Technical University of Munich

Chair of Transportation Systems Engineering

Master’s Thesis

Who Does New Trips and Why? -An Analysis Towards the Modelling of Induced Demand

Author:

Vasiliki Kalliga

Supervised by:

M.Sc. Santhanakrishnan Narayanan

Univ.-Prof. Dr. Constantinos Antoniou

July 1, 2021

Abstract

Improvements on travel conditions and infrastructure capacity influence the travel demand. More specifically, these improvements induce additional growth in traffic. Although, this can occur through a variety of behavioral mechanisms, in this study, the main focus is on the generation of new trips. The objective of this thesis is to design and implement a methodological approach to identify and model the induced demand. The first step of the proposed methodology is to identify the factors that influence the trips, which have occurred only once. To achieve the first step, discrete choice models, including both multinomial logit model (MNL) and nested logit model (NL), are developed based on a household travel survey from the city of Madrid. The trip frequencies are considered as the dependent variable. Findings reveal the importance of the sociodemographic characteristics on the decision of making a new trip. Factors such as age, household size, vehicle ownership, and number of children were strongly influential. There was also an indication that new trips are mainly taking place by taxis and shared mobility, while they tend to consist of more than one stages. The second step will utilize the factors identified from the discrete choice models to distinguish the induced and then non-induced demand. The third and final step of the methodology is to formulate a model for induced demand, which can be integrated with an existing transport model.

Acknowledgments

The completion of this thesis would not have been possible without the support and help I got during this period. I would therefore want to express my sincere gratitude to both my supervisors, Santhanakrishnan Narayanan and Prof. Dr. Constantinos Antoniou, who always guided me and provided me great support and inspiration throughout my thesis. I am very thankful for their guidance in all steps of my work, and the knowledge they provided me in the field of modeling. A special thanks to Santhanakrishnan Narayanan for his willingness to supervise my thesis, his patience and support, and his ideas and suggestions. This research would not have been possible without their interest in the topic and willingness to help and guide me.

Finally, to my family and friends, I am grateful for the unconditional love they showed me, their support and for just believing in me.

Table of Contents

1 Introduction 1
1.1 Induced travel demand 1
1.2 Problem statement and research motivation 4
1.3 Research questions and objectives 5
1.4 Expected contributions 5
1.5 Thesis outline 6
1.6 Report structure 7
2 Literature Review 8
2.1 Induced demand 8
2.1.1 Induced demand caused by infrastructure improvements 8
2.1.2 Induced demand caused by a new transport mode 10
2.2 Analysis Methods 12
2.2.1 Discrete Choice Modeling (DCM) 12
2.2.2 Nested Logit Models (NL): 14
2.2.3 Maximum Likelihood Estimation (MLE) 16
2.2.4 Multicollinearity 17
2.2.5 Model evaluation techniques 18
2.2.6 Statistical test 18
2.2.7 Classification metrics 19
2.2.8 Outlier detection 21
2.2.9 Imbalanced data 21
3 Methodology 22
3.1 Three-step proposed methodology 23
3.2 Implementation of first step 24
3.2.1 Data preparation for model development 25
3.2.2 Behavioral modeling 25
4 Data Analysis 28
4.1 Data source and descriptive statistics 28
4.2 Descriptive analysis 31
4.3 Qualitative analysis 45
5 Model Estimation 46
5.1 Variables considered 46
5.2 Behavioral Modelling 48

5.2.1 Independent variable models ... 48
5.2.2 Models with user's profiles ... 54
6 Discussion ... 61
6.1 Discussion of main results ... 61
6.1.1 Descriptive analytics findings ... 61
6.1.2 Model findings ... 62
7 Conclusion and Future Work ... 66
7.1 Conclusions ... 66
7.2 Limitations and future work ... 67
Bibliography ... 68

List of figures

Figure 1.1 Illustration of traffic theory and traffic reality (Speck, 2018)................................. 2
Figure 1.2 Induced demand's vicious cycle 4
Figure 1.3 Thesis outline.. 6
Figure 2.1 Random split of train and test set (Gareth, Witten, & Hastie, 2013).................... 18
Figure 2.2 K-fold cross-validation (Gareth, Witten, & Hastie, 2013)...................................... 18
Figure 3.1 Methodological Framework 22
Figure 3.2 Implementation of first step 24
Figure 4.1 Frequency by gender... 32
Figure 4.2 Trip frequency by age group.. 33
Figure 4.3 Frequency by nationality 34
Figure 4.4 Frequency by education level... 35
Figure 4.5 Frequency by HH size 36
Figure 4.6 Frequency by number of children in the household 37
Figure 4.7 Frequency by vehicle ownership.. 38
Figure 4.8 Frequency by transport mode.. 39
Figure 4.9 Frequency by licence and PT car availability 40
Figure 4.10 Frequency by month 41
Figure 4.11 Frequency per trip reason 42
Figure 4.12 Trip frequency per profile 43
Figure 4.13 Trip frequency per distance.. 44
Figure 5.1 Pearson's Correlation 47
Figure 6.1 Configuration of induced demand 63
Figure 6.2 Induced demand... 65
Figure A.1 Age distribution of total sample per frequency level 73
Figure A.2 Age distribution of shared mobility users... 73
Figure A.3 Trip frequency per gender and age group of shared mobility users 74
Figure A.4 Trip frequency per PT card and driving licence availability of shared mobility users 75
Figure A.5 Trip frequency per hh size and number of children of shared mobility users....... 76
Figure A.6 Trip frequency per number of vehicles owned by shared mobility users 76
Figure A.7 Trip frequency of shared mobility trips per month 77
Figure A.8 Trip frequency per shared mobility users profiles 77

List of tables

Table 4.1 Distribution of dependent variable 29
Table 4.2 Summary of individuals, household and trip characteristics................................. 29
Table 4.3 Profile description.. 42
Table 5.1 Cramer's V correlation matrix... 46
Table 5.2 Estimation Results MNL Model 1.1 and NL 1.2 48
Table 5.3 Confusion Matrix-Model 1.1.. 52
Table 5.4 Accuracy Rate-Model 1.1... 52
Table 5.5 Model 1.1 performance on individual classes 52
Table 5.6 Confusion Matrix-Model 1.2.. 54
Table 5.7 Accuracy Rate-Model 1.2... 54
Table 5.8 Model 1.2 Performance on Individual Classes.. 54
Table 5.9 Estimation Results MNL Model 2.1 and NL 2.2 55
Table 5.10 Confusion Matrix Model 2.1 58
Table 5.11 Accuracy Rate-Model 12.1.. 58
Table 5.12 Model 2.1 Performance on Individual Classes.. 58
Table 5.13 Confusion Matrix-Model 2.2.. 59
Table 5.14 Accuracy Rate-Model 2.2... 59
Table 5.15 Model 2.2 Performance on Individual Classes.. 59
Table 5.16 Evaluation of Models 60

List of Abbreviations

AIC	Akaike Information Criterion
DCM	Discrete Choice Modeling
HH	Household
IIA	Independence of Irrelevant Alternatives
IID	Independent and Identically Distributed
km	Kilometers
m	Meters
NC	Not Considered
NL	Nested Logit
OLM	Ordered Logit Models
PT	Public Transport
VMT	Vehicle Miles Travelled
UAM	Urban Air Mobility

1 Introduction

This chapter begins with a description of the thesis background followed by problem statement, research questions and contribution. Finally, thesis outline and the structure of the report are summarized.

1.1 Induced travel demand

Transport engineers tend to compare traffic to a fluid, assuming that a particular volume should flow through the road system. According to this assumption, total demand in transportation planning is forecasted based on exogenous factors such as land use, population, income and employment. When future demand is estimated based only on those factors through the application of a model, it is assumed that the total demand is influenced by neither transportation infrastructure nor new transportation modes, but is determined entirely by exogenous factors.

However it has been proved that it is more proper to compare traffic to a gas that expands to fill available space (Jacobsen, 1997). Traffic congestion tends to maintain equilibrium, with traffic volumes increases to the point that congestion delays discourage additional peak-period vehicle trips. This contrasting approach claims that additional capacity stimulates corresponding increment in demand which suggest that there are willing commuters who will express their demand for travel once the service is offered. In growing urban areas, the evidence from recent decades seemed to support this interpretation (Douglass, Klein, & Camus, 1999).

A trip can be described by a number of travel conditions, including the safety of the trip, comfort, reliability, frequency of a service, etc. In general, induced demand can be defined in terms of additional trips that would be made if travel conditions improved (less congested, lower vehicle costs or tolls). Modelling induced travel demand is a complex task due to the involvement of a high number of variables, which make the analysis complicated and difficult to generalize (Cascetta, 2009).

A transportation system consists of a set of elements interconnected by complicated relationships, such as supply sub-system, demand sub-system, residences and activities sub-system. As a consequence, whenever an action is planned on a part of a transportation system, there are unavoidable impacts on other parts, either positive or negative. Improvement within the supply sub-system, such as the introduction of a new road infrastructure providing faster or even cheaper services, more comfortable mobility service, or in general actions that increase the utility or the satisfaction of the commuters about the possibility of moving, contribute to the creation of additional travel demand (Cascetta, 2009).

According to traffic theory, in congested systems, the main constraint to driving is congestion (Speck, 2018). The question is not whether roads will be congested at rush hour, but how many lanes of congestion we are willing to have. It has been shown that for every new mile roadway that is constructed will be typically 40% filled up with new trips immediately, and totally full within four years (Salzman, 2010). His typical illustration of traffic theory is presented in the Figure 1.1: Congestion occurs when traffic (yellow line) outpaces capacity. Widening the road absorbs the extra trips. While the traffic reality indicates that the elimination of congestion induces people to drive more, and congestion returns immediately (Speck, 2018).

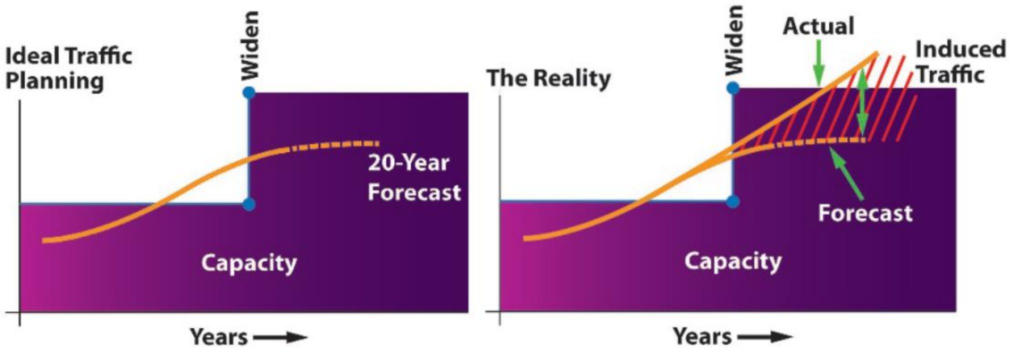

Figure 1.1 Illustration of traffic theory and traffic reality *(Speck, 2018)*

Expanding congested roads attracts *latent demand*, trips from other routes, times and modes, and encourages longer and more frequent travel. This is called *generated traffic*, and refers to additional peak-period vehicle traffic on a particular segment. It partially consists of *induced demand*, which refers to absolute increases in vehicle miles traveled (VMT) compared with what would otherwise occur (Benjamin, 2018). Generally, any intervention on a given transport section, leads to a number of induced demand related impacts, which can be split into three parts (Gorham, 2009):

I. Direct induced demand: It is generally refers to *conscious decisions* by travelers, to take advantage of changes in travel time by alternating travel patterns in a way that increases overall traffic. These decisions, in alternating travel patterns, can take place over different time frames: ***Instantaneous traveler responses***, with the travelers diverting their transportation behavior immediately (temporal, spatial or modal convergence). It can also be ***Short-run traveler responses***, where travelers alternating their trip-making patterns by either changing their trip destination, trip chaining patters or trip frequency. Finally, decision can refer to ***Long-run traveler responses***, with the commuters changing the locations where they live or work in order to take advantage of reaction of travel time. Household might be relocated, Employees might change their work location, developers might reorient or accelerate development of residential sub-visions, industrial and technology parks in outlying areas. In all of these cases, people respond to a perceived change in travel time by changing their base of operations within the metropolitan area.

II. Indirect induced demand: It is the increment in transportation activity caused by a particular transportation improvement and it occurs by the following mechanisms: ***Network effects:*** Travelers decisions are taken in response to other people's collective responses to changes in point-to-point travel time. These networks effects can be complex to represent in modelling. An example of network effect is the decision of a traveler to use an arterial, when traffic congestion along the specific arterial is alleviated in response to the opening of a new, parallel motorway segment. Indirect induced demand can also occur by: ***Lifestyle effects****:* They mainly refer to decision by travelers to make changes in response to transportation improvement that actually amount to lifestyle changes such as changes in vehicle ownership or in residential location. The desire to take advantage of a reduction in travel time by car may induce some households to purchase additional vehicles in order to substitute travels that previously took place by other modes. However, by owning a car that a household otherwise might have chosen not to buy, household members might make additional trips by the specific transport mode that are not related to the purpose for which the car was purchased, thus resulting in additional trips that are induced indirectly. The effects on the residential location is based on the theory of the classical urban economics, according to which there is a trade-off between housing and transportation costs (Alonso, 1964). There is a correlation between transportation and land cost. As a consequence, changes in point-to-point travel time, can cause shifts in land values in different parts of the city. Households could be relocated in response to these changes, so that the balance between land and travel cost is the same. Any change in the distance traveled per trip as a result of this relocation might be considered an indicator of direct-effect induced travel. Finally, indirect induced demand can occur by*:* ***Market effects***: With the same way that transportation investments can change households' calculations for accessibility and balance of land-values, it is also possible that they will affect strategic market calculations for firms. The effect of these strategic decisions can constrain households into travelling farther that they in other case would have, had such strategic decision not been induced by infrastructure changes.

III. Induced and Diverted demand: Based on researchers, induced and diverted demand corresponds to the traffic that might have occurred anyway, either at some other place or at some other time of the day on a specific network.

Summing up, induced demand is any increase in travel resulting from improved travel conditions. The vicious cycle of induced travel demand is presented in Figure 1.2.

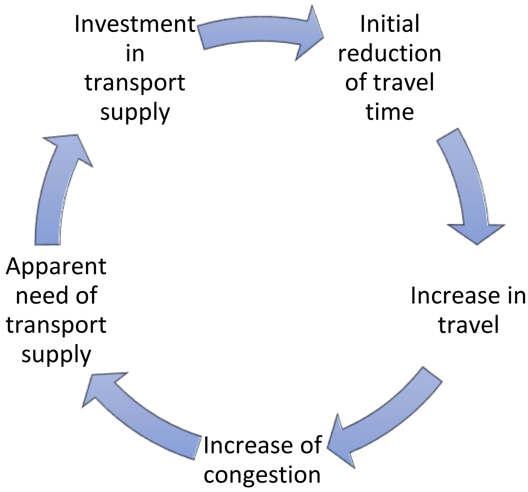

Figure 1.2 Induced demand's vicious cycle

In the vast majority of the cases, where induced demand is concerned, there is an association with travel time. That could be either a reduction in travel time, an improvement in the reliability of travel time, or both. Improved travel times are attractive to the travelers and as consequence the travel is increased. The immediate effect of this situation is the increment of congestion which leads to an apparent need of transport supply. The investment in transport supply forces the vicious cycle to start from the beginning (Gorham, 2009).

1.2 Problem statement and research motivation

Urban mobility is described as the lifeblood of modern cities, a critical economic factor and a key aspect of smart and sustainable development. Planning a smart city that delivers effective and equitable urban mobility solutions is one of the most crucial problems for cities throughout the world. Smart cities must deliver effective smart mobility solutions while encouraging innovation, facilitating a collaborative ecosystem, and meeting sustainability goals. The aforementioned challenges are part of the rapidly changing landscape of urban mobility. Two of the most well-stablished strategies that can contribute to the solution of the urban mobility problems is the improvement of infrastructures and the adoption of innovative vehicle types such as shared vehicles (Glasco, 2019).

However these strategies lead to increment in the number of trips occurring in the network. New transportation capacity induces increased travel, both due to short run effects and long rung changes in land use development patterns (Loop, Haaijer, & Willigers, 2016). On the other hand, shared mobility and microbolity offer a very comfortable, easy, quick and effective transportation choice. Therefore they can be even considered as one of the causes of the increment of vehicle kilometres travelled in a city (Glasco, 2019).

Modelling induced demand is a challenging topic, because it is difficult to estimate it with conventional models, as generation-distribution sub-models of four-step model are not able to capture the additional trips. In addition, they are not sensitive to changes in the level of service and as a consequence they are not able to capture the related effects.

The need to understand the proportion of demand that is induced, focusing on the characteristics that the new trips in a network have, acts therefore as a research motivation to study the factors associated with it.

1.3 Research questions and objectives

The main objective of this work is given a household survey to identify and model induced demand. In order to do so, a methodology should be developed with target to find out which are the influential factors associated with decision of making a new trip.

This leads us to the following research questions associated with the study:

- Is it possible to build a robust model for the identification of the characteristics of new trips based on a data set consisting of real data?
- Can characteristics of new trips be modelled using **Discrete Choice Modelling** ?
- **Multinomial Logit** or **Nested Logit models** could capture better the characteristics of the new trips?
- Can the addition of **user profiles** improve the fit of the models?
- Can the obtained factors be used to distinguish the induced from the non-induced demand?

To answer the research questions, a detailed methodology will be introduced in Chapter 3. The objective of this methodology is to classify the characteristics of trips that are taking place for first time and find out to which extend these trips contribute to the generation of induced demand.

1.4 Expected contributions

In line with our main objectives and research questions, this thesis is expected to make the following contributions:

- Methodological contributions:

  - Proposing a methodology including multiple steps to identify and model of induced demand.
  - Using the trip frequency as dependent variable to build MNL and NL models, in order to identify the significant variables for the decision of making a new trip.

- Proposing a framework for future work using the significant estimated coefficients of the discrete choice models as an input.

- Practical contributions:

    - Discussing the factors affecting the decision of making a new trip, followed by an elucidation on the insights obtained.

1.5 Thesis outline

The thesis outline is summarised in Figure 1.3.

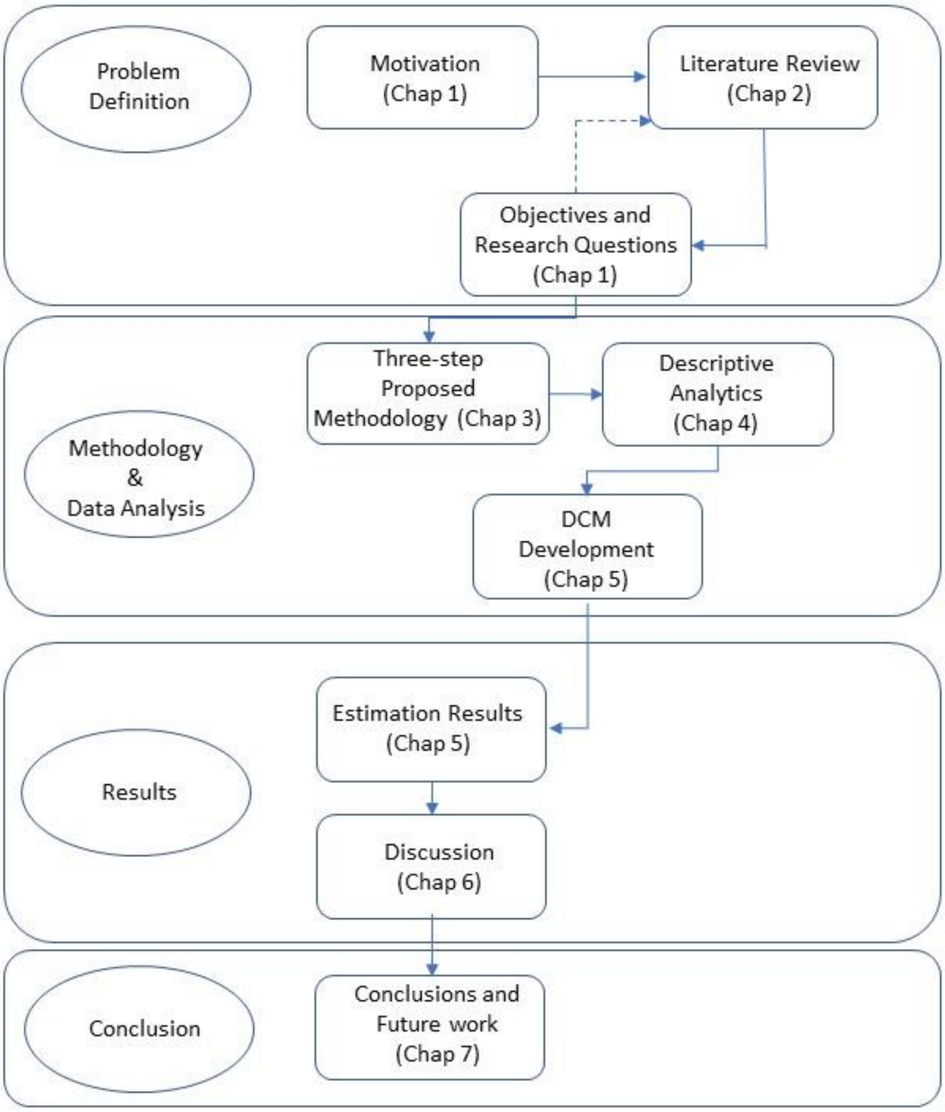

Figure 1.3 Thesis outline

1.6 Report structure

In line with the defined objective and research problems, and following the thesis framework, this report is organised as follows. The literature report will be first presented, including an overview of the reasons the can cause induced demand, and analysis methods in modelling. After that, the proposed methodology is explained, comprising both the analysis methods and model developments. Later, the descriptive statistics of the data set is presented. Thereafter, the model results elaborate on the different methods used, followed by a discussion including recommendations, limitations, but also insights for future work is given.

2 Literature Review

In this chapter, scientific articles concerning the topics of this thesis are presented and analysed. As we aim to identify the induced demand, a better understanding of induced demand in first required, followed by an extensive review of studies tackling influential factors for the aforementioned topic. These include induced demand caused by either improvements in infrastructure or by the introduction of a new transport mode. The chapter also includes relevant analysis methods used in characteristics identification studies. The overall aim is to summarise the characteristics of induced demand trips and project them to the current study.

2.1 Induced demand

2.1.1 Induced demand caused by infrastructure improvements

According to research studies the improvement of infrastructure capacity leads to increment of travel demand. The congestion reduction due to such capacity improvements is likely to be overvalued if induced demand is not considered. These effects have been measured in many road enhancement projects.

One of the earliest studies in the field of induced demand demonstrates how induced demand can be estimated for the incorporation into the evaluation process for highway expansion project (Decorla-Souza & Cohen, 1999). The study refers to a hypothetical freeway expansion analysis and the results showed that the magnitude of travel induced by highway expansion increases significantly as a relation of initial congestion levels before the expansion. Though, even under extreme cases of initial congestion and consequent forecasted induced travel, there is a positive impact regarding the congestion relief.

Since then, numerous studies using various analysis methods have quantified induced travel impacts. Cervero used data from California between 1980 and 1994 regarding freeway expansion, traffic volumes, demographic and geographic factors (Cevero, 2003). His research focused on the long-term elasticity of vehicle-mile-travelled (VMT) with respect to travel speed and resulted that 10% increase in speed results in a 6.4% increase in VMT, and that about a quarter of these results from changes in land use. The results show that approximately 80% of additional roadway capacity is filled with additional-peak period travel, a significant percentage of which (39%) can be considered the direct result of the added capacity.

Some researchers found that increase in road space or traffic signal control systems which smooth traffic flow induced additional vehicle traffic which quickly diminished any initial emission reduction benefits (Noland & Quddus, 2006). Based on another interesting study in which, the relationship between interstate highway lane kilometers and highway vehicle-kilometers travelled (VKT) in the US was investigated, they conclude that VKT increases according to highways and they specify three important sources of this extra vehicle travel including the increased driving by current residents, and inflow from current residents, and higher transport intensive production activity (Duranton & Turner, 2008). The type of data that was used is time-series, obtained for various roadway types, and the result indicate an elasticity of vehicle travel with respect to lane miles of 0.5 in the short run and 0.8 in long run. As a consequence, half of the increased roadway capacity was filled with added travel within about five years, and that 80% of the increased roadway capacity will be filled eventually. Litman's contribution (Litman, 2021) provides an analytical and comprehensive literature review on the importance of evaluating induced demand brought by road transport. One more interesting review study show that 50% - 80% of increased highway capacity was soon filled with generated traffic (Small, 1992).

Based on a comprehensive study of the impacts of urban design factors on US vehicle travel found that a 10% increase in urban road density, measured in lane-miles per square mile, increased per capita annual VMT by 0.7% (Lawrence, 2000). It is notorious in all the above mentioned studies, how induced traffic had a strong impact on the expectations of the project. The majority of the benefits, especially the ones regarding capacity increase, have been strongly restricted due to filling by induced traffic. For the projects researching induced demand caused by the improvement of infrastructure, it is important to take into account an induced demand estimation model withing the decision-making process, for which due to the great amount of factors playing (such as kind of roadway, land use, regional or urban area, short or long run effect), the appropriate adaption is essential.

Other researchers developed a model for induced demand resulting from the introduction of a HS service linking the cities of Japan namely Osaka, Nagoya and Tokyo (Yao & Moriikawa, 2005). They came up with an integrated intercity travel demand model with nested structure, including all the steps of the Four-Step-Model (trip generation, destination choice, mode choice and route choice models). The estimation of the induced demand was done by introducing an accessibility measure, as an expected maximum utility possible to capture the short run behavioral effects such as changes in the travel departure times, routes switches, modes switches, longer trips, changes of destination, and new trip generation. Additionally, elasticities of induced travel (trips and VMT) based on fares, travel time, access time and frequency of service for business and non-business trips were calculated.

Based on another study it was find out that induced demand is resulting from higher design speeds and, as a consequence by less travel time, for the High Speed 1 in UK (Pagliara & Preston, 2013). They supported that, when the improvement of the transportation system is expected to have generation effects, is more appropriate to be applied a gravity model to all O/D pairs. Since historical data was difficult to be found, they used information provided from the office of Rail Regulation (0RR), the independent safety and economic regulator for Britain's railways. They resulted that an improvement in the level of service is reflected in a customer's satisfaction increase and the related elasticities indicates that 1 unit increase results in a 3.15% increase in the number of trips.

According to a literature review study of evidence for induced demand in the US and the UK, behavioral responses have significant impacts in the congestion reduction benefits of capacity expansion projects (Noland & Lem, A review of the ecidence for induced demand and changes in transporation and enviromental policy in the US and the UK, 2002). No matter the level of congestion, VMT growth is likely to be larger with more highway capacity compared to less highway capacity. While, another study conducted again in the US indicates that induced demand occurs by either through the ability of new infrastructure to make more locations easily accessible, or through its ability to reduce urban congestion (Hymel, Small, & Dender, 2010).

Recent evidence from the Netherlands supports previous researches that new road infrastructure generates new demand, but the amount of it might be less than has been assumed thus far. It may also be that the amount of induced demand has decreased in the past few years. Increment of traffic after the improvement of infrastructure is caused also by shifts in route and departure time (Loop, Haaijer, & Willigers, 2016). Noland resulted based on his analysis that increased capacity clearly increases vehicle miles of travel beyond any short run congestion relief that may be obtained. All methods employed found statistically significant relationships between lane miles and VMT. While some more factors, such as population growth, also contributes to the increment of demand, capacity additions account for about one quarter of this growth. One more of his findings was that urban roads have a greater relationship to VMT growth than smaller rural roads (Noland, 2001).

The evidence reviewed in this section supports that induced traffic exist and may be significant in some situations. All studies concluded that, increased road capacity allows more vehicle travel to occur, indicating that there is a relationship between VMT and road capacity. Increment in demand because of improved infrastructure can be caused either by generated traffic, or by routes and time switches. Induced demand is likely to be higher for capacity improvements in urban areas or on highly congested routes. A great amount of the additional capacity is immediately filled, before the impacts of congestion relief become obvious.

2.1.2 Induced demand caused by a new transport mode

Induced demand effects are also notorious when a new transport mode is introduced. It has been observed that improvement of current transportation systems by offering higher speeds or the incorporation to the mobility of a mode such as shared or aerial vehicles can cause increment in demand.

2.1.2.1 Air Mobility

Several theoretical, empirical and statistical studies conducted by Koenig (1980) , Thill & Kim (2003), Robinson & Vickerman (2006) and others, overwhelmingly support that trip production and trip attraction are significantly affected by geographic accessibility between trip-ends. Accessibility is a context-dependent notion that needs to be approached with flexibility to situations framing the travel choice decisions (Kwan, 1998). As a consequence, providing a new transportation mode, easier accessible to the commuters leads to the existence of induced demand.

Offering to urban mobility the third dimension (air mobility) contributes to higher accessibility in various destinations. Autonomous aerial vehicles and flying offer the most innovative transportation service for urban mobility, known as Urban Air Mobility (UAM)

(Lineberger, Hussain, Mehra, & Pankratz, 2018). This third dimension to mobility is expected to improve accessibility between suburbs and cities . Since there are not still any available data from realistic applications of UAM, researchers have estimated potential market share of UAM, by including it as another alternative on the mode choice model.

According to results of recent studies, the demand for UAM proved to be higher that the available fleet could handle, particularly in the scenarios where the price is relatively low (Balać, Vetrella, Rothfeld, & Schmid, 2019). In addition, UAM capabilities have sustainable effects on the increase of the demand, especially for mid-distance trips under 40 km. Generally, low price showed to be crucial for both cases and made aerial vehicles more attractive to the users.

Overall, accessibility proved to be an influential factor for the decision of making a new trip, and hence easier accessibility induces demand. Therefore it can be concluded that aerial vehicles, that offer easier, faster and more flexible transfer will be able not only to support the current mobility situation, but also further expand the transport demand and hence, create induced traffic.

2.1.2.2 Shared Mobility

Although, the third dimension in mobility can indeed improve the transportation conditions and be the reason for causing extra kilometers travelled, other modes such as shared mobility could also be the basis for the existence of induced demand. Shared mobility consider to be a mode that provides services between private and public transport, by offering flexible and door-to-door services, without owning a vehicle (Susan, Danel, & Conrad, 1999). The most well-known definition of carsharing states that the system consist of a fleet of vehicles, available at several stations, to be used by a wide range of members (Susan, Danel, & Conrad, 1999).

Based on a research study which had as objective to estimate the impact of shared autonomous vehicles on vehicle kilometer travelled in a large metropolitan aera, it was found out that shared autonomous vehicles increased the total travelled distance by up to 8% (Moreno, Michalski, Llorca, & Moeckel, 2018). According to another recent study (Moudon, Lowry, Shen, & Ban, 2020) car-sharing and ride-hilling not only substitute vehicle trips, but also induce more travel. Induce travel could add to the traffic congestion but could also improve the accessibility to activities. Although, further research should be conducted with focus on the temporal and purpose characteristics if trips by shared mobility.

Other researchers explored the effects of shared-mobility in San Francisco, California and he found out that an approximately 7% of member's trips and more than 20% of vehicle miles traveled were by shared-use vehicles (Cervero, 2003). The model used to predict the shared-mobility impacts on travel behavior was a binomial logit one. Results indicate that access to shared cars is stimulating motorized travel. More specifically, he found out that sharing vehicles are used more for personal and social travel rather than routine travel such as to work or school. A quite interesting finding of this study is that shared-mobility cause induced demand, although it doesn't contribute to the congestion effect because they are not generally used during peak hours or to dense settings well served by transit, such as downtown. Users are increasing substantial travel-time saving and willingly pay market price for these benefits. His conclusion confirms the outcomes of a Swiss study, that the post-membership surveys, resulted in a 11.8% increase in total kilometers traveled by previously carless households (Steininger, Vogl, & Zettl, 1996).

Other studies have researched the net impacts of shared mobility on total number on trips. Based on a recent study, ride-hailing contributes to a significant growth of number of trips in New York City (Susan S. , 2017). In addition, several surveys deployed in U.S. cities estimated a net growth in total number of trips on the basis of mode substitution and ride-hailing use frequency (Clewlow & Mishra, 2017). It was also found out that, 59% of ride-hilling trips added a new vehicle on the road (Gehrke, 2017). Other researchers estimated that most driving activities by car2go added total trips because they were generally used to satisfy incidental mobility needs (Elliot & Shaheen, 2016). Though, they found that a small group of car2go users tended to decrease their driving by selling their personal vehicle and postponing a vehicle purchase. The study resulted that the net effect of car2go was to reduce total driving.

Overall, newly available shared mobility options are having a large impact on travel demand. Car- and bike-sharing and re-hailing have become increasingly viable and attractive travel modes since they have been app-based and able to link riders and vehicles in real time and space. As a consequence of this facility, there is a significant correlation between number of trips and the existence of shared mobility. Although, shared mobility does influence the congestion, it improves the accessibility to activities. This mode choice, is mainly used for personal and social travel rather than mandatory trips such as work or study ones (Wang, 2020).

2.2 Analysis Methods

2.2.1 Discrete Choice Modeling (DCM)

Discrete choice modeling is a widely used in user preferences for a given choice that uses the value of utility maximization. Each individual will choose the alternative having the highest utility, which is based on the attributed of the alternative and the decision maker (Ben-Akiva & Lerman, 1985).

For an alternative i and an individual q, the utility is a combination element V_{iq} and a random component e_{iq} (Louviere, Hensher, & Swait, 2000), which is described on the following equation:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{1}$$

where,

- U_{iq} is the utility of alternative i for individual q
- V_{iq} is the systematic componet of alternative i for individual q

- ε_{iq} is the error component which have to do with V_{iq}

V_{iq} is a mixture of components associated only with the characteristics of the alternative (differ for one individual across different choices), of the decision-maker (same for the same individual across different alternatives), and the interactions between attributes of the alternative and attributed for the person is about to take the decision. The V_{iq} also includes an alternative-specific constant for the given alternative i (Koppelman & Bhat, 2006). This variable can be written as follows (Ortuzar & Willumsen, 2011):

$$V_{iq} = \beta_{1i}X_{1iq} + \beta_{2i}X_{2iq} + \cdots + \beta_{ki}X_{kiq} \tag{2}$$

where,

- $\beta_{1i}, \beta_{2i}, \dots, \beta_{ki}$ are the parameters that are not known which should be estimated and remain constant for the individuals but may vary across alternatives.

- $X_{1iq,}\, X_{2iq,\dots,}\, X_{kiq}$ are the k independent variables including all attributes of alternative i for individual q, related to the decision maker and the alternatives.

For a specific utility, the alternative-specific constant (ASC) captures the effect of factors that are not part of the model. When this constant is added, the unobserved error term is bound to a mean of zero (Train, 2009). The only that matters in the differences in utility and as a consequense, one alternative can be normalised to zero by setting its ASC to zero. So, for i alternatives the model can at most have $i - 1$ ASCs.

Individual q will choose alternative i over j if and only if the utility of i greater than that of j. The following equations occurred (Louviere, Hensher, & Swait, 2000):

$$V_{iq} + \varepsilon_{iq} > V_{jq} + \varepsilon_{jq} \tag{3}$$

$$V_{iq} - V_{jq} > \varepsilon_{jq} - \varepsilon_{iq} \tag{4}$$

It is not possible tha the error terms will be calcilated and therefor, the porbability that $V_{iq} - V_{jq}$ is gretaer than $\varepsilon_{jq} - \varepsilon_{iq}$:

$$P_{iq} = P\left(U_{iq} \geq U_{jq}\right) \tag{5}$$

The above equation can be solved and the β coefficient can be estimated using the Maximum Likelihood Estimation method (Ben-Akiva & Lerman, 1985).

Based on the probability distribution of the error term, there are different types of discrete choice models (Ben-Akiva & Lerman, 1985). A usual assumption is that the error term is normally distributed (Koppelman & Bhat, 2006) and that leads to the formulation of the porbit models. However since they can be difficult to solve, logit models on a logisitc distribution of the error term are more commonly used.

Logisitic regression model is commonly used in regression analysis, where the independent variables are explored in terms of their relation to the dependet variables they explain (Hosmer & Lemenshow, 2013). In logit models, the discrete outcome variable is binary and the resulting model called binary logit model. Several other models following logistic regression are used in practice, and explained in the following.

Multinomial Logit Models (MNL):

Multinomial logit models (MNL) are logit models with more than two alternatives or two unordered outcomes. It is mainly assumed in this model type the alternatives are independent and irrelevant (IIA) and that they are independent and identical distributed (IID). IIA indicates that choosing one alternative is not affected by the presence or absence of other alternatives (Louviere, Hensher, & Swait, Stated choice methods: analysis and applications., 2000). Additionally, it also states that the random error terms ε_{iq}, are independent and identical distributed for different alternatives. Multinomial logit models can be expressed by the following mathematical equation:

$$P_{iq} = \frac{e^{V_{jq}}}{\sum_{j=1}^{J} e^{V_{jq}}} \qquad (6)$$

where,

- P_{iq} is the probability of choosing alternative i by individual q
- V_{jq} is the systematic component of the utility of alternative j for individual q
- V_{iq} is the systematic component of the utility of alternative i for individual q

Although they are really useful models, they have several limitations. For example, if alternatives share some similarities, other models could be used, such as nested models.

2.2.2 Nested Logit Models (NL):

Nested models have been used in various areas of transportation such as to the traditional transportation mode choice (Forinash & Koppelman, 1993). Following logistic probability distribution, nested logit models (NL) are useful when the alternatives can be grouped into subcategories, call nests (Train, 2009). Categories within a nest have a high degree of similarity than those outside the nest. In general, nested logit models partially relax MNL constraints such as the Independence from Irrelevant Alternatives (IIA) and Independent and Identically Distributed (IID) , where IIA only holds within the nest.

Each alternative i of an individual q has an utility which is the combination of a nest component that is constant across alternatives within the nest, and a variable component that is variable for alternatives within a nest. The utility in a nested model is expressed in Equation 1 (Train, 2009):

$$U_{iq} = W_{kq} + Y_{iq} + \varepsilon_{iq} \tag{7}$$

where,

- W_{KQ} it is constant across alternatives within one nest and it depends only on variables describing nest k
- Y_{iq} depends on variables describing alternative i
- ε_{iq} is the error term of alternative i for individual q

The probability of choosing i from B_k is the product of the probability that an alternative within B_k is choses and the conditional probability that i is choses given B_k. The Equation 2 expresses the previously mentioned probability:

$$P_{iq} = P_{iq|B_k} P_{qB_k} \tag{8}$$

Equation 2 can be also written as:

$$P_{qB_k} = \frac{\exp(\mathrm{W_{kq}} + \lambda_k I_{kq})}{\sum_{i=1}^{K} \exp(W_{iq} + \lambda_i I_{iq})} \tag{9}$$

$$P_{iq|B_k} = \frac{\exp(Y_{iq|\lambda_k})}{\sum_{j\varepsilon B\_k} \exp(Y_{iq}|\lambda_k)} \tag{10}$$

where,

- $I_{kq} = \ln \sum j\varepsilon B_k \ \exp(Y_{jq}|\lambda_k)$
- λ_k illustrates the degree of independence in the unobserved utility among the alternatives in the nest k
- I_{kq} is the inclusive utility
- k is the given nest
- I represents other nests

2.2.3 Maximum Likelihood Estimation (MLE)

The maximum likelihood estimation is a statistical tool used in several analytical studies, to estimate the model parameters given a set of observations (Harrell, 2015). Based on the assumption that the explanatory model variables are independent of the unobserved components of the utility, it can be applied (Train, 2009). Knowing that the choice of individuals for an alternative are not correlated with each other, the likelihood function can be expressed as follows:

$$L(\beta) = \prod_{q=1}^{Q} \prod \left(P_{iq}\right)^{y_{iq}} \tag{19}$$

where,

- β is a vector with the estimate parameters of the model
- P_{iq} is the probability that the individual q chooses alternative i
- y_{iq} is equal to one if individual q chooses i and zero otherwise

The target is to maximize the likelihood function so that it can also be maximized the probability of Y_i being one, it can be explained as the probability of success. Due to the complexity of the likelihood function, it is much easier to maximize its logarithm (Louviere, Hensher, & Swait, Stated choice methods: analysis and applications., 2000), as expressed below:

$$L(\beta) = \sum_{q=1}^{Q} \sum_{i} y_{iq} \ln\left(P_{iq}\right) \tag{20}$$

In order to maximize the above equation, its derivative it should be set equal to zero with respect to the variable parameters.

$$\frac{dL(\beta)}{d\beta} = 0 \tag{21}$$

The above formula it can be written with the use of the linear parametrization of the utility as follows:

$$\sum_{q} \sum i\left(y_{iq} - P_{iq}\right) X_{iq} = 0 \tag{22}$$

The estimation of the above equation equals to the maximization of the loglikelihood function.

2.2.4 Multicollinearity

VIF

Multicollinearity exists when independent variables are found to be highly correlated with each other. This phenomenon can be detected through correlation plots between independent variables and varies based on data type or through the Variance Inflation Factor (VIF) (Simon, Young, & Rardoe, 2018):

$$VIF_i = \frac{1}{1 - R_i^2} \tag{25}$$

where,

- R_i^2 is the pairwise coefficient of determination between two independent variables.

Research studies in statistics recommend that VIF values higher than 4 suggest further investigation and values higher than 10 indicate the existence of multicollinearity. Strategies for avoiding multicollinearity include the removal of one or more collinear variables or even, doing nothing, as sometimes it is necessary to measure the effects of the collinear variables in a model or the inclusion of those variables contributed to the interpretation of the model (Washington, Karlaftis, & Mannering, 2003).

Cramer's V

When it comes to correlation between categorical variables, one of the most common ways to check their correlation is by calculating the Cramer's V coefficients. Cramer's V is a number between 0 and 1 that indicated how strongly two categorical variables are associated. Cramer's V coefficient is calculated by the following equation (Field, 2013):

$$\varphi_c = \sqrt{\frac{x^2}{N(k-1)}} \tag{26}$$

where,

- φ_c denoted Cramer's V
- x^2 is the Pearson chi-square statistic from the aforementioned test
- N is the sample size involved in the test and
- k in the least number of categories of either variable

2.2.5 Model evaluation techniques

There is a huge variety of model evaluation techniques that are always trading off between variance, bias and computation time. The two most well-known model evaluation techniques are the train-test-split and the k-fold cross validation (Gareth, Witten, & Hastie, 2013). In the first option, the training data is randomly split into a train and test partition, commonly with a significant part of the data being retained as a training test. Most commonly used proportions in the literature are 70/30 and 80/20,though the exact ratio depends on the size of the data (Gareth, Witten, & Hastie, 2013).

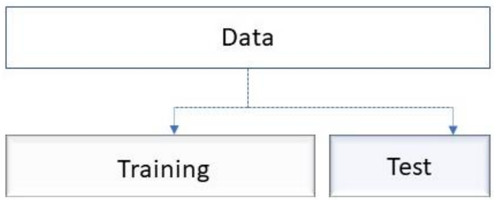

Figure 2.1 Random split of train and test set *(Gareth, Witten, & Hastie, 2013)*

A more robust alternative is the so-called k-fold cross validation. Here, the data are randomly splitted into k folds. The most important advantage over the train-test-split method is that each of the k parts is iteratively used as a test dataset with the remaining $k-1$ parts being the training set in the iteration. This process is repeated k times, so that each observation is included in both training and test partitions. The appropriate error metric is then simply calculated as a mean of all of the k folds, giving the cross-validation error (Gareth, Witten, & Hastie, 2013).

Data

1	Validate	Train	Train	Train	Train
2	Train	Validate	Train	Train	Train
3	Train	Train	Validate	Train	Train
…	Train	Train	Train	Validate	Train
$\boldsymbol{k}$	Train	Train	Train	Train	Validate

Figure 2.2 K-fold cross-validation *(Gareth, Witten, & Hastie, 2013)*

2.2.6 Statistical test

The Akaike Information Criterion (AIC)

AIC is a method which mainly used to compare models between each other. The models that has the lowest AIC value performs the best. By integrating AIC in the analysis, the best-fit model by explaining the greatest amount of variability and using the fewest possible independent variables (Bevans, 2020). Although it is a very useful method, it has a significant drawback. It allows overfitted models to be selected, as it tends to improve with a larger number of k parameters.

$$AIC = n * \ln(MSE) + 2k \tag{23}$$

where,

$$MSE = \frac{\sum_{i=1}^{n}(y - \widehat{y_i})^2}{n} \tag{24}$$

where,

- y_i is the observation of response
- $\widehat{y_i}$ is the predicted value
- n is the number of observation

2.2.7 Classification metrics

Assuming that we have trained a classification model with a response variable Y. Let C refer to a particular class of the response variable. The observations can be characterized as (Ting, Sammut, & Webb, 2011):

- **True Positive** for Class C if the model correctly predicts that the observation is of class C.
- **False Positive** for Class C is when the model wrongly predicts that the observation is of the above mentioned class.
- **True Negative** for Class C if the model correctly predicts that the observation is not of class C.
- **False Negative** for Class C if the model wrongly predicts that an observation is not of Class C, whereas in reality it is.

It is also possible to measure the performance on individual classes by defining the following metrics:

- **Sensitivity,** True Positive Rate, **Recall**, or Probability of Detection:

$$\frac{TP}{TP + FN} = \frac{TP}{Number\ of\ Actul\ C_s} \tag{11}$$

- **Specificity** or True Negative Rate:

$$\frac{TN}{TN + FP} = \frac{TN}{Number\ of\ Actual\ non - C_s} \tag{12}$$

- **Positive Predictive Value** or **Precision**:

$$\frac{TP}{TP + FP} = \frac{TP}{Number\ of\ Predicted\ C_s} \tag{13}$$

- **Negative Predictive Value**:

$$\frac{TN}{TN + FN} = \frac{TP}{Number\ of\ predicted\ non - C_s} \tag{14}$$

- **Prevalence**:

$$\frac{TP + FN}{TP + FP + TN + FN} = \frac{Number\ of\ Actual\ C_s}{Size\ of\ Total\ Population} \tag{15}$$

- **Detection Rate**:

$$\frac{TP}{TP + FP + TN + FN} = \frac{TP}{Size\ of\ Total\ Population} \tag{16}$$

- **Detection Prevalence**:

$$\frac{TP + FP}{TP + FP + TN + FN} = \frac{Number\ of\ Predicted\ C_s}{Size\ of\ Total\ Population} \tag{17}$$

- **Balanced Accuracy**:

$$\frac{Sensitivity + Specificity}{2} =$$

$$\frac{TP + TN}{2 * (TP + FP + TN + FN)} = \frac{Number\ of\ Correct\ Prediction}{2 * (Size\ of\ Total\ Population)} \tag{18}$$

2.2.8 Outlier detection

Consistent data are technically correct data that are fit for statistical analysis. They are data in which missing values, special values and outliers are removed, corrected or imputed. There is a huge variety of literature on outlier detection, and several definitions of outlier exist. A general definition by Bernett defines an outliers in a dataset as an observation or set of observations, which appear to be inconsistent with the dataset (Bernett, 1978). A quite well-known method for outlier detection, is the Tukey's box-and-whisker method. Based on this method an observation consider to be an outlier when it is larger than the so-called "whiskers" of the set of observations. The upper whisker is computed by adding 1.5 times the interquartile range to the third quartile and rounding to the nearest lower observation. The lower whisker is computed likewise (Jonge & Loo, 2013).

2.2.9 Imbalanced data

In the 1990s as continuously increasing data and applications of machine learning and data mining started to become prevalent, a crucial challenge appeared: how to achieve desired classification accuracy when dealing with data that had significantly skewed class distributions (Branco, Torgo, & Ribeiro, 2016). Among others, in 2002 proposed a novel approach which was to simply oversampling by replication, and assist the classifier to improve its generalization on the testing data. The basis of this idea was in reality to create new minority instances. This technique entitled Synthetic Minority Oversampling Technique and it is known as SMOTE (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). SMOTE tries to interpolate among neighboring minority class instances. Hence, it is possible to increase the minority class by introducing new instances and a consequence, assists the classifier to improve its generalization capacity.

3 Methodology

In this chapter, the methodology of this work is described, consisting of the proposed approach for identifying and modeling induced demand and the analysis methods, including the framework for specifying the models.

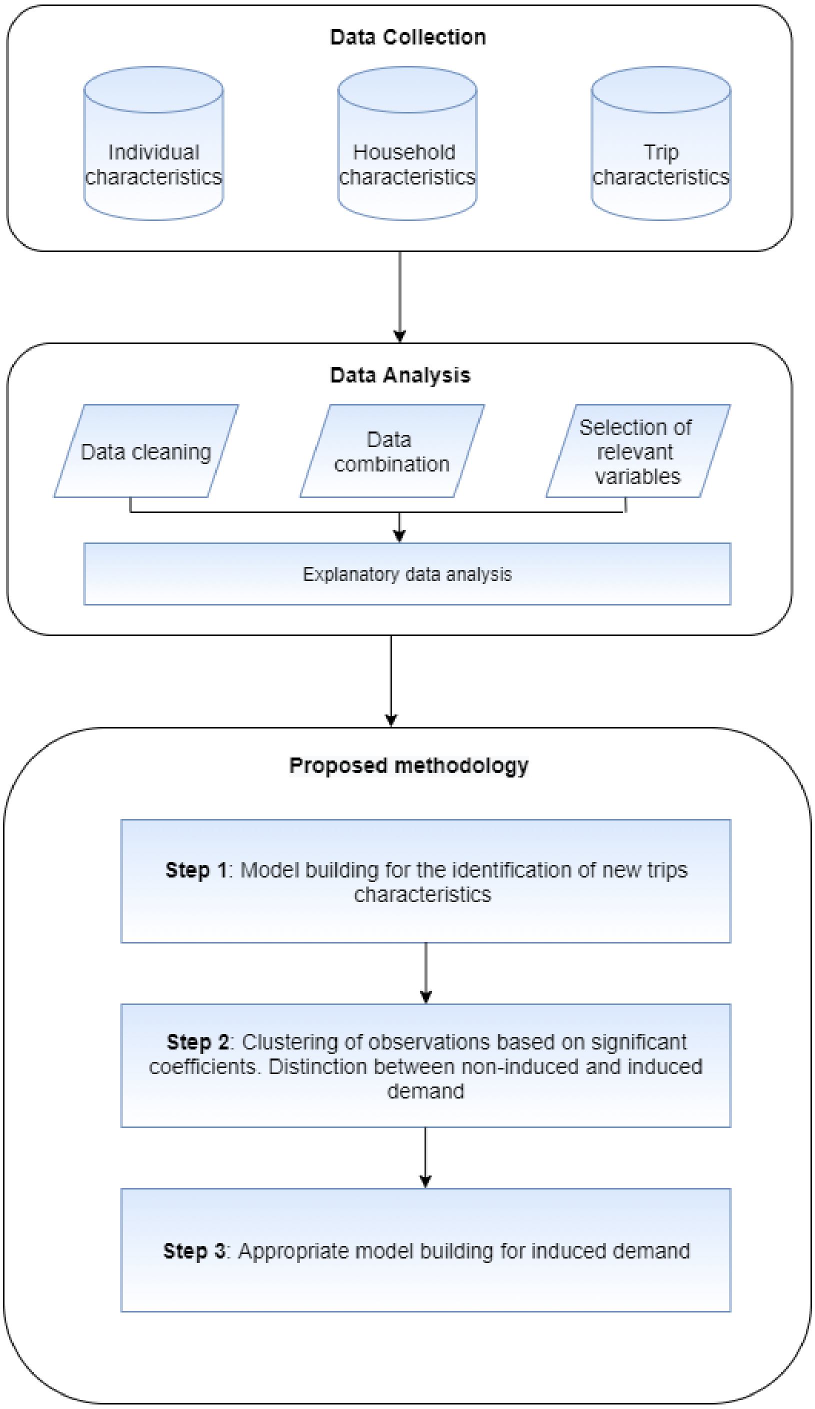

Figure 3.1 Methodological Framework

3.1 Three-step proposed methodology

As already discussed, induced demand is an important component of travel demand and increasing attention has been paid to building analytical models to identify it more precisely. This study focuses on recognising the additional trips resulting from improvements in the supply system, which can be resulted by either improved infrastructure or by the introduction of a new transport mode. The result of those actions are convenient and faster trips, but also more mobility choices beyond the traditional bookend of auto ownership and public transit. Identifying and modelling induced demand is not an easy task due to the high numbers of variables which make the analysis complicated and difficult to generalize. A transportation system is a set of elements interconnected by complex relationships and whenever an action is planned on a part of a transportation system, there are unavoidable impacts on other parts, positive or negative. Improvement within the supply system, such as the introduction of a new transport mode, that offers more comfortable movements, or in any case actions that increase the utility and the satisfaction of customers about the possibility of moving, create a new share on travel demand (Pagliara & Preston, 2013).

The methodology proposed in this study is represented in Figure 3.1 and it is composed of three steps. The first one is to identify the characteristics that the trips occurring just once in the network have. In order to achieve it, the relationship between frequency, trip and sociodemographics characteristics were explored. The most appropriate way to analyse this topic was to build multinomial and nested logit models, using as dependent variable the frequency. The analytical approach of this step in presented in Section 3.2. The significant coefficients of the best fitted model, is proposed to be used as an input for the next step.

The coefficients obtained from the first step should be appropriately clustered in order to formulate several demand groups. The main target of this step is to distinguish the non-induced from the induced demand. Based on the literature, the induced demand can be measured as either the percentage of change in VMT, or in distance travelled or even as the percentage of change in number of trips. Taking into consideration the available data for this study, induced demand should be calculated as the percentage of change in number of trips. An approach could be, to firstly subset the trips that occurring just once in the network, and then further explore the characteristics that these trips have and which proved to be significant in the first step. In the following, it should be decided whether these trips could be counted as induced or not. Assuming that trip reason is one of the significant estimated variables from the first step, an analysis on the levels of this variable should take place. Considering that trip reason consist of several levels, ex: shopping, work, study etc., we should think which of the trips should undoubtedly take place (work-, home-trips etc.) and which can be partially considered as induced demand. For example, trips concerning shopping, sports or in general private issues could consider to be partially induced trips. The idea, behind this suggestion is that, shopping- or leisuretrips could have been avoided but because of the comfort that improved supply system offers they eventually take place.

The third and final step of the proposed methodology is the development and estimation of a model for the induced demand and the appropriate integration of it within the conventional 4-step model.

3.2 Implementation of first step

This section presents an overview of the implementation of the first step of the proposed methodology. Several analysis methods were used in reaching our objective. First, data cleaning was performed, in order to obtain consistent data without outliers. Second, discrete choice models were estimated based on the cleaned data. Third, the fully interaction of the household size and the vehicle ownership was integrated as a new variable in the models. Fourth, 5-fold cross validation was used to ascertain model accuracy.

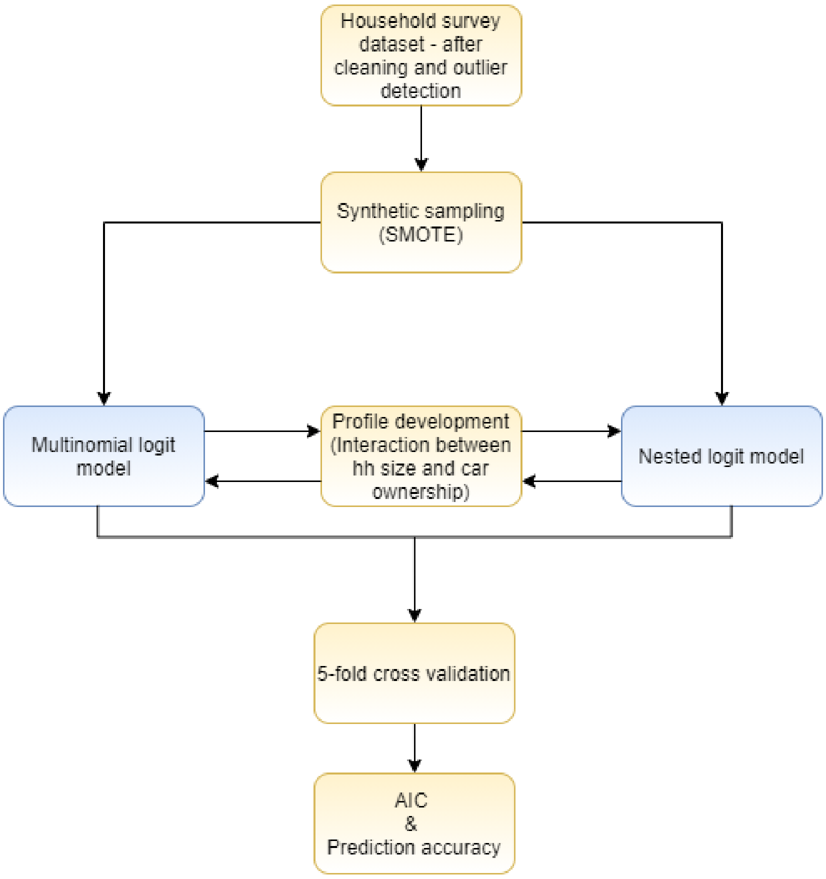

Figure 3.2 Implementation of first step

3.2.1 Data preparation for model development

The main step, before building a model, is to prepare and clean the data. One unavoidable problem in survey statistics , is the existence of outliers (Wada, 2020). They may reduce the information of survey dataset and distort estimation of each step of the survey statistics production process. As a consequence, it is necessary to detect and remove them from the data set. For this reason, boxplots where created for the numerical variables, which in this case are the age, the number of trip stages and distance. The negative values from all there variables were removed, while an upper and lower limit was set for trips distance and age. Concerning the age, individuals between the age of 18 and 70 were selected, since the age distribution graph indicate that new trips are mainly taking place from these ages. In the case of distance, trips between 200 m and 50 km were selected. The reason is that, it has been observed that the average speed in urban roads is 25-30 km/hr (The Urban Mobility Observatory, 2018). Considering a maximum two-hours travel and speed of 25 km/hr, and hence we decided to go for a cut-off of 50 km.

Based on the target of this step, which is the identification of the new trips characteristics, it was decided that, the dependent variable will be the frequency, and the rest will be considered as the independent ones. The independent variables were tested for multicollinearity. Pearson's correlation coefficients were used for continuous variables and Cramer's V was used for categorical variables. Based on the literature, if the correlation coefficient is less than 0.7, variables can considered in the analysis (Fields & Chen, 1991). A correlation above 0.7 between nationality and private vehicle usage was detected (Figure 5.1), and hence these two variables were tested separately in the models.

The dependent variable, was initially represented by five levels:

- Daily trips, Monday to Friday
- Between two and four working days per week
- Less than two working days per week
- Sometimes per month
- Trips that took place only once

Several changes were tested on those levels, but what gave most accurate results was the reassignment of the frequency into three categories: daily (first level), weekly/monthly trips (combining the second, third and fourth level) and new trips. Then, the dependent variable is severely unbalanced with only 2.1% of trips being new in the network. Therefore, the synthetic sampling technique SMOTE was used to convert the frequency levels into three equal parts.

3.2.2 Behavioral modeling

Behavioral or choice modelling was used as the analysis method for this thesis. In an aim to identify the factors affecting the decision of making a new trip, several models choice models were built with the frequency as a dependent variable. The choice were therefore the given options ranging from daily, weekly/monthly and new trips. For categorical variables like sociodemographics, binary variables were created to assess the impact they have on the models. As a first step MNL models were created to get first insights on influential and relevant factors.

MNL models: Using the available data, MNL were developed by setting the daily level as a base case. Models were built first by using all survey variables, is what is called a "structural model" method. After simplifying and eliminating the highly insignificant variables, the approach was reversed. "Empty models" were developed by adding one by one the significant variables. In the following, these models were then optimized using alternating explanatory variables, with the advantage of observing patterns of attitude estimates across different alternatives. To further investigate whether considering the household condition improves the current model fit and to which extend helps to understand the decision of making a new trip, two models were designed: Model 1.1 includes, household size and vehicle ownership independently. Model 2.1 includes the household condition represented by the full interaction of number of individuals living together and the vehicles that are available in the household

NL models: Nested logit models were developed with two different nests, one including the daily and weekly/monthly trips and another including the daily ones. The same procedure with the MNL models is followed here as well. Two nested models were developed: Model 1.2 includes, household size and vehicle ownership independently. Model 2.2 includes the household condition represented by the full interaction of number of individual living together and the vehicles that are available in the household.

The 5-fold cross validation techniques was used in order to split the dataset as train and test set appropriately. All folds show similar results, but the results from the fold that gave the highest accuracy is presented in Table 5.4 and Table 5.11. The models' goodness-of-fits were compared with the Akaike information criterion (AIC) and the prediction accuracy. The model with the smaller AIC and the biggest prediction accuracy is preferred.

The models were estimated using R Studio and the package *mlogit,* developed by Yves Croissant. The package can estimate both multinomial logit models and nested logit models (Croissant, 2011).

The *mlogit* function accepts data in wide and long format. The wide format has one row per each choice while the long format has one row per each alternative. In this study, three rows per variable record, one per each frequency level. The long data format is used here to account for the frequency specific characteristics.

The function used consist of three types of variables:

- Alternative specific variables with a generic coefficient across all alternatives
- Individual and trip specific variables
- Alternative specific variables with different coefficients for each alternative

The function then outputs estimated coefficients, statistical measures of each parameter and of the overall model, e.g. the t-statistics, the log-likelihood etc.

For nested logit models, the *mlogit* function offers the same specifications concerning the variables, with an added specification of nesting structure. The function then outputs the estimated coefficients as well as the nesting coefficient.

In all models, the variables with very low significance were immediately removed; the highly significant ones (above 95% significance level or 90% according to the importance) were kept.

4 Data Analysis

This chapter gives some general information about the dataset from Madrid and then explains the structure of the included files. In its initial form, data comprises of several files in csv format, which are combined in data analysis and model development.

4.1 Data source and descriptive statistics

Data were obtained from the Open Data CRT of the Madrid Regional (Open data CRTM , 2018). They offer substantial information about the mobility in the Community of Madrid, which are available to organisations and all citizens. More precisely, the total sample consist of 85,064 individuals, coming from 58,492 households and 222,744 trips. Since one of the focuses of this study is to identify induced demand caused by shared mobility, which only takes place in the canter of the city, only the central trips took into consideration. The zoning system of Madrid used to filter central trips, in the following outlier detection, and a final filtering of the target group based on age and trip distance was conducted in order to be created the appropriate for this study subsample. After data cleaning, the new dataset consist of one merged file with 31,428 observations.

The main findings of the sample distribution (Table 4.2) can be summarised as follows:

1. Sample is well distributed in terms of gender, which is representative of the whole Madrid's population based on the city's sociodemographics data (Madrid-Population, 2018).
2. Some levels of certain categorical variables were combined together, when they present common characteristics and they represented only a very small percentage of the sample size.
3. Based on the needs of the study population was splitted into three age groups, and it was find out that the majority of the population belongs to the age group between 35 and 54.
4. It is observed an overrepresentation of people holding a Spanish nationality and only a very small percentage of foreigners.
5. The majority of the population holds a driving licence; conversely only 21% don't know how to drive.
6. Education categories were reassigned to low educated (primary, secondary and post-secondary school) and university (bachelor, master and doctorate) levels.
7. Employment types were reassigned to students, employees, unemployed people, retired and others.
8. Private sector was overrepresented as professional activity in the sample.
9. Participants show high percentage of public transport card availability .
10. Months were reassigned to "February to March" and April to June".
11. Household size categories were reassigned to single, couple and families.

12. Number of children under the age of 4 years old and number of vehicles in the household reassigned to 0, 1 and 2 or more.
13. Trip reasons were reassigned to home, work, study, shopping, sports, private issues and others, which work trips being the dominant ones.
14. Slightly more than the half of the trips took place with private vehicle.
15. PT as commute mode was overrepresented for central subsample.
16. The majority of trips consist of only one stage and they are not part of a trip chain.

Moreover, as we are interested in the characteristics of trips taking place just once in the network, the distribution of the outcomes of this dependent variable is of great interest and shown in Table 4.1 below:

Table 4.1 Distribution of dependent variable

Outcome	**Frequency(%)**
Daily	49.5
Weekly/Monthly	48.4
New trips	2.1

As already mentioned, daily trips consist of the trips that happen at least 2 times per week, while the next category included all the trips that take place maximum one time per week and minimum sometimes during a month , and finally the last category included only the trips that happened just once in the network. The distribution of frequency show clearly that the majority of trips that are taking place in the centre of Madrid are the ones that happen daily (work-, home trips etc.) The percentage of weekly/monthly trips is about half of daily trips, while a negligible amount of trips happening only one time in the network.

Table 4.2 Summary of individuals, household and trip characteristics

Characteristics	Sample (N = 31,428)	Percentage (%)
Individual characteristics		
Gender		
Male	13757	43.8
Female	17671	56.2
Age		
18-34	6107	19.4
35-54	12750	40.6
55-70	12571	40.0
Nationality		
Spanish	29810	5.1
Other	1618	94.9
Driving license		

Yes	24843	79.0
No	6585	21.0
Education		
Primary/Secondary/Post-Secondary	11095	35.3
Vocational/Bachelor/Master/PhD	20333	64.7
Employment		
Student	3347	10.6
Employee	19486	61.9
Unemployed	3704	11.8
Retired	4531	14.4
Other	378	1.2
Professional activity		
Public sector employee	5156	16.4
Private sector employee	12601	40.1
Self-employed/freelancer	2579	8.2
Other	174	0.6
NAs	10918	34.7
Working sector		
Education	1617	5.1
Health system	1850	5.9
Public Administration	2345	7.5
Other services	12957	41.2
Industry	750	2.4
Construction/ Agricultural	991	3.1
NAs	10918	34.7
Availability of PT		
Yes	27360	87.1
No	4068	12.9
Month		
February-March	14144	45.0
April-June	17284	55.0
Restricted mobility		
Yes	826	2.6
No	30602	97.4
Household characteristics		
Household composition without members younger than 4 years old		
Single	3272	10.4
Couple	9511	30.3
Family	18645	59.3
Number of children under 4 in household		
None	29100	92.6
1	1779	5.7
2 or more	549	1.7
Number of vehicles in household		
0	5698	18.1
1	15078	48.0
2 or more	10652	33.9
Trip characteristics		

Private vehicle usage		
Yes	29810	94.9
No	1618	5.1
Trip reason		
Home	363	1.2
Work	13268	42.2
Study	1604	5.1
Shopping	3022	9.6
Private issue	8025	25.5
Sports	4908	15.6
Other	238	0.8
Main Transport mode		
Public transport	14654	46.6
Taxi	350	1.1
Vehicle	7556	24.0
Shared mobility (vehicle/ bicycle)	111	0.4
Motorbike	682	2.2
Bicycle	191	0.6
Walking	7884	25.1
Number of stages		
1	24379	77.6
2	5352	17.0
3 or more	1697	5.4

Note: The categories professional activity and working sector contain a high percentage of NAs and therefore they were not included in the model development.

4.2 Descriptive analysis

In this subsection, frequency is represented for the different demographics or attributes of interest, in order to have preliminary insights of their impacts on induced demand. The data were merged together and one file which consist of the combination of individuals, trips and household characteristics was generated. Most figures represent distributions for all outcomes. However, for some demographics, graphs were generated excluding NA values and categories representing less than 5% of the sample size, as it wouldn't be convenient or representative to observe their distribution.

Frequency by gender:

The analysis of the frequency levels choices by **gender** shows that in general this variable is expected to be indifferent whether an individual will make a decision to start a trip or not (Figure 4.1). These first insights will be further explored in the model development procedure.

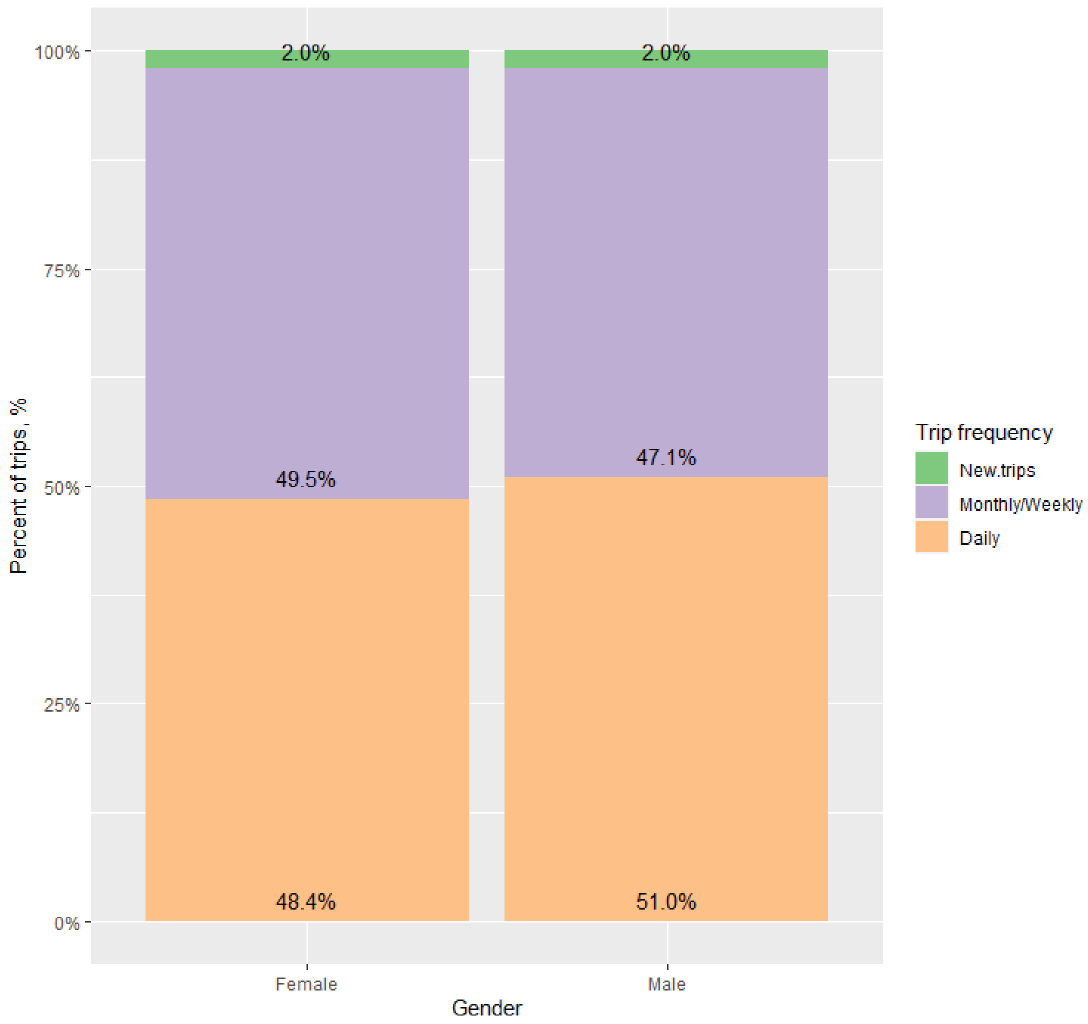

Figure 4.1 Frequency by gender

Frequency by age group:

Age was in the form of continuous variable in the beginning. After grouping the ages in three age categories it was find out that (Figure 4.1), young people under the age of 35 have the highest likelihood to make new trips, followed by the age group of 55-70 while the people between 35-55 years old are the less probable to make new trips. Concerning the daily trips are mostly probable to take place from people belonging to the age group of 35-54, while the weekly and monthly trips are mainly take place by people from 55 to 70 years old. These interesting finding will be further explored in the model analysis.

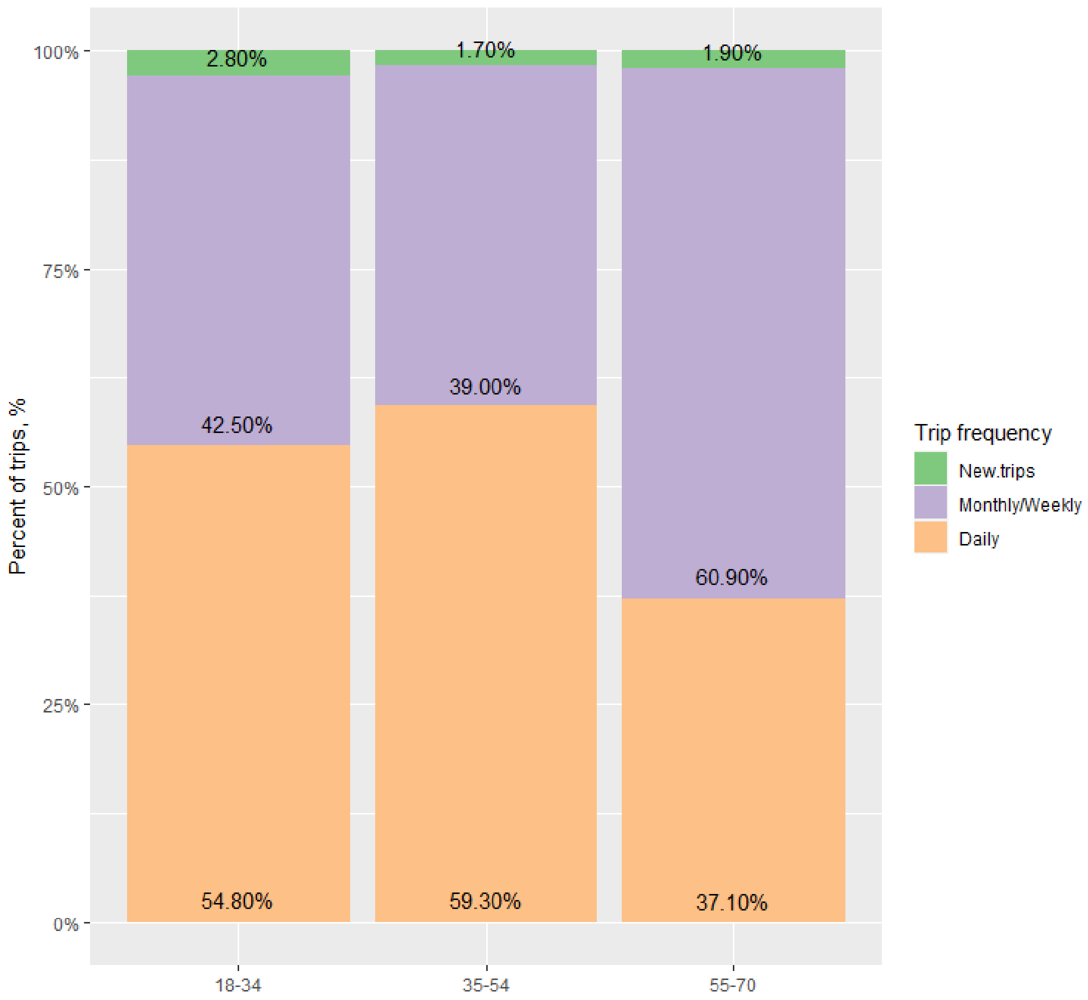

Figure 4.2 Trip frequency by age group

Frequency by nationality:

Cultural impact was not observed through the dataset, with the people wo don't hold a Spanish nationality to be slightly more probable to make daily trips (Figure 4.3). This situation, could occur because foreigners are mainly go to a new country for work reasons, and hence they commute daily trips.

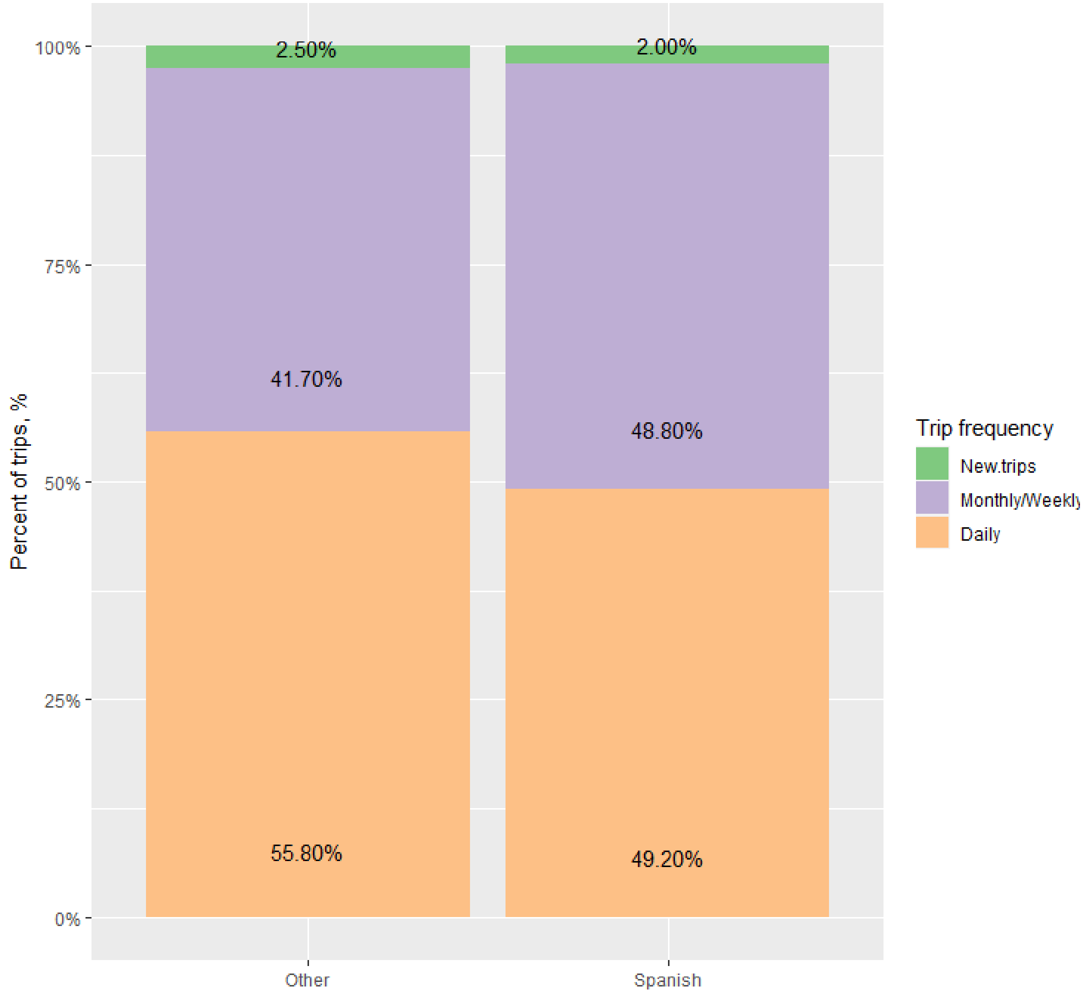

Figure 4.3 Frequency by nationality

Frequency by education level:

Education level doesn't seem to play a crucial role on the decision of making trips. The number of trips are well distributed between the two education levels. These findings are represented in Figure 4.4.

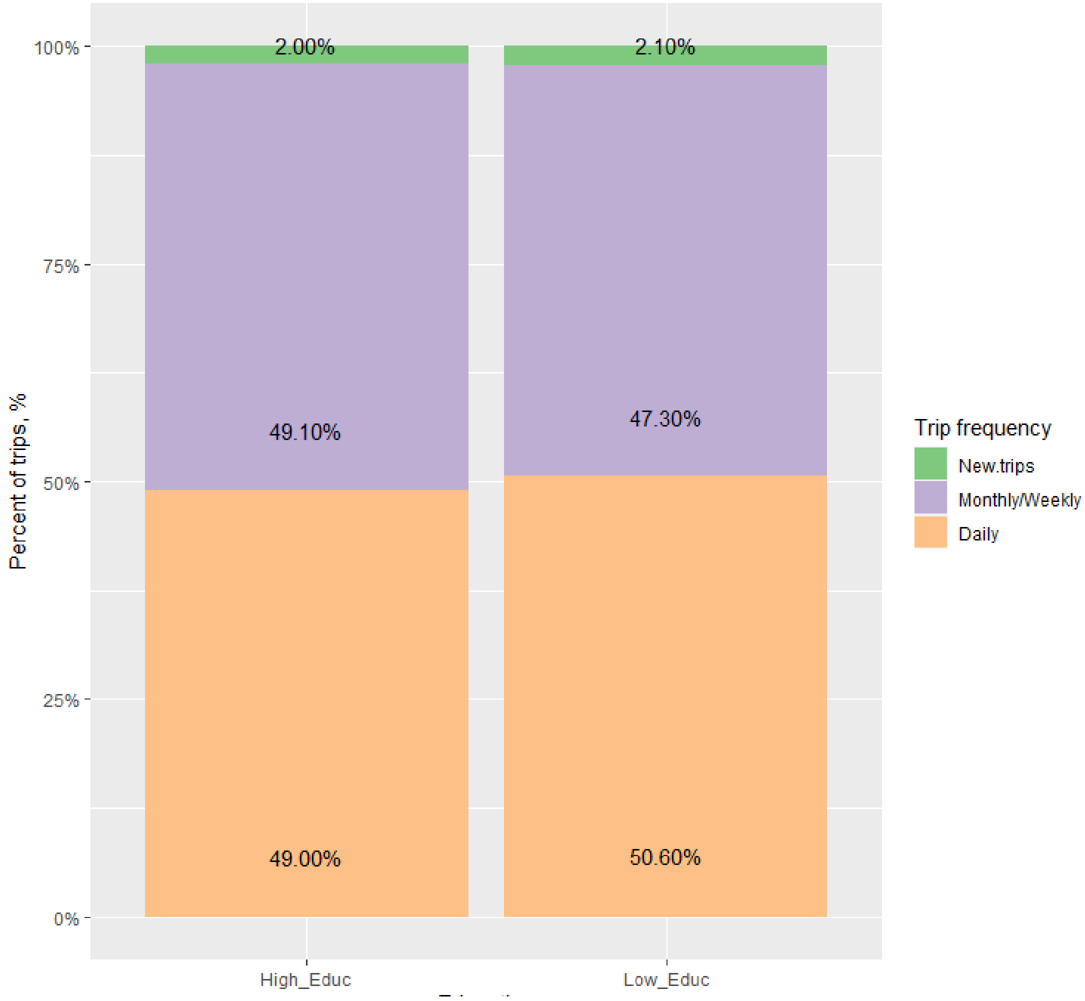

Figure 4.4 Frequency by education level

Frequency by household size:

People living with their families show a higher tendency to daily trips, while couples and singles to weekly or monthly trips (Figure 4.2). People living along or maximum with one more person show approximately the same behaviour, which is to make more weekly, monthly compared to families. These initial findings will be further explored in the model analysis.

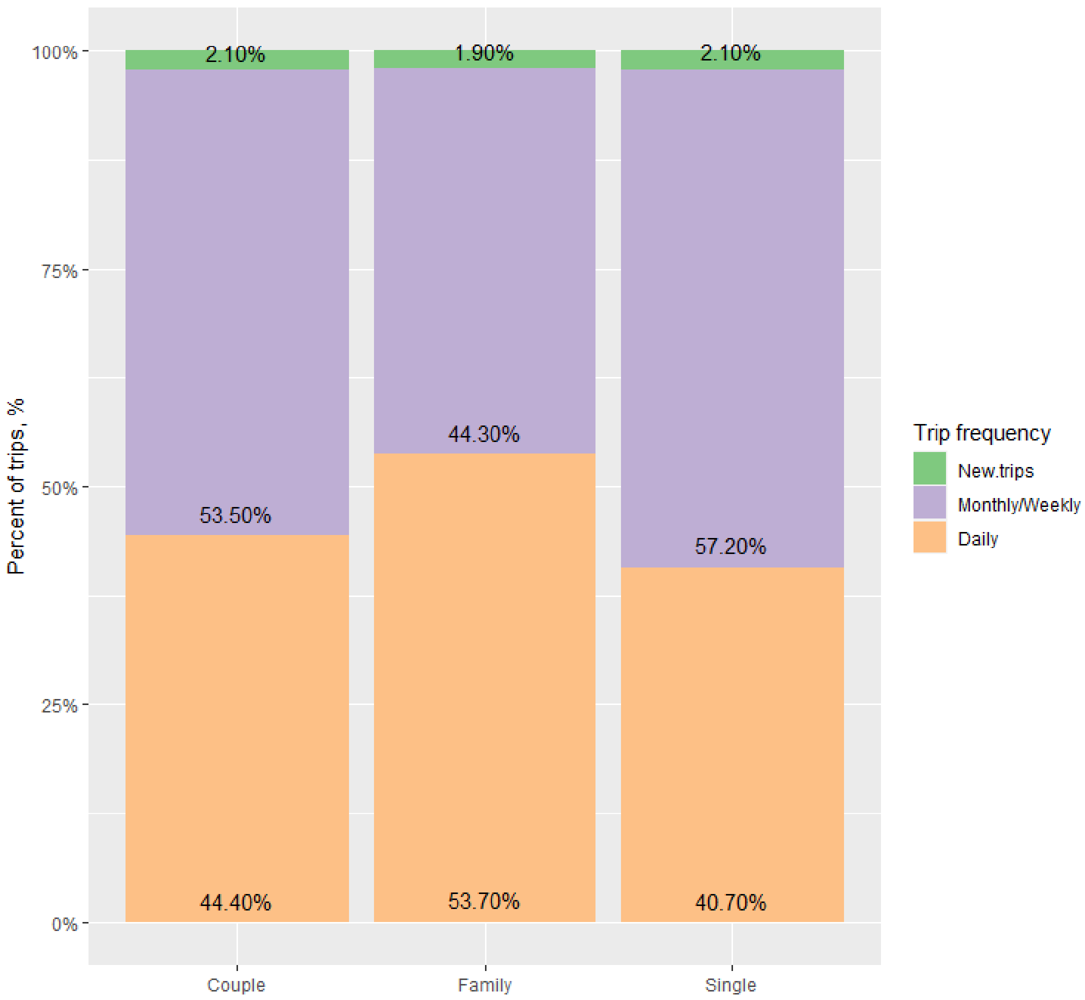

Figure 4.5 Frequency by HH size

Frequency by number of children under 4 years old:

The interesting finding that Figure 4.3 shows is as the number of young children increases, the chances of a family to make a new trips decreases.

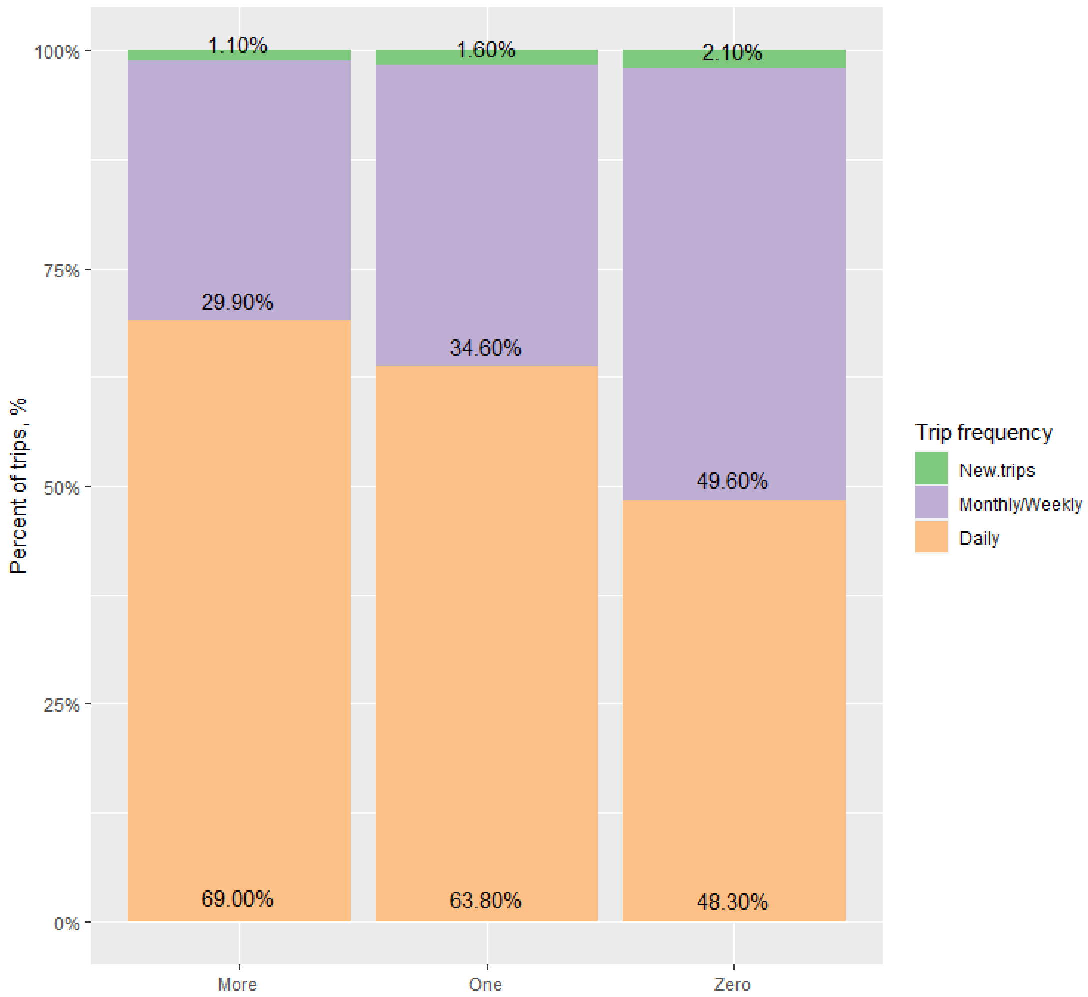

Figure 4.6 Frequency by number of children in the household

Frequency by number of vehicles:

The car ownership seem to help the decision of making trips (Figure 4.7). There is a linear relationship between the vehicle availability in the household and the new trips. More specifically, the higher the number of vehicles in the household the higher the chances that the members of the household will explore new destination.

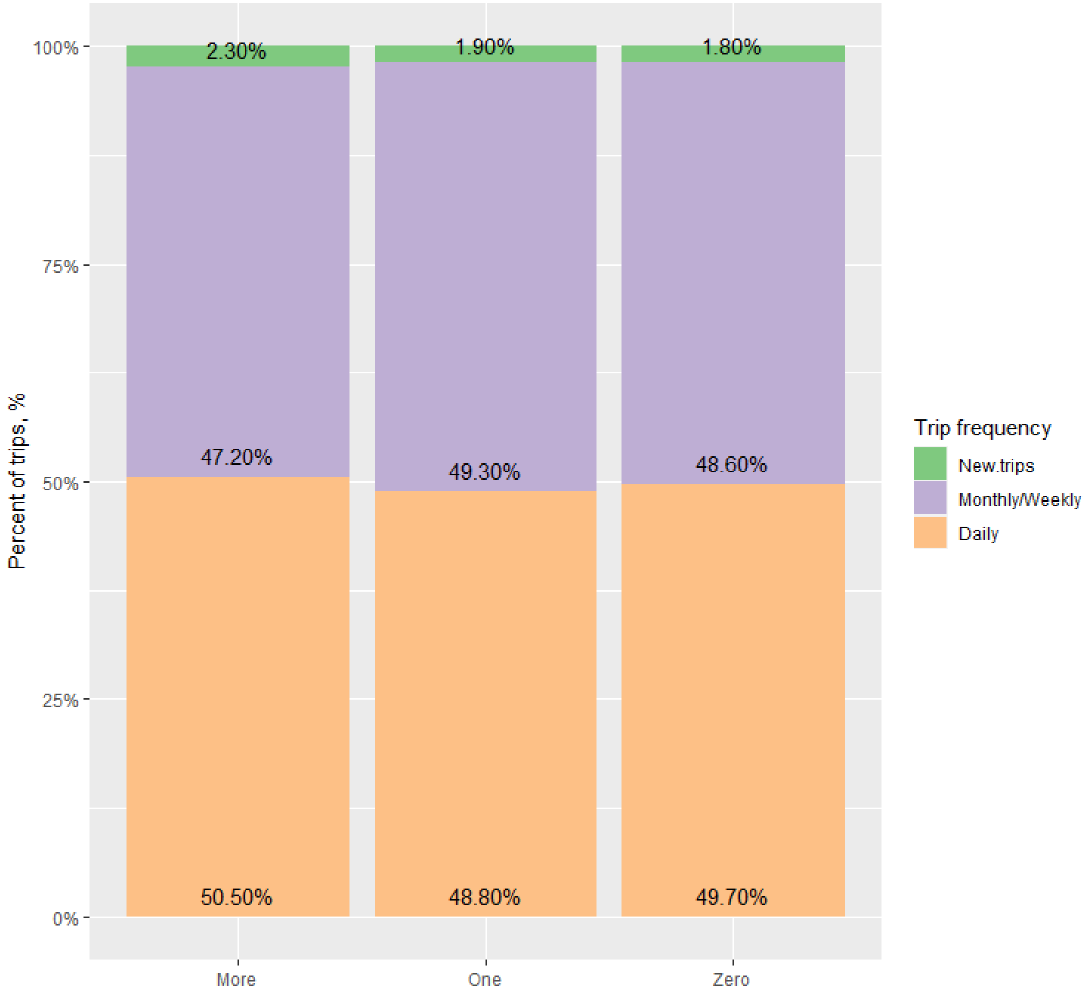

Figure 4.7 Frequency by vehicle ownership

Frequency by transport mode:

The interesting finding of the Figure 4.8 is that people using mainly shared mobility, taxis and private vehicles for making new trips. In addition, people show a high preference on using taxis for their weekly or monthly trips. Concerning the daily trips, they mainly take place by motorbikes, public transport or even vehicles, while cycling and walking are not rare options. The interpretation of those preferences will be further analyses in the model analysis.

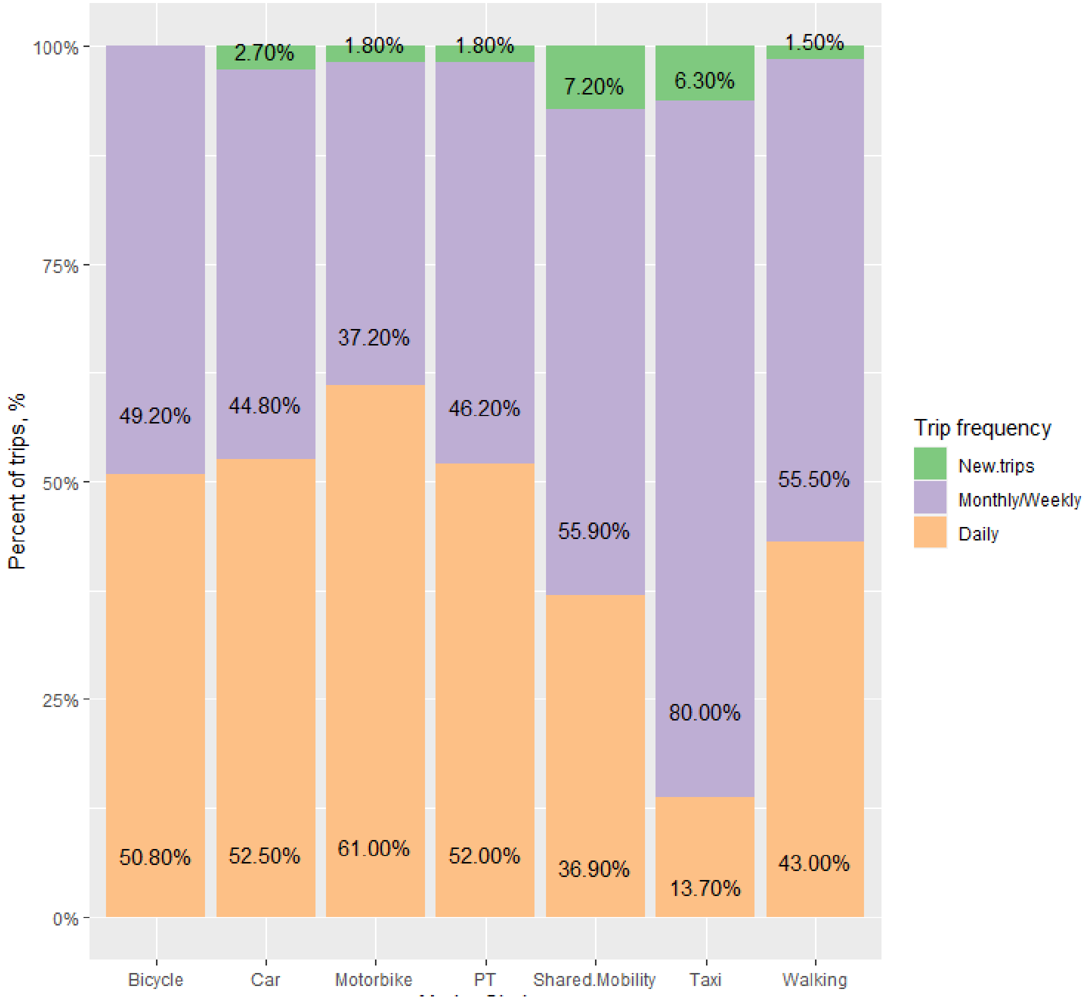

Figure 4.8 Frequency by transport mode

Frequency by PT card availability and driving licence:

A person can easily access a transport network by either holding a PT card or by having access to a vehicle. The persecute of the second option is to hold a driving licence. Figure 4.9 shows that respondents holding a **driving licence** have a tendency to make more weekly and monthly trips, while it doesn't influence the decision of making daily and new trips. In addition, people holding a **public transport card** are making more weekly or monthly trips, while the ones that don't have a public transport card are mainly make daily trips. Both, public transport card and driving licence don't seem to play a controversial role the decision of making a new trip.

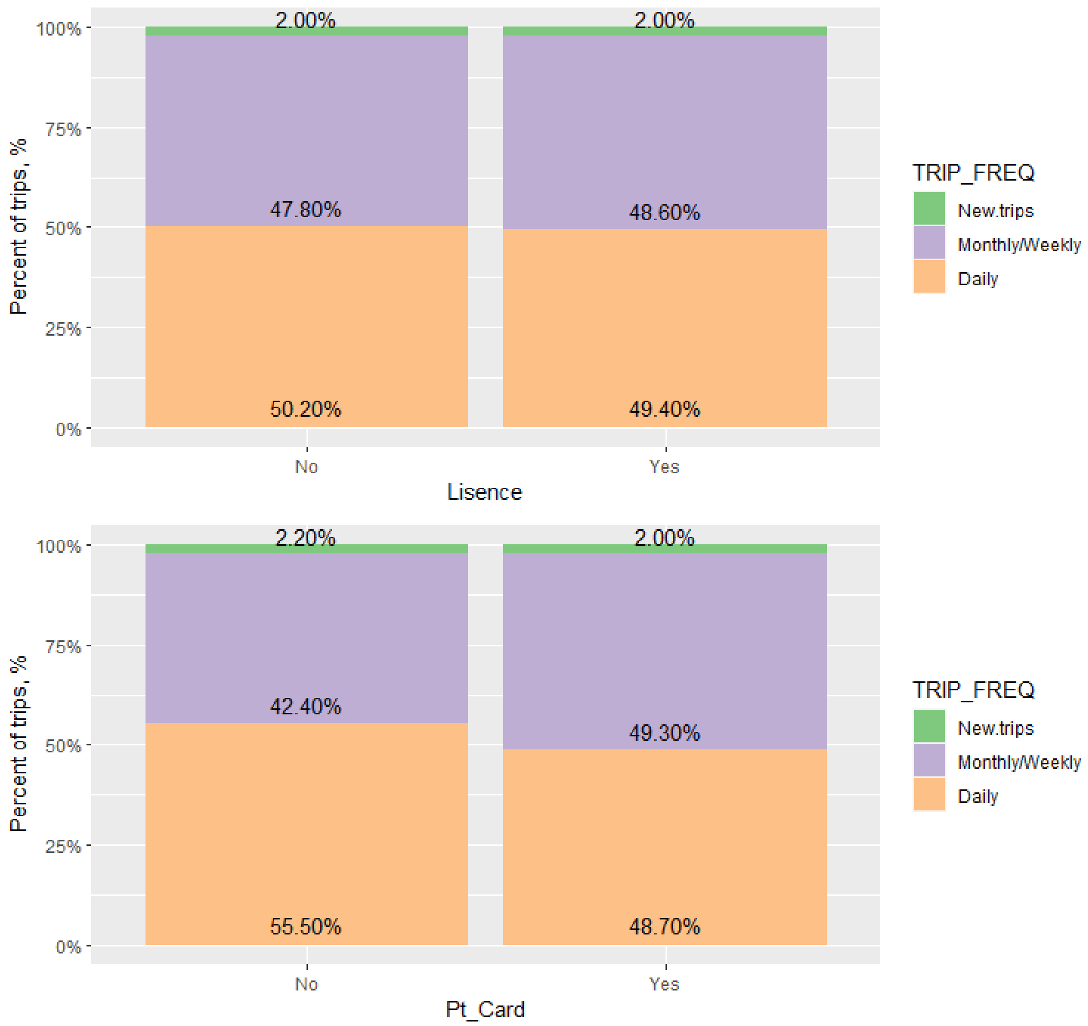

Figure 4.9 Frequency by licence and PT car availability

Frequency by month:

Months doesn't seem to be a high influential factor for trip decision making. These initial findings will be further investigated in the models (Figure 4.10).

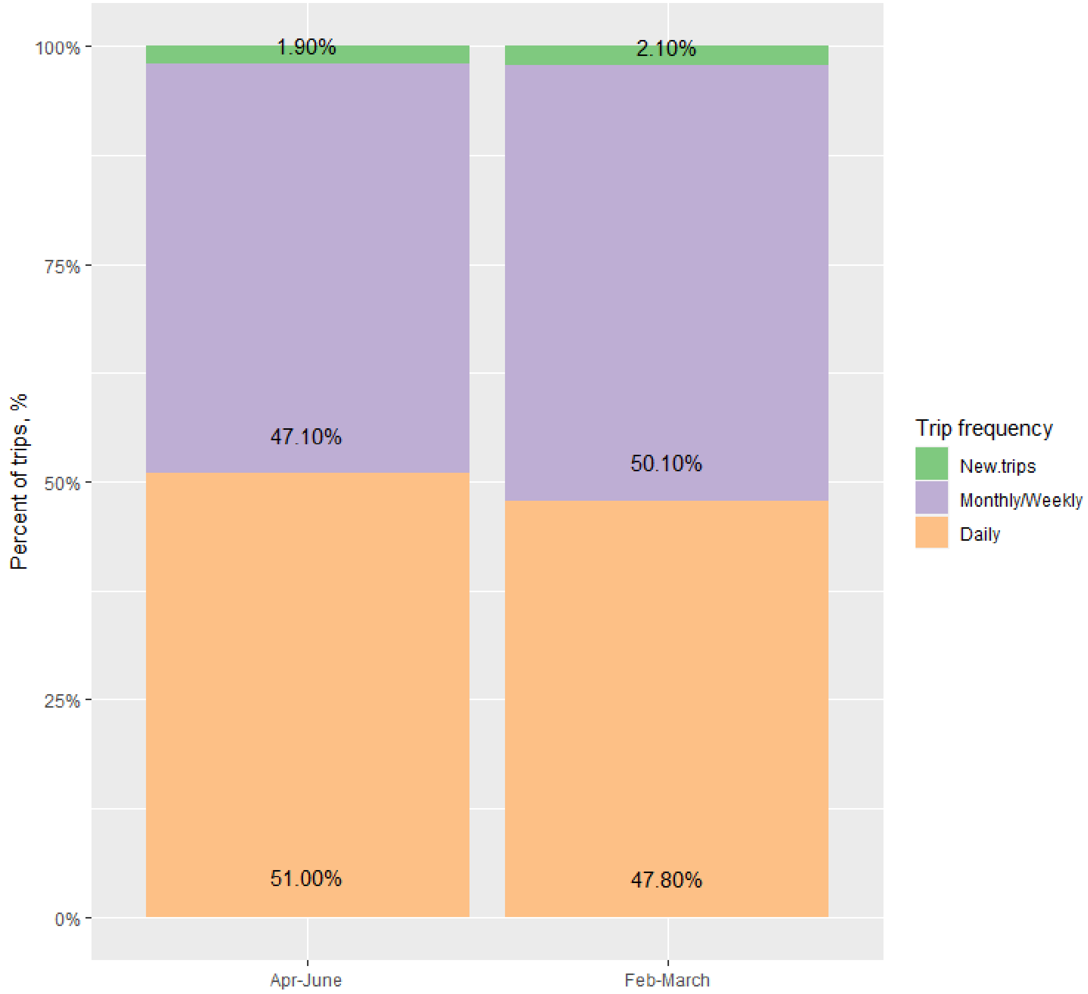

Figure 4.10 Frequency by month

Frequency by trip reason:

The provided dataset includes a high percentage of new trips that motivated from non-specified reasons. Based on the provided information, new trips are mainly taking place for private issues, including doctor visits, followed by sports and shopping reasons (Figure 4.11). The majority of daily trips are work trips, something expected. Additionally, most of the weekly/monthly are motivated from shopping reasons. A frequency which is normal when it comes for example to groceries.

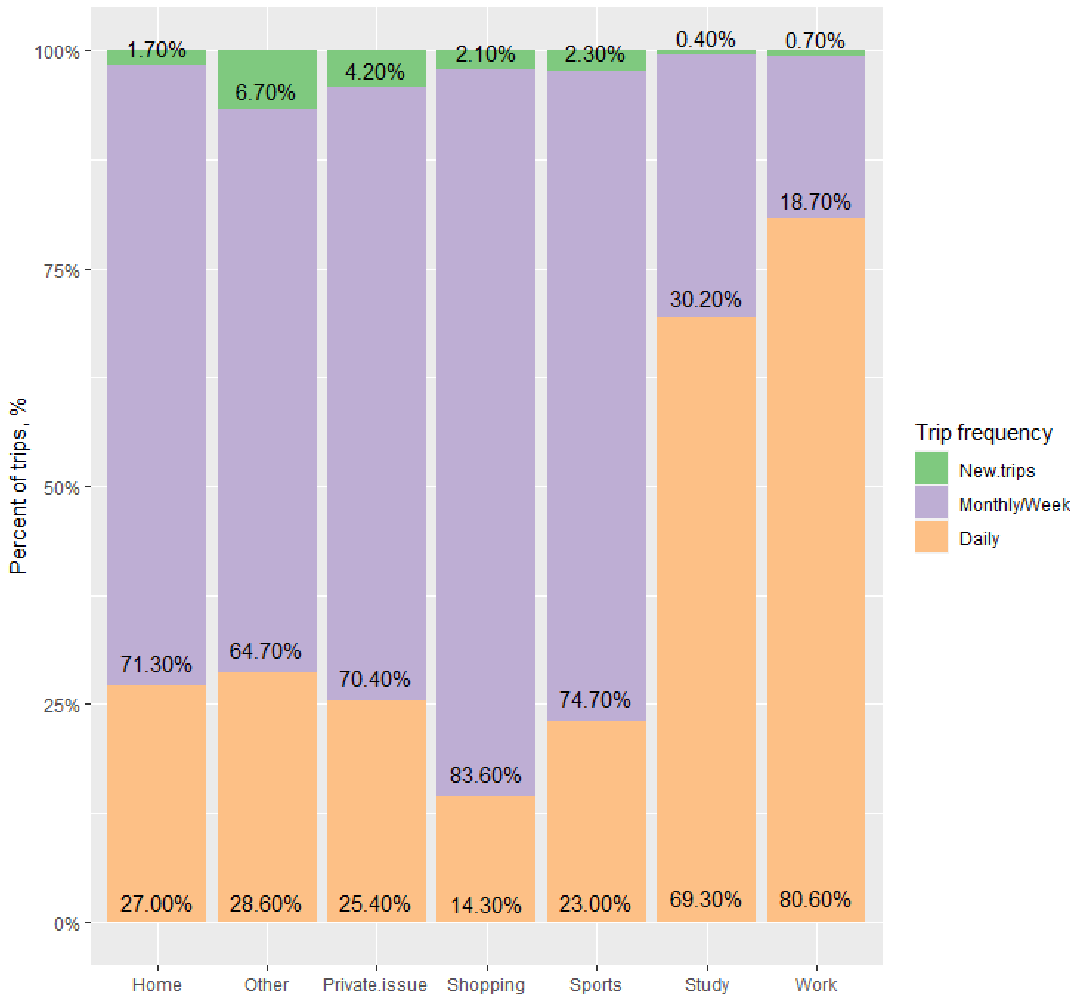

Figure 4.11 Frequency per trip reason

Frequency by user profile:

As already discussed, in order to investigate whether considering the family-life stage improves the model fit, and to which extend the car ownership conflict perspective help to understand the travel decision making, six users profiles were developed. The description of the six profiles developed are illustrated in the following:

Table 4.3 Profile description

Profiles	Description
Profile 1	Single with vehicle(s)
Profile 2	Single without vehicle
Profile 3	Couple with vehicle(s)
Profile 4	Couple without vehicle
Profile 5	Family with vehicle(s)
Profile 6	Family without vehicle

Figure 4.12 illustrates that people having a family and don't own a vehicle are the less probable to make new trips, followed by couples without vehicles. The interesting finding od this graph is that for single people, or in general people living alone is not significant the ownership of a vehicle for start making new trip. While vehicle availability plays a more controversial role on the decision of making a new trip for families and couple. These initial finding will be further explored in the model analysis.

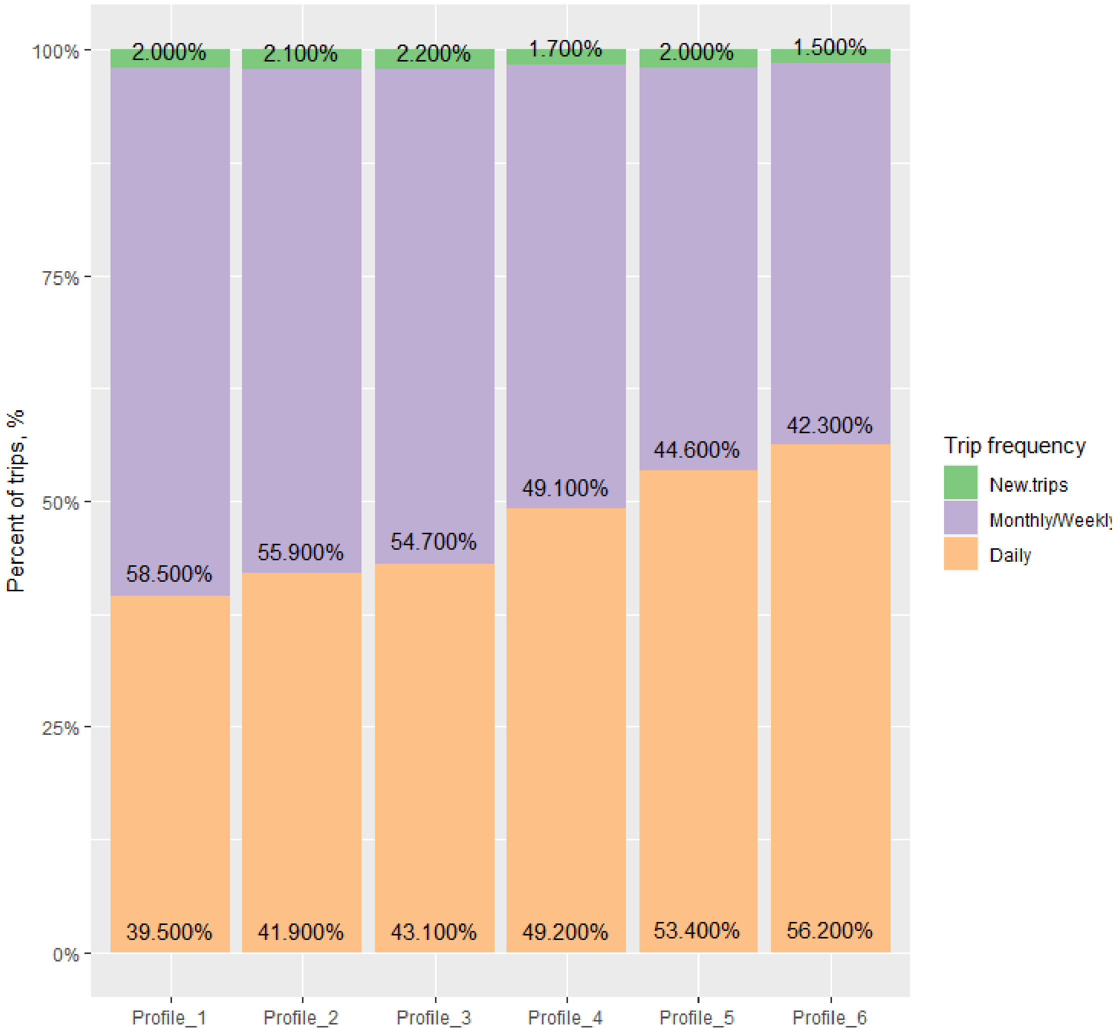

Figure 4.12 Trip frequency per profile

Frequency by trip distance:

The interesting finding from the Figure 4.14, is that daily trips have the highest mean distance. This finding can be explained by the fact that a lot of people living in Madrid commute long distances to go to their work. Slightly less is the mean distance of new trips, that mainly targets to explore new places, restaurants, stores and hence are more probable to located further from citizens' place of living. While finally the weekly and monthly trips mostly referring to shopping are the shortest trips, since people usually go for groceries or to sports facilities close to their neighborhood.

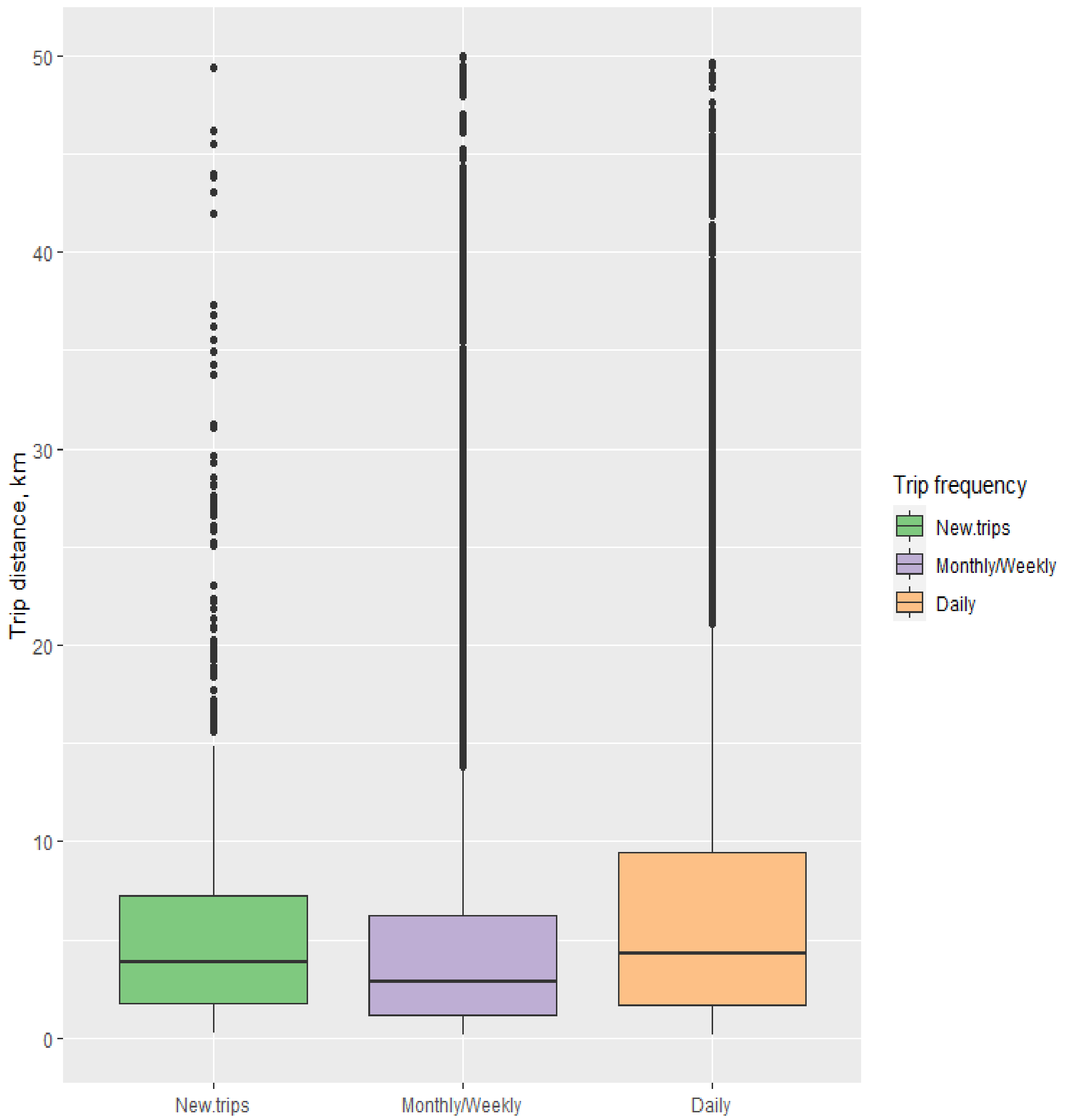

Figure 4.13 Trip frequency per distance

4.3 Qualitative analysis

In this section, a qualitative analysis of the original sample is presented through a careful examination of the graphs.

1. Gender doesn't seem to play a controversial role on the decision of making trips.
2. People over the age of 70 years old don't make use of shared mobility. New trips are mainly take place from people 18-35 and 55-70 years old.
3. Public transport card availability and driving licence don't seem to be crucial factors for the decision of making trips in Madrid. This phenomenon can be explained by the fact that a lot of trips in the city take place with cycling or even walking.
4. No preference on making trips between months was observed.
5. Education level doesn't seem to be important for trip decision making.
6. Vehicle availability in the household and the number of people leaving together seem to be a very important factor influencing the decision especially for new trips. These variables expected to be significant for the model.
7. People mainly use vehicles, taxis and shared mobility to commute to new destinations. Indication that the availability of shared mobility can indeed cause induced trips.
8. Finally, people commute long distances daily because work can be located far away, while they choose to commute short distances for their weekly and monthly needs, such as grocery shopping. New trips tend also to be long.

The data analysis in this section served as a preliminary understanding of the model building and the general directions of the research findings. Based on the general population attributes , better models can be specified in Chapter 5, leading to recommendations with strong policy implication.

5 Model Estimation

This section presents the modeling framework of this work, including the model specifications. Following the methodology discussed in Chapter 3, models were built. In this section, the results are also interpreted in terms of their meaning, and significance.

5.1 Variables considered

The variables considered were heavily dependent on the sociodemographics and in the general the characteristics of the dataset. It was observed a high correlation between private vehicle usage and Spanish nationality, so these variables were not used at the same model. The rest correlation coefficients were less than 0.7, and they considered in the analysis.

The Cramer's V correlation matrix for categorical variables is presented in Table 5.1 and Pearson's correlation for continuous variable in Figure 5.1.

Table 5.1 Cramer's V correlation matrix

	HH members older than 4	**Vehicles**	**Gender**	**License**	**PT card**	**Age group**	**Children**
HH members older than 4	1.00	0.27	0.05	0.04	0.04	0.14	0.05
Vehicles	0.27	1.00	0.07	0.36	0.16	0.07	0.03
Gender	0.05	0.07	1.00	0.13	0.08	0.04	0.01
License	0.04	0.36	0.13	1.00	0.11	0.15	0.02
PT card	0.04	0.16	0.08	0.11	1.00	0.10	0.05
Age group	0.14	0.07	0.04	0.15	0.10	1.00	0.14
Children	0.05	0.03	0.01	0.02	0.05	0.14	1.00

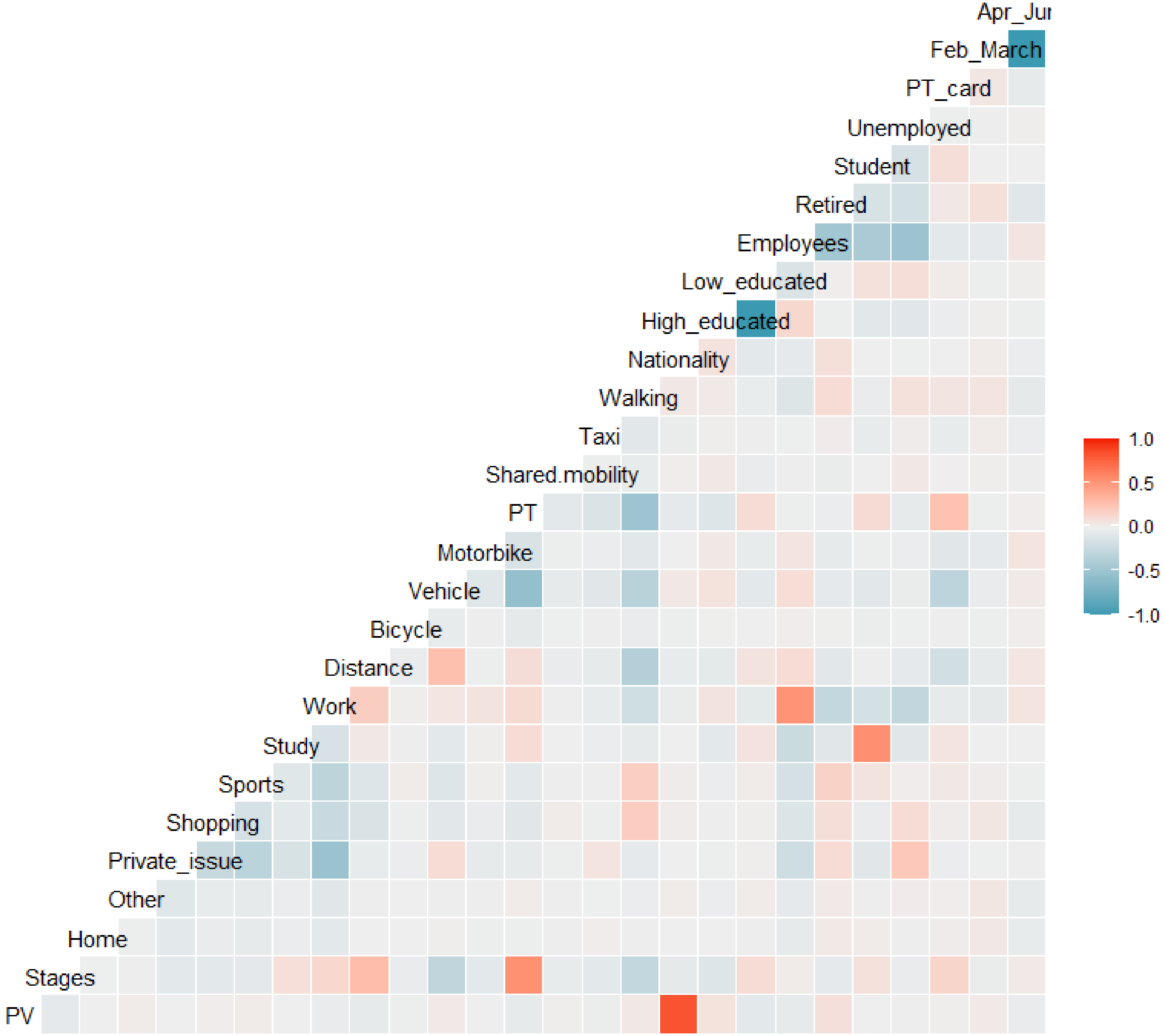

Figure 5.1 Pearson's Correlation

5.2 Behavioral Modelling

In this section, the model specification is presented. As discussed in the previous chapters, the aim is to develop models to be able to extract relevant factors in the decision making of new trips.

5.2.1 Independent variable models

The initial models included all outcomes. The resulting model gave estimate values with very large standard errors, meaning that the model was not correctly specified. The base alternative of the dependent variable was selected to be the frequency level "daily", since it is the level with the most observations and hence will help to take more robust results from the model. Modes 1.1 and Model 1.2 were estimated.

After selecting only significant (or relevant) estimates by conducting several runs, Model 1.1 and Model 1.2 was obtained as shown in Table 5.4.

Table 5.2 Estimation Results MNL Model 1.1 and NL 1.2

Variables	**Model 1.1**: Multinomial logit		**Model 1.2**: Nested Logit	
	Coefficient	S.E	Coefficient	S.E
Reference: Daily trips				
Individual characteristics				
New trips **Age group:** 18-34 (ref. = Age group 55-70)	0.10.	0.05	0.22***	0.06
Weekly/Monthly **Age group:** 18-34	-0.67***	0.05	-0.50***	0.06
New trips **Age group:** 35-54	-0.41***	0.04	-0.31***	0.05
Weekly/Monthly **Age group:** 35-54	-0.41***	0.04	-0.29***	0.04
New trips **Occupancy**: Student (ref.= Retired)	0.83***	0.06	0.73***	0.07
New trips **Occupancy:** Unemployed	1.13***	0.06	0.93***	0.07
Weekly/Monthly **Occupancy:** Unemployed	0.45***	0.06	0.28***	0.06
Weekly/Monthly **Occupancy:** Employees	-1.08***	0.04	-0.83***	0.08
Weekly/Monthly **Education:** High education (ref. =Low & No educ.)	0.13***	0.03	0.11***	0.02
New trips **February-March** (ref.= April-June)	0.08**	0.03	0.08**	0.03
Weekly/Monthly **PT card:** Yes (ref. = No PT card)	0.21***	0.05	0.16***	0.04
Household characteristics				
New trips **HH Size:** Single (ref.= Family)	0.59***	0.06	0.51***	0.07

Weekly/Monthly **HH Size:** Single	0.47***	0.06	0.33***	0.06
New trips **HH Size:** Couples	0.30***	0.04	0.26***	0.04
Weekly/Monthly **HH Size:** Couples	0.25***	0.04	0.19***	0.04
New trips **Vehicles:** One (ref.= Zero vehicles)	0.20***	0.05	0.17**	0.05
Weekly/Monthly **Vehicles:** One	0.20***	0.05	0.14***	0.04
New trip **Vehicle:** 2 or more	0.36***	0.06	0.31***	0.06
Weekly/Monthly **Vehicle:** 2 or more	0.31***	0.05	0.22***	0.05
New trip **Children:** One (ref.= Zero children)	-0.41***	0.08	-0.36***	0.08
Weekly/Monthly **Children:** One	-0.42***	0.08	-0.31***	0.06
New trip **Children:** 2 or more	-1.65***	0.18	-1.54***	0.18
Weekly/Monthly **Children:** 2 or more	-0.79***	0.14	-0.58***	0.11
Trip characteristics				
New trips **Reason:** Work (ref.= Private issue)	-1.90***	0.04	-1.87***	0.04
New trips **Reason:** Shopping	-0.52***	0.05	-0.52***	0.05
New trips **Reason:** Home	-1.07***	0.14	-1.06***	0.14
New trips **Reason:** Study	-3.00***	0.12	-3.00***	0.12
New trips **Reason:** Sports	-0.45***	0.04	-0.46***	0.04
New trips **Mode:** Shared mobility (ref.= Walking)	1.91***	0.20	1.92***	0.21
New trips **Mode:** taxi	3.15***	0.25	2.70***	0.21
Weekly/Monthly **Mode:** Taxi	2.12***	0.25	1.54***	0.24
New trips **Mode:** PT	0.38***	0.04	0.38***	0.04
New trips **Mode:** Motorbike	0.33*	0.13	0.37**	0.14
Weekly/Monthly **Mode:** Motorbike	-0.23*	0.11	-0.15.	0.09
New trips **Mode:** Vehicle	0.83***	0.05	0.85***	0.05
Weekly/Monthly **Mode:** Vehicle	-0.10*	0.04	-0.07*	0.03
Weekly/Monthly **Mode:** Bicycle	0.48*	0.20	0.28*	0.14
New trips **No of stages**	0.13	0.03	0.18***	0.04

Weekly/Monthly	-0.34	0.03	-0.24***	0.03
No of stages				
iv			0.72***	0.07
Loglikelihood	-24101		-24095	
*** Significant at the p=0.001 level ** Significant at the p = 0.01 level * Significant at the p = 0.05 level				

MNL Model 1.1:

The insights obtained from the first MNL model can be summarized as follows.

Age group: Keeping as a reference level the age group of 55-70 years old, it was found out that the youngest participants (18-34 years old) are more likely to make new trips. A finding that can be confirmed by the fact that young people are still interested in exploring new destinations and further expand their habits and activities. The weekly or monthly trips follow the opposite pattern, since older people have higher likelihood to make those kind of trips, fact that confirms the age distribution graph of the dataset. Coming to the age group of 35-54 they are in general less probable to make either new trips or weekly/monthly compared to the older ones. Something that confirms the initial finding from the descriptive analytics, which show that people 35-54 mainly do daily trips. In addition, people belonging in the second age group are more probable to have young children and in parallel work, statement that leads them to the lowest likelihood of commuting new trips.

Occupancy: Holding as reference level the retired people, students are the most probable to make new trips and slightly less are the unemployed people. Employees and unemployed people are less probable to make weekly trips compared to retired.

Education level: People holding a degree from a university or have to attend a professional training are more probable to make weekly or monthly trips.. The education level is indifferent to whether an individual will start making anew trip or not.

Months: New trips are more probable to take place during February and March compared to April until June. This can be partially explained b the fact that university students mainly have exams on spring months and this could be a reason for the preference of making new trips towards February and March. Another interpretation could that during April or May is the Easter, so people usually have holidays and hence they don't stay in Madrid to commute new trips.

Pt card: The availability of public transport card seem to be important for the weekly or monthly trips. People holding a public transport card have access partially or even entirely to the network of the city, and hence they can easily access their weekly destinations, such as sports, shopping etc.

Household size: Single people, or in general people living alone are more probable to make new trips compared to families. The same trend with a lower magnitude is followed by couples. Consequently, the higher the number of household members the more difficult is to organize a new trip. Weekly and monthly trips follow the same trend, taking place mostly by singles, then by couples and finally by families.

Vehicle ownership: People living in a household with more vehicles have higher likelihood to make new trips. Since they have the comfort of accessing their own vehicle, they can not only explore new destinations but also retain their weekly or monthly habits. The significance, the magnitude and the sign of this coefficient indicated the importance of vehicle ownership towards the decision of making a trip, either new or a frequent one.

Number of children under the age of 4: Families with children under the age of 4 years old are less probable to making not only new but also weekly or monthly trips. It should be mention that families with more than 2 young children are strongly negative correlated with the decision of making a new trip. Children, and especially young ones require a lot of care and hence, the families have less time for travelling.

Trip reason: Holding as reference level the trips motivated by private issues, it was found out that the rest reasons are less probable to motivate a new trip. Work, study and home trips are mainly take place daily or at least frequent during a week. Work and study trips are the most negative correlated variables, indicating that those trips hardly take place only once in the network. Shopping and sport trips usually take place less often compared to previous trip reasons and so the magnitude is less negative.

Transport mode: Modes, including shared mobility, taxis, public transport and private vehicles are more probable to be selected for new trips compared to walking, which was kept as the reference level. The option of cycling for new trips is not significant for the model. Concerning the weekly and monthly trips, are more probable to take place by bicycles, or even taxis compared to walking, with the rest options being less probable to be selected as mode for weekly or monthly trips. Exceptions are the public transport and shared mobility options which are non-significant for the model. Overall, taxis and shared mobility are the modes that mainly preferred for new trips and bicycles. While, bicycles and taxis are mostly selected for weekly and monthly trips.

Number of stages: There is a positive correlation between new trips and number of stages, while a negative one when it comes to weekly and monthly trips. Indicating that new trips tend to be part of a trip chain, since people can combine different activities or stops when they are exploring new different destinations. In contrary, when people do weekly or monthly trips people tend to make trips consisting of only one stage (ex. Groceries shopping, sports etc.).

Finally, the first insights from the graphical representation of the data set were confirmed, with **gender**, **driving license**, and **nationality** being indifferent to whether an individual will take a decision to commute a trip or not.

Evaluation of MNL Model 1.1

Confusion Matrix

A confusion matrix, or a misclassification matrix, was created in order to see the effectiveness of the model. The sum of each row represents the predicted number of observations while the sum of each column represents the actual number of observations. The diagonal cells are the matched observations while the non-diagonal cells show how much each level of frequency is misclassified as another level of frequency.

Table 5.3 Confusion Matrix-Model 1.1

PREDICTION	REFERENCE		
	Daily	**Monthly/Weekly**	**New trips**
DAILY	2213	22	701
MONTHLY/WEEKLY	593	28	1112
NEW TRIPS	308	76	1232

The accuracy of this first model is 55.3%. Indicating that the model can quite well predict and classify the observations to the appropriate frequency level.

Table 5.4 Accuracy Rate-Model 1.1

Accuracy	*55.3%*
95% CI	(54.0, 56.5)
No Information Rate	0.50
Kappa	0.30

In order to further explore and assess the performance of the predictive classification Model 1.1, and more specifically in individual classes some more metrics are calculated, which are presented in the following table:

Table 5.5 Model 1.1 performance on Individual Classes

	DAILY	**WEEKLY/MONTHLY**	**NEW**
SENSITIVITY/RECALL	0.71	0.22	0.40
SPECIFICITY	0.77	0.72	0.88
POS. PREDICTIVE VALUES/PRECISION	0.75	0.02	0.76
NEG. PREDICTIVE VALUES	0.73	0.98	0.61
PREVALENCE	0.50	0.02	0.48
DETECTION RATE	0.35	0.004	0.20
DETECTION PREVALENCE	0.47	0.28	0.26
BALANCED ACCURACY	0.74	0.47	0.64

The sensitivity of the model concerning the first class (daily trips) shows that the models predicts this class quite accurate, in contrast to the sensitivity of the other two classes, which is weak. The specificity among all classes of the model performs quite well, with all classes indicating a specificity rate above 72%. The precision of the model is accurate concerning the first and the third class (75% and 75% respectively), while it is weak when it comes to the second class (weekly/monthly trips). The negative predicted values indicate a trustful model in all the three classes, with the second class showing a 98% accuracy. The detection rate is weak when it comes to the second class, while the detection prevalence show a better performance on the first class. Overall, the balanced accuracy show the highest rate on the first class, followed by the third and finally the second one (with 74%, 64% and 47% respectively).

NL Model 1.2

After testing a multinomial logit model, a nested logit model was built. Two nests were created, one including the daily and the weekly/monthly trips and one only referring to the new trips. The results show that the model followed the same pattern as the Model 1.1, with the magnitudes of the estimated coefficients being a bit more smooth. The results of the 1.2 Model confirms the initial findings of the Model 1.1, regarding the significant estimated coefficients.

Overall, the variables that continue to be indifferent whether an individual will start making a new trips is the gender, the nationality, the availability of PT card and driving license. With PT card availability be significant for the weekly/monthly trips. Students, unemployed, and retired people are the most probable to make new trips. The category of employees show a negative significant coefficient only for weekly trips, confirming the initial findings that the average age of people doing weekly trips is relatively high. Education level is only significant for weekly/monthly trips, indicating a positive sign for higher educated people. Based on the data set during February and March new trips have higher likelihood to take place, although there is not available any seasonality so no further results can be extracted from this finding.

Household size, vehicle ownership and the number of children under the age of 4 years old show a monotonic relationship with trip making decision. The more the number of household members and the more the children under the age of 4 the less the chances that the household members will make the decision of making either a new trip or a more frequent one (weekly or monthly) . In contrast with the vehicle availability, which is positively correlated with the trip decision making. Indicating that the higher the number of vehicles that are available in a household, the higher the chances the members of the household will make either new or frequent trips.

New trips are more probable to be motivated from private issues, shopping or sports compared to home, work and study reasons. While they are also mostly probable to take place with shared mobility vehicles, taxis or even motorbikes. Finally, new trips have higher likelihood to be part of a trip chain, compared to weekly/monthly trips which are mostly consisting of only one stage.

Evaluation of NL Model 1.2

Confusion Matrix

The performance of the model show approximately the same results as the Model 1.1, with the accuracy being 0.05% less than the previous model. The metric concerning the performance of each class separately also show the same results as the previous model. Consequently, both multinomial and nested logit models can be selected for the predictions of the characteristics of the new trips, with slightly better performance that of multinomial logit model's.

Table 5.6 Confusion Matrix-Model 1.2

PREDICTION	REFERENCE		
	Daily	**Weekly/monthly**	**New**
DAILY	2204	22	716
WEEKLY/MONTHLY	608	28	1115
NEW	302	76	1214

Table 5.7 Accuracy Rate-Model 1.2

Accuracy	*54.8%*
95% CI	(53.6, 56.1)
No Information Rate	49.6
Kappa	0.29

Table 5.8 Model 1.2 Performance on Individual Classes

	DAILY	**WEEKLY/MONTHLY**	**NEW**
SENSITIVITY/RECALL	0.71	0.22	0.40
SPECIFICITY	0.77	0.72	0.88
POS. PREDICTIVE VALUES/PRECISION	0.75	0.02	0.76
NEG. PREDICTIVE VALUES	0.73	0.98	0.61
PREVALENCE	0.50	0.02	0.48
DETECTION RATE	0.35	0.004	0.19
DETECTION PREVALENCE	0.47	0.28	0.25
BALANCED ACCURACY	0.74	0.47	0.64

5.2.2 Models with user's profiles

To investigate whether considering the household size combined with the vehicle ownership improves the fit of the previous models , two further models were designed: Model 2.1 being a multinomial logit model including 6 new variables representing the full interaction of household size and vehicle ownership and the Model 2.2 being a nested logit model, composed by two nests and the full interaction between household size and vehicle ownership. The models' goodness-of-fits were compared with the Akaike information criterion (AIC). The model with the best combination of as small as possible AIC and as high as possible accuracy is preferred (Table 5.18).

Table 5.9 Estimation Results MNL Model 2.1 and NL 2.2

Variables	Model 2.1: Multinomial logit		Model 2.2: Nested Logit	
	Coefficient	S.E	Coefficient	S.E
Reference: Daily trips				
Individual characteristics				
New trips **Age group:** 18-34 (ref.= Age group 55-70)	0.11 *	0.05	0.23***	0.06
Weekly/Monthly **Age group:** 18-34	-0.67***	0.05	-0.49***	0.06
New trips **Age group:** 35-54	-0.42***	0.04	-0.31***	0.05
Weekly/Monthly **Age group:** 35-54	-0.42***	0.04	-0.30***	0.04
New trips **Occupancy:** Student (ref.= Retired)	0.83***	0.06	0.73***	0.07
New trips **Occupancy:** Unemployed	1.12***	0.06	0.92***	0.07
Weekly/Monthly **Occupancy:** Unemployed	0.45***	0.06	0.28***	0.06
Weekly/Monthly **Occupancy:** Employees	-1.08***	0.04	-0.83***	0.08
Weekly/Monthly **Education:** High education (ref.= Low ed.)	0.14***	0.03	0.12***	0.02
New trips **February-March** (ref.= April-June)	0.09**	0.03	0.08**	0.03
Weekly/Monthly **PT card:** Yes (ref.= No PT card)	0.21***	0.05	0.15***	0.04
New trips **Profile 1:** Single with vehicle(s) (ref.= Profile 6: Family without vehicle)	0.67***	0.11	0.57***	0.11
Weekly/Monthly **Profile 1:** Single with vehicle(s)	0.62***	0.10	0.45***	0.09
New trips **Profile 2:** Single without vehicle	0.71***	0.10	0.65***	0.10
Weekly/Monthly **Profile 2:** Single without vehicle	0.37***	0.10	0.25***	0.08
New trips **Profile 3:** Couple with vehicle(s)	0.57***	0.08	0.50***	0.08
Weekly/Monthly **Profile 3:** Couple with vehicle(s)	0.42***	0.07	0.30***	0.06
New trips **Profile 4:** Couple without vehicle	0.19.	0.10	Non-significant	
Weekly/Monthly **Profile 4:** Couple without vehicle	0.16.	0.09	0.12.	0.07
New trips **Profile 5:** Family with vehicle(s)	0.27***	0.08	0.25**	0.08
Weekly/Monthly **Profile 5:** Family with vehicle(s)	0.17*	0.07	0.25*	0.08

Household characteristics				
New trips **HH Size:** Single (ref. = Family)	NC	NC	NC	NC
Weekly/Monthly **HH Size:** Single	NC	NC	NC	NC
New trips **HH Size:** Couples	NC	NC	NC	NC
Weekly/Monthly **HH Size:** Couples	NC	NC	NC	NC
New trips **Vehicles:** One (ref.= Zero vehicles)	NC	NC	NC	NC
Weekly/Monthly **Vehicles:** One	NC	NC	NC	NC
New trip **Vehicle:** 2 or more	NC	NC	NC	NC
Weekly/Monthly **Vehicle:** 2 or more	NC	NC	NC	NC
New trip **Children:** One (ref.= Zero children)	-0.41***	0.08	-0.35***	0.08
Weekly/Monthly **Children:** One	-0.43***	0.08	-0.31***	0.06
New trip **Children:** 2 or more	-1.60***	0.18	-1.49***	0.18
Weekly/Monthly **Children:** 2 or more	-0.78***	0.14	-0.57***	0.11
Trip characteristics				
New trips **Reason:** Work (ref. = Private issue)	-1.89***	0.04	-1.87***	0.04
New trips **Reason:** Shopping	-0.52***	0.05	-0.52***	0.05
New trips **Reason:** Home	-1.06***	0.14	-1.06***	0.14
New trips **Reason:** Study	-3.00***	0.12	-2.99***	0.12
New trips **Reason:** Sports	-0.45***	0.04	-0.46***	0.04
New trips **Mode:** Shared mobility (ref. = Walking)	1.92***	0.20	1.93***	0.21
New trips **Mode:** taxi	3.16***	0.25	2.71***	0.23
Weekly/Monthly **Mode:** Taxi	2.12***	0.25	1.54***	0.24
New trips **Mode:** PT	0.38***	0.04	0.38***	0.04
New trips **Mode:** Motorbike	0.34*	0.13	0.38**	0.14
Weekly/Monthly **Mode:** Motorbike	-0.22.	0.11	Non-significant	
New trips **Mode:** Vehicle	0.87***	0.05	0.88***	0.05
Weekly/Monthly **Mode:** Vehicle	-0.07.	0.04	Non-significant	
Weekly/Monthly **Mode:** Bicycle	0.48*	0.20	0.27***	0.14

New trips **No of stages**	0.14***	0.03	0.18***	0.04
Weekly/Monthly **No of stages**	-0.34***	0.03	-0.24***	0.03
iv			0.72***	0.08
Loglikelihood	-24101		-24095	
NC: Not considered				

MNL Model 2.1

The multinomial model including the commuter's profiles overall showed similar results with the multinomial logit model including all the variables independently. Gender was found out once more to be indifferent whether an individual will make a new or even a frequent trip. People belonging in the first age group (18-34 years old) show the highest likelihood to make new trips and explore new destinations, followed by the people belonging to the last age group (55-70 years old). Finding that confirms one more the fact that young people are still interested in exploring new places and further expands their hobbies. While people of the last age group have more free time compared to people 35-54 years old, with the second ones have mainly young children and hence more family responsibilities and consequently less free time for new trips.

Concerning the occupancy, unemployed, followed by students and then by retired people are the commuter's that show higher likelihood to make new trips. The variables regarding the education level and the month that the trip is taking place show the same performance as before, with the first variable being indifferent whether will an individual commute a new trip or not and the second one show a positive magnitude on February and March for making a new trip. The availability of driving license is not significant for deciding or not make a new or a frequent trip, while people holding a public transport card show higher likelihood to commute weekly/monthly trips. Household size and car availability is not considered in this model, since new variables representing the fully interaction of those two variables was developed and tested in the model.

Children showed once more that contribute in a negative way on making not only new but also frequent trips (weekly/monthly), indicating that children add more responsibilities in a household and hence the time of trips is consequently reduced. Concerning the reasons that motivate a new trip the most dominant one are the private issues, followed by shop and sport trips. While, concerning the weekly/monthly trips the trip reasons are not significant. Shared mobility and taxis are mainly selected as transport mode for new trips, confirming once more the finding from the initial models. Number of stages is positively correlated with new trips and negative correlated with the weekly or monthly trips.

Six new variables were created for the purposes of this study, indicating as already discussed the fully interaction of the household size and the vehicle ownership. The new variables were named as Profile 1, Profile 2, Profile 3, Profile 4, Profile 5, Profile 6. Profile 6 which represents the people living with their families and don't own any vehicle kept as the base category. Based on the results all the rest profiles are more likely to commute new trips, with the single people either with vehicle or without having the highest magnitude. The next profiles that show high likelihood to make new trips are the couples. The interesting finding here, is that although in the category of singles the car availability doesn't significantly influence the decision of making a new trip, concerning the couples the vehicle ownership does increase the chances of the couple to make new trips. Profile 5, representing the people living with their families and od own a vehicle, show higher likelihood to commute new trips. Overall, singles show the highest likelihood to make not only new but also weekly trips, while the vehicle ownership does not influence the decision of making trips for this category. The next possible category to make new trips are the couples and followed by the families. In couples and families, the vehicle ownership does influence the decision of making trips.

Evaluation of Model 2.1

Confusion Matrix

The addition of the profiles in the model testing, show slightly better accuracy, indicating that the fully interaction of the household size and the vehicle ownership can improve the overall model's performance. It should be mentioned that the differences are not huge, and hence either the models with the independent variables or the ones with the profiles can be used to predict the characteristics that influence the decision of making new trip.

Table 5.10 Confusion Matrix Model 2.1

PREDICTION	REFERENCE		
	Daily	**Weekly/monthly**	**New**
DAILY	2223	24	704
WEEKLY/MONTHLY	579	28	1099
NEW	312	74	1242

Table 5.11 Accuracy rate Model 12.1

Accuracy	*55.6%*
95% CI	(54.3, 56.8)
No Information Rate	0.50
Kappa	0.30

Table 5.12 Model 2.1 performance on individual classes

	DAILY	**WEEKLY/MONTHLY**	**NEW**
SENSITIVITY/RECALL	0.71	0.22	0.41
SPECIFICITY	0.77	0.73	0.88
POS. PREDICTIVE VALUES/PRECISION	0.75	0.02	0.76
NEG. PREDICTIVE VALUES	0.73	0.98	0.61
PREVALENCE	0.50	0.02	0.48
DETECTION RATE	0.35	0.004	0.20
DETECTION PREVALENCE	0.47	0.27	0.26
BALANCED ACCURACY	0.74	0.47	0.64

NL Model 2.2

The nested model with profiles show overall similar results with the multinomial logit models with profiles. Some small differences are the following:

- Profile 4, representing the couples without vehicles is not significant for the model's performance.
- Vehicles is not significant for making weekly/monthly trips, whereas in multinomial model was significant.
- Motorbikes is not significant for making weekly/monthly trips, whereas in multinomial model was significant.

Evaluation of Model 2.2

Confusion Matrix

The addition of the profiles in the model testing, show slightly better accuracy and in the case of nested models, indicating that the fully interaction of the household size and the vehicle ownership can improve the overall model's performance. Not only in the case of multinomial logit models but also for the nested models, the differences are not huge, and hence either model with the independent variables or the model with profiles can be used to predict the characteristics that influence the decision of making new trip.

Table 5.13 Confusion Matrix Model 2.2

PREDICTION	REFERENCE		
	Daily	**Weekly/monthly**	**New**
DAILY	2223	24	703
WEEKLY/MONTLY	582	28	1098
NEW	309	74	1244

Table 5.14 Accuracy rate Model 2.2

Accuracy	*55.6%*
95% CI	(0.54, 0.57)
No Information Rate	0.50
Kappa	0.30

Table 5.15 Model 2.2 performance on individual classes

	DAILY	WEEKLY/MONTHLY	NEW
SENSITIVITY/RECALL	0.71	0.22	0.41
SPECIFICITY	0.77	0.73	0.88
POS. PREDICTIVE VALUES/PRECISION	0.75	0.02	0.76
NEG. PREDICTIVE VALUES	0.73	0.98	0.61
PREVALENCE	0.50	0.02	0.48
DETECTION RATE	0.35	0.004	0.20
DETECTION PREVALENCE	0.47	0.27	0.26
BALANCED ACCURACY	0.74	0.47	0.65

Table 5.16 Evaluation of models

Model	Log-Likelihood	McFadden R^2	AIC
1.1 MNL Independent	-24101	0.13	48291.03
1.2 Nested Independent	-24095	0.13	48281.39
2.1 MNL Profiles	-24101	0.13	48295.47
2.2 Nested Profiles	-24095	0.13	48285.99

Based only on the AIC values the model 1.2 is the best one. Based on the accuracy of the models combined with the AIC values, the best model is 2.2, a nested model including the fully interaction of household size and vehicle ownership, with an accuracy rate 55.6% and AIC equals to 48285.99. Although model 2.2 show the best performance, the differences with the rest ones is not significant, and hence both multinomial and nested logit modes could be used for the identification of the characteristics of new trips. Using profiles in the model slightly increase the overall performance of the model.

6 Discussion

This chapter discusses the findings of the thesis relates them to the research problems and objectives stated in the introduction. The main aim of this discussion is to relate the findings to research directions.

6.1 Discussion of main results

The main findings of this study can be divided into two components: the direct findings from the descriptive analytics of the dataset, and the findings obtained from the logit models.

6.1.1 Descriptive analytics findings

The descriptive analytics displayed in Table 4.2 show that the main reasons that motivated new trips are either private issues, including doctor visits, shopping or even sports. The rest trip reasons are mainly referring to daily trips such as, home-, work-, or study-trips.

The dataset results indicated a high impact of sociodemographics factors on decision making of new trips. **Younger** respondents seemed to be more enthusiastic about travelling towards a new destination and less about weekly or monthly trips (Figure 4.2). It should be mentioned, that according to a recent study (Burkhardt & Millard-Ball, 2006) in North America the same age group (people until 35 years old), show the higher likelihood to commute with shared mobility. While there are not many members below the age of 21, since they are not allowed to hold a driving license until then. So, young people until their early 30s show higher likelihood to not only commute to new destinations but also make use of new transport modes.

People without **children** show a much higher likelihood to perceive new trips (Figure 4.6), while as the number of children increased in the household, the probability for the household members to make new trips decreased accordingly. A similar effect is also observed for the people who **own one or more vehicles** , indicating that the comfort of having access to a vehicle increases the chances to make new trips (Figure 4.7). **Cultural** impact was not significant, although people who don't hold a Spanish nationality to be slightly more enthusiastic about exploring new destinations.

People show a high preference towards shared mobility and taxis for their new trips (Figure 4.8). Moreover **gender**, the availability of **driving licence** and **public transport** card didn't seem to be crucial for the decision of making a new trip.

These initial descriptive statistics of the dataset show the first insights, based on which different models were developed. The factors such as the age group, the availability of vehicle(s) and the number of children in the household were expected to be significant for the model.

6.1.2 Model findings

In the following, we summarize the model findings, which are developed based on the variables discussed not only independently but also by the fully interaction of some of them.

Based on the initial findings of the data it was found out that some of the most crucial variables for the decision of making a new trip is the vehicle ownership and the household size. For this reason the fully interaction of those two variables was represented by a new variable, and hence six commuter's profiles were created. Two kind of discrete choice models were used, the first was the multinomial logit model and the second was the nested logit model, which as a first attempt tested all the variables independently and as a second the profiles were incorporated.

Models revealed interesting findings on the characteristics that strongly influence the decision of making a new trip and support the initial findings from the initial analysis of the data. The importance of **age** and of **life stage** were highly significant, with the young people until the age of 35 years old, the ones who live alone without the responsibilities of an entire family and the people being either students of unemployed being the most probable commuters for trips that only took place once in the network. Other important factors is the number of **children** that an adult has, the chances drop accordingly to the number of babies that exist in the household. This lead us to the conclusion that not only the age but also the existence of family and children are significant for the enthusiasm of making new trips. Moreover, the main reasons that motivate new trips are the **private issues**, including the doctor visits, **sports** and **shopping**. The most dominant transport mode choices for making a new trip are **shared mobility** (vehicles and bicycles) and **taxis**. Finally, it is positively correlated the decision of exploring new destinations with the **number of stages** that a trip consist of.

Overall, the descriptive and model finding show a clear pattern of the characteristics that have the trips that took place only once. This could be generalized to other datasets, by examining the young people until the age of 34, owning one or more vehicles, living alone or at most with one more person as the most probable commuters of new trips. By applying clustering methods, the new trips that this targeted group do, can partially or even entirely assumed to be induced demand. Other crucial factors that should be examined in every case are the reasons that motivated the trips, and the modes that were used to. Private issues and shopping are the most dominant factors that can motivate a new trip. While shared mobility (vehicles and bicycles) and taxis are highly significant modes for new trips.

The findings of the models, cannot only used as generalized findings to other studies but also as an input for the second step of the proposed methodology. More specifically, the significant parameters that previously mentioned can be used in order to create some user groups for the distinction of the induced from the non-induced demand. A combination of those parameters can lead us to the distinction of induced demand.

Figure 6.1 illustrates a proposed combination of significant coefficients in order to identify the target variable, which is the increment in travel demand.

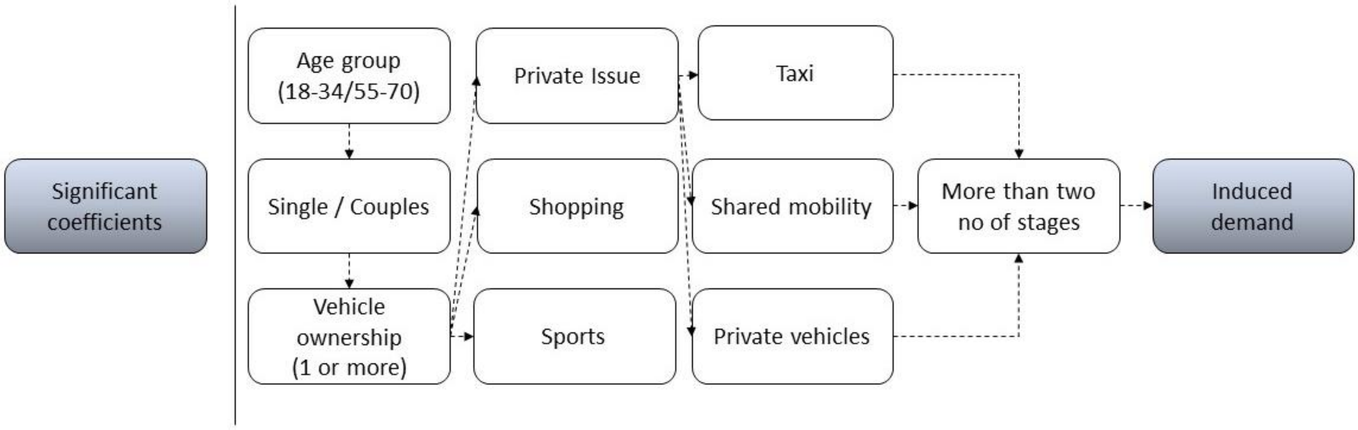

Figure 6.1 Configuration of induced demand

As already discussed, the most influential factors for the decision of making a new trips are:

- Age group
- Household size
- Vehicle ownership
- Children in the household
- Trip reason
- Transport mode
- Number of stages

By combining, the level of those variables that show the highest tendency towards new trips, induced demand could be identified. Some possible combinations of variables that could be partially or even entirely assumed to be induced demand are:

- New trips that took place from young people, were motivated from shopping, private issues or sports reason, commuted by taxi or shared vehicles/bicycles and consisted of more than one stages.

The trips that are targeted can be even more precisely identified by adding some more sociodemographics:

- New trips that took place from young people, living alone or with at most with one more person, own one or more vehicles, were motivated from shopping, private issues or sports reason, and commuted by taxi or shared vehicles/bicycles.

It should be mentioned, that these are only some proposals for combining some of the influential variables for the decision of making a new trip. Other combinations could also be used. For example, people belonging to the age group 55-70 years old are also possible commuters of new trips. It is obvious that in the proposed variables' combinations there is a gradually increment on the variables that are combined. As the numbers of variables that are combined increased, it is more probable that the result will indicate entirely induced trips.

Figure 6.2 illustrates some combination proposals of the aforementioned variables. More precisely, a subsample filtered, consisting of the age groups that are the most probable to make new trips (18-34 and 55-70 years old), including only the ones who are living alone or with one more person. Finally, the most probable reasons that motivate new trips were filtered (shopping, private issues and sports). Having already mentioned, that people who own one or more **vehicles** are the most probable commuters of new trips, 58 of the trips represented in the upper-left graph could assumed to be partially or even entirely induced demand. Consequently, by combining **age group**, **household size**, **trip reason** and **vehicle ownership** 0.18% (58 out of 31,428 observations) of the trips in the dataset can consider to be induced demand.

Another influential factor for the decision making of a new trip is **mode choice.** The most dominant choices for new trips (based on their estimated coefficient's magnitude), are the shared mobility and taxis. Hence, by combining **age group**, **household size**, **trip reason and mode choice,** the 10 trips from the upper-right graph can consider to be induced demand (0.03% from the total observations).

Finally, as already discussed the **number of stages** are positively correlated with the option of new trips. As a consequence, by combining **age group**, **household size**, **trip reason and number of stages,** 38 from the trips that presented in the lower-left graph can assumed to be induced demand (0.12% from the total observations).

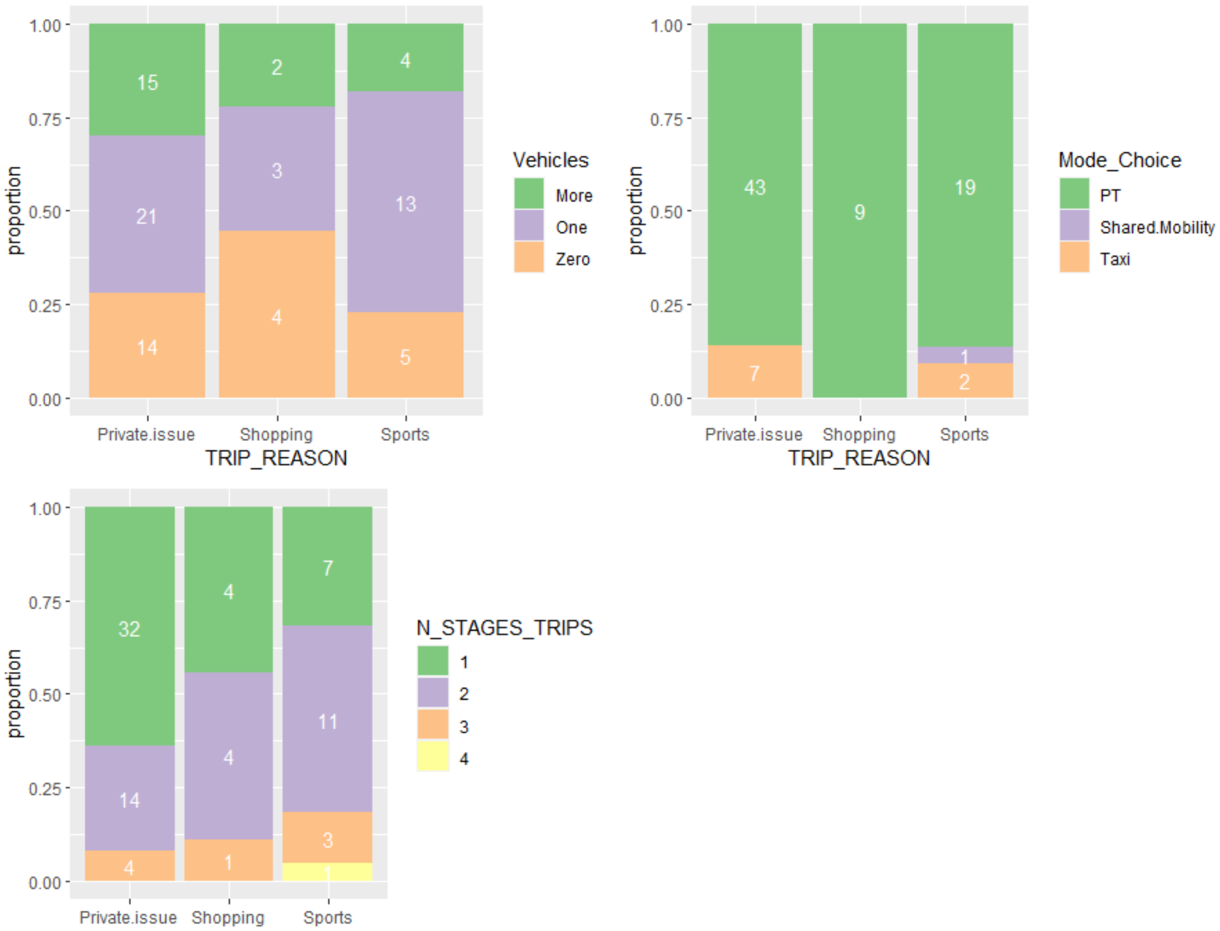

Figure 6.2 Induced demand

7 Conclusion and Future Work

In this chapter, the conclusions of this work are presented, followed by a discussion of the limitations, after which recommendations and research directions are given as suggestions for future work.

7.1 Conclusions

The findings of this thesis suggest strong indications about the main objective: identifying the factors affecting the decision of making new trips and a proposed methodology for the distinction of the induced demand from the provided demand. As there is lack of research in induced demand caused by new mode choices, such as shared mobility, the factors affecting demand extracted from the literature mostly came from the improvements of the infrastructure. The way to identify the characteristic of new trips, was to build a discrete choice model, so that the influential parameters be characterised. The dataset was provided from the Municipality of Madrid and is referring to the year of 2018. The data analysis of the provided responses gave evidence on the importance of socio-demographics groups and their attitudes in decision of making new trips. The development of statistical models aligned with the methodology framework resulted in significant multinomial and nested logit models.

The main findings of this work can be summarized in the following to answer the main research questions. Socio-demographic parameters are strongly influential in the decision making for new trips. There was a clear evidence on people between 18 and 34 years old being more enthusiastic about exploring new destinations. In addition, students show high interest on exploring new places. On the other hand, employments was negatively correlated with the decision of starting a new trip. There was not observed any cultural impact on the model, concluding that Spanish nationality is not significant for trip decision making. In addition, household parameters relieved the importance of the vehicle availability on the decision of exploring new places. People living along or with at maximum one more person show enthusiasm on making new trips. On the other hand, families and especially those with small children under the age of 4 years old showed higher scepticism of the respondents. Moreover, private issues, shopping and sports are the main reasons that can motivate the beginning of a new trip. Shared mobility, taxis and private vehicles are at most selected when it comes to new trips. Finally, trips being part of a trip chain are the most probable to be new trips.

Overall, the findings of this thesis serve to fill the gap in the literature on the identification of induced demand on a provided dataset. Most importantly, this work propose a methodology on how to closely work with a provided data and distinguish the induced from the actual demand. For the implementation of the first step of the methodology, it was used multinomial logit models, but also nested logit models with trip frequency as a dependent variable, contributing thereby to the field of research.

7.2 Limitations and future work

This work has however limitations that can be summarised as follows:

- The provided data set could be improved by having a more homogeneous sample, including more trips that took place just once in the network. Accordingly, a wider and more representative data could be gathered, as it would include more new trips, the reasons that motivated them, the modes that were used and the characteristics of the individuals that commute them.
- More shared mobility trips could have been included in the provided data set, so they could be further analysed.
- A more representative profile of the commuters could have been provided, including the income, the type of employment and the willingness of making new trips in case that shared mobility proved of the individuals.
- This study only considers new trips in Madrid. Although patterns for the characteristics were identified, it should be further explored for countries where the income and social norms regarding trip willingness are distinct from Spain.
- The model built over synthetic sample, and hence the results could have been better and more accurate with adequate sample.
- There are also some limitation in the findings. For example, it was found out that the high educated people are more probable to make weekly or monthly trips compared to low educated. A finding that cannot be appropriately explained.

Future wok could focus on trying different sampling techniques and using other models such Ordered logit models (OLM) for the identification of the characteristics of new trips. Others studies can focus on implementing the next steps of the proposed methodology. Clustering algorithms could be applied using the estimated coefficients that resulted from the discrete choice models with target to identify the induced demand. A further research motivation could be the appropriate modelling of induced and the further exploration of whether the induced growth in number of trips is beneficial or not. There may be some benefits to providing mobility and increasing access to new destinations, however, this must be weighed against the environmental and social costs associated with the increment of demand. The other concern is that providing a new more comfortable transport mode, will induce the problem of congestion. Generally, induced demand effects are real and need to be considered both by planners and policy makers.

Bibliography

Amazon Prime Air. (2016, December 7). Retrieved from https://www.amazon.com/Amazon-Prime-Air/b?node=8037720011

Alonso, W. (1964). The form of cities in developing countries. *Papers of the Regional Science Association 13*, 165-173. doi:https://doi.org/10.1007/BF01942567

Balać, M., Vetrella, A. R., Rothfeld, R., & Schmid, B. (2019). Demand estimation for aerial vehicles in urban settings. *IEEE INTELLIGENT TRANSPORTATION SYSTEMS MAGAZINE 11(3),*. doi:https://doi.org/10.3929/ethz-b-000274798

Ben-Akiva, M., & Lerman, S. (1985). Discrete choice analysis: theory and application to travel demand. *MIT press*.

Benjamin, S. (2018). *CityLab University*. Retrieved from Induced Demand: https://www.bloomberg.com/citylab

Bernett, V. (1978). The Study of Outliers: Purpose and Model. *Journal of the Royal Statistical*, 242-250.

Bevans, R. (2020). *An introduction to t-tests.*

Branco, P., Torgo, L., & Ribeiro, R. (2016). A Survey of Predictive Modeling on Imbalanced Domains. 1-50.

Burkhardt, J., & Millard-Ball, A. (2006). Article Metrics. *Journal of the Transportation Research Board*, 98-105.

Cascetta, E. (2009). *Transportation Systems Analysis.* New York: Springer.

Cervero, R. (2003). City CarShare: First-Year Travel Demand Impacts. *Transportation Research Record 1839*, 159-166. doi:https://doi.org/10.3141/1839-18

Cevero, R. (2003). Road Expansion, Urban Growth, and Induced Travel: A Path Analysis. *Journal of the American Plan- ning Association*, 145-163. doi:doi:10.1080/0194436030897630

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic minority over–sampling technique. *Journal of Artificial Intelligent Research*, 321-357.

Clavel, R., Mariotto, M., & Enoch, M. (2009). CARSHARING IN FRANCE: PAST, PRESENT AND FUTURE.

Clewlow, R. R., & Mishra, G. S. (2017). Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States .

Croissant, Y. (2011). *Estimation of multinomial logit models in R: The mlogit Packages.* Retrieved from https://cran.r-project.org/web/packages/mlogit/

Decorla-Souza, P., & Cohen, H. (1999). Estimating induced travel for evaluation of metropolitan highway expansion. *Transportation 26*, 249-262.

Douglass B. Lee, L. A. (1999). Induced Traffic and Induced Demand. *Journal of the Transportation Research Board*, 68-75. doi:https://doi.org/10.3141/1659-09

Duranton, G., & Turner, M. (2008). *The Fundamental Law of Road Congestion: Evidence from US cities.* Toronto.

Efthymiou, D., Antoniou, C., & Waddell, P. (2013). Factors affecting the adoption of vehicle sharing systems by young drivers. *Transport policy*, 6473.

Elliot, M., & Shaheen, S. (2016). Impacts of car2go on vehicle ownership, modal shift, vehicle miles traveled, and greenhouse gas emissions: An analysis of five forth American cities. *tsrc*.

Field, A. (2013). Discovering Statistics with IBM SPSS Newbury Park.

Forinash, C., & Koppelman, F. (1993). APPLICATION AND INTERPRETATION OF NESTED LOGIT MODELS OF INTERCITY MODE CHOICE. (pp. 98-106). Transportation Research Board.

Gareth, J., Witten, D., & Hastie, T. T. (2013). An Introduction to Statistical Learning .

Geert, R. (2021). *SPSS Tutorials*. Retrieved from Correlation.

Gehrke, S. (2017). A Survey of Ride-Hailing Passengers.

Glasco, J. (2019). Retrieved from SMART MOBILITY: CHALLENGES AND SOLUTIONS IN SMART CITIES: https://hub.beesmart.city/en/solutions/smart-mobility/smart-mobility-challenges-and-solutions-in-smart-cities

Gorham, R. (2009). *Demystifying Induced Travel Demand.* Retrieved April 27, 2021

Guillaume, S. (2017, June 19). *AIRBUS*. Retrieved May 1, 2021, from VSR700 demonstrator performs first autonomous flights: https://www.airbus.com/newsroom/press-releases/en/2017/06/VSR700.html

Harrell, F. (2015). *Regression Modeling Strategies.* doi:https://doi.org/10.1007/978-3-319-19425-7

Heling, M., Saphores, J.-D. M., & Samuelsen, G. S. (2009). User Characteristics and Responses to Shared-Use Station Car Program: Analysis of ZEV-NET in Orange County, California. *Transportation Research Board 88th Annual Meeting.* Washington DC, United States.

Hosmer, J., & Lemenshow, S. (2013). Applied logistic regression. *John Wiley & Sons, 398*.

Hymel, K., Small, K., & Dender, K. V. (2010). Induced demand and rebound effects in road transport. *Transportation Research Part B*, 1220–1241.

Jacobsen, P. (1997). Liquid vs Gas Models for Traffic. *Los Angeles Times*.

Jonge, E., & Loo, M. (2013). *An introduction to data cleaning with R.*

Koenig, G. (1980). INDICATORS OF URBAN ACCESSIBILITY: THEORY AND APPLICATION. *Transportation 9* , 145-172 .

Koppelman, F., & Bhat, C. (2006). A self instructing course in mode choice modeling: multinomial and nested models.

Kwan, M. (1998). Space-time and integral measures of individual accessibility: a comparative. *Geogr Anal 30*, 191-216.

Lawrence, B. (2000). Testing for the Significance of Induced Highway Travel Demand in Metropolitan Areas. *Journal of the Transportation Research Board*. doi:https://doi.org/10.3141/1706-01

Lineberger, R., Hussain, A., Mehra, S., & Pankratz, D. (2018). *Deloitte.* Retrieved May 1, 2021, from Elevating the future of mobility: https://www2.deloitte.com/us/en/insights/focus/future-of-mobility/passenger-drones-flying-cars.html

Litman, T. (2021). *Generated Traffic and Induced Travel.* Victoria Transport Policy Institute.

Loop, H. v., Haaijer, R., & Willigers, J. (2016). New findings in the Netherlands about induced demand and the benefits of new road infrastructure. *Transportation Research Procedia 13*, 72-80.

Louviere, J., Hensher, D. A., & Swait, J. D. (2000). Stated choice methods: analysis and applications. Cambridge university press. *Cambridge university press*.

Louviere, J., Hensher, D., & Swait, J. (2000). *Stated choice methods: analysis and applications.* Cambridge university press.

Madrid-Population. (2018). Retrieved from Country economy: ountryeconomy.com/demography/population/spain-autonomous-communities/madrid?year=2018

Moreno, A., Michalski, A., Llorca, C., & Moeckel, R. (2018). Shared Autonomous Vehicles Effect on Vehicle-Km Traveled and Average Trip Duration. *Journal of Advanced Transportation*. doi:https://doi.org/10.1155/2018/8969353

Moudon, A., Lowry, M., Shen, Q., & Ban, X. (2020). *The Impact of Shared Mobility Options on Travel Demand.* Retrieved from http://hdl.handle.net/1773/46637

Multinomial Logistic Regression. (2020). Retrieved from RPubs: https://rpubs.com/beane/n4_2

Noland, R. (2001). *Transportation Research Part A*, 47-72.

Noland, R., & Lem, L. (2002). *A review of the ecidence for induced demand and changes in transporation and enviromental policy in the US and the UK.* London.

Noland, R., & Quddus, M. (2006). Flow improvements and vehicle emissions: Effects of trip generation and emission control technology. *Transportation Research Part D: Transport and Environment*, 1-14.

Open data CRTM . (2018). Retrieved from https://datos.crtm.es/

Ortuzar, J., & Willumsen, L. (2011). Modelling transport.

Pagliara, F., & Preston, J. (2013). An Induced Demand Model for High Speed 1 in UK. *Journal of Transportation Technologies*, 44-51. doi:http://dx.doi.org/10.4236/jtts.2013.31005

(2012). *Perspectives on parking policy.*

Ploetner, K. O., Al Haddad, C., Antoniou, C., Frank, F., Fu, M., Kabel, S., . . . Zhang, Q. (2020). Long-term application potential of urban air mobility complementing public transport: an upper Bavaria example. *CEAS Aeronautical Journal*. doi:10.1007/s13272-020-00468-5

Program, T. C. (2005). *Carsharing: How and where it succeeds.* Washington D.C.: Federal Transit Administration.

Robinson, R., & Vickerman, R. (2006). The demand for shopping travel: a theoretical and empirical study. *Applied Economics*, 267-281. doi:https://doi.org/10.1080/00036847600000023

Salzman, R. (2010). *Build More Highways, Get More Traffic.* The Daily Progress.

Shaheen, S., & Cohen, A. D. (2005). Carsharing in North America: Market Growth, Current Developments, and Future Potential. *Transportation Research Record Journal of the Transportation Research Board*, 116-124. doi:10.3141/1986-17

Simon, L., Young, D., & Rardoe, I. (2018). *Detecting Multicollinearity Using Variance Inflation Factors.* Retrieved May 20, 2021, from https://online.stat.psu.edu/stat462/node/180/

Small, K. (1992). *Urban Transportation Economics.*

Speck, J. (2018). Understand Induced Demand. In *Walkable City Rules: 101 Steps to Making Better Places* (pp. 64-65). Washington, DC: Island Press/Center for Resource Economic.

Steininger, K., Vogl, C., & Zettl, R. (1996). Car-sharing organizations: The size of the market segment and revealed change in mobility behavior. *Transport Policy*, 177-185. doi:https://doi.org/10.1016/S0967-070X(96)00024-8

Stillwater, T., Mokhtarian, P., & Shaheen, S. (2008). Carsharing and the Built Environment: A GIS-Based Study of One U.S. Operator. 27-34.

Susan, A. S., Danel, S., & Conrad, W. (1999). A Short History of Carsharing in the 90's. *THE JOURNAL OF WORLD TRANSPORT POLICY & PRACTICE*, 16-37. Retrieved April 27, 2021

Susan, S. (2017). Shared Mobility: The Potential of Ridehailing and Pooling. 55-76.

Swiss Post . (2017, March 31). Retrieved from Swiss Post drone to fly laboratory samples for Ticino hospitals: https://www.post.ch/en/about-us/media/press-releases/2017/swiss-post-drone-to-fly-laboratory-samples-for-ticino-hospitals

The Urban Mobility Observatory. (2018, October 31). Retrieved from Eltis: https://www.eltis.org/discover/news/madrid-cuts-speed-limit-and-announces-new-rules-roads

Thill, J.-C., & Kim, M. (2003). Trip making, induced travel demand,. 229-248. doi:10.1007/s10109-005-0158-3

Ting, K., Sammut, C., & Webb, G. (2011). Encyclopedia of machine learning. doi:doi:10.1007/978-0-387-30164-8

Train, K. (2009). Discrete Choice Methods with Simulation.

UBER. (2016). *Fast-Forwarding to a Future of On-Demand Urban Air Transportation.*

Velardi, V. (2010). High Speed Rail Demand Forecasting: Italian Case Study. *European Transport Conference, 2010.*

Wang, Y. (2020). *THE IMPACT OF SHARED MOBILITY OPTIONS ON TRAVEL DEMAND.*

Washington, S. P., Karlaftis, M. G., & Mannering, F. (2003). *Statistical and Econometric Methods for Transportation Data Analysis .* New York.

Yao, E., & Moriikawa, T. (2005). A study of on integrated intercity travel demand model. *Transportation Research Part A Policy and Practice*.

Zhou, B., & Kockelman, K. (2011). Opportunities for and Impacts of Carsharing: A Survey of the Austin, Texas Market. *International Journal of Sustainable Transport*, 135-152. doi:https://doi.org/10.1080/15568311003717181

Appendix A

Additional Plots

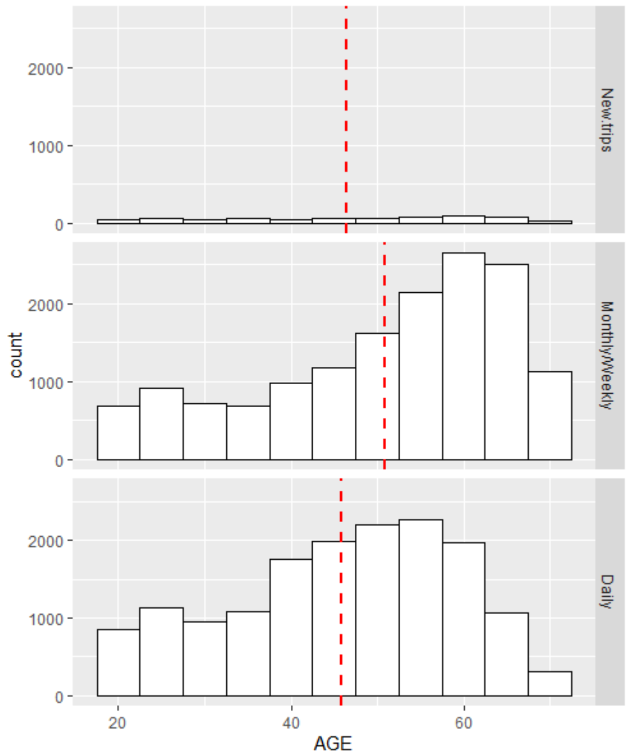

Figure A.1 Age distribution of total sample per frequency level

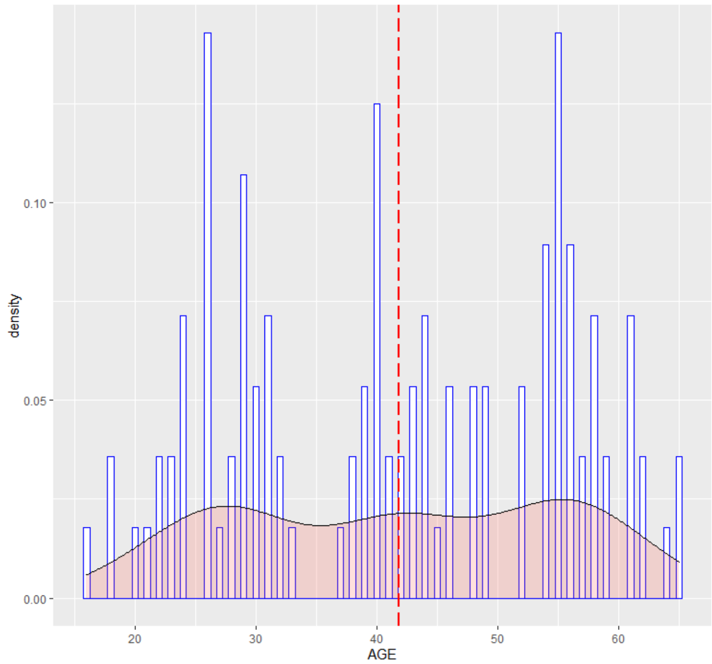

Figure A.2 Age distribution of shared mobility users

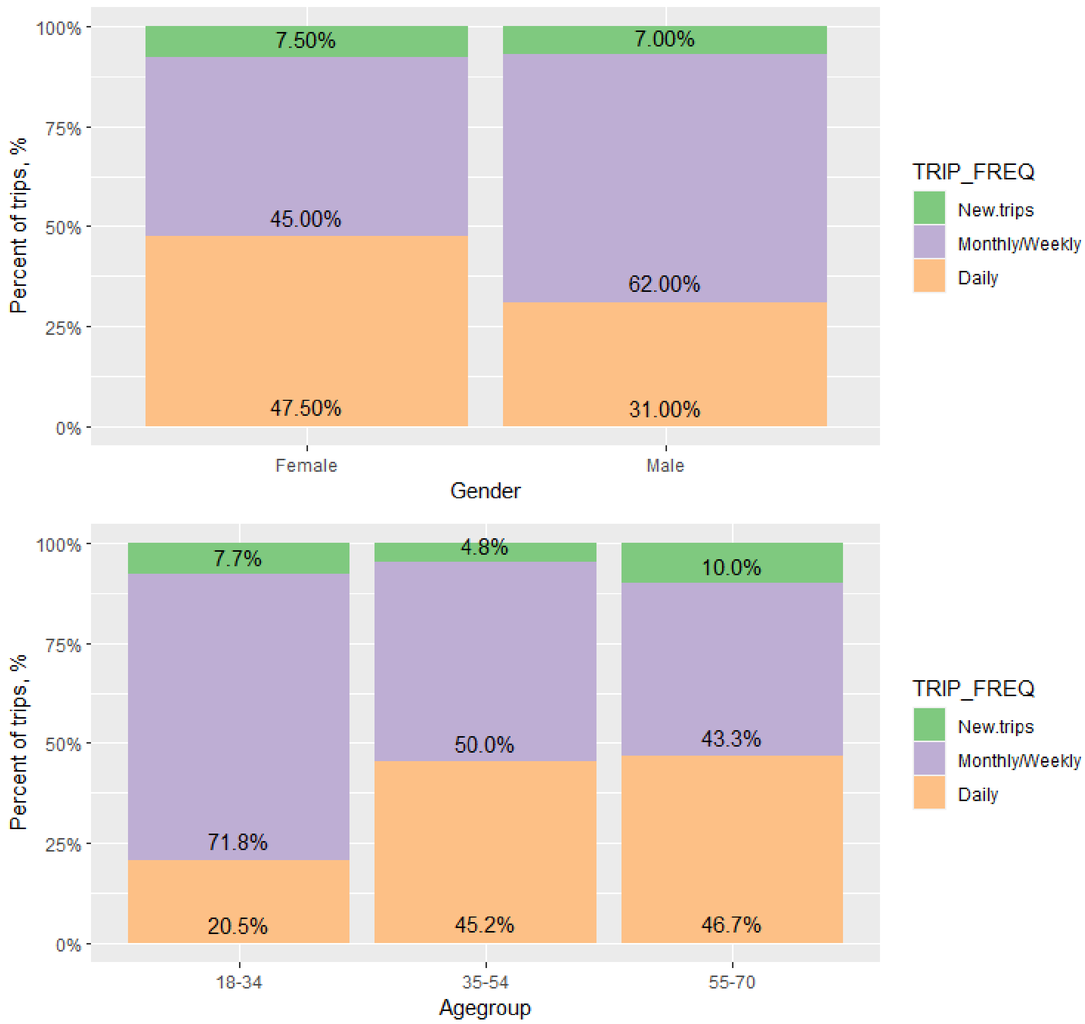

Figure A.3 Trip frequency per gender and age group of shared mobility users

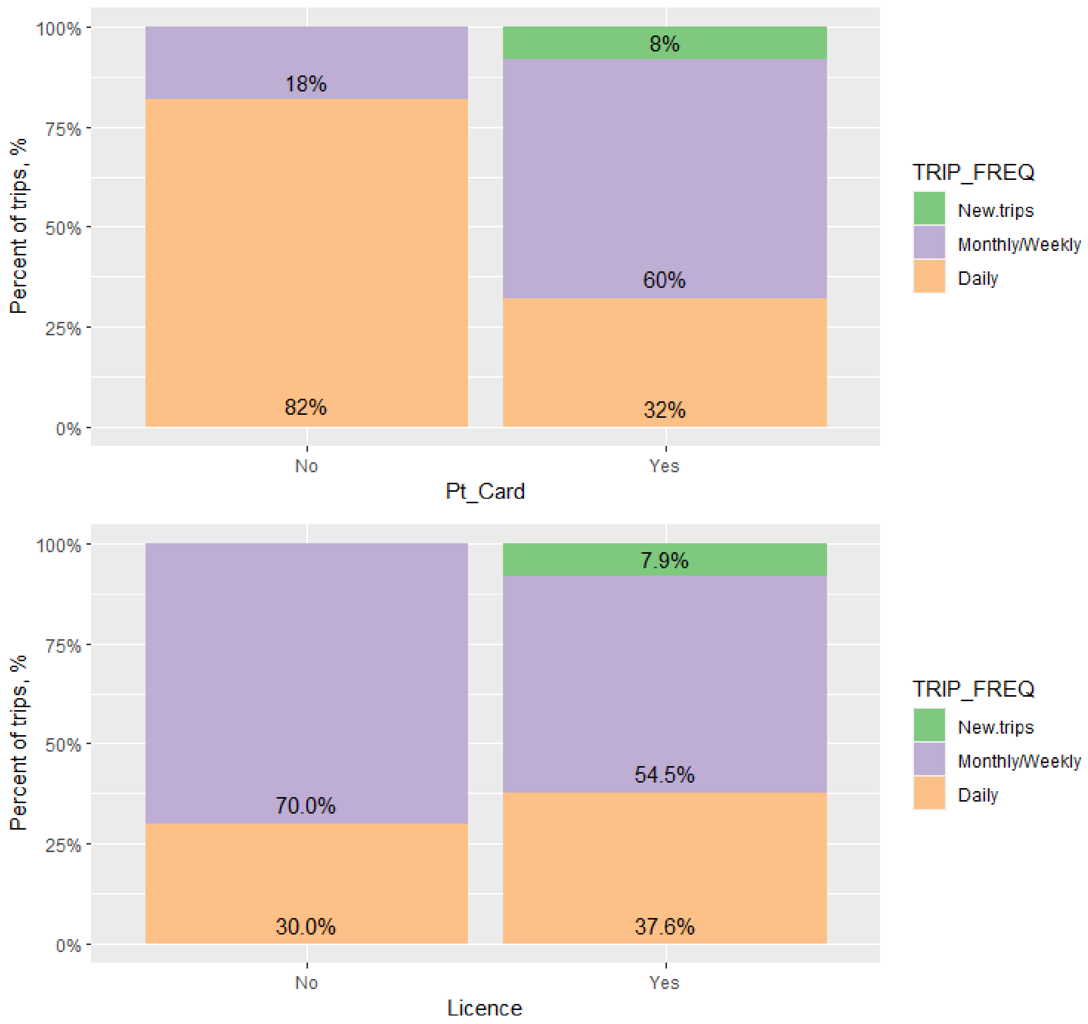

Figure A.4 Trip frequency per PT card and driving licence availability of shared mobility users

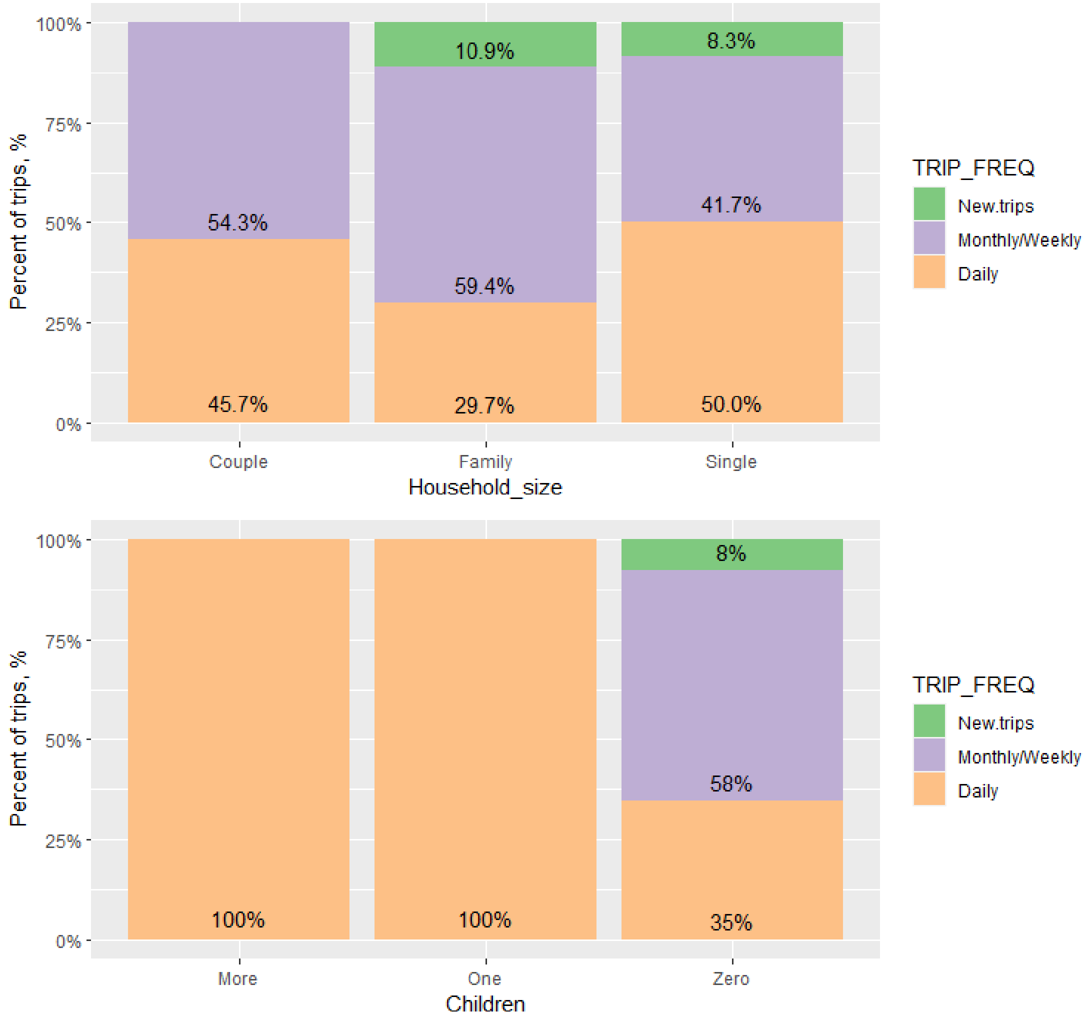

Figure A.5 Trip frequency per hh size and number of children of shared mobility users

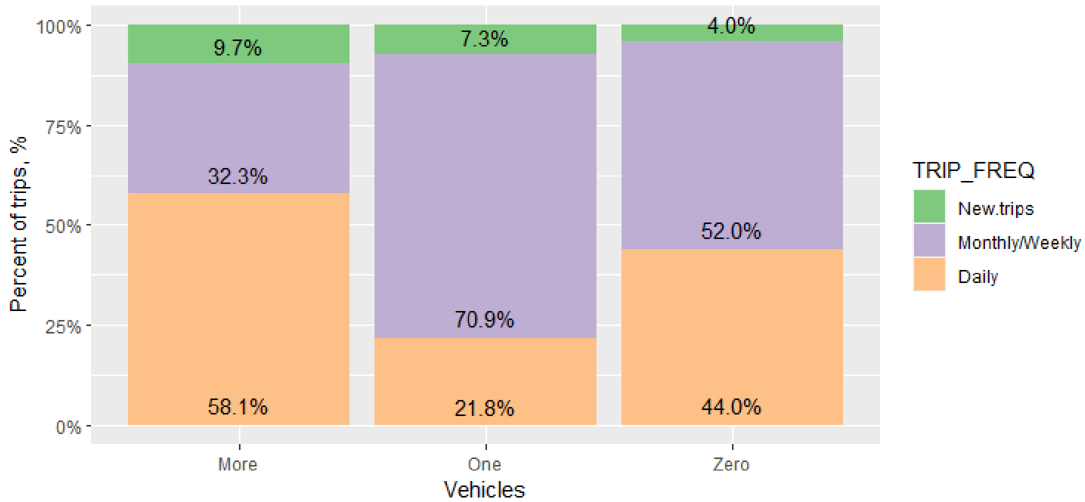

Figure A.6 Trip frequency per number of vehicles owned by shared mobility users

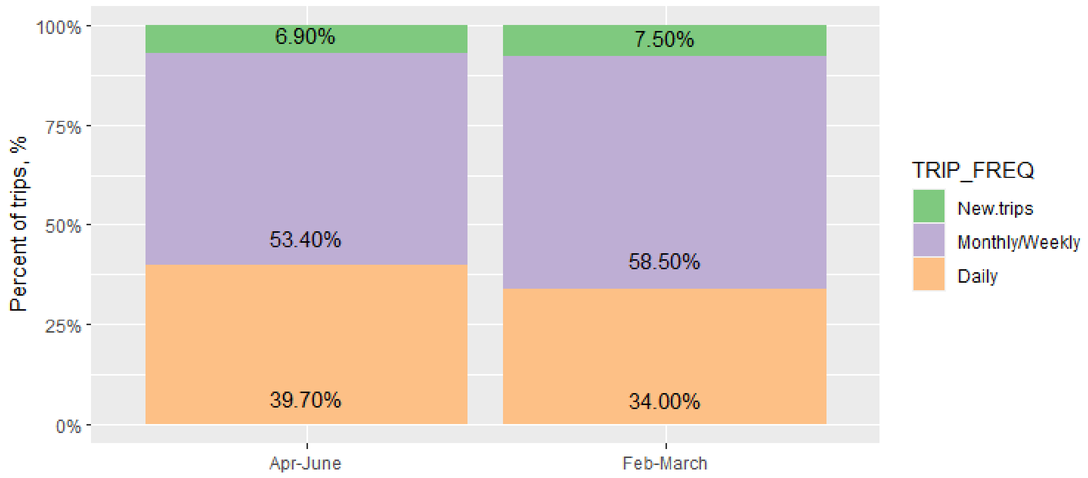

Figure A.7 Trip frequency of shared mobility trips per month

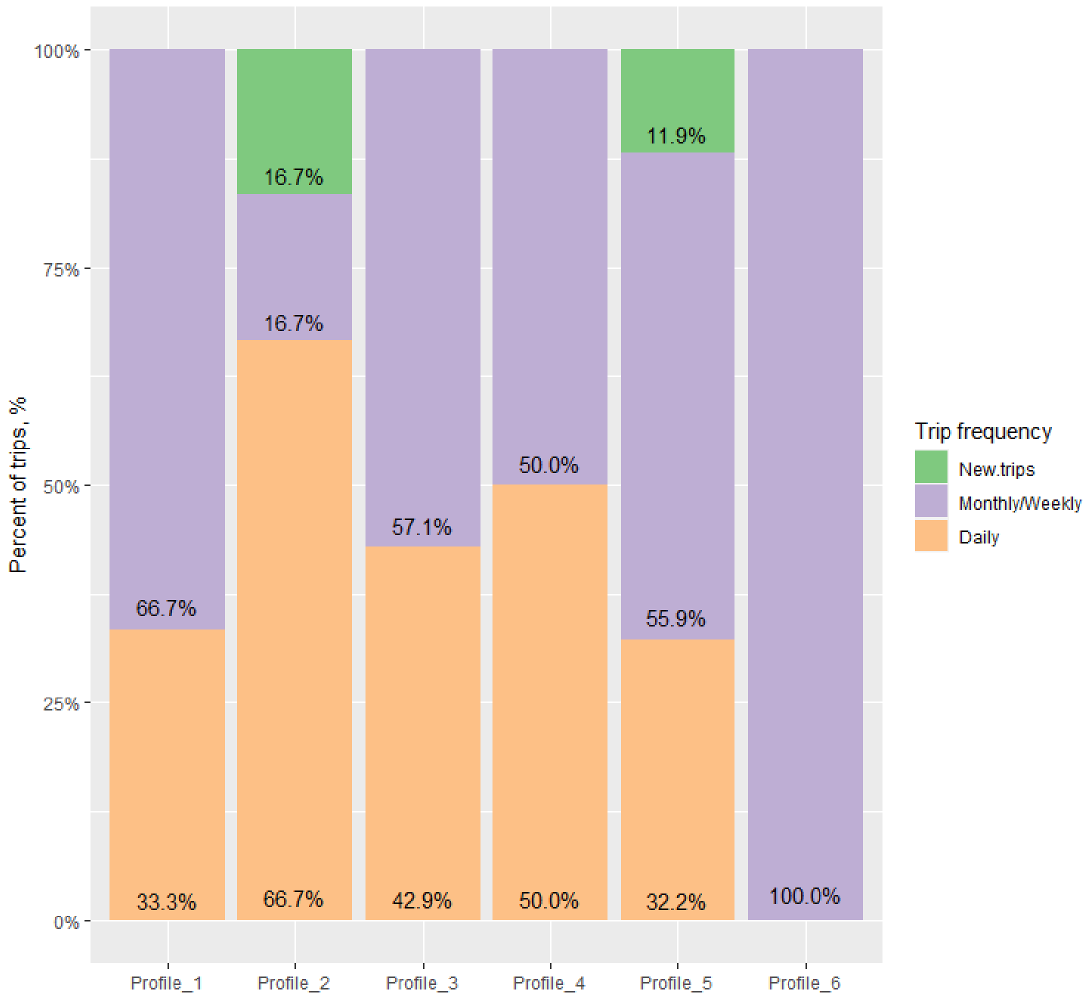

Figure A.8 Trip frequency per shared mobility users profiles

Disclaimer

I hereby confirm that this Master's thesis is my own work and I have documented all sources and materials used. This thesis has not been submitted elsewhere for purposes of assessment.

Munich, July 1st, 2021

Vasiliki Kalliga