



# Time organization for urban mobility congestion : an interdisciplinary approach for Bureaux des Temps

Yassamina Berkane

## ► To cite this version:

Yassamina Berkane. Time organization for urban mobility congestion : an interdisciplinary approach for Bureaux des Temps. Artificial Intelligence [cs.AI]. Université Paris-Saclay, 2023. English. NNT : 2023UPASG071 . tel-04426230

**HAL Id: tel-04426230**

**<https://theses.hal.science/tel-04426230v1>**

Submitted on 30 Jan 2024

**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.

# Time organization for urban mobility congestion: an interdisciplinary approach for Bureaux des Temps

*Organisation du temps pour la décongestion de la  
mobilité urbaine : une approche interdisciplinaire pour  
les Bureaux des Temps*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n° 580 Sciences et technologies de l'information et de la  
communication (STIC)

Spécialité de doctorat: Informatique

Graduate School : Informatique et sciences du numérique

Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **DAVID** (Université Paris-Saclay, UVSQ), sous la  
direction de **Leïla KLOUL**, maître de conférences (HDR), et le co-encadrement de **Yoann  
DEMOLI**, maître de conférences

**Thèse soutenue à Versailles, le 20 décembre 2023, par**

**Yassamina BERKANE**

## **Composition du jury**

Membres du jury avec voix délibérative

**Dominique BARTH**

Professeur des universités, Université de Versailles Saint-  
Quentin-en-Yvelines

Président

**Vincent KAUFMANN**

Professeur des universités, École Polytechnique Fédérale  
de Lausanne

Rapporteur & Examineur

**Jakob PUCHINGER**

Professeur des universités, EM Normandie Business  
School

Rapporteur & Examineur

**Frédéric CHAILLAN**

Docteur, Aménagement Numérique de Grand Paris Sud

Examineur

**Kamila TABAKA SIMON**

Maître de conférences, Institut d'Urbanisme et de Géo-  
graphie Alpine

Examinatrice

**Titre:** Organisation du temps pour la décongestion de la mobilité urbaine : une approche interdisciplinaire pour les Bureaux des Temps

**Mots clés:** Mobilité urbaine, approche interdisciplinaire, gestion du temps, analyse des données, apprentissage automatique, *Bureaux des Temps*.

**Résumé:** Soutenue par la croissance démographique au sein des zones urbaines, l'augmentation de la mobilité quotidienne est un défi majeur pour la gestion des villes et de leurs infrastructures routières, dans une optique de croissance économique et de développement durable. Dans cette thèse, le concept de gestion du temps est intégré dans une nouvelle approche pour réduire la congestion de la mobilité urbaine. Les méthodes traditionnelles se concentrent principalement sur les aspects spatiaux, négligeant ainsi la dimension temporelle. Dans l'approche proposée, nous nous intéressons à la prise en compte de concepts des Sciences Humaines et Sociales (SHS), en particulier les concepts sociologiques, aux Sciences et Technologies de l'Information et de la Communication (STIC) pour proposer

une solution d'aide à la décision pour les *Bureaux des Temps*, qui vise à modifier les temporalités afin de réduire la congestion urbaine. Cette approche intègre trois phases : identifier les profils de mobilité, analyser la congestion du trafic et prendre des décisions temporelles. Les profils de mobilité sont prédits à l'aide de techniques d'apprentissage en prenant en compte des critères sociologiques. L'analyse de la congestion du trafic exploite des données de l'application Waze. Enfin, l'utilisation des modèles de séries temporelles nous permettent de prédire les degrés de congestion afin de proposer des temps de départ optimisés pour éviter la congestion urbaine. La solution proposée a le potentiel d'intégrer un ensemble de données hétérogènes dans la gestion de la congestion pour des villes urbaines harmonieuses.

**Title:** Time organization for urban mobility congestion: an interdisciplinary approach for Bureaux des Temps

**Keywords:** Urban mobility, interdisciplinary approach, time management, data analysis, machine learning, time bureaus.

**Abstract:** Population growth and urbanization are major challenges for the management of cities and their resources. Urban mobility is one of these challenges because of its impact on quality of life, economic productivity, and environmental sustainability. In this thesis, the concept of time management is integrated into a new approach to reduce urban mobility congestion. Traditional methods primarily focus on spatial aspects, thus neglecting the temporal dimension. We concentrate on integrating concepts from the Social and Human Sciences (SHS), particularly sociological concepts, into Sciences and Technologies of Information and Communication (STIC) to propose a decision-

support solution for the time bureaus, which aims to modify temporalities to reduce urban congestion. The proposed approach consists of three phases: identifying mobility profiles, analyzing traffic congestion, and making temporal decisions. Mobility profiles are predicted using learning techniques that take into account sociological criteria. Traffic congestion analysis investigates Waze data. Finally, the use of time series models allows us to predict congestion levels to propose optimized departure times to avoid urban congestion. The proposed solution has the potential to integrate a set of heterogeneous data in congestion management for harmonious urban cities.

## *Dedicated to My Precious Family*

*My husband (Merouane Mokhfi) & My daughters (Lilia & Maria)*

*My Father (Abderrahim Berkane) & My Mother (Horia Ait Seguer)*

*My Brothers (Nabil, Sofiane, Adel, Lyes, Massinissa)*

*My Sisters (Lynda & Amel)*

*My dear person*

*Yassamina @ Versailles, France*

*December 20th, 2023*



## Acknowledgments

I extend my heartfelt gratitude to the individuals and institutions who have played a pivotal role in the completion of my thesis. This scientific journey has been enriched by the support and contributions of a diverse community.

I am deeply thankful to Paris-Saclay University for providing an enriching academic environment and essential resources that have been instrumental in the development of my research.

Special appreciation goes to the members of Doctoral School STIC for their continuous support and guidance, creating an environment that fosters academic development.

The collaborative and stimulating atmosphere of both the DAVID and PRINTEMPS laboratories has significantly enhanced the quality of my research. My sincere thanks to the entire team for their contributions.

I express my gratitude to my supervisors, Dr. Leïla Kloul and Dr. Yoann Demoli, for their constant support, mentorship, and insightful guidance throughout this academic journey.

The financial support from CNRS (Centre National de la Recherche Scientifique) has been a fundamental element of my research project, and I am grateful for their investment in my scholarly pursuits.

To the members of the thesis defense jury: Dominique Barth, Vincent Kaufmann, Jakob Puchinger, Frédéric Chaillan, and Kamila Tabaka Simon. I extend my appreciation for their time, expertise, and constructive feedback.

Special thanks to Professeur Vincent Kaufmann and Professeur Jakob Puchinger for their meticulous review of my dissertation, which significantly contributed to its refinement.

The collaborative spirit within the ALMOST team has been a source of inspiration and encouragement, and I am grateful for the support of my colleagues.

I also acknowledge the Grand Paris Sud Agglomeration for their partnership in our questionnaire initiative. The collaboration and the traffic data provided through this partnership have been invaluable to the success of my research.

To my friends and family members in both France and Algeria, I extend my deepest thanks for their unwavering encouragement, understanding, and emotional support.

To my husband and daughters, your patience, love, and encouragement sustained me through the challenges of this academic pursuit, and for that, I am truly grateful.

Lastly, to my mother & father, to my brothers & sisters, your constant support, guidance, and belief in my abilities have been a driving force behind my success.

This thesis stands as a testament to the collaborative efforts and support of these exceptional individuals and institutions. Each one has left an enduring mark on my academic and personal journey.



# Contents

<b>Acronyms</b>	<b>12</b>
<b>List of Figures</b>	<b>14</b>
<b>List of Tables</b>	<b>15</b>
<b>1 Introduction</b>	<b>17</b>
1.1 Urban mobility congestion	18
1.1.1 Measurements of urban mobility congestion	18
1.1.2 Causes of urban mobility congestion	19
1.2 <i>Bureaux des Temps</i>	20
1.2.1 Missions of <i>Bureaux des Temps</i>	21
1.2.2 Difficulties of <i>Bureaux des Temps</i>	22
1.2.3 The Rennes example	22
1.3 Thesis objectives and dissertation structure	23
<b>2 State of art</b>	<b>27</b>
2.1 Connected and autonomous vehicles	27
2.2 Car sharing	28
2.3 Autonomous mobility on demand	29
2.4 Modifying travel demand through schedules actions	30
2.4.1 Telecommuting	30
2.4.2 Shifting schedules to smooth congestion peaks	31
2.5 Congestion pricing	34
2.6 Flexible working	35
2.7 Data-driven traffic management	36
2.8 Conclusion	36
<b>3 Data analysis</b>	<b>39</b>
3.1 Introduction	39
3.2 Steps in data analysis	40
3.2.1 Data collection	40
3.2.2 Data cleaning and preprocessing	40
3.2.3 Data interpretation	41
3.3 Data analysis techniques	43
3.3.1 Artificial intelligence	43
3.3.2 Machine learning	44
3.3.3 Supervised learning models	46
3.3.4 Unsupervised learning models	50



3.3.5	Semi-supervised learning and reinforcement learning . . . . .	53
3.3.6	Time series forecasting . . . . .	53
3.4	Conclusion . . . . .	54
<b>4</b>	<b>The decision platform</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Input data . . . . .	57
4.2.1	The questionnaire . . . . .	59
4.2.2	Open Street Map . . . . .	62
4.2.3	Waze application programming interface database . . . . .	63
4.3	Data classification and quantification . . . . .	64
4.3.1	Data classification . . . . .	64
4.3.2	Data quantification . . . . .	65
4.4	Data analysis and processing . . . . .	65
4.5	Decision making . . . . .	66
4.6	User interface of the decision platform . . . . .	67
4.7	Conclusion . . . . .	68
<b>5</b>	<b>Mobility profiles identification and prediction</b>	<b>71</b>
5.1	Introduction . . . . .	71
5.2	The methodology . . . . .	71
5.3	Criteria for mobility profiles identification . . . . .	72
5.4	Mobility profiles prediction using supervised learning models . . . . .	77
5.4.1	Multi-layer perception neural artificial network model . . . . .	77
5.4.2	K-nearest neighbors model . . . . .	78
5.4.3	Comparison between the MLPANN and KNN methods . . . . .	78
5.5	Mobility profiles clustering using K-means model . . . . .	81
5.5.1	Optimal number of clusters . . . . .	81
5.6	Conclusion . . . . .	86
<b>6</b>	<b>Sociological interpretations</b>	<b>89</b>
6.1	Introduction . . . . .	89
6.2	Data and census . . . . .	89
6.3	Supervised techniques results . . . . .	91
6.3.1	The gender impact . . . . .	92
6.3.2	Impact of the marital status and the presence of children . . . . .	94
6.3.3	Impact of the socio-professional category (SPC) . . . . .	96
6.4	Unsupervised technique results . . . . .	99
6.5	Supervised techniques vs unsupervised approach . . . . .	100
6.6	Conclusion . . . . .	102

<b>7</b>	<b>Optimal traffic-free departure forecast</b>	<b>103</b>
7.1	Introduction . . . . .	103
7.2	The methodology . . . . .	103
7.3	Experimental results . . . . .	107
7.3.1	Waze data . . . . .	108
7.3.2	Departure and arrival points . . . . .	108
7.3.3	Routes' name generation . . . . .	109
7.3.4	Traffic jams analysis . . . . .	109
7.3.5	Congestion level calculation and prediction . . . . .	113
7.3.6	Best departure time extraction . . . . .	113
7.4	Conclusion . . . . .	115
<b>8</b>	<b>Conclusion</b>	<b>117</b>
<b>A</b>	<b>Appendix A</b>	<b>133</b>
A.1	Before the health crisis: Your mobility and organization of working hours . . . . .	133
A.2	Your trips since the start of the health crisis . . . . .	136
A.3	Your work organization since the health crisis . . . . .	137
A.4	General information . . . . .	139
<b>B</b>	<b>Appendix B</b>	<b>141</b>
<b>C</b>	<b>Appendix C</b>	<b>143</b>
<b>D</b>	<b>Appendix D</b>	<b>147</b>



## Acronyms

<b>AI</b>	<b>A</b> rtificial <b>I</b> ntelligence
<b>ACC</b>	<b>A</b> daptive <b>C</b> ruise <b>C</b> ontrol
<b>ADAS</b>	<b>A</b> dvanced <b>D</b> river <b>A</b> ssistance <b>S</b> ystem
<b>ADEME</b>	<b>A</b> gence <b>D</b> e l' <b>E</b> nvironnement et de la <b>M</b> aîtrise de l' <b>E</b> nergie
<b>API</b>	<b>A</b> pplication <b>P</b> rogramming <b>I</b> nterface
<b>ARIMA</b>	<b>A</b> uto <b>R</b> egressive <b>I</b> ntegrated <b>M</b> oving <b>A</b> verage
<b>AV</b>	<b>A</b> utonomous <b>V</b> ehicles
<b>CATRAL</b>	<b>C</b> omité pour l'étude et l' <b>A</b> ménagement des horaires de <b>T</b> RAvail et des temps de <b>L</b> oisirs
<b>CNAT</b>	<b>C</b> omité <b>N</b> ational pour l' <b>A</b> ménagement des horaires de <b>T</b> ravail
<b>DBSCAN</b>	<b>D</b> ensity- <b>B</b> ased <b>S</b> patial <b>C</b> lustering of <b>A</b> pplications with <b>N</b> oise
<b>EDA</b>	<b>E</b> xploratory <b>D</b> ata <b>A</b> nalysis
<b>FB</b>	<b>F</b> ace <b>B</b> ook
<b>F-F</b>	<b>F</b> avorable <b>F</b> lexible
<b>FN</b>	<b>F</b> alse <b>N</b> egative
<b>F-NF</b>	<b>F</b> avorable <b>N</b> ot <b>F</b> lexible
<b>FODT</b>	<b>F</b> inding <b>O</b> ptimal <b>D</b> eparture <b>T</b> ime
<b>FP</b>	<b>F</b> alse <b>P</b> ositive
<b>FW</b>	<b>F</b> lexible <b>W</b> orking
<b>GDP</b>	<b>G</b> ross <b>D</b> omestic <b>P</b> roduct
<b>GIS</b>	<b>G</b> eographic <b>I</b> nformation <b>S</b> ystems
<b>GPS</b>	<b>G</b> lobal <b>P</b> ositioning <b>S</b> ystem
<b>ICT</b>	<b>I</b> nformation and <b>C</b> ommunication <b>T</b> echnologies
<b>INSEE</b>	<b>I</b> nstitut <b>N</b> ational de la <b>S</b> tatistique et des <b>E</b> tudes <b>E</b> conomiques
<b>KBS</b>	<b>K</b> nowledge- <b>B</b> ased <b>S</b> ystems
<b>KNN</b>	<b>K</b> -Nearest <b>N</b> eighbors
<b>MLPANN</b>	<b>M</b> ulti <b>L</b> ayer <b>P</b> erceptron <b>A</b> rtificial <b>N</b> eural <b>N</b> etwork

<b>OSM</b>	<b>O</b> pen <b>S</b> treet <b>M</b> ap
<b>OSRM</b>	<b>O</b> pen <b>S</b> ource <b>R</b> outing <b>M</b> achine
<b>RER</b>	<b>R</b> éseau <b>E</b> xpress <b>R</b> égional
<b>SHS</b>	<b>S</b> ocial and <b>H</b> uman <b>S</b> ciences
<b>SPC</b>	<b>S</b> ocio- <b>P</b> rofessional <b>C</b> ategory
<b>SSH</b>	<b>S</b> ocial <b>S</b> ciences and <b>H</b> umanities
<b>TN</b>	<b>T</b> rue <b>N</b> egative
<b>TP</b>	<b>T</b> rue <b>P</b> ositive
<b>TTI</b>	<b>T</b> ravel <b>T</b> ime <b>I</b> ndex
<b>UNF</b>	<b>U</b> nfavorable
<b>UTC</b>	<b>U</b> niversal <b>C</b> oordinated <b>T</b> ime
<b>VMT</b>	<b>V</b> ehicle <b>M</b> iles <b>T</b> raveled

## List of Figures

1.1	Thesis project phases. . . . .	24
2.1	Effect of schedule staggering on metro train load in 2010. . . . .	33
3.1	Various types of artificial intelligence. . . . .	43
3.2	Taxonomy of machine learning techniques. . . . .	45
3.3	An MLPANN with one hidden layer. . . . .	47
3.4	K-nearest neighbors technique. . . . .	49
4.1	Decision platform flowchart. . . . .	58
4.2	Grand Paris Sud territory. . . . .	60
4.3	Municipalities of surveyed population residency. . . . .	62
4.4	Decision-making procedure. . . . .	67
5.1	The 3-steps of the mobility profile identification methodology. . . . .	72
5.2	Predefined profile characterization . . . . .	74
5.3	Error value as a function of the parameter K value. . . . .	79
5.4	Confusion matrices for KNN and MLPANN. . . . .	80
5.5	Supervised learning techniques results. . . . .	81
5.6	Algorithm K-means steps. . . . .	82
5.7	Elbow function results. . . . .	83
5.8	Silhouette score results. . . . .	83
5.9	K-means results using different distance metrics. . . . .	85
5.10	Comparative graphs of distortion, running time, and number of iterations for the different metrics used. . . . .	85
6.1	Representation of the surveyed persons within the real population. . . . .	92
6.2	MLPANN results according to marital status and dependent children . . . . .	94
6.3	KNN results according to marital status and dependent children . . . . .	95
6.4	Social hierarchy and regularity of work schedules . . . . .	97
6.5	MLPANN results according to the socio-professional category . . . . .	98
6.6	KNN results according to the socio-professional category . . . . .	98
6.7	Cluster's repartition according to K-means compared to KNN and MPLANN classification. . . . .	101
7.1	Methodology scheme. . . . .	104
7.2	Describing scheme of routes' time spending. . . . .	106
7.3	The network structure of Évry-Courcouronnes. . . . .	108
7.4	Departure and arrival points. . . . .	109
7.5	Rue Léopold forecast components. . . . .	110
7.6	Boulevard des Champs-Élysées forecast components. . . . .	111

7.7 Rue Bois Guillaume forecast components. . . . .	112
C.1 The descriptive welcoming window. . . . .	143
C.2 The QR code of the questionnaire window. . . . .	144
C.3 Mobility profile identification processing result. . . . .	144
C.4 GPS coordinates with preferred departure time window. . . . .	145
C.5 Trajectory and temporality scenario suggestion window. . . . .	145

## List of Tables

3.1	Example of One-hot encoding . . . . .	41
4.1	Waze data attributes . . . . .	63
5.1	Parameters of the MLPANN . . . . .	78
5.2	Comparison between KNN and ANN profiles . . . . .	79
5.3	Evaluation metrics for MLPANN and KNN . . . . .	80
6.1	Census of the surveyed population . . . . .	91
6.2	MLPANN and KNN results according to gender . . . . .	93
6.3	Clusters' repartition according to sociological criteria. . . . .	100
7.1	The congestion level predicted values for the <i>3rd</i> of September, 2022 . . . . .	114
B.1	Table showing the representative proportionality of the used sample. . . . .	141





# 1 - Introduction

In the modern world, urbanization and population growth have led to remarkable challenges in managing cities and their resources. Among these challenges, urban mobility is a significant concern, as it directly impacts urban areas' quality of life, economic productivity, and environmental sustainability. The efficient and sustainable movement of people and goods within cities is crucial for enabling a thriving urban ecosystem. Cities, however, are confronted with increasing congestion difficulties that call for creative and innovative solutions as a result of the growing demand for transportation infrastructure and the rise in vehicle traffic. There are more cars on the road as cities expand and the number of people rises, which causes traffic congestion and extends travel times. Besides the frustration generated to commuters, these have severe economic, environmental, and social implications. Congestion costs in Europe were estimated at over 200 billion euros in 2016, equivalent to 1.4% of the region's Gross Domestic Product (GDP) [33]. In the United States, such costs reached US \$160 billion in 2015 in wasted time and fuel [112] and are expected to grow steadily to US \$186.2 billion by 2030 [97]. Of the 10% of CO that is man-made, about 70% is attributed to vehicles [87]. Mitigating congestion is, therefore, one of the main goals for sustainable urban mobility.

Urban mobility, at the heart of a city's functionality, requires comprehensive planning and strategic interventions to ensure smooth and efficient transportation for its residents and visitors. As urban areas become dynamic centers of economic, cultural, and social activity, the temporal aspects of mobility have gained importance. The traditional nine-to-five work schedules, rush-hour commutes, and predictable activity patterns have become less applicable to the current urban landscape.

Several European urban areas have entrusted the time bureaus, namely *Bureaux des Temps* in France, with the mission of addressing the causes of mobility congestion during specific hours. This mission involves a strategic reevaluation of the timing of various activities, including daily routines such as work, study, commercial operations, and leisure pursuits. The underlying assumption is that by strategically reorganizing the timing of these activities, it is possible to alleviate the pressure on transportation infrastructure during peak hours and enhance the overall efficiency and quality of urban mobility.

The relationship between lifestyles and mobility is influenced by a variety of factors, including culture, social class, and education. The dynamics of mobility are unfolded through an interdisciplinary perspective that goes beyond traditional physical movements. These dynamics include virtual dimensions and social aspects. Mobility involves a spectrum of movements, incorporating

virtual elements facilitated by information technologies, shifts in social positions, economic flows, and various other forms of movement that collectively mold the modern daily life [148].

The phenomenon of social acceleration in late modernity is intricately linked to the challenges of urban mobility. As societal rhythms accelerate, individuals find themselves caught in a fast-paced lifestyle that often translates into the need to address the pressures of social acceleration, seeking more efficient ways to flow, and offering individuals the opportunity to allocate more time to meaningful activities [46].

Social times, defined by cultural norms, work schedules, and family practices, structure the perception of time in society. The influence of the past, through primary and secondary socialization, is evident in mobility choices [36]. Thus, an understanding of social mobility dynamics requires an approach that integrates various factors while acknowledging the limitations of available data and research constraints.

In the following sections, we will explore the concept of urban mobility congestion, delve into its underlying factors, discuss methods for its assessment, and subsequently shift our focus to the *Bureaux des Temps* structure, for which we aim to create a decision support solution.

## **1.1 . Urban mobility congestion**

Urban mobility congestion refers to the situation in urban areas where the movement of people and vehicles experiences delays due to excess demand for transportation services, leading to traffic congestion and reduced efficiency in transportation networks. Measuring urban mobility congestion is a critical aspect for understanding its impact and proposing effective strategies for congestion management. By analyzing travel times, vehicle speeds, and other congestion indicators, cities can identify congested areas, assess the effectiveness of congestion reduction measures, and make decisions to improve urban mobility [137].

### **1.1.1 . Measurements of urban mobility congestion**

Various key indicators are used to assess congestion levels, including travel time, vehicle speed, and traffic volume.

The Travel Time Index (TTI) is a commonly used metric for evaluating urban mobility congestion. It compares the estimated travel time during peak hours and the free-flow travel time. When TTI values are greater than one, they indicate increased travel times during peak periods, signifying congestion. For instance, a TTI of 1.2 would mean that during peak hours, a trip takes 20% longer than the same trip under free-flowing conditions. TTI is an interesting measure for quantifying the impact of congestion on travel times

and shows the efficiency of transportation networks [137].

Average vehicle speed during peak hours is another critical measure used to assess congestion levels. High congestion levels result in lower average speeds as traffic flow becomes slow and less fluid. Reduced vehicle speeds not only lead to longer travel times but also decrease the capacity of transportation systems to handle traffic demand [96].

To measure urban mobility congestion, researchers and transportation authorities often use data from traffic monitoring systems (such as cameras and sensors embedded in roadways), Global Positioning System (GPS) tracking devices (such as GPS devices in vehicles and smartphones), and mobile applications (navigation applications such as Google Maps and Waze). These data sources provide information on travel times and vehicle speeds, enabling the assessment of congestion levels across different times of the day and specific road segments [103].

### **1.1.2 . Causes of urban mobility congestion**

Urban mobility congestion is a multifaceted problem caused by a variety of factors. This section elaborates on the causes of urban mobility congestion and its implications.

Population growth and urbanization are fundamental causes of urban mobility congestion. As cities experience an increasing number of residents seeking better economic opportunities and improved lifestyles, the demand for transportation services increases. This huge demand can overwhelm existing transportation infrastructure, resulting in traffic congestion and longer travel times [89].

The effectiveness and capacity of transportation infrastructure significantly impact urban mobility congestion. Insufficient infrastructure can create narrow streets, limiting the flow of traffic and increasing congestion. Inadequate public transportation options may lead to more people opting for private cars. Poorly designed road networks may lack efficient traffic flow management, leading also to traffic jams and delays [149].

Land use patterns are critical in shaping transportation behavior and congestion levels. Decentralized land use can increase travel distances between residential areas, workplaces, and commercial centers. The increased dependence on private vehicles for daily commuting in such dispersed urban areas causes additional congestion problems [117].

The rise of e-commerce has introduced new challenges for urban mobility congestion. With the growing popularity of online shopping, there has been a surge in package delivery services, leading to increased commercial vehicle movements in urban areas [144]. Similarly, the rise of on-demand transportation services, such as ride-hailing and food delivery, has increased the number of vehicles on the road, especially in densely populated urban centers.

Concentrating on these activities during peak hours can further worsen congestion, making it necessary to explore strategies to reduce traffic congestion.

Clearly, urban mobility congestion arises from a combination of factors, including population growth, urbanization, land use patterns, transportation infrastructure, and e-commerce. Understanding the complex link between these factors is essential for proposing effective congestion management strategies. By addressing the causes of congestion, cities can create more efficient and sustainable urban mobility systems, leading to improved transportation systems and enhanced quality of life for their residents.

## **1.2 . Bureaux des Temps**

Work and non-work times are two strongly interconnected aspects of life, and this recognition has led to significant legislative efforts. The law Aubry II [72], enacted in 2002 to reduce working hours, played a pivotal role in this context and incorporated a provision to address the concept of city time.

### ***Law Aubry II.***

#### ***No. 2000-37 of January 19, 2000, regarding the negotiated reduction of working hours Article 1, Paragraph 7.***

Dans les agglomérations de plus de 50 000 habitants, le président de la structure intercommunale, en liaison, le cas échéant, avec les maires des communes limitrophes, favorise l'harmonisation des horaires des services publics avec les besoins découlant, notamment du point de vue de la conciliation entre vie professionnelle et vie familiale, de l'évolution de l'organisation du travail dans les activités implantées sur le territoire de la commune ou à proximité.

A cet effet, il réunit, en tant que de besoin, les représentants des organismes ou collectivités gestionnaires des services concernés et les met, le cas échéant, en relation avec les partenaires sociaux des entreprises et des collectivités afin de promouvoir la connaissance des besoins et de faciliter la recherche d'adaptation locale propre à les satisfaire.

The concept of *Bureaux des Temps* was born in Italy in the mid-80s in their theoretical principles and from the 90s in their concrete developments [25] and was later adopted by Germany and France. These bureaux are structures, either municipal or closely linked to the actions of the local community, that aim to consider various time dimensions within a territory while also listening to and involving the residents and users of the area in problem-solving and decision-making processes [26].

The experience of these bureaux in Italian municipalities served as inspiration for a similar approach in some French cities (Rennes, Lille, Lyon, Paris,... etc.). Rennes, for instance, innovated by integrating the concept of

city time into the issue of gender equality. In 1988, a popular initiative called "Les femmes changent les temps: une loi pour humaniser le temps de travail, les horaires dans la ville et le rythme de la vie" (women change times: a law to humanize working hours, schedules in the city, and life's rhythms) emerged, which significantly influenced an Italian law in 1990 granting municipalities the power to coordinate business and public service hours to harmonize activities and meet the needs of users [104].

As these bureaux expanded in Italy, their focus shifted to address the "double burden" carried by women, who simultaneously engage in paid work and domestic responsibilities, making them the primary victims of temporal inequalities. The aim was to enhance women's integration into urban life by acknowledging and addressing these temporal inequalities [34].

A study conducted in France in 1998/1999 examining the impact of household types on the daily lives of active women and men highlighted that when living alone, women and men have relatively comparable daily routines, with women dedicating about 40 minutes more to domestic tasks than men. However, when couples form, women's domestic time increases by 39 minutes, while men's decreases by 21 minutes. Additionally, women's professional time decreases by half an hour, while men's increases by half an hour. The arrival of children further amplifies these gaps, with women devoting more time to domestic duties and less to their professional lives, especially with the third child. Meanwhile, men's routines are minimally disturbed by the arrival of children, with their professional time even increasing with the second child [84].

The politics of urban time management are closely linked to the need to reconsider collective functioning in the context of flexible schedules and resynchronize the timings of businesses, administrative services, transportation and communication, health and educational services, leisure activities, etc. [80]. However, time management inevitably leads to stress, temporal pressure, excessive medication consumption (as evident in studies on female workers), family tensions, and fatigue. The issue is not about women's work but rather its nature and conditions [63].

Overall, the *Bureaux des Temps* represent an effort to shape more equitable and harmonious urban environments by addressing temporal inequalities and their impact on various aspects of life, particularly concerning gender roles and responsibilities.

### **1.2.1 . Missions of *Bureaux des Temps***

The mission of the *Bureaux des Temps* is to analyze the needs of citizens regarding the schedules of public services such as nurseries, educational institutions, municipal services, and transportation, in order to implement policies that facilitate equal access for all citizens to the city [104]. These *Bureaux des*

*Temps* raise the issue that citizenship is no longer solely determined by the relationship with the State but must find expression within the city's territory, which is perceived as a common good with equitable access [104]. Temporal policies have been developed with the objectives of combating all forms of discrimination, establishing proximity services across all sectors, and improving the quality of life in the agglomeration. To achieve these objectives, the *Bureaux des Temps* have been assigned the task of conducting studies and surveys to understand the lifestyles, needs, and time constraints of city users, coordinating the actors of the agglomeration to develop solutions, and collaborating with elected officials, services, and organizations from various sectors to address time-related issues [84].

### **1.2.2 . Difficulties of *Bureaux des Temps***

The *Bureaux des Temps* face several challenges, starting with their fragile organizational structures. On one hand, these bureaus suffer from inadequate funding, as their budgets are perceived as minimal by those involved. On the other hand, the limited number of persons working in these bureaus (often just one or two individuals) and their unique position within local authorities set them apart from other services. Moreover, there is a lack of significant connections with urban planners. Few urban planners are involved in the *Bureaux des Temps*, and the training of urban planners rarely addresses temporal issues. This disconnection leads to paradoxes and contradictions in practice aligned with the *Bureaux des Temps*' goals. For instance, despite acknowledging the undesirability of continuous urban development operating 24/7, many actions taken by temporal policies promote extended nightlife activities. However, these activities are mostly daytime-oriented, such as going to libraries, swimming pools, or museums. The limited resources and little credit given to the *Bureaux des Temps* by urban planners severely restrict their ability to participate in large-scale projects at the agglomeration or neighborhood level [105].

### **1.2.3 . The Rennes example**

In the city of Rennes, a notable action taken by a municipal employer was aimed at reducing atypical and fragmented working hours for cleaning staff, a profession dominated by women (90%) [84]. The existing working conditions presented multiple challenges, with only a quarter of them working full-time, half of them holding permanent positions, and dealing with scattered schedules (7 : 00 AM - 9 : 00 AM and 4 : 00 PM - 7 : 30 PM) across 47 municipal sites, resulting in high absenteeism. After a year of focused efforts, several reforms were implemented: new daily schedules (7 : 30 AM - 3 : 30 PM with lunch break from 11 : 00 AM to 12 : 00 PM or 10 : 45 AM - 6 : 45 PM with lunch break from 1 : 45 PM to 2 : 45 PM), improved working conditions through team-based collaboration and geographic site consolidation to reduce travel

time, and the establishment of full-time positions with career advancement opportunities.

The outcomes of these reforms were highly positive, with a satisfaction rate of over 70% among the staff. The cleaning staff appreciated their new schedules, the full-time positions, and the improved work environment. Moreover, the institution experienced a significant decrease in absenteeism, reducing it by 40%.

In addressing childcare needs, the city also diversified its services to meet various demands, supporting childcare with atypical hours and emergency care. Services included home-based childcare in addition to traditional facilities. An intercommunal and inter-company daycare center was established in a business area, operating from 6 : 00 AM to 9 : 30 PM, with a limit of 10 hours of care per child per day. As part of the Local Educational Plan, children could receive care between 7 : 30 AM and 6 : 30 PM. Additionally, 23 leisure centers were opened during school holidays for children aged 2 to 12, with operating hours from 7 : 30 AM to 6 : 45 PM for ages 2 to 6 and from 8 : 00 AM to 6 : 00 PM for ages 6 to 12.

The city of Rennes also focused on facilitating access to leisure activities for active women who often face time constraints. Initiatives such as "Happy hours" at the opera, offering shows at 6 : 00 PM and 8 : 00 PM with reduced ticket prices and childcare services, and "exceptional" lunchtime concerts with classical music and a meal (sandwich and drink). Additionally, "midis musées" (museum lunch breaks) from 12 : 30 PM to 1 : 15 PM became popular, with 87% attendance by women.

### **1.3 . Thesis objectives and dissertation structure**

The primary goal of our doctoral project is to propose a refined platform based on an interdisciplinary methodology for analyzing heterogeneous data to reduce traffic congestion. We recognize that urban mobility is a multi-face phenomenon influenced by many social, economic, and environmental factors. As such, collaborating with different disciplines is essential. Using Information and Communication Technologies (ICT) and incorporating key concepts from the Social and Human Sciences (SHS), our proposed platform aims to serve as a decision support solution for the *Bureaux des Temps*, which plays a pivotal role in addressing a fundamental challenge that future cities must confront—the urgent necessity to decongest transportation means and infrastructure. In response to this challenge, our research shifts the focus from spatial considerations to the dimension of time, re-imagining its role in shaping the activities and mobility patterns of urban territories.

The Social and Human Sciences (SHS) form a critical component of this interdisciplinary approach. These sciences encompass various disciplines, from



geography to psychology, sociology, and more. Sociology, the field of SHS used in our thesis, provides valuable insights into the interactions between individuals, society, and the environment, significantly shaping mobility behavior [32].

Moreover, digital disciplines are instrumental in exploiting the potential of diverse datasets generated in modern urban environments. Integrating digital innovation with sociology enables us to understand mobility patterns while developing strategies for decongesting urban mobility.

Our thesis project unfolds in three distinct phases (see Figure 1.1): mobility profile identification and prediction, traffic jams analysis, and time/space decision.

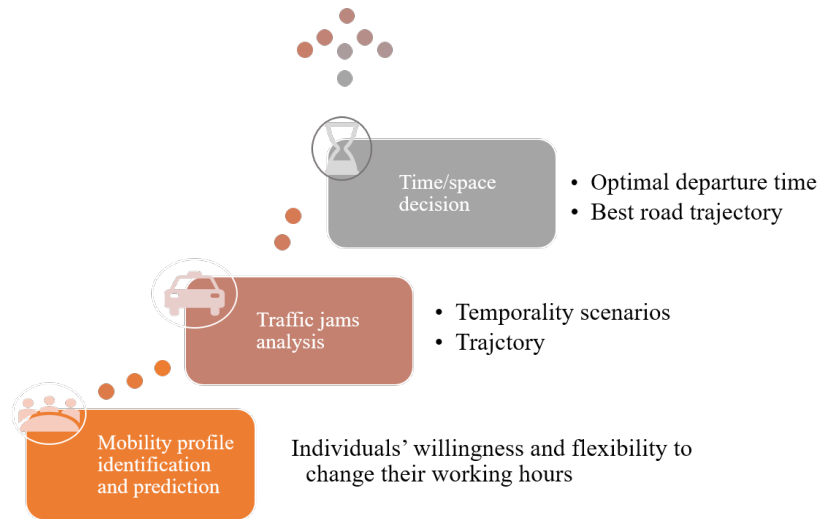


Figure 1.1: Thesis project phases.

- **Mobility profile identification and prediction:** The initial phase of our research work centers around the identification and prediction of mobility profiles. These profiles are meticulously defined and validated from a sociological perspective. Upon completion of this phase, a mobility profile about the individual willingness and flexibility to change her/his working hours to reduce traffic congestion is assigned.
- **Traffic jams analysis:** In the second phase, the examination of traffic congestion is made. The outcome of this phase provides us with an optimal trajectory in terms of distance and an array of temporality scenarios. The optimal trajectory encompasses a collection of routes that guarantee minimal traveled distance, while the temporality scenarios encompass departure times that strategically avoid peak congestion periods, ranked from the most to the least effective.

- Time decision: The third and final phase consists of extracting the most advantageous departure time. This extraction process is implanted in the ordered list from the previous phase.

Collectively, these three phases constitute a comprehensive investigation into mobility profiles, traffic congestion analysis, and the extraction of optimal departure times and routes. The thesis dissertation is thus structured as follows:

Chapter 2 (State of art) discusses various existing solutions for reducing congestion, including integrating autonomous vehicles, car sharing, connected vehicles, and autonomous mobility on demand (AMOD) systems. It also addresses the potential effectiveness of congestion pricing and different flexible working techniques.

Chapter 3 (Data analysis) overviews techniques for understanding behaviors from collected data, including supervised learning techniques with neural networks and K-nearest neighbors, unsupervised learning with k-means clustering, and time series modeling using Facebook Prophet. Evaluation metrics and their relevance to the research objectives are discussed.

Chapter 4 (The decision platform) presents the general architecture of the decision platform we propose to tackle urban mobility congestion through an integrated system. It utilizes data analytics to analyze diverse data sources and facilitate decision-making. The platform consists of five modules: input data, data classification and quantification, data analysis and processing, decision-making, and output data. By considering sociological factors, it identifies optimal temporality scenarios for congestion reduction, improving urban traffic management with personalized recommendations.

Chapter 5 (Mobility profiles identification and prediction) details a methodology to address urban traffic congestion by predicting individuals' mobility profiles using machine learning and sociological criteria [20] [21]. The defined profiles allow personalized recommendations for optimal work schedules and improved time management.

Chapter 6 (Sociological interpretations) delves into sociological interpretations of the defined mobility profiles and their influence on individuals' willingness to adjust working hours. Using machine learning, we analyze the impact of gender, marital status, dependent children, and socio-professional categories on people's willingness and flexibility to adapt their working hours to reduce congestion.

Chapter 7 (Optimal traffic-free departure forecast) presents a traffic analysis approach to reduce urban traffic congestion during rush hours. It utilizes route data from the Open Street Map, space/time traffic jam data obtained through a collaboration between Waze-connected citizens, the Grand Paris Sud urban community, and Paris Saclay University, and time series models for congestion prediction. The approach generates temporality scenarios to

identify the best departure time for minimizing traffic [22]. An illustrative example using Waze data for Évry-courcouronnes, a Paris suburb area, demonstrates the methodology's effectiveness.

Finally, Chapter 8 (Conclusion) is dedicated to conclusions and perspectives for the evolution of the proposed approach.

## 2 - State of art

Traffic congestion remains a critical challenge in urban centers worldwide, causing significant economic losses, environmental impacts, and reduced quality of life for residents. Over the years, researchers and transportation experts have been diligently working to find innovative solutions to alleviate congestion and improve urban mobility. As we enter the digital age, advancements in technology, data-driven approaches, and the emergence of intelligent transportation systems have opened up new possibilities for addressing traffic congestion. In this chapter, we explore various existing strategies to alleviate urban traffic congestion, including connected and autonomous vehicles, car sharing, autonomous mobility on demand, temporal policies, congestion pricing, flexible working arrangements, and data-driven traffic management.

### 2.1 . Connected and autonomous vehicles

Integrating autonomous vehicles (AVs) holds great promise as a solution to alleviate traffic congestion. Studies have shown that AVs can operate at reduced headways on narrower lanes, effectively increasing road capacity compared to conventional vehicles with similar capacities [152]. This novel feature enables AVs to optimize their movement, leading to smoother traffic flow and reduced congestion [100]. Furthermore, AVs can have a substantial indirect impact on congestion by reducing the need for parking spaces in dense urban areas. After dropping off passengers, AVs can return home or relocate to less congested areas, reducing the overall number of cars parked near workplaces or in high-traffic zones [110]. This aspect of AVs can significantly contribute to alleviating congestion in urban centers [100].

However, it is crucial to consider potential challenges that may arise from the widespread adoption of AVs. One concern is that the availability of driverless vehicles could lead to increased car use among individuals who are not able or qualified to drive, such as children, the elderly, and people with disabilities [110]. It is essential to carefully manage policies and regulations surrounding AV adoption to ensure that congestion reduction benefits are not compromised by unintended consequences [140]. Implementing supportive policies and considering lessons learned from early AV deployments can maximize the positive impact of autonomous vehicles on congestion mitigation.

Another solution cited in the literature is the connected vehicles, which encompass a range of applications, services, and technologies designed to establish links between a vehicle and its surroundings [146]. They represent a transformative shift in the automotive industry.

One of the prominent features of connected vehicles is adaptive cruise control (ACC), which utilizes sensor data, including radar, lidar, and cameras, to automatically adjust the vehicle's speed and following distance in response to traffic conditions [85]. ACC enhances safety and traffic efficiency by continuously monitoring nearby vehicle positions and velocities, regulating acceleration and deceleration, and reducing accident risks [154]. Its integration into connected vehicles opens avenues for cooperative driving and vehicle-to-vehicle (V2V) communication, which enable seamless interactions between vehicles and the surrounding infrastructure [90]. V2V communication allows vehicles to exchange real-time information, encompassing position, speed, and acceleration, with nearby vehicles. This information exchange authorizes vehicles equipped with ACC to anticipate other vehicles' behavior on the road, make proactive decisions, enhance traffic flow, and reduce congestion [150]. Furthermore, V2V communication enables collective awareness of potential hazards or road obstacles, expediting response times and elevating overall road safety, thereby mitigating congestion [7].

The integration of connected vehicles and ACC, however, also poses some challenges. One major concern is the need for robust and reliable vehicular communication networks to support the real-time exchange of data between vehicles and infrastructure [81]. A stable and secure communication infrastructure is essential to ensure the safety and effectiveness of ACC and other connected vehicle applications.

Additionally, issues related to data privacy and cybersecurity must be carefully addressed to safeguard sensitive information shared through connected vehicle systems [7]. Ensuring data integrity and protecting against potential cyber-attacks are paramount in maintaining public trust and confidence in connected vehicle technologies.

## **2.2 . Car sharing**

Car sharing, also known as carpooling, is a term used for a business model where the users act as drivers who share their vehicles with others. This model can be operated by companies or organizations that provide a fleet of vehicles for their members to share [135], fostering a more sustainable and efficient approach to personal transportation. Car sharing has gained attention as a potential solution to address the increasing car occupancy and congestion on roads. By encouraging multiple individuals to share a single vehicle, these models have the potential to reduce the number of cars on the road, leading to decreased traffic congestion and improved traffic flow [107]. Numerous studies have explored the impact of car sharing on congestion and found positive outcomes reporting a noticeable reduction in vehicle ownership and overall miles driven [139]. Furthermore, car sharing can promote the

use of high-occupancy vehicles, which have the advantage of using carpool lanes in some areas, further reducing travel times and congestion [14].

However, the potential benefits of car sharing are not without challenges. While these models may reduce congestion, they could also attract potential road users who previously opted for other modes of transportation. For instance, car-sharing services offering lower costs and greater convenience may lure public transport passengers, leading to a shift from buses and trains to shared-use vehicles [99]. This modal shift could potentially increase the demand for cars and taxis, partially offsetting the congestion reduction benefits achieved through car sharing. Additionally, some studies have suggested that the introduction of shared-use vehicles may lead to an increase in vehicle miles traveled (VMT), especially if individuals choose to use car-sharing services for trips they would have otherwise made by walking, cycling, or using public transport [88].

### **2.3 . Autonomous mobility on demand**

Another exciting approach is the deployment of Autonomous Mobility On Demand (AMOD). The concept behind AMOD is to offer convenient and efficient mobility options, such as self-driving taxis or shared autonomous vehicles, to meet people's travel needs without the need to own a personal car [118].

However, the successful deployment of AMOD systems is not without challenges. One significant problem that arises in AMOD implementations is the imbalance in the distribution of vehicles within the transportation network. In dense urban areas, the demand for AMOD services may be concentrated in certain regions, leading to an excess of vehicles in those locations, while other areas might experience a shortage of available vehicles [155]. This imbalance can lead to longer waiting times for users in under-served areas and decreased efficiency in resource utilization. A re-balance that can lead to crossing the same or greater distances must be made [106].

To address this challenge, vehicle rebalancing is an essential aspect of AMOD operations. Rebalancing involves dynamically redistributing the fleet of autonomous vehicles to ensure a better match between demand and supply in different regions [123]. It necessitates relocating underutilized vehicles from areas with low demand to regions with high demand, thereby optimizing the efficiency and coverage of the AMOD service. Effective rebalancing strategies are crucial to maintaining user satisfaction, enhancing the overall reliability of the service, and minimizing empty vehicle miles, which can contribute to unnecessary traffic congestion and energy consumption.

## **2.4 . Modifying travel demand through schedules actions**

Temporal policies propose to spread the demand over time to limit rush hour and smooth traffic flow without increasing infrastructure. Two prominent levers emerge: reducing transportation demand by promoting telecommuting (as well as distance learning and remote services) and spreading travel demand over time to attenuate peak periods, thereby optimizing the use of transportation infrastructure.

### **2.4.1 . Telecommuting**

Telecommuting, or remote work, offers a persuasive solution to the issue of excessive mobility. Embracing this forward-looking approach provides a real opportunity for growth and development, as demonstrated by various cities like Amsterdam and its "smart work centers," London's implementation of remote work during the 2012 Olympics, and Japan's adoption of this practice during the 2020 Olympics [41].

More than just reducing the need for physical mobility, remote work caters to the desires of modern employees seeking flexibility and agility in their careers. It enables a healthier work-life balance and enhances the overall quality of life. Studies reveal that time saved from commuting leads to a remarkable increase of 60% in work productivity and 40% in improved personal well-being. Every day of telecommuting, individuals gain an extra 45 minutes of sleep and an additional 37 minutes of valuable family time [111]. Furthermore, embracing telecommuting also has positive implications for road safety. Out of all bodily accidents, 38% involve individuals commuting between their homes and workplaces or while on business-related trips [73].

However, studies have consistently shown that remote work, while offering flexibility and convenience, can lead to social disconnection, highlighting the increased feelings of loneliness among telecommuters due to reduced face-to-face interactions with colleagues. Furthermore, prolonged isolation can negatively impact mental health, resulting in increased stress, anxiety, and burnout [94].

## **Experimentation in the Lyon metropolis**

The Lyon Metropolis, in collaboration with the ADEME (Agence De l'Environnement et de la Maîtrise de l'Énergie) [71], has engaged in an innovative initiative known as "ReducMob" [69]. This experiment focused on implementing remote work in two distinct sectors: the Valley of Chemistry, an industrial sector employing 27,000 individuals, and Part-Dieu, a service sector with 50,000 jobs. The primary objective of this project was to support 40 companies in these sectors that have voluntarily chosen to participate, representing a total of 18,000 jobs.

Based on a potential study conducted among employees, the results in-

dicating that by introducing one day of telecommuting per week in both regions, an impressive annual saving of 3.6 million kilometers and 700 tons of CO<sub>2</sub> can be achieved. To accurately quantify these environmental savings, the "ReducMob" project has developed specialized software, recognizing the complexities involved in such assessments and fulfilling the need for precise data as requested by ADEME. As of now, the database includes records of 150 telecommuters, accounting for over 5,600 telecommuting days, resulting in an impressive total of 336,000 saved kilometers, out of which 140,000 kilometers were from reduced car usage, effectively preventing 29 tons of CO<sub>2</sub> emissions [41].

### **Experimentation in the city of Paris**

The city of Paris approximated that around 9,000 employees had the potential to telecommute. An initiative started in 2015 as part of the second transport plan under the Parisian administration. To begin with, a trial phase was introduced involving 150 voluntary employees and their supervisors, who were initially unwilling about their staff to engage in remote work. Effecting this change necessitated the implementation of innovative management techniques and a project-based approach. Notably, the voluntary employees had an average daily commute of 1 to 1.5 hours. Following the trial's positive feedback from both employees and supervisors, remote work was expanded to encompass all employees with teleworkable tasks in the institution, with a goal of reaching 1,500 employees by the end of 2019. By the close of 2018, 1,300 employees were telecommuting, resulting in the saving of 275,000 hours of travel and 6.8 million kilometers [41].

Despite the advantages of telecommuting in reducing rush-hour travel, certain researchers advise that further verification is needed to assess its overall impact on mobility. Giving employees the opportunity to telecommute one or two days per week could result in some "frequent travelers" choosing to reside farther from their workplace. As a result, while the frequency of home-to-work trips might decrease, their duration could increase. Additionally, on telecommuting days, some of the time saved from commuting may be utilized for other trips, such as running errands, shopping, or engaging in leisure activities. To understand telecommuting's impact on mobility, it is essential to consider and quantify these additional effects. Nevertheless, it is important to note that these effects do not negate the positive benefits of telecommuting in reducing rush-hour travel [13].

#### **2.4.2 . Shifting schedules to smooth congestion peaks**

Another approach to manage travel and reduce congestion, especially during peak hours, is to spread the demand for mobility over different times.

All major cities experience congestion during peak periods, on roads, and



on public transportation [67]. Congestion occurs when demand exceeds supply, localized in specific spaces and times, and it becomes a chronic issue in urban transportation networks. In this context, demand refers to the number of people wanting to travel, while supply denotes the capacity of transportation systems [10]. One significant factor contributing to recurrent congestion during peak hours is the synchronization of social schedules, such as work and school hours. Historically, the traditional response to this problem was to expand transportation networks and infrastructure to increase supply. However, this approach has drawbacks, as it leads to more trips and is not sustainable in the long run since congestion inevitably returns [114].

A sustainable solution that intuitively appeals to address this issue is to stagger individuals' work schedules, thereby spreading out arrivals at workplaces and easing traffic peaks. This approach would alleviate inconveniences commuters and transportation providers face, such as wasted time, stress, accidents, and pollution. Implementing such demand spreading, aligning well with the concept of temporal policies, can help optimize traffic during peak periods without requiring investments in infrastructure.

The idea of scheduling work hours to achieve traffic smoothing is not new. There were early experiments as far back as 1941 in Washington DC, when the United States Bureau of the Budget designed a staggered work schedule plan for federal employees. Since then, many federal buildings have experimented with staggered work hours. Similarly, the British Ministry of Transport conducted a similar experiment in 1959 in London, targeting private employers, though it did not yield significant results due to employers' reluctance to participate [41].

France also gave consideration to this issue through the establishment of committees such as the CNAT (Comité National pour l'Aménagement des horaires de Travail), operating between 1958 and 1961, and the CATRAL (Comité pour l'Etude et l'Aménagement des horaires de TRAvail et des temps de Loisirs), operating between 1966 and 1971 [113]. Nevertheless, these efforts to optimize schedules fell short of taking into account employees' individual time management preferences, primarily due to the absence of quantifiable data on these past instances [41].

A notable and successful example in this regard is the "Hyperpointe Métro" experiment conducted in Rennes by the Rennes Time Bureau in 2012. The project's main objective was to investigate the potential of big data in assessing the impact of transportation projects or policies, with a focus on the specific case of adjusting the schedules at the University of Rennes 2. Back in 2010, the city of Rennes and the transport provider Keolis Rennes observed that the trains experienced severe overcrowding (reaching 200% capacity at certain stations) during the morning peak hours, primarily due to the movement of students from the University of Rennes 2. The situation was particu-

larly critical at 8:00 a.m. when all students began their classes simultaneously, leading to a saturation of the transport system. To address this issue without investing in additional trains, the Bureau des Temps of Rennes proposed a solution: staggering the class starting times. They suggested having half of the students start classes at 8:30 a.m. instead of the usual 8:15 a.m. This idea was put into action as an experiment starting in 2012 and was subsequently extended due to its resounding success and positive feedback from all users [35] (see Figure 2.1).

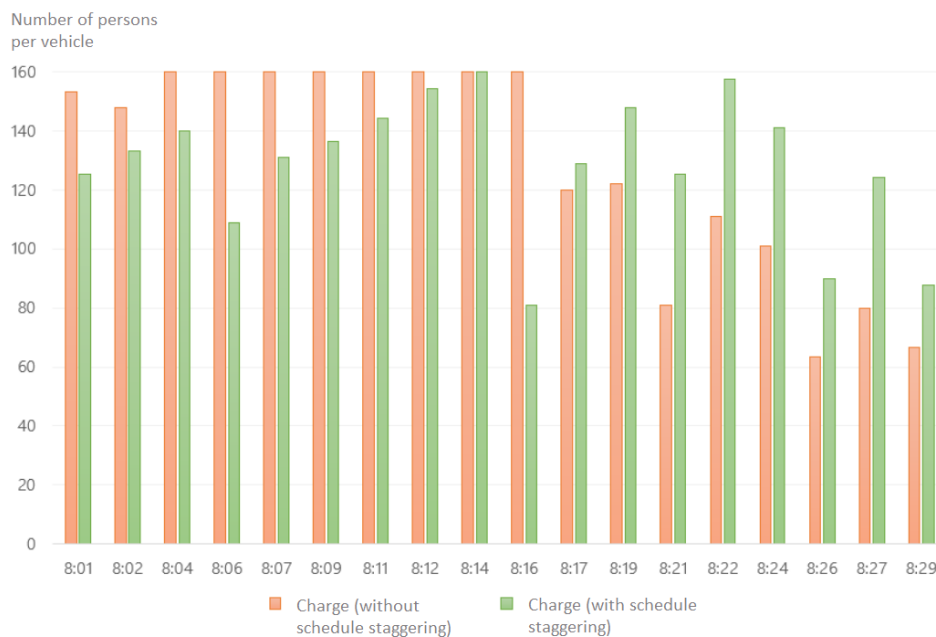


Figure 2.1: Effect of schedule staggering on metro train load in 2010.

These experiments highlight the importance of understanding users' temporal patterns in their specific area of interest. Imposing staggered hours on populations with a certain degree of freedom in their schedules is not always feasible, except in exceptional cases like factories or schools. Therefore, the focus should be on identifying the factors that influence individuals' travel schedules, which may include barriers such as synchronization constraints with spouses or colleagues, the need to accompany children to school, and societal pressures related to workplace norms. These barriers need to be adequately identified through surveys and analysis, along with assessing the potential for flexibility. Examining users' temporal patterns reveals the complexities of behavior change, which is a solution cited in the literature to reduce traffic congestion. While staggered hours face limitations, especially outside controlled environments, understanding influencing factors, like familial commitments and societal norms, helps the development of effective solutions.

## 2.5 . Congestion pricing

Congestion pricing is another potentially effective solution to reducing traffic congestion cited in the literature. It involves charging fees for driving in congested areas or during peak hours to discourage vehicle use and encourage alternative transportation modes. One notable example of successful congestion pricing is Singapore's Electronic Road Pricing (ERP) system. Singapore implemented this dynamic congestion pricing system in 1998, and it has since been recognized as a model for effective traffic management. The long-term impacts of this system have consistently reduced traffic congestion, improved travel time reliability, and enhanced overall urban mobility in Singapore [108]. The system relies on a set of sensors to charge vehicle drivers tolls based on the traffic conditions and the time of day, with higher fees during peak hours. This dynamic pricing encourages commuters to shift their travel times or choose alternative modes of transportation.

London's congestion charge is another notable example of a successful implementation of congestion pricing. In 2003, a research study was conducted about the impact of congestion charging. It has consistently demonstrated its effectiveness in reducing traffic congestion in the city center. Shortly after the introduction of the London congestion charge, a substantial reduction in traffic volumes within the charged zone, with a significant decrease in the number of vehicles and traffic delays was observed [57]. This initial success was further reaffirmed. A long-term study by Transport for London (TfL) in 2019 reaffirmed the positive impact of the congestion charge, demonstrating that it continued to be effective in reducing traffic congestion and encouraging the use of public transport and sustainable modes of transportation within the charging zone. The study also highlighted how revenue generated from the congestion charge has been reinvested into improving public transportation, cycling infrastructure, and pedestrian facilities, contributing to a more sustainable urban environment.

These examples underscore the importance of congestion pricing as an efficient policy tool for reducing congestion.

While congestion pricing has shown promising results in reducing traffic congestion in various studies, its acceptability among travelers remains a subject of investigation [19]. The example of Gothenburg, Sweden, provides one of the challenges associated with implementing congestion pricing and its acceptance by the public.

In Gothenburg, the city implemented a congestion pricing scheme by establishing 37 toll booths surrounding the city center between 2013 and 2015. The goal was to reduce traffic congestion and improve overall traffic flow. Indeed, during the initial period of implementation, traffic jams decreased by approximately 12%, indicating some success in achieving the intended objective [29]. However, despite the initial positive impact on traffic congestion,

the long-term acceptability of the congestion pricing scheme by the citizens of Gothenburg was not sustained.

After just one year of implementation, the citizens refused to pay the appropriate fines associated with congestion pricing. The opposition was largely due to perceived inequities and concerns about the financial burden on commuters. Some critics argued that the toll disproportionately affected lower-income individuals who relied heavily on private cars for their daily commute [49]. Such resistance and opposition from the public raised questions about the political and social acceptability of congestion pricing as a sustainable solution for reducing traffic congestion.

The case of Gothenburg highlights the importance of understanding public attitudes and preferences when implementing congestion pricing schemes. Effective implementation requires consideration of various factors, such as the design of the pricing scheme, the level of public engagement, and the provision of alternative transportation options [62].

None of the studies cited above consider the sociology of people who adapt their working hours to reduce congestion. Each solution proposed eventually develops a novel problem and too many challenges following its implementation.

## **2.6 . Flexible working**

Flexible Working (FW) has emerged as another promising solution in the literature to tackle traffic congestion and enhance workers' quality of life. FW refers to the practice of working without rigid constraints such as fixed working spaces, schedules, and contractual boundaries [93]. Several approaches to flexible working currently exist [93].

FW is categorized into four main categories, each offering different levels of flexibility to employees. The first category involves flexitime saving accounts or time banking, where employees can accumulate flexible hours for future use. The second category includes alternative work schedules like staggered hours arrangement and compressed workweek arrangements, enabling employees to work during less congested hours. The third category, workload flexibility, comprises job sharing, contingent work, and part-time work, offering workers more control over their schedules and reducing their commuting needs. The fourth category is the flexibility of place, involving working from home and teleworking which allows employees to work remotely, reducing the need for daily commuting [16]. Applying flexible working to workers' lifestyles may reduce congestion and consequently improve their overall quality of life.

However, it is essential to acknowledge that FW is not without potential drawbacks. The arguments against FW have mainly focused on the disproportional

tionately impacting certain groups, particularly women, who may face greater challenges in juggling work responsibilities with caregiving responsibilities [54]. Moreover, poor work-life balance can arise when employees feel the pressure to be continuously available and engaged with work due to flexible working hours [31]. Health problems, such as burnout and increased work-related stress, have also been associated with flexible working arrangements, especially when boundaries between work and personal life become blurred [17].

Additionally, the feasibility of FW depends on the nature of the work and the availability of employees. Not all jobs and industries can accommodate flexible working hours, and not all employees may have the resources or capabilities to work remotely or adjust their schedules [151]. Societal and cultural factors also play a role, as employees' willingness and ability to embrace flexible working may vary across different regions and work cultures [119].

## **2.7 . Data-driven traffic management**

Data-driven traffic management is crucial in addressing urban mobility congestion issues. An empirical evaluation compared the effectiveness of time-series analysis and supervised learning methods in predicting short-term traffic flow. The study emphasizes the significant potential of these models in forecasting traffic patterns for decongestion purposes in urban areas [98]. Moreover, a survey on data-driven intelligent transportation systems underscores the role of time series forecasting models in enhancing transportation efficiency[156]. These studies collectively demonstrate the utilization of machine learning and time series forecasting models to solve the complexities of urban mobility decongestion.

## **2.8 . Conclusion**

This chapter highlights various strategies to combat urban traffic congestion. These include connected and autonomous vehicles, car sharing, autonomous mobility on demand (AMOD), temporal policies like telecommuting and schedule shifting, congestion pricing, flexible working, and data-driven traffic management. While each solution has merits, addressing congestion requires a balanced approach, considering technological advancements, public acceptance, and societal dynamics.

It is worth noticing that the sociological aspect is often missing in the solutions cited in the literature. Our proposed work aims to fill this gap by leveraging a data-driven traffic management approach to optimize departure times while considering the sociological aspects of individuals' behavior and preferences regarding their work schedules.

The next chapter explores data analysis methods that enable the utiliza-

tion of data-driven approaches for reducing urban mobility congestion.



## 3 - Data analysis

### 3.1 . Introduction

Data analysis plays an important role in research, allowing researchers to understand patterns, trends, and relationships in data. Data analysis is the process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information and drawing conclusions [8]. By applying the right data analysis techniques, researchers can transform raw data into meaningful information, leading to valuable conclusions [127].

The applications of data analytics cover a variety of fields and industries. In business, data analytics helps businesses better understand customer preferences, market trends, and operating efficiencies. It is used to identify trends and predict future results, enabling companies to optimize their strategies and stay ahead of the competition [15][78]. In healthcare, the use of data analytics technology enables health service providers to obtain accurate, up-to-date, and complete information about patients. Benefits also include sharing electronic information with patients and health centers, effective patient diagnosis, and better interaction between patients and healthcare providers [18]. In 2014, a model was presented to produce a prediction model for healthcare centers based on machine learning algorithms and analysis methods about heart failure using a remote health monitoring system [8]. With its ability to turn raw data into actionable knowledge, data analytics has become an indispensable tool for many fields.

In this chapter, we discuss the importance of data analysis, its goals, and the different stages of the data analysis process. Additionally, we outline several commonly used data analysis techniques.

To begin, we explore the use of the one-hot encoding technique, which is instrumental in representing categorical variables obtained from a questionnaire that we distributed to gather individuals' opinions about their work organization. Furthermore, we delve into the application of supervised learning models, specifically neural networks and K-nearest neighbors (KNN), for data analysis. Additionally, we discuss the application of unsupervised learning methods, particularly the K-means technique.

To assess the effectiveness of these techniques, we discuss evaluation metrics such as confusion matrices, accuracy, precision, and recall. Moreover, through this chapter, we discuss time series models, with a particular focus on the FB Prophet model for predicting congestion levels.

The chapter concludes with a discussion of the specific techniques we use in our work and their relevance to our research objectives.



## 3.2 . Steps in data analysis

Data analysis typically involves several steps, each with specific objectives and tasks. These steps are presented in the following.

### 3.2.1 . Data collection

Data collection involves acquiring relevant data from various sources, such as surveys, experiments, observations, or databases. This stage requires careful planning to ensure the data collected align with the research objectives.

### 3.2.2 . Data cleaning and preprocessing

Once the data are collected, they need to be organized and prepared for analysis. This includes data cleaning, data transformation, and handling missing values using methods like dropping missing data, imputation, or advanced techniques [120].

Data cleaning involves the identification and removal of errors in a data warehouse. When data are collected and combined from multiple sources into a data warehouse, ensuring high data quality and consistency becomes a challenging task [131]. Common data quality problems also referred to as anomalies, include variations in data conventions among sources, such as different abbreviations or synonyms. Additionally, data entry errors, like spelling mistakes and inconsistent data formats, as well as missing, incomplete, or incorrect attribute values. Data duplication and the presence of irrelevant objects are other issues that need to be addressed during data cleaning. Data that is incomplete or inaccurate is commonly referred to as "dirty" data [45].

Data preprocessing involves preparing the data for analysis by transforming and manipulating them into a suitable format.

Data analysis techniques, by their nature, operate with real number vectors [56]. Therefore, when working with qualitative data, it becomes necessary to convert it into real number vectors. The most commonly used method for this purpose is one-hot encoding, which efficiently transforms categorical values into vectors with minimal processing [65].

Thus, when working with questionnaires, one-hot encoding represents categorical variables as binary vectors, where each question is represented by a binary vector of zeros, except for the index corresponding to the chosen response, which is marked as 1 (see Table 5.1). For example, a question with N possible responses uses an N-dimensional vector to represent each answer. However, this representation is only applicable to single-choice questions. In the case of multiple-choice questions, the index positions of all selected responses need to be changed to 1.

Table 3.1: Example of One-hot encoding

Color	Red	Green	Bleu
Red	1	0	0
Green	0	1	0
Bleu	0	0	1

## Data exploration

Along with cleaning the data, this step also involves executing an exploratory analysis. Data exploration involves examining the dataset through descriptive statistics and data visualization techniques to gain an initial understanding and identify relationships within the data [75].

Exploratory Data Analysis (EDA) techniques include data visualization, correlation analysis, hypothesis testing, and dimensionality reduction methods such as Principal Component Analysis (PCA) [40]. PCA is used to identify the most important features or components in a dataset. It helps researchers reduce the complexity of the data, identify key variables, and visualize high-dimensional data in a lower-dimensional space while preserving the maximum amount of variation in the data [58].

### 3.2.3 . Data interpretation

Data interpretation involves the extraction of patterns, trends, and conclusions from data [133]. This process encompasses various stages of data analysis, which include descriptive analysis, inferential analysis, predictive analysis, and prescriptive analysis. The subsequent sections provide detailed explanations of each of these analysis types.

1. Descriptive analysis: Also known as descriptive statistics, it is a branch of statistics that involves summarizing and presenting data in a meaningful way to understand the characteristics of a dataset. It helps analysts to explore and describe the main features of their data. It includes techniques for describing measures of central tendency (e.g., mean, median, and mode), measures of variability (e.g., range, variance, and standard deviation), and graphical methods (e.g., histograms, bar charts, and pie charts) to display the data [116].
2. Inferential analysis: In contrast to descriptive analysis, which describes the sample group, inferential analysis takes the description of the sample group to another level by using the sample data to draw conclusions about the population from which the sample was selected. The sample group must be selected randomly from the population in order to draw conclusions about the population [42].

Inferential statistics involves establishing a connection between a sample and the entire population. Since accessing the entire population is often impractical, we rely on estimations based on the sample, such as calculating the mean age of the sample, and seek to infer this estimated value accurately represents the population. To assess the accuracy of the estimator, we utilize the concept of sampling distribution, which enables us to assess the precision of our estimations [52].

3. Predictive analysis: It utilizes statistical and machine learning techniques to develop models that can predict outcomes or make future projections based on the available data [28]. This stage involves selecting an appropriate modeling technique, training the model using historical data, evaluating its performance, and applying the model to new data for predictions. Techniques such as linear regression, decision trees, random forests, support vector machines, or neural networks are commonly used in predictive modeling [91].

An example of a predictive modeling technique is cluster analysis which is an unsupervised learning technique used to group similar observations together based on their characteristics. It helps researchers identify patterns or subgroups within the data, enabling them to discover hidden structures or categorize data points. Common clustering algorithms include K-means clustering, hierarchical clustering, and density-based spatial clustering of applications with noise (DBSCAN) [91].

Another example of a predictive modeling technique is time series analysis which focuses on analyzing data collected over time to identify patterns, trends, and seasonality. It involves techniques such as decomposition, auto-correlation analysis, moving averages, exponential smoothing, and forecasting models. Time series analysis is particularly relevant when dealing with data that exhibit temporal dependencies and is commonly used in fields such as finance, economics, and demographics sciences [91].

4. Prescriptive analysis: The final stage of data analysis involves interpreting the findings and communicating the results effectively. The analysis outcomes need to be interpreted, meaningful conclusions drawn, and their implications in the context of the research objectives discussed. Clear and concise communication of the results through written reports, visualizations, or presentations is essential to disseminate the research findings to the relevant audience. Prescriptive analysis, in the context of data analysis, is the highest level of data-driven decision-making. It goes beyond Predictive analysis, which focuses on predicting future outcomes based on historical data. Prescriptive analysis aims to recommend specific actions or interventions to optimize future outcomes

based on the predictions made through predictive models [138].

### 3.3 . Data analysis techniques

In this section, we provide an overview of various data analysis techniques commonly used in research studies and a combination of the aforementioned data analysis types employed in this thesis to explore and analyze our dataset.

#### 3.3.1 . Artificial intelligence

Artificial Intelligence (AI) is a key component that can be used in various types of data analysis. It is not limited to any specific type but can be applied across multiple stages of the data analysis process to enhance efficiency, accuracy, and depth of analysis.

Artificial intelligence is a field of computing that focuses primarily on the transmission of anthropomorphic intelligence and thinking into machines that can assist humans in many ways [125]

Between 1960 and 1970, scientists studied AI using two types of systems: Knowledge-based systems (KBS) and Artificial Neural Networks (ANNs)[134]. The KBS systems function as computers that offer guidance based on predetermined rules, relying on the knowledge provided by humans. In contrast, ANNs are networks of interconnected neurons structured in multiple layers, imitating the human brain. ANNs have found applications in diverse fields such as medicine, biology, language translation engineering, law, and manufacturing [9].

We find various types of AI that include analytical, functional, interactive, textual, and visual [86], as shown in Figure 3.1. In the following, we present the scope of each category in terms of computing and real-world applications [136].

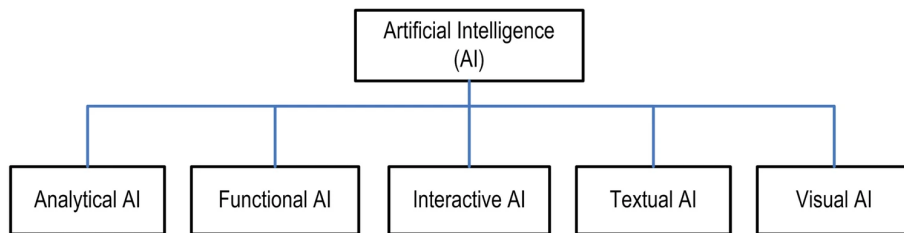


Figure 3.1: Various types of artificial intelligence.

- Analytical AI: It focuses on analyzing and interpreting data to uncover meaningful patterns and relationships. Its goal is to assist in data-driven decision-making. In the domain of business intelligence, analytical AI plays a crucial role by generating suggestions and recommendations

based on its analytical processing capabilities. Various techniques, such as machine learning, can be utilized to develop analytical AI models for solving real-world problems.

- **Functional AI:** Similar to analytical AI, functional AI also explores large datasets to identify patterns and dependencies. However, functional AI goes beyond recommendations and takes direct actions. It finds applications in robotics and IoT.
- **Interactive AI:** It facilitates efficient and interactive communication automation, which has become prevalent in various aspects of our daily lives, particularly in commercial contexts. Building chatbots and smart personal assistants is an example of where interactive AI models are beneficial. Techniques such as machine learning, frequent pattern mining, reasoning, and AI heuristic search can be employed while developing interactive AI models.
- **Textual AI:** It contains textual analytics and natural language processing, enabling businesses to leverage capabilities like text recognition, speech-to-text conversion, machine translation, and content generation. For instance, organizations can utilize textual AI to support an internal corporate knowledge repository, providing relevant services such as answering customer queries.
- **Visual AI:** It specializes in recognizing, classifying, and sorting items, as well as extracting information from images and videos. It can be considered a branch of computer science that trains machines to learn and understand visual data in a manner similar to humans. Visual AI finds applications in fields like computer vision and augmented reality.

In this work, we leverage analytical AI techniques, with a primary focus on machine learning and time series models. Machine learning enables us to develop intelligent systems that can autonomously learn from data and make predictions. Time series models, on the other hand, allow us to capture and analyze patterns and trends over time, which is essential for tasks involving sequential data.

### **3.3.2 . Machine learning**

Machine learning is a branch of data analysis that enables computers to learn patterns, make predictions from data, and inform decision-making processes. Machine learning is a concept that involves acquiring knowledge from examples rather than relying on explicit programming. Instead of writing code directly, data are inputted into an algorithm, which then constructs logical patterns based on the provided information [92]. Figure 3.2 shows the classification of machine learning methods.

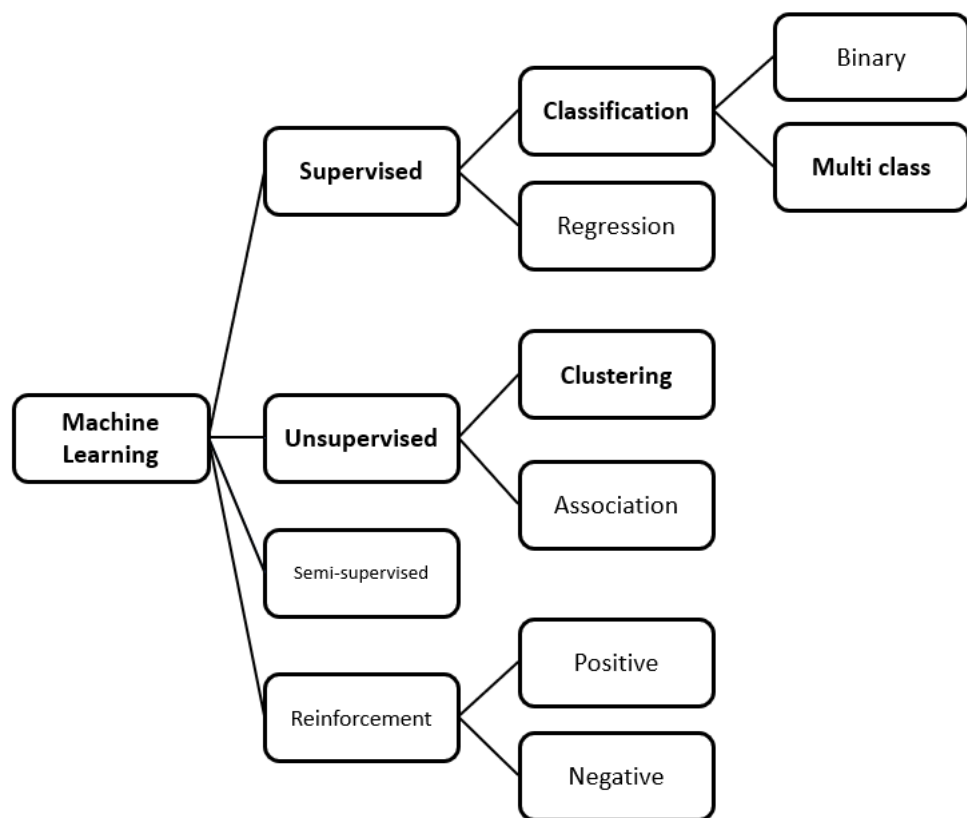


Figure 3.2: Taxonomy of machine learning techniques.

### 3.3.3 . Supervised learning models

Supervised learning is a type of machine learning where the algorithm learns to make predictions or classifications by using a labeled dataset. The output is associated with a label of the input data or the supervision. An input-output training sample is also called labeled training data or supervised data [101]. The algorithm is trained on this labeled dataset and learns to identify patterns or relationships between the input data and the target data. Once the algorithm is trained, it can make predictions on new, unseen data by using the patterns it has learned.

Supervised learning entails two distinct types of learning: regression and classification. Regression is used when the output is continuous. Continuous variables have an infinite number of possible values between any two distinct values. By contrast, classification is used when the output is categorical [47]. Binary classification means assigning instances to one of two classes. Multi-class classification means assigning instances to more than two classes.

Supervised learning can be used for a variety of tasks, such as image classification, speech recognition, and natural language processing [101]. Common supervised learning algorithms include linear regression, logistic regression, decision trees, K-nearest neighbors, and neural networks.

In this thesis, we use methods from the classification models because we have a limited set of profiles to predict. We use Artificial Neural Networks (ANN) and K-Nearest Neighbors (KNN) algorithms for supervised learning tasks. ANN was chosen for its ability to capture complex patterns in the data, while KNN is utilized for its effectiveness in making predictions for unknown mobility profiles based on their proximity to labeled instances. This combination provides a robust and accurate solution for classifying and predicting individual mobility profiles. Both techniques are presented in the following.

### Artificial neural networks

Artificial neural networks (ANN) have been discovered as an essential alternative to various standard classification methods. A good definition is given by Haykin [61], describing a neural network as a massively parallel combination of the simple processing unit that can acquire knowledge from the environment through a learning process and store the knowledge in its connections. In theory, a neural network can perform any given classification task, provided that a judicious choice of the model is made and a suitable training method is implemented.

A MultiLayer Perceptron Neural Network (MLPANN) is a backpropagation artificial neural network that generates a set of outputs from a set of inputs [115]. It comprises one input layer, one or more hidden layers, and one output

layer. While the hidden layers are used for computation (see Figure 3.3), the output layer represents the modeling purpose.

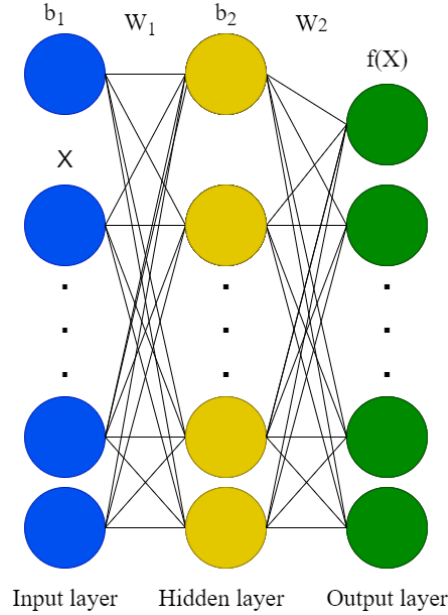


Figure 3.3: An MLPANN with one hidden layer.

MLPANN is a powerful machine learning technique and has been successfully applied in many domains, including image recognition, speech recognition, natural language processing, and financial forecasting [128]. However, training MLPANN can be computationally expensive, and the network architecture must be carefully designed to avoid overfitting and achieve good performance on new data [122].

Generally, the number of nodes composing the input layer depends on the selected factors in the data source, and the hidden layer nodes are quantified based on a specific training dataset. Moreover, each node of the input layer must be connected to the nodes of the first hidden layer, whereas each node of the output layer must be connected to all nodes in the last hidden layer. Using the network's connections, the training implementation of the MLPANN method can be divided into two phases: the forward phase and the backward phase. In the forward phase, propagation of activation occurs from the input to the output layer. The propagation of error among the actual, requested, and observed value occurs at the backward phase in which the output layer changes the bias values and weights [83]. This computation process is illustrated in Equation 3.1 by the function  $f(X)$ , where  $X$  is the input data vector,  $W_1$ , and  $W_2$  are the matrices of the connection's weights, and  $b_1$  and  $b_2$  are the bias vectors of the hidden and the output layer, respectively.



$$f(X) = b_2 + W_2 * f_A(b_1 + W_1 * X) \quad (3.1)$$

$f_A$  represents an activation function. The most popular activation function is the sigmoid function (see Equation 3.2).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.2)$$

The weight matrices  $W_1$  and  $W_2$ , as well as the bias vectors  $b_1$  and  $b_2$ , are randomly initialized first and then updated through a training process. The training parameters such as the learning rate and training epoch are set based on experience. Comparing the model's output values to the target values estimates the error value which can be minimized in the backward stage by updating the connection weights.

## K-nearest neighbors

The nearest neighbors algorithms are among the simplest of all machine learning algorithms. K-Nearest Neighbors (KNN) is a supervised machine learning algorithm that can be used for classification and regression problems [76]. For classification problems, the idea is to memorize the training set and then predict the label of any new instance on the basis of the labels of its closest neighbors in the training set. KNN classification goes through two stages; the first one allows the determination of the nearest neighbors by calculating the Euclidean or any other distance metric and the second one the determination of the class using those neighbors [124].

The KNN method is a non-parametric classification method, which is effective in many cases. However, it requires the selection of an appropriate value for parameter  $K$ , and the success of classification is very much dependent on this value. In a sense, the KNN method is biased by  $K$  [60]. A basic illustrative example is shown in Figure 3.4. If we set  $K=3$  the new data to classify (the dot in green) will be classified as a triangle, but if we choose  $K=5$  it will be considered as a square. There are many ways of choosing the  $K$  value, but a simple one is to run the algorithm with different  $K$  values and choose the one with the best performance.

One of the main advantages of KNN is that it is a simple algorithm to understand and implement. However, it can become computationally expensive as the number of training examples and dimensions of the feature space increase. Moreover, the choice of the value of  $K$  and the distance metric can greatly affect the performance of the algorithm. Thus, these parameters need to be carefully selected.

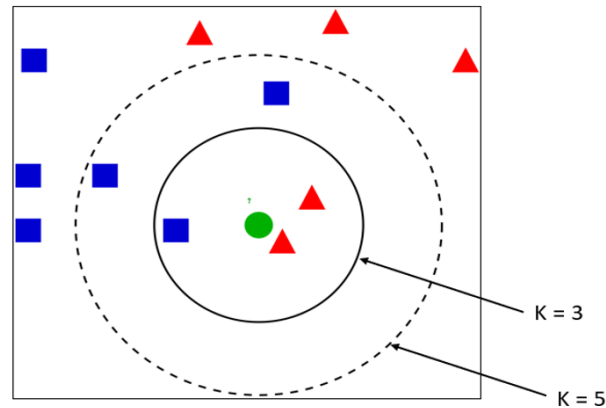


Figure 3.4: K-nearest neighbors technique.

## Supervised learning techniques evaluation

In this section, we provide a detailed explanation of how the effectiveness and accuracy of a machine learning model, which depends on classification techniques, are examined and assessed. This evaluation involves utilizing the confusion matrix to derive metrics such as accuracy, precision, and recall. Further information and a comprehensive definition of these metrics are presented.

### 1. Confusion matrix

A confusion matrix is a condensed table that shows the number of correct and incorrect predictions made by a classification model. If the number of target classes is  $N$ , then the confusion matrix is  $N \times N$ . By looking at the confusion matrix, we can determine the model's accuracy by examining the diagonal values, which indicate the number of accurately classified instances [82]. In a confusion matrix, true positive ( $TP$ ) represents instances correctly classified as the positive class, true negative ( $TN$ ) represents instances correctly classified as the negative class, false positive ( $FP$ ) denotes instances incorrectly classified as positive, and false negative ( $FN$ ) signifies instances incorrectly classified as negative in a classification scenario.

### 2. Accuracy

Accuracy is defined as the proportion of correctly classified instances out of the total number of instances. For multiclass classification, it is computed as the ratio of the number of correctly classified instances ( $TN, TP$ ) to the total number of instances (see Equation 3.3). This metric measures the overall correctness of a model's predictions across all classes.

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN} \quad (3.3)$$

### 3. Precision

Precision can be defined for each class individually or in a macro-averaged or micro-averaged form across all classes. In the individual class definition, precision for a particular class is the proportion of true positives for that class out of the total number of instances predicted as that class. In macro-averaged precision, the precision for each class is first computed, and then the average of those precisions is taken. In micro-averaged precision, the precision is computed by summing the true positives and false positives across all classes and then computing the ratio of true positives to the total number of predicted positives. It is given by Equation 3.4.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.4)$$

### 4. Recall

Recall can also be defined for each class individually or in a macro-averaged or micro-averaged form across all classes. In the individual class definition, recall for a particular class is the proportion of true positives for that class out of the total number of actual instances belonging to that class. In macro-averaged recall, the recall for each class is first computed, and then the average of those recalls is taken. In micro-averaged recall, the recall is computed by summing the true positives and false negatives across all classes and then computing the ratio of true positives to the total number of actual positives (see Equation 3.5).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.5)$$

In summary, for multiclass classification, accuracy measures the overall correctness of a model's predictions across all classes, precision measures how many of the predicted instances for each class are actually correct, and recall measures how well the model is able to identify instances for each class. These metrics can be calculated for each class individually or in a macro-averaged or micro-averaged form across all classes.

#### 3.3.4 . Unsupervised learning models

Unsupervised learning is a type of machine learning where the algorithm is trained on a dataset without any supervision or labeled data. The goal of unsupervised learning is to identify patterns or structures in the data without

being told what to look for. This makes unsupervised learning useful when there is no prior knowledge about the data or when there are too many variables to consider [59].

There are several types of unsupervised learning algorithms, including clustering and association. Clustering algorithms group similar data points together, while association rules are used to find interesting relationships in large datasets. These relationships are often represented as rules or sets of items that frequently appear together. The main purpose of association rules is to discover exciting connections between different objects in a dataset. For example, it can help us find out which items are commonly bought together by customers in a store or which stores people tend to visit together. It can also be used for other tasks like determining which products should be bundled together, making decisions about product assortment, suggesting cross-selling opportunities, and more. In a way, association rule mining is like a more advanced version of a "what-if" scenario. It helps us understand if one event occurs (if this), what other events might also happen (then that) [48]. Several methods of unsupervised learning exist such as K-means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and hierarchical clustering.

Choosing the K-means algorithm for this thesis is due to its simplicity and efficiency. Its easy-to-understand iterative process and computational effectiveness make it suitable for clustering large datasets. Moreover, K-means often produce well-separated and interpretable clusters represented by centroids, aiding result interpretation and visualization. However, it has limitations, such as the assumption of spherical clusters with similar variance and its sensitivity to initial centroid placement, potentially leading to local optima, but in our research work, as we will see in Chapter 5, we address this by evaluating the Silhouette score and the Elbow function.

## K-means

K-means is an unsupervised heuristic algorithm that partitions a data set into  $K$  clusters by minimizing the sum of squared distance in each cluster. The algorithm consists of three main steps:

- a) Initialization by setting center points (or initial centroids) with a given  $K$ ,
- b) Dividing all data points into  $K$  clusters based on  $K$  current centroids,
- c) Updating  $K$  centroids based on newly formed clusters. The more similarity among the data in clusters, the more the chance of particular data items to belong to a particular group.

A metric function or distance function is a function that defines the distance between elements/objects in a set [141]. Distance metrics play a significant role in the clustering process.

In this thesis, we consider different metrics: Euclidian, Manhattan, Minkowski, Chebyshev, Hamming, Jaccard, and Correlation. However, we only present those that provided the pertinent results. The distance between two points  $A$  and  $B$  are presented respectively for each distance metric in the following. Considering that:

- $d$  is the distance,
- $n$  is the number of dimensions of the considered points,
- $x_i$  is the coordinate of point  $A$  in the  $i$ th dimension,
- $y_i$  is the coordinate of point  $B$  in the  $i$ th dimension.

- The Euclidean distance computes the root of the square difference between the coordinates of a pair of objects.

$$d(A, B) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.6)$$

- The Manhattan distance computes the absolute difference between the coordinates of a pair of objects.

$$d(A, B) = \sum_{i=1}^n |x_i - y_i| \quad (3.7)$$

- The Minkowski distance is defined as the generalized metric distance

$$d(A, B) = \left( \sum_{i=1}^n (|x_i - y_i|)^p \right)^{1/p} \quad (3.8)$$

When  $p=2$ , the distance becomes the Euclidean distance. When  $p=1$  it becomes city block distance (Manhattan distance).

In order to obtain the optimal number of clusters  $K$ , we use the Elbow function and the Silhouette score. The Elbow function calculates the variance of the different volumes of clusters considered and then places the variances obtained on a graph. The optimal number of clusters corresponds to the point of the elbow from which the variance does not drop more significantly. The Elbow method is a visual method to test the consistency of the best number of clusters [147].

The Silhouette score is simple to interpret since it captures the trade-off between within- and between-cluster distances. The best value is 1, and the worst is  $-1$  [126].

### 3.3.5 . Semi-supervised learning and reinforcement learning

Semi-supervised learning sits between supervised and unsupervised learning. It uses a bit of labeled data and lots of unlabeled data to train models. This helps the model make better predictions by finding patterns in the unlabeled data. It's useful when getting labeled data is hard or expensive [130].

Reinforcement learning involves training agents to make decisions in an environment to maximize a cumulative reward. Positive reinforcement is a fundamental concept in reinforcement learning, where agents receive rewards for actions that lead to desired outcomes. Conversely, negative reinforcement involves penalties or decreased rewards for undesirable actions. This dual approach of positive and negative reinforcement guides agents toward learning optimal strategies through trial and error, enabling them to navigate complex environments and learn from their experiences [64].

In this thesis, we did not delve into the semi-supervised learning and reinforcement learning approaches. This choice was made because our primary objective was to classify profiles, a task that can be effectively accomplished using classification methods. As a result, there was no requirement for the utilization of semi-supervised or reinforcement learning techniques. Instead, we opted to employ unsupervised learning to identify common features within each predefined category.

### 3.3.6 . Time series forecasting

Time series forecasting is a technique for predicting events through a time sequence. It predicts future events by analyzing past trends assuming that future trends will be similar to historical trends [51].

There are several time series models, such as ARIMA (AutoRegressive Integrated Moving Average), Random Forest, and Exponential Smoothing. These models have been widely used for capturing hidden patterns and variations within time-dependent data.

However, when dealing with non-sequential data, where a temporal ordering might not be present or when the data has missing values and irregular patterns, traditional time series models might not perform optimally. This is where the Facebook Prophet (FB Prophet) model takes advantage. The FB Prophet model was specifically designed to address some of the limitations of traditional models in handling these types of non-sequential time series data [70].

Moreover, one of the interesting features of the FB Prophet model is its ability to handle various sources of non-linearity and seasonality. It incorporates a flexible framework to capture annual, weekly, and daily seasonalities along with holidays and special events. Moreover, FB Prophet can effectively handle missing data points and outliers, making it more robust in scenarios where the data is not consistently recorded. These attributes make FB

Prophet an attractive choice for our purposes.

### FB Prophet model

The FB Prophet is a machine learning developed by Facebook for forecasting time series data where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series with strong seasonal effects and several seasons of historical data [145]. FB Prophet is based on an additive model, defined as follows [74]:

$$y(t) = Trend(t) + Seasonal(t) + Residual(t) \quad (3.9)$$

where  $y(t)$  is the observed value of the time series at time  $t$ ,  $Trend(t)$  represents the long-term trend in the series,  $Seasonal(t)$  captures the recurring patterns or seasonality, and  $Residual(t)$  is the random noise or error component.

Thus, the model has several components that can be customized to improve its prediction accuracy [145].

The first component is the *trend*, which is the overall direction of the time series data. The trend component is usually modeled as a linear or nonlinear function of time. FB Prophet can automatically detect and model changes in the trend over time.

The second component is *holidays*, which are special events or occurrences that can affect the time series data. FB Prophet allows users to specify a list of holidays that can impact the forecast. Holidays can cause a peak in congestion data, which may need to be accounted for in the forecast.

The third component is the *weekly seasonality*, which captures any recurring patterns that occur every week. For example, if the data show higher congestion on weekends than on weekdays, then the weekly seasonality component can help capture that pattern.

The fourth and last component is the *daily seasonality*, which captures any recurring patterns that occur daily. For example, if the data show higher congestion during certain hours, then the daily seasonality component can help capture that pattern.

By incorporating these components into the forecasting model, FB Prophet can generate accurate predictions.

### 3.4 . Conclusion

This chapter serves as an overview for understanding the pivotal role that data analysis plays in extracting meaningful information from data. As emphasized throughout the chapter, data analysis enables researchers to uncover patterns, trends, and relationships within complex datasets. By em-

ploying a variety of techniques, we can effectively transform raw data into evidence-based conclusions.

The chapter covers various aspects of data analysis, starting with an exploration of its significance and goals. It delves into the stages of the data analysis process. Then, we discuss the commonly used techniques, such as one-hot encoding, supervised learning models, and unsupervised learning methods. Notably, the integration of evaluation metrics like confusion matrices, accuracy, precision, and recall underscores the importance of rigorously assessing the performance of analysis techniques. This ensures the reliability and applicability of the results obtained. The chapter also highlights the relevance of time series models, specifically the FB Prophet model, for predictive purposes.

In the next chapters, we use a set of the aforementioned data analysis techniques to explore and analyze our dataset. Firstly, we utilize one-hot encoding to transform categorical variables into a suitable numerical representation that machine learning models can process. We chose this technique because it efficiently transforms categorical values into vectors with minimal processing. Next, we apply the MLPANN algorithm and the K-Nearest Neighbors algorithm to perform classification task (Chapter 5). Furthermore, we use K-Means clustering to discover hidden patterns and groupings within our dataset. By partitioning the data into distinct clusters, we understand the relationships present in our dataset (Chapter 6). Lastly, we use Facebook Prophet, a time series forecasting tool, to model and predict future congestion levels based on temporal patterns (Chapter 7).





## 4 - The decision platform

### 4.1 . Introduction

A decision platform is an integrated system that uses advanced technologies and data analytics to support decision-making processes. It enables the analysis and visualization of diverse data sources to understand patterns and facilitate decision-making. In the context of urban mobility, as it is raised in our research work, the decision platform assists users in understanding the complexities of urban mobility and proposing effective strategies to mitigate congestion. By using an interdisciplinary approach and incorporating sociological considerations, the objective of this decision platform is to enhance the understanding of urban mobility dynamics and enable the identification of optimal temporality scenarios for congestion reduction. Ultimately, the decision platform serves as a valuable tool for improving urban traffic management.

The objective of our doctoral project is to propose an interdisciplinary approach to understanding urban mobility in its complexity, going beyond simple flow measurement. This understanding is a prerequisite for the interdisciplinary approach of ICT (Information and Communication Technologies) and SHS (Social and Human Sciences) that we propose.

Our decision platform uses diverse data sources, including a questionnaire capturing individuals' mobility choices and work organization, Open Street Map (OSM) for trajectory patterns, and Waze application programming interface (API) for road congestion data. The platform classifies and quantifies collected data, then conducts analysis to generate optimal departure time scenarios, considering both traffic conditions and sociological factors. By offering personalized recommendations and options, the platform aims to mitigate congestion and improve urban traffic management, making it an interesting tool for enhancing urban mobility and individual quality of life.

To achieve this goal, we propose an architecture for the decision platform designed for the time bureaus (see Figure 4.1). It consists of five modules, each of which is explained in detail in the following sections.

### 4.2 . Input data

The input data module in the decision platform architecture focuses on collecting and integrating data from various heterogeneous sources. These sources may include traffic sensors, public transport systems, GPS data, social media, and other relevant datasets. By combining these diverse sources, the platform aims to create a comprehensive and accurate understanding of

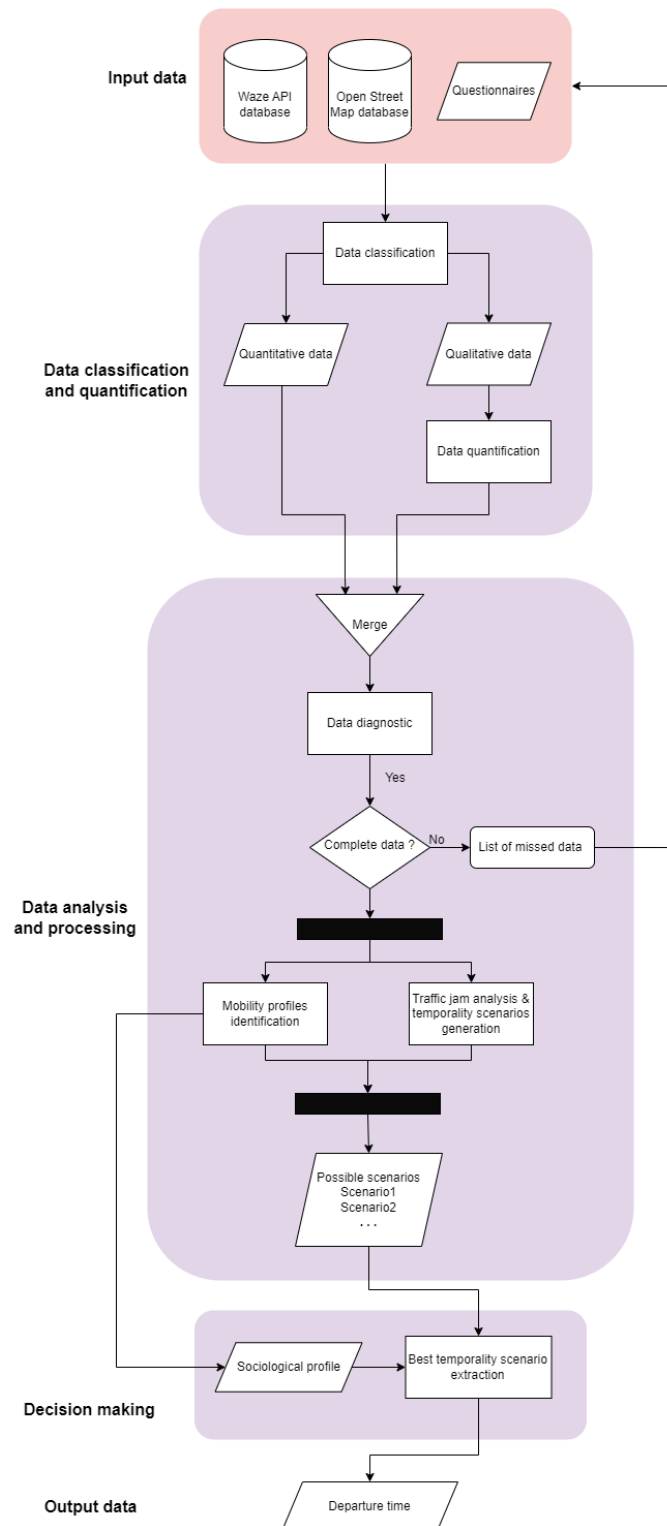


Figure 4.1: Decision platform flowchart.

urban mobility patterns. The module is responsible for harmonizing the heterogeneous data to create a comprehensive view of urban mobility. The input data module involves data cleaning, data transformation, and data fusion to ensure the accuracy and reliability of the integrated datasets.

In our proposed platform, the input data module incorporates several key components. Firstly, we have developed a questionnaire. This questionnaire is designed to capture the expectations of individuals regarding their mobility choices and the organizational dynamics of their work. This input from individuals is instrumental in understanding their specific needs and preferences. Additionally, the Open Street Map (OSM) database is utilized within this module. By leveraging OSM, the Open Street Routing Machine (OSRM) can access detailed information about trajectory patterns that individuals are likely to traverse, aiding in route optimization and analysis. Furthermore, the input data module integrates the Waze API. By accessing the Waze API, relevant data on road congestion are obtained, facilitating thorough analyses of traffic conditions that individuals are likely to encounter along their routes. This comprehensive integration of the questionnaire, OSM, and Waze API databases performs the platform's ability to gather critical information, optimize routing, and conduct detailed traffic analyses to support decision-making processes. These three data sources of the input data module are detailed in the following sections.

#### **4.2.1 . The questionnaire**

The questionnaire has been defined in collaboration with Île-de-France Region and Grand Paris Sud Urban Community.

Île-de-France Region is involved in most areas that concern the actions of the 12 million residents of Île-de-France: transportation, high schools, apprenticeship programs, economic development, environment, and more. Moreover, it plans and develops a territory that is both urban and rural, representing 2% of France and housing 18% of its population. The Ile-de-France Region contributes to nearly 30% of the national Gross Domestic Product (GDP) [2].

The Grand Paris Sud territory is the largest community of municipalities in France in terms of population, with a population of 355,010 inhabitants. Located in the southern part of Île-de-France, it brings together 23 municipalities spanning two departments, Essonne and Seine-et-Marne (see Figure 4.2) [1].

As a major economic hub in the Paris region, Grand Paris Sud hosts over 31,000 businesses from big companies (such as Safran, Arianespace, and Truffaut), and five renowned regional shopping centers. Grand Paris Sud is also a cultural hub. The area boasts 11 venues dedicated to performing arts, including two national stages, 22 media libraries, a dense network of conservatories and music schools, multiplex cinemas, as well as seven art-house cinemas. Grand Paris Sud is governed by 83 elected officials. Every day, Grand Paris

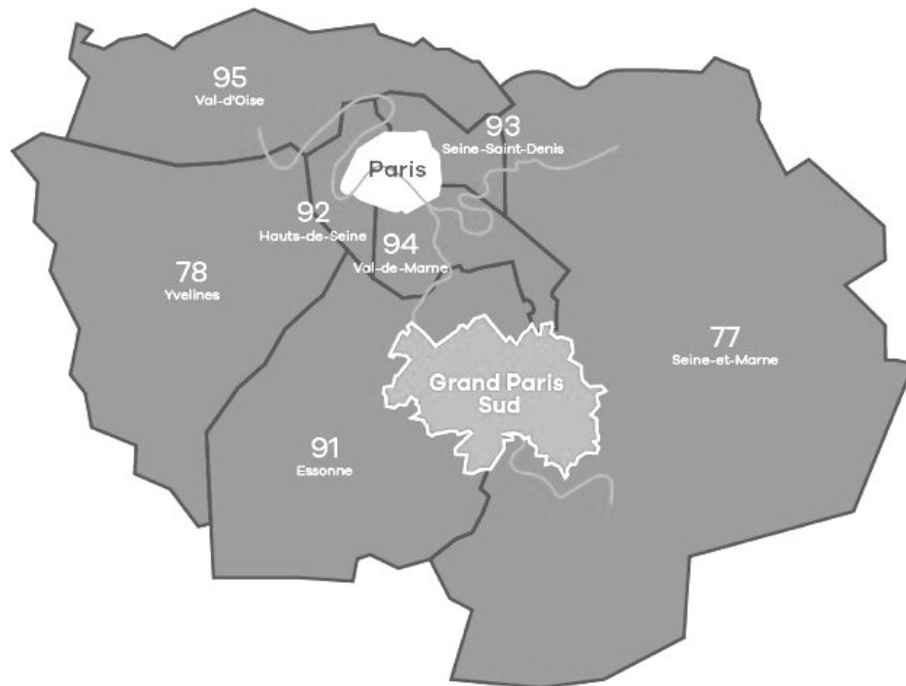


Figure 4.2: Grand Paris Sud territory.

Sud strives to enhance residents' quality of life, support education for their children, and foster economic development [1].

The aforementioned factors make Grand Paris Sud an appealing territory for conducting our research work. The questionnaire was distributed to the employees of more than fifty (50) companies in this Paris suburb area between May and October 2021. It deals with the evolution of mobility and work organization before and after the sanitary crisis. It consists of the following four main sections.

**1) Mobility and work organization before the health crisis:** This first section of the questionnaire focuses on mobility and work organization before the health crisis. It gathers questions about employees' transportation habits and work arrangements. It begins by inquiring about the modes of transportation used by employees to commute to their workplace and the frequency of their travels. Additionally, it explores the employees' perceptions of the pleasantness of their commute and asks for their opinion on the key criteria that contribute to a pleasant journey. Furthermore, participants are asked about their ideal mode of transportation. In this section, we also ask about aspects such as journey times and distances traveled between home and work, as well as the typical arrival time at the workplace and the reasons behind it. Finally, this section addresses the topic of teleworking, seeking information about employees' teleworking practices before the health crisis. By capturing

these information, the questionnaire provides us with a comprehensive understanding of employees' pre-crisis mobility patterns and work organization, which is crucial for analyzing and proposing temporality scenarios to enhance urban mobility and reduce congestion.

**2) Mobility since the health crisis:** The second section of the questionnaire focuses on the changes in mobility patterns since the health crisis. It collects information about the transportation modes used by employees and the frequency of their usage. The section also explores whether employees express a desire to change their current transport mode and, if so, the reasons behind this. By investigating the differences in transportation preferences and understanding the underlying motivations, this section provides a comprehension of the evolving mobility choices and priorities of the surveyed employees. These information also contribute to a better understanding of the potential changes in travel behavior and help us to intervene in individuals' schedules with the aim of improving urban mobility and reducing congestion.

**3) Organization of work since the health crisis:** This third section focuses on the organization of work since the health crisis, with a particular emphasis on the initial three confinement phases. It explores how employees adapted to the changing work environment during this period. The questionnaire asks employees about their work practices, such as remote working or on-site attendance, during the crisis. It further delves into aspects such as the frequency and duration of remote work, any challenges faced, and the level of satisfaction with the work organization implemented by their respective companies. By capturing employees' experiences and perceptions of work arrangements during the health crisis, this section shows the effectiveness of different work practices and allows the evaluation of employee satisfaction with the terms and conditions of their work organization. This set of information can inform us about future strategies for flexible work arrangements, considering information learned from the crisis and aiming to optimize work organization while addressing employee needs and preferences.

**4) Personal information about the respondents:** This last section includes questions about the individual's gender, age, marital status, family situation, and socio-professional categories. These information are crucial as they allow the identification and understanding of the sociological profiles of the surveyed individuals. By capturing these details, the questionnaire enables the identification of specific factors that may influence work temporalities and preferences. Understanding the personal context of the respondents allows us the development of more targeted and effective strategies for work organization, considering factors such as age, family obligations, and professional categories. By taking these aspects into account, the decision platform can provide personalized recommendations and solutions that accommodate

the needs and preferences of different individuals within the surveyed population.

The whole questionnaire is provided in ??.



Figure 4.3: Municipalities of surveyed population residency.

The distribution of the surveyed individuals is shown in Figure 4.3 regarding their residential municipalities. The largest slice represents Essonne, with 513 people, followed by Seine et Marne, with 224 people. The other communes are much smaller, with Hauts-de-Seine having 43 people, Paris having 72 people, and Val-de-Marne having 70 people who answered the questionnaire.

#### 4.2.2 . Open Street Map

Open Street Map (OSM) is a remarkable example of participatory Geographic Information Systems (GIS) [79]. OSM is a collaborative web-mapping project that collects geospatial data to create and distribute online maps. The OSM is freely available to anyone with an Internet connection. It allows Internet users to contribute and edit geospatial data. OSM is maintained by volunteer cartographers worldwide who use GPS devices, portable cameras, and laptops for field mapping. The governmental and commercial sources complete the collected data with digitized open-source aerial photography and free maps [39].

In this thesis, we use the Open Source Routing Machine (OSRM) for generating route names and the traveled distance of each route. OSRM is an

open-source routing engine that calculates the shortest paths between two points on a map using the OSM database. OSRM uses modern C++ and runs on various platforms, including Linux, macOS, and Windows. It can be used as a standalone or a library in other applications. OSRM supports several traveling modes, including car, bicycle, and pedestrian. It uses Open Street Map data as its source of map information, meaning it can be used in any area where Open Street Map data is available. One of the key features of OSRM is its ability to generate routes and travel times in real-time [102].

#### 4.2.3 . Waze application programming interface database

Historical data is made available by Waze [3] as part of the connected citizen partnership with Grand Paris Sud. The Grand Paris Sud agglomeration community supports open data initiatives and provides its engineering and expertise to the municipalities of Grand Paris Sud regarding the opening of public data through two mechanisms: hosting data on the *data.grandparissud.fr* portal and providing support to elected officials and municipal employees (training, workshops). Grand Paris Sud has granted us access to this dataset, which can be used through an API that allows searching and downloading records based on multiple criteria.

We use the dataset of congestion reported by Waze. Table 4.1 describes the attributes present in a dataset of congestion reported by Waze users. This dataset is essential for our work because it contains information that allow us to analyze the traffic situation at different times.

Table 4.1: Waze data attributes

Attribute	Description
utc_date	The UTC time of the jam's occurrence
speed	Current average speed on jammed segments meters/seconds
length	Jam length in meters
delay	Delay of jam compared to free flow speed in seconds
day of the week	Day of the week

- **utc\_date:** It represents the Universal Coordinated Time (UTC) when a jam report occurred. The timestamp is essential for understanding the exact timing of congestion events. Analyzing congestion patterns across different times of the day and days of the week can provide pertinent information about peak traffic hours and potential recurring patterns.
- **speed:** It indicates the current average speed on the segments affected by the jam, measured in meters per second. This information helps us



assess the severity of the congestion. Lower average speeds usually indicate heavier traffic and potentially more significant delays.

- length: This attribute provides the length of the jam in meters. This gives us an understanding of how far the congestion extends along the road.
- delay: It represents the delay of the jam in comparison to the free-flow speed, measured in seconds. It quantifies how much longer it takes to traverse the jammed segment compared to normal traffic conditions. This helps us assess the impact of congestion on travel time.
- day of the week: This attribute indicates the day of the week when the jam occurred. Analyzing congestion patterns across different days of the week can reveal trends, such as higher congestion during weekdays due to work commutes or variations during weekends.

### **4.3 . Data classification and quantification**

The second module of the decision platform focuses on data classification and quantification to facilitate effective data processing.

#### **4.3.1 . Data classification**

Data classification is the process of categorizing data to maximize their effectiveness and efficiency in analysis [157]. Accurate classification is crucial for ensuring accurate analysis. This is why making the choice of classification method is significant [11]. This module's first part objective is to split the data into qualitative and quantitative data.

#### **Qualitative data**

Qualitative data refers to non-numerical data that provide information about subjective experiences, opinions, behaviors, and perspectives. Unlike quantitative data, which focus on measuring quantities, qualitative data concerns the qualitative aspects of a phenomenon. It involves methods such as interviews, observations, and textual analysis to collect rich and descriptive information [68]. Moreover, qualitative data analysis involves identifying themes, patterns, and meanings within the data to generate a comprehensive understanding of a phenomenon. It is particularly useful in social sciences, psychology, anthropology, and other fields that aim to explore the meaning of human behavior and social interactions [37]. Through qualitative research, researchers gain a deeper understanding of individuals' lived experiences and uncover nuances that quantitative data alone may not capture [109].

In our work, obtaining qualitative data are essential as they offer crucial information about individuals' subjective experiences, opinions, and behaviors regarding their mobility and work organization. Qualitative data uncover nuances that quantitative data alone cannot capture, facilitating a comprehensive exploration of the phenomena under study.

## **Quantitative data**

Quantitative data refer to information that are expressed in numerical form and can be measured or counted. They provide a quantitative measure of various attributes or characteristics, allowing for objective analysis and statistical treatment [50]. By utilizing quantitative data, scientists can have a better vision of a phenomenon and make predictions [129].

In our work, obtaining quantitative data from sources like Waze and OpenStreetMap (OSM) is crucial. These data provide numerical information about traffic patterns and trajectory details.

### **4.3.2 . Data quantification**

Data quantification is the process of assigning numerical values to qualitative data, enabling researchers and analysts to measure, analyze, and draw meaningful conclusions. It plays a crucial role in various scientific disciplines, including statistics, economics, and research methodologies [12]. In numerous areas of the social sciences, there has been a shift toward using quantitative methods. In these fields, comprehending a phenomenon involves measuring and expressing its observed characteristics using numerical values [142].

Data quantification is essential in our work as it converts qualitative information gathered from the questionnaire and the data extracted from traffic patterns and routes, which are represented as text and sequences of text, into numerical values, facilitating data analysis.

## **4.4 . Data analysis and processing**

The initial step of this module involves merging the data to create a single file containing all the necessary information for the entire processing. The first process within this module is to verify if the available data are complete, ensuring that all the required information are available for the processing to proceed successfully. This verification process can have two distinct outcomes.

If the data are found to be incomplete, the process provides a report listing the missing data that should be completed if possible. Conversely, if the data are complete, two parallel processing tasks are initiated:

- The identification of mobility profiles which focuses on addressing ur-

ban traffic congestion by identifying and characterizing individuals' mobility profiles. Using survey data, the methodology classifies individuals into predefined mobility profiles: "Unfavorable," "Favorable flexible," and "Favorable not flexible."

1. Unfavorable (UNF): This category encompasses individuals refusing to alter their work arrangements.
2. Favorable flexible (F-F): Within this category, we find individuals willing to modify their working hours to alleviate traffic congestion. Their flexibility is attributed to the absence of personal or professional constraints.
3. Favorable non flexible (F-NF): Individuals in this category are in favor of altering their work arrangements, yet due to personal or professional limitations, their flexibility is constrained

Additionally, this process explores the use of clustering technique, revealing common features within each cluster and aligning with predefined profiles. The approach provides insights for optimizing work schedules and reducing traffic congestion.

These works are presented in Chapter 5 (Mobility profiles identification and prediction).

- The analysis of traffic jams and the generation of temporality scenarios: this processing introduces a multi-step process to optimize departure times. The methodology involves gathering route information from Open Street Map, training a traffic congestion model using real-time data from Waze API, and utilizing time series forecasting models for congestion prediction.

These works are presented in Chapter 7 (Optimal traffic-free departure forecast).

The output of this processing consists of an ordered list of departure times, starting from the one resulting in the minimum congestion and extending to the worst case.

## **4.5 . Decision making**

The primary objective of this module is to finalize the comprehensive processing framework proposed within the platform. Therefore, our initial focus is on the mobility profile of individuals. If an individual's profile is assessed as unfavorable for changes in her/his daily mobility, no scenarios are proposed. However, if an individual is classified as favorable and flexible, we

recommend departure times that minimize congestion. This recommendation is made without taking her/his preferred departure times into account. If multiple congestion-avoidance options are available with the same impact, we suggest the option that is closest to her/his desired time. This approach grants them the freedom to choose from a set of potential scenarios.

On the contrary, when an individual is classified as favorable but lacks flexibility, we incorporate their preferred departure time. We then recommend the nearest alternative departure time that avoids congestion. This acknowledgment stems from the understanding that even a minor time adjustment, such as 5 minutes, can yield a substantial impact on avoiding congestion peaks.

Upon completing all processes, the output module establishes the optimal departure time. This determination is made while accounting for the sociological elements tied to individuals. Figure 4.4 illustrates the explained dynamic.

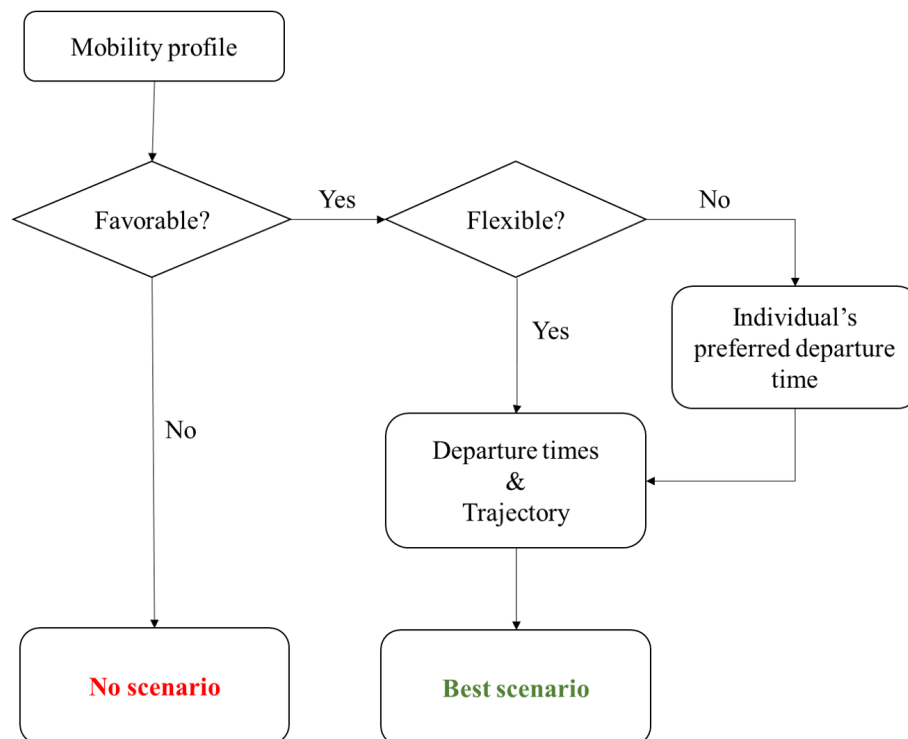


Figure 4.4: Decision-making procedure.

#### 4.6 . User interface of the decision platform

The user interface of the decision platform consists of interactive components that allow users to interact with the platform ergonomically and effec-

tively.

- **Dashboard:** The dashboard serves as the main interface, providing an overview of the platform's capabilities and options.
- **Input section:** Users can input their data, including responses to the questionnaire and preferred departure times.
- **Data analysis section:** This section displays visualizations generated from the analysis of collected data. It shows mobility profiles, congestion patterns, and suggested departure times.
- **Decision recommendations:** The platform presents personalized recommendations based on an individual's mobility profile and preferences. For flexible individuals, it suggests optimal departure times, and for non-flexible individuals, it offers departure times closest to their preferences.
- **Output section:** Users receive the final output, which includes ordered lists of departure times optimized to avoid congestion.
- **Navigation bar:** A navigation bar provides easy access to different sections of the platform.

The model of the platform that we propose is detailed in Appendix 3.

## **4.7 . Conclusion**

In this chapter, we have presented an overview of the decision platform designed to address the issue of congestion in urban mobility. The platform consists of five key modules: Input data, Data classification and quantification, Data analysis and processing, Decision making, and Output data.

In the Input data module, we have described the sources of data for the platform, including the questionnaire distributed to employees in collaboration with the Ile-de-France Region and Grand Paris Sud Urban Community. This questionnaire captures interesting features in mobility choices, work organization, and preferences of the surveyed individuals. Additionally, we have discussed the integration of Open Street Map (OSM) and Waze API databases, which provide detailed information about trajectory patterns, road congestion, and traffic conditions.

The data classification and quantification module plays a crucial role in categorizing and assigning numerical values to the collected data. We have discussed the distinction between qualitative and quantitative data, highlighting

the significance of qualitative data in capturing subjective experiences, opinions, and behaviors, while quantitative data allow for objective analysis and measurement.

In the data analysis and processing module, we have outlined the workflow of merging, verifying, and analyzing the data. This module includes the identification of mobility profiles, analysis of traffic jams, and the generation of temporality scenarios. The following chapters elaborate on these processes. The output of this module is an ordered list of departure times, providing options based on congestion levels.

In the Decision-making module, we focus on considering the individual's mobility profile and sociological constraints to identify the most suitable scenario. This module takes into account flexibility in departure times and offers personalized recommendations. For individuals with specific preferences, the platform suggests the closest departure time to their preferred choice, while for others, it provides the earliest optimal departure time to avoid congestion.

Finally, the Output data module delivers the optimal departure time and best trajectory, considering both traffic congestion and sociological factors. By combining the previous modules, the platform provides information to time management offices, enabling them to propose scenarios aimed at reducing congestion and improving urban mobility.

Our proposed decision platform offers an interdisciplinary approach to understanding and managing urban mobility. By integrating data from various sources, classifying and quantifying the data, and conducting a thorough analysis, the platform provides recommendations for decision-making. The platform's ability to consider both quantitative and qualitative aspects, along with sociological factors, enhances its effectiveness in addressing congestion in urban mobility.

In the next chapter, we present in detail the identification of mobility profiles processing as mentioned earlier.



## 5 - Mobility profiles identification and prediction

### 5.1 . Introduction

Mobility profiles refer to patterns of movement and transportation behaviors of individuals or groups of people. These profiles can include various types of transportation modes, such as walking, biking, driving, and public transportation. This can be useful in various applications, such as urban planning, transportation management, and marketing. Indeed, predicting mobility profiles can help us understand people's travel behavior, such as their preferred modes of transportation, travel patterns, and timings. This can help city planners and transportation authorities optimize public transportation systems, reduce congestion, and improve overall mobility. It can also be used by businesses to target customers based on their travel patterns and preferences.

Our research work addresses the challenge of alleviating congestion in urban mobility by proposing new temporalities for city users, aiming to optimize their departure times to workplaces. To achieve this goal, in this chapter, we propose a methodology for identifying individuals' mobility profiles.

This methodology consists of three steps: data collection, profile identification, and the learning and prediction process. These are presented in Section 5.2.

In Section 5.3, we discuss the sociological criteria that have been established, defining the mobility profile classes. To provide a comprehensive understanding of each class, we illustrate them with two examples, highlighting the characteristics and behaviors associated with each category.

Section 5.4 shows how we apply supervised machine learning methods, the Multi-layer Perception Neural Networks (MLPANN) and K-nearest neighbors (KNN), to predict individual mobility profiles.

Section 5.5 investigates the application of unsupervised learning using the K-means algorithm to cluster individuals with similar responses.

### 5.2 . The methodology

The proposed methodology follows a three-step process, as illustrated in Figure 5.1.

It first involves gathering survey data from over 50 companies located in



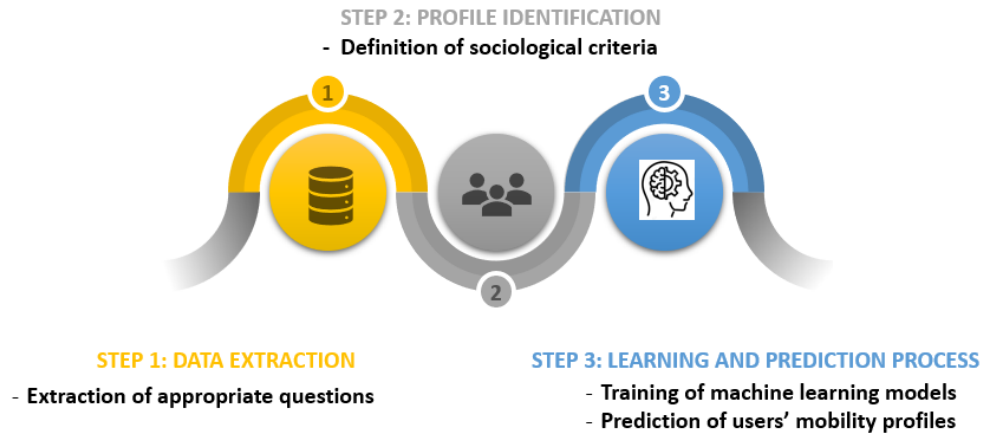


Figure 5.1: The 3-steps of the mobility profile identification methodology.

the suburbs of Grand Paris Sud, as discussed in detail in Chapter 4 (The decision platform). The survey data provides insights into employees' preferences and flexibility regarding their work schedules and commuting times. This step entails the extraction of the relevant questions from the questionnaire required for our processing.

In the second step of our methodology, we define three distinct classes of mobility profiles: "Favorable," "Favorable flexible," and "Favorable not flexible." These profile classes are predefined based on sociologically validated criteria. These criteria take into consideration factors such as personal circumstances, family obligations, and professional constraints that influence individuals' willingness and ability to adapt their arrival and departure times.

In the third and final step of the methodology, we employ machine learning techniques to analyze the survey data and predict the unknown profiles of employees according to the predefined classes.

Thus, the objective of our methodology is to enable the identification and prediction of individuals' mobility profiles, providing valuable and necessary information for our ultimate goal that is proposing new temporalities to alleviate congestion in urban mobility.

### 5.3 . Criteria for mobility profiles identification

Following the acquisition of data from the discussed questionnaire in Chapter 4 (The decision platform), which aims to gather individuals' opinions and expectations regarding their mobility and work organization, the second step of the methodology involves selecting appropriate questions from the questionnaire. We define a set of sociological criteria that rely on the sociological profile and the personal and professional constraints of the surveyed per-

sons extracted from their responses to the selected questions. Accordingly, we consider that a person can belong to one of the following profiles:

- *Unfavorable (UNF)*: this profile concerns persons who are happy with their current work organization and therefore do not want any changes.
- *Favorable flexible (F-F)*: those are employees who are in favor of changes in their current work organization and are flexible because they have no personal constraints.
- *Favorable not flexible (F-NF)*: this profile groups persons who favor changes but are not flexible because of personal constraints.

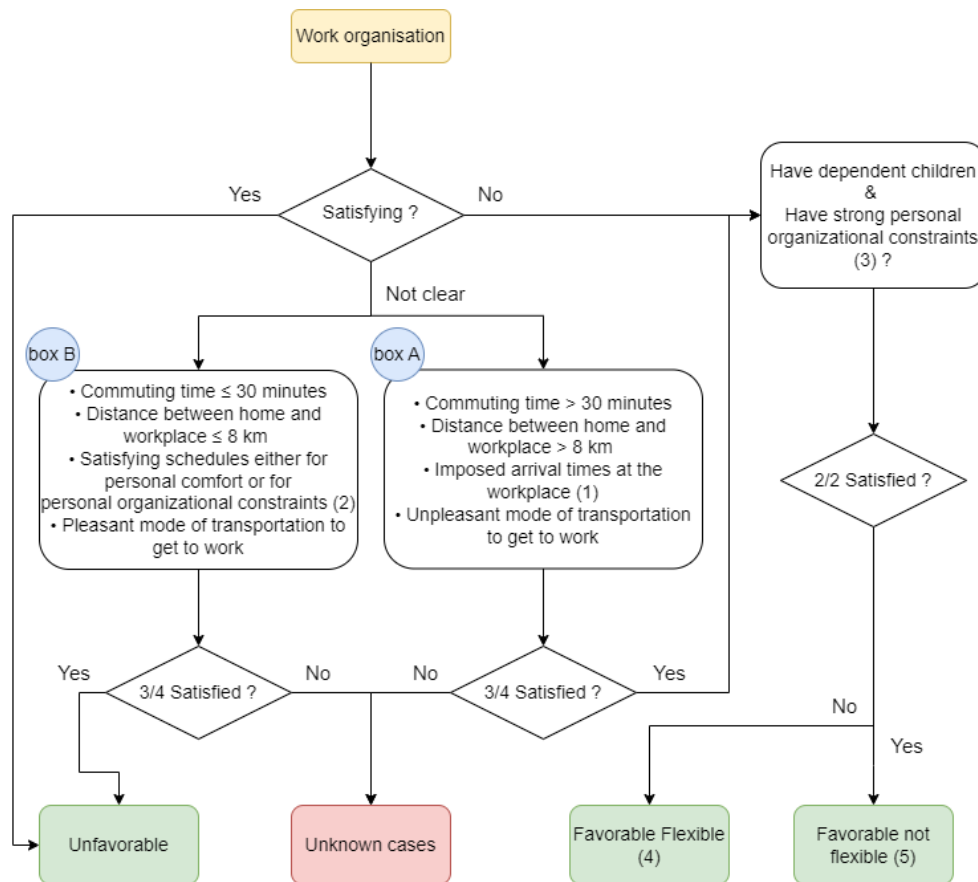
To be able to predict the profile of a person, we need to characterize each of the predefined profiles. For that, for each type of profile, we associate a set of appropriate criteria that have to be satisfied by a person to belong to this profile. The satisfaction of a criterion depends on the answer(s) given to one or more questions in the questionnaire.

Figure 5.2 summarizes the predefined profile characterization process using the criteria considered. As we can see, the first criterion we consider is *Work organization*. One of the questions a respondent should answer is "*When you return to your workplace, what would you like to change in your current work organization?*". To this question, the person must answer "*Nothing, the current organization suits me*" and "*No change desired in my work organization*" to consider that she/he is not in favor of changes in her/his work organization, that is the shift of her/his working hours. This person will be classified as *Unfavorable (UNF)*.

In the case where the person does not choose these answers, we need to investigate her/his answers to other questions in the questionnaire. This investigation allows us to check if the criteria provided in box *A* or box *B* are verified (see Figure 5.2). We consider that three out of four of them (3/4) must be satisfied to say whether the person is in favor (box *A*) or not (box *B*) for the shift of her/his working hours.

If a person is considered in favor of changes in her/his work organization, the second processing step consists of predicting whether this person is flexible or not. We set two criteria for that: *having dependent children* and *having strong personal organizational constraints*. If a person satisfies both criteria, then that person is considered as *Favorable not flexible (F-NF)*. In revenge, if, at least, one of these criteria is not satisfied, then the person is considered as *Favorable flexible (F-F)*.

The following are examples of conditions that correspond to different mobility profiles. To be considered to belong to a particular profile, an individual must meet at least one of the associated conditions. A condition is essentially a group of specific criteria that must be met.



- 1) Accompaniment of children, sport, shopping, medical constraints
- 2) Playing sports, shopping, work-life balance, picking up children at fixed times
- 3) Either by the company or by transport schedules or by transport saturation or traffic congestion
- 4) Is not a priority to meet their ideal time of arrival at workplace
- 5) Priority to meet their ideal time at the workplace

Figure 5.2: Predefined profile characterization

- Unfavorable:

1. Example1: To meet the first criterion, which is having a satisfactory work organization, the questioned individuals need to respond positively to the question *"Is your journey pleasant?"* with the answer "pleasant itinerary." They also need to answer negatively to the question *"Do you want any changes in your current work organization?"* with the response "No, the current work organization is satisfying for me." Additionally, if asked why they would like to change their mode of transportation, they should respond by saying *"There are no desired changes."* Lastly, if asked why they have not changed their current work organization, they should reply, *"I have no desired changes."*

2. Example2: One of the requirements we establish to classify someone as being in the *Unfavorable* category is that their commuting time must be 30 minutes or less, and the distance between their home and workplace must be 8 km or less. We verify whether these conditions are met by checking the response against the predetermined time and distance values. Additionally, we assess whether the individual satisfies the "satisfying schedules" criterion by asking them the question, *"What are the reasons that affect your arrival time at your workplace?"* their response must include *"personal comfort," "personal organizational constraints (e.g., accompanying children, sports, etc.), and "medical constraints."*

- Favorable flexible:

1. Example1: One of the criteria for belonging to this category is dissatisfaction with one's work organization and not having dependent children or strong personal organizational constraints. To determine whether someone meets these requirements, they must answer "unpleasant" when asked if their commute is enjoyable and provide specific factors that would make their commute more pleasant, such as minimum correspondences, low fare, comfortable mode of transportation, and low traffic. They must also answer whether they wish to change their work organization and provide reasons for doing so, such as reducing the number of weekly trips, improving work-life balance, increasing efficiency and productivity, optimizing their workday, and avoiding congestion. Lastly, they are requested to specify whether they have any dependent children, and the response must be "negative."

2. Example2: To belong to this category, one must have a commute time of more than 30 minutes, live more than 8 km away from work, have imposed arrival times, and not have dependent children or strong personal organizational constraints. To confirm whether the commute time and distance requirements are met, we compare the given answers to the predetermined time and distance values. When asked about the reasons for their arrival time at work, the individual must provide answers such as "company-imposed schedule," "transport schedules," "to avoid transportation saturation," or "to avoid road congestion."

- Favorable not flexible:

1. Example1: In order to be classified within this category, certain criteria must be met. Firstly, the individual must respond negatively to the question regarding the pleasantness of their com-

mute, and they must have a minimum commuting time. Secondly, when asked about the factors that make their journey pleasant, they must cite the minimum commuting time. Thirdly, when asked about the reasons for their arrival at work, they must provide answers such as working hours imposed by their company or transportation schedules. Fourthly, if asked why they want to change their mode of transport, they must cite personal organizational constraints, such as childcare, sports, or errands. Fifthly, if asked why they want to change their work organization, they must respond with factors such as improving their work-life balance or optimizing their work day with professional travel and meetings outside the office. Finally, if asked about the barriers preventing them from changing their work organization, they may mention personal organizational constraints, such as needing to pick up their children at fixed times.

2. Example2: The person must respond with "unpleasant itinerary" when asked about the pleasantness of their commute. The distance between their home and workplace must be over 8 km and take more than 30 minutes. When asked about the reasons for wanting to change their mode of transportation, the individual must cite strong personal constraints. Finally, when asked why they want to change their work organization, the individual must respond with a desire to optimize their work day with professional travel and meetings outside the offices.

Given that the questions corresponding to the criteria entries are multiple-choice, we can have two or more sets of answers belonging to two categories at the same time. For that we set a priority order as follows: *UNF*, *F-F*, *F-NF*. For example, if a person answers the question: *"Why do you want these changes in your work organization?"* by both statements *"I would be more efficient and productive"* which is a response corresponding to an *F-F* profile, and *"It will be good for my work-life balance"* which is a response which may correspond to an *F-NF* profile, we consider, using the priority order set, that this person has an *F-F* profile.

For the unknown cases, that is, persons who may not satisfy all the criteria and the conditions defined for each profile, we rely on the machine learning-based prediction approach to predict to which category or profile the persons in these cases may belong.

## **5.4 . Mobility profiles prediction using supervised learning models**

In the final phase of our approach, we use supervised machine learning techniques to predict unknown employee profiles based on the predefined categories. As explained before, these unknown profiles may arise from conflicting responses to the questionnaire that deviate from our established conditions.

To prepare our data for the learning process, we first clean and preprocess them. We initially gathered 1178 observations from the questionnaire, each representing an individual's complete responses. To ensure the quality of our research work, we removed 214 observations with missing responses, resulting in a total of 964 observations.

For preprocessing, considering that the questionnaire responses are independent, do not have ordinal relationships, and are non-gradual, we opt for a one-hot encoding approach as the coding scheme for input data in our machine learning models.

The input for the employed machine learning methods consists of 28 one-hot encoded entries, each corresponding to an individual's response to the selected questions.

### **5.4.1 . Multi-layer perception neural artificial network model**

Our multi-layer perception neural artificial network (MLPANN) consists of 28 nodes (corresponding to the hot encoded responses) in the input layer and one hidden layer of 42 nodes, determined through experimentation by taking the  $3/2$  the number of entries. Because of the number of classes we have defined (3), the output layer consists of 3 nodes.

This network has a total of 1.347 parameters, distributed as follows: 1.218 parameters at the level of the hidden layer and 129 parameters in the output layer. The number of parameters in the hidden layer is obtained by considering the connection weights between the first two layers ( $28 * 42$ ) plus 42, the bias values of the hidden layer.

Similarly, the number of parameters in the output layer is obtained by considering the connection weights between the last two layers ( $42 * 3$ ) and the bias values (3) of the output layer. Table 5.1 summarizes the parameters of our MLPANN model.

During the training phase, our dataset comprises 14.748.145 labeled cases from a total pool of 16.258.033, leaving 1.509.888 cases categorized as unknown, which are synthesized based on predefined criteria.

The case distribution includes 9.674.014 unfavorable cases, 4.998.042 cases categorized as favorable flexible, and 76.088 cases categorized as favorable not flexible, resulting in a comprehensive total of 16.258.032 cases.

Table 5.1: Parameters of the MLPANN

Layer(type)	Output shape	Parameters
Input layer	28	
Hidden layer	42	1218
Output layer	3	129
Total parameters: 1347		

Applying the MLPANN method on the survey data classifies 531 persons as *Favorable flexible (F-F)* and 396 persons as *Unfavorable (UNF)*. Only a tiny proportion was considered as *Favorable not flexible (F-NF)*. On the 964 surveyed persons, this represents 55%, 41%, and 4%, respectively.

In order to identify what are the factors that have an impact on this classification, we have analyzed these results considering gender, the socio-professional category, marital status, and having or not having small children. All the results of this analysis are detailed in Chapter 6 (Sociological interpretations).

#### 5.4.2 . K-nearest neighbors model

As described in Chapter 3 (Data analysis), the K-nearest neighbors method is a simple and effective method for classification but requires the setting of an appropriate value of parameter  $K$ . For that, we run the algorithm several times with different  $K$  values. This experimentation allowed us to select the best  $K$  value, that is the one that raises the smallest error value. According to this experimentation,  $K = 13$  is the best value to consider in our case. The error value as a function of  $K$  value is plotted in Figure 5.3.

The processing delivers for the KNN method 552 *F-F* cases out of 964, against 403 *UNF* cases, and a tiny portion of favorable non flexible cases representing: 57%, 42%, and 1%.

#### 5.4.3 . Comparison between the MLPANN and KNN methods

The aim of this part is first to discuss the obtained results after applying KNN method and then to compare them to the MLPANN results. The results of these methods can be compared considering two different aspects: the technical aspect and sociological interpretation. In this chapter, we focus only on the technical aspect. The comparison in terms of

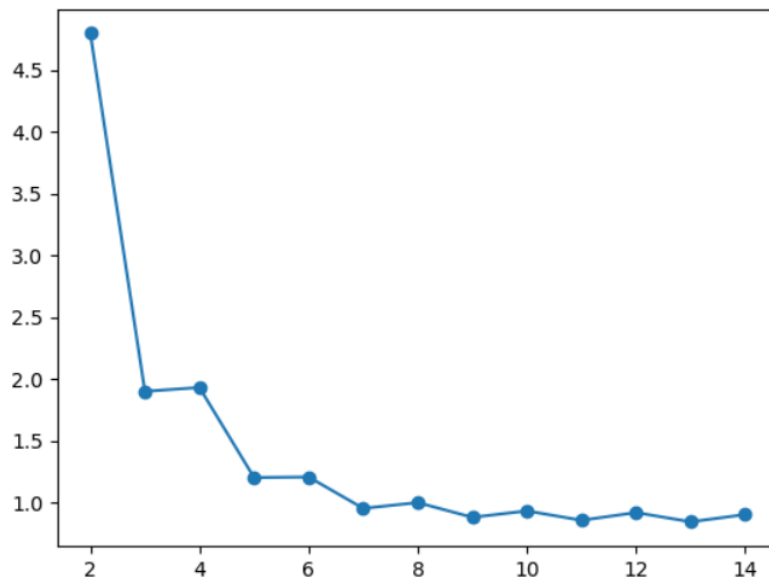


Figure 5.3: Error value as a function of the parameter K value.

sociological criteria is presented in Chapter 6 (Sociological interpretations).

As we can see in Table 5.2 the classification results obtained with the two methods are very similar except for the F-NF profile. We maintain 97% similarity and 3% difference according to profiles.

Table 5.2: Comparison between KNN and ANN profiles

Profiles	ANN vs KNN similarities (%)
<i>UNF</i>	98,26%
<i>F-F</i>	96,19%
<i>F-NF</i>	24,32%

However, the implementation of the two supervised methods shows some differences between them. Several conclusions can be drawn:

- The calculation time for the KNN learning process is less than MLPANN because the mathematical computations involved are not difficult.
- KNN takes more time to make predictions compared to MLPANN as it doesn't learn from the training data; it simply memorizes it and then uses that data to classify the new input.



- Using 10000 random test observations, the two algorithms are evaluated by calculating their Accuracy, precision, and recall from their respective confusion matrices (see Figure 5.4).

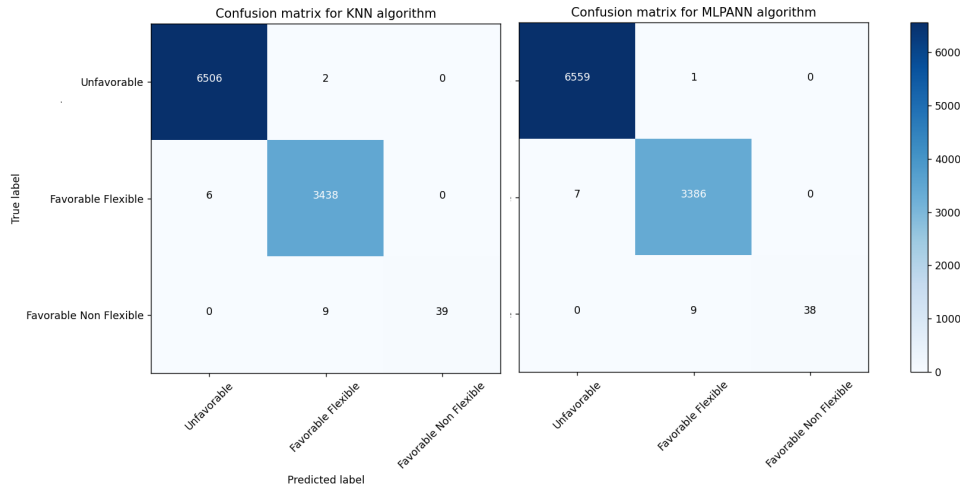


Figure 5.4: Confusion matrices for KNN and MLPANN.

- The two algorithms give a very high accuracy (99,47% for MLPANN and 99,8% for KNN), see Table 5.3.

Table 5.3: Evaluation metrics for MLPANN and KNN

	MLPANN			KNN		
	Prec.	Recall	Acc.	Prec.	Recall	Acc.
UNF	0.9998	1	0.9973	0.9997	1	0.9997
F-F	0.9979	1	0.9973	0.9983	1	0.9991
F-NF	1	0.8085	0.9957	0.8125	1	0.9981

The models have clearly achieved a remarkably high level of precision, which can be attributed to the vast number of known cases with approximately 90% of the cases we have at our disposal. These models are unbiased, and we confirm this by doing a cross-validation for both methods.

However, for the F-NF profile, we observe in all comparison metrics a difference between the results of the two methods. Given the minor representation of the *F-NF* profile within the dataset (76.088 out of 16.258.032 cases representing 0,000468%), the substantial dissimilarity in the percentages (see Figure 5.5) for this profile becomes explicable.

Indeed, this difference can be attributed to the limited sample size affecting both KNN and MLPANN's ability to learn the distinctive features

of the *F-NF* profile, justified by KNN's sensitivity to sparse neighbors and MLPANN's potential overfitting due to the neural network's complexity [153].

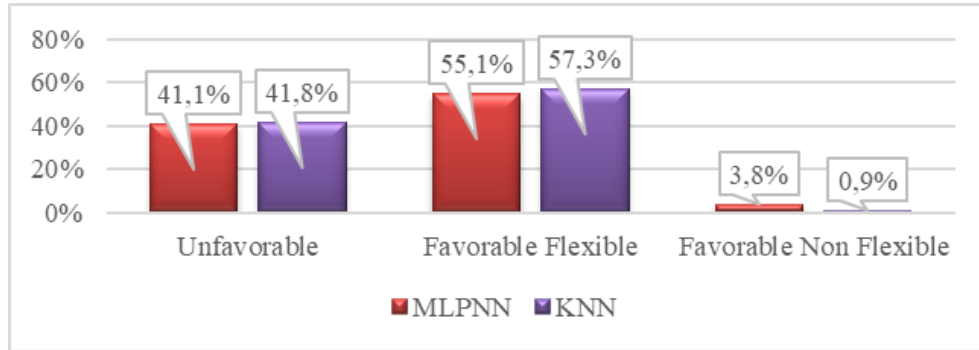


Figure 5.5: Supervised learning techniques results.

To address this imbalance, several strategies exist. One of them is the ensemble method. They are techniques in machine learning where multiple individual models (base models) are combined to create a stronger and more accurate predictive model. The idea is that by aggregating the predictions of multiple models, the ensemble can compensate for the weaknesses of individual models, leading to improved overall performance and generalization [30].

## 5.5 . Mobility profiles clustering using K-means model

Within this section, we use the K-means unsupervised machine learning technique to partition individuals exhibiting a strong correlation. Afterward, we thoroughly examine the contents of each cluster to identify the predominant characteristics that bring together each distinct group of individuals (refer to chapter Sociological interpretations). The K-means algorithm consists of three main steps, selecting the number  $K$  of clusters, computing the distance between each point from selected cluster centers using a specific metric, and grouping based on the minimum distance (see Figure 5.6).

### 5.5.1 . Optimal number of clusters

For our study, we found a value of  $K = 3$  when applying both the Elbow function and Silhouette score. Figure 5.7 and Figure 5.8 show that  $K = 3$  is the optimal number of clusters using both methods. Three clusters emerge, aligning with the predefined categories.

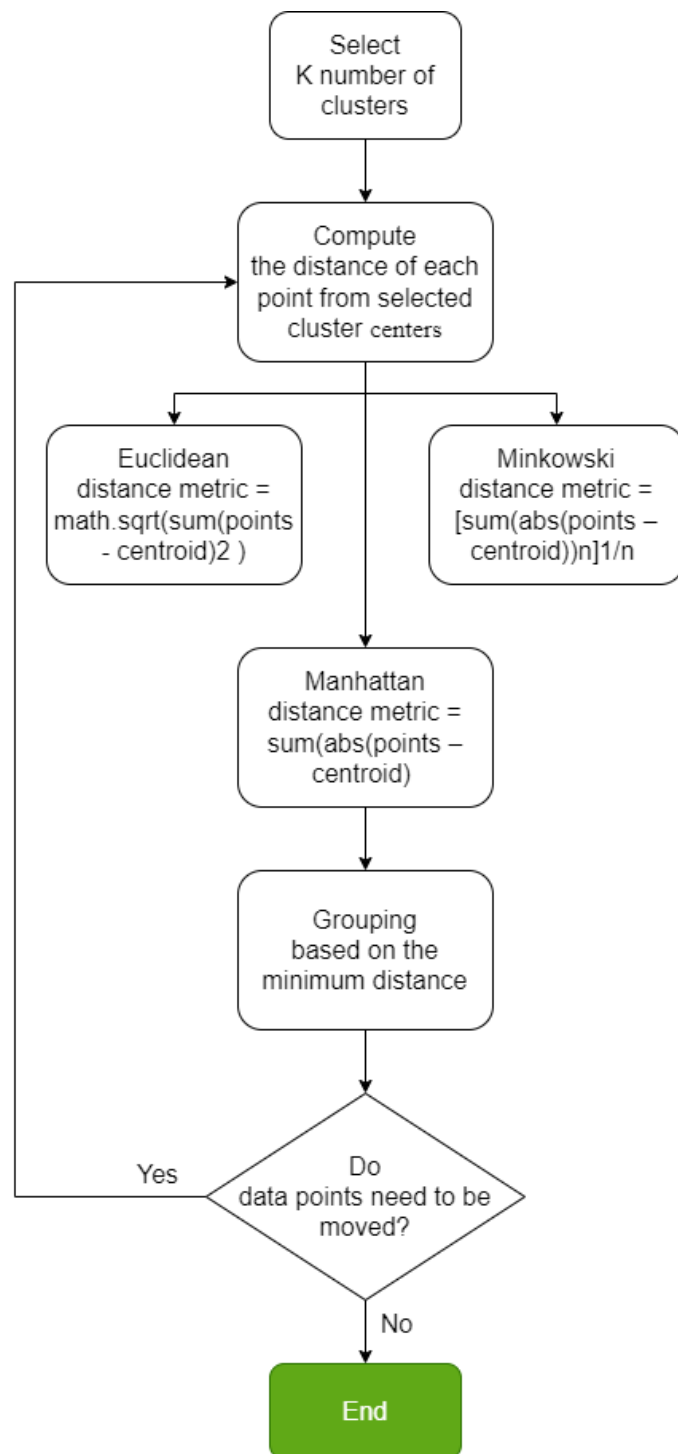


Figure 5.6: Algorithm K-means steps.

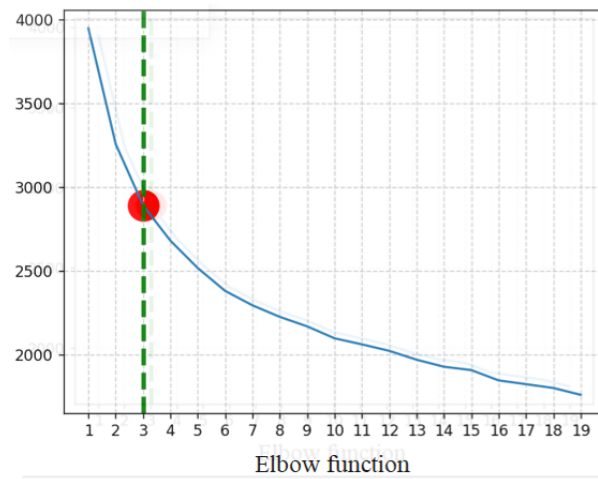


Figure 5.7: Elbow function results.

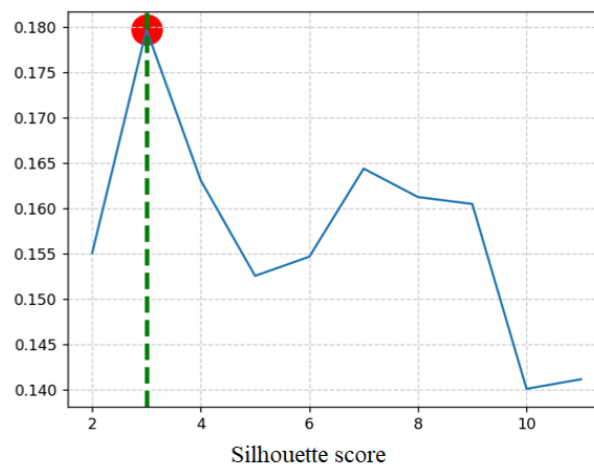


Figure 5.8: Silhouette score results.

Once the number of clusters was fixed to  $K=3$ , the results on different distance metrics were evaluated. Many runs were carried out for each case to achieve perfect results. The distortion and the number of iterations were accounted for, and the running times were measured.

Cluster distortion, also known as intra-cluster distance or intra-cluster variation, is a measure of how tightly the data points within each cluster are packed together. It quantifies the similarity or dissimilarity among the data points within a cluster. A lower cluster distortion indicates that the points within each cluster are closely related, while a higher distortion suggests greater variability within the cluster [66].

The number of iterations needed to achieve stability in clustering depends on the algorithm being used and the nature of the dataset. In general, clustering algorithms aim to optimize an objective function, which is called a convergence function, such as minimizing cluster distortion. The iterations involve an iterative process of updating the cluster assignments and optimizing the objective function until convergence or stability is reached [77].

The clusters' repartition obtained using Euclidean distance, Manhattan distance, and Minkowski distance metrics are shown in Figure 5.9. The plot shows that Euclidean and Minkowski distances give exactly the same clusters' repartition. Even if the results of the K-means with Manhattan are different, the difference is very small.

All the experiments incorporate the questionnaire data. Figure 5.10 displays the comparative graph of distortion, running time, and the number of iterations needed to achieve stability inside a cluster for all the distance metrics used.

The comparative figure shows that the K-means using Euclidian distance metric gives better results than with Manhattan K-means: with both distance metrics, the clusters have almost the same values (number of items), but the distortion in Manhattan K-means is greater than in Euclidian K-means. Figure 5.10 also shows the K-means algorithm results using the Minkowski distance metric. As the value of  $P$  increases, the value of the distortion oscillates, and the results start converging at  $P \geq 10$  with the same distortion as for  $P=3$ . The Minkowski distance metric gives similar results for larger values of  $P$  but converges slowly compared to the Euclidian K-means algorithm. Thus, in terms of performance, Euclidian K-means gives the best results.

The performances of K-means approach with the three metrics are also measured and compared in terms of execution time using the same dataset. These experiments show that with Manhattan distance it per-

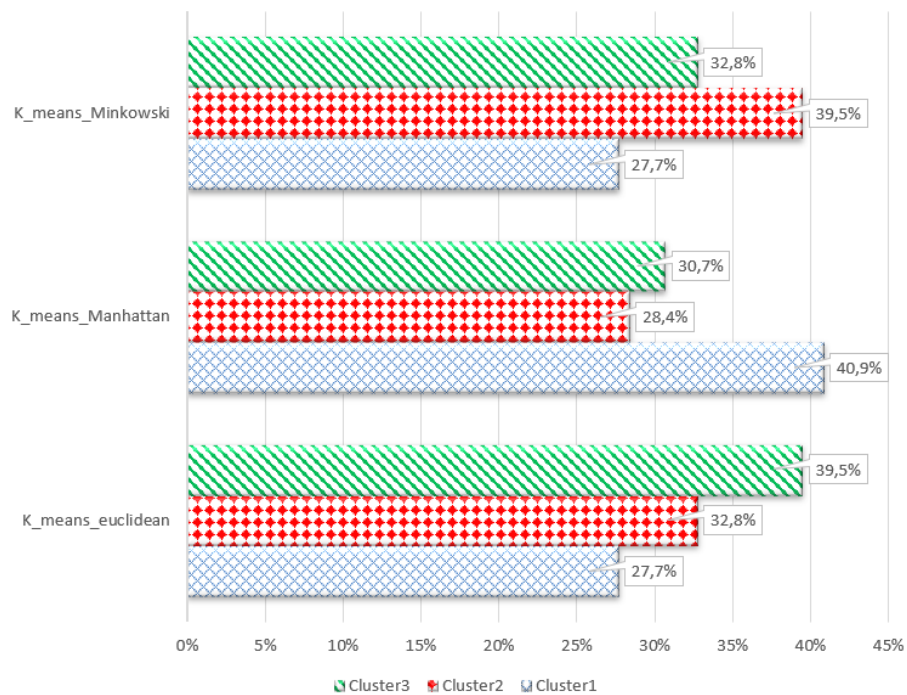


Figure 5.9: K-means results using different distance metrics.

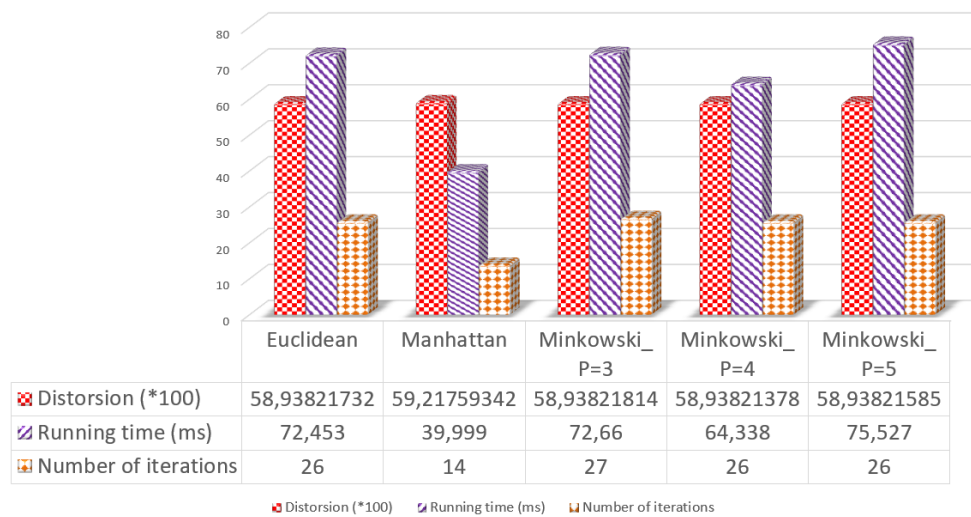


Figure 5.10: Comparative graphs of distortion, running time, and number of iterations for the different metrics used.

forms better, while Minkowski distance performs better for  $P=4$  compared to Euclidian K-means.

These results show that the distance metrics play a critical role in clustering when these metrics are implemented and should therefore be selected carefully.

As explained before, the comparison between the supervised and unsupervised techniques results in terms of sociological criteria are discussed in Chapter 6 (Sociological interpretations).

## **5.6 . Conclusion**

This chapter has presented a methodology for identifying and predicting mobility profiles of individuals in order to address traffic congestion in urban areas. The methodology involves gathering survey data from companies and employing machine learning techniques for analysis and prediction. Three distinct mobility profiles, namely "Unfavorable," "Favorable flexible," and "Favorable not flexible," were defined based on sociological criteria.

The application of the supervised learning techniques, Multi-layer Perception Neural Network (MLPANN), and K-nearest neighbors (KNN) resulted in a classification of individuals into the predefined mobility profiles. The MLPANN method classified the majority as Favorable flexible (55, 1%), followed by Unfavorable (41, 1%), and a small proportion as Favorable not flexible (3, 8%). Similarly, the KNN method yielded a similar distribution with 57, 3% Favorable flexible, 41, 8% Unfavorable, and 0, 9% Favorable not flexible.

A comparison between the MLPANN and KNN methods demonstrated a high level of similarity (97%) in the classification results across the profiles. Both methods exhibited high accuracy and achieved precision and recall values close to 1, indicating their effectiveness in predicting mobility profiles. The MLPANN method was found to be faster in making predictions, while KNN excelled in the learning process.

Furthermore, this chapter explored the application of the unsupervised learning technique K-means algorithm for clustering individuals based on their mobility profiles. The optimal number of clusters was determined to be three, and the Euclidean distance metric provided the best results compared to the other distance metrics used.

In summary, the methodology presented in this chapter offers a comprehensive approach to identify, predict, and cluster mobility profiles of individuals.

In the next chapter, we interpret the sociological impact on the classification we made.





## 6 - Sociological interpretations

### 6.1 . Introduction

Mobility profile prediction, as it is raised in our work, involves using machine learning algorithms to predict people's willingness and flexibility to adapt their working hours based on sociological criteria. To identify the factors influencing our predefined classification (as discussed in the previous Chapter), we analyze individuals' sociological circumstances within each mobility profile category, employing various learning techniques. We consider factors such as gender, socio-professional category, marital status, and the presence of small children. By conducting a cross-analysis of these sociological factors, we understand why certain individuals are more inclined to change their working hours to alleviate congestion, as well as their level of flexibility in doing so. Additionally, we examine the differences in predictions made by different machine learning techniques.

The chapter is structured as follows: firstly, we discuss aligning the questionnaire data with the actual population, comparing it to the enumeration given by INSEE (Institut National de la Statistique et des Etudes Economiques) [5]. Then, we explore the impact of different sociological criteria on the obtained results of each learning technique used.

### 6.2 . Data and census

After enumerating the population under investigation, we conduct a comparative analysis between this population and the census data obtained from the residential municipality provided by INSEE. Our research work focuses on the residential departments of Paris, Seine et Marne, Essonne, Hauts de Seine, and Val de Marne. The reason for including these departments is that they correspond to the locations where the surveyed individuals live (see Chapter 4: The decision platform). An analysis that compares the census data obtained from INSEE with the data collected from our sample for each of these departments covers various demographic factors, including gender, age group, socio-professional category, marital status, and children in care.

The INSEE in France classifies individuals into different socio-professional categories based on their occupation and socio-economic status. Here is a description of each category as defined by INSEE [4]:

1. **Agriculteurs exploitants (Farmers):** This category includes individuals who own and operate agricultural holdings. They engage in activities related to farming, animal husbandry, and the cultivation of crops.
2. **Artisans, commerçants, chefs d'entreprise (Craftsmen, traders, business leaders):** This category comprises self-employed individuals who run small businesses, such as craftsmen, shopkeepers, and entrepreneurs. They are engaged in various trades and commercial activities.
3. **Cadres et professions intellectuelles supérieures (Executives and higher intellectual professions):** This category includes professionals with higher-level responsibilities and intellectual work. It encompasses executives, managers, senior professionals, and high-level employees in various sectors.
4. **Professions intermédiaires (Intermediate professions):** This category includes individuals in intermediate-level occupations that require specialized skills and knowledge. It includes technicians, supervisors, office managers, and other skilled workers.
5. **Employés (Employees):** This category consists of individuals who work under the supervision of others, performing tasks that do not require specialized qualifications. It includes administrative staff, salespersons, and other similar roles.
6. **Ouvriers (Workers):** This category comprises manual laborers and workers engaged in physical tasks that require less specialized skills. It includes construction workers, factory workers, machine operators, and laborers in various industries.
7. **Retraités (Retirees):** This category includes individuals who have reached the legal retirement age and are no longer actively employed. They receive pension benefits and are no longer part of the active workforce.
8. **Autres personnes sans activité professionnelle (Other persons without professional activity):** This category includes individuals who do not fall into any of the above categories. It encompasses students, individuals without paid employment, unemployed individuals actively seeking jobs, and those who are not part of the labor force.

Each category is assigned corresponding values that indicate its representation within the population (see Table 6.1).

Table 6.1: Census of the surveyed population

Variable	Category	Count	Percentage (%) (out of total)
Gender	Men	461	47.8
	Women	503	52.2
SPC	2	1	0.1
	3	557	57.7
	4	172	17.9
	5	198	20.6
	6	24	2.5
	8	12	1.2
Marital status	Married	791	82.0
	Single	173	18.0
Dependent children	No	401	41.6
	Yes	563	58.4
Age	<30	117	12.2
	30-40	278	28.9
	40-50	286	29.7
	50-60	254	26.4
	>60	30	3.1

Figure 6.1, presents the Cleveland diagram visually demonstrating the representative proportionality of the sample. For more details, please consult the table provided in ??.

Our sample becomes more realistic as the proportions approach 1. If the proportion exceeds 1, the category is over-represented, whereas if it falls below 1, the category is underrepresented. The plot indicates that the proportions for gender and dependent children are close to 1. However, individuals between the ages of 30 and 60, married individuals, and those belonging to the socio-professional category (SPC) 3 are over-represented. Conversely, all other categories are underrepresented. We can, therefore, conclude that our research concerns a limited and biased subset of the overall population under comparison. The respondents in our survey tend to be older, more highly educated, and married.

### 6.3 . Supervised techniques results

As seen in Chapter 5 (Mobility profiles identification and prediction) applying MLPANN method on the survey data classifies 531 individuals as

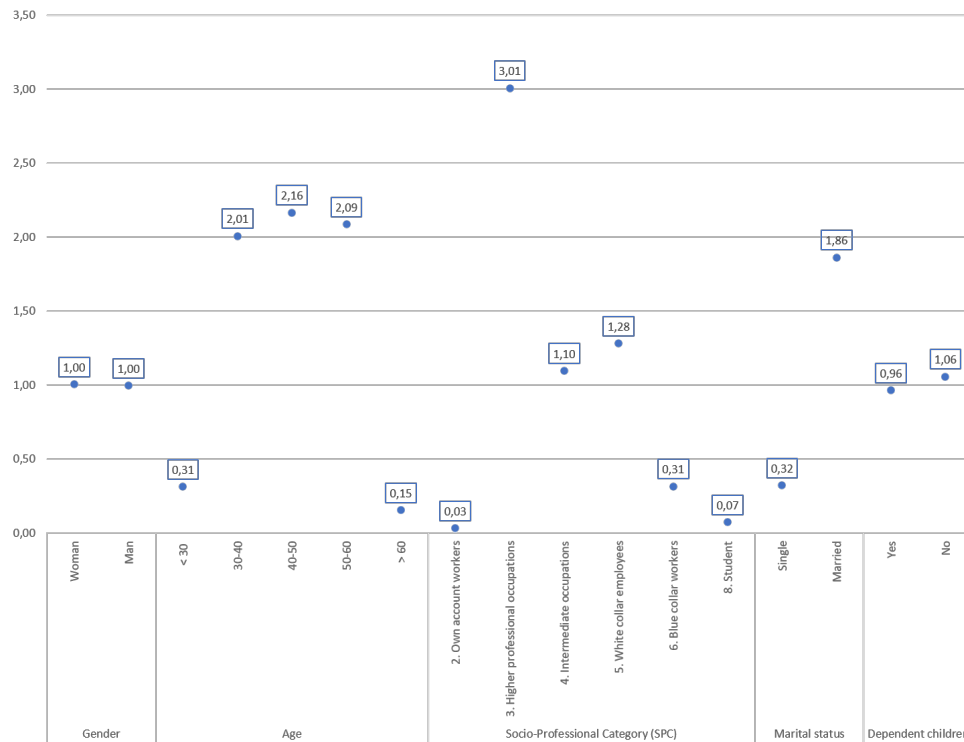


Figure 6.1: Representation of the surveyed persons within the real population.

Favorable flexible (F-F) and 396 individuals as Unfavorable (UNF). Only a tiny proportion was considered as Favorable not flexible (F-NF). Of the 964 surveyed individuals, this represents 55,1%, 41,1%, and 3,8%, respectively. The results obtained with KNN method are very similar to those of MLPANN method as we obtain 552 F-F cases out of 964, against 403 UNF cases, and a tiny proportion of F-NF cases representing, respectively, 57,3%, 41,8%, and 0,9%.

In this section, our objective is first to analyze the obtained results after using MLPANN and KNN methods to predict mobility profiles, investigating the impact of the sociological factors on the outcomes derived from predicting an individual's mobility profile.

### 6.3.1 . The gender impact

The first analysis concerns the impact of gender (man, woman) on the classification of individuals. Table 6.2 provides an overview of the distribution of mobility profiles based on gender using supervised learning techniques. These are row percentages obtained according to gender for both MLPANN and KNN methods.

Upon examining the MLPANN model results, it is evident that both men

Table 6.2: MLPANN and KNN results according to gender

	MLPANN			KNN		
	UNF	F-F	F-NF	UNF	F-F	F-NF
Men	39.3%	56.5%	4.2%	40.0%	59.6%	0.4%
Women	42.9%	53.6%	3.4%	43.8%	54.8%	1.4%
Not mentioned	30.8%	61.5%	7.7%	30.8%	69.2%	0.0%

and women exhibit a considerable proportion of favorable and flexible preferences for adjusting their working hours to avoid traffic congestion. Indeed, among the surveyed men, approximately 40% are categorized as Unfavorable (UNF), while 56.48% belong to the Favorable Flexible (F-F) group. Similarly, for women, around 43% are labeled as UNF, and 53.62% are classified as F-F. Women are less flexible than men because their daily mobility, in particular mothers, has already been described as ambivalent given that they must ensure, in addition to work, trips related to domestic work. The daily mobility of parents is strongly gendered. Women's movements are largely linked to what could be called mobile domestic work which can consist of an activity in itself (accompaniment of children for example) or an activity with a view to carrying out another task (e.g., shopping or administrative procedures). [44].

Our research work's results suggest that gender has minimal impact on changing working hours preferences, as the profiles remain relatively consistent across genders according to both models, albeit with less flexibility observed among women.

The KNN model further supports these findings, with approximately 59% of men identified as F-F compared to 54% of women. Additionally, around 40% of men are classified as UNF compared to 43% of women. This indicates that men tend to be more receptive to change, while women who exhibit favorable preferences display slightly less flexibility, according to the KNN model.

Once again, gender demonstrates limited influence on preferences for modifying working hours, as the overall distribution of profiles (UNF, F-F, and F-NF) remains largely similar across genders in the KNN model, despite the slight reduction in flexibility for women.

In conclusion, both the MLPANN and KNN models consistently demonstrate a prevalent preference for changing working hours among individuals, irrespective of gender. The majority of respondents exhibit an F-F profile inclination toward modifying their start and end times of work to reduce traffic congestion. While gender has minimal influence

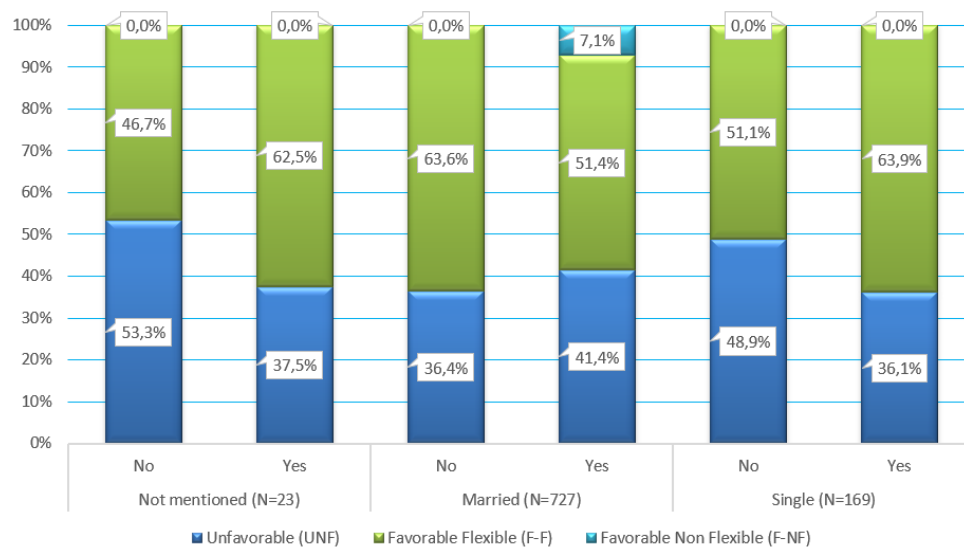


Figure 6.2: MLPANN results according to marital status and dependent children

on these preferences, it is important to consider the context in which women's mobility is shaped by additional responsibilities, such as domestic work. This may explain the relatively lower flexibility observed among women. Thus, the table results suggest that women's preferences for work-hour flexibility are influenced by other factors, such as personal constraints, work-life balance, and the obligation to accommodate various responsibilities. In a recent study [43], an analysis of travel purposes and changes provides an understanding of women's mobility. Comparing men's and women's travel purposes, a typology with four classes is constructed: leisure, work, social, and domestic trips. Over time, there is a trend towards equalization, with men dominating leisure trips and women over-represented in domestic trips. The increase in women's driving access has led to a redistribution of trips, highlighting the influence of personal constraints and responsibilities on women's mobility choices. This underscores the less flexibility about women's preferences for work-hour flexibility.

### 6.3.2 . Impact of the marital status and the presence of children

A second analysis concerns the impact of marital status and the fact of having dependent children on the classification of individuals.

Research indicates that the marital status, the presence of children, and the work-life balance play significant roles in shaping individuals' flexibility preferences. Studies have found that marital status can influence

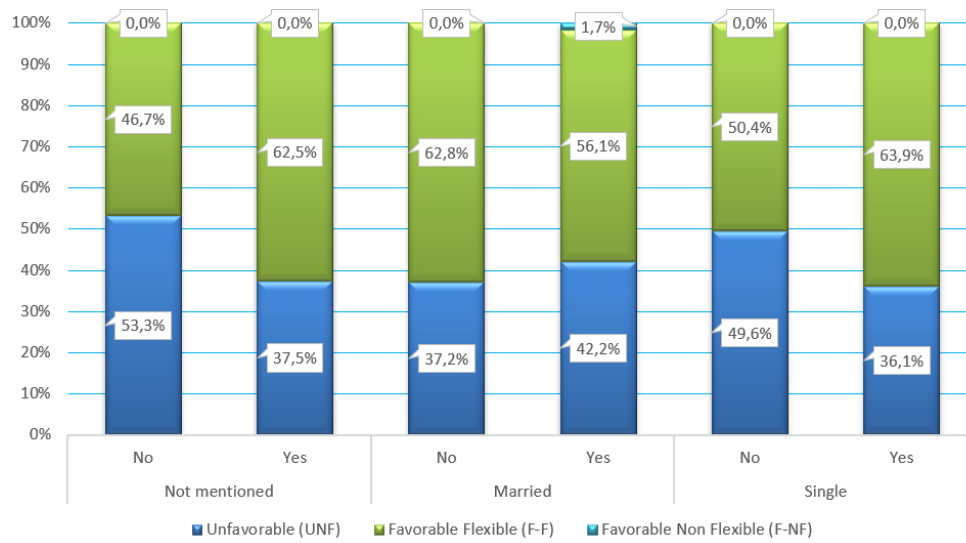


Figure 6.3: KNN results according to marital status and dependent children

flexibility choices, with married individuals often prioritizing stability in their work schedules [23]. Conversely, single individuals may exhibit greater flexibility preferences due to fewer familial responsibilities [53]. Moreover, the presence of children can significantly impact flexibility preferences, as individuals with children often seek greater flexibility to manage the demands of parenting and employment [55]. These studies highlight the multifaceted nature of flexibility preferences, which are shaped by marital status, the presence of children, and the need to balance work and family responsibilities.

Figure 6.2 and Figure 6.3 present the distribution of profiles we have obtained according to family situations (living with a partner or not) and whether they have dependent children or not using MLPANN and KNN models results. The y-axis displays the percentages of people in a specific category, and the x-axis shows the marital situation of the person. We distinguish between persons having children or not with the first line of the bar (*no* for persons who do not have children, and *yes* for those who have).

The distribution of mobility profiles in relation to marital status and having dependent children reveals interesting dependencies. Among single individuals with no dependent children, 51% exhibit a favorable flexible preference, while 49% are unfavorable. In contrast, the category of favorable not flexible exclusively applies to individuals who are in a couple and have dependent children, as observed in both the MLPANN and KNN model results.

However, an intriguing finding emerges when examining single individ-



uals with children, as they display the highest proportion of favorable flexible profiles. This finding contradicts previous studies, but it can be attributed to the biased sample in this research work, where the age range of participants is primarily between 40 and 50, and their children are older and do not require significant care. This highlights the importance of considering sample characteristics when interpreting the results. On the other hand, we can conclude that travel choices are not purely individual arbitrations but involve all members of the household; this is why single people are more often favorable.

### **6.3.3 . Impact of the socio-professional category (SPC)**

The role of the socio-professional category is important in shaping individuals' judgments and behaviors. The ways in which individuals perceive and evaluate social phenomena, such as culture, taste, and lifestyle choices, are extremely correlated with the SPC. Individuals' social positions and socio-professional categories heavily influence their dispositions and preferences [27].

One of the studies cited in the literature dives into the impact of the socio-professional category on the flexibility of individuals by employing factor analysis of correspondences. This technique is a graphical representation of the correlations between social positions, measured by classifying professions and socio-professional categories into 24 groups and work schedules. The analysis aims to create a map of temporal inequalities.

The first dimension identified in this analysis, represented vertically in Figure 6.4 [95], highlights a contrast between self-employed workers and service sector employees, who adhere to new work schedule norms, and employees whose workdays align with the traditional industrial organization. The second dimension, represented horizontally, examines the role of social hierarchy within these different socio-economic worlds. Occupying higher positions typically corresponds to standard working hours, with average durations increasing for liberal professions and executives. In contrast, individuals in lower social positions experience non-standard working hours characterized by shifts, fragmented schedules, and shorter durations.

The specific form of non-standard working hours varies depending on the nature of the work. Industrial workers, for instance, often have "3x8" shift schedules, while service sector workers may have afternoon and evening shifts that are fragmented and reduced [95].

The impact of the socio-professional category on individuals' experiences of balancing work and personal life is a challenging issue faced

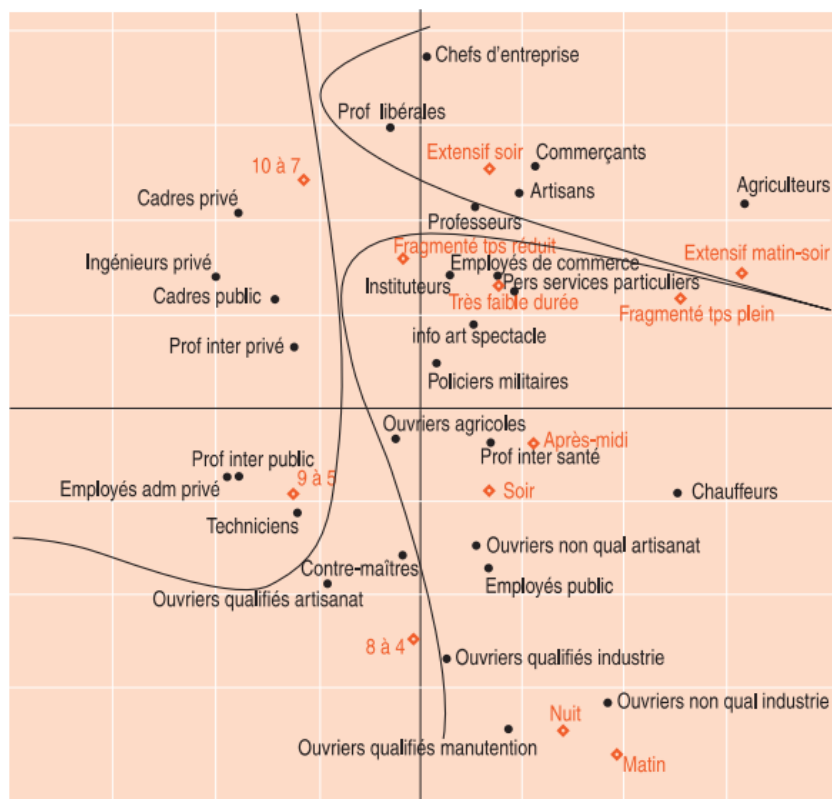


Figure 6.4: Social hierarchy and regularity of work schedules

by some SPC groups, considering factors such as working hours, job autonomy, and family responsibilities [38]. The literature suggests that different socio-professional categories may have varying levels of access to flexible work arrangements and resources, which can influence individuals' preferences for flexibility. Understanding the role of the socio-professional category is crucial for designing effective policies and practices that promote work-life integration and accommodate diverse needs across different professional contexts.

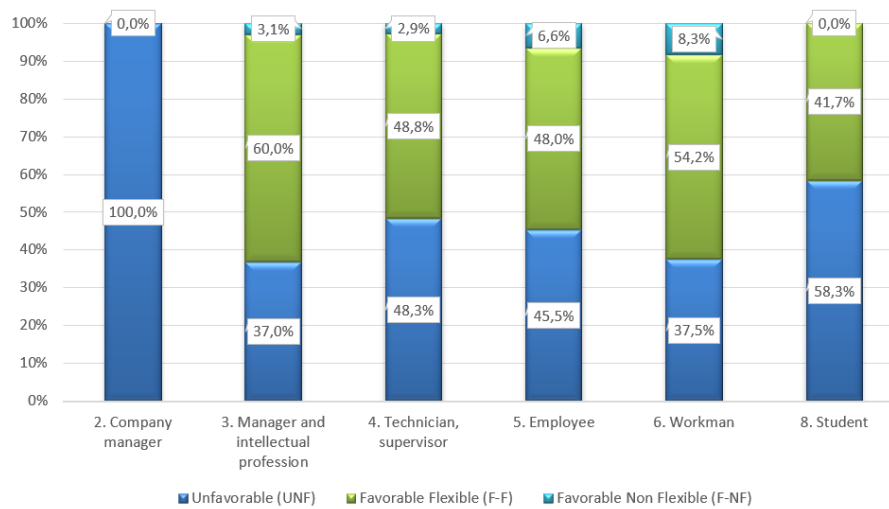


Figure 6.5: MLPANN results according to the socio-professional category

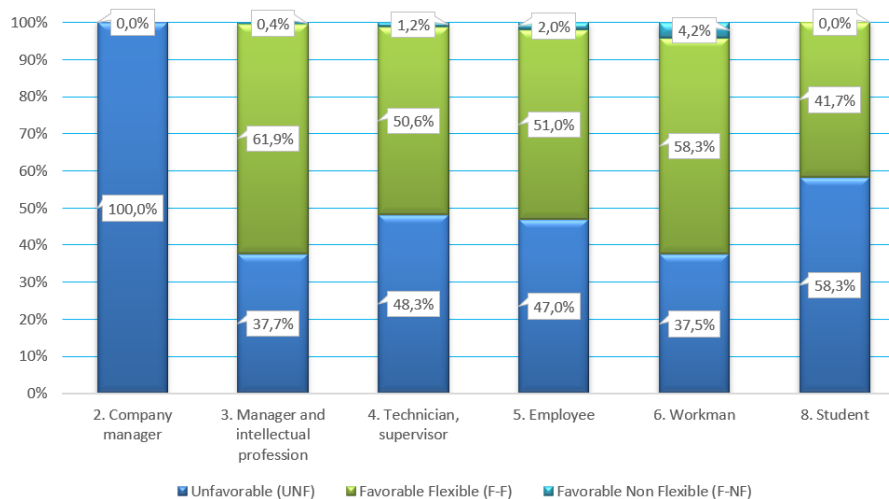


Figure 6.6: KNN results according to the socio-professional category

The socio-professional category, as it is raised in our work, has a notable impact on individuals' flexibility and willingness to shift their work-

ing hours to reduce traffic congestion. Figure 6.5 and Figure 6.6, which depict the distribution of profiles based on different socio-professional categories. Building on previous research, these figures show the relationship between the SPC and flexibility preferences.

Among the surveyed individuals in our research work, a significant majority (57.78%) belong to the "Manager, intellectual professions" category according to the INSEE classification in France. Within this category, most individuals (59.96%) are classified as F-F, indicating their willingness to change their working hours. In contrast, a smaller proportion of individuals (3.05%) fall into the F-NF category, while 36.98% are categorized as UNF to change their working hours. The "Employees" category demonstrates a more balanced distribution, with 47.98% classified as F-F and 45.45% as UNF. Interestingly, most students express an unfavorable position towards changing their start and end working hours. These findings indicate that individuals in managerial and intellectual professions exhibit a higher propensity for flexibility, whereas students tend to resist such changes. The obtained findings support the former studies cited.

#### **6.4 . Unsupervised technique results**

In this section, we investigate the results achieved using the unsupervised learning technique K-means with Euclidean distance which gave the best results (see Chapter 5: Mobility profiles identification and prediction). By employing this method on the target population, the results give three distinct subcategories (Cluster1, Cluster2, Cluster3).

Table 6.3 presents the content of the three clusters according to sociological criteria of gender, socio-professional category (SPC), marital status, dependent children, and age.

The table shows that according to gender, the clusters have an equitable repartition. Regarding the socio-professional category, SPC 2 (Craftsmen, traders, business leaders) is not very present in the sample and is in Cluster2. The most present SPC in the sample is SPC 3 (Executives and higher intellectual professions). Cluster1 includes mainly people who are F-F, and we see in this table that they are also people without dependent children, even though there are 19.55% married. Cluster2 includes a large majority of UNF people, and we see on the table that this concerns SPC 2, 5, 6 compared with the other clusters (1, 3). Cluster2 contains more married people with dependent children. Cluster3 retains the majority of people F-F. Unlike Cluster1, Cluster3 includes people belonging to SPC3 and those married with dependent children. It is

Table 6.3: Clusters' repartition according to sociological criteria.

		Cluster1	Cluster2	Cluster3
Gender	Men	53.8%	46.0%	45.2%
	Women	46.2%	54.0%	54.8%
SPC	2	0	0.3%	0
	3	68.2%	41.8%	63.8%
	4	15.4%	21.8%	16.3%
	5	14.2%	28.2%	18.6%
	6	0.4%	5.7%	1.3%
	8	1.9%	2.2%	0
Marital status	Married	70.8%	78.8%	92.5%
	Single	29.2%	21.2%	7.5%
Small children	No	99.2%	42.7%	0.3%
	Yes	0.8%	57.3%	99.8%
Age	<30	26.0%	13.0%	1.6%
	30-40	29.1%	24.1%	32.5%
	40-50	12.1%	28.3%	43.1%
	50-60	27.9%	29.8%	22.2%
	>60	4.9%	4.8%	0.5%

also noticeable that this Cluster contains more women than men.

The category of F-NF person is not very present, but it is found more present in Cluster2 with low rates.

For the age, Table 6.3 shows that people between 30 and 60 are the most present in the sample. In Cluster1, there are more people between 30 and 40. In Cluster2 are people between 50 and 60, and in Cluster3, people between 40 and 50.

## 6.5 . Supervised techniques vs unsupervised approach

A preliminary evaluation compares the outcomes of K-means with the distribution of mobility profiles obtained from the implementation of MLPANN and KNN. Figure 6.7 displays the composition of each cluster and their respective mobility profiles as determined by the MLPANN and KNN algorithms.

Cluster 1 and Cluster 3 collectively encompass a majority of individuals belonging to the F-F category. This suggests that these clusters consist of individuals who exhibit mobility profiles indicating a preference for flexible schedule options. On the other hand, Cluster 2 is predominantly composed of individuals categorized as UNF, specifically those who do

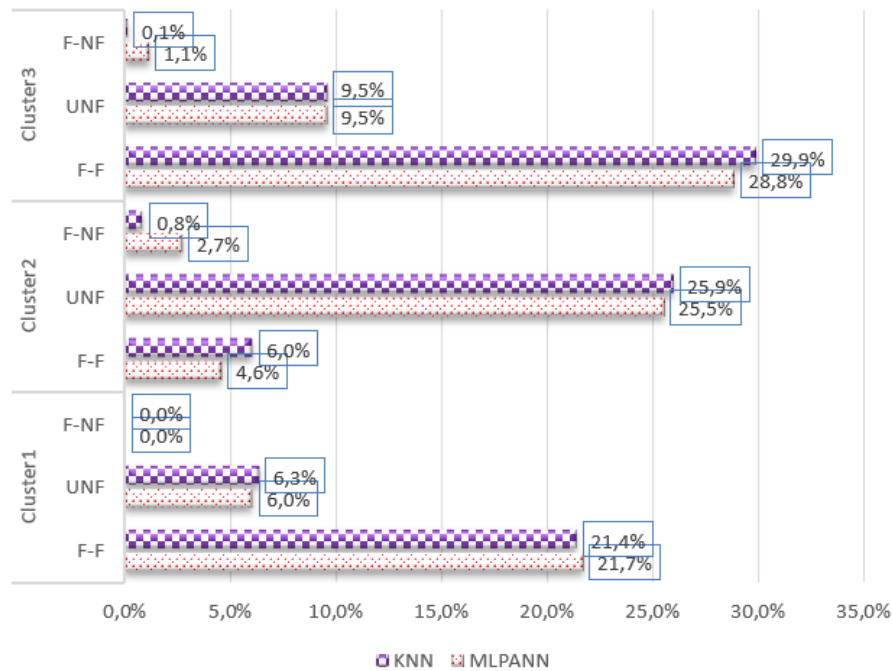


Figure 6.7: Cluster's repartition according to K-means compared to KNN and MPLANN classification.

not express any wish to change their working hours because they are happy with their work organization. Notably, almost all individuals in Cluster 2 fall under this UNF profile.

The comparison between the MLPANN and KNN results shows a remarkable similarity, demonstrating the reliability of both methods in identifying and categorizing individuals based on their mobility profiles. Such information can support the development of targeted transportation strategies and policies to address the varying needs and constraints of different mobility profile groups.

The clustering analysis based on the predicted mobility profiles reveals distinct patterns among the surveyed individuals. The obtained findings can contribute to informing decision-making processes in transportation planning and policy development, aiming to enhance mobility options and address the specific needs of different mobility profile groups.

## 6.6 . Conclusion

This chapter presents the prediction of mobility profiles using machine learning algorithms, specifically supervised techniques (MLPANN and KNN) and an unsupervised approach (K-means). The analysis reveals critical sociological factors influencing individuals' willingness to adapt working hours. Despite the limited representativeness of the sample, the analysis shows the impact of marital status, dependent children, and socio-professional categories on individual flexibility. Supervised techniques indicate that while gender has minimal influence, married individuals and those with dependent children are less flexible. Socio-professional categories also play a crucial role, with managers and intellectual professionals showing a higher tendency for flexibility. Unsupervised clustering categorizes individuals into distinct groups based on similarity among the individuals' responses to the questionnaire. The comparison of the results of supervised and unsupervised methods reinforces the reliability of understanding mobility behaviors. For policymakers and planners, the understanding gained from this analysis shows guidance in adapting transportation strategies to meet the needs of different population segments.

In the next chapter, we present the developed methodology to find the optimal departure time to avoid urban congestion based on traffic analysis.

## **7 - Optimal traffic-free departure forecast**

### **7.1 . Introduction**

Our thesis aims to address the issue of urban mobility congestion during rush hours in urban areas by conducting a traffic analysis. Our primary objective is to minimize urban mobility congestion by organizing time, thus, adjusting the departure time for commuting for individuals who are in favor of changing their working hours to reduce traffic congestion. To achieve this, we use a multi-step approach. This approach involves generating route names and corresponding distances from the Open Street Map (OSM) database. To enhance traffic congestion analysis, space/time traffic jam data is analyzed. Time series models are employed to predict congestion levels during peak hours by analyzing historical traffic data. The focus then shifts to generating temporality scenarios using the time series model to predict congestion levels across routes. The goal is to determine the most optimal departure time, minimizing time spent in traffic by exploring various scenarios and identifying the best option.

The chapter is structured as follows: First, we present the methodology. The second part of the chapter is an illustrative application example of the developed approach.

### **7.2 . The methodology**

The methodology is dedicated to investigating traffic jam times and adapting the departure time to go to the workplace to reduce urban mobility congestion. Initially, our approach involved collecting a list of route names from the Open Street Map (OSM) database and the distance traveled within each route.

To have a robust space/time traffic congestion analysis, we leverage space/time traffic jam data. These data are obtained through a collaboration between Waze-connected citizens and Grand Paris Sud urban community, ensuring an accurate representation of the traffic patterns. By utilizing this dataset, we can capture the situation of the traffic within the city.

To predict congestion levels for each route during the peak hours, we use time series models in the training process. These models analyze



historical traffic data to identify patterns and trends, enabling us to forecast congestion levels. By understanding the congestion levels on different routes at different times, we gain insights into the traffic patterns during rush hours.

Finally, we focus on generating temporality scenarios to optimize the sum of congestion levels across the various routes. We aim to determine the most optimal departure time that minimizes time spent in traffic. To accomplish this, we use the Prophet package. This package enables us to predict congestion levels at different times. This allows us to explore various departure time scenarios and identify the best option.

The approach is built upon seven steps (see Figure 7.1):

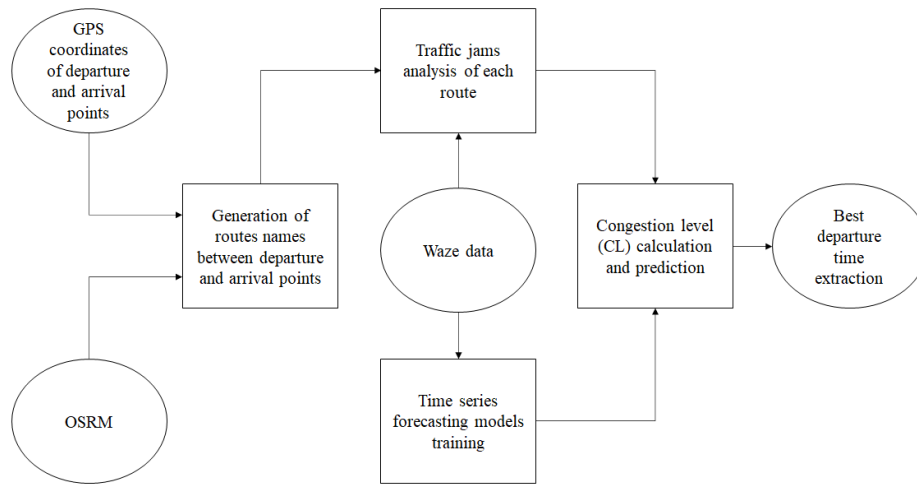


Figure 7.1: Methodology scheme.

- **Step 1: GPS coordinates.** It involves obtaining the Global Positioning System (GPS) coordinates. The user provides the coordinates of her/his departure and arrival points (notably home and workplace GPS coordinates).
- **Step 2: routes names generation.** Using the GPS coordinates provided by the user, this step aims to collect a list of routes from the Open Street Map (OSM) database by identifying the nodes that individuals pass through and generating the route's name based on the corresponding node. However, this approach does not produce the best results, particularly at intersections where some nodes might belong to a different road. Consequently, we

shifted to using the Open Street Routing Machine (OSRM) to directly gather geographical information about an individual's itinerary based on a distance optimizer. These routes represent the paths individuals traverse to reach their workplaces within the city. From this first processing, we formulate a predetermined itinerary with the route names and the traveled distance of each route.

- **Step 3: time series data forecasting.** Time series forecasting is a technique for predicting events through a time sequence. It predicts future events by analyzing past trends assuming that future trends will be similar to historical trends [51]. We consider the FB Prophet model, which is a machine learning developed by Facebook for forecasting time series data where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series with strong seasonal effects and several seasons of historical data [145].
- **Step 4: traffic jams analysis.** This step involves extracting data for each individual route from a set of routes and then processing that data. The impact of various parameters is taken into account in order to create a proof of concept for the proposed work. For this, we analyze the FB prophet components. As we have seen in Chapter 3 (Data analysis) the model has several components that can be customized to improve its prediction accuracy [145].
- **Step 5: congestion factor calculation and prediction.** The congestion factor at time  $k$ ,  $c(k)$  is defined as the ratio of extra travel time  $T(k)$  with respect to the free flow travel time  $T_f$  [121]:

$$c(k) = \frac{\Delta T(k)}{T_f} = \frac{T(k) - T_f}{T_f} \quad (7.1)$$

This step also predicts the congestion factor for different timings using the time series model.

- **Step 6: best departure time extraction.** The last step of the approach involves extracting the optimal departure time by finding the time at which the total congestion factor for all the routes the individual will take is the lowest possible. We refer to this optimal departure time as a "*temporality scenario*". Figure 7.2 shows that when traveling along route  $i$ , a person's expenditure of  $t_i$  is determined by multiplying the distance  $D_i$  covered along that route with the estimated speed at the corresponding time.

This process is based on Equation 7.2 and Equation 7.3, which outline how we determine the ideal departure time for minimizing traffic congestion during travel.

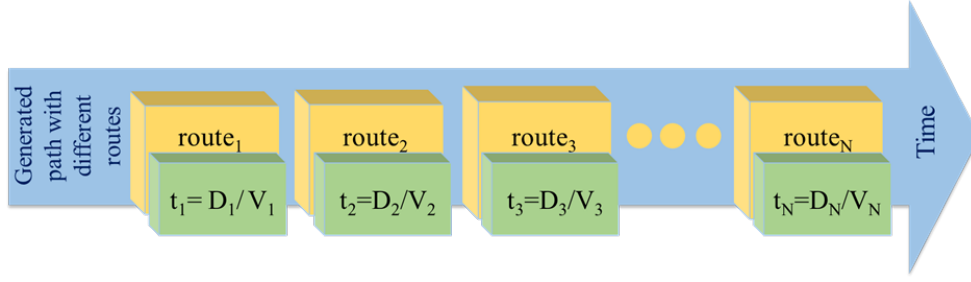


Figure 7.2: Describing scheme of routes' time spending.

Equation 7.2 represents the minimum value among a set of functions  $f_1(t), f_2(t), f_3(t), \dots, f_m(t)$ . Each function  $f_i(t)$  represents a different temporality scenario. It represents the congestion level for the  $i$ -th temporality scenario at time  $t$ . For example,  $f_1(t)$  represents the congestion level at time  $t$  for the first scenario,  $f_2(t)$  for the second scenario, and so on. The goal of Equation 7.2 is to determine the minimum congestion level across all temporality scenarios at each time  $t$ , allowing travelers to choose the departure time that offers the least traffic congestion during their trip.

$$f(t) = \min (f_1(t), f_2(t), f_3(t), \dots, f_m(t)) \quad (7.2)$$

Equation 7.3 is divided into two cases based on the comparison between the total traveling time ( $T$ ) and the time interval between observations ( $\Delta t$ ). This distinction is crucial in determining how the congestion factor is calculated for each scenario. If  $T$  is less than or equal to  $\Delta t$ , the congestion factor  $f_j(t)$  for the  $j$ -th scenario at time  $t$  is determined by summing the congestion levels for all  $N$  routes at time  $t$  ( $c(t_j)$ ). If  $T$  is greater than  $\Delta t$ , the congestion factor  $f_j(t)$  for the  $j$ -th scenario at time  $t$  is calculated by summing the congestion values for all  $N$  routes at a specific time  $t_s \% (\Delta t)$ . Here,  $t_s$  represents the summation of the times accumulated at each iteration (route) up to time  $t$ .

$$\begin{cases} \text{If } T \leq \Delta t : f_j(t) = \sum_{i=1}^N c(t_j) \\ \text{If } T > \Delta t : f_j(t) = \sum_{i=1}^N c(t_s \% (\Delta t)) \end{cases} \quad (7.3)$$

Here:

- \*  $m$ : Number of possible temporality scenarios
- \*  $T$ : Total traveling time
- \*  $\Delta t$ : Time interval between the observations
- \*  $j$ : Temporality scenario number

- \*  $N$ : Number of routes traveled for a trip
- \*  $c(t)$ : Congestion factor at time  $t$
- \*  $t_s$ : Summation of the times accumulated at each iteration (route)

The algorithm of the proposed approach is given by Algorithm 1 (FODT-Finding Optimal Departure Time).

---

**Algorithm 1** Finding Optimal Departure Time (FODT) algorithm

---

```

1: procedure FindOptimalDepartureTime
2:   Inputs: Speed data, Congestion-factor data,  $\Delta t$ : time interval
      between each observation (mn),  $I$ : time interval // from 7 a.m. to 10
      a.m.
3:   Output: Optimal departure time
4:   Group data by road name
5:   Create the time matrix and congestion factor matrix for each
      road
6:   Perform computations based on congestion factors
7:   for each time interval  $i \in I$  do
8:      $S \leftarrow 0$  // accumulated travel times
       $Y \leftarrow 0$  // accumulated congestion factors
9:     for each road do
10:      Check if  $\lfloor \frac{S}{\Delta t} \rfloor$  is within the valid range
11:      if within valid range then
12:        Get the travel time  $E$  and congestion factor  $D$  from the
        respective matrices
13:        Update  $S$  with the accumulated travel time and  $Y$  with the
        accumulated congestion factor
14:      else
15:        Set  $S$  to -1 to indicate an invalid interval and break the loop
16:      if  $S \neq -1$  then
17:        Find the minimum value of the summation of congestion fac-
        tors
18:   Extract the optimal departure time based on the index of the mini-
      mum sum

```

---

### 7.3 . Experimental results

This section presents the numerical results of implementing the proposed methodology. It uses traffic Waze data of an encounter trajectory. We focus on the city of Évry-courcouronnes due to the significant

number of employees from this city who participated in our survey [20]. So we choose departure and arrival points from this region. However, it is important to note that the developed methodology is applicable to data from any other city as well, as soon as the necessary data are available.

### 7.3.1 . Waze data

Our analysis begins with the Waze jam and Waze traffic data, composed of 207.037 and 162.699 rows, respectively. The historical data available for the selected example goes from February 1, 2021, at 03 : 27 to August 29, 2022, at 16 : 12.

We create the network structure of Évry-Courcouronnes boundary and its street network data as an example from Open Street Map (see Figure 7.3) using the library OSMnx of Python. OSMnx is a Python package to retrieve, model, analyze, and visualize street networks from Open Street Map. Users can download and model walkable, drivable, or bikeable urban networks [24].

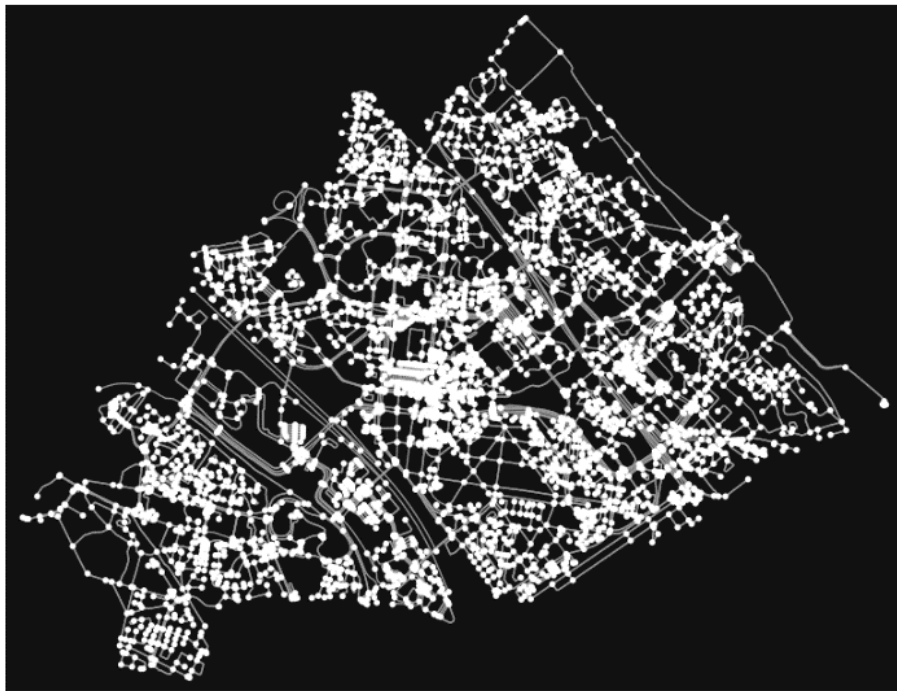


Figure 7.3: The network structure of Évry-Courcouronnes.

### 7.3.2 . Departure and arrival points

First, we choose a random departure point A and a random arrival point B on the Évry-courcouronnes city.

The departure point A is located at the address:

*Square du Long Près, Les Aunettes, Évry-Courcouronnes, Essonne, Île-de-France, France métropolitaine, 91000, France, with the GPS coordinates (48.6259, 2.4264).*

The arrival point B is situated at the address:

*8 Rue du Bois Guillaume, Évry-Courcouronnes, Essonne, Île-de-France, France métropolitaine, 91000, France, with the GPS coordinates (48.6400, 2.4257).*

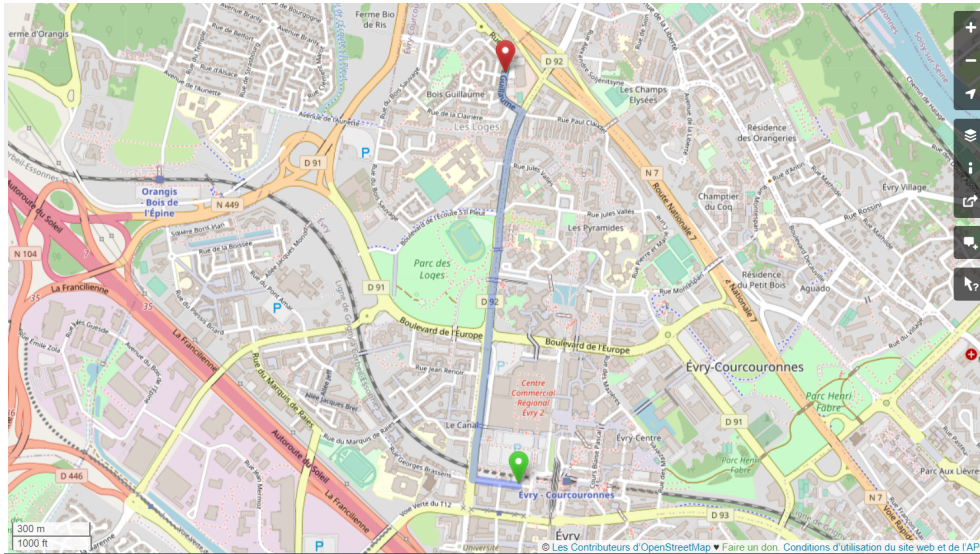


Figure 7.4: Departure and arrival points.

### 7.3.3 . Routes' name generation

To determine the routes that a person would take to travel from point A to point B, it is essential to convert the nodes extracted from the OSRM database into their respective route names. Upon completion, we obtain the following route names:

- Rue Léopold Sédar Senghor
- Boulevard des Champs-Élysées
- Rue du Bois Guillaume

### 7.3.4 . Traffic jams analysis

In this section, we perform a traffic analysis utilizing the data accessible through the Waze API for every road traversed by the individual. Subsequently, we employ the FB Prophet model to train with this dataset and make predictions about future timings. As part of our work's proof

of concept, we conduct an examination of the components of the FB Prophet model.

- **Rue Léopold Sédar Senghor:** We collect from the waze API 755 observations for *Rue Léopold Sédar Senghor*. Its FB Prophet model components are shown in Figure 7.5

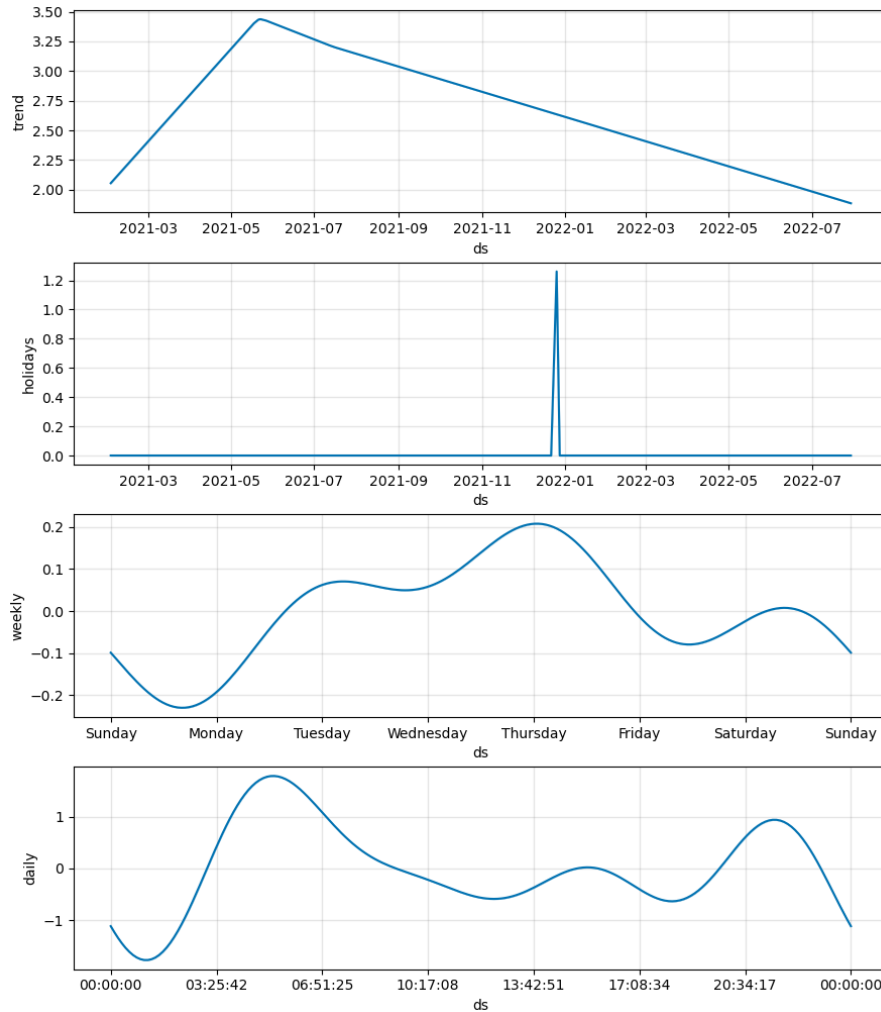


Figure 7.5: Rue Léopold forecast components.

During the examination of this street, we identify a notable upward trend between February 1<sup>st</sup>, 2021, and May 20<sup>th</sup>, 2021. This indicates a consistent rise in congestion level throughout that period. Conversely, we observe a decline from June 2021 to August 2022. Furthermore, Tuesdays and Thursdays exhibit the highest congestion levels, while daily analysis reveals peak congestion at 5 a.m. and 9 : 30 p.m.

- **Boulevard des Champs-Élysées:** A total of 1000 data points for *Boulevard des Champs-Élysées* are gathered from the Waze API, covering the period from May 2022 to August 2022. The components of the FB Prophet model can be seen in Figure 7.6

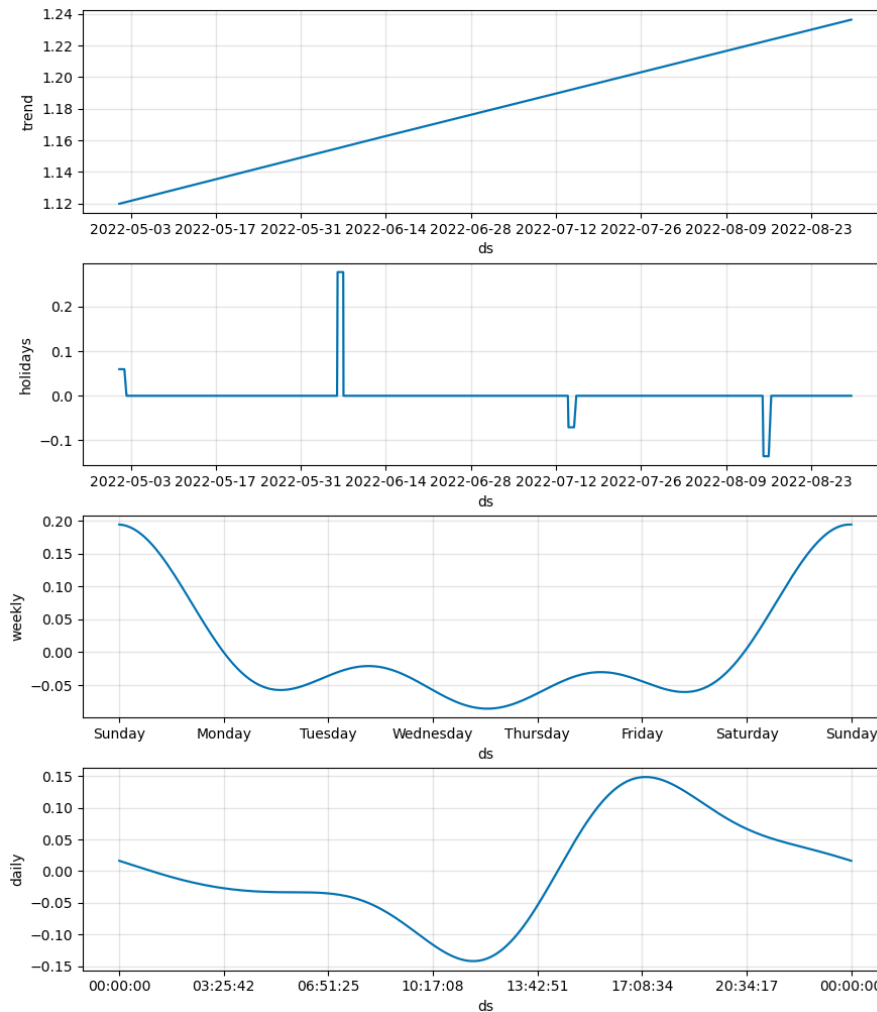


Figure 7.6: Boulevard des Champs-Élysées forecast components.

Regarding this Boulevard, we observe an overall increase in activity over the analyzed period using the FB Prophet trend component. For the holiday component, we notice that there are occasional peaks on some holidays while others show a decrease. In terms of weekly patterns, there are noticeable spikes in activity on Saturdays and Tuesdays. Furthermore, our daily analysis reveals that there is a significant surge in activity at 8 a.m. and 5 p.m.

- **Rue du Bois Guillaume:** We collect from the waze API 162 obser-



uations for *Rue du Bois Guillaume* covering the period from February 1<sup>st</sup>, 2021, to August 1<sup>st</sup>, 2022. The FB Prophet model components are shown in Figure 7.7.

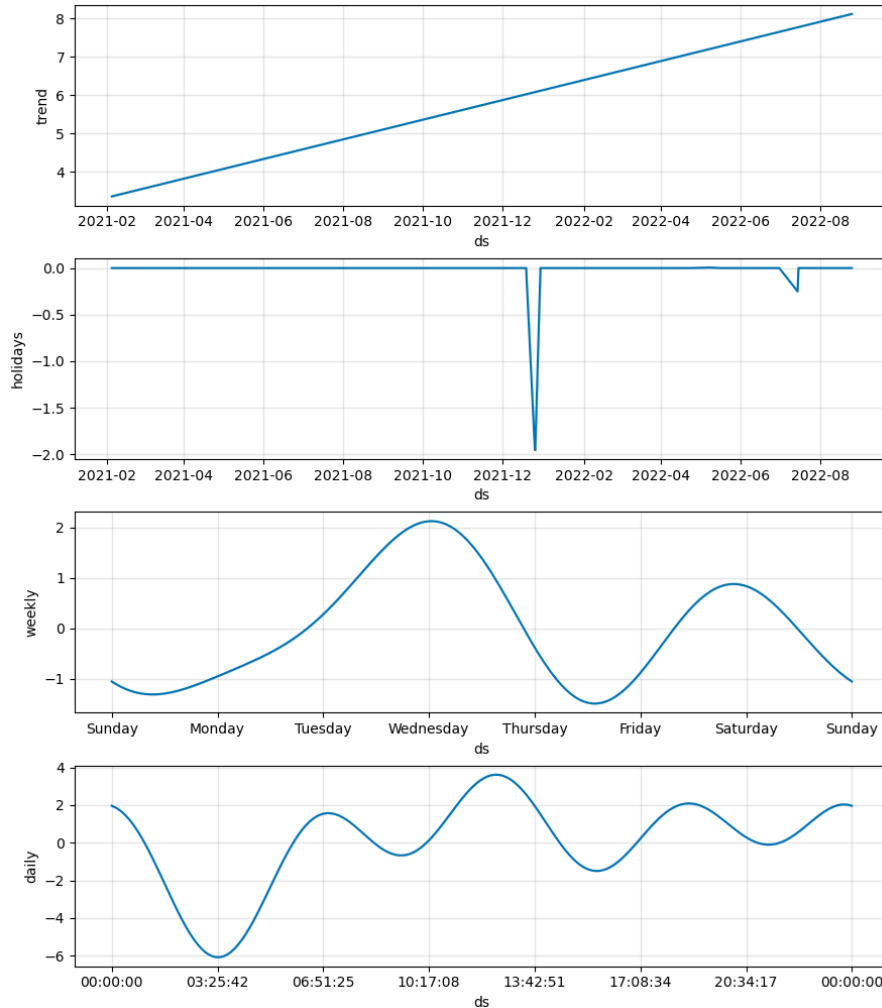


Figure 7.7: Rue Bois Guillaume forecast components.

For this street, we notice a peak on Saturdays and Wednesdays. For daily analysis, we notice a peak at 7 a.m., 12 : 30 a.m., and 6 : 30 p.m.

The traffic analysis results demonstrate the importance of our research on traffic analysis. The components of the FB Prophet model suggest that weekly and daily variations influence traffic congestion. The congestion level tends to increase at specific times and decrease at others and fluctuate depending on the day of the week. These findings serve as a proof of concept of our work.

### 7.3.5 . Congestion level calculation and prediction

In this step, we calculate the congestion level for the available data, and we forecast for a future date.

In our use-case, we utilize data ranging from February 1st, 2021 to August 30th, 2022 for training purposes, and our forecast is for September 3rd, 2022. The results are presented in Table 7.1.

To analyze and interpret the congestion factors obtained, we can consider several aspects. such as the magnitude and the comparison to a threshold.

- The magnitude of congestion factors: Compare the values of congestion factors across different roads and times. Relatively higher values indicate higher levels of congestion [10]. For example, at 07 : 00 : 00, the *Rue du Bois Guillaume* has a congestion factor of 10.1935, indicating significant congestion on that road at that time.
- Comparison with threshold: Determine if there is a predefined threshold for congestion beyond which the traffic is considered heavy. If such a threshold exists, you can assess if any of the congestion factors exceed it. For instance, if the threshold is set at 7, then at 07 : 00 : 00, both *Rue Léopold Sédar Senghor* and *Rue du Bois Guillaume* have congestion factors exceeding the threshold, indicating heavy congestion. Determining threshold values involves various approaches, with one of the widely recognized methods being the maximization of mutual information between traffic parameters and traffic state [143]. Traffic parameters, encompassing metrics like volume, speed, and occupancy, are treated as continuous variables. The likelihood distributions for these parameters can be established based on historical traffic flow data.

### 7.3.6 . Best departure time extraction

Once we have acquired the results displayed in the previous section's table, we must estimate the time taken to travel each route by making predictions on the speed at various timings. After this, we can utilize the proposed algorithm (FODT) to determine the optimal departure time that circumvents traffic congestion.

For the particular selected scenario, we ascertain that the optimal departure time from point A to point B to evade traffic congestion is 9 : 25 a.m. then 9 : 20, then 9 : 30, then 9 : 15, and the worst departure time is 7 : 00.

Table 7.1: The congestion level predicted values for the 3<sup>rd</sup> of September, 2022

	<b>Rue Léopold Sédar Senghor</b>	<b>Boulevard des Champs-Élysées</b>	<b>Rue du Bois Guillaume</b>
07:00:00	2,756139839	1,283905424	10,19350364
07:05:00	2,701353373	1,28432135	10,18271671
07:10:00	2,647177253	1,284671028	10,15642639
07:15:00	2,593819796	1,284950225	10,11539443
07:20:00	2,541470705	1,285154836	10,06048551
07:25:00	2,490300391	1,285280914	9,992659963
07:30:00	2,440459474	1,285324704	9,912965846
07:35:00	2,392078455	1,285282673	9,822530335
07:40:00	2,345267566	1,285151543	9,722550606
07:45:00	2,300116788	1,284928328	9,614284215
07:50:00	2,256696034	1,284610361	9,499039062
07:55:00	2,215055506	1,284195331	9,378163026
08:00:00	2,175226195	1,283681311	9,253033336
08:05:00	2,137220542	1,283066794	9,125045763
08:10:00	2,101033237	1,282350716	8,995603721
08:15:00	2,066642147	1,28153249	8,866107346
08:20:00	2,034009372	1,28061203	8,737942647
08:25:00	2,003082409	1,279589779	8,612470801
08:30:00	1,973795411	1,278466726	8,491017685
08:35:00	1,94607054	1,277244433	8,374863702
08:40:00	1,919819386	1,275925049	8,265234003
08:45:00	1,894944448	1,274511327	8,163289155
08:50:00	1,871340662	1,273006637	8,070116335
08:55:00	1,848896953	1,271414972	7,986721119
09:00:00	1,827497815	1,269740961	7,914019916
09:05:00	1,807024881	1,267989862	7,852833111
09:10:00	1,787358492	1,266167571	7,80387896
09:15:00	1,768379229	1,26428061	7,767768291
09:20:00	1,74996942	1,262336123	7,745000039
09:25:00	1,732014577	1,260341864	7,735957656
09:30:00	1,714404785	1,258306178	7,740906412
09:35:00	1,697036004	1,256237985	7,759991617
09:40:00	1,679811282	1,254146757	7,793237758
09:45:00	1,66264188	1,252042486	7,840548579
09:50:00	1,64544828	1,249935659	7,901708078
09:55:00	1,628161078	1,24783722	7,976382432

## **7.4 . Conclusion**

This chapter presents a methodology aimed at addressing urban traffic congestion during rush hours. The primary objective is to provide commuters with recommendations to optimize their departure times.

The methodology starts with route generation based on Open Street Map data and progresses through traffic analysis utilizing traffic jam data collected from Waze. Using time series forecasting, particularly the FB Prophet model, enables us to predict congestion levels for different timings.

The extraction of optimal departure times, referred to as "temporality scenarios," is done by evaluating the summation of congestion levels across various routes. The presented FODT algorithm helps in identifying the departure time that offers the least congestion for a given journey.

The methodology demonstrates its applicability through an illustrative example involving Évry-Courcouronnes city, located in the GPS region showing the process from data collection and analysis to the extraction of optimal departure time (temporality scenario).



## 8 - Conclusion

This doctoral dissertation explores existing solutions to mitigate urban traffic congestion by proposing a decision platform. Acknowledging the multifaceted nature of urban congestion, this platform relies on an interdisciplinary approach, coordinating data analysis and sociological considerations to forge a path toward sustainable and efficient urban mobility solutions.

The initial chapter of the state of the art outlined various congestion-alleviating strategies, including connected and autonomous vehicles, car sharing, and temporal policies. In our knowledge, existing solutions lack a sociological dimension. Our research works fill this gap by considering sociological factors influencing individual mobility choices. Leveraging diverse techniques such as one-hot encoding, supervised and unsupervised learning models, and time series forecasting with FB Prophet, our research works offer the groundwork for a robust analytical approach. The decision platform introduced a modular architecture including classification, analysis, and decision-making modules. The platform exploits heterogeneous data, including survey responses obtained through the developed questionnaire with information from Open Street Map and Waze API. Significantly, the decision-making module takes into consideration individual sociological constraints, offering adapted and flexible recommendations to them. In the identification and prediction of mobility profiles using machine learning techniques (MLPANN, KNN, and K-means clustering), we investigated the impact of sociological factors on mobility choices. The integration of supervised and unsupervised methods provides a refined understanding of mobility behaviors, providing policymakers with information to adapt transportation strategies for diverse segments of the population.

Through sociological interpretations, the link between machine learning predictions and sociological factors was identified. The impact of marital status, dependent children, and socio-professional categories on flexibility in working hours emerged as critical factors. This sociological perspective not only improves the decision platform but can also guide policymakers seeking to align transportation strategies with the diverse needs of urban inhabitants. A methodology for optimal traffic-free departure forecasts was presented, translating theoretical insights into practical recommendations. Utilizing data from Open Street Map and Waze, coupled with the predictive capabilities of the FB Prophet model, the developed FODT algorithm offers commuters optimized de-

parture times.

Looking forward, numerous avenues for future research can be explored.

- One critical aspect to consider is the replicability of our work to other regions. It's important to note that while our works consider the Île-de-France region, they may vary significantly in other territories with distinct characteristics, such as limited public transportation or different occupational demographics. Researchers can explore the potential of replicating our research works by employing a similar questionnaire, utilizing the data we have gathered, and using the techniques we have used.
- Another potential avenue for future research involves exploring alternative data analysis techniques and evaluating their efficacy. This could include investigating ensemble techniques for predicting F-NF category profiles. For example, the voting classifier technique, which is an ensemble technique categorized under bagging, combines predictions from diverse base models through majority or probability-based voting [132]. This approach could be particularly pertinent in our case when dealing with F-NF.  
Another technique worth trying is the association method to uncover patterns and rules connecting various sociological criteria and predefined mobility profiles. Our previous analysis was conducted manually with the guidance of my sociological supervisor, but it may be beneficial to assess the performance of automated approaches. Additionally, exploring different forecasting methods to predict congestion levels could provide interesting information.
- Another perspective worth exploring is the promotion of active and sustainable mobility by introducing alternative modes of transportation, such as walking or cycling. Indeed, our focus in our research works has been exclusively on the utilization of cars.
- Another avenue for further investigation is to consider additional data sources, including EMD (Enquête Ménage déplacement), available on the INSEE website [5], EGT (Enquête Globale de Transport), EOD (Enquêtes Origine Destination), and Declared Preference surveys, available on the ADISP (Archives de Données Issues de la Statistique Publique) website [6].
- Incorporating telecommuting as an alternative for users experiencing significant congestion, particularly during morning hours, could serve as a valuable addition to congestion solutions.
- Finally, validating and implementing the proposed decision support solution for the *Bureaux des Temps* in real urban environments

is essential to assess its effectiveness in reducing congestion and enhancing the quality of urban mobility.





## Bibliography

- [1] . URL <https://www.grandparissud.fr/>.
- [2] . URL <https://www.iledefrance.fr/>.
- [3] . URL <https://www.waze.com/waze>.
- [4] <https://infonet.fr/lexique/definitions/categories-socio-professionnelles/>: :text=une.
- [5] <https://www.insee.fr/fr/accueil>. .
- [6] [http://www.progedo-adisp.fr/enquetes\\_aonloc.php](http://www.progedo-adisp.fr/enquetes_aonloc.php).
- [7] G. Abdelkader, K. Elgazzar, and A. Khamis. Connected vehicles: Technology review, state of the art, challenges and opportunities. *Sensors*, 21(22):7712, Nov 2021.
- [8] S. Abdul-Jabbar and A. Farhan. Data analytics and techniques: A review. *ARO-The Scientific Journal of Koya University*, 10:45–55, 10 2022.
- [9] R. Abduljabbar, H. Dia, S. Liyanage, and S. Bagloee. Applications of artificial intelligence in transport: An overview. *Sustainability*, 11:189, 01 2019.
- [10] M. Aftabuzzaman. Measuring traffic congestion- a critical review. *30th Australasian Transport Research Forum*, 01 2007.
- [11] C. Aggarwal. *Data Classification: Algorithms and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery. CRC Press, 2020.
- [12] S. Aguadé-Bruix and J. Castell-Conesa. Methods of quantification. 01 2001.
- [13] A. Aguilera, V. Lethiais, A. Rallet, and L. Proulhac. Le télétravail, un objet sans désir ? *Rev. Econ. Reg. Urbaine*, Février(1):245–266, Feb. 2016.
- [14] S. C. A. Alexandre, Shaheen. The societal value of carpooling: The environmental and economic value of sharing a ride, 2018.
- [15] H. Alsghaier, M. Akour, I. Shehabat, and S. Aldiabat. The importance of big data analytics in business: A case study. *American Journal of Software Engineering and Applications*, Vol. 6:pp. 111–115, 10 2017.
- [16] S. Amirul, S. Amirul, and R. Mail. Strategic flexible working arrangement: The realignment between human resource and management accounting. *Humanities Social Sciences Reviews*, 8, 09 2020.

- [17] O. Azam. Flexible timing and its effects, 02 2018.
- [18] G. B G, I. M. Umesh, and R. B. Varma. A study on applications of data analytics in non-business domains. *International Journal of Science, Engineering and Technology Research (IJSETR)*, 6:1249–1251, 08 2017.
- [19] L. J. Basso, C. A. Guevara, A. Gschwender, and M. Fuster. Congestion pricing, transit subsidies and dedicated bus lanes: Efficient and practical solutions to congestion. *Transport Policy*, 18(5):676–684, 2011.
- [20] Y. Berkane, L. Kloul, and Y. Demoli. Time organization for urban mobility decongestion: A methodology for people's profile identification. *International Journal of Urban and Civil Engineering*, 16(7):186–194, 2022. ISSN eISSN: 1307-6892. URL <https://publications.waset.org/vol/187>.
- [21] Y. Berkane, L. Kloul, and Y. Demoli. Mobility profiles identification using machine learning techniques. 2023. URL <https://osf.io/6jxr2>.
- [22] Y. Berkane, L. Kloul, and Y. Demoli. Optimal traffic-free departure forecast. 2023. URL <http://www.icitt.org/index.html>.
- [23] S. M. Bianchi, L. C. Sayer, M. A. Milkie, and J. P. Robinson. Housework: Who Did, Does or Will Do It, and How Much Does It Matter? *Social Forces*, 91(1): 55–63, 09 2012.
- [24] G. Boeing. Osmnx: A python package to work with graph-theoretic open-streemap street networks. *The Journal of Open Source Software*, 2, 04 2017. doi: 10.21105/joss.00215.
- [25] J. Boulin. Les temps de la ville: Au-delà d'un aspect fonctionnel (l'organisation des services publics), les politiques des temps visent à un "bien-être temporel", offrant disponibilité et prévisibilité à tous. *Revue Projet*, 273:64–72, 2003.
- [26] J.-Y. Boulin. Les temps de la ville. *Projet*, 273(1):64, 2003.
- [27] P. Bourdieu. *La distinction: Critique sociale du jugement*. Les Éditions de Minuit. 1998.
- [28] D. Brown, A. Abbasi, and R. Lau. Predictive analytics introduction. *Intelligent Systems, IEEE*, 30:6–8, 03 2015.
- [29] M. Börjesson and I. Kristoffersson. The gothenburg congestion charge. effects, design and politics. *Transportation Research Part A: Policy and Practice*, 75, 05 2015.

- [30] P. Bühlmann. Bagging, boosting and ensemble methods. *Handbook of Computational Statistics*, 01 2012.
- [31] E. A. Cech and L. T. O'Connor. 'like second-hand smoke': the toxic effect of workplace flexibility bias for workers' health. *Community, Work & Family*, 20(5): 543–572, 2017.
- [32] M. Colleoni. *A Social Science Approach to the Study of Mobility: An Introduction*, pages 23–33. 12 2016.
- [33] E. Commission, D.-G. for Mobility, Transport, H. Essen, D. Fiorello, K. El Beyrouthy, C. Bieler, L. Wijngaarden, A. Schroten, R. Parolin, M. Brambilla, D. Sutter, S. Maffii, and F. Fermi. *Handbook on the external costs of transport : version 2019 – 1.1*. Publications Office, 2020.
- [34] G. Costa, S. Prunier-Poulmaire, and B. Barthe. Horaires de travail et société : quels constats, quelles évolutions en europe et en france ? 09 2013.
- [35] N. Coulombel, E. Côme, L. Oukhellou, and C. Richer. Mobilletic - données billettiques et analyse des mobilités urbaines : le cas rennais, 09 2017.
- [36] T. Cresswell. Towards a politics of mobility. *Environment and Planning D: Society and Space*, 28(1):17–31, 2010. doi: 10.1068/d11407.
- [37] J. Creswell. *Research Design: Qualitative, Quantitative, and Mixed-Method Approaches*. 01 2009.
- [38] R. Crompton and C. Lyonette. Work-life 'balance' in europe. *Acta Sociologica - ACTA SOCIOLOGICA*, 49:379–393, 12 2006.
- [39] K. Curran, G. Fisher, and J. Crumlish. Openstreetmap. *International Journal of Interactive Communication Systems and Technologies*, 2:69–78, 01 2012.
- [40] S. Daele and G. Janssenswillen. *Identifying the Steps in an Exploratory Data Analysis: A Process-Oriented Approach*, pages 526–538. 03 2023.
- [41] C. Dameron, L. Verchère, D. Royoux, E. MUNCH, and L. Gendreau. *Prendre en compte le temps dans les mobilités*. 05 2019.
- [42] L. Delaney. Inferential analysis: Powerful storytelling. *African Journal of Midwifery and Women's Health*, 5:31–35, 01 2011.
- [43] Y. Demoli. Les femmes prennent le volant. *Trav. Genre Soc.*, n° 32(2):119–140, Oct. 2014.
- [44] Y. Demoli and M. K. Gilow. Mobilité parentale en belgique : question de genre, question de classe. *Espaces Soc.*, n° 176-177(1):137–154, Oct. 2019.

- [45] R. Deshmukh and V. Wangikar. Data cleaning: Current approaches and issues. 01 2011.
- [46] B. Drevon. Accélération. une critique sociale du temps. *Idées économiques et sociales*, 177(3):78–79, 2014.
- [47] S. Dridi. Supervised learning - a systematic literature review, 12 2021.
- [48] S. Dridi. Unsupervised learning - a systematic literature review a preprint. [salimdridi.info](https://salimdridi.info), December 2021. Preprint.
- [49] J. Eliasson and L.-G. Mattsson. Equity effects of congestion pricing: Quantitative methodology and a case study for stockholm. *Transportation Research Part A: Policy and Practice*, 40:602–620, 02 2006.
- [50] A. N. M. A. Fakir. *Quantative Data Analysis in Social Science Research*. 01 2016.
- [51] J. Faraway and C. Chatfield. Time Series Forecasting with Neural Networks: A Comparative Study Using the Air Line Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 47(2):231–250, 01 2002.
- [52] D. Farren. Inferential statistic, 06 2014.
- [53] V. Gash. Preference or constraint? part-time workers’ transitions in denmark, france and the united kingdom. *Work Employment and Society - WORK EMPLOY SOC*, 22:655–674, 12 2008.
- [54] L. Golden. *Flexible Work Schedules and Their Impact on Employees*, pages 122–137. 01 2003.
- [55] L. Golden, S. Lambert, J. Henly, and B. Wiens-Tuers. Working time in the employment relationship: Perceived control and work-life balance. *Research Handbook on the Future of Work and Employment Relations*, 05 2010.
- [56] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.
- [57] D. Graham and S. Glaister. The demand for automobile fuel: A survey of elasticities. *Journal of Transport Economics and Policy*, 36(1):1–25, 2002.
- [58] M. Greenacre, P. Groenen, T. Hastie, A. Iodice D’Enza, A. Markos, and E. Tuzhilina. Principal component analysis. 2:100, 12 2022.
- [59] D. Greene, P. Cunningham, and R. Mayer. *Unsupervised Learning and Clustering*, pages 51–90. 01 2008.
- [60] G. Guo, H. Wang, D. Bell, and Y. Bi. Knn model-based approach in classification. 08 2004.

- [61] E. Güresen and G. Kayakutlu. Definition of artificial neural networks with comparison to other networks. In *WCIT*, 2011.
- [62] T. Gärling, C. Bergstad, P. Loukopoulos, and S. Fujii. *Acceptability of Road Pricing*, pages 193–208. 02 2008.
- [63] M. Haicault. Des pratiques temporelles du travail aux temporalités urbaines : 20 années de recherches sur la thématique des temps sociaux. 04 2005.
- [64] A. Hammoudeh. A concise introduction to reinforcement learning, 02 2018.
- [65] J. Hancock and T. Khoshgoftaar. Survey on categorical data for neural networks. *Journal of Big Data*, 7, 04 2020.
- [66] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 28(1):100, 1979.
- [67] M. Hasan. A framework for intelligent decision support system for traffic congestion management system. *Engineering, Scientific Research, USA.*, Vol. 2:, pp. 270–289, 04 2010.
- [68] R. Heiss. *Data, Types of*. 11 2017.
- [69] <https://reducmob.alternatives developpement.fr/>.
- [70] [https://research.facebook.com/blog/2017/2/prophet-forecasting-at scale/](https://research.facebook.com/blog/2017/2/prophet-forecasting-at-scale/).
- [71] <https://www.ademe.fr/>.
- [72] <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000398162>.
- [73] <https://www.onisr.securite routiere.gouv.fr>.
- [74] R. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 3rd edition, 2021.
- [75] S. Idreos, O. Papaemmanouil, and S. Chaudhuri. Overview of data exploration techniques. pages 277–281, 05 2015.
- [76] S. Imandoust and M. Bolandraftar. Application of k-nearest neighbor (knn) approach for predicting economic events theoretical background. *Int J Eng Res Appl*, 3:605–610, 01 2013.
- [77] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323, sep 1999.
- [78] D. Jelonek. Big data analytics in the management of business. *MATEC Web of Conferences*, 125:04021, 01 2017.

- [79] J. Jokar Arsanjani, A. Zipf, P. Mooney, and M. Helbich. *An Introduction to Open-StreetMap in Geographic Information Science: Experiences, Research, and Applications*, pages 1–15. 03 2015.
- [80] A. Junter. Le temps des villes à rennes : retour sur une expérience. *Inf. Soc.*, n° 153(3):88–96, July 2009.
- [81] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, and T. Weil. Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions. *IEEE Commun. Surv. Tutor.*, 13(4):584–616, 2011.
- [82] Z. Karimi. Confusion matrix, 10 2021.
- [83] A. Katal and N. Singh. *Artificial Neural Network: Models, Applications, and Challenges*, pages 235–257. 01 2022.
- [84] M. Kergoat and D. Touchard. La Ville et le temps des femmes : l'exemple de la Ville de Rennes. *Espace Populations Sociétés*, (2-3):411–420, 2007.
- [85] A. Kesting, M. Treiber, and D. Helbing. Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philos. Trans. A Math. Phys. Eng. Sci.*, 368(1928):4585–4605, Oct. 2010.
- [86] H. Khan. Types of ai | different types of artificial intelligence systems [fosskguru.com/types-of-ai-different-types-of-artificial-intelligence-systems](https://fosskguru.com/types-of-ai-different-types-of-artificial-intelligence-systems). 9: 50, 09 2021.
- [87] A. Kibangou, T. Moyo, and W. Musakwa. Dynamic Linear Model for Urban Essential Traffic Congestion and Emissions. In *CCA 2021 - Control Conference Africa*, pages 1–6, Magaliesburg, South Africa, Dec. 2021.
- [88] A. Kolleck. Does car-sharing reduce car ownership? empirical evidence from germany. *Sustainability*, 13(13):7384, Jul 2021.
- [89] A. Koźlak and D. Wach. Causes of traffic congestion in urban areas. case of poland. *SHS Web of Conferences*, 57:01019, 01 2018.
- [90] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker-Walz. Recent development and applications of sumo - simulation of urban mobility. *International Journal On Advances in Systems and Measurements*, 34, 12 2012.
- [91] V. Kumar and M. L. Predictive analytics: A review of trends and techniques. *International Journal of Computer Applications*, 182:31–37, 07 2018.

- [92] V. Kushwah and A. Bajpai. Machine learning and its algorithms a research. *International Journal of Innovative Technology and Exploring Engineering*, 8:801–803, 10 2019.
- [93] A. A. Latip and S. R. Amirul. Conceptualising the roles of human resource in managing flexible working arrangements. *Labuan Bulletin of International Business and Finance (LBIBF)*, 20(1):74–86, Aug. 2022.
- [94] J. C. Lengen, A. C. Kordsmeyer, E. Rohwer, V. Harth, and S. Mache. Social isolation among teleworkers in the context of the covid-19 pandemic: Insights for designing telework with regards to social needs. *Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie*, 71(2):63–68, 2021.
- [95] L. Lesnard. Flexibilité des horaires de travail et inégalités sociales. 2006.
- [96] H. Levinson and J. Falcocchio. Urban development and traffic congestion. pages 948–956, 03 2011.
- [97] S. Li, G. Li, Y. Cheng, and B. Ran. Urban arterial traffic status detection using cellular data without cellphone gps information. *Transportation Research Part C: Emerging Technologies*, 114:446–462, 05 2020.
- [98] M. Lippi, M. Bertini, and P. Frasconi. Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *Intelligent Transportation Systems, IEEE Transactions on*, 14:871–882, 06 2013.
- [99] T. Litman. Evaluating carsharing benefits. *Transportation Research Record*, 1702:31–35, 01 2000.
- [100] T. Litman. Autonomous vehicle implementation predictions: Implications for transport planning. 2015.
- [101] Q. Liu and Y. Wu. Supervised learning. 01 2012.
- [102] D. Luxen and C. Vetter. Real-time routing with openstreetmap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '11*, pages 513–516, New York, NY, USA, 2011. ACM.
- [103] Z. Lv and W. Shang. Impacts of intelligent transportation systems on energy conservation and emission reduction of transport systems: A comprehensive review. *Green Technologies and Sustainability*, 1(1):100002, 2023.
- [104] E. Lépinard. Une occasion perdue : l'exemple français des politiques publiques des « bureaux des temps ». *Cahiers du Genre*, 33:113, 12 2002.



- [105] S. Mallet. Aménager les rythmes: politiques temporelles et urbanisme. 01 2013.
- [106] K. A. Marczuk, H. Soh, C. L. Azevedo, D.-H. Lee, and E. Frazzoli. Simulation framework for rebalancing of autonomous mobility on demand systems. 2016.
- [107] E. Martin and S. Shaheen. Impacts of car2go on vehicle ownership, modal shift, vehicle miles traveled, and greenhouse gas emissions: An analysis of five north american cities. *Transportation Sustainability Research Center (TSRC)*, July 2016.
- [108] G. menon and S. Guttikunda. Electronic road pricing: Experience lessons from singapore, 01 2010.
- [109] S. Merriam. *Qualitative Research: A Guide to Design and Implementation*. Higher and adult education series. John Wiley & Sons, 2009.
- [110] D. Metz. Developing policy for urban autonomous vehicles: Impact on congestion. *Urban Science*, 2(2), 2018.
- [111] . Ministry in charge of industry.
- [112] A. Mondschein and B. D. Taylor. Is traffic congestion overrated? examining the highly variable effects of congestion on travel and accessibility. *Journal of Transport Geography*, 64:65–76, 2017.
- [113] E. Munch. L’harmonisation des horaires de travail: en finir avec l’heure de pointe? *Métropolitiques. eu*, page 5, 2014.
- [114] E. Munch. *Mais pourquoi arrivent-ils tous à la même heure ? : le paradoxe de l’heure de pointe et des horaires de travail flexibles : enquête sociologique auprès de cadres franciliens*. Theses, Université Paris-Est, Dec. 2017.
- [115] J. Nazzal, I. El-Emary, and S. Najim. Multilayer perceptron neural network (mlps) for analyzing the properties of jordan oil shale. *World Applied Sciences Journal*, 5, 01 2008.
- [116] P. Newbold, W. L. Carlson, and B. Thorne. *Statistics for Business and Economics*. Pearson, 2019.
- [117] T. M. C. Nguyen, H. Kato, and L. B. Phan. Is built environment associated with travel mode choice in developing cities? evidence from hanoi. *Sustainability*, 12(14):5773, Jul 2020.

- [118] D. Q. Nguyen-Phuoc, M. Zhou, M. Hong Chua, A. Romano Alho, S. Oh, R. Seshadri, and D.-T. Le. Examining the effects of automated mobility-on-demand services on public transport systems using an agent-based simulation approach. *Transportation Research Part A: Policy and Practice*, 169:103583, 2023.
- [119] H. Nijp, D. Beckers, S. Geurts, P. Tucker, and M. Kompier. Systematic review on the association between employee worktime control and work-non-work balance, health and well-being, and job-related outcomes. *Scandinavian journal of work, environment health*, 38:299–313, 06 2012.
- [120] S. Niraula. A review of research process, data collection and analysis. *Insights in Biology and Medicine*, 3:001–006, 01 2019.
- [121] B. Parasumanna Gokulan, X. German, and D. Srinivasan. Urban traffic signal control using reinforcement learning agents. *Intelligent Transport Systems, IET*, 4:177–188, 10 2010.
- [122] Y.-S. Park and S. Lek. *Artificial Neural Networks*, volume 28, pages 123–140. 12 2016.
- [123] M. Pavone. *Autonomous Mobility-on-Demand Systems for Future Urban Mobility*, pages 399–416. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.
- [124] L. Peterson. K-nearest neighbor. *Scholarpedia*, 4:1883, 01 2009.
- [125] A. S. Pothen. *Artificial Intelligence and its Increasing Importance*, pages 74–81. 01 2022.
- [126] M. Quintana, P. Arjunan, and C. Miller. Islands of misfit buildings: Detecting uncharacteristic electricity use behavior using load shape clustering, 10 2019.
- [127] K. Raheja, A. Dubey, and R. Chawda. Data analysis and its importance in health care. *International Journal of Computer Trends and Technology*, 48:176–180, 06 2017.
- [128] A. Rana, A. Rawat, A. Bijalwan, and H. Bahuguna. Application of multi layer (perceptron) artificial neural network in the diagnosis system: A systematic review. pages 1–6, 08 2018.
- [129] J. Rana, P. Luna Gutierrez, and J. Oldroyd. *Quantitative Methods*, pages 1–6. 06 2021.
- [130] Y. Reddy, V. Pulabaigari, and E. B. Semi-supervised learning: a brief review. *International Journal of Engineering Technology*, 7:81, 02 2018.
- [131] F. Ridzuan and W. M. N. Zainon. A review on data cleansing methods for big data. *Procedia Computer Science*, 161:731–738, 01 2019.

- [132] P. S P, R. C., and S. K S. Classifying the unstructured it service desk tickets using ensemble of classifiers. pages 221–227, 12 2018.
- [133] R. Sabah, G. Scholar, University, P. Karachi, M. Hassan, S. Syed, and S. Qadri. Research process and steps involved in data analysis. 07 2022.
- [134] A. W. Sadek. Artificial intelligence applications in transportation. *Transportation research circular*, 2007.
- [135] A. Samaha and H. Mostofi. Predicting the likelihood of using car-sharing in the greater cairo metropolitan area. *Urban Science*, 4(4), 2020.
- [136] I. Sarker. Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems, 01 2022.
- [137] D. Schrank, B. Eisele, and T. Lomax. 2020 urban mobility report. Technical report, Texas A&M Transportation Institute, 2020.
- [138] P. Selvaraj and P. Marudappa. A survey on various applications of prescriptive analytics. *International Journal of Intelligent Networks*, 1:76–84, 01 2020.
- [139] S. Shaheen and A. Cohen. Carsharing and personal vehicle services: World-wide market developments and emerging trends. *International Journal of Sustainable Transportation - INT J SUSTAIN TRANSP*, 7, 06 2012.
- [140] D. Shoup. The trouble with minimum parking requirements. *Transportation Research Part A: Policy and Practice*, 33:549–574, 09 1999.
- [141] A. Singh, A. Yadav, and A. Rana. K-means with three different distance metrics. *International Journal of Computer Applications*, 67:13–17, 04 2013.
- [142] B. Smith, E. Jensen, and B. Wagoner. *Quantification*. 10 2016.
- [143] Z. Sun, W. Gu, J. Feng, and X. Zhu. Threshold value based traffic congestion identification method. 04 2012.
- [144] E. Taniguchi and Y. KAKIMOTO. *Modelling effects of E-commerce on urban freight transport*, pages 135–146. 01 2004.
- [145] S. Taylor and B. Letham. Forecasting at scale, 09 2017.
- [146] E. Uhlemann. Introducing connected vehicles [connected vehicles]. *Vehicular Technology Magazine, IEEE*, 10:23–31, 03 2015.
- [147] E. Umargono, J. Suseno, and S. Gunawan. K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula. 01 2020.

- [148] J. Urry. *Mobilities*. Polity Press, 2007.
- [149] D. H. van Amelsfort and V. Swedish. Introduction to congestion charging: A guide for practitioners in developing cities. 2015.
- [150] C. Wang, S. Gong, A. Zhou, T. Li, and S. Peeta. Cooperative adaptive cruise control for connected autonomous vehicles by factoring communication-related constraints . *Transportation Research Part C Emerging Technologies*, 113, 04 2019.
- [151] M. Weideman and K. Hofmeyr. The influence of flexible work arrangements on employee engagement: An exploratory study. *SA Journal of Human Resource Management*, 18, 04 2020.
- [152] M. Yavuz, H. Özen, and O. Şahin. Effects of connected and autonomous vehicles on traffic and mobility. 12 2021.
- [153] X. Ying. An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168:022022, 02 2019.
- [154] L. Yu and R. Wang. Researches on adaptive cruise control system: A state of the art review. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 236(2-3):211–240, 2022.
- [155] G. Zardini, N. Lanzetti, M. Pavone, and E. Frazzoli. Analysis and control of autonomous mobility-on-demand systems: A review, 06 2021.
- [156] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen. Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639, 2011.
- [157] R.-M. Ştefan. A comparison of data classification methods. *Procedia Economics and Finance*, 3:420–425, 12 2012.

-



## A - Appendix A

### The questionnaire

This questionnaire is intended for employees of the territory of Grand Paris Sud. It deals with the evolution of mobility and ways of organizing work before and since the sanitary crisis. The Île-de-France Region, the Grand Paris Sud urban community, and Paris Saclay University invite you to reply to this questionnaire to collect your expectations on these themes. Your company may have initiated a process to improve mobility (e.g., smoothing of peak hours, inter-company mobility plan, etc.) for which these elements will be valuable. The results of this survey will be used to produce an overall report and detailed reports by the entity at the end of June. The data collected are anonymized and are not personal. The information entered are only used in the context of mobility projects.

#### A.1 . Before the health crisis: Your mobility and organization of working hours

(Before March 2020)

1. In which municipality of Grand Paris Sud do you work?
2. Before March 2020, how did you get to your workplace? Please indicate your answer for each category.
3. If you arrive by public transport, what is your arrival station?
  - (a) Evry (RER D)
  - (b) Bras de Fer (RER D)
  - (c) Corbeil Essonnes (RER D)

	Every day of the week	2 or 3 times a week	2 or 3 times a month	Less often	Never
Car with combustion engine					
Car with electric/hybrid motor					
Metro					
RER					
Train					
Motorcycle					
Bus					
Bike					
Scooters					
Walking (more than 15 min)					
Combine several modes					

- (d) Evry Val de Seine (RER D)
  - (e) I don't take public transport to go to work
4. On a scale of 1 to 6, do you consider this journey to be pleasant?
- |                 |   |   |   |               |   |
|-----------------|---|---|---|---------------|---|
| 1               | 2 | 3 | 4 | 5             | 6 |
| Very unpleasant |   |   |   | Very pleasant |   |
5. In your opinion, what is the most important criterion for a pleasant journey?
- (a) A minimum of connections
  - (b) Shortest time
  - (c) Lowest price
  - (d) Mode comfort
  - (e) Low affluence of the mode
  - (f) Other
6. Ideally, what mode of public transport would you prefer to take to get to work?
- (a) RER
  - (b) Bus
  - (c) Tramway
  - (d) None
7. Before March 2020, what was your commuting time (one way, in minutes)?
- (a) Less than 15 minutes
  - (b) From 15 to 30 minutes
  - (c) From 30 to 45 minutes
  - (d) From 45 minutes to 1 hour
  - (e) From 1h to 1h30
  - (f) More than 1h30
8. What is the distance between your home and your workplace? (one way, in km)
- (a) Less than 3km
  - (b) From 4 to 7 km
  - (c) From 8 to 15 km
  - (d) From 15 to 30 km
  - (e) More than 30km
  - (f) I don't know

9. Before March 2020, at what time do you usually arrive at your workplace?
- (a) Before 8 a.m.
  - (b) Between 8 a.m. and 8:30 a.m.
  - (c) Between 8:30 a.m. and 9 a.m.
  - (d) Between 9 a.m. and 9:30 a.m.
  - (e) Between 9:30 a.m. and 10 a.m.
  - (f) After 10 a.m.
10. For what reasons?
- (a) For personal convenience
  - (b) Hours imposed by your company
  - (c) Transport schedules
  - (d) Personal organization constraints (accompaniment of children, sport, shopping,...)
  - (e) Medical constraints
  - (f) To avoid transport congestion
  - (g) To avoid traffic congestion
  - (h) Other
11. Before March 2020, can you describe the organization of your typical working week?
- (a) Arrivals and departures at fixed times (for example, 8:30 a.m. to 9:30 a.m. in the morning and 5 p.m. to 6 p.m. in the evening)
  - (b) Arrivals at fixed times and departures at flexible times
  - (c) Arrivals at flexible times and departures at fixed times
  - (d) Arrivals and departures at flexible times
  - (e) Shift work schedules (ex: 3x8)
  - (f) Night work hours
  - (g) Other
12. Before March 2020, how often did you telework during the week?
- (a) 1 day per week
  - (b) 2 days a week
  - (c) 3 days a week
  - (d) More than 3 days per week
  - (e) Occasionally, several times a month
  - (f) Occasionally, several times a year



	Yes	No	I don't know
Bike mileage indemnity			
Secure parking for bicycles			
Integration of hybrid or electric vehicles in its vehicle fleet			
Access to a bike rental or loan			
Charging stations for electric vehicles			
A carpool organization system			
The possibility of telecommuting or remote work 1 day per week minimum			
The possibility of having more flexible hours			
Reimbursement of the Navigo Pass beyond 50%			
Better real-time information on transport schedules			

(g) Never

13. Before March 2020, what specific actions/solutions were implemented by your company to promote sustainable transport modes and new ways of working?

## A.2 . Your trips since the start of the health crisis

The evolutions in your mobility since the start of the health crisis

14. Have you or do you want to change your mode of transport since the health crisis to get to your workplace?
- (a) No, I will use the same mode of transport back to my workplace
  - (b) Yes, I have changed or will change my mode of transport back to my workplace
  - (c) I no longer travel to my workplace
15. By which mode do you have or do you wish to change mode of transport?
- (a) Car with combustion engine
  - (b) Hybrid/electric car
  - (c) Carpooling
  - (d) Motorcycle/scooter
  - (e) RER
  - (f) Bus
  - (g) Bike
  - (h) Walking (more than 15 minutes)
  - (i) Other
16. Why would you want to change the mode of transport?
- (a) Due to the health crisis and the lack of distancing in transport
  - (b) Necessity in the context of my professional activity

- (c) Personal organization constraints (accompanying children, sport, shopping, ...)
- (d) The previous mode no longer suits me (discomfort, travel time, ...)
- (e) For economic reasons
- (f) To reduce my carbon footprint
- (g) Other

### **A.3 . Your work organization since the health crisis**

17. How did you work during the first phase of confinement? (March, April, May 2020)
  - (a) 100% teleworking
  - (b) Teleworking and on-site
  - (c) 100% on-site
  - (d) Partial unemployment and teleworking
  - (e) Partial unemployment and work on site
  - (f) On sick leave
  - (g) Other
18. How did you work during the second phase of confinement? (November to mid-December 2020)
  - (a) 100% teleworking
  - (b) Teleworking and on-site
  - (c) 100% on-site
  - (d) Partial unemployment and teleworking
  - (e) Partial unemployment and work on site
  - (f) On sick leave
  - (g) Other
19. How did you work during the third phase of confinement?
  - (a) 100% teleworking
  - (b) Teleworking and on-site
  - (c) 100% on-site
  - (d) Partial unemployment and teleworking
  - (e) Partial unemployment and work on site
  - (f) On sick leave
  - (g) Other

20. Since the start of the health crisis, are you satisfied with the work organization methods put in place by your company?
- (a) Yes
  - (b) No
  - (c) Partially
21. For people in the tertiary sector, when you return to your workplace, what would you like to change in relation to your current work organization?
- (a) More telework
  - (b) Have more flexible hours (come and leave outside peak hours)
  - (c) Part-time working
  - (d) Full time working
  - (e) Take advantage of coworking or a third place near my home
  - (f) Nothing, the current organization suits me
  - (g) My position does not correspond to the tertiary sector, these changes cannot apply to my position
  - (h) Other
22. Why do you want this/these change/s in your work organization?
- (a) It would reduce the number of weekly trips
  - (b) It would be favorable for my work-life balance
  - (c) I would be more efficient and productive
  - (d) This would optimize my working day: business trips, meetings outside, etc.
  - (e) I would avoid transport congestion
  - (f) I would avoid traffic congestion
  - (g) No changes desired
  - (h) Other
23. What is stopping you from making this/these change/s?
- (a) This possibility is not offered by my company
  - (b) My position is not eligible or suitable
  - (c) It is possible within my company, but my direct hierarchy refuses
  - (d) My personal organizational constraints prevent me from doing so (I have to pick up my children at fixed times, etc.)
  - (e) No changes desired
  - (f) Other
24. In the morning, what would be your ideal arrival time on site?

- (a) Before 8 a.m.
  - (b) Between 8 a.m. and 8:30 a.m.
  - (c) Between 8:30 a.m. and 9:00 a.m.
  - (d) Between 9:00 a.m. and 9:30 a.m.
  - (e) Between 9:30 a.m. and 10:00 a.m.
  - (f) Between 10:00 a.m. and 10:30 a.m.
  - (g) After 10:30 a.m.
25. To improve your commuting, what solutions would be of interest to you when you next return to the workplace?
- (a) Better real-time information about the state of traffic in public transportation
  - (b) A better transportation offer (frequency, line)
  - (c) Secure cycling facilities
  - (d) Additional car parking lots
  - (e) Secure parking for bicycles
  - (f) Access to a bike/scooter rental or loan by your company
  - (g) The possibility of traveling with my bike on public transport
  - (h) A simple carpooling organization system
  - (i) Charging stations for electric vehicles
  - (j) I am satisfied with my current organization
  - (k) Actions in my company promoting the use of bicycles (shower layout, bike repair, etc.)
  - (l) The implementation of the Sustainable Mobility Package
  - (m) Other
26. To improve your work organization, what solutions would interest you when you return to the workplace?
- (a) The possibility of telecommuting one day per week
  - (b) The possibility of telecommuting two days per week
  - (c) The possibility of telecommuting three days per week
  - (d) The possibility of having more flexible hours (arrivals and departures outside peak hours)
  - (e) I am satisfied with my current organization
  - (f) Coworking spaces/third places near my home/on my routes
  - (g) Other

#### **A.4 . General information**

27. What is the name of your company?

28. What is your socio-professional category?
- (a) 2. Craftsmen, traders, entrepreneurs
  - (b) 3. Executives and intellectual professions (liberal professions, civil service executives, company salespeople, engineers and technical company executives)
  - (c) 4. Intermediate professions (technician, supervisor, etc.)
  - (d) 5. Employees
  - (e) 6. Workers
  - (f) 8. Students
29. What is your residential postal code? (This information is useful for us to understand your travel choices)
30. What is your age group?
- (a) Under 30
  - (b) Between 30 and 40 years old
  - (c) Between 40 and 50 years old
  - (d) Between 50 and 60 years old
  - (e) Over 60
  - (f) I don't want to specify
31. You are?
- (a) A man
  - (b) A woman
  - (c) I don't want to specify
32. Do you live alone?
- (a) Yes
  - (b) No
  - (c) I don't want to specify
33. Do you have dependent children? (This information is useful for us to understand your travel and work organization choices)
- (a) Yes
  - (b) No
  - (c) I don't want to specify

## B - Appendix B

Table B.1: Table showing the representative proportionality of the used sample.

		Department of Paris (75) (%)			Department of Seine et Marne (77) (%)		
		Sample	Population	Sample/Population	Sample	Population	Sample/Population
Gender	Female	37.14	53.00	0.70	50.68	51.32	0.99
	Male	62.86	47.00	1.34	49.32	48.68	1.01
Age Group	<30	26.39	37.38	0.71	10.71	39.53	0.27
	30-40	34.72	15.38	2.26	31.25	13.47	2.32
	40-50	25.00	13.05	1.92	27.68	14.01	1.98
	50-60	11.11	12.01	0.92	26.34	13.06	2.02
	>60	2.78	22.16	0.13	4.02	19.93	0.20
SPC	2	0.00	3.70	0.00	0.00	3.20	0.00
	3	83.33	29.80	2.80	54.46	10.50	5.19
	4	13.89	14.40	0.96	21.88	17.80	1.23
	5	1.39	12.00	0.12	20.98	19.20	1.09
	6	0.00	4.20	0.00	2.68	11.40	0.24
	8	1.39	17.50	0.08	0.00	16.50	0.00
Marital Status	Alone	27.94	49.10	0.57	17.73	60.60	0.29
	Couple	72.06	50.90	1.42	82.27	39.50	2.08
Dependent children	Yes	37.50	46.35	0.81	61.16	66.57	0.92
	No	62.50	53.65	1.16	38.84	33.43	1.16

		Department of Essonne (91) (%)			Department of Hauts-de-Seine (92) (%)		
		Sample	Population	Sample/Population	Sample	Population	Sample/Population
Gender	Female	59.84	51.04	1.17	34.15	52.43	0.65
	Male	40.16	48.96	0.82	65.85	47.57	1.38
Age Group	<30	9.63	39.75	0.24	12.20	38.30	0.32
	30-40	24.95	13.34	1.87	36.59	15.17	2.41
	40-50	30.45	13.79	2.21	34.15	14.12	2.42
	50-60	31.83	12.89	2.47	17.07	12.40	1.38
	>60	3.14	20.23	0.16	0.00	20.01	0.00
SPC	2	0.19	2.80	0.07	0.00	3.30	0.00
	3	50.10	13.70	3.66	95.35	26.00	3.67
	4	20.08	17.40	1.15	0.00	15.50	0.00
	5	25.73	17.20	1.50	0.00	14.20	0.00
	6	2.73	10.00	0.27	2.33	5.40	0.43
	8	1.17	17.40	0.07	2.33	16.70	0.14
Marital Status	Alone	17.26	59.50	0.29	20.00	55.60	0.36
	Couple	82.74	40.50	2.04	80.00	44.50	1.80
Dependent Children	Yes	61.99	65.80	0.94	48.84	60.96	0.80
	No	38.01	34.20	1.11	51.16	39.04	1.31

		Department of Val-de-Marne (94) (%)			Total (%)		
		Sample	Population	Sample/Population	Sample	Population	Sample/Population
Gender	Female	41.43	51.91	0.80	52.16	51.94	1.00
	Male	58.57	48.09	1.22	47.84	48.06	1.00
Age Group	<30	17.14	39.40	0.44	12.11	38.87	0.31
	30-40	41.43	14.43	2.87	28.81	14.36	2.01
	40-50	27.14	13.54	2.00	29.65	13.70	2.16
	50-60	11.43	12.68	0.90	26.30	12.61	2.09
	>60	2.86	19.96	0.14	3.13	20.46	0.15
SPC	2	0.00	3.20	0.00	0.10	3.24	0.03
	3	74.29	16.10	4.61	57.78	19.22	3.01
	4	7.14	16.30	0.44	17.84	16.28	1.10
	5	12.86	17.60	0.73	20.54	16.04	1.28
	6	2.86	8.70	0.33	2.49	7.94	0.31
	8	2.86	18.70	0.15	1.24	17.36	0.07
Marital Status	Alone	17.39	55.00	0.32	17.96	55.96	0.32
	Couple	82.61	45.00	1.84	82.04	44.08	1.86
Dependent Children	Yes	48.57	63.46	0.77	58.40	60.63	0.96
	No	51.43	36.54	1.41	41.60	39.37	1.06



## C - Appendix C

### The user interface of the proposed decision platform

The appendix provides a depiction of the user interface of the proposed decision platform. This interface is designed, to incorporate interactive components that facilitate ergonomic user interactions.

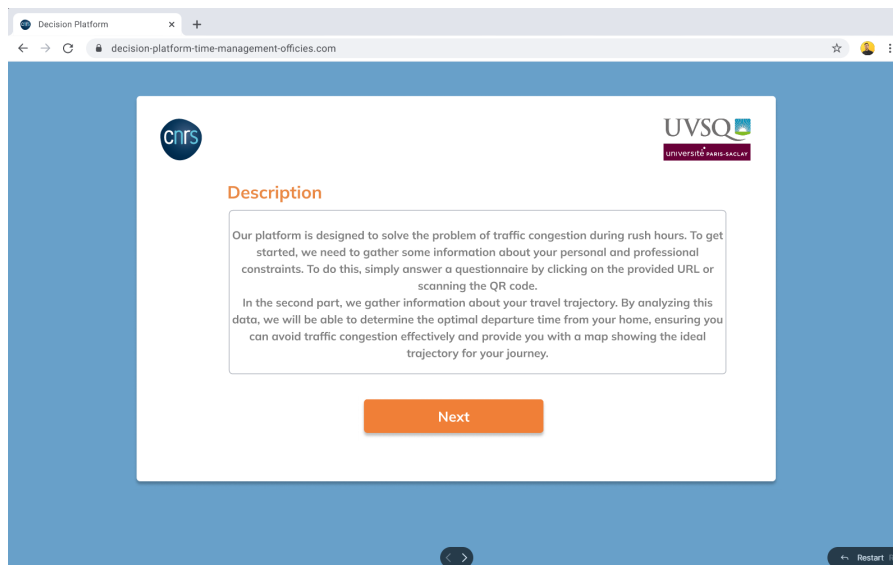


Figure C.1: The descriptive welcoming window.



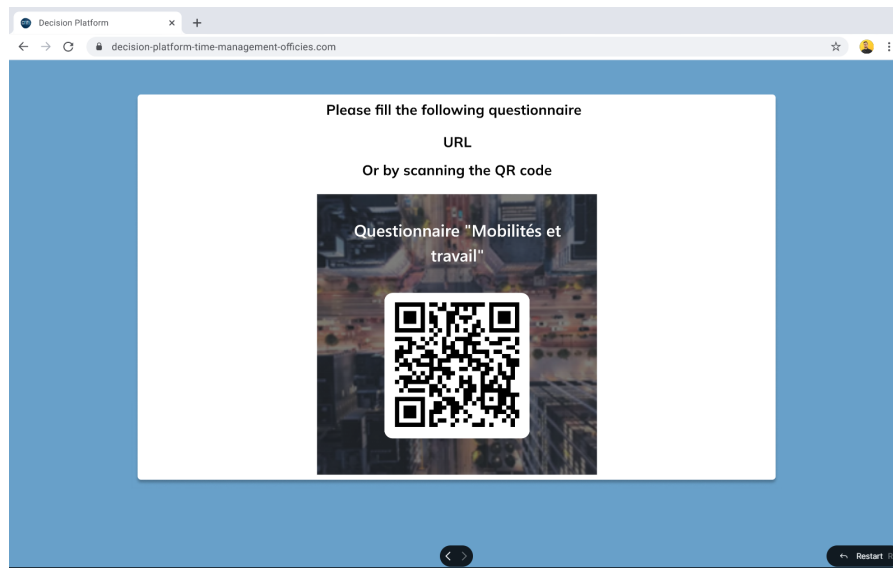


Figure C.2: The QR code of the questionnaire window.

The welcoming window of the platform's user interface in Figure C.1 is dedicated to the presentation of the decision platform, unfolding its design through a series of steps. The second window in Figure C.2 refers to the QR code associated with the questionnaire on the platform.

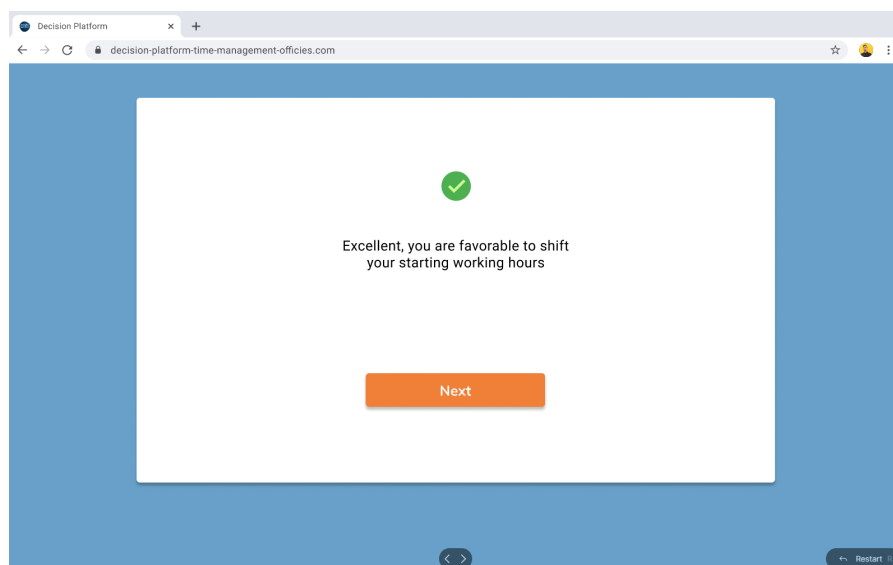


Figure C.3: Mobility profile identification processing result.

Decision Platform

decision-platform-time-management-offices.com

Start position

Arrival position

Preferred departure time

Next

Restart

Figure C.4: GPS coordinates with preferred departure time window.

The third interface window in Figure C.3 is designed to showcase the outcomes of mobility profile identification processing. Meanwhile, the fourth window, presented in Figure C.4, allows the user to provide her/his GPS coordinates and her/his preferred departure time information.

Decision Platform

decision-platform-time-management-offices.com

The best departure time is : 8h50

SRM

Rue de l'Internationale, Rue Henri Rochefort

1.3 km, 8 min

Head southeast on Rue de l'Internationale

Turn right onto Square Léo Frankel

Keep left onto Square Léo Frankel

Make a slight right

Turn right onto Boulevard Louise Michel

Turn left onto Rue Léo André

Turn right onto Square Charles Amégroux

Other selection

Restart

Figure C.5: Trajectory and temporality scenario suggestion window.

The last interface window (see Figure C.5) is designed to display the final result of the processing by suggesting a trajectory and the best temporality scenario among a list of others.



## D - Appendix D

### Résumé en Français

L'urbanisation et la croissance démographique posent d'importants défis pour la gestion des villes et de leurs ressources. Parmi ces défis, la mobilité urbaine suscite une préoccupation de tous les pays en raison de ses externalités négatives directes sur la qualité de vie, la productivité économique et la durabilité environnementale des zones urbaines. La congestion de la mobilité urbaine découle de la demande croissante de déplacements, entraînant une nécessité accrue d'infrastructures de transport et de gestion du trafic. Cela souligne l'urgence de trouver des solutions innovantes pour assurer une mobilité urbaine durable.

Cette thèse se concentre spécifiquement sur la décongestion de la mobilité urbaine pour une gestion efficace et durable des déplacements de personnes au sein des villes. Notre objectif est de développer une solution d'aide à la décision destinée aux Bureaux des Temps. Ces bureaux ont été inspirés de l'expérience italienne et adoptés en France pour coordonner les horaires des services publics avec les besoins des citoyens. Ils visent à résoudre les problèmes de congestion en réorganisant stratégiquement les horaires des activités quotidiennes pour éviter les pics de congestion.

Cette solution est soutenue par une plateforme de décision, qui repose sur une approche interdisciplinaire, combinant des concepts des sciences humaines et sociales, notamment la sociologie, avec les technologies de l'information et de la communication. Elle vise à améliorer la mobilité urbaine en exploitant des données hétérogènes notamment les réponses à un questionnaire lancé auprès de plus de 50 entreprises dans une banlieue parisienne entre mai et octobre 2021 dans un objectif de recueillir les opinions et les attentes des individus concernant leur mobilité et leur organisation du travail avant et après la crise sanitaire. La plateforme intègre également des données d'OpenStreetMap (OSM) pour la génération des trajectoires et des noms des routes et finalement les données de Waze pour l'analyse du trafic et de la congestion routière. Les modules de la plateforme comprennent la collecte de données, la classification et la quantification, l'analyse des données, et la prise de décision.

Les travaux de la thèse suivent une méthodologie composée de trois phases distinctes. La première phase identifie et prédit les profils de

mobilité, permettant des recommandations personnalisées des horaires de travail optimaux en fonction du profil sociologique, des contraintes personnelles et professionnelles des individus. La deuxième phase analyse la congestion à l'aide de données de trafic et de modèles de séries temporelles. Enfin, la troisième phase extrait le meilleur moment de départ en fonction des résultats des phases précédentes.

Nos travaux de thèse incluent également une revue de la littérature sur les solutions existantes pour réduire la congestion, un examen des techniques d'analyse de données. En outre, les analyses menées s'appuient sur des études sociologiques pour comprendre l'influence des caractéristiques socio-démographiques telles que, le genre, la situation matrimoniale, la catégorie socio-professionnelle, et avoir ou pas des enfants à charge sur la volonté des individus d'ajuster leurs horaires de travail pour diminuer la congestion de la mobilité urbaine.

En conclusion, cette thèse aborde le problème de la congestion de la mobilité urbaine en intégrant des composantes sociologiques et techniques. Les résultats escomptés comprennent des recommandations personnalisées pour la gestion du temps de travail, des analyses de trafic optimisées et une plateforme décisionnelle pour les Bureaux des Temps afin de favoriser une mobilité urbaine plus efficace et durable.

De nombreuses pistes de recherche peuvent être explorées pour étendre ces résultats. Un aspect crucial à considérer est la reproductibilité de notre travail pour d'autres régions. Il est important de noter que, bien que nos travaux portent sur la région Île-de-France, ils peuvent varier considérablement dans d'autres territoires présentant des caractéristiques distinctes telles qu'un transport public limité ou des démographies professionnelles différentes.

Une autre piste potentielle d'extension de nos travaux de recherche implique l'exploration de techniques alternatives d'analyse des données et l'évaluation de leur efficacité, telles que les techniques d'ensemble ou d'association.

L'intégration de la mobilité active et durable, l'exploration de modes de transport alternatifs, l'incorporation du télétravail comme solution pour réduire la congestion, et l'utilisation d'autres sources de données non biaisées, sont également des perspectives de recherche intéressantes. Enfin, le déploiement de la solution proposée dans des environnements urbains réels est essentiel pour évaluer l'efficacité de l'approche.