	92,17%
KH10 prédit / observé	2 158	85,81%	91,44%	981	84,13%	90,75%
KH11 prédit / observé	1 356	86,53%	91,87%	569	85,18%	91,63%
KH12 prédit / observé	4 494	90,04%	90,68%	1 999	89,52%	89,08%
KH13 prédit / observé	3 119	79,18%	89,81%	1 330	78,93%	87,44%
KH14 prédit / observé	1 939	85,49%	66,15%	852	85,89%	66,20%
KH15 prédit / observé	1 467	81,73%	68,68%	644	80,30%	69,10%
KH16 prédit / observé	3 768	87,65%	86,07%	1 620	87,38%	85,49%
KH17 prédit / observé	3 124	94,12%	89,51%	1 364	92,29%	90,27%
KH18 prédit / observé	719	56,93%	61,04%	289	54,43%	57,57%
KH19 prédit / observé	1 188	61,62%	71,22%	525	61,40%	69,63%
KH20 prédit / observé	478	49,53%	68,09%	193	46,73%	66,10%
KH21 prédit / observé	2 728	83,68%	94,36%	1 193	83,54%	93,13%
KH22 prédit / observé	3 422	89,42%	83,20%	1 458	89,17%	83,70%
KH23 prédit / observé	587	57,66%	78,27%	206	51,63%	73,84%
KH24 prédit / observé	7 281	96,83%	95,41%	3 114	96,71%	95,64%
KH25 prédit / observé	4 873	92,57%	91,75%	2 136	92,19%	91,60%
KH26 prédit / observé	1 244	88,04%	81,79%	540	89,11%	83,98%
KH27 prédit / observé	372	50,96%	71,68%	144	47,06%	65,75%
KH28 prédit / observé	1 833	84,55%	88,72%	757	82,82%	88,33%
KH29 prédit / observé	4 739	86,92%	87,79%	2 030	86,09%	85,51%
KH30 prédit / observé	291	38,34%	74,42%	120	34,99%	65,93%
KH31 prédit / observé	3 085	90,10%	91,49%	1 341	89,52%	90,12%
KH32 prédit / observé	9 610	98,06%	94,91%	4 078	97,96%	94,38%
KH33 prédit / observé	1 609	76,84%	69,87%	678	75,00%	69,18%
KH34 prédit / observé	2 700	92,75%	93,95%	1 113	93,06%	93,45%
KH35 prédit / observé	659	95,65%	93,34%	275	91,67%	92,28%
KH36 prédit / observé	865	97,19%	96,43%	368	96,34%	96,08%
KH37 prédit / observé	8 174	98,58%	95,93%	3 515	98,43%	95,75%
KH38 prédit / observé	7 029	99,77%	99,79%	2 988	99,70%	99,87%
KH39 prédit / observé	3 551	98,69%	98,67%	1 580	99,06%	98,32%
KH40 prédit / observé	11 664	98,02%	99,31%	4 929	97,76%	99,42%

Tableau 101 : Taux de réaffectation correcte par les réseaux Bayésiens

Dans le tableau ci-dessus, les groupes qui sont pauvrement réaffectés, contiennent un petit nombre de consommateurs.

Même pour ces groupes, la fiabilité reste au-dessus des 65%

C/ Réallocation par un arbre de décision

Le taux de validation croisé est de 94.2% d'individus correctement affectés aux groupes.

L'échantillon test confirme ce taux.

Il y a une forte convergence entre les deux méthodes d'apprentissage supervisées, les Réseaux Bayésiens et le C5.0 sont capable de réaffecter correctement les individus dans les 40 classes.

La robustesse de la classification est également validée par cette méthode.

	APPRENTISSAGE			TEST		
	EFFECTIF	PRECISION	FIABILITE	EFFECTIF	PRECISION	FIABILITE
Total	149 241	97,82%		64 335	94,56%	
KH01 prédit / observé	7 575	99,19%	96,87%	3 208	96,86%	95,25%
KH02 prédit / observé	4 761	97,94%	96,10%	1 972	92,93%	90,67%
KH03 prédit / observé	7 322	99,52%	96,38%	3 018	96,08%	93,35%
KH04 prédit / observé	8 014	98,91%	98,77%	3 292	96,65%	97,34%
KH05 prédit / observé	2 060	89,72%	93,94%	735	80,07%	83,33%
KH06 prédit / observé	976	87,46%	95,03%	352	73,33%	80,00%
KH07 prédit / observé	1 613	92,01%	94,16%	663	83,19%	84,24%
KH08 prédit / observé	2 508	96,80%	93,41%	1 061	90,14%	87,18%
KH09 prédit / observé	2 527	96,23%	98,25%	1 055	93,28%	95,39%
KH10 prédit / observé	2 417	96,10%	96,53%	1 090	93,48%	92,84%
KH11 prédit / observé	1 464	93,43%	97,28%	591	88,47%	93,22%
KH12 prédit / observé	4 883	97,84%	97,39%	2 113	94,63%	94,16%
KH13 prédit / observé	3 908	99,21%	98,41%	1 646	97,69%	96,82%
KH14 prédit / observé	2 225	98,10%	98,98%	959	96,67%	95,33%
KH15 prédit / observé	1 717	95,65%	95,71%	714	89,03%	89,14%
KH16 prédit / observé	4 209	97,91%	95,99%	1 721	92,83%	89,82%
KH17 prédit / observé	3 253	98,01%	98,67%	1 413	95,60%	97,52%
KH18 prédit / observé	1 173	92,87%	95,99%	411	77,40%	81,23%
KH19 prédit / observé	1 830	94,92%	95,46%	688	80,47%	84,42%
KH20 prédit / observé	813	84,25%	97,72%	270	65,38%	78,95%
KH21 prédit / observé	3 206	98,34%	97,27%	1 369	95,87%	94,74%
KH22 prédit / observé	3 743	97,81%	97,40%	1 542	94,31%	94,37%
KH23 prédit / observé	944	92,73%	96,23%	337	84,46%	81,20%
KH24 prédit / observé	7 495	99,68%	99,87%	3 185	98,91%	99,31%
KH25 prédit / observé	5 151	97,85%	97,28%	2 188	94,43%	94,43%
KH26 prédit / observé	1 346	95,26%	93,28%	546	90,10%	87,50%
KH27 prédit / observé	663	90,82%	96,37%	230	75,16%	84,87%
KH28 prédit / observé	2 161	99,68%	99,68%	901	98,58%	98,26%
KH29 prédit / observé	5 252	96,33%	95,14%	2 144	90,92%	88,23%
KH30 prédit / observé	648	85,38%	93,64%	235	68,51%	72,76%
KH31 prédit / observé	3 322	97,02%	98,46%	1 420	94,79%	94,79%
KH32 prédit / observé	9 776	99,76%	99,56%	4 126	99,11%	99,14%
KH33 prédit / observé	2 007	95,85%	98,24%	820	90,71%	94,36%
KH34 prédit / observé	2 855	98,08%	99,37%	1 147	95,90%	96,88%
KH35 prédit / observé	682	98,98%	99,42%	285	95,00%	98,96%
KH36 prédit / observé	872	97,98%	98,31%	366	95,81%	95,81%
KH37 prédit / observé	8 278	99,83%	99,86%	3 559	99,66%	99,55%
KH38 prédit / observé	7 035	99,86%	99,86%	2 988	99,70%	99,87%
KH39 prédit / observé	3 549	98,64%	99,27%	1 547	96,99%	97,11%
KH40 prédit / observé	11 757	98,80%	99,18%	4 928	97,74%	97,93%

Tableau 102 : Le poids des classes de clones entre l'ensemble d'apprentissage et l'ensemble de test

D/ La connectivité des super classes

Nous utilisons une analyse par réseau, qui identifie un petit nombre de variables qui sont les plus importantes pour la classification.

Nous analysons un tableau de contingence entre les 40 groupes et les variables qui contribuent au réseau à plus de 10% des capacités explicatives.

- La situation familiale
- C.S.P.
- R.F.M. à 3 mois
- P.M.G à 3 mois
- Le type d'habitat
- Les classes d'âges
- Le chiffre d'affaire filtré en cumul
- Les classes de séniorité des clients

Sur cette table, l'échelle utilisée et la distance du Khi^2 (Otto, 2007, Meunier et al, Romesburg, 2004) et une méthode d'agrégation de Ward est appliquée (Clarke and Sun, 1997, Barnier 2008).

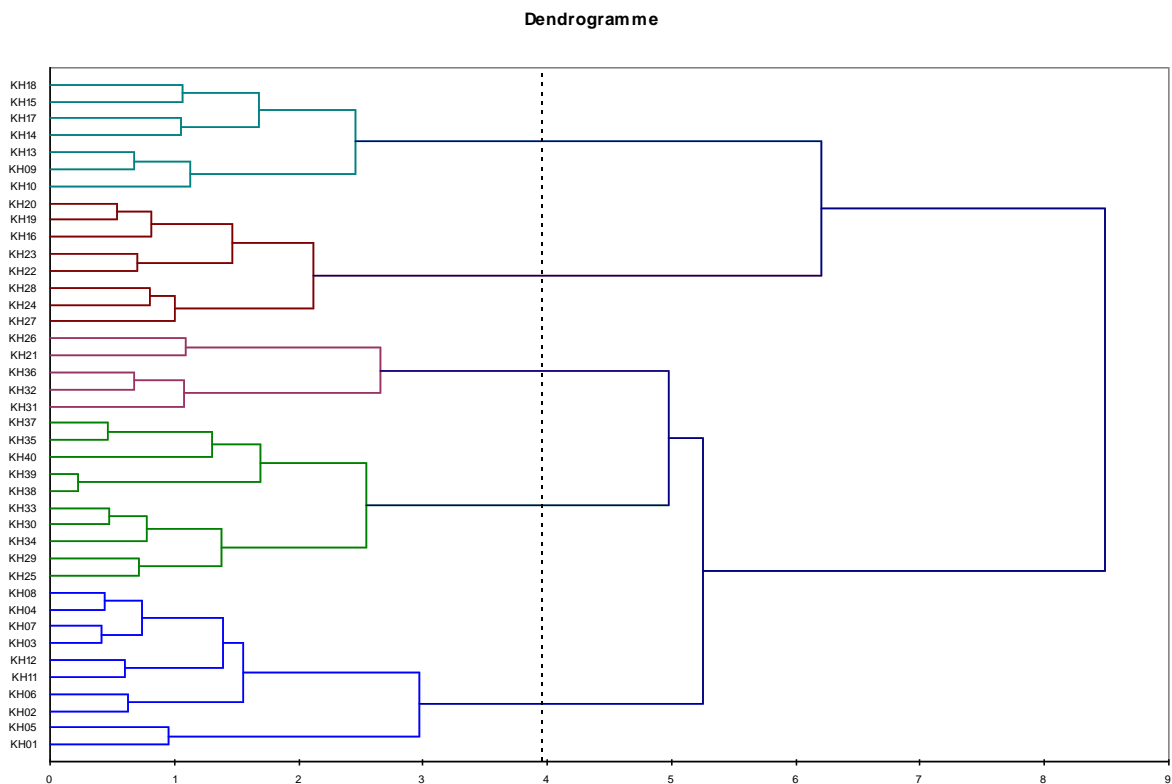


Figure 61 : Le dendrogramme des super classes

Les césures des écarts types pour une classification optimale :

Intra-groupes	89790172,495
Inter-groupes	25701803,959
Total	115491976,454

Distances entre les centres des classes :

	1 (KH08)	2 (KH14)	3 (KH19)	4 (KH31)	5 (KH39)
1 (KH08)	0	7116,532	6162,291	9155,711	9754,960
2 (KH14)	7116,532	0	5004,158	8103,255	11777,617
3 (KH19)	6162,291	5004,158	0	8507,614	11356,060
4 (KH31)	9155,711	8103,255	8507,614	0	9403,860
5 (KH39)	9754,960	11777,617	11356,060	9403,860	0

Résultats par groupe :

Classe	1	2	3	4	5
Objets	10	7	8	5	10
Somme des poids	10	7	8	5	10
Variance intra-classe	72493506,744	25490092,048	67169820,393	125787548,900	151548331,778
Distance minimale au barycentre	3535,855	3419,845	3465,065	3277,114	4143,086
Distance moyenne au barycentre	7323,184	4577,213	6560,727	8500,865	10606,386
Distance maximale au barycentre	14057,216	5976,284	16549,137	18792,278	22881,067
	KH01	KH09	KH16	KH21	KH25
	KH02	KH10	KH19	KH26	KH29
	KH03	KH13	KH20	KH31	KH30
	KH04	KH14	KH22	KH32	KH33
	KH05	KH15	KH23	KH36	KH34
	KH06	KH17	KH24		KH35
	KH07	KH18	KH27		KH37
	KH08		KH28		KH38
	KH11				KH39
	KH12				KH40

Une vérification est réalisée pour s'assurer que le respect des limites inférieures et supérieures de l'ordre des groupes : le clone 40 n'est pas regroupé avec le clone 3. C'est un des critères de « qualité » d'une carte de Kohonen.

En conclusion, la forme de la classification obtenue par l'algorithme de Kohonen, satisfait les critères de stabilité et de reproductibilité qui garantissent un potentiel robuste et durable.

2.64.4 Calcul des potentiels

Nous divisons le chiffre d'affaires annuel (le chiffre d'affaires sur 12 mois) en déciles.

Pour chacune des classes obtenues par la méthode de Kohonen, nous avons calculé un potentiel basé sur le chiffre d'affaires.

Nous divisons le chiffre d'affaires total de chaque classe de clones en déciles, nous calculons la médiane de chaque décile.

Nous allouons par groupe une valeur de chiffre d'affaires potentiel déduite du calcul des taux de potentiel d'accroissement entre les médianes et les déciles.

Pour chaque groupe de clones, le taux de croissance de la mesure d'accroissement du chiffre d'affaires, va de la limite inférieure du décile, à la limite supérieure.

20 taux d'accroissement par groupe de clones sont déterminés :

- Entre la limite inférieure du premier décile et la médiane du premier décile : Tx01
- Entre la médiane du premier décile et limite supérieure du premier décile du second décile : Tx02
- Entre la limite inférieure du premier décile et la médiane du second décile : Tx03
- ...
- Entre la médiane du huitième décile et limite supérieure du huitième décile : Tx18
- Entre la limite inférieure du neuvième décile et la limite médiane du neuvième décile : Tx19
- Entre la médiane du neuvième décile et limite supérieure du dixième décile : Tx20

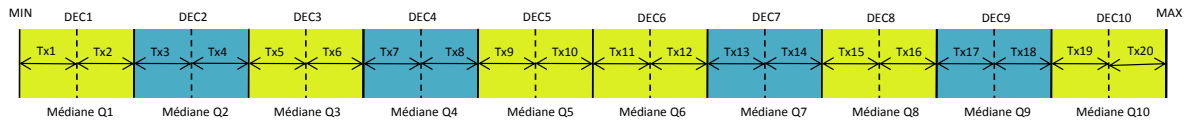


Figure 62 : Césurage détaillé des classes de clones

Pour chaque classe de Kohonen, un client pour lequel le chiffre d'affaire filtré, se situe entre le minimum et la médiane du premier décile, se verra attribué un taux d'évolution de 1 (Tx01).

Un client pour lequel le chiffre d'affaires est compris entre la médiane et la limite supérieure de premier décile, se verra attribuer un taux d'accroissement égal au taux de 2 (Tx02) etc....

Chaque client se voit attribuer un taux d'évolution. Le taux multiplié par chiffre d'affaires nous permet d'estimer un potentiel de chiffre d'affaires à chaque client.

4.1 – Les limites et les médianes des déciles par groupe de Kohonen :

CA en euros	CLASSE	EFF	Moyenne		1er Décile		2ème Décile		3ème Décile		4ème Décile		5ème Décile		6ème Décile		7ème Décile		8ème Décile		9ème Décile		10ème Décile	
			Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup	Médiane	Borne sup
KH01	00	271 944	683	354	16,6	34,6	56,4	81,9	111,9	146,5	187,5	234,1	289,4	354,5	431,5	523,7	634,9	769,0	935,8	1 151,5	1 409,5	1 788,9	2 359,2	62 857,0
KH02	01	171 396	730	396	19,0	39,9	64,5	93,1	126,3	164,1	209,3	262,8	323,4	396,1	482,8	584,4	705,7	847,6	1 023,5	1 245,5	1 521,7	1 895,8	2 430,8	12 322,8
KH03	02	261 136	770	414	22,2	45,8	72,5	102,8	137,3	177,3	223,6	277,3	340,4	414,1	501,1	603,4	724,0	867,4	1 041,4	1 266,4	1 570,2	1 978,8	2 582,6	40 591,8
KH04	03	289 912	1 353	658	31,3	65,4	104,5	148,8	201,6	262,2	335,5	423,3	529,1	658,0	812,4	1 000,2	1 230,8	1 510,2	1 864,7	2 314,6	2 895,5	3 689,7	4 995,4	30 647,2
KH05	10	80 239	471	303	17,8	35,4	55,7	79,8	106,7	136,4	170,6	209,7	253,7	303,4	359,7	425,9	499,4	582,3	675,7	784,9	911,0	1 058,1	1 415,4	11 734,8
KH06	11	49 698	452	312	19,7	39,1	61,4	89,5	113,9	146,0	180,7	217,6	262,3	311,7	367,8	436,3	502,8	584,4	677,0	775,0	885,0	1 007,7	1 152,1	8 501,0
KH07	12	64 515	469	360	23,9	49,5	77,3	107,4	140,9	177,0	217,5	261,5	308,6	360,1	417,2	480,0	549,0	622,9	705,6	797,8	900,8	1 012,2	1 139,9	6 455,0
KH08	13	93 686	697	467	27,9	57,4	91,0	127,5	168,3	215,1	268,2	326,7	391,7	466,7	550,4	643,3	748,6	862,3	995,6	1 146,6	1 381,5	1 712,6	2 144,6	23 753,5
KH09	20	95 415	432	313	20,5	39,6	61,7	86,1	113,2	144,8	179,9	219,5	263,6	313,0	369,2	430,3	499,9	576,5	662,0	756,4	863,1	978,8	1 103,9	11 486,2
KH10	21	91 169	519	452	32,7	65,5	101,0	140,9	183,5	229,8	280,1	333,2	391,0	451,7	516,7	581,1	651,4	722,7	797,6	877,1	957,5	1 042,2	1 130,6	7 150,8
KH11	22	57 384	580	454	27,6	56,8	92,0	129,3	170,7	219,8	271,1	327,2	390,2	454,2	525,6	600,9	683,3	771,7	865,6	965,8	1 071,6	1 221,2	1 631,8	5 563,9
KH12	23	161 891	850	670	37,7	80,4	129,2	184,3	245,6	314,3	390,1	475,0	567,6	669,9	783,8	905,1	1 037,8	1 180,3	1 330,4	1 501,9	1 694,4	1 925,0	2 202,6	8 916,9
KH13	30	142 728	988	779	39,4	85,9	144,8	214,2	293,1	377,6	472,0	571,3	674,2	799,0	888,3	1 000,1	1 115,8	1 257,1	1 429,2	1 627,0	1 853,7	2 138,4	2 541,7	16 298,7
KH14	31	83 298	2 971	2 592	1 268,3	1 393,2	1 525,2	1 658,7	1 802,7	1 946,8	2 099,5	2 257,7	2 430,3	2 591,5	2 764,1	2 946,5	3 156,4	3 396,3	3 675,6	4 009,5	4 422,9	5 029,0	6 052,2	33 406,4
KH15	32	65 365	2 488	2 079	1 209,3	1 295,8	1 383,0	1 472,9	1 564,9	1 660,3	1 759,2	1 863,6	1 971,1	2 078,9	2 189,0	2 304,7	2 422,7	2 594,9	2 876,1	3 239,7	3 701,3	4 336,4	5 423,0	40 744,4
KH16	33	152 665	4 203	3 745	1 584,7	1 976,5	2 325,5	2 579,3	2 746,6	2 917,8	3 107,6	3 304,4	3 520,0	3 745,0	3 991,8	4 256,8	4 549,6	4 879,7	5 264,2	5 707,4	6 261,8	6 995,1	8 178,9	23 610,6
KH17	40	119 559	2 646	2 254	1 186,1	1 270,9	1 377,5	1 486,5	1 604,0	1 722,5	1 843,4	1 971,9	2 110,0	2 253,5	2 405,7	2 574,7	2 784,9	3 027,5	3 304,2	3 625,9	4 041,6	4 607,9	5 558,8	24 745,9
KH18	41	45 360	3 263	2 888	1 336,7	1 512,5	1 681,1	1 859,9	2 035,6	2 215,0	2 397,4	2 588,8	2 720,1	2 888,3	3 069,4	3 267,3	3 488,7	3 738,1	4 025,8	4 393,4	4 841,6	5 400,8	6 493,2	25 585,5
KH19	42	73 151	4 554	4 065	2 621,0	2 755,0	2 893,3	3 034,8	3 177,2	3 334,5	3 500,5	3 673,4	3 862,7	4 064,4	4 274,6	4 499,5	4 752,2	5 056,5	5 391,7	5 793,1	6 303,2	7 000,8	8 172,5	28 501,8
KH20	43	35 914	4 571	4 070	2 570,7	2 705,2	2 850,9	3 004,1	3 155,6	3 312,4	3 487,7	3 669,4	3 861,8	4 069,6	4 284,9	4 527,1	4 797,5	5 103,9	5 447,5	5 861,8	6 382,0	7 126,6	8 333,2	40 421,4
KH21	50	120 165	2 291	1 943	1 212,4	1 282,5	1 356,5	1 432,3	1 510,3	1 590,1	1 674,2	1 760,4	1 849,6	1 942,5	2 038,5	2 141,3	2 246,1	2 355,7	2 469,4	2 762,9	3 232,9	3 893,1	4 958,5	15 664,1
KH22	51	137 752	4 572	4 109	2 515,0	2 664,9	2 817,2	2 978,6	3 143,5	3 319,0	3 501,2	3 691,9	3 891,3	4 109,4	4 339,2	4 594,2	4 870,6	5 186,9	5 539,8	5 956,9	6 479,9	7 167,4	8 304,3	33 452,0
KH23	52	36 215	4 336	3 849	2 599,6	2 709,0	2 823,5	2 941,5	3 069,1	3 201,5	3 344,2	3 504,2	3 667,1	3 848,9	4 045,0	4 260,8	4 498,7	4 770,8	5 088,1	5 474,5	5 962,4	6 610,9	7 681,7	26 524,0
KH24	53	267 939	4 633	4 099	2 637,0	2 772,2	2 912,9	3 057,5	3 208,4	3 365,3	3 528,8	3 707,9	3 896,9	4 099,3	4 320,3	4 562,6	4 827,9	5 128,9	5 477,8	5 896,2	6 428,4	7 154,5	8 393,3	83 222,3
KH25	60	193 624	552	436	24,0	50,9	83,8	121,2	163,1	210,5	260,4	314,0	374,5	439,1	502,9	573,0	646,0	722,1	802,9	888,8	977,2	1 070,2	1 219,1	16 014,5
KH26	61	50 454	1 692	1 669	161,2	1 182,5	1 234,0	1 285,9	1 342,0	1 400,7	1 463,2	1 528,9	1 595,8	1 668,7	1 741,8	1 821,0	1 904,3	1 990,9	2 082,9	2 178,5	2 281,0	2 391,4	2 517,7	10 099,3
KH27	62	26 271	2 766	2 552	1 348,1	1 538,1	1 709,0	1 852,4	1 978,5	2 103,6	2 221,7	2 339,1	2 452,7	2 551,7	2 652,2	2 772,0	2 905,3	3 059,1	3 241,7	3 468,9	3 781,4	4 245,2	5 031,9	29 806,6
KH28	63	76 724	3 224	2 933	1 632,4	1 913,0	2 120,9	2 299,2	2 463,6	2 559,5	2 640,9	2 728,8	2 827,7	2 933,2	3 052,2	3 184,3	3 340,5	3 517,5	3 732,8	3 997,6	4 345,7	4 831,4	5 676,7	43 822,0
KH29	70	199 372	365	276	22,6	43,0	64,6	87,8	112,6	140,2	169,7	202,1	237,2	276,0	320,2	368,9	424,1	486,6	557,6	641,4	738,9	852,2	989,6	1 160,4
KH30	71	28 913	598	565	58,5	119,2	174,8	228,6	281,5	335,2	390,3	444,8	502,3	564,5	622,8	684,5	744,3	808,2	873,5	940,2	1 009,9	1 079,9	1 149,5	4 207,3
KH31	72	124 878	1 292	1 442	25,9	62,7	118,5	215,8	498,0	1 194,3	1 251,5	1 313,0	1 375,6	1 441,9	1 512,9	1 589,5	1 672,5	1 763,4	1 859,4	1 965,2	2 078,4	2 206,0	2 347,0	56 362,3
KH32	73	347 565	1 220	1 415	14,7	32,2	59,1	102,6	184,8	445,0	1 200,4	1 267,9	1 339,4	1 414,7	1 493,6	1 577,2	1 665,5	1 759,2	1 859,9	1 967,5	2 084,2	2 212,6	2 351,1	42 716,8
KH33	80	75 304	527	466	72,8	124,0	169,2	212,6	257,3	301,7	348,6	395,7	444,1	494,5	547,2	602,7	659,6	719,7	783,0	849,3	917,2	989,5	1 072,6	1 160,4
KH34	81	103 658	417	334	36,0	62,9	90,2	118,1	146,9	178,4	211,3	248,3	289,0	334,2	383,6	437,9	498,8	567,4	643,6	728,1	821,9	926,2	1 037,9	1 160,4
KH35	82	24 658	458	345	16,7	51,6	82,6	115,2	149,1	184,6	219,3	257,4	297,8	345,3	395,3	449,8	507,5	575,5	647,2	729,4	822,4	930,0	1 057,4	8 368,6
KH36	83	31 206	982	1 199	0,9	3,8	9,1	15,0	22,1	32,4	47,8	76,5	115,9	1 199,2	1 264,9	1 335,1	1 417,4	1 509,6	1 622,1	1 759,1	1 930,6	2 162,3	2 502,6	57 514,7
KH37	90	301 456	623	619	151,9	214,8	269,7	322,0	371,7	421,1	470,2	520,2	569,5	619,2	669,1	719,9	771,5	823,8	876,4	930,6	985,9	1 042,5	1 100,5	1 160,4
KH38	91	252 820	322	241	27,8	47,1	66,0	86,4	107,8	130,7	155,3	181,6	209,5	241,0	275,7	314,9	358,7	409,4	467,9	537,8	624,2	738,1	897,9	1 160,4
KH39	92	130 904	265	171	21,4	35,3	48,6	62,5	77,3	92,6	109,4	127,5	147,5	170,5	196,9	228,3	263,3	306,0	357,8	423,1	509,1	628,7	813,8	11 915,8
KH40	93	425 926	143	62	5,5	9,6	13,7	18,2	23,3	29,0	36,4	43,1	52,0	62,4	74,8	89,8	108,4	131,6	163,2	206,9	271,4	377,9	584,3	26 849,0
TOTAL		5 373 026	1 390	687	31,2	2 772,2	90,0	3 057,5	181,3	3 365,3	316,4	3 707,9	461,9	4 109,4	632,4	4 594,2	780,9	5 186,9	1 055,1	5 956,9	1 556,4	7 167,4	2 282,5	83 222,3

CLASSE	EFF.	TX1	TX2	TX3	TX4	TX5	TX6	TX7	TX8	TX9	TX10	TX11	TX12	TX13	TX14	TX15	TX16	TX17	TX18	
KH01	00	5 789	108,63	63,13	45,26	36,55	30,96	27,95	24,86	23,62	22,48	21,74	21,36	21,22	21,13	21,69	23,05	22,41	25,49	33,37
KH02	01	1 580	109,62	61,89	44,29	35,62	29,93	27,54	25,55	23,07	22,50	21,87	21,06	20,76	20,11	20,74	21,69	22,18	24,58	28,22
KH03	02	3 067	106,41	58,41	41,79	33,61	29,12	26,10	24,00	22,76	21,65	21,02	20,42	19,98	19,80	20,06	21,60	23,99	26,03	30,51
KH04	03	2 345	109,01	59,80	42,32	35,48	30,08	27,94	26,19	24,98	24,36	23,48	23,11	23,05	22,70	23,48	24,13	25,23	27,30	35,39
KH05	04	1 781	99,49	57,27	43,20	33,71	27,81	25,14	22,89	20,97	19,60	18,54	18,42	17,25	16,61	16,03	16,17	16,06	16,15	33,77
KH06	05	4 534	98,18	56,94	40,91	31,68	28,12	23,79	20,39	20,55	18,84	18,01	16,98	16,84	16,24	15,86	14,47	14,19	13,86	14,33
KH07	10	688	106,73	56,18	38,95	31,22	25,67	22,87	20,23	17,99	16,71	15,86	15,05	14,38	13,45	13,28	13,06	12,91	12,36	12,62
KH08	11	378	105,92	58,47	40,16	32,00	27,76	24,72	21,80	19,91	19,14	17,94	16,89	16,39	15,16	15,46	15,16	20,49	23,96	25,23
KH09	12	398	92,99	55,61	39,55	31,49	27,96	24,22	22,00	20,07	18,76	17,97	16,54	16,18	15,32	14,83	14,25	14,10	13,41	12,78
KH10	13	472	100,21	54,17	39,50	30,19	25,21	21,90	18,98	17,34	15,53	14,38	12,46	12,09	10,96	10,36	9,96	9,18	8,84	8,49
KH11	14	627	106,06	62,07	40,46	32,05	28,76	23,36	20,69	19,26	16,38	15,72	14,34	13,70	12,94	12,17	11,58	10,96	13,96	33,63
KH12	15	1 388	113,10	60,81	42,60	33,25	28,00	24,11	21,77	19,49	18,02	17,02	15,47	14,66	13,74	12,72	12,89	12,82	13,61	14,42
KH13	20	4 594	118,30	68,56	47,95	36,82	28,83	25,00	21,04	18,01	15,55	14,03	12,58	11,58	12,66	13,69	13,84	13,93	15,36	18,86
KH14	21	3 075	9,85	9,47	8,76	8,68	7,99	7,84	7,53	7,65	6,63	6,66	6,60	7,19	7,53	8,22	9,08	10,31	13,70	20,35
KH15	22	420	7,16	6,73	6,50	6,25	6,09	5,96	5,93	5,77	5,47	5,30	5,28	5,12	7,11	10,84	12,64	14,25	17,16	25,06
KH16	23	3 872	24,73	17,66	10,91	6,49	6,23	6,50	6,33	6,52	6,39	6,59	6,64	6,88	7,26	7,88	8,42	9,71	11,71	16,92
KH17	24	959	8,99	8,39	8,05	7,76	7,39	7,02	6,98	7,00	6,80	6,75	7,02	8,16	8,71	9,14	9,73	11,26	14,23	20,64
KH18	25	2 661	13,15	11,15	10,64	9,44	8,81	8,23	7,15	5,89	6,18	6,27	6,45	6,78	7,15	7,69	9,13	10,20	12,58	19,12
KH19	30	883	5,11	5,02	4,89	4,69	4,95	4,98	4,94	5,15	5,22	5,17	5,26	5,68	6,34	6,74	7,34	8,81	11,07	16,74
KH20	31	1 037	5,23	5,39	5,38	5,04	4,97	5,29	5,21	5,24	5,38	5,29	5,65	5,97	6,39	6,73	7,61	8,87	11,67	16,93
KH21	32	66	5,78	5,77	5,58	5,45	5,28	5,29	5,15	5,07	5,02	4,94	5,04	4,89	4,88	4,83	11,89	17,01	20,42	27,37
KH22	33	1 429	5,96	5,71	5,73	5,53	5,58	5,49	5,45	5,40	5,60	5,59	5,88	6,02	6,49	6,81	7,53	8,78	10,61	15,87
KH23	34	381	4,21	4,23	4,18	4,34	4,31	4,46	4,78	4,65	4,96	5,10	5,34	5,58	6,05	6,65	7,59	8,91	10,88	16,20
KH24	35	3 392	5,12	5,07	4,96	4,94	4,89	4,86	5,08	5,10	5,19	5,39	5,61	5,82	6,23	6,80	7,64	9,03	11,30	17,32
KH25	40	1 508	111,69	64,82	44,56	34,56	29,06	23,70	20,59	18,95	16,75	15,12	14,15	12,73	11,79	11,19	10,69	9,95	9,52	13,91
KH26	41	1 340	633,59	4,35	4,21	4,36	4,37	4,47	4,49	4,37	4,57	4,38	4,55	4,58	4,55	4,62	4,59	4,70	4,84	5,28
KH27	42	1 525	14,09	11,11	8,39	6,81	6,33	5,61	5,28	4,86	4,04	3,94	4,52	4,81	5,29	5,97	7,01	9,01	10,27	18,53
KH28	43	956	17,19	10,87	8,40	7,15	3,89	3,18	3,33	3,63	3,73	4,06	4,33	4,91	5,30	6,12	7,09	8,71	11,18	17,50
KH29	44	684	90,17	50,40	35,81	28,31	24,52	21,03	19,11	17,34	16,37	16,02	15,22	14,95	14,74	14,58	15,04	15,19	15,33	16,13
KH30	45	2 217	103,76	46,45	30,93	23,13	19,06	16,44	13,96	12,92	12,38	10,34	9,90	8,74	8,59	8,08	7,63	7,41	6,93	6,45
KH31	50	3 839	143,57	89,13	81,91	130,78	139,82	4,80	4,91	4,77	4,82	4,92	5,06	5,23	5,43	5,44	5,69	5,76	6,14	6,39
KH32	51	277	118,82	83,58	73,58	80,03	140,84	169,74	5,62	5,64	5,62	5,58	5,60	5,60	5,63	5,70	5,80	5,93	6,16	6,26
KH33	52	1 649	70,22	35,66	26,38	21,03	17,27	15,53	13,53	12,21	11,36	10,67	10,13	9,45	9,11	8,80	8,47	7,99	7,89	8,40
KH34	53	799	74,74	43,50	30,94	24,35	21,45	18,44	17,50	16,40	15,64	14,78	14,16	13,91	13,76	13,43	13,12	12,88	12,70	12,05
KH35	54	392	208,99	60,21	39,48	29,41	23,76	18,82	17,38	15,70	15,94	14,46	13,78	12,83	13,39	12,46	12,71	12,76	13,08	13,70
KH36	55	4 055	322,22	139,61	65,18	46,88	46,76	47,38	60,19	103,73	669,03	5,48	5,54	6,17	6,50	7,46	8,45	9,74	12,00	15,74
KH37	60	353	41,38	25,55	19,40	15,44	13,28	11,67	10,62	9,49	8,73	8,05	7,60	7,16	6,78	6,38	6,18	5,94	5,75	5,66
KH38	61	1 000	69,11	40,25	30,82	24,84	21,22	18,77	16,95	15,39	15,04	14,40	14,20	13,93	14,13	14,28	14,93	16,07	18,24	21,65
KH39	62	904	64,60	37,69	28,67	23,61	19,82	18,09	16,55	15,76	15,56	15,50	15,91	15,33	16,25	16,93	18,25	20,33	23,48	29,45
KH40	63	1 670	73,32	43,87	32,68	27,54	24,82	22,54	21,23	20,67	20,03	19,75	20,16	20,60	21,49	24,02	26,72	31,21	39,22	54,61
MOYENNE			39,16	38,60	29,60	23,07	18,69	16,44	15,15	13,07	12,38	11,70	11,36	11,18	11,19	11,45	12,07	12,98	14,62	19,14

Tableau 103 : Taux d'évolution de CA par classe de clone

Ce tableau nous donne le taux d'évolution qui est affecté à chaque consommateur en fonction de son groupe de clones et de son chiffre d'affaires actuel.

En bas du tableau, nous pouvons voir la moyenne de ces taux : si nous obtenons un taux d'évolution aberrant (>100%), nous remplaçons cette valeur excessivement élevée par la moyenne du taux calculée sur tous les groupes sauf ceux contenant une valeur extrême. Le tableau ci-dessous monte les corrections effectuées.

Par exemple, si un client appartient au groupe de clones 00 (KH01) et a un chiffre d'affaires en dessous de 21.5 euros, il aura un taux d'évolution égale au taux 1 : soit 89%.

Si un client appartient à la classe de clones 00 (KH01) et a un chiffre d'affaires entre 21.5 et 40.5 euros, nous aurons alors un taux d'évolution égal au taux 2, soit 54%.

Les taux d'évolution donnés pour les deux exemples sont très élevés, mais ils concernent très peu de clients.

Chaque client se voit affecter un taux d'évolution. Le taux multiplié par le chiffre d'affaires nous permet d'estimer un chiffre d'affaires potentiel pour chacun des clients.

CORRECTION DES TAUX ABERRANTS

CLASSE	EFF.	TX1	TX2	TX3	TX4	TX5	TX6	TX7	TX8	TX9	TX10	TX11	TX12	TX13	TX14	TX15	TX16	TX17	TX18	
KH01	0	5 789	39,16	63,13	45,26	36,55	30,96	27,95	24,86	23,62	22,48	21,74	21,36	21,22	21,13	21,69	23,05	22,41	25,49	33,37
KH02	1	1 580	39,16	61,89	44,29	35,62	29,93	27,54	25,55	23,07	22,50	21,87	21,06	20,76	20,11	20,74	21,69	22,18	24,58	28,22
KH03	2	3 067	39,16	58,41	41,79	33,61	29,12	26,10	24,00	22,76	21,65	21,02	20,42	19,98	19,80	20,06	21,60	23,99	26,03	30,51
KH04	3	2 345	39,16	59,80	42,32	35,48	30,08	27,94	26,19	24,98	24,36	23,48	23,11	23,05	22,70	23,48	24,13	25,23	27,30	35,39
KH07	10	688	39,16	56,18	38,95	31,22	25,67	22,87	20,23	17,99	16,71	15,86	15,05	14,38	13,45	13,28	13,06	12,91	12,36	12,62
KH08	11	378	39,16	58,47	40,16	32,00	27,76	24,72	21,80	19,91	19,14	17,94	16,89	16,39	15,16	15,46	15,16	20,49	23,96	25,23
KH10	13	472	39,16	54,17	39,50	30,19	25,21	21,90	18,98	17,34	15,53	14,38	12,46	12,09	10,96	10,36	9,96	9,18	8,84	8,49
KH11	14	627	39,16	62,07	40,46	32,05	28,76	23,36	20,69	19,26	16,38	15,72	14,34	13,70	12,94	12,17	11,58	10,96	13,96	33,63
KH12	15	1 388	39,16	60,81	42,60	33,25	28,00	24,11	21,77	19,49	18,02	17,02	15,47	14,66	13,74	12,72	12,89	12,82	13,61	14,42
KH13	20	4 594	39,16	68,56	47,95	36,82	28,83	25,00	21,04	18,01	15,55	14,03	12,58	11,58	12,66	13,69	13,84	13,93	15,36	18,86
KH25	40	1 508	39,16	64,82	44,56	34,56	29,06	23,70	20,59	18,95	16,75	15,12	14,15	12,73	11,79	11,19	10,69	9,95	9,52	13,91
KH26	41	1 340	39,16	4,35	4,21	4,36	4,37	4,47	4,49	4,37	4,57	4,38	4,55	4,58	4,55	4,62	4,59	4,70	4,84	5,28
KH30	45	2 217	39,16	46,45	30,93	23,13	19,06	16,44	13,96	12,92	12,38	10,34	9,90	8,74	8,59	8,08	7,63	7,41	6,93	6,45
KH31	50	3 839	39,16	89,13	81,91	130,78	139,82	4,80	4,91	4,77	4,82	4,92	5,06	5,23	5,43	5,44	5,69	5,76	6,14	6,39
KH32	51	277	39,16	83,58	73,58	80,03	189,69	16,44	5,62	5,64	5,62	5,58								

La table ci-dessus montre le remplacement des taux aberrants par la moyenne (>100%).

2.64.5 Principaux résultats

	CA an (CA filtré 12 mois)	Potentiel de CA	Taux d'augmentation
Total	7 467 726 175 €	953 284 621 €	12,77%
Moyenne	1 390 €	177 €	-

Tableau 105 : CA total, CA potentiel et taux d'augmentation

Le taux incrémental du potentiel client est de 12.77%.

Ce distributeur peut gagner 12.77% de chiffre d'affaires, sur ses clients.

Les clients affectés d'un chiffre d'affaires ≤ 0 ne sont pas retenus pour le score de potentiel puisque nous ne sommes concernés que par les clients actifs.

La situation présente vis-à-vis des clients à plus forte valeur semble être plus de la rétention client.

Les classes de potentiel :

On regroupe les taux de potentiel par classe de clones en trois catégories.

	COMPTES		CA		CA Pot		CA moyen	CA potentiel moyen
	Effectif	%	Montant	%	Montant	%	Montant	Montant
Bronze	2 248 626	41,9%	4 924 641 623 €	65,9%	366 263 710 €	38,4%	2 190 €	163 €
Argent	760 979	14,2%	1 121 962 901 €	15,0%	191 042 568 €	20,0%	1 474 €	251 €
Or	2 363 421	44,0%	1 421 121 650 €	19,0%	395 978 343 €	41,5%	601 €	168 €
Total	5 373 026	100%	7 467 726 175 €	100%	953 284 621 €	100%	1 390 €	177 €

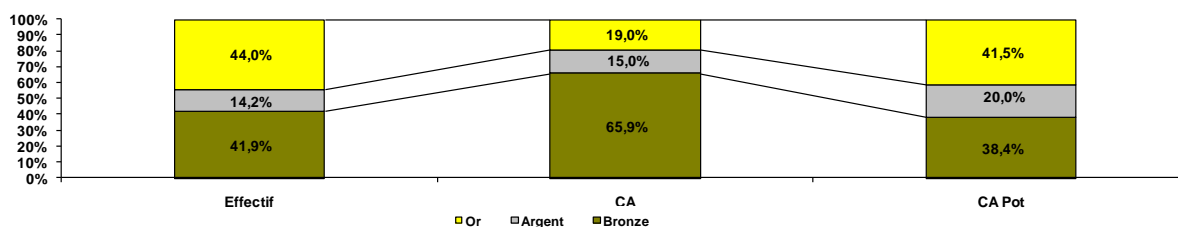


Figure 63 : Effectif, CA et CA potentiel par classe de clones regroupés en Bronze, Argent, Or

41.9% des clients du distributeur ont un potentiel "Bronze" générant un chiffre d'affaires annuel de 65.9% et comptant pour 38.4 % du chiffre d'affaires potentiel.

D'un autre coté 44% des clients sont des potentiels « Or », générant seulement 19% du chiffre d'affaires mais générant 41.5% du chiffre d'affaires potentiel.

Le regroupement en segment PMG :

	Bronze		Argent		Or		TOTAL	
	Effectif	%	Effectif	%	Effectif	%	Effectif	%
NV	4 894	0,2%	28 646	3,8%	211 738	9,0%	245 278	4,6%
P	706 262	31,4%	583 689	76,7%	1 797 002	76,4%	3 086 953	57,6%
M	759 003	33,8%	45 042	5,9%	210 731	9,0%	1 014 776	18,9%
G	778 193	34,6%	103 485	13,6%	133 952	5,7%	1 015 630	18,9%
TOTAL	2 248 352	100%	760 862	100%	2 353 423	100%	5 362 637	100,0%

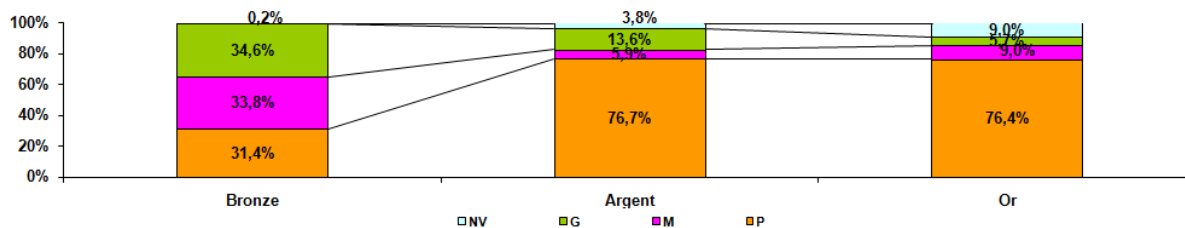


Figure 64 : Croisement des segments Bronze, Argent, Or par PMG en effectif

Les consommateurs P comptent pour une proportion élevée des potentiels « Or », basé sur le chiffre d'affaires annuel. Il y a des clients M parmi les potentiels « Bronze », et des clients G parmi les potentiels « Bronze » et « Argent ». La plupart d'entre eux ont une marge de progression intéressante.

En chiffre d'affaires annuel (en €):

	Bronze		Argent		Or		TOTAL	
	Montant	%	Montant	%	Montant	%	Montant	%
NV	2 612 344 €	0,1%	12 311 209 €	1,1%	25 441 780 €	1,8%	40 365 333 €	0,5%
P	507 518 386 €	10,3%	252 572 370 €	22,5%	425 060 654 €	29,9%	1 185 151 410 €	15,9%
M	1 334 543 548 €	27,1%	79 117 654 €	7,1%	358 381 176 €	25,2%	1 772 042 377 €	23,7%
G	3 079 967 346 €	62,5%	777 961 668 €	69,3%	612 238 041 €	43,1%	4 470 167 054 €	59,9%
TOTAL	4 924 641 623 €	100%	1 121 962 901 €	100%	1 421 121 650 €	100%	7 467 726 175 €	100,0%

Les clients P sont sur représentés parmi les potentiels "Or", avec 29,9% du chiffre d'affaires potentiel.

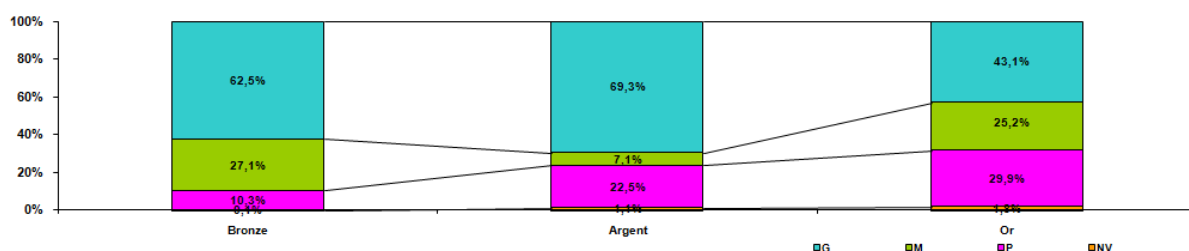


Figure 65 : Croisement des segments Bronze, Argent, Or par PMG en CA

	Bronze	Argent	Or	TOTAL
	Montant moyen	Montant moyen	Montant moyen	Montant moyen
NV	534 €	430 €	120 €	165 €
P	719 €	433 €	237 €	384 €
M	1 758 €	1 757 €	1 701 €	1 746 €
G	3 958 €	7 518 €	4 571 €	4 401 €
TOTAL	2 190 €	1 475 €	604 €	1 393 €

Tableau 106 : Croisement des segments Bronze, Argent, Or par PMG en CA en moyen

En chiffre d'affaires potentiel (en k€) :

	Bronze		Argent		Or		TOTAL	
	Montant	%	Montant	%	Montant	%	Montant	%
NV	194 573 €	0,1%	2 201 431 €	1,2%	10 402 290 €	2,6%	12 798 293 €	1,3%
P	53 232 096 €	14,5%	42 569 788 €	22,3%	118 298 788 €	29,9%	214 100 672 €	22,5%
M	95 035 722 €	25,9%	13 600 554 €	7,1%	93 761 586 €	23,7%	202 397 862 €	21,2%
G	217 801 319 €	59,5%	132 670 795 €	69,4%	173 515 680 €	43,8%	523 987 794 €	55,0%
TOTAL	366 263 710 k€	100%	191 042 568 k€	100%	395 978 343 k€	100%	953 284 621 €	100,0%

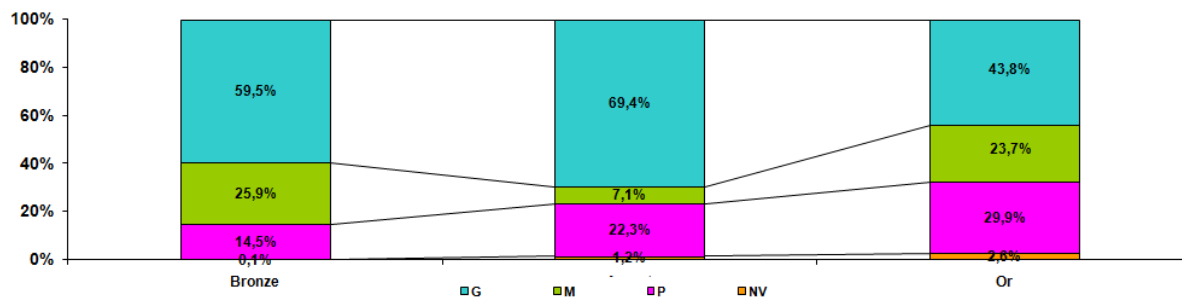


Figure 66 : Croisement des segments Bronze, Argent, Or par PMG en CA potentiel

	Bronze	Argent	Or	TOTAL
	Montant moyen	Montant moyen	Montant moyen	Montant moyen
NV	40 €	77 €	49 €	52 €
P	75 €	73 €	66 €	69 €
M	125 €	302 €	445 €	199 €
G	280 €	1 282 €	1 295 €	516 €
TOTAL	163 €	251 €	168 €	178 €

Tableau 107 : Croisement des segments Bronze, Argent, Or par PMG en CA potentiel moyen

Les clients G ont le plus important potentiel en valeur absolue, bien qu'ils ne soient pas les plus forts taux d'évolution. Ils sont dans cette situation car ils ont un chiffre d'affaires nettement plus significatif que les segments P et M.

La distribution par la RFM du distributeur et par les catégories de potentiel.

En effectifs :

	Bronze		Argent		Or		TOTAL	
	Effectif	%	Effectif	%	Effectif	%	Effectif	%
Sans Statut	21 253	0,9%	30 590	4,0%	347101	14,7%	398 944	7,4%
INACTIF 3 MOIS	31 524	1,4%	29 084	3,8%	400059	17,0%	460 667	8,6%
NOUVEAU	4 896	0,2%	28 654	3,8%	211983	9,0%	245 533	4,6%
M-F--	245 481	10,9%	329 558	43,3%	738141	31,4%	1 313 180	24,5%
M-F-	728 010	32,4%	185 532	24,4%	388396	16,5%	1 301 938	24,3%
M-F+	174 538	7,8%	22 490	3,0%	51591	2,2%	248 619	4,6%
M+F-	516 991	23,0%	53 058	7,0%	140234	6,0%	710 283	13,2%
M+F+	525 659	23,4%	81 896	10,8%	75918	3,2%	683 473	12,7%
TOTAL	2 248 352	100%	760 862	100%	2 353 423	100%	5 362 637	100%

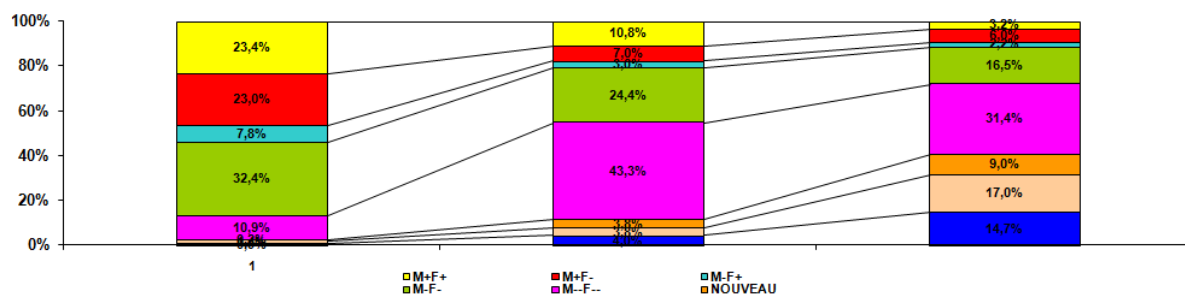


Figure 67 : Croisement des segments Bronze, Argent, Or par RFM en effectif

En CA annuel (en €):

	Bronze		Argent		Or		TOTAL	
	CA	%	CA	%	CA	%	CA	%
Sans Statut	28 482 104 €	0,6%	15 908 974 €	1,4%	69 877 281 €	4,9%	114 268 358 €	1,5%
INACTIF 3 MOIS	38 084 339 €	0,8%	10 476 055 €	0,9%	42 374 499 €	3,0%	90 934 893 €	1,2%
NOUVEAU	2 613 184 €	0,1%	12 313 384 €	1,1%	25 462 408 €	1,8%	40 388 976 €	0,5%
M--F--	206 804 797 €	4,2%	128 061 467 €	11,4%	210 286 948 €	14,8%	545 153 212 €	7,3%
M-F-	964 638 682 €	19,6%	134 255 494 €	12,0%	316 010 822 €	22,2%	1 414 904 998 €	18,9%
M-F+	288 843 514 €	5,9%	25 721 163 €	2,3%	70 733 841 €	5,0%	385 298 518 €	5,2%
M+F-	1 407 713 946 €	28,6%	189 736 462 €	16,9%	393 166 652 €	27,7%	1 990 617 060 €	26,7%
M+F+	1 987 461 057 €	40,4%	605 489 903 €	54,0%	293 209 200 €	20,6%	2 886 160 159 €	38,6%
TOTAL	4 924 641 623 k€	100%	1 121 962 901 k€	100%	1 421 121 650 k€	100%	7 467 726 175 €	100,0%

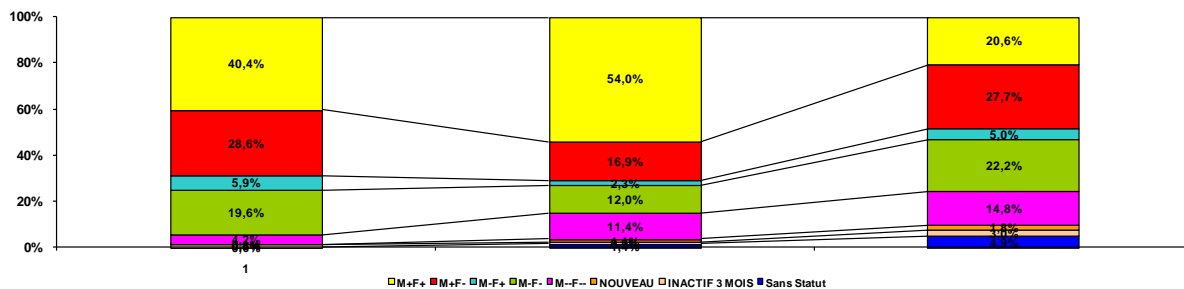


Figure 68 : Croisement des segments Bronze, Argent, Or par RFM en CA

	Bronze	Argent	Or	TOTAL
	Montant moyen	Montant moyen	Montant moyen	Montant moyen
Sans Statut	1 340 €	520 €	201 €	286 €
INACTIF 3 MOIS	1 208 €	360 €	106 €	197 €
NOUVEAU	534 €	430 €	120 €	164 €
M--F--	842 €	389 €	285 €	415 €
M-F-	1 325 €	724 €	814 €	1 087 €
M-F+	1 655 €	1 144 €	1 371 €	1 550 €
M+F-	2 723 €	3 576 €	2 804 €	2 803 €
M+F+	3 781 €	7 393 €	3 862 €	4 223 €
TOTAL	2 190 €	1 475 €	604 €	1 393 €

Tableau 108 : Croisement des segments Bronze, Argent, Or par RFM en CA moyen

Logiquement les potentiels les plus importants sont les plus présents en valeur absolue dans les segments RFM +.

Les catégories de potentiel :

Les taux de potentiel par classe de clones sont regroupés en quatre catégories :

- P0: Pas de CA potentiel
- P1: Potentiel supérieur à 20 %
- P2: Potentiel supérieur ou égale à 15 et inférieur à 20 %
- P3: Potentiel inférieur à 15%

	COMPTES		CA		CA Pot		CA moyen	CA potentiel moyen
	Effectif	%	Montant	%	Montant	%	Montant	Montant
P0	268 675	5,0%	1 035 180 496 €	13,9%	0 €	0,0%	3 853 €	0 €
P1	2 242 226	41,7%	999 789 812 €	13,4%	268 786 859 €	36,6%	446 €	120 €
P2	699 709	13,0%	672 185 375 €	9,0%	114 265 871 €	15,6%	961 €	163 €
P3	2 162 416	40,2%	4 760 570 492 €	63,7%	350 848 889 €	47,8%	2 202 €	162 €
Total	5 373 026	100%	7 467 726 175 €	100%	733 901 619 €	100%	1 390 €	137 €

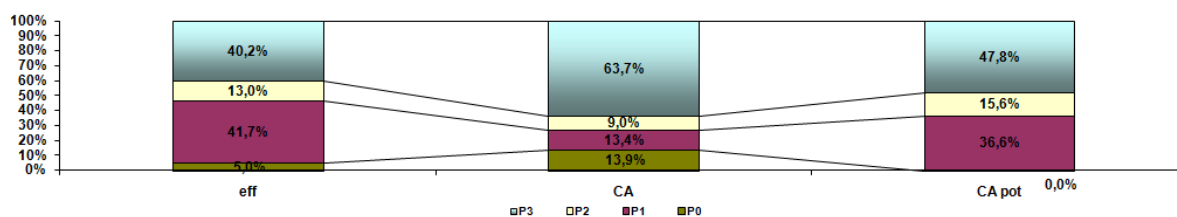


Figure 69 : Effectifs et CA par classes de potentiel regroupées

40% des clients créent 63% du chiffre d'affaires et 47.8% du CA potentiel. En moyenne, ils réalisent un chiffre d'affaires de 2 202€ pour un potentiel moyen de 162€. Ces clients qui ont déjà une contribution substantielle, sont les plus à même (pour le moindre effort perçu) d'atteindre leur potentiel.

Le regroupement des segments PMG :

	P0		P1		P2		P3		TOTAL	
	Effectif	%	Effectif	%	Effectif	%	Effectif	%	Effectif	%
NV	117	0,0%	211 720	9,5%	28 646	4,1%	4 795	0,2%	245 278	4,6%
P	78 682	29,3%	1 757 427	78,7%	573 720	82,0%	677 124	31,3%	3 086 953	57,6%
M	61 962	23,1%	198 393	8,9%	45 032	6,4%	709 389	32,8%	1 014 776	18,9%
G	127 914	47,6%	64 688	2,9%	52 194	7,5%	770 834	35,7%	1 015 630	18,9%
TOTAL	268 675	100%	2 232 228	100%	699 592	100%	2 162 142	100,0%	5 362 637	100,0%

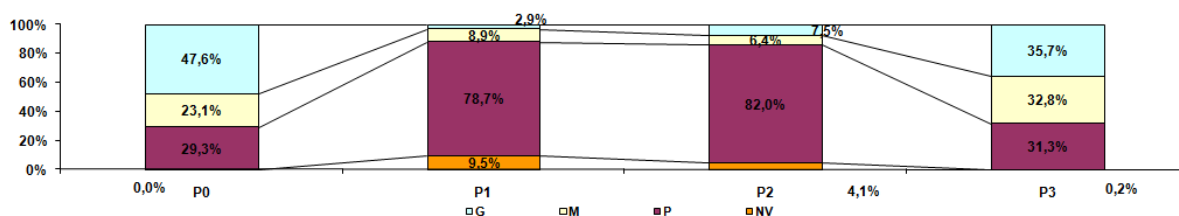


Figure 70 : PMG en effectif par classes de potentiel regroupées

Les clients P représentent une part élevée de P1 et de P2 basée sur le CA annuel généré par les potentiels P1. Nous trouvons des clients M principalement par les potentiels P3, alors que les clients G pour leur part se retrouvent dans P0 mais aussi dans P3.

En chiffre d'affaires annuel (en €) :

	P0		P1		P2		P3		TOTAL	
	Montant	%	Montant	%	Montant	%	Montant	%	Montant	%
NV	489 934 €	0,0%	25 376 536 €	2,5%	12 311 209 €	1,8%	2 187 654 €	0,0%	40 365 333 €	0,5%
P	78 985 251 €	7,6%	389 503 691 €	39,0%	241 887 660 €	36,0%	474 774 808 €	10,0%	1 185 151 410 €	15,9%
M	132 044 286 €	12,8%	331 934 951 €	33,2%	79 092 619 €	11,8%	1 228 970 521 €	25,8%	1 772 042 377 €	23,7%
G	823 661 025 €	79,6%	252 974 633 €	25,3%	338 893 887 €	50,4%	3 054 637 509 €	64,2%	4 470 167 054 €	59,9%
TOTAL	1 035 180 496 €	100%	999 789 812 €	100%	672 185 375 €	100%	4 760 570 492 €	100,0%	7 467 726 175 €	100,0%

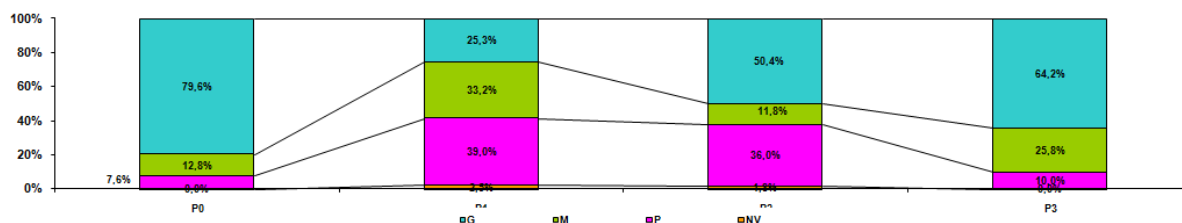


Figure 71 : PMG en CA annuel par classes de potentiel regroupées

	P0	P1	P2	P3	TOTAL
	Montant moyen	Montant moyen	Montant moyen	Montant moyen	Montant moyen
NV	4 187 €	120 €	430 €	456 €	165 €
P	1 004 €	222 €	422 €	701 €	384 €
M	2 131 €	1 673 €	1 756 €	1 732 €	1 746 €
G	6 439 €	3 911 €	6 493 €	3 963 €	4 401 €
TOTAL	3 853 €	448 €	961 €	2 202 €	1 393 €

Tableau 109 : PMG en montant moyen de CA par classes de potentiel regroupées

En CA potentiel (en €) :

	P1		P2		P3		TOTAL	
	Montant	%	Montant	%	Montant	%	Montant	%
NV	10 381 965 €	3,9%	2 201 431 €	1,9%	165 329 €	0,0%	12 748 726 €	1,7%
P	104 608 676 €	38,9%	40 846 282 €	35,7%	50 583 326 €	14,4%	196 038 284 €	26,7%
M	85 568 776 €	31,8%	13 596 614 €	11,9%	85 102 427 €	24,3%	184 267 817 €	25,1%
G	68 227 442 €	25,4%	57 621 544 €	50,4%	214 997 807 €	61,3%	340 846 793 €	46,4%
TOTAL	268 786 859 €	100%	114 265 871 €	100%	350 848 889 €	100%	733 901 619 €	100,0%

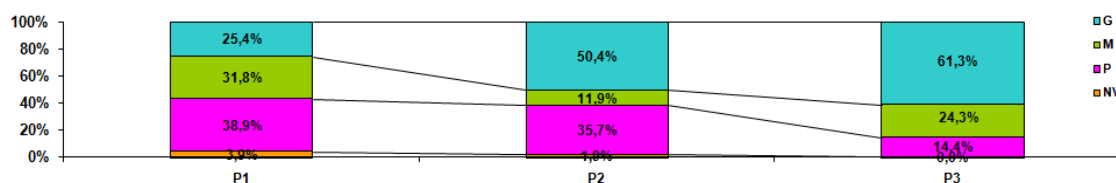


Figure 72 : PMG en CA potentiel par classes de potentiel regroupées

Les clients P sont sur représentés dans la catégorie "P1" avec un chiffre d'affaires potentiel de 38,9% pour ce segment. En terme absolu, ce sont vraiment les clients G qui ont le potentiel le plus élevé. C'est avec le «gros» client que nous pouvons accroître le chiffre d'affaire et que celui-ci a le plus de chance d'être atteint, que dans n'importe quel autre segment. Le budget marketing et l'intensité des offres marketing peuvent donc être alloués sur la base d'un chiffre d'affaire moyen par potentiel. Les deux concepts sont complémentaires dans la définition des mécanismes de fidélisation/rétention.

2.64.6 Résumé des résultats

Le réseau de Kohonen nous permet de regrouper les clients en 50 classes de clones. La matrice 4X10 n'a pas de groupe vide, donc nous la retenons.

Les clients dans un même groupe se ressemblent entre eux sur des caractéristiques socio démographiques et de consommation.

En utilisant la méthode des déciles, nous affectons un taux d'évolution de chiffre d'affaires à chaque client de la base.

Nous créons les scores de potentiels suivants :

- « Or » : taux d'évolution supérieur à 20%

- « Argent » : taux d'évolution supérieur ou égale à 15% inférieur à 20%
- « Bronze » : taux d'évolution inférieur à 15%

Nous calculons le chiffre d'affaires potentiel à partir de ce taux et du chiffre d'affaires réel.

Notre base est composée de 5 373 026 clients générant un chiffre d'affaires annuel de 7.46 milliards d'euros, et représentant un chiffre d'affaires potentiel de 953.2 millions d'euros.

Le distributeur peut gagner presque 12.77% de chiffre d'affaires sur ses clients.

En termes de taux, la base est composée de 41.9% de clients « Bronze », de 14.1% clients « Argent » et de 44% de clients « OR ». Dans la réalité, il doit être admis que les meilleurs clients sont ceux avec les plus fortes valeurs absolues de potentiel.

76.6% sont des consommateurs « Or » et P et ils génèrent 29.9% du chiffre d'affaire annuel et 38,4% du chiffre d'affaire potentiel.

65% sont des clients « Or », inactifs 3 mois, et RFM -- et RFM- et ils génèrent 40% du chiffre d'affaires annuel et 38,4% du chiffre d'affaires potentiel.

Logiquement les clients avec les taux d'évolution les plus élevés, se retrouvent parmi les clients avec les plus petits chiffres d'affaires en valeur.

A la fin de l'échelle, les clients avec le chiffre d'affaires le plus élevé, ont le potentiel le plus fort en valeur absolue.

	P1	P2	P3	TOTAL
	Montant moyen	Montant moyen	Montant moyen	Montant moyen
NV	49 €	77 €	34 €	52 €
P	60 €	71 €	75 €	64 €
M	431 €	302 €	120 €	182 €
G	1 055 €	1 104 €	279 €	336 €
TOTAL	120 €	163 €	162 €	137 €

Tableau 110 : PMG en CA potentiel moyen par classes de potentiel regroupées

2.65 Les procédures de validation de nos modèles

La validité interne

Nous avons mené plusieurs tests sur notre modèle :

- La division de notre population en sous populations pour vérifier la cohérence d'allocation des classes de clones
- Le benchmark entre différentes techniques de classification
- La réaffectation des groupes par des modèles supervisés (C5.0, Réseaux Bayésiens)
- La connectivité des super classes

Les méthodes de validation devront bien s'être complétées

Validé externe

Le taux de nourriture par client selon TNS est de 24%. Sur, l'ensemble de la consommation, le potentiel atteignable sur le taux de nourriture la fait croître de 28%. Un différentiel de 2% de taux de nourriture entre le réel et le prédit semble tout à fait admissible.

Nous aurions pu dédupliquer notre base avec le panel Home Scan de Nielsen pour vérifier si les ventes s'accroissent réellement, mais cela n'était pas possible dans ce contexte précis.

2.66 Conclusions

2.66.1 Discussion des résultats de l'étude

Les résultats de notre recherche doivent être replacés dans le contexte de la détermination du potentiel client d'une entreprise donnée : déterminer le potentiel client va impacter fortement le choix d'investissement en marketing direct ou en promotion d'une société. La plupart des grands programmes de fidélisation sont basés sur cette notion. Nous allons regarder comment notre approche comparée à d'autres méthodes, nous permet de déterminer des résultats convergents pour nos questions de recherches :

- La technique de classification de Kohonen (SOM) est utilisée pour identifier des clients qui sont similaires et pour définir un potentiel réaliste
- Notre méthode de céurage en déciles des classes de clones donne un potentiel par client, plus fins que d'autres méthodes : un potentiel atteignable
- Nous pouvons estimer la stabilité des groupes de différentes manières et elles montrent une stabilité interne de la méthode
- Nous avons développé une approche pragmatique qui est une méthode de détermination des potentiels : « la méthode des clones »

2.66.2 Les limites et la contribution de notre étude

Nous n'avons utilisé que certaines techniques de classifications particulières dans le but de valider notre méthode. Nous avons montré les éventuelles limites statistiques de notre approche en termes de complexité ou de fiabilité des modèles utilisés.

Pour des raisons de faisabilité, nous n'avons travaillé que sur un seul secteur industriel : la grande distribution alimentaire en France, et nous n'avons pas utilisé d'autres données venant d'autres secteurs.

Nous n'avons pas eu accès pour le moment à des données de distributeurs étrangers.

Le calcul du CA potentiel dans les groupes est très empirique et devrait être plus scientifiquement justifié.

2.66.3 Pistes de recherche

Il y a plusieurs façons d'améliorer notre recherche :

- Affiner le choix de nos variables,
- Déterminer une méthode plus empirique que la méthode des déciles/médianes pour estimer le potentiel par groupe,
- Faire plus de rotations du modèle sur d'autres secteurs industriels ; nous avons déjà fait cela, et la méthode s'applique plutôt bien, mais il est important que d'autres le testent,
- Valider les résultats dans le temps, en observant la réalité des valeurs de potentiels prédites sur les ventes réelles.

Nous espérons que par son impact stratégique sur les résultats des entreprises, et le fait que son calcul ne soit basé que sur des données clients internes déjà en possession de l'entreprise, cette « méthode de clones » trouvera un usage important.

Conclusion générale

Nous avons démontré que le taux de nourriture est borné, que l'on peut déterminer le potentiel d'un client en le comparant à des clients avec des comportements de consommations similaires et en estimant leur différentiel de consommation, enfin que la valeur d'un client à vie est bien la somme de ces valeurs à l'instant T.

Un élément essentiel est de déterminer ce que la marque peut prendre à ses concurrents sur la durée de vie du consommateur : cette notion de taux de captation sur le taux de nourriture est omniprésente dans nos trois travaux, c'est la clef du potentiel client.

Les panels de home scanning, les BDD clients avec le détail des consommations à l'acte d'achat sont les données « idéales » pour faire cela. A l'avènement du big data ces données vont être amenées à se multiplier.

Les modèles statistiques agrégés ou désagrégés permettent de modéliser ces phénomènes de manière plus ou moins convergente.

On peut estimer que modèles et données vont permettre de faire mieux dans les années qui viennent, là nous avons travaillé que sur quelques marchés avec quelques modèles mais le chercheur qui voudrait poursuivre cette analyse pourra dans un proche avenir travailler sur tous les marchés avec des quantités de consommateurs et d'actes d'achats encore plus importantes sur des durées très longues.

Nous espérons que le chercheur le fera....

Bibliographie :

Aguilera, P. A., Frenich, A. G., Torres, J. A., Castro, H., Vidal, J. L. M., and Canton, M., Application of the cohune neural network in coastal water management: Methodological development for the assessment and prediction of water quality. *Water Research*, 35(17):4053–4062, 2001.

Ailawadi Kusum L., Lehmann Donald R. and Neslin Scott A., “Market response to a major policy change in the marketing mix: learning from procter and gamble’s value pricing strategy”, *The journal of Marketing*, p.44-61, Janvier 2001.

Allard C, *Le Management de la Valeur Client*, Dunod, 2002.

Anderson, B. Kohonen neural networks and language. *Brain and Language*, 70(1):86–94, 1999.

Badoc M, Lavayssiere B et Copin E, *E Marketing de la banque et de l’assurance*, Ed d’Organisation, p 100-101, 1999.

Baran, Stanley J. *Theories of Mass Communication*, Wadsworth, 2012

Bell D., Chiang J. et Padmanaban V., *The Decomposition of Promotional Response : an Empirical generalisation*, *Marketing Science*, 18, 4. p 504-26, 1999.

Bénavent C. et Crié D., *Mesurer l’efficacité des cartes de fidélité*, *DM*, Septembre-décembre, p 83-90, 1998.

Beran, R., Discussion of Wu, C.F.J.: Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). *Ann. Statist.*, 14:1295-1298, 1986.

Berend Wierenga and Gerrit Harm van Bruggen, *Marketing Management, Springer Support Systems: Principles, Tools, and Implementation*, Springer, 2000.

Berger, Paul D. and Nada I. Nasr, "Customer lifetime value: Marketing models and applications," *Journal of Interactive Marketing*, 12 (1), p.17 – 30, 1998.

Berges F, Hassan D et Monier-Dilhan S, *Les déterminants de la demande en marques de distributeurs : étude à partir de deux produits de consommation courante*, *Toulouse School of Economics (GREMAQ-INRA)*, 2007.

Berry L.L., *Relationship marketing of Services. Growing interest, Emerging Perspectives*, *Journal of the Academy of Marketing Science*, vol. 23, n° 4, p. 236-245, 1995.

Bertrand Clarke et Dongchu Sun, *Reference priors under the Chi-Squared distance: The Indian Journal of Statistics*, Volume 59, Series A, Pt. 2, 215-231, 1997.

Bhattacharya C.B., “Is your brand’s loyalty too much, too little, or just right ? Explaining deviations in loyalty from the Dirichlet norm”, *International Journal of Marketing*, p.421-435, 1997.

Bhattacharya C.B., Fader Peter S., Lodish Leonard M., Desarbo Wayne S., “The relationship between the marketing mix and share of category requirements”, 1996.

Blattberg et al, *Customer Equity : Building and Managing Relationship as Valuable Assets*, Harvard Business School Press, 2001.

- Blattberg, R., Managing the firm using lifetime-customer value, *Chain Store Age*, January, p 46-49, 1998.
- Blattberg, R., Research Opportunities in Direct Marketing. *Journal of Direct Marketing*, 1986.
- Bolton R., Kannan P. et Bramlett M., Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value, *Journal of the Academy of Marketing Science*, 28, 1, Winter. p 95-108, 2000.
- Boos, D.D., Introduction to the bootstrap world. *Statist. Science*, 18:168-174, 2003.
- Borko, H. and Bernick, M., 'Automatic document classification', *Journal of the ACM*, 10, 151-162, 1963.
- Bourbonnais R et Usunier JC, *Prévision des Ventes : Théorie et pratique*, Economica, 2007.
- Bouroche J.M. et Saporta G., *L'analyse des données*, PUF, 1980
- Bremer and Joyce, *Human Judgment, The SJT View*, North-Holand, 1988.
- Brown G., Brand Loyalty – Fact or fiction, *Advertising Age*, 24, p75-76, Janvier 1953.
- Bruce Cooil B, Keiningham T, Aksoy L, Hsu M., A Longitudinal Analysis of Customer Satisfaction and Wallet share: Investigating the Moderating Effect of Customer Characteristics. *Journal of Marketing* 71:1, 67-83, 2007.
- Brusset O., « Segmentation Cibler, scorer, analyser, une seule limite, les rendements » *Marketing Direct* N°92, p.2, 2005.
- Casin P., *Analyse des données et des panels de données*, De Boeck, 1999
- Carguillo-Charles E., *Gestion de la rentabilité des clients bancaires*, DM, p 25-37, Janvier-Avril 1999,
- Chen-Kuo Tsao E, Bezdek James C and Pal Nikhil R, "Fuzzy Kohonen clustering networks" Published by Elsevier Science B.V., 1994.
- Chestnut R.W and Jacoby J., *Brand Loyalty : Measurement and Management*, J. Wiley and Sons, NY, 1978.
- Chintagunta Pradeep K, Jain Dipak C. and Vilcassim Naufel J., "Investigating heterogeneity in brand preferences in logit models for panel data", *Journal of Marketing Research*, 417-428, Novembre 1991.
- Chintagunta Pradeep K., "Inertia and Variety Seeking in a Model of Brand-Purchase Timing", *Marketing Science*, p.253-270, 1998.
- Ciampi, A. and Lechevallier, Y., Clustering large, multi-level data sets: an approach based on Kohonen self-organizing maps. In *Principles of Data Mining and Knowledge Discovery. 4th European Conference, PKDD. Proceedings (Lecture Notes in Artificial Intelligence Vol.1910)*. Springer-Verlag, Berlin, Germany, pages 353–8, 2000.
- Collectif, "Recherche sur la Distribution moderne" p.64, éd: l'Univers du Livre, 2007.
- Cook W.D. et Kress M., Ordinal ranking with intentensity of of Preference, *Management Science* 31, p 26-32, Janvier 1985.
- Cooper LG et Nakanishi, *Market-share analysis: Evaluating competitive marketing effectiveness*, Kluwer Academic Press, 1989

- Cooper, L. and Nakanishi, M., *Market-Share analysis*, Kluwer, Norwell, Mass, 1988.
- Crié D. , *Rétention de clientèle et fidélité des clients*, DM, p 25-30, Janvier-Avril 1996.
- Cunningham R., *Brand Loyalty-what, where, how much*, Havard Business Review, 3. p 116-128, 1956.
- Curry J. et Stora P., *La Client Capital de l'Entreprise*, Ed d'Organisation, p 42, 1993.
- Dahbur, K. and Muscarello, T., *Hybrid Kohonen neural network in data mining*. In *Proceedings of the IASTED International Conference. Artificial Intelligence and Applications*. ACTA Press, Anaheim, CA, USA, pages 30–3, 2001.
- Dall'Olmo Riley F., Ehrenber A.S.C., Castleberry S.B., Barwise T.P. and Barnard N.R., "The variability of atitudinal repeat-rates", *International Journal of Marketing*, p.437-450, 1997.
- David Huff, University of Texas Austin, "A Retrospective View of the Huff Model and its Application to Spatial Interaction Analysis" *University of Redlands/ESRI Colloquium Series*, 18 Juin 2003.
- Dekimpe Marnik G., Steenkamp, Mellens and Abeele., "Decline and variability in brand loyalty", *International Journal of Marketing*, p.405-420, 1997.
- Desmet P., *Marketing Direct*, Dunod, p 20-21, 2001.
- Desmet P., *Promotion des ventes*, Dunod, p 238 et p 260, 2002.
- Distefano J., *The decisionning frontier : optimizing the value of your customer portfolio*, DM Review, Février 2000.
- Dorofeyuk, A.A., 'Automatic Classification Algorithms (Review)', *Automation and Remote Control*, 32, 1928-1958, 1971.
- Dowling G.R et Uncles M., *Do Customer Loyalty Programs Really Work ?*, *Sloan Management Review*, Summer. p 71-82, 1997.
- Drèze X. Hoch S. et Purk M., *EDLP, Hi-Lo and Margin Arithmetic*, *Journal of Marketing*, 58, p 16-27, Octobre 1994.
- Dwyer F, *Customer life time valuation to support decision marketing*, vol 3, n°4, p 8-15, 1989.
- Dwyer, R.F., "Customer lifetime valuation to support marketing decision making", *Journal of Direct Marketing*, Vol. 11 No.4, p.6-13, 1997.
- Efron B., *Non parametric estimates of standard error: the jackknife, the bootstrap and other methods*. *Biometrika* 68. pp 589—599, 1981.
- Ehrenberg A.S.C, "Empirical Generalisations, Theory, and Method", *Marketing Science*, 1995.
- Ehrenberg A.S.C. et Goodhardt GJ., *Understanding Buyer Behaviour*, New York : J. Walter Thompson and the Market Research Corporation of America, 1977.
- Ehrenberg A.S.C., Goodhardt J.G et Hammond K., *The After-Effect of Price-Related Consumer Promotions*, *Journal of Advertising Research*, p11-21, Juillet-Aout 1994.

- Ehrenberg A.S.C., Uncles M. et Hammond K., *Patterns of Buyer Behavior : Regularities, Models and Extensions*, *Marketing Science*, 14. p 71-78, 1995.
- Fader Peter S. and Lodish Leonard M., “A cross-category analysis of category structure and promotional activity for grocery products”, *Journal of Marketing Research*, Octobre 1990.
- Fader Peter S. and Schmittlein David C. “Excess Behavioral Loyalty for High-Share Brands: Deviations from the Dirichlet Model for Repeat Purchasing”, *Journal of Marketing research*, p.478-493, Novembre 1993
- Fang, K.; He, S. The problem of selecting a given number of representative points in a normal population and a generalized mill’s ratio. Technical report, Department of Statistics; Stanford University: 1982.
- MacQueen J. Some methods for classification and analysis of multivariate observations. *Proceedings 5th Berkeley Symposium on Mathematics, Statistics and Probability*, 3:281–297, 1967.
- Gabin Philippe, *Analyse des données et des panels de données*, De Boeck, 1999.
- Gehrlein W. V. “General mathematical programming formulations for the statistical classification problem” *Operations research letters* ISSN 0167-6377 CODEN ORLED5, 1986.
- Ghosh, A., Neslin, S. and Shoemaker, R., A comparison of market share models and estimation procedures, *Journal of Marketing Research*, Vol. 21 No. 2, p 202-210, 1984.
- Gordon W. ,Retailer Brands – The Value Equation for Success in the 90s, *Journal of the Market Research Society*, 36, 3., p 165-181, 1994.
- Gordon W., Taking Brand Repertoires Seriously, *Journal of Brand Management*, 1994, 2, 1., p 25-30.
- Guadagni P.M et Little J.C, A Logit Model of Brand Choice Calibrated on Scanner Data, *Marketing, Marketing Sciences*, vol. 27 no. 1, p 29-48, Janvier /Février 2008.
- Gupta Sunil and Zeithaml Valarie, “Customer Metrics and Their Impact on Financial Performance”, *Marketing Science*, p.718-739, Novembre-Décembre 2006.
- Hallberg G., *All consumers are not created equal*, Wiley, 1995.
- Harris, M.J. and N. Blisard, “Characteristics of the Nielsen Homescan Data.” Working paper. Washington, DC: U.S. Department of Agriculture, Economic Research Service, . 1995.
- Hartigan JA, Wong MA. A k-means clustering algorithm. *Applied Statistics*, 28:100–108, 1979.
- Hauser John R. and Wernerfelt Birger., “The Competitive Implications of Relevant-Set/Response Analysis”, *Journal of Marketing Research*, p.391-405, November 1989.
- Haykin S.O., *Neural Networks and Learning Machines*, 3rd Edition, Pearson, 2008.
- Herniter D.J. et Howard A.R., *Stochastic Marketing Models*. Progress in Operations Research, vol II, New York, John Wiley and Sons, 1964.
- Howard et Seth., *The theory of buyer behavior* Prentice Hall, 1969.
- Howard J., *Buyer behavior in marketing strategy*, Prentice Hall, p 128-129, 1989.
- Hsue Cheng and You-Shyang Chen, Classifying the segmentation of customer value via RFM model and RS theory Expert Systems with Applications, In Press, Corrected Proof, Available online, 16 Avril 2008.

http://en.wikipedia.org/wiki/Lifetime_value

Jackson, D., Determining a customer's lifetime value, *Direct Marketing*, p 60-62, Mars 1989.

Jain Dipak C. and Vilcassim Naufel J., "Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity", *Journal of Marketing Research*, p.29-4, Février 1991.

Jajuga K."Classification, Clustering and Data Analysis : Recent Advances and Applications", Lavoisier, 2002.

Jambu M., Estimation et prédiction de la fidélité, de la durée de vie et de la valeur économique des clients par des techniques de data mining, *Revue Française du Marketing*, n° 170, p 67-80, 1998.

James R . et Hughes A., How banks use profitability to segment their customers, *Journal of database marketing*, Vol 6 n°2, p 145-155, 1998.

Jensen F. V, "Introduction to Bayesian Networks, 1st edition", Springer-Verlag New York, Inc, 1996.

John A. McCartya, and Manoj Hastak "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression" *Journal of Business Research*, Volume 60, Issue 6, Pages 656-662, Juin 2007.

Jolivot et Chandon, Contribution à l'étude de la taille de l'ensemble de considération: une application aux enseignes de restauration, *Congrès sur les Tendances du marketing en Europe*, 2, Paris, ESCP-EAP, 2002

Juha Vesanto, "The SOM in data mining: analysis of world pulp and paper technology", 1997.

Julien Barnier « Tout ce que vous n'avez jamais voulu savoir sur le Chi2 sans jamais avoir eu envie de le demander » Groupe de Recherche sur la Socialisation CNRS – UMR 5040, 15 avril 2008.

Jung Sang-Uk, Gruca Thomas S. and Rego Lopo L., "Excess loyalty in CPG Markets: A Comprehensive Examination", *Journal of empirical generalisations in marketing science*, 1995.

Kahn, B. E. et R. Ratner, Variety for the Sake of Variety? Diversification Motives in Consumer Choice, *Inside Consumption: Frontiers of Research on Consumer Motives, Goals, and Desires*, ed. S. Ratneshwar and David Glen Mick, London: Routledge, 2005.

Kaski, S., "Data exploration using self-organizing maps. *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 82*, Espoo 1997.

Kearney T.J (), *Frequent Flyer Programs : A Failure in competitive strategy, with lessons of management*, *The Journal of Consumer Marketing*, Winter, 7, 1., p 31-40, 1990.

Kim, Min et Shin, Calculating lifetime values of business customers for a telecommunication company, *Journal of database marketing*, Vol 7 n° 3, p 254-264, 1999.

Kohonen, T., *Self-Organization and Associative Memory* , New York : Springer-Verlag, 1988

Kumar Nanda, Rajiv Surendra and Jeuland Abel,"Effectiveness of Trade Promotions: Analyzing the Determinants of Retail Pass Through", *Marketing Science*, p.382-404, automne 2001.

Laisney C et Soyeux S, La consommation alimentaire à l'épreuve de la crise, *Analyse*, INSEE, N°10, Juillet 2009.

- Larose D et Vallaud T. Des données à la connaissance. Une introduction au data mining, Vuibert, 2005
- Larose D et Vallaud T. Exploration des données à la connaissance. Méthodes et modèle du data mining, Vuibert, 2012
- Lefébure R. et Venturi P. , Gestion de la relation client, Eyrolles, p 119-126, Janvier-avril 1996.
- Leforestier B. Le marketing relationnel des distributeurs, DM, p 47-52, Janvier-avril 1996.
- Lehu JM, Encyclopédie du Marketing , Editions d'Organisation, 2012.
- Lerman, I.C., Les Bases de la Classification Automatique, Gauthier-Villars, Paris, 1970.
- Liquet JC. et Crié D., Mesurer la durée de vie des clients, DM, p 75-84, Janvier-avril 1998,.
- Loddish L., F. Melac., If Brands Are Built Over Years, Why are they Managed Over Quarters, 2006
- Magson N., Database workshop : determining and mesuring customer value, Journal of database marketing, Vol 6 n° 3, p 24-33, 1998.
- Manuel B. & Xardel D. , La marketing direct en France, Dalloz, p 107, 1980.
- Mattias Otto “ChemometricsStatistics and Computer Application in Analytical Chemistry”, Wiley-VCH, 2007.
- Mc Dowall et all, Interrupted time series analysis, Sage, 1990.
- Meunier B, Dumas E, Picc I, Bechet D, Hebraud M, ... - Proteome Res J, Assessment of hierarchical clustering methodologies for proteomic data mining - les 4 versions » aseanbiotechnology.info, 1997.
- Meyen-Waarden L. La fidélisation client : stratégies, pratiques et efficacité des outils du marketing relationnel, Lavoisier, 2004
- Meyer-Waarden L. et Benavent C., Programmes de fidélisation : stratégies et Pratiques, Actes du 17ème congrès de l'Afm, Deauville, Mai, 2001.
- Morgan R.M et Hunt S.D, The commitment-Trust Theory of Relationship Marketing, Journal of Marketing, 58, 3., p 20-38, 1994.
- Mrazek J., Fully integrated customer value management, DM Review, Juin 2002.
- Naert, P. and Leeflang, M., Building implementable marketing models, Martinus Nijhoff, Leiden, 1978.
- Nako S.M., Frequent Flyer Programs and Business Travellers : an Empirical Investigation, Logistics and Transportation Review, 28, 4., p 395-410, 1997.
- Neslin S., Henderson C. et Quelch J., Consumer Promotions and Acceleration of Product Purchases, Marketing Science, 4,2., p 147-165, 1985.
- Nielsen, Inc., “Understanding the Homescan Advantage.” Presentation by Liz Crews and Ed Groves, Nielsen at RTI International, Research Triangle Park, NC, Mai 2006.
- O'Brien B. et Jones C., Do Rewards Really Create Loyalty, Harvard Business Review, p 75-82, Mai Juin 2005.

- Pena M. Vanegas A. Valencia “Digital Hardware Architectures of Kohonen's Self Organizing Feature Maps with Exponential Neighboring Function” IEEE International Conference on Reconfigurable Computing and FPGA's J., pp. 1-8, 2006
- Peppers D et Rogers M, The one to one report : customer valuation, DM Review, September 1998.
- Pfeifer P., On the use of customer lifetime value as a limit on acquisition spending, Journal of database marketing, Vol 7 n°1, p 81-86, 1999.
- Plastria F, Static competitive facility location: An overview of optimisation approaches European Journal of Operational Research, Volume 129, Issue 3, Pages 461-470, 16 mars 2001
- Pourret O., Naim P. et Marcot B., Bayesian Networks: A Practical Guide to Applications. Chichester, UK: Wiley. ISBN 978-0-470-06030-8, 2008.
- Quinlan J.R, Improved use of continuous attributes in c4.5. Journal of Artificial Intelligence Research, 4:77-90, 1996.
- Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
- Quinlan, R., “Data mining tools see5 and c5.0”, 2004.
- Rajaneer Ranjan “Encyclopaedia of Marketing Research”, Anmol Publications PVT. LTD., p.585, 2002.
- Reichheld F.F. et Sasser W.E., Zero Defections : Quality Comes to Services, Harvard Business Review, 68, p 105-111, Septembre-Octobre 1990.
- Reichheld F.F., The loyalty Effect, the Hidden Force Behind Growth, Profits and Lasting Value, Harvard Business School Press, 1996.
- Reilly, W.J., The law of retail gravitation, New York, 1931.
- Reinartz Werner, and Kumar V., “The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration”, Journal of Marketing, p.77-99, Janvier 2003.
- Rex Yuxing Du, Wagner A. Kamakura, Carl F. Mela., Size and Share of Customer Wallet, Journal of Marketing, Volume: 71, N° 2, p 94-113, 2007
- Romesburg C, “Cluster Analysis for Researchers”, Lulu press p.135, 2004
- Roux M., Algorithmes de classification Editions Masson, Paris, 1985.
- Rungie Cam, Uncles Mark and Laurent Gilles, “Consumer Characteristics and Behavioral Loyalty”, Marketing Science, décembre 2009.
- S. Kaski, J. Nikkila, and T. Kohonen “ Methods for Exploratory Cluster Analysis” Intelligent Exploration of the Web De Piotr S. Szczepaniak, Springer, 2003.
- Sharp A., Loyalty Programs and their impact on Repeat purchase Loyalty Patterns, International Journal of Research in Marketing, N° 14. p 473-486, 1997.
- Sharp B et Sharp A., Loyalty Programs and their impact on repeat purchase Loyalty Patterns : a replication and extension, 28th European Marketing Academy Conference Proceedings, Vol. CD; Berlin, Institute of Marketing II, Humboldt-University, 1999.
- Sharp Byron, “Loyalty limits for repertoire markets”, 2007

- Shepherdson N., Holding all the cards, American Demographics, Ithaca, Feb, 22, 2., p 35-37, 2000.
- Tan, Peter J.,Dowe David L., Dix Tevor I, Building classification model in two steps 1997
- Taylor, L., Telecommunications demand and theory and practice, Kluwer Academic Publishers, The Netherlands, 1994.
- Teuvo Kohonen. "Self-Organization and Associative Memory." Springer-Verlag, Berlin, 3rd edition, 1989.
- Timothy L. Keiningham, Bruce Cooil, Lerzan Aksoy, Tor W. Andreassen, Jay Weiner., The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet. *Managing Service Quality* 17:4, 361-384, 2007.
- Todd A. Stephenson "An Introduction to Bayesian Network Theory and Usage" IDIAP-RR 00-03, 2000.
- Trinquecost JF. , Fidéliser le consommateur : un objectif marketing prioritaire, DM, , p 17-23, Janvier-avril 1996.
- Tufféry Stéphane, Data mining et statistiques décisionnelles, Technip, 2010.
- Vallaud T. Estimation du potentiel individuel de chiffre d'affaires en utilisant des données issues d'une base de données clients, JePublie, 2012
- Vallaud T., La fidélisation rentable, la proposition d'un modèle composite entre la valeur d'un client à l'instant T et sa valeur à vie, JePublie, 2012
- Vallaud T., La fidélisation rentable : la proposition du modèle composite, www. numlog.com, 2003.
- Venkatesan, Rajkumar and V. Kumar, "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68, p.106-125, Octobre 2004.
- Verhoef Peter C., « Understanding the effect of customer relationship management efforts on customer retention and customer share development », *Journal of Marketing*, p.30-45, October 2003.
- Volle P., Promotion et choix du point de vente, Vuibert, 1999, p 221, p 222. Woolf B, Le marketing de la différence, Village Mondial, p28-29, 53-54, 1999.
- Wansink B. et Deshpande R. , Out of Sight, out of Mind Panty and Brand-Usage Frequency, *Marketing letters*, p 91-100, 5 janvier 1994.
- Wayland et Cole, Customer connections: new strategies for growth, Harvard Business Press, 1997
- Wheelwright, S. and McGee, V., *Forecasting : Methods and applications*, 2nd edition, John & Sons, Inc, 1983.
- Woolf B, Le marketing de la différence, Village Mondial, p28-29, 1999.
- Zufryden Fred S. and Jones J.Morgan, "Adding Explanatory Variables to a Consumer Purchase Behavior Model: An Exploratory Study", *Journal of Marketing Research*, p.323-334, aout 1980.

ANNEXES

Annexe 1 : Question sur la mesure de la valeur client par les entreprises françaises

Bonjour,

Nous réalisons une étude avec Direct Marketing News sur la mesure de la valeur client par des entreprises avancées en matière de marketing de base de données.

Merci de répondre par mail aux quelques questions ci-dessous.

Les résultats seront évidemment anonymes mais si vous êtes gênés par une question en particulier, n'hésitez à passer à la question suivante.
Merci d'avance pour votre collaboration.

Questions :

Segmentez-vous vos clients en fonction de leurs valeurs ?

Oui Non NSP

Pouvez-vous décrire brièvement les critères de calcul de cette valeur ?
.....
.....
.....
.....

Modélisez-vous dans cette valeur, les ventes croisées de produits et services ?

Oui Non NSP

Modélisez-vous dans cette valeur : les coûts d'acquisition ? Oui Non NSP

Les coûts de fidélisation ? Oui Non NSP

Les coûts de rétention ? Oui Non NSP

Pondérez-vous la valeur de vos clients par ces coûts ?

Oui Non NSP

Pourquoi ?
.....
.....
.....
.....

Tenez-vous compte de l'inflation dans l'estimation de la valeur client ?

Oui Non NSP

cette modélisation est faite : en interne ?

par un prestataire externe ?

Y a-t-il une structure 20/80 ? 20% de vos clients représentant 80% de votre MB ?

Oui Non NSP

Savez-vous si vos meilleurs/gros clients saturent en termes de consommation à votre marque ou sont-ils élastiques ? Oui Non NSP

S'ils sont élastiques comment estimez-vous cette élasticité ?

.....
.....
.....
.....

La valeur de vos clients est-elle calculée sur l'ensemble des produits ou services que vous proposez :
Est-ce :

en global, la somme de tous vos produits et services ?

par produit et service ?

NSP

La valeur de vos clients est-elle calculée que sur des sources internes : bases de données, autres.... ?

Oui Non NSP

Utilisez-vous des sources externes pour estimer la valeur de vos clients ? Si oui quel genre de sources et pourquoi ?

1°.....2°.....
3°.....4°.....

Pourquoi :

.....
.....
.....
.....

Connaissez-vous la contribution des produits (les plus contributifs) à la valeur de vos clients ?

Oui Non NSP

Connaissez-vous la durée de vie moyenne de vos clients :

via votre base de données ? Oui Non NSP

ou par une autre méthode ? Oui Non NSP

Si oui laquelle ?

.....
.....
.....
.....

Tenez-vous compte de la durée de vie de vos clients dans l'estimation de leurs valeurs ?

Oui Non NSP

L'avez-vous déterminé ? Oui Non NSP

Tenez-vous compte de l'augmentation de cette valeur :

En fonction des moments de vie de vos clients ? Oui Non NSP

Des événements clients ? Oui Non NSP

Appliquez-vous un raisonnement tenant compte des produits ou services que votre client peut acheter à vos concurrents ?

Oui Non NSP

Brièvement, de quelle façon, procédez-vous ?

.....
.....
.....

Y a-t-il un âge ou un moment de vie où vos clients sont au maximum de leur valeur ?

Oui Non NSP

L'avez-vous déterminé ?

Oui Non NSP

Connaissez-vous les méthodes, le modèle utilisé pour mesurer les valeurs de vos clients ?

Oui Non NSP

Précisez le ou lesquels si vous les connaissez de nom (s) :

1° 2°

3° 4°

Une approche de la valeur des clients est-elle utilisée à date de façon opérationnelle dans vos actions marketing ?

Oui Non NSP

Avez-vous des remarques par rapport à ce que l'on vient d'évoquer si oui lesquelles ?

.....
.....
.....

Merci de votre collaboration. Nous vous ferons parvenir les résultats de cette étude dans environ 5 semaines.

Annexe 2 : Glossaire CRM.

- A

Appels entrants :

Dans un centre d'appels, ce sont les appels qui sont reçus et répondus par les agents. Encore appelés inbound. Par opposition aux appels sortants.

Appels sortants :

Dans un centre d'appels, ce sont les appels qui sont émis par les agents vers l'extérieur. Encore appelés Outbound. Par opposition aux appels entrants.

- B -

Base de données :

Ensemble structuré de données relatives à un thème déterminé, et associé à un logiciel permettant son entretien et une utilisation efficace.

Base de données relationnelle :

Bases de données s'appuyant sur un modèle relationnel. Les données sont stockées dans des tables sous forme de lignes et de colonnes. SQLBase, Oracle et Sybase sont des exemples de SGBDR (Système de Gestion de Base de Données Relationnelle).

BtoB:

Abréviation de « Business to Business ». Désigne l'ensemble du commerce des entreprises vers d'autres entreprises.

BtoC:

Abréviation de « Business to Consumers ». Désigne l'ensemble du commerce des entreprises vers les particuliers.

- C

Campagne Marketing :

Ensemble d'actions visant à atteindre un objectif de marketing opérationnel (ex : qualification de données clients, prospection appliquée à une cible, en s'appuyant sur un médium (pub, mailing, contact téléphonique, événement et/ou un script)).

Centre d'appels :

Opérateur spécialisé de l'entreprise ou d'un prestataire externe qui, à partir d'une plate-forme téléphonique, émet (prospection, qualification) ou reçoit des appels (numéro vert, réclamations,...). Certains centres d'appels peuvent mettre en place des passerelles qui permettent à un internaute de se retrouver en communication téléphonique en cliquant sur un bouton de site web de l'entreprise.

Ciblage :

Technique consistant, par utilisation d'analyses de variables et de données internes, à concentrer un effort de promotion sur une population supposée intéressée par une offre donnée.

Cible :

En marketing, c'est le segment de marché que vous voulez atteindre à travers une stratégie commerciale déterminée.

Client actif :

Client ayant passé commande dans une période récente en général, moins de deux ans.

Clients les plus précieux (CPP) :

Les clients qui ont la valeur actualisée la plus élevée pour l'entreprise. Ce sont les clients avec lesquels l'entreprise fait le plus fort chiffre d'affaires, qui dégagent les marges les plus élevées, qui collaborent le plus volontiers et qui sont les plus fidèles.

Clients non rentables (CNR) :

Ils sont au niveau le plus bas dans la hiérarchie de la clientèle : ils coûtent plus cher à servir qu'ils ne rapporteront jamais. Ce groupe illustre à l'envers le Principe de Pareto : 20% des clients (appartenant à la tranche inférieure) représentent 80% des pertes, des ennuis, des appels de recouvrement, etc.

Conquête :

Voir RECRUTEMENT

Coût au mille :

Ensemble des coûts directs de mille messages affranchis.

Coût par commande :

Coût d'un mailing ,ou d'une annonce, divisé par le nombre de commandes reçues.

Coûts commerciaux :

En marketing direct, les coûts d'achat d'espace pour les annonces, la réalisation des mailings, les campagnes de télémarketing, seront considérés comme coûts commerciaux et non comme coûts de publicité.

CRM (Customer Relationship Management), CRM / eCRM /IRM :

Le CRM recouvre l'ensemble des fonctions de l'entreprise visant à conserver sa clientèle. Ce terme regroupe la gestion des opérations de marketing, l'aide à la vente, le service client, le centre d'appel et le helpdesk. Les progiciels de CRM du marché déclarent assurer une ou plusieurs de ces fonctions. En français, cette solution s'appelle GRC, « Gestion de la relation Client ».

Cross-Selling (vente croisée) :

Technique visant à tirer parti d'une transaction pour en effectuer une autre sur un produit/service complémentaire ou lié à un autre besoin lors d'un appel téléphonique.

Customer Care :

Soin du client. Technique de marketing qui permet d'accroître le CRM afin d'atteindre la satisfaction client maximum. Ensemble des moyens qui permettent d'augmenter la connaissance client ainsi que sa satisfaction globale vis à vis des produits et services qui lui sont proposés.

- D -

Datamart :

Ensemble de données limité à un domaine fonctionnel de l'entreprise. Un datamart peut être un composant du datawarehouse.

Data Mining :

Les outils dits de Data Mining (« forage des données » ou « prospection des données ») permettent de sélectionner dans l'ensemble des informations stockées par l'entreprise un certain nombre de données sur le client.

- E

ECR (Efficient Consumer Response) :

Traduit en français par « Efficacité et Réactivité au service du Consommateur » l'ECR est une stratégie Industrie/Commerce dans laquelle fournisseurs et distributeurs travaillent ensemble pour apporter une meilleure satisfaction au consommateur final. Les deux axes de ce système sont un échange d'information de qualité s'appuyant sur le principe du zéro-papier et un flux de marchandises sans rupture. Celui-ci, appliqué depuis la chaîne de fabrication jusqu'à la ligne de caisse du point de vente, doit minimiser les ruptures, les pertes et toutes les non-valeurs dans la relation entre les partenaires commerciaux.

EIS :

Abréviation d'Executive Information System. Terme désignant un frontal d'analyse marketing.

Enregistrement :

Ensemble de champs en relation. Dans une base de données relationnelle un enregistrement est une ligne. Dans une fiche, un enregistrement est l'ensemble des champs affichés. Ensemble de champs, d'information se rapportant à un même individu, à un même « objet », repéré par un identifiant.

ERM (Enterprise Relation Management) :

Ensemble de pratiques visant la gestion des relations entre une entreprise et ses tiers (clients, fournisseurs, partenaires...).

Expiré :

Client qui a été actif et qui désormais cesse de l'être.

- F

FAQ (Frequently Asked Question) :

Traduit en français par « Foire aux questions », ce terme désignait à l'origine le document contenant les réponses aux questions les plus fréquemment posées dans un serveur de news. Le FAQ est un document qui regroupe les questions et les réponses les plus courantes associées à un sujet.

Fichier de comportement :

Par oppositions aux fichiers de compilation, se dit de fichier regroupant des données de comportement d'achat ou demande de documentation qui qualifient ainsi les prospects et les clients. Les fichiers de la vente par correspondance, par exemple, sont des fichiers de comportement.

Fidélisation :

Technique de marketing qui concentre les efforts de l'entreprise sur le maintien de la clientèle actuelle à travers la Customer Satisfaction.

Free ride :

Se dit des offres de « VENTES CROISEE » véhiculées par un message dont la fonction principale est différente : facturation, relevé de compte, annonce de différé, remerciements, information complémentaire.

Fréquence :

Nombre d'achats effectués par une personne dans un délai donné (voir aussi RECENCE et MONTANT).

- G -

Géomarketing :

Maillage d'informations géographiques et socio-économiques, démographiques ou comportementales. La mise en œuvre du géomarketing nécessite des cartes, des BDD, et des logiciels SIG (Système d'Information Géographique). Exemples d'application : politique d'implantation idéale des points de vente, ciblage des campagnes de Marketing Direct.

GRC (Gestion de la relation client) :

Idem CRM, voir définition.

- H

Hotline/Help desk :

Ligne téléphonique réservée, uniquement utilisée pour conseiller ou dépanner à distance l'utilisateur d'une application. Certains Help Desk s'appuient également sur une messagerie. D'autres sur des applications de gestion des réclamations et demandes, permettent d'enregistrer, spécifier et documenter l'appel.

- L

Life Time Value :

Valeur d'un client calculée à partir de son comportement et ses habitudes d'achat au regard de ses préférences futures sur une longue période. L'intérêt de l'analyse peut être de repérer des clients encore peu rentables qui ont le potentiel de le devenir.

- M

Marketing direct :

Ce sont toutes les techniques qui permettent une relation one to one: mailing, phoning, fax-mailing, e-mailing, ...

Marketing One-to-One :

Personnalisation des actions de promotion et de communication en fonction des attentes et du profil de chaque client. A chaque client doit correspondre une action particulière. Cette personnalisation poussée n'est possible que si l'entreprise possède une base de données suffisamment qualifiée.

Marketing relationnel/situationnel :

Marketing relationnel :

Action par laquelle l'entreprise bâtit une relation continue et régulière avec ses clients (ou prospects) en dehors des relations ponctuelles de ventes. Cette relation permet d'améliorer la proximité client-entreprise.

Marketing situationnel :

Adaptation des actions de l'entreprise en fonction du moment ou du lieu. Sur internet le marketing situationnel se pratique essentiellement sur le critère temporel. Un fleuriste en ligne américain utilise par exemple des bandeaux publicitaires spécifiques dans les jours précédant le secretary-day.

Mass customization :

Personnalisation de la relation tout en traitant un nombre massif de clients.

Matrice de segmentation des clients :

Matrice positionnant les clients selon différents critères (valeur pour l'entreprise, comportement d'achat...).

Ménage :

Ensemble de personnes qui habitent dans une même résidence principale. Le nombre de ménages est égal au nombre de résidences principales.

Modélisation :

Technique statistique permettant d'expliquer un phénomène en fonction des données dont on dispose. On pourra modéliser les ventes d'un produit par rapport au nombre d'habitants et leur répartition en âge par exemple. Une modélisation peut donner des résultats plus ou moins précis. Cette précision est donnée par le pouvoir explicatif.

Une modélisation se définit également par sa sensibilité aux données qu'on a utilisées. Ainsi, le résultat d'une modélisation pourra augmenter dans les endroits où il y a plus de personnes âgées (sensibilité positive) et diminuer dans les endroits où il y a des ménages de taille importante (sensibilité négative)

MRC (Marketing Relationnel Ciblé) :

Terme récent employé pour désigner le marketing direct orienté par l'utilisation des bases de données, dans l'optique d'envoyer le message pertinent à la bonne personne.

-N-

Nettoyage de la base de données (cleaning) :

Opération consistant à effacer de la base de données des informations périmées sans intérêt pour l'exploitation. Cette opération donne lieu à un transfert sur bande magnétique pour stockage (et récupération toujours possible pendant quelques années).

Nouveaux entrants :

Se dit des clients nouvellement acquis (recrutés) par un moyen ou un autre (mailing, annonce presse...). Certains praticiens mettent en œuvre une communication des offres spécifiques pour ces clients particuliers et ce pendant un laps de temps variable (aussi longtemps que leur comportement est discriminant par rapport à la population de base).

-O-

Offre :

Élément de base du marketing direct, qui doit être l'objet d'une réflexion préalablement à toute commercialisation. L'offre regroupe les notions de produit, prix, garantie, service, consommation, mix média, conditions de paiement, période, cadeaux, incitations diverses... La réflexion sur l'offre conditionne le plan de tests et donc de succès de toute communication directe.

Outils décisionnels :

(= outils intelligents) : ensemble d'outils technologiques permettant d'extraire, de stocker et d'analyser un ensemble vaste de données afin de déterminer des actions adéquates et ciblées sur un marché. Une infrastructure « intelligente » intègre notamment la maîtrise de l'amont et des opérations (fournisseurs), le suivi et le pilotage des activités, enfin la connaissance et la satisfaction du client.

-P-

Panel :

Nombre limité de clients dont les réactions ou comportements sont étudiés afin de préparer un mailing ou un test.

Part de client :

A la différence de la part de marché, « la part de client » représente le pourcentage de CA que détient l'entreprise sur le flux d'activités d'un client en particulier, au cours de la durée de vie de ce client dans l'entreprise.

Pénétration :

Pourcentage du nombre de clients obtenus sur un fichier par rapport à ceux que l'on pourrait encore obtenir dans l'avenir sur ce même fichier.

Point mort :

En marketing Direct, le point mort est le rendement calculé comme suit : coût de promotion / marge brut produit.

Potentiel :

Le potentiel caractérise les volumes de vente ou le nombre de clients ou de prospects qu'il est raisonnable d'atteindre sur les différents endroits d'un territoire que l'on examine. Il se définit à l'aide de vos propres données de vente ou de prospection et des données qui caractérisent le territoire (caractéristiques de la population ou des entreprises).

Pouvoir explicatif :

Pourcentage qui définit la précision d'une modélisation.

Prescripteur :

Acteur en situation d'influence sur le choix d'un service ou produit (relation non contractuelle).
Processus : Enchaînement d'opérations visant un objectif donné transversalement à une organisation.

Profiling :

En français : personnalisation. Sur un site web, stratégie consistant à proposer à l'internaute un environnement de navigation familier et adapté à ses besoins (personnalisation à l'accueil, d'environnement, ou outils personnalisés). L'utilisation d'identifiants, de cookies et de mots de passe va permettre cette personnalisation.

Prospect :

Client potentiel « qualifié » destiné à être amené jusqu'à l'acte d'achat à travers les étapes du processus de vente représentées par le Carnet d'Affaires.

Push/pull :

La méthode « traditionnelle » de consultation d'un site Web consiste, pour l'utilisateur, à taper l'adresse du site qui l'intéresse sur le réseau et, à l'aide de son navigateur, à charger les pages recherchées afin de les afficher. Cette méthode, le « pull », consiste à aller chercher l'information ou à la tirer vers soi. La méthode du « push » fonctionne de façon inverse. En effet, plusieurs logiciels permettent désormais de déterminer le profil de l'utilisateur et de lui proposer un contenu correspondant précisément à ses goûts ou intérêts. Contrairement aux cookies, ces techniques ne violent pas la vie privée de l'utilisateur car elles répondent à un mode déclaratif et sont utilisées à son initiative. Elles permettent notamment d'établir un tableau de veille dans lequel l'utilisateur référence ses préoccupations particulières (suivi de résultats sportifs, cours des actions en Bourse...). Lorsqu'il interroge ce tableau, l'utilisateur voit automatiquement apparaître les informations relatives aux préoccupations qu'il a déclarées.

- Q**Qualification :**

Affectation au prospect, d'un code d'intérêt pour la ou les offres proposées par l'entreprise. Cette qualification peut se faire plus ou moins directement. Indirectement via l'analyse d'une base de données ou l'ajout de données externes, plus directement par une offre de documentation ou d'information, très directement par télémarketing.

Questionnaire de qualification (carte typo) :

Utilisé parfois auprès de clients actifs ou prospect pour les qualifier avec quelques questions clefs.

- R

Récence :

Le dernier achat effectué par une personne figurant sur un fichier déterminé.

Recrutement :

Opération de recherche de clients par voie postale.

Relance :

Opération complémentaire succédant à un mailing afin de recruter de nouveaux clients sur un même fichier.

Relation d'apprentissage :

Définition One to One : Une relation privilégiée qui se noue entre une entreprise et son client et qui se nourrit d'une remontée d'informations régulière et suivie de la part du client. L'entreprise est alors en mesure de satisfaire de plus en plus précisément les besoins individuels du client. Elle utilise la technologie de personnalisation de masse pour créer sur mesure un produit ou un service qui corresponde aux besoins réels du client. Lorsque le client et l'entreprise nouent une relation d'apprentissage, chaque nouveau cycle d'interaction suivi de personnalisation montre au client combien il lui est facile de traiter avec cette entreprise. En définitive, cela renforce la fidélité du client vis-à-vis de l'entreprise. Car si le client devait recréer une relation analogue avec une autre société, il devrait réapprendre au concurrent ce que cette entreprise connaît déjà. A mesure que la relation d'apprentissage s'intensifie, le client y trouve des avantages accrus, ce qui cimente sa fidélité et protège par conséquent la mare de l'entreprise.

Rendement/pull :

Exprime, en pourcentage, le rapport entre le nombre de prospections envoyées et le nombre de réponses ou de commandes reçues. La signification du résultat obtenu dépend de la taille de l'échantillon et du niveau de rendement.

Retours :

Se dit des retours de produits, soit dans le cadre de la garantie « satisfait ou remboursé » soit dans celui des offres à l'essai. Les retours et leur observation sont des éléments fondamentaux de la stratégie des entreprises de ventes à distance. Les retours de coupons réponses ou de bons de commandes sont plus souvent appelés REPONSES OU RENDEMENT.

Revenus :

Autre terme parfois employé pour désigner le chiffre d'affaires d'une offre ou d'une opération de prospection.

RFM (Récence, Fréquence, Montant) :

Mode de sélection, voire de segmentation fondé sur une combinaison de 3 facteurs essentiels de la relation client/fournisseur (récence d'achat, fréquence d'achat, cumul des montants achetés).

ROI (Return On Investment) :

Retour sur investissement.

-S**SCM – Supply Chain Management :**

En français : Gestion de la Chaîne Logistique. Il s'agit d'outils et de fonctions décisionnelles appliquées à l'ensemble des opérations d'approvisionnement, de distribution et de stockage d'une entreprise.

Scoring :

Importée des Etats-Unis, cette méthode statistique consiste à attribuer une « note » à un client afin d'évaluer son potentiel commercial ou son niveau de risque (dans le cadre d'un dossier de crédit par ex). L'évaluation s'effectue par la combinaison de plusieurs critères qui aboutit à une grille de notation.

Segment de marché : Un segment de marché correspond à un groupe déterminé de clients présentant des caractéristiques similaires.

Segmentation :

Fragmentation d'un fichier en segments contenant chacun des adresses possédant des caractéristiques similaires.

SFA – Sales Force Automation :

Terme anglais pour Automatisation de la force de vente. Il s'agit d'optimiser l'efficacité des forces de vente en automatisant les flux d'information au moyen de progiciels appropriés.

SIAD :

Terme désignant les systèmes d'aide à la décision dans les frontaux d'analyse marketing.

SIG (Système d'Information Géographique) :

Palette d'outils logiciels et matériels réalisant des analyses spatiales et affichant des images géographiques, telles que des cartes, au moyen d'une base de données géoéconomique qui modélise la réalité. Il est très souvent utilisé comme système d'informations capables de résoudre les besoins marketing et commerciales. D Interactive intègre le SIG MapInfo comme une composante dans la réflexion globale du besoin en système d'information marketing du client.

Smart card :

Carte intelligente, carte à puce.

Source :

Origine du fichier et du matériel utilisé pour une prospection donnée. Le code source, élément essentiel du marketing direct, reste attaché au client tout au long de son enregistrement dans une base de données.

Split-run :

Publication de deux annonces différentes dans le même nombre d'exemplaires d'un même journal ou magazine.

Stratégie one to one:

Cette stratégie vise à donner au client ou au prospect, le sentiment d'être reconnu en tant qu'individu et de resserrer le lien avec l'entreprise. Cette politique a pour effet d'augmenter la satisfaction et la fidélisation des clients ; elle donne à l'entreprise, par l'échange régulier d'informations, une meilleure connaissance de ses clients. Enfin elle permet d'optimiser les coûts dans le cadre de campagnes marketing par ex.

Supply chain :

Indique l'ensemble des unités organisationnelles qui interviennent dans le processus de production d'un bien ou d'un service prêt à satisfaire les exigences d'un utilisateur final

SVI (Serveur Vocal Interactif) :

Cet outil vocal utilisé dans les centres d'appels, permet d'identifier les appelants, de les aiguiller vers des services automatiques (de type consultation des comptes bancaires, horaires SNCF ou des agents compétents). Permet de désengorger le centre notamment au niveau des demandes basiques des appelants.

Système d'information marketing :

C'est un outil qui permet de se mesurer et de se comparer par rapport à un marché et des concurrents. En outre, il optimise l'organisation territoriale et permet le couplage avec l'action et le contrôle.

- T**Tableau de bord :**

C'est le résultat de votre étude que vous allez pouvoir télécharger sur votre ordinateur et manipuler à loisir. Un tableau de bord est composé de tableaux, de graphes ou de cartes.

Taux d'attribution :

Se dit du taux de chute d'un effort de promotion par rapport à celui qui le précède, dans un système de série ouverte, d'option négative, de facturation, de renouvellement.

Taux de conversion :

Terminologie souvent utilisée dans les offres à 2 tours. Ces offres consistent, par exemple via un coupon de presse, à proposer une documentation ou un échantillon du produit avant de proposer le produit lui-même. Le taux de conversion est le pourcentage de produit ou de service commandé à la suite de l'envoi de l'échantillon ou de l'information demandée.

Taux de pénétration :

Pourcentage définissant la pénétration des ventes d'un produit. Si, parmi une commune de 10 000 ménages, 50 sont clients du produit, le taux de pénétration est de 0.5%.

Taux de renouvellement :

Par exemple pourcentage d'abonnés qui renouvellent leur abonnement l'année A+1, A+2... C'est du fait de la chute de taux que les entreprises de presse doivent prospecter chaque année de nouveaux clients pour maintenir leur « CIRCULATION ».

Temps de réponse :

Intervalle de temps qui passe entre une demande de l'utilisateur à l'entreprise et la réponse à cette dernière.

Tests :

Fondement même de toutes les techniques de marketing direct. Presque tout peut être testé : offre, produit, copie,... et, dans des limites raisonnables de rentabilité, doivent l'être.

- U**Up-Selling :**

Concept marketing visant à vendre des produits d'une gamme supérieure à celle déjà réalisée.

Annexe 3 : Tableaux, graphiques et figures

Variables	Description
ID Client	Pour lier les tables entre elles sur un même client
Ratio Autre	% du chiffre d'affaire dans la catégorie de produit
Ratio Bazar	% du chiffre d'affaire dans la catégorie de produit
Ratio BOF/APF	% du chiffre d'affaire dans la catégorie de produit
Ratio Charcuterie LS	% du chiffre d'affaire dans la catégorie de produit
Ratio Clients Animaux	% du chiffre d'affaire dans la catégorie de produit
Ratio client bébés	% du chiffre d'affaire dans la catégorie de produit
Ratio clients boucherie	% du chiffre d'affaire dans la catégorie de produit
Ratio boulangerie	% du chiffre d'affaire dans la catégorie de produit
Ratio charcuterie	% du chiffre d'affaire dans la catégorie de produit
Ratio dietetiques bio	% du chiffre d'affaire dans la catégorie de produit
Ratio fromage	% du chiffre d'affaire dans la catégorie de produit
Ratio fruits et légumes	% du chiffre d'affaire dans la catégorie de produit
Ratio poisson	% du chiffre d'affaire dans la catégorie de produit
Ratio surgelés	% du chiffre d'affaire dans la catégorie de produit
Ration vins	% du chiffre d'affaire dans la catégorie de produit
Ratio DPH	% du chiffre d'affaire dans la catégorie de produit
Ratio épicerie	% du chiffre d'affaire dans la catégorie de produit
Ratio liquide	% du chiffre d'affaire dans la catégorie de produit
Ratio textile	% du chiffre d'affaire dans la catégorie de produit
Ratio ultra frais	% du chiffre d'affaire dans la catégorie de produit
Ratio volaile	% du chiffre d'affaire dans la catégorie de produit
Ratio Premier Prix	% du chiffre d'affaire dans la catégorie de produit
Ratio Marque Distributeur 1	% du chiffre d'affaire dans la catégorie de produit
Ratio Marque Distributeur 2	% du chiffre d'affaire dans la catégorie de produit
Nombre d'enfants au foyer	
CA Filtré 12 mois	Chiffre d'affaire diminué de quelques dépenses hors mag
CA Total	Chiffre d'Affaire total
CA Promo Annuel	CA annuel réalisé sous promo
Nb de points transformés sur 12 mois	Nb de point "promotionnels" utilisés sur une année
CM transformé sur 12 mois	Part CA ou le client a utilisé des points promotionnels
Cumul nb BA pris	Cumul des bons d'achat que le client a utilisé
PMG 12 mois	Segmentation
RFM 3 mois	Segmentation

Annexe 2 : Détail du premier audit des données

L'ensemble des données a été audité en deux étapes : une première étape pour déterminer toutes les données utiles à l'analyse dans la base de données originale, et une seconde étape pour déterminer les données disponibles pour calculer un potentiel. Dans cette annexe, seule la seconde étape est présentée.

Analyse des tables "Potential_Ratio" et Potentiel_SOCIO »

La table Potentiel_Ratio contient 5 373 048 observations (comptes client)
Elle est composée de 26 champs

La table Potentiel_Socio contient 5 373 056 observations (comptes client)
Elle est composée de 18 champs

Cet audit est basé sur la combinaison des deux tables soit 5 373 048 observations

Le format des données

C'est le format original des données reçues. Nous avons dû changer certains formats pour mieux atteindre nos objectifs de modélisation.

Variables	Description
Identifiant Client	Pour lier les tables entre elles sur un même client
Ratio Autre	% du chiffre d'affaire dans la catégorie de produit
Ratio Bazar	% du chiffre d'affaire dans la catégorie de produit
Ratio BOF/APF	% du chiffre d'affaire dans la catégorie de produit
Ratio Charcuterie LS	% du chiffre d'affaire dans la catégorie de produit
Ratio Clients Animaux	% du chiffre d'affaire dans la catégorie de produit
Ratio client bébés	% du chiffre d'affaire dans la catégorie de produit
Ratio clients boucherie	% du chiffre d'affaire dans la catégorie de produit
Ratio boulangerie	% du chiffre d'affaire dans la catégorie de produit
Ratio charcuterie	% du chiffre d'affaire dans la catégorie de produit
Ratio dietetiques bio	% du chiffre d'affaire dans la catégorie de produit
Ratio fromage	% du chiffre d'affaire dans la catégorie de produit
Ratio fruits et légumes	% du chiffre d'affaire dans la catégorie de produit
Ratio poisson	% du chiffre d'affaire dans la catégorie de produit
Ratio surgelés	% du chiffre d'affaire dans la catégorie de produit
Ration vins	% du chiffre d'affaire dans la catégorie de produit
Ratio DPH	% du chiffre d'affaire dans la catégorie de produit
Ratio épicerie	% du chiffre d'affaire dans la catégorie de produit
Ratio liquide	% du chiffre d'affaire dans la catégorie de produit
Ratio textile	% du chiffre d'affaire dans la catégorie de produit
Ratio ultra frais	% du chiffre d'affaire dans la catégorie de produit
Ratio volaile	% du chiffre d'affaire dans la catégorie de produit
Ratio Premier Prix	% du chiffre d'affaire dans la catégorie de produit
Ratio Marque Distributeur 1	% du chiffre d'affaire dans la catégorie de produit
Ratio Marque Distributeur 2	% du chiffre d'affaire dans la catégorie de produit
Nombre d'enfants au foyer	
CA Filtré 12 mois	Chiffre d'affaire diminué de quelques dépenses hors mag
CA Total	Chiffre d'Affaire total
CA Promo Annuel	CA annuel réalisé sous promo
Nb de points transformés sur 12 mois	Nb de point "promotionnels" utilisés sur une année
CM transformé sur 12 mois	Part CA ou le client a utilisé des points promotionnels
Cumul nb BA pris	Cumul des bons d'achat que le client a utilisé
PMG 12 mois	Segmentation
RFM 3 mois	Segmentation

RFM 3 mois la variable est vide, pour cette raison elle est écartée

Analyse variable par variable

Les variables discrètes

RFM 3 mois	Effectif	%
Nouveaux	247 326	4.60%
Anciens clients	400 236	7.45%
Inactifs	465 107	8.66%
M--F--	1 315 873	24.49%
M-F-	1 302 089	24.23%
M-F+	248 619	4.63%
M+F-	710 311	13.22%
M+F+	683 487	12.72%
Total	5 373 048	100.00%

Comparaison avec le 1 er audit
3.74%
19.15%
21.00%
26.00%
6.06%
11.65%
12.40%
100.00%

Statut familial	Effectif	%
Couple	1 557 871	28.99%
Célibataire	642 374	11.96%
Vide	3 172 803	59.05%
Total	5 373 048	100.00%

Comparaison avec le 1 er audit
26.19%
10.81%
62.99%
100.00%

PMG sur 12 mois	Effectif	%
NA	7 816	0.15%
Nouveau	245 278	4.56%
Inconnu	2 573	0.05%
P	3 086 955	57.45%
M	1 014 777	18.89%
G	1 015 649	18.90%
Total	5 373 048	100.00%

Type d'habitat	Effectif	%
Appartement	880 722	16.39%
Maison et appartement	1 300	0.02%
Maison	1 576 829	29.35%
Vide	2 914 197	54.24%
Total	5 373 048	100.00%

Comparaison avec le 1 er audit
18.76%
0.00%
34.98%
65.02%
100.00%

Nombre d'enfants dans le foyer	Effectif	%
0	4 088 282	76.09%
1	510 448	9.50%
2	508 967	9.47%
3	196 520	3.66%
4	48 644	0.91%
5	11 482	0.21%
> 5	8 705	0.16%
Total	5 373 048	100.00%

Comparaison avec le 1 er audit
75.72%
9.57%
9.54%
3.79%
1.01%
0.22%
0.14%
100.00%

Catégories sociales	Effectif	%
Fermier	49 892	0.93%
Artisan	86 807	1.62%
Autre	84 696	1.58%
Directeur	188 017	3.50%
Employé	737 469	13.73%
Etudiant	85 928	1.60%
Femme au foyer	211 003	3.93%
Fonctionnaire	233 386	4.34%
Travailleur indépendant	42 913	0.80%
Ouvrier	138 099	2.57%
Retraité	664 304	12.36%
Sans emploi	147 700	2.75%
Technicien	91 956	1.71%
Vide	2 610 877	48.59%
24	1	0.00%
Total	5 373 048	100.00%

Comparaison avec le 1 er audit
0.96%
1.79%
1.39%
3.50%
14.42%
1.28%
4.38%
3.70%
0.72%
2.73%
14.05%
3.28%
2.08%
45.71%
0.00%
100.00%

La valeur 24 est une erreur, nous l'avons éliminé.

Age	Effectif	%
0 à 18 ans	8 604	0.16%
19 à 29 ans	317 371	5.91%
30 à 39 ans	537 010	9.99%
40 à 49 ans	652 497	12.14%
50 à 59 ans	649 038	12.08%
60 à 69 ans	458 669	8.54%
70 ans et plus	592 227	11.02%
Vide	2 157 632	40.16%
Total	5 373 048	100.00%

Comparaison avec le 1 er audit
0.21%
6.00%
10.76%
12.59%
12.30%
8.07%
10.92%
39.15%
100.00%

Historique client	Effectif	%
0 à 2 mois	194 269	3.62%
3 à 5 mois	238 267	4.43%
6 à 8 mois	182 733	3.40%
9 à 11 mois	221 313	4.12%
12 à 17 mois	354 680	6.60%
18 à 23 mois	231 513	4.31%
24 à 35 mois	517 972	9.64%
36 à 47 mois	396 706	7.38%
48 à 59 mois	403 389	7.51%
60 mois et plus	2 631 717	48.98%
Vide	489	0.01%
Total	5 373 048	100.00%

Comparaison avec le 1 er audit
2.76%
2.54%
3.20%
3.26%
6.39%
6.17%
9.58%
7.96%
10.01%
44.60%
3.53%
100.00%

Temps depuis le dernier achat	Effectif	%
0 ta 2 mois	4 213 358	78.42%
3 à 5 mois	562 908	10.48%
6 à 8 mois	312 721	5.82%
9 à 11 mois	241 093	4.49%
12 à 17 mois	42 968	0.80%
Total	5 373 048	100.00%

Les variables numériques

	Ratio autre	Ratio bazar	Ratio BOF / APF	Ratio charcuterie allégée	Ratio Pet	Ratio beauté/maquillage	Ratio bébé	Ratio boucherie
Effectif	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055
Moyenne	77.1	7.0	9.0	6.5	1.2	0.7	1.4	6.5
Min	-34 035.0	-27 666.7	-195.2	-1 580.0	-15.3	-61.2	-23.5	-125.2
Max	294.9	5 328.9	9 685.6	3 186.7	1 206.5	574.5	3 614.4	4 193.0
SD	22.2	16.3	8.4	6.7	3.7	2.4	5.8	8.6

	Ratio boulangerie	Ratio charcuterie	Ratio nourriture diététique	Ratio fromage	Ratio fruits et légume	Ratio poisson	Ratio surgelé	Ratio vin
Effectif	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055
Moyenne	2.5	2.6	0.4	1.5	8.4	1.9	3.1	2.1
Min	-138.5	-74.5	-11.3	-29.4	-105.5	-200.0	-308.3	-530.3
Max	1 247.6	1 373.1	2 428.6	262.7	17 364.3	1 542.9	11 528.6	610.2
SD	4.9	4.7	2.3	2.9	12.0	4.5	7.6	5.4

	Ratio produits d'entretien	Ratio épicerie	Ratio liquide	Ratio textile	Ratio ultra frais	Ratio volaille	Ratio premier prix	Ratio marque distributeur 1	Ratio marque distributeur 2
Effectif	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055	5 373 055
Moyenne	8.9	16.8	10.8	2.0	4.7	1.7	6.2	14.6	2.0
Min	-1 472.8	-1 463.6	-2 843.2	-1 250.0	-372.3	-37.5	-393.8	-194.9	-30.9
Max	14 007.1	16 122.2	6 680.4	2 713.2	5 814.3	2 822.8	26 242.0	7 501.1	2 031.8
SD	12.4	14.9	13.8	5.4	6.4	3.8	15.0	11.5	3.4

Les champs monétaires n'ont pas de décimal. Nous avons divisé le CA par 100.

PMG 12 mois	Effectif	% de clients	Chiffre d'affaire filtré sur 12 mois	% de chiffre d'affaire filtré sur 12 mois	Chiffre d'affaire filtré sur 12 mois: Moyenne par client
NA	7 816	0.15%	0 €	0.00%	0.0
New	245 278	4.56%	40 365 377 €	0.54%	164.6
I	2 573	0.05%	0 €	0.00%	0.0
S	3 086 955	57.45%	1 185 154 005 €	15.87%	383.9
M	1 014 777	18.89%	1 772 044 857 €	23.72%	1 746.2
L	1 015 649	18.90%	4 471 932 029 €	59.87%	4 403.0
Total	5 373 048	100.00%	7 469 496 268 €	100.00%	1 390.2

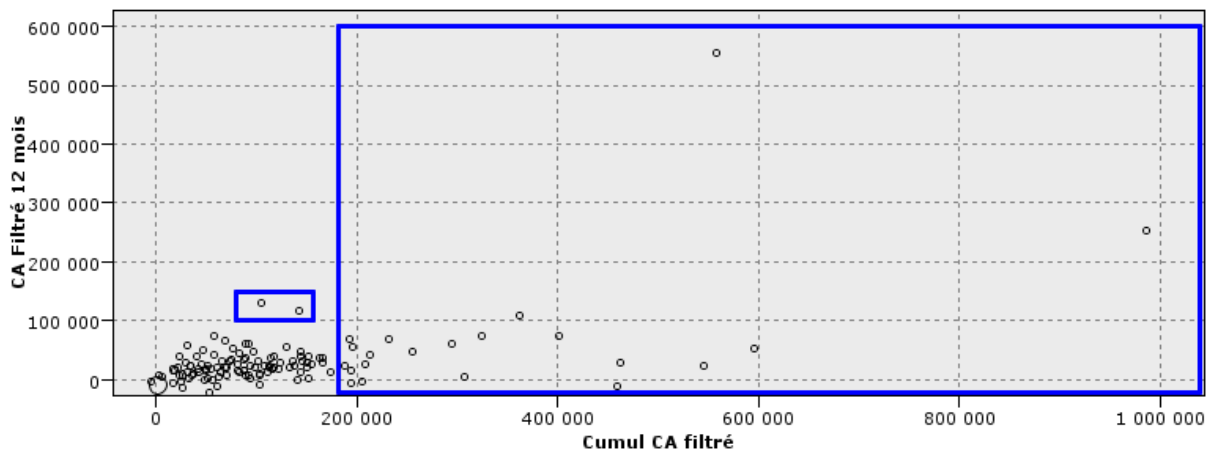
PMG 12 mois	Chiffre d'affaire total	% du chiffre d'affaire total	Chiffre d'affaire total: Moyenne par client	Chiffre d'affaire filtré cumulé	% du chiffre d'affaire filtré	Chiffre d'affaire filtré : Moyenne par client
NA	29 189 799 €	0.06%	3 734.6 €	16 060 558 €	0.05%	2 054.8 €
New	118 781 921 €	0.25%	484.3 €	91 122 432 €	0.31%	371.5 €
I	19 120 403 €	0.04%	7 431.2 €	9 701 975 €	0.03%	3 770.7 €
S	11 506 545 207 €	23.89%	3 727.5 €	6 400 520 698 €	21.51%	2 073.4 €
M	10 946 321 611 €	22.72%	10 786.9 €	6 841 635 615 €	22.99%	6 742.0 €
L	25 549 213 252 €	53.04%	25 155.6 €	16 398 620 825 €	55.11%	16 146.0 €
Total	48 169 172 193 €	100.00%	8 965.0 €	29 757 662 104 €	100.00%	5 538.3 €

PMG 12 mois	Chiffre d'affaire annuelle sous promo	% du chiffre d'affaire annuelle sous promo	Chiffre d'affaire annuelle sous promo : Moyenne par client	Total des bons de réduction pris (BA)	% des bons de réduction pris (BA)	Total des bons de réduction pris (BA): Moyenne par client
NA	173 429 €	0.04%	22.2	1 529	0.07%	0.20
New	4 749 065 €	1.02%	19.4	4 156	0.18%	0.02
I	16 358 €	0.00%	6.4	727	0.03%	0.28
S	84 017 089 €	18.12%	27.2	235 034	10.11%	0.08
M	109 248 453 €	23.56%	107.7	469 769	20.21%	0.46
L	265 468 817 €	57.25%	261.4	1 613 749	69.41%	1.59
Total	463 673 210 €	100.00%	86.3	2 324 964	100.00%	0.43

PMG 12 mois	Nb. de points transformés sur les 12 mois	% de points transformés sur les 12 mois	Nb. de points transformés sur les 12 mois : Moyenne par client	Nb de CM transformé dans les 12 mois	% de CM transformé dans les 12 mois	Nb de CM transformé dans les 12 mois : Moyenne par client
NA	5 916 575	0.08%	756.98	7 870.7	0.03%	1.01
Nouveau	13 551 700	0.18%	55.25	127 290.4	0.54%	0.52
Inconnu	4 396 785	0.06%	1 708.82	501.4	0.00%	0.19
P	850 847 714	11.31%	275.63	6 024 886.1	25.34%	1.95
M	1 628 377 733	21.64%	1 604.67	5 079 909.9	21.37%	5.01
G	5 020 142 644	66.73%	4 942.79	12 531 310.3	52.72%	12.34
Total	7 523 233 151	100.00%	1 400.18	23 771 769.0	100.00%	4.42

PMG 12 mois	Nb de point bonus utilisés	% de point bonus utilisés	Nb de point bonus utilisés : Moyenne par client
NA	516 210	0.03%	66.05
Nouveau	3 972 503	0.23%	16.20
Inconnu	50 615	0.00%	19.67
Petit	115 501 391	6.80%	37.42
Moyen	376 564 221	22.18%	371.08
Gros	1 201 136 053	70.75%	1 182.63
Total	1 697 740 993	100.00%	315.97

La recherche des clients aberrants



Dans ce nouveau fichier, obtenu après le premier audit nous n'avons pas le nombre d'achats (corrélé avec le chiffre d'affaires) ; dans le but d'identifier les individus aberrants, nous avons utilisé les deux versions du chiffre d'affaires. Ci-dessous les ID des clients aberrants.

ID Client
2012027201917
2012030213679
2012031991910
2012033265293
012036595809
2012043773283
2012047323217
2012049821087
2012059641187
3228461599988
3228465186504
3228465989082
3547800078848
3572567264395
3594036148395
3594037801794
3597040504980
3597045375189
3597055994080
3597059399003
3600060213265
3600067130046

La rentabilité de la fidélisation du consommateur

Résumé :

Dans cette thèse sur travaux, l'auteur part de deux travaux précédents sur la rentabilité de la fidélisation et la détermination du potentiel client pour faire un constat : une partie de la rentabilité de la fidélisation et du potentiel pour une marque est basée sur la part captable du chiffre d'affaire fait à la concurrence ; le taux de captation.

Dans ce nouveau travail il s'agit de montrer que le taux de captation est basé sur l'élasticité du taux de nourriture. A partir d'une analyse de la littérature et de plusieurs modélisations sur des données de panel scannées, l'auteur démontre, sur plusieurs marchés, que l'élasticité du taux de nourriture est contrainte et prévisible.

C'est donc en tenant compte de cet écart limité qu'une marque peut estimer le taux de captation et donc la rentabilité de la fidélisation ainsi que le potentiel client.

Mots Clés :

fidélisation, taux de nourriture, élasticité, potentiel client, LTV, taux de captation, méthode des clones, prévision, panel de home scanning, ROI, valeur client, gros volume de données, modèle désagrégé

Résumé :

In this thesis based on works the author goes from two previous studies on the profitability of loyalty and customer potential determination to make a statement : part of the profitability of loyalty and of the potential for a brand is based on the reachable share of turnover done by the competition ; the catch rate.

In this new work it is shown that the catch rate is based on the elasticity of the share of category requirement. From a review of the literature and several modeling on scanning panel data the author demonstrates on several markets that elasticity of the share of category requirement is limited and predictable.

Then it's in taking into account this small difference that a brand can estimate the "catchable" rate and therefore the profitability of loyalty and potential of a customer.

Mots Clés :

Loyalty, share of requirement, elasticity, customer's potential, LTV, catch rate, clones method forecasting, home scanning panel home, ROI, customer value, large volume of data (big data), disaggregated model