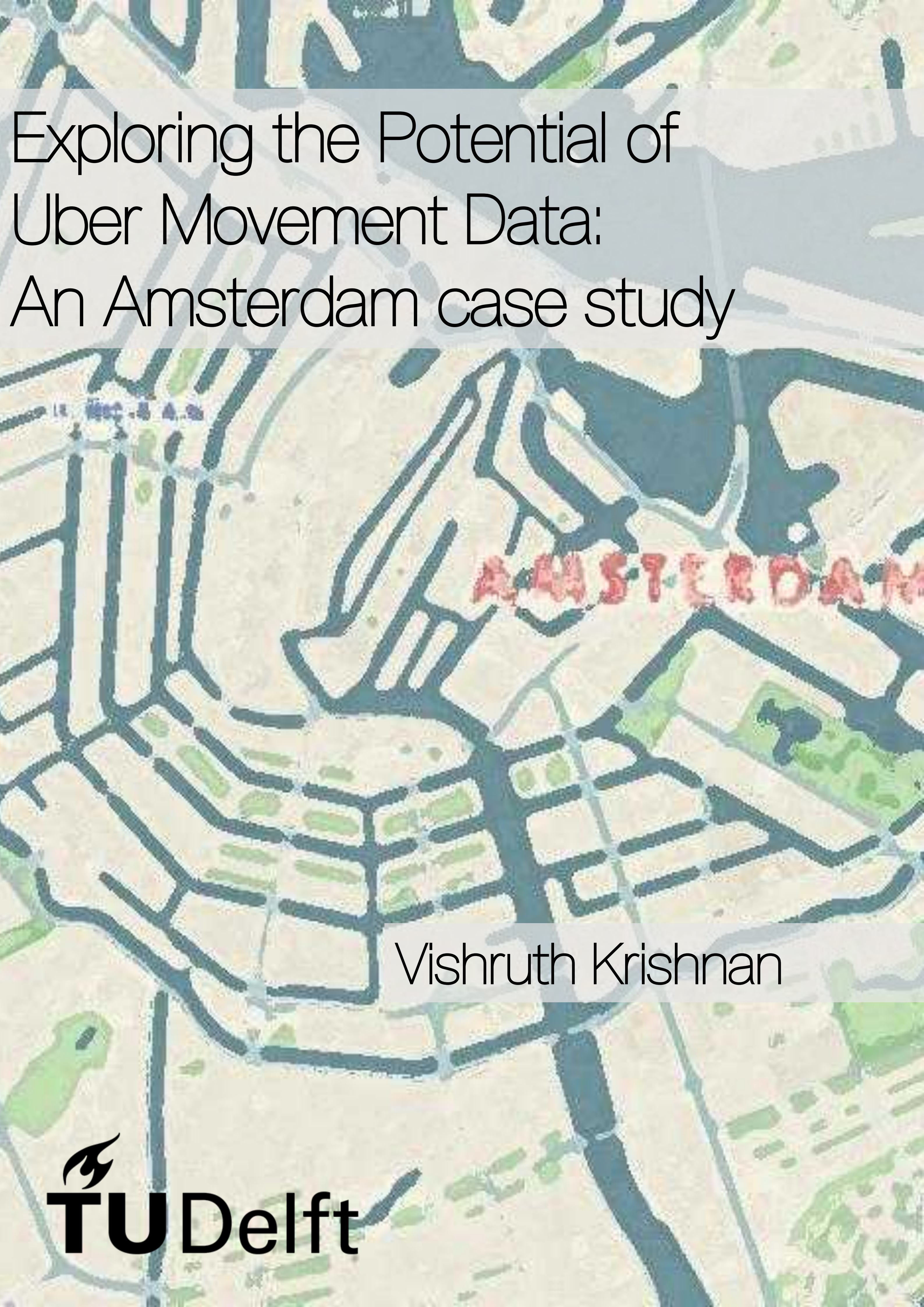
Exploring the Potential of Uber Movement Data: An Amsterdam case study

Vishruth Krishnan



TU Delft

EXPLORING THE POTENTIAL OF UBER MOVEMENT DATA - AN AMSTERDAM CASE STUDY

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Civil Engineering

by

Vishruth Krishnan

2019

Vishruth Krishnan: *Exploring the Potential of Uber Movement Data - An Amsterdam Case Study* (2019)

The work in this thesis was made in the:

Department of Transport & Planning
Faculty of Civil Engineering & Geosciences
Delft University of Technology

Gemeente Amsterdam
CTO, Mobilab

Supervisors: Prof.dr.ir. J.W.C van Lint (CEG)
Dr. ir. S.C. Calvert (CEG)
Prof. dr. ir. Alessandro Bozzon (IDE)
Tom Knijff (Gemeente Amsterdam)

ABSTRACT

With the increasing use of Big Data in varied applications to improve decision making and provide new insights, the research explores the potential of the Uber Movement data set released by Uber comprising of travel times from one zone to the other. A better understanding of the potential of the dataset could lead to the addition of the existing tool kit of Transport planners and city officials at the municipality of Amsterdam. Moreover, it would be the first of a kind data set enabling an understanding of taxi movement in the city.

The Uber Movement Travel Time comprises of the average travel time between two wijken, where the 'sourceid' and 'dstid' do not correspond to the origin and destination of a trip but simply represent the directionality of the travel time measured. The data is aggregated across different levels of temporal detail and the number of data points directly corresponds to the level of temporal aggregation. For instance, if the quarterly aggregated data for the different days of the week is downloaded, the number of data points between a 'sourceid' and 'dstid' cannot exceed seven.

Three aspects of the data set were explored: 1) ability to capture the demand for Ubers 2) ability to capture recurrent congestion and 3) ability to capture non-recurrent congestion. While the data according to the Uber Movement and previously used instances, the data is suited for performance (recurrent congestion and non-recurrent congestion) and impact-related studies of the network. The absence of route related information limits the applications of the data. The potential of the data is also limited by the data sparsity. The potential of the data was best revealed through demand studies which indicated a skewed user group of tourists, airport users (to and fro), work-related trips and users using Ubers late at night. In addition, with respect to the goals of the municipality in managing traffic activity across different zones and time periods, by implementing and extending an existing model in the form of adding 'occupancy related measures' and 'shortest path'. Thus, based on the data penetration levels and travel time data, the model developed offers insights at a strategic level to the city in the form of Spatio-temporal concentration of Uber vehicles, occupancy levels through the day. The potential of the data lies in its ability to offer strategic insights to the city of Amsterdam and the greater Amsterdam region in the form of the unique Spatio-temporal spread of Uber vehicles across different hours of the day.

PREFACE

This document is the end product of 8 months of research at Delft University of Technology & Gemeente Amsterdam. The thesis is aimed at introducing a data set to the Urban Traffic Data ecosystem by highlighting its unique applicability. As big data continues to expand its footprint in the field of transport and planning, it was considered essential to address the dataset. The context and results from the research are largely Amsterdam specific. An exploratory approach was taken to understand the implications of the open source Uber Movement travel time data. The data set is tested for different possibilities related to its ability to capture taxi demand and reflect traffic congestion. Chapter 1 forms the basis of the research and Chapter 2 offers the state of the art. Readers can focus on the tables in chapter 3 at the end of each section to have an overview on the exploratory part of the research. The final phase of the research, involves the implementation of a model to offer the spatio-temporal distribution of Uber taxis. The model results are coherent with the conclusions derived from the demand studies in the exploratoy part of the research. The final chapter of recommendations lists the data attributes, the municipality should aim for, to enable an extensive analysis of Uber taxis in the city and to enable modelling their spatio-temporal distribution. The motivation behind focusing on the spatio-temporal distributon stems from the goals of the municipality to introduce access control in parts of the city to prevent taxis from causing congestion and pollutuion. Additionally, the goals of the municipality are to ensure electrification of all taxi fleets. The model results can form a basis for priortising charging infrastructure in the city. The research and the model are an essential first step to gain an understanding related to the user of Ubers. For broader implications of the research, the author suggests the 'Implications for Urban Data data ecosystem' section in chapter 6. An interview with someone who drives for Uber is also offered in Appendix C . The interview is not meant for scientifc validation of the study, but offers another perspective. My profound gratitude to Ottmar Francisca for agreeing to talk to me.

Vishruth Krishnan
Delft, December 2019

"All data is wrong but some are useful."
-A play on words on the original quote by George E.P Box

ACKNOWLEDGEMENTS

Eight months ago when I was about to start my Thesis, it dawned upon me, this might be my last chance to study, research, critique and discuss in a space where there are minimal consequences to things going wrong. The topic at hand being exploratory offered me the flexibility to try new things and in the truest sense, I got to work on a new problem every day (what dream jobs are made of!). In this chaos, Dr Simeon Calvert ensured I did not sway too far from the goal at hand, with almost incisive feedback on what can be done better and differently for which I am ever so thankful.

I would like to thank my supervisor at the Gemeente, Tom Knijff, who ensured I had all the resources in terms of data, the right people and made my experience at the Gemeente, a pleasant one. Tom, I sincerely hope to imbibe some of the efficiency you so easily exude. I would also like to thank Professor Hans van Lint, whose comments and opinions I looked forward to, for one they sometimes made me understand my own work better and made me realise the expansiveness of what was possible. My sincere gratitude to Professor Alessandro Bozzon, whose comments have enhanced the work carried out by ensuring its wider applicability and help shape the report.

My profound gratitude to Chris Olbert (Gemeente Amsterdam), whom I could always rely on for help, Panchamy Krishnakumari (DiTT lab) for all the help and meetings, Sanjay for helping with the programming, colleagues at Mobilab (Gemeente Amsterdam) for making the Gemeente an enjoyable and enriching experience and Adithya for coffee breaks and conversations during the long hours spent at the library.

I want to thank my parents, who offered me unwavering support, gave me the freedom of choice, the space to make mistakes and most importantly learn from them, my elder brother Vivek, who has probably had an impact on my likes and dislikes more than anyone, my friends at Transport & Planning, my housemates at House no. 8, Aart van der Leeuwlaan who offered me respite from the daily grind. Lastly, I want to dedicate this thesis to my Grandparents whom I miss dearly. They shaped me as a person.

To life long learning!

Vishruth Krishnan
Delft, December 2019
...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Algorithms

ACRONYMS

1 INTRODUCTION

The chapter provides the motivation, research questions, scope and scientific/social relevance of the research undertaken. The chapter is concluded with the outline of the report.

1.1 BACKGROUND, OBJECTIVES & MOTIVATION

Data dominates discourse. As the field of Transportation engineering continues to evolve, emerging data streams can potentially solve new and old issues in the field. There is an increase in the availability of data, thereby offering previously unavailable insights. Antoniou et al. [2019], attributes the enormous increase in data sources to a spurt in non-conventional data sources such as the internet of things, crowd-sourcing and superior computational means to store and gather large sources of data. Another important and well-acknowledged development is the omnipresence of personal mobile phones equipped with internet and GPS. Large scale data collected from GPS sources can now unravel mobility patterns within a city [Grauwin et al., 2015], they offer network-wide information on weather-related incidents, road closures, and their effect on traffic. The era of Big Data, Laney [2001] refers to as data sets which are characterised by volume (large size of the data set), velocity (the speed at which data is logged) and variety (range of data sources and types) could offer new insights in transportation. Big data extends itself well to transportation as it is dynamic, transient and stochastic. Transport operations involve a multitude of agents interacting with each other over space and time. Big data can capture the invisible patterns [Gkania and Dimitriou, 2018] and also address previously held notions about how mobility in cities work. The challenge is to develop a semantic which can be used by Transportation engineers to understand and apply this data. Three major impediments in developing better tools to understand data are; firstly, privacy concerns, data often needs to be aggregated and anonymised, which then leads to a loss of information; secondly only a few data sets are open source and can be accessed by researchers and city officials; and thirdly the lack of understanding and complexity of the data makes it difficult to use.

The objective of the research is to add a tool to the Transportation Planner's toolbox in the form of a new data set offered by Uber Technologies Inc., (simply referred to as Uber) under the Uber Movement initiative [Movement, 2019a]. Uber is a multinational company offering ride-hailing services, food delivery (Uber Eats), and bicycle sharing (JUMP). The Uber Movement initiative, at the time of writing the report, offered average travel time data from one zone to the other (administrative spatial unit or Traffic Analysis Zone) for thirty-six cities across the world. The travel times are derived from over two billion rides. The data is open source and aggregated to ensure the privacy of their riders. Figure 1.1 represents the visualisation of zones and travel times in the Uber Movement website. The website provides an easy to use interface, where one can download the travel time data between two zones for a particular date in the year or across all zones for a quarter of the year. Uber derived the travel time data from the GPS location of their drivers carrying passengers. The Methodology behind the same can be found in [Movement, 2019b].

Due to the method of aggregation, the trajectory or route of trips is lost.

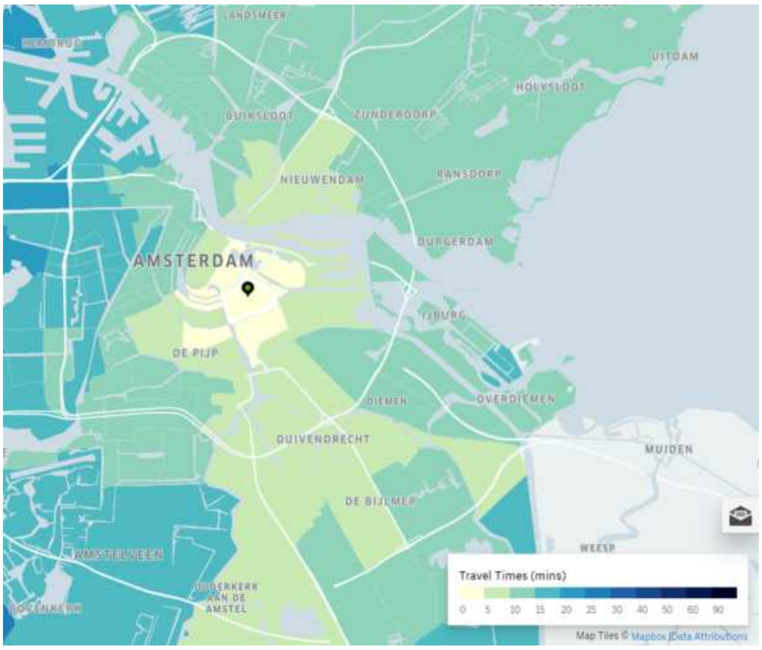


Figure 1.1: Visualisation of Travel time for Amsterdam in Uber Movement [Movement, 2019a]

The motivation of the research stems from the emergence of ride-hailing services such as Uber, Didi-Chuxing (China), Lyft (United States of America), Grab (Singapore) and Ola (India) who have capitalised on technological advancements with the advent of the platform economy. Ride-hailing has become a mainstream option in cities across the world and this has led to an important addition in terms of choices to the urban traveller/commuter. Their emergence can be attributed to advancement in smart-phones (enabling a mobile phone-based application), GPS (real-time position of the vehicles) and social media networks (user-driver ratings and interaction). Uber offered easy payment options and a door to door service without the hassle of finding a parking place or paying for it, in cities where parking can be scarce and expensive. These simplifications proved crucial in the rise of the ride-hailing service. On the user's end, users have been open to ride-sharing (the platform economy) and are open to sharing data for services [Deloitte, 2018]. The rise of transportation network companies suggests an increase in the availability of data-sets generated by ride-hailing services and the research tries to establish the potential of one such dataset through exploratory research.

In addition to this, there is a growing crop of datasets derived from users offered by software companies. The intended audience of these datasets is city and transport planners. An example of such a company is Strava, which through its Strava Metro initiative, has released counts (aggregated and anonymised) of cyclists, and pedestrians commuting or pursuing recreational activities while using the Strava application. Strava, claims the dataset can be effectively used to understand the movement patterns of such users in cities, offering cities a way to evaluate and possibly construct new infrastructure targeting pedestrians and cyclists [Strava-Metro, 2019]. Populus.ai is another organisation offering real-time data to cities from shared mobility modes such as bikes, scooters, and ride-hailing services in the United States of America [Populus.ai, 2019]. However, these datasets are not open-source unlike Uber Movement but are derived from the platform economy. Thus, the rise of the platform economy and the subsequent generation and availability of data generated has presented an opportunity to understand the context in which the Uber Movement data set can fit.

According to Uber, the dataset is meant for Transportation Planners and Policy Makers to understand congestion patterns in the city. As part of the exploratory research, three different objectives are formulated to understand the potential of the data set. The first objective relates to identifying how the data can fit in, considering the larger urban traffic data ecosystem. It is important to understand the context in which the dataset is available i.e. the usage patterns of Uber would be different across cities, the data is offered. The scope of the research is restricted to the Amsterdam Uber Movement data set. Thus, the first objective of the research can be thus stated as

Research Objective I: Establishing the advantages, limitations and the unique value of the Uber movement data for a Transport Planner in the Amsterdam context.

To further enhance the insights gained from a data set, multi-source data fusion is adopted. The need for data fusion results from the need to improve data output quality. Data sources such as Public Transport smart card data, Mobile phone data, Bluetooth data, Automatic Vehicle Location data, Social Media data from sites such as Twitter (in terms of Traffic Engineering), Navigation systems such as WAZE & Tom-Tom host information about road infrastructure and congestion patterns. Integration with one or more of these data sets can be especially valuable in the case of travel time data between zones which is an indicator of the travel costs involved between two points. Thus, the intention is to identify how the Uber Movement data set can be integrated with other relevant data sets. The second research objective can be stated as follows:

Research Objective II: Establishing the possible synergy gains by fusing Uber Movement Data with other relevant data sets.

Data needs to be translated into information. The information needs to be relevant to the problem it is addressing and needs to be based on the findings of the first and second research objective, a potential application of the data will be implemented to better realise the usefulness of the data. The third research objective is addressing the absence of a tool or model which utilises the unique value addition offered by Uber Movement to provide a valuable insight which leads to an application and can be stated as follows:

Research Objective III: Identifying and estimating a model for translating Uber Movement Data into actionable insights

Data is finding its way to make traffic models more realistic and accurate. Understanding the significance of Uber Movement data in traffic congestion analysis and management, will lead to better utilisation of existing infrastructure and prevent information excess in the era of big data. An in-depth exploratory analysis would serve as a blueprint for the use of data gathered from ride-hailing services while also creating a new tool for the city of Amsterdam derived from the added value of the data set. Thus, the main objective of this research can be succinctly stated as identifying the potential of the Uber movement Travel Time Data for the city of

Amsterdam, leading to the development of a tool or model which can lead to a Congestion analysis or Traffic Management application.

1.2 RESEARCH QUESTIONS

The research objectives and motivation of the project have been specified in the first section. The second section will elaborate on the formulation of research questions. The research objectives specified establish the need for an exploratory study to understand what the data is capable of offering. This leads to the following main research question:

Main Research Question: What is the potential of using Uber Movement Travel Time Data, either singularly or in fusion with other data sets for Traffic Congestion Analysis and management in the city of Amsterdam?

To answer the main research question, a list of sub-questions are formulated. The first sub-question formulates the need to highlight the state of the art, concerning the use of Uber Movement data and other Taxi GPS data. A review of related work and literature study can offer an overview of methods to use the data and help identify possible applications. Besides, data sources which can be fused with Uber Movement for synergy analysis can also be identified. Thus, the first research question can be formulated as:

Sub Question 1: What is the current state of the art concerning the use of Uber Movement Travel Time Data and other Taxi GPS data sources for applications in Traffic congestion analysis and management?

To understand the potential of the data, one needs to understand the context of the data set. Answering the question of 'Why' and 'Who' are the people using Uber in Amsterdam, is crucial to understanding the data and its eventual application. The second sub-question can thereby be formulated as:

Sub Question 2: Which user groups are likely to use Uber in Amsterdam and for what purpose?

Congestion can be recurrent or non-recurrent as specified in Skabardonis et al. [2003]. Recurrent congestion is caused by demand fluctuation during certain times of the day due to the design of the road, or operation measures. Non-recurrent congestion, on the other hand, is incident related such as an accident, an unexpected Jam or bad weather. To perform a thorough congestion analysis with Uber Movement Data, both need to be looked at separately. This leads to the third and fourth sub research questions:

Sub Question 3: To what extent can recurrent congestion analysis be carried out using Uber Movement Travel Time data, either singularly or fused?

Sub Question 4: To what extent can non-recurrent congestion analysis be carried out using Uber Movement Travel Time data, either singularly or fused?

The fifth and final sub-question looks to answer the need to identify the unique value of the data set in addition to the existing data sets being used in Amsterdam for traffic congestion analysis and management.

Sub Question 5: What is the unique value addition of the Uber Movement Travel Time data set to Transport Planners and officials at the city of Amsterdam?

1.3 SCOPE

The current section defines the scope of the research. The research utilises Travel Time data from Uber movement for the city of Amsterdam, from 2016 to 2018. At the time of writing the report, travel time data was available until the first quarter of 2019. However, the decision was made to analyse the data for three full years. While the data exclusively comprises of trips made by Uber's taxi services, the company claims it has validated the data to ensure it represents the regular movement in a city. The data includes the arithmetic mean, geometric mean and standard deviations for aggregated travel time over a select date-range between two zones or a quarter of the year between all origin and destinations in the city.

The spatial granularity of the data is limited and is defined at the *Wijk* or the neighbourhood level for the city of Amsterdam. The data stems from urban traffic in the city of Amsterdam and the larger Metropolitan region of Amsterdam[1]. The analysis of the data will be across different temporal levels (hours, days, weeks, months and quarters of a year) to derive regular and irregular patterns of congestion. The applications to be suggested will be based on the context of Traffic congestion Analysis and the traffic management of taxis in Amsterdam. The dataset will also be fused with other data sets to improve the possible insights gained from the data. The data can understandably be used for different purposes. However, the research would focus on analysis and solution development from the perspective of the municipality. The analysis will be done at a macroscopic scale i.e. at a network-wide level dictated by the spatial detail of the data.

1.4 SCIENTIFIC & SOCIETAL RELEVANCE

Enriching existing data sources to better analyse and quantify mobility and accessibility problems is important for cities [Casadei et al., 2018]. The need to make justifiable decisions using limited resources requires insights based on data-driven models. Making use of data can help identify the pertinent problems in the city and help prioritise projects and measures. Moreover, gathering this data requires limited resources from the city's perspective. While motorways and major arterial roads are usually equipped with loop detectors, inner roads and streets often lack the same. This is a challenge when estimating travel time in Urban networks. Understanding Uber data will also lead to a better understanding of data available from ride-hailing (and sharing) services. Their relevance signifies a continued stream of data.

From the perspective of congestion Analysis, understanding mobility patterns in a city can help in ITS (Intelligent Transportation Systems) applications for managing traffic, and route guidance. Use of taxi trip data, for traffic flow, has been extensively researched.[Liu et al., 2013] explored taxi trip patterns obtained from four days worth of taxi data, [Zhu et al., 2017a] explored Spatio-temporal mobility

1 The metropolitan region of Amsterdam: https://en.wikipedia.org/wiki/Amsterdam_metropolitan_area

patterns at the street level (data collected over a period of three months), and [Cui et al., 2016] looked at urban accessibility using taxi GPS data from a period of two months. The Uber data enables one to derive patterns based on data which has been gathered over a long period (3 years). The ubiquity of Uber Taxis (Figure 1.2) and longitudinal nature, leads to the potential of using the data in developing traffic models to evaluate emission levels, accessibility in a city, and predictive models to estimate travel times or congestion patterns such as in [Castro et al., 2012].

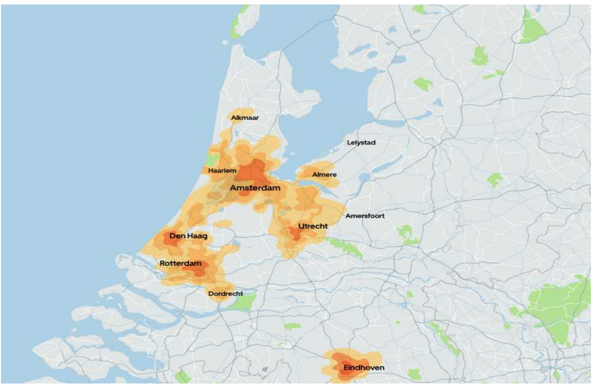


Figure 1.2: Uber service areas across the Netherlands [Uber Nederland, 2019]

Uber as one of the most valued private companies in the world holds immense data resources and is likely to expand its Uber movement portfolio with the addition of street speeds and data on it's shared biking company-Jump. It is important to acknowledge, Uber has experienced controversies and criticism. The company has been banned from operating in certain countries, found itself amid a 2016 data breach, and has been criticised for unfair treatment of its drivers [Pelzer et al., 2019]. However, Uber and other ride-hailing services are unlikely to disappear due to their ease of use and thereby remain relevant in the field of mobility.

1.5 REPORT OUTLINE

The outline of the report is as illustrated in Figure 1.3. In Chapter 2, the state of art concerning the work done using Uber Movement travel time data along with the broader applications of Taxi GPS data is summarised. Chapter 3, forms the principal data analysis and exploratory part of the research. The emphasis is on the usage patterns and demand for Ubers followed by recurrent and non-recurrent congestion analysis using the dataset. Chapter 4, utilises the understanding gained from Chapter 3, to rationalise the selection and application of a model employed in the research based on the data's unique value. The model is extensively described in the chapter. Chapter 5, offers case studies based on the model. Chapter 6 comprises of conclusions and the final recommendations are included in chapter 7.

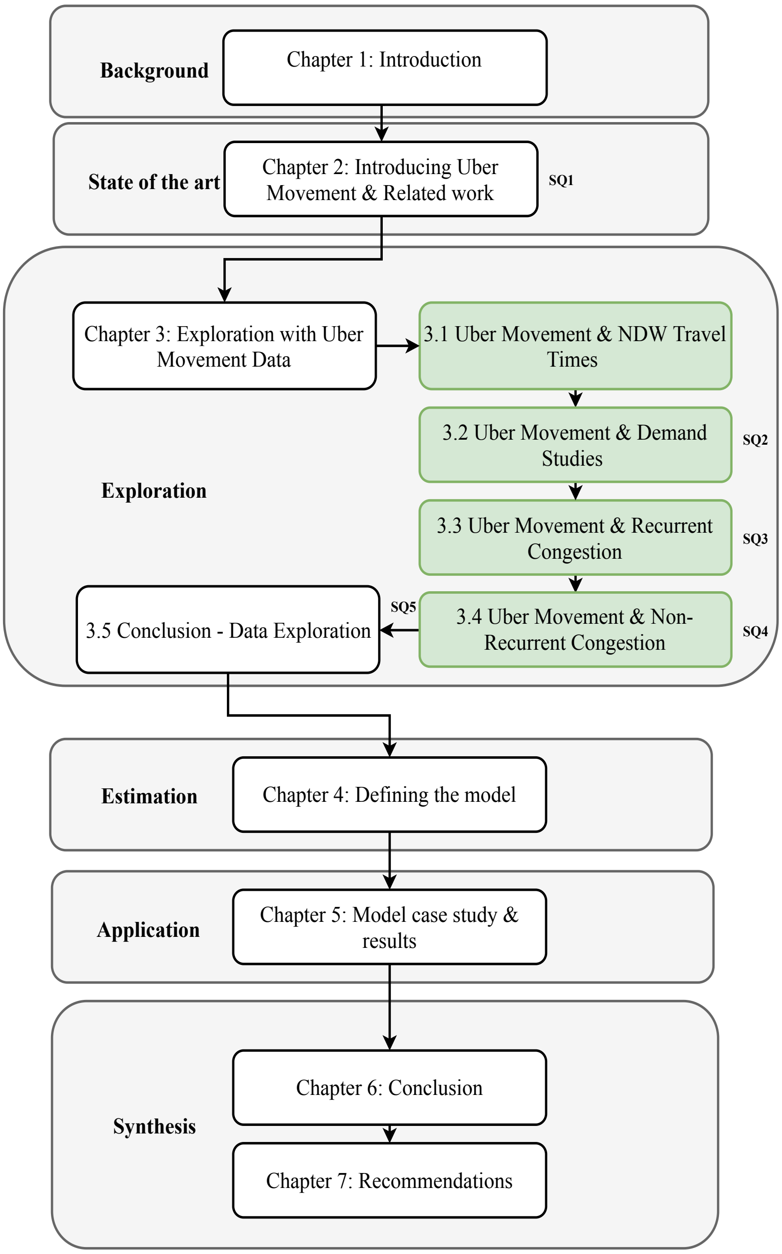


Figure 1.3: Report Outline

2 INTRODUCING UBER MOVEMENT & RELATED WORK

The chapter introduces the Uber Movement travel time data for Amsterdam and the related descriptive statistics in the first section. The introduction of the data facilitates an early understanding to streamline the literature review. The second section comprises of the literature survey on the existing applications and use of the data-set. The third section reviews the applications of other taxi GPS data in academia. The chapter is concluded with a section on the research methodology.

2.1 UBER MOVEMENT DATA

The section explains the Uber Movement Data in terms of the attributes of the data, spatial detail and temporal detail and offers the descriptive statistics of the dataset. The Uber Movement data was released in 2017 (2018 in Amsterdam) to offer transport planners and city officials, open-source historical travel time data from Ubers operating in cities. Uber Movement offers travel time from one wijk to the other for the greater Amsterdam metropolitan region across different levels of temporal detail. The data is collected from Uber taxis travelling across different wijken and comprises exclusively of occupied trips i.e trips with a passenger. The travel time calculation method is offered in the Uber Movement Methodology Report. Travel times are calculated from the Uber trip app which generates GPS signals, offering the trajectory and thereby the realised travel time across different wijken. The GPS signals are matched to a wijk and the trajectory information is lost [Movement, 2019b].

The Uber Movement data can be expressed as a four tuple (o_i,d_i,t_i,T_i); a data point comprises of the origin o_i at time period T_i, heading to destination d_i realised in Travel Time t_i. These attributes are typical of taxi GPS data. A fifth possible element is a route, which is absent from the Uber Movement data. The data is collected from Uber taxis travelling across different wijken and comprises exclusively of those trips when there was a passenger.

The attributes o_i corresponds to the 'sourceid', d_i corresponds to the 'dstid', t_i corresponds to the Time period: 'hour of day', 'day of the week' or 'month' for the quarterly data and across time periods available in the disaggregated data set which include the early morning period (00:00 to 06:00), the A.M peak (07:00 - 09:00), the midday (10:00 to 15:00), the P.M peak (16:00 to 18:00) and the evening time period (19:00 to 23:00). The different attributes of the data set are tabulated in Table 2.1.

The number of Ubers in Amsterdam is estimated to be 2100 (35% of the total 6000 taxis registered at Amsterdam) as of 2018 [Gemeente Amsterdam, 2019]. The number of Ubers offers an overview of the penetration rate which has a direct impact on the volume and quality of the data. Ubers are found to be limited in the Dutch Taxi market, owing to the existence of alternative modes and a well developed public transit system.

The data from the Uber Movement Website can be derived in two separate ways. One is by specifying an Origin wijk, Destination wijk followed by the desired date and time-period for which the travel times are required. For instance, travel time

Table 2.1: Attributes for the quarterly aggregated Uber Movement Data

Attributes	Description
sourceid	Origin wijk - Does not correspond to the pick-up point
dstid	Destination wijk - Does not correspond to the destination point
hod or dow or month	Temporal detail at which the quarterly data can be downloaded hod - Hour of Day dow - Day of week month
mean_travel_time	Mean Travel Time between origin wijk and destination wijk in seconds
standard_deviation_travel_time	Standard deviation with respect to the mean travel time
geometric_mean_travel_time	Geometric mean travel time between origin wijk and destination wijk in seconds
geometric_standard_deviation_travel_time	Standard deviation with respect to the geometric Mean Travel Time

between an Origin wijk and a Destination wijk on the 03/08/2018 for the morning peak. The second approach is to download the aggregated data which is aggregated at the quarter of a year and is available for all Origin-Destination pairs. Disaggregated data can be downloaded as four different '.csv' files as shown in Figure 2.1.

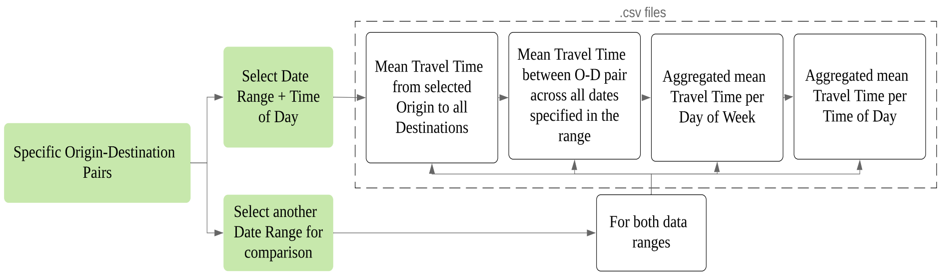


Figure 2.1: Deriving disaggregated Travel Time data by specifying an Origin and Destination problem

The aggregated data, on the other hand, can be downloaded for multiple Origin-Destination pairs. The data is available at different levels of temporal detail as shown in Figure 2.2. For instance, travel time by per hour of the day gives the mean travel time between an Origin wijk and Destination wijk between 08:00 to 09:00 for the second quarter of the year. Similarly, the mean travel time on a Monday for the fourth quarter of the year or the mean travel time between two wijken for a month of the year in 2018. Additionally, it is possible to distinguish between weekday and weekend data for the hour of day and months. The data points for disaggregated and aggregated data are released under the condition there are five unique trips and two unique drivers. This was done to ensure rider and driver privacy.

As specified previously, the research conducted utilises data from 2016 to 2018. However, the following chapters will utilise the 2018 data set due to it being the most recent data set available for a full year, has larger data coverage spatially (data points for more wijken). The descriptive statistics for the weekday 2018 data is tabulated in Table 2.2 and for the weekend data Table 2.3. The datasets do not require pre-processing as the data set does not contain empty fields, and despite the

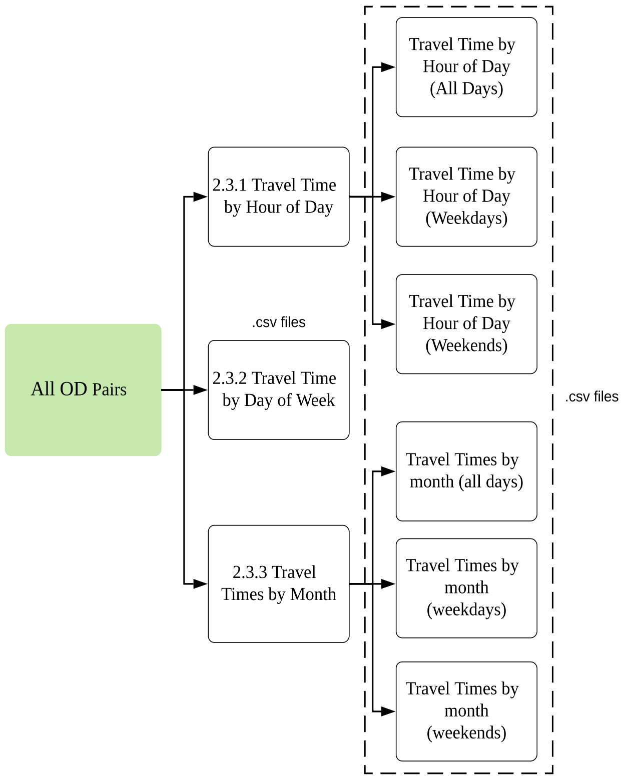


Figure 2.2: Deriving aggregated Travel Time data for multiple Origin and Destinations

presence of outliers, there is no rationale to remove them for analysis. The maximum standard deviation of mean travel time observed in the 2018 weekday data is 4608.28 seconds (76 minutes) between Slotermeer Norddoost and Weesperbuurt at 13:00 for the third quarter of 2018. This would mean a vehicle travelling between the two wijken could experience $\pm$76.76 minutes in addition to the mean travel time of 3519.33 seconds (58 minutes). However, for analysis, this reveals the variability in travel time experienced. A high standard deviation can indicate, travel time variability on the route caused by conditions on the route such as the opening and closing of a bridge, or a lack of data points.

The dataset also offers the geometric mean. While the arithmetic mean is the sum of the travel times divided by the total number of elements, the geometric mean is the product of the travel times followed by the under root as shown in Equation 2.1 where x_n represents the travel time between an origin and destination wijk. The geometric mean is useful in finding the central tendency of skewed data as it reduces the influence of large values. The travel times in an Urban network are an example of skewed data, with a long-tailed distribution where most of the values are normally distributed except for a few large values resulting in the long tail.

$$X_{geometric} = \sqrt[n]{\prod_{i=1}^{n}} = \sqrt{x_1.x_2.....x_n} \tag{2.1}$$

On conducting a Hypothesis test, to check if the distribution of travel times in the Uber Movement data set follows a heavy-tailed distribution, a Kolmogorov-Smirnov test was conducted. The null hypothesis of the test being, the distribution is normal and the alternative hypothesis being the method is significantly different

Table 2.2: Descriptive statistics for 2018 Hour of day weekday Uber Movement data

	sourceid	dstid	hod	mean_travel_time	standard_deviation_travel_time
count	1056845	1056845	1056845	1056845	1056845
mean	-	-	-	793.75 seconds	289.05 seconds
std	-	-	-	377.25 seconds	151.14 seconds
min	1	1	0	9.11 seconds	1.54 seconds
25%	-	-	-	518.29 seconds	197.98 seconds
50%	-	-	-	767.80 seconds	264.04 seconds
75%	-	-		1033.82 seconds	347.70 seconds
max	181	181	23	3737.71 seconds	4608.28 seconds

	geometric_mean_travel_time	geometric_standard_deviation_travel_time
count	1056845	1056845
mean	744.65 seconds	149.52 seconds
std	373.21 seconds	381.91 seconds
min	3.31 seconds	1.02 seconds
25%	471.28 seconds	1.29 seconds
50%	721.97 seconds	1.39 seconds
75%	984.72 seconds	1.56 seconds
max	3100.33 seconds	20.64 seconds

from a normal distribution and follows a log-normal distribution. The p-value was significant at 1%, proving the null-hypothesis as False and the distribution is indeed not normal. The Figure 2.3 represents the distribution of arithmetic mean travel time from all data points for 2018; both weekdays and weekends, indicating a heavy-tailed distribution.

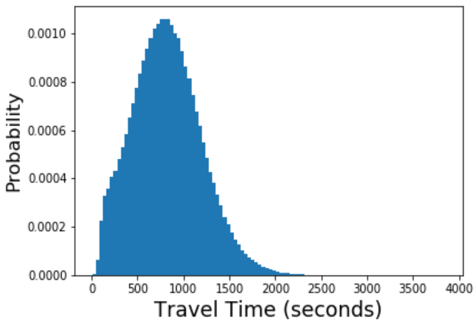


Figure 2.3: Log-normal travel time distribution for travel time in 2018

The geometric mean also has expectedly smaller standard deviations, as it is less sensitive to extreme values compared to the arithmetic mean. The use of the geometric mean for analysis will be a logical choice considering the nature of the data. The spatial detail of the Uber Movement Data set is at the wijk level. The wijk is a combination of neighbourhoods at which land-use data is provided by the Central Bureau of Statistics in the Netherlands. For a lack of an English equivalent, the research will refer to the spatial unit as a wijk and wijken in the plural. The wijken are not limited to the municipality of Amsterdam and extend over to the Greater Amsterdam Metropolitan region. The availability of data for the wijken varies temporally. Figure 2.4 represents the wijken and their area in km². As mentioned in Movement [2019b], the mean travel time from one wijk to another wijk, and the pick-up/drop-off point or the trajectory followed when passing through the wijken is lost in aggregation. This would suggest spatial coarseness as the wijk can be about 0.46 km² (Tuindorp Buiksloot) to 117 km² (Amstelveen). Thus, the Uber Movement data offers mean travel time from one wijk to the other along with the standard deviation. The section establishes the attribute, descriptive statistics and

Table 2.3: Descriptive statistics for 2018 Hour of day weekend Uber Movement Data

	sourceid	dstid	hod	mean_travel_time	standard_deviation_travel_time
count	964435	964435	964435	964435	964435
mean	-	-	-	729.43 seconds	267.84 seconds
std	-	-	-	337.14 seconds	132.23 seconds
min	1	1	1	7.17 seconds	1.46 seconds
25%	-	-	-	483.725 seconds	186.56 seconds
50%	-	-	-	712.16 seconds	246.04 seconds
75%	-	-	-	948.89 seconds	321.29 seconds
max	181	181	23	3246.570 seconds	4465.780 seconds

	geometric_mean_travel_time	geometric_standard_deviation_travel_time
count	964435	964435
mean	682.92 seconds	1.50 seconds
std	335.24 seconds	0.39 seconds
min	3.31 seconds	1.01 seconds
25%	437.46 seconds	1.29 seconds
50%	667.89 seconds	1.4 seconds
75%	904.12 seconds	1.57 seconds
max	3183.82 seconds	13.09 seconds

Figure 2.4: wijken available in the Uber Movement Data set and their area in km²

spatial/temporal detail of the dataset. The main insights from the section include; the data offers travel time data across different temporal details (Hour of Day, Day of Week, Month) for weekdays and weekends, it is possible to derive travel times on a specific date, however for a time period of the day i.e. the AM peak (07:00 to 09:00), PM peak (16:00 to 18:00), midday (09:00 to 16:00), evening (18:00 to 00:00) and early morning (00:00 to 07:00). The Geometric mean is a more relevant measure than the arithmetic mean for the uber movement travel time. The standard deviation is a possible measure to establish the validity of travel time variability i.e. travel times exceeding the mean ±standard deviation and be established as a case of when travel times were higher or lower than expected. In terms of some limitations identified; the spatial detail is coarse, especially the quarterly data, there is a lack of trajectory information, and the reliability of the travel time values (how the sample travel time available differ from the population) cannot be calculated due to the unknown number of Uber vehicles which resulted in the data point. The following section looks at the previous work done with the data set.

2.2 PREVIOUS WORK WITH UBER MOVEMENT DATA

The data source has not been actively explored in academia. We thereby explore case studies available at the Uber Movement website in addition to the research available from academia. The limited research with the data set could be attributed to the aggregated nature of the data, and the data has been made available publicly only in 2017. The first two subsections discuss the previous work done with the different aggregation levels of the data.

2.2.1 Congestion Analysis using disaggregated Uber Movement Data

The subsection discusses the case studies [Uber Movement, b] carried out using the Uber Movement data set. The case studies looked at Travel Time Variability caused due to infrastructure closure (closure of the London Tower Bridge and Disruption of the Washington Metro Rail Service), Event-related (Festival in New Delhi, India), and weather-related events (Flash floods in Nairobi, Kenya and Rainstorm in Pittsburgh, USA).

These applications use a base date range and time of day for comparison with the date range the non-recurrent event took place. For instance, when the London Tower Bridge was closed for three months (from 1st October to 31st December 2016). For comparison, the first month of the closure was compared to the first month of the year, across the PM peak period. A 65% increase in Travel Time was found travel South of the bridge and a 30% increase for travel North of the bridge. The spatial units selected were located at the entry and exit of the bridge. The rise in travel time was attributed to the limited number of links across the river [Uber Movement, a]. The table Table 2.4 summarises the different case studies conducted using Uber Movement.

Table 2.4: Uber Movement Case Studies done previously [Uber Movement, b]

Case Title	Type of disruption	Extent of travel time change
Analysing trends in 2015 holiday travel conditions	Holiday season travel patterns leading to longer travel times to the airport and shopping districts. Especially during off peak hours.	i) 70 minutes compared to 16 minutes (CBD to CBD) ii) 58 minutes compared to 37 minutes (CBD to CBD)
The effects of DC metrorail disruption	Effects of a system wide shutdown followed by a single line shutdown	10-30% higher than typical
Examining the impact of London Tower Bridge Closure	Infrastructure closure for a month	i) 65% increase in South bound travel times ii) 30% increase in North bound travel times
How March Floods affected Nairobi Travel Times	Effects of adverse weather	Travel time increase of 124% from the CBD to the South and East zones

The case studies utilise the web user interface (as shown in Figure 2.5) by Uber Movement to download data for a particular date or a range of dates on which the event of interest has occurred and compares it against a base range. Travel time variability caused by specific events has been investigated. Researchers and traffic operators term these types of incidents as incidents of non-recurrent congestion. Non-recurrent congestion is caused by unexpected events like traffic accidents, bad weather, jams, large-scale events, and infrastructure-related closures [Anbaroglu, 2013],[Dowling et al., 2004a]. Figure 2.6 represents the causal relationships leading to recurrent and non-recurrent congestion. Traffic Control Devices such as signal control affect the physical capacity of the network and on interaction with Demand and volume leads to recurrent or bottleneck congestion. Other causes for recurrent congestion are daily/seasonal variation, for example in the Uber Movement case studies; the holiday travel conditions in Manilla, Planned events such as the closure of the London Tower Bridge for one month and effect of a line shutdown in Washington DC for a period of fifteen days fall under the Special events category affecting the demand/volume of traffic. The effect of weather, incidents and work zones have also been identified, in the Uber movement case studies, non-recurrent congestion is caused by Weather in the Nairobi case, and the system-wide shutdown of the Washington DC metro rail for a day.

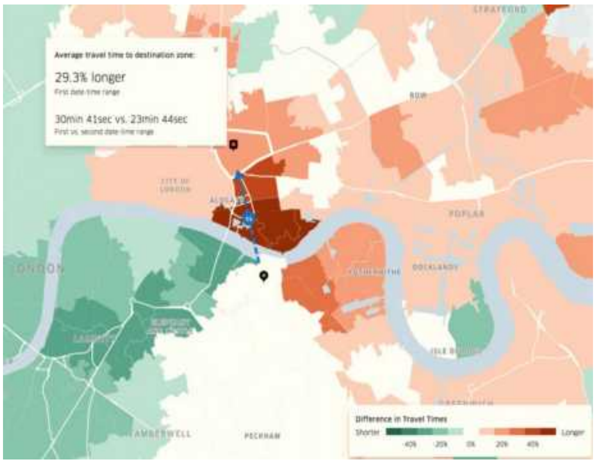


Figure 2.5: Travel Time comparisons before and after closing of the London Tower bridge - Uber Movement Web Interface [Uber Movement, a]

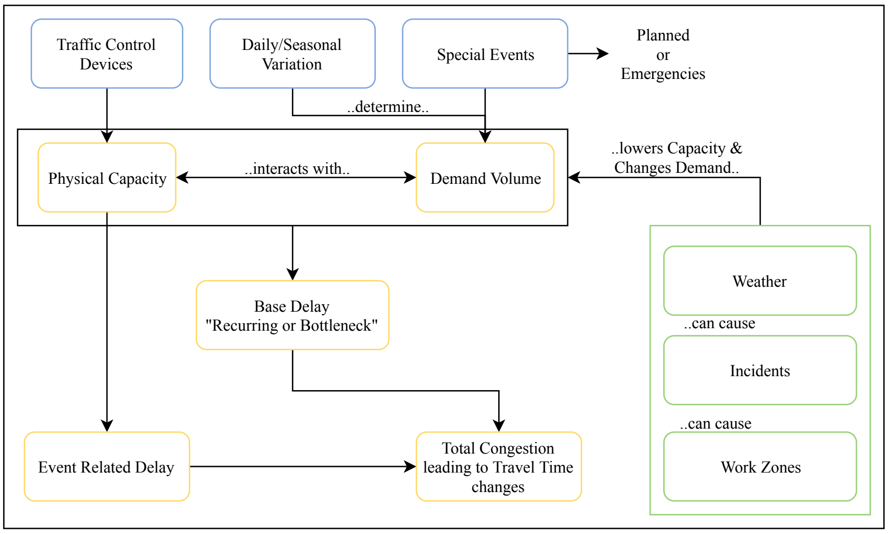


Figure 2.6: Factors affecting recurrent and non-recurrent congestion(adapted from Administration et al. [2004])

Thus, the data set has been used for looking at Travel Time variability. However, the method of checking for travel time variability relied on creating an arbitrary baseline comparison. For instance, in the London bridge closure case the travel times were compared to that of the month before the closure. The travel time changes in the case of localised incidents such as accidents have also not been explored. Thus, this creates the exploratory possibility of looking at the potential of the Uber movement data set for an ex-post analysis on the travel time changes caused recurrent and non-recurrent congestion.

2.2.2 Congestion Analysis using aggregated Uber Movement Data

The subsection discusses the use of quarterly aggregated data. Pearson and Samaniego [2017] attempted at representing temporal patterns of cities using Uber Movement Data. The Traffic Analysis zones (the Spatial unit of the dataset can vary across cities) were represented by a single node at the centroid, the travel time data was used to weigh the edges connecting the centroids, which enabled creating a static Spatial graph which represented the Euclidean distances between nodes and a dynamic temporal graph weighted by the travel time between the centroids. The indicators calculated include Degree centrality, Betweenness Centrality, Closeness centrality and Page Rank and HITS for both spatial and temporal graph. The research also utilises the community detection tool, which enables identifying latent communities based on distance and time using the Girvan Newman algorithm [Girvan and Newman, 2002]. The notion of communities is used to identify latent communities which exist in a social network. The latent communities are identified by identifying those communities which are connected more densely among each other than other sub-communities. Due to computational complexity, the method has shown to work better with sparser networks. A list of other methods used for community detection can be found in [Csardi and Nepusz, 2005]. There is no single best method to approach community detection. Thus, the research looked at how the mobility patterns change temporally over a day in comparison to a static spatial network. The paper offers a method to macroscopically represent Uber Movement data for cities based on network indicators.

The paper also attempts at discovering congestion hotspots, by weighing the Indegree using the mean Travel Time between two Traffic Analysis zones. Thereby, the resulting summation of travel time represents nodes which take higher travel time to reach. However, there was no mention of normalisation. Higher travel times can be a function of the distance. Also, if there are more edges required to reach a node, it would result in higher travel time in total. We will overcome the limitation by normalising the weighted in-degree by distance and the number of edges incident on the node.

Redelosa and Lim [2018] calculated degree centrality for the city of Boston and Manilla. The degree centrality for different zones was compared by plotting a cumulative distribution function. The centrality here is referred to as 'Used centrality' for indicating the usage across the nodes representing zones in the city. The research utilises the notion, that only after a certain number of traversals through a node can the travel time be made available (since Uber has to aggregate and anonymise). The underlying assumption being, if the number of travel time observations are higher for a node (Across weeks, months, quarters and years), the more traversed is the node.

> While the Uber Movement data does not offer the number of pick-ups and drop-offs in a spatial unit, the data penetration levels can suggest the extent of how often a spatial unit is traversed.

Both papers rely on Graph theory, which provides a framework for representing complex networks. A graph (G) can be composed of Nodes (V) and edges that connect the edges (E) i.e G = (V, E). The edges can be weighed and the nodes can represent spatial units. As both papers have already illustrated, network graphs are especially suited to represent the Uber Movement data. Degree centrality refers to the number of edges incident on a node. While, the measure offers insights on how frequently a node is visited, to describe the flow, the betweenness centrality is used. The indicator identifies the nodes crucial to the network, by indicating the number of shortest passing through the node. For instance, nodes in the centre of a city are likely to have higher betweenness centrality. The measure is sensitive to how the edges are weighed. In an unweighted graph (where the weight of all edges equal one), can result in a different set of nodes with high betweenness centrality compared to that of one weighted by distance or time.

> The Uber movement data is essentially travel time data which should be able to reflect travel time variability under the event of recurrent or non-recurrent congestion either singularly or when fused with another data set. Also, the network graphs are a useful way to represent the Uber movement network.

2.2.3 Uber Movement Travel Time Comparison

Wu [2018] compared Google Map Travel Time obtained from the Google Distance Matrix API with the Uber Movement Data Travel Times. It was found, Uber Movement Travel Times were systematically lower, a ratio of Travel Time from Google API to Uber Movement ranged between zero and three, with the heavier tail of the distribution ranging towards higher travel time.

The author speculates two possible causes for the longer travel time. The first one relates to how Google API travel times were aggregated spatially. The travel times derived were based on the address found at the actual centroid of the zone. However, in Uber's case, the driver may not have driven up to the centroid but somewhere in the zone. The other cause relates to how Google Maps, provides an Estimated travel time to the user and might include buffer time to ensure the satisfaction of the user. Whereas the Uber Movement Data is based on actually realised travel times by Uber vehicles. Thus, integrating data sources requires one to investigate the differences (mainly spatial and temporal in the case of Taxi GPS data) to be able to use and compare them. The article aggregated the spatial unit of origin-destination pairs data from the Google API as per that of Uber Movement for Sydney and also discarded intra-zonal trips. The centroid of the zone was selected for determining travel times in the case of the Google API. Thus, validating the differences between the data set. The work remains the only one which compares the Uber Movement Data set with another data set to test its validity.

> It is possible to establish the approximate error-bounds for travel times caused by the spatial aggregation of the Uber data set by comparing it with travel time data available at the coordinate level. another data set.

2.2.4 Travel Time Prediction

Uzel [2018] offered a method to use Uber Movement Data using a machine-learning approach for Travel Time prediction. For computational efficiency a sub-set of the origin-destination pairs were selected, actual distances between centroids of a zone were determined from Open Street Maps and combined with Travel Time from Uber Movement data. An algorithm termed Random Forest is then used for machine learning. Seventy per cent of the data set is used for estimation and the rest 30% hold out data for validation. The algorithm was able to estimate travel times with a 6% error rate. Considering the Google Maps Travel Time to be the base truth. The article suggested separate models for inner districts and the outskirts to reduce the prediction error. Especially since the inner spatial units are smaller and hence have a smaller error rate compared to the outer zones.

> The Uber Movement data set cannot only be used for ex-post analysis but also prediction.

2.3 TAXI GPS DATA SOURCES AND THEIR APPLICATIONS

As specified earlier, Taxi GPS data can take the four or five tuple format in terms of attributes. Depending on the source, collection method, and privacy issues, the data set may contain all or select dimensions. The following subsections comprise of the applications for which historical taxi GPS data has been utilised. Open Source Taxi GPS Data can be crucial in unravelling a city's mobility pattern. New York and Chicago have published GPS data from Taxi companies operating in the city for the use of traffic engineers, city planners and researchers. The data-set from Porto was made available for *Kaggle Taxi Travel Time prediction challenge*, and the T-Drive data set collected by Microsoft research. The Geo-Life data set for Beijing was also collected by Microsoft research. The San Francisco data set emerged from a Taxi GPS data collection initiative called Cabspotting and was collected to reveal mobility patterns across the city. Table 2.5 tabulates the different data sets on the basis of the collection period, Spatio-temporal granularity, number of taxis, thereby the penetration rate, number of trips and trajectory information (string of GPS coordinates offering trajectory information).

Table 2.5: Taxi GPS data from alternative sources and their attributes compared to Uber Movement

Dataset	Period of Collection	Temporal Unit	Pick-up Location
Amsterdam (Uber Movement)	01/2016-12/2018	Per hour	At the wijk Level
Chicago	01/2013 - 05/2017	15 minutes	Coordinates
Porto	07/2013 - 01/2014	Seconds	Coordinates
New York	01/2009 - 12/2018	Seconds	Coordinates
T-Drive(Beijing)	02/2008	Average - Seconds	Coordinates
Geo-Life(Beijing)	04/2007-08/2012	1.5 Seconds	Coordinates (GPS trajectories)
Cabspotting (San Francisco)	05/2008-06/2008	seconds	Coordinates

	Drop-Off Location	Taxis	Trips
Amsterdam (Uber Movement)	At the wijk Level	-	-
Chicago	Coordinates	>6000	>100 million
Porto	Coordinates	442	1.7 million
New York	Coordinates	>12000	>1 billion
T-Drive(Beijing)	Coordinates	10,357	15 million
Geo-Life(Beijing)	Coordinates (GPS trajectories)	178	25 million
Cabspotting(San Francisco)	Coordinates	536	11 million

	Trajectory information	Other information in the dataset
Amsterdam (Uber Movement)	No	Realised travel times and standard deviations
Chicago	No	Taxi fare, Taxi company, Payment Type
Porto	Yes	N.A
New York	No	Taxi fare, Taxi company, Payment Type
T-Drive(Beijing)	Yes	N.A
Geo-Life(Beijing)	Yes	N.A
Cabspotting (San Francisco)	Yes	Occupancy

On comparison with other open-source Taxi GPS data, the apparent lack of Spatio-temporal detail in the Uber movement data is evident. The New York and Chicago taxi GPS data sets would enable a much more disaggregated analysis. For instance, exploring the actual pick-up and drop-off points enables one to derive the activities pursued by the user, track actual trajectories and thereby the distance travelled. The Porto, Beijing, and San Francisco data sets have not been updated and in certain cases only subsets of the data are available. The Uber movement data is updated every quarter and comprises data for several major cities across the world. Applications built using Uber Movement Data can enable applications for the different cities.

In a broader sense, taxi GPS data is a type of Urban Traffic Data. The Urban Traffic data can be categorised as Supply Data, Demand Data, Performance Data and Impact data [Huang, 2003] for Transport and Planning decisions. Based on the level of detail available in the data, Taxi GPS data can fall under each of these categories. Following subsections discuss the applications of Taxi GPS data under each of the data types.

2.3.1 Taxi GPS data for Supply and Demand related applications

Taxi GPS data in terms of supply related applications can offer information related to the physical infrastructure i.e. the transport network. Zhu et al. [2017b] looked at the resilience of the New York transport network during adverse weather conditions by utilising the New York Taxi GPS data set and Subway ridership data. Variations in the usage were suggestive of the parts of the network which were usable and the extent to which they could be used. Liu et al. [2018b] utilised Beijing's taxi GPS data and employed AP (affinity propagation) and K-means clustering to identify hot-spots for Taxis which in turn can be used as car-sharing depots, to ensure an adequate supply of vehicles in urban environments. Bock et al. [2017] proposed a methodology to validate the parking sensor measurements with Taxi trajectories in San Francisco. A model was built to estimate the parking tendency of taxis and thereby derive the number of taxis parking at a location. This was compared to the parking sensor measurements. Thus, the taxi GPS data was used as crowd-sensing data.

Taxi GPS data in terms of supply can also relate to the availability of taxis in a certain area during a period. Thereby, the number of taxis available for riders at a point in time. Maintaining the supply-demand equilibrium for taxis is a complex problem. In Hangzhou, the trajectories of 5,500 taxis were analysed to determine taxi drivers best strategies to pick up passengers at a given time and location [Li et al., 2011]. A study in Shenzhen explored taxi-drivers operation patterns with the focus on differences between the behaviour of top drivers and ordinary ones. Ensuring the supply of taxis to meet the varying demand Spatio-Temporally is beyond the scope of the research [Liu et al., 2010].

Bischoff et al. [2015] analysed Taxi GPS data in Berlin and evaluated the demand-supply equilibrium, and found the peaks for taxi demand was well-matched with the supply. Additionally, the demand patterns for Berlin suggested most trips were between the city centre and Tegel Airport, most trips were found to be exclusively in the city centre, and central zones were likely to be destinations of taxi trips but outskirts were also found to be a popular destination. The research also seems to conclude points of special interest are a destination for taxi trips, such as train stations, the fairgrounds and major event locations. Temporally, it was found taxi demand is the highest during weekdays with peaks at 09:00 and a smaller one during the afternoon. During weekends the demand for taxis was found to be generally lower. In terms of asymmetry for taxi demand (taxi flow incoming and outgoing) which can prompt empty trips, the asymmetry was found to be the highest at the airport.

> Thus, taxi data can be evaluated in terms of demand which can vary spatiotemporally and help identify points of interest and also if there exists asymmetry in demand. The usage of taxis is often city-specific and highly context-dependent.

Yang et al. [2018] developed two linear regression models for pick-ups and drop-offs with land-use data as explanatory variables using Taxi GPS data from Washington DC. The research offers a method to assess taxi demand based on land-use data such as Industrial density, residential density etc. and measured an entropy factor which defines how monofunctional is the zone (a higher entropy is suggestive of varied land-usage) and can take a value between zero to one.

2.3.2 Taxi GPS data for Performance related applications

The second category of applications noted was the use of Taxi GPS data to evaluate the performance of the road network. Kuang et al. [2015], converted the road network in Harbin in China, to virtual nodes, with a defined traffic demand at every node. These nodes were then assigned a subregion and traffic anomalies (for instance: accidents) were detected at when the traffic flow conditions deviated from the usual. The modal share of taxis account for 23% in the city and was assumed to be representative of not only routes traversed by taxis but also other traffic. Zhan et al. [2013] employed the New York taxi data set to estimate travel times at the link level and [Castro et al., 2012] developed a predictive model for predicting future traffic states based on historical observations. The GPS data was sampled at every minute, obtained from 5000 taxis over a period of one month in Shanghai, resulting in 300 million GPS entries. The authors hypothesized a regularity in traffic flow obtained over a month enables prediction of future traffic states. The network was composed into edges and their orientation based on Taxi GPS pick-up and drop-off as the authors found trajectory points resulted in a cluttered plot. A probabilistic matrix (for the transition to a future state) was defined for traffic propagation from one edge-orientation pair to the other based on the Markov property of how an edge-orientation pair in the future is dependent solely on the immediate states of its neighbouring edge orientation pairs. The transition (time-step) took place at every fifteen minutes as the distribution variability between every minute was very high. So, the probabilistic matrix offered the probability of current density at one edge-orientation pair to flow to another edge-orientation pair. The model was based on one week's worth of data and was validated based on the other three weeks worth of data. In order to test the validity of the model, the mean density was calculated for every 15 minutes and compared with the predicted density by calculating the error involved.

The paper also offered a method to determine congestion as calculation of densities was considered an incomplete understanding of the city's dynamics, by determining the ratio of GPS vehicles with high speed to those in low speed, and set a ratio of 0.4 as an indicator for congestion. This can be interpreted as the number of vehicles moving in slower speeds (defined at 20 km/h) is 2.5 times more than vehicles at higher speeds. Thereby, indicating congestion.

The use of individual Taxi Trajectories, to identify vehicle traffic patterns has been carried out by [Keler et al., 2017]. The research utilised Taxi GPS coordinates and velocity data from seven thousand to ten thousand taxis in Shanghai to plot trajectories of individual vehicles. Trajectory Intersection Points (TIP) were derived by intersecting trajectories. It was found, a high-speed difference among the vehicles and a large number of trajectories, indicated a point of congestion in the city. The capacity of the road was quantified by plotting nodes on intersections using Open Street Maps and in turn using it as a dummy variable for capacity (more lanes, more nodes thereby indicating a higher capacity), as the authors were unaware of the actual capacity of roads in the network.

> Thus, Taxi GPS data can be used for evaluating the performance of the network but also the future states with respect to congestion and travel time prediction. The high penetration rates of taxis can enable a network-wide performance analysis.

2.3.3 Taxi GPS data for Impact related implications

The third category of impact-related implications involves measuring negative externalities such as pollution on the network. Lu et al. [2017] fused taxi GPS data, carbon emission data, road networks, point of interests in a city and meteorological data to predict carbon emissions. The paper offers a method to calculate emissions on a grid-based level and extends emission estimation from not only taxis but also other vehicles by estimating the flow of traffic in the trajectories followed by the taxis. The emission model developed enables estimation based on traffic patterns on road segments and showed the emissions are consistent with the peak and off-peak traffic flows in the city.

2.4 CONCLUSION AND NEXT STEPS

Three distinct application categories can be found for taxi GPS data from the literature. This includes demand and supply related applications, performance-related applications and impact-related implications. To establish the unique potential of the Uber Movement data set, the data set will be explored in terms of its ability to establish the demand for Ubers as part of demand studies and performance of the network in the form of recurrent and non-recurrent congestion studies. Impact related potential is not explored as it is a by-product of the data set being able to offer performance-related solutions. The outline for the exploratory phase is shown in Figure 2.7.

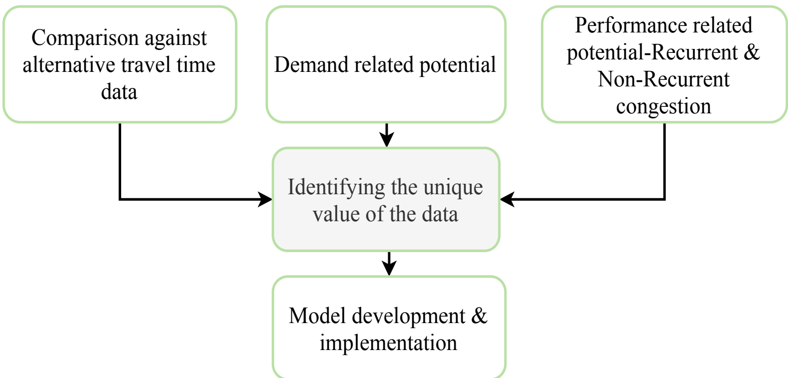


Figure 2.7: Next steps to identify the unique value of the data

3 EXPLORATION WITH UBER MOVEMENT DATA

The current chapter explains the exploratory part of the research. The insights gathered from the previous chapter with respect to the applications of Taxi GPS data are utilised here. For benchmarking the data, the data set is first compared against the Nationale Databank Wegverkeersgegevens (NDW) travel time data. This is followed by looking at the demand and supply related application of the data in Section 3.2 and after which the performance-related applications are evaluated in Section 3.3 and Section 3.4.

3.1 UBER MOVEMENT DATA & NDW TRAVEL TIMES

As the first step of data exploration, the travel times are compared against the NDW through the Dexter Web Interface which enables one to download travel time for the national highways (A - Autosnelwegen) and major arterial roads (the N - Provincial roads and the S- City roads) at different levels of temporal detail (minute, quarter, hour and day). The travel time data is collected from different sensors which include ANPR (Automated Number Plate Recognition), floating car data (sourced from mobile operators), road detector data for A roads (from RWS).

The travel times from Uber Movement data are realised from Uber trips with passengers in them. Liu et al. [2018a] illustrated taxi drivers knowingly take longer routes (detours) with non-local passengers in New York City from the airport whereas Uber drivers tend to take longer routes during surge pricing. Dynamic surge pricing is a phenomenon where the Uber fares go up during peak hours, due to limited supply and greater demand. The phenomenon is not limited to the peak hour and can be induced whenever there are limited supply and greater demand. The difference in route choice behaviour of Uber drivers will affect the travel times observed in the Uber Movement Data set. However, there is no empirical evidence of this happening in Amsterdam. The travel times are compared against the NDW travel times The comparison will help validate the data against a dataset adopted by policymakers in the Netherlands and also establish the approximate error bounds caused by the spatial aggregation.

The comparison of these two datasets requires Spatio-temporal matching. The difference between the two datasets is demonstrated in Table 3.1. The spatial matching is achieved by determining starting and end coordinates of the polyline from the NDW data and placing it in the wijk. For temporal matching, the NDW data is downloaded at per hour detail for all dates in a quarter. The mean travel time is calculated across all dates per hour of the day. For instance, the travel time now becomes an average of travel time at 00:00 across all dates in a quarter of the year. The standard deviation is also calculated for every hour of the day. The methodology for spatial and temporal matching is depicted in Figure 3.1.

The Spatio-temporal matching of the datasets is not straightforward. The direction of travel for which travel times are obtained need to be adhered to. Also, the

Table 3.1: Uber Movement and NDW Spatio-Temporal Detail

	Uber Movement	NDW
Spatial Detail	Wijken (administrative spatial units - combination of neigbourhoods)	Polylines represeting road segments
Temporal Detail	Travel Time per hour of the day per quarter of the year	Can range from one minute to a day. Available detail: 1 minute, Quarter, Hour and Day

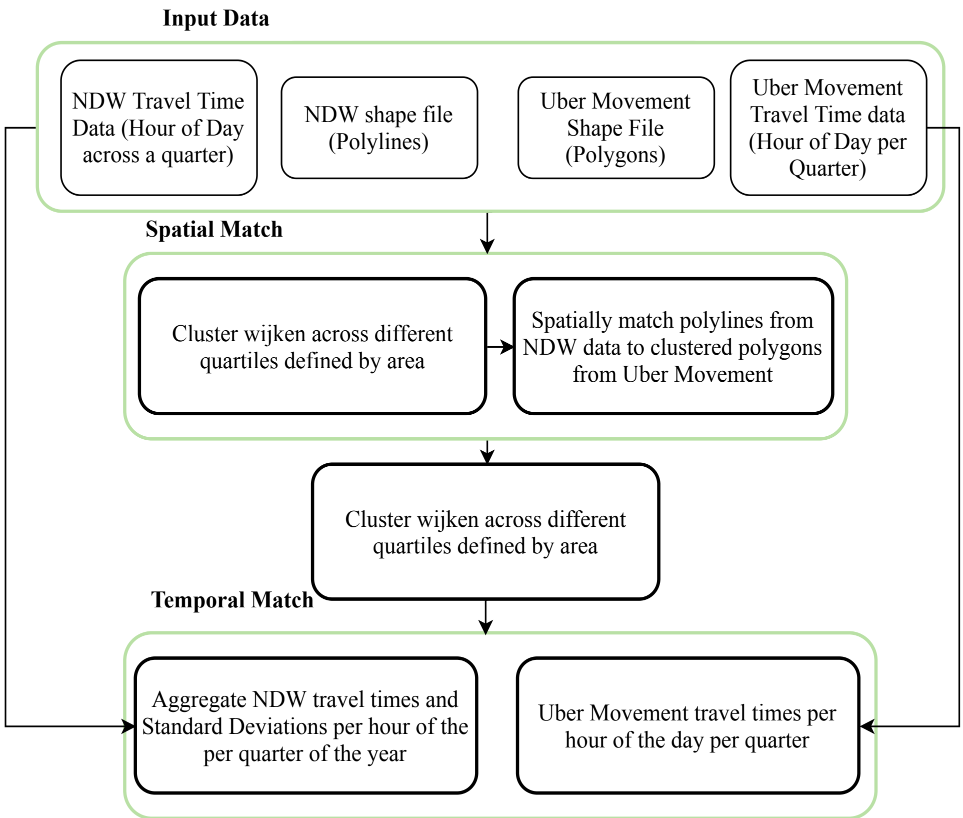


Figure 3.1: Methodology for comparison of Uber Movement and NDW travel time data

trajectory obtained will not necessarily correspond to the route taken by the Uber drivers and this is a limitation to be considered when interpreting the results. The decision to cluster the wijken according to the area in the Uber Movement dataset stemmed from a need to determine error bounds caused by spatial aggregation. The four clusters were based on the four area-based quartiles. The trajectory identified from the first quartile was from Burgwallen Nieuwe Zijde to Burgwallen Oude Zijde, the second quartile has a trajectory was from Sloterdijk to Spaarndammer, the third quartile has a trajectory from Frankendael to Overamstel and the fourth quartile has a trajectory from Badhoevedorp to Sloter and Riekerpolder. Each of the trajectories is presented in Figure 3.2.

For simplicity, the trajectories chosen were such that they lead from one adjacent wijk to others. The travel time differences have been tabulated in Table 3.2 while the speed differences can be seen in Table 3.3. It can be noted the travel time differences are the highest in the first quarter, while the speed differences are the lowest. These speeds become nearly equal, as the distance between centroids is higher than the length of the NDW trajectory (1.2km >0.518km). The different clusters do not suggest a consistent pattern of higher travel time differences in larger spatial units and smaller travel time differences in smaller wijken because the travel time errors are linked dependent. The length, capacity of the link, number of vehicles and the

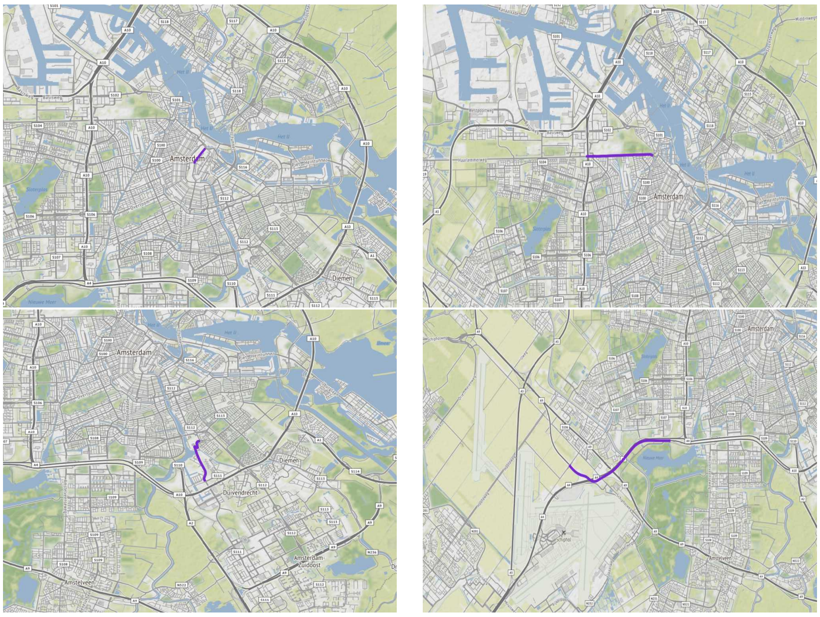


Figure 3.2: Selected segments from the NDW data- from left to right (a) Burgwallen Nieuwe Zijde to Burgwallen Oude Zijde (b) Sloterdijk to Spaarndammer/Zeeheldenbuurt (c)Frankendael to Overamstel (d) Badhoevedorp to Sloter/Riekerpolder

route taken would need to be established for the data to be comparable. The extent of noise in the Uber Movement data renders it difficult to establish the spatial error bounds. Thus, the comparison of the Uber Movement and NDW data while being an effective way to establish the supposed travel time differences, the unknown route taken by Uber vehicles is a significant limitation.

Table 3.2: Comparison of Uber Movement Travel Time data with NDW travel time data for trajectories across four clusters

	Trajectory	Mean travel time difference across all hours of the day (s)	% Mean travel time difference across all hours of the day	Distance between Geographical centroids from OSRM (km)	Length of the trajectory from NDW (km)
Cluster I	**Burgwallen Nieuwe Zijde to Burgwallen Oude Zijde**	118.35	59%	1.2	0.518
Cluster II	**Sloterdijk to Spaarndammer and Zeldenbuurt**	-96.55	-45%	3.65	2.58
Cluster III	**Frankendael to Overamstel**	40.85	17%	4.4	2.15
Cluster IV	**Badhoevedorp to Sloter and Riekerpolder**	-56.08941667	-34%	2.95	5.04

3.2 UBER MOVEMENT DATA – DEMAND STUDIES

The number of trips across an Urban area made by taxis can be valuable information in deriving the proportion of taxis in traffic and areas frequented by them. An understanding of the demand is fundamental to be able to manage the Uber Traffic in Amsterdam. The section aims to determine the usage patterns for Uber through the Uber Movement Data set i.e. the section will aim to answer 'why' are Ubers being used in Amsterdam. The insights gathered from this can help answer the second sub-question.

Table 3.3: Resulting speed differences through comparison of travel time data from NDW and Uber Movement

	Trajectory	Mean Speed difference across all hours of the day (km/h)	% Speed difference across all hours of the day
Cluster I	**Burgwallen Nieuwe Zijde to Burgwallen Oude Zijde**	-2.13	-9%
Cluster II	**Sloterdijk to Spaarndammer and Zeldenbuurt**	31.81	50%
Cluster III	**Frankendael to Overamstel**	27.781	40%
Cluster IV	**Badhoevedorp to Sloter and Riekerpolder**	-19.61	-29%

SQ2: Which user groups are likely to use Uber in Amsterdam and for what purpose?

An understanding of the demand and is fundamental to be able to manage the Uber Traffic in Amsterdam. To determine the user groups, the research first determines where are the Ubers being used by utilising the notion of data penetration to identify wijken which are frequently visited by Uber users in Amsterdam. The 'where' in turn is used to derive 'why' by fusing with Land-Use data.

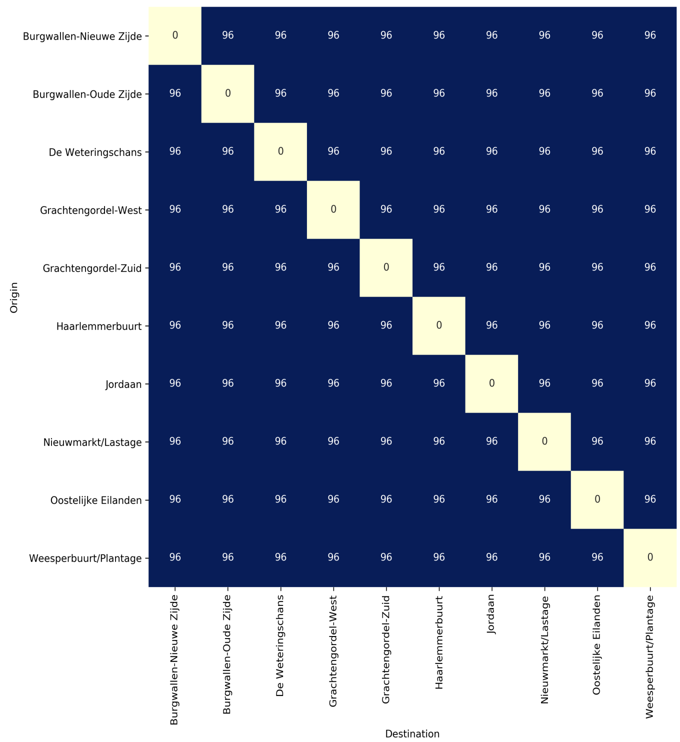
- The number of trips per hour or the production and attraction between one wijk to another wijk.
- The activities performed by passengers using the Uber Movement data.

The above two are determined across the following two subsections:

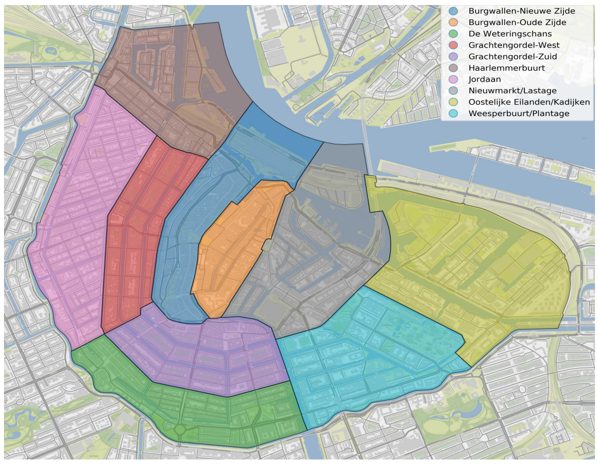
3.2.1 The production and attraction between wijken

To determine the production i.e. the number of Ubers originating from a wijk and the attraction i.e the number of Ubers ending their trip at a wijk, there is a need for the pick-up and drop-off location. Both of which is not made available in the Uber Movement data set as the aggregation method followed by Uber Movement results in a loss of all trajectory related information. This renders it impossible to derive the pick-up and drop-off locations. Exploratory analysis of the production and attraction revealed the data points are aggregated by the temporal detail i.e only one data point comprising of the source ID, destination ID and average travel time is released per temporal aggregation level. For example, a travel time data point is released between two wijken in the city centre per hour of the day per quarter of the year. Therefore, it is not possible to derive the actual number of Uber trips. Figure 3.3 indicates the data points between the wijken as ninety-six. This is a consequence of the twenty-four hours and the four quarters of the year.

The threshold for a data point to be released is five, and the number of unique drivers is equal to two. The absence of a data point leads to a qualitative indication, about the intensity of trips between two wijken. To better understand, the qualitative insights which can be gathered, origin-destination matrices were constructed representing the sum of production and attraction. An illustration of the same can



(a) Data points available for 2018 where 'sourceid' and 'dstid' were within the central ring - here the Origin-Destination pair represents a single data point



(b) Wijken represented in the central ring

Figure 3.3: Data penetration in the centre of Amsterdam

be seen for the morning peak in Figure 3.4. This can be especially useful to understand the data penetration during different time periods of the day. The maximum possible data points are 96 in the matrix. This relates to the four quarters of the year and 24 hours of the day. The matrix can help understand the relative flow between wijken. For instance, 'Amsterdamsewijk' has lower data points leading to Delftwijk compared to the movement to 'Haarlmerhoutkwartier' which resulted in the maximum possible data points of 96. This can indicate greater movement, on a qualitative basis.

The qualitative interpretation consequently results in a caveat. A data point can represent any number of vehicles, a single data point can represent 'n_i' number of vehicles and two data points can represent a number 'n_j' which can be less than

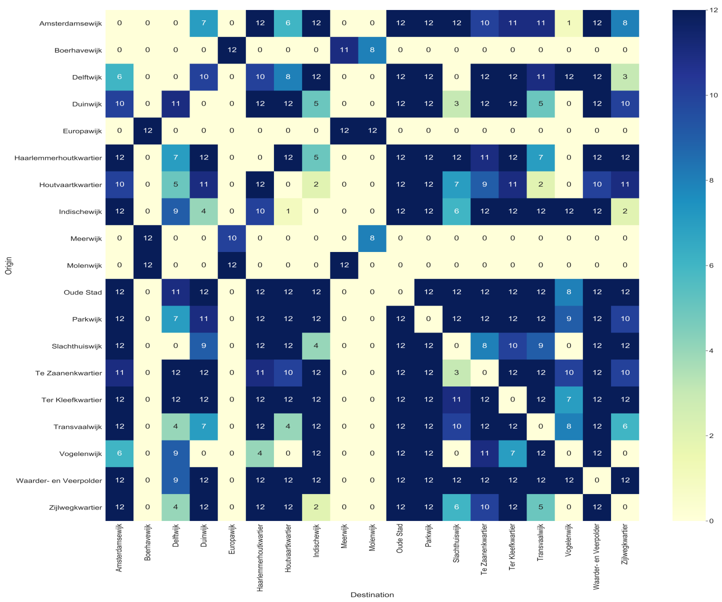


Figure 3.4: Data penetration for Haarlem - AM peak

'n_i'. Thus, there is limited applicability of the Uber Movement data in determining the demand apart from offering a qualitative indication. However, the research will utilise the qualitative insights and derive activities, and thereby the trip purpose in the next subsection.

3.2.2 Activities performed by Uber passengers in Amsterdam

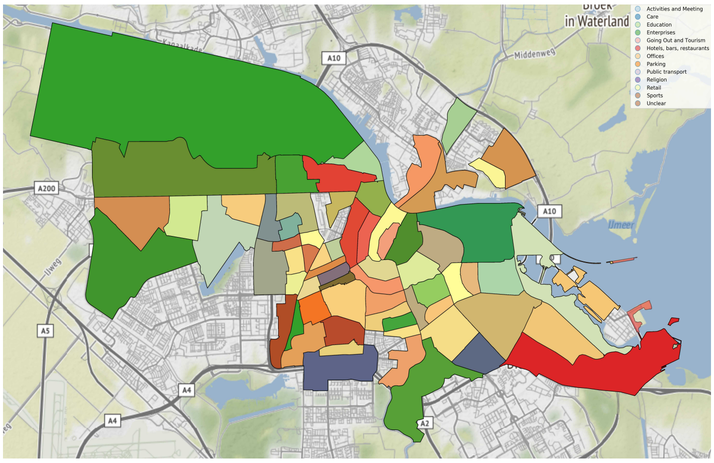
The activities performed by Uber passenger relate to the motives behind which Uber vehicles are used in the city. As was already determined in the previous section, the number of vehicles arriving or leaving a wijk cannot be determined. Instead, the research uses data penetration i.e. the number of data points made available (a summation of the 'sourceid' and 'dstid' attributes of the data). The data penetration across wijken is fused with the publicly available land-use data for the city of Amsterdam. The land-use data offers information on the functions of different buildings in the city. The functions may include retail, public transport, care, activities and meeting, education, sports etc.

Table 3.4: Uber Movement and Amsterdam open-source land-use data spatial detail

	Uber Movement	**Amsterdam Open Source Land-Use data**
Spatial Detail	Wijken (administrative spatial units - combination of neighbourhoods) across the Amsterdam metropolitan region	Building level function detail for the area within the municipality

The land-use data was integrated with the data penetration levels between 08:00 to 09:00 for the first quarter of 2018 as presented in Figure 3.5. The darker shades of green indicating a greater number of instances when the wijk had a 'sourceid' data point. The top three functions revealed include, retail at 32.1%, Hotels, bars and restaurants at 19% and Offices at 16.3%. However, this is not necessarily indicative of how Ubers are used due to the lack of pick-up/drop-off points and the spatial aggregation.

(a) Data penetration between 08:00 to 09:00 for the first quarter of 2018



(b) Function of wijken post fusion with Land-Use data

Figure 3.5: Integrating Land-Use data and comparing it with data penetration

The land-use split is not necessarily indicative of how Ubers are used due to the lack of pick-up/drop-off points and the spatial aggregation. The spatial aggregation is a drawback because a wijk in the centre of the city tends to have a variety of functions. On aggregation, the most dominant function is assigned i.e. if the majority of the businesses are retail-related, the wijk is assigned the land use function of retail. On calculating the land-use entropy, used to define the extent to which a zone has mixed land use, the entropy was found to be high for the wijken in the centre of Amsterdam. These have been tabulated in Table 3.5. A value closer to one indicates greater entropy and is calculated as shown in Equation 3.1. Here, P_k is the proportion of total land area of k^{th} land use category found in the tract being analysed and k is the number of land-use categories. The 'k' value can be constant and can be applied for every wijk, or depending on the availability of land-use categories per wijk, the value can vary. Both possible 'k' values were implemented and the tabulated results confirm the spatial aggregation to the wijk level prevents one from establishing the purpose of the trips made by Uber passengers. Additionally, the land-use data available was limited to the municipality boundaries of Amsterdam. wijken with activity outside the municipality has not been captured when integrated with land-use data.

$$\text{LandUseEntropy} = -\sum_k P_k \times \frac{ln(p_k)}{ln(K)} \tag{3.1}$$

Table 3.5: Land-Use entropy for wijken in the central ring of Amsterdam

Wijk	Land-Use Entropy with a constant k	Land-Use Entropy with a variable k
Burgwallen Nieuwe Zijde	0.67	0.71
Burgwallen Oude Zijde	0.60	0.62
De Weteringschans	0.58	0.60
Grachtengordel-West	0.60	0.61
Grachtengordel-Zuid	0.56	0.58
Haarlemmerbuurt	0.65	0.68
Jordaan	0.80	0.83
Nieuwmarkt/Lastage	0.75	0.75
Oostelijke Eilanden /Kadijken	0.73	0.78
Weesperbuurt	0.65	0.70

The Uber Movement data covers the greater Amsterdam Metropolitan Region and comprises of neighbouring municipalities. This is an advantage of the data set as it does not treat Amsterdam as an island and captures the movement across different municipalities which can account as longer trips than the ones carried out in the centre of the city. For instance, the data can capture travel times to Schiphol airport. It lies in the municipality of Haarlemmermeer and airports are traditionally an important spot for taxi movement.

As a next step, a heatmap for movement across municipalities from the 2018 hour of day data was plotted as shown in Figure 3.6. The heatmap is a normalised matrix, controlling the data points for the number of wijken in a municipality. The total number of data points were divided by the number of wijken. This is to avoid a skewed matrix where data points from Amsterdam dominate. Amsterdam is likely to have higher data points and as it has the maximum number of wijken. The figure while being labelled with an origin and destination only indicates if the 'sourceid' and 'dstid' data point belonged to the municipality.

The heatmap while once again qualitative indicate activity in Haarlem, Haarlemmermeer, Velsen, Waterland, Zaanstad and Zandvoort. The municipality of Haarlemmermeer comprises of Schiphol airport, Hoofddorp, Badhoevedorp, Bloemendaal and Vijfhuizen among others which are wijken with hotels and places of work near the airport. Additionally, Zaanstad is a tourist destination North of the IJ river. Zandvoort is also a popular beach destination away from the centre of the city. Bloemendaal also indicates activity and has a beach along with a major steel refinery. Activity in areas such as Heemstede and Velsen could just be due to the route taken to reach Zandvoort and Bloemendaal respectively. The heatmap suggests the use of Uber to tourist destinations especially those which are way from the centre of the city. This does seem logical as one would tend to walk in the dense centre of Amsterdam and rely on alternative modes to reach areas further away from the city. The trips to Schiphol does seem to suggest a principle user group of airport users, and those unfamiliar with the city.

To explore the temporal detail of the activities carried out in the wijken, the data penetration across a different hour of days was plotted along with the data points

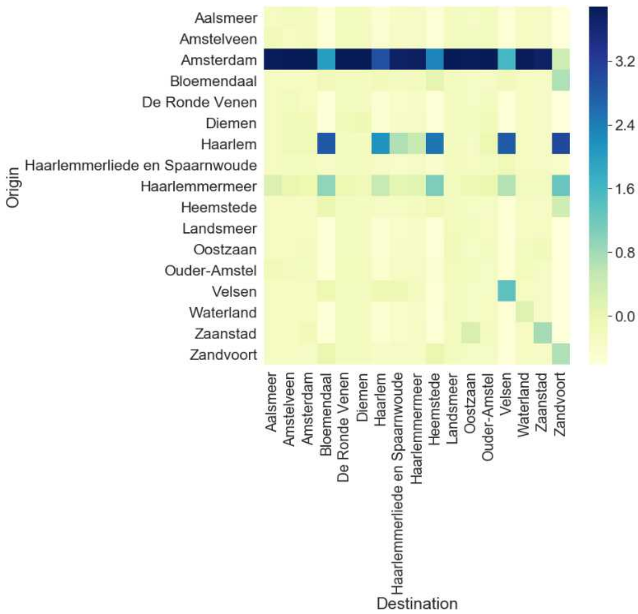


Figure 3.6: Data penetration across different municipalities (normalised by the number of wijken in a municipality)

in different quarters of the year. Figure 3.7 represents the number of data points available per hour of the day in 2018 for the municipality of Zaanstad for weekdays and weekends. For the data penetration during weekdays, it increases over the day, with the lowest point being between 07:00 to 09:00 and the highest at 11:00. For weekends, the number of data points peak between 12:00 to 02:00 and gradually reduce till 06:00. Both graphs suggest increased activity during night time but more in the case of the weekend. The figure also represents the increasing Uber activity across different quarters of 2018 for weekdays and weekends respectively. The data penetration progressively increases, pointing to increased usage or wider coverage resulting in a greater number of data points.

The number of data points progressively increases in Zandvoort during the day and decreases post-midnight. Figure 3.8 reveals a higher number of data points post-midnight in the case of weekends and shows minimal data penetration during the morning peak. The lack of data penetration in the morning peak in the Zandvoort and Zaanstad case would suggest, the activity carried out by Uber passengers are not work-related. Another indication of the municipality being visited for the beach is the data penetration across different quarters in 2018. The data penetration is higher for the second and third quarter during the warmer months of the year, which might be a result of Uber usage by beachgoers. The increased data penetration post-midnight like in the case of Zaanstad could once again point to a lack of public transport connectivity during these hours during weekends.

In the municipality of Haarlemmermeer, despite a peak in data points during midnight, the data penetration remains fairly constant through the day as shown in Figure 3.9. Schiphol airport has a single operational runway at midnight, and the frequency of flights is lesser during these hours. Despite the limited arrivals and departures, a peak in data points for Uber indicates the lack of Public Transport connectivity during these hours. However, it is important to note Schiphol airport has frequent trains operated by NS (Nederlandse Spoorwegen - Dutch railways) at night. Last-mile connectivity might be an issue despite the availability of trains. The data points also peak in the second and third quarter of the year. The peak might be a result of the increased visitor/tourist activity during the second and third quarters of the year.

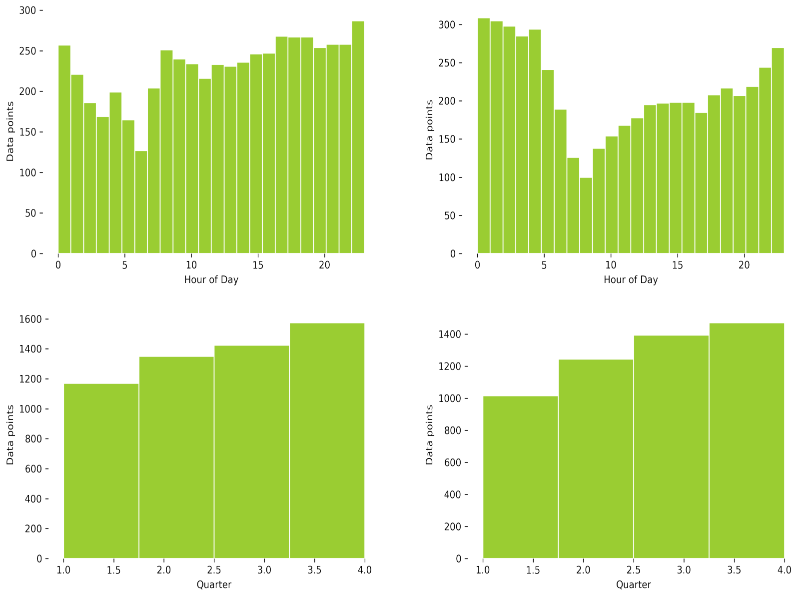


Figure 3.7: Data penetration for Zaanstad (from left to right) across (a) Weekdays (b) Weekends (c) Quarter of the year on weekdays (d) Quarter of the year on weekends

The temporal analysis reveals a variation in data penetration across different hours of the day. The variation in data penetration temporally suggests a difference in the number of Uber vehicles originating and ending at the municipalities. This could suggest the prevalence of Uber drivers travelling empty kilometres to pick-up passengers. Distance travelled without passengers while the taxi driver is either searching for a ride hailer or driving up to a pick-up point can contribute to congestion among other negative externalities such as emissions. These are referred to as empty-vehicle kilometres or deadheaded trips. Figure 3.11a represent the data points for when the 'sourceid' was in Zaanstad and 'dstid' in Zaanstad. The figure suggests a greater number of data points with 'dstid' in Zaanstad, post-midnight. This is possibly an indication of limited Public Transport connectivity to and from Zaanstad post-midnight during weekends. A higher number of 'dstid' data points can be seen post-midnight for weekdays and weekends. During the middle of the day, more data points emerge for 'sourceid' indicating a greater number of trips originating during these hours. The weekend graph suggests the movement of Ubers originating from Zaanstad to be higher during the middle of the day. Figure 3.11b represents data points with the 'sourceid' and 'dstid' in Zandvoort. All graphs indicate a disbalance in the number of trips originating and ending at Zaanstad and Zandvoort. For example, on weekdays in Zandvoort, the difference in data points between 12:00 and 01:00 for 'sourceid' and 'dstid' is 1,220 and 1,661 data points. This results in a difference of 441 data points (a 36% increase compared to sourceid points). If every data point is assumed to represent five trips, made by five different vehicles, it can result in a total of 2,205 empty trips to pick-up the passenger from Zaanstad, assuming they do not originate at Zaanstad. The magnitude of dead headed trips are higher during the evening and late at night during both weekdays and weekends. The finding indicates the potential of the dataset to determine the empty or deadheaded trips by assuming the trip distance, the magnitude of trips every data point represents. The pioneering research by King and Saldarriaga [2018] used the New York Cab Taxi Data set to estimate the dead-headed trips to the business district of Manhattan, John F. Kennedy airport, La Guardia airport

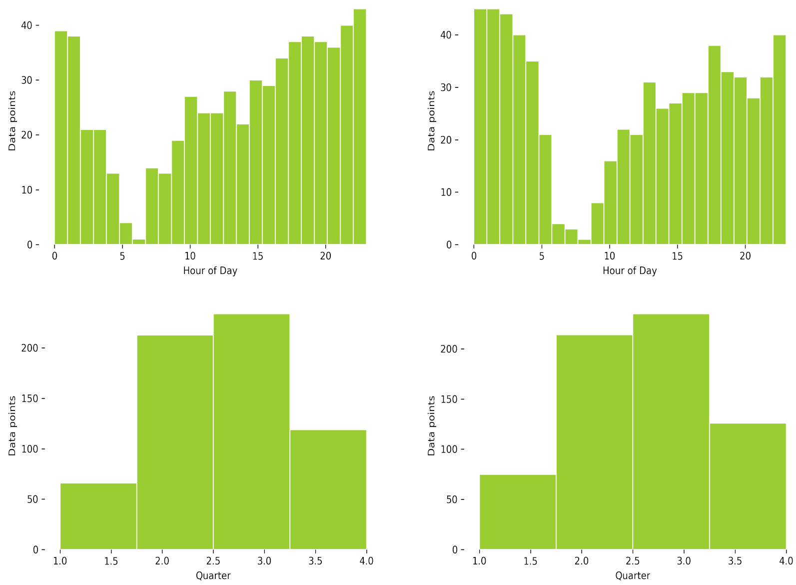


Figure 3.8: Data penetration for Zandvoort (from left to right) across (a) Weekdays (b) Weekends (c) Quarter of the year on weekdays (d) Quarter of the year on weekends)

and Newark airport. The proportion of empty vehicle kilometres was found to be 20.8%. The research was however not based on the actual route but the shortest route determined through the Open Street Routing Machine as the route taken by the cabs were not available. Figure 3.11c represents the data penetration for Haarlemmermeer with Schiphol airport.The difference in data points across 24 hours of the day is 8.15% and 11.16% for weekdays and weekends respectively. This is considerably lesser than Zaanstad and Zandvoort. The reason behind this could be the frequent pick-up and drop-off at the airport resulting in fewer dead-headed trips.

3.2.3 Conclusion – Uber Demand Studies

The current section explored the potential of the Uber Movement Data to capture the demand for Uber taxis in terms of production and attraction between wijken, its integration with land use data to reveal activity patterns, the macro mobility patterns between municipalities to understand the 'where' and 'purpose' behind the use of Ubers. The temporal differences of 'sourceid' and 'dstid' points suggest the prevalence of empty vehicle kilometres as illustrated in the previous subsection.

In terms of production and attraction, the Uber Movement data does not offer the number of trips between wijken. The data is aggregated by the temporal detail at which it is available. The maximum number of data points possible between a sourceid and dstid is 96 in a year for data aggregated by per hour of the day and per quarter of the year (24 x 4). Due to the unknown magnitude of trips and also the precise pick-up/drop-off points the Uber Movement data set cannot act as a data source for the estimation or validation of taxi trips across the city of Amsterdam, instead it offers a qualitative understanding on the wijken frequented by Uber as data points are not released till the minimum threshold of five unique trips and two unique drivers are met. Thus, the aggregation of the data set prevents one from estimating the pick-ups and drop-offs but nevertheless offers only a qualitative indication.

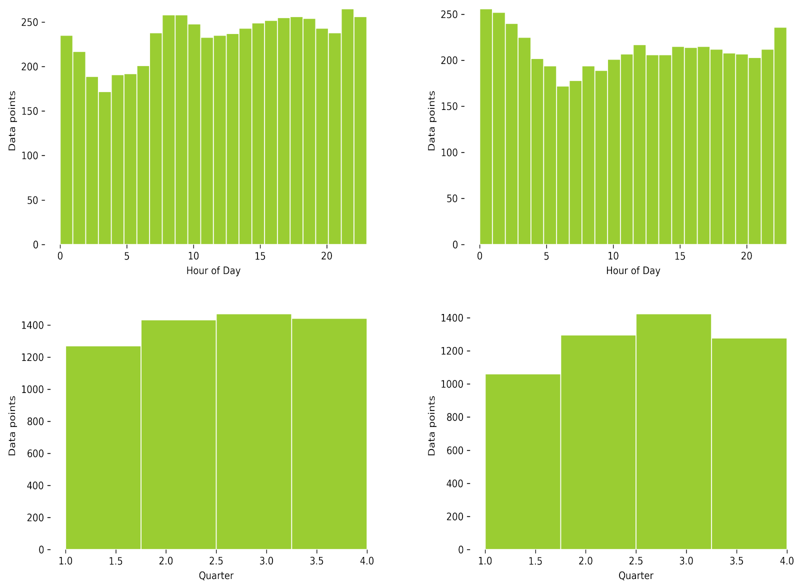


Figure 3.9: Data penetration for Haarlemmermeer (from left to right) across (a) Weekdays (b) Weekends (c) Quarter of the year on weekdays (d) Quarter of the year on weekends

On integration with land-use data which has the land-use function at the building level to evaluate the activities performed by Uber passengers, the aggregated spatial detail of the Uber Movement data renders the insights gathered to be inconclusive as to the dominant function, tends to act as the land-use function of the wijk. To derive activity patterns, one would require relatively precise pick-up and drop-off points which are absent in the Uber Movement data. Thus, trip purpose cannot be determined using the Uber movement data. However, macroscopic movement patterns across municipalities such as Haarlemmermeer, Zaanstad, Zandvoort, Velsen, Haarlem and Bloemendaal indicate the use of Uber vehicles to reach farther off destinations from the city. The municipality of Haarlemmermeer with Schiphol and nearby hotels is a place for airport users visiting the city for work and leisure. Zaanstad and Zandvoort can be considered as destinations for tourism. The quarterly data penetration (which was higher during the warmer months) for Zandvoort further reinforced the notion of using Uber for visiting the beach. Thus, based on the exploratory study, the likely users of Ubers are airport users, visitors for leisure and tourism who are either unfamiliar with the city or have to travel long distances.

There was also an imbalance in the 'sourceid' and 'dstid' data points suggestive of asymmetry in demand across the different hours of the day. The difference in data points was higher in the night for Zaanstad and Zandvoort. The higher data points can reveal approximate dead-headed trips over per hour of the day per quarter of the year. While the precise the origin of the dead headed trips are not known the distance in km from the assumed origin point to the assumed pick-up point can be determined by the Open Street Routing machine. This can be a valuable insight while managing taxi traffic in the city and regulating its movement. The main conclusions from the demand studies have been tabulated in Table 3.6

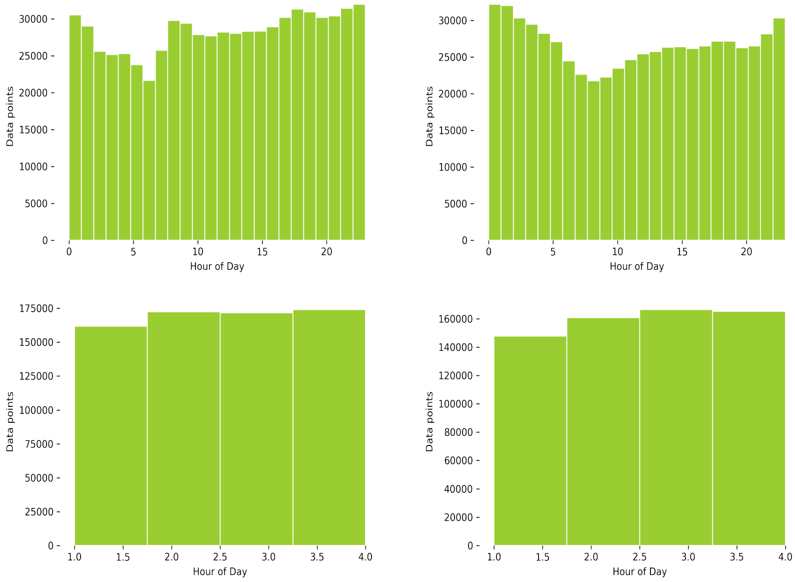


Figure 3.10: Data penetration for Amsterdam (from left to right) across (a) Weekdays (b) Weekends (c) Quarter of the year on weekdays (d) Quarter of the year on weekends

3.3 UBER MOVEMENT DATA & RECURRENT CONGESTION

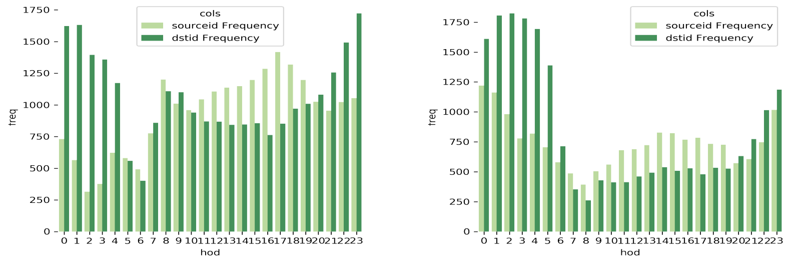
Recurrent congestion relates to congestion patterns which are frequent and occur over an extended period, as the flow of traffic exceeds the design capacity. This is especially during peak and off-peak hours. As a means for the abstraction of the Uber network in Amsterdam, network graphs were constructed to be able to derive congestion patterns. The section explains the method of construction, indicators identified and used to explain network characteristics to explore the ability of the data to capture recurrent congestion and thereby answer the following sub-questions:

SQ3: To what extent can recurrent congestion analysis be carried out using Uber Movement Travel Time data, either singularly or fused?

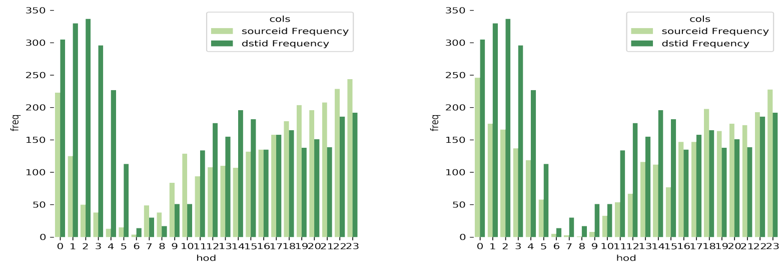
3.3.1 Construction of network graphs

The network graphs were constructed using the '.geojson' shapefile made available on the Uber Movement website. The shapefile comprises the geometry of wijken in the form of polygons and their respective attributes. The polygons in the centre of Amsterdam can be represented as shown in Figure 3.12. The centroids for each of the wijken are calculated using the geopandas library in python [Jordahl et al., 13]. The centroids are now assumed to represent a wijk and thereby as a node in the network graph. The edges are constructed based on an adjacency matrix. An adjacency matrix represents a matrix with nodes as the column names and row names. Binary coding is used to indicate a connection between two nodes. If there exists a connection, the cell value is attributed with one and zero otherwise.

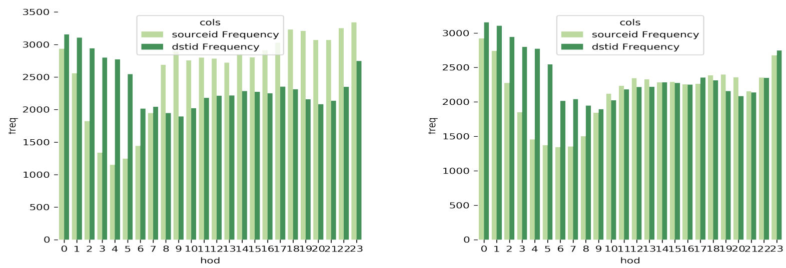
The connection is defined by a shared boundary. For instance, wijk 27 in the figure shares its boundaries with wijk 28, 24 and 22 and have direct edges. wijk 27 is connected to wijk 25 by wijk 24 or any of the other possible routes. A spatial graph and a temporal graph is constructed. The temporal graph comprises of directed edges and nodes. The graph is directed because the travel time from an origin to



(a) Data points with 'sourceid' in Zaanstad (light green) and data points with 'dstid' in Zaanstad (dark green) across different hours of the day - Weekday (left) and Weekend (right) 2018



(b) Data points with 'sourceid' in Zandvoort (light green) and data points with 'dstid' in Zandvoort (dark green) across different hours of the day - Weekday (left) and Weekend (right) 2018



(c) Data points with 'sourceid' in Haarlemmermeer (light green) and data points with 'dstid' in Haarlemmermeer (dark green) across different hours of the day - Weekday (left) and Weekend (right) 2018

Figure 3.11: Difference in 'sourceid' and 'dstid' data points suggestive of empty vehicle kilometers

a destination can be different from the return trip. Two types of spatial graphs are constructed. First is an unweighted spatial graph where all edges are weighted by a value of 1. The second graph is a weighted graph, where the edges are weighted by the Open Street Map distances. The visualisation of the unweighted and weighted spatial graph can be seen in Figure 3.13. Here, the nodes correspond to the actual geographic centroid of wijken in the data set. For constructing a weighted graph, the values of one are replaced by the respective weight of the edge in the adjacency matrix.

The spatial graphs are static in nature i.e. the weights do not change temporally and can be represented as shown in Equation 3.2 where V_s, E_s and w_s represent the nodes, edges and weights respectively. Temporal graphs can be represented as shown in Equation 3.3 where where V_t, E_t and w_t represent the nodes, edges and weights respectively of temporal graphs. Temporal graphs are subsets of spatial graphs Equation 3.4. The reason behind this is the sparsity of travel time data. Data points for origin-destination pairs could be missing during certain hours of the day. Temporal graphs have their edges weighted by travel time. As a rule, geometric mean travel time was used to weigh the edges of the temporal graph.

Table 3.6: Principle conclusions from Uber Movement Demand exploration

	Applicability of the data	Limitations
Production & Attraction	Qualitatively suggests wijken which are more frequently visited as the Uber passes through, picks-up or drops-off a passenger	Absence of route, pick-up, drop-off points and aggregation by temporal detail resulting in limited data points.
Purpose of Trip	At the municipality level (with a dominant land use function), high data penetration rates are indicative of 'tourist' and 'airport goer' user groups	At the wijk level, it is difficult to determine the purpose of the trip. due to the absence of pick-up and drop-off points, especially in the case of mixed land-use in the city centre.
Dead headed trips	Difference in 'sourceid' and 'dstid' points are indicative of temporally asymmetrical demand thereby suggestive of Ubers redistributing spatially.	The 'sourceid' and 'dstid' data points are not the same as origin and destination points and the applicability relies on the assumption, the higher 'sourceid' data points are suggestive of greater origin trips and 'dstid' data points of destination points.

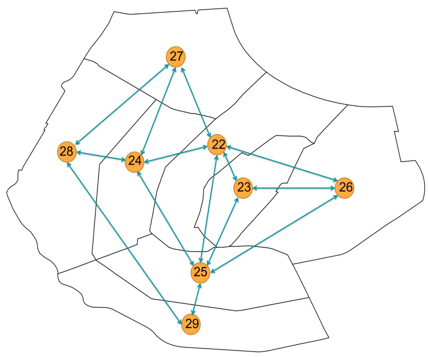


Figure 3.12: Construction of edges based on shared boundaries between wijken derived from the adjacency matrix

This is because the geometric mean travel time is a better measure for averaging travel times in an Urban network due to the heavy-tailed distribution of travel time as explained in Section 2.1. The next subsection discusses temporal graphs in detail.

$$G_s = (V_s, E_s, w_s) \tag{3.2}$$

$$G_t = (V_t, E_t, w_t) \tag{3.3}$$

$$G_t \subset G_s \tag{3.4}$$

3.3.2 Temporal Graphs

Temporal graphs as already discussed in the previous sub-section are dynamic in nature. The research employs dynamic graphs to calculate the following network indicators:

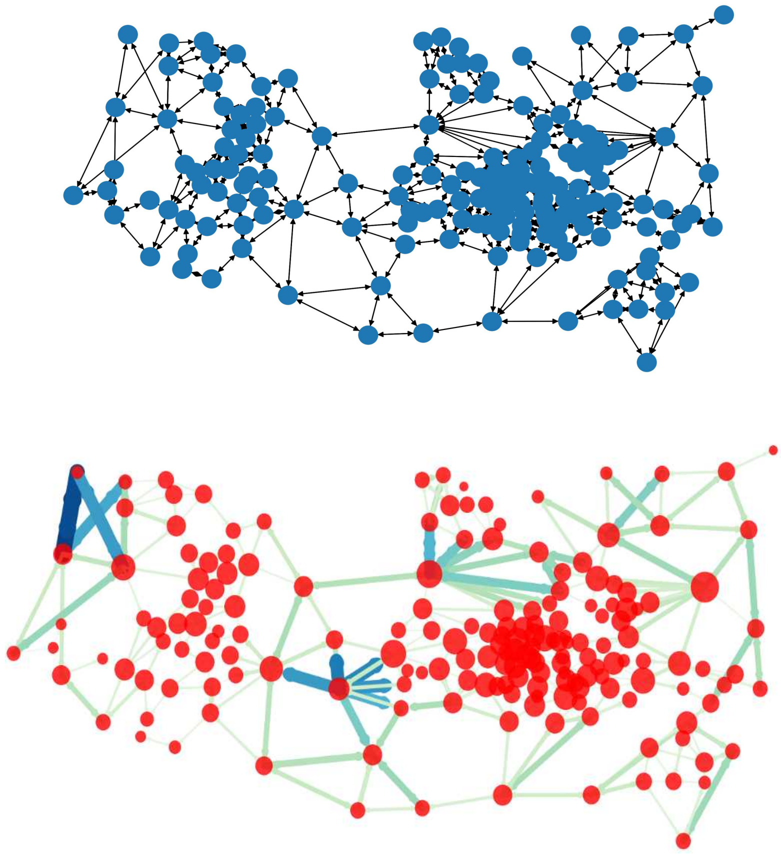


Figure 3.13: Spatial Graph for the Uber Movement Network (from left to right) (a) Unweighted (b) Weighted

- Indegree and Outdegree
- Betweenness Centrality

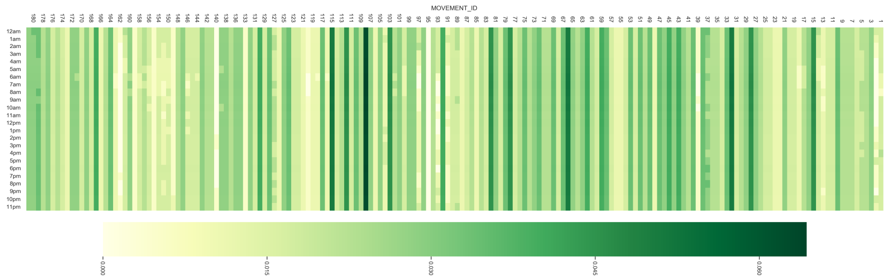
3.3.2.1 *Indegree/Outdegree*

In a directed graph, degree centrality is measured in terms of indegree and outdegree. Indegree accounts for the number of edges incident on a node and outdegree is a measure of the number of edges projecting from the node. In the graph G_t, the in-degree value is calculated as the number of incident edges representing the travel time between a 'dstid' and 'sourceid' from adjacent wijken and vice-versa for outdegree.

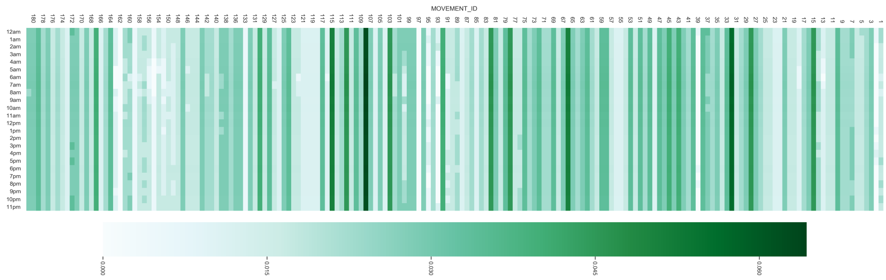
$$ind(V_i) = \frac{\sum V_n}{(ind_{max} V_j - 1)} \tag{3.5}$$

$$out(V_i) = \frac{\sum V_n}{(out_{max} V_k - 1)} \tag{3.6}$$

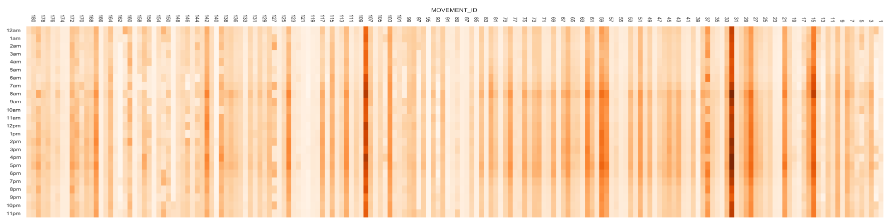
To visualise and better understand the indegree and outdegree centrality values for the network, a heatmap representing the indegree and outdegree for 2018 across different wijken for 24 hours of the day using weekday data has been plotted in Figure 3.14a and Figure 3.14b respectively. The uniformity in the vertical lines exists due to the limited number of data points possible per 'sourceid' and 'dstid' per hour of the day. The blank spots in the heatmap (wijken 148 to 161) suggest limited data penetration. The indegree and outdegree can be used to understand data penetration in the city and the greater Amsterdam region changing temporally i.e. the indegree and outdegree are better suited for demand studies. For recurrent congestion studies, the weighted indegree Figure 3.14c is better suited and utilises the sum of travel times incident on the edges. The heatmap captures the travel time variability across different hours of the day for every wijk. Temporally, longer travel times (indicated by darker lines) darker can be observed between 06:00 to 19:00. Spatially, wijken numbered 15 (Vijfhuizen), 21 (Amstelveen), 32 (Bedrijventerrein Sloterdijk), and 108 (Waterland) show longer travel times across the heatmap compared to the rest of the wijken.



(a) Indegree for 2018



(b) Outdegree for 2018



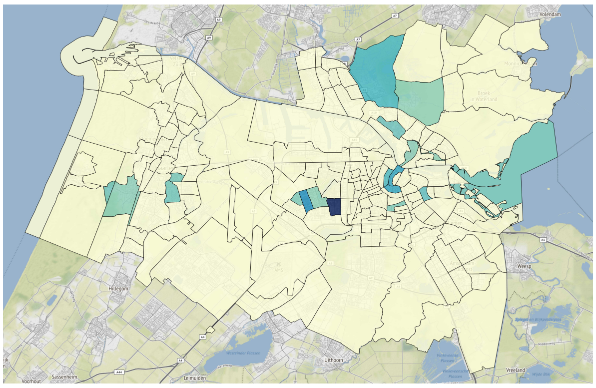
(c) Weighted Indegree for 2018

Figure 3.14: Weighted Indegree (pickups) at wijken (numbered according to Movement ID) across different hours of day for different years

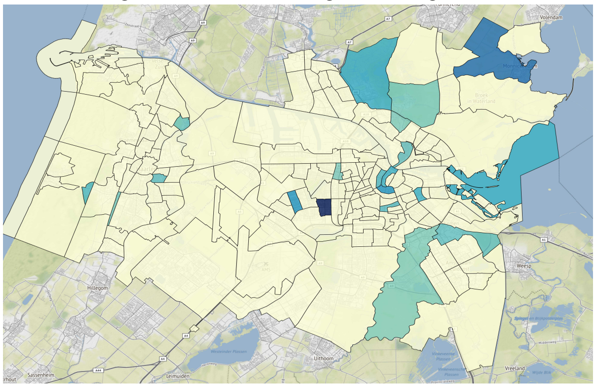
Weighted indegree is the sum of edges weighted by travel times incident on the node. It helps identify the wijken in the city which takes longer travel times to arrive at. The weighted indegree can thereby be interpreted as stress points in the city. The research normalises the weighted indegree in two steps. First, it divides the sum of edges by the number of incident edges, followed by the sum of distance. A higher number of incident edges would lead to a higher sum of travel times, and longer distances would naturally lead to longer travel times. The normalisation process prevents both the biases. Table 3.7 depicts the wijken with the ten highest normalised weighted indegree across, 24 hours of the day, the morning peak and the evening peak respectively. Figure 3.15a, Figure 3.15b & Figure 3.15c represent the wijken with the twenty most highest normalised weighted indegree across the same.

Table 3.7: Wijken with the highest normalised Weighted Indegree(2018)

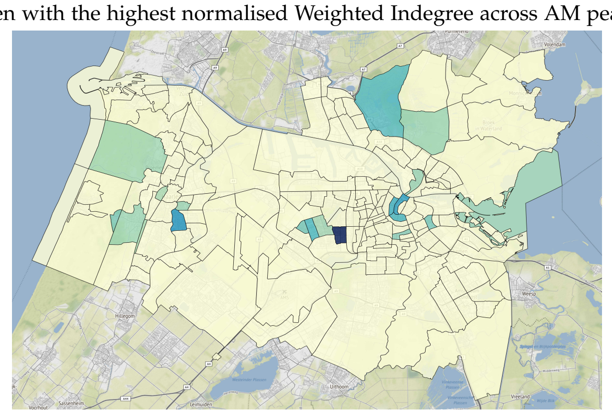
	MOVEMENT_ID	Across 24 hours	MOVEMENT_ID	AM peak	MOVEMENT_ID	PM peak
1	3	Europawijk (Haarlem)	2	Spaarndam	3	Europawijk (Haarlem)
2	19	Heemstede West van de Spoorbaan	19	Heemstede West van de Spoorbaan	19	Heemstede West van de Spoorbaan
3	23	Burgwallen Nieuwe Zijde	23	Burgwallen Nieuwe Zijde	23	Burgwallen Nieuwe Zijde
4	24	Grachtengordel-West	24	Grachtengordel-West	24	Grachtengordel-West
5	25	Grachtengordel-Zuid	25	Grachtengordel-Zuid	25	Grachtengordel-Zuid
6	37	Landsmeer	37	Landsmeer	37	Landsmeer
7	49	Nieuwe Pijp	49	Nieuwe Pijp	49	Nieuwe Pijp
8	53	Dapperbuurt	53	Dapperbuurt	53	Dapperbuurt
9	56	Oostzaan	56	Oostzaan	56	Oostzaan
10	86	Volewijck	61	Ouder Amstel	86	Ouder Amstel



(a) Wijken with the highest normalised Weighted Indegree across 24 hours (2018)



(b) Wijken with the highest normalised Weighted Indegree across AM peak (2018)



(c) wijken with the highest normalised Weighted Indegree across PM peak(2018)

Figure 3.15: Wijken with the highest normalised weighted indegree(2018)

Most wijken consistently features across the different levels of temporal detail. Europawijk, Haarlem features as the wijk with the highest normalised indegree across 24 hours of the day and the evening peak. The wijk has the important arterial road Europaweg connecting the centre of Haarlem and Haarlem Schalkwijk. Spaarndam, located in the municipality of Haarlem, is a commuter town with the A9 passing through it. Majority of the residents work in Amsterdam and Haarlem. While it is unknown if trips originate at Spaarndam, or is it simply traversed, the wijk has a normalised weighted indegree value of zero during the PM peak. Thus, the travel time values are influenced by data penetration. Thus, the wijken featured here are influenced by the usage of Uber. This can be treated as an important limitation of the data when evaluating cases of recurrent congestion through the dataset. According to the data set, the most congested wijken in Amsterdam include Burgwallen Nieuwe Zijde, Grachtengordel West, Grachtengordel-Zuid, Dapperbuurt and Volewijck. The wijken Dapperbuurt and Volewijck are the ones located outside the city centre. Dapperbuurt, located in the Eastern district and Volewijck located North of IJ river are suggestive of accessibility issues. Incidentally, Volewijck borders the Nord station of the Nord-Zuid metro line. The next subsection discusses the betweenness centrality derived from the network graph.

3.3.2.2 *Betweenness Centrality*

Betweenness centrality is the measure of the number of shortest paths passing through a node. The centrality indicator can help identify nodes where the majority of the flow passes through. A higher betweenness centrality indicates the importance of the node in terms of how much flow does the node receive. The weighted betweenness centrality is used where the edges passing through the nodes are weighted by travel time. Thereby, leading to betweenness centrality where the number of shortest edges based on travel time is calculated. The centrality measure can be expressed as shown in Equation 3.7.

$$C_B(V) = \sum \frac{d_V(i,j)}{D(i,j)} \tag{3.7}$$

Where,

d_V(j,j) = number of shortest paths between node i and node j through node V, and
D(i,j) = number of shortest paths between node i and node j
C_B(V) = Betweenness Centrality of node V

The betweenness centrality is valuable in highlighting the flow of 'traffic' which might not be the actual centres of flow but instead the centres through which the shortest paths of the network graph pass. The weighted betweenness centrality reveals the points (weighted by travel time) through which much of the flow passes across different hours of the day. wijken with the highest betweenness centrality for 2018 includes Westelijk Havengebied, Vijfhuizen, and Buiteveldert West. Westelijkhaven Gebied is the port area located North-West of Amsterdam suggesting a concentration of data points around the harbour area from the centre. For instance, Amsterdam Sloterdijk with offices is located just South of Westelijk Havengebied. Vijfhuizen is located North of Schiphol and is also adjacent to Badhoevedorp, which comprises of the P5 Parking terminal at Schiphol, next to which Uber vehicles required to wait before picking up passengers from the arrival area. wijken adjacent to Vijfhuizen also have hotels and places of work. Buitenveldert West is located South of the Zuidas NS train station and is an area dominated by office space. Zuidas as such features in area with high betweenness centrality across different hours of the day. The heatmap for wijken with the highest betweenness centrality in 2018 is visualised in Figure 3.16.

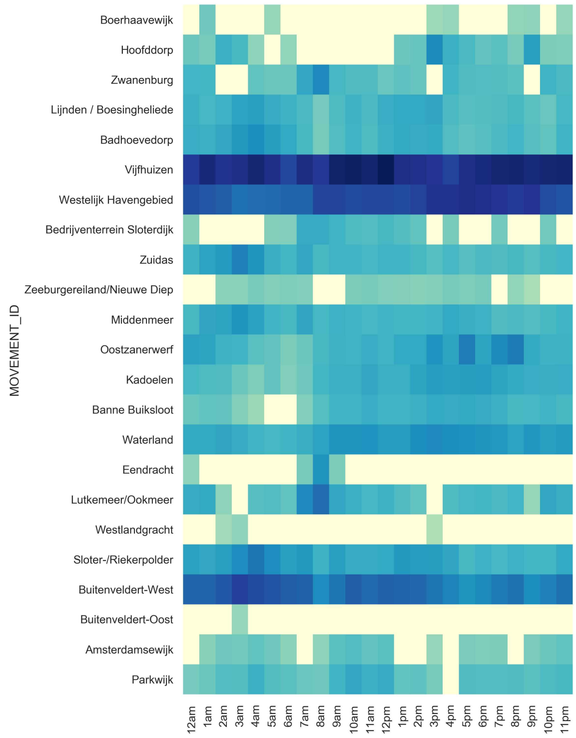


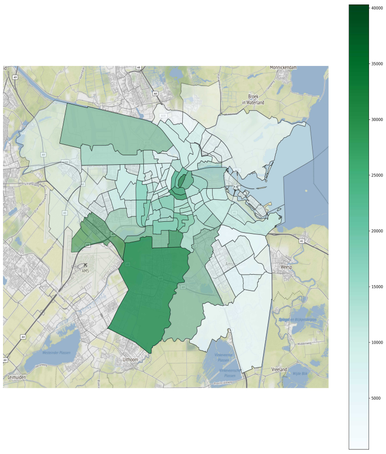
Figure 3.16: Heatmap showing wijken with the highest betweenness centrality (2018)

The other wijken which feature include wijken located in the municipality of Zaandam, Zandvoort and Bloemendaal. The findings are coherent with which was found concerning data penetration in the previous chapter. This suggests user groups, who frequently use the airport, live and work in areas close to the airport, the Harbour and South Amsterdam. The third user group being visitors to Bloemendaal, Zaandam and Zandvoort. wijken with the highest betweenness centrality have been visualised in Figure 3.17a. In order to validate the indication of Uber users being those unfamiliar with the city and foreign, the cellphone data with information on the number of international sims across different days of November 2018 were plotted and can be visualised as shown in Figure 3.17b and the spatial attributes of the two data sets have been tabulated in Table 3.8. The magnitude of international sims is the highest in the old centre of the city (tourist movement), the harbour area, the Sloterdijk area, South Amsterdam, Amstelveen (residential area for expats) and Badhoevedorp, North of Schiphol. Each of these areas (except Amstelveen) have shown high betweenness centrality. This is further indicative of the usage of Uber by visitors to the city. The sim card data was limited to the wijken in Amsterdam and thereby cannot validate the activity seen in Zaanstad, Zandvoort, Haarlem and Bloemendaal. Wijken with the highest international sims has been tabulated in Table 3.9.

Table 3.8: Spatial detail and temporal range for fusing Uber Movement Data and Cellphone Data

	Uber Movement Data	**Cellphone Data**
Spatial Detail	wijken (administrative spatial units combination of neigbourhoods)	Postal Code Level Four
Temporal range	The whole of 2018	November 2018

(a) Wijken with the highest betweenness centrality (2018)



(b) Wijken with the highest concentration of international and european sims (Nov 2018)

Figure 3.17: Wijken with the highest normalised weighted indegree(2018)

3.3.3 Conclusion - Uber Movement Data & Recurrent Congestion

The Uber Movement data can offer insights on the congestion patterns in the city i.e the Uber Movement data can capture Recurrent congestion patterns between two wijken across different hours of the day, day of the week per quarter of the year and monthly for 2016, 2017 and 2018. Network graphs offer a method for the abstraction of the Uber Movement network. The weighted indegree is an effective method to establish wijken which take longer travel times to reach i.e. areas which face frequent congestion. This can enable an accessibility analysis study, for taxis, and the variation of stress points across different quarters of the year and urban planning decisions.

However, as illustrated in the case of stress points varying across different time periods for the 2018 data, the results of the wijken with the most congestion are directly influenced by the areas frequented by Ubers. The data points for Spaarndam did not feature in the PM peak while it was the wijk with the highest weighted indegree during the AM peak. It was found data points were absent to Spaarndam during the PM peak. This can lead to faulty interpretations on which particular wijken are congested during different times of the day.

In terms of ex-ante analysis or prediction of future congestion patterns, the number of data points available per OD pair per hour is 12 data points for three years. The number of data points further reduces in the case of the day of the week and month. This leads to an insufficient number of data points for reliable predictions. The conclusions from the section are tabulated in Table 3.10

Table 3.9: International sim card penetration for November 2018

	International + EU cellphone sim cards
Burgwallen-Oude Zijde	40301
Burgwallen-Nieuwe Zijde	40301
Grachtengordel-Zuid	40301
Nieuwmarkt/Lastage	40301
Weesperbuurt/Plantage	40301
Oostelijke Eilanden/Kadijken	40301
Amstelveen	37136
Westelijk Havengebied	32914
Bedrijventerrein Sloterdijk	32914

Table 3.10: Principle conclusion for the use of Uber Movement Travel time to evaluate recurrent congestion

	Applicability of the data	Limitations
Ex-post Analysis for recurrent congestion	1. The data set can be abstracted as network graphs which enables the calculation of additional indicators. 2. Indegree weighted by travel time can represent congested wijken across different levels of temporal detail in the city. 3. Can be used to understand travel times experienced by Ubers.	Data sparsity for certain wijken can lead to false interpretations.

3.4 UBER MOVEMENT DATA & NON-RECURRENT CONGESTION

The fourth aspect of the data explored is the ability of data to capture travel time changes due to non-recurrent events or incidents. One of the performance indicators for recurrent congestion travel time [OECD, 2007]. The occurrence of non-recurrent congestion is identified as a measure which deviates from the usual travel time taken (defined by Historical observations) on a route. For instance, Dowling et al. [2004b] measured travel times on a route using ANPR (Automated Number Plate Recognition) data and considered recurrent congestion to be 1.2 times the travel times. This would imply non-recurrent congestion occurs when travel times are greater than a factor of 1.2.

Data sources for traffic incident detection can be classified as traffic surveillance data i.e. loop detectors, CCTV, camera, ANPR data; non-transportation related reports such as police reports and crowdsourced data from social networks such as twitter, and WAZE [Amini et al., 2016]. The research utilises the WAZE data set. The WAZE data is a particularly advantageous data set which offers incident data as part of its connected citizen programme where cities share planned network closures with WAZE and in return the company shares incidents reported by the users and the ones provided by the city government in shapefiles. The incidents can include accidents, road-closures, unexpected jams on the network or large scale events. Previous research has highlighted the advantage of the data as an essential source of crowdsourced data for traffic managers [Niforatos et al., 2015]. In the Netherlands, WAZE provides incident data information as part of the programme for the road network and is not exclusive to Amsterdam. This becomes especially advantageous when fusing with Uber data which comprises of travel time from the greater Amsterdam metropolitan region.

The section looks at the exploring travel time changes caused by traffic incidents reported in WAZE data i.e. to it describes to what extent does the different type of incidents reported in WAZE data correspond with travel time changes in Uber movement? Additionally, the data set explores the ability of the data to capture events of non-recurrent congestion caused by large scale events. This is explained through the Amsterdam marathon case.

The question stems from a hypothesis related to the data fusion process undertaken in the section i.e. the Uber movement data set is essentially travel time data which can be investigated and should reflect travel time changes under the event of non-recurrent congestion incidents reported in the WAZE data and thereby answering the subquestion:

SQ4: To what extent can non-recurrent congestion analysis be carried out using Uber Movement Travel Time data, either singularly or fused?

The following subsection, explains the methodology applied to derive an event of non-recurrent congestion when fused with WAZE data. Section 3.4.3 explains the methodology for large scale events separately.

3.4.1 Methodology - WAZE & Uber Movement Data

The WAZE data is understandably at a finer temporal and spatial level of detail than, the Uber Movement data. The methodology to fuse the WAZE data and check for travel time variability in the Uber data can be found in Figure 3.18. To create grounds for comparison, the WAZE data is brought to the same level of Spatio-temporal detail as the Uber Movement data. This can be achieved by, using the Spatial join tool available in the Python Geopandas package, assigning the incident reports at the wijk level and dissolving the coordinate level detail. The data provided was in the form .geojson shapefiles and was parsed and analysed using the Geopandas and Pandas library available in Python. The columns contained in the shapefile can be seen to provide the information for jam and alert. For the study, the WAZE alert data is utilised, as it offers information on the different type of events under the column label: 'type', these include: Accidents, Road closures, Jams, Weather hazards and policeman. The 'policeman' type of incident is not considered as it is irrelevant to the study undertaken.

The temporal detail of the Uber Movement data, when derived from the Web interface at movement.uber.com is provided as time periods of the day which include ; Early Morning (12:00 to 06:00), AM peak (07:00 to 09:00), midday (10:00 to 15:00), PM peak (16:00 to 19:00) and evening (20:00 to 23:00). The incident reported in the WAZE data is matched to one (or more) of these periods. For instance, if an accident was reported at 06:00 in the morning, the incident is matched to the early morning time period and a road closure incident could occur over more than one period of the day and possibly the daily average i.e. mean travel time between the OD pair for the entire day.

Post the temporal and spatial match, each type of incident: Accident, Road closures, Jams and Weather Hazards are returned using a query based on the highest reliability and check for different road types available. The attributes of the WAZE data are tabulated in Table A.1. If two reports with the same type of incident and reliability are returned, report with the higher number of Thumbs Up is selected. For

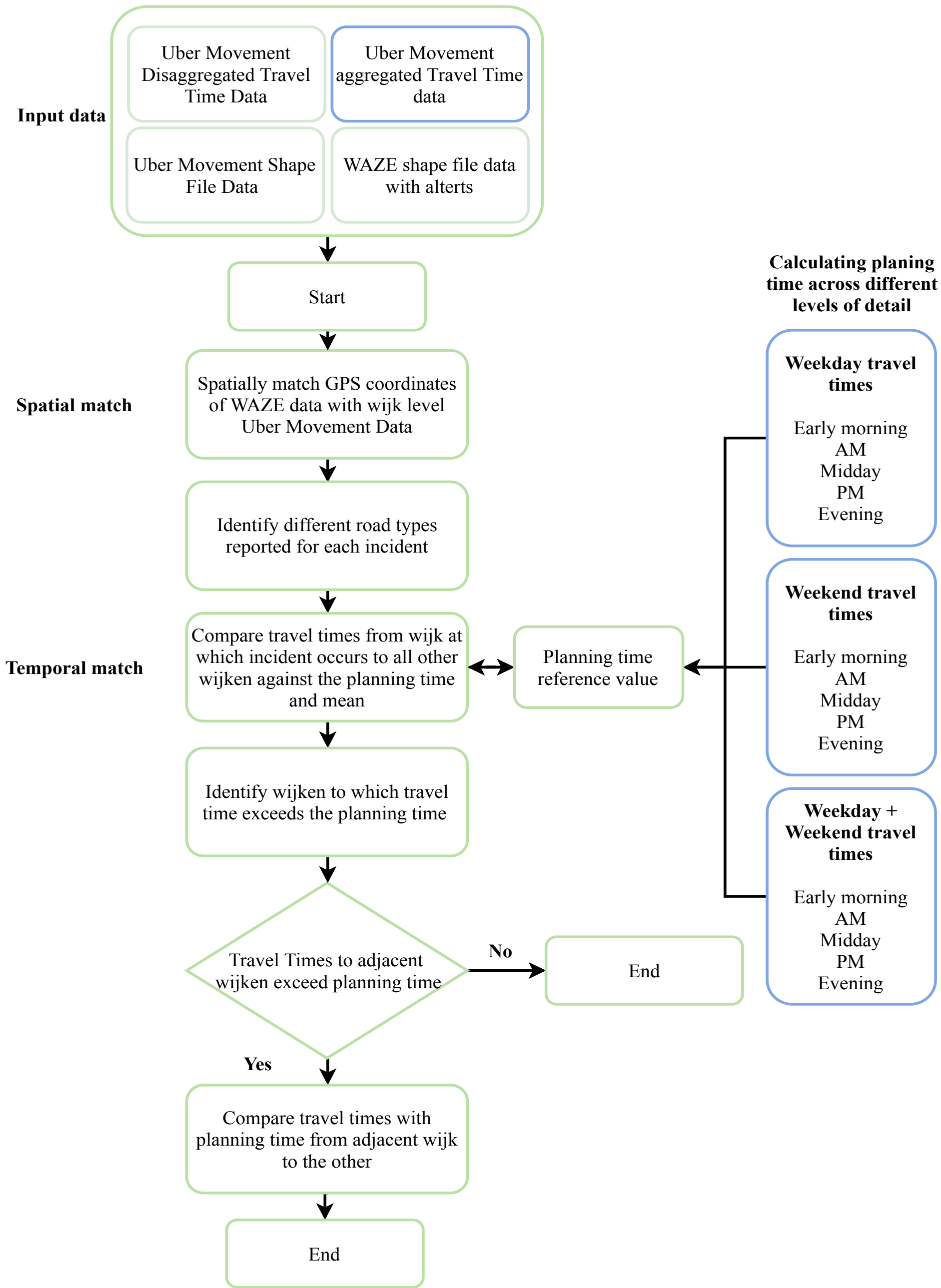


Figure 3.18: Methodology for fusing Uber Movement and WAZE data

example, if two road closure incidents with the same reliability and type of road are found, the decision is made to check for travel time variability in the incident with a higher number of thumbs up. The reliability is used as a measure for choosing incidents as it relates to the verification of the incident occurring by fellow users. Higher reliability would suggest, a greater number of users have approved of the incident's occurrence. The approach has also been employed while testing the reliability of WAZE data in the state of Iowa, United States of America [Amin-Naseri et al., 2018].

To test if the travel times were higher than what is considered free flow or average, the Planning Time is used (the 95th percentile travel time). The travel measure is recommended by the Cambridge Systematics and Texas Transportation. However, the measure is used for freeways. Travel time reliability studies in Urban networks are done through microscopic (vehicle to vehicle interaction is studied) and mesoscopic levels (packet of vehicles defined by the speed-density function) of simulation where the effect of the traffic of signals, right of way and access restrictions are considered to establish expected reliability of arriving at a destination. This can be noted as a limitation of the study. However, due to the continuous nature of the data, absence of information on the number of vehicles per time period, and the route, the measure is be deemed suitable for the study.

The planning time measure can be derived from distributions of the continuous data available from the 'Hour of day' data available for every quarter from 2016-2018. Moreover, the hour of days can be combined to form different time periods of the day as available in the web interface. The planning time measure cannot be used across trips i.e. trips will have different lengths, different routes can be taken between an OD pair and directionality also needs to be taken into account. Therefore, the planning time is derived for every possible Origin-Destination pair available in the data and for every time period of the day. An extra layer of temporal detail is the planning time can be derived for the weekday and weekends separately. Thus, the planning time of an incident during the Monday morning peak would differ from the Saturday Morning peak for an OD pair. As an illustration, the planning times for Waterland and Waterlandpleinbuurt for the early morning period is shown in Figure 3.19.

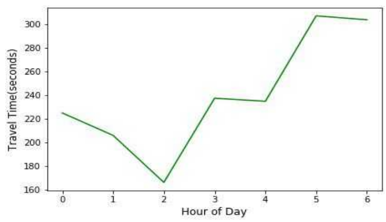


Figure 3.19: Planning time between Waterland and Waterlandpleinbuurt for the early morning period

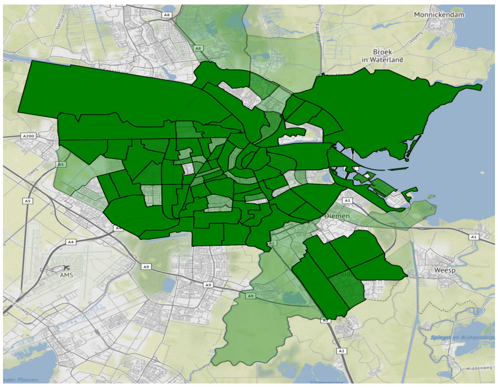
The disadvantage of using the planning time here can stem from how the data is aggregated, the travel times available for every OD pair, per hour of the day is aggregated over a quarter. The data points are instances of 'mean travel times' and not the actual travel times. This implies the data points are the only representative of average conditions. The 95th percentile travel time is derived from a total of 12 data points for every hour (4 quarters x 3 years) and depending on the time period, this can go up to 72 data points when a time period is considered. Thus, an incident may have occurred at 05:00 in the morning, but the planning time for the morning period comprises of data points (from 00:00 to 06:00) and it is likely the travel time variability may be overestimated or underestimated here. However, the strictest possible measure of the 95th percentile is considered, and the aggregated data will be from Uber taxis.

3.4.1.1 *Data Fusion*

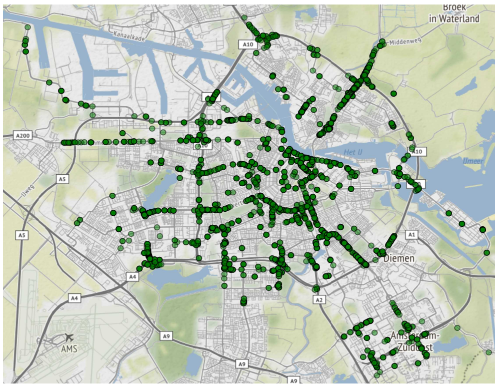
The higher level of Spatio-temporal aggregation of the Uber Movement results in a minimal level of detail, however when fused with a more detailed data source such as WAZE it is possible to reveal additional insights. Table 3.11 tabulates the difference in temporal and spatial of the two data sources. Figure 3.20 depicts the GPS coordinates of reports in the WAZE data from July to September and the spatial detail at which travel time data is available from Uber Movement. The following sections will explain the steps to fuse the data and the methodology used to study the travel time variability.

Table 3.11: Spatial detail and temporal range for fusing Uber Movement Data and WAZE data

	Uber Movement Data	WAZE Data
Spatial Detail	wijken (administrative spatial units combination of neigbourhoods)	GPS coordinate pings at the location of the incident
Temporal range	Time periods of day defined as Early Morning - 12:00 to 06:00, AM peak – 07:00 to 09:00, MD – Midday – 10:00 to 15:00, PM – 16:00 pm to 18:00, Evening – 19:00 to 23:00)	15 minute periods from July 2018



(a) Uber Movement



(b) WAZE

Figure 3.20: Spatial detail of (a) Uber Movement at the wijk level and b) WAZE at the coordinate level

3.4.2 Incidents

The subsection describes the incidents based on the methodology described in the previous subsection. For brevity, the report discusses a single incident from accidents, road closures, jams and weather hazards.

3.4.2.1 *Accidents*

On querying incidents which were accidents with the highest possible reliability (maximum value = 10), four different accident reports were returned as shown in Table A.2. It can be seen, two reports of the same road type (S108 Hobbemakade and S112 Gooiseweg) can be found, in this scenario, the report with the higher number of Thumbs Up is investigated (thereby, S112) as it is reflective of the number of WAZE users approving the incident. The available road types (with accidents as incidents) are 3, 6 and 7 which correspond to an A road (national highways), N road (provincial roads) and an S road (arterial roads) respectively. Three accidents are derived from this table; the first one belonging to the wijk: Waterland, with an accident at N247 Slochterweg during the early morning hours on the 6th of July. The second being an accident in Bijlmer Oost along the S112 – Gooiseweg during the evening peak and an accident along the A9 at Nellestein during the early morning hours.

The case of the N247 Slochterweg is discussed. Travel times were derived from the Uber Movement website after specifying the origin, as the name of the wijk from the fused data set and selecting the appropriate date and time period. The destination was arbitrarily chosen as a neighbouring wijk. This was because the interface offers a '.csv' file with travel times from the specified origin to all other destinations. The next step was to specify the origin as adjacent wijken of the specified origin wijk and downloading travel time files for each of them to understand the wider network impacts. The wijken adjacent to Waterland include; Landsmeer (wijk no. 37), Nieuwendammerdijk and Buiksloterdijk (wijk no. 92), Waterlandpleinbuurt (wijk no. 99), Buikslotermeer (wijk no. 101), Noordelijke IJ-oevers Oost (wijk no. 107), Elzenhagen (wijk no. 109), Broek in Waterland (wijk no. 157) and Watergang (wijk no. 161).

Here, Waterland was specified as the origin wijk in the web interface. The date range was set to (07/06/2018) and the travel period was specified as early morning (12:00 to 06:00). The destination was taken as a neighbouring wijk, travel times were downloaded, and this was followed by specifying each adjacent wijk as an origin and downloading their respective travel time files. To understand if the accident along N247 Slochterweg impacted travel times, the obtained travel time values will be compared against the planning time. The planning time, mean travel time and the travel time have been plotted in Figure 3.21. To ascertain, if the incident had an impact, the travel times to and from adjacent wijken were found and can be tabulated as shown in Table A.4. Table A.3 illustrates the adjacent wijken where travel time to and from Waterland exceeded the planning time. The wijken where the travel times to and from adjacent wijken did exceed the planning time can be visualised as shown in Figure 3.22.

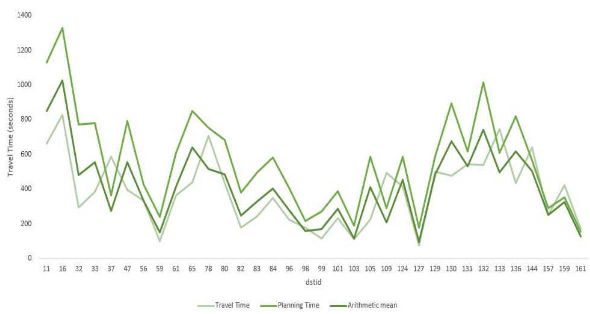
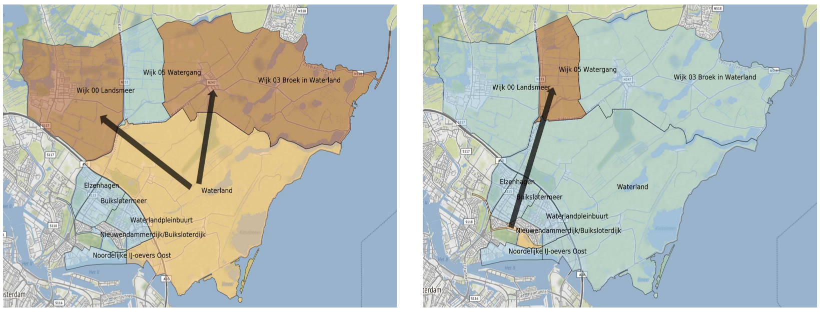


Figure 3.21: Travel times form Waterland (N247 Slochterweg) to other wijken – 6th July 2018 – Early morning period

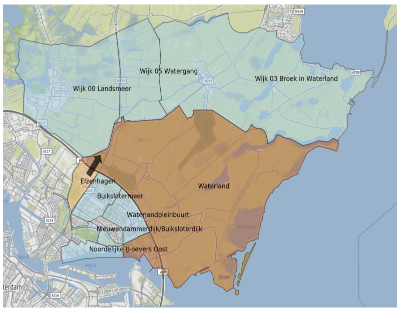
Two of the three accident incidents with the highest reliability correlated in travel time changes greater than the planning time. The accident at the A9 in the Nellestein wijk did not lead to enough data points to perform an analysis. The results of the travel time variability analysis for accidents have been tabulated in Table 3.12.

3.4.2.2 *Road Closures*

The Road closure incidents tend to have a spatial and temporal spread. Whereas, jams and accidents tend to be localised. Road closures often last a few days and can extend over more than one wijk. A query was run to reveal road closure incidents



(a) Waterland to Landsmeer and Broek in Waterland (left) & Broek in Waterland to Watergang (right)



(b) Elzenhagen to Waterland and IJ-Oever oost

Figure 3.22: wijken to and from Waterland where travel time exceeded the planning time

Table 3.12: Results for accident incidents with the highest reliability

Type of Incident	Name	Type of Road	Time Period	Exceeds planning time to neighbouring wijken
Accident	Waterland - N247 Slochterweg	6 Provincial	6th July – Early Morning	Yes
	Nellestein – A9	3 National	3rd July – Early Morning	No*
	Bijlmer Oost – S112 Gooiseweg	7 City	12th July - PM	Yes

***Insufficient data points**

with the highest reliability and the results are displayed in Table A.5. Multiple reports of the same streets ('Singel','Lijnbaansgracht' & 'Reijnier Winkeleskade') were reported with the same reliability. For conciseness, only the first and last report (the first and last date of the report) are included in the table for each street. The road closure reports made no mention of the road type, and the three reports returned are from streets in the inner city.

As mentioned earlier due to the nature of road closures, an extra step was taken to test if the road closure incidents were reported in more than one wijk. For instance, it was found, the road closure at Singel was reported across wijken; Burgwallen Nieuwe Zijdge and Haarlemebuurt. Thereby, a spatial spread was noted and travel times from both wijken are investigated. In terms of the temporal detail, it was assumed the road closure lasted from the first and till the last report of the incident. For example, the first report of the road closure at Singel was 3rd July 2018 and it was last reported on 19th July 2018. The planning time or the percentile travel times were compared against this period using the daily average for both weekdays and weekends. The report will discuss the road closure across the two wijken.

Singel lines the Singelgraacht (Singelcanal) on either side. The road closure incident reported in Singel was found to be from two wijken in the same time period. The locations reported by WAZE user and the spatially matched wijken with the road closure incident can be visualised as shown in Figure 3.23. The travel time variability has been compared against the 95th percentile daily average travel time of the investigated OD pairs.

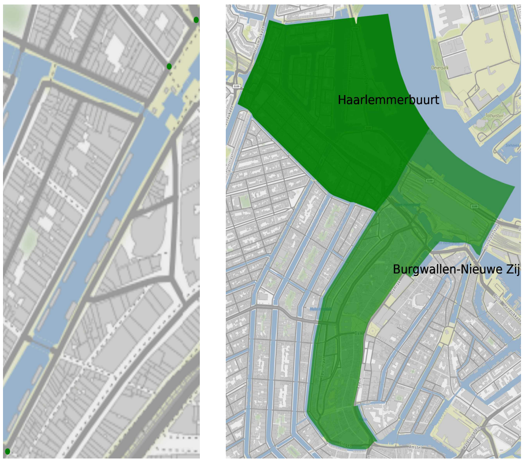
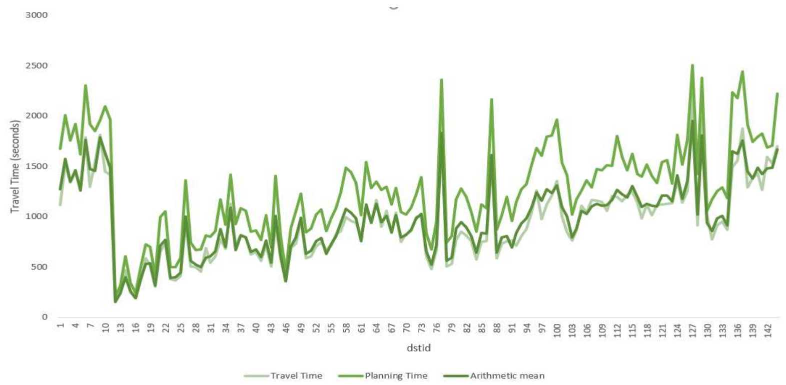


Figure 3.23: GPS pings by WAZE users for the road closure (left) & wijken in Uber Movement with the road closure (right)

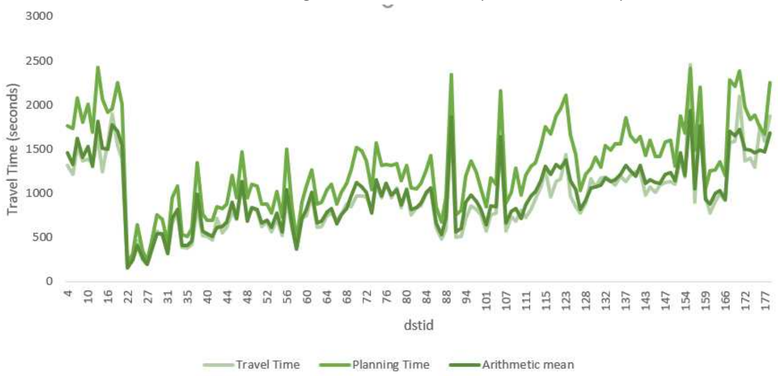
It was found the travel time did not increase to any of the neighbouring wijken. As indicated previously, it was found 3rd July 2018 was the first instance of the report in July, and the 19th of July 2018 was the last instance. An obvious limitation would be its inability to capture the temporal extent of the road closure i.e. if it started before the reporting date and if it lasted after the reporting date. However, one can be certain the road closure event lasted in the time period investigated and thereby the daily average travel times for the date range were compared against the planning time and mean as shown in Figure 3.24a.

Figure 3.24b depicts the travel times from Monday 9th July to Friday 13th July. The intention was to verify if higher travel times were observed during weekdays and the travel times were compared against the planning time of weekday traffic. Similarly, Figure 3.24c depicts the travel times during weekends of the 14th and 15th July compared against, planning time for weekend traffic conditions. In both scenarios, the travel time to adjacent wijken from Burgwallen Nieuwe Zijde did not exceed the planning time. The adjacent wijken include Burgwallen-Oude Zijde (wijk no. 22), Grachtengordel-West (wijk no. 24), Grachtengordel-Zuid (wijk no. 25), Nieuwmarkt/Lastage (wijk no. 26) & Haarlemmerbuurt (wijk no. 27).

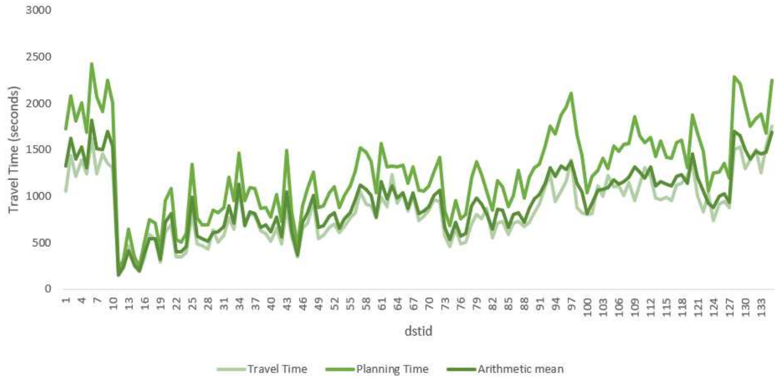
Travel times from the AM and PM peaks during the weekdays of 9th to 13th July 2018 were also explored. No instances of the travel time exceeding the planning time was found in the AM peak except at wijk 50 (Zuid Pijp). In the PM peak, travel times to Hoofddorppleinbuurt (wijk no. 70) and wijk 115 (Lutkemeer/Ookmeer) exceeded the planning time. Neither of the wijken is adjacent to Burgwallen Nieuwe Zijde. Thus, in the case of the road closure incident reported at Burgwallen Nieuwe Zijde, the travel times from Uber Movement did not exceed the planning time. While the aggregation and noise of the data could be a reason behind this, the non-increase could also be attributed to multiple alternative routes and traffic management measures in place which prevented increased travel times. Due to the spatial spread of the road closure, travel times from Haarlemmerbuurt were also investigated and the findings have been discussed in the next paragraph.



(a) Comparisons of Travel Time from Burgwallen Nieuwe Zijde to other wijken from 3rd to 19th July



(b) Comparisons of Travel Time from Burgwallen Nieuwe Zijde to other wijken on the weekdays between 9th to 13th July 2018

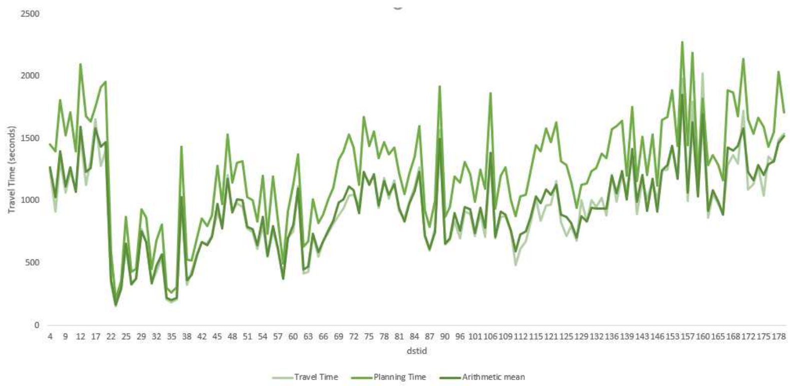


(c) Comparisons of Travel Time from Burgwallen Nieuwe Zijde to other wijken on the weekends of 14th and 15th July 2018

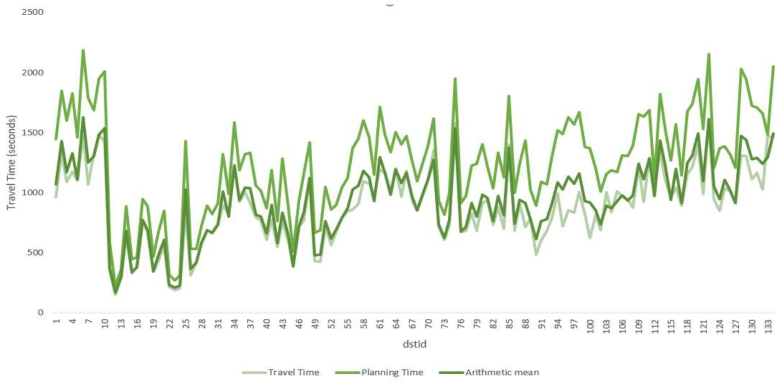
Figure 3.24: Comparisons of Travel Time from Burgwallen Nieuwe Zijde to other wijken

Haarlemmerbuurt is adjacent to Burgwallen Nieuwe Zijde. The first and last reported date of the road closure was the same (3rd and 19th July respectively). A similar set of temporal levels are investigated against the planning time as in the previous wijk. Daily average travel times were compared against the planning time from both weekday and weekend data (the date range comprises of both weekdays and weekends) and have been depicted in Figure 3.25a. Travel times during the weekdays (9th to 13th July) and weekends (14th to 15th July) have also been investigated separately and depicted in Figure 3.25b and Figure 3.25c. As was previously seen in the Burgwallen Nieuwe Zijde's case, travel times from Uber movement did not exceed the planning time when considering trips originating at Haarlemmerbuurt to all other destinations except for wijken which were not adjacent to the wijken under consideration and significantly further away from the road closure incident.

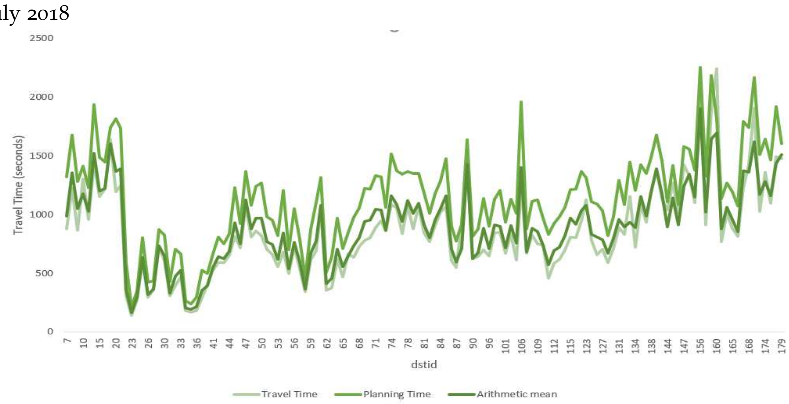
While road closures do not necessarily lead to localised travel time changes, it is difficult to ascertain if the road closure led to the travel time increase in the exceptions mentioned.



(a) Comparisons of Travel Time from Haarlemmerbuurt to other wijken from 3rd to 19th July



(b) Comparisons of Travel Time from Haarlemmerbuurt to other wijken on the weekdays between 9th to 13th July 2018



(c) Comparisons of Travel Time from Haarlemmerbuurt to other wijken on the weekends of 14th and 15th July 2018

Figure 3.25: Comparisons of Travel Time from Haarlemmerbuurt to other wijken

None of the road closure incidents correlated with travel time changes greater than the planning time. The Lijnbaansgracht road closure did result in travel times to neighbouring wijken greater than the 90th percentile. The decision to use the 95th percentile would ensure a stricter indicator as the travel time data points from quarterly data are essentially average travel times per hour of the day from different quarters of 2016 to 2018. The road closure incidents tend to have a spatial spread as in the case of Burgwallen Nieuwe Zijde and Haarlemmerbuurt, where Singel extended into both wijken. They also have a temporal spread and are not restricted to a few hours or a day. Road closure incidents can occur all through the day from few to several months, or during certain times of the day such as off-peak hours. It was not possible to derive the precise period at which the road closure incident occurred. Instead, the first and last reported date of the incident as treated as the period of road closure and daily average travel times are looked at. In addition, to this, the peak hours (AM & PM), Weekdays and weekends are additionally explored. The date and time of the incidents reported are also investigated. None of the temporal levels resulted in travel time changes greater than the planning time. This could be attributed to road closures being planned and the occurrence of these road closures at streets in inner parts of the city leaving sufficient alternatives routes. The results from the road closure incidents have been tabulated in Table 3.13. The next subsection discusses travel time variability caused by jams reported in the WAZE data.

Table 3.13: Results for road closure incidents with the highest reliability

Type of Incident	Name	Type of road	Time Period	Exceeds planning time to neighbouring wijken
Road Closures	Burgwallen Nieuwe Zijde /Haarlemmerbuurt - Singel	Street	3rd to 19th July – Weekday and weekend Daily averages/Weekday/Weekend/ AM/PM/Specific days at which incident was reported	No
	Jordaan - Lijnbaansgracht	Street	17th July to 26th July – Weekday and weekend Daily averages/Weekday/Weekend/ AM/PM/Specific days at which incident was reported	No
	Museum Kwartier - Reijnier Vinkeleskade	Street	3rd July to 26th July Weekday and weekend Daily averages/Weekday/Weekend/ AM/PM/Specific days at which incident was reported	No

3.4.2.3 *Jams*

The WAZE data contains incidents of jams reported by WAZE users. On running the query of returning Jam incidents with the highest reliability, only five incidents were returned and are displayed in Table A.6. Three different road types were found in the data; the road type N (Provincial roads), S (City arterial roads) and A (National Highways). Jam incidents with the highest reliability were found, followed by the ones with the highest thumbs up for each road category. For the S road – two incidents are returned, due to a higher number of Thumbs Up; the Centrale Markt (S105 Jan van Galenstraat) is selected. The A9 case is discussed in the report.

The A9 national highway is the Southern by-pass of the city of Amsterdam and runs till Alkmaar, and a jam related incident was reported on the 13th of July 2018 during the midday period. The incident was spatially matched to the wijk Holendrecht in the Uber movement data set. The travel time from Holendrecht to all other wijken have been plotted in Figure 3.26. It can be visually noted, at wijk 145 & 147, the travel times higher than the planning the time and has been tabulated in Table A.7.

Travel times from adjacent wijken have also been investigated and tabulated in Table Table A.8. The wijken with higher travel time than the planning time has also been visualised in Figure 3.27. It can be seen travel time from Holendrecht to Nellestein is higher, and the wijk Nellestein also has the A9 passing through it. However, on investigating travel times for adjacent wijken it can be noted, travel times tend to be higher Southward as well. This can be seen in Figure 3.27b (Bijlmer Centrum to Bullewijk/Bijlmer Oost to Bijlmer Centrum, Bullewijk & Nellestein) and Figure 3.27c (Nellestein to Bullewijk). It is difficult to ascertain the jammed direc-

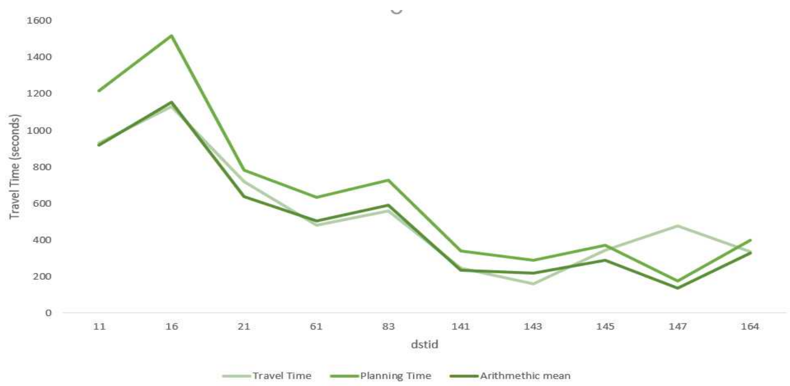


Figure 3.26: Travel time from Holendrecht to all other wijken compared against mean travel time and planning time

tion as travel times tend to be higher Northwards and Southwards. Travel time to or from Abcoude, (while an adjacent wijk) did not exceed the planning time, the wijk does not connect to the A9, unlike other wijken. Thus, it is likely the jam at the A9 has resulted in higher travel times.

All jam incidents correlated with higher travel time changes, interestingly the Jam incident at N247-Slochtetweg was caused by an accident, also reported in the WAZE data and offered the only incident at a national highway which correlated with travel time changes. The Jam incidents have subtypes such as JAM_HEAVY_TRAFFIC and JAM_MODERATE_TRAFFIC. However, it is difficult to ascertain the severity of the Jam as they tend to be a subjective response of the user. The results for the jam incidents have been tabulated in Table 3.14.

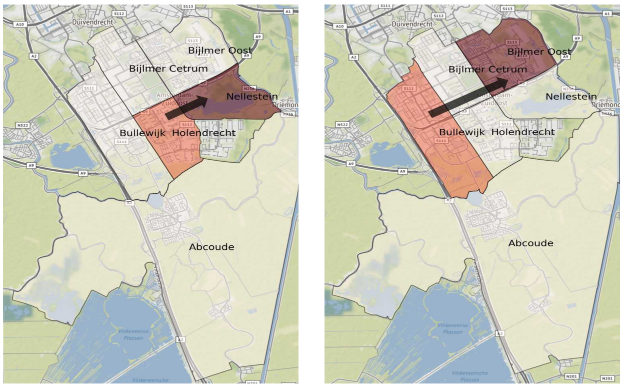
Table 3.14: Results for jam incidents with the highest reliability

Type of Incident	Name	Type of road	Time Period	Exceeds planning time to neighbouring wijken
Jams	Waterland – N247 Slochterweg	6 Provincial	6th July Early Morning	Yes
	Holendrecht – A9	3 National	13th July Midday	Yes
	Centrale Markt – S105 Jan van Galenstraat	7 City	6th July Midday	Yes

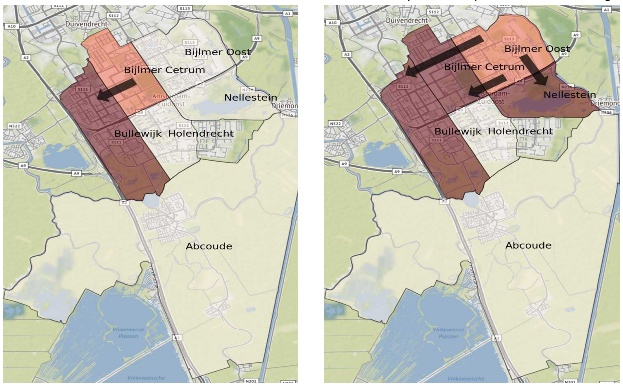
3.4.2.4 *Weather Hazards*

The fourth type of incident offered in the WAZE data is the weather hazard. The weather hazard was the most frequently reported type of incident from the July to September data set and has subtypes including Hazard due to road construction, hazard on road due to pothole etc. The incident does not describe the level of precipitation or wind speed etc., instead, it mentions a hazardous situation caused by the weather. As was done in the previous three types, the intention is to identify reports with the highest reliability. Here, the weather hazards for the same location have been reported across multiple days. For instance, in the case of Anderlechtlaan, diverse road works were planned from March 2018 to late 2019 [Bouw en verkeersprojecten, 2018]. Weather-related incidents such as heavy rain would have led to potentially hazardous situations, requiring vehicles to travel at lower speeds and an incident would have been reported whenever it rained.

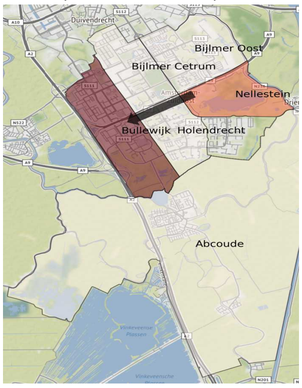
Three road types were obtained, after running a search query with the type of incident as 'WEATHER_HAZARD' and sorting the data by reliability for the month



(a) Holendrecht to Nellestein (left) & Bullewijk to Bijlmer Oost (right)



(b) Bijlmer Centrum to Bullewijk (left)& Bijlmer Oost to Bijlmer Centrum, Bullewijk & Nellestein(right)



(c) Nellestein to Bullewijk

Figure 3.27: Wijken to and from Holendrecht where the travel time exceeded the planning time

of July. The ones selected have been tabulated in Table A.9. Unlike the previous cases, it was possible to derive travel time data from Uber Movement for a national highway (A10). Anderlechtlaan is an arterial road connecting the on and off-ramps of the A4, North of Schiphol airport.

The third weather incident investigated, is at S112 Prins Bernhardplein located adjacent to Amsterdam Amstel station Figure 3.29 The S112 has been spatially matched to Frankendael, and the reports primarily posted at 18:00 and were selected as the temporal period for the investigation. Travel times for the 7th of July evening from Frankendael to all other wijken compared against the mean and planning time is as shown in Figure 3.28.

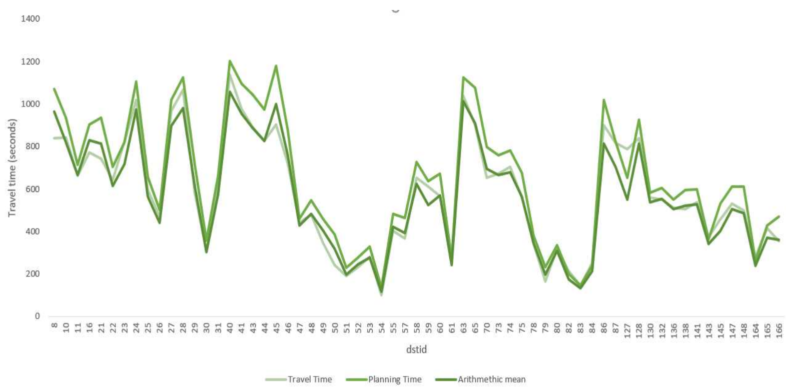


Figure 3.28: Travel time from Frankendael to all other wijken compared against mean travel time and planning time

The travel times exceed the planning time to the wijken for the adjacent wijken as shown in Table A.10. wijken 82 (Middenmeer), 83 (Betondorp) and 84(Omval/Overamstel) are neighbouring wijken and does reflect longer travel times. Noticeably wijken 143 (Bijlmer Centrum), and wijk 164(Diemen Zuid) are both wijken accessible via the S112. wijk 23 (Burgwallen Nieuwe-Zijde-comprises of metro station Rokin) is located within the 17th-century ring and is not directly accessible by the S112.

Travel times from the adjacent wijken was also looked at and has been tabulated in Table A.11. All adjacents wijken had higher travel times except wijk 51 and 54. However, it's important to take note of the directionality. Travel times to wijken 80, 81 and 84 from wijk 54 (Transvalbuurt) were also found to be higher and each of this wijken is accessible by the S112, suggesting the accident did have an impact on the travel times. Similarly travel times from wijk 51 to wijken adjacent (80 and 84) to Frankendael were found to be higher.

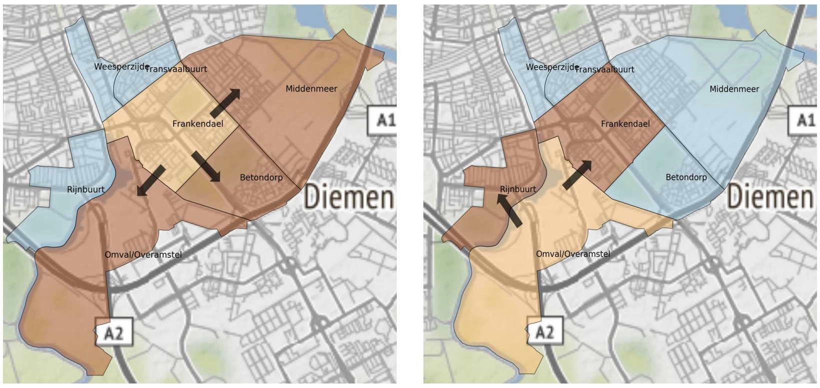
If one looks at the subtypes for weather hazards, they refer to roadworks or hazard caused by stopped vehicles. Thus, in principle, these could not be related to the weather at all. To ascertain if the weather did play a part, one would need to match the data with weather data. The weather hazards are the most common type of incident reports from the July, August and September data. The weather hazards are the most common type of incident reports from the July, August and September data. However, two of the three weather incidents did result in travel times greater than the planning time. These have been tabulated in Table 3.15.

Table 3.15: Results for weather hazard incidents with the highest reliability

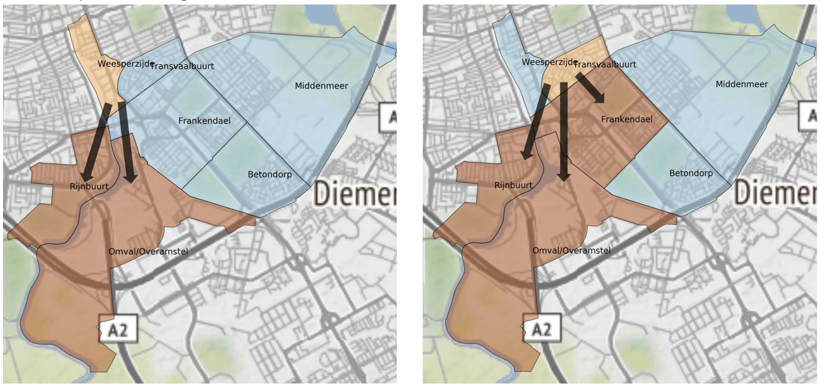
Type of Incident	Name	Type of road	Time Period	Exceeds planning time to neighbouring wijken
Weather Hazard	Sloter and Riekerpolder Anderlechtlaan	2	4th July Midday	Yes
	Middenmeer A10	3 National	9th July AM & Early Morning	Yes
	Frankendael S112 Prins Bernhardplein	7 City	7th July Evening period	Yes

3.4.3 Events

Events can widely vary in how they impact congestion. Their similarities lie in the ability to cause non-recurrent stress in terms of capacity reduction, demand surge and reduced safety [Amini et al., 2016]. Unlike the incidents discussed in Section 3.4.2, large scale events are predictable in terms of their spatio-temporal impacts i.e. where and when the event is occurring and relatively predictable in



(a) Frankendael to Omval-Overamstel, Betondorp and Middenmeer (left) & Omval-Overamstel to Frankendael and Rijnbuurt (right)



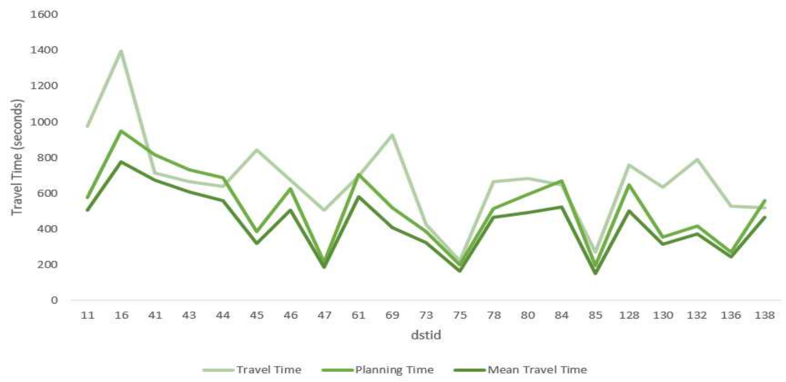
(b) Weesperzijde to Rijnbuurt and Omval-Overamstel (left)& Transvaalbuurt to Frankendael, Omval-Overamstel & Rijnbuurt (right)

Figure 3.29: Wijken to and from Frankendael where the travel time exceeded the planning time

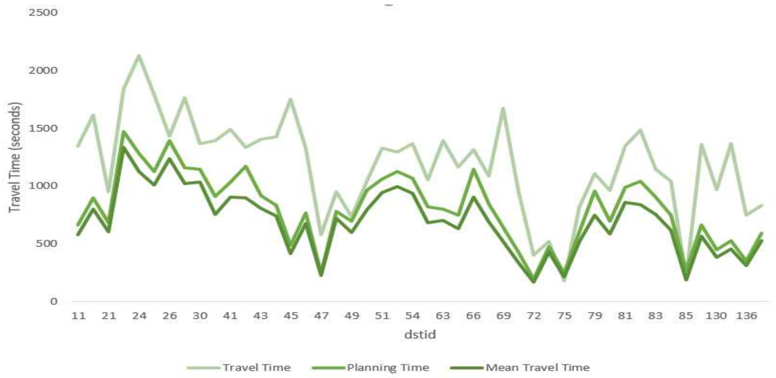
terms of the demand fluctuation. This creates the potential for implementing traffic management measures in advance. The Uber Movement data can identify the effectiveness of traffic management measures with travel time as a performance indicator. The limitation of estimating the precise route taken by the visitors to the event and those impacted by the event remains.

The current subsection identifies the usability of the Uber Movement data to evaluate travel time changes during large scale events. The subsection will specifically discuss the case of the 2017 Amsterdam Marathon. The case is especially interesting as specific routes in the city are closed to create routes for runners. In 2017, the Amsterdam marathon was held on the 15th of October. The travel times data are downloaded from movement.uber.com with Stadionbuurt as an origin wijk. The marathon was held between 09:30 to 17:00, the travel times for the time periods; Morning peak (07:00 - 09:00) i.e. before the start of the marathon, midday (09:00 to 16:00) i.e. during the marathon and pm peak (16:00 to 18:00) i.e. during the closure of the marathon are downloaded. The travel times and its comparison against the mean travel time for the period and the planning time is visualised as shown in Figure 3.30a, Figure 3.30b & Figure 3.30c. The travel times for the AM peak suggest the wijken exceed the planning time from Stadionbuurt (the wijk at which the marathon starts) to wijken Zuidas (wijk no. 47), Apollobuurt (wijk no. 75), Prinses Irenebuurt (wijk no. 85) & Buitenveldert-West (wijk no. 136). Travel times for the midday period show travel times exceed the planning time to wijken Zuidas (wijk no. 47), Hoofddorppleinbuurt (wijk no. 70), Willemspark (wijk no. 72), & Buitenveldert-West (wijk no. 136). In the PM peak, while there are instances of wijken exceed the planning time, none of the values exceeds the mean +standard deviation values.

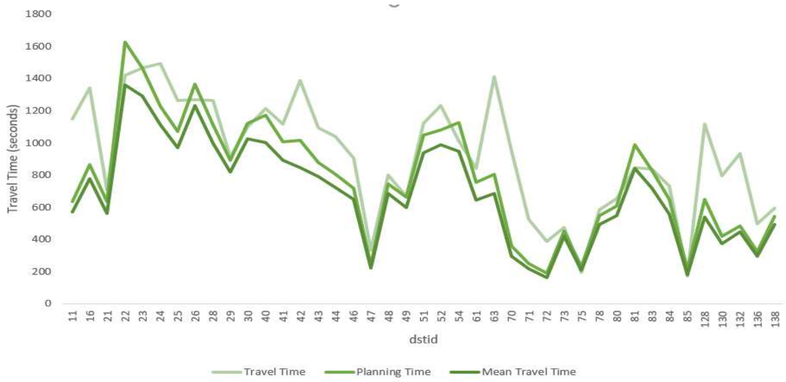
The travel times, planning time, mean and the differences illustrated in the figures have been tabulated in Table A.12,Table A.13 & Table A.14 for AM peak, midday and PM peak respectively.



(a) Comparisons of Travel Time from Stadionbuurt to other wijken on 15/10/2017 during the AM peak



(b) Comparisons of Travel Time from Stadionbuurt to other wijken on 15/10/2017 during midday



(c) Comparisons of Travel Time from Stadionbuurt to other wijken on 15/10/2017 during the PM peak

Figure 3.30: Comparisons of Travel Time from Stadionbuurt to other wijken against mean travel time and planning time

The travel times to adjacent wijken during the midday period and morning peak is a clear indication of the travel times being impacted by the Amsterdam marathon. Additionally, the travel times return to normalcy in the PM peak after the marathon is over, is suggestive of the opening of roads closed during the marathon. On an average travel time to adjacent wijken increased by 50.11 % & 54 for the AM peak, and midday respectively.

3.4.4 Conclusion - Uber Movement Data & Non-Recurrent Congestion

The WAZE data (specifically WAZE alert data) has been investigated concerning the type of incidents reported and post-spatiotemporal matching, travel times were investigated from the Uber Movement data set. Despite the aggregated nature of the Uber Movement data set, an incident reported in the WAZE data triggered travel time changes according to Uber Movement in the cases of accidents, jams and weather hazards. The data was unable to reflect travel time changes due to road closures. This might be an indication of a traffic management measure during the road closure event. The Amsterdam marathon case study also revealed the ability of the data to capture travel time changes during large scale events.

Thus, the Uber Movement data can be used to perform an ex-post analysis of the travel time changes caused by non-recurrent congestion in the city. A methodology has been developed to perform the same. The WAZE data can be replaced by any type of incident data. The methodology can be used to evaluate travel time changes caused by incidents at the streets and inner Urban roads and not exclusively for the S (City arterial roads), N (Provincial roads) and A (National motorways) roads in Amsterdam. Thus, revealing insights which were not previously available.

A major disadvantage of using the Uber Movement travel time data for an ex-post analysis of the travel time changes is the limited data penetration of the data when compared to existing data-sets used by policymakers at the municipality of Amsterdam. For instance, the NDW travel time data includes (but not limited to) Automatic Number Plate Recognition, Sensors and Floating car data whereas, the Uber Movement Data is derived from a smaller sample of approximately 5000 uber vehicles in Amsterdam which cover limited routes during different times of the day.

While capturing travel time changes is an important application of the data set, it does not necessarily perform better than existing data sets at the disposal of the municipality. The unknown route, between two wijken (to ascertain if the incident and route can be spatially matched) and the absence of the number of vehicles (to estimate the reliability of travel times), are both limitations. Additionally, the data is aggregated according to time periods compared to the 15 minute time periods of the WAZE data, resulting in further loss of the detail. The strength of the data while evaluating non-recurrent congestion lies in the simplicity of using the data set, which can be downloaded also using a web scraper and additionally offers insights in terms of Origin-Destination pairs and not route segments as in the case of the NDW data. The principal conclusions from the section have been tabulated in Table 3.16.

Table 3.16: Principle conclusion for the use of Uber Movement Travel time to evaluate non-recurrent congestion

	Applicability of the data	Limitations
Ex-post Analysis for non-recurrent congestion	1. Can be spatio-temporally fused with incident data to provide travel time variability for Origin-Destination pairs 2. Offers travel time variability for Jams, accidents & weather hazards.	1. Limited data penetration as data is only representative for Uber vehicles. 2. Temporal and Spatial aggregation necessitated by the Uber Movement data set leads to loss of detail.

3.5 CONCLUSION – DATA EXPLORATION

The chapter concludes the data exploration phase of the research. The objective of the chapter was to find the usability and unique value of the data through demand studies, travel time variability due to recurrent congestion and travel time variability due to non-recurrent congestion. The conclusions will be used to answer:

Sub Question 5: What is the unique value addition of the Uber Movement Travel Time data set to Transport Planners and officials at the city of Amsterdam?

The unique value of the data set determined from the demand studies include; the usage of Ubers by airport goers, tourists visiting places further away from the city such as Zaandam, Zandvoort & Bloemendaal. Additionally, it was revealed, there exists a disbalance in the 'sourceid' and 'dstid' data points for these locations. These are suggestive of empty trips being made to pick-up clients. The externalities of these dead-headed trips include congestion and related emissions.

Exploratory studies with recurrent congestion have revealed, network graphs are an effective method for the abstraction of the network and revealing congested hotspots across different levels of temporal detail quarterly. This can aid in accessibility analysis and Urban planning decisions. An important limitation relates to the data sparsity as not all wijken are visited by Ubers. Thus, the results of the congestion analysis may simply be a symptom of the skewed Spatio-temporal distribution of Uber vehicles. Thereby, the unique value revealed is the difference in the wijken frequented by Uber which may differ from other traditional data sets which offer a more comprehensive overview of the congestion patterns in the city (For instance, the NDW data).

Concerning non-recurrent congestion, it was found, the Uber data reflected travel time changes caused by large scale events (the Amsterdam marathon), accidents, jams and weather hazards. The data sparsity and the absence of the route taken, once again limit the applicability of the data. There is also the issue of spatial and temporal aggregation necessitated by the Uber Movement data set. Thus, the unique value of the Uber Movement lies in its skewed user base dictating a skewed Spatio-temporal distribution of taxi cabs. Additionally, the next step of the research will focus on developing a model which utilises this aspect and create a tool to predict the Spatio-temporal distribution of taxis across different wijken. Based on the conclusions from data exploration and confining to the objectives of deriving the latent potential of data such that it can meet the goals of the municipality in managing taxi traffic, the next chapter focuses on the development of a model which focuses on the Spatio-temporal distribution of taxis. The model will use the insights gained from demand studies for building the model.

4 DEFINING THE MODEL FOR SPATIO-TEMPORAL DISTRIBUTION OF UBERS

The chapter defines and estimates a model based on the unique value identified in the previous chapter. The results from data exploration suggest the Spatio-temporal distribution of Ubers revealed by data penetration, the temporal asymmetry in demand can offer additional insights which are crucial to meeting the municipality's goals concerning taxis. The intention is to employ travel time data to estimate the demand between wijken. The intended result can be best described as Origin-Destination matrices for throughout the day depicting the production and attraction of Ubers. Most methods rely on traffic counts, travel diaries with an established pick-up and drop-off point. This is unavailable in the Uber Movement data. [Krishnakumari et al., 2019] utilises flow and speeds to estimate production and attraction. It was shown that given the outgoing flows of zone i during time period t and incoming flows of zone j during period t, realised travel times and link counts, the production and attraction can be determined by placing a set of constraints on the possible routes and employing a logit model with realised travel times as an explanatory variable for determining the proportion of flow on each route and is provided with a path size factor penalising overlapping of routes. Ultimately the method of estimating the Spatio-temporal distribution is dependent on the available input data and thereby the assumptions required to build the model. The data available for the current problem is mean travel time data between two centroids of a wijk and standard deviation.

To unlock the intrinsic value of the data, the research utilises a model developed by Aryandoust et al. [2019] which utilises Uber Movement data for estimating the traffic activity per zone in Melbourne, Australia. The research will from here on citing the model as the reference model. The chapter will explain the conceptual model and the modelling steps involved. The model is extended in the current research by introducing the ability to introduce occupancy followed by the route taken by Ubers. The reference model was built with the intention of not only simulating Ubers but all traffic activity in the city. The intention in the current research is to develop a model exclusively for simulating the movement of Uber vehicles.

The model is based on Markov chains where the transition from one state to the other depends only on the previous state. This is called the memoryless nature or the Markov property. The conditions governing transitions are conditional probabilities which are stochastically sampled. Markov chains has been extensively used in transport-related and especially traffic prediction related applications (Faizrahnemoon et al. [2015] & Guoqiang Yu et al. [2003]). The following sections discuss the modelling steps.

4.1 INTRODUCING THE MODEL

The intention to model Spatio-temporal distribution of taxis originates from a need to understand the taxi activity in cities which in turn creates the scope for managing them and implementing solutions at a strategic level. The current section explains the various modelling steps and mentions instances when the modelling process has differed from that followed by the reference model which utilised the Julia programming language. The model in the current research is developed using the Python programming language. The rationale behind the model relates to increased flow between two wijken if the travel times between them are higher relative to the other wijken. The data is revealed in nature establishes areas which have been visited and differ from a route choice problem where one would minimise travel costs. Higher travel times are assumed to be indicative of greater congestion (and thereby higher density) when normalised by distance. In addition to this, an assumption is made concerning the fleet size 'C' as this is absent in the Uber Movement data. The 'C' number of Uber vehicles need to now drive and travel between two wijken. In order to achieve this p_drive (probability an Uber vehicle drives) is established as a binomial distribution for each wijk at a time period T and p_dest (probability a wijk is chosen as a destination) is in the form of a multinomial distribution as the interaction now represents the popularity of choosing one wijk over all the others. The probability that an Uber vehicle is occupied (p_occ) is also estimated in the form of a binomial distribution for each zone but is dependent on the data penetration. As a next step, each of these probabilities can be stochastically sampled at every time step T. This ensures the model is not deterministic in nature. It is stochastically sampled by checking if the probability (p_drive or p_dest) is greater than a randomly generated number. If the Uber vehicle now drives (or decides to park or travel between a wijk), and chooses a destination wijk, a likely route in the form of wijken traversed can be found by using a spatial graph. Thus, the model holds the ability to be modified and extended. For instance, in addition to occupancy, the probability a vehicle is charged can be modelled to identify wijken for charging infrastructure. The modelling framework can be visualised as shown in Figure 4.1. The steps have been described in the following sub-sections.

4.2 PREPARING THE DATA SETS

Preparing the data sets to be utilised in the model is done over two steps. The first, preparing the data matrices comprised of travel times and standard deviation. The second, being the distance matrix with distances between a 'sourceid' and 'dstid'. The steps can be explained as follows:

4.2.1 Preparing the data matrices

2018, the hour of the day, aggregated over a quarter of the year is utilised. The rationale behind choosing the 2018 data set lies in it being the most recent and complete data set available at the time of the research. The 'hour of the day' data is the finest temporal detail available in the data. The columns: 'sourceid', 'dstid', 'mean_travel_time', 'standard_deviation_travel_time' are retained. This is followed by converting the data in the form of matrices of the dimensions N x N x T x 2 where N is the number of zones (181 in the case of Amsterdam), T refers to the time period (hour of day in this case), and 2 refers to separate matrices of one comprising of mean travel time and the other standard deviations. The data sparsity is checked for using the formula in Equation 4.1:

$$Data_{sparsity} = 100 \times \left(1 - \frac{length(\forall t_{ij} \epsilon T_{ij})}{T_{ij}}\right) \tag{4.1}$$

Here, t_{ij} is the travel time data point between 'sourceid' i and 'dstid' j for the time period and T_{ij} is the maximum possible data points between sourceid 'i' and dstid 'j' for the time period. The maximum possible data points are determined based on the time period. For instance, for the hour of the day, one-quarter of the year data,

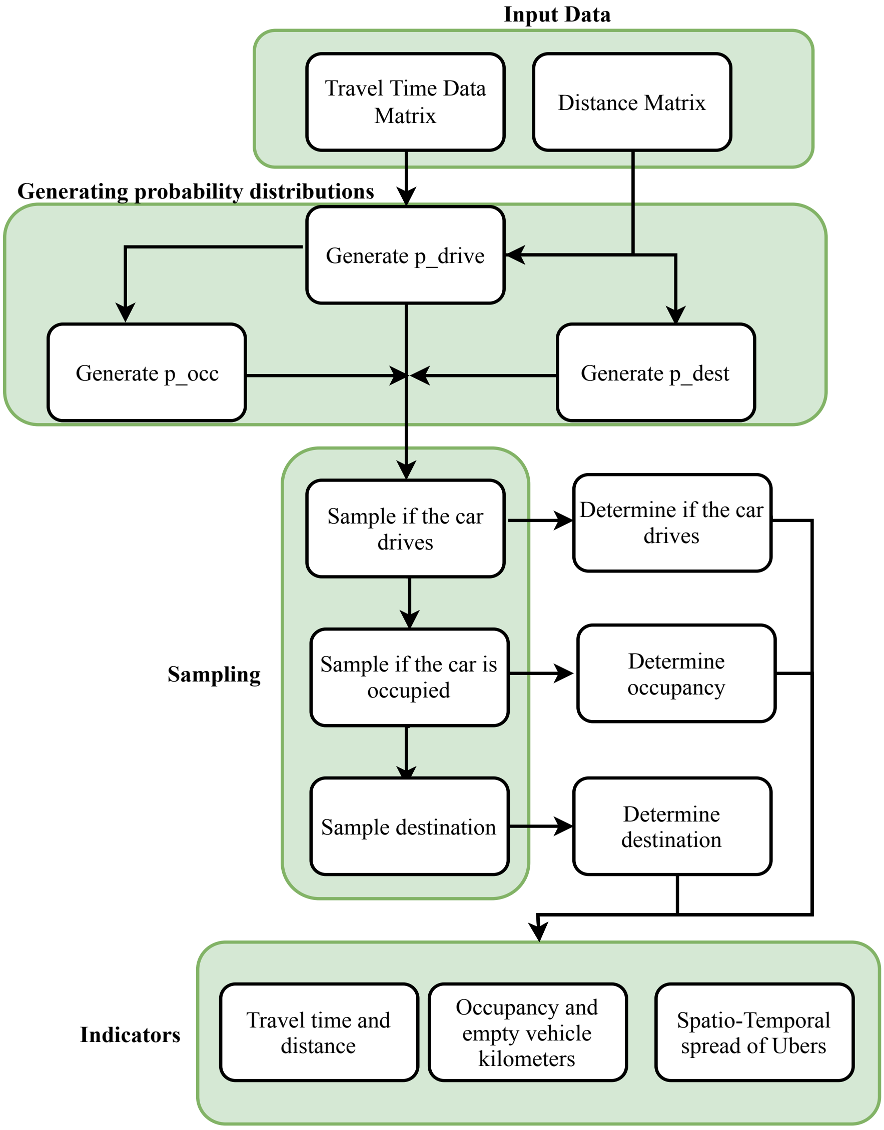


Figure 4.1: Conceptual framework of the model

the maximum possible data points between a 'sourceid' and 'dstid' cannot exceed 24 for a quarter and 96 for the whole year. The length of the available data points is divided by the maximum possible data points and the percentage in the current research for the 2018 data was found to 60.14%. The implication of the data sparsity is the limited coverage of the data. However, it will also supplement the model in estimating the frequently visited wijken in Amsterdam. The data sparsity could stem from the limited usage of Uber in Amsterdam or the concentration of Ubers spatially. For reference, the Melbourne data set had a data sparsity of 91%.

4.2.2 Preparing the distance matrix

The distances between a 'sourceid' and 'dstid' point were determined using the OSMNX package available in Python. The package utilises Open street maps for calculating distances between two coordinates. First, the centroid of each wijk is calculated and every wijk is assigned a centroid. As a result every 'sourceid' and 'dstid' points now have coordinates. In OSMNX, the coordinates are matched to the closest possible road network and the shortest path between them is calculated using real-world distances available in the map. The distances calculated are comparatively realistic than adopting Euclidean distances (as was done in the reference model) as it adheres to a road network. This is a computationally intensive step but needs to be run only once. The permutations of all origin-destination pairs and the distances between them are stored. The distances calculated are not entirely realistic. The route taken does not acknowledge access restrictions. Moreover, the actual pick-up and drop-off coordinates will vary from the centroid calculated for the wijk. Figure 4.2 visualises the road network available in OSMNX and the matching which took place between two coordinates.

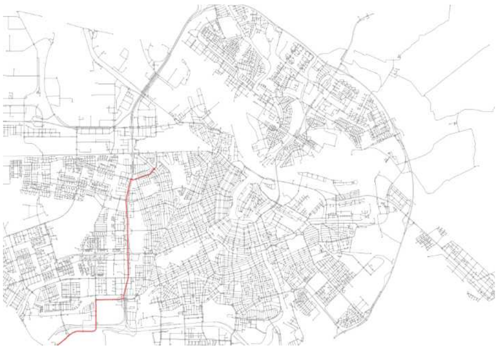


Figure 4.2: OSMNX package in Python determines distances between coordinates by matching the Amsterdam road network

The distances are generated for every 'sourceid' and 'dstid' combination, resulting in a permutation of 181 data points and 2 samples which equals to 32,580 data points. Thus, the distance from A to B is not the same as the distance from B to A. The distance matrix is a 181 x 181 matrix. The Coordinate Reference System (CRS) employed is the EPSG: 3256 which enables the calculation of distance in meters.

4.3 DEFINING PROBABILITY DISTRIBUTIONS

The subsection discusses the process of defining probability distributions. The reference model had probability distributions related to the probability of driving and the probability of choosing a destination. The current research extends upon it by introducing a probability of occupancy i.e the probability of an Uber being occupied in binary terms. We define the process of obtaining each of the distributions below:

4.3.1 Generating the probability of driving - p_drive

We define the probability of driving from a particular zone at a particular hour of the day. Thus, resulting in a N x T matrix where N is 181 and T is 24. The first step is to calculate the sum of the (mean) travel time 't' out of every sourceid 'i' to all other dstid 'j' i.e. $\sum t_i \forall$ j=1,2,3.. for t = 1,2,3..., this is followed by determining the minimum and maximum sum of travel time to all other zones for every hour of the day. The probability is then calculated using Equation 4.2. The sum of mean travel times from a zone to all other zones at a particular hour of the day is subtracted from the minimum sum of travel time out of at that hour of the day for all zones. This is divided by the subtraction of the maximum and minimum sum. The result is then multiplied by the difference of p_{max} and p_{min} set to 0.9 and 0.1 respectively. The value has been retained from the reference model. The result of p_{drive} is then exponentially increased by a factor e_{drive}. The value is set to two as was done in the reference model. If the vehicle does not drive, it is assigned a value of zero and is assumed to be in a parked state. The probability of parking can be simply expressed as shown in Equation 4.3. An additional value introduced is the d_{ij} which is the distance between sourceid 'i' and dstid 'j' obtained from the distance matrix. The addition of distance ensures higher travel times are not a direct result of longer distances. The p_{drive} across 24 hours of the day for different Wijken can be visualised as visible in Figure 4.3.

$$p_{ij}^{drive} = \left\{ \begin{array}{ll} p_{min} - (p_{max} - p_{min}) \times \dfrac{\sum_n^{j=1} \frac{\mu_{i,j,t}}{dij} - min_t\left\{\sum_N^{j=1} \frac{\mu_{i,j,t}}{dij}\right\}}{max_t\left\{\sum_N^{j=1} \frac{\mu_{i,j,t}}{dij}\right\} - min_t\left\{\sum_N^{j=1} \frac{\mu_{i,j,t}}{dij}\right\}} & max_t\left\{\sum_N^{j=1} \frac{\mu_{i,j,t}}{dij}\right\} > 0 \\ 0 & max_t\left\{\sum_N^{j=1} \frac{\mu_{i,j,t}}{dij}\right\} = 0 \end{array} \right\} \tag{4.2}$$

$$p_{ij}^{park} = 1 - p_{ij}^{drive} \tag{4.3}$$

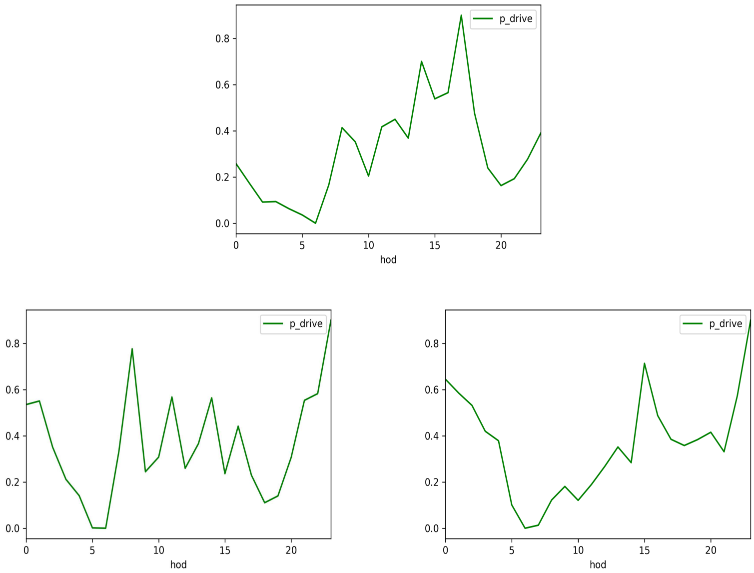


Figure 4.3: Probability of driving across 24 hours of the day as determined for; (a) Haarlemmerliede en Spaarnwoude (b) Zwaanshoek (c) Delftwijk

4.3.2 Generating the probability of occupancy - p_occ

Post defining if the car drives or not, we determine if the vehicle is occupied. We define the probability of an Uber being occupied if it is driving and is an extension made over the reference model. The occupancy is determined in binary terms and not the actual number of passengers. Once again, the resulting probabilities will be stored in a 181 x 24 matrix. The sum of data points is determined from one sourceid 'i' to all other dstid 'j' for a particular time period. The minimum and maximum values are also obtained for every hour of the day. The weighting parameter e_{occ} is set to 2 and the values p_{max} and p_{min} are set to 0.9 and 0.1 as was done previously in p_{drive}. The formula for determining p_{occ} is shown in Equation 4.4. Here t refers to one data point from sourceid 'i' to dstid 'j'. The formula only differs from p_{drive} in the use of data penetration points over travel time and is not normalised by distance. The probabilities generated can be visualised as shown in Figure 4.4. The probability of the vehicle not being occupied will be the supplement of p_{occ} as shown in Equation 4.5.

$$p_{ij}^{occ} = \left\{ \begin{array}{ll} p_{min} - (p_{max} - p_{min}) \times \frac{\sum_{n}^{j=1} t_{i,j,t} - min_t\left\{\sum_{N}^{j=1} t_{i,j,t}\right\}}{max_t\left\{\sum_{N}^{j=1} t_{i,j,t}\right\} - min_t\left\{\sum_{N}^{j=1} t_{i,j,t}\right\}} & max_t\left\{\sum_{N}^{j=1} t_{i,j,t}\right\} > 0 \\ 0 & max_t\left\{\sum_{N}^{j=1} t_{i,j,t}\right\} = 0 \end{array} \right\} \tag{4.4}$$

$$p_{ij}^{unocc} = 1 - p_{ij}^{occ} \tag{4.5}$$

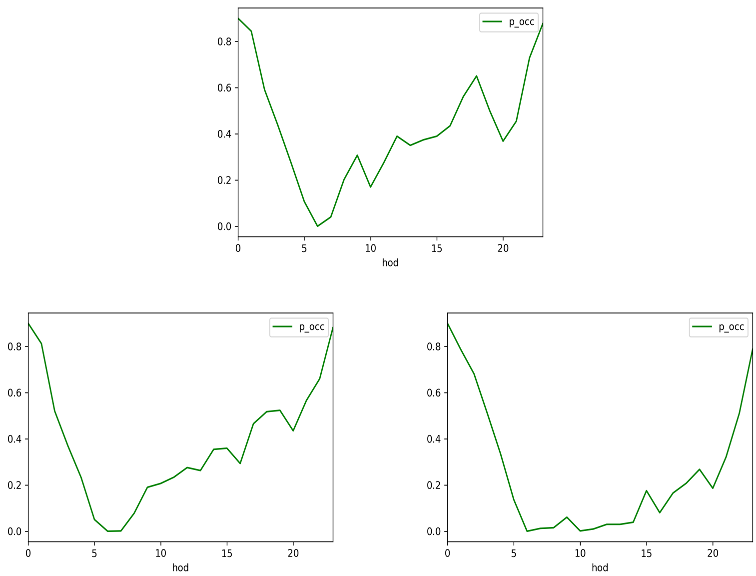


Figure 4.4: Probability of occupancy across 24 hours of the day as determined for; (a) Haarlemmerliede en Spaarnwoude (b) Zwaanshoek (c) Delftwijk

4.3.3 Generating the probability of choosing a destination - p_dest

The final probability distribution to be generated is the probability a vehicle at sourceid 'i' will choose 'dstid' j at a given time of the day. Due to the directionality involved, we now generate N x N x T matrices (181 x 181 x 24) resulting in 786,264 probabilities. The p_{dest} is calculated by obtaining the difference of the travel time between a sourceid 'i' and dstid 'j' at a particular hour of the day and the minimum travel time between that sourceid 'i' and dstid 'j' across all 24 hours of the day. This is then divided by the difference of the maximum and minimum travel time for the sourceid-dstid pair across all hours of the day. The weighting parameter e_{drive} is set to 2. Thus, p_{dest} can be calculated as shown in Equation 4.6. The probability of choosing a certain destination from multiple 'sourceid' at a certain hour of the day is obtained and the resulting multinomial probability distribution can be visualised as shown in Figure 4.5. The rationale behind the formula can be expressed as; if the travel time at an instant is closer to the maximum travel time to a zone across all hours of the day, then the probability the driver chooses the destination is higher.

$$p_{ij}^{dest} = \begin{cases} \frac{1}{N_{it}} \frac{\mu_{i,j,t} - min\left\{\mu_{i,j,t}\right\}}{max\left\{\mu_{i,j,t}\right\} - min\left\{\mu_{i,j,t}\right\}} & max\left\{\mu_{i,j,t}\right\} > 0 \\ 0 & max\left\{m_{i,j,t}\right\} = 0 \end{cases} \tag{4.6}$$

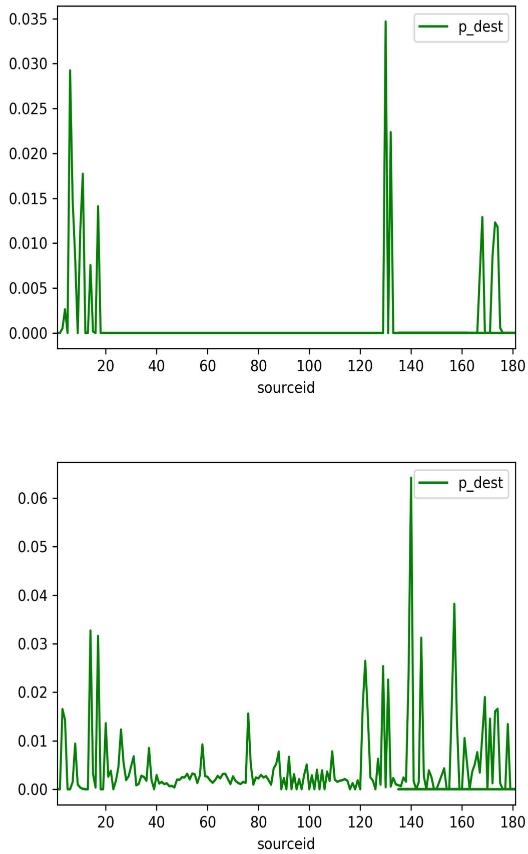


Figure 4.5: Probability of choosing destination at hod = 0 for (a) Molenwijk (b) Burgwallen-Nieuwe Zijde

4.4 SAMPLING THE DISTRIBUTIONS AND DETERMINING THE INDICATORS

The sampling steps involve establishing if a vehicle drives, if yes, is it occupied and where does it drive to. Besides, the travel time, route and distance are calculated. The inclusion of occupancy and route are an extension over the reference model. A state matrix is defined of the dimensions N x c x T representing the number of cars across Wijken and 24 hours. Here, c refers to the number of vehicles to be simulated per wijk. The state matrix can be represented as shown in Equation 4.7

$$S_{c,t} = \begin{Bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \cdots & c_{1,T} \\ c_{2,1} & c_{2,2} & c_{2,3} & \cdots & c_{2,T} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{C,1} & c_{C,2} & c_{C,3} & \cdots & c_{C,2T} \end{Bmatrix} \tag{4.7}$$

This is followed by generating transition matrices, which contain the information for the next state at time period t+1. The transition matrices have the same dimensions as the state matrix. Five transition matrices are prepared for storing if a car drives, the occupancy, the destination is chosen, the travel time, route and the fifth one for distance. These matrices are filled in with new values, as the Markov property is memoryless in nature and does not retain information on the previous state. The sampling process involves comparing the probabilities generated against a randomly generated number.

4.4.1 Sampling p_drive and p_occ

If the p_{drive} is higher for a sourceid 'i' at hour of day t is higher than a randomly generated number, the car drives and the transition matrix retains a value of 1. If p_{drive} is lower than the randomly generated number, a value of zero implying the car does not drive is entered. Algorithm 4.1 indicates the algorithm employed for sampling if a car drives from the probability distribution. Algorithm 4.2 formulates the algorithm for sampling occupancy. The difference from sampling p_{drive} lies in the condition if the car is driving. Thereby, the occupancy is only determined for those vehicles which are driving.

Algorithm 4.1: Sampling if a car drives

```
c ← C
t ← T
origin = state_matrix[c, t]
pdrive = pdrive_matrix[origin, t]
if pdrive ≥ random() then
    transition_matrix_drive[c, t] = 1
else
    transition_matrix_drive[c, t] = 0
```

Algorithm 4.2: Sampling if a car is occupied

```
c ← C
t ← T
drive = state_matrix[c, t]
origin = state_matrix[c, t]
if drive = 1 then
    pocc = pocc_matrix[origin, t]

    if pocc ≥ random() then
        transition_matrix_occ[c, t] = 1

    else
        transition_matrix_occ[c, t] = 0
```

4.4.2 Sampling p_dest

The sampling of the probability of destination determines the destination, the vehicle will drive at a given hour of the day from a wijk, conditional to the car having decided to drive. The probability is first sampled from a probability distribution as shown in Section 4.3.3. The sampling of destination is independent of occupancy. If the sum of distribution is found to be higher than zero, the randomly generated number is compared against a lower range and an upper range equivalent to the probability of choosing the destination. In the event, the number is found to be between the upper range and lower range, the destination at the instance is assigned to the transition matrix. If the sum of distribution equals to zero, the vehicle is assumed to drive within the Wijk. The sum of distribution tends to be zero, due to the absence of data points in the original data set. The sampling process can be algorithmically described as shown in Algorithm 4.3. Concerning the entire sampling process, post the completion of one loop at time period T, the next loop receives an updated state matrix at T+1 with the vehicles in their new locations, which in turn is chosen as the origin. The process of sampling for driving, occupancy and destination is repeated with the updated state matrix.

Algorithm 4.3: Sampling if a destination is chosen

```
c ← C
t ← T
drive = transition_matrix_drive[c, t]
if drive = 1 then
    origin = state_matrix[c, t]
    range_up = 0
    range_low = 0
    distribution = pdest_matrix[origin, :, t]

    if distribution.sum() = 0 then
        destination = origin
        transition_matrix_dest[c, t] = destination
    if distribution.sum() > 0 then
        j ← J
        if range_low < random() < range_up then
            destination = j
            transition_matrix_dest[c, t] = destination
            break
            range_low = range_up
            destination = origin
            transition_matrix_dest[c, t] = destination
state_matrix[c, t + 1] = transition_matrix_dest[c, t]
```

4.4.3 Determining Travel Time

After the determination of the origin and destination, the travel time can be estimated from the travel time data matrix. The origin-destination, and hour of the day are offered as an input along with mean travel time and standard deviation for the pair. The uber movement travel time data set for Amsterdam follows lognormal distribution(Section 2.1). Consequently, the travel times are sampled from a lognormal distribution. Travel times for trips within the wijk are not modelled due to the absence of data points.

4.4.4 Determining route and distance

The route and distance are determined through network graphs and the distance matrix. Determining the route is an extension made over the reference model. The spatial graph was constructed with the wijken represented as nodes. A connection between two nodes exists if the wijk are adjacent to each other. This enabled the development of an adjacency matrix. The edges are weighted by distances obtained from the distance matrix. As a next step, the shortest path algorithm was used to determine the route between an origin wijk and a destination wijk. The edges which form the shortest path are then determined and the distance is simply a summation of the weights (distances). The Uber Movement data can be abstracted as a network graph. This creates the potential for determining the route using the shortest path algorithm. An additional dimension added when determining the distance is occupancy. This enables determining the empty vehicle kilometres. Each car ID with an occupancy value of 1 or 0 is identified. The movement of these vehicles across T=24 is determined and the total distance travelled when occupied and unoccupied can be determined.

4.5 DEFINING THE PARAMETERS: E_DRIVE, E_OCC AND E_DEST

The parameters e_drive, e_occ and e_dest are the exponential parameters utilised in the equations which define the probability for driving, occupancy and choosing a destination respectively. On testing different values for e_drive, it was revealed an exponent closer to zero, lead to a sharper gradient and the gradient smoothens as the parameter value increases. This also holds when changing the parameter for e_occ. The curves have been visualised in Figure 4.6 for a wijk across 24 hours of the day. The e_dest parameter was tested with different values and the probability of choosing the destination Burgwallen Nieuwe Zijde from different 'sourceid' with an hour of the day as zero. The results have been visualised as shown in Figure 4.7. The parameters are responsible for smoothening the gradient. The reference model decided on the parameters based on empirical data. The open-source data on the number of cars on the streets in Melbourne across different hours of the day is used to estimate the parameter by implementing a gradient descent optimisation algorithm. The reference model found the optimal parameter for e_drive to be 0.5. For parameter e_dest, the measured parking density across different zones is used to improve the value. The e_dest parameter for the reference model was found to be 0.59. Besides, the parameters, p_min and p_max were found to be 0.1 and 0.5 respectively and relate to the minimum and maximum probability of driving. The value used in the current research equals to 2. Validating the 'e' parameters for the traffic activity in Amsterdam can fail to reveal taxi patterns in the city. At the moment no such empirical data for taxis in Amsterdam exists. The goal of the reference model was to simulate the general traffic activity across zones for all types of vehicles, while in the current research the focus is to retain the movement of taxis brought about by the data penetration levels in the Uber Movement data set.

4.6 INITIAL VALUE PROBLEM

The initial value problem relates to the initial state describing the distribution of vehicles per wijk. The model is initially run with the vehicles equally distributed across all wijken. This is not an ideal initial state. To solve the initial value problem, the model is simulated from T=0 to T=24. The time steps can vary based on the temporal aggregation of the data. It could vary between the hour of the day, day of the week and month. The T=24 state is set as the initial state. The model is then re-run with a new initial state.

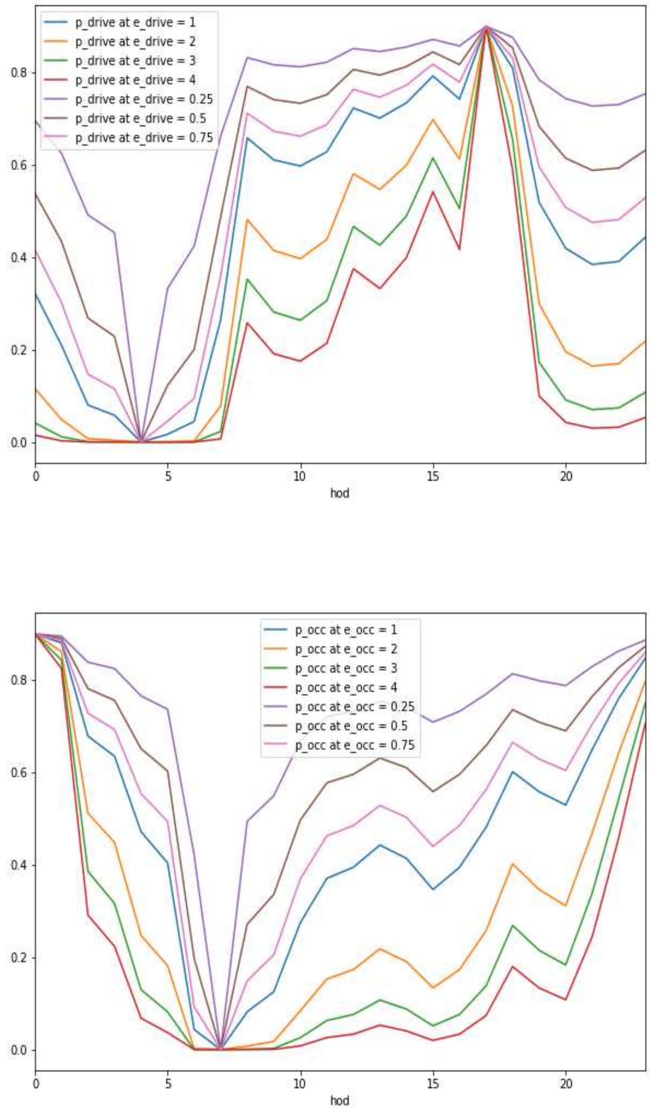


Figure 4.6: Varying parameters (a) e_drive and (b) e_occ for Burgwallen Nieuwe Zijde

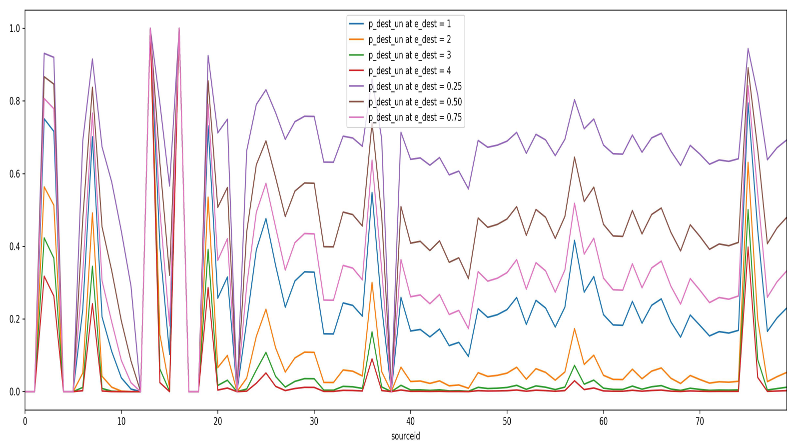


Figure 4.7: Varying e_dest values for destination Burgwallen Nieuwe Zijde at hour of day = 0

5 MODEL CASE STUDY & RESULTS

The current chapter explains the implementation of the model through a case study and discusses the results obtained. Also, results and discussion of the the modelling steps undertaken in the previous chapter is presented. The results from the model include the spatio-temporal spread of Uber vehicles, the occupancy levels, number of empty trips, the empty kilometres travelled and the travel time experienced by Uber vehicles. In addition to this, the most frequently traversed wijken have been presented.

Three scenarios are simulated. The first scenario is with 1810 Uber vehicles i.e. every wijk (a total of 181) is initally assigned ten vehicles resulting in 1810 Uber vehicles. and as per the solution to the initial value problem discussed in the previous chapter, the model is first simulated till the time step T=24. This is then treated as the inital state and the results from the model are obtained including the properties such as travel time, distance and occupancy between states. The second and third scenarios are simulated with 905 (5 Uber vehicles per wijk initially) and 3620 (20 Uber vehicles per wijk initially) respectively. The spatio-temporal distribution for the latter two scenarios can be found in Figure B.1 and Figure B.2. For brevity, only the first scenario is discussed in this chapter. The simulation of the second and third scenarios shows the relative proportion of Uber vehicles across wijken remains the same only the magnitude of Uber vehicles increases or decreases depending on the fleet size. The following sections discuss each of the results for the first scenario in detail.

5.1 SPATIO-TEMPORAL SPREAD OF UBERS

On simulating 1810 Uber vehicles post solving the initial value problem the spatio-temporal spread of Uber vehicles i.e. the number of vehicles per wijk per hour of the day was obtained. The spatial spread across the time periods, early morning (00:00 - 06:00) Figure 5.1, AM peak (07:00 - 09:00) Figure 5.2, midday (10:00 - 15:00) Figure 5.3, PM peak (16:00 to 18:00) Figure 5.4 and the evening peak (19:00 to 00:00) can be visualized as shown in Figure 5.5. The vehicles concentrate in areas around Schiphol, the centre of the city (the 17th-century ring), Sloterdijk, the port area of Westelijk Havengebeid and the wijken, North of the IJ river, of Landsmeer and Ilpendam.

The number of vehicles across the different time periods are as shown in Table 5.1 and Table 5.2. The AM peak sees Lijnden and Boesingheliede with the highest number of Uber vehicles. The wijk is located just North of Schiphol and has the major arterial Schipholweg passing through it. Badhoevedorp is also located North of Schiphol and comprises of hotels close to the airport. Schiphol Rijk, located South of Schiphol with the third-highest number of Ubers comprises of freight forwarding services near the airport and also hotels. The wijk Burgwallen Oude Zijde with the Amsterdam central station also features and this is followed by Amstelveen, a wijk close to the Southern Amsterdam business district dominated by expat residences and hotels. For the PM peak, Heemstede Centrum which neighbours Hoofddorp

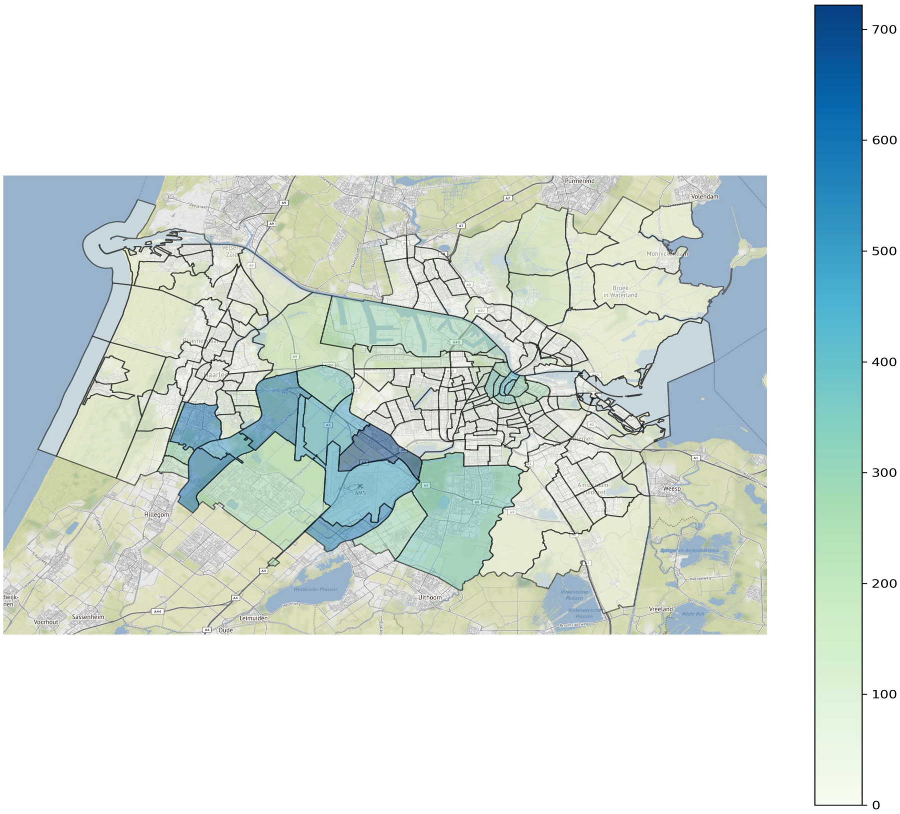


Figure 5.1: Spatial spread - Early morning (00:00 to 06:00)

Figure 5.2: Spatial spread - AM peak (07:00 to 09:00)

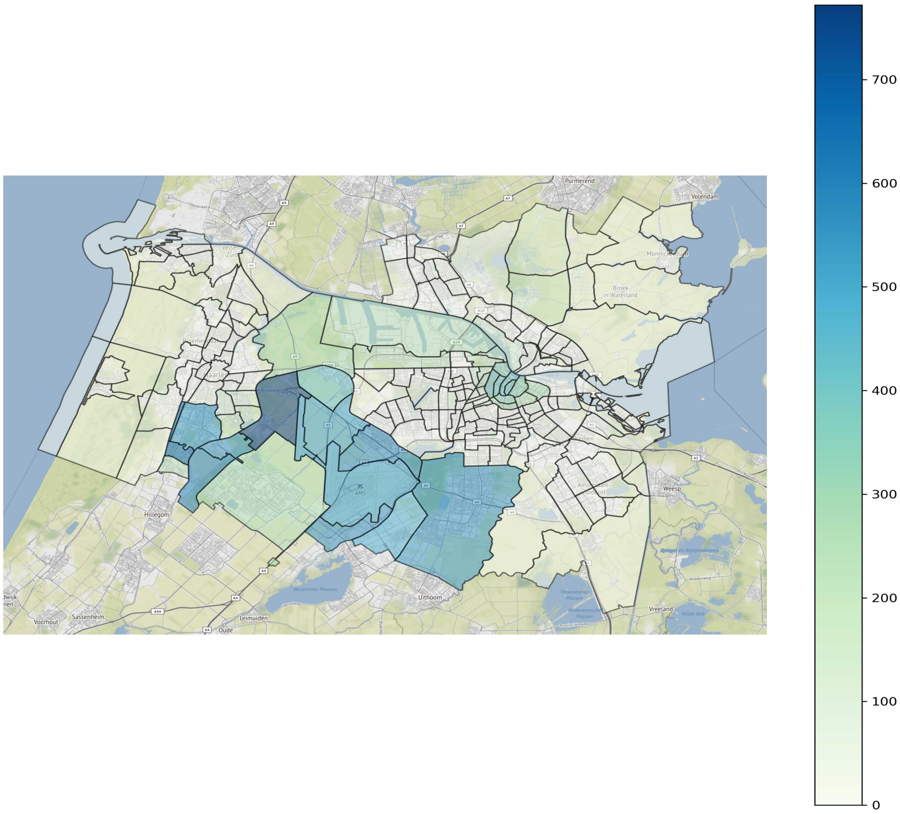


Figure 5.3: Spatial spread - Midday (10:00 - 15:00)

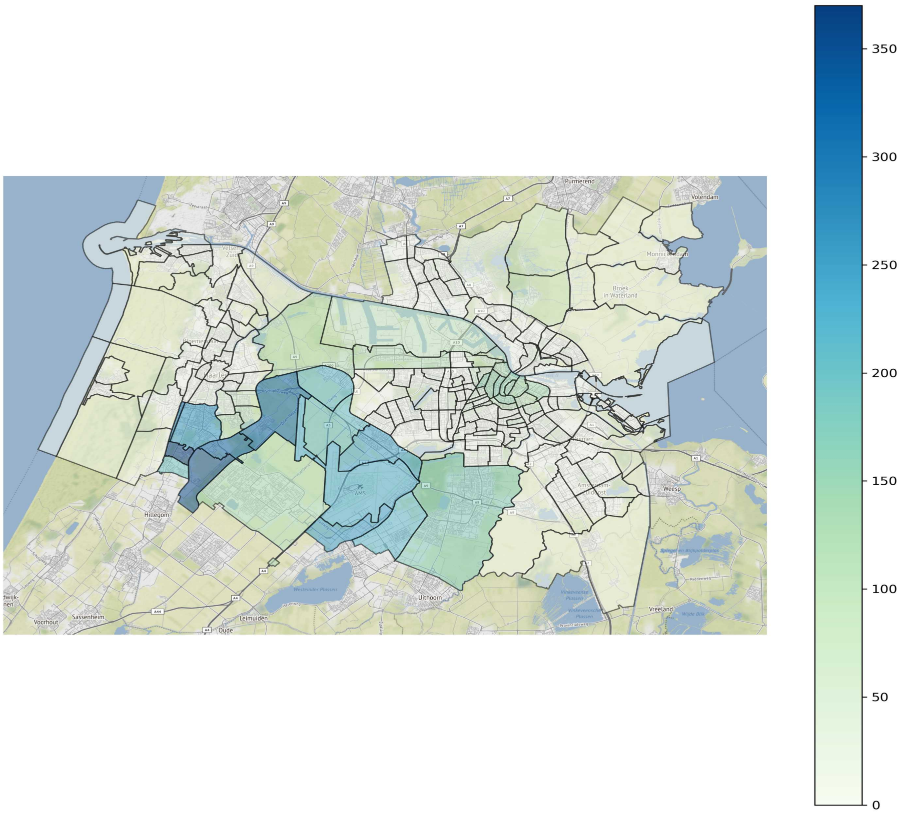


Figure 5.4: Spatial spread - PM peak (16:00 to 18:00)

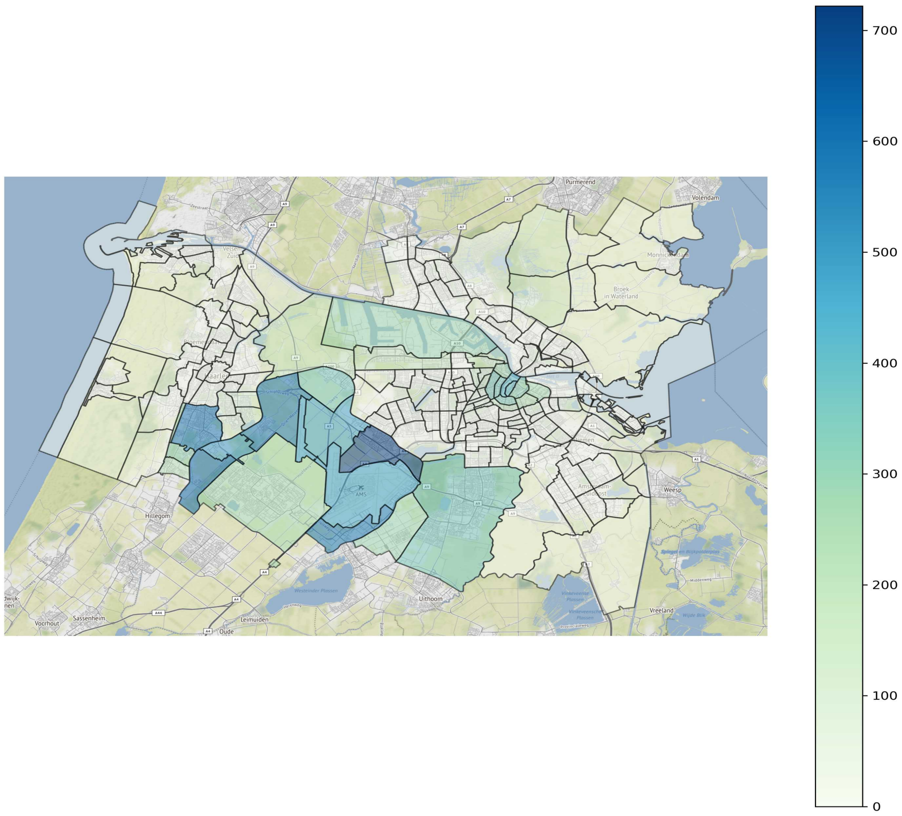


Figure 5.5: Spatial spread - Evening (19:00 to 00:00)

and Schiphol Rijk feature with the highest number of Uber vehicles. Zwaanshoek, also next to Hoofddorp has the third-highest number of Uber vehicles as per the model. None of the Wijken in the centre feature for the PM peak. The wijken in the centre of Amsterdam: Burgwallen Nieuwe Zijde, Bugwallen Oude Zijde and Grachtengordel West also have a significant number of Ubers (221, 189 and 184 respectively in the AM peak and 125, 122 and 133 respectively in the PM peak). Table 5.2 with wijken comprised of the Uber vehicles shows Zwaanshoek, Cruiquis and Lijnden/Boesingheliede as the Wijken with the most number of Ubers. The Wijken with the highest number of vehicles tend to remain consistent across all time periods. This was also noted in the midday and evening period. Wijken from the peak periods have been indicated in the tables (along with the early morning period)

Table 5.1: AM peak and PM peak Wijken with the most Uber vehicles (driving and parked)

AM peak		PM peak	
Wijk	**Ubers**	**Wijk**	**Ubers**
Lijnden / Boesingheliede	383	Heemstede-Centrum	370
Badhoevedorp	338	Schiphol Rijk	359
Schiphol Rijk	309	Zwaanshoek	332
Cruquius	302	Cruquis	326
Schiphol	260	Heemsted Zuid	306
Vijfhuizen	259	Schiphol	261
Zwanenburg	254	Badhoevedorp	254
Zwaanshoek	249	Lijnden / Boesingheliede	243
Burgwallen-Oude Zijde	243	Vijfhuizen	229
Amstelveen	210	Hoofddorp	216

Table 5.2: Wijken with the highest number of vehicles in the early morning period

Early Morning	
Wijk	**Ubers**
Lijnden / Boesingheliede	722
Schiphol Rijk	636
Badhoevedorp	594
Schiphol	586
Cruquius	584
Zwaanshoek	500
Vijfhuizen	478
Zwanenburg	473
Burgwallen-Oude Zijde	419
Amstelveen	401

Discussion - Spatio Temporal spread

The results suggest, Ubers tend to concentrate in and around Schiphol, followed by the centre of the city and the port area of Westelijk Havengebeid. The Ubers are used to travel from or to wijken near Schiphol. For instance, wijken such as Hoofddorp and Schiphol Rijk comprise of hotels Figure D.1. Wijken such as Lijnden/Boesingheliede and Badhoevedorp are wijken with traversals but are unlikely areas of pick-ups and drop-offs. Cruquis and Zwaanshoek also, have a concentration of Uber likely due to the frequent traversals and the data penetration in these areas being higher.

Amsterdam Sloterdijk and Westelijk Havengebied are wijken with enterprises and hotels. Amsterdam Sloterdijk is especially monofunctional concerning it's land-use and the concentration of 77 and 76 Ubers in the AM and PM peak respectively is suggestive of Ubers being used for work-related trips in these Wijken. Amstelveen is also suggestive of work-related trips. In terms of the applicability, the results offer spatio-temporal distribution of Ubers. This can in turn offer insights to enable access control for taxis during cetrain hours of the day. For instance, in the centre of the city, the number of Ubers across the AM and PM peak have been visualized as shown in Figure 5.6. The varying number of Ubers can be evaluated to analyse time periods or wijken when the entry of Ubers can be restricted.

The spatial concentration of Ubers across different time periods is suggestive of the limited purposes behind Uber usage. The results from the model offer only a strategic overview of the dynamics of taxi distribution in the greater Amsterdam metropolitan region and the model cannot simulate precise spatio-temporal distribution due to the aggregation of the data and the absence of an indication of the actual number. The model instead simulates based on the data penetration levels and the movement across wijken are a function of travel time divided by distance. The results from the model are consistent with what was found in data exploration. For instance, Amstelveen receives a large number of Uber vehicles 210 Ubers and 401 Ubers in the AM and early morning period respectively. Amstelveen has been found to have a large expatriate population, thereby indicating the primary users of Ubers as foreigners and tourists. The cellphone data also indicates a large number of international sims in Amstelveen Table 3.9.

The modelling decisions have a direct impact on the results. One of the assumptions made is concerning the probability of driving increasing as the travel time to a wijk relative to the others increases. The consequence of such a decision is the increase in p_drive during the AM (07:00 to 09:00) and PM peak (16:00 to 18:00) as travel times are higher during the peak hours. However, in the wijken Zwaan-

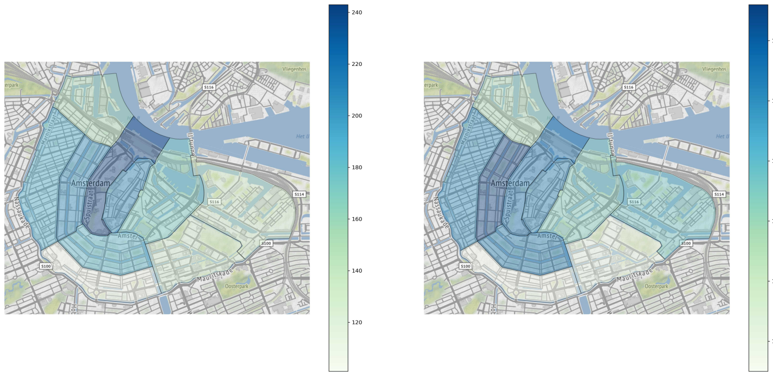


Figure 5.6: Number of Ubers in the centre AM peak (left) & Number of Ubers in the PM peak(right)

shoek and Delftwijk, the p_drive is found to be the highest at midnight. This is a consequence of the higher data penetration during midnight. Figure 5.7 depicts the travel time in seconds across the day, the data penetration points and the travel time divided by distance for Haarlemmerliede and Spaarnwoude. Thus, the probability of driving and consequently the movement of Ubers across wijken is a function of both travel time and data penetration.

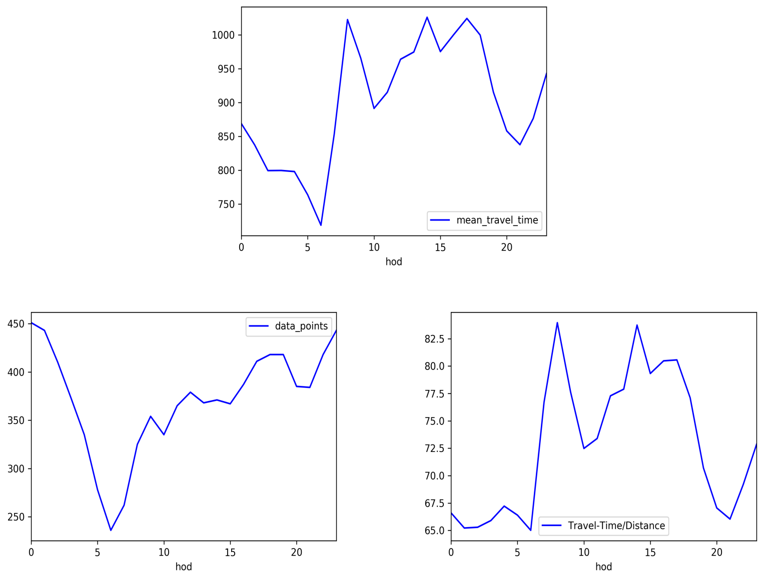
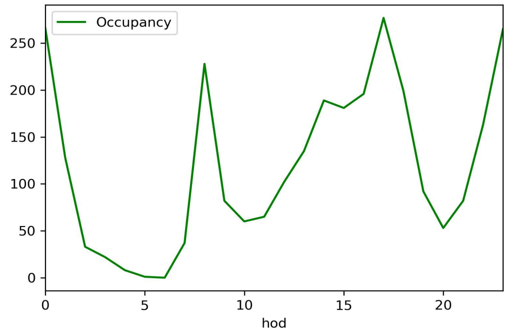


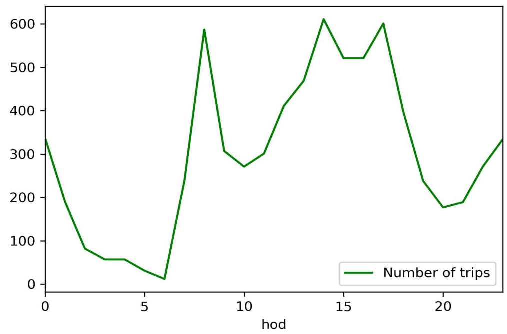
Figure 5.7: Haarlemmerliede and Spaarnwoude (from left to right) - (a) Mean Travel Time in seconds across the day (b) Data points across the day (c) Travel time divided by distance across the day

5.2 OCCUPANCY LEVELS

The occupancy level indicates if the Uber vehicle is occupied at the instance when it is driving. A binary coding system is followed. If the Uber is occupied, it receives a value of '1', and '0' otherwise i.e. the occupancy is not the actual number of passengers. The occupancy levels across different hours of the day have been plotted in Figure 5.8a. The figure suggests occupancy levels are the highest at midnight and the evening peak. The number of trips at midnight are the lowest as shown in Figure 5.8b.



(a) Occupancy levels across the day



(b) Number of trips across the day

Figure 5.8: Occupancy and number of trips as indicated by the model

Discussion - Occupancy levels

The higher occupancy at midnight is indicative of greater demand. The limited frequency of Public Transport could be one of the principal causes. Uber especially becomes advantageous as it offers a door to door service. This is consistent with what was observed in demand studies, where the 'sourceid' and 'dstid' numbers went up during midnight. Additionally, the service area of Uber spans the greater Amsterdam region. Connectivity to areas outside the centre may especially become sparse, prompting greater usage. Thus, the model result suggesting greater occupancy during these hours only seems logical. The number of trips, however, are lower at night compared to the peaks and midday period. The lower number of trips could also pertain to limited empty trips during these hours as Ubers are not required to relocate to pick-up passengers during these hours of the day. This was also found when empty vehicle kilometres and occupied vehicle kilometres were compared in the next section.

5.3 EMPTY VEHICLE KILOMETERS

Each Uber and its occupancy level was tracked across different hours of the day in the model. This enabled the calculation of total 'Occupied vehicle km' i.e the total distance travelled by an occupied Uber and 'Empty vehicle km' is the number of kilometres travelled while a vehicle was without a passenger either in the process of searching for a passenger or on its way to pick-up a passenger. The empty vehicle kilometres accounted to about 48,498 km whereas the occupied vehicle kilometres was considerably lesser at 35,189 km according to the model. Thus, according to the model, only 42% of the vehicles kilometres were with a passenger. The proportion of empty vehicle kilometres was found to be higher during the AM peak and mid-day. The proportion of occupied and empty kilometers can be visualised as seen in Figure 5.9.

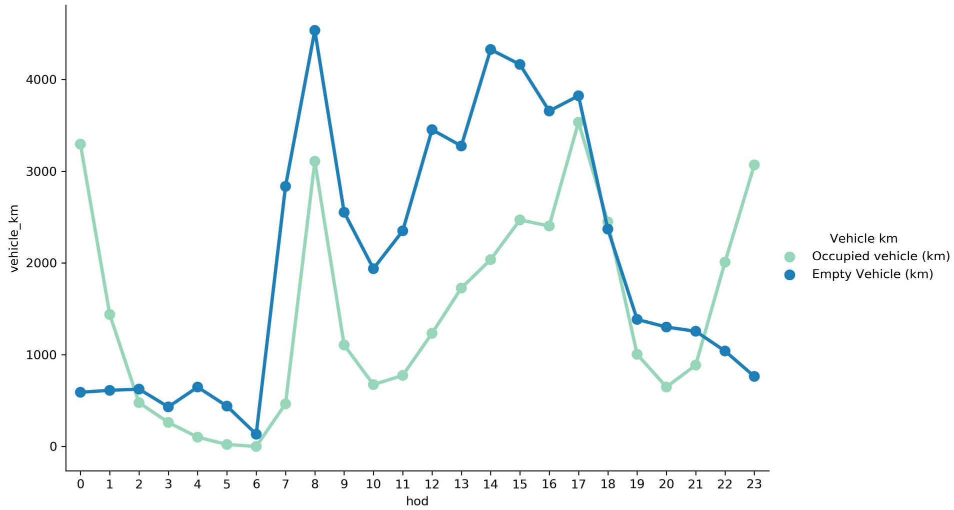


Figure 5.9: Total occupied vehicle kilometres and empty vehicle kilometres

Discussion - Empty vehicle kilometers

The results of the empty vehicle kilometres and occupied vehicle kilometres only when the occupancy levels and the number of trips are observed. According to the model, the demand-supply equilibrium is better matched during midnight and is caused by the higher demand. The relocation, in turn, contributes to empty kilometres. While these aspects have not been captured in the model, the extent of empty vehicle kilometres is expected to be significant as it is caused by the asymmetry in demand and can be noted in the data penetration Figure 3.11. The empty vehicle kilometers will reduce if Equation 4.2 is normalised by the data penetration points, as the asymmetry will then be account for when p_drive is calculated. Occupancy is a function of the data points, so these peak at midnight as was seen in the data penetration levels and reduce. Consequently empty vehicle kilometers is a result of this. Normalisation will lead to reduction in empty vehicle kilometers. Thus, the results must be interpreted with due consideration of the modelling decisions and data in this case.

5.4 FREQUENTLY TRAVERSED WIJKEN

An extension of the reference model produces the route followed between an origin and destination by a network graph and the shortest path algorithm. Besides, occupancy and the route it takes is included as extensions. The intention is to identify routes which are frequnted by those vehicles which drive empty. The first type of routes generated are those with no intermediate Wijken. The 'sourceid' and 'dstid' pairs with no intermediate Wijken and no occupancy has been visualised as a matrix as shown in Figure 5.10. The origin-destination pair of Heemstede Zuid and Heemstede Centrum have the highest number of empty trips occurring between them. All of the Wijken which feature in the matrix are in and around Schiphol airport.

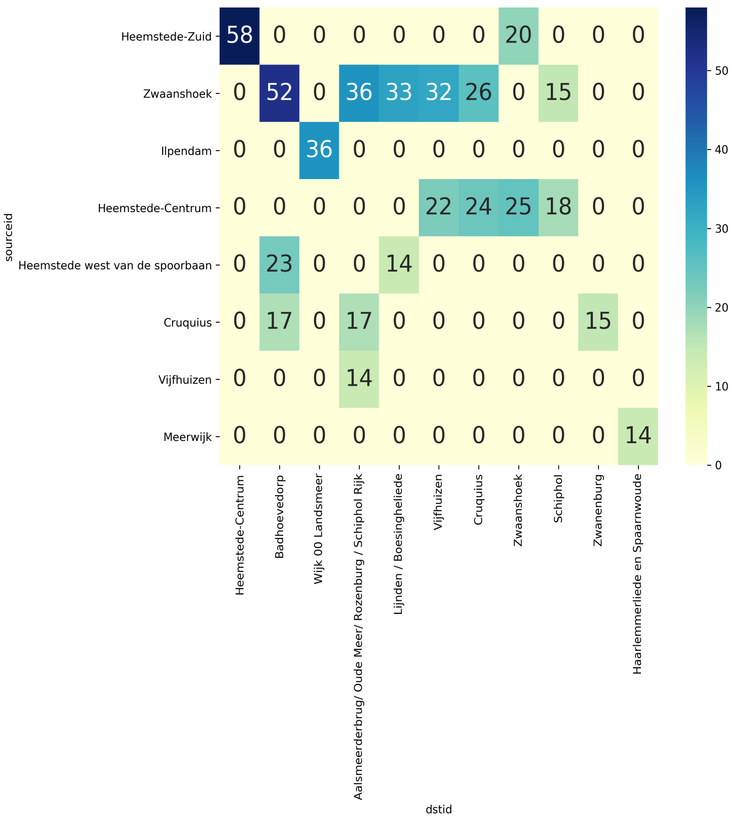


Figure 5.10: Most frequent 'sourceid'-'dstid' pairs with empty trips

Table 5.3 tabulates the routes where there exists one intermediate Wijk between the 'sourceid' and 'dstid' for empty and occupied trips respectively. The most frequent origin-destination pairs with empty trips are once again in areas around Schiphol whereas, the occupied trips are at the centre of the city or the North of the IJ river (Ilpendam). The trend remains consistent in Table 5.4 with two intermediate wijken as well.

Table 5.3: The paths most frequented with one intermediate wijk

Empty trips across an intermediate wijk			
Origin	**Intermediate wijk**	**Destination**	**Number of trips**
Cruquius	Heemstede-Centrum	Meerwijk	15
Heemsted-Zuid	Heemsted-Centrum	Heemstede West	13
Occupied trips across an intermediate wijk			
Origin	**Intermediate wijk**	**Destination**	**Number of trips**
Haarlemmerbuurt	Lijnden/Boesingheliede	Burgwallen Nieuwe Zijde	12
Ilpendam	Cruquius	Meerwijk	9

Discussion-Frequently Traversed Wijken

Network graph offering routes are not the most realistic tool as edges can directly connect nodes even during the absence of such edges in the real world. Here if a Wijk is adjacent to another, an edge is established. This is a simplification done but due to the aggregated nature and absence of routes in the data. This was the best possible measure. This is especially notable in the Oostelijke Eilanden to Schiphol

Table 5.4: The paths most frequented with two intermediate wijken

Empty trips across two intermediate wijken				
Origin	**Intermediate wijk - I**	**Intermediate wijk - II**	**Destination**	**Number of trips**
Heemstede-Zuid	Heemstede Centrum	Cruquis	Amstelveen	10
Heemstede-Zuid	Haarlemmerhoutkwartier	Landlust	Grachtengordel Zuid	5
Occupied trips across two intermediate wijken				
Origin	**Intermediate wijk - I**	**Intermediate wijk - II**	**Destination**	**Number of trips**
Oostelijke Eilanden	Weesperbuurt/ Plantage	Hoofddorppleinbuurt	Schiphol	5
Vondelkwartier	Vogelenwijk	Slachthuiswijk	Meerwijk	2
Cruquius	Vijfhuizen	Landlust	Grachtengordel-West	2

case, where the number of wijken one would traverse would be higher than two. The higher occupancy trips in the centre seem logical considering these are dense areas where an Uber is likely to get a trip compared to the sparser areas near the airport where there might be higher asymmetry in demand. The number of trips and traversals are insignificant in number to derive conclusions on if these add to congestion.

5.5 VALIDATION

Total vehicle kilometers travelled by taxis according to sectoranalysis mobility report – 460,000 km [Gemeente Amsterdam, 2019]. 35% of the total taxis in operation are platform vehicles. If assumed 35% of the vehicle kilometers are travelled by Uber, it results in a total of 126,000 km. According to the model, the distance travelled by 1810 Uber vehicles results in a total of 79,534 km resulting in a percentage difference of 37%. The distance calculated from Uber vehicles is based on the shortest path algorithm. Therefore, according to sectoraanlyse mobiliteit, the per capita vehicle kilometers travelled is 60km a day per vehicle compared to 44 km per day per vehicle. So a factor difference of 1.36 can be observed between total vehicle kilometers for Ubers suggested in the report and model results.

A better method for validation could be to employ Automated Number Plate Recognition (ANPR) for blue coloured number plates (taxis in the Netherlands have a blue coloured number plate) compare the number of taxis across different hours of the day against the model results. For instance, this can be specifically done for the center of the city and the number of Ubers depicted in Figure 5.6 can be validated. To specifically determine if the blue plate is an Uber vehicle, one can check the registration and remove number plates registered as a taxi at the Toegelaten Taxi Organisaties (TTO).

5.6 LIMITATIONS

As with every model, there are limitations and, the most important ones are explained here. The first one is concerning the nature of the data itself. The model utilises data penetration to impact the choice of wijken. However, a single data point can be representative of a minimum of five Uber vehicles, and the number can go up to the total number of vehicles in Amsterdam. Thus, in reality a single data point might capture more Ubers than multiple data points. This would significantly impact the spatio-temporal distribution generated by the model. The second limitation is a symptom from the use of Markov chains. The model cannot vary the fleet size as it progresses from one time step to the other. This is an important limitation as in reality the supply of Ubers can vary throughout the day. For instance, [Brodeur and Nield, 2018] found the supply of Ubers increased during rains, as drivers were encouraged by higher taxi fares (due to the dynamic pricing prevalent in Ubers). The third limitation relates to the underlying assumption that occupancy and consequently empty vehicle kilometers is based on the number of data points. The interpretation of the results must be based on the understanding that they have been determined as a consequence of modelling choices and are not directly based on empirical data.

6 CONCLUSION

The chapter offers the main conclusions through answers to the sub-questions and the main research question. This is followed by the implications section. Answer to the first sub-question is derived from Chapter 2.

Sub Question 1: What is the current state of the art with respect to the use of Uber Movement Travel Time Data and other Taxi GPS data sources for applications in Traffic congestion analysis and management?

The current state of the art concerning use of Uber Movement data has been limited to case studies where the Uber Movement interface was utilised to evaluate travel time changes by benchmarking it against a baseline travel time at a data before the incident. The incident or disruption included effects of bad weather, infrastructure closure, public transport shut down and holiday season traffic. Thus, events impacting the physical capacity of the network or seasons of higher demand and its impact on travel time. In academia, previous work has been done with network graphs, for identifying congestion patterns and using degree centrality as 'used degree centrality' to quantify data penetration levels for the spatial units.

Concerning other Taxi GPS data sources, the applications can be broadly categorised as demand-related, supply related, performance-related and impact related. Taxi GPS data can be evaluated in terms of demand which can vary spatiotemporally and help identify points of interest and also if there exists asymmtery in demand. For supply and performance-related application, taxi GPS data can help identify deficiencies in urban networks as they frequently traverse different parts of the city. Impact related applications relate to the ability of the Taxi GPS data to identify negative externalities such as air pollution caused by congestion. Thus, Taxi GPS data has wide variety of applications and their applicability will be determined by the context in how taxis are used, penetration levels of the taxi, and spatiotemporal detail offered in the taxi data set.

The answer to the second subquestion is based on the demand studies carried out in Section 3.2.

Sub Question 2: Which user groups are likely to use Uber in Amsterdam and for what purpose?

The user groups likely to use Ubers are those unfamiliar with the city and travellers to or from the airport. The demand studies have revealed higher data penetration in areas such as Bloemendaal, Zandvoort and Zaanstad. Bloemendaal, Zandvoort are locations dominated by tourists. Thus, the purpose behind Uber trips to these wijken is leisure. The wijk Haarlemmermeer also resulted in high data penetration due to the presence of Schiphol airport and the surrounding Wijken are dominated by hotels serving the airport. On fusion with land-use data for wijken within the centre of Amsterdam, it was revealed the trip purposes are difficult to derive due to the mixed land-use patterns and the spatial aggregation of the Uber data. An exception to this was Amsterdam Sloterdijk, which indicated high data penetration and is monofunctional concerning its land-use as it primarily comprises of offices.

Thereby, identifying another user group of people using Uber for work-related trips. Those unfamiliar with the city are likely to be foreigners. Validation from cellphone data suggests the concentration of international sim cards and data penetrations tend to overlap in wijken. For instance, in the case of Sloterdijk.

In addition to this, the model developed showed the wijken with the highest number of Uber vehicles across different hours of the day to concentrate in and around Schiphol along with the centre of the city and the office areas of Sloterdijk. The model also indicates higher demand for Ubers at night, possibly caused by limited public transport during late hours. Thus, four broad user groups can be identified: tourists (and those using Uber for trips to places of leisure), travellers to and from the airport, working professionals working in the business districts of Sloterdijk or Zuidas, and lastly those using Ubers as an alternative to public transport during the later hours of the day. It is important to acknowledge these user groups are not entirely distinct from each other and on the contrary overlap. For instance, tourists are also likely to use Uber to travel to the airport.

The answer to the third subquestion has been based on the studies with respect to recurrent congestion carried out in Section 3.3.

Sub Question 3: To what extent can recurrent congestion analysis be carried out using Uber Movement Travel Time data, either singularly or fused?

Ex-post recurrent congestion analysis using the Uber Movement Travel Time data set can be carried out by an abstraction of the network as a network graph and using normalised weighted Indegree which can help highlight congestion points in the city. A normalised weighted Indegree is the sum of the edges weighted by the travel time and normalised by distance and the number of edges. The normalisation prevents a higher indegree value from being a function of the Wijk being located at a further away distance. The weighted indegree can be dynamically analysed for different hours of the day, thereby, revealing congestion spots across the day. This can be done for one quarter of the year or multiple quarters. Europawijk followed by Heemstede West were the wijken with the highest indegree across 24 hours of the day for 2018. Both wijken are located in Haarlem. Other wijken which featured such as Burgwallen Nieuwe Zijde, Grachtengordel-West and Grachtengordel-South are located in the centre of the city, in the 17th-century ring. A limitation for the metric stems from the data penetration of the data set. Unavailability of a datapoint can suggest lack of congestion, either by not adding an edge (and thereby weight in the form of travel-time) or the absence of a node as a whole. Besides, the data sparsity for 2018 was found to be 60.14%. Thus, the data for most wijken are not potentially available across different time periods. However, congestion points frequented by Ubers can be identified. Thus, recurrent congestion analysis is best carried out through network graphs and weighted indegree, however, the data penetration limits the application of the insights concerning policy decisions. Additionally, the aggregated nature of the results (Hour of Day across a quarter) limits its applications at a tactical or operational level.

The answer to the fourth subquestion has been derived from the fusion of Uber Movemetn and WAZE data done in Section 3.4.

Sub Question 4: To what extent can non-recurrent congestion analysis be carried out using Uber Movement Travel Time data, either singularly or fused?

Ex-post non-recurrent congestion analysis using the Uber Movement Data set necessitates fusing it with an additional data set which offer incident-related information. This could be the Coördinatiestelsel Werken aan de weg (CORA) data-set or crowd-sourced data such as the WAZE dataset. Due to the low spatio-temporal resolution of the Uber Movemetn data, the incident data must be brought to the same resolution at the expense of lost detail. In the research, the WAZE data set was utilised due to its reliability indicators, enabling the researcher to identify if the event actually took place and also determine the spatial and temporal spread. Four types of events were investigated; these include Accidents, Road closures, jams and Weather Hazards. The Uber Movement data set was able to reflect travel time changes in all incidents except road-closures. Road closures tend to have a spatial and temporal spread and Traffic management measures are usually in place before road closures when compared to accidents, jams and weather hazards which tend to be localised temporally and spatially and cannot be predicted and prepared for. Post-spatiotemporally matching, the travel times to Wijken adjacent of the Wijk of interest (where the incident occurred) against the planning time (95th percentile travel time). It was found, planning times are an effective benchmark to compare travel time changes caused by incidents. The use of planning time is a significant improvement over comparing the travel times at a date and time period assumed to be representative of normal conditions.

Travel time changes though observed when fused with WAZE data, the data does not ascertain if the incident led to the travel time change. Limitations are mainly brought by the aggregation levels and the data sparsity. The temporal aggregation is an issue as the travel times are available for a time period of the day whereas incidents such as jams or accidents do not span the entire time period. Also, while the WAZE data offers coordinates on the incident location, aggregation to the Wijk level and absence of routes in the travel time data prevent one from understanding the actual cause of travel time change. For instance, the lane or direction of driving at which the accident took place and how the travel times varied on the routes cannot be determined. Thus, limiting the applicability of the Uber Movement data to identify measures to solve incidents of non-recurrent congestion. With respect to identification, the Uber Movement data can be used and also understand incidents of non-recurrent congestion experienced by Uber vehicles in regions frequented by them.

Answer to the fifth subquestion is based on the conclusions from Section 3.5 and the model results in Section 5.1.

Sub Question 5: What is the unique value addition of the Uber Movement Travel Time data set to Transport Planners and officials at the city of Amsterdam?

The unique value of the data set is latent in its ability to represent skewed user groups which in turn have a skewed spatial and temporal distribution compared to the average traveller. The data sets available at the municipality despite higher levels of detail, and applicability captures the average traveller. With the increased number of visitors to the city of Amsterdam, the Uber data offers unique insights into user groups who are less aware of the city and their spatial concentration. Although, the intention of the data is released for performance-related or impact-related analysis existing data sets with the municipality such as the TomTom data, ANPR and Floating car data from NDW are better suited. Demand studies and the model results strongly suggest the Uber user does not represent the average traveller performing work-related trips during peak hours, thereby adding to the existing data-ecosystem by the people who use Uber and the purpose they use it for. Additionally, results generated from the model can offer the spatio-temporal distribution of Ubers across different hours of the day. The impact of fleet sizes, and occupancy levels can additionally be analysed to implement traffic management solutions to restrict or enable the movement of Ubers. Thus, the data set offers a strategic overview on the usage of Uber in the city, where it is used, and the broad reasons behind using it (such as visiting the airport). The data set serves as a bearing point for further studies on taxi management and its data requirements.

Main Research Question: What is the potential of using Uber Movement Travel Time Data, either singularly or in fusion with other data sets for Traffic Congestion Analysis and management in the city of Amsterdam?

The Uber Movement data is available for multiple cities across the world and being exclusively derived from Uber trips, presents an opportunity to understand the travel times experienced by the vehicles and their skewed user base in the Amsterdam context. The research has revealed the possible user groups through data penetration. The results are qualitative in nature. However, based on the model built there is strong evidence to suggest the usage of Uber is primarily at Schiphol, the centre of the city and Amsterdam Sloterdijk. The potential of the data set for Amsterdam, lies in its ability offer the dynamics of spatio-temporal distribution at a macroscopic scale for Ubers in the greater Amsterdam region through results from the model. Additional insights, such as occupancy and empty vehicle kilometers offer added value, however, these are subject to future validation with empirical data of taxis in Amsterdam. Thus, the data set offers information on a user base not captured by other data sets at the municipality. With respect to other applications such as recurrent congestion,the Uber Movement travel time data set has limited potential for the city of Amsterdam. The spatial aggregation at the wijk level results in a significant loss of detail for applications to be developed. For example, in the case of non-recurrent congestion, a localised incident needs to be aggregated at the wijk level. Similarly, the temporal detail is either the time period of a day or aggregated across a quarter of the year. This leads to noise in the data and, consequently ascertaining the underlying mechanism which has led to the travel time in the data set becomes difficult. On evaluating the applicability of the data with respect to recurrent and non-recurrent congestion, it was found the data can offer congestion hot spots and reflect travel time changes, however, the existing data sets at the municipality with a finer temporal and spatial resolution are better suited for such analysis. Especially, as these lead to policy decisions and investments, the Uber Movement data is unreliable. The limited data penetration which stems from the limited use of Ubers in Amsterdam (also, the spatial concentration of where Ubers are used) and the absence of trajectory or route data further limits the potential of the data. Due to the absence of the number of vehicles which resulted in the data, traffic management applications are also limted.

6.1 IMPLICATIONS FOR THE URBAN DATA ECOSYSTEM

The Uber Movement Travel time data is available for multiple cities across the world. For transport planners and cities to derive usability from the data, a clear objective will be essential. For instance, an investigation needs to be done to check if the data is better for validation purposes or estimation of travel time. This will ensure the usability of the data. The data set does not necessarily enhance the data ecosystem of a city. For instance, it is possible usage patterns do not differ from the usual vehicle traffic. In such a situation, the data's temporal and spatial aggregation severly limits its usage and existing alternatives might already have better utility.

The data set reflects one of the larger issues with big data. The privacy concerns and aggregation levels can severely limit its usage. Aggregation due to privacy concerns and interests of the data provider is likely to be prevalent in instances when the data is provided for free. The research does offer a method to work at the aggregation level, however, the applications become limited. Thus, there needs to be a balance between addressing privacy concerns while also rendering the data as usable. As was shown in the research, the data needs to be translated into actionable insights. To achieve this, one needs to find the context, iron out the complexities and create analytical models which can turn data into information and eventually information into policy decisions. Thus, the introduction of Uber Movement data is context dependent, requires analytical models, and fusion with other data-sets.

For the city of Amsterdam, the research also indicates the need to standardise attributes released in open data by different stakeholders. The standardisation needs to ensure the data has usability while ensuring privacy concerns are addressed. The next chapter offers recommendations on the data attributes which could be made available to enable a comprehensive analysis.

7 RECOMMENDATIONS

The chapter comprises of the recommendations for improving the model and identifying the possible data attributes which can be obtained to enable a more comprehensive analysis of Ubers in the city from the perspective of congestion studies and demand studies. The chapter is concluded with recommendations on future research.

7.1 RECOMMENDATIONS REGARDING THE MODEL

The first recommendation relates to the use of the modelling parameters: p_max, p_min, e_drive, e_dest and e_occ as these need to be validated for the taxi movement in Amsterdam. Empirical data can be first collected, and these parameters can then be optimised to represent the cyclic activity of taxis across different hours of the day. This would better ensure realistic flows and origin-destination patterns. Alternatively, these paramters can be validated using Tom Tom data which offers traffic activity at the street level and can be aggregated to the wijk level. However, this step would be to risk loosing the taxi flow patterns which are a consequence of the data penetration. The model can be simulated with the 2019 data set (to be) offered in Uber Movement and is likely to be less sparse than the 2018 data. The data penetration levels have been consistently higher from 2016 to 2018 , thus offering insights on areas frequented and thereby improving the model results. The probability of occupancy in taxis has been crudely modelled. A better way would be to adopt clustering methods, which consider spatial and temporal correlation of zones with higher and lower demands. As a result, the taxi will be more likely to receive a passenger in a zone based on the demand for taxis at a wijk belonging to a cluster of high, medium or low demand. The data for clustering should ideally be based on taxi demand patterns from other empirical data. However, considering that this can be challenging, the data penetration levels in Uber Movement can be used. The initial value problem for the model has been solved by setting the final simulation result as the initial step. This can be improved once again using empirical data offering the proportion of taxis across zones. The determination of routes in the model is based on the application of the shortest path algorithm in network graphs. A broader understanding of the possible routes or the wijken traversed can be derived from determing N* , i.e. the 'n' possible number of routes and be validated against the routes preferred by vehicles from data sets such as TomTom data.

7.2 RECOMMENDATIONS REGARDING THE DATA

The recommendations act as a guideline for the municipality to establish the minimum data requirements to carry out congestion studies from Uber data and demand studies which can eventually enable access control of the taxis. This is explained across the following two subsections. The data requirements could vary according to the precise policy goals of the municipality. The following two subsections establish the basic requirements derived from the unavailable attributes of the Uber Movement travel time data set. The recommendations are based on the possibility of the data being historical and not real-time.

7.2.1 Congestion studies

The first data attribute required would be the route undertaken by the taxis. This will be essential for establishing recurrent and non-recurrent congestion. Additionally, one can understand if the travel time for taxis are different from average vehicle traffic as taxis can make use of bus-lanes in the city of Amsterdam. Thus, the data could suggest free-flow travel times even during congested periods. If it is not possible to obtain the route, the most popular route between two spatial units can be obtained. The most popular route across different hours of the day is an improvement over the complete absence of the route. The second data attribute would be the number of Uber vehicles between two spatial units. This can be obtained at the per hour of the day temporal detail. The different levels of temporal detail and the attribute levels have been tabulated in Table 7.1.

Table 7.1: Recommended detail for data attributes to perform congestion studies

Route	**Route/vehicle**	**Route/'n' vehicles**			
	Route taken by each vehicle	Per hour/day	Per hour/weekday & Per hour/weekend	Per hour/month	Per hour/quarter of the year
Travel Time (seconds)	Per hour/day	Per hour/weekday & Per hour/weekend	Per hour/month	Per hour/quarter of the year	
Number of Ubers	Per hour/day	Per hour/weekday & Per hour/weekend	Per hour/month	Per hour/quarter of the year	

The spatial attribute can vary depending on the required level of detail. This can vary between a coordinate point to a polygon at a postal code level or Traffic analysis zones. The precise pick-up and drop-off location is not relevant to congestion studies rather the data do not need to establish if a pick-up or drop-off happened as is the case in the Uber Movement data. The attributes mention in the table enable the calculation of average speed (Distance travelled on the route / time), rate of flow (vehicles/temporal unit) and density (vehicles/spatial unit). The fundamental quantities for traffic flow studies expand the usability of the data and enable congestion studies. Data fusion can offer additional value. For instance, the data can be fused with Waze data to understand the impact of incidents on travel time differences or emissions caused by taxis. Studying emissions caused by taxis can indeed be a useful barometer to decide the position of environmental zones in the city and if it can be dynamic across the day. It is important to reiterate that the data requirements will vary according to the precise policy goals and the level (Strategic, tactical or operational) at which the decisions need to be made. For instance, one can also obtain the fleet specifications such as the type of vehicles, vehicle age as is done by the New York Taxi & Limousine commission [NYC-Open-Data, 2018]. The data attributes recommended address the crucial attributes missing in the data.

7.2.2 Demand studies and access control

The demand studies to derive the spatio-temporal distribution of taxis would require the number of Ubers travelling between two spatial units across different hours of the day. [Huang, 2003] mentions space and time as scarce resources which constrain the movement of people. Thus, the two most important data attribute is the number of Ubers travelling across different temporal units and different traffic analysis zones. Interpreting the demand for Ubers also requires one to take aspects such as the weather, dynamic pricing employed by ride-sharing services, waiting time for passengers and events taking place in zones. These can serve as attributes while estimating mode preference in a transport model through a discrete choice model. Additional attributes can also include, the number of passengers transported. Since, Uber does not have a carpooling taxi segment yet, this can be avoided. The spatial detail for the data should be such that the Land-Use entropy value is closer to zero unlike what was found in Section 3.2.2. A land-use entropy value closer to zero is suggestive of a dominant land-use function, thereby adding to the understanding of the behaviour and purpose of Uber users. Table 7.2 tabulates the recommended attributes and the level of detail.

Table 7.2: Recommended detail for data attributes to perform demand studies

	Increasing aggregation (left to right)		
Pick-up & Drop-off location	Coordinates Longitude & Latitude	Postal code level 4	Traffic analysis zone consistent with transport demand model
Temporal unit	Seconds	15 minutes	1 hour
Trajectory information		Polyline between pick-up & drop-off	

Concerning access control, a model can be developed to predict the number of Ubers across different hours of the day across zones. The input data for this model should consist of the number of Ubers across different hours of the day and different zones collected longitudinally over a time period. A model can predict the demand for Ubers as a function of the land-use, weather (rainy weather can prompt increased usage), large-scale events, time of the day, weekday or weekend, month of the year (to capture the tourist season) and public-transport availability (based on the General Transit Feed specification data). The list is not exhaustive. Regression, and machine learning techniques can be employed for predicting the demand, based on the available data attributes.

7.3 RECOMMENDATIONS FOR FURTHER RESEARCH

The recommendations for further research include, the need to collect Taxi data from other operators (possibly registered at TTO). This will serve two purposes, one it can help understand if the users of the 65% taxi market (Uber comprises of 35%) behave differently when compared to Uber users. Additionally, it can be used for validation of the model results. The data can also be used to compare Uber movement travel times with travel times from other taxis. This can help establish if the usage of bus and taxi lanes results in travel time differences compared to an average vehicle user. The other recommendation relates to the observed peak in data penetration points at midnight. It would be relevant to understand if the data penetration is indeed caused by a lack of public transport during these hours of the day. The model implemented can also be extended to determine charging infrastructure. Extensions such as the charge carried by vehicles can integrated into the model. The travel times in the Uber Movement data can also be explored for accessibility studies and its potential in highlighting accessibility issues. To enable a disaggregated analysis, a web scraper can be built to scrape travel time data across time periods for several dates in a year from the Uber Movement website. Another recommendation is to estimate separate models using the weekday and weekend travel time data and explore if the model results vary from each other as the travel times and data penetration could be different.

BIBLIOGRAPHY

Administration, U. S. F. H., Systematics, C., and Institute, T. T. (2004). *Traffic Congestion and Reliability: Linking Solutions to Problems*. Cambridge Systematics, Incorporated.

Amin-Naseri, M., Chakraborty, P., Sharma, A., Gilbert, S. B., and Hong, M. (2018). Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from waze. *Transportation Research Record*, 2672(43):34–43.

Amini, S., Papapanagiotou, E., and Busch, F. (2016). *Traffic Management for Major Events*, pages 187–197.

Anbaroglu, B. (2013). *Spatio-temporal Clustering for Non-Recurrent Traffic Congestion Detection on Urban Road Networks*. PhD thesis.

Antoniou, G., Batsakis, S., Davies, J., Duke, A., McCluskey, T. L., Peytchev, E., Tachmazidis, I., and Vallati, M. (2019). Enabling the use of a planning agent for urban traffic management via enriched and integrated urban data. *Transportation Research Part C: Emerging Technologies*.

Aryandoust, A., Vliet, O., and Patt, A. (2019). City-scale car traffic and parking density maps from uber movement travel time data. *Scientific Data*, 6.

Bischoff, J., Maciejewski, M., and Sohr, A. (2015). Analysis of berlin's taxi services by exploring gps traces. pages 209–215.

Bock, F., Martino, S. D., and Sester, M. (2017). Data-driven approaches for smart parking. In *ECML/PKDD*.

Booking.com (2019). Available hotels across the greater Amsterdam region in Booking.com. www.booking.com.

Bouw en verkeersprojecten, G. A. (2018). loterweg en omstreken: diverse werkzaamheden.

Brodeur, A. and Nield, K. (2018). An empirical analysis of taxi, lyft and uber rides: Evidence from weather shocks in nyc. *Journal of Economic Behavior & Organization*, 152:1–16.

Cambridge Systematics, I. and Texas Transportation, I. (2005). Traffic Congestion and Reliability. Technical report, Federal Highway Administration.

Casadei, G., Bertrand, V., Gouin, B., and Canudas-de Wit, C. (2018). Aggregation and travel time calculation over large scale traffic networks: An empiric study on the Grenoble City. *Transportation Research Part C: Emerging Technologies*.

Castro, P. S., Zhang, D., and Li, S. (2012). Urban traffic modelling and prediction using large scale taxi gps traces. In *Pervasive*.

Csardi, G. and Nepusz, T. (2005). igraph reference manual.

Cui, J., Liu, F., Janssens, D., An, S., Wets, G., and Cools, M. (2016). Detecting urban road network accessibility problems using taxi gps data. *Journal of Transport Geography*, 51:147–157.

Deloitte (2018). The rise of the platform economy.

Dowling, R., Skabardonis, A., Carroll, M., and Wang, Z. (2004a). Methodology for measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record*, 1867:60–68.

Dowling, R., Skabardonis, A., Carroll, M., and Wang, Z. (2004b). Methodology for measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record*, 1867:60–68.

Faizrahnemoon, M., Schlote, A., Maggi, L., Crisostomi, E., and Shorten, R. (2015). A big-data model for multi-modal public transportation with application to macroscopic control and optimisation. *International Journal of Control*, 88(11):2354–2368.

Gemeente Amsterdam, C. (2019). Sectoranalyse Mobiliteit: Amsterdam.

Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826.

Gkania, V. and Dimitriou, L. (2018). A Back-Engineering Approach to Explore Human Mobility Patterns Across Megacities Using Online Traffic Maps. In *Mobility Patterns, Big Data and Transport Analytics*.

Grauwin, S., Sobolevsky, S., Moritz, S., Gódor, I., and Ratti, C. (2015). *Towards a Comparative Science of Cities: Using Mobile Traffic Records in New York, London, and Hong Kong*, pages 363–387. Springer International Publishing, Cham.

Guoqiang Yu, Jianming Hu, Changshui Zhang, Like Zhuang, and Jingyan Song (2003). Short-term traffic flow forecasting based on markov chain model. In *IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No.03TH8683)*, pages 208–212.

Huang, Z. (2003). Data integration for urban transport planning.

Jordahl, K., van den Bosche, J., et al. (2013–). Geopandas: Open source scientific tools for gis analysis Python. [Online; accessed: 30th June 2019].

Keler, A., Krisp, J. M., and Ding, L. (2017). Geo-spatial Information Science Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points. 20(4):333–344.

King, D. and Saldarriaga, J. F. (2018). Spatial regulation of taxicab services: Measuring empty travel in new york city. *Journal of Transport and Land Use*, 11.

Krishnakumari, P., Lint, J., Djukic, T., and Cats, O. (2019). A data driven method for od matrix estimation. *Transportation Research Part C Emerging Technologies*.

Kuang, W., An, S., and Jiang, H. (2015). Detecting traffic anomalies in urban areas using taxi gps data. *Mathematical Problems in Engineering*, 2015:1–13.

Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group Res Note 6*, 6.

Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., and Yang, Q. (2011). Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 63–68.

Liu, L., Andris, C., and Ratti, C. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems*, 34(6):541 – 548. GeoVisualization and the Digital City.

Liu, M., Brynjolfsson, E., and Dowlatabadi, J. (2018a). Technology , incentives , and service quality : the case of taxis and uber.

Liu, X., Gong, L., Gong, Y., and Liu, Y. (2013). Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 43.

Liu, Z., Jia, Y., and Zhu, X. (2018b). Deployment strategy for car-sharing depots by clustering urban traffic big data based on affinity propagation. *Scientific Programming*, 2018:3907513:1–3907513:9.

Lu, X., Ota, K., Dong, M., Yu, C., and Jin, H. (2017). Predicting transportation carbon emission with urban big data. *IEEE Transactions on Sustainable Computing*, 2(04):333–344.

Movement, U. (2019a). Uber Movement.

Movement, U. (2019b). Uber Movement Travel Time Calculation Methodology Report. Technical report.

Niforatos, E., Vourvopoulos, A., and Langheinrich, M. (2015). Weather with you: Evaluating report reliability in weather crowdsourcing. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, MUM '15, pages 152–162, New York, NY, USA. ACM.

NYC-Open-Data (2018). NYC Taxi & Limousine commission. https://www1.nyc.gov/site/tlc/about/fact-book.page.

OECD, Organisation (ECMT, E. (2007). Managing urban traffic congestion. *Managing Urban Traffic Congestion*, pages 1–296.

Pearson, M. J. and Samaniego, S. (2017). Traffic flow analysis using uber movement data.

Pelzer, P., Frenken, K., and Boon, W. (2019). Institutional entrepreneurship in the platform economy: How uber tried (and failed) to change the dutch taxi law. *Environmental Innovation and Societal Transitions*.

Populus.ai (2019). Populus Resources.

Redelosa, C. K. and Lim, M. (2018). Centrality measurement of traffic analysis zones using uber movement data. In *Proceedings of the Samahang Pisika ng Pilipinas*, volume 36, pages SPP–2018–3D–05.

Skabardonis, A., Varaiya, P., and Petty, K. (2003). Measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record*, 1856.

Strava-Metro (2019). Strava Metro: Creating better cities by using big data. https://metro.strava.com/.

Uber Movement. Examining the Impact of the London Tower Bridge Closure.

Uber Movement. Medium Case Studies - Uber Movement.

Uber Nederland, P. (2019). Waar is Uber actief in Nederland.

Uzel, A. (2018). Putting "Uber Movement" Data into Action — Machine Learning Approach.

Wu, H. (2018). Comparing Google Maps and Uber Movement Travel Time Data. *Transport Findings*.

Yang, Z., Franz, M. L., Zhu, S., Mahmoudi, J., Nasri, A., and Zhang, L. (2018). Analysis of washington, dc taxi demand using gps and land-use data. *Journal of Transport Geography*, 66:35 – 44.

Zhan, X., Hasan, S., Ukkusuri, S., and Kamga, C. (2013). Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies*, 33:37–49.

Zhu, D., Wang, N., Wu, L., and Liu, Y. (2017a). Street as a big geo-data assembly and analysis unit in urban studies: A case study using beijing taxi data.

Zhu, Y., Xie, K., Ozbay, K., Zuo, F., and Yang, H. (2017b). Data-driven spatial modeling for quantifying networkwide resilience in the aftermath of hurricanes irene and sandy. *Transportation Research Record: Journal of the Transportation Research Board*, 2604:9–18.

A NON-RECURRENT CONGESTION

Table A.1: WAZE alert data attributes

Attribute	Example	Description
reportBy	Wegestatus.nl	ID of WAZE user reporting the incident
country	NL	Country
nThumbsUp	4	Number of Thumbs Up to confirm the alert
reportRating	3	Rating of the report, determined by number of Thumbs Up
confidence	1	Confidence determined by the number of Thumbs up. Relates to the reliability of the incident.
nReliability	6	Can take a value between 0 to 10. Usually between 5 to 10. Relates to the number of users who confirm the presence of the incident
nImages	0	Number of Images posted with respect to the incident by the reporter
type	ROAD_CLOSED	Type of incident reported by the user
speed	0	Speed in km/h
reportMood	1	Mood of the reporter displayed through an icon on WAZE
showFacebookPic	No	No if no Facebook images with respect to the event are found
subtype	ROAD_CLOSED_EVENT	Further describing type of incident. For example, in the case of a jam it can specify if the jams are light, moderate or heavy.
location	{"x":4.872711,"y":52.379103}	Location of the incident
nComments	1	Number of comments on the incident by WAZE users
City	Amsterdam	Specifies the city in which the incident took place
Street	Postjesweg	Name of the street at which the incident has occurred
roadType	2	WAZE assigns roads from different levels a number. For instance, S-roads get a value of 7. A higher value indicating a road higher in the network hierarchy
reportDescription	Ontscheping cruiseschip	Description of the incident
imagesUrl	https:\s3.amazonaws.com\waze.photos \ac8acfa4-3892-4026-bae5-aabcf390740d	URL to displaying the image of the incident
geometry	POINT (4.848504 52.398977)	Coordiantes of the incident

Table A.2: Accident incidents reported in the WAZE data with the highest reliability

MOVEMENT_ID	WK_NAAM	nThumbsUp	reportRati	confidence	reliabilit	type	subtype	street	roadtype	Date	Time
108	Waterland	3	0	2	10	ACCIDENT	None	N247-Slochterweg	6	2018-07-06	05:45:00
108	Waterland	4	0	2	10	ACCIDENT	None	N247-Slochterweg	6	2018-07-06	06:00:00
73	Museumkwartier	4	1	1	10	ACCIDENT	ACCIDENT_MINOR	S108 - Hobbemakade	7	2018-07-26	17:15:00
145	Bijlmer Oost	8	4	3	10	ACCIDENT	ACCIDENT_MAJOR	S112 - Gooiseweg	7	2018-07-12	17:15:00
147	Nellestein	8	5	4	10	ACCIDENT	ACCIDENT_MINOR	A9	3	2018-07-03	06:15:00
145	Bijlmer Oost	10	4	4	10	ACCIDENT	ACCIDENT_MAJOR	S112 - Gooiseweg	7	2018-07-12	17:45:00

Table A.3: Travel Time comparisons and differences - Waterland N247 - Slochterweg

dstid	Travel Time (TT)	Planning Time (PT)	90th	85th	80th	mean	TT-PT	TT-90th	TT-85th	TT-80th	TT-mean
37	587	363.1475	350.83	344.385	329.22	273.0021	223.8525	236.17	242.615	257.78	313.9979
56	334	426.052	388.062	368.04	356.354	332.4353	-92.052	-54.062	-34.04	-22.354	1.5647
78	707	751.507	705.672	655.249	586.838	516.2371	-44.507	1.328	51.751	120.162	190.7629
98	175	213.867	201.194	188.942	177.676	156.9537	-38.867	-26.194	-13.942	-2.676	18.0463
109	493	288.96	270.82	252.17	246.46	208.0389	204.04	222.18	240.83	246.54	284.9611
129	499	588.09	555.04	540.63	534	483.8161	-89.09	-56.04	-41.63	-35	15.1839
131	542	614.98	599.628	578.784	566.104	532.0027	-72.98	-57.628	-36.784	-24.104	9.9973
133	744	610.095	583.775	575.1425	556.67	496.9717	133.905	160.225	168.8575	187.33	247.0283
144	641	565.6835	552.643	544.9665	536.134	502.5331	75.3165	88.357	96.0335	104.866	138.4669
159	422	352.102	339.478	337.261	335.436	325.1256	69.898	82.522	84.739	86.564	96.8744
161	169	155.097	143.981	142.423	139.244	127.2601	13.903	25.019	26.577	29.756	41.7399

Table A.4: Travel times from Waterland (N247-Slochterweg) to adjacent Wijks – 6th July 2018 – early morning

sourceid	Origin wijk	dstid	Destination wijk	Travel Time	Planning Time	Arithmetic mean
108	Waterland					
		37	Landsmeer	587	353.732	249.9909
		92	Nieuwendammerdijk and Buiksloterdijk	-	-	-
		99	Waterlandpleinbuurt	116	201.2655	144.9526
		101	Buikslotermeer	230	387	286.8167
		107	Noordelijke IJ-oevers Oost	-	-	-
		109	Elzenhagen	493	286.459	216.9234
		157	Broek in Waterland	249	281.207	254.1446
		161	Watergang	169	143.108	126.7611
37	Landsmeer					
		92	Nieuwendammerdijk and Buiksloterdijk	-	774.807	542.1467
		99	Waterlandpleinbuurt	-	725.204	570.306
		101	Buikslotermeer	-	592.7025	458.194
		107	Noordelijke IJ-oevers Oost	-	-	-
		108	Waterland	-	725.047	545.3471
		109	Elzenhagen	-	566.143	414.6921
		157	Broek in Waterland	-	-	-
		161	Watergang	-	-	-
92	Nieuwendammerdijk and Buiksloterdijk					
		37	Landsmeer	499	610.596	439.3986
		99	Waterlandpleinbuurt	289	352.181	257.7725
		101	Buikslotermeer	82	153.8	95.25321
		107	Noordelijke IJ-oevers Oost	-	258.066	176.9816
		108	Waterland	215	331.178	234.7031
		109	Elzenhagen	131	219.406	163.448
		157	Broek in Waterland	-	463.4605	410.0777
		161	Watergang	389	360.015	291.5844
99	Waterlandpleinbuurt					
		37	Landsmeer	-	915.5	655.7633
		92	Nieuwendammerdijk and Buiksloterdijk	-	378.013	286.0963
		101	Buikslotermeer	115	226.764	177.2632
		107	Noordelijke IJ-oevers Oost	-	422.479	330.8321
		108	Waterland	-	134.5	101.8531
		109	Elzenhagen	-	421.997	304.7334
		157	Broek in Waterland	-		
		161	Watergang	-	486.8945	421.1425
101	Buikslotermeer					
		37	Landsmeer	423	513.13	379.6808
		92	Nieuwendammerdijk and Buiksloterdijk	141	214.994	159.022
		99	Waterlandpleinbuurt	133	181.0335	135.843
		107	Noordelijke IJ-oevers Oost	-	349.414	250.0667
		108	Waterland	131	212.457	159.0808
		109	Elzenhagen	120	197.468	134.3065
		157	Broek in Waterland	-	402.1965	350.6381
		161	Watergang	268	279.05	234.1582
107	Noordelijke IJ-oevers Oost					
		37	Landsmeer	-	741.3435	639.18
		92	Nieuwendammerdijk and Buiksloterdijk	186	243.0425	181.8837
		99	Waterlandpleinbuurt	-	506.425	402.6742
		101	Buikslotermeer	147	339.3875	274.0195
		108	Waterland	310	540.175	436.5263
		109	Elzenhagen	-	607.974	465.326
		157	Broek in Waterland	-		
		161	Watergang	-	836.67	836.67
109	Elzenhagen					
		37	Landsmeer	-	550.3905	430.396
		92	Nieuwendammerdijk and Buiksloterdijk	101	140.245	96.17325
		99	Waterlandpleinbuurt	380	401.745	262.2878
		101	Buikslotermeer	209	265.881	154.6627
		107	Noordelijke IJ-oevers Oost	-	439.192	308.4012
		108	Waterland	597	433.5875	308.385
		157	Broek in Waterland	-	560.2165	398.296
		161	Watergang	-	399.5525	263.55
157	Broek in Waterland					
		37	Landsmeer	-	-	-
		92	Nieuwendammerdijk and Buiksloterdijk	-	-	-
		99	Waterlandpleinbuurt	-	-	-
		101	Buikslotermeer	-	-	-
		107	Noordelijke IJ-oevers Oost	-	-	-
		108	Waterland	-	387.665	296.3783
		109	Elzenhagen	-	445.982	403.238
		161	Watergang	-	266.63	164.6586
161	Watergang					
		37	Landsmeer	-		
		92	Nieuwendammerdijk and Buiksloterdijk	-	502.46	391.1321
		99	Waterlandpleinbuurt	-	714.9575	584.0867
		101	Buikslotermeer	-	632.044	469.77
		107	Noordelijke IJ-oevers Oost	-		
		108	Waterland	-	259.05	169.2678
		109	Elzenhagen	-	460.8165	302.0185
		157	Broek in Waterland	-	177.0525	145.8629

Table A.5: Road closed events with the highest reliability reported in WAZE

MOVEMENT_ID	WK_NAAM	nThumbsUp	reportRati	confidence	reliabilit	type	subtype	street	roadtype	Date	Time
27	Haarlemerbuurt	4	0	2	10	ROAD_CLOSED	ROAD_CLOSED_EVENT	Singel	NaN	2018-07-17	10:00:00
23	Burgwallen Nieuwe Zijdge	4	0	2	10	ROAD_CLOSED	ROAD_CLOSED_EVENT	Singel	NaN	2018-07-18	20:15:00
27	Haarlembuurt	4	0	2	10	ROAD_CLOSED	ROAD_CLOSED_EVENT	Singel	NaN	2018-07-19	16:15:00
28	Jordaan	5	0	2	10	ROAD_CLOSED	ROAD_CLOSED_EVENT	Lijnbaansgracht	NaN	2018-07-17	09:45:00
28	Jordaan	5	0	2	10	ROAD_CLOSED	ROAD_CLOSED_EVENT	Lijnbaansgracht	NaN	2018-07-26	11:00:00
73	Museumkwartier	9	0	4	10	ROAD_CLOSED	ROAD_CLOSED_EVENT	Reijnier Winkeleskaade	NaN	2018-07-03	12:45:00
73	Museumkwartier	9	0	4	10	ROAD_CLOSED	ROAD_CLOSED_EVENT	Reijnier Winkeleskaade	NaN	2018-07-26	11:00:00

Table A.6: Jam incidents reported in the WAZE data with the highest reliability

MOVEMENT ID	WK_NAAM	nThumbsUp	reportRati	confidence	reliabilit	type	subtype	Street	roadtype	Date	Time
108	Waterland	1	2	3	10	JAM	JAM_HEAVY_TRAFFIC	N247-Slochterweg	6	2017-07-06	05:15:00
148	Holendrecht	1	0	2	10	JAM	JAM_MODERATE_TRAFFIC	A9	3	2018-07-13	14:15:00
87	IJplein/Vogelbuurt	2	3	2	10	JAM	JAM_HEAVY_TRAFFIC	S116 Niuewe Leeuwarderweg	7	2018-07-07	14:15:00
38	Centrale Markt	3	1	3	10	JAM	JAM_STAND_STILL_TRAFFIC	S105 – Jan van Galenstraat	7	2018-07-06	13:45:00
38	Centrale Markt	3	1	3	10	JAM	JAM_STAND_STILL_TRAFFIC	S105 – Jan van Galenstraat	7	2018-07-06	13:15:00

Table A.7: Travel Time differences to adjacent Wijks – Holendrecht (Jam at A9)

dstid	Travel Time	Planning Time	90th	85th	80th	mean	TT-PT	TT-90th	TT-85th	TT-80th	TT-mean
21	722	783.1375	756.86	728.73	704.66	636.268	-61.1375	-34.86	-6.73	17.34	85.732
145	345	370.5375	353.841	334.6975	318.906	290.1399	-25.5375	-8.841	10.3025	26.094	54.8601
147	476	173.789	162.379	158.2375	152.67	137.0789	302.211	313.621	317.7625	323.33	338.9211

Table A.8: Travel time comparisons and differences to other Wijk(s) - Holendrecht (Jam at A9)

sourceid	Origin wijk	dstid	Destination wijk	Travel Time	Planning Time	Arithmethic mean
148	**Holendrecht**					
		141	Amstel III/Bullewijk	246	341.05	236.1997
		143	Bijlmer Centrum (D,F,H)	162	288.869	217.4392
		145	Bijlmer Oost (E,G,K)	345	370.5375	290.1399
		147	Nellestein	476	173.789	137.0789
		153	Abcoude	-	615.23	504.6083
141	**Amstel III/Bullewijk**					
		143	Bijlmer Centrum (D,F,H)	197	291.0275	250.8349
		145	Bijlmer Oost (E,G,K)	461	522.4035	468.7774
		147	Nellestein	817	436.13	337.9811
		148	Holendrecht	182	279.427	233.7232
		153	Abcoude	-	582.33	464.2554
143	**Bijlmer Centrum (D,F,H)**					
		141	Amstel III/Bullewijk	313	287.967	254.2319
		145	Bijlmer Oost (E,G,K)	217	230.7915	197.5042
		147	Nellestein	-	350.0975	276.724
		148	Holendrecht	176	263.0305	219.1083
		153	Abcoude	-	976.8105	723.6816
145	**Bijlmer Oost (E,G,K)**					
		141	Amstel III/Bullewijk	703	509.095	420.2808
		143	Bijlmer Centrum (D,F,H)	224	210.1985	175.2379
		147	Nellestein	499	312.5655	246.0868
		148	Holendrecht	277	366.076	286.6994
		153	Abcoude	-	903.824	719.0828
147	**Nellestein**					
		141	Amstel III/Bullewijk	599	512.229	335.7667
		143	Bijlmer Centrum (D,F,H)	-	332.005	272.6667
		145	Bijlmer Oost (E,G,K)	278	288.33	245.2019
		148	Holendrecht	155	171.125	129.8899
		153	Abcoude	-	695.587	521.1453
153	**Abcoude**					
		141	Amstel III/Bullewijk	317	382.7765	281.7649
		143	Bijlmer Centrum (D,F,H)	-	851.205	698.2036
		145	Bijlmer Oost (E,G,K)	-	996.556	789.0797
		147	Nellestein	-	747.2445	540.1054
		148	Holendrecht	-	631.6015	489.2236

Table A.9: Weather Hazard incidents reported in the WAZE data with the highest reliability

MOVEMENT ID	WK_NAAM	nThumbsUp	reportRati	confidence	reliabilit	type	subtype	Street	roadtype	Date	Time
132	Sloter/Riekpolder	29	2	5	10	WEATHER_HAZARD	HAZARD_ ON_ROAD_ CONSTRUCTION	Anderlechtlaan	2	2018/07/04	15:15:00
82	Middenmeer	8	4	4	10	WEATHER_HAZARD	HAZARD_ON_ SHOULDER_ CAR_STOPPED	A10 Parallel	3	2018/07/02	10:00:00
82	Middenmeer	6	3	3	10	WEATHER_HAZARD	HAZARD_ON_ SHOULDER_CAR _STOPPED	A10 Parallel	3	2018/07/31	06:15:00
81	Frankendael	40	2	5	10	WEATHER_HAZARD	HAZARD_ON_ROAD _CONSTRUCTION	S112 – Prins Bernhardplein	7	2018/07/07	21:00:00
81	Frankendael	40	2	5	10	WEATHER_HAZARD	HAZARD_ON_ROAD _CONSTRUCTION	S112 – Prins Berhardplein	7	2018/07/08	07:45:00

Table A.10: Travel Time comparisons and differences to other Wijks - Frankendael

dstid	Travel Time	Planning Time	90th	85th	80th	mean	TT-PT	TT-90th	TT-85th	TT-80th	TT-mean
23	824	820.526	807.684	785.6175	769.102	717.762	3.474	16.316	38.3825	54.898	106.238
27	968	1019.142	970.536	955.3565	948.418	896.0177	-51.142	-2.536	12.6435	19.582	71.9823
28	1067	1124.051	1096.402	1033.603	1026.024	980.4355	-57.051	-29.402	33.397	40.976	86.5645
31	628	661.0265	644.094	633.794	626.72	576.9902	-33.0265	-16.094	-5.794	1.28	51.0098
40	1137	1203.602	1173.449	1143.609	1134.492	1056.201	-66.602	-36.449	-6.609	2.508	80.799
59	611	636.9335	627.473	599.1255	566.99	524.454	-25.9335	-16.473	11.8745	44.01	86.546
61	270	272.3915	264.666	257.3125	254.19	241.5988	-2.3915	5.334	12.6875	15.81	28.4012
82	214	199.7685	192.61	186.184	180.7	174.803	14.2315	21.39	27.816	33.3	39.197
83	145	145.113	142.218	140.616	138.734	131.4985	-0.113	2.782	4.384	6.266	13.5015
84	252	237.4905	236.74	230.61	227.204	214.5082	14.5095	15.26	21.39	24.796	37.4918
86	900	1017.805	930.85	903.84	879.4	814.4927	-117.805	-30.85	-3.84	20.6	85.5073
87	818	824.036	808.352	769.8145	760.486	704.6088	-6.036	9.648	48.1855	57.514	113.3912
127	787	651.609	637.608	618.8065	605.168	549.5992	135.391	149.392	168.1935	181.832	237.4008
128	839	926.416	886.585	871.2305	865.61	814.6123	-87.416	-47.585	-32.2305	-26.61	24.3877
143	373	365.26	358.697	355.7365	351.31	340.8795	7.74	14.303	17.2635	21.69	32.1205
145	456	531.937	438.358	428.456	418.462	400.9992	-75.937	17.642	27.544	37.538	55.0008
164	281	260.024	256.806	253.498	250.628	239.2495	20.976	24.194	27.502	30.372	41.7505
165	414	429.3525	413.433	404.1335	399.106	371.5313	-15.3525	0.567	9.8665	14.894	42.4687

Table A.11: Frankendael (Evening) 7th July 2018 – Travel times to and from adjacent Wijks

sourceid	Origin wijk	dstid	Destination wijk	Travel Time	Percentile Travel Time	Arithmetic mean
81	**Frankendael**					
		80	Rijnbuurt	320	335.6555	308.8262
		82	Middenmeer	214	199.7685	174.803
		83	Betondorp	145	145.113	131.4985
		84	Omval/ Overamstel	252	237.4905	214.5082
		51	Weesperzijde	192	229.9425	196.1682
		54	Transvaalbuurt	101	137.035	117.005
80	**Rijnbuurt**					
		81	Frankendael	211	265.3915	238.6665
		82	Middenmeer	263	318.1435	284.4658
		83	Betondorp	199	232.264	205.9818
		84	Omval/ Overamstel	110	113.843	94.29583
		51	Weesperzijde	369	438.62	400.0633
		54	Transvaalbuurt	297	375.2595	345.5617
82	**Middenmeer**					
		80	Rijnbuurt	343	413.832	354.1228
		81	Frankendael	174	218.9605	183.924
		83	Betondorp	141	170.3725	141.8133
		84	Omval/ Overamstel	233	278.933	242.8882
		51	Weesperzijde	463	604.861	520.8905
		54	Transvaalbuurt	210	298.3955	246.1813
83	**Betondorp**					
		80	Rijnbuurt	225	262.1465	229.4443
		81	Frankendael	131	167.139	138.6417
		82	Middenmeer	115	128.987	113.5968
		84	Omval/ Overamstel	139	158.462	138.0305
		51	Weesperzijde	268	331.4335	265.8472
		54	Transvaalbuurt	180	242.7105	207.0842
84	**Omval/ Overamstel**					
		80	Rijnbuurt	387	137.2905	116.2393
		81	Frankendael	312	275.493	246.814
		82	Middenmeer	209	225.435	203.0453
		83	Betondorp	127	155.38	135.5522
		51	Weesperzijde	387	438.2725	393.2345
		54	Transvaalbuurt	312	365.127	333.1345
51	**Weesperzijde**					
		80	Rijnbuurt	475	428.8695	395.5097
		81	Frankendael	175	186.5465	170.7328
		82	Middenmeer	415	473.5605	421.66
		83	Betondorp	254	276.186	254.6463
		84	Omval/ Overamstel	390	382.5775	352.6058
		54	Transvaalbuurt	215	232.2185	203.3768
54	**Transvaalbuurt**					
		80	Rijnbuurt	601	549.4845	492.2402
		81	Frankendael	158	147.5775	133.2088
		82	Middenmeer	160	180.826	159.456
		83	Betondorp	313	336.8095	299.655
		84	Omval/ Overamstel	533	456.5925	415.698
		51	Weesperzijde	107	138.0905	110.3258

Table A.12: Travel time comparisons and differences to other Wijken - Stadionbuurt - AM peak - 15/10/2017

dstid	Travel Time	Planning Time	90th	85th	80th	Mean Travel Time	TT-PT	TT-90th	TT-85th	TT-80th	TT-Mean
11	977	577.2825	549.42	531.7575	521.61	505.4886111	399.7175	427.58	445.2425	455.39	471.5113889
16	1399	948.8175	906.105	876.135	824.54	778.0719444	450.1825	492.895	522.865	574.46	620.9280556
41	715	815.199	768.445	745.988	731.4	676.1141176	-100.199	-53.445	-30.988	-16.4	38.8858824
43	666	729.446	695.278	679.192	664.582	608.3268571	-63.446	-29.278	-13.192	1.418	57.6731429
44	640	688.69	653.875	628.495	607.7	559.7380556	-48.69	-13.875	11.505	32.3	80.2619444
45	844	384.7625	381.24	362.225	355.14	318.6319444	459.2375	462.76	481.775	488.86	525.3680556
46	674	624.6025	597.195	561.1125	538.59	504.4433333	49.3975	76.805	112.8875	135.41	169.5566667
47	505	212.17	201.965	197.3025	195.21	184.1916667	292.83	303.035	307.6975	309.79	320.8083333
61	696	707.0585	687.047	679.9085	664.94	578.8559375	-11.0585	8.953	16.0915	31.06	117.1440625
69	925	519.94	498.09	473.9	462.08	408.295	405.06	426.91	451.1	462.92	516.705
73	426	385.8175	367.195	355.4275	350.3	324.5208333	40.1825	58.805	70.5725	75.7	101.4791667
75	220	201.2625	189.45	176.855	176.73	163.6547222	18.7375	30.55	43.145	43.27	56.3452778
78	667	514.9325	506.04	501.1375	497.06	467.51	152.0675	160.96	165.8625	169.94	199.49
80	683	592.877	575.188	540.95	519.73	493.3088571	90.123	107.812	142.05	163.27	189.6911429
84	649	669.578	643.535	610.063	592.078	521.8428125	-20.578	5.465	38.937	56.922	127.1571875
85	272	196.5625	186.16	175.365	173.48	152.0888889	75.4375	85.84	96.635	98.52	119.9111111
128	759	646.9375	633.845	613.5175	565.1	502.2891667	112.0625	125.155	145.4825	193.9	256.7108333
130	633	354.18	343.54	341.04	336.19	315.1519444	278.82	289.46	291.96	296.81	317.8480556
132	788	416.3125	412.845	397.3625	391.65	372.2258333	371.6875	375.155	390.6375	396.35	415.7741667
136	529	270.4675	265.035	262.165	255.62	243.4075	258.5325	263.965	266.835	273.38	285.5925
138	517	556.445	536.805	517.5775	504.85	467.3155556	-39.445	-19.805	-0.5775	12.15	49.6844444

Table A.13: Travel Time comparisons and differences to other Wijken - Stadionbuurt - Midday - 15/10/2017

dstid	Travel Time	Planning Time	90th	85th	80th	Mean Travel Time	TT-PT	TT-90th	TT-85th	TT-80th	TT-Mean
11	1347	663.004	637.951	634.5675	631.504	580.7954167	683.996	709.049	712.4325	715.496	766.2045833
16	1614	896.7425	879.491	867.277	861.058	798.9659722	717.2575	734.509	746.723	752.942	815.0340278
21	951	682.023	671.899	662.731	646.796	605.9179167	268.977	279.101	288.269	304.204	345.0820833
23	1843	1474.7265	1451.019	1447.215	1412.584	1335.146111	368.2735	391.981	395.785	430.416	507.8538889
24	2128	1286.3615	1235.475	1207.0495	1188.302	1129.656111	841.6385	892.525	920.9505	939.698	998.3438889
25	1789	1128.2295	1098.639	1086.618	1066.24	1009.920278	660.7705	690.361	702.382	722.76	779.0797222
26	1433	1396.6635	1373.916	1358.335	1328.528	1240.226111	36.3365	59.084	74.665	104.472	192.7738889
28	1763	1157.8695	1109.156	1087.558	1075.364	1024.684583	605.1305	653.844	675.442	687.636	738.3154167
30	1367	1144.987	1117.477	1109.925	1086.126	1038.640139	222.013	249.523	257.075	280.874	328.3598611
33	1396	913.096	882.31	853.314	835.306	755.4728986	482.904	513.69	542.686	560.694	640.5271014
41	1490	1034.676	979.041	965.306	959.632	907.4581944	455.324	510.959	524.694	530.368	582.5418056
42	1336	1169.62	1095.33	990.64	977.47	897.7614754	166.38	240.67	345.36	358.53	438.2385246
43	1410	916.867	880.373	865.2395	853.36	808.2933333	493.133	529.627	544.7605	556.64	601.7066667
44	1427	834.3965	821.936	807.5605	797.552	742.2445833	592.6035	605.064	619.4395	629.448	684.7554167
45	1750	489.8155	474.055	467.345	460.634	417.03625	1260.1845	1275.945	1282.655	1289.366	1332.96375
46	1330	766.651	759.265	730.492	723.35	676.3638889	563.349	570.735	599.508	606.65	653.6361111
47	579	248.3735	246.857	243.9195	241.636	229.7006944	330.6265	332.143	335.0805	337.364	349.2993056
48	954	784.5715	768.464	761.333	752.812	720.2859722	169.4285	185.536	192.667	201.188	233.7140278
49	733	694.2705	658.955	644.0685	636.602	600.6895833	38.7295	74.045	88.9315	96.398	132.3104167
50	1042	963.99	937.068	904.31	893.26	793.5920408	78.01	104.932	137.69	148.74	248.4079592
51	1330	1061.31	1034.235	1006.8585	996.528	944.5881944	268.69	295.765	323.1415	333.472	385.4118056
52	1297	1129.939	1085.779	1066.764	1056.268	998.5620833	167.061	211.221	230.236	240.732	298.4379167
54	1368	1069.3495	1044.281	1024.4235	1008.3	936.7765517	298.6505	323.719	343.5765	359.7	431.2234483
61	1055	821.83	789.694	759.88	732.864	683.6504167	233.17	265.306	295.12	322.136	371.3495833
63	1394	799.98	781.073	760.5085	745.398	700.59375	594.02	612.927	633.4915	648.602	693.40625
65	1168	747.589	705.406	699.9095	684.222	633.1931944	420.411	462.594	468.0905	483.778	534.8068056
66	1317	1147.659	1106.404	1041.574	1023.542	905.2185106	169.341	210.596	275.426	293.458	411.7814894
67	1090	848.8	790.42	787.44	763.5	689.4239344	241.2	299.58	302.56	326.5	400.5760656
69	1677	641.2435	604.65	595.7425	574.22	520.7263889	1035.7565	1072.35	1081.2575	1102.78	1156.273611
70	959	427.2345	415.888	400.8775	383.834	337.7516667	531.7655	543.112	558.1225	575.166	621.2483333
72	404	195.2985	186.286	184.991	179.326	166.3502778	208.7015	217.714	219.009	224.674	237.6497222
73	518	475.31	467.445	462.591	456.046	430.0847222	42.69	50.555	55.409	61.954	87.91527778
75	180	244.276	236.831	230.4935	226.538	211.8211111	-64.276	-56.831	-50.4935	-46.538	-31.82111111
78	820	605.4315	575.624	564.625	557.134	521.2013889	214.5685	244.376	255.375	262.866	298.7986111
79	1107	956.3115	933.131	862.836	814.768	748.2375	150.6885	173.869	244.164	292.232	358.7625
80	967	695.6305	646.796	637.8655	629.058	585.5088889	271.3695	320.204	329.1345	337.942	381.4911111
81	1351	988.694	977.904	949.482	919.48	859.2853623	362.306	373.096	401.518	431.52	491.7146377
82	1488	1041.448	995.148	939.36	919.206	839.3310769	446.552	492.852	548.64	568.794	648.6689231
83	1146	906.6925	888.445	853.82	825.09	752.2260606	239.3075	257.555	292.18	320.91	393.7739394
84	1039	749.4575	703.03	689.565	674.228	619.0375	289.5425	335.97	349.435	364.772	419.9625
85	205	263.2035	234.17	215.3685	208.292	190.0290278	-58.2035	-29.17	-10.3685	-3.292	14.97097222
128	1359	662.5185	654.89	623.644	616.616	566.3943056	696.4815	704.11	735.356	742.384	792.6056944
130	971	448.538	436.108	421.0005	408.958	385.6641667	522.462	534.892	549.9995	562.042	585.3358333
132	1371	527.8675	510.032	505.1735	496.728	457.8363889	843.1325	860.968	865.8265	874.272	913.1636111
136	752	350.819	341.317	331.8015	330.336	312.4879167	401.181	410.683	420.1985	421.664	439.5120833
138	831	594.4775	578.962	566.79	557.352	529.3983333	236.5225	252.038	264.21	273.648	301.6016667

Table A.14: Travel Time comparisons and differences to other Wijken - Stadionbuurt - PM - 15/10/2017

dstid	Travel Time	Planning Time	90th	85th	80th	Mean Travel Time	TT-PT	TT-90th	TT-85th	TT-80th	TT-Mean
11	1151	637.3975	624.59	610.0775	606.29	572.7247222	513.6025	526.41	540.9225	544.71	578.2752778
16	1344	864.7775	840.52	819.9725	811.09	776.545	479.2225	503.48	524.0275	532.91	567.455
21	711	635.1925	624.6	616.7425	604.74	563.8988889	75.8075	86.4	94.2575	106.26	147.1011111
22	1421	1626.795	1535.465	1505.5525	1484.05	1362.025	-205.795	-114.465	-84.5525	-63.05	58.975
23	1467	1464.1525	1392.115	1380.7225	1370.44	1290.644722	2.8475	74.885	86.2775	96.56	176.3552778
24	1495	1226.185	1197.545	1189.0725	1184.2	1112.791111	268.815	297.455	305.9275	310.8	382.2088889
25	1266	1070.51	1055.16	1047.2975	1027.45	970.3486111	195.49	210.84	218.7025	238.55	295.6513889
26	1267	1366.662	1304.624	1280.541	1278.678	1230.963429	-99.662	-37.624	-13.541	-11.678	36.03657143
28	1266	1110.8675	1068.445	1055.4175	1050.62	996.2002778	155.1325	197.555	210.5825	215.38	269.7997222
29	911	890.31	868.34	858.71	849.64	817.1419444	20.69	42.66	52.29	61.36	93.85805556
30	1101	1122.425	1075.508	1072.294	1052.934	1024.410286	-21.425	25.492	28.706	48.066	76.58971429
40	1214	1170.489	1125.536	1101.572	1066.932	1001.351143	43.511	88.464	112.428	147.068	212.6488571
41	1118	1007.7775	998.135	985.9875	965.94	894.0788889	110.2225	119.865	132.0125	152.06	223.9211111
42	1390	1016.564	990.571	976.878	922.446	844.6158824	373.436	399.429	413.122	467.554	545.3841176
43	1096	879.27	868.415	862.1275	855.4	793.1227778	216.73	227.585	233.8725	240.6	302.8772222
44	1038	804.62	794.4	786.2075	774.45	723.9536111	233.38	243.6	251.7925	263.55	314.0463889
46	905	720.1175	705.91	693.9075	688.4	649.9497222	184.8825	199.09	211.0925	216.6	255.0502778
47	333	239.5575	236.58	233.8775	228.97	220.4202778	93.4425	96.42	99.1225	104.03	112.5797222
48	801	747.815	739.09	729.61	725.19	687.8791667	53.185	61.91	71.39	75.81	113.1208333
49	666	663.8925	648.39	642.745	630.14	598.3644444	2.1075	17.61	23.255	35.86	67.63555556
51	1122	1047.9525	1017.4	1002.77	998.21	937.0877778	74.0475	104.6	119.23	123.79	184.9122222
52	1230	1080.595	1068.055	1040.8975	1032.69	989.9363889	149.405	161.945	189.1025	197.31	240.0636111
54	1010	1127.5055	1080.592	1027.0575	1018.144	947.2675	-117.5055	-70.592	-17.0575	-8.144	62.7325
61	843	756.7275	719.63	695.8875	686.23	645.0694444	86.2725	123.37	147.1125	156.77	197.9305556
63	1411	804.1175	758.565	750.485	738.45	687.2213889	606.8825	652.435	660.515	672.55	723.7786111
70	951	360.97	338.31	328.3975	322.24	296.3905556	590.03	612.69	622.6025	628.76	654.6094444
71	525	250.4225	240.1	236.9925	231.45	217.625	274.5775	284.9	288.0075	293.55	307.375
72	387	187.965	175.88	173.82	171.5	162.8311111	199.035	211.12	213.18	215.5	224.1688889
73	474	453.2825	447.2	445.785	435.43	419.2169444	20.7175	26.8	28.215	38.57	54.78305556
75	195	230.7775	219.805	217.1125	213.44	207.1183333	-35.7775	-24.805	-22.1125	-18.44	-12.11833333
78	585	549.655	535.74	525.5575	522.97	493.3444444	35.345	49.26	59.4425	62.03	91.65555556
80	653	609.4275	598.1	593.56	579.85	550.2397222	43.5725	54.9	59.44	73.15	102.7602778
81	847	988.744	951.316	914.156	874.292	842.2557576	-141.744	-104.316	-67.156	-27.292	4.744242424
83	839	825.962	806.038	784.998	760.264	718.0984848	13.038	32.962	54.002	78.736	120.9015152
84	734	649.85	632.735	623.425	619.36	558.9161111	84.15	101.265	110.575	114.64	175.0838889
85	171	210.9675	207.85	204.5525	199.15	182.7419444	-39.9675	-36.85	-33.5525	-28.15	-11.74194444
128	1116	648.785	616.405	588.4875	582.13	540.7980556	467.215	499.595	527.5125	533.87	575.2019444
130	798	418.2925	409.035	395.925	393.55	374.3177778	379.7075	388.965	402.075	404.45	423.6822222
132	933	484.1675	478.59	473.115	471.04	447.1377778	448.8325	454.41	459.885	461.96	485.8622222
136	497	324.615	321.23	319.59	315.3	297.7658333	172.385	175.77	177.41	181.7	199.2341667
138	595	543.0925	541.145	529.345	528.59	493.6936111	51.9075	53.855	65.655	66.41	101.3063889

B | SUPPLEMENTARY MODEL RESULTS

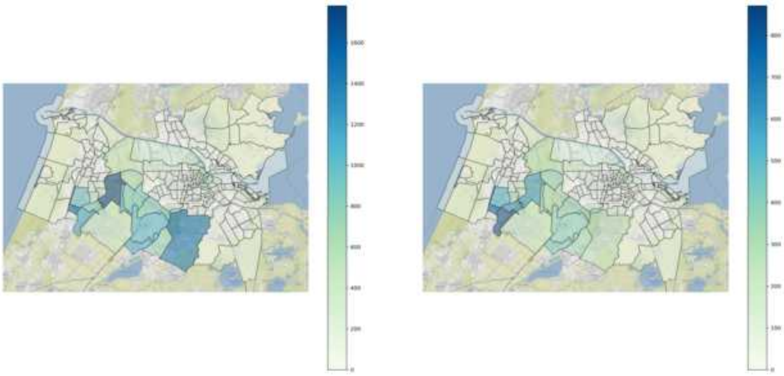
(a) Spatial spread - Early morning (left) & Spatial spread - AM peak (right)

(b) Spatial spread - Midday (left) & Spatial spread - PM peak (right)

(c) Spatial spread - Evening

Figure B.1: Spatial spread of Ubers across different time periods on simulating the model with 905 Uber vehicles

(a) Spatial spread - Early morning (left) & Spatial spread - AM peak (right)



(b) Spatial spread - Midday (left) & Spatial spread - PM peak (right)

(c) Spatial spread - Evening

Figure B.2: Spatial spread of Ubers across different time periods on simulating the model with 3620 Uber vehicles

C INTERVIEW

The following is an interview with Ottmar Francisca who drives for Uber. The questions relate to the spatial-temporal aspects and type of users. The questions have accordingly been split.

Temporal Aspects

1. What period of the day do you tend to get most trips (Early monring - 00:00 to 06:00, AM peak - 07:00 to 09:00, Midday - 09:00 to 15:00, PM peak - 16:00 to 18:00 and evening - 19:00 to 23:00)

AM and PM peak. They are both peak periods. Morning and evening (as well).

2. Do trips around midnight tend to be especially high?

If you mean high by euros (money) than not. Depends on the surge, the trips get higher. That can be when there are lot of requests in an area. Can be because of rain, events or promotion.

3. What hours of the day are you travelling around searching for a passenger the most - (Early monring - 00:00 to 06:00, AM peak - 07:00 to 09:00, Midday - 09:00 to 15:00, PM peak - 16:00 to 18:00 and evening - 19:00 to 23:00)?

My working time is 06.00 Am till 18.00 pm approximately.

4 .Is there a seasonal trend for Uber usage? Do you tend to get more trips in certain months compared to the others?

November, 1st to 15th December and January are a bit lower than normal. But still good earnings.

5. Are weekdays or weekends busier?

At the times I work, the best days are Sundays and Mondays. Sundays tourists are going home, seems there are more trips to the airport. Mondays is the start of the week.

Spatial aspects

6. What are the most popular areas where you have pick-ups and drop-offs (For instance: Schiphol, Central Station, Blijmer arena) - Does this change according to the time of the day, day of the week and month? If yes, can you also please give me some examples?

Airport is always popular. This because the trip from Amsterdam to the Airport is about 15 to 25 km and from the airport you get someone back. Other locations are not very special only if there is something special like an event.

7. Are there a lot of trips to and from Haarlem?

There are trips to and from Haarlem. I see that the trips on the peak hours getting more each day. But seems there are not a lot of drivers in Haarlem. This is the reason the surge goes up during peak hours.

8. Do you also have trips to Bloemendaal, Zandvoort, and Zaandam?

Yes, especially in the warm months. For Bloemendaal and Zandvoort: these are normally not tourists. For Zaandam (Zaanseschans), there are a lot of tourists dropped there.

9. Are most of your trips short distance (0-5 km) or long-distance trips (above 5 km)?

There is a mix. I think 60% short and 40% long

Passengers using Uber

10. What is the purpose of most travellers to use Uber? For example: Going to work, hotels, tourism, shopping etc.?

85 % are tourist. Local people only use Uber when they are late, when they go out or when it rains.

11. If most of the people using Ubers are tourists? Which places do you have to drop or pick them?

Picked up from hotels and dropped at hotels and tourist spots

D SUPPLEMENTARY IMAGE

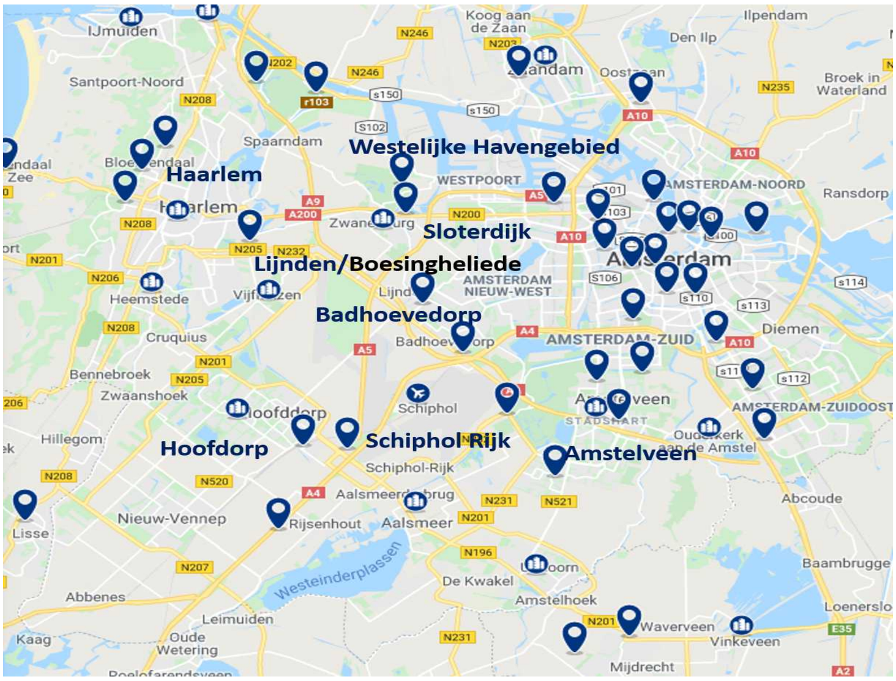


Figure D.1: Spread of hotels across the greater Amsterdam region as per listings on Booking.com [Booking.com, 2019]

COLOPHON

This document was typeset using LaTeX. The document layout was generated using the arsclassica package by Lorenzo Pantieri, which is an adaption of the original classicthesis package from André Miede.