MSc thesis in Geomatics for the Built Environment

Data driven sustainable mobility analysis in the city of Amsterdam

C.A.N.L. Duynstee

2020

TU Delft

DATA DRIVEN SUSTAINABLE MOBILITY ANALYSIS IN THE CITY OF AMSTERDAM

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

Master of Science in Geomatics for the Built Environment

by

C.A.N.L. Duynstee

July 2020

The work in this thesis was made in the:

OTB-Research for the built environment
Department of OTB and Department of Urbanism
Faculty of Architecture & the Built Environment
Delft University of Technology

TU Delft Supervisors:	Drs. C.W. Quak
	Dr.ir. M. Snelder
Co-reader:	Dr.ir F.M.Welle Donker
TNO Supervisors:	Dr.ir M.Snelder
	ir. E.J van Ark

ABSTRACT

This research focuses on the analysis of Floating Car Data (FCD) data to understand sustainable transportation behavior in the area of Amsterdam. Using data collected by mobile devices, trips are analyzed by their distance and transportation method.

Sustainability is a relevant factor to the social and scientific community. With increasing population and growing cities, the impact of travel on the environment is also increasing. New policies are required to stimulate a more sustainable approach to transportation. Research on sustainable travel behavior provides input for the policy makers. The main research question is defined by what extent FCD can be used to provide insights in the sustainable mobility behavior in Amsterdam.

In the existing literature, there are four objectives for sustainable mobility: hazards reduction, travel reduction, modal shift and accessibility. The modal shift (e.g. replacing care usage by public transport) is one of the key drivers for this research. To quantify the behavior, different sustainable mobility indicators are identified to determine the sustainable direction related to the mode of transportation. In the literature there are many methods to quantify sustainable transportation behavior ranging from the traditional methods like counting and surveys to modern approaches including smartphone and sensor data.

The methodology used starts with the problem statement and literature review. The data sets are selected and analyzed, providing the results and conclusions. A decision tree is used for categorizing different trips people make, where a difference between short and longer distance trips is made. This research makes use of a wide range of tools ranging from PostgreSQL databases to advanced features of ESRI to visualize data.

Five different available data sets are analyzed for their suitability for this research. Based on several requirements like availability, having Origin and Destination (OD) information, usability and documentation, the data sets are assessed and two data sets (Google OD and LMS) are chosen to be analyzed in more detail. The data is filtered and cleaned to make sure it fits the scope (Amsterdam) and the two sets are compared to each other. The trips are split into short and long distance trips, where for both a detailed analysis is performed on the trips which can be easily replaced by more sustainable trips.

Regarding the short distance trips, analysis shows that walking and biking are the most common option in busy areas as the city center and the business district. The less sustainable car trips for short distances show several patterns in the city. Using interactive maps, these patterns are identified and both data sets are compared. For longer distance car trips, train transportation is the more sustainable option. There is however a tradeoff between the most sustainable and least time-consuming transportation option. This makes that train trips are not always the most logical or even sustainable choice for transportation. A method is developed and applied to the data set to test if a train trip is a realistic alternative.

The results show for both data sets that there is a significant amount of car trips which could easily be replaced by the train. During the analysis of the data for long distances, a problem has been found with the Google data set. It seems that this data set is showing most of the time the same noise. This phenomenon was found by comparing the different modes of transportation available in the Google data set.

The error is not in all the data, but it seems to have an impact on the data set.

Conclusion is that the data provides many insights on travel behavior and different available data sets can be linked together to provide deeper insights. The Google data set shows interesting results for shorter distances but gives less reliable results for longer distances. The LMS data set is used to compare the results for both short and long distances. The use of FCD data to study and stimulate sustainable mobility behavior seems very promising, but the quality of the available data sets has a large influence on the usability.

ACKNOWLEDGEMENTS

This document is the result of my Master Thesis Project of the Master's program Geomatics for the Built Environment at Delft University of Technology. The thesis was carried out in collaboration with the company Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek (TNO), department of Sustainable Urban Mobility and Safety.
While writing this thesis, I received help from a number of people. I want to thank Wilko Quak as my main mentor and reviewer. Thank you for your patience and guidance along the way.

I would like to thank my second mentor and supervisor from TNO Maaike Snelder for her guidance, inspiring thoughts and valuable suggestions during my thesis. During my time at TNO I learned a lot in the field of mobility. I would also like to thank Ernst Jan van Ark who guided me in the first period of my internship at TNO and Fieke Beemster who introduced me to TNO and supported me during the process. I am very happy that I did my thesis in collaboration with TNO.

I would also like to thank Jorge Gil for the discussions on sustainable mobility. His background in sustainable mobility was very valuable as a basis of the report and my literature study.

In the last part of my graduation, Frederika Welle Donker became involved as a fellow reader for my thesis. Thank you for taking the time to read this thesis and to provide helpful suggestions.

Last but not least, I would like to thank my family and friends for their support. I would especially like to thank my boyfriend Hidde Westra who always supported me during the entire journey of my dissertation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

API Application Programming Interface . . . 42
CBS Statistics Netherlands . . . 11
ETL Extraction, Transformation and Load . . . 17
FCD Floating Car Data . . . 1
GDPR General Data Protection Regulation . . . 12
GNSS Global Navigation Satellite System . . . 12
GPS Global Positioning System . . . 1
IoT Internet of Things . . . 12
KIM Kennisinstituut voor Mobiliteitsbeleid . . . 43
LMS Landerlijk Model Systeem . . . 13
NRM Nederlands Regionaal Model . . . 22
NS Nederlandse Spoorwegen . . . 42
OD Origin-Destination . . . 1
OViN Onderzoek Verplaatsingen in Nederland . . . 11
ROI Region Of Interest . . . 43
TNO Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek vii
UTC Coordinated Universal Time . . . 32
VF verplaatsingstijdfactor . . . 10

1 INTRODUCTION

This chapter provides the context of the research, states the problem with the research question and subquestions. It clarifies the scope of the research and provides an overview of the rest of the chapters.

1.1 CONTEXT

Nowadays cities are developing rapidly. This urbanization has several positive and negative effects on the cities and the people who live in these cities. Urbanization may lead to economic and cultural development, which is important for the economy of a country. Mobility is an ingredient of a well-functioning society. Mobility of employees and goods make the economy more productive and other forms of mobility help sustain the social and cultural network [OECD, 2015]. Although there are positive effects, the growing size of cities and increasing population is resulting in a rapid increase in the number of vehicles on the roads [Djahel et al., 2015]. Transportation can have different impacts: environmental, social equity, economic, cultural, land use and urban form are the most important ones [Schiller et al., 2010]. The increased amount of traffic results in high levels of energy consumption, CO_2 emissions, noise, air pollution, degradation of the urban landscape with infrastructure and the use of open space for car parking [Gil, 2016]. In order to deal with these issues new policies are required for a more sustainable approach towards mobility. This gives many challenges for road traffic management authorities and urban planners. Sustainable mobility aims at promoting better and healthier ways of meeting individual and community transportation needs. It also reduces the social and environmental impacts of current mobility practices [Schiller et al., 2010]. To visualize the sustainable mobility performance and to monitor the impact of policies taken towards sustainable mobility, research on sustainable mobility is needed. With this graduation research the sustainable mobility performance is studied by making use of different data sets collected in the Netherlands.

1.2 PROBLEM STATEMENT

To analyze the mobility and sustainable mobility behavior, data on the transportation modes of the population needs to be collected. Currently, traffic models are based on countings, surveys, statistics and estimations which do not, or only sample based, provide information on the origin and destination of the traffic [Vries, 2012]. These methods give a good insight in the travellers behaviour, but are limited by the sample size and accuracy of the data gathering. With the introduction of mobile phones equipped with Global Positioning System (GPS) sensors, new opportunities of collecting traffic data became available using Origin-Destination (OD) matrices. These methods yield larger data sets, covering a significant portion of the trips, but give less insight in personal preferences. The data generated by mobile phones and mobile GPS sensors in other devices is called Floating Car Data (FCD). Nowadays a lot of FCD data is collected in the Netherlands. With FCD data real traffic behavior can be analyzed over a broader group of users. With this data, traffic from and to different neighborhoods can be analyzed. One of the challenges is how to process the amount of data and data sources. In this research an effort is made to investigate to which degree this data can be used to provide information on the sustainable mobility in cities. Combining FCD data with demographic data can provide more insight the general incentives of choices for different means of transportation.

1.3 SOCIAL AND SCIENTIFIC RELEVANCE

This research will make a contribution to the existing body of knowledge on sustainable mobility data. This will be done by adding empirical research on sustainable mobility performance on neighborhood level, using passively collected data which is collected by smartphone applications with a big group of participants providing data. Because the passively collected FCD data is a relative new type of data, less research is done on this type of data. Using this passive GPS tracking, people have not explicitly approved the use of their geo data (e.g. via apps on their smartphones). This makes that the data needs to be anonomized by using either aggregation or cutting off beginning and end points. This is in contrast with active GPS tracking where people are aware that their location is tracked, this however gives a more limited data set and the behavior may be biased. At this moment, researchers are searching for ways to use this new type of data. With this research, a new way is developed to use the FCD data in the research towards traffic behavior and sustainable mobility. One of the benefits of the use of passive FCD data is that it provides a cost reduction with respect to more traditional ways of obtaining data, by using counters or detection loops in roads. Another benefit is that FCD also adds ODs not present in counts.

This research provides information for urban planners and provides extra tools to map sustainable mobility performance. This can be used to improve policies towards sustainable mobility, improve the CO_2 reduction, and validate the regulations. The research aims to identify which relationships are contributing to a sustainable mobility performance of a neighborhood.

1.4 RESEARCH QUESTIONS

The aim of this research is to give new insights in sustainable mobility using FCD data. For this research the following research question will be answered:

To what extent can Floating Car Data (FCD) be used to give an insight in the sustainable mobility behavior in Amsterdam?

To facilitate this, the following set of sub questions is defined:

1. What is sustainable travel behavior and why is it important?
2. Which data sets are commonly used and available to analyze travel behavior?
3. What are the differences between these data sets and which ones suited best?
4. Which short distance car trips in Amsterdam could be replaced by more sustainable opportunities like walking or cycling ?
5. Which long distance car trips in Amsterdam could be replaced by more sustainable opportunities like public transportation?

1.5 SCOPE OF THE RESEARCH

This thesis will focus on mapping the sustainable mobility performance. As a use case the neighborhoods of Amsterdam are investigated (see figure 1.1 for the area in scope). Considered in this research will be the traffic which has an origin and/or a destination in one of the neighborhoods of Amsterdam (e.g. a trip starting at the Rijksmuseum and ending in Delft will be taken into account). Because of the size of the data set, the research is focused on a specific area in the Netherlands. As Amsterdam is big enough to draw conclusions, an extension can be made for other parts of the Netherlands.

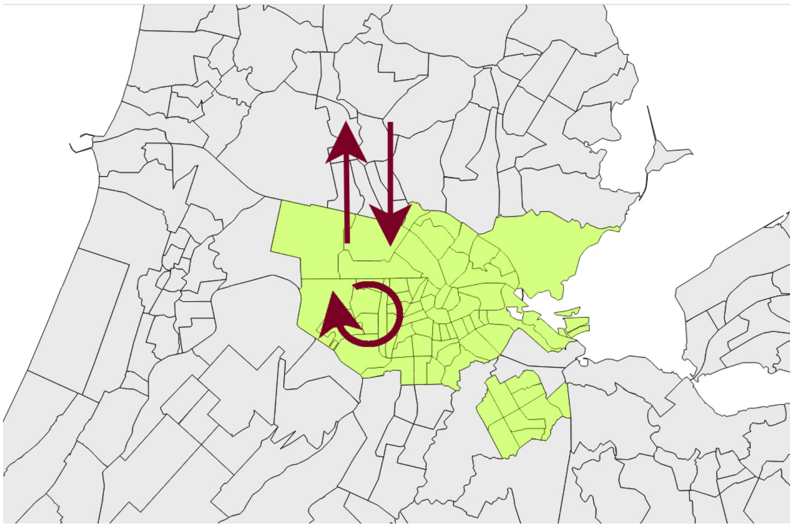


Figure 1.1: Research area

To test and explain the sustainable mobility performance in the different neighborhoods in Amsterdam, different aspects will be investigated. The different aspects can be tested because of the available data that is collected. Although it can give new insights on the sustainable performance, some aspects cannot be investigated from the data that is available for this research. Due to scarce data availability, these aspects are out of scope:

- Efficiency of the cars that are in the data sets (some cars use less fuel for the same distance)
- Use of electric car vehicles
- Carpooling
- Use of shared cars (i.a. Greenweels etc)
- Use of other forms of public transport than the train

Finally, it should be noted that the analysis that will be done in this thesis gives an insight of sustainable mobility of Amsterdam in a specific time window (2015-2016). The method that is used for this analysis can be used for future work, and can be enriched with new data and new methods when the built environment changes.

1.6 THESIS OUTLINE

This thesis consists of the following chapters.

- In Chapter 2 related work will be reviewed by performing an literature study on the available material on this topic.
- In Chapter 3 the research methodology is explained.
- In Chapter 4 the data selection process is explained and the reasons for choosing certain data sets is explained
- In Chapter 5 the data preparation that is needed to do the experiments is described.
- In Chapter 6 the implementation of the experiments with the chosen data is described
- In Chapter 7 the results of te experiments are presented.
- In Chapter 8 the main conclusions are summarized and recommendations for further research are given.

2 RELATED WORK

Numerous studies have been done on sustainable mobility. This chapter gives a summary of related research found in the literature. First, the definition of sustainable mobility is explained, then a deeper analysis of current traffic modelling is performed. The data classification model is provided. The influence of behavior and the urban form on sustainable mobility is discussed. The final parts of this Chapter will deal with the sustainable mobility regulations and ethics.

2.1 SUSTAINABLE MOBILITY

There are multiple ways how sustainable mobility is defined in the literature. Generally speaking, sustainable mobility deals with the transportation of people and goods in relation to environmental and social impact. For this research, a definition for sustainable transport has been used, which has been implemented in several studies [Gil, 2016; Gudmundsson et al., 2016; Kowalczyk, 2010; Hall and Sussman, 2006; Toth-Szabo et al., 2011]. This definition is based on an adapted version of European Council [2001] of an expanded definition by Centre for Sustainable Transportation [1997]:
"A sustainable transport system [is] defined as one that:

- allows the basic access and development needs of individuals, companies and societies to be met safely and in a manner consistent with human and ecosystem health, and promotes equity within and between successive generations;
- is affordable, operates fairly and efficiently, offers choice of transport mode, and supports a competitive economy, as well as balanced regional development;
- limits emissions and waste within the planet's ability to absorb them, uses renewable resources at or below their rates of generation, and, uses non-renewable resources at or below the rates of development of renewable substitutes while minimizing the impact on the use of land and the generation of noise." ([European Council, 2001] pp.15-16)

Sustainable mobility aims at promoting better and healthier ways of meeting individual and community transportation needs. It also reduces the social and environmental impacts of current mobility practices [Schiller et al., 2010]. Nowadays urban areas face serious problems linked to the current car based travel patterns, affecting the environment and the socio-economic fabric of society [Gil, 2016]. Since the 1970s, concerns about negative impacts of transportation started to grow. Highway oriented planning around cities and the associated increasing pollution caused the topic to be on the agenda [Schiller et al., 2010]. From this stage planning policy and guidance has been concerned with these problems and environmental awareness has increased [Gil, 2016]. In absolute sense there is no mode of transportation that is 100% sustainable [Banister, 2009]. Different modes of transportation will have different impact on the environment. In Table 2.1 different modes of transportations and their environmental impact are shown. The impact of these different modes of transportations is not only based on their use of oil/energy, but also the impact on

Mode	Seats/space	MJ/vehicle km	MJ/seat km	MJ/passenger km
Air Boeing 727	167	243	1.45	2.42
Rail electric/diesel	377	168	0.45	1.65
Metro underground	555	141	0.25	1.69
Tram light rail	265	79.8	0.30	0.90/1.20
Bus	48	14.7	0.34	0.92/1.53
lorry				2.94
car	4	3.7	0.92	2.10
Motorcycle	2	1.9	0.95	1.73
Cycling	1	0.06	0.06	0.06
Walk	1	0.16	0.16	0.16

Table 2.1: Impact of different modes of transportation on the environment [Banister, 2009]

the urban space and the occupancy rate of the transportation mode is taken into account.

In terms of sustainable mobility, walking and cycling are considered as most sustainable because they use very little non-renewable energy [Banister, 2009]. Public transport with a high occupancy rate is considered as the best alternative when it comes to sustainable transportation. It should be noticed that the occupancy rate is very important for the energy consumption of this type of transport. Finally, car, lorry and air transportation can be seen as less sustainable mode of transportations [Banister, 2009].

Traffic calming for personal motor vehicles and pedestrianization (excluding personal motor vehicles from certain streets) may have many benefits for mobility and the environment. It increases the numbers of people walking and using public transport and it will decrease the traffic related injuries, especially those of pedestrians and bicyclist [Schiller et al., 2010].

2.2 CLASSIFICATION METHODS OF SUSTAINABLE MOBILITY

As described in section 2.1 sustainable mobility is highly related to the mode of transportation used by an individual. The mobility performance will be measured by classifying different modes of transportations and travel distance. A base for the classification method for this research is the method that is developed by Gil 2016. This section provides a literature review discussing the main objectives of sustainability mobility. The overview of this literature study is shown in Table 2.2. This is explained in more detail in the remaining of this section.

General objectives	Specific objectives
Hazards reduction	Reduce CO2emissions Reduce air pollution Reduce land consumption Reduce urban landscape degradation Reduce noise Reduce accidents
Travel reduction	Reduce energy consumption Reduce congestion Reduce distance travelled Reduce need to travel
Modal shift	Reduce car use in urban areas Increase walking and cycling Increase share of public transport Replace medium and long distance car travel by rail
Accesability	Maintain or increase accessibility (while reducing mobility) Narrow the accessibility divides

Table 2.2: Summary of the objectives of sustainable mobility [Gil, 2016] based on the work of [Banister, 2005; Black, 2010; Bruun et al., 2012; Centre for Sustainable Transportation, 2002; European Commission, 2007; World Business Council for Sustainable Development, 2001]

The first general objective of this literature review is the hazard reduction. The focus of this objective is the reduction of different factors that have impact on the environment of a city like land consumptions, urban landscape degradation and accidents. But also more general objectives like CO_2 emissions, air pollution, noise should be reduced. The second general object is focused on travel reduction. Travel reduction can be realized by reducing the need to travel , but also the travelled distance. Travel reduction can lead to a reduction of traffic congestion and less energy consumption.

The third, model shift is an important base for this research. It focuses on the way citizens travel in a city. By reducing the car use and increasing the share of public transport, walking and cycling the impact in terms of hazards, energy and congestion will be lower and it will improve accessibility in the city [Gil, 2016]. For this reduction it is important to understand the modal shift in a city and analyze which car trips could be replaced by more sustainable alternatives, like for example walking, cycling or public transport, which will be the focus of this research.

The fourth objective is about accesability and includes aspects of land use and transport infrastructure [Bertolini et al., 2005; Curtis, 2011]. Accessibility relates to people and goods rather individual modes of transportation [Halden, 2002]. Although accessibility is an important objective, this research will be focused on the third objective model shift.

To go more in depth on the objective of model shift, different sustainable mobility indicators can be described. The overview of this selected indicators related to modal shift are shown in Table 2.3. The focus of this indicators is to decrease the use off unsustainable car use, and increase the use of more sustainable opportunities, like cycling and public transport.

Modal shift	Indicators	Sustainability direction
Non-motorised share	Neighbourhood walking share share Neighbourhood cycling share City cycling share	increase
car share	Neighbourhood car share City car share Regional car share	decrease
Public transport share	Neighbourhood transit share City transit share Regional transit share	increase

Table 2.3: Selected sustainable mobility indicators related to modal shift [Gil, 2016]

The selected indicators indicate which modal shift contributes to a positive sustainability direction. A classification method based on these objectives is shown in Table 2.4.

The classification is tested with empirical data from the Netherlands Mobility Survey called OVIN (Onderzoek Verplaatsingen in Nederland) which is conducted between 2004 and 2009. It makes use of different modes of transportations and makes a distinction between distances between different trips (differentiated in 1. short (< 1.5 km), 2. medium (> 1.5 and <10km) and 3. long (>10km)). This classification will be used as a base for this research, to provide insights in the sustainable mobility performance in the different neighborhoods in Amsterdam. The sustainability direction shows the impact of the indicators on sustainable mobility objectives, as defined in section 2.1.

Indicator	Sustainability Direction
Share of short walk journeys	+ + +
Share of walk journeys	+ + +
Share of short cycle journeys	+ + +
Share of medium cycle journeys	+ + +
Share of cycle journeys	+ + +
Share of short car journeys	- - -
Share of medium car journeys	- -
Share of long car journeys	-
Share of car journeys	- - -
Share of car distance	- -
Share of car duration	- -
Share of medium local transit journeys	+ +
Share of local transit journeys	
Share of long train journeys	+ +
Share of train journeys	+ +
Share of transit distance	+ +
Share of transit duration	+ +
Mean journey distance	-
Mean daily distance per person	-
Mean daily journeys per person	-

Table 2.4: Selection of sustainable mobility indicators. The Sustainability direction shows the intended direction of the indicator in relation to general sustainable mobility objectives. [Gil, 2016]

Using the different indicators with the sustainability directions shown in table 2.4, mode of transportation can be analyzed based on travel distance. The indicators are classified into three different distance groups: long, medium and short. In addition to these distance groups, there is also a general direction per modality shown, but this is not used further because the classification per distance is more specific than the general direction. Each indicator has a different sustainability direction, which make it easier to show more sustainable alternatives for unsustainable options, based on the distance groups that are made. For each distance, the most sustainable option is shown in figure 2.1.

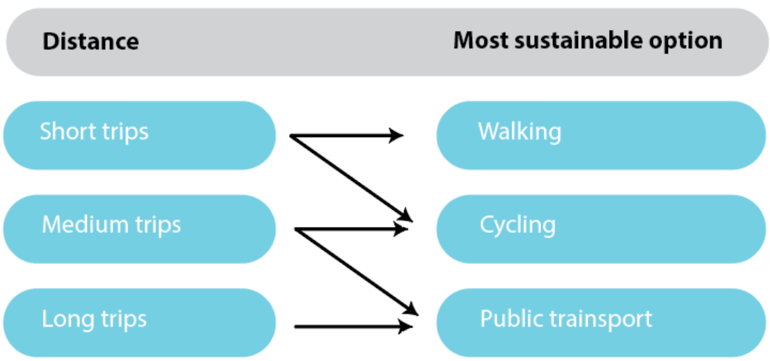


Figure 2.1: Sustainable option for the different classes in distance

2.3 ANALYSIS OF TRAVEL BEHAVIOUR

Understanding travel behavior and the reasons for choosing one mode of transport over another is a complex phenomenon. For every journey, new choices between different transport modes need to be made. The choice of one specific mode of transportation can vary over time and with the type of journey that is made [Beirão and Cabral, 2005]. The main quality elements for choosing car or public transport according to Van Goeverden and Van Den Heuvel [1993] are:

- driving time (in means of transport)
- transfer time (only for public transport)
- parking search time (only with car)
- availability by location (proximity to stop or parking space)
- availability by time (with public transport: frequency per period of the day or week)
- reliability
- comfort
- number of transfers (only for public transport)
- provision of information
- rate / payment system

The travel time determined by the first five quality elements and the convenience mainly by the last four. The travel time determined by the first five quality elements and the convenience mainly by the last four. A study done by McKinsey, where passengers were asked about their choice of transport mode, shows an indication of the importance of the different quality elements. The quality of the transport turned out to be much more important for the choice of transport mode than the price. The travel time was mentioned the most and to a lesser extent convenience [McKinsey, 1989].

Using the transit time factor (verplaatsingstijdfactor (VF)) the accepted travel time difference between public transport and car trips is calculated. The VF value is a simple measure of the quality of public transport in relation to that of the car. The smaller the VF becomes, the higher quality of public transport. In figure 2.2 the theoretical relationship between the VF and the share of public transport is shown.

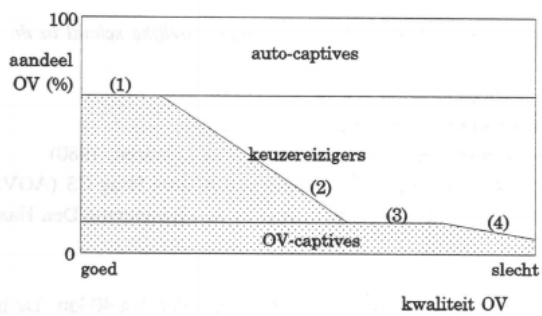


Figure 2.2: Theoretical relationship between VF and the share of public transport [Van Goeverden and Van Den Heuvel, 1993]

Figure 2.2 shows that there is always a fixed group that only uses the car or only the public transport. A certain percentage can or will not use public transport, no matter how good it is (auto captives). This also applies the other way around, where people only travel by public transport because they for example do not own a car. In addition to these two groups, there are also optional travellers. The percentage of optional travellers who opt for public transport is highly dependent on the relative quality of the public transport offer. Below a certain quality level, optional travellers will no longer use public transport and only public transport captives will remain [Van Goeverden and Van Den Heuvel, 1993]. The VF curve that shows the relationship of VF values and the share of Public transport is shown in figure 2.3.

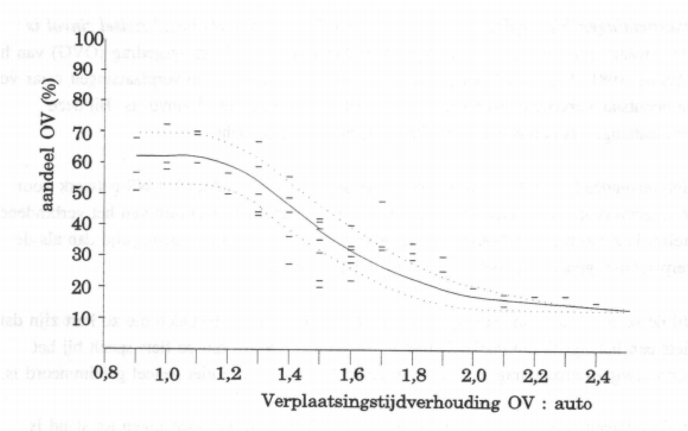


Figure 2.3: Observed relationship between VF value and the share of public transport (the "VF curve") on a number of regional relations in the Randstad [Van Goeverden and Van Den Heuvel, 1993]

With a VF value of 1.4, about half of the optional travellers choose public transport. At a VF value of 1.8, that is less than a quarter, and with a VF value of 2.4, no optional travellers wants to travel by public transport anymore. To analyse sustainable travel behaviour for optional travellers, it is therefore important to take into account the VF value which should not exceed 2.4.

2.3.1 Traditional data sets and methodologies

Most of the research on travel behavior is related to psychological and social science. These studies tries to combine personal diaries with socio economic and demographic statistics [Oliveti, 2015]. Traditional travel behavior research use data collected through paper or phone call surveys. The participants are asked to describe their travel behavior on an average day or to reconstruct their travel behavior in the previous period. It is proven that this traditional type of research deviates systematically from actual behavior. Participants often underreport the small trips as well the trips that are not starting or ending at their home. Deviations can also be found in the travel time reported. Car drivers are more likely to underestimate their travel time than participants who use the public transportation [Bohte, 2010]. Since 1978 research has been conducted by the Statistics Netherlands (CBS) into mobility of people in the Netherlands under the name Onderzoek Verplaatsingen in Nederland (OViN). The aim of this research is to to provide adequate information on the daily mobility of the Dutch population. This information will be used in the development and testing of the traffic and transport policy in the Netherlands [CBSl, 2015]. The OViN research is based on sampling. The OViN sample contains 0.3% of the total population of the Netherlands. Information of all participants is collected on travel behaviour on a particular day of the year. To be able to draw conclusions for the entire year and for the entire population based on OViN, an extrapolation is done on the OViN sample. By adding weighting variables from collected background characteristics of the participants, selectivity in the sample can be reduced. Background characteristics that are included in the weighting variables that are important for the movement behaviour are: age, gender, income, level of urbanization and vehicle ownership [CBSl, 2015]. Because the OViN research is based on a small sample of the population, many uncertainties are appearing in this dataset. Despite this limitations, the OViN research is used in the development and testing of the traffic and transport policy in the Netherlands.

2.3.2 State-of-the-art data sets and methodologies

Recently more and more big data sets are collected in the urban environment. A significant portion of this big data is geospatial data, and the size of such data is growing rapidly, at least by 20% every year [Lee and Kang, 2015]. Due to this fact, more and other traffic data is becoming available. There has been an increase in the request for embedded devices, such as sensors and smartphones that are connected through the Internet of Things (IoT), in which all devices are capable of interconnecting and communicating with each other over the Internet [Rathore et al., 2016]. This is resulting in a phenomenal amount of data collected in the urban environment.
The traffic data collected by smartphones provides a cost-effective way to collect this data [Herrera et al., 2010]. Smartphones are able to collect information about the location of the user by making use of Global Navigation Satellite System (GNSS), Wi-Fi and inertial measurement units. Because a many will carry their smartphone with them all day, traffic data can be collected in this way [Lee et al., 2014; Moloo and Digumber, 2011]. This data is passively collected, as people are not explicitely aware that their data it used. Before the General Data Protection Regulation (GDPR) it was possible to use this type of collected data. After the implementation of the GDPR on 25th may 2018 it is not possible to collect and use this data passively. Because this research was started in 2016 the GDPR had not yet come into effect.

Data collected using bluetooth techonology, Wi-Fi techonology or cell phone data have been used in previous research [Alexander et al., 2015; Duynstee et al., 2016; Braggaar, 2018; DATMobility, 2013]. For bluetooth data gathering, scanners are placed on strategic locations and cell phones connecting to bluetooth are registered. Wi-Fi tracking can make use of the same principle, but it can use also the existing Wi-Fi infrastructure, which is available on most locations. The use of cell phone data using cell phone tower triangulation has also been investigated, but the main limiting factor was the level of accuracy (hundreds of meters) [Ahson and Ilyas, 2010]. Using GPS the level of accuracy is improved (meters) and it is not required to place devices on several locations.

The need of placing devices on locations is not required. The drawback of using passive GPS data is that the data needs to be anonymized for the research by either aggregation or cutting off beginning and end points [Gambs et al., 2014].

Different companies provide applications such as Google maps and Flitsmeister collect traffic data and store this data in their databases. Because the majority of the people carry a smartphone, the penetration rate of the sampling will be much higher than the OViN research. The use of this new type of big data has been proposed as an alternative traffic sensing infrastructure, as they usually provide a cost-effective way to collect traffic data [Herrera et al., 2010; Moloo and Digumber, 2011]. This makes this data sets interesting for further research. In this research data collected by smartphones will be referred to as FCD data.

Another well-known research method is to collect data by making use of a sample group carrying GPS trackers around for a fixed amount of time (active GPS tracking). Examples of this kind of research are the graduation projects of Biljecki, Oliveti, Van der Winden which make use of the same GPS dataset collected by the University of Delft [Oliveti, 2015; Biljecki, 2010; van Winden, 2014; Bohte, 2010]. This dataset contains GPS points with coordinates (x, y, z) and time (t) recorded every 5 seconds. The sample was taken by more than 800 people in several cities during a time span of two weeks [van Winden, 2014]. With these GPS tracks, individual traffic behaviour can be analysed. This method gives a more realistic view on traffic behaviour than the traditional method by making use of paper and phone call surveys. Because the participants will carry all the time a GPS receiver, small trips will be recognized and will not be missing in the dataset. Another benefit of this type of research is that background information about the participants can be collected by questionnaires. This gives information which can be assigned to the traffic behaviour that is analysed.

Compared to the active GPS research FCD can be collected over a longer period of time and gives this information over a wider group of users. Although a bigger group can be analysed over a longer period, there is less information available on the background of the group that is analysed. In GPS research normally the sample group provides personal information via questionnaires, which is in practise less feasible for FCD data research.

2.4 ETHICS

In this research, there are several considerations on ethics which need to be taken into account. The data is provided by Google and Landerlijk Model Systeem (LMS) data are provided by Rijkswaterstaat via TNO. The researcher has permission to use this data and tools for this research. The data received from Google is an anomonized set, the raw data cannot be shared as this may contain privacy sensitive information (e.g. GPS track data in combination with other geospatial data). The data must always be represented in an aggregated manner to ensure that data cannot be attributed to (several) individuals.

3 METHODOLOGY

In this chapter, the research methodology is described. First the workflow is described. Then the concept trip sustainability is introduced, which is an essential part of this research. Then the available data sources are evaluated and an initial exploration is done. Finally, the tools which are used are mentioned.

3.1 WORKFLOW

In fig 3.1 the workflow of this research is shown. It depicts the logical steps taken. It starts with the problem statement. This problem statement results in several research questions and subquestions, which this research aims to answer. From the literature review, where existing research is analyzed, a data selection process is performed to find the most suitable data available capable of answering the research quetions. From the data selection, the implementation of data analyses used are described. The results from the implementation phase are discussed, which are summarized in the conclusion. The conclusions are compared to the initial research questions to check if the questions have been answered, or if additional research is required.

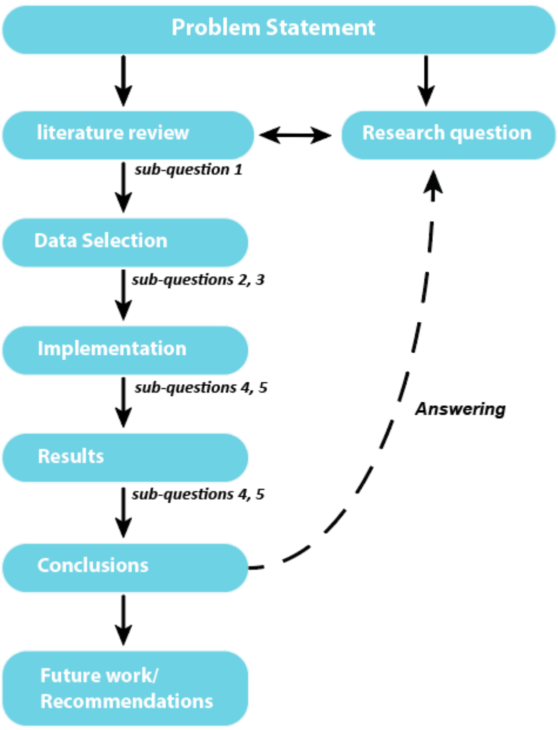


Figure 3.1: Structure of this thesis (own image)

To answer the main research question of this research, a subset of questions is setup in chapter 1.4. To find the answer to the first question, a literature study in chapter 2 is done. The second and third subquestion are done by comparing different available data sets and reading different literature studies about different data sets used in other research projects, this subject will be discussed in chapter 4. The answer to the final two subquestions is found by doing experiments with the available data. The implementation of this experiments will be described in chapter 6 and the results of these experiments will be shown in chapter 7. After all research questions are investigated, a conclusion in chapter 8 is given. In this chapter all answers to the sub questions will be analyzed, so that in the end an answer can be formulated to the main question of this research.

3.2 TRIP SUSTAINABILITY

Based on literature classification method Table 2.4, there are different possibilities to choose the most sustainable option for a trip. In this research a distinction is made between modes of transportation an distance travelled. The short trips smaller than 10 km and the long trips are representing the tips longer than 10km. In Figure 3.2 the decision tree is depicted, which is used in this research. For the short distance category both walking and biking are sustainable alternatives. Although the walking is not always seen as an alternative in this category. Because 10 km walking trip is far for daily journeys, a decision is made to make a distinction when walking is a sustainable alternative and when cycling is chosen as the most sustainable option. To do this, the travel time, that is calculated with the Google API, is used. If a walking trip is taking not more than 15 min travel time, it should be seen as a realistic alternative. Trips where origin and destination are the same, and no travel time could be calculated, will also seen as walkable trips. This assumptions is made because the areas in Amsterdam are not very large and inside an area a trip should always be a walkable distance. For long trips (> 10 km), where a train station is available in the neighborhood of the origin and destination of the trip, the train is considered to be the most sustainable option. The decision tree is explained in more detail in chapter 6.

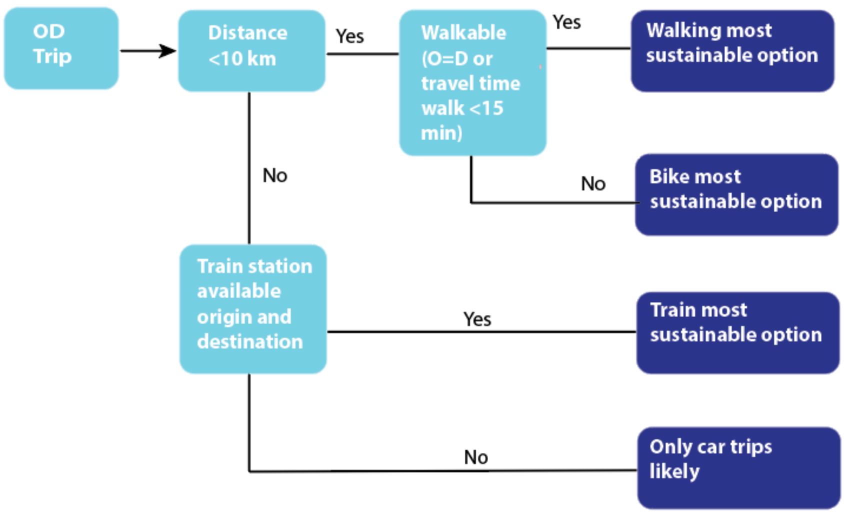


Figure 3.2: High level decision tree

3.3 AVAILABLE DATA FOR THIS RESEARCH

For this research different data sets will be explored if they can be used in the experiments. These data sets were made available by TNO and TU Delft. In chapter 4 the different available data sets will be described and analysed there usability. Different types of floating car data FCD will be analysed. Floating car data is data that is collected by smartphones, representing the travel behaviour of individuals or on a aggregate level. In addition to this floating car data, another type of aggregate origin destination data from the Landelijk Systeem Model will be analysed. To be able to do the analyses a set of requirement will be set up, which the data must meet. In the end al data sets will be compared whether they meet the requirements and the best suited dataset(s) will be chosen.

3.4 USED TOOLS

This section provides a short description of the tools used in this research. A combination of open and comercial software is used.

1. Google Cloud platform: used to get the data and query the data in the cloud. Google provided the data in this cloud format. A fast way to query a big dataset, but due to the limitations on the query posibilities, the data is exported to a PostgreSQL database.
2. PostgreSQL is an open source relational database management sytem. PostgreSQL is used to query and combine the different data sets and aggegate the data on certain levels.
3. QGIS is used to visualize the spatial data and create the images presented in this thesis
4. FME is a spatial Extraction, Transformation and Load (ETL) tool that is not open source. For this research a student licence was available to perform analysis and data transformations with FME. In the research for example it is used to tranform data into different formats, make a buffer around a train station, calculating the midpoints of the LMS areas etc.
5. Matlab is a programming language which is used to make a script to obtain the Google API travel time and distance information.
6. ESRI Insights is an extension in the Arcgis enterprise environment of ESRI software. In this research it is used to make the visualisation of the results of the experiments. The commercial license for this software was available at Tensing.
7. The Adobe creative cloud is used to make visualizations and graphs in this report. In this suite two different applications are used: Adobe Photoshop and Adobe Illustrator. A student license for these programs has been purchased for this research.
8. LaTeX is a document preparation system that is used to make the report template and layout of the report. To combine the different scripts an open source editor called TeXworks is installed. A template for the report was made available by the geomatics department of the TU Delft.

4 DATA SELECTION

As introduced in the methodology, the data selection process is one of the key topics to be addressed. There are many different data sources all with their limitations and advantages. These are related to the way the data is received, data privacy limitations and/or sample sizes. A selection of data sets is discussed in more detail in this chapter. To make a selection a set of requirements is set up. With this requirements a selection is done, and a comparison between these data set is set up. In the end two data sets are chosen to use in the case studies done by experiment in this research.

4.1 REQUIREMENTS DATA SELECTION

To select a suitable data set for this research a list of requirements has been prepared. These requirements are necessary to determine whether a data set is suitable for use in the experiments. The use cases are needed to answer the last two research sub-questions and play an important role in answering the main research question. The following requirements are set:

- The data set must be available and free of use for this research.
- The data must show origin and destination information
- The data set is sufficiently documented for the use in the experiments.
- The data should cover at least the area of Amsterdam, but preferable available for a larger area.

4.2 AVAILABLE DATA SETS

Based on the first requirement five data sets are considered in this research. This data sets are made available by TNO or by the TU Delft. In this section the different data sets will be introduced. Different types of floating car data (FCD) will be described. Floating car data is data that is collected by for example smartphones, representing the travel behavior of individuals or on a aggregate level. In addition to this floating car data, another type of aggregated data from the LMS will be introduced. The data sets considered are:

- Google
- Flitsmeister
- Ring Ring
- GPS data collected by the TU Delft
- Landelijk Model Systeem

4.2.1 Google data

This data set is collected by Google, which is a large provider of applications worldwide. Google collects data from smartphones and provided data to TNO to use this data in scientific and applied research. There are two kind of data one with origin destination (OD) information and one with intensity on street level.

Google Origin Destination table

The data is given on an aggregate level, in a way that no individual patterns can be recognized. The level of aggregation is on zone level on a request of TNO. These zones are largely the same as in the LMS network. In total, TNO requested 1030 LMS zones and one extra Non-LMS zone that is called "the rest of the Netherlands" [Bakri, 2016]. The requested areas are shown in figure 4.1.

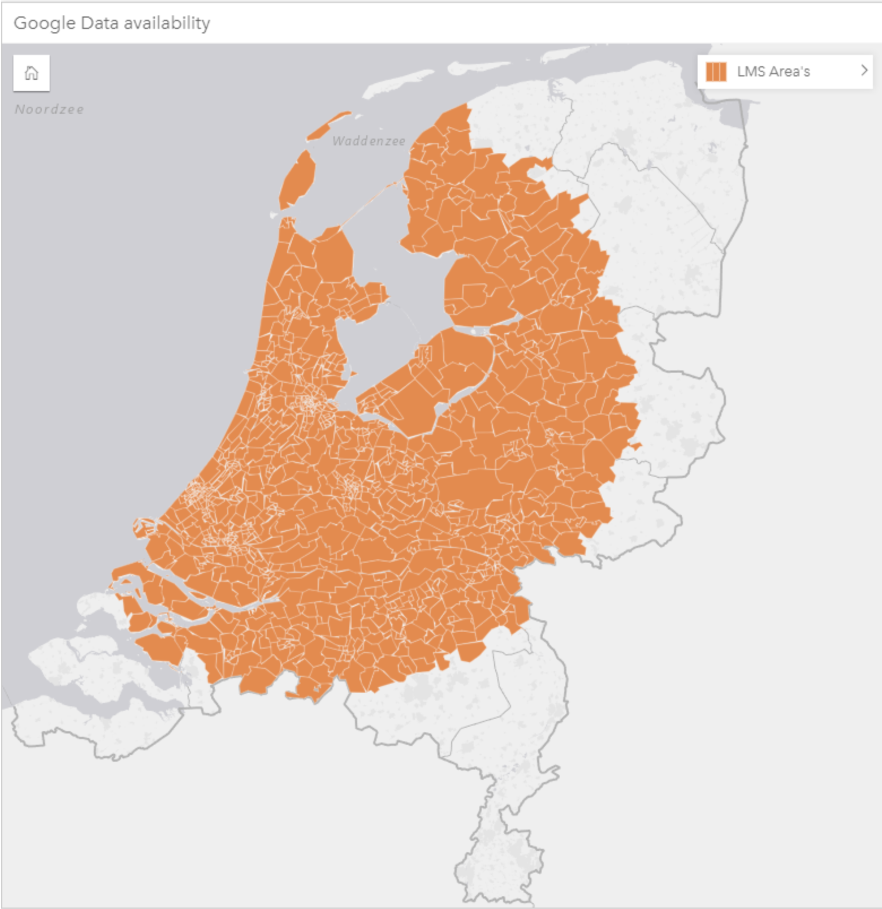


Figure 4.1: Available area's in the google orgin destination data.

After selecting traffic which has an origin or a destination in one of the neighborhoods of Amsterdam, aggregated on an hourly bases, 31.5 GB of data is available. As the data will be classified on mode of transportation, it is important that the data provide this kind of information. These different types of modality should be in the data set according to the getting started guide [van Grieken, 2016]:

- Walking: People that are walking, hiking or running.
- Vehicle: People that are driving or riding a motorcycle.
- Cycle: People who cycle.
- Public Transit: People that used a bus, tram, metro or other means of public transport.

The group public transport should be in the data set but is not visible after processing the data. This is a limitation of this data set which was discovered during the graduation proces.

Google Flow and Speed

Google flows and speed are based on the same smartphone data as described in the previous data set. However this data set gives information per road segment. It shows a relative flow level in a range of zero to ten of every road, with a link-ID to Google maps. The speed information provides two types of information, first the estimated mean speed in meters per second, second the variance of the estimated main speed in meters per second [van Grieken, 2016]. Next to the flow and speed data, a table with turn fractions is available. The inbound and outbound road ids area linked to a turn count, expressed a number of observations per hour [van Grieken, 2016]. Although this data is very valuable for analysing traffic and hotspots there is no information on the origin and destination of people, making it not suitable for analysing trips.

4.2.2 Flitsmeister

Flitsmeister is a mobile application which focuses on giving actual information about speed traps and traffic jams in the Netherlands. It collects GPS tracks of a wide group of users. These GPS tracks can be used for this research. The data is stored in binary files which can be approached by MATLAB scripts developed by TNO. In this way trips and tracks can be extracted. To extract the origins or destinations in Amsterdam a script is required to be built on top of this code. Because of the aim of this application, only car traffic is available from this data source. This car traffic information can be compared with the google data set, to see if the found results are comparable. To avoid privacy issues an aggregate data set has been requested to Flitsmeister.

4.2.3 RingRing

Ring Ring is a mobile application especially for cyclists. Ring Ring is a service that promotes the use of bicycles using bicycle miles. These bicycle miles give participants benefits for different entrepreneurs, employers, cities and health insurers. The data that is collected consists of GPS tracks where raw points are stored in binary files. This data set can be approached in the same way as the data collected by flitsmeister. Because of privacy concerns the first and last 500 meters of the trip are removed from the data set. This gives problems by generating the origin and destinations on an aggregated level. For example a trip can start in a different neighborhood, but due to this cutting of, the origin will be represented inaccurately. This real origin can be calculated by drawing a circle of 500m around the first or last point. The origin/destination of the trip should be in this circle. By comparing the surface of the circle with defined neighborhoods, an estimation can be made on how likely it is that an origin/destination is in a certain neighborhood.

4.2.4 GPS data set Tu Delft

Because this research is done at the Technical University of Delft, it was possible to look at the data that is available for research in the department of Geomatics. Several researchers and students on the faculty did research in the field of mobility [Biljecki, 2010; Bohte, 2010; van de Coevering and Maat, 2013; Gil, 2016; Michailidou, 2019; Oliveti, 2015; van der Spek et al., 2009; van Winden, 2014]. A common used data set in this studies is the GPS data set collected in 2013 by Paul van de Coevering. Different households joint this research which took place for the first time in 2007 (by Bothe) and repeated in 2013 (by van de Coevering) with the same households. The collection of the data set took place in Amersfoort, Veenendaal and Zweewolde. In addition to GPS registration, in-depth interviews have also been conducted with the participant [van de Coevering and Maat, 2013]. The data contains GPS points with coordinates (x, y, z) and time (t) , every 5 seconds by more than 800 people. The data consists of 40 million GPS points, 1,5 million made by bicycle and 3,7 million made by car, with a track length 385.000 kilometres total [van Winden, 2014].

4.2.5 Landelijk Systeem Model

The LMS together with the Nederlands Regionaal Model (NRM) are used to forecast the development of mobility in The Netherlands. It is originating from Rijkswaterstaat and is used to determine the effect of policy making on the mobility and environment topics [rijkswaterstaat, 2018]. It has been developed in the late 1980's and developed further with the NRM system in the 1990's. The LMS is updated yearly with new input data, regarding public transport, roadway network, social economic developments, policy starting points. Every four years, the basis year is updated. This year forms the basis on which the prognoses are made for the year which is considered [rijkswaterstaat, 2017]. The most current basis year during this research is 2014.
The model is based on three pillars (see figure 4.2), the input data, the calculation and output data. For the input, different data sets are used, such as zonal data, features of the roadway system, train network features, station relation matrix and the basis matrices for cars and trucks. For the calculation part, there are different modules in the software (population module, mobility module, foreign traffic module, growth factor module and the multiplication module). As an output the prognosis matrix for cars, and other matrices are given.

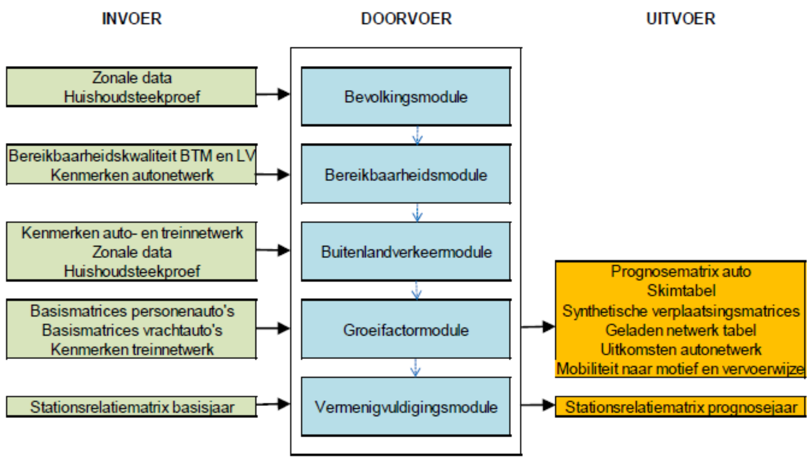


Figure 4.2: Overview three pillars, the input data, the calculation and output data LMS/NRM [L.Tavasszy et al., 2012]

The LMS base year OD matrix for cars is used in this research. This matrix is constructed with trip generation, destination choice, mode choice and departure time choice models amongst others. The model parameters are estimated based on a large scale survey called OViN [CBSl, 2015]. The resulting car matrix is calibrated based on traffic counts of different road sections.

4.3 SELECTION OF DATA SETS

Now the selection of data described in section 4.2 is made based on the first requirement, it is time to review all other requirements listed in section 4.1. In this section all data sets will be reviewed per requirement.

4.3.1 Origin and destination information

To detect if a car trip could be replaced by a more sustainable alternative, it is important to know the orgin and detination information of the car trip. Because of privacy concerns this can be not always the case. In this section all data sets will be listed.

Google (Flow- and OD data set)

As described in the previous section the Google data set contain two different parts. The origin/destination matrices and the Google flow data set. The origin and destination information in the first data set are distributed on zone level. This gives a clear area where the trip is started and ended. The Google origin destination matrices does meet this requirement. The flow data set is related to the road segments and does not contain origin and destination information. Therefore it does not meet the requirement.

Flitsmeister

This data set contains GPS tracks from individual users of the application. After analyzing a subset of the data, question marks have have been raised to whether this data embraces the entire trip. When the track was plotted on the map, it often appeared to start or end on a highway. There is a suspicion that the beginning and end parts of it, are cut off for privacy reasons or that users turn on the application on the highway themselves. This is also confirmed in the privacy statement, on the Flitsmeister website. It states that at least 500 m from the ride is cut off and the start location after this action is not easily traceable [Flitsmeister, 2020]. As a result, the correct origin and destination information may be missing.

RingRing

Similar to Flitsmeister, individual GPS tracks are visible, but due to privacy concerns the first and last part of the trip is cut off. In the privacy statement of Ring Ring it is noted that randomly 100 to 500 m is cut off of the trips [RingRing, 2020]. The influence of this is more significant for cycling trips than for car trips, because it mostly contain shorter trips. Orgin and destination can hardly be related by this operation that has been done.

GPS data set TU Delft

This data set contains simular to Flitmeister and Ring Ring GPS tracks of individuals. The main difference is the way it has been collected. In contrast to the other two this data set is not passivly collected. Participants choose to share their gps tracks for research purposes. With this in mind, it is possible to show the full track, with origin and destination information.

Landelijk Model Systeem

Simular to the Google matrices, origin and destination is available on a aggragate level. Therefore it will meet the requirement. The Google data set and the LMS data set are aggregated on the same areas. This makes a comparison possible.

4.3.2 Documentation data sets

For alll data sets, documentation is available. The amount of documentation differs per data set. The LMS model is more intensely described in the literature than newer data set as Flitsmeister and Ring Ring.

4.3.3 Availability in the research area

All data sets except the GPS tracks of the TU Delft are available in the Area of Amsterdam. The TU Delft data set is available in Amersfoort, Veenendaal and Zweewolde. This means that this data set is not suitable unless the reseach area is changed. The Google data set is available in a circle of 200km around Amsterdam. This mean that almost all of the Netherlands is available. The LMS data set covers all of the Netherlands and is the most complete in this requirement. Flitsmeister and Ring Ring are based on individual tracks, that could be everywhere the Netherlands. The distribution of the track is not reviewed. It is known from the Ring Ring dataset that there are many tracks in the IJburg area in Amsterdam due to a pilot that started in that neighborhood. The distribution of the trips can possible lead to misleading conclusions.

4.3.4 Potential use of the data sets

After analysing the data set based on the requirements, it is time to compare the data sets on their potential use, and whether this can be combined with the use in the case study in this research. In table an overview can be founded of the data sources, the type of data and usability for sustainability research.

Data set	Type	Examples of use in sustainability research
Flitsmeister	Individual GPS tracks Lost of origin- destination information Individual routes visable	Analyse routes and heavy traffic. Identify potential risks for air- and/or noise polution due to car traffic.
Ring Ring	Individual GPS tracks Lost of origin- destination information Individual routes visable	Analyse cycle routes. Analyze the use of cycle paths. Analyse the willingness to use the bicycle. analyze which bicycle routes are more appropriate.
Google Flow	Data linked to road segments Speed per road segment during the day	identify heavy traffic. Identify potential risks for air- and/or noise polution due to car traffic. detect traffic delays and traffic jams.
Google OD	Information of travel patterns between different origins and destinations.	Analyze travel behaviour Differences in use of modality in different area's.
LMS	Information of travelpatterns between different origins and destinations.	Analayse travel behaviour Differences in use of modality in different area's
TU Delft	Individual GPS tracks Origin- destination information Individual routes visable Information of travel patterns between different origins and destinations. questionairs provide extra information.	Analyze travel behaviour Differences in use of modality in different area's.

Table 4.1: Overview of the data and the potential use in sustainability research

Flitsmeister and Ring Ring both provide individual GPS tracks, but as they cut off the start and end of every trip, the origin and destination information is lost. Analyses on travel behaviour can be a challenge. The benefits of both data sources for sustainability research lies in the ability to identify areas with high and low intensities, both for car (Filtsmeister) and cycling (Ring Ring) traffic. Google flow data can be used to show peak traffic times on different roads, therefore analyzing areas where traffic can be responsible for air and noise polution. The potential use of the last two data sets, Google OD and LMS, are the most suited to use in the case studies. The Google OD data set provides information on travel patterns between different origins and destinations. LMS also provides information on the travel behaviour through origin destination information. Both data sets use the same aggregation level, making a comparison very interesting. The TU Delft data set is very interesting to analyse traffic behavour because it has origin and destination information, attached with information from questionairs.

4.4 COMPARISON BASED ON THE REQUIREMENT AND SELECTION OF THE DATA SET

Based on the list of requirements the best suited data can be choosen. In table the result of analysing the different data sets is shown. Every data set is analysed on every individual requirement.

Requirement	Flitsmeister	Ring Ring	TU Delft GPS	Google Flow	Google OD	LMS
Availability	v	v	v	v	v	v
OD information	x	x	v	x	v	v
documentation	v	v	v	v	v	v
research area	v	v	x	v	v	v
Useability	x	x	v	x	v	v

Table 4.2: requirements

Based on this list of requirements, the Google OD data set and the LMS data set are choosen to be futher analysed in this research. These data sets meet all the requirements set for this research. Both data sets will be used in the case studies, and the results will be compared.

5 DATA PREPERATION

In the previous chapter the focus was to answer the 2nd and 3rd subquestion of this research, which is resulting in two selected data sets based on different requirements. This data sets will be used in the experiments that are described in the coming chapters. To use this data some preperation is needed to compare the data sets in the experiments. In this chapter the two data sets are desribed in more detail and the necessary preperation is explained.

5.1 GOOGLE DATA SET

The Google data is provided via the Google Cloud and can be entered though Google BigQuery. This cloud based platform has a good performance for large data sets, which made it easy to query results fast. During the process some limitations came up. Some rows in the data set where provided as strings instead of integers (origins and destinations). It was not possible to make this kind of changes in the cloud and also some other limitations to work directly from BigQuery came up. Due to this limitations it has been decided to export the data related to Amsterdam using the Google SDK program. This resulted in multiple CSVs that needed to be joined. The resulting CSV file was entered to PostgreSQL, where the data could further be analysed.

Table 5.1 shows the columns in the data set. Google has provided OD data for the period from July 1 through December 31 2015. This data is delivered at one hour resolution and contains normalized counts, called weights. This means that all observations are divided by the largest observation for this period. The weights will therefore all have to be between 0 and 1. This has been tested and is correct [Bakri, 2016].

Column name	**Data information**
Origin	Origin of trip , number corresponding to the LMS areas
Destination	Destination of trip , number corresponding to the LMS areas
Weight	Weight compared to the highest trip, highest trip = 1
start_timeinterval	Unix in epoch starttime interval
end_timeinterval	Unix in epoch endtime interval
Modality	’in passenger vehicle’, ’cycling’ or ’walking’

Table 5.1: This is the information provided in the Google data set

5.1.1 Characteristics of the data set

In the documentation of the Google data set, different characteristics are described [van Grieken, 2016]. The specifications that are described in the documentation are stated in Table 5.2.

Characteristic	Specifications
Group	Unknown group of users of the google application.
Time	Data every hour for several months, weekdays and weekend
Modality	Different modalities: walking, cycling and in passenger vehicle
Values	No absolute values of travelers, weights compared to the highest value

Table 5.2: Specifications of the google data set

As desribed in section 4.2.1 and Table 5.2, Google should have the following modalities: walking, vehicle, cycling and public transport [van Grieken, 2016]. However, after processing and analyzing the data, three different mode of transportations came up that are different from the description of the guide provided by Google. These types are:

- Walking: people that are walking, hiking or running.
- Cycling: people who cycle.
- In passenger vehicle: people that are driving a car or riding a motorcycle.

5.1.2 Available area of the data set

After processing the data set, all available areas are plotted on the map. The available areas are shown in figure 5.1. In total there are 1030 LMS zones available and one extra Non-LMS zone that is called "the rest of the Netherlands". In the data set this area is indicated as LMS zone 0.

Figure 5.1: Availability of the Google data set

5.2 LMS DATA SET

From Rijkswaterstaat different LMS data sets are made available for this research. The following files and tables were provided 5.3:

LMS_file	transportation means	motive	daypart	year
TRP9_114_Hoog_LMS_uur	N/A	freight	morning rush hour	2014
TRP9_214_Hoog_LMS_uur	N/A	freight	rest day	2014
TRP9_314_Hoog_LMS_uur	N/A	freight	evening rush hour	2014
TRP22114_Hoog_LMS_uur	car driver	home-work	morning rush hour	2014
TRP22214_Hoog_LMS_uur	car driver	home-work	rest day	2014
TRP22314_Hoog_LMS_uur	car driver	home-work	evening rush hour	2014
TRP23114_Hoog_LMS_uur	car driver	home-business	morning rush hour	2014
TRP23214_Hoog_LMS_uur	car driver	home-business	rest day	2014
TRP231314_Hoog_LMS_uur	car driver	home-business	evening rush hour	2014
TRP25114_Hoog_LMS_uur	car driver	home-other	morning rush hour	2014
TRP25214_Hoog_LMS_uur	car driver	home-other	rest day	2014
TRP25314_Hoog_LMS_uur	car driver	home-other	evening rush hour	2014

Table 5.3: The table indicates the different files provided for the LMS, where the first three files were not used as freight traffic is not in the scope of this research.

The data sets have different subsets. A distinction can be made for cars and trucks, trip purpose and time of the day. Different from the Google data set, the LMS data set has only 3 types of time differentiation: morning rush hour, evening rush hour and rest day. It gives an average of different days, so that differences on particular days cannot be visualized.

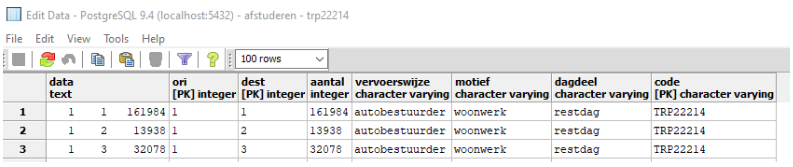
5.2.1 Processing the files

The files from Rijkswaterstaat are delivered in ASC format. Every datafile contains only information about origin, destination and count. To import these files in sql, first a table is created with one column data. After that it was possible to import the ASC file in the Postgres database in PGAdmin. All data was imported in one data column. For further use, it is recommended to spit the origin, destination and count data, in a way that the columns can be queried individually. With the following sql statement, shown in Figure 5.2, is was possible to split the data in separate columns.

```
create table TRP25314 as
select *,substring(data, 1 ,4)::int as ori,substring(data,5,7)::int as dest ,
substr(data,10)::int as aantal from lmsimportasc
```

Figure 5.2: SQL query to split the original LMS data in separate rows for further analysis

To get a full overview of the data set some extra columns are added with the information of the file. Information about the modality, travel purpose and time of the day are added. This information was hidden in the file name, and is important to keep when the data is merged in a later stage. The example of the table is shown in Figure 5.3 .


Edit Data - PostgreSQL 9.4 (localhost:5432) - afstuderen - trp22214

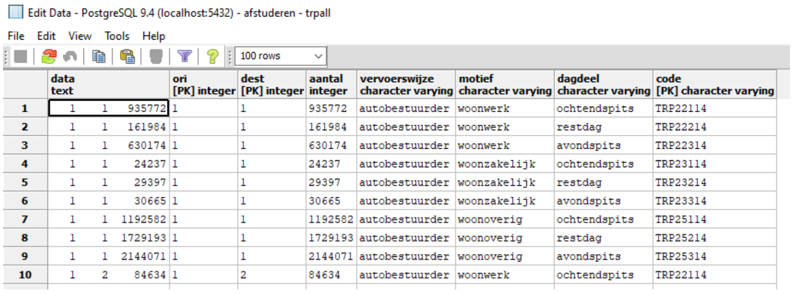
File Edit View Tools Help

100 rows

	data text	ori [PK] integer	dest [PK] integer	aantal integer	vervoerswijze character varying	motief character varying	dagdeel character varying	code [PK] character varying
1	1 1 161984	1	1	161984	autobestuurder	woonwerk	restdag	TRP22214
2	1 2 13938	1	2	13938	autobestuurder	woonwerk	restdag	TRP22214
3	1 3 32078	1	3	32078	autobestuurder	woonwerk	restdag	TRP22214

Figure 5.3: Example of LMS data set, after importing the TRP22214 file in the database and adding rows

After processing all the different files, all tables are combined in one big table as shown in Figure 5.4. The data column is the origininal information and the others are added. In this table only the transportation mean car driver is seleted, since truck drivers are out the scope of this research.


Edit Data - PostgreSQL 9.4 (localhost:5432) - afstuderen - trpall

File Edit View Tools Help

100 rows

	data text	ori [PK] integer	dest [PK] integer	aantal integer	vervoerswijze character varying	motief character varying	dagdeel character varying	code [PK] character varying
1	1 1 935772	1	1	935772	autobestuurder	woonwerk	ochtendspits	TRP22114
2	1 1 161984	1	1	161984	autobestuurder	woonwerk	restdag	TRP22214
3	1 1 630174	1	1	630174	autobestuurder	woonwerk	avondspits	TRP22314
4	1 1 24237	1	1	24237	autobestuurder	woonzakelijk	ochtendspits	TRP23114
5	1 1 29397	1	1	29397	autobestuurder	woonzakelijk	restdag	TRP23214
6	1 1 30665	1	1	30665	autobestuurder	woonzakelijk	avondspits	TRP23314
7	1 1 1192582	1	1	1192582	autobestuurder	woonoverig	ochtendspits	TRP25114
8	1 1 1729193	1	1	1729193	autobestuurder	woonoverig	restdag	TRP25214
9	1 1 2144071	1	1	2144071	autobestuurder	woonoverig	avondspits	TRP25314
10	1 2 84634	1	2	84634	autobestuurder	woonwerk	ochtendspits	TRP22114

Figure 5.4: Example of LMS data set, after importing the TRP22214 file in the database and adding rows

5.2.2 Characteristics of the data set

Just like the google data set, the characteristics about the usage of groups, time, modality and values are given in an overview. The characteristics of the LMS data is provided in Table 5.4 :

Characteristic	Specifications
Group	Motivation of users are known: Home-work, home-business, home-other Focus on people who live in the area.
Time	Only an avarage weekday. Three catagories : rush-hour morning, rush hour evening , rest of the day. Time of the rush hour is unknown, use peaks in the data to calculate, no exact time.
Modality	Car drivers and truks
Values	Every Origin Destination has an absolute amount of vehicles.

Table 5.4: Specifications of lms od matrices

The time catagories that are specified in the data set rush-hour morning, rush hour evening and rest of the day are not extensively explained in the literature. This makes it hard to relate to exact times.

5.2.3 Available area of the data set

After processing the data, the data set appears to be available for the whole of the Netherlands and some neighboring countries. Figure 5.5 shows a map of the available areas.

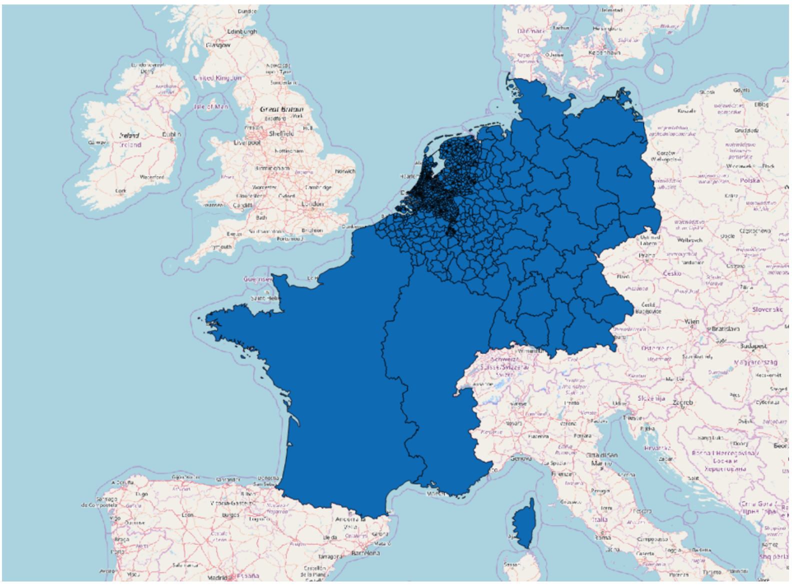


Figure 5.5: Availability of the LMS data set

5.3 HARMONIZATION OF THE TWO DATA SETS

Although there are a lot of similarities between the two data set, also some differences have to overcome. The following items should be aggregate in te same way:

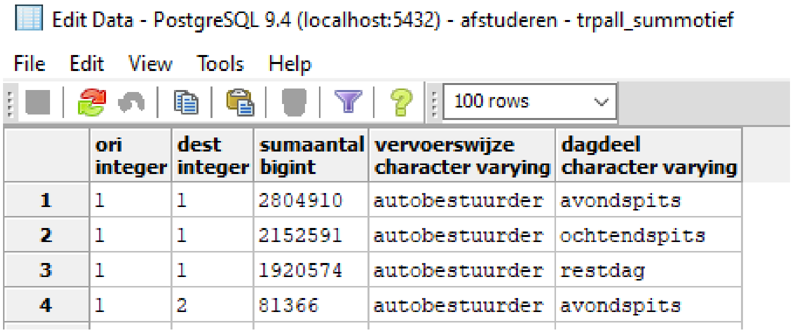
- *Area*: The LMS data set has far more areas data available than the Google data set, which has only in a circle around Amsterdam. While the number of areas varies, the corresponding areas do have the same shape and size.
- *Purpose*: The LMS data set provides the purpose: home-work, home-business, home-other, the google data set does not provide this information.
- *Time*: The Google data set provides information for every hour in unix epoch time, in a particular period. The LMS data only provides information of an average day.
- *Research area*: Select only the trips that have an origin or destination in Amsterdam.
- *Counts*: The Google data set contains normalized counts, called weights. While the LMS data set provide absolute values.

5.3.1 Harmonization Area

Since the LMS data set and Google did not have the same areas available, a selection needs to be made from the LMS data set. All the available areas in Google are selected in the LMS data set, and areas that where not in the Google data set are removed from the LMS data set. The area 'zero' in the Google data set is removed from the google data set because it was not useful for this research.

5.3.2 Harmonization Motive

This step is only done in the LMS data set, to remove the motives. An agreggation is made, where all the different counts for every origin destination combination are summed. This creates a total of all trips per time catagory. The result of this operation is shown in figure 5.6.


Edit Data - PostgreSQL 9.4 (localhost:5432) - afstuderen - trpall_summotief

File Edit View Tools Help

100 rows

	ori integer	dest integer	sumaantal bigint	vervoerswijze character varying	dagdeel character varying
1	1	1	2804910	autobestuurder	avondspits
2	1	1	2152591	autobestuurder	ochtendspits
3	1	1	1920574	autobestuurder	restdag
4	1	2	81366	autobestuurder	avondspits

Figure 5.6: Harmonization motive of the LMS data set

5.3.3 Harmonization time

As the LMS data only provides information for an average day, this means that the Google data needs to be converted from hourly to an average day. Firstthe unix epoch time is converted to the Coordinated Universal Time (UTC). From that timestamp 2 different columns were made:

- *Day of the week*: Shows the days of the week corresponding to the timestamp. For example: Monday or Sunday.
- *Hour of the day*: Shows the start hour of the orgin destination combination.

With this information an aggregation on time is possible. In the LMS data set only contains only work days, so all records with the value Saturday or Sunday could be deleted from the Goolgle data set. An aggregation is then made, resulting in one value for each combination of origin destination with an added weight.

The LMS data set is therefore aggregated, because there are three different time indications. The result is one row for each combination of origin destination for an average day.

5.3.4 Harmonization research area

For the Google data set a selection of the origin and destination that have a relationship with Amsterdam is made with the export form of the Google Query table. The LMS data set on the other hand, still contains all trips throughout the Netherlands. A filter has been applied, so all trips that are not connected to the research area are removed from the database.

5.3.5 Harmonization counts

The two available data sets (Google and LMS) are made comparable using the heighest weight for the car trips available in both data sets and normalizing all weights to correspond to this value. This means that for both the LMS and the Google data, the connection with the highest amount of car trips has a weight of one. Using this method, the LMS and Google data can be compared.

$$x_{\text{new}} = \frac{x_{\text{old}} - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{5.1}$$

The formula for normalisation [Freedman et al., 2007] is used in both data sets to normalized the values of the counts and weights in the data sets. The formulas for this two data sets are shown in 5.2 and 5.3.

$$\text{LMS}_{\text{new}} = \frac{\text{LMS}_{\text{old}} - \text{LMS}_{\text{min}}}{\text{LMS}_{\text{max}} - \text{LMS}_{\text{min}}} \tag{5.2}$$

$$\text{Google}_{\text{new}} = \frac{\text{Google}_{\text{old}} - \text{Google}_{\text{min}}}{\text{Google}_{\text{max}} - \text{Google}_{\text{min}}} \tag{5.3}$$

After this step, the new weights need to be calculated for both data sets. For the Google data there are 3 tables. The car table is used to calculate the highest value and this is set to 1 (data is normalized to the highest value of the car trip). The walk and cycling values are re-calculated according to the car normalization.

5.4 CORRELATION BETWEEN THE TWO HARMONIZED DATA SETS

To determine the comparibility of both data sets, the correlation between the LMS and the Google data sets will be determined using the correlation formula [Freedman et al., 2007]:

$$C = \frac{\sum_i^N (G_i - \overline{G})(L_i - \overline{L})/(N-1)}{\sigma(G)\sigma(L)} \tag{5.4}$$

where G and L are the Google and LMS weights, the bar represents the average value and the σ is the standard deviation. The total number of connections in the data set is represented by N. The correlation is scaled between zero and one.

To determine the validity of the data sets, the correlation between the LMS and Google data sets is calculated. Using the ares in scope (Amsterdam), for every short distance connection the correlation coefficient is 0.5, meaning that there is a moderate correlation between the two data sets. In chapter 6 and 7 the correlation between these data sets will be furter analysed in different cases during the experiments. Also a small experiment is done to analyse the top 20 highest car weights in the Google and LMS data set. The analysis shows that there are similar origin destination combinations in both data sets. The results are shown in Appendix A.

5.5 DATA SET LIMITATIONS

There are also some limitations to the Google and LMS data sets. Not all trips are shown in the data, for example public transport is not available in these data sets, see Figure 5.7. This makes the data more suitable for short distances, as walking an cycling are an option for these trips. For longer distances, these options become less relevant and more data is required to get meaningful results.

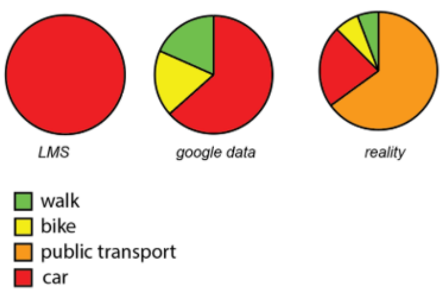


Figure 5.7: Comparison of the amount of trips for the the two data sets (LMS and Google data) and the possible reality where are all trips are considered (Not based on real number).

6 IMPLEMENTATION

Various sub-questions have been prepared in this study. To answer sub-questions 4 and 5 it is necessary to set up experiments. The design of these experiments is described in this chapter. The data sets, chosen in Chapter 4 are the basis for the experiments, with the aim of answering sub-questions 4 and 5. This chapter describes the implementation and data analysis of the chosen data sets. The goal of these experiments is to analyze which car trips could be replaced by a more sustainable alternative.

6.1 IMPLEMENTATION OF THE EXPERIMENTS

A decision tree is introduced, based upon which trips are categorized in short and longer trips. The basis for this decision tree originated on the literature study in Chapter 2. The decision tree reflects on the most sustainable option, which is shown in Figure 2.1 in Chapter 2. For this research, short and medium journeys are combined. Because the duration of the trips is also calculated, an extra dimension is added to make a decision on the most sustainable option. The trips where the duration of walking is less than 15 minutes, walking should be seen as most sustainable option for short trips. For the remaining trips in the short (and medium) category, cycling is considered the most sustainable. This is because cycling is seen as more sustainable than public transport according to the literature Chapter 2 shown in Table 2.1. For all short trips, public transport should be seen as second or third most sustainable alternative to car.

For longer trips, with a distance longer than 10 km, public transport is the most sustainable option. In the case study in Amsterdam it is likely that a long distance trip is made by train. The accessibility of an area can be different, this depends whether a train station is close by the origin and destination. The method on how the distances are calculated are explained in the following sections.

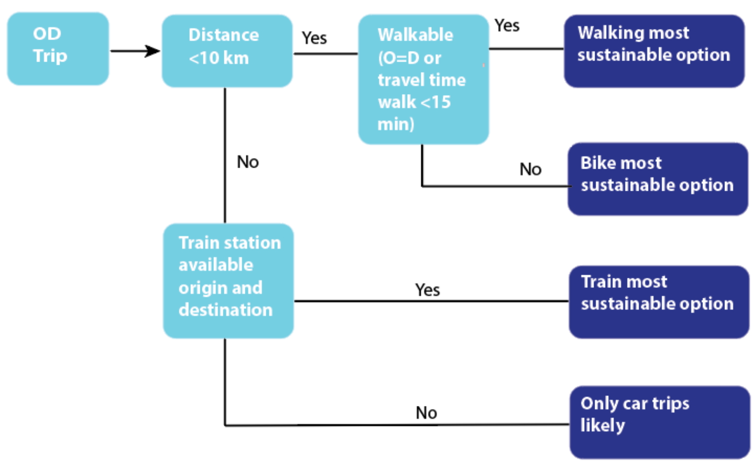


Figure 6.1: Decision tree for the most sustainable option depending on the trip distance, travel time and availability

To be able divide the data in long and short trips, the method to calculate distances is described. Last, two separate sections will describe the experiments for short (sub-question 4) and long distances (sub-question 5).

6.2 DISTANCE CALCULATION

To calculate the distances between different areas, the midpoints of the polygons (representing the zones/areas) are used, because the data on individual trips is not provided due to privacy reasons, and only the aggregated data is available.

There are different methods and tools to calculate the midpoint of a polygon. QGIS and FME are frequent used tools for this. For this research, FME is chosen to calculate the midpoints as this provides the most features. For example, one of the challenges is that when the center of gravity is chosen, the midpoint can be outside of the polygon (e.g. L-shaped polygons), FME has a feature to solve this issue.

Figure 6.2 shows the FME script used to calculate the midpoints from the shapefile.

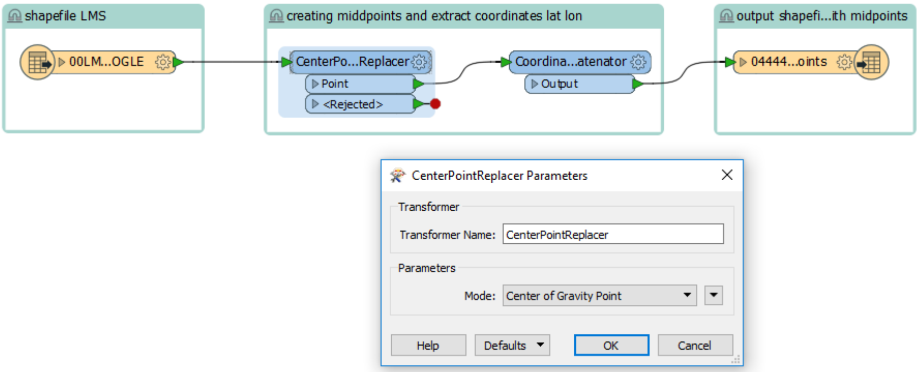


Figure 6.2: FME script used to calculate midpoints

The output is shown in figure 6.3, where all midpoints of the areas and zones used in this research are calculated and indicated by an orange circle.

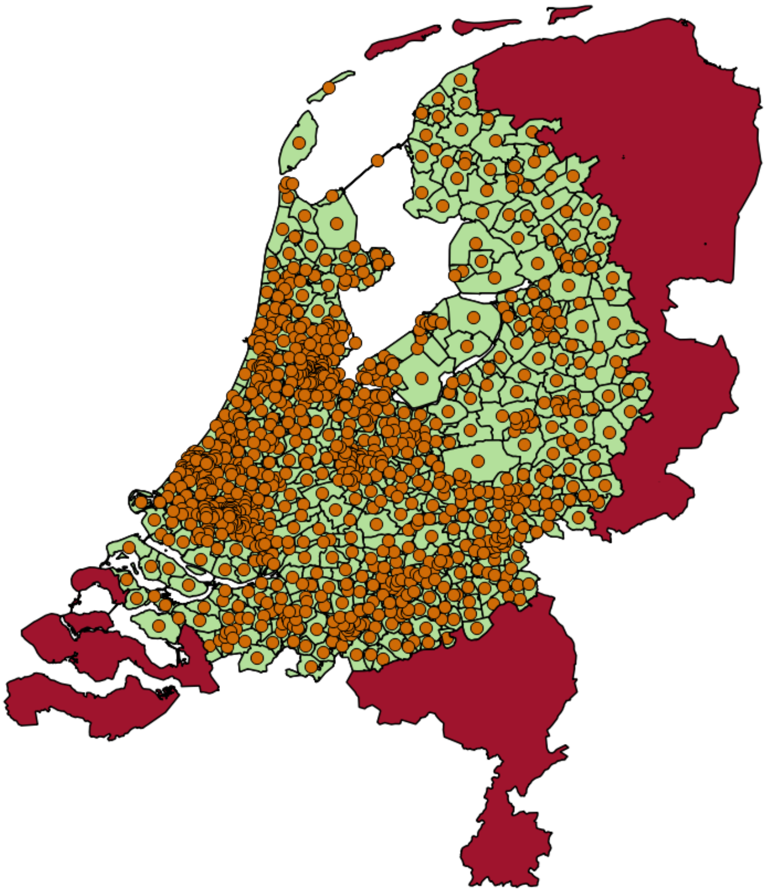


Figure 6.3: Map showing the midpoints (orange circles)

6.2.1 Google API

To determine the most sustainable option for each origin destination combination, all distances between origins and destinations need to be calculated. One approach could be to calculate the distance as a straight line between the origin and destination, but this is not a realistic situation as people need to use roads and infrastructure to travel. For this reason the Google API is used. This is an API from Google, where you can provide the origin and destination coordinates and the API will return the distance calculated by the Google Maps engine for each transportation method. With this API it is also possible to calculate the duration of the trip. To retrieve this data, a MATLAB script has been made to call this API and write the data in a csv file. The csv file is then imported into the SQL database, where it is merged with the other origin destinations information such as the Google and LMS data set. After calculating the distances, the origin and destination combinations can be divided into short and long trips. The duration of the trips is also used in the experiments to analyse the data. The following sections describe the experiments for short and long distances.

6.3 SHORT DISTANCES

After the calculation of the distances and durations, the short distance trips are analyzed further in order to answer the 4th sub-question of this research. Short trips (< 10 km) can be done by car, cycle, walk or public transport. For this research it is chosen to investigate the car, cycle and walk trips. This is due to the available data. To determine the most sustainable option for short distances, a specific decision tree is created and shown in figure 6.4. This decision tree is more specific than the overall decision tree. It assumes that when both the origin and destination are in the same zone, biking and walking are the most sustainable options. For longer distances where the walking time is more that 15 minutes, biking is considered as the most sustainable option when the time difference between biking an taking the car is less than 15 minutes. In other cases, taking the car is considered as the most logical option. Taking public transport into account for short distances is complex and no Google or LMS data is available, so this transportation means is not considered for the short distances.

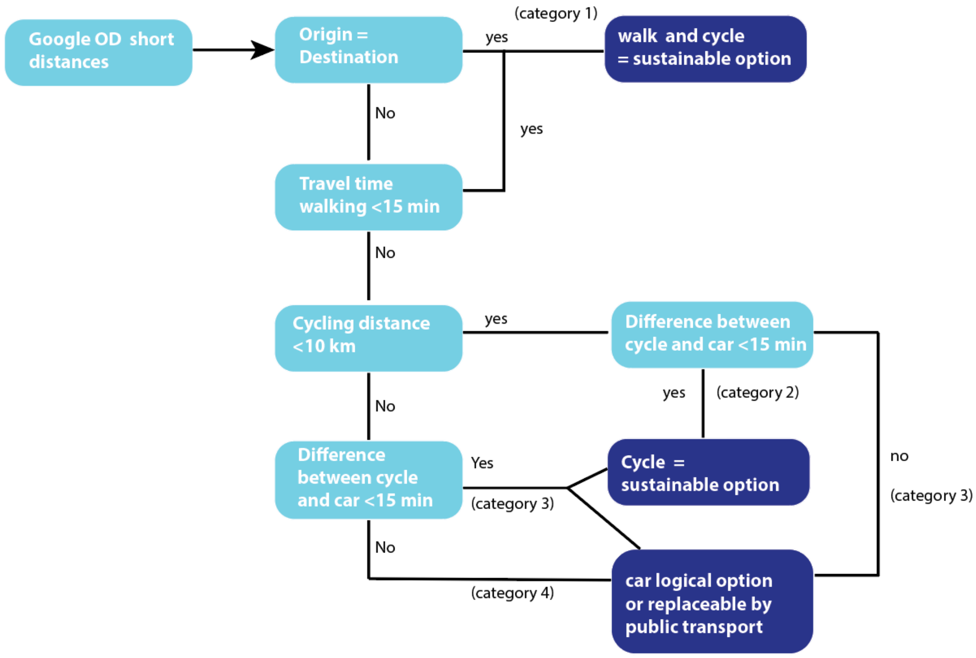


Figure 6.4: Short distance decision tree

Based on the decision tree and the traveled time and distance, calculated with the Google API, a potential of replacement is calculated. The following categories are represent:

- Catagory 1: Easily replaceable by walk and cycle
- Catagory 2: Easily replaceable by cycle
- Catagory 3: Maybe replaceable by cycle or public transport (public transport outside the scope of this research)
- Catagory 4: Not replaceable by walk or cycle (public transport potential but outside the scope of this research)

These catagories will be assigned to every origin destination combination and is added to the sql tabels with the data, so that it could be compared to the weights to see if the trips are replaceable.

Focussing on the region under investigation (the Amsterdam area), a buffer around Amsterdam is shown in Figure 6.5. This area is seen as the research area for short distance trips from Amsterdam.

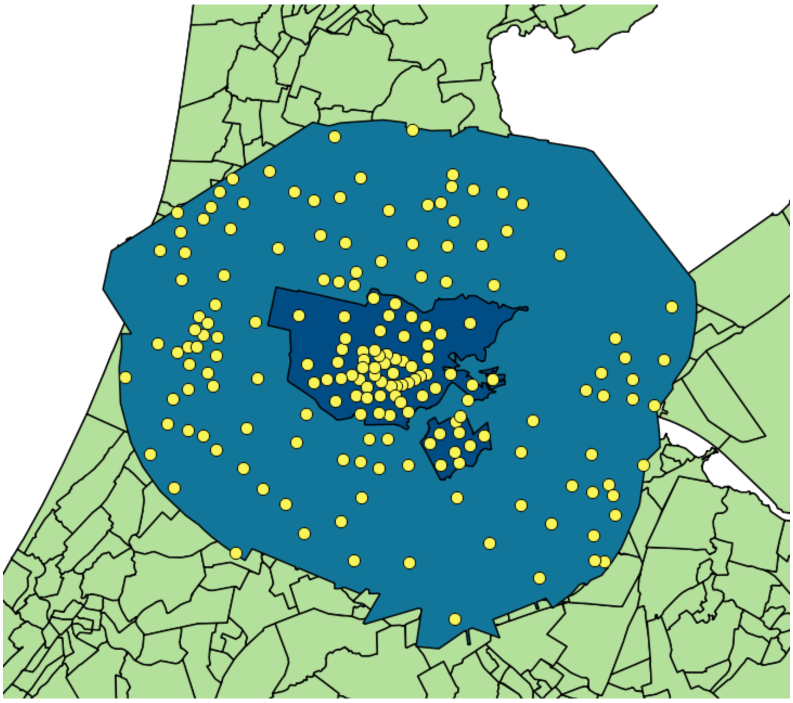


Figure 6.5: Buffer around Amsterdam to select the midpoints of the areas in the short distance data set

The Google distance API is run for the different options as shown in a example in figure 6.6. The walk, cycle and car duration and distance are calculated by the Google API from an origin and/or destination in Amsterdam. In this example it is shown that cycling is the most sustainable option. It has a similar duration (18 or 19 min so 1 min difference) and walking is really too far and too long compared to cycle and car. For every combination in the research area for short distances, these analyses are done and the most sustainable is chosen for every trip. This information is added to the sql tables and form the base for the results in the experiments.

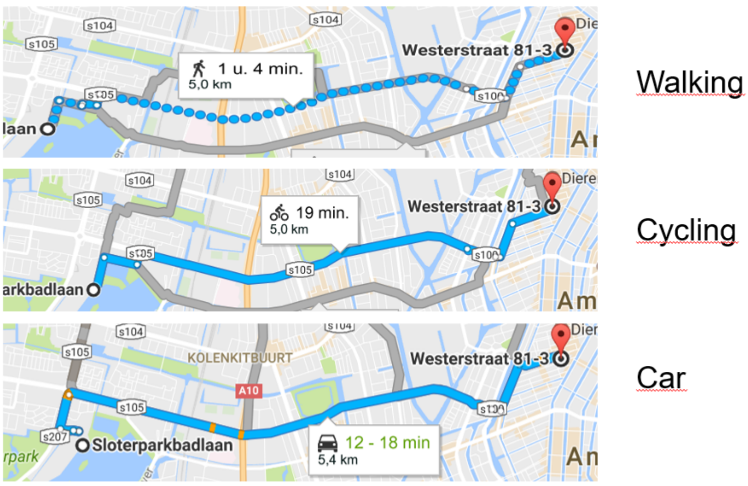


Figure 6.6: Use of the Google Api for one single short distance trip with different modes of transportation

6.3.1 Analyses of non car data

Because the Google data includes walking and cycling data it is interesting to see what these data sets looks like. First the walking trips are analysed. It is expected that walking trips with for example the same origin as destination have higher weights than longer trips. This will be analysed and the results are visualised. In the beginning of this chapter different catagories are described in which term a trip is replaceable. The first catagory is easy replaceable by walking. This catagory is used to see if this origin destination combinations have heigher walking trips. For cycle all different catagories as described in the beginning of this capter are analysed and compared with the weights in the cycle data set. Also the travel time, calculated with the Google API, is compared to the car travel time.

The walking and cycling data is used to see if the replaceable origin destination combinations calculated as result from the catagories, are corresponding to the amount of trips made with these modalities.

6.3.2 Favorite in google

To compare the different modalities a new colum 'favorite' is added. This colum compares the values of the weight colums from walking, cycle and car data in the Google data set. This is done by calculating the percentage for each modality compared to the sum of all weights in the google data set for that particular origin destination. In figure 6.7 an example is shown of the colums that are added to the data set. The following catagories are assigned in the favorite table.

- All: if all modalities differ at most 1 percent from each other
- Car: if the share of the car trips is the highest
- Walk: if the share of the walking trips is the highest
- Cycle: if the share of the cycle trips is the highest

	origin integer	destination integer	google_car_weight numeric	google_cycle_weight double precision	google_walk_weight double precision	sumgoogleweights double precision	percentage_googlecar double precision	percentage_googlecycle double precision	percentage_googlewalk double precision	favorite character varying	favorite2 numeric
421	615	562	0.0399447475230047	0.0398895859688451	0.0398896007080714	0.119723934199921	33.3640452011065	33.3179712439415	33.3179835549521	all	4
422	629	507	0.0883864034867332	0.0794483008742408	0.0794483549180703	0.247283059279044	35.7430079296271	32.1284851076628	32.1285069627101	car	1
423	638	639	0.103283280478824	0.133021614972386	0.17075957606603	0.40706447151724	25.3727081840032	32.678267026493	41.9490247895038	walk	3
424	648	597	0.0733242408117178	0.0672552071937807	0.0672552170199315	0.20783466502543	35.2800822724862	32.3599564998224	32.3599612276914	car	1

Figure 6.7: Added colums to show the favorite modality in google data set

First the catagories car, walk and cycle are assigned by calculating if the percentages is higher than the others. The category all is later added because it happened that in some situations the percentages where almost the same in all catagories. For this cases the catagory 'all' is introduced. To calculate this catagory the following sql query is used as shown in figure 6.8.

```
UPDATE tableshortdistance
SET favorite  = 'all'
WHERE  (percentage_googlecar-percentage_googlewalk) between -1 and 1
and (percentage_googlecycle-percentage_googlewalk) between -1 and 1
```

Figure 6.8: SQL query to calculate the 'all' category

For this catagory 'all' the difference between all the percentages of the modalities are less than 1 percent.

6.3.3 Analysis replaceable car trips

To answer the 4th subquestion as desribed in Section 1.4, analyses of the short car trips are needed. In the end an answer is given on the following sub question:

- Which short distance car trips in Amsterdam could be replaced by more sustainable opportunities like walking or cycling ?

First the correlation between the Google and LMS data set as described in Section 5.4 will be calculated. This is showing if the data set are corresponding to each other on the short distance trips. This will als be done for the trips that are classified as replaceable according to the decision tree. After this comparison it is time to look deeper at the catagories of replaceability as described in the beginning of this chapter. For every catagory the amount of trips will be shown in the short data set as well as the percentage of the total of all short trips. The result is the amount of replaceable trips in the short data set, which later can be compared with the weights in the Google and LMS data set. First the Google favorite column is compared to the replaceable categories. To see how many replaceable trips have the sustainable alternative as favorite or not. To see the difference in the both data sets, analyses on both Google and LMS are done. The visualisation of these maps and graphs are made with ESRI insights. This program can visualize the data in an interactive way. In the results the screenshots of these maps and graphs are shown.

6.4 LONG DISTANCES

In this chapter the implementation for long distances trips (> 10 km) is described. In this category, the train is a more sustainable option than taking the car. However, not for all trips, the train is the logical and most sustainable option. For example when there is no train station near the origin and destination, then the car is the most logical option. This is further explained in the decision tree in figure 6.9.

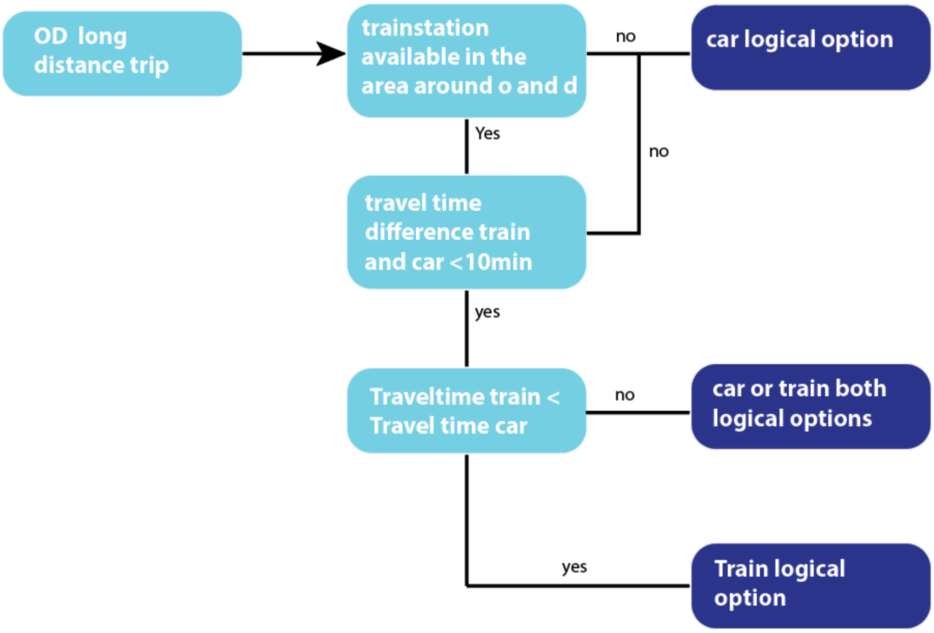


Figure 6.9: Decision tree long OD distances

For this research, information of all train stations is gathered. This excel is an export of the Nederlandse Spoorwegen (NS) Application Programming Interface (API) platform, which includes all train stations in the Netherlands. The data contained the following information:

- station name
- station code
- coordinates: lattitude and longitude
- address, including street, homenumber, zipcode, cityname and country
- name intercity station: filled if it is an intercity station, otherwise empty

Using the exported excel file with this information, a shapefile is created in FME and the data is loaded into the database.

6.4.1 Trainstation locations

After creating the shapefile with the train stations and comparing it to the shapes of the areas the following problem arises. The train stations can not be related to one LMS area in most of the times as it lays on the border of two areas. In the following example, the Amsterdam South station lies in area 650, but is related to different areas in the neighboorhood, see Figure 6.10.

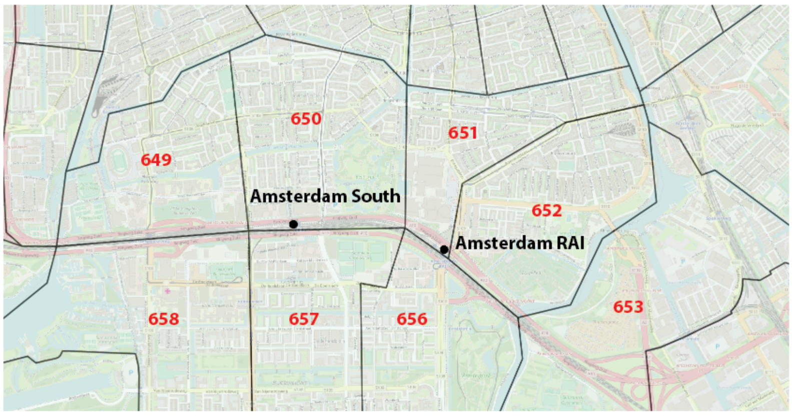


Figure 6.10: trainstation are on borders LMS areas

As shown in the Figure 6.11, Amsterdam south station has an entrance in two LMS areas. The station is thus related to both area 650 and 657. The solution to this problem will be described in the following paragraph.

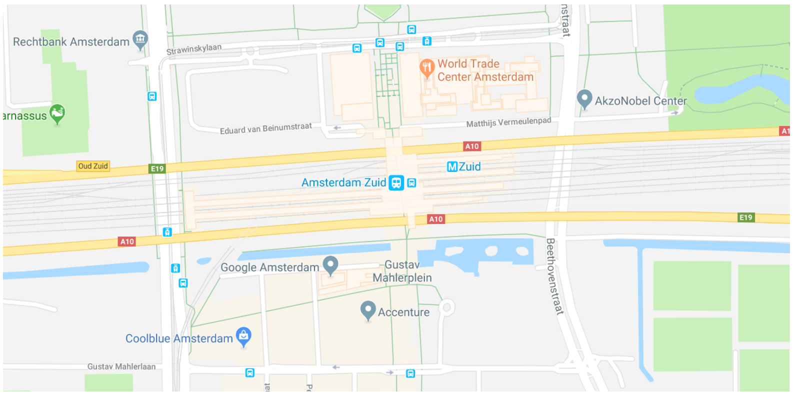


Figure 6.11: situation Amsterdam South station, entree of the train station in two different LMS zones (google maps)

6.4.2 Region of interest trainstation

To deal with these situations, a Region Of Interest (ROI) is introduced. This is a circle around the train station. The area of this ROI is based on the type of trainstation, as shown below:

- intercity stations: travel distance to the station max 2,5 km
- non intercity stations: travel distance to the station max 1,5 km

The ROI is an area where someone can easily get acces to a trainstation. In this area a person can easily cycle to a station or get acces to public transport to a train station. Research published by Kennisinstituut voor Mobiliteitsbeleid (KIM) shows that a distance of 1 to 3 kilometers cycling as the crow flies between the residential location and the station is widely accepted by train users in the Neterlands. The average cycling distance between residential location and station is 2.4 kilometers as the crow flies [Jonkeren et al., 2018]. For this research this number is rounded to 2,5 kilometers for intercity stations.

Intercity stations are better accessible (e.g. more options to reach with public transport) and provide faster and better connections to other stations. In the choice for a station, bicycle-train users, but also train users who have another travel preference to and from the train, have a clear preference for stations that are high in the station hierarchy (Intercity stations). Even if those train travelers have another station (smaller, with a lower operating status) available at a shorter distance from their home location, this preference occurs. In general it is the case that bicycle-train travelers who live within cycling distance of an IC station (up to about 5 kilometers), are less inclined to also use other stations that are within an acceptable cycling distance [Jonkeren et al., 2018]. For this reason a lower distance is chosen for non-intercity stations.

The figure 6.12 shows how the ROIs are located on the map of the Netherlands. Amsterdam is covered by many ROIs of different stations, with an exception of the nothern part of Amsterdam.

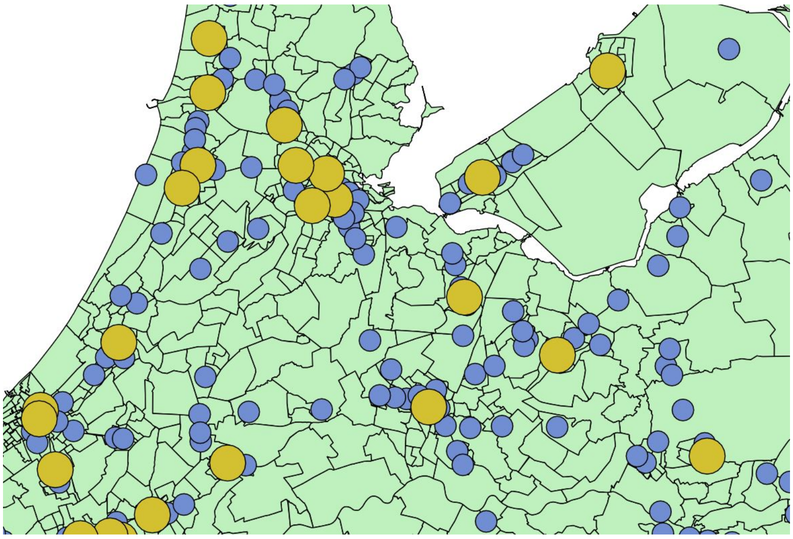


Figure 6.12: different type region of interest

Another reason to work with a ROI is due to lack of the Google API to support most logical multimodal options. An example of this phenomena is the travel time to a train station, which is shown in figure 6.13. In this example the Google API returns an option to walk to a tram station and then transfer to the train (which causes some extra time, because tram and train times not always connected). Instead people took more often the bicycle to the train station in the Neterlands [Brands et al., 2014]. During this research this is not a standard chosen option in the Google API. If a route from the midpoint of the polygon was chosen, it will probaly lead to an unrealistic travel time to the trainstation, because the option biking to the train station is not available. Therefore a ROI is more suited to the train trips.

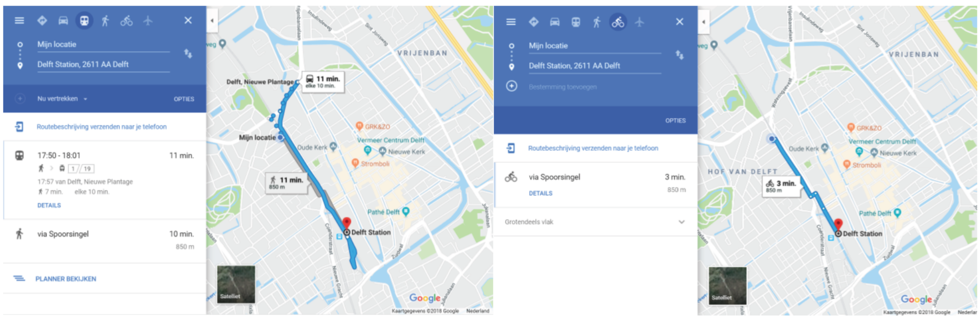


Figure 6.13: Left: what the Google api thinks you should do, right: what people in reality do

The ROI are clipped with FME and the different ROI parts can be assigned to the overlapping areas. A list is be made of LMS areas and stations that belong to that area. This can result that one LMS area is assigned to multiple ROI (multiple trainstations) where its connected to as shown in figure 6.14: for example area 624 connected to: Amsterdam Science park, Amsterdam Muiderpoort, Amsterdam Central station and Amsterdam Amstel.

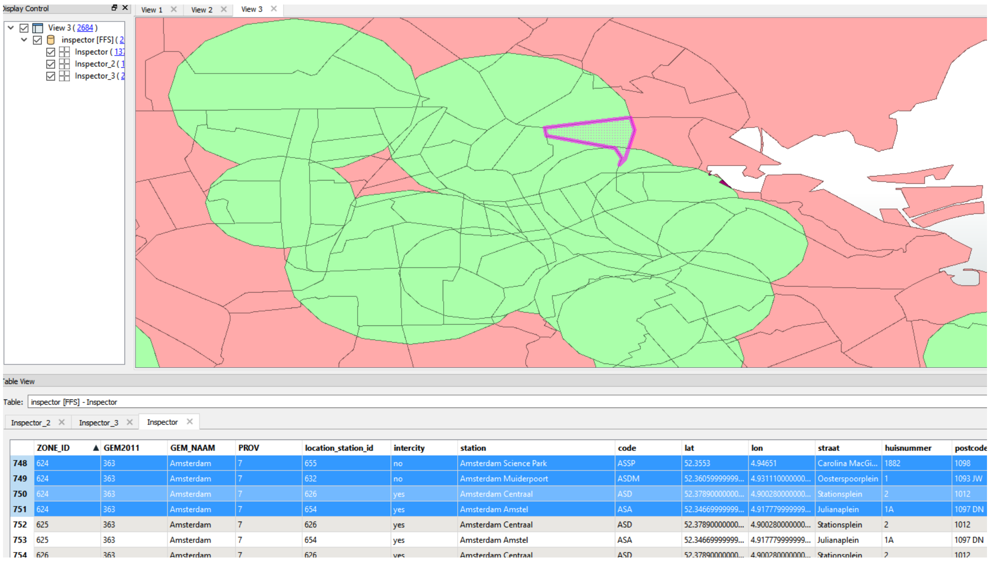


Figure 6.14: One area is related to 4 different stations

6.4.3 Accessibility of train stations in different areas

To determine if a sustainable option is realistic, the accessibility of a train station close to the area is important. Different catagories for the accessibility to a train station are determined based on the availability in the area, ranging from 1 (Best accessibility) to 6 (no trainstation), see fig 6.15.

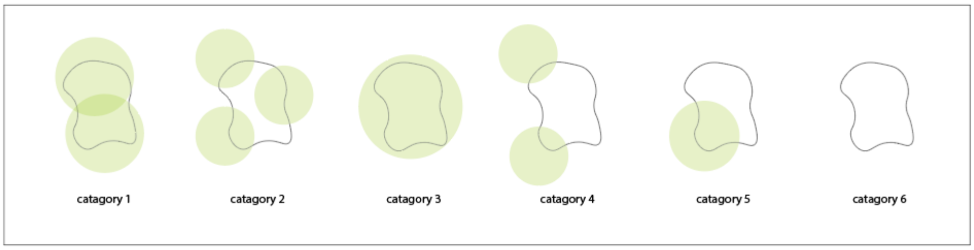


Figure 6.15: The 6 categories of accessibility

- Category 1: ROI from multiple stations overlapping > 75% of the LMS area
- Category 2: ROI from multiple stations partially overlapping (between 25% and 75%) the LMS area
- Category 3: ROI from one station is overlapping at least 75% of the LMS area
- Category 4: ROI from one or multiple stations overlapping a very small part (<25%) of the LMS area
- Category 5: ROI from one station is overlapping partly (between 25% and 75%) the LMS area
- Category 6: No ROI from any station are overlapping the LMS area (no station available)

6.4.4 Different methods to calculate overlapping area with the ROI

There are several ways to calculate the overlapping area ROI as shown in figure 6.16. This percentage is used to categorize the LMS areas.

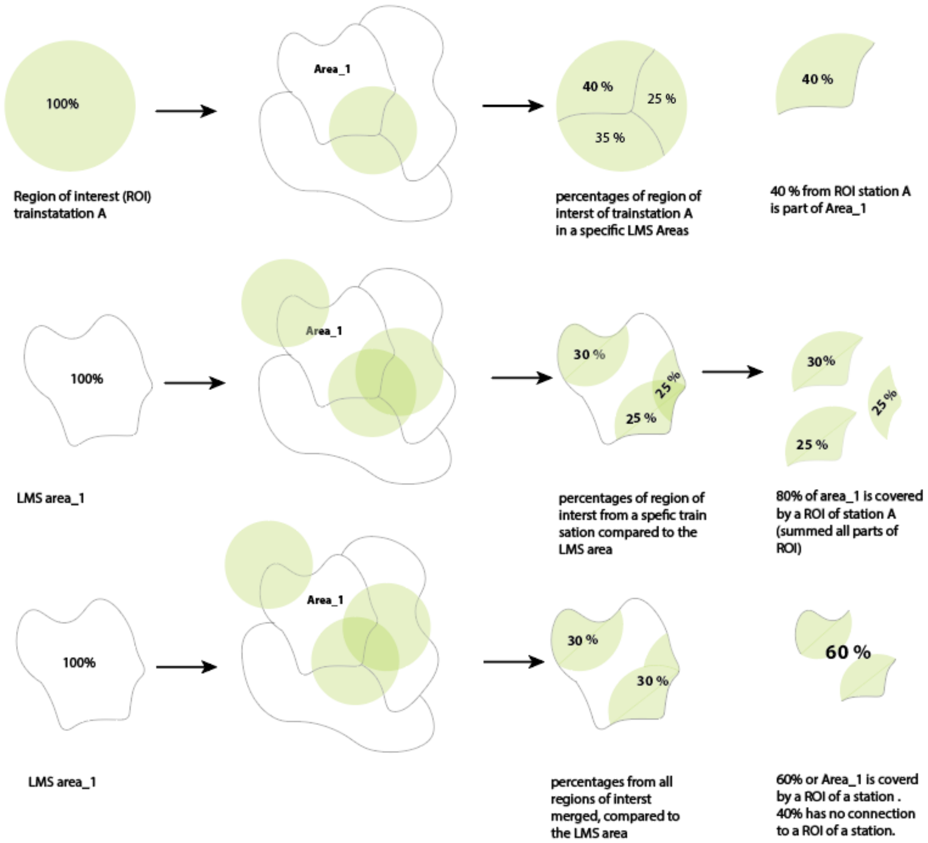


Figure 6.16: Different ways to calculate the percentage of the Region of interests

The third and last option is used to calculate the percentage of the merged ROIs. Using option 2, the percentages can be over 100 procent for areas which are overlapping, and option 1 provides information for every station on the precentage of the travellers travelling to the different areas. All options are calculated and added to the data set. During the proces only the 3rd option is used. The other calculation options maybe interesting for further research.

6.4.5 Google api for long distance

In addition to the accessibility of a train station, the route to be followed is also important. There are cases when a train trip is faster than a car, but also the opposite can be true. Illustration of this is the example of travelling from Hindeloopen to Amsterdam Sloterdijk (figure 6.17). By car its only 1 hour and 30 min drive, while the train option is at least double the duration. In this example it is not very likely that people will take the more sustainable option: the train.

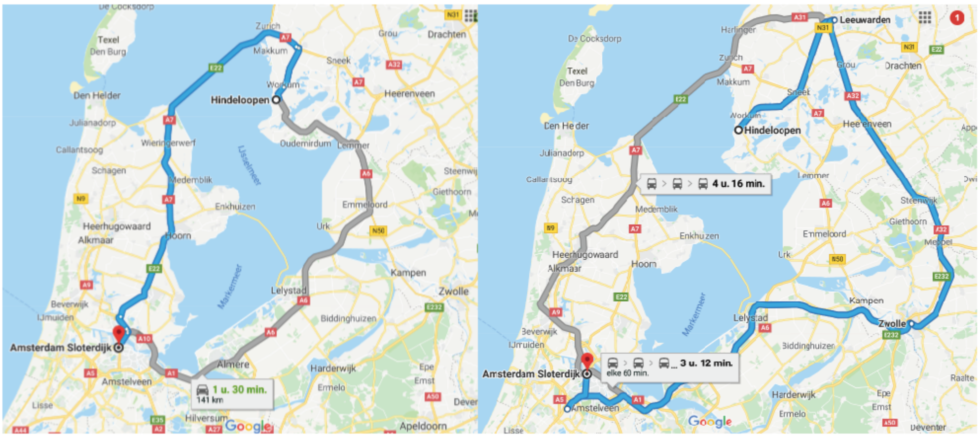


Figure 6.17: Example of travelling from Hindelopen to Amsterdam Sloterdijk

To obtain the duration and distance of the traintrips with the Google API the following steps are taken. First all combinations from and to stations in Amsterdam are made in an SQL table. After creating the posible combinations, the data needs to be enriched with the lattitude and longitude cooridinates of the trainstations. Finally a CSV file can be exported and loaded in the Matlab script for Google API.

The results can be catagorized as in following figure: 6.18

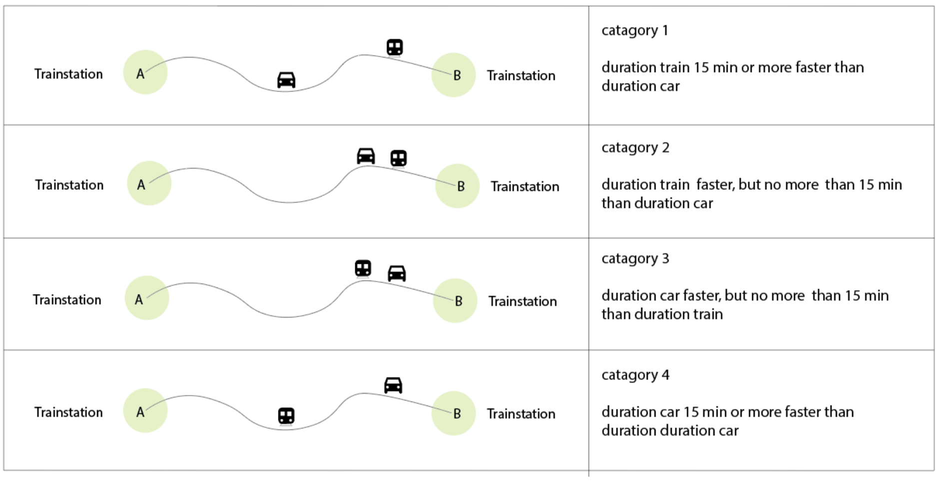


Figure 6.18: Categories for connections that are calculated by the Google API

The car trips are now related to the train stations. Most trips may start more from the centre of the polygon. For this situation the Google api is also run from midpoint to midpoint for long distance (only for car) see figure 6.19.

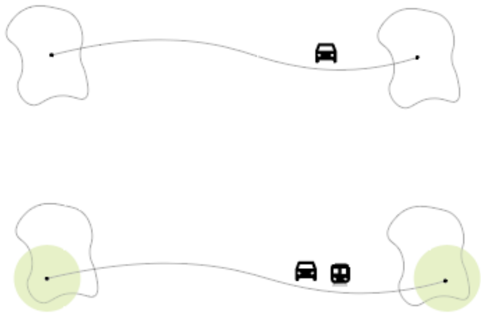


Figure 6.19: Google API for trainstations and midpoints

The complexity of the different travel options are shown in figure 6.20. There are several possible routes between origin destination combinations. It is possible for example that an origin or destination has multiple train stations. It is also possible that a train connection has a transfer which makes the train route much longer. The different combinations are compared and the shortest travel times are compared for the analysis. The shortest travel time by train is compared to the car journey from midpoint to midpoint.

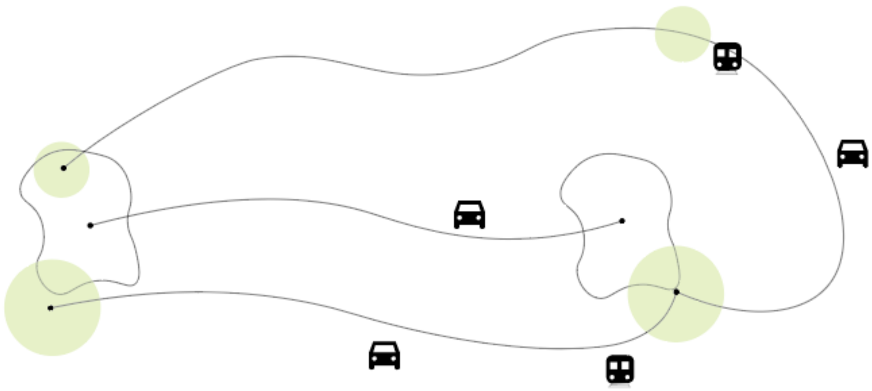


Figure 6.20: Complexity, multiple connections by train, different durations. Car is logical to look at the midpoint to midpoint, more representable for the origin destination combination

Due to the large number of OD combinations and possible connections to train stations, only a selection of the data is used to show the results: all areas in Amsterdam (68 LMS areas) and multiple different municipalities (366 LMS areas). The locations of the used areas shown in figure: 6.21. In Appendix C all municipalities in the selection are listed.

These cities are chosen because of the differences in categories, station accessibility, size of the city and population density of the area. This selection was needed to make to limit the use of the Google API.

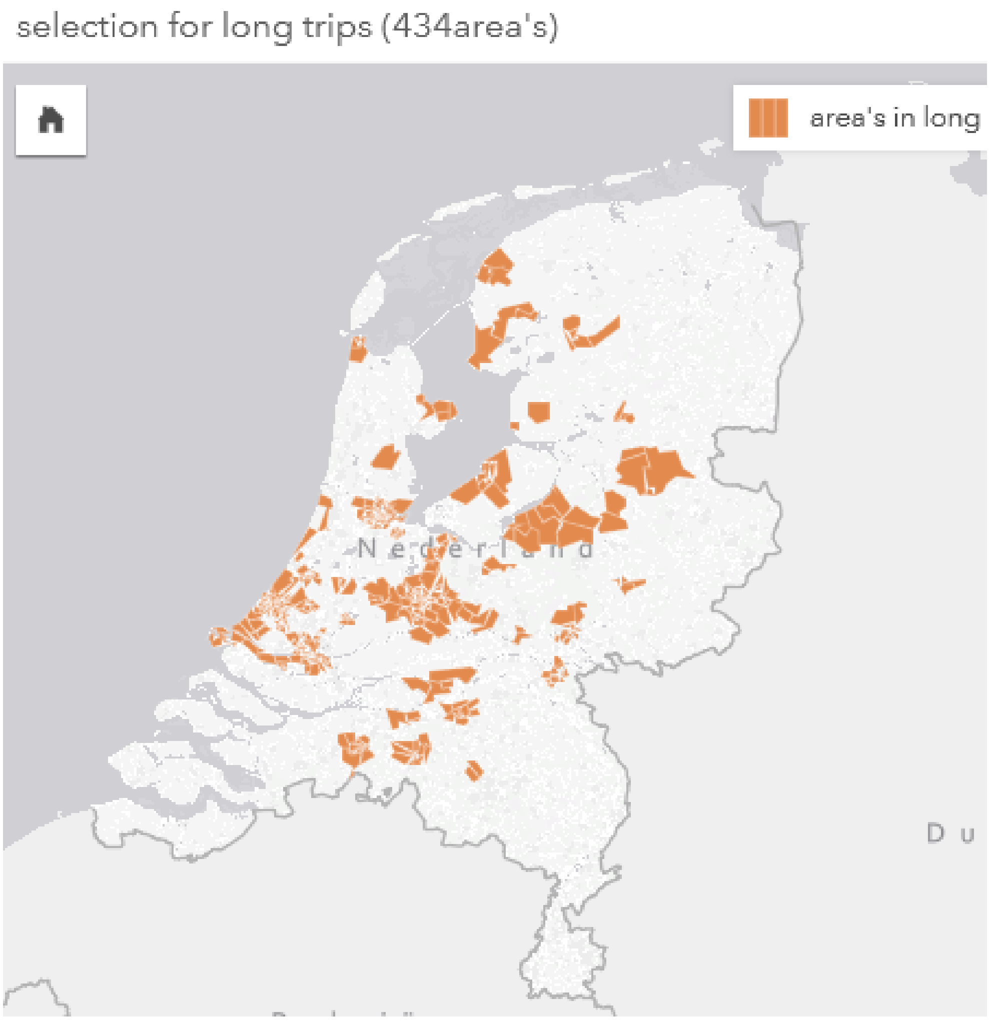


Figure 6.21: Areas that have been used to analyse the data and show the results

An example of the data is shown in figure 6.22. One origin destination combination can have multiple train connection combinations as shown in the example from LMS area 60 to 637 (Franeker to Amsterdam). This is because polygon 637 overlaps with four regions of interests of different train stations. In this example: Amsterdam Rai, Amsterdam Centaal, Amsterdam Zuid and Amsterdam Amstel.

combicat text	combiod text	from_zone_id integer	to_zone_id integer	fromstation character varying(255)	tostation character varying(255)	duration_1 real	duration_2 real	distance_1 real	distance_2 real	o in
5>1	60>636	60	636	Franeker	Amsterdam RAI	82.4	156.267	134.96	215.185	
5>1	60>637	60	637	Franeker	Amsterdam RAI	82.4	156.267	134.96	215.185	
5>1	60>637	60	637	Franeker	Amsterdam Centraal	80.05	149.667	122.565	212.766	
5>1	60>637	60	637	Franeker	Amsterdam Amstel	84.6167	163.8	134.992	231.545	
5>1	60>637	60	637	Franeker	Amsterdam Zuid	80.3833	149.933	130.658	214.161	
5>1	60>638	60	638	Franeker	Amsterdam Amstel	84.6167	163.8	134.992	231.545	

Figure 6.22: Example of traincombinations of one origin destination combination

6.4.6 Replaceability of a long distance trip

To decide if a trip has a realistic sustainable alternative, different aspects are involved: availabilty and duration. Both are included and discussed in this research. A replaceable option is a connection between two areas, were both areas are related to a ROI of a trainstation. This means that all connections to or from an area with accessibility catagory 6 are not counted as a sustainable alternative. Another aspect that is taken into account is the duration of the train connection. By subtracting the car duration (midpoint to midpoint) from the duration with public transport, the differences in traveltime are calculated.

In section 2.3 the VF value is introduced. This VF value give the accepted travel time difference between public transport and car trips. The smaller the VF becomes, the higher quality of public transport. To analyse sustainable travel behaviour for optional travellers, it is important to take into account that the VF value should not exceed 2.4. To calculate the VF value it is importand to calculate all the travel time including access and egress transportation as shown in figure 6.23.

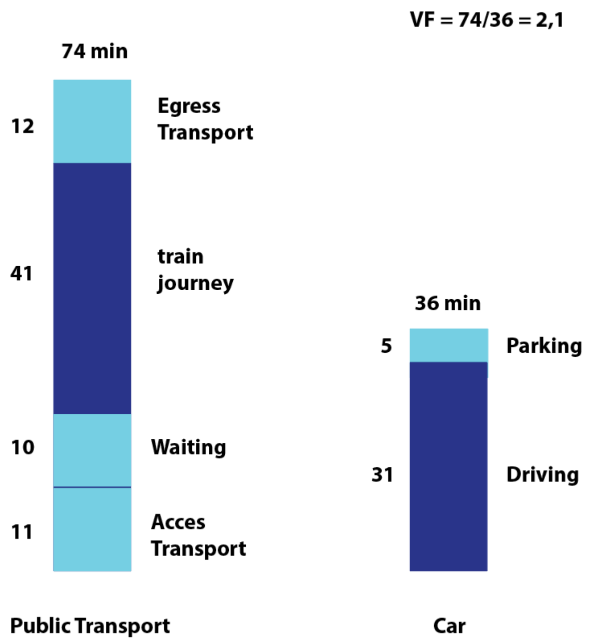


Figure 6.23: Structure of the elements to calculate the VF value for public transport (left) and car (right)

Because this study uses ROIs instead of exact travel times including access and egress transportation, it is not possible to calculate the exact VF values. In this research only the dark blue elements of figure 6.23 are calculated. Only looking at the travel time difference of the dark blue values it shows that a time difference of 10 min can lead to a VF value of 2,1. Because the VF value sould not be above 2,4 it is chosen to set a limit of 10 minutes difference between the calculated travel time of public transport and the travel time by car.

Four different replacabability categories can be identified, namely:

- ’yes’: if the calculated difference is smaller than 0 (train faster than car)
- ’maybe’: if the calculated difference is between 0 and 10 (train maximum 10 min slower than car)
- ’no’: if the calculated difference is higher than 10 (train at least 10 min slower than car)
- ’no data’: there is not a calculated value, due to the fact that one or two areas does not have a trainstation connection. In this case no alternative could be found.

In the case that there are more train connection posibilities, the shortest travel time by train is taken into account. Both no and ’no data’ category shows trips that are not replaceable. Because there are different reasons why they are not replaceable it is chosen to separate these categories.

7 RESULTS

In this chapter the results of the data analysis obtained from the implementation are presented. This chapter is divided in two parts. The first section 7.1 presents the results of the analysis of short distances trips obtained from the selection of data. Section 7.2 describes the results of the long distances trips.

7.1 SHORT DISTANCE

In this chapter only trips are selected that are a result from short distance selection made in chapter 6. To analyse the short distance trips, the following data is analysed:

- Walking (Google OD)
- Cycling (Google OD)
- Car (Google OD)
- Car (LMS OD)

For the distance calculation the Google API is used for the 3 different modalities. For each modality the travel time and travel distance is calculated. The results for the different modalities are presented in the following paragraphs, starting with walking, cycling and finally car trips.

7.1.1 Walking

In this section only trips are selected that are a result from short distance selection made in chapter 6.3. The modality walking could only be analysed from the Google OD data set. For walking trips, looking at the traveled distance, the most logical trips are the trips were the origin is in the same area as the destination. This is also shown in the Google OD data, where the highest weights are found when the area code of the origin is the same as the destination.

Note that some values are higher than 1, meaning that there are more trips than the highest number of car trips in the data. The result of the highest values walking (higher than 1) is shown in table 7.1.

Origin	Destination	Weight
626	626	7.51
629	629	3.19
672	672	2.47
676	676	1.61
630	630	1.47
625	625	1.42
674	674	1.33
620	620	1.30
638	638	1.26
658	658	1.08
657	657	1.04

Table 7.1: Walking trips that have a higher value than the highest in the car data set

The areas that have a higher value than 1 are shown in figure 7.1.

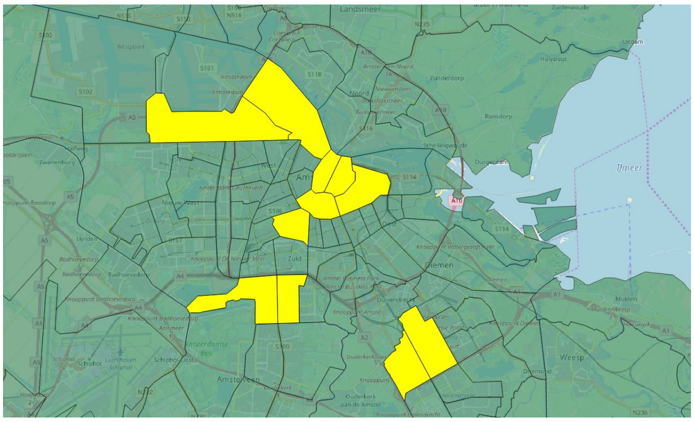


Figure 7.1: Areas where walking weights have values higher than 1

Note that from the figure 7.1 the areas with the most trips can be linked to important places in the city of Amsterdam:

- historical city centre
- Museum Square
- Zuidas, business area
- Ziggodome, Bijlmer arena, AFAS live

There seems to be a high correspondence between the Google walking data and the activities in the city. To see if the walkable distances have high values of walking trips, a comparison is made between the different modalities. A selection is made that shows only trips with walkable distance as described as category 1 in chapter 6.3. This selection is than compared to the colum favorite in Google. The result will show the favorite modality in the Google data of the walkable trips in category 1. The result for the OD combinations that are walkable are shown in tabel 7.2.

Favorite modality in Google	Count origin-destination pair
Walk	96
Car	13
Cycle	2

Table 7.2: Walkable distance vs. favorite in the Google OD data set

There are also examples found in the data set where walkabe should not be an option, but from the Google OD data set walking is still the favorite option.
This phenomenon occurs 84 times in the data set. These OD combinations have a duration for walking that is higher than 15 minutes.
Trips that have not the label of walkable from the dicision tree, but have walk as a favorite in the Google data set, are shown in figure 7.2. In this figure the duration of the walking trips from midpoint tot midpoint is shown.

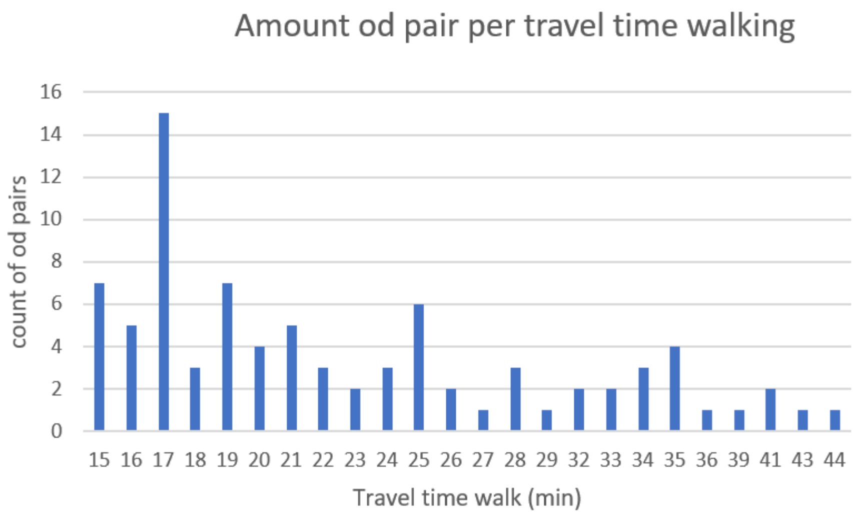


Figure 7.2: Unexpected walk favorite in Google and travel time walking

In figure 7.2 a peak is shown at 17 min. Probably there are in these areas more trips around the borders of the area. The distances of these unexpected walking trips is shown in table 7.3.

Distance walking (round)	Count origin-destination pair
1	30
2	37
3	15
4	2

Table 7.3: Distances of the unexpected walk trips

For most unexpected walking trips the distances are 2 km or less. Again it should be mentioned that this is the distance measured from midpoint to midpoint, and in reality the walking trips could be shorter or longer.

7.1.2 Cycling

In this section only cycling trips are selected that are a result from short distance selection as made in chapter 6.3. The modality cycling could only be analysed from the Google OD data set as it is not available in the LMS data set. The travel time of the cycling trips are compared to the travel time of the car trips for every origin destination combination in the short data set. The result is categorized and shown in table 7.4.

duration_cycle-duration_car	count	cycle is aternative	min distance_cycle	max distance-cycle	avg distance_cycle	number of cycling trips >10 km	cycling trips <10km
-15 tot -10 min	11	+++	1,508	5,116	2,98	0	100%
-10 tot -5 min	45	+++	0,96	6,418	3,48	0	100%
-5 tot 0 min	458	+++	0,57	7,955	3,41	0	100%
0 tot 5 min	1132	++	0,651	23,52	4,92	8	99,3%
5 tot 10 min	931	++	2,17	24,308	7,50	64	93,1%
10 tot 15 min	993	+/-	4,27	26,474	9,44	324	67,4%
>15 min	16898	-	6,054	50,502	23,51	16597	1,9%

Table 7.4: Results of Google API compare traveltime cycle and car

It shows that there are many trips that have less travel time by bicycle than by car. These OD combinations have a potention that car trips could be replaced by cycling trips. Also notable from this table is that the cycling trips that have a duration longer than 15 minutes compared to the car trip duration, are almost all trips with a distance longer than 10 km (98,1%). OD combinations where cycling is an sustainable alternative to the car trips mostly have a distance lower than <10 km. The reason why these trips can be longer than 10 km, is because the selection of the short data set is made with a buffer of 10 km around Amsterdam. The distance calculation of the trips is later calculated with the Google API per modality. It can be noted that the trip distance is longer by bicycle, as for example a brigde over the water needs to be taken with a detour.

In table 7.5 trips are selected, that should be bikeable comform the discision tree shown in figure 6.4. This results in the following categories where cycling is as an alternative to car trips:

- yes: o=d or traveltime walking <15 min , distance cycle <10km and difference between bicycle and car <15 min
- maybe: distance longer dan 10 km, but difference between bicycle <15 min
- slow but short : difference between bicycle and car >15 min but distance bicycle <10 km.
- no: distance bicycle >10km and difference between car and bicycle >15 min (car faster than bicycle)

Note that in the category 'yes' also trips are included that could be replaced by either walking or cycling. Here is an overlap, because walking and cycling are both sustainable options, cycling should also be a sustainable alternative for trips shorter than 10 km.

Cycle is option	Favorite in Google OD	Appearance OD pairs	Percentage
Yes	Walk	180	0,9%
Yes	Cycle	724	3,5%
Yes	Car	2116	10,3%
Yes	all	221	1,1%
Maybe	Cycle	4	0,0%
Maybe	Car	281	1,4%
Maybe	All	99	0,5%
Slow, but short	Cycle	1	0,0%
Slow, but short	Car	270	1,3%
Slow, but short	All	49	0,2%
No	Cycle	9	0,0%
No	Car	6638	32,3%
No	All	9944	48,4%

Table 7.5: Favorite Google vs. the expected from the Google api

For the category yes, where trips are shown that have cycling as an sustainable alternative, the Google data shows only 724 od combinations of the 3241 in the category yes that have the favorite of cycling.

Notable in table 7.5 is that there are cycle trips which should not be counted as a cycle alternative, that still have more cycling trips than car trips. These od combinations are shown in appendix B. The distance of these unexpected cycle trips are in 7 cases between 10.3 and 13.3 km cycling distance.

7.1.3 Car trips

In this section car trips are selected that are a result from short distance selection as made in chapter 6.3. For this modality both Google and LMS data sets can be used.

Correlation LMS and Google OD for short distances

To determine the validity of the data sets, the correlation between the LMS and Google OD data sets using the method described in section 5.4 is calculated. Using the areas in scope (Amsterdam), for every short distance connection the correlation coefficient for car trips is 0.5, meaning that there is a moderate correlation between the two data sets. When only considering the short distance connections, which are easily replaceable with sustainable method (see next section), the correlation coefficient is 0.6.

Potential trips to be replaced by sustainable transportation means

Based on the decision tree (as shown in figure 6.4) and the traveled time and distance, calculated with the Google API, a potential of replacement is calculated. The following categories are represent:

- Category 1: Easily replaceable by walk and bicycle.
- Category 2: Easily replaceable by bicycle .
- Category 3: Maybe replaceable by bicycle or public transport. Travel time difference between bicycle and car less than 15 minutes, but distance bicycle longer than 10 km.
- Category 4: Maybe replaceable by bicycle or public transport. Distance longer dan 10 km, but differece between bicycle and car less than 15 minutes.
- Category 5: Not replaceable by walk or bicycle (public transport potential but outside the scope of this research).

These catagories are used to test if a origin destination combination is replaceable by a sustainable option. The result for the short distance data set is described in table 7.6.

category	walk-option	cycle-option	in short data-set	percentage	reparable
1	yes	yes	111	0,5%	easy, by cycle and walk
2	no	yes	3130	15,2%	easy, by cycle
3	no	slow but short	320	1,6%	maybe by cycle (or public transport)
4	no	maybe	384	1,9%	maybe by cycle (or public transport)
5	no	no	16591	80,8%	not by walk or cycle (public transport most sustainable option)

Table 7.6: Potential replaceable OD combinations from Google api for the selection short distances

As shown in table 7.6, there are 5 different categories defined, where the category 1 and 2 are easy replaceable by bicycle, 3 and 4 have a smaller potential to be replaced by cycling, but a higher potential to be replaced by public transport. This is outside the scope of this research, because there is no data for this trips available in the chosen data sets. For category 5 trips, these are only replaceable by public transport, but this is as mentioned outside the scope of this research. After looking at all categories, the three main categories of replaceability are shown: easy (category 1 and 2), maybe (category 3 and 4) and not likely (category 5). These categories are compared to the favorite column to see if this corresponds to the values in the Google data set. The results are shown in tabel 7.7.

Replaceable Google API	Favorite in Google	Count OD in Google	percentage	Replace potential
Easy	walk	180	0,9 %	sustainable option already favorite
Easy	cycle	724	3,5%	sustainable option already favorite
Easy	car	2116	10,3%	high potential to replace car trips
Easy	all	221	1,1%	unclear favorite in Google data
Maybe	cycle	5	0,0%	sustainable option already favorite
Maybe	car	551	2,7%	medium potential
Maybe	all	148	0,7%	unclear favorite in Google OD data
Not likely	cycle	9	0,0%	unexpected sustainable favorite
Not likely	car	6638	32,3%	logical car use
Not likely	all	9944	48,4%	unclear favorite

Table 7.7: Potential replaceable trips vs actual favorite in Google OD data set

In all replaceable categories, exept from walking, all the different favorites are present for some origin destination combinations. Only walking is present in the easy replaceable category and not in the Maybe and Not likely. For the easily replaceable category, 904 combinations have the the sustainable option as favorite in the Google data set . For 2116 origin destination combination trips car is still the favorite option. This trips have a high potential to be replaced by a more sustainable alternative. For the maybe category, the car is the most often chosen as favorite. In the not likely replaceable category the all category is mostly represented.

Comparison of the Google and LMS data set

For this research, the interesting car trips are the ones that are easily replaceable by cycling or walking. To analyse these replaceable trips, a sub set is made with 3241 OD pairs that should be easily replaceable by cycling and walking according to the Google API distances and duration. This selection is based on the result in table 7.6 and is the summedcount of category 1 and category 2. To compare these data sets, selections are made with SQL and this data is exported to ESRI Insights to create interactive maps. The resulting maps are shown and described in this section. First all data of the short distance table is compared. Per origin and per desination the weight values of the Google car trips and the weight values of the LMS cartrips are summed. The result of this analysis are shown in figure 7.3

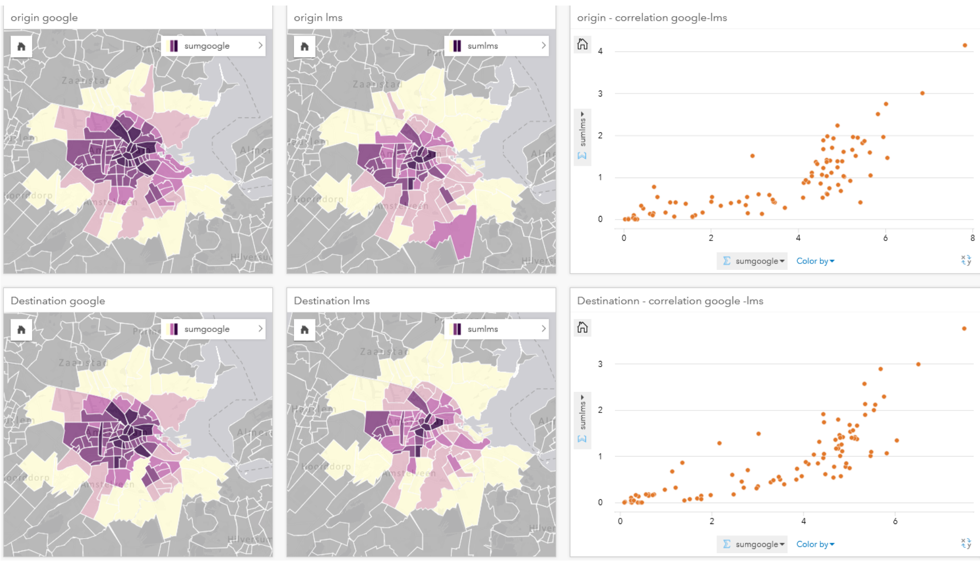


Figure 7.3: Summed weights for orgins and destinations from the Google and LMS data sets for car trips for short distances and the correlation between them

In these interactive maps data can be selected for example in the scatterplots and the selection will be applied in the maps next to it. By selecting the highest weight values in the scatterplot, the origins and destion can be shown that have the highest summed values. The result of the selection is shown in figure 7.4. The highest summed weights are mostly concentraded around the citycentre of Amsterdam.

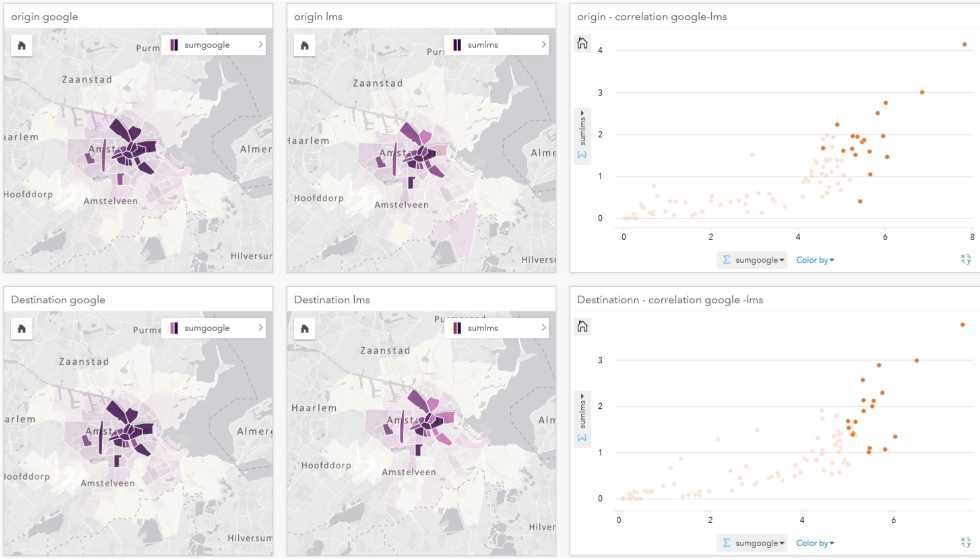


Figure 7.4: Selection made in summed weights for orgins and destinations from the Google and LMS data sets for car trips for short distances and the correlation between them

To compare the Google and LMS data sets an origin destination matrix is made where the different weights for every origin and destinations are ploted. The visualisation of the Google weight is shown in figure 7.5 and the visualisation of the LMS weight is shown in figure 7.6.

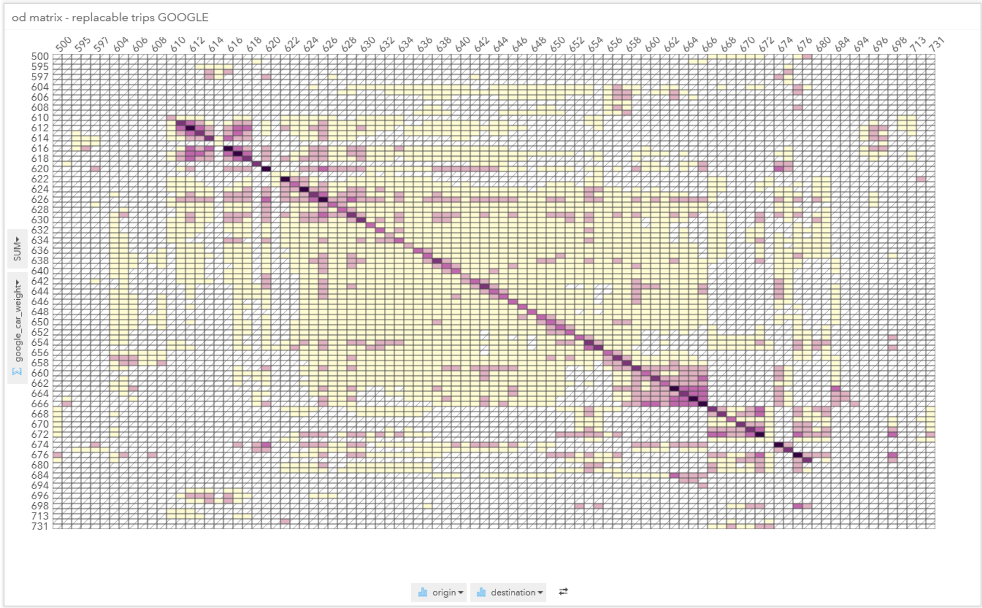


Figure 7.5: OD matrix for replaceable trips, visualizing the Google weigths by coloring the data with natural break method

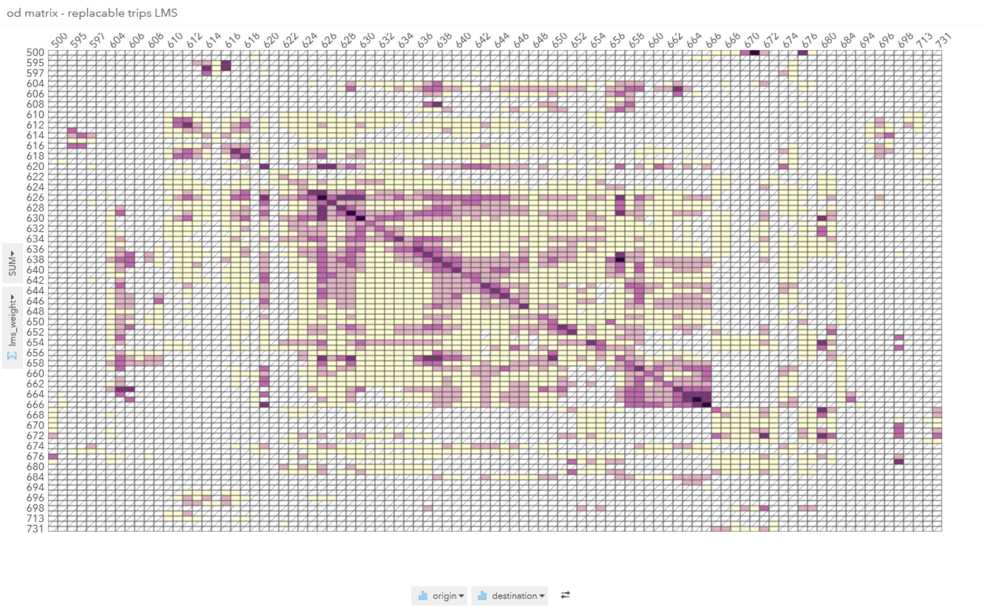


Figure 7.6: OD matrix for replaceable trips, visualizing the LMS weigths by coloring the data with natural break method

In both matrices a diagonal line occurs. This line is caused by the high weights of OD trips that have the same origin as destination. In both data sets these short trips produce the highest weights in the data set. In the cases where origin and destination is not simular, trips on the borders are more common to have high values. For the Google weights, the highest weights can be visualized with lines on the map by selecting the darkest color in the matrix (figure 7.7)

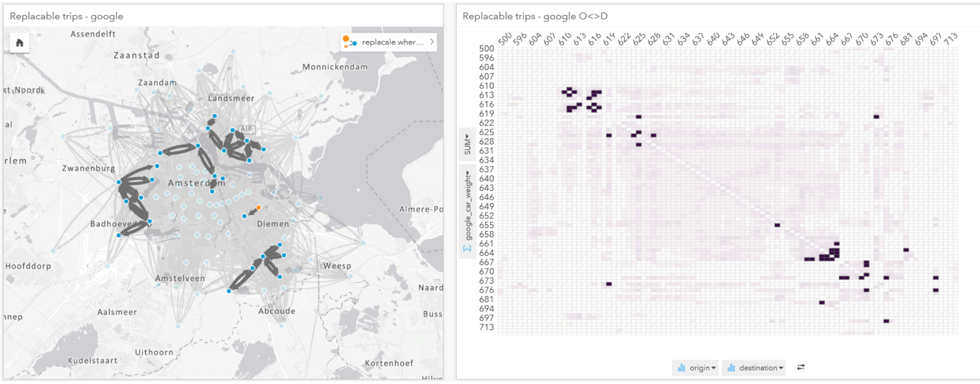


Figure 7.7: Highest weight values of Google data sets for car where origin is not destination

If this result is compared to the data of LMS with this similar approach (figure 7.8), different results come up.

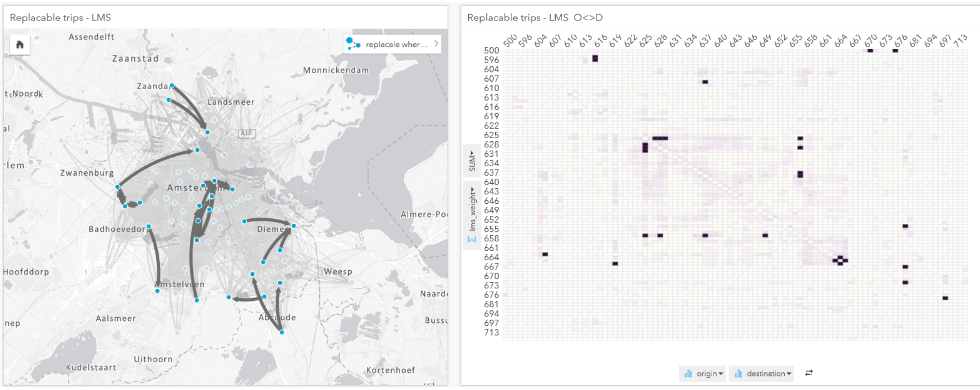


Figure 7.8: Highest weight values of LMS data sets for car where origin is not destination

In the LMS data set, car trips from the city center connected to the south areas in the city center are more present than the Google data set. Another difference in both figures is that the Google data set shows shorter distances compared to the LMS highest values.

Another way to analyse the data is by selecting values from a correlation table. In this way differences can be easaly found. In figure 7.9 trips that have a high correspondence in the Google data set but a low correspondence in the LMS data set are shown.

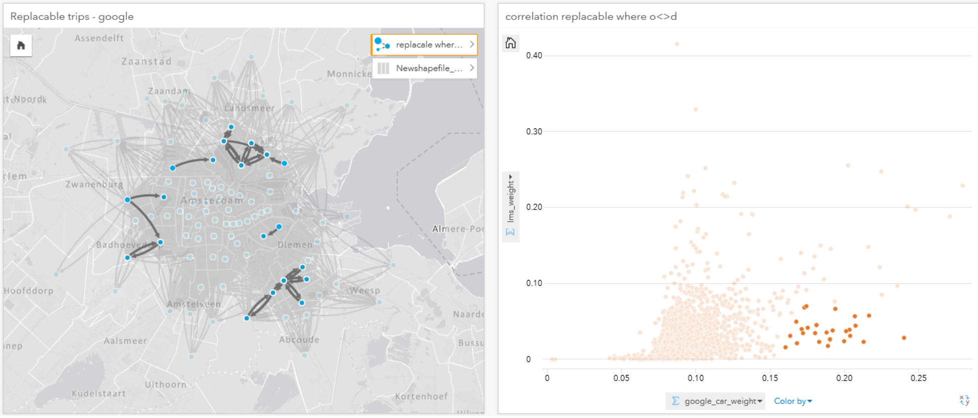


Figure 7.9: Replaceable trips that have a high weight in the Google data set but not in the LMS data set

On the other hand, the LMS data sets also have origin destination combinations that have a high weight, but show a low value in the Google data set. This result is shown in figure 7.10. In this case a very strong relationship between the centre and the south part is visible in the LMS data set, which is not clear in the Google data set. Also, the distances of the highest values in the lms data set seem to be slightly longer than in the Google data set. Where the highest values in the Google data set are mainly adjacent areas, in the LMS data set this is somewhat further in some cases.

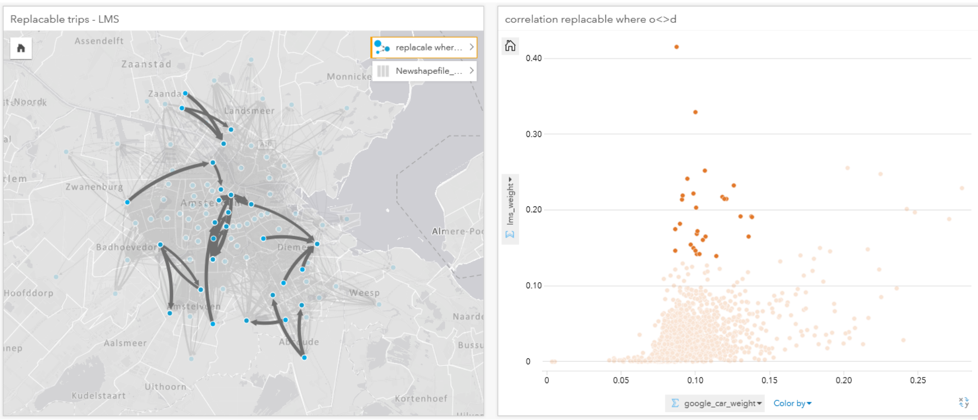


Figure 7.10: Replaceable trips that have a high weight in the LMS data set but not in the Google data set

There are also combinations that are high in both data sets, which are visualized in figure 7.11. These combinations are in the center and on the west side of the city.

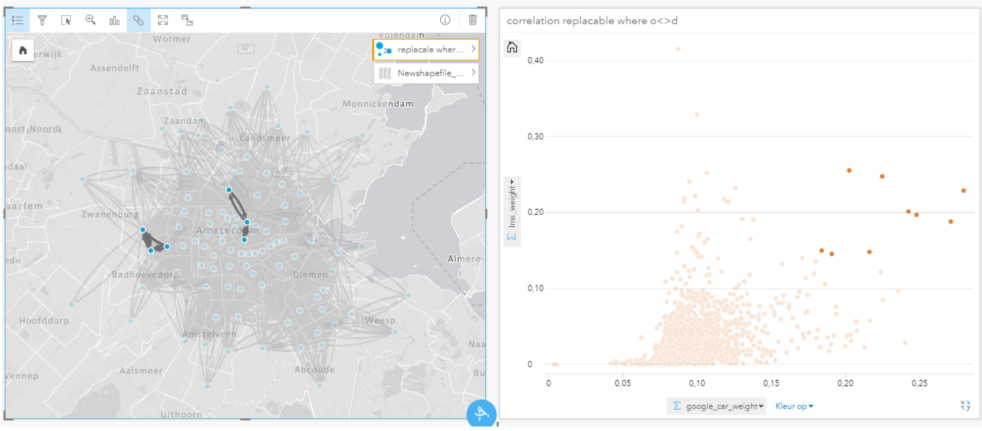


Figure 7.11: Replaceable trips that have a high weight in the LMS data set and in the Google data set

7.2 LONG DISTANCE

After studying the short distances trips now the rest of the origin destination data is analyzed. In this section only trips are selected that are a result from long distance selection as shown in figure 6.21. To analyse the long distance trips, the following data is analysed:

- Car (Google OD)
- Car (LMS OD)

In this section Google cycle and Google walk will not be analysed. The reason to do this is because they are not considered as a realistic sustainable alternative for long distance trips.

For the distance and duration calculation, the Google api is used for the 3 different approaches based on figure 6.20

- Car: from midpoint tot midpoint of an area
- Public transport: from train station to train station
- Car: from trainstation to trainstation

7.2.1 Correlation LMS and Google OD

To determine the validity of the data sets, the correlation between the LMS and Google OD data sets are determined, using the method described in Section 5.4. Using the areas in scope (Amsterdam), for every long distance connection the correlation coefficient is 0.15, meaning that there is a weak correlation between the two data sets for longer distances. This means, for longer distances, there are more differences between the LMS and Google data sets compared to the shorter distances.

7.2.2 Accesability categories

For each modality the travel time and travel distance is calculated. In addition for each area a accessibility category is calculated based on the presence of a train station as shown in catagories area's accessibility in LMS areas, explained in section 6.4.3.

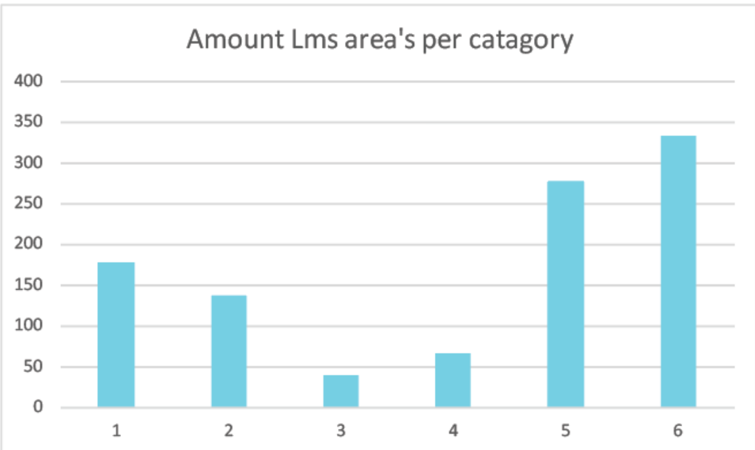


Figure 7.12: Amount of areas per accessibility category for all 1030 areas

In the selection made for al 1030 available areas as shown in figure 7.12, remarkable is the amount of areas in accessibility category five and category six. These accessibility catagories representing the lowest accessibility by all accessibility categories.

The subset of 434 areas which is used in the analysis for long distance trips figure 7.13 gives a higher response on values in accessibility category 1 which are representing the best accessible areas by train.

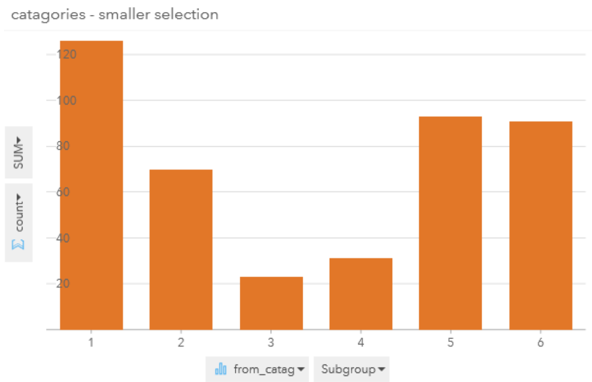


Figure 7.13: Amount of areas per accessibility category for the subset that is used for long distances

If the 68 areas of Amsterdam are only selected, a even higher response is visible in the first accessibility category. The other accessibility catagories are less representive as shown in figure 7.14. This high response is probably because there are 9 train stations in Amsterdam at the time of this study. As a result, ROI overlap easily and many areas are rerelated to at least one ROI.

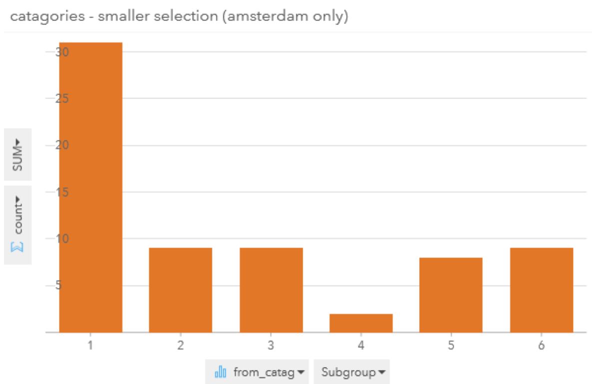


Figure 7.14: Amount of areas per accessibility category for all areas in Amsterdam

When looking at the combinations of different trips in the selected data set in figure 7.15, there are relatively many combinations from an area with accessibility category 1 to an area with also accessibility category 1. A lot of combinations are connected to an area with the accessibility category 1. This can be the result because all trips need to have a connection with the research area in Amsterdam where there are relatively a lot of areas in accessibility category 1.

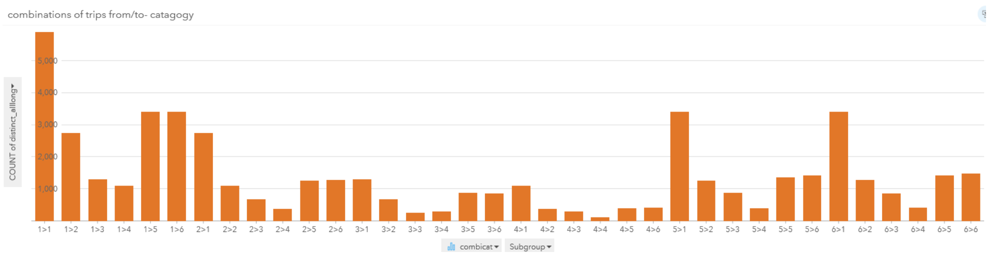


Figure 7.15: Number of OD combinations per accessibility catagory combination in the long data set

Analyses were done to see whether there is a relationship between accessibility and the weights in the LMS and Google data set. The average weight value is calculated for each combination as shown in figure 7.15. The LMS areas do not show a very clear relationship as shown in figure 7.16. It can be seen that combinations that include a trip with a connection to an area in accessibility catogory 3 have slightly lower weights on average. Journeys where there is a connection from or to a accessibility category 2 area, score a little higher on average.

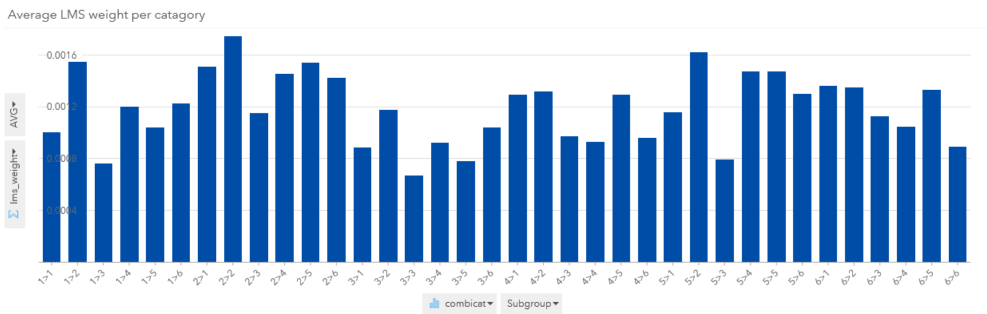


Figure 7.16: LMS average weight per accessibility category combination

The Google data set shows a completely different story. All combinations look similar to each other. Only small differences are visible in the combinations from a accessibility category 6 area, where there is no train station available.

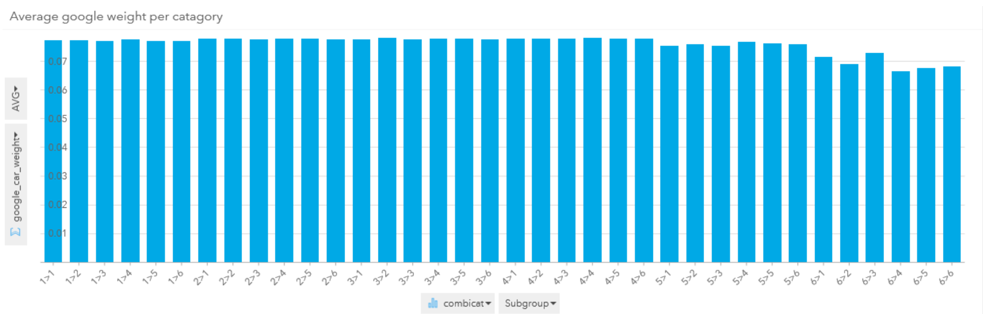


Figure 7.17: Google average weight per accessibility category combination

To compare the two data sets, both results are plotted in a matrix (figure 7.18). On the x-axis the origin accessibility category is shown and on the y-axis the destination accessibility category is shown. The colors show the difference in average weights, where the light color corresponds to a low average and a dark color to a high average.

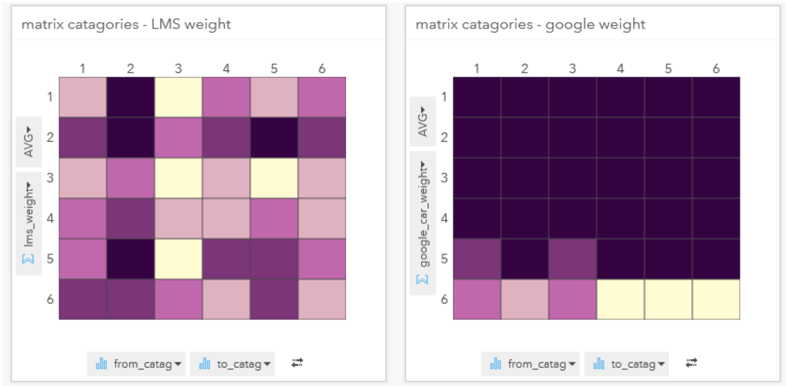


Figure 7.18: Matrices for combinations for different accessibility categories compared to the LMS average weights (left) and the average Google weights (right)

To look more in detail in the LMS data set, it is split in two parts, one for trips with an origin in Amsterdam, and one for trips with a destination in Amsterdam. The results of accessibility are visible in figure 7.19 where on the left matrix the trips that have a origin in Amstertdam are shown, and right matrix the trips to Amsterdam. Both on the x-axis the origin is show and on the y-axis the destination.

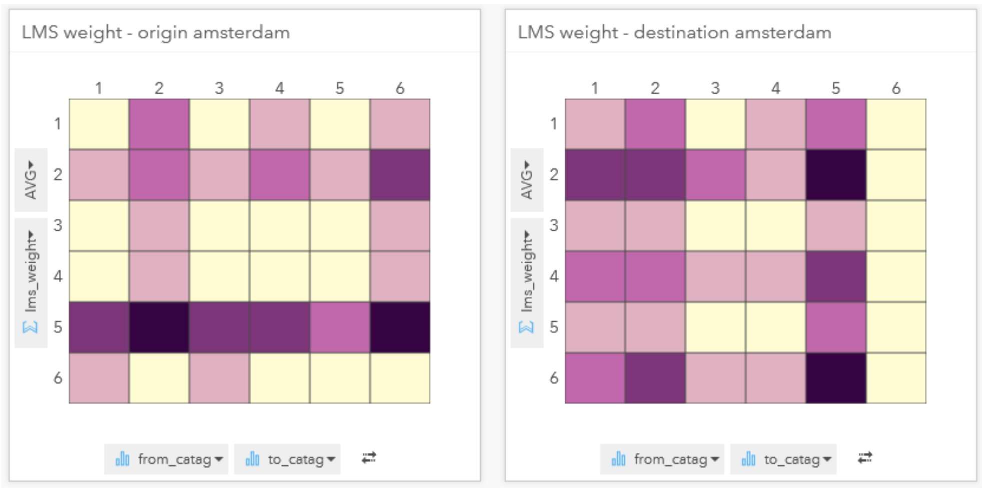


Figure 7.19: Matrices for combinations for different accessibility categories compared to the LMS average weights for origins and destinations to Amsterdam

In the left matrix there is a clear horizontal line visible at row five. This corresponds to a high average weight for trips with an origin in a accessibility category 5 area in Amsterdam. Also origins in Amsterdam in accessibility category 2 are higher than others. Also trips from Amsterdam to a accessibility category 6 area have a higher average weight than others. The right matrix, where incoming trips in Amsterdam are shown gives a different result. Destinations in Amsterdam have that have the highest weights on incomming trips are from accessibility category 5, 2 or 1. But it is striking that areas in Amsterdam in accessibility category 6, where there is no train station, have a lower average car weight than others.

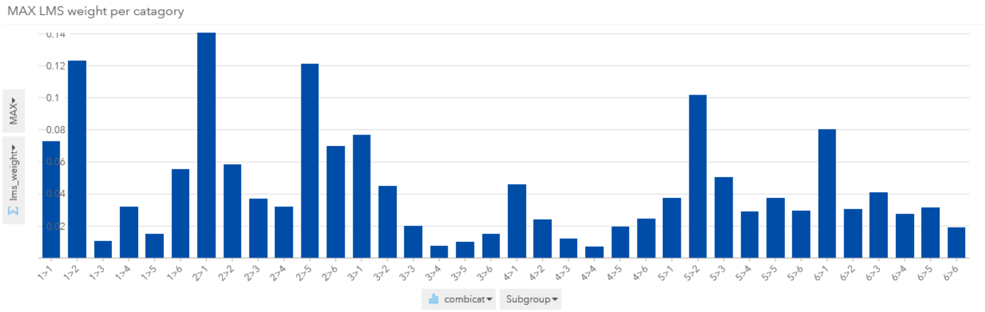


Figure 7.20: LMS maximum weight per accessibility category combination

To compare the highest weights in all accessibility category combination, a graph is made in figure 7.20. This figure shows more diversity than the average weight graph. But again connections in accessibility category 2 seems to have higher max weight values than other accessibility catagories. A combination with accessibility category 2 and 5 or accessibility category 1 and 2 are having the highest weights.

7.2.3 Travel Time

The travel time is calculated for both train and for car. The travel time from train station to train station is compared to the travel time from midpoint to midpoint. An OD combination can have multiple train combinations. In the previous mentioned case, the shortest travel time by train will be selected. Some notable cases will be discussed in this section.

During the analysis of the travel time of the trip, results showed that there are connections that are >15 min faster by train than by car, these connections are shown in table 7.8

From station	To station
Rotterdam Centraal	Amsterdam Centraal
Gouda Goverwelle	Amsterdam Holendrecht
Amsterdam Holendrecht	Gouda Goverwelle
Amsterdam Centraal	Haarlem
Zwolle	Amsterdam Centraal
Amsterdam Centraal	Halfweg-Zwanenburg
Amsterdam Centraal	Overveen
Gouda Goverwelle	Amsterdam Bijlmer ArenA
Amsterdam Centraal	Utrecht Centraal
Rotterdam Centraal	Amsterdam Lelylaan
Amsterdam Centraal	Rotterdam Centraal
Utrecht Centraal	Amsterdam Centraal
Amsterdam Centraal	Zandvoort aan Zee
Castricum	Amsterdam Centraal
Amsterdam Bijlmer ArenA	Gouda Goverwelle
Amsterdam Centraal	Castricum
Hilversum	Amsterdam Centraal
Haarlem	Amsterdam Centraal
Amsterdam Bijlmer ArenA	Utrecht Centraal
Amsterdam Centraal	Heemstede-Aerdenhout

Table 7.8: connections that are >15 min faster by train than by car

The results of these connections as shown in table 7.8 are sometimes expected. For example the connections to the different big stations, where direct intercity connections are facilitated (Rotterdam Centraal, Utrecht Centraal), or connections to stations close by amsterdam (Haarlem, Heemstede-Aerdenhout, Halfweg-Zwanenburg, Overveen). Unexpected were the connections outside these scope like: Gouda Goverwelle, Castricum, Zwolle and Zandvoort aan zee.

7.2.4 Replaceable trips

To answer question 5 it is important to know more about the replaceable trips. An explanation of when a trip is considered replaceable is given in section Section 6.4. Given the selection made for long distances, figure 7.21 shows the different categories for all od combinations. The no data category response to the situation where origin and / or destination is in accessibility category 6, where no ROI of a train station is in that area. The maybe category represents the original destination combinations where the travel time difference between car and train is between 0 and 10 minutes. This means that the train is no more than 10 minutes slower than the travel time of the car. If this is the case, the od combination comes in the no category. A threshold of 10 minutes is chosen because access and egress transport to and from the station is not calculated in the travel time of the train.

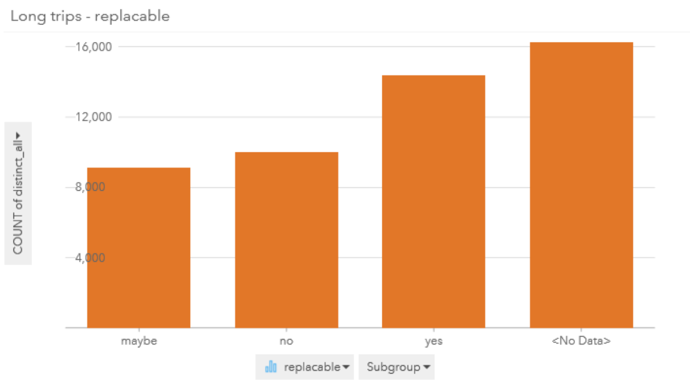


Figure 7.21: The number of trips replaceable by a sustainable aternative

In the selected areas for long distance trips, most trips are in the category 'no data' or 'yes' as shown in figure 7.21. Both LMS (figure 7.22) and Google (figure 7.23) data set are compared on there average weights of all categrories of replaceability.

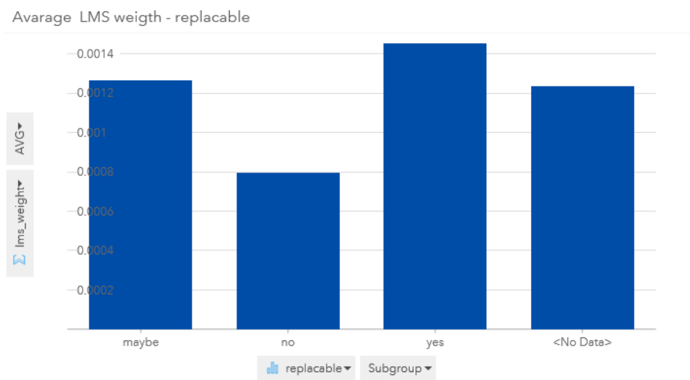


Figure 7.22: Average weight of LMS trips replaceable by a sustainable aternative

The average weight for all the different replaceable options for the LMS data set shows that trips with a higher potential for replacement by a more sustainable option have a higher average weight in the LMS data set. Trips that are less likely replaced by public transport have a lower average weight in LMS.

The average weights of the Google data set for all the replacabilty catagories is more striking. The average weight is almost similar in all categories. In the case of the accessibility categories this was also the case. This phenomena will be discussed further in section 7.2.8.

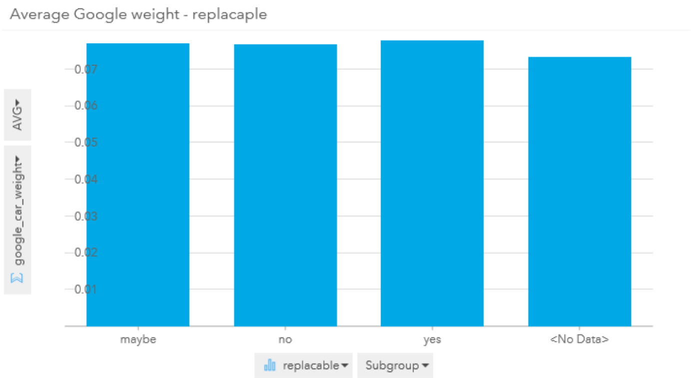


Figure 7.23: Average weight of Google trips replaceable by a sustainable aternative

7.2.5 Relationship weights to distance and duration of the trips

To see if there is a relationship between the weights of the trips measured in the LMS and Google data set, and the duration of the train connections to the corresponding origin destination combination, a correlation table is set up for both data sets. Figure 7.24 is showing the correlation between LMS weights and duration of train trips. It seems in this case that shorter train connection have a higher weight in LMS. This trend is also seen if the distance from midpoint to midpoint (calculated by Google API, along roads) is analysed (figure 7.25)

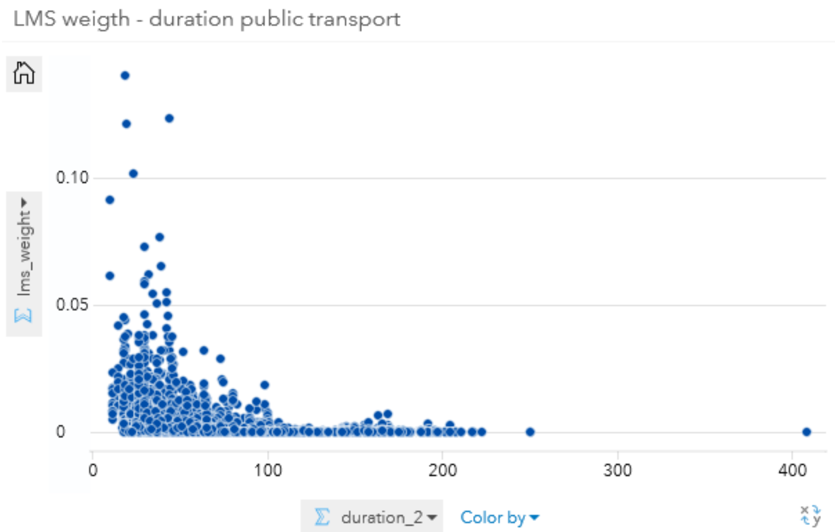


Figure 7.24: Correlation between the duration public transport and LMS weigthts

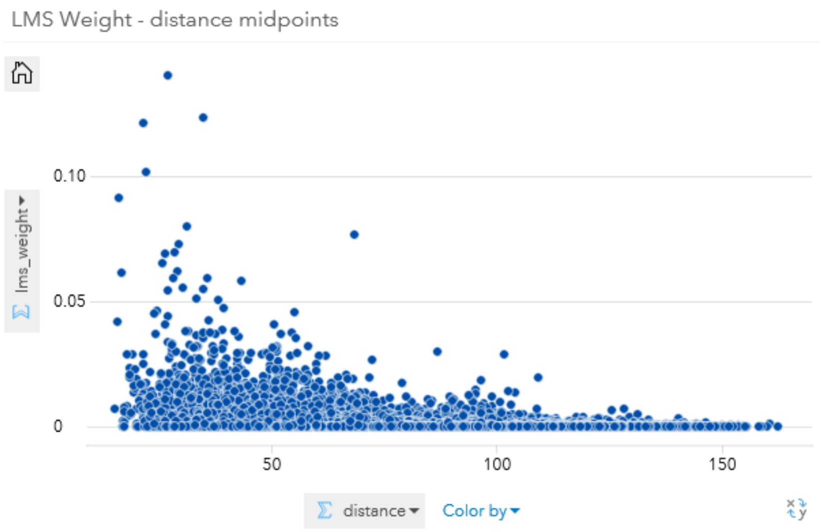


Figure 7.25: Correlation between the distance midpoints and LMS weights

For the Google data set the patterns as described before are different. The correlation between distance from midpoint to midpoint (figure 7.26) and the correlation between duration of the train connection(figure 7.25) is weak.

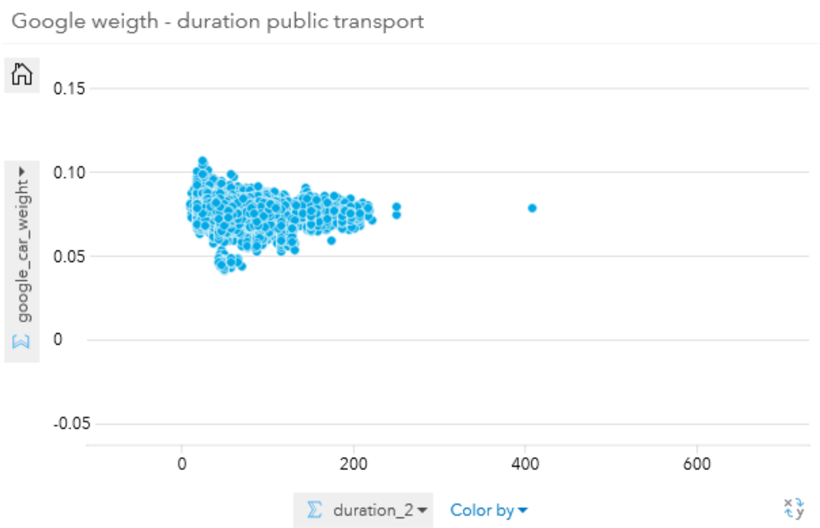


Figure 7.26: Correlation between the duration public transport and Google weights

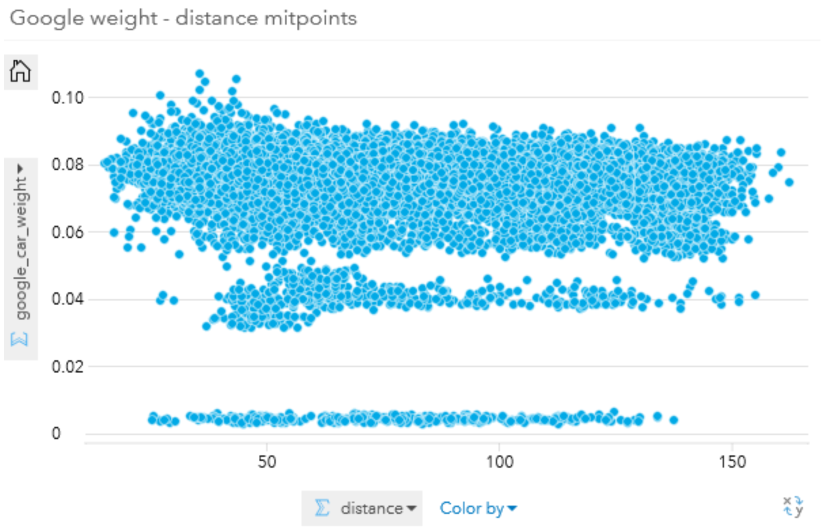


Figure 7.27: Correlation between the distance midpoints and Google weights

7.2.6 Highest weights in LMS and Google data sets

In this section the highest weight of trips in the LMS data set (figure: 7.28) and the highest values in the Google data set (figure: 7.29) are visualized. After the visalisation a comparason is done , and overlapping values are shown in figure: 7.30.

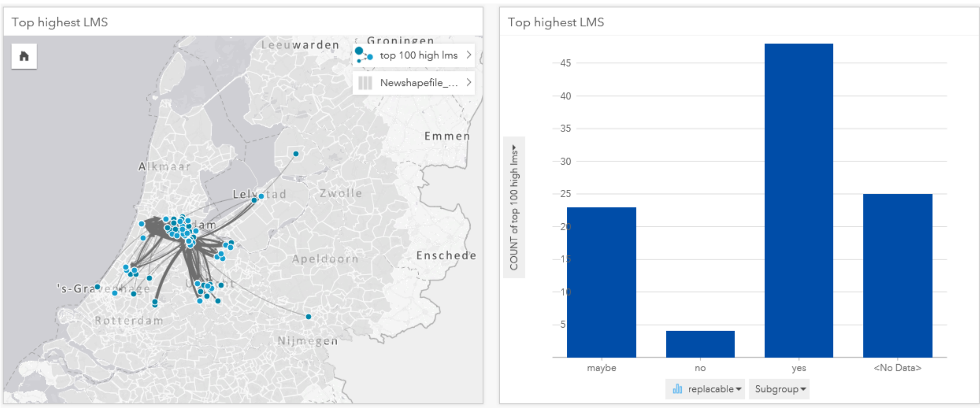


Figure 7.28: Top100 highest LMS weights in the longsdistance data set

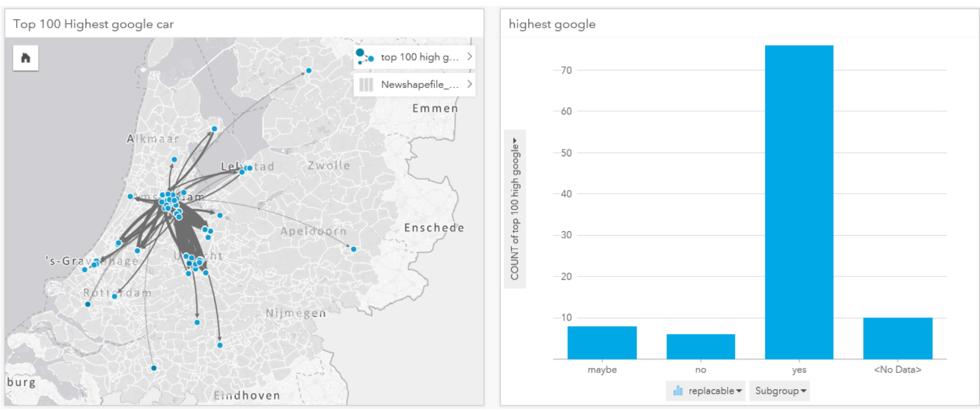


Figure 7.29: Top100 highest Google weights in the longsdistance data set

For both LMS and Google data set, the rides with the highest weights are most consistent with the origin destination combinations in the replaceable category. In the top 100 of the LMS data set, there are also more trips in the category maybe and no data than the Google data set. In both data sets, few journeys occur in the no category.

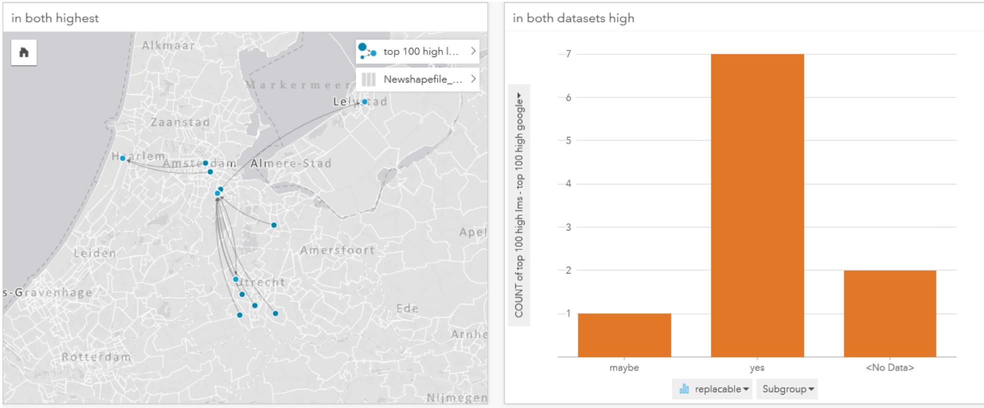


Figure 7.30: Corresponding trips in Top100 highest LMS weights and Top 100 higest Google trips in the longsdistance data set

Looking at the corresponding trips in both top 100 highest weight values, only ten are simular in both datsets. The trips that are corresponding are mosty trips that have a connection with the event area (Ziggodome, Bijlmer arena, AFAS live) and the area of Utrecht.

7.2.7 Lowest weights in LMS and Google data sets

After comparing the highest weights, now also the lowest weights are compared in the same way. In the LMS data set, the lowest weights all have value zero. It amounts to 1084 values with a zero value. It is therefore decided to compare the lowest 1084 values of both data sets with each other. The lowest values in the LMS data set are shown in figure 7.31. It shows that almost all trips with value zero are in a category that is seen as not replaceable ('No' and 'no data').

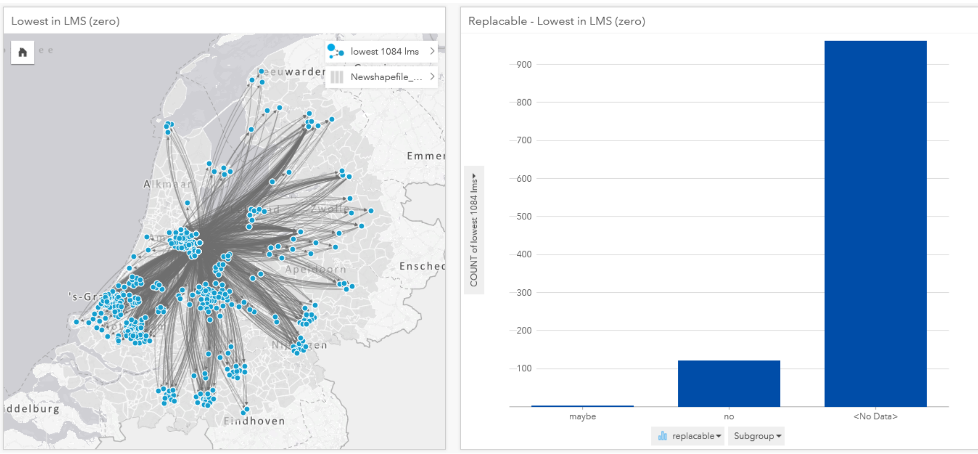


Figure 7.31: Lowest LMS weights in the longsdistance data set

The lowest values in the Google data set, shown in figure 7.33, give similar results. Most values are corresponding to the not replaceable category of 'no data'. This category represents the origin destination combinations where no traintrip is possible.

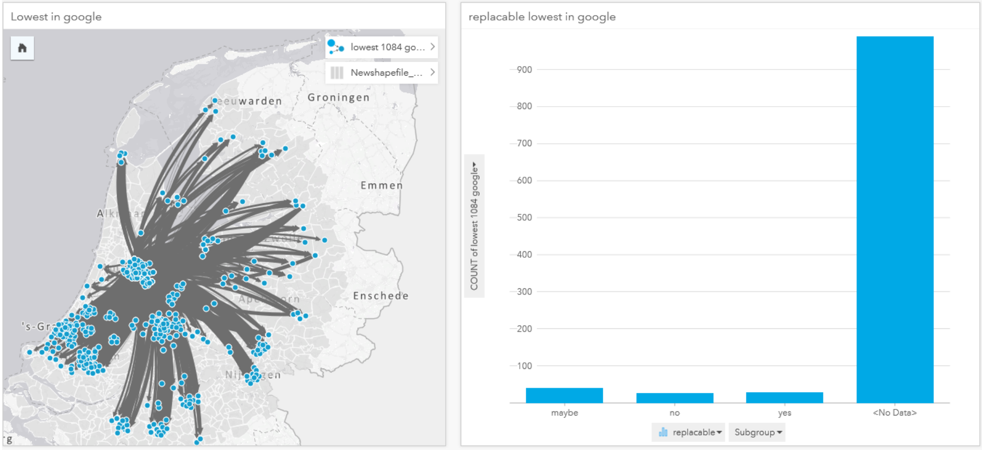


Figure 7.32: Lowest Google weights in the longsdistance data set

Looking at both high and low values of the two data sets, differences can be found between high and low values. In both high and low there is an overlap in the data that is shown in figure 7.33. The overlapping trips that have both high weights in both data set are mostly from the replaceable category. The overlapping trips in the lowest weights part are all not replaceable trips in the category 'no data', corresponding to no train connection available.

Figure 7.33: Corresponding highest and lowest values in LMS and Google data set

7.2.8 Google data remarkable notion

As shown in figure 7.34, there is something unexpected in the Google data set for long distance trips. Because of the remarkable results in figures: 7.23, 7.26 and 7.27, the Google data set is further analysed. Although it was outside the scope of the long distance trips, the column with the weights for cycling trips are still attached to the data set. By looking at the data a strong correlationship between these data is found. The long distance trips should not contain walking and cycling trips, but they are still in this data set. They have a very strong correlation which indicateds that there is some added noise to to the data.

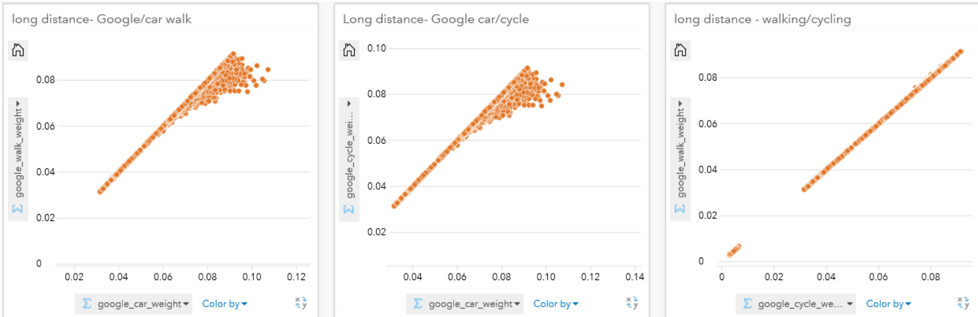


Figure 7.34: Correlation between: Google car/Google walk, Google car/Google cycle, Google walk/Google cycle

Analyzing the correlations related to the Google car weight in Figure 7.34, almost the same correlation is found. Some of the weights in the Google car weights differ a bit from the standard noise pattern, but most of the data is similar to the noise error. This value of the Google data looks less reliable than it was in the short distance results.

8 CONCLUSION AND DISCUSSION

To conclude this thesis, the main conclusions are presented in this chapter. First an answer to the research questions is given. Secondly the recommendations are given for further research in section 8.7.

In summary, the availability of travel data generated by mobile devices FCD provides researchers with new ways to understand transportation behavior and provides insight in the way people make choices for their transportation means. When understanding the decision making process, new ways can be found to stimulate more sustainable choises for transportation. Due to the amount of data sources available, this research has been focussing on answering five research questions related to FCD in relation to sustainable mobility. The questions raised in this thesis are:

1. What is sustainable travel behavior and why is this important?
2. Which data sets are commonly used and available to analyze travel behavior?
3. What are the differences between these FCD data sets and which ones suited best?
4. Which short distance car trips could be replaced by more sustainable opportunities like walking or cycling ?
5. Which long distance car trips could be replaced by more sustainable opportunities like public transportation?

In the coming sections each of these questions are answered. The answer to the first question has been found through a literature study. The second and third sub-question are done by comparing different available data sets and reading different literature studies. The answer to the final two sub-questions is found by doing experiments with the available data. After answering these sub-questions the general conclusion is stated and the conclusions will be discussed to finalay come to an answer on the main research question of these thesis:

To what extent can FCD be used to give an insight in the sustainable mobility behavior in Amsterdam?

After the discussion some recommendations will be discussed for futher research.

8.1 SUSTAINABLE TRAVEL BEHAVIOR AND IMPORTANCE

Due to the urbanisation, mobility issues are getting more important. Mobility of employees and goods make the economy more productive and other forms of mobility help sustain the social and cultural network [OECD, 2015]. Although there are positive effects, the growing size of cities and increasing population is resulting in a rapid increase in the number of vehicles on the roads [Djahel et al., 2015]. Transportation can have different impacts: environmental, social equity, economic, cultural, land use and urban form are the most important ones [Schiller et al., 2010].

There are multiple ways how sustainable mobility is defined in the literature. Generally speaking, sustainable mobility deals with the transportation of people and goods in relation to environmental and social impact. The definition used in this research is based on an adapted version of European Council [2001] of an expanded definition by Centre for Sustainable Transportation [1997]: "A sustainable transport system [is] defined as one that:

- allows the basic access and development needs of individuals, companies and societies to be met safely and in a manner consistent with human and ecosystem health, and promotes equity within and between successive generations;
- is affordable, operates fairly and efficiently, offers choice of transport mode, and supports a competitive economy, as well as balanced regional development;
- limits emissions and waste within the planet's ability to absorb them, uses renewable resources at or below their rates of generation, and, uses non-renewable resources at or below the rates of development of renewable substitutes while minimizing the impact on the use of land and the generation of noise." ([European Council, 2001] pp.15-16)

Sustainable mobility aims at promoting better and healthier ways of meeting individual and community transportation needs. It also reduces the social and environmental impacts of current mobility practices [Schiller et al., 2010].

The reason why sustainable travel behavior is important is the fact that different transportation methods have different impacts on society and the environment. With the increased demand for transportation over the years, the understanding of travel behavior and how this behavior can be influenced, is developed as a field of research. In terms of the usage of non-renewable energy usage the transportation methods of walking and cycling are considered the most sustainable, followed by public transportation. Travelling by car or airplane are the least sustainable options [Banister, 2009].

8.2 COMMONLY USED DATA TO ANALYZE TRAVEL BEHAVIOR

Understanding travel behavior and the reasons for choosing one mode of transport over another is a complex phenomenon. For every journey, new choices between different transport modes need to be made. The choice of one specific mode of transportation can vary over time and with the type of journey that is made [Beirão and Cabral, 2005]. Most of the research on travel behavior is related to psychological and social science. It tries to combine personal diaries with socio economic and demographic statistics [Oliveti, 2015]. While these methods provide some data to understand travel behavior, it is known that there is a gap between what correspondends report and the actual trips they made [Bohte, 2010]. With mobile devices present in almost every traveller's pocket, new data sets based on floating car data FCD become available [Lee et al., 2014; Moloo and Digumber, 2011]. Smartphones are able to collect information about the location of the user by making use of GNSS, Wi-Fi and inertial measurement units. Data collection using bluetooth techonology, wifi techonology or cell phone data have been used in previous research [Alexander et al., 2015; Duynstee et al., 2016; Braggaar, 2018; DATMobility, 2013]. For bluetooth data gathering, scanners are placed on strategic locations and cell phones connecting to bluetooth are registered. WIFI tracking can make use of the same principle, but it can use also the existing WIFI infrastructure, which is available on most locations. The use of cell phone data using cell phone tower triangulation has also been investigated, but the main limiting factor was the level of accuracy (hundreds of meters) [Ahson and Ilyas, 2010]. Using GPS the level of accuracy is improved (meters) and the need of placing devices on locations is not required. The drawback of using passive GPS data is that the data needs to be anonymized for the research by either aggregation or cutting off beginning and end points [Gambs et al., 2014]. Different companies provide applications such as Google maps and Flitsmeister collect traffic data and store this data in their databases. Because the majority of the people will carry a smartphone, the penetration rate of the sampling will be much higher than the OVIN research. The use of this new type of big data has been proposed as an alternative traffic sensing infrastructure, as they usually provide a cost-effective way to collect traffic data [Herrera et al., 2010; Moloo and Digumber, 2011]. This makes this data sets interesting for further research. In this research data collected by smartphones will be referred to as FCD data.

Another well-known research method is to collect data by making use of a sample group carrying GPS trackers around for a fixed amount of time (active GPS tracking). Examples of this kind of research are the graduation projects of Biljecki, Oliveti, Van der Winden and which makes use of the same GPS dataset collected by the University of Delft [Oliveti, 2015; Biljecki, 2010; van Winden, 2014; Bohte, 2010]. This data set contains GPS points with coordinates (x, y, z) and time (t) recorded every 5 seconds. The sample was taken by more than 800 people in several cities during a time span of two weeks [van Winden, 2014]. With this GPS tracks, individual traffic behaviour can be analysed. This method gives a more realistic view on traffic behaviour than the traditional method by making use of paper and phone call surveys. Because the participants will carry all the time a GPS receiver, small trips will be recognized and will not be missing in the data set. Another benefit of this type of research is that background information about the participants can be collected by questionnaires. This gives information which can be assigned to the traffic behaviour that is analysed.

Compared to the active GPS research FCD data can be collected over a longer period of time and gives this information over a wider group of users. Although a bigger group can be analysed over a longer period, there is less information available on the background of the group that is analysed. In GPS research normally the sample group provides personal information via questionnaires, which is in practise less feasible for FCD data research.

8.3 DIFFERENT DATA SETS AND BEST FITTING

There are many mobile applications available which store travel data of users and which can be used to explain travel behavior, all with their limitations and benefits. These are related to the way the data is received, data privacy limitations and/or sample sizes. To select a suitable data set for this research a list of four requirements has been prepared. First requirement is that the dat must be available and free of use for this research. Based on this requirement five data sets are considered in this research: Google, Flitsmeister, Ring Ring, GPS data collected TU Delft and Landelijk Model Systeem. The data sets were made available by TNO or by the TU Delft. The second requirement is that the data must show origin and destination information. This requirement is only met by three data sets: GPS data of TU Delft, Google OD and Landelijk Model Systeem. The data set of Flitsmeister and Ring Ring are both anonymized by cutting off the first and / or last part of the trip, and therefore do not met this requirement. For all data sets, documentation is available. So all data sets meet the third requirement. The amount of documentation differs per data set. The LMS dastaset is more intensely described in the literature than newer data set as Flitsmeister and Ring Ring. The last requirement states that the data should

cover at least the area of Amsterdam, but preferable available for a larger area. This requirement is only not met by the GPS data set of the TU Delft. That data set is only available in Amersfoort, Veenendaal and Zweewolde.
Based on the list of requirements, the Google OD and LMS are chosen to be used for further analysis because these data sets met all the requirements formulated in this research.

8.4 REPLACEABLE SHORT DISTANCE TRIPS

From the Google and LMS data set the short trips are selected and used in the experiments. The results of this experiment of short distance trips could be found in Section 7.1. With the help of the calculated distances and travel time with the Google API it was possible to calculate the replaceable origin destination combinations in the short data set based on the disision tree described in Section 6.3.
First the walking trips are analyzed from the Google data set. For walking trips the highest weights are founded on trips that have the same origin and destination. There seems to be a high correspondence between the Google walking data and the activities in the city. The areas with the highest activity of walking are linked to the historical city centre, Museum Square, Zuidas business area, Ziggodome, Bijlmer arena, AFAS live. The relationship between google data and the Google api for walking is evident. From the 111 options where walking was an option, the Google OD data set showed also more walking trips. Only in 13 cases of the calculated walkable options, the car was more chosen option in the Google OD data set. In two cases another sustainable option (cycling) was chosen. There are also examples found in the data set where walking should not be an option, but from the Google OD data set walking is still the favorite option. For most unexpected walking trips the distances are 2 km or less. This phenomenon occurs 84 times in the data set. These OD combinations have a duration for walking that is higher than 15 minutes. In the analysis, the acceptable walking time has a maximum of 15 minutes, but after looking at the data a peak at 17 min is shown. Looking at the results in the Google OD table, the boundary of the dicision tree shown in figure 6.4 should be a little higher than 15 minutes. The difference between the expected travel time in the disision tree and the action measurements in the Google origin destination matrix can be caused by the way the travel time is measured. The travel time with the Google api is only measured from midpoint to midpoint of an area. The measured travel time in this case can be longer than the actual travel time of the trip that is made.

The selection of cycle trips in the Google data set shows that there are many trips that have less travel time by bicycle than by car. These OD combinations have a potential that car trips could be replaced by cycling trips. The cycle trips that have a duration longer than 15 minutes compared to the car trip duration, are almost all trips with a distance longer than 10 km. This is caused by the way the selection is made for short distance trips and the possibility that a detour is taken because for example a bridge over the water needs to be crossed. Based on the decision tree, different replaceable categories are set up for cycling trips. For the category yes, where trips are shown that have cycling as an sustainable alternative, the google data shows only 724 od combinations of the 3241 in the category yes that have the favorite of cycling. It is striking that some origin-destination combinations where cycling did not seem an alternative, still shows more bicycle trips than car trips. The distance of these unexpected cycle trips are in 7 cases between 10.3 and 13.3 km cycling distance. This could be the same problem as for the unexpected walking trips, where the distance is measured from midpoint to midpoint. The distance for replaceable cycle trips could be a little bit higher in further research to avoid this cases.

For the car trips it was possible to find the replaceable origin destination combinations with the setup rules from the implementation. Five categories have been created to show when a trip is considered replaceable. With the selection made for short distances, most replaceable trips are started or ended in the city centre of Amsterdam. This result is visible in both Google and LMS data sets. The highest weights of the replaceable trips are those with the same origin and destination. This phenomenon is more visible for the Google data set than for the LMS data set, where other origin-destination combinations may also be higher. For the data where origin and destination are not comparable, Google and LMS show different patterns. In Google the replaceable trips with high car weights are concentrated in the north, west and east part, outside the citycentre of Amsterdam. In the LMS data set a totally different connection is visible for replaceable car trips. The replaceable car journeys with high weights are usually between the center and Amsterdam south. A few trips have corresponding high weights for replaceable trips in both data sets. These are in the northpart of the city centre and Amsterdam West. The heighest weights are in the Google datatset mostly shorter trips than the LMS data set.

8.5 REPLACEABLE LONG DISTANCE TRIPS

For the long distance, a smaller selection of the data is made. This selection was necessary to limit the use of the Google API. Different cities have been selected due to differences in categories, accessibility of stations, city size and population density of the area. Looking at the replaceability of long-distance travel, train travel is seen as the most sustainable alternative for long distance car journeys. There is however a trade off between the most sustainable and least time consuming transportation option. This makes that train trips are not always the most logical or even sustainable choise for transportation. In this research a method is developed and applied to the data set to test if a train trip is a realistic alternative.
First the correlation is calculated between the Google and LMS data set. Using the areas in scope (Amsterdam), for every long distance connection the correlation coefficient is 0.15, meaning that there is a weak correlation between the two data sets for longer distances.

Looking at the accessibility for the selected areas, there is a higher response to category one values compared to all areas in the Netherlands. This means that the selected areas are more connected to the train network than the average in the Netherlands. If the areas of Amsterdam are only selected, a even higher response is visible in the first catagory. This high response is probably because there are 9 train stations in Amsterdam at the time of this study. As a result, ROIs overlap easily and many areas are rerelated to at least one ROI. The research area is one that could have potential more replaceable trips than the average area in the Netherlands. Analyses were done to see whether there is a relationship between accessibility and the weights in the LMS and Google data set. The average weight value is calculated for each combination. The LMS areas do not show a very clear relationship. It can be seen that combinations that include a trip with a connection to an area in catogory 3 have slightly lower weights on average. Journeys where there is a connection from or to a catagory 2 area, score a little higher on average. The Google data set shows a completely different story, all combinations look similar to each other. There seems no relationship between high weights in the data sets and a low accesability for train journeys. It was expected that connections from or to a catorgy 6 area (no trainstation) have more car trips, but this is not found in the data. Looking at the maximum weight values in the LMS data set for all combinations it is found that combinations with category 2 and 5 or catagory 1 and 2 are having the highest weights. These areas have one or more ROIs overlapping with the selected area for 25% till 100%.

To see if a trip is replaceable, travel time is an important measurement. The travel time is calculated for both train and for car. In this analysis it is found that there are some connections that are more than 15 minutes faster than a car trip. These connections are mostly connections to big cities or areas close to Amsterdam.

Four different catagories are set up to check the replacability: yes, no, maybe, no data. To setup these catagories the travel time for train and car are compared. Most of the trips are in the no data category, which means that at least one area is not connected to a ROI of a trainstation. Both LMS and Google data sets are compared on there average weights of all categrories of replaceability. LMS data set shows that: trips with a higher potential for replacement by a more sustainable option have a higher average weight in the lms data set. Trips that are not likely replaced by public transport have a lower average weight in LMS. This means that there are more car trips on potential replaceable trips than on not likely replacale trips. The average weights of the Google data set for all the replacabilty catagories is striking. The average weight is almost similar in all categories so that no relation can be found.

Whether there is a connection between the weights of the journeys measured in the LMS and Google data set, and the duration of the train connections, analyses have been carried out. In the LMS dataset it seems that shorter train connection have a higher weight in LMS. This trend is also seen if the distance from midpoint to midpoint is analysed. For the Google data set the patterns are different. The correlation between distance from midpoint to midpoint and the correlation between duration of the train connection is weak.

Comparing the highest values in both Google and LMS data sets, it seems that LMS has higher weights on shorter trips than the google data set. For both LMS and Google data set, the rides with the highest weights are most consistent with the origin destination combinations in the replaceable category. As said before there is a low correlation between the Google and LMS data set. This is also visible when the top 100 highest values are compared for long distances. Only a few trips are corresponding, mostly connections with the event area (Ziggodome, Bijlmer arena, AFAS live) and the area of Utrecht.

Comparing the lowest values of both data sets, the most journeys are in the no data catagory. These are trips to areas where no connection to a train station is possible. The corresponding lowest values are all over the Netherlands but always found in the no data catagory, which are representing the best accesable areas by train.

During the analysis of the data for long distances, a problem has been found with the Google data set. It seems that this data set is showing most of the time the same noise. This phenomenon was found comparing the different mode of transportations available in the google data set. The error is not in all of the data, but it seem to have an impact on the data set.

As described before, some unexpected results came up anlalysing the Google data set for long distances. Looking at the data a strong correlation between the car, walk and cycle data is found. This correlation indicates that there is added some noise to the data.

8.6 GENERAL CONCLUSION AND DISCUSSION

The aim of this thesis is to investigate whether FCD data can be used to gain insight into the sustainable mobility behavior in Amsterdam. Both Google and LMS data sets are analysed for short and long distances. The method that is developed in this reseach to identify replaceable trips for unsustainable mobility behavior could be used for long and short distance trips. With distance and travel time calculation different catagories could be identified. Also the accesability of trainstations for every area is succesfully calculated. With this information the FCD data could be furher analysed and compared to the LMS model. For short distances there seems a correlation between the two data sets. Most replaceable trips are started or ended in the city center of Amsterdam according to the Google and the LMS data set. In both data sets is found that the highest car weights of the replaceable trips are those with the same origin and destination. For replaceable trips where the origin and destion is not simular, some different patterns are found but also some corresponding. For long distances, the reliability of the data seems less than for short distances in both data sets. The Google data set seems to have a lot of noise in the long distance selection. This makes it hard to use in the research to get insights in the sustainable mobility behavior in Amsterdam. To get more insights on this noice futher reseach needs to be done. The Google data set is orginally aggregated hourly. It could be possible that trips longer than hour are cut off or not visible in the data set. This should be futher investigated. The LMS data set has also very low weights on long distance trips. This is not futher analysed in this research but could be interesting for coming research.

Altogether it looks like both data set are usefull for analysing the sustainable mobility behavior in Amsterdam for short distance trips and for long distance trips furter research is needed. The method developed to identify replaceable trips could still be used for this analysis.

The data provides many insights on travel behavior and different availalbe data sets can be linked together to provide deeper insights. The Google data set shown interesting results for shorter distances, but gives less reliable results for longer distances. For further research it would be interesting to further investigate the other data sets available and find a standardized way of connecting them together to create a better picture of the reality. Also taking data from public transportation providers into account would improve the results.

There are several things to consider when analyzing FCD data. Interpreting the results should be done carefully. Distinguishing the cause and effect can be difficult for some cases, especially when correlating demographic data with behavior. Another topic is the reliabilty of the data itself. As data sets are obtained on an already aggregated level, the source data and key metrics are not reveiled (like exact coutings and personal tied data).

8.7 RECOMMENDATIONS FOR FURTHER RESEARCH

Based on the results of this research, several recommendations for further reasearch have been identified. The following recommendations could be given for futher research:

- Due to the results for the Google data set in the long distance selection, further research could be done to investigate the noise problem. Having a deeper understanding of this noise may make it possible to perform further data cleansing and may provide better data. This can give probably more insights how usable this data set is for longer distances.

- For further research it could be interesting to see why some origin destination combinations have walking as favorite in the Google data set, while the walking time >30 min.
- Both data sets have a low response on longer distances, it could be investigated more to see how reliable the long distance data is.
- This thesis gives a general overview of the replaceability of trips in the data set. It would be interesting to look more in to specific cases where trips could be replaced.
- It could be intresting to see if there is a relationship between the travel data and the demographic data of the different areas.
- For this thesis it was not possible to analyse public transport data, it would be interesting to add this data if it is available.
- It could be intresting to investigate the trip purpose for the different replaceable trips. This data is available for the LMS data and could be furter investigate.
- The influence of the weather was outside the scope of this research, but has probaly an influence on the choise towards sustainable transportation. This meteorological data is interesting to investigate in future research.

BIBLIOGRAPHY

Ahson, S. A. and Ilyas, M. (2010). *Location-based services handbook: Applications, technologies, and security*. CRC Press.

Alexander, L., Jiang, S., Murga, M., and González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies*, 58:240–250.

Bakri, T. (2016). Aanpak Assessment OD data Google Beschrijving van de OD data zoals ingewonnen door Google. Technical report, tno.

Banister, D. (2005). *Unsustainable transport: city transport in the new century*. Routledge.

Banister, D. (2009). Sustainable transport and public policy. Transportation Engineering and Planning-Volume II, 192. II.

Beirão, G. and Cabral, J. S. (2005). 9 TH CONFERENCE ON COMPETITION AND OWNERSHIP IN LAND TRANSPORT MODELLING SERVICE QUALITY FOR PUBLIC TRANSPORT CONTRACTS : ASSESSING USERS '. pages 1–15.

Bertolini, L., Le Clercq, F., and Kapoen, L. (2005). Sustainable accessibility: a conceptual framework to integrate transport and land use plan-making. two test-applications in the netherlands and a reflection on the way forward. *Transport policy*, 12(3):207–220.

Biljecki, F. (2010). automatic segmentation and classification of movement trajectories for transportation modes.

Black, W. R. (2010). *Sustainable transportation: problems and solutions*. Guilford Press.

Bohte, W. (2010). *Residential self- selection and travel*.

Braggaar, R. C. (2018). Wi-fi network-based indoor localisation, the case of the tu delft campus.

Brands, T., De Romph, E., Veitch, T., and Cook, J. (2014). Modelling public transport route choice, with multiple access and egress modes. *Transportation research procedia*, 1(1):12–23.

Bruun, E., Schiller, P. L., and Litman, T. (2012). An introduction to sustainable transportation: Policy, planning and implementation.

CBSl (2015). Onderzoek Verplaatsingen in Nederland 2014, Centraal Bureau voor de Statistiek. pages 1–47.

Centre for Sustainable Transportation (1997). Definition and vision of sustainable transportation.

Centre for Sustainable Transportation (2002). Definition and vision of sustainable transportation. toronto, canada: The centre for sustainable transportation, mission statement.

Curtis, C. (2011). Integrating land use with public transport: The use of a discursive accessibility tool to inform metropolitan spatial planning in perth. *Transport reviews*, 31(2):179–197.

DATMobility (2013). Verbeter de toegankelijkheid, veiligheid en leefbaarheid. analyseer, monitor en plan gebaseerd op meerdere databronnen.

Djahel, S., Doolan, R., Muntean, G.-M., and Murphy, J. (2015). A communications-oriented perspective on traffic management systems for smart cities: Challenges and innovative approaches. *IEEE Communications Surveys & Tutorials*, 17(1):125–151.

Duynstee, C. A. N. L., Haayen, M. J., Murga, Kyritsis, D., Ortega-Cordova, L. M., and Samat, S. N. N. (2016). Synthesis project dordrecht final report smart city dordrecht – identification of pedestrian movement patterns with wi-fi tracking sensors.

European Commission (2007). *Towards a new culture for urban mobility*. Office for Official Publications of the European Communities.

European Council (2001). 2340th Council meeting—Transport/telecommunications— Luxembourg, April 4–5, 2001, pp. 15–16.

Flitsmeister (2020). Privacystatement Flitsmeister. https://www.flitsmeisterapp.com/#/nl/privacy. Accessed: 2020-01-09.

Freedman, D., Pisani, R., and Purves, R. (2007). *Statistics: Fourth International Student Edition*. International student edition. W.W. Norton & Company.

Gambs, S., Killijian, M.-O., and del Prado Cortez, M. N. (2014). De-anonymization attack on geolocated data. *Journal of Computer and System Sciences*, 80(8):1597–1614.

Gil, J. (2016). *Urban Modality: Modelling and evaluating the sustainable mobility of urban areas in the city-region*.

Gudmundsson, H., Marsden, G., and Josias, Z. (2016). Sustainable transportation: Indicators, frameworks, and performance management.

Halden, D. (2002). Using accessibility measures to integrate land use and transport policy in edinburgh and the lothians. *Transport policy*, 9(4):313–324.

Hall, R. P. and Sussman, J. M. (2006). Promoting the concept of sustainable transportation within the federal system-the need to reinvent the us dot.

Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., and Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568–583.

Jonkeren, O., Harms, L., Jorritsma, P., Huibregtse, O., and Bakker, P. (2018). Waar zouden we zijn zonder de fiets en de trein? *Kennisinstituut voor Mobiliteitsbeleid*.

Kowalczyk, U. (2010). Study report on IMPLICATIONS OF THE EU TRANSPORT POLICY ON DEVELOPMENT OF SUSTAINABLE TRANSPORT.

Lee, J.-G. and Kang, M. (2015). Geospatial Big Data: Challenges and Opportunities. *Big Data Research*, 2(2):74–81.

Lee, S., Tewolde, G., and Kwon, J. (2014). Design and implementation of vehicle tracking system using gps/gsm/gprs technology and smartphone application. In *2014 IEEE world forum on internet of things (WF-IoT)*, pages 353–358. IEEE.

L.Tavasszy, M.Snelder, M.Duijnisveld, R.Haaijer, H.Meur, van R.Nes ; E.Verroen, Schie, C., J.Bates, and B.Jansen (2012). Audit LMS en NRM Syntheserapport.

McKinsey (1989). Kiezen voor openbaar vervoer: "ov maal twee".

Michailidou, G. (2019). The influence of the visible views on cyclists' route choices: A geospatial approach for the measurement of the determinants in the urban environment based on 3D isovists and cyclists' GPS trajectories in Amsterdam. *thesis tu delft*.

Moloo, R. K. and Digumber, V. K. (2011). Low-cost mobile gps tracking solution. In *2011 International Conference on Business Computing and Global Informatization*, pages 516–519. IEEE.

OECD (2015). *Environmental Performance Reviews, The Netherlands*.

Oliveti, M. (2015). *Analysis of mobility patterns in different neigbourhoods, integrating GPS tracks with OpenStreetMap data*. PhD thesis.

Rathore, M. M., Ahmad, A., Paul, A., and Rho, S. (2016). Urban planning and building smart cities based on the Internet of Things using Big Data analytics. *Computer Networks*, 101:63–80.

rijkswaterstaat (2017). Verkeer en vervoer : Landelijk Model Systeem (LMS).

rijkswaterstaat (2018). Nederlands regionaal model (nrm) en landelijk model systeem (lms).

RingRing (2020). Privacystatement Ring Ring. https://ring-ring.nu/disclaimer/. Accessed: 2020-01-22.

Schiller, P. L., Bruun, E. C., and Kenworthy, J. R. (2010). *An introduction to sustainable transportation: Policy, planning and implementation*. Earthscan.

Toth-Szabo, Z., Várhelyi, A., Koglin, T., and Angjelevska, B. (2011). *Measuring Sustainability of Transport in the City: A Development of an Indicator-set*. Department of Technology and Society, Lund University.

van de Coevering, P. and Maat, K. (2013). De invloed van temporele dynamiek op de relaties tussen attitudes, de gebouwde omgeving en verplaatsingsgedrag. *CVS: Rotterdam*.

van der Spek, S., van Schaick, J., de Bois, P., and de Haan, R. (2009). Sensing Human Activity: GPS Tracking. *Sensors*, 9(4):3033–3055.

Van Goeverden, C. and Van Den Heuvel, M. (1993). De verplaatsingstijdfactor in relatie tot de vervoerwijzekeuze. *LVV rapport, VK 5304-301*.

van Grieken, S. (2016). Getting Started Guide. pages 1–25.

van Winden, K. (2014). Automatically Deriving and Updating Attribute Road Data from Movement Trajectories. *thesis tu delft*.

Vries, J. (2012). Provinciale Toepassingen voor Wegverkeersgegevens: Gebeurtenisdetectie op basis van NDW Verkeersdata.

World Business Council for Sustainable Development (2001). Mobility 2001: World mobility at the end of the twentieth century and its sustainability.

A | LMS AN GOOGLE TOP 20 ORIGIN DESTINATION RELATIONSHIPS

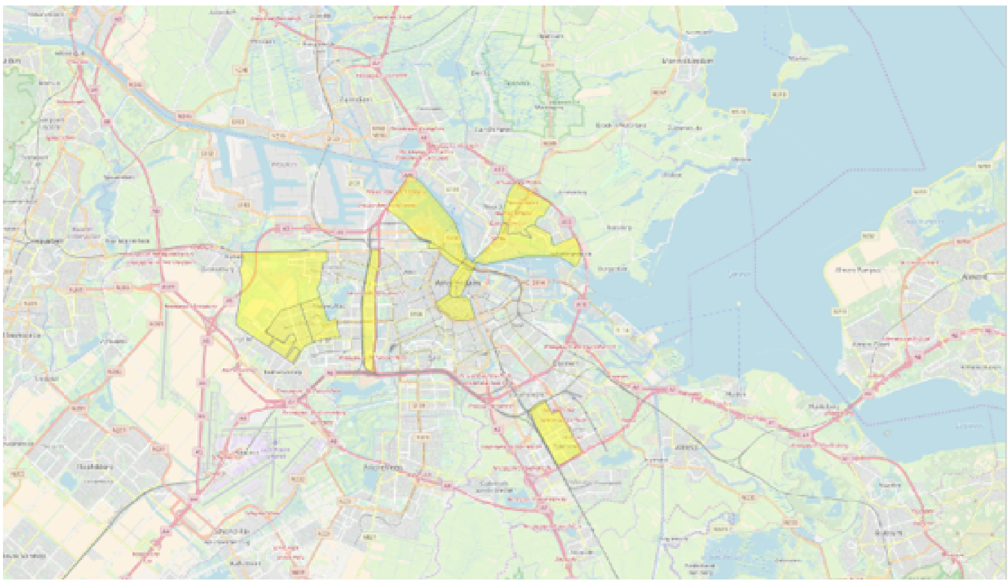
	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS

Top 20 Highest origin destination combinations in LMS and google.



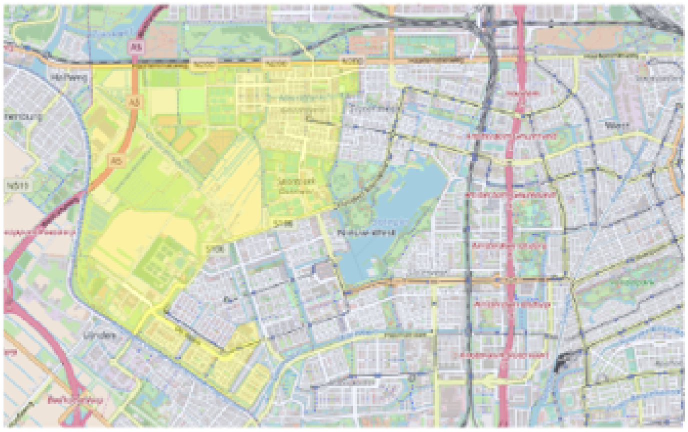
Overlapping results in Google and LMS top 20.

	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS



	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS

	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS

	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS

	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS

	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS

	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

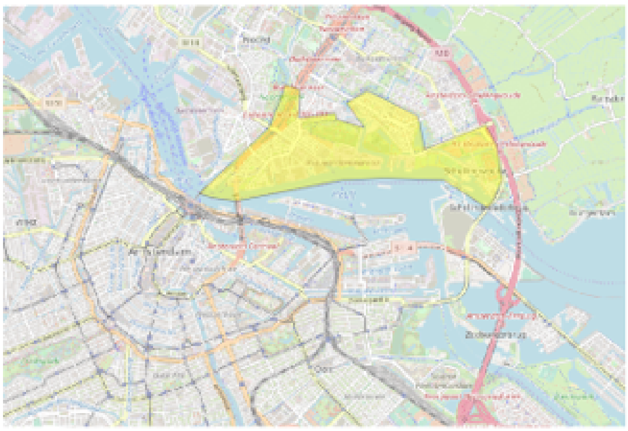
LMS

	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

LMS



	origin integer	destination integer	newweight numeric
1	626	626	1.00000
2	672	672	0.90306
3	666	666	0.85539
4	616	616	0.77434
5	663	663	0.76088
6	620	620	0.73363
7	674	674	0.65197
8	612	612	0.64668
9	676	676	0.64331
10	617	617	0.61114
11	624	624	0.59724
12	665	665	0.58400
13	622	622	0.57291
14	618	618	0.53721
15	664	664	0.53092
16	654	654	0.51161
17	625	625	0.49495
18	659	659	0.48977
19	619	619	0.47890
20	629	629	0.44717

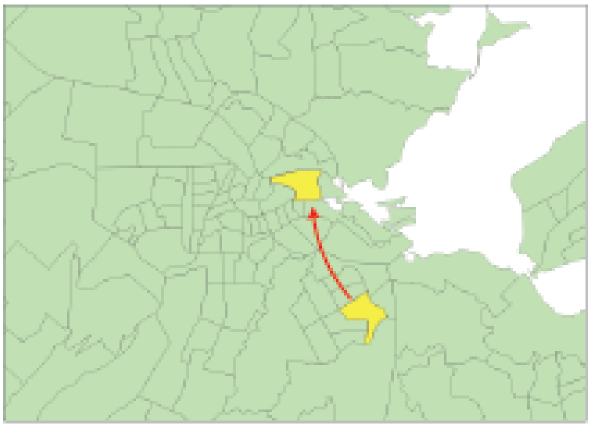
Google

	ori integer	dest integer	newweight numeric
1	676	689	1.00000
2	675	687	0.73629
3	666	666	0.65414
4	626	626	0.64339
5	629	629	0.52799
6	665	665	0.46276
7	500	671	0.41504
8	630	630	0.40295
9	669	689	0.33971
10	638	657	0.32903
11	638	638	0.32729
12	673	689	0.32132
13	618	618	0.31777
14	664	664	0.31562
15	672	672	0.30367
16	612	612	0.28764
17	659	659	0.27643
18	620	620	0.26087
19	690	663	0.26016
20	626	629	0.25574

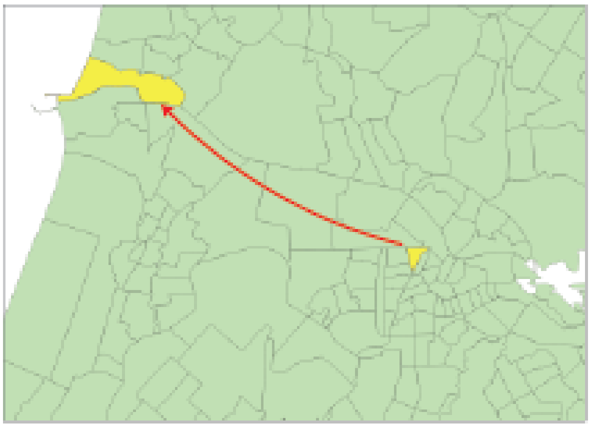
LMS

B CYCLING- UNEXPECTED HIGH TRIPS

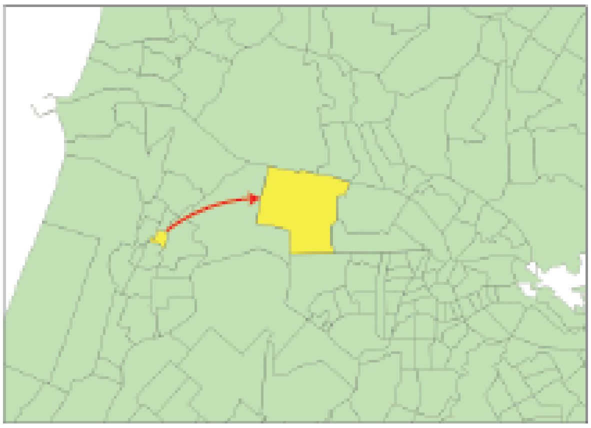
where bikeoption = no and favorite = cycling (unexpected high cycling)



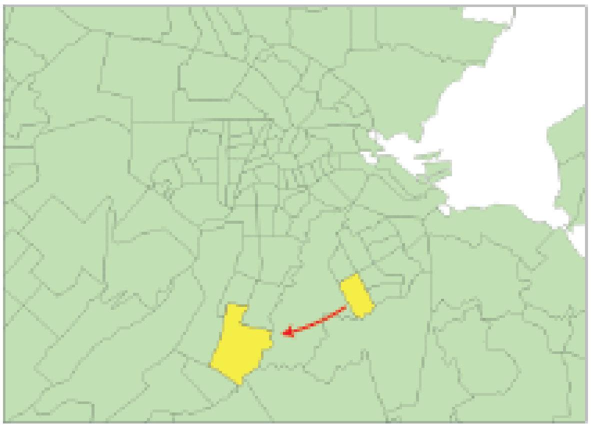
A. 669 –> 624 , distance 11.1 km, 39 min biking



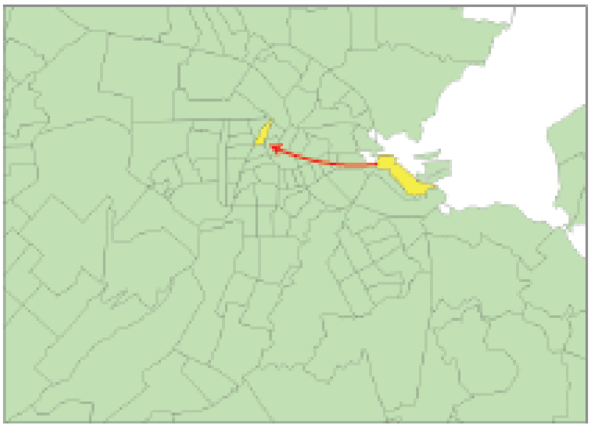
B. 642 –> 567 , distance 25.7 km, 87 min biking



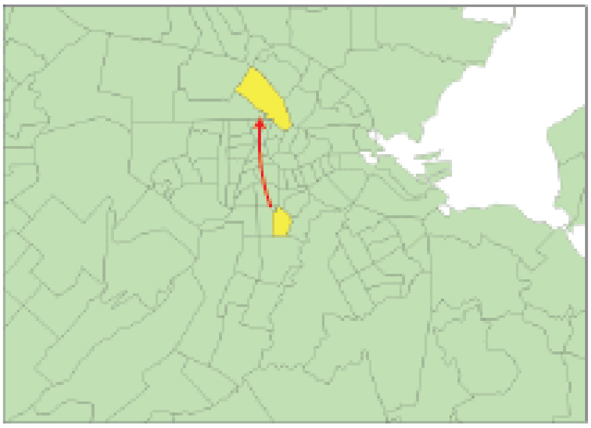
C. 578 –> 619 , distance 12.4 km, 37 min biking



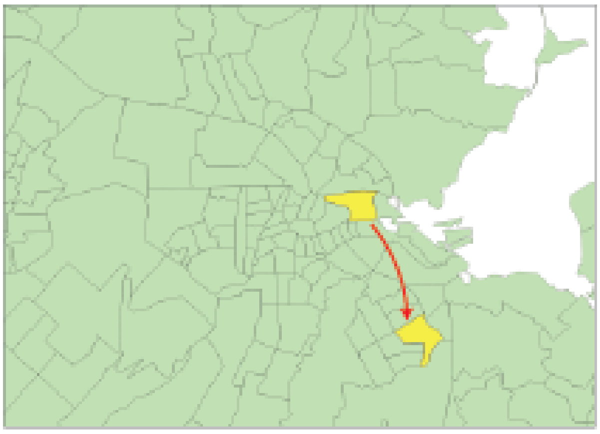
D. 677 –> 607, distance 12,1 km, 37 min biking



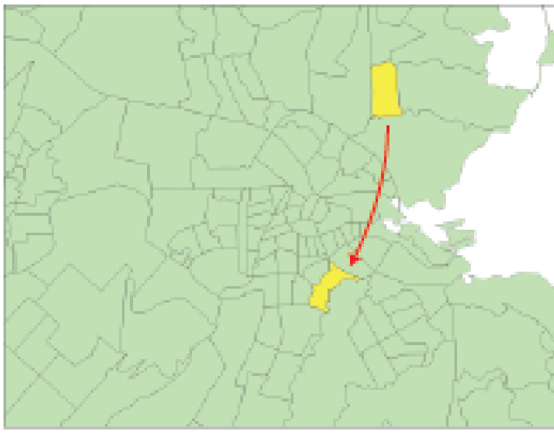
E. 622–> 641, distance 11.3 km, 39,7 min biking



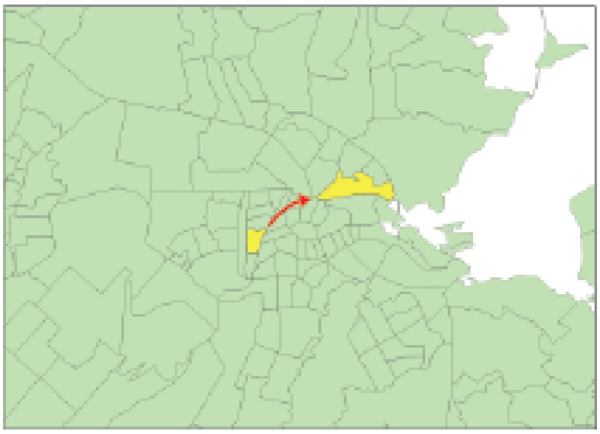
F. 656 –> 620, distance 11,0 km, 37 min biking



G. 624 –> 669 , distance 11,7 km, 37 min biking



H. 713 –> 653 , distance 17,3km, 55,6 min biki



I. 646 –> 618 , distance 10,3 km, 34,6 min biking

C SELECTED MUNICIPALITIES FOR EXPERIMENTS LONG DISTANCE TRIPS

The following municipalities are selected in this research for long distance trips:

- ’s-Gravenhage,
- ’s-Hertogenbosch
- Alphen aan den Rijn
- Arnhem
- Barendrecht
- Beemster
- Best
- Blaricum
- Bloemendaal
- Breda
- Bunnik
- Dalfsen
- De Bilt
- Delft
- Den Helder
- Elburg
- Enkhuizen
- Epe
- Ermelo
- Franekeradeel
- Gouda
- Harderwijk
- Harlingen
- Heerenveen
- Hilversum
- Houten
- Huizen
- IJsselstein

- Laren
- Leiden
- Leiderdorp
- Leidschendam-Voorburg
- Lelystad
- Medemblik
- Meppel
- Montfoort
- Neerijnen
- Nieuwegein
- Nijkerk
- Nijmegen
- Noordoostpolder
- Noordwijk
- Nunspeet
- Oegstgeest
- Olst-Wijhe
- Ommen
- Ridderkerk
- Rijswijk
- Rotterdam
- Rozendaal
- Sudwest Fryslan
- Stede Broec
- Stichtse Vecht
- Tilburg
- Urk
- Utrecht
- Utrechtse Heuvelrug
- Vianen
- Waalwijk
- Wageningen
- Wassenaar
- Westland
- Woerden

- Zaltbommel
- Zeist
- Zoetermeer
- Zutphen

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.