Potential use of public transport in the context of air pollution reduction |

Development of a tool to determine and localize potential bus users.

Final Colloquium

Final Presentation Date	|	25th of February 2015
Location	|	Auditorium Room 4, TU/e Campus

Author

Graduation candidate	|	Hannely Hortensius BSc
Student ID	|	0714758

Graduation Committee

Prof. dr. ir. W.F. (Wim) Schaefer	|	Chairman Master Program CME, TU/e
Dr. ing. P.J.H.J. (Peter) van der Waerden	|	First graduation supervisor, TU/e
Dr. Q. (Qi) Han	|	Second graduation supervisor, TU/e
Ir. J.J.F. (Joris) Hoogenboom	|	Consultant, BonoTraffics bv.

Institutional information

Study Program	|	Construction Management &Engineering
Faculty	|	Faculty of the Built Environment
Institute	|	Eindhoven University of Technology

Company

Company name	|	BonoTraffics bv
Address	|	Ijsselkade 60, 8261 AH, Kampen

PREFACE

This graduation project concludes my master 'Construction Management and Engineering' at Eindhoven University of Technology. Before this master track I obtained my Bachelor 'Built Environment' (Bouwkunde) at this same university.

My main supervisor during this graduation research was dr. ing. P.J.H.J. van der Waerden, my second supervisor dr. Q. Han and the chairman of the supervisory was prof. dr. ir. W.F. Schaefer. During this project I was supported by BonoTraffics bv. represented by ir J.J.F. Hoogenboom.

Hereby, I would like to take the opportunity to thank you all very much. Especially Peter, who was involved from the beginning, with the defining of the research, and helped me very much during the whole process with his never-ending enthusiasm.

The subject regarding the effect public transport can have on the environment came first to mind on the 14th of March 2014, when Paris decided to offer public transport services for free for a whole weekend to avoid people to take their car. My curiosity was aroused by this measure, could this really be a remedy of the air pollution? However my research proposal was very broad, and no specific scientific issue could be withdrawn from this topic. Peter helped narrowing the subject down, and this resulted in a research which Peter already was involved with in the past. With an eye to the practical value of the thesis BoboTraffics bv. got involved. Joris helped multiple times during the process with new insights and brainstorm sessions.

Furthermore, I would like to thank (all of) my parents, which supported me during my whole study time, and which ensured I could enjoy my student years to the fullest. And last but certainly not least my dear Jeroen, who sometimes had to keep up with some bad mood swings, but he kept supporting me all the way to the end.

If there is one thing I learned during my study years it is the following:

"You can be creative in anything – in math, science, engineering, philosophy – as much as you can in music or in painting or in dance." – Sir Ken Robinson

I came to Eindhoven to study architecture, because of the creative aspects. Now, I am a graduated engineer, and can still implement my creativity in all projects.

Although, I am really glad I did graduate at this moment, I enjoyed working together with all people who were involved during my study time. I hope you enjoy reading the result of these years in this thesis.

Hannely

CONTENTS

Chapter One | Introduction

This first chapter gives an introduction into the problem which is examined during this research. The background for the problem is given, from this the problem definition was subtracted and the research questions were formulated to substantiate the problem discussed in this research. Next to this, the relevance of the research is emphasized and the expected results are stated. The chapter finishes with a reading guide for the remaining of the thesis.

One.1 | Problem Introduction

In Western Europe very high air pollution levels are measured and since the need for transport only grows, higher levels are expected. In the near past especially France, Belgium and Germany suffered from very bad air quality, but if no measures are taken the Netherlands will surely follow their paths (European Environment Agency, 2014).

Also worldwide, the air pollution in growing cities is a significant problem (Sierra-Vargas and Teran, 2012). Due to air pollution more and more people suffer from morbidity and, even worse, mortality because of the effects of air pollution on health (Miller, 2004)(European Environment Agency, 2013)(Künzli et al., 2000)(Knol, 2005). By 2020 an estimated 80 percent of the population will live in urban areas, which is 5 percent more than in 2010 (Chen et al., 2006)(European Environment Agency, 2010). This increasing number of citizens moving to cities will most likely worsen the case, because more people result into more required transport, which leads to more emissions. These are harmful situations, which should be resolved.

Sustainability has a growing interest all over the world, companies are profiling themselves as green and sustainable, and also governments are joining this trend. Both the European Union and the government of the Netherlands set targets stating that by 2030 the air pollution has to be decreased (European Environment Agency, 2010)(ANP, 2013). Therefore, national government and local governance, like the municipalities, are taking steps to reach this goal. The government is already using all kind of policies regarding emission sources (Granier et al., 2011), engine techniques (Bellekom et al., 2012), and different uses of transport (Hossain and Davies, 2013). In this research the focus will be only on the latter policy: transportation.

Motor vehicles are a very significant source of urban air pollutions (Health Effects Institute, 2010). In 2011, a third of all final energy consumption in Europe was accounted by transport, this regards to all transport not only road traffic. This transport-consumption was responsible to more than one fifth of the greenhouse gas emissions (European Environment Agency, 2011). So, one option to tackle the air pollution problem is to make an adjustment regarding the use of transport modes.

One.2 | Problem Definition

To reduce the emissions in transport, two possibilities come to mind: more sustainable driving or less driving. The first has to do with the kind of car and the amount of emissions from this car, but also the way of driving. To stimulate drivers to drive more sustainable, several campaigns are held in the Netherlands and apps are developed to helps people drive more sustainable (Ministerie van Infrastuctuur en Milieu, n.d.). One example of a campaign is 'Het Nieuwe Rijden' ('The New Driving'), with tips regarding e.g. switching gears before 2500 rpm (Het Nieuwe Rijden, 2012).

In contrast, the second possibility to reduce emission can be stimulated by the attracting of green transport modes. Public transport can be seen as such a transport mode, because of, among other, the higher occupancy rate than cars. Public transport users have less emissions per kilometer than private car users, only when a car has five passengers the car is cleaner (KpVV CROW, 2014). So, public transport seems to be more sustainable than private car use, and therefore an important issue regarding sustainable mobility (Redman et al., 2013)(Bérénos, 2009). However, in many western countries the public transport sector is facing a decreasing market share since the private car is seen as increasingly attractive (Redman et al., 2013)(Waerden et al., 2005).

Since the privatization of the public transport sector in the Netherlands the interest about the effectiveness of several planning and marketing measures towards (potential) public transport users has grown (Waerden et al., 2007)(Ongkittikul and Geerlings, 2006)(Martens, 2007). The attitudes of travelers towards different transportation modes have repeatedly been investigated (Redman et al., 2013)(Bérénos, 2009)(Wang and Chen, 2012)(Ettema et al., 2012), also the effects of measures are investigated (Waerden et al., 2007)(Ongkittikul and Geerlings, 2006)(Martens, 2007)(de Grange et al., 2012)(Waerden et al., 2010) and the potential for public transportation in certain areas has been mapped (Waerden et al., 2005)(Waerden et al., 2007).

Different stakeholders are participating in this development, to make transport more sustainable, but they have diffusing goals. Public transport companies' goal is to attract more travelers to their services and therefore, they aim on both car-users and bicyclists or pedestrians. On the other hand, local government only is interested in the increasing sustainable transport, and therefore only wants to reduce car-use but do not want to change the latter group to public transport (Kennisplatform Verkeer en Vervoer, 2014)(Waerden et al., 2007).

Besides the stakeholders' issue, it is very hard to trigger people in leaving their private cars and travel by bicycle or public transport (Redman et al., 2013). There are several initiatives to stimulate this so-called switching-behavior towards the public transport. Examples for this are: 'Goed bezig BUS' ('Doing well BUS') in the Province of Noord-Brabant (HVR group, 2010) and 'Ervaar het OV' ('Experience the Public Transport') in Gelderland; Overijssel and Flevoland (Flowresulting, 2011). 'Twents' in the province of Overijssel and surroundings, which upgraded their busses and information system (Syntus, 2014), and 'Eurokaartje' in Groningen and Drenthe which introduced more accessible tickets (Qbuzz, n.d.).

Next to this, many papers are written about the attitude of travelers, but still the public transportation agencies have little to no insights towards the potential of their services. Important questions for these businesses are whether they reach all the possible travelers, where potential users are vested and how sensitive their customers are towards different facilities. A tool is missing which can be used to find potential public transport users. This tool is necessary when public transport is seen as an important solution for air pollution reduction. If the existence and location of potential users is unknown, public transportation agencies cannot reach these persons with their marketing strategies.

ONE.3 | RESEARCH INTRODUCTION

The aim of this research is the development of the previously discussed missing tool to determine and localize potential public transportation users. This is done by analyzing the travelers' attitude towards switching behavior towards public transportation. The methodology of the development of this tool will be the main focus of the research.

In order to carry out this research a demarcation should be made. As stated the air quality in the Netherlands is worst in cities, and the forecasted increase of citizens living in cities will only worsen the situation. 74 percent of the NO_x emissions in Dutch cities is caused by only 14 percent of driven kilometers. The most of this is from freight and busses (Verbeek et al., 2012), therefore the bus services will be investigated in this research.

The research's focus lies with the methodology. No survey was hold to create new a dataset, but secondary data is used from a dataset out of the 'verplaatsingsonderzoek Nuenen 2009' ('travel survey Nuenen 2009'). This survey was performed in the context of the Transumo project 'Regionale Openbaar Vervoer Bereikbaarheid Consumentgericht (ROVBECO)' ('Regional Public Transport Accessibility Project'). This was the most recent and comprehensive dataset about transport mode choices, available at the university.

ONE.4 | RESEARCH QUESTIONS

In this paragraph the research questions will be stated. The target of the research is to develop a tool which can give public transport agencies more information about the potential users of their bus services. This tool should determine and localize potential bus users, so the agencies can anticipate on this knowledge, and the amount of private car-users can be reduced. The research question that follows this target is:

"Which research methodology should be used to develop a tool which determines and localizes potential bus users?''

A research methodology is the process used to collect information and data for the purpose of making business decisions. The methodology may include publication research, interviews, surveys and other research techniques, and could include both present and historical information (BusinessDictionary.com, n.d.). So, this question regards both to which model should be used to develop the tool, and the type of data which should be obtained to run the model.

This model should be made to identify potential public transport users from several characteristics. The research question following from this topic is:

“Which personal, trip and areal characteristics should be used to model potential users of public transport?’’

The characteristics refer to the behavior of travelers. The target of this question is to create a link between the behavior of travelers and the potential public transport users, which are needed for the tool. However, to answer this question, first this so-called potential should be defined. This implies to state when someone is a potential bus user, the sub-question regarding this subject is:

“What is the definition of potential bus users?

There are probably many ways to define potential. This questions includes the topic of gradations of potential; should a scale be implemented in levels of the potential? Related to this general definition of potential is the following sub-question:

“Which independent variables should be formulated to describe the potential of a person?”

It is assumed that the potential of someone depends on his behavior towards traveling. In this research an attempt will be made to describe the potential by linking behavior of travelers to the area, including characteristics of the area; the persons living there and the trips they make. The logical sub-question here is:

“Can the potential of travelers be explained by characteristics of an area?”

If this is the case the potential of an area can be modeled, and maps can be created that show the potential of this area. But maybe the characteristics of an area are too generic for one person’s potential, in that case possibly a hierarchical model should be used which disconnects the individual characteristics from the general data. Next to the scale of a person’s behavior, the scale of the area is important to consider. Perhaps the research is not meaningful for every level of aggregation.

The last research question which is discussed in this research, goes back to the social trigger of the research topic: ‘Air Quality’, the question is:

“Can the Air Quality in the Netherlands be improved by stimulating the use of Public Transport?”

In order to answer all research questions several dependent and independent variables with their own utilities should be found and several models could be estimated.

ONE.5 | RELEVANCE OF THE RESEARCH

The theoretical relevance of the research was already explained in the beginning of this chapter. The use of more public transport can reduce the amount of private car-users, which will reduce emissions and will lead to a cleaner and healthier environment. Public transport agencies can locate potential users by the use of the developed tool. They can anticipate to the new information by changing their services and facilities in their operating areas which can switch the potential users into actual users.

The practical relevance can be for example the implementation of the model when a municipality is expanding. If a municipality wants to expand a city or village a consultant could use the model to visualize the number of future potential public transport users. From this information a plan can be made about the public transport facilities in the new neighborhood which will fit the prospective demand.

ONE.6 | EXPECTED RESULTS

The expected result for this research is to obtain a tool, which will not only state whether there are potential public transport users, but also where they are located. The tool will determine whether potential users are present and which policy could be implemented to reach a certain effect where the potential public transport users will become actual public transport users. This tool will rely on a model, when this model is made all the potential users can be localized and the expected need for bus lines and maybe even bus stops can be modeled. This could be used by traffic accounting agencies which advice governments with new public transport routes in existing or future neighborhoods.

ONE.7 | READING GUIDE

The remainder of this thesis is organized as follows.

First, the Theoretical Framework of the research is described in chapter Two. Which handles the sustainability in the Netherlands including air quality, transportation's impact on this issue and durable public transport possibilities. Next to the environmental literature, public transport is discussed; a definition will be given, trends are examined, transport mode choice is looked at and potential public transport users are defined. The last topic discussed in this chapter is the marketing aspects; general marketing is covered, then specific marketing for public transport and finally Geo-Marketing. The chapter concludes with stating findings during the literature study.

The third chapter describes the research approach of this thesis. The theory used during this study is discussed. After this the research method is described, which states previous work and how the research proceeded. The models mentioned in the research method are then separately explained, these models are the multinomial logistic regression model and the hierarchical regression model. Finally conclusions are drawn about which models are used in the further investigation.

The data needed to create and run the models are discussed in chapter Four. The enquiry which is used for the research is described, together with Nuenen, the place where the survey took place. After this areal data is discussed which is needed next to the previously mentioned data. Both types of data are analyzed using the Golden Standard, which is a calibration instrument to ascertain whether a sample fits the actual population. The chapter ends with findings regarding the described data and the analysis.

In the fifth chapter the theoretical models are being combined with the previously described data to create functional models. Starting with correlation analyses to find fitting independent variables to predict the dependent variable ‘Bus Potential’. Then the different multinomial logistic regression models which have been studied are discussed. Finally conclusions are made about the findings during the modelling.

The last chapter of this thesis concerns the conclusions of the overall research. Next to this conclusions also recommendations for further study are made and a discussion is started about how the research could be done in other ways.

Chapter Two | Theoretical Framework

In this chapter the theoretical framework of the research will be discussed. In the previous chapter the problem definition and the origin of the problem are stated. This chapter will explain the nature of the problem further. This will be done by analyzing the most important aspects of the problem separately: ‘sustainability’, ‘public transport’ and ‘marketing’. At the end of this chapter the findings from all separately analyzed aspects, which occurred during this literature study, will be presented.

Two.1 | Sustainability

The first aspect of the literature study, sustainability, is chosen because it regards the social problem from which this research is abstracted. As stated in paragraph One.1 the air quality in Europe is getting worse and governments and other parties are trying to increase the sustainability in several ways. In this paragraph the actual influence of transportation to the environment is analyzed with special attention to the impact which public transport can have regarding this problem.

Two.One.1 | Air Quality

The air quality depends on the presence of air pollutants. If the concentrations of this pollutants are too high air pollution can be harmful for the health. The most important pollutants are: particulate matter (PM_x), nitrogen dioxide (NO_2), ammonia (NH_3); sulfur dioxide (SO_2); nitrogen oxides (N_xO), heavy metals and hazardous substances which are difficult to degrade. The first pollutant is the collective term for harmful particles in the air which are one of the main causes of smog. The second pollutant is mainly derived from traffic and affects the lungs of people (Atlas Leefomgeving, 2013). The third type of pollutants are responsible for the acidification and eutrophication of the soil. The fourth and fifth pollutants are part of ‘Smog’, a Dutch term to indicate a period of increased air pollution (Rijksinstituut voor Volksgezondheid en Milieu, n.d.). The last type can also be called Persistent Organic Pollutants (POPs) and are pollutant because they do not dissolve easily. The most important causes for the pollution of the air are traffic, unsustainable companies/factories, agricultural and livestock farms, consumers and forest fires and volcanic eruptions (Rijksoverheid, 2014a).

The European Union (EU) has limits and target values for several pollutants. The limits apply to sulfur dioxide, particulate matter, nitrogen dioxide, lead, benzene and carbon monoxide. The targets are for ozone, arsenic, cadmium, nickel and benzo (a) pyrene. The limits must not be exceeded, and all EU Member States should ensure to remain below the target values.

The Netherlands meets almost all European air quality limit values, except for a few places where the concentrations of particulates and nitrogen dioxide are still too high. This occurs especially in areas with many intensive farming or industry (fine dust) and within major cities (nitrogen dioxide) (Rijksoverheid, 2014b). To ensure the EU-targets will be met in the future the ‘Nationaal Samenwerkingsprogramma Luchtkwaliteit’ (NSL) (‘National Air Quality Cooperation’) is founded (Rijkswaterstaat, 2014). The Dutch air quality standards are divided into five concepts instead of two, next to the limits and the target values also alarm and information thresholds are defined and a long-term objective is set (Compendium voor de Leefomgeving, 2013).

The main target set by the EU acquired by the Dutch government is the 20 percent reduction in 2020 (Ministerie van Infrastuctuur en Milieu, 2011). To reach these targets for the pollutants and to safeguard the air quality in the Netherlands several measures are taken. As stated before the NSL is founded to meet the target values In the future, but also the government has to annually monitor the air quality (Rijksoverheid, 2014c).

Next to this, cities can close certain areas for polluting trucks by the use of environmental zones (Rijksoverheid, 2014c). This is already implemented in 13 big cities in the Netherlands including Amsterdam, Eindhoven and Utrecht (Centrum Milieuzones, 2014). This measure is proven to have positive effects on the air quality in the specific zones. A shift has taken place from high polluting vehicles towards cleaner ones, which improves the air quality and with this the public health. This leads to the conclusion that the environmental zoning is one of the most promising measures to improve the air quality next to flow measures, which can only be implemented in limited areas (Buck Consultants International and Goudappel Coffeng, 2009).

The last measure is the regulation 'Besluit gevoelige bestemmingen (luchtkwaliteit)' ('Decree sensitive locations (air quality requirements)'). Establishment or expansion of sensitive destinations, like schools, are only permitted if the limits for air quality are met(Rijksoverheid, 2014c). This measure focusses on people who have increased susceptibility to particulate matter and nitrogen dioxide, like children; elderly and sick people (Kenniscentrum InfoMil, 2014).

TWO.ONE.2 | TRANSPORTATION

Road traffic causes a large part of the emission (Buck Consultants International and Goudappel Coffeng, 2009) in the Netherlands (KpVV Dashboard, 2014)(Milieu Centraal, n.d.). In 2012 the National Institute for Public Health and the Environment published a report with all large-scale concentration and deposition maps for the Netherlands (Rijksinstituut voor Volksgezondheid en Milieu, 2012). The aim for this report is to research emissions in different sectors of the country. Based on this report Figure 1 is drawn, the concentrations from four of the main pollutions[1] are visualized in the chart. The pollutions due to sea salt, soil particles and other countries are disregarded, because these are aspects the government cannot control. Road traffic is the main source for all pollutions except for SO_2, but this concentration is very low in general.

[1] These pollutions are used in this charts because these are not yet below the target values of the government, other pollutions are disregarded because they are already on legal level, or because the source is only located in one of the subjects (e.g. NH_3 derives only from agriculture).

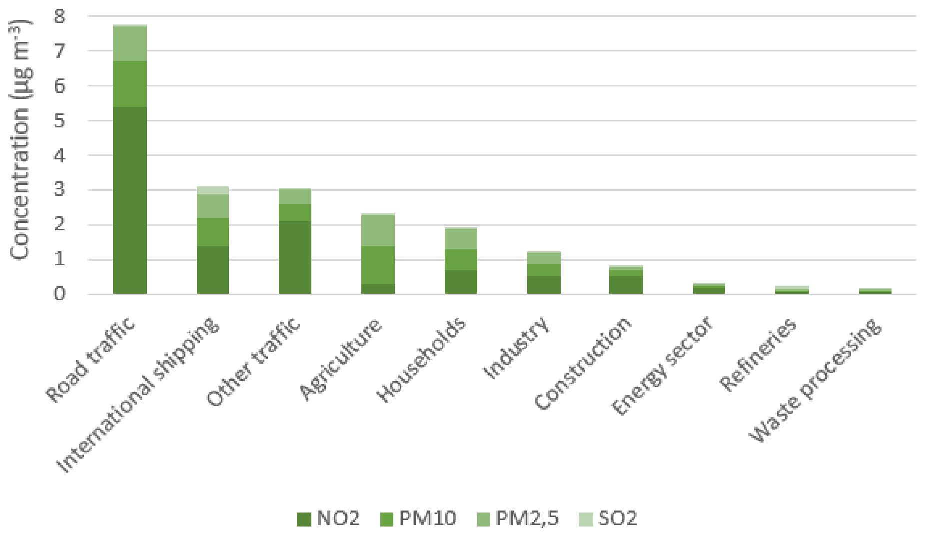


FIGURE 1 | POLLUTION CONCENTRATION CHART (RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU, 2012)

The government has to meet the limits and target values from the EU as stated above. Because the traffic in the Netherlands has the biggest share of the emissions, this is still an important issue on the governmental agenda (Ministerie van Infrastuctuur en Milieu, 2011).

TWO.2 | PUBLIC TRANSPORT

As stated before changes of transport modes can have a positive impact to the air quality and the overall environment. One option is to stimulate public transport, so less private-cars will be used, therefore the public transport will be described in this paragraph. The public transportation reviewed in this research will be defined, the trends that can be seen in the public transport are shown, the negative attitude from travelers towards the public transport will be investigated and the definition of the potential public transport users will be given.

TWO.TWO.1 | DEFINITION PUBLIC TRANSPORTATION

There are many different forms of public transportation, each having different impact on the environment. The forms of public transportation typical for a metropolitan city are visualized in Figure 2. Bold lines represent the high-capacity and high-speed services, like rapid trail or rapid bus transit systems. These services connect multiple activity centers, such as suburban areas and surrounding villages. The more local services, like local bus networks or other forms of ridesharing vehicles, like the metro or tram, primarily serve communities or individual employment centers. These services do not reach all the activity centers but are more centered towards the central city (Kutz, 2008).

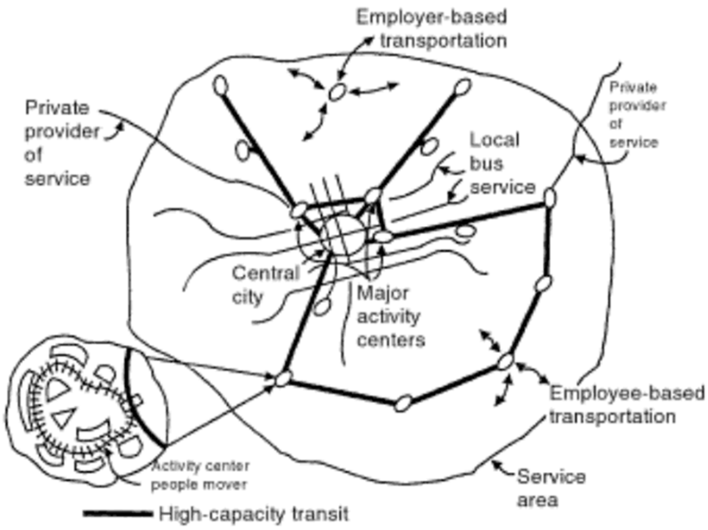


FIGURE 2 | TYPICAL METROPOLITAN TRANSIT SYSTEM (KUTZ, 2008)

In order to carry out this research a demarcation should be made. As stated in the introduction of this thesis the air quality in the Netherlands is worst in cities, and the forecasted increase of citizens living in cities, mentioned in paragraph One.1, will only worsen the situation. The research introduction from Paragraph One.3 states that the most pollution emissions in cities are from freight and busses. Therefore the bus networks will be investigated in this research, this refers not only to the local inner-city networks, but also to rapid bus transport systems which transfer between cities and/or villages.

Multiple bus types can be determined in the Netherlands. The main types are the “Streekbussen, Stadsbussen en Buurtbussen”, resp. Intercity busses, City busses and Neighborhood busses. The first two types are often about the same size, but the first are regional busses and the second are inner city busses. Both bus types run a fixed route with a fixed timetable and fixed stops, the passenger has to request the bus to stop, otherwise the stop will be skipped. Intercity busses run once or twice per hour and stop only at relative big activity centers, however when train connections are missing these busses can run much more frequent. City busses run overall more often, depending the destination, and have many stops at e.g. train stations, shopping centers, hospitals and living and business areas. In some cases high capacity vehicles are used on busy lines, these are busses which are 18 or even 24 meters long. The Neighborhood busses are the smallest of the types and run also on a fixed route with fixed stops. But unlike the Intercity and City busses the Neighborhood busses can either run on a fixed timetable or the passenger can use the bus on demand. The latter case implies that the traveler has to call and reserve a place on the bus, otherwise the bus will not drive pass the stop. On some calm traffic moments (e.g. evenings and weekends) also other bus types can drive on demand instead of the fixed timetable (Qbuzz, 2014)(OV-bureau Groningen Drenthe, 2014).

TWO.TWO.2 | SUSTAINABLE PUBLIC TRANSPORT

Public transport is often seen as an important air pollution reduction strategy (Vincent and Jerram, 2006). This is not only because of the strategies which can be implemented in the systems, which will be mentioned later on, but also about the amount of people transporting per vehicle. When more travelers travel by public transport, the less personal car-trips will be used and the air pollution will be reduced. Car-transport can be relatively clean, but the occupancy rate should be sufficiently high. A car with only one person is per travelers-kilometer more polluting than public transport. When it comes to marginal emissions, CO_2 emissions per car outside peak hours is much higher than per public transport. During rush hour the difference is smaller, but still in favor of the public transport (Boer et al., 2008). So, the target is not only to make transit service more appealing but mainly to lower the attention of the automobile (Kutz, 2008). Buses emit less emissions per passenger than an average car, if the occupancy rate is higher than 5. The average occupancy of a bus is 11 passengers, this varies per region but is overall higher than 7 (KpVV CROW, 2014).

When looking at the emissions per passenger, the bus is a more sustainable transport mode than the car. However, busses in the public transport are in total responsible for 470 million vehicle kilometers, which is 0.4 percent of the total vehicle kilometers from the road traffic in the Netherlands. Even though this seems to be a small amount, the contribution of polluting emissions is very high, 74 percent of the NO_x emissions in Dutch cities is caused by only 14 percent of driven kilometers. The most of this is from freight and busses (Verbeek et al., 2012). The bus is, when looking at the occupancy rates, more sustainable than cars, however the systems can be improved.

One of the possible solutions to make busses more sustainable is to make their actuator more sustainable. Traditionally busses are powered by diesel engines with pollutant emissions: particulate matter (PM_{10}) and nitrogen oxides (NO_x). Although the diesel engines are getting cleaner by the use of incentive programs, there are more sustainable solutions like alternative fuels. These alternative fuels can be natural or green gas, the latter is highly subsidized by the national and regional governments (Kennisplatform Verkeer en Vervoer, 2009). Green Gas can reduce the CO_2 emissions by at least 73 percent compared to gasoline and diesel and improves the air quality significantly, compared to diesel, by greatly reducing particulate and NO_x emissions (Stichting Groen Gas Nederland, 2014). In the Netherlands already 400 public transport busses, divided over approximately 12 cities, drive on Green Gas (GroengasMobiel, 2014). Other alternative actuators are hybrids, which are at the moment far too expensive, and new sustainable drive forms which are not yet developed (Verbeek et al., 2012).

Next to the innovations in fuels and actuators of the busses, other energy saving strategies are available for busses. One of these are lightweight busses which are made with alternative materials and can reduce emissions up to 10 percent. Also the rolling resistance from the busses can be reduced by the use of different tires or another alignment of the vehicles. The passenger occupancy can be optimized as is done by the so-called “Fit For Purpose” busses. Next to this the driving behavior of the chauffeurs can reduce a lot of the emissions by changing to Eco-driving or “New Driving”, and (part of) the driving task could be automated. The last strategy is already implemented through whole of the Netherlands and entails the usage of bus lanes which increase the flow of the vehicles (Verbeek et al., 2012).

TWO.TWO.3 | PUBLIC TRANSPORT USAGE

According to European legislation in the field of public transport, regional and local public transport must in principle be arranged through public tender. This is usually done through a concession contract. A concession is the exclusive right to provide public transport in a given area during a given period, a concession period for bus services has a maximum of ten years. The 'Nederlandse Wet Personenvervoer 2000 (Wp2000)' ('Dutch Act 2000') states that a competent authority must grant concession before the transporter may perform public transport. In the Netherlands this authority is often a municipality or, in case of regional public transport by train, the province or region (Europa decentraal, 2015).

This concession contract system has been used in the Netherlands since 1994, when the first tender trials where hold. These trials where organized and evaluated by the 'ministerie van Verkeer en Waterstaat' ('Ministry of Transport'). January 2001, the 100th award of a concession to a transportation agency was granted (Kennisplatform CROW, 2013). So, the system seems to be a successful one and has not be changed for quite some time.

But how does the bus services do in comparison with all transport modes in the Netherlands. The numbers of the transportation of the last few years are subtracted from the 'Onderzoek Verplaatsingen in Nederland (OViN)' ('Travel Research in the Netherlands'). This is an annual research performed by the 'Centraal Bureau voor de Statistiek (CBS)' ('Central Statistical Office'). The research was initiated by the CBS in 1978 by the name 'Onderzoek Verplaatsingsgedrag (OVG)' (Travel Survey), in 2004 Rijkswaterstaat took over the research and called it 'Mobiliteitsonderzoek Nederland (MON)' (Mobility Survey Netherlands). As of 2010 the implementation of the mobility research is back at CBS, now referred to as OViN. For this paragraph the data from 2011 till 2013 will be used (Centraal Bureau voor de Statistiek, 2014a)(Centraal Bureau voor de Statistiek, 2013)(Centraal Bureau voor de Statistiek, 2014b).

Figure 3 shows the mean travel distance per person per day from 2011 till 2013, in kilometers. The mean is taken because the amounts are very stable across the years. The data is categorized by transport modes. As can be seen in the diagram, the car is by far the most popular transport mode, after this the train and bicycle, and only after these comes the bus/tram/metro.

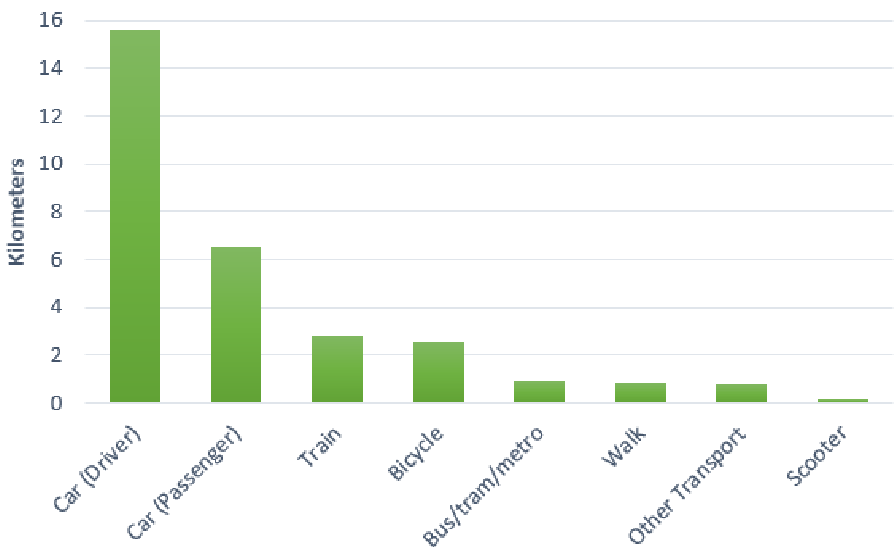


FIGURE 3 | TRAVEL DISTANCE PER PERSON PER DAY (KM)

From this information can be concluded that most kilometers are traveled per day by car and most cars have an occupancy of just one person, because the Drivers' kilometers are far more than the Passengers' kilometers. This is a bad thing because of the fact stated in paragraph Two.Two.2: "Car-transport can be relative clean, but the occupancy rate should be sufficiently high."

Two.Two.4 | Transport Mode Choice

As seen in the previous subparagraph the car is by far the most used transport mode, see Figure 3, the reason behind this will be discussed next. When the reason for transport mode choice is clear, measures can be taken to trigger switching behavior. Only the variables which can affect the choice between public transport and car use are dealt with because private car usage is the mean competitor for the public transport sector.

Regarding to Cliquet (2006) distance is no longer the only variable for mobility, but especially time, money and the behavior of the traveler are big assets (Cliquet, 2006). From several papers and publications, which will be mentioned later on, all kinds of variables which affect travelers' mode choice are summarized in Table 1 and 2.

TABLE 1 | VARIABLES TRAVEL MODE CHOICE PUBLIC TRANSPORT (ETTEMA ET AL., 2012)(SUMAEDI ET AL., 2014)(BEIRÃO AND SARSFIELD CABRAL, 2007)(ROJAS-RUEDA ET AL., 2012)(ABOU-ZEID ET AL., 2012)

	Advantages	**Disadvantages**
Public Transport	Cost	Waste of Time
	Less stress	Too crowded
	No need to drive	Lack of comfort
	Be able to relax	Unreliability
	Travel time on bus lanes	Lack of control
	More sustainable	Long waiting times
	Contact with other people	Need of transfers
	Able to work/study	Traffic
		Lack of flexibility
		Accessibility
		Service
		Noise, disturbance

An important advantage of public transport is the ability to do other activities besides travelling, such as working; studying and relaxing (Ettema et al., 2012). Unfortunately, there is often noise when traveling with the public transport, either from the service itself or from fellow travelers, which can disturb these activities (Sumaedi et al., 2014). Other disadvantages are the lack of flexibility, comfort and control. Besides this there are often long waiting times, and the need for transfers causes a relative long travel time, which many travelers experience as a waste of time (Beirão and Sarsfield Cabral, 2007)(Rojas-Rueda et al., 2012).

TABLE 2 | VARIABLES TRAVEL MODE CHOICE CAR (BEIRÃO AND SARSFIELD CABRAL, 2007)(ROJAS-RUEDA ET AL., 2012)(STEG, 2005)

	Advantages	**Disadvantages**
Private Car	Freedom/independence	Cost
	Flexibility	Difficulty finding parking place
	Convenience	Costs of parking
	Rapidity	Driving stress
	Comfort	Congestion
	Safety	Pollution
	Private space	Accidents
	Music	Isolation
	Hobby	
	Status	
	Weather	
	Luggage	

In contrast to the public transport a private car can give someone much more flexibility which leads to a feeling of freedom and independence. Public transport services can be inconvenient for some persons to use because of the distances between home and the service and the service and the destination (Abou-Zeid et al., 2012). Next to this, a car can be seen as a symbol of status, many car-owner like their car and some experience their car as a hobby (Steg, 2005). Other aspects favoring the car-usage are the minimum outside traveling with bad weather and the ability to bring much luggage with you during your trip (Steg, 2005).

Next to the attributes that affect the travel mode choice there are motivations and barriers that can affect this choice, these aspects can be found in Table 3. The motivations are aspects which can affect the travel mode choice if they would be implemented or aspects which the traveler can use to motivate itself to use the public transport. The latter could be saving money or contribute to the environment aspects. The barriers are aspects which are already issues and which should be eliminated to encourage people to use the public transport. The motivations and barriers are seperated in several quality attributes adapted from Redman et al. (2013). These attributes are: reliability, frequency, price, speed, accessibility, comfort and convenience (Redman et al., 2013).

TABLE 3 | MOTIVATIONS AND BARRIERS TO PUBLIC TRANSPORT (REDMAN ET AL., 2013)(WAERDEN ET AL., 2007)(WAERDEN ET AL., 2010)(SUMAEDI ET AL., 2014)(BEIRÃO AND SARSFIELD CABRAL, 2007)

	Motivations	**Barriers**
Reliability	Be certain that the timetables are performed	Buses' unreliability
	More information available and easy to understand	Poor information
Frequency	Increase of frequency	Lack of availability of buses
		Limited bus frequency
Price	Save money	
	Increase costs car use	
	Increase of parking costs	
Speed	Travel time public transport shorter	Travel time ratio
	Increase travel time car and bicycle	
	Introduction of more speed limits for cars	
Accessibility	Direct transport from home to work	Long distance to bus-stop
	Decrease in distance to bus stop	
Comfort	Better service	Service (not customer friendly)
	More comfort and air-conditioning on vehicles	Buses are smelly and crowded
	Aesthetics	Negative feeling towards public transport
	Vehicle conditions	Few Seats
	Higher guarantee of seat	Bad waiting conditions
Convenience	Not having a parking space	Not having alternative to car
	Contribute to a better environment	Lack of direct transport
	Easy transfers	Do not known what to expect
	Acceptable waiting time	Having to use more than one Transport
	Easy to find	Habit of driving
		No direct connection
		Not safe at the bus

As can be seen from the table most barriers are formed by the aspects of comfort and convenience. The lack of a direct transport and long waiting times are aspects often seen in the literature (Waerden et al., 2010)(Beirão and Sarsfield Cabral, 2007). The price, speed and accessibility attributes have the most motivations opposite of the least barriers. If these aspects will be implemented the chance is great to have immediate effect. These motivations that can trigger the use of the public transport are mostly concerned with discouraging travelers to use the car: increasing car use costs, parking prices and car travel time; introducing more speed limits (Waerden et al., 2007).

The quality attributes about comfort and convenience have many barriers, but these also have quit a lot of motivations. A high garantuee of seats and acceptable waiting times are examples of aspects that can motivate travelers to use the public transport within the comfort and convenience attributes(Waerden et al., 2007)(Sumaedi et al., 2014). Finally, it seems that the aesthetics and the overall vehicle conditions are also an important aspect of motivating travelers (Redman et al., 2013).

Reasons for using the transport modes are examined next, these variables are translated into policy measures. Table 4 shows policies that influence mode choice with possible actions that can be taken summarized by Kutz (Kutz, 2008) and De Grange (de Grange et al., 2012).

TABLE 4 | POLICIES AND MEASURES (DE GRANGE ET AL., 2012)(KUTZ, 2008)

	Measures
Transportation Investment Policy	Infrastructure spending directly affects the relative attractiveness of each mode
	Transit operating assistance can help maintain, improve or expand services
	Research and development funding provides innovations in the provision of transportation services
Transportation Pricing Policy	Taxes and tolls make automobile use more expensive
	Fare policy determines cost of transit trips
	Local policies dictate taxi fare and, indirectly, service levels
	Local parking pricing and availability are very important components of the costs of driving
Environment Policy	Federal/state emissions standards increase new car prices
	Local air quality mandates require programs to reduce single-occupant vehicle use
	Local policies influence development patterns and transportation planning
Energy Policy	Minimum average fuel economy standards increase new car prices, decrease operation costs
	Alternative fuel vehicles are unlikely to affect modal choice
Tax Policy	Income taxes affect economic activity and disposable income, hereby influencing the affordability of various travel choice
	Preferential parking cost deductions promote automobile commuting over transit
	Sales taxes affect automobile costs, and may support public transport
	Mortgage interest deductions influence housing location choice
	Property taxes may support local roadway infrastructure
Land Use Policy	Provisions of zoning laws (lot size, use) affect viability of public transport
	Design reviews and other restrictions can require definitive plans for addressing transportation issues in new developments

Several measures have been implemented in the public transport market, but only sometimes the effect of this measures are investigated after the implementation. And only in a very limited number of cases the potential effects are estimated before the implementation of the measure (Waerden et al., 2007).

TWO.TWO.5 | POTENTIAL

Many studies are performed on the actual use of public transport, but not on the potential use of public transport per area (Waerden et al., 2005). The potential users of public transport can be defined in multiple ways, three of these will be discussed here.

The first definition of potential public transport users is purely theoretical. Of the public transport services will become faster/ cheaper/ more frequent/ etc., with how many per cent will the usage increase. In order to answer this, many studies had been done towards bus demand elasticities. For example, Nijkamp & Pepping (1998), which tried to assess the aspects that influence the sensitivity of travelers towards travel costs of the public transport in Europe, by doing a comparative analysis of elasticity values of transport demand resulting from studies in multiple countries (Nijkamp and Pepping, 1998). More recent Holmgren (2007) studied not only the costs of the public transport, but also the influence of vehicle-kilometers, income of the travelers, petrol price and car ownership (Holmgren, 2007). However interesting, these models have no relation to individual people. The information is not location bound, and therefore no geo-marketing can be used with the knowledge of these models.

The second definition of potential public transport users is purely physical. Potential could be defined by the distance someone has from, for example, a bus stop. In this way someone who lives between 100 meters of a bus stop has more potential than someone who lives 500 meters from the bus stop. This distance can be measured in several ways, like, as the crow flies, travel time or actual walking distance, and graphs can be made to visualize the potential in areas. Next to the distance of a bus service also other physical ways can be thought of regarding bus potential. For example, facilities around bus stops, both origin and destination stops can be looked at. However, also in this case no personal touches are taken into account, and therefore no geo-marketing can be used with this physical knowledge.

The third, and most realistic, definition of potential public transport users depends on a person's switching behavior. This switching behavior is a person's individual opinion, and can both refer to someone's ability to switch, or someone's willingness to switch. In this definition potential public transport users are the travelers who, at the present, do not use the public transport, but which could/would switch towards it when certain measures are taken. Such a measure can vary from actual change of bus lines to only marketing of the present supply. The switching possibility depends on several aspects which mostly relates to the willingness of people to switch. The aspects are about the distance between origin or destination and bus stops, frequency of busses, whether the connection can be traveled directly or there is a need for transfers, ratio between travel time/distance/costs from bus or other transport mode, the ability to bring belongings and other comfort aspects like crowdedness and service (Waerden and Bérénos, 2010).

TWO.3 | MARKETING

The previous paragraph discussed variables affecting the use of public transport and possible policies and measures. But nothing will change regarding to transport mode choices from travelers if they are not aware of the newly implemented measures. The steps to get this information to the potential users are researched in this paragraph. First the definition of marketing is given, followed by the specific marketing for public transport. The final topic that will be discussed is another aspect of marketing called Geo-Marketing and the possibilities to implement this for public transport.

TWO.THREE.1 | GENERAL MARKETING

When searching for marketing, dozens of definitions can be found. The broadest one found is given by the 'American Marketing Association (AMA)' and states:

> *"Marketing is the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large"* (American Marketing Association, 2013).

'Centrum Vernieuwing Openbaar Vervoer (CVOV)' (Innovation Centre Public Transport) describes marketing as:

> *"a cyclical process that aims to reduce disparities between supply and demand by tuning the internal and external activities optimally to the needs of (potential) customers"* (Centrum Vernieuwing Openbaar Vervoer, 2002).

This latter definition is more related to this research because of the addition of potential customers. The AMA only refers to the present users for the service given, and not to the research of new potential users. So, depending on this second definition marketing is all about the challenge to make a customer-focused and profitable service. Also according to the CVOV, the steps to design such a marketing strategy are as following:

I. External and internal situation analysis (*Where are we now?*)
II. Strategic alternatives (*Where can we go?*)
III. Marketing objectives (*Where are we going?*)
IV. Marketing strategy (*How do we get there?*)
V. Monitoring (*How do we measure the qualitative and quantitative results?*)

TWO.THREE.2 | MARKETING PUBLIC TRANSPORT

The role of marketing in the public transport has increased much in the recent past because of the evolution of transit. First, most people had to use the public transport because of few alternatives. Now is the time of high car ownership and the public transport has to make an effort to attract customers. Therefore, regarding to Vuchic (2005), marketing is implemented in the public transport sector to achieve the following major goals: Promote services and inform public; Ensure the maximum number of potential riders is attracted; Ensure the transit service is provided where it has the maximum overall benefit (Vuchic, 2005). The first of these goals can be done by the implementation of general marketing to give people knowledge about the services. The second target is trickier, to know whether the maximum number of potential riders is attracted, first the knowledge is needed who are potential riders. For the third goal not only the presence of potential riders is needed, but also the location. These target could be reached with the use of the tool discussed in Paragraph One.3. In this Paragraph the present marketing techniques in the public transport sector will be discussed, they only regard the first goal, about promoting services and informing public.

The development of marketing of the public transport was slowly because public transport agencies did not comprehend the competition of the car. However, inefficient marketing will lead to loosing passengers in the long term (Vuchic, 2005).The traditional approach to attract potential public transport users is the implementation of 'hard policies', like the provision of transport services and infrastructure, including pricing and longer term land use policies (Brög et al., 2002). These policies are necessary to increase the use of the transport system, but they are of almost no value if people are unaware of these improvements. Empirical surveys show there is a lack of communication and this information does not reach the target group. This communication and information can also be referred to as 'soft policies' which can be implemented faster and cheaper as the before mentioned 'hard policies' (Brög, 2000)(Taniguchi and Fujii, 2006).

Vuchic defines four different marketing strategies for public transport: undifferentiated marketing, differentiated marketing, concentrated marketing and individualized marketing. The first is the simplest and cheapest strategy, but also the least effective if implemented as sole strategy. Next to this, the characteristics of the area need to be uniform to affect all potential customers. The second strategy specifies different needs for a more diverse public, this is a more expensive option but the revenues are proven to be better. The third strategy can be used if a single component of a service should be promoted by a segmented market, so this strategy can only be used for specialized services. The fourth and last strategy regarding the individualized marketing is multiple times proven to be very effective, but is also the most labor and costs expensive. For this strategy first the potential should be selected and then these people have to be contacted directly (Vuchic, 2005).

TWO.THREE.3 | GEO-MARKETING

To implement the individualized marketing strategy as explained before, the potential costumers should be selected and contacted directly (Vuchic, 2005). But to do this, it is not only necessary to know who are the possible users, but also where they are located (Waerden et al., 2005). When the location of potential is not known it is not possible to directly focus the marketing to the potential. This location bounded marketing can also be referred to as 'Geo-Marketing' (Cliquet, 2006)(GfK Geomarketing GmbH, 2013).

Currently Geo-Marketing is more important than ever. It is necessary to quickly adapt to the rapidly changing market by taking advantage of any open niches and staying one step ahead of competitors. The six most important advantages of a Geo-Marketing approach regarding to

GfK GeoMarketing GmbH (2013) are stated and explained next (GfK Geomarketing GmbH, 2013).

1. Speed & transparency
2. Efficiency and savings
3. Location of new potential
4. Effective management and controlling tool
5. Overview of entire market and close-up of specific areas
6. Worldwide applicability

The first advantage regarding speed and transparency is coming from the ability to communicate all results effortlessly on a digital basis, which ensures employees are always kept informed. In the past, large wall maps, pins and strings where necessary for market visualizations. This approach would take a lot of time, and it showed difficulties about ensuring the maps reflected the latest status. Nowadays, these tasks are best done by the use of specialized software that visualizes the market's structure directly on digital maps, plan new optimized structures and perform measures.

The efficiency and savings are better with Geo-Marketing because the wasting of resources is avoided in direct marketing campaigns; the Geo-Marketing tool already analyses where the advertisements should be positioned.

A very important aspect of Geo-Marketing is the possibility to locate new potential. This potential and geographical characteristics can vary much from area to area, good marketing management needs to focus on these differences. This attention can lead to the knowledge about where new customers can be acquired.

The fourth advantage which entails an effective management and controlling point is made possible by the creation of a very detailed portrait of the target group. Geo-Marketing makes it possible to check whether the envisioned target group corresponds to the profile of the actual purchasing customers and which areas have high concentrations of these target group.

The fifth advantage overarches the before mentioned aspects, like potential and the target group, in the overview of the entire market and close-ups of specific areas. Important insights in this market are the market exploitation, customer distribution and new customer potential. The market exploitation deals with the turnover potential, the customer distribution visualizes spot areas where customers are concentrated and the new customer potential reveals the gap between market operations; for example an area where high potential is measured, but where only a few customers are located.

The last advantage is the worldwide applicability, there are no location bounded aspects needed for the implementation of Geo-Marketing, the advantage of this is self-evident (GfK Geomarketing GmbH, 2013).

The tool which will be developed, based on the model made in this research, will determine and localize potential public transport users. Therefore, the location of this potential will be known and Geo-Marketing would be the best marketing technique to stimulate the potential to become actual bus users.

TWO.4 | FINDINGS

Air quality in is getting worse and the Netherlands does not meet the European standards yet. The main source of the air pollution is Road Traffic. The public transport sector can contribute to a cleaner environment, not because they drive more sustainable, but because it has higher occupancy rates. Public transport is at the moment one of the least used transportation services. Several measures have been devised to attract travelers to the public transport, however this should become common known to people by the use of marketing techniques. If the location of potential public transport users is known, Geo-Marketing would be the best strategy to stimulate potential users to actually take the bus. This is a marketing technique which is proven to be functional and has multiple advantages compared to general marketing.

Chapter Three | Research Approach

This chapter describes the theory behind the research and the methods used to create the model which will be discussed in Chapter Five.

Three.1 | Theory

In this research behavioral theory is used, the purpose is to understand why people choose a specific transport mode and to get to know where potential users are located. With this knowledge the public transport agencies can implement Geo-Marketing, and participate to the behavior of the travelers and thereby promote, among others, the bus usage. The research of behavior could be referred to as psychology. Regarding R.D. Luce (1959) a big amount of psychology, including motivation, has the common theme: choice. Stated preference and choice models can be created with the use of multiple theories, from these the random utility theory is often considered to be the most appropriate (Borgers and Timmermans, 1993). Individuals are assumed to maximize their utility, in this research they choose the transport mode that fits their needs the best. The survey used during this research deals with personal trips of the respondents, they choose which transport mode they use and whether alternatives are realistic and therefore individual choice behavior is used. Which is, again referring to Luce, best described as probabilistic and not algebraic (Luce, 1959).

However, the individual choice behavior is not the only used technique in the research. The research is about whether travelers take either the car, the bicycle or the bus. Because these are all different alternatives the model will be a discrete choice model. Discrete choice modelling is about understanding and predicting choices between alternative products, facilities, services, etc. (Train, 2003). So, in this case the transport modes are the different alternatives. From the research empirical proven links between variables can be substantiated.

The last theory that will be discussed is the actual choice behavior. The respondents are asked when they will use another travel mode like public transport, also referred to as switching behavior. The theory behind switching behavior is about understanding how, why and when consumers use a certain aspect and what aspects influence this choice (Pookulangara et al., 2011). The model will use actual behavior with a link to analytical behavior, the model will transfer from revealed choice towards stated adaptation.

Three.2 | Research Method

Earlier attempts on modelling public transport were almost always about the actual use of public transport instead of the potential use, nevertheless there are some exceptions. Sugiki et al. developed a potential traffic demand, not only public transport, model using a binary logit model, with outcomes: ‘potential’ or ‘no potential’. However, this model is very limited because it only focusses on the Household Life Stage, with the use of only three aspects: household age, household type and residence type (Sugiki et al., 2001). Another attempt, regarding public transport, was made by Zhou et al. using a two-leveled nested logit model to estimate a mode choice model that differentiates between regular and enhanced transit services. Eight different market segments were identified, which were linked with several areas, and the market share per market area was calculated (Zhou et al., 2004). In this model no directly relation is made between the area characteristics and the market shares (Waerden et al., 2005).

In 2005, Van der Waerden et al. tried to make public transportation potential maps (Waerden et al., 2005). The research questions are:

1. Where is he potential demand for public transport located?
2. Is the potential demand related to characteristics of areas?
3. How can this potential demand easily be presented to planners and policy makers?

Data was collected in a medium sized city in the South of Netherlands. A questionnaire was used to find data of the travelers' behavior, and for the areal data six position ZIP code data was obtained from Bridgis. A logistic regression model was estimated to describe the relation between bus usage and available area characteristics. This paper uses a simple approach to identify potential customers of public transport by the use of GIS-based public transport potential maps. However, this research and following researches involving Van der Waerden et al. only use variables with areal characteristics and none about personal behavior of the subjects (Waerden et al., 2005)(Waerden et al., 2008)(Bérénos, 2009).

In this research this earlier attempts will be used and expanded, because the similarity in the research questions, and the conclusions of Van der Waerden et al. (2005). These conclusions declare that the first attempt for transport potential maps can be an adequate tool for public transport companies. However, the research will be extended with personal and trip variables. At first a model that only uses possible personal, trip or areal variables will be made. When successful the model will be expended by using combinations of these variable groups to see whether the model improves. Also several variables can be implemented both as an areal variable or personal variable. In such a case both scales will be checked and then the most suitable scale will be used in the final model.

THREE.3 | MODELS

Because several research methods are used to describe potential in the past it is not clear which method is the best. Therefore, to get the most out of this research, several models can be created with the use of different research approaches. These possible models are described below.

This research is performed with secondary data. In the research where the data is subtracted from, the potential bus users are already defined, this definition is adopted in this research. The potential bus users are described like the third definition from Two.Two.5, the bus potential is defined by the ability of the respondents to switch non-bus trips to bus trips, and four different categories are used to describe this ability to switch. This dependent variable and all independent variables will be discussed extensive in the following Chapter Four: Data.

THREE.THREE.1 | MULTINOMIAL LOGISTIC REGRESSION

The dependent variable in this model will be the potential for replacing a non-bus trip by a bus-trip. This is one of the questions asked in the enquiry and therefor the answers are clear. Now this variable should be connected to the independent variables: the personal, trip, and areal characteristics. Because the dependent variable, bus potential, is categorical, this can be done by the use of logistic regression. Logistic regression is a special version of regression in which the outcome variable is categorical. If the categorical variable has exactly two categories the analysis is called binary logistic regression, and when the outcome has more than two categories it is called multinomial logistic regression (Field, 2013a). The dependent variable 'Bus Potential' has four different categories, so the latter will be used.

The most important principle behind regression is the prediction of an outcome variable from a predictor variable (simple regression) or predictor variables (multiple regression). This can be seen in Equation 1.

EQUATION 1 | REGRESSION FORMULA (FIELD, 2013B)

$$Outcome_i = (Model) + \ Error_i$$

So the outcome that is going to be predicted is described by the fitted model and an error. The equation for a simple regression is stated in Equation 2, Equation 3 illustrated this for a multiple regression:

EQUATION 2 | SIMPLE REGRESSION FORMULA (FIELD, 2013B)

$$Y_i = (\beta_0 + \beta_1 X_{1i}) + \ \varepsilon_i$$

EQUATION 3 | MULTIPLE REGRESSION FORMULA (FIELD, 2013B)

$$Y_i = (\beta_0 + \beta_1 X_{1i} + \ \beta_2 X_{2i} + \cdots + \beta_n X_{in}) + \ \varepsilon_i$$

To use an ordinary regression model some assumptions have to be met, these assumptions are stated in Table 5.

TABLE 5 | ASSUMPTIONS ORDINARY REGRESSION (FIELD, 2013B)

Assumptions	**Explanation**
Additivity and Linearity	The outcome variable should, in reality, be linearly described to any predictors and, with several predictors, their combined effect is best described by adding their effects together.
Independent errors	For any two observations the residual terms should be uncorrelated (i.e., independent).
Homoscedasticity	At each level of the predictor variable(s), the variance of the residual terms should be constant.
Normally distributed errors	It is assumed that the residuals in the model are random, normally distributed variables with a mean of zero.
Predictors are uncorrelated with 'external variables'	External variables are variables that haven't been included in the regression model and that influence the outcome variable.
Variable types	All predictor variables must be quantitative or categorical.
No perfect multicollinearity	If the model has more than one predictor then there should be no perfect linear relationship between two or more of the predictors.
Non-zero variance	The predictors should have some variation in value (1.e., they do not have variances of 0).

However, when the outcome variable is categorical, no ordinary regression is used. In that case logistic regression is used, and not the value for Y is predicted from a predictor variable(s) X, but the probability of Y occurring given known value(s) of X. The general equation for binary logistic regression can be found in Equation 4.

EQUATION 4 | BINARY LOGISTIC REGRESSION FORMULA (FIELD, 2013A)

$$Probability\ (Y) = \frac{1}{1 + e^{-\,(Model)}}$$

The equations for the simple and multiple logistic regression models are the same as the equations of the ordinary regression models. The logistic regression equations are visualized in Equation 5 and Equation 6.

EQUATION 5 | SIMPLE BINARY LOGISTIC REGRESSION FORMULA (FIELD, 2013A)

$$P(Y) = \frac{1}{1 + e^{-\,(\beta_0 + \beta_1 X_{1i})}}$$

EQUATION 6 | MULTIPLE BINARY LOGISTIC REGRESSION FORMULA (FIELD, 2013A)

$$P(Y) = \frac{1}{1 + e^{-\,(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{in})}}$$

When using the logistic regression model less assumptions need to be taken into account. This reduced list of assumptions can be found in Table 6.

TABLE 6| ASSUMPTIONS LOGISTIC REGRESSION (FIELD, 2013A)

Assumptions	Explanation
Linearity	Assumes there is a linear relationship between any continuous predictors and the logit of the outcome variable.
Independence of errors	The same as for ordinary regression; cases of data should not be related.
Multicollinearity	Predictors should not be too highly correlated

As stated it is possible to use logistic regression to predict dependent variables with more than two variables, this is called Multinomial regression. This form of logistic regression works essentially the same as the binary logistic regression explained before. The equations used are the same, however the analysis breaks the outcome variable down into series of comparisons between two categories. The category which is compared to all other categories is called the reference category. The model estimates the possibility of a category getting chosen over the reference category (Field, 2013a).

If these kind of logistic regression models are used, the model does not take the multiple levels of characteristics into account. The personal and areal variables are both used in the same level. When this occurs several aspects of the reality are not captured in the model. Personal variables are all individual and on the same (lower) level. The areal variables are on a higher level, because there can be groups of individual clusters by the area they originated from. To implement these higher and lower levels, a multi-layered hierarchical regression model should be used, which will be discussed in the following paragraph.

THREE.THREE.2 | HIERARCHICAL REGRESSION (MULTILEVEL LINEAR MODELS)

Previous model treats the data at a single level. However regarding Andy Field (2013), in the real world data is often hierarchical. This means some variables are clustered in another variable, also referred to as nested variables (Field, 2013c). The higher level variable will be called the contextual variable, in the case of this research this will be the area a respondent lives in. The hierarchical structure of the data is visualized in Figure 4.

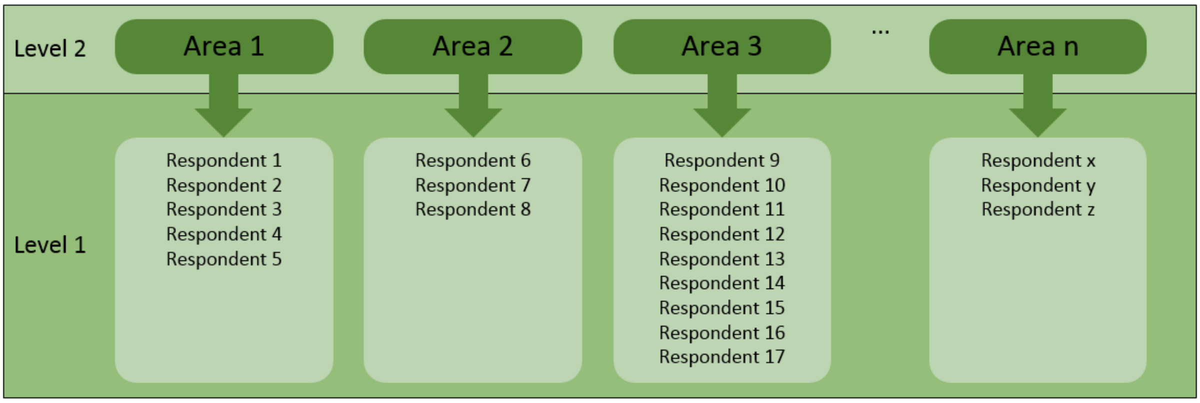


FIGURE 4 | HIERARCHICAL STRUCTURE (BASED ON: FIELD, 2013C)

The introduction of these different levels is important for several reasons. One of the reasons can be explained by the assumptions discussed with the logistic regression model. One of the assumptions states that errors are independent, so it is assumed that there is no correlation between residual scores of multiple respondents. However, when respondents come from the same area, so have the same context they live in, this assumption of independence is improbable to be the truth. Next to the bypass of the wrong assumption, multilevel models have several other benefits. The biggest of these benefits regards missing data in the dataset. Logistic regression models, like the multinomial regression model, cannot work with missing data and techniques to correct or impute missing data are often complicated. However multilevel models do not require complete datasets because parameters can be estimated by the use of the available data (Field, 2013c).

Another difference between the multinomial regression model and a multilevel linear model is the usage of fixed or random effects. Most regression models only use fixed variables with fixed intercepts and slopes, the variable and coefficients are not supposed to change over time. However with the use of a multilevel linear model a distinction should be made between fixed and random coefficients (Field, 2013c). The three possible scenarios using fixed and random coefficients are illustrated in Appendix I. The figures shows a dashed trend line from the total dataset and filled trend lines from the separate areas. As can be seen none of the areas follow the exact same trend line, so probably the area the respondents live in does affect their answers.

The latter of the three scenarios is the most realistic when trying to model the real world (Field, 2013c). But when using random coefficients the simple regression equations discussed in the previous paragraph have to be adjusted. Equation 2 from the previous paragraph is the starting point for the new equations.

The first scenario where the intercept is random and the slope is fixed needs a factor added to the intercept to make it random. The new equation is stated in Equation 7.

EQUATION 7 | REGRESSION FORMULA WITH RANDOM INTERCEPT AND FIXED SLOPE (FIELD, 2013C)

$$Y_{ij} = (\beta_0 + u_{0j}) + \beta_1 X_{ij} + \varepsilon_{ij}$$

The second scenario with a fixed intercept and a random slope, this factor is added to the slope instead of the intercept. This is shown in Equation 8.

EQUATION 8 | REGRESSION FORMULA WITH FIXED INTERCEPT AND RANDOM SLOPE (FIELD, 2013C)

$$Y_{ij} = \beta_0 + (\beta_1 + u_{1j}) X_{ij} + \varepsilon_{ij}$$

The last scenario with both intercept and slope random, combines the two equations above. Combined the equation can be found in Equation 9.

EQUATION 9 | REGRESSION FORMULA WITH RANDOM INTERCEPT AND SLOPE (FIELD, 2013C)

$$Y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j}) X_{ij} + \varepsilon_{ij}$$

Because the last equation is a bit messy and not very clear to instantly understand the additions are separately explained in formulas. This more orderly format of the formula is stated in Equation 10.

EQUATION 10 | REGRESSION FORMULA WITH FIXED INTERCEPT AND RANDOM SLOPE, ORDERLY (FIELD, 2013C)

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$$
$$\beta_{0j} = \beta_0 + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}$$

Because a multilevel linear model is an extension of regression all the assumptions for regression apply to this model. As stated above the assumption of independent errors can be solved by the use of a hierarchical regression model. However, also an assumption is added when using such a model because of the use of random coefficients. When these are used the intercepts and slopes are assumed to be normally distributed around the overall model. Finally multicollinearity can be a problem, if there are cross-level interactions, but this can be solved by centring predictors (Field, 2013c).

No measures to change the transport mode of the user will be taken into account in this model. The model will only identify potential bus users, regarding the characteristics of the circumstance.

THREE.4 | CONCLUSIONS

This research uses behavioral theories to describe travelers transport mode choice. To collect information for public transport agencies about which measures to implement to which customers, two models should be developed. The first model is a choice of transport model which models the location of potential bus users. The second model, which will be omitted for the time being in this research, regards the choice of route. The potential model can be made in several ways, the first attempt will be made with a multinomial regression model. If this model can be made, the second model that will be tried to create is a hierarchical regression model, this model has multiple advantages compared to the multinomial regression model. The last possible model is a nested logit model, this model can be used to decide which measures will be successful for certain individuals or areas.

CHAPTER FOUR | DATA

The data used for this research is mainly from an enquiry about travelers' behavior from a case study performed in 2009. This study and the area where it has been executed will be discussed in this chapter. Next to this, other data is needed about areal and travel aspects to create the bus potential models, these data is also described in this chapter. Finally a data analysis will be done to check whether the data fits for this research.

Despite the fact that in this report first the data is discussed and afterwards the models, this process was iterative. The prepared data could not immediately be implemented in a model and many modifications have been made to the data. New variables were made, new categories have been specified and all variables had to be recoded multiple times. The data discussed in this chapter is mostly the final form of the dataset which is used in the following chapter which discusses the models. The most important exception of this statement are the four position ZIP code data, which will be discussed in this chapter, but not again in Chapter Five.

During the literature study multiple influences, motivations, and barriers were found. But most of these aspects regard the service of the public transport itself e.g. sustainability, ability to work/study, lack of control, comfort. These aspects will not be taken into account in this model, because this model concerns the influence of personal and areal characteristics. Some of the aspects can be used, like the accessibility of the public transport. This aspect is represented by the distance to a bus stop and the service will be defined by the amount of bus lines in the area. If the costs of the transport mode are an important factor, this can be related to the income or education of the inhabitants. In that way also this aspect will be taken into account.

FOUR.1 | ENQUIRY DATA

This research has been done based on existing data. The most recent and comprehensive dataset is from 'verplaatsingsonderzoek Nuenen 2009' ('travel survey Nuenen 2009') presented by van der Waerden and Bérénos (Waerden and Bérénos, 2010). This survey was performed in the context of the Transumo project 'Regionale Openbaar Vervoer Bereikbaarheid Consumentgericht (ROVBECO)' ('Regional Public Transport Accessibility Project') (Waerden et al., 2009).

FOUR.ONE.1 | NUENEN

Because the survey which is used for this research was executed in Nuenen, Nuenen will be explored in this paragraph. First general information about Nuenen is given, after this the public transport is examined. For this last topic also changes like the HOV and how inhabitants react on these developments will be looked at.

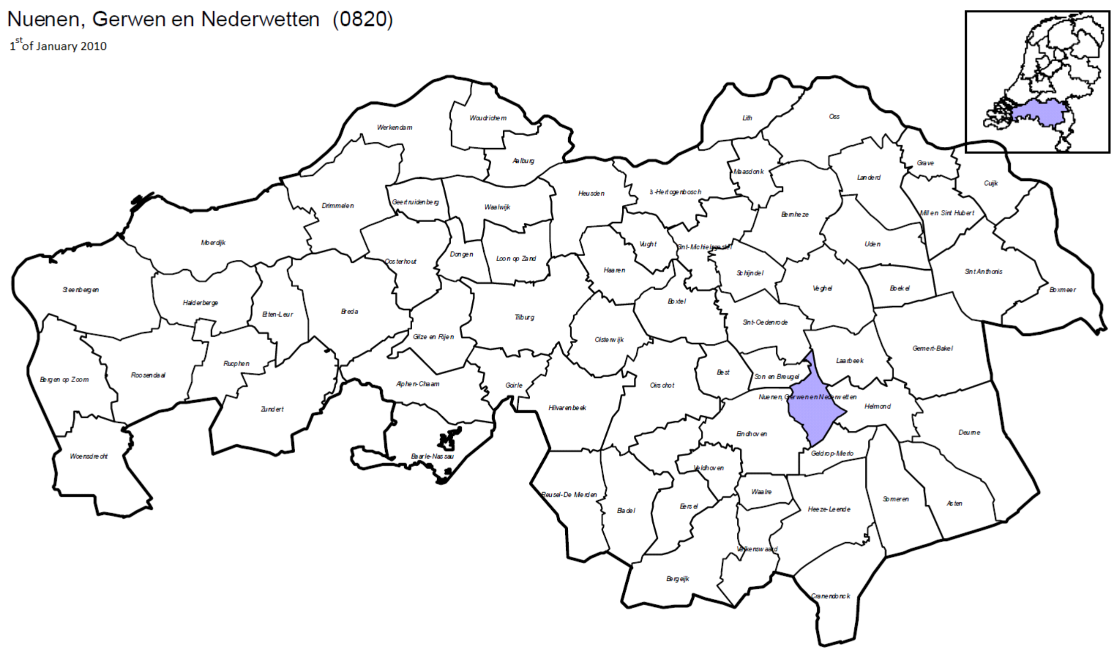


FIGURE 5 | LOCATION MUNICIPALITY OF NUENEN (CENTRAAL BUREAU VOOR DE STATISTIEK, 2011)

Nuenen is a village in "Noord-Brabant" located between Eindhoven and Helmond. The exact location of the municipality can be found in Figure 5. The town has four different four position ZIP code areas divided in ten neighborhoods, which are visualized in Figure 6. The municipality is a vibrant town with 20.000 inhabitants, where citizens; companies and institutions engage in many initiatives (Gemeente Nuenen, 2009).

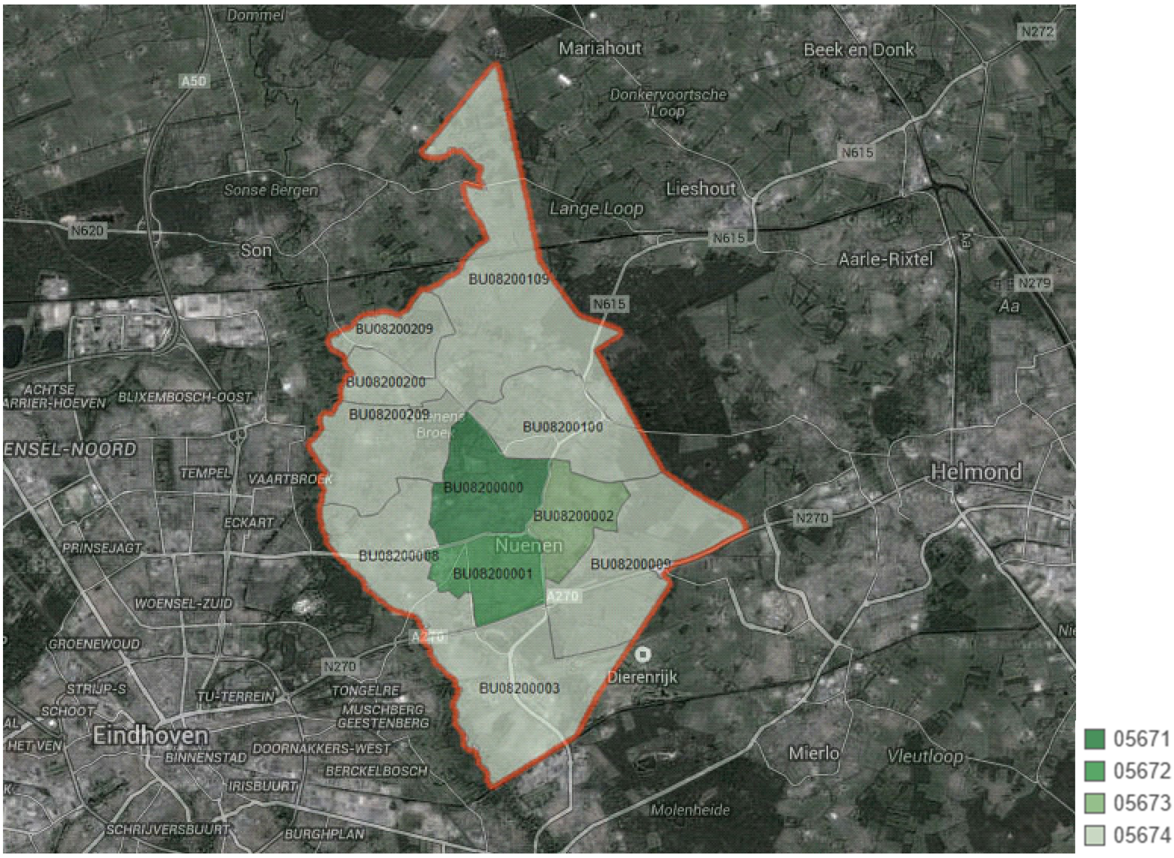


FIGURE 6 | NEIGHBORHOODS NUENEN GERWEN EN NEDERWETTEN (BASED ON: GOOGLE, 2015)

To lead all these new developments into the right direction the municipality is working on multiple spatial plans, like the 'Verkeersstructuurplan (VSP)' ('Transport Master Plan') determined in 2009. This master plan demonstrates a vision on the main traffic structure of Nuenen is described for the medium term. One of the mean measures to reduce the traffic density in the village is to stimulate the usage of bus and bicycle, this is already done by the implementation of a 'Hoogwaardig Openbaar Vervoer (HOV)' ('High Quality Public Transport') route (Movares Nederland B.V., 2012).

However, half of the inhabitants does not know about this new public transport system. Overall locals assess the accessibility of the examined city centers good by bike, however the accessibility by bus or car is experienced as middling. Accessibility of the center of Nuenen is assessed better with bus or bicycle than by car. In contrast to the reluctance of the bus accessibility to the city centers, the reachability to the nearest bus stop is well rated. The most common reasons for not going by bus is the unfavorable time travel ratio of car and bus, the absence of a direct connection and busses which do not drive frequently enough (Waerden and Bérénos, 2010).

FOUR.ONE.2 | DATA SURVEY

As stated before this research uses secondary data from 'verplaatsingsonderzoek Nuenen 2009' ('travel survey Nuenen 2009') presented by van der Waerden and Bérénos (Waerden and Bérénos, 2010). Because secondary data is used the potential bus users are already defined by the authors of the earlier research. The potential bus users are defined like the third definition discussed in paragraph Two.Two.5, so regarding switching-behavior. The respondents are asked whether the bus is an actual switching possibility for their trips. The general information about the dataset can be found in Table 7.

TABLE 7 | GENERAL INFORMATION DATASET NUENEN

	All Respondents	Non Bus-Users
Number of Respondents	1203	1007
Number of Trips	2552	2391
Number of 6 position Zip Codes	435	428
Male (%)	64	62
Mean Age	55	53
Average Family Size	2.2	2.7
Driver License (%)	97.1	97.8
Bus Trips (%)	2.6	-
Potential Bus Trips (%)		
Yes	-	12.1
Yes, But not Realistic	-	38.4
Don't Know	-	10.7
No Possibility	-	38.3

The first column shows the numbers of all respondent in the entire dataset. Because respondents who already use the bus are no potential bus users, the second column shows the numbers for all non-bus users. These Non Bus-Users data is the data which is used in all the models.

As can be seen in the table the 'Bus Potential' is divided in four different categories. 'Yes' regards to trips which could be done by bus, 'Yes, But not Realistic' regards to trips which could be done by bus but are not likely to switch, 'No Possibility' regards to trips which cannot be replaced by bus and the last category 'Don't Know' regards to trips from which the respondent does not know whether it could be done by bus. The last category could be a possible switch to bus-usage, so this category is counted as potential. In which case 61.2 percent of all trips could be changed to bus trips. This is a very big opportunity for the public transport sector.

The bus potential categories are visualized on a map in Figure 7.

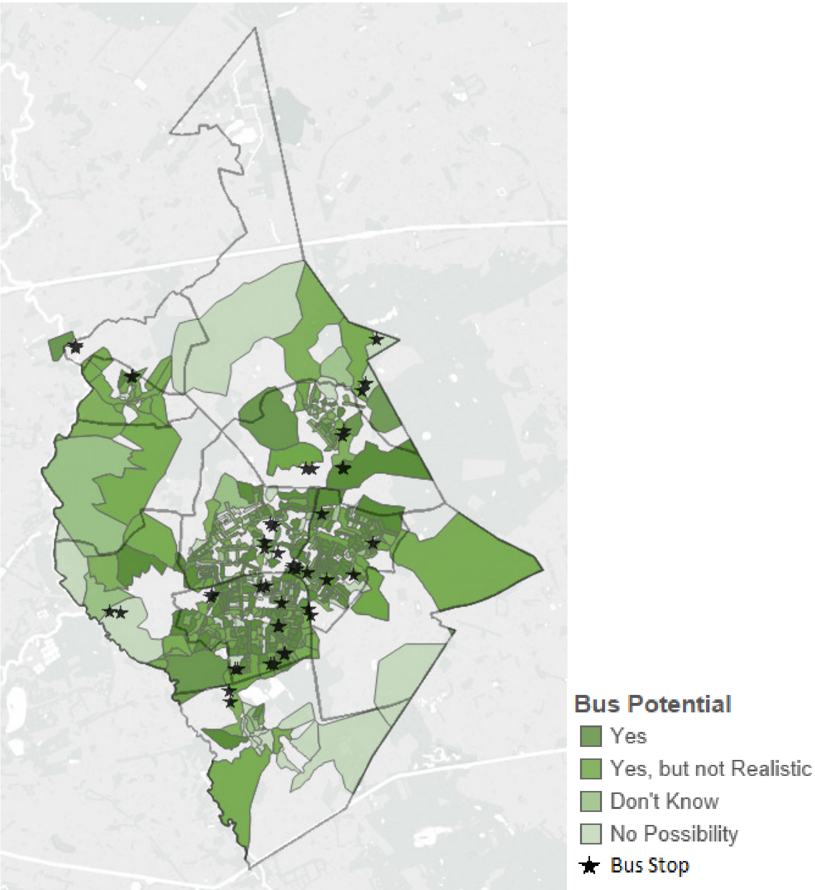


FIGURE 7| BUS POTENTIAL IN SIX POSITION ZIP CODE AREAS[2]

In the figure can be seen that most potential is near the center of Nuenen, this can probably be explained by the lack of bus stops in the outer areas of the village.

[2] Nuenen is not entirely represented in this map, because not all six position ZIP code areas where available and probably not all areas do have respondents.

In the research only regular journeys are analyzed because respondents know more about their frequent trips than of rarely made movements. Several kind of trips are questioned, these types are from home to work, from home to shops and from home to leisure activities. From all these trips the frequency, distance and travel time are known.

Table 8 and 9 show the variables obtained from the enquiry which will be examined to be significant in the model. Table 8 shows all possible personal variables, all these variables are from the individual level.

A Complete table with all possible variables, categories per variable and coding can be found in Appendix II. The variables are coded, using effect coding, to ensure all variables are equally distributed into the models. Dummy coding could also be used, because no interactions of variables were tested, however if this would be done in the future the variables are already prepared.

TABLE 8 | POSSIBLE PERSONAL VARIABLES

Variables	**Description**
Gender	Gender of the individual
Age	Age of the Respondent
Education	Level of Education
Family Size	Number of People in the Family
Family Structure	Composition of the Family
Housing Type	Type of Housing
Estate	Type of House Ownership
Driver's License	Obtained License
Owned Cars	Numbers of Cars Owned by Household
PT Usage	Frequency of Public Transport Usage
PT Subscription	Type of Public Transport Subscription

Table 9 states the possible trip variables, these trips are made by the previous discussed individuals.

TABLE 9 | POSSIBLE TRIP VARIABLES

Variables	**Description**
Motive	Motive for the Trip
Transport Mode	Transport Mode of the Trip
Departure Time	Time of Departure
Trip Frequency	Frequency of the Trip
Trip Distance	Distance of the Trip
Trip Duration	Time Duration of the Trip

FOUR.2 | AREAL DATA

Next to the data of the questionnaire, data about the characteristics of the area are needed. This data will be taken from open source databases, like CBS, processed by Irias Informatiemanagement (Irias, n.d.), additional data about households is obtained from Bridgis

(Bridgis, n.d.), and data of bus services in an area comes from Goudappel Coffeng (Goudappel Coffeng, n.d.) combined with open data.

Nuenen only exists of four different four position ZIP code areas, from which three of the areas form the center and the rest of Nuenen is the fourth area. When using these four areas the level of detail is quite gross and little variance is present in the variables. However initially only four position ZIP code areal variables where available from Irias Informatiemanagement. After only a few tries to make a multinomial model with these variables it became clear that these variables will have no effect on the model when used together. The first variable used would have a significant effect on the model, but because all variables have the same very small variance all additional areal variables had no effect anymore. Therefore these four position ZIP code variables were discarded and the remaining of the research was done with another scale, six position ZIP code variables where sought and used.

Nuenen has 603 different six position ZIP code areas, in the dataset only 428 were represented. Table 10 shows the possible areal variables at six position ZIP code scale obtained from Bridgis. Either the general number, the average of the area or the most frequent (dominant) category is available.

TABLE 10 | POSSIBLE AREAL VARIABLES SIX POSITION ZIP CODE

Variables	**Description**
Surface	Surface of the Area
Inhabitants	Number of People in the Area
Education	Dominant Level of Education in the Area
Households	Number of Households in the Area
Family Size	Average Number of People per Household in the Area
Family Structure	Dominant Composition of a Family in the Area
Housing Type	Dominant Housing Type in the Area
Year of Construction	Average Year of Construction in the Area
Property Value	Average Property Value in the Area
Estate	Dominant Type of House Ownership in the Area

Next to the general data of the area a public transport service level will be specified. The variables with regard to this service are stated in Table 11. These variables are also on the six position ZIP code scale.

TABLE 11 | POSSIBLE BUS VARIABLES SIX POSITION ZIP CODE

Variables	**Description**
Bus Stops	Number of Bus Stops Available in the Area
Bus Lines	Number of Bus Lines Available at Stops in the Area
Bus Lines Area	Number of Bus Lines Available in the Area
Bus Distance	Distance from Centre of the Area to Nearest Bus Stop

All these possible areal variables for the model can also be found in the complete variable description from Appendix II.

FOUR.3 | DATA ANALYSIS

One of the most important aspects of a survey is that the sample used represents the actual population of the study area. Therefore the CBS and MOA developed the 'Gouden Standaard' (Golden Standard), which is a unique calibration instrument to ascertain whether a sample fits the actual population (MOA, 2014). The idea of the instrument is to use it to construct the survey, but in this case the data is already collected from a survey. Therefore this collected data will be compared to the Golden Standard. However this instrument is very extensive and will not be fully investigated in this thesis, only five of the variables used in the Golden Standard will be examined. These five variables are chosen because they are the key variables in the Golden standard and they also are used in the 'Mobiliteitspanel Nederland' (Mobility Panel Netherlands) (Kalter et al., 2014). The Mobility Panel uses six variables, but the 'Level of Urbanization' cannot be used in this case, because the survey does not handle this subject. Therefore two of the variables regard households, the other three of the variables are personal.

To make the charts all respondents of the survey are counted and settled into percentages. The data from Nuenen is either obtained from Bridgis or Irias Informatiemanagement. This data is only available as an average per zip code area and therefore had to be rescaled to the household or personal level. With this in mind there is a possibility that the differences in the charts can be explained by this rescaling.

Figure 8 shows the household variables 'Family Structure' and the 'Amount of Cars'.

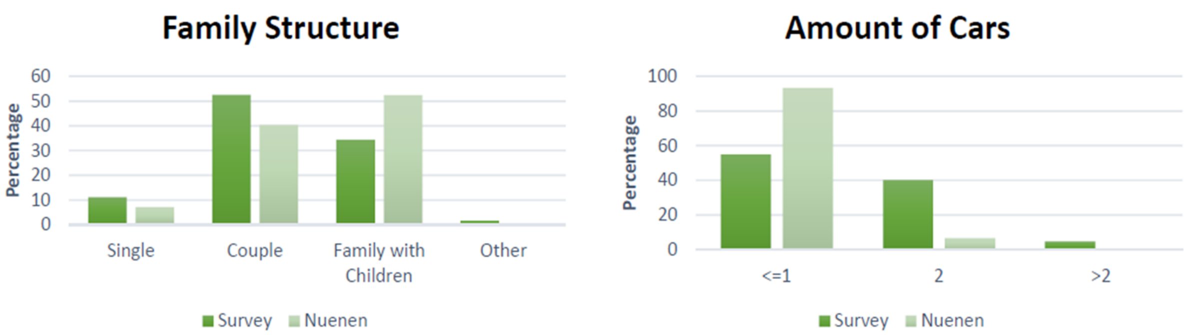


FIGURE 8| GOLDEN STANDARD HOUSEHOLD VARIABLES

As can be seen in the charts, the survey is not equally distributed as the actual population in Nuenen. A sample can never by perfectly representative for the population, but in this case there are rather large differences. There were significantly more couples represented in the survey instead of the family with children which are more present in the actual population. The distribution of the amount of cars owned by a household follow the same course in both the survey and Nuenen. However the households with only 1 car or less are less prominent in the survey, this difference in number is found at the 2 cars category.

Figure 9 states the charts with the personal variables of the survey and the actual population. The same facts about the rescaling apply to these charts.

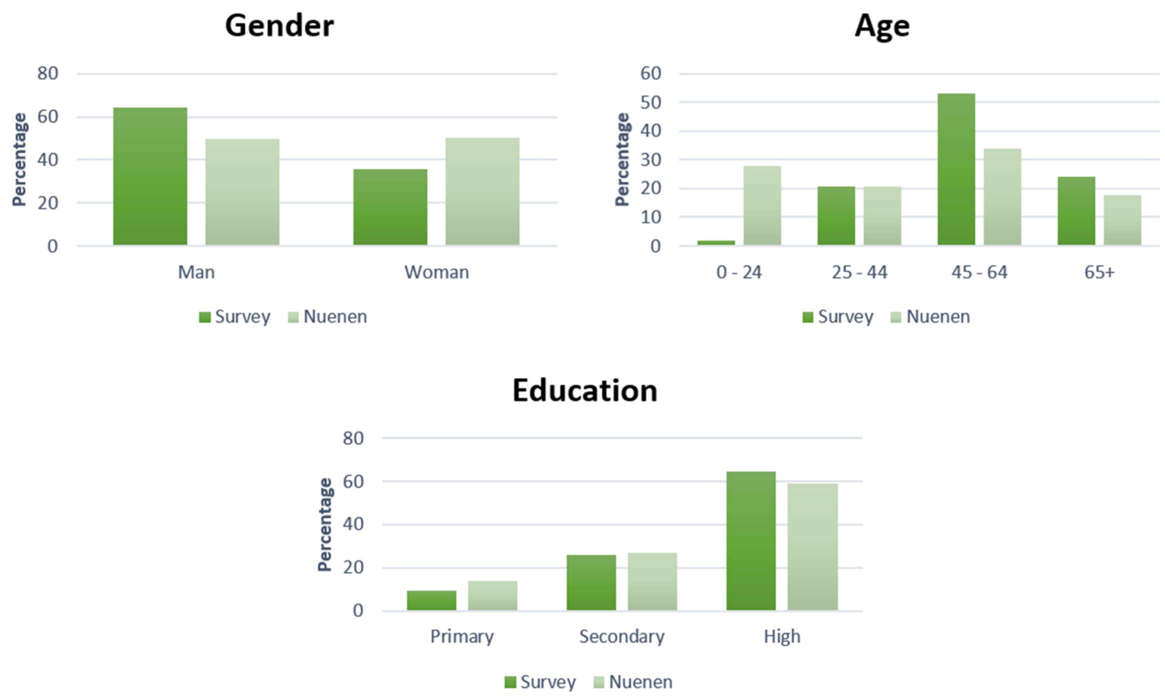


FIGURE 9 | GOLDEN STANDARD PERSONAL VARIABLES

The education level is almost the same in the survey as in the actual population, so here no irregularities can be expected. The same cannot be told about the gender and age representations. In reality the population of Nuenen is equally divided between men and women, but the survey is completed by significantly more men. From the age chart can be seen that the youth is not enough represented by the survey and the category from 45 to 64 is overrepresented.

One of the causes of these differences between the survey and the actual population could be a selective non-response group (Kalter et al., 2014). Maybe men are overall more willing to cooperate in surveys, and possibly young people did not fill in the questionnaire because they have not obtained their driver's license yet and are not interested in the outcome of the research.

FOUR.4 | FINDINGS

The data which would be used in this research comes from a survey presented by the 'Verplaatsingsonderzoek Nuenen 2009' ('travel survey Nuenen 2009') . This survey was performed in the context of the Transumo project 'Regionale Openbaar Vervoer Bereikbaarheid Consumentgericht (ROVBECO)' ('Regional Public Transport Accessibility Project'). The area described in by the dataset is Nuenen, a village in Noord-Brabant, between Eindhoven and Helmond. The village has four different four position ZIP code areas, which is too gross to perform a model on. Therefore sic position ZIP code areas will be used.

The bus potential is determined to have four categories in the enquiry. These categories will adopted in this research, the categories are: 'Yes', Yes, but not Realistic', 'Don't Know' and 'No Possibility'. Next to this dependent variables, possible independent variables are determined.

The independent variables are divided in the following four groups: ‘Personal’, ‘Trip’, ‘Areal’, and ‘Bus’.

The sample composition is analyzed by the use of the golden standard, and it is not equally distributed as the actual population in Nuenen. However a sample can never be perfectly representative for the population, but in this case there are rather large differences. It can be doubted whether the data is a good representative of the population.

CHAPTER FIVE | MODELS

The models created to predict the potential public transport users in an area are described in this chapter. First, crosstabs with correlations between all variables are discussed, after this the multinomial regression models are described.

To make a hierarchical regression model, first the variables that should be used had to be selected. This is done by the use of crosstabs which show the correlation between different variables and the use of multinomial logistic regression models. When the variables are picked using these methods, the chosen variables should be checked whether they fit the right assumptions, as discussed earlier in paragraph Three.Three.1.

FIVE.1 | CORRELATION ANALYSES

The variables which will be examined in this research are presented in Tables 8 to 11 of the previous chapter. However not all of these variables are taken into account. Whether the respondent has a driver's license cannot be used in the model because only 2.1 percent of all the respondent has no obtained license, see Table 7. This will result in a violation of the second assumption for the use of the Chi-square test, which represents the correlations between variables. The lack of respondents without a driver's license can be explained by the fact that the dataset has few young respondents as is explained in Paragraph Four.3.

To study the correlations between the possible variables IBM SPSS Statistics is used. An analysis is done with the Descriptive Statistics Crosstabs. The Chi-square is the most important aspect in this analysis because this will tell whether the correlation is significant. If this is the case, it may be better not to use the two variables in the Crosstab together in the model because they could affect each other.

All kind of model analyses have to meet certain assumptions, the Chi-square test is no exception to this. However, only two assumptions have to be met:

1. For the Chi-square to be meaningful there have to be an independence of residuals, each person; item or entity contributes to only one cell of the contingency table.
2. No expected value in the table should be below five when using a two by two contingency table, so two variables with two categories each. When using variables with more categories, all expected counts should be greater than one and no more than 20 percent of the expected counts should be less than five (Field, 2013d).

To be able to implement the Chi-Square test all continuous variables are binned into categorical variables in such a way that all categories are more or less equally represented. This will lead to a minimal amount of violations of the second assumption.

The correlation matrixes are constructed as following. If the Pearson Chi-Square has a significance of 0.050 or less the combination of variables gets a 'c' for that case, this indicates a significant correlation between the variables. If the test shows more than 20 percent of the cells to have expected count less than 5, so more than 20 percent of the cells violate the second assumptions of the Chi-square test, no model could be made and the combination of variables will get a '-' in that case. If two possible variables correlate this indicates to a 'Bad Match', when no significant correlation is found the case is a 'Good Match'. Combinations between a possible variable and the 'Bus Potential' are just the opposite, if there is a correlation it indicates a 'Good Match', otherwise a 'Bad Match'. When no test was possible the case will get a 'Not Possible' indication.

The first variables that are checked using the Chi-square test are the personal variables from Table 8. As mentioned before the variable 'Driver's License' cannot be used in the analysis because of the lack of respondents with no obtained license. The correlation matrix can be seen in Table 12.

TABLE 12 | CORRELATIONS PERSONAL VARIABLES

	Gender	Age	Education	Family Size	Family Structure	Housing Type	Housing Estate	Owned Cars	PT Usage	PT Subscription
Gender	-	-	-	-	-	-	-	-	-	-
Age	c	-	-	-	-	-	-	-	-	-
Education	c	c	-	-	-	-	-	-	-	-
Family Size	c	c	c	-	-	-	-	-	-	-
Family Structure	c	c	c	c	-	-	-	-	-	-
Housing Type	c	c	c	c	c	-	-	-	-	-
Housing Estate	c	c	c	c	c	c	-	-	-	-
Owned Cars	c	c	c	c	c	c	c	-	-	-
PT Usage	c	c	c	c	c	c	c	c	-	-
PT Subscription	x	c	c	c	c	x	x	c	c	-
Bus Potential	x	c	c	c	c	c	x	x	c	c

c = Correlation x = No Correlation = Good Match = Bad Match

The table shows that almost all variables correlate with each other, so it is probably not possible to use all the personal variables in the model. Only 'PT Subscription' does not correlate to three of the other variables. So, it is possible to use these variables together in the model. However, the bottom row showing the correlation between the variables and the bus potential states that 'Gender' and 'Housing Estate' do not correlate to 'Bus Potential'. This can be interpreted in such a way that these variables will probably have no significant effect in the model.

'Amount of Cars' as well has no significant correlation to the 'Bus Potential'. It is assumed that it is important whether a person has a car to predict potential of the public transport. However the category 'no cars owned' cannot be taken into account because only 16 of the 2391 trips are filled in by a respondent with no car. When this category is used the variable 'Cars Owned' does correlate with the 'Bus Potential', but because of the small amount in the 'no car' category the second assumption of the Chi-Square test is violated and the validity of the test in uncertain.

The significance of the other variables for the model are not totally unexpected. 'Age' is assumed to correlate with the 'Bus Potential', because young people have no car and are therefore forced to use the public transport and old people get discount so will probably use the bus more often. This will probably lead to a parabolic function, but because the young people are disregarded a linear relation is detected. 'Education' is assumed to correlate with the 'Bus Potential', because higher educated people are believed to have their own car and do not like to use public transport services. The correlation with the 'Household Size' and 'Family Structure' can be explained by the fact that the car gets cheaper when more people drive in it, so households with children will probably take the car. The 'PT Usage' and 'PT Subscription' have very obvious correlations with the 'Bus Potential'. If someone already uses the public transport often this person is probably more willing to take the bus and can quickly be counted as potential. The same applies to a person with a public transport subscription.

So, from this first correlation matrix conclusions can be made about the variables which could affect the model. 'PT Subscription' and 'Housing Type' can be used together, and 'Gender'; 'Housing Estate' and 'Amount of Cars' probably have no significant effect on the model.

The second crosstab stating the correlations between the trip variables, from Table 9, can be found in Table 13.

TABLE 13 | CORRELATIONS TRIP VARIABLES

	Motive	Transport Mode	Departure Time	Trip Frequency	Trip Distance	Trip Duration
Motive	-	-	-	-	-	-
Transport Mode	c	-	-	-	-	-
Departure Time	c	c	-	-	-	-
Trip Frequency	c	c	c	-	-	-
Trip Distance	c	c	c	c	-	-
Trip Duration	c	c	c	c	c	-
Bus Potential	c	c	c	c	c	c

c = Correlation x = No Correlation = Good Match = Bad Match

From this crosstab nothing can be concluded about the combination of trip variables that should be used, because, all variables correlate with each other. However all variables also do correlate with the 'Bus Potential', probably all of them will separately have an effect on the model. These correlations with the 'Bus Potential' come as no surprise, since it is assumed that trip specifics have great influences on the transport mode choice.

So, nothing can be said about this variable type at the moment. All trip variables will probably have an effect on the model, but because all variables also correlate to each other they can probably not all be used together in the model.

The correlation matrix regarding the six position ZIP code variables, from Table 10, can be found in Table 14.

TABLE 14 | CORRELATIONS SIX POSITION ZIP CODE AREAL VARIABLES

	Surface	Inhabitants	Education	Households	Family Size	Family Structure	Housing Type	Year of Construction	Property Value	Estate
Surface	-	-	-	-	-	-	-	-	-	-
Inhabitants	c	-	-	-	-	-	-	-	-	-
Education	c	c	-	-	-	-	-	-	-	-
Households	c	c	c	-	-	-	-	-	-	-
Family Size	c	c	c	c	-	-	-	-	-	-
Family Structure	c	c	c	c	c	-	-	-	-	-
Housing Type	c	c	c	c	c	c	-	-	-	-
Year of Construction	c	c	c	c	c	c	c	-	-	-
Property Value	c	c	c	c	c	c	c	c	-	-
Estate	c	c	c	x	c	x	c	c	c	-
Bus Potential	c	x	c	x	x	c	c	c	x	x

c = Correlation x = No Correlation = Good Match = Bad Match

In this table almost all variables correlate with each other, it is assumed that they cannot all be used together in the model. Only 'Estate' could be used in the model together with 'Households' and 'Family Structure'. However 'Estate' and 'Households' both do not correlate with the 'Bus Potential', so will probably not significantly affect the dependent variable. But five other variables do significantly correlate with the 'Bus Potential'. These variables regard the surface, dominant education level, dominant family structure, dominant housing type and the dominant year of construction in the area.

The number of inhabitants and households seem to have no effect on the bus potential of the area. This could be explained by the fact that a transport mode choice is personal and trip related, but has nothing to do with the amount of people surrounding a person. The size of a family also does not correlate to the potential, this is against expectation because this variable did correlate to the bus potential on the personal scale. Also the dominant property value of the area does not correlate to the 'Bus Potential'. This is less expected, because it is assumed that the bus potential is influenced by the income of people and the income will probably correlate to the property value of the house a person lives in. Next to this 'Education' does correlate to the 'Bus Potential', and it is assumed that education affects the income, which affects the property value. Perhaps this irregularity can be explained by the fact that the dominant property value category is used in the dataset and not the average property value of the area. The last variable which probably has no significant effect on the model is 'Estate', this variable also did not correlate to the 'Bus Potential' on the personal scale, so there is no contradiction here.

The last group of variables also regards six position ZIP code areas, but is all about the public transport service. The bus variables from Table 11 are examined in the correlation matrix stated in Table 15.

TABLE 15| CORRELATION BUS VARIABLES

	Bus Stops	Bus Lines at Stops	Bus Lines in Area	Distance to Bus Stop
Bus Stops	-	-	-	-
Bus Lines at Stops	-	-	-	-
Bus Lines in Area	c	c	-	-
Distance to Bus Stop	c	c	c	-
Bus Potential	x	x	x	x

c = Correlation x = No Correlation = Not Possible = Bad Match

In this case all variables, except for the distance variable, are divided in yes or no variables. So whether there are stops, lines at stops or lines in the area. Probably because far more areas do not have stops a violation of the second assumption of the Chi-Square Test occurs as can be seen by the yellow 'Not Possible' box in the matrix. Disregarding this box all variables correlate with each other and therefore can probably not be used all together in a model. Furthermore can be seen that none of the variables seem to correlate to the 'Bus Potential'. Which means none of the variables will presumably have a significant effect on the model.

Because the assumption is that the public transport service will be of importance for the bus potential model the previously used variables are also examined with another categorization. Most of the areas do not have a bus stop, therefore only those areas with bus stops are taken into account. The same applies to the lines at the stops and through the area, only those cases where the value does not equal zero are taken into account. Also the 'Distance to Bus Stop' will get other categories. An explanation for the lack of correlation can be that the relation is not linear. Maybe the first 100 meters do have impact, but the difference between 500 and 1000 meters have no effect on the potential. In this case the relationship could be exponential instead of linear. Therefore the linear categorization is disregarded and an exponential one will be used. The correlation matrix with these newly categorized variables can be found in Table 16.

TABLE 16| CORRELATIONS BUS VARIABLES 2

	Bus Stops	Bus Lines at Stops	Bus Lines in Area	Distance to Bus Stop
Bus Stops	-	-	-	-
Bus Lines at Stops	c	-	-	-
Bus Lines in Area	-	-	-	-
Distance to Bus Stop	c	c	-	-
Bus Potential	-	-	x	c

c = Correlation x = No Correlation = Not Possible = Good Match = Bad Match

Possibly due to the disregard of all zero cases here the second assumption for the use of the Chi-Square Test is violated at half of the cases, this can be seen by the 'Not Possible' indications. Like the first matrix all other possible variables correlate with each other, but one of the variables does also correlate to the 'Bus Potential'.

So, the new categorization of the bus variables did not contribute to the correlations, it only seemed to make it worse because now more cases did violate the second assumption. However, the new categorization of the 'Distance to Bus Stop' causes this variable to have a possible effect on the model.

Because all other possible bus related variables do not correlate to the 'Bus Potential' the variable 'Distance to Bus Stop' will be put in with the areal variables. The new areal correlations matrix is stated in Table 17.

TABLE 17| CORRELATIONS AREAL VARIABLES 2

	Surface	Inhabitants	Education	Households	Family Size	Family Structure	Housing Type	Year of Construction	Property Value	Estate	Distance to Bus Stop
Surface	-	-	-	-	-	-	-	-	-	-	-
Inhabitants	c	-	-	-	-	-	-	-	-	-	-
Education	c	c	-	-	-	-	-	-	-	-	-
Households	c	c	c	-	-	-	-	-	-	-	-
Family Size	c	c	c	c	-	-	-	-	-	-	-
Family Structure	c	c	c	c	c	-	-	-	-	-	-
Housing Type	c	c	c	c	c	c	-	-	-	-	-
Year of Construction	c	c	c	c	c	c	c	-	-	-	-
Property Value	c	c	c	c	c	c	c	c	-	-	-
Estate	c	c	c	x	c	x	c	c	c	-	-
Distance to Bus Stop	c	c	c	c	c	c	c	c	c	c	-
Bus Potential	c	x	c	x	x	c	c	c	x	x	c

c = Correlation x = No Correlation = Good Match = Bad Match

The new variable does correlate to any of the other possible areal variables, so no new conclusion about the usage of combinations of areal variables in the model can be made.

The conclusion that can be made from all the correlation matrices is that it is still not clear which variables should be used in a model to describe 'Bus Potential'. However, 19 of the 30 possible variable did show a correlation with the 'Bus Potential', so could probably be used to describe the dependent variable in the model.

FIVE.2 | MULTINOMIAL LOGISTIC REGRESSION

As mentioned before: MNL is Logistic regression in which the outcome variable has more than two categories (Field, 2013a). This regression model will be used to check the significance of the variables analyzed in the previous paragraph. Also in this case all variable types will initially be analyzed separately. The first models will be with individual variables, to see whether the correlations found in the previous paragraph are also significant in a multinomial regression model. Then integral models with different selection methods are made to see if a working model can be made using all variables. If no working model comes from the integral models, step-by-step models are tested. In this case all variables will be put into the model manually. If a successful model is made, the variables used should be checked by the assumptions of the multinomial regression method, which will be done at the end of this paragraph. The variables used in the multinomial regression model will be proven to be significant for the bus potential model and should be used in a hierarchical regression model.

FIVE.TWO.1 | INDIVIDUAL MODELS

Initially all single variables are individually put in a multinomial regression model with dependent variable 'Bus Potential'. As explained in Chapter Four the only non-potential category from the variable 'Bus Potential' is the 'No Possibility', therefore this category is used as 'Reference Category' in the model. This means the model checks whether the other three categories are more or less likely to be picked above the 'No Possibility' category.

All significant variables will be discussed in the next section, except variables with a warning in the model. It seems all continuous variables get the same warning regarding empty subpopulations, this warning is illustrated in Figure 10.

Warnings

There are n (x %) cells (i.e., dependent variable levels by subpopulations) with zero frequencies.

FIGURE 10| ZERO FREQUENCIES WARNING

This warning is given when the model has to use subpopulations with no cases. Because these empty subpopulations are practically inevitable when using covariates the warning is often ignored, however empty cells can create problems (Field, 2013a). Table 18 shows the Crosstab with ‘Bus Potential’ and the continuous variable ‘Family Size’.

TABLE 18 | CROSSTAB 'BUS POTENTIAL' AND 'FAMILY SIZE'

		Bus Potential				Total
		Yes	Yes, but not Realistic	Don't Know	No Possibility	
Family Size	0	0	4	2	1	7
	1	38	97	35	88	258
	2	137	435	112	496	1180
	3	35	115	38	111	299
	4	52	202	57	153	464
	5	23	56	12	69	160
	6	4	6	0	7	17
	8	0	3	0	0	3
	12	0	1	0	2	3
Total		289	919	256	927	2391

The subpopulations with zero frequencies are found by a family size with six or more members. For example no trips for which the bus is a realistic replacement are made by a respondent with a family of eight or twelve people. This warning disappears when the categorical variable is used, because the last category refers to families with five or more members.

Table 19 to 22 show the summaries of all individual models with respectively personal, trip, areal and bus variables. The tables show the variable used, the type of variable, the start and final -2 Log Likelihoods, the Chi-square, the degrees of freedom, and the significance of the model. The final column in the tables states the warnings given when estimating the mode. The Chi-Square is the difference between the -2 Log Likelihoods, the lower the Final -2 Log Likelihood, the higher the Chi-Square, the better the model. The degrees of freedom of a model says how many aspects of the model can be altered, when making a tool high degrees of freedom are preferable. A model is seen as a ‘Significant Model’ when the significance is 0.050 or less, otherwise the model will be ‘No Significant Model’.

When compared to the correlation crosstabs from Paragraph Five.1, can be concluded that the same variables prove to be significant. All models with continuous variables give a warning about zero frequencies, which was explained above. The manner in which the remaining variables have a significant effect on the model will be discussed next. This will be done by the use of the model’s parameter estimates which can be found in Appendix III.

The first group of variables regards the personal characteristics of the respondents, summaries of the individual models are stated in Table 19 and the parameter estimates can be found in Appendix III.1.

TABLE 19 | MNL INDIVIDUAL PERSONAL VARIABLES

VARIABLE	Type	-LL2 Start	-LL2 Final	Chi²	df	Sign.	Warning
Gender	Categoric	46.676	40.869	5.806	3	0.121	-
Age	Continuous	890.821	845.749	45.073	3	0.000	16.2% with zero frequencies
	Categoric	105.725	67.495	38.230	9	0.000	-
Education	Categoric	98.337	55.342	42.995	6	0.000	-
Family Size	Continuous	138.846	136.749	2.097	3	0.553	19.4% with zero frequencies
	Categoric	111.273	85.42	25.850	12	0.011	-
Family Structure	Categoric	87.110	67.193	19.918	9	0.018	-
Housing Type	Categoric	92.716	71.933	20.783	9	0.014	-
Housing Estate	Categoric	44.959	37.627	7.333	3	0.062	-
Owned Cars	Categoric	61.498	54.034	7.465	6	0.280	-
PT Usage	Categoric	124.427	66.051	58.376	9	0.000	-
PT Subscription	Categoric	56.768	39.688	17.079	3	0.001	-

= Significant Model	= No Significant Model

The first variables which proofs to be significant for the model is 'Age'. This variable will be discussed extensively, therefore the Parameter Estimates of the model with variable 'Age' are besides Appendix III.1 also stated in Table 20.

TABLE 20 | PARAMETER ESTIMATES MNL AGE MODEL

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.303	.151	74.701	1	.000			
	[Personal_Age=1]	.674	.463	2.119	1	.146	1.962	.792	4.862
	[Personal_Age=2]	.356	.209	2.903	1	.088	1.428	.948	2.152
	[Personal_Age=3]	.087	.176	.247	1	.619	1.091	.773	1.540
	[Personal_Age=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.401	.110	13.264	1	.000			
	[Personal_Age=1]	.688	.359	3.679	1	.055	1.990	.985	4.021
	[Personal_Age=2]	.833	.148	31.843	1	.000	2.301	1.723	3.073
	[Personal_Age=3]	.329	.126	6.781	1	.009	1.390	1.085	1.780
	[Personal_Age=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.590	.169	88.221	1	.000			
	[Personal_Age=1]	.268	.588	.209	1	.648	1.308	.413	4.138
	[Personal_Age=2]	.674	.222	9.221	1	.002	1.962	1.270	3.031
	[Personal_Age=3]	.262	.194	1.826	1	.177	1.299	.889	1.900
	[Personal_Age=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

Because all intercepts are negative and significant, it can be concluded that all bus potential categories are less promising than 'No Possibility'. The 'Yes' bus potential category is not influenced by the age of people, however the second and third bus potential categories, respectively 'Yes, but not Realistic' and 'Don't Know', have significant age categories which can improve the chance of choosing this category. In both cases age category two is significant and positive, within bus potential category two, as well age category three is significant and positive. So, people between 25 and 64 years old are more willing to choose the 'Yes, but not

Realistic' than the 'No Possibility' bus potential category. The people between 25 and 44 have are as well more willing to choose the 'Don't Know' over the 'No Possibility' bus potential category.

The remaining variables will be more generally described, so not as detailed as 'Age'. Most of the models with personal variables give an equal probability for bus potential categories two and four, respectively 'Yes, but not Realistic' and 'No Possibility'. This is not contrary to expectations because these two answers were the most given in the enquiry, as can be seen in Table 7 from Paragraph Four.One.2. The only variable which influences the 'Yes' category of bus potential is 'PT usage' , however this affect is negative so will only lower the potential. However all variables except for 'Housing Type' do influence the 'Don't Know' category, 'Housing Type' only affects the 'Yes, but not Realistic' category. Next to the previous discussed variable 'Age', also 'Education' and 'Family Size' have a slightly positive influence on the second and a quite high positive influence on the third category of bus potential. The only variable which does not seem to have any effect on the 'Bus Potential' is 'Family Structure', all of the intercepts are not significant, and so all categories have the same chance.

The individual models with Trip variables are shown in Table 21 and the parameter estimates are stated in Appendix III.2. Only the conclusions which can be drawn for these models will be discussed, for the manner of interpretation is the same as described before.

TABLE 21 | MNL INDIVIDUAL TRIP VARIABLES

VARIABLE	Type	-LL2 Start	-LL2 Final	Chi²	df	Sign.	Warning
Motive	Categoric	140.119	80.420	59.698	12	0.000	-
Transport Mode	Categoric	85.957	40.344	45.613	3	0.000	-
Departure Time	Categoric	220.807	86.711	134.096	12	0.000	-
Frequency	Categoric	135.257	81.692	53.565	12	0.000	-
Distance	Categoric	418.452	71.351	347.101	9	0.000	-
Duration	Categoric	466.566	85.500	381.066	12	0.000	-

= Significant Model

Also here most of the models have an equal chance for 'Yes, but not Realistic', 'Don't Know' and 'No Possibility'. The intercepts of these two categories, 'Yes, but not Realistic' and 'Don't Know', are not significant. Therefore, they can be set to zero, and are not more or less likely to be chosen over the reference category 'No Possibility'. The other models only tend to 'No Possibility'. Except for the model regarding 'Motive', this is, together with the 'Year of Construction model' which will be discussed with the areal variables, the only model which has the biggest probability for the second bus potential category 'Yes, but not Realistic'. The same as with the personal variables all trip variables have a very diverse effect on the 'Bus Potential'. However, the influences from the trip variables are even more various because quite a lot of variable categories have positive and negative impact on the bus potential categories.

The next individual models regard the areal, the general information of the models is stated in Table 22 and the parameter estimates can be found in Appendix III.3. Also here only the conclusions which can be drawn for these models will be discussed, for the manner of interpretation is the same as described before.

TABLE 22 | MNL INDIVIDUAL SIX POSITION ZIP CODE AREAL VARIABLES

VARIABLE	Type	-LL2 Start	-LL2 Final	Chi²	df	Sign.	Warning
Surface	Continuous	2856.578	2831.122	25.455	3	0.000	40.6% with zero frequencies
	Categoric	111.065	87.843	23.222	12	0.026	-
Inhabitants	Continuous	1090.174	1085.930	4.244	3	0.236	14.8% with zero frequencies
	Categoric	89.996	73.105	16.892	9	0.050	-
Education	Categoric	70.873	54.919	15.954	6	0.014	-
Households	Continuous	628.262	624.930	3.332	3	0.343	8.9% with zero frequencies
	Categoric	82.474	73.586	8.888	9	0.448	-
Family Size	Continuous	1772.589	1765.959	6.630	3	0.085	28.2% with zero frequencies
	Categoric	66.625	54.735	11.890	6	0.064	-
Family Structure	Categoric	76.329	54.782	21.546	6	0.001	-
Housing Type	Categoric	89.175	69.007	20.168	9	0.017	-
Year of Construction	Categoric	118.103	71.833	46.270	9	0.000	-
Property Value	Categoric	99.265	84.699	14.566	12	0.266	-
Estate	Categoric	44.281	39.079	5.202	3	0.158	-

= Significant Model	= No Significant Model

All the individual models with areal variables, except for the earlier mentioned 'year of Construction' have the same probability for the 'Yes, but not Realistic', 'Don't Know' and 'No Possibility' bus potential categories. Also at these models the intercepts of these two categories, 'Yes, but not Realistic' and 'Don't Know', are not significant. Therefore, they can be set to zero, and are not more or less likely to be chosen over the reference category 'No Possibility'. The 'surface' and 'Education' of an area have some positive influences on the 'Bus Potential', but the other three variables only negatively affect the models. 'Year of construction' even influences all the categories negatively.

The last group of variables are the Bus variables. The summary of the models is shown in Table 23. Only one of the variables result in a significant model, the parameter estimates of this model are stated in Appendix III.4.

TABLE 23 | MNL INDIVIDUAL BUS VARIABLES

VARIABLE	Type	-LL2 Start	-LL2 Final	Chi²	df	Sign.	Warning
Bus Stops	Continuous	55.724	54.462	1.261	3	0.738	25% with zero frequencies
	Categoric	23.508	22.179	1.329	3	0.722	-
Bus Stops 2	Categoric	36.044	35.620	0.424	3	0.935	-
Lines at Stops	Continuous	77.135	75.710	1.425	3	0.700	17.9% with zero frequencies
	Categoric	23.491	20.388	3.103	3	0.376	-
Lines at Stops 2	Categoric	36.803	34.801	2.002	3	0.572	-
Lines in Area	Continuous	146.540	144.996	1.544	3	0.672	22.7% with zero frequencies
	Categoric	84.136	68.221	15.915	12	0.195	-
Lines in Area 2	Categoric	48.471	40.904	7.566	3	0.056	-
Distance	Continuous	2856.578	2827.148	29.430	3	0.000	40.5% with zero frequencies
	Categoric	109.328	81.002	28.326	12	0.005	-

= Significant Model	= No Significant Model

The highest probability for the bus potential in this model is 'No Possibility'. However the 'Bus Distance' does have a positive effect on all of the other categories, which makes this variables the most promising of all variables over all the variable groups.

FIVE.TWO.2 | INTEGRAL MODELS

With all variables examined individually the first and logical step is to put all variables into a model and see if it works. Probably this will give some problems, however there is a change a working model will be created.

Also here the variable groups are initially checked separately. First, all continuous variables are put into the model, together with categorical variables which do not have a continuous background. After this, only categorical variables are used, and then the effect coded variables from these categorical variables are examined. The last integral models are models which only use the variables which were significant in the individual models.

Next to the multiple combinations of variables, several selection methods are used. The first selection method is forced entry, which just forces every variable into the model whether it will have a significant effect or not. The second method is the forward selection, the model starts empty and picks step by step significant variables until the new variables will not significantly improve the model anymore. The last selection method is backward elimination, in this case the model starts with all variables and discards the variables which influence the model the least until the model significantly decreases. Tables with all attempted integral models are stated in Appendix IV. Also here, model is seen as a 'Significant Model' when the significance is 0.050 or less, otherwise the model will be 'No Significant Model'. When no model could be estimated the table will state 'No Model'.

All the models give the same warning as explained in the previous paragraph about zero frequencies. Next to this warning, multiple of the models with personal or bus related variables give another warning about unexpected singularities in the Hessian matrix. This warning is stated in Figure 11.

Warnings

There are n (x) cells (i.e., dependent variable levels by subpopulations) with zero frequencies.
Unexpected singularities in the Hessian matrix are encountered. This indicates that either some predictor variables should be excluded or some categories should be merged.
The NOMREG procedure continues despite the above warning(s). Subsequent results shown are based on the last iteration. Validity of the model fit is uncertain.

FIGURE 11 | WARNING UNEXPECTED SINGULARITIES IN THE HESSION MATRIX

This warning is even more disturbing than the previous one, because now the model indicates that the validity of the model fit is uncertain. Therefore, the Chi-square cannot be trusted anymore and models cannot easily be compared.

Because of the warning the integral models are discarded and "step by step models" are examined, which are discussed in the next paragraph.

FIVE.TWO.3 | STEP BY STEP MODELS

The final attempt to make a working model during this research will be a step by step model. It starts with an empty model and manually variables will be inserted. The step by step models are made in the following manner:

A starting variable is chosen which is already proven to be significant to the model. A second variable is added; if the model still turns out to be significant and no warnings were detected a third variable is added. If the model was not significant anymore or did get warnings due to the second variable a new second variable is chosen. If all possible second variables are tried in the model a new starting variable is picked and the process starts again. New variables will be added until the model is not significant anymore or warnings appear, if this happens variables are removed and tried until the model is significant again.

This process is also illustrated in a chart flow diagram which can be found in Appendix V. All models which were significant and did not deliver any warnings are summarized in Appendix VI.

The step by step models which combine variable groups are made in a similar way as described above. The only difference is that the model does not start with one starting variable, but with a combination of variables from the models in Appendix VI. The significant models without a warning are stated in Table 24, the parameter estimates of these models can be found in Appendix VII.

TABLE 24 | STEP BY STEP MODELS COMBINED VARIABLE CATEGORIES

CATEGORY	Variable 1	Variable 2	Variable 3	Variable 4	-LL2 Start	-LL2 Final	Chi²	df	Sign.
Area and Personal	-								
Area and Trip	-								
Personal and Trip	Education	PT Subscription	Transport Mode	x	287.176	186.473	100.703	12	0.000
	Family Size	PT Subscription	Transport Mode	x	357.695	276.638	81.057	18	0.000
	Housing Type	PT Subscription	Transport Mode	x	318.403	236.941	81.462	15	0.000

= Significant Model

As can be seen in the table, the only models which could be made were combinations of personal characteristics with trip characteristics. All three models have the same significance, and all three models have eight significant parameters. However, the second model has significant parameters in all three bus potential categories besides 'No Possibility', therefore this model is chosen to be discussed as the final model. Furthermore, this model has the highest degrees of freedom, which implies that this model has the most aspects to influence in the final tool.

The parameter estimates of the final model are stated in Appendix VII as well as in Table 25.

TABLE 25 | PARAMETER ESTIMATES FINAL MODEL

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
Yes	Intercept	-1.004	.092	118.906	1	.000			
	FamilySize_EC1	.219	.171	1.648	1	.199	1.245	.891	1.741
	FamilySize_EC2	-.229	.111	4.264	1	.039	.795	.640	.988
	FamilySize_EC3	-.061	.170	.127	1	.721	.941	.674	1.314
	FamilySize_EC4	.035	.149	.054	1	.816	1.035	.773	1.387
	PTSubscription_EC1	.152	.079	3.690	1	.055	1.164	.997	1.360
	TransportMode_EC1	.024	.072	.111	1	.739	1.024	.889	1.181
Yes, but not Realistic	Intercept	-.126	.072	3.077	1	.079			
	FamilySize_EC1	.056	.129	.184	1	.668	1.057	.820	1.362
	FamilySize_EC2	-.144	.078	3.390	1	.066	.866	.742	1.009
	FamilySize_EC3	-.008	.119	.005	1	.946	.992	.785	1.253
	FamilySize_EC4	.231	.102	5.104	1	.024	1.260	1.031	1.541
	PTSubscription_EC1	-.004	.059	.004	1	.948	.996	.887	1.119
	TransportMode_EC1	.326	.054	36.133	1	.000	1.385	1.245	1.540
Don't Know	Intercept	-1.503	.120	157.182	1	.000			
	FamilySize_EC1	.367	.180	4.128	1	.042	1.443	1.013	2.055
	FamilySize_EC2	-.176	.123	2.039	1	.153	.839	.659	1.068
	FamilySize_EC3	.168	.172	.955	1	.329	1.183	.844	1.658
	FamilySize_EC4	.227	.152	2.245	1	.134	1.255	.932	1.690
	PTSubscription_EC1	-.238	.101	5.613	1	.018	.788	.647	.960
	TransportMode_EC1	.183	.082	5.031	1	.025	1.201	1.023	1.409

a. The reference category is: No Possibility.

As mentioned, the model has eight parameters which are significant, the rest of the parameter will be treated as best guess instead of setting to zero. All three bus potential categories will be discussed, starting with ‘Yes’.

The intercept of the ‘Yes’ category is negative, which means that this option is less likely to be chosen than ‘No Possibility’. However, this category is positively affected by the variables ‘Public Transport Subscription’ and ‘Transport Mode’. The third variable, ‘Family Size’, has both a positive and negative influence on the bus potential category.

The second bus potential category, ‘Yes, but not Realistic’, has the almost same polarity of effects form the variable as the previous category. The exception is ‘Public Transport Subscription’, which has in this case a negative influence. ‘Don’t Know’ is again similar to category two, except ‘Family Size’ category two.

For all variables the part-worth utility is calculated by the use of Equation 11. This states how much a category of a variable influences the model.

EQUATION 11 | FORMULA PART-WORTH UTILITY

$$U(Y) = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_{in}$$

U(Y) means the part-worth utility (U) of the category (Y). The X’s are one, zero or minus one, regarding the coding of the category. The coding can be found in the table of possible variables in Appendix II. For example the part-worth utility of a ‘Family Size’ of one person for the bus potential category ‘Yes’ is calculated, this calculation is stated in Equation 12.

EQUATION 12 | FORMULA PART-WORTH UTILITY

$$U(Yes) = 0.219 * 1 + -0.229 * 0 + -0.061 * 0 + 0.035 * 0 = 0.219$$

All part-worth utility values are illustrated in Figure 12.

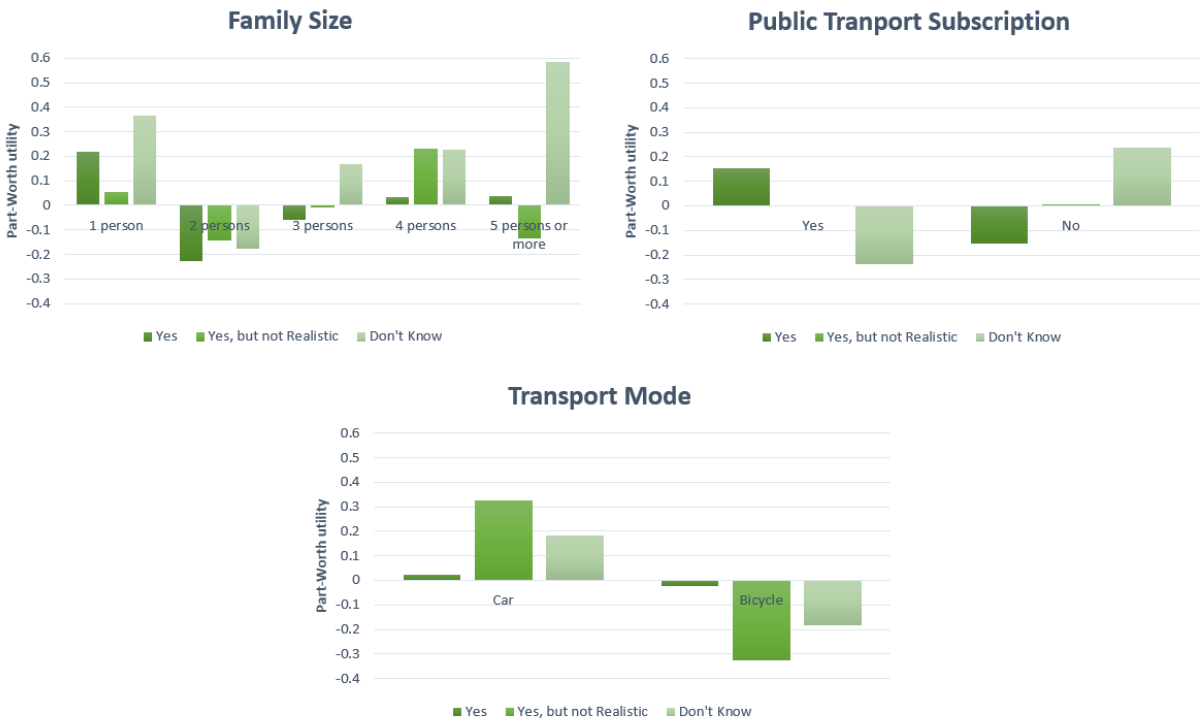


FIGURE 12 | PART-WORTH UTILITY

The part-worth utilities of 'Family Size' fluctuate quite a lot over the different variable categories, and also over the bus potential categories. It seems that families with 1 or 4 persons are more likely to be potential bus users than families with 2 persons. Families with 3 or 5 persons are more likely to pick the 'Don't Know' category over the other categories. 'PT Subscription' and 'Transport Mode' only have two categories, the part-worth utility of the first category is the negative opposite to the second category. The part-worth utilities from the 'Public Transport Subscription' are logically explainable, persons with a subscription are more likely to be potential than persons without a subscription, and persons without a subscription have not many knowledge about the bus services so are more likely to pick the 'Don't Know' category. The car category of 'Transport Mode' does have a positive utility over all the bus potential categories, in this case the conclusion can be made that car-drivers are more likely to be potential bus users than bike-riders.

FIVE.TWO.4 | ASSUMPTIONS

The previous two paragraphs tried to identify the variables which could be used in the potential bus users model because they demonstrated to have a significant influence on the dependent variable. But to use these variables in the model they have to meet the assumptions introduced in Table 6 from paragraph Three.Three.1.

The first assumption concerns the linearity, it assumes there is a linear relationship between any continuous predictors and the logit of the outcome variable (Field, 2013b). In essence, no continuous variables are used in the final model, so this assumption about linearity can be discarded. When continuous variables are used, a check should be done towards outliers. This could be done by the use of standardized residuals and scatterplots (The RMUoHP Biostatistics Resource Channel, 2013).

The second assumption regards the independence of errors. All residual terms of all observations should be independent, or uncorrelated. So, there cannot be any autocorrelation. This assumption is already violated, most of the respondents describe multiple trips. Which makes these trips dependent, because they were made by the same individual. Next to this, persons can be correlated because they live in the same area, and in the multinomial regression model no levels are included in the model. Nevertheless, this assumption of independence of errors can be checked by the use of the Durbin-Watson test, which tests for serial correlations between errors. The value of the Durbin-Watson can vary between zero and four, in which two will mean the residuals are uncorrelated. The overall opinion is that the score is good if the value lies between one and three (Field, 2013b).

The third and last assumptions is about multicollinearity, this can affect the parameters of a regression model. No variables may have a perfect collinearity, this occurs when one predictor is a perfect linear combination of the others and the correlation coefficient equals one. When perfect collinearity is found in the model the β's are untrustworthy, the R is limited and it becomes difficult to assess the influence of predictors. Multicollinearity causes this kind of problems in the model if correlation coefficients are above 0.8 or 0.9, the Variance Inflation Factor (VIF) is 10 or higher and the Tolerance (1/VIF) is between 0.1 and 0.2. Furthermore, the model maybe biased it the average VIF is substantially greater than 1.

Unfortunately, IBM SPSS Statistics has no option to check these assumptions within a logistic regression model, therefore a linear regression model is made to check the assumptions. The output of this model is stated in Appendix VIII.

The model summary from Appendix VIII.1 shows a value of 1.778 for the Durbin-Watson test, this value is between 1 and 3 and close to 2. Therefore, it seems there is an independence of errors, despite the earlier mentioned reasons why this assumption is probably violated. Appendix VIII.2 shows all correlations in the Coefficient Correlations table, which are all below 0.8. Therefor this does not indicate to a problem regarding multicollinearity. The table with Coefficients, in Appendix VIII.3, shows the Tolerance and VIF values. All tolerance values are between 0.9 and 1.0, so certainly not between 0.1 and 0.2. The VIF values are all close to 1, which means none is above 10, so also here no problem was detected. The last check is for the average VIF value, which is 1.033, this value is close to 1. It can be concluded that multicollinearity is no problem in this model.

FIVE.3 | CONCLUSIONS

The possible variables are initially analyzed by the use of correlation analyses, based on the Chi-square test. Almost all variables, over all categories, do correlate to each other. Therefore, it can be said that most variables cannot be combined in a model, because they correlate to each other. However quite a lot of variables, 19 out of 30, do significantly correlate to the dependent variable 'Bus Potential'. At first none of the variables regarding bus characteristics did correlate to the dependent variable, but after a new categorization also one of these variables seemed to have a significant effect on the 'Bus Potential'.

After the correlation tests, several multinomial regression models were estimated. First all individual variables were tested into a model. All continuous variables did get a warning regarding subpopulations with zero frequencies, this problem could be solved by the use of only categorical variables. All significant individual models were put together in integral models, however none of these models could be used because of even more warnings. The last attempt for a model was done by step by step models, these models were made by manually putting variables in the model. Only three models which combined multiple variable groups could be made, and none of these included areal variables. Therefore the aim of the research, to combine areal and personal characteristics into a model, was not achieved.

The final model has three variables: 'Family Size', 'Public Transport Subscription' and 'Transport Mode'. The first two variables are personal, the last variable regards trip characteristics. When analyzing the part-worth utility of all variable's categories, the model does not seem to be very stable. However the model does not violate any of the assumptions regarding multinomial regression models.

Due to lack of time and tools no hierarchical regression model was investigated.

Chapter Six | Conclusions

This final chapter of the thesis will discuss conclusions which can be made from the research, recommendations for future research and a discussion.

Six.1 | Conclusion

This paragraph is divided in two sections, the first section will give the research answers by the research questions. The second sub-paragraph will deal with remaining conclusions which can be made after the done research.

Six.One.1 | Research Answers

In this paragraph the research questions established in Paragraph One.3 are answered.

The first research question concerns the tool which would be developed during the research, the question is:

"Which research methodology should be used to develop a tool which determines and localizes potential bus users?''

As stated in paragraph One.4, a research methodology is the process used to collect information and data for the purpose of making business decisions. The methodology may include publication research, interviews, surveys and other research techniques, and could include both present and historical information. So, this question regards both to which model should be used to develop the tool, and the type of data which should be obtained to run the model.

To be able to develop a tool to estimate bus potential in an area, the actual situation should be examined, and the switching-behavior of travelers towards bur usage should be interpreted. This can be done by the use of a model which combines the personal, trip and areal characteristics to determine the presence and location of potential bus users. The data needed on the personal level can be acquired by interviews and surveys, the areal data can mostly be obtained by open data sources.

The type of model which should be used depends strongly on the definition of the dependent variable, in his case bus potential. During this research secondary data was used, and the bus potential was defined as a categorical variables, which concerned the ability of travelers to switch to bus trips (Three.3; Four.One.2). Because of the categorical variable a multinomial logistic regression model was used during this research (Three.Three.1). This model only estimates linear relationships between the dependent and independent variables, but with the use of the right categorization techniques also other relations can be estimated. For example the variable 'Distance to Bus Stop' was used with two different categorizations, from which only one had a significant effect on the bus potential (Five.1).

When the independent variables are estimated a GIS-based tool should be developed to determine and localize bus potential in other areas. When the location of potential is known Geo-Marketing can be implemented to turn these potential users into actual bus users (Two.Three.3).

The second research question regards the model which had to be estimated to develop the tool:

“Which personal, trip and areal characteristics should be used to model potential users of public transport?’’

This question is answered later on, because to respond to this question the following sub-answers are needed. The first sub-question needed to answer this main question is:

“What is the definition of potential bus users?”

The answer for this first sub-question can be found in Paragraph Two.Two.5. This states the possible definitions of potential in a theoretical or physical way, or depending on the travelers’’ switching-behavior. The latter is the most realistic way of defining potential, the definition of bus potential would be: “The potential users of public transport are the users who, at the present, do not use the public transport, but which could/would switch towards it when certain measures are taken.”

Paragraph Four.One.1 discusses the data available for this research. Secondary data was used, and the bus potential was already defined in two different ways.

In the first definition, which was used during this research, the potential is scaled, four different scales of ability to switch towards bus usage where established. An attempt was made to use only a Yes/No variant of the bus potential variables. However, the assumption was made that someone who relates to a switch from car to bus as possible, has more potential than someone who states it is possible but not very realistic. Therefore all four the bus potential categories were used in the research.

From the used dataset also a second definition of bus potential can be extracted. This definition was not used in this research, but could be examined in the future. This definition also uses the switching-behavior of a traveler, but combines this with measures which will affect this behavior.

The second sub-question regarding this research question regards the variables needed for a model, and is formulated as:

“Which independent variables should be formulated to describe the potential of a person?”

To answer this question, again a link to Chapter Four is made, because it discusses the available data. All possible variables mentioned here were formulated and coded to the final dataset stated in Appendix II. All variables which were proven to have a significant effect on the 'Bus Potential' in Paragraph Five.Two.1 could be referred to as variables which can describe the potential of a person.

However, the final model which was made during this research uses only three independent variables, regarding family size, public transport subscription and present transport mode (Five.Two.3). So, it can be concluded that these variables are definitely significant for the estimation of the dependent variable 'Bus Potential'.

The next sub-question regarded the behavior of travelers within a certain area, the question was:

"Can the potential bus travelers be explained by characteristics of an area?"

It was assumed that the potential of someone lies in his switching-behavior towards bus usage (Two.Two.5). In this research an attempt was made to describe the potential by linking behavior of travelers to the area, including characteristics of the area; the persons living there and the trips they make (Three.3). However, this attempt did fail because no model could be made by using both the characteristics of the area and the personal characteristics of the travelers. The personal characteristics could be combined with the trip characteristics, but because these variables both came from the same respondent in the same survey nothing can be concluded about the area they live in (Five.Two.3). The answer to this sub-question could still be positive, however this could not be proven in this research.

From this last sub-question back to the second research question:

"Which personal, trip and areal characteristics should be used to model potential users of public transport?''

During the correlation analyses 19 of the 30 possible independent variables seem to have a significant effect on the dependent variable 'Bus Potential' (Five.1). The final multinomial logistic regression model used three of these variables, regarding family size, public transport subscription and present transport mode (Five.Two.3). So, it can be concluded that these variables are significant for the estimation of the dependent variable 'Bus Potential'.

However, the aim was to combine personal, trip and areal characteristics into the model. This final model only uses two personal characteristics and a trip characteristic. Therefore, it is unfortunately not possible to answer this research question completely at the moment. Probably the dataset used in this research is not fitting for the model, at least not for a multinomial regression model. Therefore, a hierarchical regression model could be tried to make (Three.Three.2). Otherwise other data should be obtained, which is a very time-consuming and expensive alternative. Recommendations about the future of this research will be given in Paragraph Six.2.

The third and final research question deals with the social problem of air pollution:

"Can the Air Quality in the Netherlands be improved by stimulating the use of Public Transport?"

A large part of the emissions in the Netherlands is caused by road traffic (Two.One.2), and a change of transport modes can have a positive effect on the air-quality and the overall environment (Two.2). Public transport is often seen as an important solution to reduce air pollution, because of the lower emissions per passenger than an average car has (Two.Two.2). In Nuenen 61.2 of the non-bus trips, had potential to be changed to a bus trip (Four.One.2), this is a very big opportunity for the public transport sector if these kind of numbers can be found in whole of the Netherlands.

Therefore, the answer for this final research question is: yes, the air quality in the Netherlands can probably be improved by the stimulation of public transport use. However, to stimulate the use of public transport, knowledge is needed about the presence and location of potential public transport users, which can be a time and cost consuming task. This research attempted to develop a tool which will do precisely this, but the modelling of bus potential users has proven to be difficult.

SIX.ONE.2 | REMAINING CONCLUSIONS

Next to the sub-questions, options were mentions about the scale that should be used for the characteristics of the area. As encountered very early in the modelling process the four position ZIP code scale was too gross and could not lead to a significant model for the bus-potential, this was stated in Paragraph Four.2. A smaller scale should be used, in this research the six position ZIP code areas were used. However none of the bus related characteristics did initially have a significant effect on the model, as illustrated in Paragraph Five.1. This could be because very little areas have bus stops when using the six position ZIP code. For this reason perhaps medium scales for the areal variables could be examined, like neighborhoods.

Besides the scale of the areal characteristics, the usage of areal characteristics in the same level as personal behavior was doubted. To solve this a hierarchical model could be used, however due to time constraints and lack of tools such a model was not made during this research. The lack of tools refers to the software available via the university. Statistical models are usually done in IBM SPSS Statistical, but this is no ideal program for hierarchic models. HLM will be a more fitting program, but could not be figured out in the short amount of time available. However the hierarchical regression model could still be a good solution to create a fitting model for both areal and personal variables. The warnings about missing data will be solved this way because, like discussed in Paragraph Three.Three.2, hierarchical regression models can estimate parameters by the use of the available data.

Altogether, the research did show signals which were detected before by Waerden et al. In this research for 'Regionaal Openbaar Vervoer Bereikbaarheid Consumentgericht (ROVBECO)' ('Regional Public Transport Accessibility Consumer Oriented'), also an attempt was made to realize a planning tool for bus potential. The data used in this thesis was obtained by this particular research. However the model made by ROVBECO also contained only three significant variables (Waerden et al., 2009).

SIX.2 | RECOMMENDATIONS

Next to the conclusion drawn in the previous paragraph, recommendations will be made. Firstly pitfalls of this research will be given, so future research will not make the same mistakes.

SIX.TWO.1 | PITFALLS

The first pitfall of this research was the missing of tools to perform a hierarchical regression model. So, the first recommendation will be to try to make such a model and see if more variables could be implemented in a model, and if areal and personal variables could be combined in this way.

The second pitfall of this research was probably the dataset itself. It was from an enquiry which was done five years ago, the data was raw and many adjustments had to be made for the data to be useful in a model. After analyzing the data with the golden standard showed the data did not really fit the actual population of the area. Next to this the bus potential was already specified, it was classified by whether a respondents thinks their non-bus trip could be done by bus. However whether it is possible to take the bus, does not mean people will actually make this switch. Finally the survey was held in Nuenen, a small village which can be referred to as a car town. Because the village lies between the cities of Eindhoven and Helmond, many people live in Nuenen, but travel to the city for their work. Nuenen does not have a very extensive public transport service yet.

Therefore, if the first recommendation about the hierarchical regression model did not result in a better model, the advice will be to disregard this dataset and use other data to make a model. However it is very time consuming and expensive to make such an extensive dataset as used in this research. But if the intention is to make a new dataset by the use of an enquiry several recommendations can be given: pick a place with a great varied public transport offer, preform a digital enquiry where respondents can mark their travels on a map and reconsider the best way to describe bus potential.

SIX.TWO.2 | REMAINING RECOMMENDATIONS

Probably the primary recommendation when investigating bus potential in the future is the use of open data. As stated before, to obtain a dataset by the use of a survey is very expensive and time consuming, open data continues to extend. It is available, so why not use it? The largest open data set about travelers behavior is the before mentioned 'Onderzoek Verplaatsingen in Nederland (OViN)' (Travel Research in the Netherlands). Which is an annual research performed by the 'Centraal Bureau voor de Statistiek (CBS)' (Central Statistical Office from the Netherlands). Every year approximately 50,000 people participate with this research. However this results in an average of 127 respondent per municipality, which is a too small amount to do research on this scale. When research is done in a larger scale, for example a provincial scale, this dataset could probably be used very well. Next to this, the OViN has no separately data about bus usage, this is only combined with metro and tram. The OViN survey is comparable with the enquiry used in Nuenen. It deals with all kind of personal characteristics, and the respondents have to fill in several trips they made, so from this trip characteristics are taken. Also the OViN dataset does not deal with areal characteristics, so they should again be obtained from other open data sources.

Next to het OViN, new developments are made with open data in the field of traffic. November 2014 the first results of the 'Mobiliteitspanel Nederland (MPN)' ('Dutch Mobility Panel'), performed by 'Het kennisinstituut voor Mobiliteitsbeleid' ('Knowledge Institute for Mobility Policy'), the University of Twente and Goudappel Coffeng, were presented. This research does not only focus on present travel behavior, but aims to increase the understanding of factors which play a role in changing travel behavior. This is done by monitoring a fixed group of households over several years. With the knowledge gained by this survey, differences between groups of respondents can be examined, like young people, elderly people, families with and families without kids. The first try of the research got response from 3,572 households, including 6,126 persons (Kalter et al., 2014).

SIX.3 | DISCUSSION

This discussion in meant to reveal some points which could probably be improved when performing this research. Most of these points were already mentioned in the previous paragraphs of this chapter, so will only be cited shortly again here.

The research is about the creation of a bus potential model, however the definition of the 'bus Potential' was copied from previous researches. When going on with investigating this subject this definition should be carefully looked at. However, the theoretical and physical definitions of public transport discussed in Paragraph Two.Two.5, do not seem to be useful when geo-marketing techniques will be implemented in the future, because of the lack of personal and location bounded knowledge. Therefore, the third definition regarding switching-behavior of travelers should be used. In this research the ability of non-bus users to switch to bus trips was investigated. However, also the availability for this switch, or the sensibility of travelers towards several measures could be used to define bus potential. The dataset used during this research, from Nuenen, also has data about respondents' attitude towards measures which can be implemented to stimulate travelers to travel by bus. This sensitivity can also be used as the dependent variable in a bus potential model.

The dataset for this research was collected with an enquiry by previous research. However data will probably be more fitting for a particular research if collected during this research. Due to the time and expenses aspects of such a survey, open data should be considered even if this means a bigger scale of the areas should be used.

In the previous paragraph the MPN was already discussed, but when doing a new research all new developments in the transportation research sector should be considered.

One of the main goals of the model was to combine areal and personal characteristics. However, multiple researches did not succeed in this. Maybe the areal characteristics should be disregarded and only personal characteristics should be considered.

The research is carried out because of an important social problem, air pollution. Therefore, the target to reduce private-car usage by the stimulation of public transport seems important enough to continue the research to develop a bus potential tool. Maybe the air quality problem will in the future not be as stressing as nowadays, because all transport continuously becomes more sustainable, and maybe in the future green transportation is achieved by the use of alternative fuels or electric driving. However, the stimulation of public transport can have multiple other positive effect on society. For example solutions for congestions, parking problems and livability of residential areas.

Finally can be concluded that this research did explore one path to make a bus potential model, the multinomial regression model. However many paths are still out there, ready to be discovered.

REFERENCES

Abou-Zeid, M., Witter, R., Bierlaire, M., Kaufmann, V. and Ben-Akiva, M. (2012), "Happiness and travel mode switching: Findings from a Swiss public transportation experiment", *Transport Policy*, Elsevier, Vol. 19 No. 1, pp. 93–104.

American Marketing Association. (2013), "Definition of Marketing", *About AMA*, available at: https://www.ama.org/AboutAMA/Pages/Definition-of-Marketing.aspx (accessed 3 November 2014).

ANP. (2013), "Nederland wil veel minder CO2-uitstoot Europa", *NU.nl*, available at: http://www.nu.nl/economie/3579330/nederland-wil-veel-minder-co2-uitstoot-europa.html.

Atlas Leefomgeving. (2013), "Stikstof dioxide", *Lucht*, available at: https://www.atlasleefomgeving.nl/meer-weten/lucht/stikstofdioxide (accessed 6 September 2014).

Beirão, G. and Sarsfield Cabral, J. a. (2007), "Understanding attitudes towards public transport and private car: A qualitative study", *Transport Policy*, Vol. 14 No. 6, pp. 478–489.

Bellekom, S., Benders, R., Pelgröm, S. and Moll, H. (2012), "Electric cars and wind energy: Two problems, one solution? A study to combine wind energy and electric cars in 2020 in The Netherlands", *Energy*, Vol. 45 No. 1, pp. 859–866.

Bérénos, M. (2009), *Transumo-project " Regionale openbaar vervoerbereikbaarheid consumentgericht " (ROVBECO)*, Zoetermeer, Netherlands, pp. 1–19.

Boer, L.C. (Eelco) den, Brouwer, F.P.E. (Femke) and Essen, H.P. (Huib) van. (2008), *Studie naar TRansport Emissies van Alle Modaliteiten*, Delft, Netherlands, p. 180.

Borgers, A. and Timmermans, H. (1993), "Transport facilities and residential choice behavior: A model of multi-person choice processes", *The Journal of the Regional Science Association International*, Vol. 72 No. August 1991, pp. 45–61.

Bridgis. (n.d.). "Bridgis Geoservices", available at: http://www.bridgis.nl/.

Brög, W. (2000), *Switching to public transport*, *UITP: 2nd Asia Pacific Congress, Melbourne*, München, Germany, available at: http://www.kontiv.info/info/Switching to Public Transport.pdf (accessed 25 September 2014).

Brög, W., Erl, E. and Mense, N. (2002), *Individualised Marketing - Changing Travel Behaviour for a better Environment*, *Paper presented at the OECD Workshop: Environmentally Sustainable Transport*, Berlin, Germany, pp. 1–21.

Buck Consultants International and Goudappel Coffeng. (2009), *Effectstudie Milieuzones Vrachtverkeer - Stand van zaken 2009*, Nijmegen, Netherlands, available at: http://www.milieuzones.nl/sites/default/files/Effectstudie 2009.pdf.

BusinessDictionary.com. (n.d.). "Research Methodology", available at: http://www.businessdictionary.com/definition/research-methodology.html (accessed 6 February 2015).

Centraal Bureau voor de Statistiek. (2011), *Gemeente Op Maat, Nuenen, Gerwen en Nederwetten*.

Centraal Bureau voor de Statistiek. (2013), *Publicatie Onderzoek Verplaatsingen in Nederland 2012*, Den Haag, Netherlands.

Centraal Bureau voor de Statistiek. (2014a), *Publicatie Onderzoek Verplaatsingen in Nederland 2011 - Revisiebestand*, Den Haag, Netherlands.

Centraal Bureau voor de Statistiek. (2014b), *Publicatie Onderzoek Verplaatsingen in Nederland 2013*, Den Haag, Netherlands.

Centrum Milieuzones. (2014), "Locaties Milieuzones", *Locaties Milieuzones*, available at: http://www.milieuzones.nl/locaties-milieuzones (accessed 8 October 2015).

Centrum Vernieuwing Openbaar Vervoer. (2002), *Marketing in het openbaar vervoer - Rapport 16*, Rotterdam, Netherlands, available at: http://publicaties.minienm.nl/documenten/marketing-in-het-openbaar-vervoer.

Chen, M., Ho, T. and Jan, C. (2006), "A System Dynamics Model of Sustainable Urban Development : Assessing Air Purification Policies at Taipei City", Vol. 4 No. 1.

Cliquet, G. (2006), *Geomarketing; Methods and Strategies in Spatial Marketing*, ISTE Ltd, London, United Kingdom.

Compendium voor de Leefomgeving. (2013), "Nationale luchtkwaliteit: overzicht normen", *Luchtkwaliteit*, available at: http://www.compendiumvoordeleefomgeving.nl/indicatoren/nl0237-Nationale-luchtkwaliteit:-overzicht-normen.html?i=14-65 (accessed 5 September 2014).

Ettema, D., Friman, M., Gärling, T., Olsson, L.E. and Fujii, S. (2012), "How in-vehicle activities affect work commuters' satisfaction with public transport", *Journal of Transport Geography*, Elsevier Ltd, Vol. 24, pp. 215–222.

Europa decentraal. (2015), "Concessies openbaar vervoer", *Vervoer en aanbesteden*, available at: http://www.europadecentraal.nl/onderwerpen/vervoer/vervoer-en-aanbesteden/concessies-ov/ (accessed 9 February 2015).

European Environment Agency. (2010), *The European Environment State and Outlook 2010*.

European Environment Agency. (2011), "Transport", available at: http://www.eea.europa.eu/themes/transport/intro (accessed 5 May 2014).

European Environment Agency. (2013), *Air pollution fact sheet 2013 Netherlands*.

European Environment Agency. (2014), "Very high air pollution levels across Western europe", available at: http://www.eea.europa.eu/highlights/very-high-air-pollution-levels (accessed 5 May 2014).

Field, A. (2013a), "Logistic Regression", *Discovering Statistics Using IBM SPSS Statistics*, Sage Publications Ltd, London, United Kingdom, 4thed., pp. 760 – 813.

Field, A. (2013b), "Regression", *Discovering Statistics Using IBM SPSS Statistics*, Sage Publications Ltd, London, United Kingdom, 4thed., pp. 293 – 356.

Field, A. (2013c), "Multilevel Linear Models", *Discovering Statistics Using IBM SPSS Statistics*, Sage Publications Ltd, London, United Kingdom, 4thed., pp. 814 – 866.

Field, A. (2013d), "Categorical Data", *Discovering Statistics Using IBM SPSS Statistics*, Sage Publications Ltd, London, United Kingdom, 4thed., pp. 720–759.

Flowresulting. (2011), "Marketing van Ervaar het OV", available at: http://www.flowresulting.nl/uds-portfolio/ervaarhetov/ (accessed 21 September 2014).

Gemeente Nuenen, G. en N. (2009), "Structuurvisie Gemeente Nuenen, Gerwen en Nederwetten", *Structuurvisie*, available at: file:///C:/Users/Hannely/Dropbox/Afstuderen/literatuur/Nuenen/structuurevisie nuenen.html (accessed 19 January 2015).

GfK Geomarketing GmbH. (2013), *Geomarketing in practice*, Bruchsal, Germany, available at: www.gfk-regiograph.com/.../geomarketing_in_practice_compact.pdf.

Google. (2015), "Google Maps", *Nuenen*, available at: maps.google.nl (accessed 11 February 2015).

Goudappel Coffeng. (n.d.). "Goudappel Coffeng", available at: http://www.goudappel.nl/.

De Grange, L., Troncoso, R. and González, F. (2012), "An empirical evaluation of the impact of three urban transportation policies on transit use", *Transport Policy*, Vol. 22, pp. 11–19.

Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., Denier van der Gon, H., Frost, G.J., Heil, A., et al. (2011), "Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period", *Climatic Change*, Vol. 109 No. 1-2, pp. 163–190.

GroengasMobiel. (2014), "Groengas", *Feiten en Cijfers*, available at: http://www.groengasmobiel.nl/alles-over-groen-gas/feiten-en-cijfers/ (accessed 15 November 2014).

Health Effects Institute. (2010), *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*, Boston, Massachusetts, USA, available at: pubs.healtheffects.org/getfile.php?u=553.

Het Nieuwe Rijden. (2012), "Rijstijltips", *Tips voor het nieuwe rijden voor personenauto's*, available at: http://www.hetnieuwerijden.nl/wat-kunt-u-doen/rijstijltips/personenautos/ (accessed 6 February 2015).

Holmgren, J. (2007), "Meta-analysis of public transport demand", *Transportation Research Part A: Policy and Practice*, Vol. 41, pp. 1021–1035.

Hossain, A.K. and Davies, P. a. (2013), "Pyrolysis liquids and gases as alternative fuels in internal combustion engines – A review", *Renewable and Sustainable Energy Reviews*, Elsevier, Vol. 21, pp. 165–189.

HVR group. (2010), "Goed bezig BUS", *Provincie Noord-Brabant*, available at: http://www.hvrgroup.nl/werk/marketing-ov-brabant/ (accessed 23 September 2014).

Irias. (n.d.). "Irias innovatief informatiemanagement", available at: https://www.irias.nl/.

Kalter, M.O., Hoogendoorn, S. and Geurs, K. (2014), "De eerste resultaten van het Mobiliteitspanel Nederland", *Colloquium Vervoersplanologisch Speurwerk*, Goudappel, Eindhoven, Netherlands, available at: http://www.goudappel.nl/media/files/uploads/cvs2014_Eerste_resultaten_Mobiliteits_Panel.pdf.

Kenniscentrum InfoMil. (2014), "Besluit gevoelige bestemmingen", *Luchtkwaliteit*, available at: http://www.infomil.nl/onderwerpen/klimaat-lucht/luchtkwaliteit/wettelijk-kader/besluit-gevoelige (accessed 6 November 2014).

Kennisplatform CROW. (2013), “100ste gunning ov-concessie onder Wp2000 een feit”, *Kennispagina*, available at: http://www.crow.nl/vakgebieden/verkeer-en-vervoer/bibliotheek/kennisdocumenten/100ste_gunning_ov-concessie_onder_wp2000_een_feit?Zoekterm=100+aanbestedingen&page=1&searchsort=score&pagesize=10&parenturl=/Vakgebieden/Verkeer-en-Vervoer/Bibliotheekf (accessed 9 February 2015).

Kennisplatform Verkeer en Vervoer. (2009), *Lucht voor schone bussen*, Utrecht.

Kennisplatform Verkeer en Vervoer. (2014), *Duurzaam mobiliteitsbeleid voor gemeenten*, Ede.

Knol, A. (2005), *Trends in the environmental burden of disease in the Netherlands*, Bilthoven.

KpVV CROW. (2014), “Prestaties van vervoerwijzen”, *Dashboard duurzame en slimmem mobiliteit*, available at: file:///C:/Users/Hannely/Dropbox/Afstuderen/literatuur/Prestaties van vervoerwijzen.html (accessed 7 February 2015).

KpVV Dashboard. (2014), “Autoverkeer veroorzaakt groot deel van emissie”, *Schone Lucht*, available at: http://kpvvdashboard-5.blogspot.nl/2011/07/autoverkeer-veroorzaakt-groot-deel-van.html (accessed 26 September 2014).

Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., et al. (2000), “Public-health impact of outdoor and traffic-related air pollution: a European assessment.”, *Lancet*, Vol. 356 No. 9232, pp. 795–801.

Kutz, M. (2008), *Environmentally Concious Transportation*, John Wiley & Sons, London, United Kingdom.

Luce, R.D. (1959), *Individual Choice Behavior*, John Wiley & Sons, New York, USA.

Martens, K. (2007), “Promoting bike-and-ride: The Dutch experience”, *Transportation Research Part A: Policy and Practice*, Vol. 41 No. 4, pp. 326–338.

Milieu Centraal. (n.d.). “Luchtvervuiling”, *Klimaat en milieuproblemen*, available at: http://www.milieucentraal.nl/themas/klimaat-en-milieuproblemen/gezondheid-en-milieu/luchtvervuiling (accessed 17 November 2014).

Miller, J.G.T. (2004), *Living in the environment: principles, connections, and solutions*, Brooks/Cole.

Ministerie van Infrastuctuur en Milieu. (2011), *Kabinetsaanpak Klimaatbeleid op weg naar 2020*, Den Haag, Netherlands, pp. 1–15.

Ministerie van Infrastuctuur en Milieu. (n.d.). “Duurzaam doen”, *reizen*, available at: http://www.duurzaamdoen.nl/reizen/ (accessed 6 February 2015).

MOA. (2014), “Gouden Standaard”, *Methoden en Technieken*.

Movares Nederland B.V. (2012), *Uitvoeringsprogramma Verkeersstructuurplan*, Eindhoven, Netherlands, pp. 0–34.

Nijkamp, P. and Pepping, G. (1998), “Meta-Analysis for Explaining the Variance in Public Transport Demand Elasticities in Europe”, *Journal of Transportation and Statistics*, Vol. 1, pp. 1–14.

Ongkittikul, S. and Geerlings, H. (2006), “Opportunities for innovation in public transport: Effects of regulatory reforms on innovative capabilities”, *Transport Policy*, Vol. 13 No. 4, pp. 283–293.

OV-bureau Groningen Drenthe. (2014), “Soorten openbaar vervoer”, *Concessies*, available at: http://www.ovbureau.nl/ov-bureau/concessies/soorten-openbaar/ (accessed 25 October 2014).

Pookulangara, S., Hawley, J. and Xiao, G. (2011), “Explaining consumers’ channel-switching behavior using the theory of planned behavior”, *Journal of Retailing and Consumer Services*, Elsevier, Vol. 18 No. 4, pp. 311–321.

Qbuzz. (2014), “Waarmee-reis-ik”, *Onderweg*, available at: http://qbuzz.nl/GD/onderweg/waarmee-reis-ik/ (accessed 25 November 2014).

Qbuzz. (n.d.). “Eurokaartje”, *Mijn reis betalen*, available at: http://qbuzz.nl/GD/mijn-reis-betalen/kaartje-de-bus/eurokaartje/ (accessed 25 November 2014).

Redman, L., Friman, M., Gärling, T. and Hartig, T. (2013), “Quality attributes of public transport that attract car users: A research review”, *Transport Policy*, Vol. 25, pp. 119–127.

Rijksinstituut voor Volksgezondheid en Milieu. (2012), *Grootschalig concentratie en depositie kaarten Nederland*, *Rapportage 2012*.

Rijksinstituut voor Volksgezondheid en Milieu. (n.d.). “Smog”, *Smog*, available at: http://www.rivm.nl/Onderwerpen/S/Smog.

Rijksoverheid. (2014a), “Luchtvervuiling”, *Luchtkwalitiet*, available at: http://www.rijksoverheid.nl/onderwerpen/luchtkwaliteit/luchtvervuiling (accessed 7 September 2014).

Rijksoverheid. (2014b), “Normen luchtkwaliteit”, *Luchtkwaliteit*, available at: http://www.rijksoverheid.nl/onderwerpen/luchtkwaliteit/normen-luchtkwaliteit (accessed 7 September 2014).

Rijksoverheid. (2014c), “Maatregelen tegen luchtvervuiling”, *Luchtkwaliteit*, available at: http://www.rijksoverheid.nl/onderwerpen/luchtkwaliteit/maatregelen-tegen-luchtvervuiling (accessed 7 September 2014).

Rijkswaterstaat. (2014), “Nationaal Samenwerkingsprogramma Luchtkwalitiet (NSL)”, *Luchtkwaliteit*, available at: http://www.infomil.nl/onderwerpen/klimaat-lucht/luchtkwaliteit/nsl/ (accessed 7 September 2014).

Rojas-Rueda, D., de Nazelle, A., Teixidó, O. and Nieuwenhuijsen, M.J. (2012), “Replacing car trips by increasing bike and public transport in the greater Barcelona metropolitan area: a health impact assessment study.”, *Environment international*, Elsevier Ltd, Vol. 49, pp. 100–9.

Sierra-Vargas, M.P. and Teran, L.M. (2012), “Air pollution: impact and prevention.”, *Respirology (Carlton, Vic.)*, Vol. 17 No. 7, pp. 1031–8.

Steg, L. (2005), “Car use: lust and must. Instrumental, symbolic and affective motives for car use”, *Transportation Research Part A: Policy and Practice*, Vol. 39 No. 2-3, pp. 147–162.

Stichting Groen Gas Nederland. (2014), “Duurzame Mobiliteit”, *Overijssel verdubbelt groengastankstations*, available at: http://groengas.nl/tag/duurzame-mobiliteit/ (accessed 30 October 2014).

Sugiki, N., Furusawa, K. and Aoshima, N. (2001), “Location Behavioral Model Considering Transit Characteristics in Life Stage of Households”, *9th World Conference of Transportation Research, Seoul, Korea*, Gunma University, Gunma, Japan.

Sumaedi, S., Bakti, I.G.M.Y., Astrini, N.J., Rakhmawati, T., Widianti, T. and Yarmen, M. (2014), *Public Transport Passengers ’ Behavioural Intentions Paratransit in Jabodetabek – Indonesia*.

Syntus. (2014), “Twents - Service van Syntus”, *Wie zijn we*, available at: http://www.syntus.nl/over-syntus/wie-zijn-we.aspx (accessed 12 May 2014).

Taniguchi, A. and Fujii, S. (2006), “Promoting Public Transport Using Marketing Techniques in Mobility Management and Verifying their Quantitative Effects”, *Transportation*, Vol. 34 No. 1, pp. 37–49.

The RMUoHP Biostatistics Resource Channel. (2013), “How to Use SPSS:Logistic Regression”, available at: https://www.youtube.com/watch?v=yorCGUmLh8Y (accessed 1 January 2014).

Train, K.E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, available at: http://elsa.berkeley.edu/~train/.

Verbeek, R., Smokers, R., Kadijk, G., Bolech, M., Driever, H. and Research Group Sustainable Transport and Logistics. (2012), *Schone en duurzame bussen in het openbaar vervoer : wat is de route ?*, available at: http://repository.tudelft.nl/view/tno/uuid:fa28f128-4614-4fa9-8b3f-129fdb327d54/.

Vincent, W. and Jerram, L. (2006), “The Potential for Bus Rapid Transit to Reduce Transportation-Related CO~ 2 Emissions”, *Journal of Public Transportation*, Vol. 9 No. 3, pp. 219–237.

Vuchic, V.R. (2005), *UrbanTransit Operations, Planning and Economics*, John Wiley & Sons, New Jersey, USA.

Waerden, P. van der and Bérénos, M. (2010), *Houd Nuenen Bereikbaar; Verplaatsingsonderzoek Nuenen 2009*, Eindhoven, Netherlands.

Waerden, P. Van Der, Bérénos, M. and Timmermans, H. (2010), “Travelers’ Sensitivity to Public Transport Related Planning Measures”, *12th World Conference on Transport Research*, Lisbon, Portugal.

Waerden, P. Van Der, Bérénos, M. and Tromp, H. (2009), “Regionaal Openbaar Vervoer Bereikbaarheid Consumentgericht”, *Colloquium Vervoersplanologisch Speurwerk*, Antwerpen, Belgium.

Waerden, P. Van Der, Silva, A.N.R. da, Timmermans, H., Bérénos, M. and Rocha, G.R. (2008), “Public Transport Planning Evaluation Tools and their Data Requirements and Availability”, *9th International Conference on Design & Decision Suport Systems*, Leende, Netherlands, available at: http://cumincad.architexturez.net/system/files/pdf/ddss2008-45.content.pdf (accessed 23 September 2014).

Waerden, P. Van Der, Timmermans, H. and Bérénos, M. (2005), “In search of the Public Transport Users : Towards Public Transportation Potential Maps”, *Tranportation Research Board*, Washington D.C., USA.

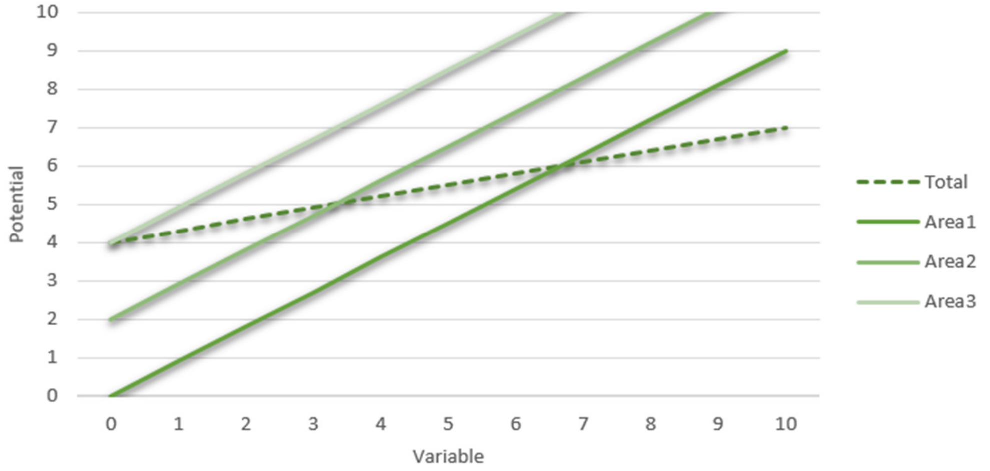
Waerden, P. van der, Timmermans, H., Bérénos, M. and Vermeulen, G. (2007), “Potential effects of public transport planning measures”, *ETC Conference, PTRC London*, London, United Kingdom.

Wang, T. and Chen, C. (2012), “Attitudes, mode switching behavior, and the built environment: A longitudinal study in the Puget Sound Region”, *Transportation Research Part A: Policy and Practice*, Elsevier Ltd, Vol. 46 No. 10, pp. 1594–1607.

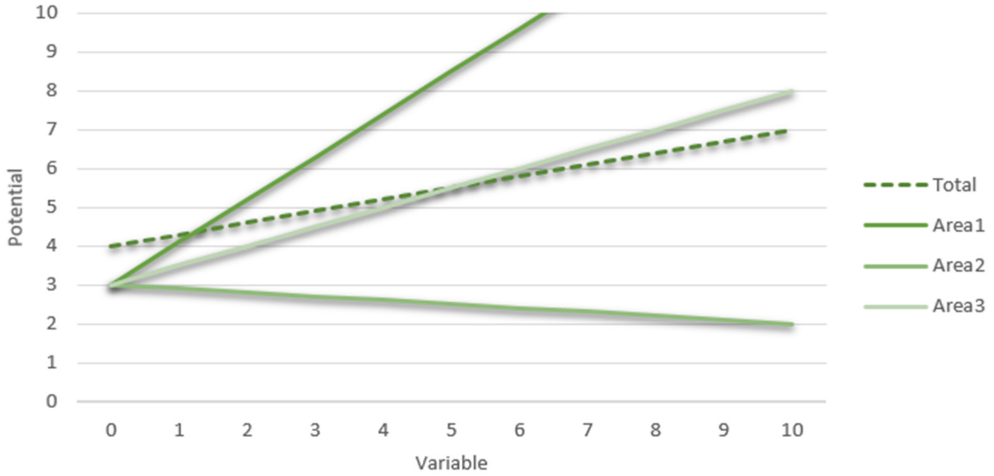
Zhou, Y., Viswanathan, K., Popuri, Y. and Proussaloglou, K.E. (2004), “Transit Customers – Who , Why , Where , and How : A Market Analysis of the San Mateo County Transit District”, *83rd Annual Meeting of the Transportation Research Board and Publication in the Transportation Research Record.*, Chicago, USA, available at: http://www.camsys.com/pubs/SAMTRANS_Market_Analysis.pdf.

APPENDIX I | FIXED AND RANDOM COEFFICIENTS SCENARIOS[3]

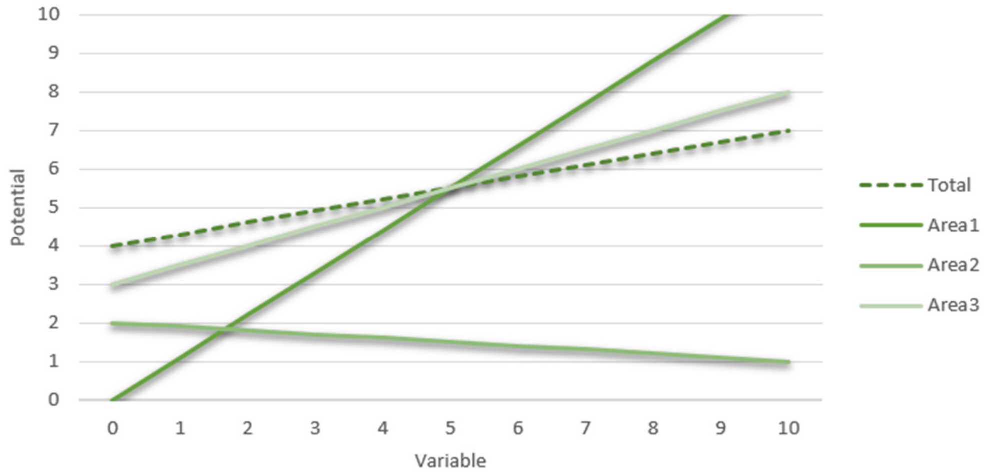
APPENDIX I. 1 | RANDOM INTERCEPTS, FIXED SLOPE



APPENDIX I. 2 | FIXED INTERCEPT, RANDOM SLOPES



APPENDIX I. 3 | RANDOM INTERCEPTS AND SLOPES



[3] The scenarios are completely imaginary, so they are not linked to actual data or reality.

APPENDIX II | TABLE WITH ALL POSSIBLE VARIABLES

APPENDIX II. 1| DEPENDENT VARIABLE

DEPENDENT VARIABLE	Description	Categories	Number	Effect Coding
Bus Potential	Whether the bus is a good replacement for the trip	Yes	1	1 0 0
		Yes, But not Realistic	2	0 1 0
		No Possibility	3	0 0 1
		Don't Know	4	-1 -1 -1

APPENDIX II. 2 | PERSONAL VARIABLES

PERSONAL VARIABLES	Description	Categories	Number	Effect Coding
Gender	Gender of the Individual	Man	1	1
		Woman	2	-1
Age	Age of the Respondent	Continuous		
	Categorical Age of yhe Respondent	0-24 years	1	1 0 0
		25-44 years	2	0 1 0
		45-64 years	3	0 0 1
		65+ years	4	-1 -1 -1
Education	Level of Education	Primary	1	1 0 0
		Secondary	2	0 1 0
		High	3	0 0 1
		Other	4	-1 -1 -1
Family Size	Number of People in Family	Continuous		
	Categorical Number of People in Family	1 person	1	1 0 0 0
		2 persons	2	0 1 0 0
		3 persons	3	0 0 1 0
		4 persons	4	0 0 0 1
		5+ persons	5	-1 -1 -1 -1
Family Structure	Composition of the Family	Single	1	1 0 0
		Couple	2	0 1 0
		Family with Children	3	0 0 1
		Other	4	-1 -1 -1
Housing Type	Type of Housing	Teracced House	1	1 0 0
		Semi-Detached or Corner House	2	0 1 0
		Detached House	3	0 0 1
		Other Type	4	-1 -1 -1
Estate	Type of House Ownership	Personal Estate	1	1
		Rental House	2	-1
Driver License	Obtained License	Yes	1	
		No	2	
Owned Cars	Number of Cars Owned by Family	1 or less cars	1	1 0
		2 cars	2	0 1
		3 or more cars	3	-1 -1
PT Usage	Frequency of Public Transport Usage	Never	1	1 0 0
		Rarely	2	0 1 0
		Regularly	3	0 0 1
		Often	4	-1 -1 -1
PT Subscription	Type of Public Transport Subscription	Discount or Total Subscription	1	1
		No Subscription	2	-1

APPENDIX II. 3 | TRIP VARIABLES

TRIP VARIABLES	Description	Categories	Number	Effect Coding
Motive	Motive for the Trip	Single Purpose*	1	1 0 0 0
		Work and Shopping	2	0 1 0 0
		Work and Leisure	3	0 0 1 0
		Shopping and Leisure	4	0 0 0 1
		Work, Shopping and Leisure	5	-1 -1 -1 -1
Transport Mode	Transport Mode of the Trip	Car	1	1
		Bicycle	2	-1
Departure Time	Time of Departure	Between 24:00 and 7:00 uur	1	1 0 0 0 0
		Between 7:00 and 9:00 uur	2	0 1 0 0 0
		Between 9:00 and 12:00 uur	3	0 0 1 0 0
		Between 12:00 and 16:00 uur	4	0 0 0 1 0
		Between 16:00 and 18:00 uur	5	0 0 0 0 1
		Between 18:00 and 24:00 uur	6	-1 -1 -1 -1 -1
Trip Frequency	Frequency of the Trip	Less than once a week	1	1 0 0 0
		1-3 times per week	2	0 1 0 0
		4-5 times per week	3	0 0 1 0
		6-7 times per week	4	0 0 0 1
		More than 7 times per week	5	-1 -1 -1 -1
Trip Distance	Distance of the Trip	Less than 6 kilometers	1	1 0 0
		6-10 kilometers	2	0 1 0
		11-30 kilometers	3	0 0 1
		More than 30 kilometers	4	-1 -1 -1
Trip Duration	Time Duration of the Trip	Less them 6 minutes	1	1 0 0 0
		6-10 minutes	2	0 1 0 0
		11-20 minutes	3	0 0 1 0
		21-45 minutes	4	0 0 0 1
		More than 45 minutes	5	-1 -1 -1 -1

* Work, Shopping or Leisure

APPENDIX II. 4 | BUS VARIABLES

BUS VARIABLES	Description	Categories	Number	Effect Coding
Bus Stops	Number of Bus Stops Available in the Area	Continuous		
	Categorical Number of Stops Available in the Area	1 stop	1	1
		2 or more stops	2	-1
	Bus Stops Available in the Area	Yes	1	1
		No	2	-1
Bus Lines	Number of Bus Lines Available at Stops in the Area			
	Categorical Number of Lines at Stops in the Area	1 - 5 lines	1	1
		6 or more lines	2	-1
	Lines Available at Stops in the Area	Yes	1	1
		No	2	-1
Bus Lines Area	Number of Bus Lines Available in the Area			
	Categorical Number of Lines Available in the Area	1 - 2 lines	1	1 0 0 0
		3 - 4 lines	2	0 1 0 0
		5 - 6 lines	3	0 0 1 0
		7 - 8 lines	4	0 0 0 1
		9 lines or more	5	-1 -1 -1 -1
	Lines Available in the Area	Yes	1	1
		No	2	-1
Bus Distance	Distance from Centre of the Area to Nearest Bus Stop	Less than 50 meters	1	1 0 0 0
		50 - 99 meters	2	0 1 0 0
		100 - 249 meters	3	0 0 1 0
		250 - 499 meters	4	0 0 0 1
		500 or more meters	5	-1 -1 -1 -1

APPENDIX II. 5 | AREAL VARIABLES

AREAL VARIABLES	Description	Categories	Number	Effect Coding
Surface	Surface of the Area	Continuous		
	Categorical Surface of the Area	Less than 6000 square meter	1	1 0 0 0
		6000 - 7999 square meter	2	0 1 0 0
		8000 -14999 square meter	3	0 0 1 0
		15000 - 19999 square meter	4	0 0 0 1
		20000 or more square meter	5	-1 -1 -1 -1
Inhabitants	Number of People in the Area	Continuous		
	Categorical Number of People in the Area	Less than 30 persons	1	1 0 0
		30-49 persons	2	0 1 0
		50-69 persons	3	0 0 1
		70 or more persons	4	-1 -1 -1
Education	Dominant Level of Education in the Area	Primary	1	1 0
		Secondary	2	0 1
		High	3	-1 -1
Households	Number of Households in the Area	Continuous		
	Categorical Number of Households in the Area	Less than 15 households	1	1 0 0
		15-19 households	2	0 1 0
		20-24 households	3	0 0 1
		25 or more households	4	-1 -1 -1
Family Size	Average Number of People per Household in the Area	Continuous		
	Categorical Average Household Size in the Area	1 person	1	1 0
		2 persons	2	0 1
		3 persons or more	3	-1 -1
Family Structure	Composition of the Family	Single	1	1 0 0
		Couple	2	0 1 0
		Family with Children	3	0 0 1
		Other	4	-1 -1 -1
Housing Type	Dominant Composition of a Family in the Area	Terraced house	1	1 0 0
		Semi-detached or corner house	2	0 1 0
		Detached house	3	0 0 1
		Other type	4	-1 -1 -1
Year of Construction	Average Year of Construction in the Area	Before 1970	1	1 0 0
		Between 1970 and 1979	2	0 1 0
		Between 1980 and 1989	3	0 0 1
		After 1990	4	-1 -1 -1
Property Value	Average Property Value in the Area	Less than 200.000 euros	1	1 0 0 0
		Between 200.000 and 250.000 euros	2	0 1 0 0
		Between 250.000 and 300.000 euros	3	0 0 1 0
		Between 300.000 and 400.000 euros	4	0 0 0 1
		More than 400.000 euros	5	-1 -1 -1 -1
Estate	Dominant Type of House Ownership in the Area	Personal Estate	1	1
		Rental House	2	-1

APPENDIX III | PARAMETER ESTIMATES, INDIVIDUAL MULTINOMIAL REGRESSION MODELS

APPENDIX III.1 | PERSONAL MULTINOMIAL REGRESSION MODELS

APPENDIX III.ONE. 1 | AGE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.303	.151	74.701	1	.000			
	[Personal_Age=1]	.674	.463	2.119	1	.146	1.962	.792	4.862
	[Personal_Age=2]	.356	.209	2.903	1	.088	1.428	.948	2.152
	[Personal_Age=3]	.087	.176	.247	1	.619	1.091	.773	1.540
	[Personal_Age=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.401	.110	13.264	1	.000			
	[Personal_Age=1]	.688	.359	3.679	1	.055	1.990	.985	4.021
	[Personal_Age=2]	.833	.148	31.843	1	.000	2.301	1.723	3.073
	[Personal_Age=3]	.329	.126	6.781	1	.009	1.390	1.085	1.780
	[Personal_Age=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.590	.169	88.221	1	.000			
	[Personal_Age=1]	.268	.588	.209	1	.648	1.308	.413	4.138
	[Personal_Age=2]	.674	.222	9.221	1	.002	1.962	1.270	3.031
	[Personal_Age=3]	.262	.194	1.826	1	.177	1.299	.889	1.900
	[Personal_Age=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.ONE. 2 | EDUCATION

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.153	.082	199.938	1	.000			
	[Personal_Education=1]	-.064	.246	.067	1	.796	.938	.579	1.520
	[Personal_Education=2]	-.034	.163	.044	1	.834	.966	.701	1.331
	[Personal_Education=3]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.051	.057	.786	1	.375			
	[Personal_Education=1]	-.249	.180	1.927	1	.165	.779	.548	1.108
	[Personal_Education=2]	.223	.109	4.164	1	.041	1.250	1.009	1.548
	[Personal_Education=3]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.605	.098	270.154	1	.000			
	[Personal_Education=1]	.994	.211	22.169	1	.000	2.703	1.787	4.089
	[Personal_Education=2]	.650	.163	15.995	1	.000	1.916	1.393	2.634
	[Personal_Education=3]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.ONE. 3 | FAMILY SIZE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.061	.223	22.573	1	.000			
	[Personal_FamilySize=1]	.221	.296	.559	1	.455	1.247	.699	2.228
	[Personal_FamilySize=2]	-.226	.243	.861	1	.353	.798	.495	1.285
	[Personal_FamilySize=3]	-.093	.296	.100	1	.752	.911	.510	1.626
	[Personal_FamilySize=4]	-.018	.275	.004	1	.947	.982	.573	1.683
	[Personal_FamilySize=5]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.167	.167	.998	1	.318			
	[Personal_FamilySize=1]	.264	.223	1.408	1	.235	1.303	.842	2.016
	[Personal_FamilySize=2]	.036	.180	.040	1	.842	1.036	.729	1.474
	[Personal_FamilySize=3]	.202	.214	.897	1	.343	1.224	.805	1.861
	[Personal_FamilySize=4]	.445	.199	5.016	1	.025	1.560	1.057	2.303
	[Personal_FamilySize=5]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.872	.310	36.438	1	.000			
	[Personal_FamilySize=1]	.950	.369	6.629	1	.010	2.585	1.255	5.327
	[Personal_FamilySize=2]	.384	.327	1.375	1	.241	1.468	.773	2.787
	[Personal_FamilySize=3]	.800	.363	4.866	1	.027	2.225	1.093	4.529
	[Personal_FamilySize=4]	.884	.347	6.506	1	.011	2.422	1.227	4.778
	[Personal_FamilySize=5]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.ONE. 4 | FAMILY STRUCTURE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-.606	.508	1.426	1	.232			
	[Personal_FamilyStructure=1]	-.234	.543	.185	1	.667	.792	.273	2.296
	[Personal_FamilyStructure=2]	-.674	.517	1.704	1	.192	.509	.185	1.402
	[Personal_FamilyStructure=3]	-.523	.519	1.013	1	.314	.593	.214	1.641
	[Personal_FamilyStructure=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	.375	.392	.915	1	.339			
	[Personal_FamilyStructure=1]	-.277	.418	.439	1	.507	.758	.334	1.721
	[Personal_FamilyStructure=2]	-.535	.397	1.813	1	.178	.586	.269	1.276
	[Personal_FamilyStructure=3]	-.240	.399	.363	1	.547	.786	.360	1.718
	[Personal_FamilyStructure=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-.318	.465	.470	1	.493			
	[Personal_FamilyStructure=1]	-.604	.506	1.424	1	.233	.547	.203	1.474
	[Personal_FamilyStructure=2]	-1.173	.476	6.058	1	.014	.310	.122	.788
	[Personal_FamilyStructure=3]	-.888	.478	3.441	1	.064	.412	.161	1.051
	[Personal_FamilyStructure=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.ONE. 5 | HOUSING TYPE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.204	.219	30.106	1	.000			
	[Personal_HousingType=1]	.070	.268	.069	1	.793	1.073	.635	1.814
	[Personal_HousingType=2]	.061	.242	.064	1	.800	1.063	.662	1.709
	[Personal_HousingType=3]	-.017	.255	.004	1	.948	.983	.597	1.621
	[Personal_HousingType=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.474	.170	7.771	1	.005			
	[Personal_HousingType=1]	.439	.202	4.747	1	.029	1.552	1.045	2.304
	[Personal_HousingType=2]	.521	.184	8.017	1	.005	1.684	1.174	2.417
	[Personal_HousingType=3]	.505	.191	6.974	1	.008	1.656	1.139	2.409
	[Personal_HousingType=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.034	.206	25.243	1	.000			
	[Personal_HousingType=1]	-.014	.254	.003	1	.956	.986	.599	1.622
	[Personal_HousingType=2]	-.330	.234	1.984	1	.159	.719	.455	1.138
	[Personal_HousingType=3]	-.453	.251	3.249	1	.071	.636	.389	1.040
	[Personal_HousingType=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.ONE. 6 | PUBLIC TRANSPORT USAGE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.179	.404	8.499	1	.004			
	[Personal_PTUsage=1]	-1.167	.482	5.869	1	.015	.311	.121	.800
	[Personal_PTUsage=2]	.098	.413	.057	1	.812	1.103	.491	2.479
	[Personal_PTUsage=3]	.340	.426	.639	1	.424	1.405	.610	3.238
	[Personal_PTUsage=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.039	.280	.020	1	.889			
	[Personal_PTUsage=1]	-.102	.302	.114	1	.736	.903	.499	1.633
	[Personal_PTUsage=2]	.086	.286	.090	1	.764	1.090	.622	1.910
	[Personal_PTUsage=3]	-.028	.299	.009	1	.926	.973	.541	1.749
	[Personal_PTUsage=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-2.565	.734	12.218	1	.000			
	[Personal_PTUsage=1]	1.524	.749	4.140	1	.042	4.593	1.058	19.946
	[Personal_PTUsage=2]	1.346	.739	3.313	1	.069	3.840	.902	16.354
	[Personal_PTUsage=3]	.810	.758	1.142	1	.285	2.249	.509	9.941
	[Personal_PTUsage=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.ONE. 7 | PUBLIC TRANSPORT SUBSCRIPTION

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
Yes	Intercept	-1.226	.079	242.754	1	.000			
	[Personal_PTSubscription=1]	.256	.153	2.795	1	.095	1.292	.957	1.744
	[Personal_PTSubscription=2]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	.025	.053	.225	1	.635			
	[Personal_PTSubscription=1]	-.162	.114	2.024	1	.155	.851	.681	1.063
	[Personal_PTSubscription=2]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.184	.077	233.803	1	.000			
	[Personal_PTSubscription=1]	-.585	.196	8.880	1	.003	.557	.379	.819
	[Personal_PTSubscription=2]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.2 | TRIP MULTINOMIAL REGRESSION MODELS

APPENDIX III.TWO. 1 | MOTIVE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
Yes	Intercept	-1.097	.088	154.167	1	.000			
	[Trip_Motive=1]	.249	.295	.713	1	.398	1.283	.719	2.289
	[Trip_Motive=2]	.677	.283	5.737	1	.017	1.968	1.131	3.423
	[Trip_Motive=3]	.378	.293	1.656	1	.198	1.459	.821	2.592
	[Trip_Motive=4]	-.543	.168	10.400	1	.001	.581	.417	.808
	[Trip_Motive=5]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	.190	.060	10.093	1	.001			
	[Trip_Motive=1]	-.401	.238	2.833	1	.092	.670	.420	1.068
	[Trip_Motive=2]	-.082	.240	.115	1	.734	.922	.575	1.476
	[Trip_Motive=3]	.018	.224	.006	1	.936	1.018	.657	1.578
	[Trip_Motive=4]	-.703	.112	39.519	1	.000	.495	.398	.616
	[Trip_Motive=5]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.151	.090	162.956	1	.000			
	[Trip_Motive=1]	-.022	.330	.004	1	.947	.978	.512	1.868
	[Trip_Motive=2]	.160	.337	.226	1	.634	1.174	.606	2.273
	[Trip_Motive=3]	.126	.324	.151	1	.697	1.135	.601	2.142
	[Trip_Motive=4]	-.561	.173	10.485	1	.001	.571	.407	.801
	[Trip_Motive=5]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.TWO. 2 | TRANSPORT MODE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.171	.116	102.532	1	.000			
	[Trip_TransportMode=1]	.008	.142	.003	1	.955	1.008	.763	1.332
	[Trip_TransportMode=2]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.488	.091	28.612	1	.000			
	[Trip_TransportMode=1]	.659	.106	38.308	1	.000	1.933	1.569	2.381
	[Trip_TransportMode=2]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.613	.138	136.598	1	.000			
	[Trip_TransportMode=1]	.460	.161	8.191	1	.004	1.584	1.156	2.171
	[Trip_TransportMode=2]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.TWO. 3 | DEPARTURE TIME

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.786	.208	73.720	1	.000			
	[Trip_DepartureTime=2]	1.637	.240	46.630	1	.000	5.139	3.213	8.221
	[Trip_DepartureTime=3]	.191	.247	.598	1	.439	1.210	.746	1.964
	[Trip_DepartureTime=4]	.521	.262	3.946	1	.047	1.684	1.007	2.818
	[Trip_DepartureTime=5]	-.274	.361	.580	1	.446	.760	.375	1.541
	[Trip_DepartureTime=6]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	.037	.110	.110	1	.740			
	[Trip_DepartureTime=2]	.567	.150	14.376	1	.000	1.763	1.315	2.364
	[Trip_DepartureTime=3]	-.354	.139	6.482	1	.011	.702	.535	.922
	[Trip_DepartureTime=4]	-.150	.155	.932	1	.334	.861	.635	1.167
	[Trip_DepartureTime=5]	-.385	.189	4.128	1	.042	.681	.469	.987
	[Trip_DepartureTime=6]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.038	.154	45.387	1	.000			
	[Trip_DepartureTime=2]	-.114	.226	.255	1	.613	.892	.573	1.390
	[Trip_DepartureTime=3]	-.472	.201	5.528	1	.019	.624	.421	.925
	[Trip_DepartureTime=4]	-.168	.220	.581	1	.446	.846	.550	1.301
	[Trip_DepartureTime=5]	-.409	.274	2.219	1	.136	.665	.388	1.138
	[Trip_DepartureTime=6]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.TWO. 4 | TRIP FREQUENCY

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.792	.382	22.014	1	.000			
	[Trip_Frequency=1]	1.290	.425	9.203	1	.002	3.632	1.578	8.355
	[Trip_Frequency=2]	.212	.397	.285	1	.593	1.236	.568	2.691
	[Trip_Frequency=3]	1.016	.399	6.499	1	.011	2.763	1.265	6.038
	[Trip_Frequency=4]	.709	.458	2.399	1	.121	2.032	.829	4.985
	[Trip_Frequency=5]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.504	.235	4.590	1	.032			
	[Trip_Frequency=1]	.628	.283	4.916	1	.027	1.873	1.076	3.262
	[Trip_Frequency=2]	.459	.244	3.545	1	.060	1.582	.981	2.551
	[Trip_Frequency=3]	.647	.251	6.632	1	.010	1.909	1.167	3.123
	[Trip_Frequency=4]	.248	.304	.666	1	.414	1.281	.707	2.324
	[Trip_Frequency=5]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-2.079	.433	23.062	1	.000			
	[Trip_Frequency=1]	.884	.494	3.202	1	.074	2.421	.919	6.377
	[Trip_Frequency=2]	.855	.443	3.726	1	.054	2.352	.987	5.605
	[Trip_Frequency=3]	.769	.455	2.857	1	.091	2.158	.884	5.264
	[Trip_Frequency=4]	.660	.520	1.613	1	.204	1.935	.699	5.362
	[Trip_Frequency=5]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.TWO. 5 | TRIP DISTANCE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.464	.199	53.948	1	.000			
	[Trip_Distance=1]	-.976	.245	15.816	1	.000	.377	.233	.610
	[Trip_Distance=2]	1.777	.237	56.346	1	.000	5.913	3.718	9.404
	[Trip_Distance=3]	1.189	.263	20.467	1	.000	3.283	1.962	5.495
	[Trip_Distance=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	.072	.120	.360	1	.549			
	[Trip_Distance=1]	-.650	.138	22.273	1	.000	.522	.398	.684
	[Trip_Distance=2]	.762	.167	20.769	1	.000	2.142	1.544	2.972
	[Trip_Distance=3]	.806	.180	20.075	1	.000	2.238	1.573	3.184
	[Trip_Distance=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.401	.194	51.998	1	.000			
	[Trip_Distance=1]	-.165	.217	.573	1	.449	.848	.554	1.299
	[Trip_Distance=2]	.881	.251	12.307	1	.000	2.413	1.475	3.948
	[Trip_Distance=3]	.528	.284	3.459	1	.063	1.696	.972	2.960
	[Trip_Distance=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.TWO. 6 | TRIP DURATION

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.565	.229	46.590	1	.000			
	[Trip_Duration=1]	-2.457	.471	27.174	1	.000	.086	.034	.216
	[Trip_Duration=2]	-.551	.307	3.225	1	.073	.576	.316	1.052
	[Trip_Duration=3]	1.086	.264	16.982	1	.000	2.962	1.767	4.965
	[Trip_Duration=4]	1.850	.264	49.203	1	.000	6.361	3.793	10.668
	[Trip_Duration=5]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.116	.139	.692	1	.406			
	[Trip_Duration=1]	-.662	.170	15.215	1	.000	.516	.370	.719
	[Trip_Duration=2]	-.196	.173	1.290	1	.256	.822	.586	1.153
	[Trip_Duration=3]	.544	.173	9.895	1	.002	1.724	1.228	2.419
	[Trip_Duration=4]	1.061	.181	34.313	1	.000	2.888	2.025	4.118
	[Trip_Duration=5]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.442	.218	43.751	1	.000			
	[Trip_Duration=1]	-.501	.267	3.510	1	.061	.606	.359	1.023
	[Trip_Duration=2]	.220	.259	.723	1	.395	1.247	.750	2.072
	[Trip_Duration=3]	.526	.265	3.946	1	.047	1.692	1.007	2.844
	[Trip_Duration=4]	.797	.275	8.375	1	.004	2.218	1.293	3.804
	[Trip_Duration=5]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.3 | AREAL MULTINOMIAL REGRESSION MODELS

APPENDIX III.THREE. 1 | SURFACE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.473	.162	82.942	1	.000			
	[Area_Surface=1]	.464	.214	4.714	1	.030	1.590	1.046	2.417
	[Area_Surface=2]	.511	.230	4.960	1	.026	1.668	1.063	2.616
	[Area_Surface=3]	.347	.205	2.870	1	.090	1.415	.947	2.113
	[Area_Surface=4]	.287	.256	1.255	1	.263	1.333	.806	2.203
	[Area_Surface=5]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.060	.100	.362	1	.548			
	[Area_Surface=1]	.039	.143	.075	1	.784	1.040	.785	1.378
	[Area_Surface=2]	.165	.155	1.132	1	.287	1.179	.870	1.598
	[Area_Surface=3]	.109	.133	.679	1	.410	1.115	.860	1.447
	[Area_Surface=4]	-.144	.175	.680	1	.409	.865	.614	1.220
	[Area_Surface=5]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.797	.185	94.137	1	.000			
	[Area_Surface=1]	.633	.237	7.145	1	.008	1.884	1.184	2.998
	[Area_Surface=2]	.816	.247	10.866	1	.001	2.261	1.392	3.673
	[Area_Surface=3]	.459	.230	3.992	1	.046	1.583	1.009	2.484
	[Area_Surface=4]	.752	.264	8.097	1	.004	2.121	1.264	3.561
	[Area_Surface=5]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.THREE. 2 | EDUCATION

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.308	.085	238.577	1	.000			
	[Area_Education=1]	.489	.250	3.821	1	.051	1.631	.999	2.663
	[Area_Education=2]	.436	.155	7.902	1	.005	1.547	1.141	2.096
	[Area_Education=3]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.106	.057	3.498	1	.061			
	[Area_Education=1]	.248	.187	1.766	1	.184	1.282	.889	1.848
	[Area_Education=2]	.324	.110	8.624	1	.003	1.383	1.114	1.717
	[Area_Education=3]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.379	.087	250.521	1	.000			
	[Area_Education=1]	.479	.257	3.469	1	.063	1.615	.975	2.674
	[Area_Education=2]	.234	.168	1.947	1	.163	1.264	.910	1.757
	[Area_Education=3]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.THREE. 3 | FAMILY STRUCTURE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.034	.095	118.402	1	.000			
	[Area_FamilyStructure=1]	.435	.254	2.937	1	.087	1.545	.939	2.540
	[Area_FamilyStructure=2]	-.418	.145	8.300	1	.004	.659	.496	.875
	[Area_FamilyStructure=3]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	.058	.068	.719	1	.396			
	[Area_FamilyStructure=1]	-.139	.213	.426	1	.514	.870	.573	1.322
	[Area_FamilyStructure=2]	-.150	.097	2.412	1	.120	.860	.712	1.040
	[Area_FamilyStructure=3]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-1.089	.097	126.043	1	.000			
	[Area_FamilyStructure=1]	-.204	.317	.413	1	.521	.816	.438	1.518
	[Area_FamilyStructure=2]	-.423	.148	8.131	1	.004	.655	.490	.876
	[Area_FamilyStructure=3]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.THREE. 4 | HOUSING TYPE

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-.799	.401	3.958	1	.047			
	[Area_HousingType=1]	-.269	.420	.411	1	.521	.764	.335	1.740
	[Area_HousingType=2]	-.361	.417	.748	1	.387	.697	.308	1.579
	[Area_HousingType=3]	-.555	.421	1.739	1	.187	.574	.251	1.310
	[Area_HousingType=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.357	.348	1.048	1	.306			
	[Area_HousingType=1]	.525	.359	2.145	1	.143	1.691	.837	3.415
	[Area_HousingType=2]	.312	.357	.763	1	.382	1.366	.678	2.753
	[Area_HousingType=3]	.164	.359	.208	1	.648	1.178	.583	2.380
	[Area_HousingType=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-.598	.375	2.536	1	.111			
	[Area_HousingType=1]	-.481	.395	1.483	1	.223	.618	.285	1.341
	[Area_HousingType=2]	-.722	.395	3.353	1	.067	.486	.224	1.052
	[Area_HousingType=3]	-.985	.400	6.056	1	.014	.373	.170	.818
	[Area_HousingType=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.THREE. 5 | YEAR OF CONSTRUCTION

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-.800	.143	31.345	1	.000			
	[Area_YearOfConstruction=1]	-1.075	.251	18.296	1	.000	.341	.209	.559
	[Area_YearOfConstruction=2]	-.476	.181	6.942	1	.008	.621	.436	.885
	[Area_YearOfConstruction=3]	-.006	.208	.001	1	.978	.994	.661	1.496
	[Area_YearOfConstruction=4]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	.216	.107	4.063	1	.044			
	[Area_YearOfConstruction=1]	-.488	.157	9.697	1	.002	.614	.451	.835
	[Area_YearOfConstruction=2]	-.374	.131	8.115	1	.004	.688	.532	.890
	[Area_YearOfConstruction=3]	.129	.153	.707	1	.400	1.138	.842	1.537
	[Area_YearOfConstruction=4]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-.952	.151	39.862	1	.000			
	[Area_YearOfConstruction=1]	-.923	.256	13.016	1	.000	.397	.241	.656
	[Area_YearOfConstruction=2]	-.333	.187	3.178	1	.075	.716	.497	1.034
	[Area_YearOfConstruction=3]	.010	.219	.002	1	.963	1.010	.658	1.552
	[Area_YearOfConstruction=4]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX III.4 | BUS MULTINOMIAL REGRESSION MODELS

APPENDIX III.THREE. 6 | DISTANCE BUS STOP

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.508	.213	50.286	1	.000			
	[Bus_Distance=1]	.564	.380	2.200	1	.138	1.757	.834	3.701
	[Bus_Distance=2]	-.219	.439	.249	1	.618	.803	.340	1.898
	[Bus_Distance=3]	.549	.236	5.394	1	.020	1.731	1.089	2.750
	[Bus_Distance=4]	.300	.240	1.560	1	.212	1.349	.843	2.159
	[Bus_Distance=5]	0[b]	.	.	0	.	.	.	.
Yes, but not Realistic	Intercept	-.410	.143	8.166	1	.004			
	[Bus_Distance=1]	.381	.277	1.892	1	.169	1.464	.850	2.522
	[Bus_Distance=2]	.186	.266	.493	1	.483	1.205	.716	2.028
	[Bus_Distance=3]	.531	.161	10.805	1	.001	1.700	1.239	2.333
	[Bus_Distance=4]	.407	.162	6.319	1	.012	1.502	1.094	2.062
	[Bus_Distance=5]	0[b]	.	.	0	.	.	.	.
Don't Know	Intercept	-2.239	.292	58.899	1	.000			
	[Bus_Distance=1]	.958	.461	4.312	1	.038	2.607	1.055	6.440
	[Bus_Distance=2]	1.140	.417	7.475	1	.006	3.128	1.381	7.085
	[Bus_Distance=3]	1.140	.311	13.434	1	.000	3.128	1.700	5.756
	[Bus_Distance=4]	.962	.313	9.441	1	.002	2.617	1.417	4.834
	[Bus_Distance=5]	0[b]	.	.	0	.	.	.	.

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX IV | INTEGRAL MODELS

APPENDIX IV. 1 | INTEGRAL PERSONAL VARIABLES

COMBINATIONS	Selection Method	-LL2 Start	-LL2 Final	Chi^2	df	Sign.	Used Variables	Warning
All Continuous (and Categoric) Variables	Main Effects	4385.436	4203.456	181.980	54	0.000	Gender, Age, Education, Family Size, Family Structure, Owned Cars, PT Usage, PT Subscription, Housing Type, Housing Estate	56.6% with zero frequencies
	Forward Entry	4385.436	4247.446	137.970	18	0.000	Age, PT Usage, Education	56.6% with zero frequencies
	Backward Elimination	4385.436	4231.551	153.885	27	0.000	Education, PT Usage, Housing Type, Age	56.6% with zero frequencies
All Categoric Variables	Main Effects	3329.581	3134.525	195.056	66	0.000	Gender, Age, Education, Family Size, Family Structure, Owned Cars, PT Usage, PT Subscription, Housing Type, Housing Estate	50.4% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
	Forward Entry	3329.581	3198.627	130.954	24	0.000	PT Usage, Education, Age	50.4% with zero frequencies
	Backward Elimination	3329.581	3145.571	184.010	51	0.000	Family Size, Age, Education, Family Structure, PT Usage, Housing Type	50.4% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
All Effect Coded Variables	Main Effects	3329.581	3134.525	195.056	66	0.000	Gender EC1, Age EC1; 2; 3, Education EC1; 2; 3, Family Size EC1; 2; 3; 4, Family Structure EC1; 2; 3, Owned Cars EC1; 2, PT Usage EC1; 2; 3, PT Subscription EC1, Housing Type EC1; 2; 3, Housing Estate EC1	50.4% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
	Forward Entry	3329.581	3187.765	141.816	21	0.000	Age EC2, PT Usage EC1, Eduction EC3, PT Usage EC3, Education EC2, Housing Type EC3, Family Size EC1	50.4% with zero frequencies
	Backward Elimination	3329.581	3153.302	176.279	39	0.000	PT Usage EC2, Family Structure EC3, Family Structure EC1, Family Size EC3, Family Size EC1, Education EC2, Age EC2, Education EC1, Family Size EC2, Family Size EC4, PT Usage EC1, PT Usage EC3, Housing Type EC3	50.4% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
Significant Variables	Main Effects	4074.260	3880.889	193.371	48	0.000	Age, Education, Family Size, Family Structure, PT Usage, PT Subscription, Housing Type	54.2% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
	Forward Entry	4047.260	3897.995	176.265	36	0.000	Age, PT Usage, Education, Family Size, Family Structure	54.2% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
	Backward Elimination	4074.260	3881.635	192.625	45	0.000	Family Structure, Education, Family Size, PT Usage, Housing Type	54.2% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
Significant Effect Coded Variables	Main Effects	4074.260	3880.889	193.371	48	0.000	Age, Education EC1; 2; 3, Family Size EC1; 2; 3; 4, Family Structure EC1; 2; 2, PT Usage EC1; 2; 3, PT subscription EC1, Housing Type EC1; 2; 3	54.2% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
	Forward Entry	4074.260	3920.017	154.243	21	0.000	Age, PT Usage EC1, Education EC3, PT Usage EC3, Education EC1, Family Size EC1, Housing Type EC3	54.2% with zero frequencies
	Backward Elimination	4047.260	3888.978	185.282	33	0.000	Housing Type EC3, PT Usage EC3, PT Usage EC1, Family Structure EC2, Family Size EC2, Education EC3, Education EC1, Age, Family Size EC1, Family Structure EC1, PT Usage EC2	54.2% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain

= Significant Model

APPENDIX IV. 2 | INTEGRAL TRIP MODELS

COMBINATIONS	Selection Method	-LL2 Start	-LL2 Final	Chi²	df	Sign.	Used Variables	Warning
All Categoric Variables	Main Effects	3183.587	2576.771	606.816	60	0.000	Motive, Transport Mode, Departure Time, Frequency, Distance, Duration	58.7% zero frequencies
	Forward Entry	3183.587	2576.771	606.816	60	0.000	Duration, Distance, Motive, Transport Mode, Frequency, Departure Time	58.7% zero frequencies
	Backward Elimination	3183.587	2576.771	606.816	60	0.000	Duration, Frequency, Transport Mode, Motive, Departure Time, Distance	58.7% with zero frequencies, Stepwise has no effect on the initial model
All Effect Coded Variables	Main Effects	3183.587	2576.771	606.816	60	0.000	Motive EC1; 2; 3; 4, Transport Mode EC1, Departure Time EC1; 2; 3; 4; 5, Frequency EC1; 2; 3; 4, Distance EC1; 2; 3, Duration EC1; 2; 3; 4	58.7% zero frequencies
	Forward Entry	3183.587	2618.133	565.454	27	0.000	Duration EC4; 1; 2, Distance EC2, Motive EC4, Tranport Mode EC1, Departure Time EC2, Frequency EC1, Duration EC2; 3	58.7% zero frequencies
	Backward Elimination	3183.587	2607.652	575.935	30	0.000	Duration EC4, Distance EC1, Frequency EC1, Departure Time EC2, Transport Mode EC1, Motive EC2; 4, Distance EC2, Duration EC1; 3	58.7% zero frequencies

= Significant Model

APPENDIX IV. 3 | INTEGRAL AREAL MODELS

COMBINATIONS	Selection Method	-LL2 Start	-LL2 Final	Chi²	df	Sign.	Used Variables	Warning
All Continuous (and Categoric) Variables	Main Effects	1685.146	1581.953	103.509	57	0.000	Surface, Inhabitants, Households, Family Size, Education, Family Structure, Housing Type, Year of Construction, Property Value, Estate	41.2% with zero frequencies
	Forward Entry	1685.461	1636.455	49.006	12	0.000	Year of Construction, Surface	41.2% with zero frequencies
	Backward Elimination	1685.461	1618.635	66.827	21	0.000	Family Size, Year of Construction, Family Structure	41.2% with zero frequencies
All Categoric Variables	Main Effects	1644.508	1505.731	138.777	81	0.000	Surface, Inhabitants, Households, Family Size, Education, Family Structure, Housing Type, Year of Construction, Property Value, Estate	40.8% with zero frequencies
	Forward Entry	1644.508	1568.105	76.403	30	0.000	Family Size, Year of Construction, Family Structure, Housing Type	40.8% with zero frequencies
	Backward Elimination	1644.508	1539.907	104.601	48	0.000	Family Size, Inhabitants, Year of Construction, Family Structure, Housing Type, Households	40.8% with zero frequencies
All Effect Coded Variables	Main Effects	1644.508	1505.731	138.777	81	0.000	Surface EC1; 2; 3; 4, Inhabitants EC1; 2; 3, Housholds EC1; 2; 3, Family Size EC1; 2, Education EC1; 2, Family Structure EC1; 2; 3, Housing Type EC1; 2; 3, Year of Construction EC1; 2; 3, Property Value EC1; 2; 3; 4, Housing Type EC1; 2; 3	40.8% with zero frequencies
	Forward Entry	1644.508	1600.246	44.262	12	0.000	Households EC3, Surface EC4, Family Structure EC2, Education EC1	40.8% with zero frequencies
	Backward Elimination	1644.508	1545.554	98.954	36	0.000	Family Size EC2, Households EC3, Surface EC4, Property Value EC3; 1, Year of Construction EC2, Family Structure EC2, Education EC1, Year of Construction EC1; 3, Inhabitants EC1, Family Size EC1	40.8% with zero frequencies
Significant Variables	Main Effects	2607.367	2513.996	93.371	33	0.000	Surface, Education, Family Structure, Housing Type, Year of Construction	39.7% with zero frequencies
	Forward Entry	2607.367	2526.666	80.701	21	0.000	Surface, Housing Type, Year of Consruction	39.7% with zero frequencies
	Backward Elimination	2607.367	2526.666	80.701	21	0.000	Surface, Housing Type, Year of Construction	39.7% with zero frequencies
Significant Effect Coded Variables	Main Effects	2607.376	2513.996	93.371	33	0.000	Surface, Education EC1; 2, Family Structure EC1; 2; 3, Housing Type EC; 2; 3, Year of Construction EC1; 2; 3	39.7% with zero frequencies
	Forward Entry	2607.376	2544.120	63.247	12	0.000	Year of Construction EC 3; 1, Surface, Education EC2	39.7% with zero frequencies
	Backward Elimination	2607.376	2531.513	75.855	18	0.000	Year of Construction EC 3; 1, Housing Type EC2, Surface, Housing Type EC3, Year of Construction EC2	39.7% with zero frequencies

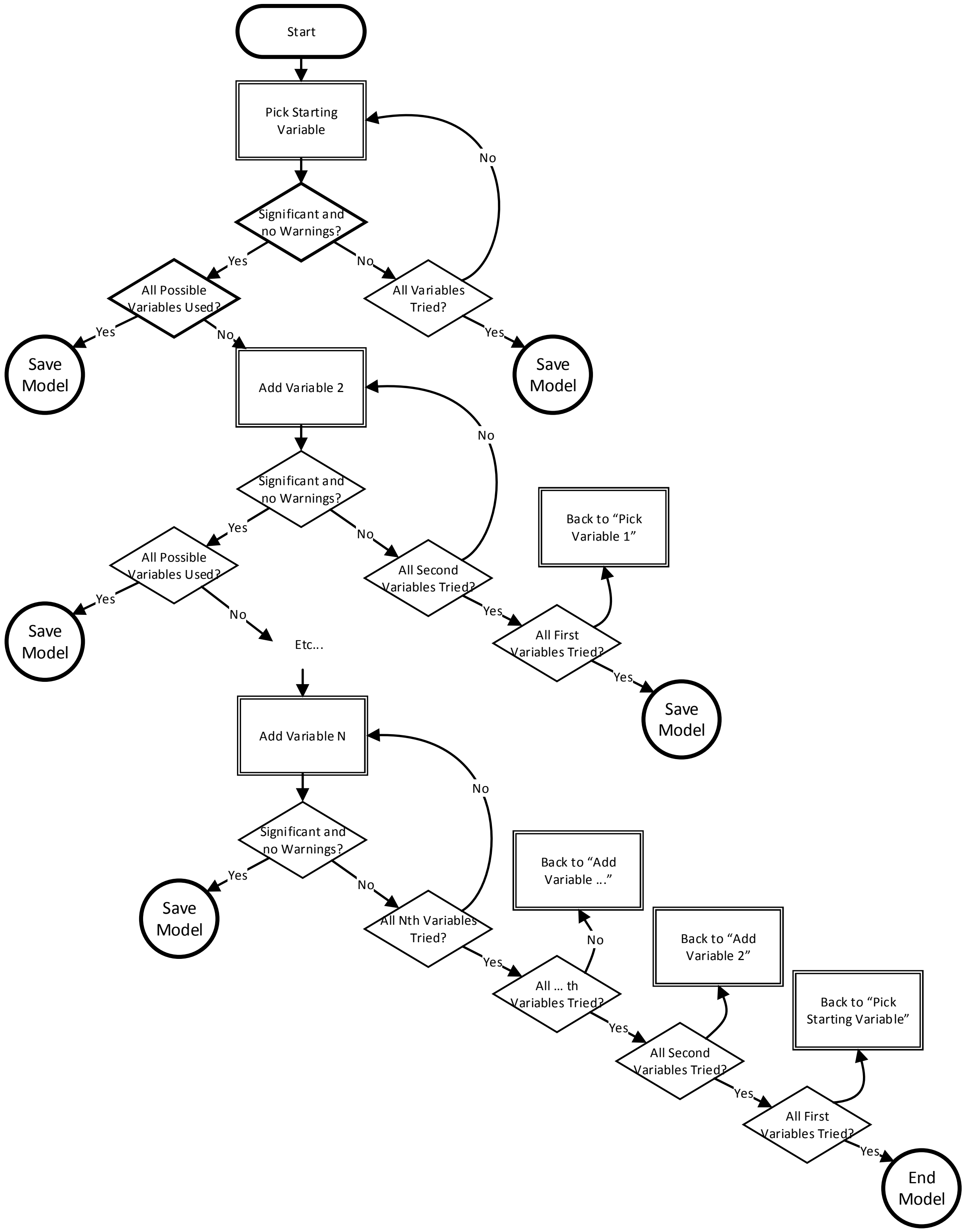
= Significant Model

APPENDIX IV. 4 | INTEGRAL BUS MODELS

COMBINATIONS	Selection Method	-LL2 Start	-LL2 Final	Chi²	df	Sign.	Used Variables	Warning
All Continuous Variables	Main Effects	2856.578	2821.133	35.444	12	0.000	Stops, Lines, Lines Area, Distance	40.5% with zero frequencies
	Forward Entry	2856.578	2827.148	29.430	3	0.000	Distance	40.5% with zero frequencies
	Backward Elimination	2856.578	2827.148	29.430	3	0.000	Distance	40.5% with zero frequencies
All Categoric Variables	Main Effects	84.493	60.711	23.782	24	0.474	Stops, Lines, Lines Area, Distance	37.5% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
	Forward Entry	84.493	84.493	0.000	0	-	Intercept	37.5% with zero frequencies, Stepwise has no effect on the initial model
	Backward Elimination	84.493	65.812	18.681	12	0.097	Lines Area	37.5% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
All Effect Coded Variables	Main Effects	84.493	60.711	23.782	24	0.474	Stops EC1, Lines EC1, Lines Area EC1; 2; 3; 4, Distance EC1; 2; 3; 4	37.5% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
	Forward Entry	84.493	74.534	9.958	3	0.019	Lines Area EC4	37.5% with zero frequencies
	Backward Elimination	84.493	74.534	9.958	3	0.019	Lines Area EC4	37.5% with zero frequencies, Unexpected singularities in the Hessian matrix, Validity of the model fit is uncertain
All Categoric Variables 2	Main Effects	223.936	187.721	36.215	18	0.007	Stops, Lines, Lines Area, Distance	7.1% with zero frequencies
	Forward Entry	223.936	195.610	28.326	12	0.005	Distance	7.1% with zero frequencies, Stepwise has no effect on the initial model
	Backward Elimination	223.936	195.610	28.326	12	0.005	Distance	7.1% with zero frequencies
All Effect Coded Variables 2	Main Effects	223.936	187.721	36.215	18	0.007	Stops EC1, Lines EC1, Lines Area EC1, Distance EC1; 2; 3; 4	7.1% with zero frequencies
	Forward Entry	223.936	205.704	18.231	3	0.000	Distance EC3	7.1% with zero frequencies
	Backward Elimination	223.936	205.704	18.231	3	0.000	Distance EC3	7.1% with zero frequencies

= No Model	= Significant Model	= No Significant Model

APPENDIX V | FLOW CHART DIAGRAM STEP BY STEP MODELS



APPENDIX VI | STEP BY STEP MODELS BY VARIABLE GROUP

APPENDIX VI. 1 | STEP BY STEP PERSONAL MODELS

CATEGORY	Variable 1	Variable 2	Variable 3	-LL2 Start	-LL2 Final	Chi²	df	Sign.
From Correlations	PTSubcsription	Housing Type	x	164.980	126.945	38.035	12	0.000
Age	Age	PT Subscription	x	171.408	119.220	52.188	12	0.000
Education	Education	Housing Type	x	249.955	191.388	58.567	15	0.000
	Education	PT Subscription	x	164.051	106.802	57.249	9	0.000
Family Size	Family Size	PT Subscription	x	200.667	160.222	40.445	15	0.000
Family Structure	Family Structure	Housing Type	x	273.674	234.472	39.201	18	0.003
	Family Structure	PT Subscription	x	166.959	131.622	35.337	12	0.000
Housing Type	Housing Type	PT Subscription	x	164.980	126.945	38.035	12	0.000

= Significant Model

APPENDIX VI. 2 | STEP BY STEP TRIP MODELS

CATEGORY	Variable 1	Variable 2	Variable 3	-LL2 Start	-LL2 Final	Chi²	df	Sign.
From Correlations	-							
Motive	Motive	Transport Mode	x	252.721	149.970	102.751	15	0.000
Transport Mode	Transport Mode	Departure Time	x	347.144	168.071	179.073	15	0.000
	Transport Mode	Distance	x	516.280	135.055	381.225	12	0.000
	Transport Mode	Duration	x	603.043	181.755	421.287	15	0.000

= Significant Model

APPENDIX VI. 3 | STEP BY STEP AREAL MODELS

CATEGORY	Variable 1	Variable 2	Variable 3	-LL2 Start	-LL2 Final	Chi²	df	Sign.
From Correlations	-							
Surface	Surface	Year of Construction	x	367.258	307.590	59.668	21	0.000
Education	Education	Family Structure	x	170.884	138.949	31.935	12	0.001
Family Structure	Family Structure	Year of Construction	x	240.119	184.304	55.815	15	0.000

= Significant Model

APPENDIX VII | PARAMETER ESTIMATES, STEP BY STEP MODELS, COMBINED VARIABLE GROUPS

APPENDIX VII. 1| MODEL 1

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.097	.089	150.961	1	.000			
	Education_EC1	-.053	.247	.046	1	.830	.948	.584	1.539
	Education_EC2	-.019	.164	.014	1	.906	.981	.711	1.353
	Education_EC3	0[b]	.	.	0	.	.	.	.
	PTSubscription_EC1	.141	.078	3.275	1	.070	1.151	.988	1.342
	TransportMode_EC1	.041	.073	.312	1	.577	1.041	.903	1.201
Yes, but not Realistic	Intercept	-.202	.068	8.742	1	.003			
	Education_EC1	-.320	.182	3.098	1	.078	.726	.508	1.037
	Education_EC2	.191	.111	2.997	1	.083	1.211	.975	1.504
	Education_EC3	0[b]	.	.	0	.	.	.	.
	PTSubscription_EC1	-.020	.059	.113	1	.737	.981	.874	1.100
	TransportMode_EC1	.342	.054	39.516	1	.000	1.407	1.265	1.565
Don't Know	Intercept	-1.795	.120	224.372	1	.000			
	Education_EC1	.938	.212	19.543	1	.000	2.555	1.686	3.874
	Education_EC2	.603	.163	13.636	1	.000	1.828	1.327	2.518
	Education_EC3	0[b]	.	.	0	.	.	.	.
	PTSubscription_EC1	-.221	.100	4.880	1	.027	.802	.660	.975
	TransportMode_EC1	.178	.082	4.695	1	.030	1.194	1.017	1.402

a. The reference category is: No Possibility.

b. This parameter is set to zero because it is redundant.

APPENDIX VII. 2 | MODEL 2

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.004	.092	118.906	1	.000			
	FamilySize_EC1	.219	.171	1.648	1	.199	1.245	.891	1.741
	FamilySize_EC2	-.229	.111	4.264	1	.039	.795	.640	.988
	FamilySize_EC3	-.061	.170	.127	1	.721	.941	.674	1.314
	FamilySize_EC4	.035	.149	.054	1	.816	1.035	.773	1.387
	PTSubscription_EC1	.152	.079	3.690	1	.055	1.164	.997	1.360
	TransportMode_EC1	.024	.072	.111	1	.739	1.024	.889	1.181
Yes, but not Realistic	Intercept	-.126	.072	3.077	1	.079			
	FamilySize_EC1	.056	.129	.184	1	.668	1.057	.820	1.362
	FamilySize_EC2	-.144	.078	3.390	1	.066	.866	.742	1.009
	FamilySize_EC3	-.008	.119	.005	1	.946	.992	.785	1.253
	FamilySize_EC4	.231	.102	5.104	1	.024	1.260	1.031	1.541
	PTSubscription_EC1	-.004	.059	.004	1	.948	.996	.887	1.119
	TransportMode_EC1	.326	.054	36.133	1	.000	1.385	1.245	1.540
Don't Know	Intercept	-1.503	.120	157.182	1	.000			
	FamilySize_EC1	.367	.180	4.128	1	.042	1.443	1.013	2.055
	FamilySize_EC2	-.176	.123	2.039	1	.153	.839	.659	1.068
	FamilySize_EC3	.168	.172	.955	1	.329	1.183	.844	1.658
	FamilySize_EC4	.227	.152	2.245	1	.134	1.255	.932	1.690
	PTSubscription_EC1	-.238	.101	5.613	1	.018	.788	.647	.960
	TransportMode_EC1	.183	.082	5.031	1	.025	1.201	1.023	1.409

a. The reference category is: No Possibility.

APPENDIX VII. 3 | MODEL 3

Parameter Estimates

Bus Potential[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	Upper Bound
Yes	Intercept	-1.114	.089	157.988	1	.000			
	HousingType_EC1	.047	.134	.124	1	.725	1.048	.805	1.365
	HousingType_EC2	.039	.107	.133	1	.715	1.040	.843	1.283
	HousingType_EC3	-.059	.121	.236	1	.627	.943	.743	1.196
	PTSubscription_EC1	.143	.078	3.398	1	.065	1.154	.991	1.344
	TransportMode_EC1	.030	.073	.169	1	.681	1.030	.893	1.189
Yes, but not Realistic	Intercept	-.284	.070	16.253	1	.000			
	HousingType_EC1	.103	.097	1.141	1	.286	1.109	.917	1.340
	HousingType_EC2	.179	.077	5.421	1	.020	1.196	1.029	1.391
	HousingType_EC3	.106	.085	1.551	1	.213	1.112	.941	1.314
	PTSubscription_EC1	-.030	.058	.260	1	.610	.971	.866	1.088
	TransportMode_EC1	.338	.055	38.345	1	.000	1.402	1.260	1.560
Don't Know	Intercept	-1.484	.108	187.448	1	.000			
	HousingType_EC1	.201	.132	2.325	1	.127	1.223	.944	1.583
	HousingType_EC2	-.120	.112	1.149	1	.284	.887	.713	1.104
	HousingType_EC3	-.263	.129	4.143	1	.042	.769	.597	.990
	PTSubscription_EC1	-.249	.099	6.274	1	.012	.780	.642	.947
	TransportMode_EC1	.214	.082	6.779	1	.009	1.239	1.054	1.455

a. The reference category is: No Possibility.

APPENDIX VIII | OUTPUT TABLES LINEAR REGRESSION MODEL

APPENDIX VIII. 1| MODEL SUMMARY

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.081[a]	.007	.005	1.093	1.778

a. Predictors: (Constant), T: Transport Mode of the Trip, P: Number of People in the Family, P: Type of Public Transport Subscription

b. Dependent Variable: Bus Potential

APPENDIX VIII. 2 | COEFFICIENT CORRELATION

Model			T: Transport Mode of the Trip	P: Number of People in the Family	P: Type of Public Transport Subscription
1	Correlations	T: Transport Mode of the Trip	1.000	-.033	.150
		P: Number of People in the Family	-.033	1.000	-.160
		P: Type of Public Transport Subscription	.150	-.160	1.000
	Covariances	T: Transport Mode of the Trip	.003	-3.279E-5	.000
		P: Number of People in the Family	-3.279E-5	.000	.000
		P: Type of Public Transport Subscription	.000	.000	.003

a. Dependent Variable: Bus Potential

APPENDIX VIII. 3 | COEFFICIENTS

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.458	.132		18.624	.000		
	P: Number of People in the Family	-.024	.020	-.024	-1.182	.237	.974	1.026
	P: Type of Public Transport Subscription	.068	.056	.025	1.214	.225	.953	1.049
	T: Transport Mode of the Trip	.190	.050	.078	3.769	.000	.977	1.023

a. Dependent Variable: Bus Potential

English Summary

POTENTIAL USE OF PUBLIC TRANSPORT IN THE CONTEXT OF AIR POLLUTION REDUCTION |

Development of a tool to determine and localize potential bus users.

H. (Hannely) Hortensius

Graduation program:
Construction Management and Urban Development 2014-2015

Graduation committee:

Prof. dr. ir. W.F. (Wim) Schaefer	|	Chairman Master Program CME, TU/e
Dr. ing. P.J.H.J. (Peter) van der Waerden	|	First graduation supervisor, TU/e
Dr. Q. (Qi) Han	|	Second graduation supervisor, TU/e
Ir. J.J.F. (Joris) Hoogenboom	|	Advisor, BonoTraffics bv.

Date of graduation:
25-02-2015

ABSTRACT

Since the air pollution situation in Europe is getting worse, some measures should be taken. More than a fifth of the greenhouse gas emissions is due to traffic-consumption and therefore this research will make an attempt to optimize the transport effectiveness by supporting public transport. This will be done by developing a model which determines and localizes potential bus users. To make this tool a model is estimated with the use of personal, trip, and areal characteristics. The tool could be used by public transport agencies or traffic engineering consultancies to optimize the existing bus network or implement the model by creating future networks in new neighborhoods.

Keywords: Air quality; Environment & Health; Public Transport; Geo-Marketing; Multinomial Regression Model.

INTRODUCTION

In Western Europe very high air pollution levels are measured, and since the need for transport only grows, higher levels are expected. In the near past, especially France, Belgium and Germany suffered from very bad air quality, but if no measures are taken the Netherlands will surely follow their paths (European Environment Agency 2014).

Motor vehicles are a very significant source of urban air pollutions (Health Effects Institute 2010). In 2011 a third of all final energy consumption in Europe was accounted by transport, this traffic-consumption was responsible to more than a fifth of the greenhouse gas emissions (European Environment Agency 2011). So, one option to tackle the air pollution problem is to make an adjustment regarding the use of transport.

Problem Definition and Research Questions

Many papers are written about the attitude of travelers, but still the public transportation agencies have little to no insights towards the potential of their nodes. Important questions for these businesses are whether they reach all the possible travelers, where potential users are vested and how they can be reached and how sensitive their customers are towards different facilities. Therefore this research will combine the elements of attitude towards transport modes, the effects of measures and the potential for customers in a certain area.

The target of the research is to develop a tool, which gives public transport agencies more information about the potential users of public transport. Then, they can anticipate on this, and the amount of private car-users can be reduced. The research questions that follow this target are:

“Which research methodology should be used to develop a tool which determines and localizes potential bus users?’’

“Which personal, trip and areal characteristics should be used to model potential users of public transport?’’

“Can the Air Quality in the Netherlands be improved by stimulating the use of Public Transport?”

To answer these main questions several sub-questions are drafted concerning the definition of potential, behavior of travelers explained by areal characteristics and measurements to attract travelers towards the public transport.

Relevance of the Research
The use of more public transport can reduce the amount of private car-users, which will reduce emissions and will lead to a cleaner and healthier environment. Public transport agencies can locate potential users by the use of the tool. They can anticipate to the new information by changing their services and facilities in their operating areas which can switch the potential users into actual users.

The practical relevance can be, for example, the implementation of the model when a municipality is expanding. If a municipality wants to expand their city a consultancy could use the model to visualize the number of future potential public transport users. From this information a plan can be made about the public transport facilities in the new neighborhood which will fit the prospective demand.

Expected Results
The expected result for this research is the obtaining of a tool, which will not only state whether there are potential public transport users, but also where they are located. The tool will determine whether potential users are present and which policy could be implemented to reach a certain effect where the latent potential will become manifest potential. When this model is made all the potential users can be localized and the expected need for bus lines and maybe even bus stops can be modeled. This could be used by traffic accounting agencies which advice governments with new public transport routes in existing or future neighborhoods.

THEORETICAL FRAMEWORK

Sustainability

The actual influence of transportation to the environment is analyzed with special attention to the impact which public transport can have regarding this problem. The air quality depends on the presence of air pollutants. If the concentrations of this pollutants are too high air pollution can be harmful for the health. The most important pollutants are: particulate matter (PM_x), nitrogen dioxide (NO_2), ammonia (NH_3); sulfur dioxide (SO_2); nitrogen oxides (N_xO), heavy metals and hazardous substances which are difficult to degrade.

The European Union (EU) has limits and target values for several pollutants. The limits apply to sulfur dioxide, particulate matter, nitrogen dioxide, lead, benzene and carbon monoxide. The targets are for ozone, arsenic, cadmium, nickel and benzo (a) pyrene. The limits must not be exceeded, and all EU Member States should ensure to remain below the target values.

The Netherlands meets almost all European air quality limit values, except for a few places where the concentrations of particulates and nitrogen dioxide are still too high. Road traffic causes a large part of the emission in the Netherlands. Because the traffic in the Netherlands has the biggest share of the emissions, this is still an important issue on the governmental agenda.

Public transport is often seen as an important air pollution reduction strategy. This is not only because of the strategies which can be implemented in the systems, but also about the amount of people transporting per vehicle. When more travelers travel by Public Transport, the less personal car-trips will be used and the air pollution will be reduced. The target is not only to make transit service more appealing but also to lower the attention of the automobile.

Public Transport

In order to carry out this research a demarcation should be made. The air quality in the Netherlands is worst in cities, and the increase of citizens living in cities which predicted will only worsen the case. The most pollution emissions in cities are because of freight and busses. Therefore the bus networks will be investigated in this research, this refers not only to the local inner-city networks, but also to rapid bus transport systems which transfer between cities and/or villages.

The trends from the transportation of the last few years are subtracted from the ‘Onderzoek Verplaatsingen in Nederland (OViN)’ (Travel Research in the Netherlands). This is an annual research performed by the ‘Centraal Bureau voor de Statistiek (CBS)’ (Central Statistical Office from the Netherlands). This researches show that the car is by far the most popular transport mode, after this the train and bicycle, and only after these three the bus/tram/metro. The most important motive is commuting and after this visit trips or recreational trips.

Multiple advantages and disadvantages can be found in literature for both public transport and private car usage. The most important advantage of the private car is the flexibility which it gives to a driver. Also motivations and barriers are found for public transport users. A conclusion here was that most barriers are formed by aspects regarding comfort and convenience. Also several measurements have been implemented in the public transport market, but only some of the time the effect of this measures are investigated after the implementation. And only in a very limited number of cases the potential effects are estimated before the implementation of the measure.

The potential users of public transport are the travelers who, at the present, do not use the public transport, but which could switch towards it when certain measures are taken. This switching possibility depends on several aspects which mostly relates to the willingness of people to switch. The aspects are about the distance between origin or destination and bus stops, frequency of busses, whether the connection can be traveled directly or there is a need for transfers, ratio between travel time/distance/costs from bus or other transport mode, the ability to bring belongings and other comfort aspects like crowdedness and service (Waerden & Bérénos 2010).

Marketing

The previous discussed variables affecting the use of public transport and possible policies and measures. However, nothing will change regarding to transport mode choices from travelers if they are not aware of the newly implemented measures. To deliver this knowledge marketing can be used.

The role of marketing in the Public Transport has increased much in the recent past because of the evolution of transit. First, most people had to use the Public Transport because of few other alternatives. Now is the time of high car ownership and the Public Transport has to make an effort to attract customers.

To implement an individualized marketing strategy, the potential costumers should be selected and contacted directly. To do this, it is not only necessary to know who are possible users, but also where they are located. When the location of potential is not known it is not possible to directly focus the marketing to the potential. This location bounded marketing can also be referred to as 'Geo Marketing'. This is a marketing technique which is proven to be functional and has multiple advantages compared to general marketing.

RESEARCH APPROACH

Theory

In this research behavioral theory is used, the purpose is to understand why people choose a specific transport mode and to get to know where potential users are vested. With this knowledge the public transport agencies can participate to the behavior of the travelers and thereby promote, among others, the bus usage. Individual choice behavior, discrete choice modelling and switching-behavior are the most important theories used during the research.

Research Method

Earlier attempts on modelling public transport were almost always about the actual use of public transports instead of the potential users, nevertheless there are some exceptions. Sugiki et al. developed a potential traffic demand model using a binary logit model, but this model is very limited because of the use of only three variables which are all on the same level: household age, household type and residence type (Sugiki et al. 2001). Another attempt was made by Zhou et al. using a two-leveled nested logit model, eight different market segments were identified which were linked with several areas and the market share per market area was calculated (Zhou et al. 2004). However, in this model no directly relation is made between the areal characteristics and the market shares.

In 2005 Van der Waerden et al. made an attempt to make public transportation potential maps (Waerden et al. 2005). This paper uses a simple approach to identify potential customers of public transport by the use of GIS-based public transport potential maps. However this research and following researches involving Van der Waerden et al. only use variables with areal characteristics and none about personal behavior of the subjects.

In this research this earlier attempts will be used and expanded. At first a model using only the possible personal, trip or areal variables will be made and after this combinations of these levels are included to see whether the model improves. Also several variables can be implemented both as an areal variable or personal variable. In such a case both scales will be checked and then the most suitable scale will be used in the final model. The intention is to firstly make a multinomial regression model and after this the more complex hierarchic regression model.

Potential Models

Because several research methods are used to describe potential in the past it is not clear which method is the best. Therefore, to get the most out of this research, several models can be created with the use of different research approaches.

The dependent variable in this model will be the potential for replacing the car-trip by a bus-trip. This is one of the questions asked in the enquiry and therefor the answers are clear. Now this variable should be connected to the independent variables: the personal -, trip -, and areal characteristics. Because the dependent variable, public transport potential, is categorical, this can be done by the use of logistic regression. If the categorical variable has exactly two categories the analysis is called binary logistic regression, and when the outcome has more than two categories it is called multinomial logistic regression (Field 2013). The dependent variable 'Bus Potential' has four different categories, so the latter will be used.

If these kind of logistic regression models are used, the model does not take the multiple levels of characteristics into account, the personal and areal variables are both used in the same level. When this occurs several aspects of the reality are not captured in the model. Personal variables are all individual and on the same level, however the areal variables are also individual but there can be groups of individual clusters by the area they come from. To implement these multiple levels a multi-layered hierarchic regression model should be used like a hierarchic regression model.

No measurements to change the transport mode of the user will be taken into account in this model. The model will only identify potential bus users, regarding the characteristics of the circumstance. So, the two models described before do not take the measurements to switch potential into actual users into account.

DATA

The data used for this research is mainly from an enquiry about travelers' behavior from a case study performed in 2009. Next to this, other data was needed about areal and travel aspects to create the bus potential models. A data analysis was done to check whether the data fits for this research.

Enquiry Data
The most recent and comprehensive dataset is from 'verplaatsingsonderzoek Nuenen 2009' ('travel survey Nuenen 2009') presented by van der Waerden and Bérénos (Waerden & Bérénos 2010). This survey was performed in the context of the Transumo project 'Regionale Openbaar Vervoer Bereikbaarheid Consumentgericht (ROVBECO)' ('Regional Public Transport Accessibility Project') (Waerden et al. 2009).

The models were created by the use of multiple possible variables divided in four different categories. Two of these categories are extracted from the enquiry dataset, these categories refer to the personal and trip characteristics.

Areal Data
Next to the data of the questionnaire, data about the characteristics of the area were needed. This data was taken from open source databases like CBS, processed by Irias Informatiemanagement. Additional data about households was obtained from Bridgis, and data of bus services in an area came from Goudappel Coffeng combined with open data.

Nuenen only exists of four different four position ZIP code areas, from which three of the areas form the center and the rest of Nuenen is the fourth area. When using these four areas the level of detail is quite gross and little variance is present in the variables. Therefore these four position ZIP code variables were discarded and the research was done with another scale, six position ZIP code variables where sought and used. The two categories of variables from the areal data regard the areal and bus characteristics.

Data Analysis
The data was analyzed by the use of the golden standard, and it is not equally distributed as the actual population in Nuenen. However a sample can never be perfectly representative for the population, but in this case there are rather large differences. It can be doubted whether the data is a good fit for the area.

MODELS
To make a hierarchic regression model, first the variables that should be used had to be selected. This is done by the use of crosstabs which show the correlation between different variables and the use of multinomial logistic regression models. When the variables are picked using these methods, the chosen variables should be checked whether they fit the right assumptions.

Correlation Crosstabs
To study the correlations between the possible variables IBM SPSS Statistics was used. An analysis is done with the Descriptive Statistics Crosstabs. The Chi-square is the most important aspect in this analysis because this will tell whether the correlation is significant. If this is the case, it may be better not to use the two variables in the Crosstab together in the model because they could affect each other.

Almost all variables, over all categories, do correlate to each other. Therefore, it can be said that most variables cannot be combined in a model, for they correlate to each other. However quite a lot of variables, 19 out of 30, do significantly correlate to the dependent variable 'Bus Potential'. At first none of the variables regarding bus characteristic did correlate to the dependent variable, but after a new categorization also one of these variables seemed to have a significant effect on the 'Bus Potential'.

Multinomial Logistic Regression

After the correlation tests, multinomial regression models were made. First all individual variables were tested into a model. All continuous variables did get a warning regarding subpopulations with zero frequencies, this problem could be solved by the use of only categorical variables. All significant individual models were put together in integral models, however none of these models could be used because of even more warnings. The last attempt for a model was done by step by step models, these models were made by manually putting variables in the model. Only three models which combined multiple variable groups could be made, and none of these included areal variables. Therefore the aim of the research, to combine areal and personal characteristics into a model, was not achieved.

Findings

The final model has three variables: 'Family Size', 'Public Transport Subscription' and 'Transport Mode'. The first two variables are personal, the last variable regards trip characteristics. When analyzing the part-worth utility of all variable's categories, the model does not seem to be very stable. However the model does not violate any of the assumptions regarding multinomial regression models.

Due to lack of time and tools no hierarchic regression model was made.

CONCLUSIONS

Conclusions which can be made from the research, recommendations for future research and a discussion will be discussed in this final paragraph.

Conclusion

The methodology to develop a tool which determines and localizes potential bus users includes a model which regards the switching-behavior of travelers. This switching behavior can be extracted from personal variables, and areal variables are needed to link the potential to their location.

The final multinomial logistic regression model used three of these variables, regarding family size, public transport subscription and present transport mode. So, it can be concluded that these variables are significant for the estimation of the dependent variable 'Bus Potential'. However, the aim of the model was to combine personal, trip and areal characteristics into the model, this was not succeeded. Probably the dataset used in this research is not fitting for the model, at least not for a multinomial regression model.

Public transport is often seen as an important solution to reduce air pollution, because of the lower emissions per passenger than an average car has. In Nuenen 61.2 of the non-bus trips, had potential to be changed to a bus trip, this is a very big opportunity for the public transport sector if these kind of numbers can be found in whole of the Netherlands. Therefore, the air quality in the Netherlands can probably be improved by the stimulation of public transport use.

As encountered very early in the modelling process the four position ZIP code scale was too gross and could not lead to a significant model for the bus-potential. A smaller scale should be used, in this research the six position ZIP code areas were used. However none of the bus related characteristics did initially have a significant effect on the model. This could be because very little areas have bus stops when using the six position ZIP code. For this reason perhaps medium scales for the areal variables could be examined, like neighborhoods.

Besides the scale of the areal characteristics, the usage of areal characteristics in the same level as personal behavior was doubted. To solve this a hierarchical model could be used, however due to time constraints and lack of tools such a model was not made during this research. The lack of tools refers to the software available via the university. Statistical models are usually done in IBM SPSS Statistical, but this is no ideal program for hierarchic models. HLM will be a more fitting program, but could not be figured out in the short amount of time available. However the hierarchic regression model could still be a good solution to create a fitting model for both areal and personal variables. The warnings about missing data will be solved this way because, hierarchic regression models can estimate parameters by the use of the available data.

Altogether, the research did show signals which were detected before by Waerden et al. In this research for ‘Regionaal Openbaar Vervoer Bereikbaarheid Consumentgericht (ROVBECO)’ (‘Regional Public Transport Accessibility Consumer Oriented’), an attempt was made as well to realize a planning tool for bus potential. The model made by ROVBECO also contained only three significant variables (Waerden et al. 2009).

Recommendations

The first pitfall of this research was the missing of tools to perform a hierarchic regression model. So, the first recommendation will be to try to make such a model and see if more variables could be implemented in a model, and whether areal and personal variables could be combined in this way.

The second pitfall of this research was probably the dataset itself. It was from an enquiry which was done five years ago, the data was raw and many adjustments had to be made for the data to be useful in a model. After analyzing the data with the gulden standard showed the data did not really fit the actual population of the area. Next to this the bus potential was already specified, it was classified by whether a respondents thinks their non-bus trip could be done by bus. However, whether it is possible to take the bus, does not mean people will actually make this switch. Finally, the survey was held in Nuenen, a small village which can be referred to as a car town. Because the village lies between the cities of Eindhoven and Helmond, many people live in Nuenen, but travel to the city to work. Nuenen does not have a very extensive public transport service yet. Therefore, if the first recommendation about the hierarchical regression model did not result in a better model, the advice will be to disregard this dataset and use other data to make a model. However, it is very time consuming and expensive to make such an extensive dataset as used in this research.

Probably the primary recommendation when investigating bus potential in the future is the use of open data. Open data continues to extend and it is available, so why not use it? The largest open data set about travelers behavior is the before mentioned 'Onderzoek Verplaatsingen in Nederland (OViN)' (Travel Research in the Netherlands). Which is an annual research performed by the 'Centraal Bureau voor de Statistiek (CBS)' (Central Statistical Office from the Netherlands). However, the dataset does not individually discussed bus transport, but only bus/tram/metro. Next to this the research scale should be expanded, for example provincial scale, because of the limited amount of respondents.

Next to het OViN, new developments are made with open data in the field of traffic. November 2014 the first results of the 'Mobiliteitspanel Nederland (MPN)' ('Dutch Mobility Panel'), performed by 'Het kennisinstituut voor Mobiliteitsbeleid' ('Knowledge Institute for Mobility Policy'), the University of Twente and Goudappel Coffeng, were presented. This research does not only focus on present travel behavior, but aims to increase the understanding of factors which play a role in changing travel behavior. Therefore, this can be an opportunity to investigate switching-behavior towards public transport usage.

One of the main goals of the model was to combine areal and personal characteristics. However, multiple researches did not succeed in this. Maybe the areal characteristics should be disregarded and only personal characteristics should be considered.

The research is carried out because of an important social problem, air pollution. Therefore, the target to reduce private-car usage by the stimulation of public transport seems important enough to continue the research to develop a bus potential tool. Maybe the air quality problem will in the future not be as stressing as nowadays, because all transport continuously becomes more sustainable, and maybe in the future green transportation is achieved by the use of alternative fuels or electric driving. However, the stimulation of public transport can have multiple other positive effect on society. For example solutions for congestions, parking problems and livability of residential areas.

Finally can be concluded that this research did explore one path to make a bus potential model, the multinomial regression model. However many paths are still out there, ready to be discovered.

REFERENCES

European Environment Agency, 2011. Transport. Available at: http://www.eea.europa.eu/themes/transport/intro [Accessed May 5, 2014].

European Environment Agency, 2014. Very high air pollution levels across Western europe. Available at: http://www.eea.europa.eu/highlights/very-high-air-pollution-levels [Accessed May 5, 2014].

Field, A., 2013. Logistic Regression. In *Discovering Statistics Using IBM SPSS Statistics*. London, United Kingdom: Sage Publications Ltd, pp. 760 – 813.

Health Effects Institute, 2010. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*, Boston, Massachusetts, USA. Available at: pubs.healtheffects.org/getfile.php?u=553.

Sugiki, N., Furusawa, K. & Aoshima, N., 2001. Location Behavioral Model Considering Transit Characteristics in Life Stage of Households. In *9th World Conference of Transportation Research, Seoul, Korea*. Gunma, Japan: Gunma University.

Waerden, P. van der & Bérénos, M., 2010. *Houd Nuenen Bereikbaar; Verplaatsingsonderzoek Nuenen 2009*, Eindhoven, Netherlands.

Waerden, P. Van Der, Bérénos, M. & Tromp, H., 2009. Regionaal Openbaar Vervoer Bereikbaarheid Consumentgericht. In *Colloquium Vervoersplanologisch Speurwerk*. Antwerpen, Belgium.

Waerden, P. Van Der, Timmermans, H. & Bérénos, M., 2005. In search of the Public Transport Users : Towards Public Transportation Potential Maps. In *Tranportation Research Board*. Washington D.C., USA.

Zhou, Y. et al., 2004. Transit Customers – Who , Why , Where , and How : A Market Analysis of the San Mateo County Transit District. In *83rd Annual Meeting of the Transportation Research Board and Publication in the Transportation Research Record.* Chicago, USA. Available at: http://www.camsys.com/pubs/SAMTRANS_Market_Analysis.pdf.

H. (Hannely) Hortensius, BSc.

This thesis is the result of a six month graduation project carried out in collaboration with Eindhoven university of Technology and Bonotraffics B.V., innovative traffic consultancy. This report represents also the end of my five and a half years of studying at the TU/e, beginning with the bachelor program 'Bouwkunde' and finalized with the master program of 'Construction Management and Engineering'. The first I started for it combined technical and creative aspects, but it resulted in a master which concerns the process and project management of construction projects. I do not know yet where my career will take off and where it will lead me, I only hope I will keep doing what I love to do, and that I can keep learning all the way.

Educational Curriculum Vitae

2002 – 2009 VWO, 'Natuur en Techniek', CSG Het Noordik
2009 – 2013 Bachelor 'Bouwkunde', Management and Urban Design, TU/e
2013 – 2015 Master 'Construction Management and Engineering', TU/e
2014 – 2015 Graduation Internship, BonoTraffics B.V.

Nederlandse Samenvatting

POTENTIEEL GEBRUIK VAN HET OPENBAAR VERVOER IN HET KADER VAN VERMINDERDE LUCHTVERONTREINIGING |
Ontwikkelen van een tool om potentiële bus gebruikers te bepalen en lokaliseren.
H. (Hannely) Hortensius

Afstudeerrichting:
Construction Management and Urban Development 2014-2015

Afstudeercommissie:

Prof. dr. ir. W.F. (Wim) Schaefer	|	Hoofd Master Programma CME, TU/e
Dr. ing. P.J.H.J. (Peter) van der Waerden	|	Eerste afstudeerbegeleider, TU/e
Dr. Q. (Qi) Han	|	Tweede afstudeerbegeleider, TU/e
Ir. J.J.F. (Joris) Hoogenboom	|	Adviseur, BonoTraffics bv.

Afstudeerdatum:
25-02-2015

Samenvatting
De luchtvervuiling in Europa verslechterd en hiervoor moeten maatregelen worden ondernomen. Meer dan één vijfde van de uitstoot van broeikasgassen is te wijten aan het verkeergebruik. Om deze reden zal dit onderzoek een poging doen de effectiviteit van het vervoer te optimaliseren door openbaar vervoer gebruik aan te moedigen. Dit zal worden gedaan door het ontwikkelen van een tool die de aanwezigheid en locatie van potentiële gebruikers van het openbaar vervoer schat. De uiteindelijke tool zal kunnen worden toegepast, door OV-bedrijven of verkeerskundig adviesbureaus, om het bestaande busnetwerk te optimaliseren of om toekomstige netwerken in nieuwe wijken te voorspellen.

Trefwoorden: Luchtkwaliteit; Milieu & Gezondheid; Openbaar Vervoer; Geo-Marketing; Multinomiale Logistische Regressie Modellen.

INTRODUCTIE
In West-Europa worden zeer hoge luchtvervuilingsniveaus gemeten, en aangezien de behoefte voor vervoer blijft groeien worden nog hogere niveaus verwacht inde toekomst. Kort geleden hadden vooral Frankrijk, België en Duitsland te lijden onder zeer slechte luchtkwaliteit, maar als er geen maatregelen worden getroffen zal Nederland zeker hun voetsporen volgen (European Environment Agency 2014).

Motorvoertuigen zijn een zeer belangrijke bron van stedelijke lucht vervuilingen (Health Effects Institute 2010). In 2011 was één derde van alle energieverbruik opgegaan aan transport, dit verkeergebruik was verantwoordelijk voor meer dan één vijfde van de uitstoot van broeikasgassen (European Environment Agency 2011). Dus een optie om het probleem van de luchtvervuiling aan te pakken is het aanpassen van het vervoersgebruik.

Probleemstellig en Onderzoeksvragen
Er zijn al veel artikelen geschreven over de houding van reizigers, maar toch hebben openbaar vervoerders nog weinig tot geen inzicht van de potentie die hun knooppunten geven. Belangrijke vragen voor deze bedrijven zijn of ze alle mogelijke reizigers al bereiken, waar

potentiële reizigers gevestigd zijn en hoe gevoelig klanten zijn voor verschillende faciliteiten. Daarom richt dit onderzoek zich op de houding van reizigers ten opzichte van vervoerswijzen, de effecten van verschillende maatregelen en potentiële reizigers in een bepaald gebied.

Het doel van het onderzoek is het ontwikkelen van een tool die OV-bedrijven meer informatie verschaft over potentiele klanten, zodat ze hierop kunnen anticiperen en op deze manier autogebruik kan worden verminderd. De onderzoeksvragen die dit doel ondersteunen zijn:

“Welke onderzoeksmethode moet worden gebruikt om een tool te ontwikkelen die potentiële bus gebruikers bepaald en lokaliseert?”

"Welke persoonlijke, reis, en gebieds-kenmerken moeten worden gebruikt om potentiële gebruikers van het openbaar vervoer te modelleren?''

"Kan de luchtkwaliteit in Nederland worden verbeterd door het stimuleren van het gebruik van openbaar vervoer?"

Om deze hoofdvragen te kunnen beantwoorden zijn meerder sub vragen opgesteld met betrekking tot de definitie van potentie, gedrag van reizigers en maatregelen die reizigers kunnen aantrekken tot het openbaar vervoer.

ONDERZOEKSMETHODE

Eerdere pogingen om openbaar vervoer te modelleren gingen bijna altijd over het feitelijke gebruik en niet over het potentiele gebruik. In 2005 hebben Van der Waerden et al. Een poging gedaan om openbaar vervoer potentie kaarten te maken (Waerden et al. 2005). Echter worden in dat onderzoek alleen gebiedskenmerken gebruikt in het model. In dit onderzoek zullen eerdere pogingen gebruikt en uitgebreid. Eerst zal getracht worden een model getracht te maken met enkel persoonlijk, reis of gebiedskenmerken. Hierna zullen combinaties worden getest. De bedoeling is te beginnen met multinomiale regressie modellen en daarna naar de meer complexe hiërarchische regressiemodellen toe te werken.

DATA

De data gebruikt in het onderzoek komt van het ‘Verplaatsingsonderzoek Nuenen 2009’, gepresenteerd door Bérénos en van der Waerden (Waerden & Bérénos 2010). Dit onderzoek werd uitgevoerd in het kader van het Transumo project ‘Regionale Openbaar Vervoer Bereikbaarheid Consumentgericht (ROVBECO)’. De persoonlijke en reiskenmerken uit de modellen worden uit de enquête van dit onderzoek gehaald. Naast de gegevens van de vragenlijst waren er ook gebiedskenmerken nodig om de gevonden bus potentie aan een locatie te koppelen, deze zijn verkregen via open data bronnen.

Data Analyse

De verkregen data is geanalyseerd met gebruik van de gouden standaard, de data is helaas niet hetzelfde verdeeld als de werkelijke populatie in Nuenen. Echter is een steekproef nooit perfect representatief voor een bevolking, maar in dit geval zijn er wel vrij grote verschillen. Er kan worden betwijfeld of de gegevens genoeg representatief zijn voor het gebied.

MODELLEN

Allereerst zijn de correlaties tussen alle mogelijke variables getest met behulp van IBM SPSS Statistics, de analyse werd gedaan aan de hand van de Chi-square test. Wanneer twee variabelen significant met elkaar correleren is het beter deze variabelen niet samen in een model te stoppen, omdat deze effect op elkaar kunnen uitoefenen.

Uit de correlatie kruistabellen werd geconcludeerd dat bijna alle variabelen, over alle variabel groepen, met elkaar correleren. Hieruit ka worden opgemaakt dat de meeste variabelen niet gecombineerd kunnen worden in een model. Naast dat de variabelen met elkaar correleren, laten veel variabelen, 19 van de 30, ook een significante correlatie zien met de 'Bus Potentie'.

Multinomiale Logistische Regressie

Na de correlatie tests zijn multinomiale logistische regressie modellen gemaakt. Dit model werd gebruikt omdat de uitkomstvariabele, 'Bus Potentie', vier verschillende uitkomst categorieën heeft. Eerst zijn alle individuele variabelen in modellen gedaan. Echter kregen alle continue variabelen een waarschuwing met betrekking tot sub bevolkingen zonder waarde, dit probleem kon worden opgelost door het gebruik van enkel categorische variabelen. Alle significante variabelen zijn samen in integrale modellen gestopt, geen van deze modellen kon worden gebruikt door het opdoemen van nog meer waarschuwingen. De laatste poging tot het krijgen van een werkend model was het maken van stap voor stap modellen, deze modellen zijn gemaakt door het handmatig toevoegen van variabelen in een leeg model. Drie modellen konden worden gemaakt welke combinaties van variabel groepen gebruiken, geen van deze modellen gebruikt de gebieds-variabelen. Hierdoor is het doel van dit onderzoek, om persoons- en gebiedskenmerken te combineren in een model, niet bereikt.

Bevindingen

Het uiteindelijke model heeft drie variabelen: 'Gezinsgrootte', 'Openbaar Vervoer Abonnement' en 'Vervoerswijze'. De eerste twee variabelen zijn persoonlijk, de laatste is een reis kenmerk. Bij het analyseren van het deelnut van alle categorieën per variabele lijkt het model niet stabiel te zijn. Echter schendt het model geen van de aannames die gemaakt zijn voor het gebruik van multinomiale regressie modellen.

Wegens gebrek aan tijd en middelen is er geen hiërarchisch regressie model gemaakt.

CONCLUSIE EN AANBEVELINGEN

De onderzoeksmethode om een tool te ontwikkelen die bus potentie bepaald en lokaliseert omvat een model die wisselings-gedrag schat. Dit gedrag kan worden afgeleid van persoonlijke kenmerken, gebieds-kenmerken zijn nodig om de potentie aan een locatie te koppelen.

Het uiteindelijke model bevatte drie variabelen, waarvan geen een gebiedskenmerk was, hierdoor is het doel van het model om potentie te lokaliseren niet geslaagd. Waarschijnlijk paste de dataset niet goed genoeg bij het onderzoek. Echter kan nog wel geprobeerd worden een hiërarchisch regressie model te maken met deze dataset, omdat hiërarchische modellen minder problemen hebben met missende data.

Als met de dataset ook geen werkend hiërarchisch regressie model gemaakt kan worden is het advies om te stoppen met deze dataset en andere data te verkrijgen. Echter is het maken van een nieuwe dataset een dure en tijdrovende bezigheid. Als dit toch wordt gedaan is het advies een gebied te kiezen met een gevarieerd openbaar vervoer aanbod, een digitale enquête te maken waar respondenten hun reizen op aan kunnen geven en de definitie van bus potentie te heroverwegen.

De belangrijkste aanbeveling is waarschijnlijk om in de toekomst open data bronnen te gebruiken. Momenteel is de grootste open dataset over reizigersgedrag van het 'Onderzoek Verplaatsingen in Nederland (OViN)', een jaarlijks terugkomend onderzoek uitgevoerd door 'Centraal Bureau voor de Statistiek (CBS)'. Voor een kleinschalig onderzoek zijn er echter te weinig respondenten, maar op grotere schaal kan deze dataset goed worden gebruikt. Naast het OViN zijn er ook andere ontwikkelingen op het gebied van verkeerskundige open datasets, zoals het 'Mobiliteitspanel Nederland (MPN)'.

Al vroeg in het onderzoek bleek dat een vier positie postcodegebied te grof was om te gebruiken, in dit onderzoek is daarom voor zes positie postcodegebieden gekozen, In de toekomst is het aan te raden een tussenschaal te kiezen, omdat in dit geval in veel gebieden geen bushalte aanwezig was.

Al met al, heeft dit onderzoek signalen laten zien die eerder ook al waren gedetecteerd door van der Waerden et al, in het ROVBECO onderzoek (Waerden et al. 2009). Er kan worden geconcludeerd dat dit onderzoek een pad heeft doorlopen en afgevinkt over het hoe te maken van een bus potentie model, namelijk het pad van het multinomiale regressie model. Echter, vele paden zijn nog niet belopen, klaar om ontdekt te worden.

REFERENTIES

European Environment Agency, 2011. Transport. Available at: http://www.eea.europa.eu/themes/transport/intro [Accessed May 5, 2014].

European Environment Agency, 2014. Very high air pollution levels across Western europe. Available at: http://www.eea.europa.eu/highlights/very-high-air-pollution-levels [Accessed May 5, 2014].

Health Effects Institute, 2010. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*, Boston, Massachusetts, USA. Available at: pubs.healtheffects.org/getfile.php?u=553.

Waerden, P. van der & Bérénos, M., 2010. *Houd Nuenen Bereikbaar; Verplaatsingsonderzoek Nuenen 2009*, Eindhoven, Netherlands.

Waerden, P. Van Der, Bérénos, M. & Tromp, H., 2009. Regionaal Openbaar Vervoer Bereikbaarheid Consumentgericht. In *Colloquium Vervoersplanologisch Speurwerk*. Antwerpen, Belgium.

Waerden, P. Van Der, Timmermans, H. & Bérénos, M., 2005. In search of the Public Transport Users : Towards Public Transportation Potential Maps. In *Tranportation Research Board*. Washington D.C., USA.