UNDERSTANDING CHARACTERISTICS OF GROCERY SHOPPING TRIPS

BACHELOR THESIS REPORT

JEAN MORILLO | S2209705

DATE: 22-08-2022

UNIVERSITY OF TWENTE.

SUPERVISORS: BARAN ULAK (TRANSPORT ENGINEERING AND MANAGMENT)

JOHAN KOOLWAAIJ (MOBIDOT)

Preface

In this report, I present the findings of my bachelor's thesis, which I wrote as part of the requirements to obtain my bachelor's degree in Civil Engineering at the University of Twente. This bachelor thesis is about “Understanding characteristics of grocery shopping trips”. This study was performed from April to August 2022 and was commissioned by the company Mobidot at Enschede with the aim to have a better view of the insights behind supermarket purchasing excursions.

First and foremost, I would want to thank Johan Koolwaaij, my supervisor at Mobidot, for providing me the opportunity to work for such a prestigious company. In addition, he was always present at presentations, where his criticism was always incredibly helpful because he gave me ideas, I could not have anticipated from a civil engineering standpoint. I also want to thank Jop Zwienenberg for his support within the organization. He was always willing to give his perspective on challenges I found during my research and provided me with helpful feedback so I could continue my investigation. In addition, he provided me with fantastic advice when I joined the company.

I would also like to thank Baran Ulak, my supervisor at the University of Twente, for his guidance and support throughout the entire research process, as well as for our weekly meetings during which he always shared his perspective on the study and provided me with extremely helpful advice regarding which methods were best to use on my research and how certain results should be explained. Finally, I would want to express my gratitude and acknowledge the Secretaria Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) of Ecuador as the sponsor institution of my study.

Summary

One of the most significant reasons for a trip is to do grocery shopping. In fact, according to the Dutch National Travel Survey (NTS), there are more trips made for shopping than for work and almost three-quarters of these shopping excursions are for groceries (Veenstra, Thomas, & Tutert, 2010). But little is known about this type of trip in the Netherlands, therefore, this thesis is about to describe characteristics of grocery shopping trips.

The main goal of this research assignment is to properly analyse the data acquired from a dataset provided by Mobidot using statistical approaches in order to gain a better understanding of grocery shopping trips. In order to accomplish so, the variables affecting the frequency of grocery shopping and modes of transportation choice for this activity were examined.

A negative binomial regression model was performed to evaluate the relation between the frequency of trips and the independent variables. The goal is to identify with certainty which socio-demographic and spatial factors are most important, which can be not taken into account, and how these factors relate to predicting the frequency of grocery shopping trips. Besides that, a multinomial logit model allowed for determining the probability that an individual chooses an alternative mode of transportation based on the independent variables.

Among the primary findings was that the distance to the supermarket has a substantial impact not only on the number of trips to the supermarket but also on the mode choice. For example, when the distance is less than 2 kilometers, the frequency of trips increases, and greener modes of transportation are more likely to be chosen. Furthermore, as the time spent at the supermarket increases, fewer trips are done per month and car became the most probable mode of transportation. From the socio-demographic characteristics, it was found that annual income has an inverse relation to the frequency of trips and that individuals that have a lower income are more likely to use bikes or to walk for this activity.

Finally, it is suggested to have more in-depth analysis for this type of trip because other factors may also influence this topic. For instance, the growth of e-bikes could have a positive impact on the mode choice since they are more cost-effective for short trips. Additionally, as the Netherlands focuses on promoting cycling, it could be that in a near future, the predominant transportation mode for short-distance related activities changes from car to bike. Another factor that could affect this type of trtrip is the increasing delivery services offered by the supermarkets in the Netherlands.

Table of Contents

Table of figures

Table of tables

1. Introduction

The Dutch National Travel Survey (NTS) shows that there are more trips with the purpose of shopping (24%) than to work (20%). Of these shopping trips, 72% are related to grocery shopping (Veenstra, Thomas, & Tutert, 2010). In other words, grocery shopping is one of the most important trip purposes. However, according to Veenstra, Tutert, and Thomas (2010), there is not a lot of information concerning this function. But, new technologies, such as smartphones, have allowed the collection of accurate data regarding detailed information about trips made by individuals and this can be used to evaluate the characteristics of grocery shopping trips. Figure 1 describes some facts about mobility in the Netherlands where although the bicycle is not the main mode of transportation, it is highly used by its inhabitants for different trip purposes.

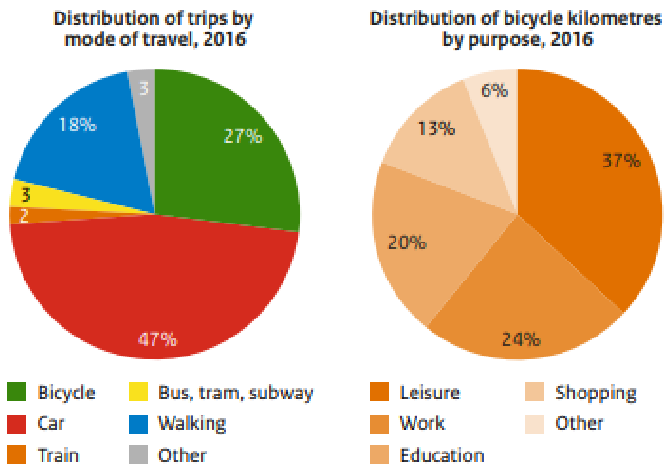


Figure 1.- Facts about mobility in the Netherlands. Taken from: Netherlands Institute fir Transport Policy Analysis

Previous studies have shown the influence of different factors on the trips made with the main purpose of grocery shopping and how this affects the travel behaviour of individuals. However, most of those studies were carried out using survey data collected through phone calls or interviews. This way to collect travel data could have led to wrong outcomes (Chen, Ma, Susilo, Liu, & Wang, 2016). Moreover, those research focused on a household level, thus, considering that the inhabitants of a certain household may have the same travel behaviour within the living place, which is not the case. Finally, few studies have centered their research on the Netherlands and this is an important factor because this country has a significant different infrastructure and culture compared with other countries, for instance, the Dutch infrastructure promotes the use of slower but greener ways to commute and regarding their culture, they have adopted a cycling culture (Hoed & Jarvis , 2021). Hence, at least in this scenario, the transport mode and frequency of trips can have a relevant different result compared with previous studies out of the Netherlands.

In Figure 2 below, the conceptual framework of the research can be seen. It aims to illustrate the process of the whole research. On the left side (green boxes), the main inputs of the models are located, these are characteristics of the users that are found on the dataset. After that, on the blue boxes, there are the models that are expected to analyse the information. Following them, are the respective situation that they intend to evaluate. With the results of both models, the characteristics of grocery shopping trips are expected to be described.

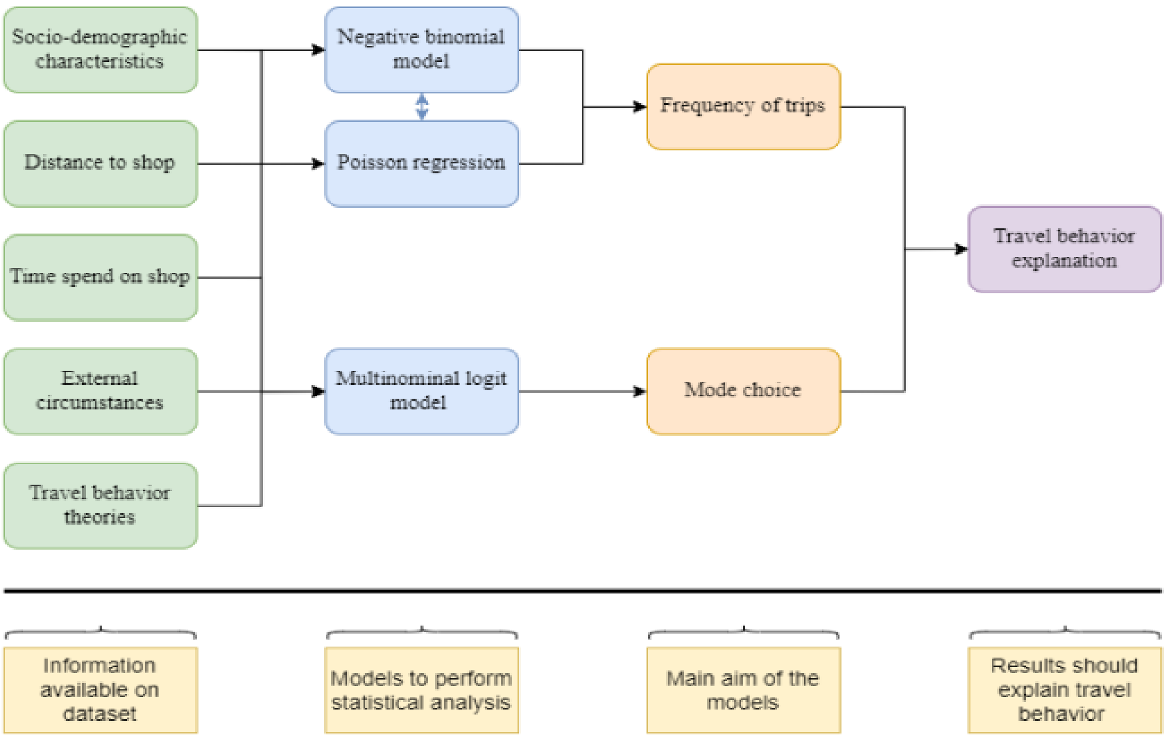


Figure 2.- Conceptual framework

Therefore, the remainder of this paper will start providing the context of the research. Then, the research objective and research questions will be described. After that, a summary of the prior research on the aforementioned subject is presented. The methodology used for this research will next be examined in depth. Then, the gathered findings will be analysed, and the report will conclude with the study's conclusions and recommendations.

2. Project context

Mobidot is the Dutch leading personal mobility data service provider. They support their clients by providing detailed data and information regarding the trips made by individuals. With their services, more human-centered approaches to addressing the social and economic effects of mobility become possible. Moreover, capturing personal travel and activity patterns gives chances not only for public but also private organizations to take advantage of "floating people data" and to provide benefits on both, individual and collective levels (Mobidot, 2022).

Among the most prominent features included in the dataset are the trip distances, trip time, time spent in the location, purpose of the trip, city, destination type, and many more. In addition to this dataset, another document was provided that detailed the socio-demographic characteristics of the users. The merging of these two datasets yields a comprehensive and envied final dataset that could lead to have a better view of factors affecting the travel behaviour of individuals.

However, in order to understand travel behaviour, it is necessary to analyse whether there is any factor that do influence travel choices, for instance, when travel distances are short, the frequency of those trips would be higher and greener modes of transportation may be chosen over other modes as car (Esztergár-Kiss, Shulha, Aba, & Tettamanti, 2021). Therefore, to explain how the relationship between those factors and important variables, regression models are used. Nevertheless, as there are multiple regressions that could be used, it is important to choose the correct regression analysis to have the right conclusions.

3. Research aim and research questions

The main purpose of this thesis assignment is to adequately evaluate the data gathered from a dataset supplied by Mobidot, using statistical techniques with the aim to have a better insight of grocery shopping. In order to do so, two main characteristics of these travels will be deeply analysed, name them: frequency of grocery shopping trips and modes of transportation chosen by the individuals. Therefore, on the one hand, count regression analysis will allow to show the relationship between the frequency of trips and the demographic characteristics of the users. On the other hand, a discrete mode choice model will be used to assess and forecast the transportation mode choices of individuals.

To reach the stated objectives a main research question has been defined for this thesis. The key research question for this study might be expressed as follows:

<u>What are the main characteristics of grocery shopping trips of individuals within the Netherlands?</u>

There are numerous aspects that can have an influence on grocery shopping and this report aims to analyse which of them have a significant influence over others and whether or not these factors may be used to predict the monthly number of grocery shopping trips and the mode choice used for this activity.

Furthermore, the following sub-questions will be used to help answer the main question:

- How often do individuals living in the Netherlands travel with the purpose of grocery shopping and is there any factor that highly influence this activity?
- Which factors do affect the transportation mode choice for the purpose of grocery shopping of individuals in the Netherlands?

By merging the two preceding sub-questions, it is possible to determine whether there is a relation between socioeconomic, demographic, and regional features and people's travel frequency and mode of transportation preference. For instance, it could be that houses with more members need to spend more time on the shop or that they buy more frequently as more products are needed. Moreover, the influence of car ownership could have a great impact on choosing car as mode of transportation but at the same time if the grocery shop is located near the origin point, a greener transportation mode as bike may be used.

4. Preliminary literature review

From the available literature about this type of human mobilization, most of the research was performed by using survey data. This type of data might have resulted in erroneous conclusions since they underestimate the number of trips taken by inhabitants of a study region and participants may not have recorded all their excursions because they consider them unimportant or simply forgot about them (Clarke, Dix, & Jones, 1981).

In addition, a similar study was conducted by Simma et all (2004), by analysing the factors influencing the shopping behavior in Switzerland, but this study was conducted at the "household" level, resulting in a loss of information since various household members may report different travel behaviors (Simma, Cattaneo, Baumeler, & Axhausen, 2004). Lastly, comparable articles have examined this problem; however, those papers pertain to locations outside the Netherlands, where not only the transportation infrastructure but also culture is notably different from that one in The Netherlands (e.g., Byung & Kyungdo, 2017; Recker & Kostyniuk, 1978). For instance, as clearly seen in Figure 3, The Netherlands is the undisputed number one country for cycling, this is due not only to its magnificent road infrastructure but also to the policies implemented that encourage the use of greener ways to commute (Harms & Kansen, 2018). As a result of the aforementioned factors, it is anticipated that Dutch residents would have distinct buying habits, and they should be investigated separately.

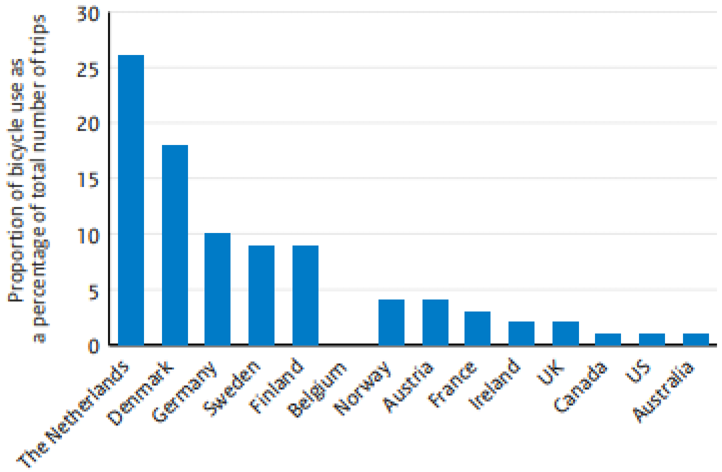


Figure 3.- Proportion of bicycle use as a percentage of total number of trips in several countries (source: Ministry of Infrastructure and Water Management, 2018).

This section of the research will begin by explaining the factors that have a substantial impact on the frequency with which individuals visit supermarkets. After that, it will explain why a person is more inclined to choose one form of transportation over another.

Summarizing up, this portion of the paper will explain the findings of prior studies that studied the factors linked with people' shopping frequency and preferred transportation mode regarding grocery shopping.

4.1 Trip frequency

One of the objectives of this research is to determine if there are indicators that may predict the number of grocery shopping visits a person makes. Diverse investigations have been conducted in various regions of the globe, from which some conclusions have been drawn.

Firstly, Kahn and Schmittlein (1989) tried to objectively characterize trip behavior using data from shopping trips. According to their findings, the frequency of food purchases follows a pattern based on seven-day cycles. They imply that the choice to go shopping is heavily influenced by the day of the week. In addition, they divide customers into two categories depending on the amount of time and money spent in the store: rapid shoppers and regular shoppers. However, since the data do not include information regarding the amount spent on travels, this research will not go into depth about this final aspect.

Secondly, according to Simma et al. (2004), socioeconomic characteristics such as age, gender, weekly working hours, and family income are more influential than spatial variables in influencing an individual's purchasing behavior (Simma, Cattaneo, Baumeler, & Axhausen, 2004). These findings are also corroborated by Kara and Mkwizu, who did similar research targeting specific demographics in order to identify tourism-boosting elements in Tanzania (Kara & Mkwizu, 2020). Moreover, Thiagarajan and Natarajan (2016) report that the majority of customers visit a food store once each month. In addition, they discovered that there is an inverse association between income and frequency of purchase, meaning that those with lower incomes shop for groceries more often than those with higher incomes (Thiagarajan & Natarajan, 2016).

In conclusion, there is evidence that demographic and socioeconomic parameters, as well as those linked to supermarket proximity, influence the frequency of shopping visits. In the subsequent chapters of this study, the issue will be elaborated upon more in order to provide a deeper comprehension of the aforementioned conclusions.

4.2 Mode choice

To understand the features of the previously described categories of trips, it is also essential to know the mode of transportation selected by people and to determine whether there is a correlation between the chosen transportation mode choice and demographic, socioeconomic, distance, or travel duration parameters. Analysing and explaining the relation between them might lead to the right adoption of policies or improvements in infrastructure with the goal of shifting the mode choice of individuals toward more environmentally friendly modes for their grocery shopping.

Handy began by analysing the general shopping mode preference amongst genders and household kinds. She observed that income, the presence of children, and their ages influenced mode selection more than gender (Handy, 1996). Furthermore, Bhat states that jobless individuals were less inclined to drive for general shopping than employed individuals because they cannot afford the expenditures associated with a vehicle excursion, such as gas and occasionally parking fees (Bath, 1998). Most importantly, Chen et al. found that the availability of a vehicle and the price of public transit had a substantial effect on shopping mode choice, therefore, users that do have a car are most likely to use it as main transportation mode for grocery shopping despite the price involved (Chen, Ma, Susilo, Liu, & Wang, 2016).

Regarding shop location, Hagberg and Johan argued that the distance to the shop is the most significant element in explaining the mode of transportation where for larger distances, the preferred mode was car, however, they did not take into consideration that the distances to the nearest shop significantly changes according to the country (Hagberg & Holmberg, 2017). Thus, the results obtained in this research could lead to different results from previous studies, for example, conducted on other countries such as USA or Switzerland where the average distance to the nearest supermarket is 1.4 km (Rhone & Ploeg, 2019) and 2.3 km (Simma, Cattaneo, Baumeler, & Axhausen, 2004) respectively. Meaning that the distances to the shop are larger compared to the one in the Netherlands which is 0.9 km (CBS, 2010).

Overall, the research demonstrated that gender, job position, wealth, the presence and ages of children in the household, and automobile ownership had a substantial influence on individuals' shopping travel mode preferences. In addition, Hagberg and Holmberg studied the travel mode preference for grocery shopping in Sweden and concluded that the most prevalent means of mobility is the car, followed by cycling and walking. They note that possessing a vehicle is a crucial aspect of buyers' travel mode due to its convenience and time-saving qualities (Hagberg & Holmberg, 2017).

Concluding, the literature revealed that there are significant elements that might explain the mode of transportation used for grocery shopping. However, the majority of previous studies indicate that the automobile is the favoured form of transportation due to the convenience and time savings it offers to people.

4.3 Modelling approaches

Finding the variables or factors that have an effect on a topic of interest can be done with accuracy using regression analysis. It can confidently establish which elements are most important, which ones can be ignored, and how these factors interact when you do a regression. In other words, regressions allow to understand which independent variables (hypothesized factors that have an impact on the dependent variable) significantly predict a dependent variable (the main factor being studied). As mentioned on the chapter above, this research aims to understand the frequency of grocery trips and the transportation mode chosen for this, therefore, the dependent variables will be the ones just mentioned.

4.1.1 Regression for trip frequency

Regression models must be adequately chosen and in order to do so, it is important to recognize the type of data that will be analyse. For the frequency of grocery shopping trip, the response variable is a nonnegative integer or count. Literature shows that Poisson regression is the starting point for count data. However, the assumptions behind this type of regression are difficult to fulfil, especially the one related with the mean = variance, where in most of the cases is not the case. In order to avoid the under or overdispersion that can be caused by the assumption just mentioned, some other regressions can handle this problem, for instance, quasi-poisson or negative binomial regressions may be used (Johnson, Ott, & Dogucu, 2021). In the following sections, this will be explained more in detail.

4.1.2 Discrete choice models

In the field of traffic and transportation, these type of models aims to investigate the behaviour of transportation users in terms of mode selection and risk of transportation, as well as to predict changes in mode characteristics or socioeconomic indicators of the decision-maker. The multinomial logit model is the most widely used discrete choice model, with applications in a variety of disciplines. This disaggregated model seeks to investigate the choice of an event's value or the perception of its worth among a group of mutually incompatible alternatives (Aloulou, 2018).

5. Methodology

The methodology that will be used for this assignment is quantitative research with a correlational approach because the research seeks to determine the extent of a relationship between the dependent and the independent variables using statistical data. This chapter will start by describing the study area. After that, it will outline the data used for the analysis. Then, the methods used to analyse the characteristics related not only with the frequency of grocery shopping trips but also the mode choice will be discussed in detail.

5.1 Study area

As mentioned on previous chapters, this study focuses on the grocery shopping trips made within the Netherlands which can be seen in Figure 4. The Netherlands' infrastructure is focused on lowering greenhouse gas emissions. For example, numerous measures have been introduced in an effort to decrease the usage of automobiles, particularly where the destination can be reached by foot or bicycle (Harms & Kansen, 2018). This is why it is assumed that the travel behaviour of individuals in the Netherlands may differ, given that this trip is deemed short distance.

Figure 4.- Study area: the Netherlands

5.2 Data

The dataset that will be used for this research is provided by Mobidot. It is an excel file that contains records about the trips of 7678 users over a period of 91 days (from 28-02-2022 till 29-05-2022). The trips are registered by the trip modality, which means, that each time that the user starts a new trip, the data record it as a new row where the features of the trip are saved. The ones of interest for this research are the travel distance and time, trip modality, and destination type. Besides this excel file, another file is given where the socio-demographic factors of each user on the file can be found.

One of the drawbacks of this dataset is that some users do not actively participate in the recognition of their trips. This means that there are certain users who are disconnected from the application that allows data collection. Therefore, only users who appear for more than 90% percent of days in the dataset would be taken into consideration in the analysis. This decision was made so that the conclusions acquired in the study could be more reliable. Furthermore, as indicated before, the dataset contains records of all of the user's travels, but the emphasis of this research is on trips made for grocery shopping. Therefore, any visits with non-shopping aims and destinations other than a supermarket were removed from the dataset.

Upon displaying the data using tools such as Python, discrepancies were discovered. For instance, when the amount of time spent at a store is excessive. It is hypothesized that this may have occurred as a result of the supermarket's position in a complex multifunctional location, where many activities may be conducted. Therefore, it is believed that despite the fact that a trip was identified as a shopping trip by the data, it was in reality a trip with the main aim of visiting and the location of the visit was above a supermarket. In an effort to address this issue, excursions with a shopping duration over 3840 seconds, which is twice the average grocery shopping duration in the Netherlands (32 minutes according to eurostat), will not be considered. In addition, only a tiny percentage of trips recorded large distances, thus, this research would only include trips within 25 kilometres (see Appendix A.- Sample description for more details). This distance was taken as reference with the help of visualization tools which described the frequency of the variables and only an insignificant number of grocery shopping trips were larger than 25 kilometres, this larger trips may be explained as extreme preference of an specific product, for instance, an international individual that want to purchase a product that was manufactured in their home country, however in The Netherlands, this particular product can only be purchased from one particular retailer.

Following the modifications described in the preceding paragraphs, the final dataset was obtained, and it includes trip records of 3003 individuals which will be analysed in the following chapters of this research.

5.3 Variables

5.3.1 Dependent variables

Two different analyses will be performed; hence, two different dependent variables will be used, On the one hand, the dependent variable for the study of the frequency of supermarket visits is the number of supermarket trips per month. On the other hand, for the analysis of the choice mode, the dependent variable is the mode of transport chosen by the user. It should be noted that, despite the fact that in the original dataset, “public transport” and “other” modes of transportation were also options in the mode choice, they will not be considered because they represented less than 1% of the choices.

5.3.2 Independent variables

The independent variables not only contain sociodemographic characteristics but also features of the trip itself. Table 1 displays the explanatory factors employed in these analyses. However, in order to prevent multicollinearity, Pearson's correlation was used to analyse the relationship between variables. No significant correlation was found between the numerical variables. Nevertheless, a strong correlation was found between vehicle ownership and driving license, meaning that all users who have a license also own a car; thus, only car ownership will be considered a variable (see Appendix B.- Correlation between variables for a complete view of the calculations).

Table 1.- Independent variables

Variables	**Data type**
Preferred mode	Categorical
Weighted avg distance	Numerical
Weighted avg travel time	Numerical
Weighted avg shopping time	Numerical
Preferred day	Categorical
Gender	Categorical
Age	Numerical
City density	Categorical
House composition	Categorical
Education level	Categorical
Profession	Categorical
Car ownership	Categorical
Household size	Numerical
Annual income	Categorical

Driving license	Categorical

The variables that can cause some confusion in the analysis are described below as follows (see Appendix A.- Sample description for detailed data):

- Preferred mode: It refers to the most used transportation mode of the users.
- Weighted average distance: It is the distance from the origin point to the supermarket. But, as an user may have visited different supermarkets with different distances, the weighted average was taken. In order to calculate the weights, the frequency of the store was taken into account. Thus, if a certain shop was visited more times, that distance got more weight in the calculation.
- Weighted average travel time: It follows the same logic at the previous variable. The difference is that it is the time it takes to reach the supermarket.
- Weighted average shopping time: It also follows the same logic as last two variables regarding the weights. This variable refers to the time that the user spent on the supermarket.
- Preferred day: this is the day where most of the trips were made.
- City density: it is the number of people inhabiting a given area. However, In this context three categories can be found, namely them: high, medium and low. They will be used to refer to the proximity of the amenities to houses.
- House composition: it refers to the type of family living in a house. For instance, it could be that a single person lives there, or that a family that has a young or adolescent son. Considering adolescent to individuals between 10-19 year old and young as 15-24 year old.
- Education level: it refers to the level of education that individuals have. For example, a bachelor's degree or superior, or just have achieved a high school level.
- Profession: it describes which profession the user has, there is plenty of options.
- Household size: it refers to the total number of people living in the same house.
- Annual income: it refers to the total amount of money earned in a yer by the user. It is divided in three categories: low, modal and high. However, no monetary value can be given as that information was not provided by the dataset. Therefore, the monetary values will be taken from the available literature and are assumed that low = < €24000, modal = > €24000 and < €34000, high = > €34000.

5.4 Frequency of trips

On the one hand, regression will be used to evaluate the relation between the frequency of trips and the independent variables. Poisson regression will be used for the mentioned analysis since it is not only supported by the literature as one of the most suitable regressions for the analysis of this “count” data (Coxe, West, & Aiken, 2009) but also for the suggestion of the supervisor of this thesis. The goal is to identify with certainty which factors are most important, which may be ignored, and how these factors interact to predict the frequency of grocery shopping trips.

In equation 1, the structure of the Poisson regression may be seen. To use this regression, however, certain assumptions must hold true. To start with, the response variable must initially be a count. In this instance, it is the number of monthly shopping excursions. The observations must also be independent of one another. To satisfy this assumption, the data must be modified appropriately. Consequently, just one record per person was used. This was done taking into consideration the user's favourite day of purchase, their preferred mode of transportation, the weighted average distance and travel time to the supermarket, and their selected mode of transportation (the weights were given by taking into account the occurrence a shop was visited, the larger the number the visits to that shop, the larger the weight).

$$PMF\ (\mathrm{yi}|\mathrm{xi}) = \frac{e^{-\lambda \mathrm{i}} * \lambda \mathrm{i}^{yi}}{yi!} \qquad \text{Eq. 1}$$

Where:

- PMF = probability of seeing count yi given the regression vector xi
- λi = event rate for the ith sample

However, Poisson regression assumes that the mean equals the variance, and this premise is not entirely met since mean = 3.42 and variance = 9.70, this can be seen as an issue of overdispersion of the dataset. Then, negative binomial regression is used because it is a common variant of Poisson regression since it relaxes the Poisson model's extremely restrictive condition that the variance equals the mean. Poisson-gamma mixed distribution provides the foundation for the conventional negative binomial regression model. This popular model models the Poisson heterogeneity with a gamma distribution. Equation 3 shows the fundamental negative binomial regression model. Here, the mean of y is determined by the exposure period t and a collection of k regressor variables (the independent variables). The relational expression between these quantities is shown in Equation 2 below.

$$\mu_i = \exp\ (\ln(t_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}) \qquad \text{Eq. 2}$$

$$\Pr(Y = y_i|\mu_i \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \qquad \text{Eq. 3}$$

Although negative binomial regression seems to suit this research better, both regressions will be tested in this report and the likelihood ratio test will be used to determine which regression works better.

5.5 Mode choice

On the other hand, a discrete choice model will be performed to analyse the preferred transportation modes of individuals and to look for patterns in the decisions they make. Therefore, the multinomial logit model will allow determining the probability that an individual chooses an alternative mode of transportation based on the independent variables, in this case, their demographic characteristics and the characteristics of the trip (distance to the store and time spend on store). This regression is an extent of the logistic regression, shown in Eq. 4, which can support only binary categorical outcomes.

$$\ln\left(\frac{\mathrm{P}}{1-\mathrm{P}}\right) = \beta 0 + \beta 1\mathrm{X}1 + \ldots \beta \mathrm{k}\mathrm{X}\mathrm{k} \qquad \text{Eq. 4}$$

Where:

- Xk: The k^th predictor variable
- $\beta\mathrm{k}$: The coefficient estimate for the k^th predictor variable
- P: The probability

Instead of the log odds, the output of multinomial logistic regression is the log odds of the probability of one category over the probability of the base category. In this analysis, three categories will be used, name them: walking, car and bike. The mode used as a base category will be car. This model will allow to evaluate what are the most important factors that affect the mode choice of individuals. Therefore, the 2 equations of the multinomial logit regression can be seen in Eq. 5 (Category bike compared with category car) and Eq. 6 (Category foot compared with category car) below.

$$\ln\left(\frac{P_{Category\ bike}}{1-P_{Category\ car}}\right) = \beta 0 + \beta 1\mathrm{X}1 + \ldots \beta \mathrm{k}\mathrm{X}\mathrm{k} \qquad \text{Eq. 5}$$

$$\ln\left(\frac{P_{Category\ foot}}{1-P_{Category\ car}}\right) = \beta 0 + \beta 1\mathrm{X}1 + \ldots \beta \mathrm{k}\mathrm{X}\mathrm{k} \qquad \text{Eq. 6}$$

It must be mentioned that the multinomial logit regression follows the assumption of the Independence of irrelevant alternatives (IIA). It states that the relative likelihood of being in one category compared to the baseline category would not change if another category were added.

6. Results

The findings of the study in their entirety will be presented in this section of the report. It will begin with a quick study of the data's descriptive characteristics. Following that, the results of the regressions that were used to determine which factors affect the frequency of trips to the grocery store will be presented, and this chapter will come to a close with the results of the multinomial logistic regression that was used to predict the transportation mode chosen by the users to do their grocery shopping.

6.1 Descriptive analysis

With the use of data visualization tools, some inferences may be derived from the resulting graphs. Figure 5 and Figure 7 illustrate, for instance, how the distance to get to the store, the trip duration to the shop as well as the time spent shopping, influence the number of food shopping visits.

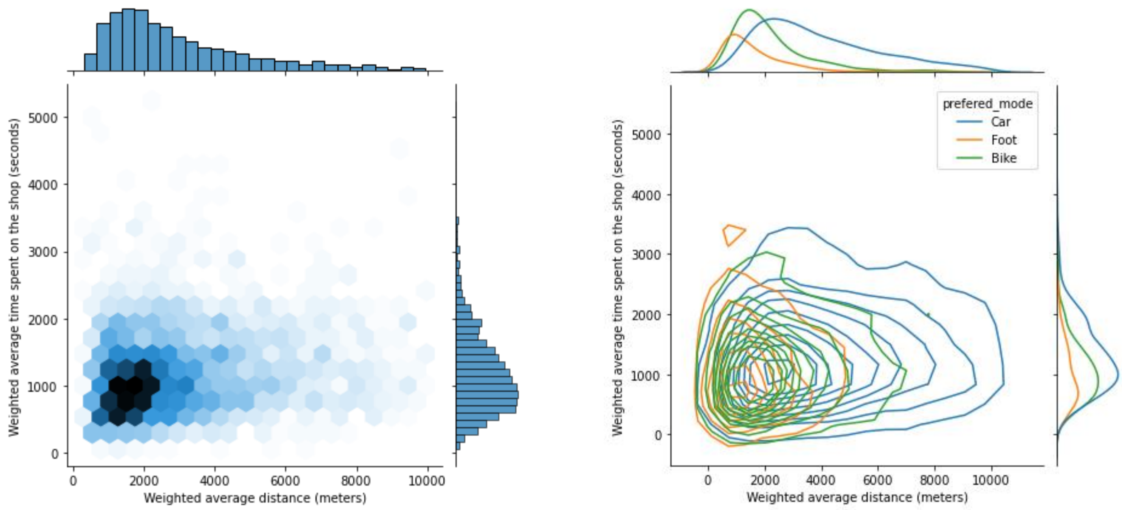


Figure 5.- Relation between distance to the shop, time spent on the shop and mode used

Figure 5 above indicates the relation between the distance to the supermarket, the time spent shopping and the mode used. In the left figure, the darker the colour the higher the amount of trips. As seen above, the majority of grocery shopping trips are within a 2-kilometer radius. This may be the case because the policies in the Netherlands focus on promoting sustainable transportation modes, thus, their infrastructure is built in a way that most of the necessary amenities can be reached by walking or cycling. For example, on average, Dutch residents live 0.9 kilometres from the largest supermarket, but also, eight in ten Dutch inhabitants live within one kilometre of a supermarket (CBS, 2022). Nevertheless, the distance to the supermarket changes according to their province as can be seen on Figure 6.

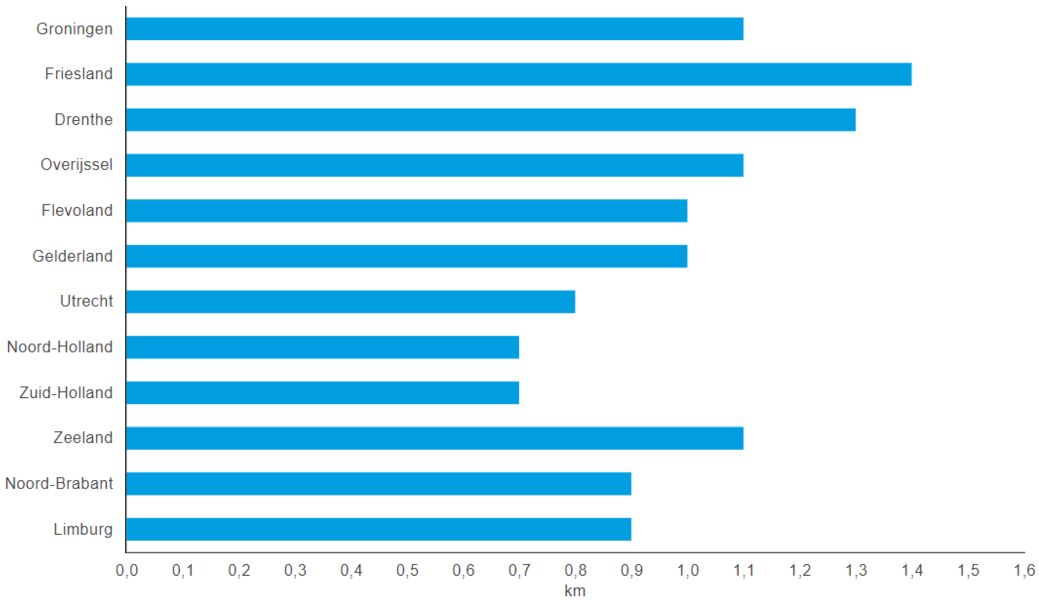


Figure 6.- Average distance to a large supermarket - 2020 – Provinces. Taken from: https://crow.databank.nl/viewer/?workspace_guid=3fb1ac06-1ded-4faf-ae93-55bfcfc65d0f

Furthermore, from the right side of Figure 5 above, it can also be seen that the distance has a high influence on the transportation mode chosen by individuals. Within 2 km to the supermarket, it is clear that all three means of transportation are equally viable; however, as the distance increases beyond this point, the likelihood of using a car increases significantly. Additionally, despite the fact that literature indicates that the average Dutch shopper spends around 32 minutes on shopping-related activities (no information was found about grocery shopping time in the Netherlands) (eurostat, 2021), the data indicates that the average time spent for purchasing groceries is 15 minutes (see Appendix C.- Descriptive analysis of dataset).

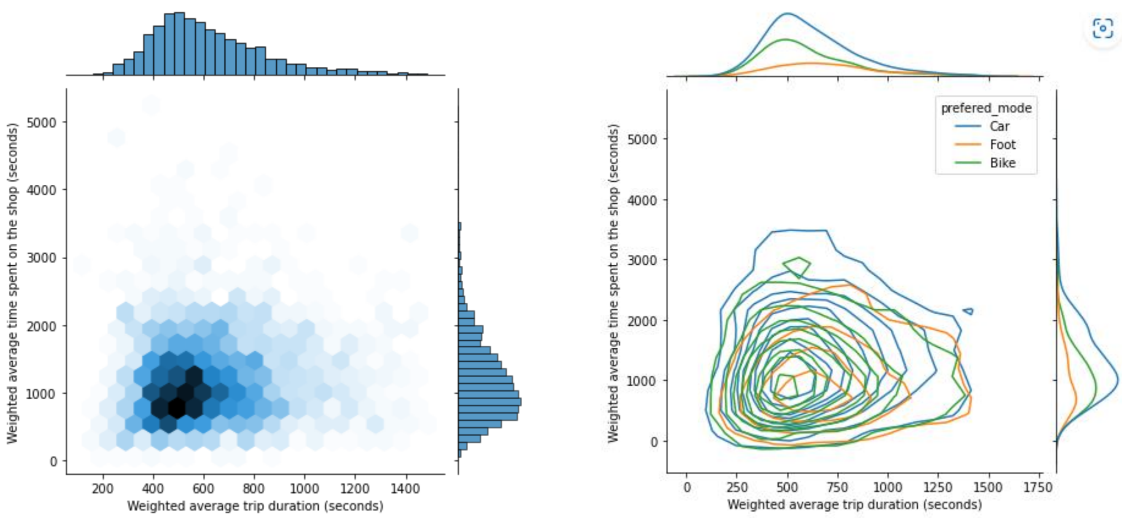


Figure 7.- Relation between trip duration, time spent on the shop and mode used

Regarding the trip duration, despite of the mode of transportation used, Figure 7 shows that individuals spent on average around 10.71 minutes to reach the store. It also shows that no users are willing to travel more than 25 minutes to buy their monthly or weekly groceries. It should be emphasized that only visits within a radius of 25 kilometers were evaluated for this research, since larger journeys were categorized as special trips. Another interesting fact about the figures above is that people traveling by car is willing to spend more time doing their grocery shopping. This may be related to the fact that possessing a vehicle makes it possible to purchase many more items and transport them home in a more comfortable way, something that would be impossible by bicycle or by foot.

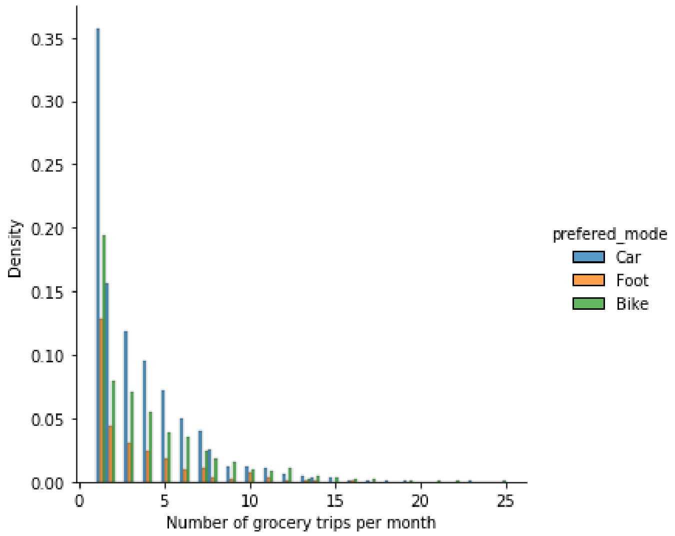


Figure 8.- Number of grocery shopping trips per month and per mode

Figure 8 displays the monthly amount of grocery shopping trips. The monthly average number of grocery shopping trips was found to be 3.41 in the dataset. Nevertheless, the figure above implies that the majority of individuals do their food shopping once every month. However, this contradicts what was found in the literature, since 61% of respondents in a poll performed in the Netherlands stated that they do their grocery shopping two to four times each week (Gelder, 2021). It can be explained by the fact that this study solely evaluates travels to physical stores, but the poll may possibly include responses from individuals who opted to purchase food online as well.

Furthermore, despite the fact that the Netherlands is a nation where cycling is greatly promoted, it has not been able to make the bicycle the major form of transport for grocery shopping trips yet. It can be seen in Figure 8 where despite the monthly number of trips, car is the most common transportation mode used.

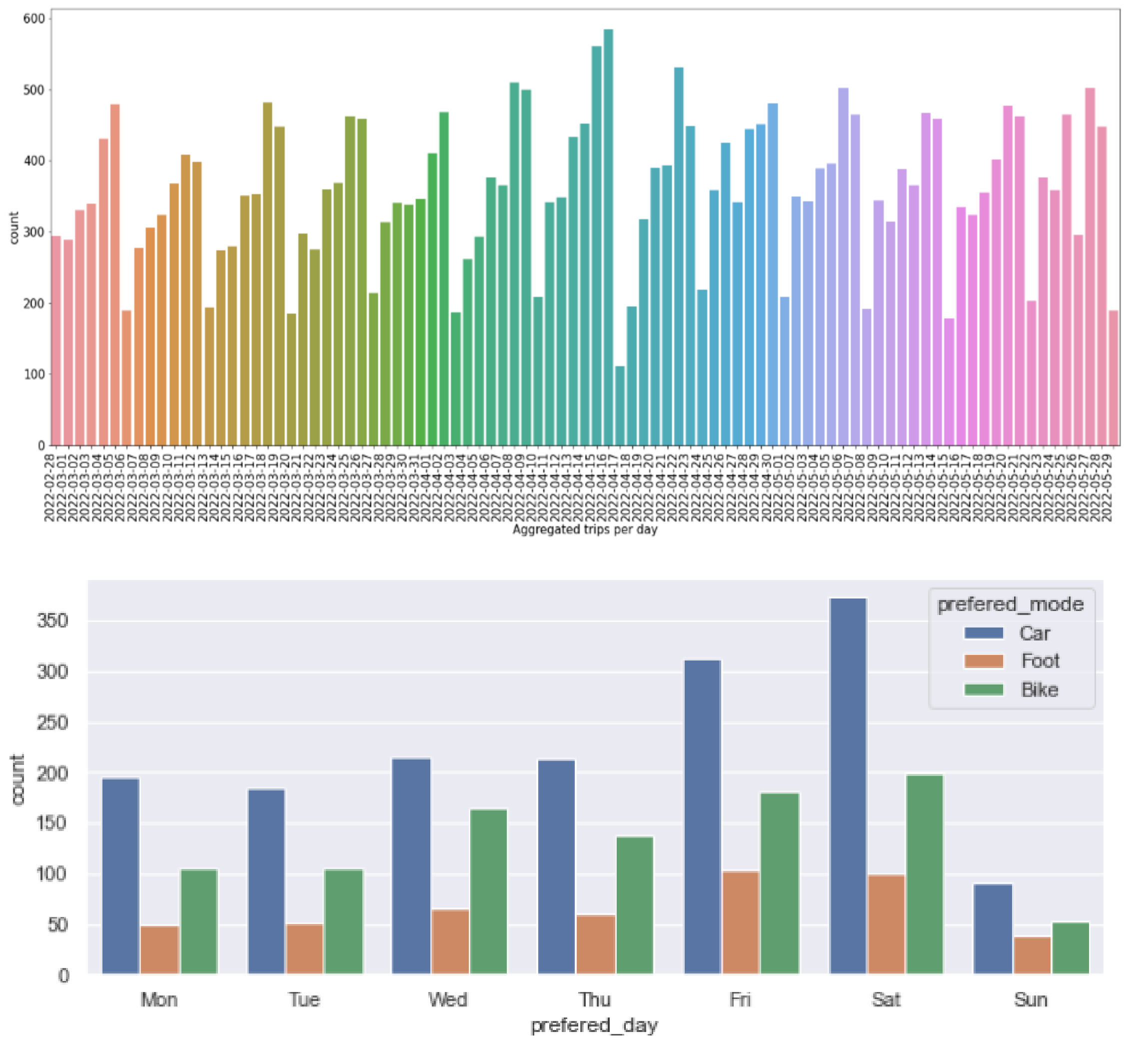


Figure 9.- Patterns in grocery shopping trips

The diagrams above confirm the concept of seven-day cycles discussed in the literature. It can be seen that the number of trips begins to rise from Monday through Saturday, but then rapidly decreases on Sunday. Research indicates that Saturday is the most popular day for grocery visits, it is claimed that this is due to the fact that most people are not employed on Saturdays and so have more time to shop (Mitova, 2021). Sundays, according to the statistics, are the days on which fewer travels are reported. This could be due to the fact that although individuals have free time like on Saturdays, the opening hours of most supermarkets are shorter than on other days of the week, ranging from midday to 5:00 p.m. in most of them.

6.2 Trip frequency

Several inferences were drawn from the data's graphical output. However, to get a more precise knowledge of the link between the variables outlined in the preceding chapters and trip frequency, two regressions were conducted. As mentioned in Section 5 above, the main intention was to use a Poisson regression model, which can predict count data, but, while looking at the data, it was found that the mean and variance of the sample is quite large. Therefore, a negative binomial model was also performed. The two models were compared by using residual plots and Likelihood ratio test to find out which one fits the data better.

From Figure 6, it can be seen that the residuals are more spread out in the Poisson regression model compared with the negative binomial one. The residuals are the difference between an observed value of the response variable and the value predicted by the regression. In the Poisson, some residuals are smaller than -2 and larger than 3. Thus, the reduced residuals associated with a negative binomial regression model indicate that this model is more suitable.

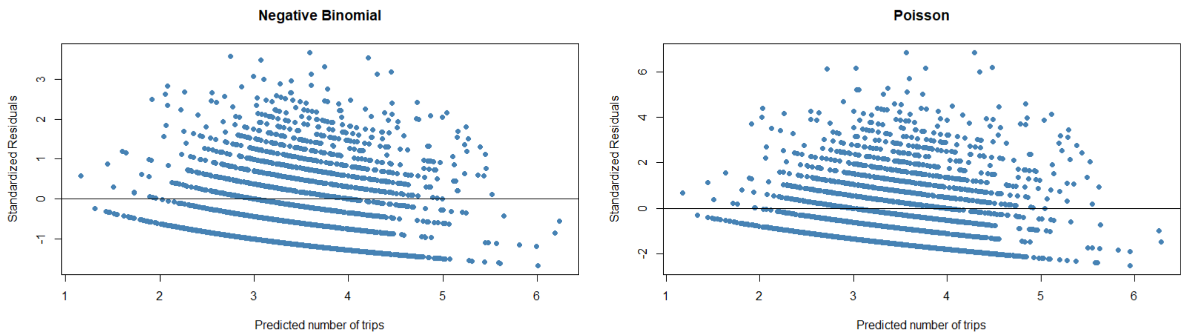


Figure 6.- Residual plots from Poisson and Negative binomial regressions

Besides the graphical results, the log likelihood and AIC value of the two models were compared to evaluate whether one model is more statistically significant than other. The results are as shown in Table 2. On the one hand, the log likelihood value of the models can be compared because they have the same number of predictor variables and the model with the bigger value is the one that fits better the data. On the other hand, the AIC value is also compared, if a model is more than 2 AIC units lower than another, it is said to be considerably better. Therefore, the Negative Binomial model also offers a better fit comparing these values.

Table 2.- Loglikelihood and AIC

	Negative Binomial model	Poisson model
Log likelihood	-6.691	-7.434
AIC	13432	14888

Concluding, the negative binomial models has demonstrated not only graphically but also statistically that fits this dataset better than the Poisson regression model, therefore, the results of the NB model will be used (for complete results of both models, see Appendix C.- Regression results).

Table 3.- Parameters estimation for Negative binomial model

Parameter estimates								
term	Estimate	std.error	statis-tic	p.value	95% Con. Interval		Exp(Esti-mate)	
(Intercept)	1.6589	0.1654	10.0280	0.0000	1.3321	1.9861	5.2537	***
Trip distance	0.0000	0.0000	-4.7450	0.0000	-0.0001	0.0000	1.0000	***
Trip duration	-0.0001	0.0001	-1.9594	0.0501	-0.0003	0.0000	0.9999	.
Time spend on supermarket	-0.0001	0.0000	-3.7399	0.0002	-0.0001	0.0000	0.9999	***
Preferred mode (Car)								
Bike	-0.0637	0.0372	-1.7141	0.0865	-0.1373	0.0098	0.9383	.
Foot	-0.2502	0.0488	-5.1286	0.0000	-0.3471	-0.1533	0.7786	***
City density (High density)								
Low density	-0.0530	0.0452	-1.1732	0.2407	-0.1415	0.0356	0.9484	
Medium density	-0.1169	0.0351	-3.3326	0.0009	-0.1858	-0.0482	0.8896	***

In order to examine the model, we discussed and concentrated on independent factors connected to the dependent variable that have a statistical significance of less than 0.05, as indicated in Table 3 (as the final fitted regression model is too large, it can be seen in Appendix C.- Regression results). Results show that the most significant variables that help to predict the frequency of grocery shopping trips are the time it takes to go from the origin point to the supermarket, the time spent on the supermarket, the mode of transportation and the annual income.

To start with, the results suggest that as the time needed to go to the supermarket increases by 1 unit (seconds), the expected log count of the number of trips decreases by a factor of 0.00014. In other words, the time it takes to go to the supermarket is inversely proportional to the frequency of shopping trips. It is also the same case with the distance to the supermarket, as the distance increases by 1 meter, the log count of the trip's frequency decreases by a factor of 0.00004. This is supported by Gustat et al., who studied the association between frequency of grocery shopping and the distance to the supermarket in New Orleans. They also concluded that the distance to the supermarket highly influence the number of grocery trips made per month, while the greater the distance, the fewer trips were made (Gustat, O'Malley, Luckett, & Johnson, 2015).

Secondly, regarding the transportation mode, it was shown that they are also statistically significant variables that predict the trip's frequency to the grocery store. For instance, the indicator variable bike is the expected difference in log count between people going by bike and the reference group (people going by car). The expected log count is 0.064 lower than the expected log count for people choosing car. It is the same with people who walks to the supermarket, the expected log count is 0.24 lower than the expected log count for individuals using car. This was also proven in the sections above, where it was shown that car was the predominant mode of transportation for grocery shopping trips.

With reference to the day of the week, the results show that the only statistically significant variable is shopping on Sundays. It is highly probable that users do not do their grocery shopping activities in the mentioned day. In terms of numbers, the difference in the logs of expected counts is expected to be 0.13 unit lower for people buying on Sundays compared with people buying on Saturdays, while keeping the other variables constant. Although the other days of the week were not statistically significant, they show to have a negative effect on the frequency of grocery shopping trips, meaning that Saturday is the preferred shopping day for groceries. This goes as explained on section 6 where Mitova explained that this happens because individuals do have more time for this activity on Saturdays because most people do not have to work or study (Mitova, 2021).

About the socio-demographic characteristics, the first statistically significant variable is the density of the city where the user lives. The results of the regressions show that users living in higher density cities are doing more trips to the supermarket. This could be the case because areas of higher density do more less long-distance trips because the amenities are closer to each other (Czepkiewicz1, Heinonen, & Ottelin, 2018). For instance, as grocery shopping is considered as a short trip, it makes sense that in cities with higher densities have more grocery shopping trips.

Regarding professions, retired and people working for the government are important to predict the frequency of this trips and they are doing grocery trips less frequent than employed individuals. Besides that, the output of the regression suggests that only people who works at home such as homemakers travel more frequently to the supermarket than employed people, hence, the difference in the logs of expected counts is expected to be 0.04 units higher for housemakers than employed.

The education level, gender, house composition, annual income and car ownership are variables that are not that useful to predict how often people do their grocery shopping. However, results show that people with low income do more trips to the supermarket than those with higher income. It is possible that this happen because people with higher income value time over money, therefore, they prefer to do not take their time doing grocery shopping but to order it online (Angelovska, 2019). Moreover, the online grocery shopping in the Netherlands has been increasing in the last years, currently, most of the popular supermarkets offer delivery services to their clients, some offer monthly subscriptions, meanwhile others just charge user for each trip (Vondráčková, 2022).

Besides that, grocery shopping trips seems to do not relate with gender, thus, male and female are equally likely to do the same number of trips to the supermarket. Finally, owning a vehicle positively affects the frequency of this type of trips. This finding is also supported by Gustat et al. (Gustat, O'Malley, Luckett, & Johnson, 2015).

6.3 Multinomial logistic regression

The multinomial logit model was used to determine the likelihood of selecting a specific mode from among multiple exclusive possibilities. In this study, three possible outputs are considered, namely them: car, bike and foot. In this instance, car is treated as the referent group and therefore estimated a model for bike relative to car and for foot relative to car. Consequently, since the parameter estimates are relative to the referent group, the standard interpretation of the multinomial logit is that for a unit change in the predictor variable, the logit of outcome m (where m is the number of levels of the outcome variable) relative to the referent group is expected to change by its respective parameter estimate (which is in log-odds units), assuming all the other variables in the model remain constant. The results of the multinomial logit regression model will be discussed in this section, however, the table with the results can be seen in Appendix C.- Regression results because the table was too large to attach it here. The results show that most of the independent variables are statistically significant to predict the mode of transportation used for grocery shopping trips. In order to have a better understanding of the variables, the marginal effect of the independent variables will be plotted because it shows the relationship between the respective independent variable and mode choice after taking all other variables into consideration.

To start with, distance to the supermarket is statistically significant to predict transportation mode at a 0.05 level. Results show that if a the distance to the supermarket were to increase by one unit, the multinomial log-odds for choosing bike to car would be expected to increase by 0.0005 units while keeping the other variables constant. Although in previous sections it was mentioned that car was the preferred mode of transportation, it was also discussed that when the distance to the supermarket is short, greener ways to travel were more likely to be chosen, and the data shows that 39% of all grocery trips are shorter than 2 kilometres. Figure 10, shows the marginal effects of trip distance to the supermarket on transportation mode choice. It can be seen that within 500 meters the probabilities of travel by walking to the supermarket are higher. After 1 km until 2 km the predominant mode is bike and finally, after 2 km it looks like people are less likely to use greener modes of transportation, since the probability of using car starts to significantly increase as the distance increases. This is supported by Jiao, Moudon and Drewnowski,, who studied how individuals and built environments are related to the choice of travel mode, their results suggests that reducing the distance between homes and grocery stores will lead to decreasing the likelihood of individuals choosing cars as transport mode (Jiao, Moudon, & Drewnowski, 2011).

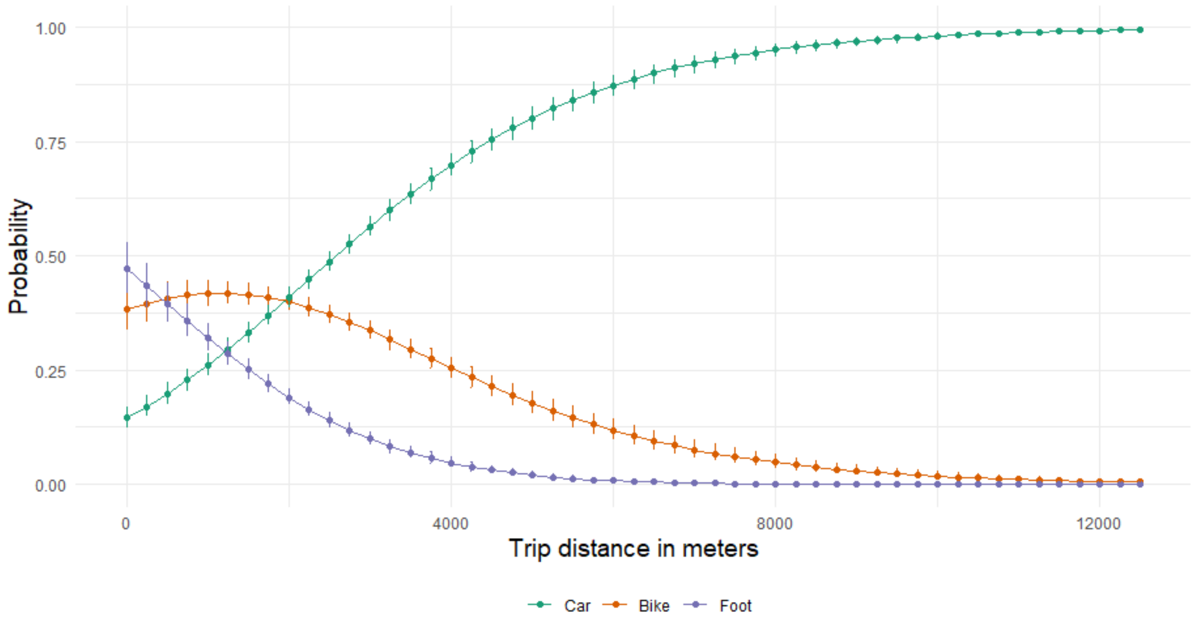


Figure 10.- Marginal effect of distance to the shop on transport mode choice

Shopping time spent on the supermarket was also found to be statistically significant in this research. Figure 11 describes the probability of choosing a travel mode depending on the time spent on the shop. It is believed that the shopping time is positively related with the number of products bought. Hence, as more time is spent on the supermarket, it is assumed that more products are been bought. Therefore, when shopping for a longer period of time and purchasing a significant quantity of groceries, a larger vehicle is required to convey them to their final destination. In addition, individuals value the automobile's superior comfort over other modes of transportation, thus, it is understandable that it is more likely to choose car when the shopping time is larger.

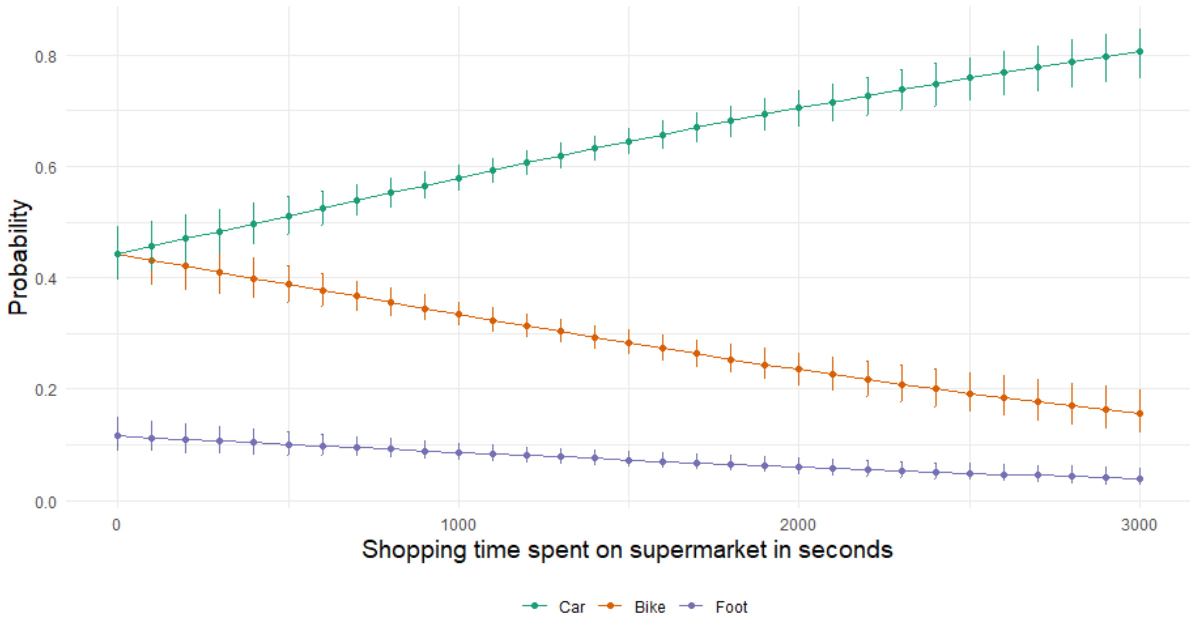


Figure 11.- Marginal effect of shopping time on transport mode choice

Regarding the age, if an individual were to increase their age by one year, the log-odds of preferring bike and walking to car are expected to decrease by 0.016 and 0.023 units respectively, given that the other variables on the model are keep constant. In other words, considering two individuals with the same socio-demographic characteristics, the individual with more age is more likely to prefer greener ways of transportation than car. However, when the other variables are controlled as well, the figure below is obtained.

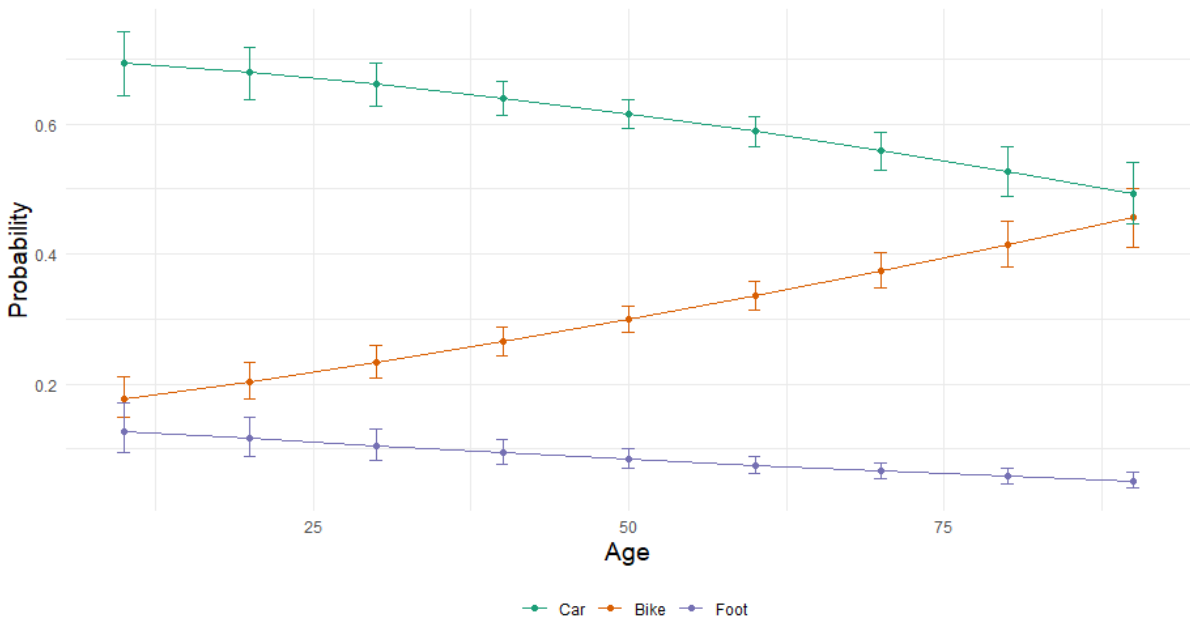


Figure 12.- Marginal effect of age on transport mode choice

This can be caused for two major reasons in the Netherlands. Firstly, the growth of e-bikes, in fact, Veldhoven assures that in the Netherlands, more e-bikes than conventional bicycles have been sold since 2018. In 2019, a record of 420,000 new e-bikes were sold in the Netherlands, accounting for about 70% of the total bicycle sales revenue (Veldhoven, nd). Secondly, because of traffic-calmed streets, separate cycling networks, one of the most secure cycling infrastructures around the world and all the benefits that cycling bring to health has increase the number of trips made by bike. For instance, a study conducted by Hoed and Jarvis showed that individuals after their 40s start to use their bikes more frequently because they see it as a good activity to keep them healthy (Hoed & Jarvis , 2021). Hence, by combining these two points, a relation can be found between elderly people cycling more as e-bikes allow elderly to keep cycling even when their stamina starts to decrease.

About the day of the week, the only statistically significant variables are Wednesday and Sunday, where from some not known reason the relative risk of using bike is relatively lower compared with other days of the week. No reason was found on the literature for this. Additionally, the multinomial logit for females relative to males is 0.14 and 0.16 units higher for preferring bike and foot respectively, relative to car, meaning that females are slightly more likely to prefer greener ways to travel. Regarding the density of the city, although, in general, car is the most common mode choice, when individuals live in high density cities, the probability of using bike slightly increases compared with lower density cities. This was discussed in section 5 above, where it was found that high densities have shorter travel distances to reach amenities and when the distance is shorter, individuals tend to cycle more because it is less costly (Bath, 1998). About statistics, the log-odds for high density cities relative to low density cities is 0.14 and 0.03 units lower for preferring bike and walking compared to car ((see Figure 17, Figure 18 and Figure 19 in Appendix E.- Marginal effects of independent variables on mode choice).

About the household composition, the reference group was taken to be households with young children, On the one hand, the results of the regression show that it is less likely to choose bike or walk relative to car for single households or adult households compared with the reference group (households with young children). On the other hand, the relative probability of choosing bike rather than car is 39% higher for households with adolescent children than for households with young children. Moreover, it was also found that people with only high school level education are more likely to choose bike than people with higher levels of education. This makes sense, since highly levels of education are related with more earning, therefore, they are the ones that can afford to buy a car, meanwhile, lower income would prefer to save the costs that traveling to the supermarkets could have. This is also supported by the results of the regressions, since it was found that for low-income individuals related to high income, the relative risk for preferring biking and walking to driving would be expected to increase by a factor of 1.09 and 1.16 respectively, given the other variables are kept constant (see Figure 20, Figure 21 and Figure 22 in Appendix E.- Marginal effects of independent variables on mode choice).

Finally, all profession but unemployed were also found to be statistically significant to predict mode choice at a 0.05 level. The output says that retired, homemakers, and entrepreneurs are more likely to bike or walk to the supermarket than employed individuals. Opposite to this, for people who work for government institutions and students relative to employed individuals, the relative risk for preferring bike to car would be expected to decrease by a factor of 0.43 and 0.83 respectively found to prefer car rather than greener modes of transportation. And to conclude, car ownership was demonstrated to be statistically significant as well, in fact, the multinomial logit for individuals that do not have a car relative to the ones that do have one is 1.32 and 0.47 units higher for preferring bike and walking to use car. This is obvious since the probability to do this type of trips by car when individuals do not own one are almost zero, however, it is not completely 0 because it could be the case that a friend who owns a car take you to the grocery store (see Figure 23 in Appendix E.- Marginal effects of independent variables on mode choice).

7. Conclusions and recommendations

This research aimed to contribute to have a better understanding of the characteristics of shopping grocery trips in the Netherlands. In order to do so, two principal factors were analysed: the frequency of this trips and the transportation mode choice. This was done by performing two regression models which helped to determine which socio-demographic and trip features variables were statistically significant helpful to predict the number of trips per month and the mode choice.

First, descriptive statistics enabled the identification of data patterns. It was noticed that grocery shopping follows a seven-day cycle, that the number of trips increases from Monday through Saturday, but then drops drastically on Sunday, and that Saturday is the most popular shopping day since people have more free time to do their groceries. In addition, the car was discovered to be the most popular form of transportation, and the majority of users purchase their groceries once per month.

Secondly, a negative binomial regression model revealed that the distance to the supermarket, the time it takes to get there, the method of transportation used, and the annual income have a substantial impact on the frequency of grocery shopping trips. Initially, research showed that grocery shopping frequency increases as the distance to the supermarket decreases. Moreover, despite all the policies implemented by the provinces and municipalities in the Netherlands to reduce car use, it has not yet been possible to make the bicycle the predominant mode of transportation for these types of trips; as a result, the car continues to be the predominant mode of transportation due to the comfort it provides. In addition, it was shown that there is an inverse link between income and shopping frequency, such that individuals with lower incomes go grocery shopping more often than those with higher incomes. It is believed that this is due to the rise of grocery delivery services in the Netherlands, as well as the fact that individuals with greater incomes value their time more than their money, and hence prefer not to waste their time going to do their grocery shopping.

Thirdly, a multinomial logit regression model was employed to investigate whether certain variables influence the preference for one mode of transportation over another. The bulk of sociodemographic and topographical characteristics are statistically significant in determining mode selection, according to this study. One of the most intriguing findings was that as individuals age in the Netherlands, their choice for greener modes of transportation grows. Additionally, it was discovered that those with a greater income are less likely to pick biking or walking as their method of transportation, as their income allows them to spend more money on grocery shopping trips. Regarding the density of the density, it was discovered that individuals living in cities with higher densities are more likely to choose to walk or bike to the supermarket. This is because higher densities indicate that amenities are located closer together, making walking or biking more economical than driving. It was also discovered that females are marginally more likely to adopt greener modes of transportation. It was shown that car ownership greatly reduces the likelihood of opting to cycle or walk. Regarding trip features, it was found that the distance to the supermarket greatly influences transportation mode choice, since within 2 km distance, the three modes of transportation are slightly equal probable to be chosen, however, after this distance, the probability of choosing car over bike or walk increases significantly. Also, as more time is spent in the supermarket, the probability of choosing car increases because it is assumed that more products are been bought and a more comfortable way to deliver them to home must be chosen.

In conclusion, this study evaluated how specific factors influence the frequency of grocery shopping trips and mode of transportation. However, the travel behaviour behind these journeys is more complex than it appears. Many other circumstances could have a substantial impact on this, such as whether, which is truly significant, since it is not expected that people will go grocery shopping by bicycle while it is raining. Moreover, the proliferation of e-bikes may increase the likelihood of choosing cycling as the primary mode of transportation for this activity, as cycling is more cost-effective than driving and e-bikes lessen the physical exertion required to cycle to the grocery. Regarding the frequency of trips, it would be advisable to analyse in greater detail how the increasing availability of online grocery stores has affected the frequency of grocery shopping, given that the Netherlands offers a large number of carriers that will deliver their groceries to their home for a very reasonable cost.

8. References

Aloulou, F. (2018). The Application of Discrete Choice Models in Transport. 85-102. doi:10.5772/intechopen.74955

Angelovska, N. (2019). *Forbes*. Retrieved from Who Are The Top Online Grocery Shoppers In Europe?: https://www.forbes.com/sites/ninaangelovska/2019/04/27/who-are-the-top-online-grocery-shoppers-in-europe/?sh=4456b8e15b46

Bath, C. R. (1998). Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological, 32*(6), 361-371. doi:10.1016/s0191-2615(98)00004-6

Byung, K. D., & Kyungdo, P. (1997). Studying patterns of consumer's grocery shopping trip. *Journal of Retailing, 73*(4), 501-517. doi:https://doi.org/10.1016/S0022-4359(97)90032-4

CBS. (2010, 09 07). Retrieved from Supermarket within walking distance for most Dutch people: https://www.cbs.nl/en-gb/news/2010/36/supermarket-within-walking-distance-for-most-dutch-people#:~:text=On%20average%2C%20people%20in%20the,kilometres%20from%20their%20nearest%20supermarket.

CBS. (2022). Retrieved from Supermarket within walking distance for most Dutch people: https://www.cbs.nl/en-gb/news/2010/36/supermarket-within-walking-distance-for-most-dutch-people

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility). *Transportation research. Part C, Emerging technologies*, 285-299. doi:10.1016/j.trc.2016.04.005

Clarke, M., Dix, M., & Jones, P. (1981). Error and uncertainty in travel surveys. *Transportation, 10*(2). doi:10.1007/BF00165261

Coxe, S., West, S., & Aiken, L. (2009). The Analysis of Count Data: A Gentle Introduction to Poisson Regression and its Alternatives. *Journal of personality assissment, 91*(2), 121-136. doi:10.1080/00223890802634175

Czepkiewicz1, M., Heinonen, J., & Ottelin, J. (2018). Why do urbanites travel more than do others? A review of associations between urban form and long-distance leisure travel. *Environmental Research Letters*. doi:10.1088/1748-9326/aac9d2

Esztergár-Kiss, D., Shulha, Y., Aba, A., & Tettamanti, T. (2021). Promoting sustainable mode choice for commuting supported by persuasive strategies. *Sustainable Cities and Society, 74*(2210-6707), 103-264. doi:https://doi.org/10.1016/j.scs.2021.103264.

Gelder, K. v. (2021). *Statisca*. Retrieved from How often do you go to the supermarket to buy groceries?: https://www.statista.com/statistics/886964/supermarket-visit-frequency-in-the-netherlands/#:~:text=In%202021%2C%20a%20significant%20share,least%20two%20times%20per%20week.

Gustat, J., O'Malley, K., Luckett, B. G., & Johnson, C. C. (2015). Fresh produce consumption and the association between frequency of food shopping, car access, and distance to supermarkets. *Preventive Medicine Reports, 2*, 47-52. doi:https://doi.org/10.1016/j.pmedr.2014.12.009.

Hagberg, J., & Holmberg, U. (2017). Travel modes in grocery shopping. *InternationaL Journal of Retail & Distribution Management, 45*, 991-1010. doi:10.1108/ijrdm-08-2016-0134

Handy, S. (1996). Non-Work Travel of Women: Patterns, Perceptions and. *WOMEN'S TRAVEL ISSUES*, (pp. 315-336).

Harms, L., & Kansen, M. (2018). Cycling facts. *Netherlands Institute for Transport Policy Analysis (KiM)*.

Hoed, W. d., & Jarvis , H. (2021). Normalising cycling mobilities: an age-friendly approach to cycling in the NetherlandsApplied Mobilities. *Applied Mobilities*, 1-21. doi:10.1080/23800127.2021.1872206

Jiao, J., Moudon, A., & Drewnowski, A. (2011). Grocery Shopping: How Individuals and Built Environments Influence Choice of Travel Mode. *Transportation research record, 2230*, 85-95. doi:10.3141/2230-10

Johnson, A. A., Ott, M. Q., & Dogucu, M. (2021). *Bayes Rules! An Introduction to Applied Bayesian Modeling*, Poisson & Negative Binomial Regression.

Kahn, B., & Schmittlein, D. (1989). Shopping Trip Behavior: An Empirical Investigation. *Marketing Letters, 1*, 55-69. doi:10.1007/BF00436149

Kara, N. S., & Mkwizu, K. H. (2020). Demographic factors and travel motivation among leisure tourists in Tanzania. *International Hospitality Review, 34*(1), 81-103. doi:10.1108/IHR-01-2020-0002

Mitova, T. (2021, august 20). *SPENDMENOT*. Retrieved from 21+ Grocery Shopping Statistics for Every CUSTOMER in 2021: https://spendmenot.com/blog/grocery-shopping-statistics/

Mobidot. (2022, 07 20). Retrieved from https://mobidot.nl/en/

Recker, W., & Kostyniuk, L. (1978). Factors influencing destination choice for the urban grocery shopping trip. *Transportation, 7*(1), 19-33. doi:10.1007/BF00148369

Rhone, A., & Ploeg, M. V. (2019, 06 03). *USDA Economic Research Service*. Retrieved from U.S. Shoppers' Access to Multiple Food Stores Varies by Region: https://www.ers.usda.gov/amber-waves/2019/june/us-shoppers-access-to-multiple-food-stores-varies-by-region/

Simma, A., Cattaneo, P., Baumeler, M., & Axhausen, K. (2004). *Factors influencing the individual shopping behaviour: The case of Switzerland.*

Thiagarajan, T., & Natarajan, P. (2016). Influence of Demographic Factors on Grocery Buying Behaviour. *Asian Journal of Research in Social Sciences and Humanities, 6*, 544-562. doi:10.5958/2249-7315.2016.00078.2

Veenstra, S., Thomas, T., & Tutert, B. (2010). *Trip generation of grocery shopping trips using survey data: A case study in the Netherlands.*

Veldhoven, S. v. (n.d.). *Cycling and Dutch national infrastructure: Working towards a more structural approach to incorporating cycling in national-level projects.*

Vondráčková, J. (2022). *dutchreview*. Retrieved from Online grocery shopping: a guide to Dutch grocery delivery services: https://dutchreview.com/expat/online-grocery-shopping-netherlands/

9. Appendix

9.1 Appendix A.- Sample description

Table appendix 1.- Sample description

Variables	Data type	Attribute	
Preferred mode	Categorical	Car	53%
		Bike	31%
		Foot	16%
Weighted avg distance	Numerical	Mean	3310 meters
Weighted avg travel time	Numerical		
Weighted aveg shopping time	Numerical	Mean	1199 seconds
Preferred day	Categorical	Mon	12%
		Tue	11%
		Wed	15%
		Thu	14%
		Fri	20%
		Sat	22%
		Sun	6%
Gender	Categorical	Male	43%
		Female	57%
Age	Numerical	Mean	54 years
City density	Categorical	High density	56%
		Medium density	17%
		Low density	27%
House composition	Categorical	Single	21%
		With young children	18%
		With adolescent children	7%
		Adult	54%
Education level	Categorical	Middle school or below	10%
		High school	23%
		Bachelor education or above	67%
Profession	Categorical	Entrepreneur	5%
		Employed	48%
		Governmental	7%
		Incapacitated	7%
		Unemployed	4%
		Retired	20%
		Studying	3%
		At home	6%

Car ownership	Categorical	Yes	55%
		No	4%
		Unknown	41%
Household size	Numerical	Mean	2,46 per/house
Annual income	Categorical	High income	54%
		Modal	19%
		Low income	13%
		Unknown income	14%
Driving license	Categorical	Yes	55%
		No	4%
		Unknown	41%

9.2 Appendix B.- Correlation between variables

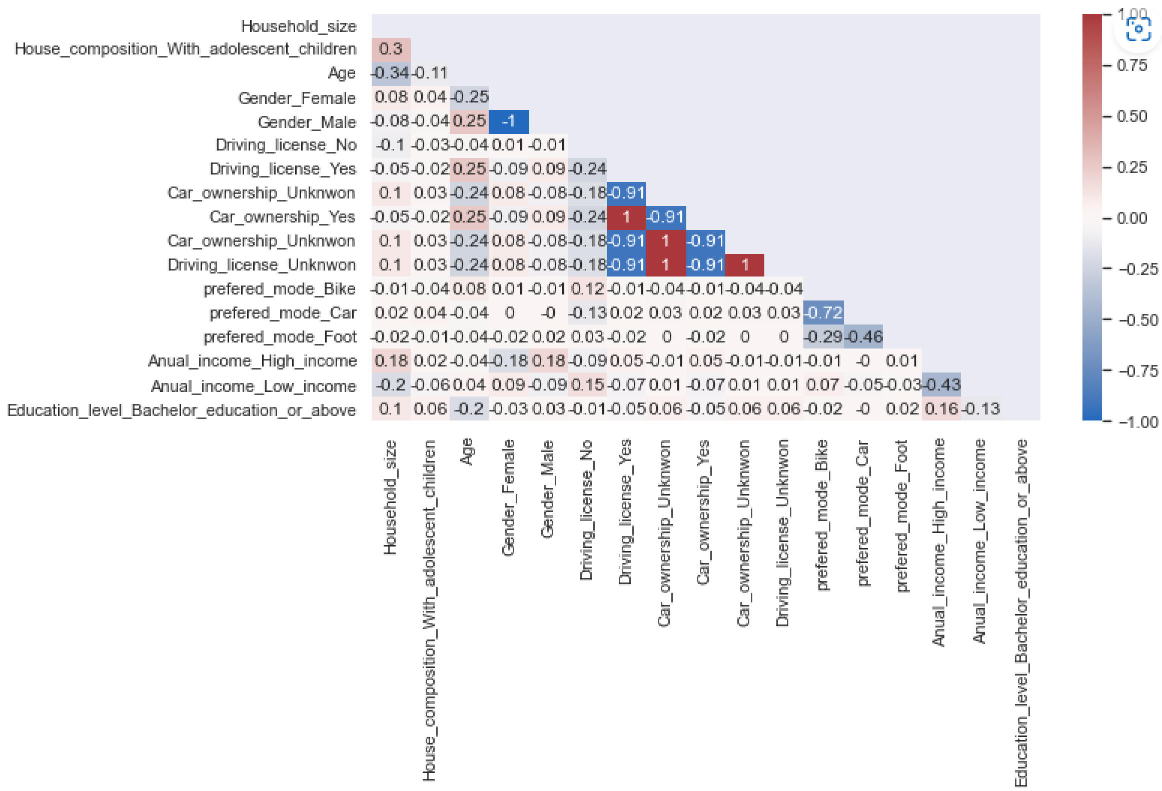


Figure 13.- Correlation of socio-demographic features

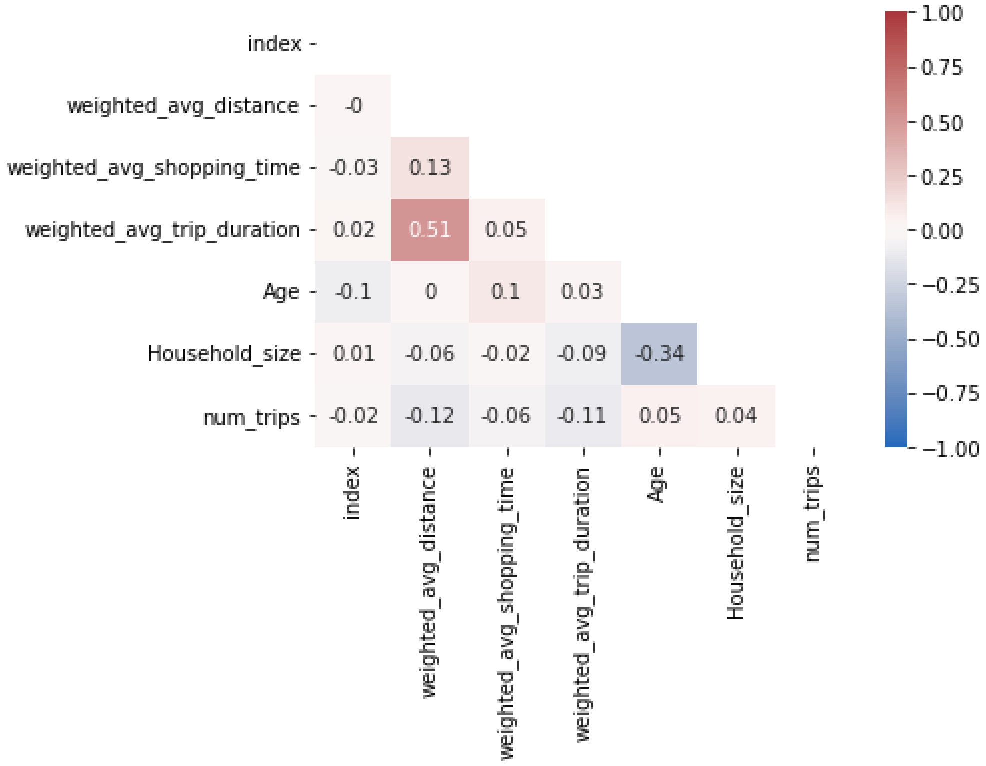
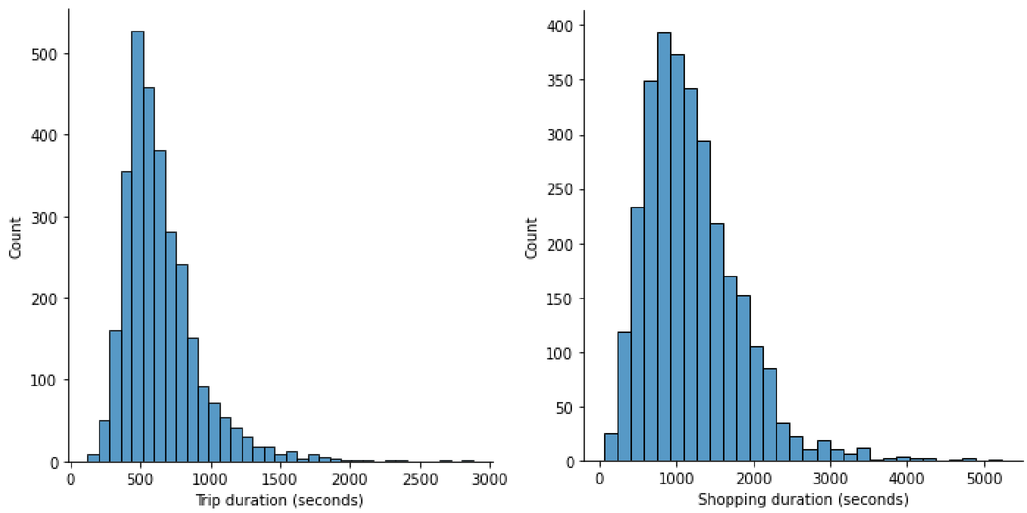


Figure 14.- Correlation of numeric variables

9.3 Appendix C.- Descriptive analysis of dataset



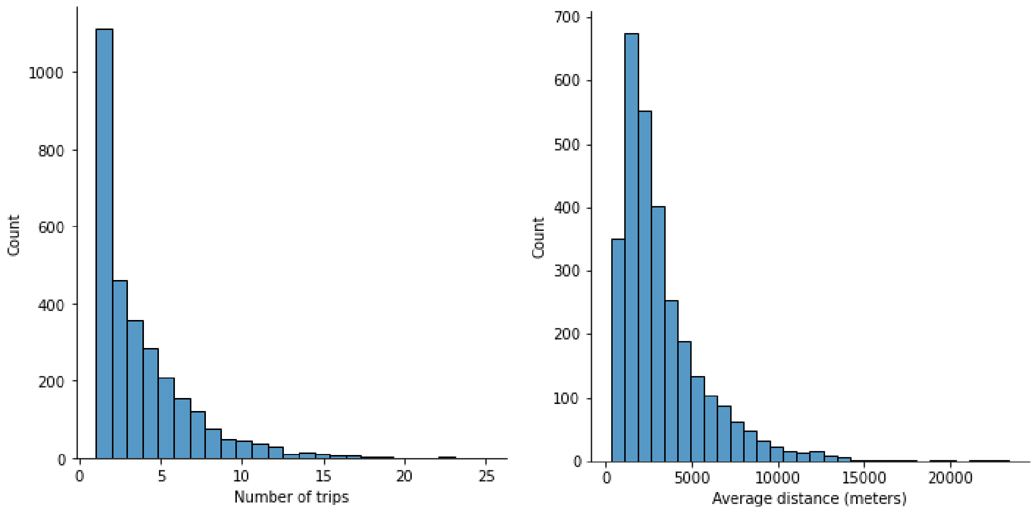


Figure 15.- Histograms of different numerical variables

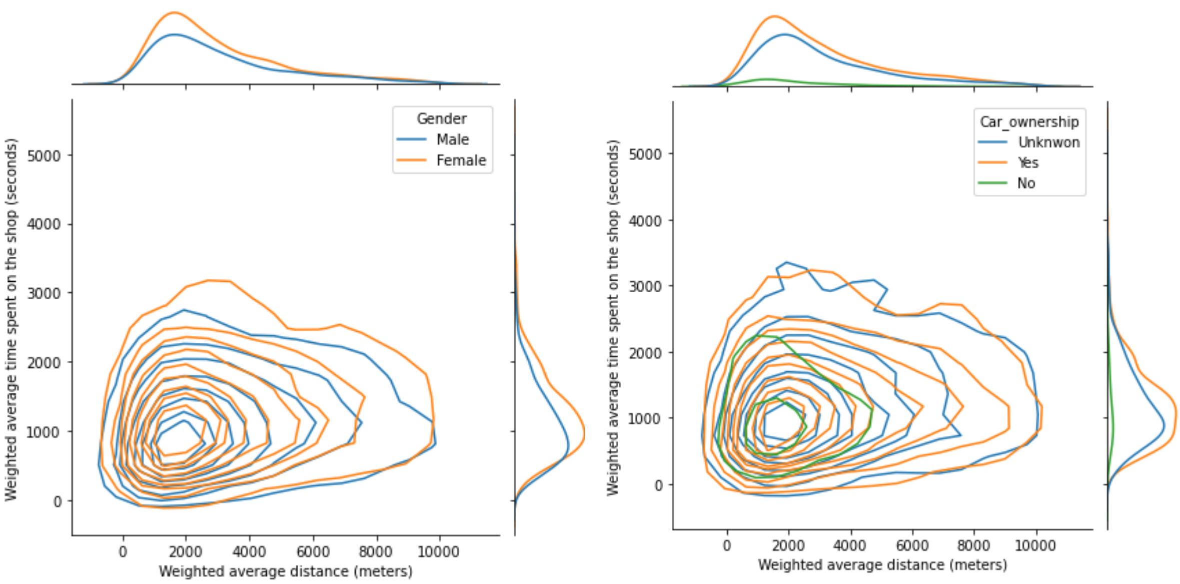


Figure 16.- Relation between distance, time spent on the shop, gender and car ownership

9.4 Appendix C.- Regression results

Table 4.- Parameter estimations for multinomial regression model

	Parameter estimates							
term	Estimate	std.error	statistic	p.value	95% Con. Interval		Exp(Estimate)	
(Intercept)	1.6589	0.1654	10.0280	0.0000	1.3321	1.9861	5.2537	***
Trip distance	0.0000	0.0000	-4.7450	0.0000	-0.0001	0.0000	1.0000	***
Trip duration	-0.0001	0.0001	-1.9594	0.0501	-0.0003	0.0000	0.9999	.
Time spend on supermarket	-0.0001	0.0000	-3.7399	0.0002	-0.0001	0.0000	0.9999	***
Age	0.0022	0.0017	1.2776	0.2014	-0.0012	0.0056	1.0022	
Household size	0.0259	0.0211	1.2251	0.2206	-0.0158	0.0676	1.0262	
Preferred mode (Car)								
Bike	-0.0637	0.0372	-1.7141	0.0865	-0.1373	0.0098	0.9383	.
Foot	-0.2502	0.0488	-5.1286	0.0000	-0.3471	-0.1533	0.7786	***
Preferred day (Monday)								
Tuesday	0.0038	0.0614	0.0611	0.9513	-0.1166	0.1241	1.0038	
Wednesday	-0.0978	0.0585	-1.6719	0.0945	-0.2124	0.0168	0.9069	.
Thursday	-0.0358	0.0591	-0.6053	0.5450	-0.1517	0.0801	0.9649	
Friday	0.0813	0.0543	1.4983	0.1341	-0.0250	0.1875	1.0847	
Saturday	0.0805	0.0532	1.5126	0.1304	-0.0237	0.1846	1.0838	
Sunday	-0.1395	0.0767	-1.8178	0.0691	-0.2902	0.0112	0.8698	.
Gender (Male)								
Female	-0.0184	0.0330	-0.5588	0.5763	-0.0833	0.0465	0.9817	
City density (High density)								
Low density	-0.0530	0.0452	-1.1732	0.2407	-0.1415	0.0356	0.9484	
Medium density	-0.1169	0.0351	-3.3326	0.0009	-0.1858	-0.0482	0.8896	***
House composition (With young children)								

Adult household	0.0302	0.0503	0.6000	0.5485	-0.0685	0.1289	1.0306	
With adolescent children	0.0676	0.0678	0.9975	0.3185	-0.0655	0.2009	1.0699	
Single	0.0562	0.0588	0.9565	0.3388	-0.0595	0.1719	1.0578	
Education level (bachelor's degree or above)								
High school	-0.0372	0.0565	-0.6588	0.5100	-0.1482	0.0736	0.9635	
Middle school or below	-0.0526	0.0526	-0.9995	0.3176	-0.1559	0.0505	0.9488	
Unknown	0.4036	0.5496	0.7342	0.4628	-0.6606	1.5503	1.4971	
Profession (Employed)								
Retired	-0.1783	0.0663	-2.6886	0.0072	-0.3086	-0.0485	0.8367	**
Entrepreneur	-0.1670	0.0893	-1.8703	0.0614	-0.3421	0.0081	0.8462	.
Incapacitated	-0.0579	0.0829	-0.6978	0.4853	-0.2201	0.1042	0.9438	
At home	0.0451	0.0800	0.5645	0.5724	-0.1117	0.2018	1.0462	
Unemployed	-0.0920	0.0723	-1.2725	0.2032	-0.2345	0.0502	0.9121	
Governmental	-0.4749	0.1675	-2.8352	0.0046	-0.8063	-0.1467	0.6219	**
Studying	-0.0796	0.0999	-0.7962	0.4259	-0.2752	0.1163	0.9235	
Unknown	0.0799	0.2504	0.3189	0.7498	-0.4172	0.5866	1.0831	
Car ownership (Yes)								
No	-0.0158	0.0748	-0.2114	0.8326	-0.1627	0.1307	0.9843	
Unknown	0.0204	0.0744	0.2740	0.7841	-0.1257	0.1661	1.0206	
Annual income (High income)								
Modal	-0.0313	0.0512	-0.6124	0.5403	-0.1319	0.0693	0.9691	
Low income	0.0223	0.0410	0.5447	0.5860	-0.0580	0.1028	1.0226	
Unknown	-0.0371	0.0459	-0.8086	0.4188	-0.1270	0.0529	0.9636	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial (2.8863) family taken to be 1)
Null deviance: 3015.9 on 2991 degrees of freedom
Residual deviance: 2804.0 on 2956 degrees of freedom
AIC: 13432

9.5 Appendix D. Multinomial logit regression results

Table 5.- Parameter estimates for multinomial regression model

Parameter estimates												
	Bike						**Foot**					
term	estimate	std.error	p.value	Exp(Estimate)	conf.low	conf.high	estimate	std.error	p.value	Exp(Estimate)	conf.low	conf.high
(Intercept)	-1.8536	0.0180	0.0000	0.1567	-1.8889	-1.8183	0.1210	0.0152	0.0000	1.1286	0.0911	0.1509
Trip distance	0.0005	0.0000	0.0000	1.0005	0.0004	0.0006	-0.0005	0.0001	0.0000	0.9995	-0.0006	-0.0004
Time spend on supermarket	0.0005	0.0001	0.0000	1.0005	0.0004	0.0007	0.0000	0.0001	0.9260	1.0000	-0.0002	0.0002
Age	-0.0160	0.0024	0.0000	0.9841	-0.0207	-0.0114	-0.0233	0.0027	0.0000	0.9769	-0.0287	-0.0180
Household size	-0.0612	0.0317	0.0535	0.9407	-0.1232	0.0009	-0.0406	0.0396	0.3056	0.9603	-0.1181	0.0370
Preferred day (Monday)												
Tuesday	-0.0336	0.0179	0.0609	0.9670	-0.0687	0.0015	-0.0191	0.0100	0.0567	0.9811	-0.0387	0.0005
Wednesday	-0.3713	0.0778	0.0000	0.6898	-0.5238	-0.2189	-0.1389	0.0377	0.0002	0.8703	-0.2128	-0.0650
Thursday	-0.1364	0.0737	0.0643	0.8725	-0.2810	0.0081	-0.0752	0.0366	0.0396	0.9275	-0.1469	-0.0036
Friday	-0.0633	0.0707	0.3703	0.9387	-0.2018	0.0752	0.1248	0.0411	0.0024	1.1329	0.0442	0.2053
Saturday	-0.0727	0.0705	0.3024	0.9299	-0.2109	0.0655	-0.0249	0.0377	0.5094	0.9754	-0.0988	0.0491
Sunday	-0.1148	0.0075	0.0000	0.8915	-0.1295	-0.1001	0.1586	0.0072	0.0000	1.1719	0.1445	0.1728
Gender (Male)												
Female	0.1414	0.0705	0.0451	1.1518	0.0031	0.2796	0.1566	0.0418	0.0002	1.1695	0.0746	0.2385
City density (High density)				1.0000								
Low density	-0.1356	0.0397	0.0006	0.8732	-0.2134	-0.0578	-0.0316	0.0197	0.1091	0.9689	-0.0702	0.0070
Medium density	-0.2855	0.0554	0.0000	0.7516	-0.3940	-0.1770	0.1251	0.0377	0.0009	1.1333	0.0512	0.1991
House composition (With young children)												
Adult household	-0.1379	0.0540	0.0107	0.8712	-0.2438	-0.0320	0.1329	0.0371	0.0003	1.1422	0.0603	0.2056
With adolescent children	0.3275	0.0096	0.0000	1.3875	0.3088	0.3463	0.0633	0.0056	0.0000	1.0653	0.0524	0.0741
Single	-0.0982	0.0246	0.0001	0.9065	-0.1463	-0.0501	-0.4652	0.0137	0.0000	0.6280	-0.4920	-0.4384
Education level (bachelor's degree or above)												
High school	0.1207	0.0460	0.0088	1.1283	0.0304	0.2109	0.3087	0.0491	0.0000	1.3617	0.2125	0.4049

Middle school or below	-0.0243	0.0557	0.6624	0.9760	-0.1334	0.0848	0.2295	0.0637	0.0003	1.2580	0.1046	0.3545
Unknown	10.3353	0.0001	0.0000	3.08E+04	10.3351	10.3354	12.3969	0.0001	0.0000	2.42E+05	12.3968	12.3971
Profession (Employed)												
Retired	0.3004	0.0625	0.0000	1.3504	0.1778	0.4229	0.7412	0.0448	0.0000	2.0984	0.6535	0.8289
Entrepreneur	0.3547	0.0089	0.0000	1.4257	0.3373	0.3720	1.2692	0.0066	0.0000	3.5579	1.2563	1.2821
Incapacitated	0.2185	0.0154	0.0000	1.2442	0.1883	0.2486	0.7770	0.0093	0.0000	2.1750	0.7588	0.7952
At home	0.2371	0.0240	0.0000	1.2676	0.1900	0.2842	0.3219	0.0104	0.0000	1.3798	0.3015	0.3423
Unemployed	0.0338	0.0301	0.2608	1.0344	-0.0251	0.0928	0.8720	0.0200	0.0000	2.3916	0.8327	0.9112
Governmental	-0.8707	0.0024	0.0000	0.4186	-0.8754	-0.8660	0.9613	0.0028	0.0000	2.6152	0.9558	0.9668
Studying	-0.1814	0.0097	0.0000	0.8341	-0.2005	-0.1623	0.6136	0.0063	0.0000	1.8471	0.6012	0.6260
Unknown	-0.1813	0.0010	0.0000	0.8342	-0.1833	-0.1793	1.7475	0.0010	0.0000	5.7404	1.7457	1.7494
Car ownership (Yes)												
No	1.3177	0.0486	0.0000	3.7347	1.2223	1.4130	0.4682	0.0609	0.0000	1.5971	0.3487	0.5876
Unknown	1.3429	0.0490	0.0000	3.8300	1.2468	1.4389	0.4424	0.0605	0.0000	1.5564	0.3238	0.5610
Annual income (High income)												
Modal	-0.2043	0.0503	0.0000	0.8152	-0.3030	-0.1057	-0.3600	0.0244	0.0000	0.6977	-0.4078	-0.3122
Low income	0.0854	0.0722	0.2369	1.0892	-0.0561	0.2269	0.1519	0.0399	0.0001	1.1640	0.0737	0.2301
Unknown	0.2609	0.0464	0.0000	1.2981	0.1699	0.3519	0.1253	0.0263	0.0000	1.1335	0.0738	0.1769

9.6 Appendix E.- Marginal effects of independent variables on mode choice

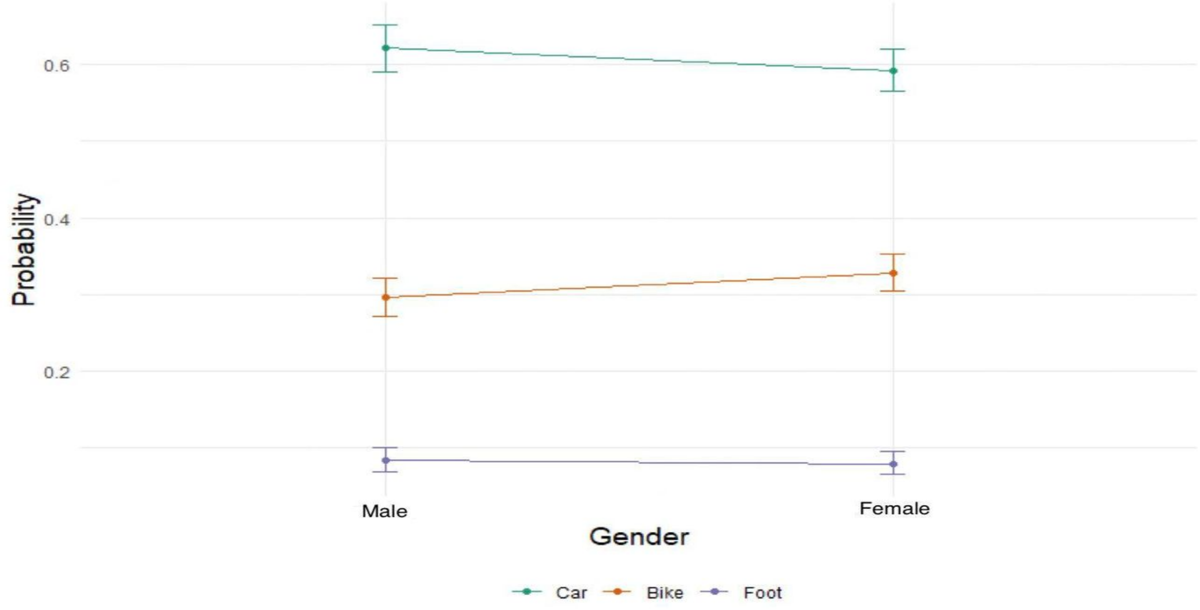


Figure 17.- Marginal effect of gender on mode choice

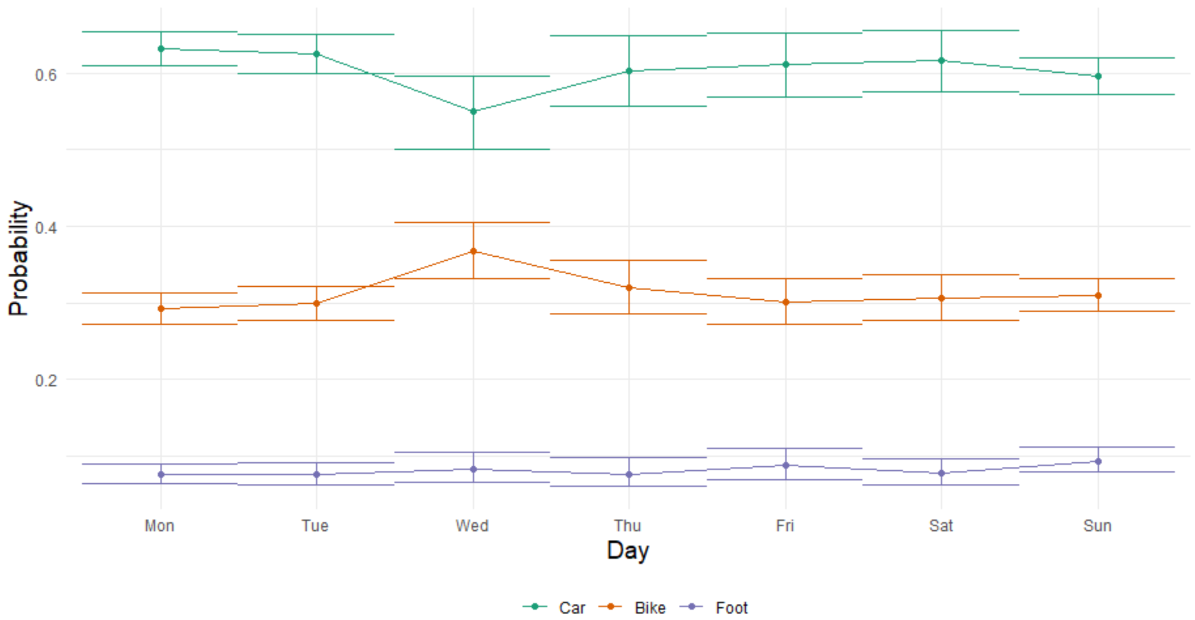


Figure 18.- Marginal effect of day on mode choice

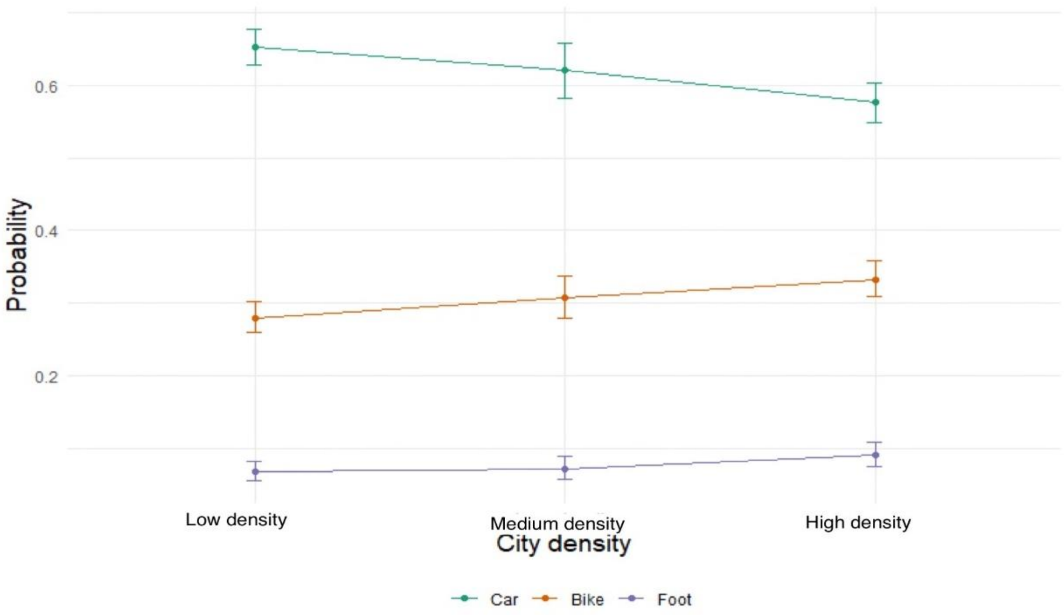


Figure 19.- Marginal effect of city density on mode choice

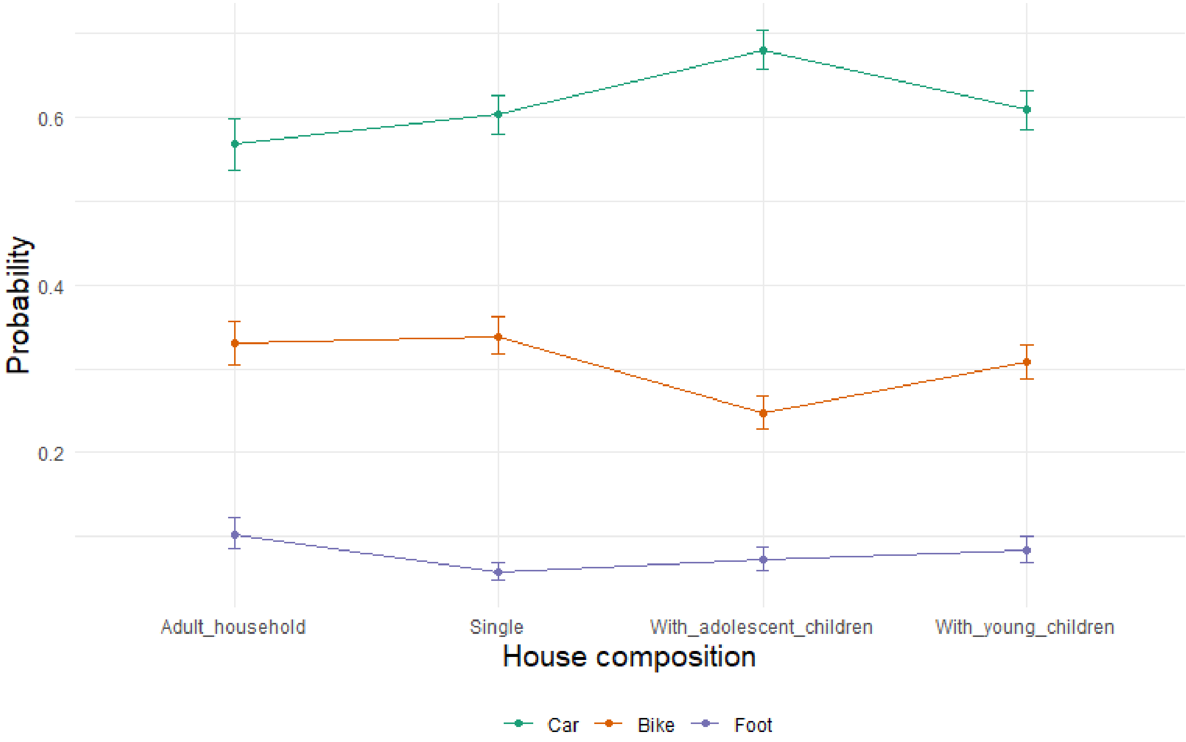


Figure 20.- Marginal effect of house composition on mode choice

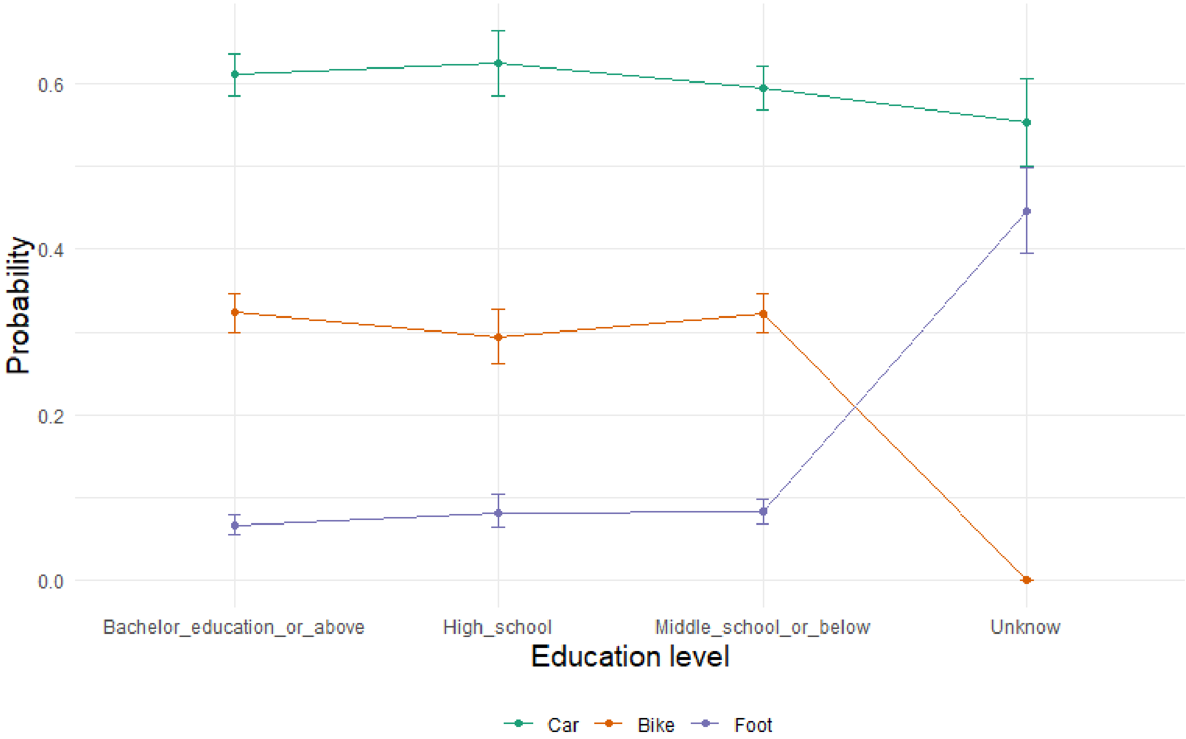


Figure 21.- Marginal effect of education level on mode choice

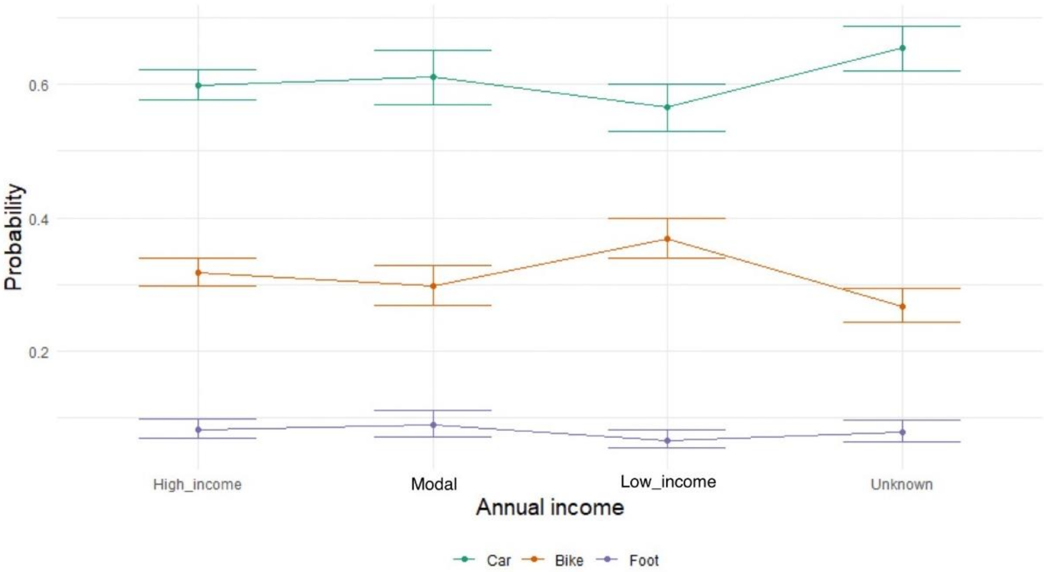


Figure 22.- Marginal effect of annual income on mode choice

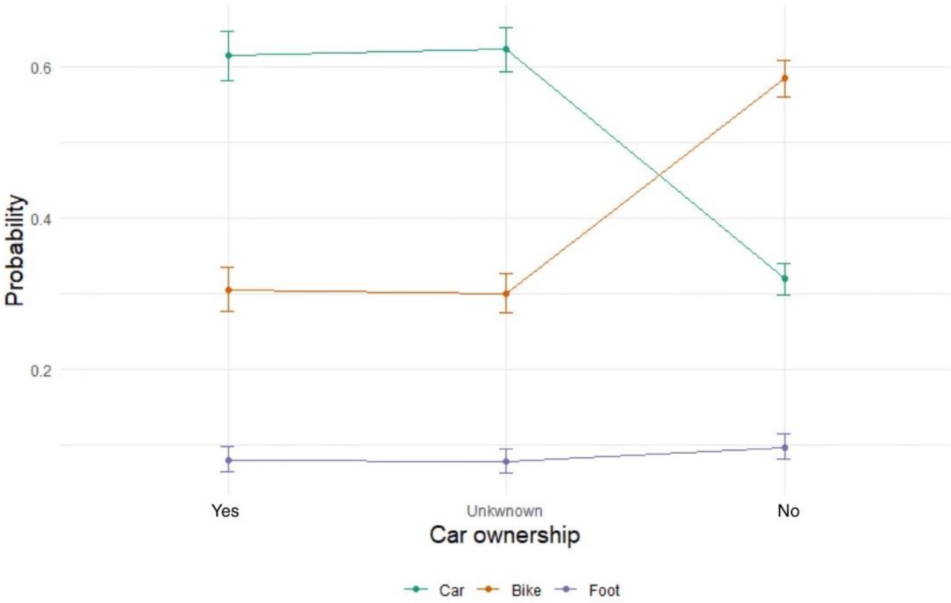


Figure 23.- Marginal effect of car ownership on mode choice