BSR lifestyles and mobility behaviour

UNIVERSITY OF TWENTE
FACULTY OF ENGINEERING TECHNOLOGY
CIVIL ENGINEERING AND MANAGEMENT

DAT.MOBILITY
GOUDAPPEL COFFENG

MUNICIPALITY OF ROTTERDAM

ELS BOSMA
27/01/2017

UNIVERSITEIT TWENTE.

dat mobility

Preface

A scientific research project is part of the study program of the Bachelor Civil Engineering and Management at the University of Twente. This thesis is submitted to fulfill that final part. I conducted this research at DAT.Mobility from November 2016 till January 2017.

I would like to thank my supervisors, Lissy La Paix Puello from the University of Twente for her statistical support and critical thinking. Luuk Brederode, who is a consultant at DAT.Mobility and my daily supervisor. For the general support, practical criticism and new ideas.

I liked to work at DAT.Mobility along with my fellow interns and colleagues. I enjoyed the lunches and coffees with cake, which proved a lot of moral support. I also want to thank my friends, family and flatmates for the encouragements and wonderful dinners.

Lastly, I would like to thank my contact from the municipality of Rotterdam, Will Clerx, without the initiative from Rotterdam, this research would never have taken place.

Els Bosma,
January 2017

Abstract

The Brand Strategy Research model (BSR model) has different uses for the municipality of Rotterdam. The statistical department raised the question whether lifestyles of the BSR model and mobility behaviour have correlations. This raised the idea to do a research in cooperation with DAT.Mobility. There are no datasets with both BSR-lifestyles and mobility behaviour. That is why the first step of this research is to explain the BSR lifestyles with quantitative measurable household indicators. This way, BSR-lifestyles can be added to datasets with mobility behaviour like OViN and can be evaluated further. This research concentrates on this first step.

A Multinomial Logistic Regression model (MLR model) is used to explain the BSR-lifestyles in household indicators. This model has different advantages, it does not make any assumptions about normality, linearity or homoscedasticity of the variables. It can also handle more than two categories of the dependent variable. In this case, the dependent variable is the BSR-model with four categories, its lifestyles. The main goal was to use the MLR model to describe the BSR model with household indicators.

A performance higher than 50%, based on the correct predicted lifestyles, was not possible. This has various reasons. Firstly, a very high performance, around 100%, would detract the psychographic classification of the BSR model. Also, people are never completely one lifestyle but have traits of all lifestyles, there is a distribution of lifestyles for every person. The lifestyle that is most present is assigned to that person, this is the dominant lifestyle. Sometimes, the dominant lifestyle is just for 40% present in data of a person. But the data we have, only contains the dominant lifestyles, not the distribution of lifestyles. That causes the fact that there can easily be made mistakes in predictions. Lastly, SmartAgent, the company that created the BSR model, uses a similar method to predict lifestyles. They accept a performance of 46%. This is a good indication for the maximum reachable performance of the model in this research.

At first sight, the performance of 47% from the MLR model seems limited, but it is still applicable in other datasets. In this case, the model is applied to a dataset with mobility data, OViN 2015. There is no elaboration on these results because it is not the aim of the research and also because of time constrains. But the first results are analysed and show some plausible correlations between mobility behaviour and lifestyles. For example the blue lifestyle, who wants to show its status, has clearly more vehicles in its household. Also, the red lifestyle, often young people who live a free life, make clearly more trips per day.

It is interesting to better analyse the relations between mobility indicators and lifestyles in future research. It is important to focus on the possibilities of different uses for BSR lifestyles in mobility models. Firstly, the lifestyles can be used as explanatory variables in models. They can add a new dimension to trip distributions. But because of a high number of insignificant factors while adding lifestyles to datasets and connecting them to mobility behaviour, the results will be unreliable. It is better to use the lifestyles as a way of communication with other fields, where BSR lifestyles are already used. For the municipality of Rotterdam, it is interesting to add BSR lifestyles as non-explanatory attributes to respondents. Policies can be adapted to the behaviour of the different groups. The lifestyles can be approached in an appropriate and appealing way, similar to the marketing strategies for which the BSR model was originally intended.

Samenvatting

Het Brand Strategy Research model (BSR-model) heeft voor de gemeente Rotterdam verschillende toepassingen. Vanuit de statistiek afdeling ontstond de vraag om te onderzoeken of er relaties bestaan tussen het BSR model met levensstijlen en mobiliteitsgedrag. Het idee ontstond om dit samen met DAT.Mobility te gaan onderzoeken. Omdat er geen datasets bestaan waarin BSR-levensstijlen en mobiliteitsgedrag samen voorkomen, is de eerste stap van dit onderzoek het verklaren van BSR-levensstijlen op basis van meetbare huishoudindicatoren, die in datasets met een structureel karakter voorkomen (MON/OViN/CBS). Hierdoor kunnen de BSR-levensstijlen worden toegevoegd aan datasets met mobiliteitsgedrag, die verder kunnen worden onderzocht. In mijn onderzoek heb ik mij op deze eerste stap geconcentreerd.

Om de BSR-levensstijlen in huishoudindicatoren uit te drukken is er gebruik gemaakt van een multinomiaal logistisch regressie model (MLR model). Dit model heeft verschillende voordelen, het doet geen aannamen over normaliteit, lineaire verdeeldheid of homogeniteit van variabelen, ook kan het rekening houden met meer dan 2 categorieën van de afhankelijke variabelen. In dit geval de BSR-levensstijlen, met 4 categorieën. Het hoofddoel is om het MLR model te gebruiken om het BSR model in huishoudindicatoren uit te drukken.

Al snel bleek dat een performance hoger dan 50%, gemeten in percentage goed voorspelde levensstijlen, niet mogelijk is. Dit heeft verschillende redenen. Ten eerste zou een hoge performance, tegen de 100%, afbreuk doen aan de psychografische segmentatie van het BSR model. Ook zijn mensen niet ingedeeld bij een levensstijl, maar verdeeld over verschillende levensstijlen, waarvan de levensstijl die het meest voorkomt aan de persoon wordt gegeven, dit is de dominante levensstijl. In de dataset die wordt gebruikt is alleen de dominante levensstijl bekend, niet de verdeling van levensstijlen. Ook wordt alleen de dominante levensstijl voorspeld. Soms is de dominante levensstijl maar voor 40% aanwezig. Waardoor er in de voorspellingen snel fouten worden gemaakt. Als laatste gebruikt SmartAgent, die het BSR model hebben ontwikkeld, zelf een vergelijkbare methode, waar zij akkoord gaan met een performance van 46%. Dit is een goede indicator voor de maximaal haalbare performance in dit onderzoek.

Het MLR model met een performance van 47% lijkt beperkt werkend, maar toch is het goed toepasbaar. Als het model wordt toegepast op een dataset met mobiliteitsgegevens, in dit geval OViN, blijkt het ook plausibele resultaten te geven. Omdat dit niet het doel van het onderzoek is en om tijdsredenen, zijn deze resultaten niet ver genoeg uitgewerkt om er met zekerheid conclusies uit te kunnen trekken. Maar de resultaten die er zijn, laten zien dat er logische verbanden aanwezig zijn tussen mobiliteitsgedrag en levensstijlen. Zo heeft de blauwe levensstijl, die graag status wil laten zien, duidelijk meer vervoermiddelen in het huishouden. Ook maakt de rode levensstijl, vaak jonge mensen die ongedwongen leven, duidelijk meer ritten per dag dan andere levensstijlen.

Het is interessant om in een vervolgonderzoek een betere analyse te doen van de relatie tussen mobiliteitsgedrag en BSR levensstijlen. Het is belangrijk om ook te focussen op de mogelijke toepassingen van de BSR levensstijlen in verkeersmodellen. Ten eerste kunnen de levensstijlen worden gebruikt als verklarende variabelen in modellen. Ze kunnen een nieuwe dimensie toevoegen aan de ritverdelingen. Maar er zijn veel insignificante factoren betrokken bij het toevoegen van levensstijlen aan datasets en het verbinden van deze levensstijlen en mobiliteitsgedrag. Dit zal deze resultaten en voorspellingen erg onbetrouwbaar maken. Het is beter om de levensstijlen te gebruiken als een manier om met andere vakgebieden, die al gebruik maken van het BSR model, te communiceren. Voor de gemeente Rotterdam is het interessant om de BSR levensstijlen aan respondenten toe te voegen als niet-verklarende variabelen. Op deze manier kunne

beleidsmaatregelen worden aangepast op de gedragingen van de verschillende groepen. Ook kunnen de levensstijlen op een passende manier worden aangesproken, op een manier die lijkt op de marketing strategieën waarvoor het BSR model oorspronkelijk bedoeld is.

Content

1. Introduction

When constructing policies for traffic plans in cities, policy makers and engineers rely on statistical data of cities. This data consists of the estimations of numbers of flows on links in passenger- or goods transportation systems. This way, the public transportation network or the public road network can be managed.

To create predictions of these trips, e.g. used when planning future neighbourhoods, only quantitative data and information about services of a neighbourhood are being used. This information can be age, income, vehicle ownership, the number of supermarkets etc.

This quantitative data does not always contribute to a reliable model of the future. These classical models assume that people only make rational choices about the fastest and cheapest way to travel. There may be other, more reliable ways to predict transportation models. This can include the personal preferences of people. Several studies showed that factors like social background (Bamberg, Hunecke, & Blöbaum, 2007) and social status (Scheiner & Holz-Rau, 2007) can influence the attitude towards different travel mode choices, which influence traffic flows.

The city of Rotterdam took the initiative to investigate if the estimation of behavioural traffic models can be constructed in another way. By examining the lifestyles of households in Rotterdam and comparing them to transportation data, there may be found relationships that predict better and more reliable data than the classical way of predicting. These predictions can include the preferences of people related to e.g. status, comfort or environmental awareness.

The Brand Strategy Research model (BSR model) is a lifestyle segmentation model and is used for multiple purposes. It was originally designed by the SmartAgent Company as a general way of customer segmentation (Lamme, 2003/2010). This could help to generate marketing strategies (Reitsma, 2012). But in Rotterdam, there is performed a research where the connection between the BSR model and the housing market is made (Smartagent, 2013). That raised the question if there is a connection between the lifestyles of the BSR model and mobility behaviour. The topic of connecting the BSR model with mobility behaviour is not completely new. In 2011 the province of Utrecht commissioned a study on the possibilities of car-sharing (Smartagent, 2011).

Another reason for the use of the BSR-model is the communication between different fields of the municipality. The municipality of Rotterdam uses the BSR model already in the fields of spatial planning and urban development. These fields work together a lot with the field of mobility. One way to make the communication between these fields easier is the ability to make them talk about the same groups of people. This makes communication a lot clearer and faster. (Goudappel Coffeng, 2012)

DAT.Mobility performs the investigation of lifestyle data in relation to transportation data. The model and information conducted from this research will be used by the municipality of Rotterdam to predict mobility behaviour in their city. The research in this bachelor thesis will focus on the relation between BSR lifestyles and household indicators.

1.1. Research Aim

The aim of the overall research is to connect the BSR lifestyles to mobility behaviour. Because there is no dataset with both mobility behaviour and the BSR lifestyles, the first step of this research is to explain the BSR lifestyles with quantitative measurable household indicators. Due to time constraints, I cannot do the whole research, so I concentrate on the first step. This results in the following research questions:

1. What are the characteristics of the different lifestyles in Rotterdam?

This first question is needed to provide information about the BSR lifestyles. It is important to get a general idea about the different lifestyles and get a feeling of how they complement each other. Also, the more general ideas about the BSR model need to be investigated. The answer to this research question can be found in chapter 2.2.2.

2. How can, according to literature, lifestyles of households be related to quantitative household indicators?

This questions relates strongly to the first question but needs a deeper investigation into the aspects of lifestyles. Also, the databases need to be analysed to know which variables can be used. The explanation of this research question can be found in the final part of chapter 2 and chapter 3. There are made choices regarding the used models and methodology.

3. Is there a relationship between lifestyles and quantitative indicators?

This research question is the most important research question and takes the most time to answer. To create predictions about lifestyles, it is important to know which indicators relate to certain lifestyles. Therefore, every individual indicator is analysed and there are made single variable models. Lastly, there is made a final model, where all variables are combined. The variables that are used need to be present in OViN and CBS datasets. This makes implementation of the model easier and creates the possibility to apply the BSR model nationwide. The answer to this question can be found in chapters 3 and 4.

4. Is there a relationship between lifestyles and mobility behaviour?

This question is asked to give an early prediction of the relationship between lifestyles and mobility behaviour. It can only be answered when there is enough time left while doing this research. The aim is to make the follow-up research easier and to investigate if the follow-up research would be meaningful and necessary. It will be used to give an advice about the next steps in the research. There can also be given advice about possibilities of applying the lifestyles to transportation models. There will be made a distinction between using the lifestyles as explanatory variables and using lifestyles to label model results to make communication easier. Existing literature shows that the attitude of people towards mobility is just as important as traditional variables. When there is found a significant relationship between mobility and BSR lifestyles, there is created a way to interpret these attitudes.

5. Which information is needed to complete the datasets that are going to be used in the next part of the research?

In the next phase of the research, predictions are made regarding the relationship between lifestyles and mobility behaviour. When question four is difficult to answer, there needs to be done an investigation about how this can be answered in the future and which data is needed for that.

1.2. Relevance

Traffic flow simulation is important when managing problems in the fields of urban planning and traffic prediction in the urban area. Travel mode choice models are an essential part of estimating traffic flows. These models mostly rely on factors like time value and travel costs. Several studies showed that attitude is at least as important as these measurable factors. Social context and thereby economic and social-cultural background influence the attitude towards public transportation (Bamberg, Hunecke, & Blöbaum, 2007). Also social status, age and family phase relate to the choice of travel mode (Scheiner & Holz-Rau, 2007). For elderly, the amount of travel and car use is influenced by sociodemographic indicators (Hildebrand, 2003), also their feelings of car dependency are influenced by other factors than only their age (Benes, 2015).

Different lifestyle models and indicators are used in these studies. In this research, the Brand Strategy Research (BSR) model is used, which is already used for managing the housing market in the Netherlands (Hammouch, 2007) and in a range of other fields (van Hattum, 2010). The municipality of Rotterdam uses the BSR model in their housing market and as a way to communicate between different disciplines in the municipality. In earlier studies, there are found general relations between the lifestyles of the BSR model and travel mode choice (de Graaf, 2011) or shared car use (Smartagent, 2011). This multi-functionality of the BSR model is a good reason to assume that the lifestyles of the model relate to a certain kind of mobility behaviour. The BSR model can also make great distinctions in terms of social status and attitude. If the fourth research question can be answered, the results are expected to be similar to the studies described above.

2. Methodology

2.1. Datasets

To clarify the following chapter, the used datasets will be summarized first. In this research, there are used three data sources. More information about the variables is provided in chapter 3.

The first source that is used is *De Grote Woontest* (DGW) (The Big Housing Survey). This survey is conducted in 2004, 2008 and 2012 in Rotterdam. The datasets from 2008 and 2012 are used in this research. These datasets contain measurable household variables (like age, income), neighbourhoods and lifestyles of individual respondents living in the region of Rotterdam. The dataset from 2008 consists of 16500 respondents, 2012 consists of 18000 respondents, but both datasets contain nearly 3000 respondents where a lot of the variables are unknown. Because the variables in both datasets are not exactly the same, the datasets cannot be merged instantly. *Matlab* is used to merge the datasets and create one consistent dataset where every variable has the same characteristics. This results in a usable dataset with 30.000 respondents.

An important variable in the DGW dataset is the BSR lifestyles, this is the dependent variable. For every respondent, there is a distribution of all different lifestyles. People are never 100% one lifestyle but have traits of all different lifestyles. Only the dominant lifestyle is present in the dataset from DGW, not the distributions of lifestyles. This causes a lot of uncertainty in the final model and makes it harder to make predictions.

The dataset that contains data from DGW is completed with data from Centraal Bureau Statistiek (CBS) (Statistics Netherlands). This data contains address and population densities and mean distances to services. The data is also merged making use of *Matlab*. Because the CBS data is on neighbourhood level, the variable neighbourhood is used to merge the datasets. CBS neighbourhood data is used because it covers the whole population of the Netherlands. When applying the model nationwide, it is important to have easily accessible information about all households in the Netherlands.

Lastly, the *Onderzoek Verplaatsingen in Nederland* (OViN) (Research Mobility in the Netherlands) is used for two purposes. To make sure that the variables used from DGW are present in this dataset and to get a first indication of the relation between lifestyles and mobility behavior. This dataset contains data from respondents living in all regions of the Netherlands with their measurable household indicators, vehicle ownership and trips made in one day. It covers the whole population of the Netherlands, just like CBS. It is important to use variables that are present in OViN and CBS. That creates the possibility to apply the BSR model nationwide.

2.2. Models

Different models are present in this research. Firstly, the conceptual models will be presented. After that, the Brand Strategy Research model with its lifestyles will be introduced. Lastly, the mathematical multinomial logistic regression model will be presented.

2.2.1. Conceptual models

Figure 1 visualizes the data flow in this research. Firstly, the different datasets from 'De Grote Woontests' are combined with data from CBS. After that, there is made a multinomial logistic regression model (MLR model) that predicts the BSR lifestyles of respondents. The literature and data studies that are conducted to answer the first two research questions are used to maintain these processes. The MLR model is made to answer the third research question about the correlation between household indicators and the BSR model. Lastly, there is made a start with the implementation of the MLR in a dataset with mobility information, in this case, a dataset from OViN. This is done to answer the last two research questions about the correlations between mobility behaviour and BSR lifestyles and about the possible applications of this.

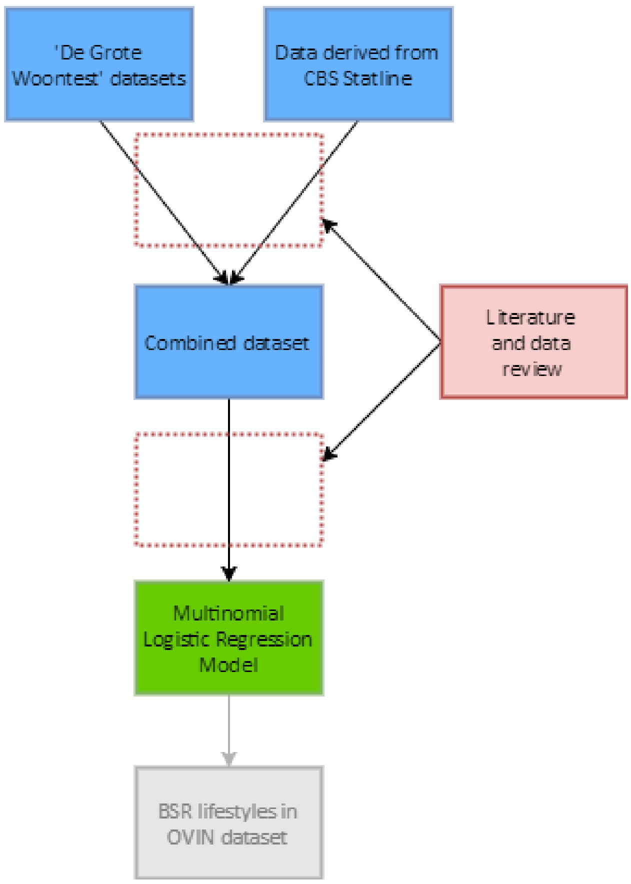

Figure 1: Information flow model

The data fusion model can be seen as the final part of the information flow model. When conduction data fusion, two datasets are fused, dataset A and dataset B. Dataset A includes information X. Dataset B also includes information X, but also includes information Y. There is a relationship between information X and Y. The data fusion model fuses these two models by extracting the relationship between data X and Y. It uses X to describe Y. That way, information Y can be added to dataset A. This dataset can be called dataset A^+. In this case, dataset A is the dataset with mobility behaviour information in it, so the OViN dataset. Dataset B is the combined dataset. Dataset A^+ will be the OViN data where the lifestyles of respondents are added.

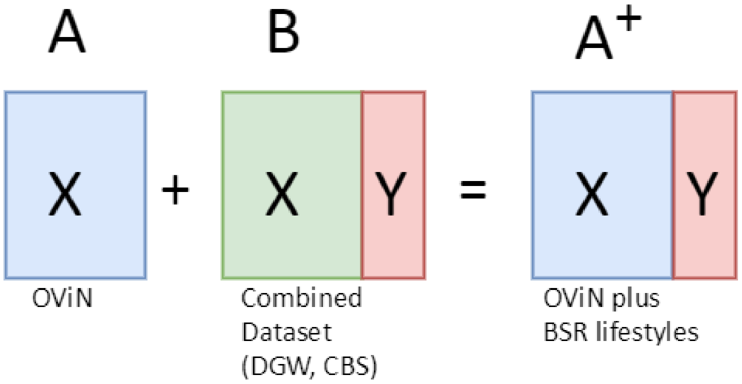

Figure 2: Data fusion model (Hattum & Hoijtink, 2008)

This data fusion model will be created with the use of a multinomial logistic regression model.

2.2.2. Brand Strategy Research Model and lifestyles

The Brand Strategy Research model (BSR-model) describes the four lifestyles that are evaluated in this research. The BSR model is developed by The SmartAgent Company. A short description and interpretation of the model and the lifestyles are needed to clarify and justify the assumptions about the significance of variables in the model.

The BSR-model is developed by the SmartAgent company (Lamme, 2003/2010). The model classifies people based on their values and character traits. The basis of the BSR model is a grid with four quadrants, which are the four lifestyles. The lifestyles differ from each other in two dimensions. The first dimension, on the x-axis, distinguishes sociological elements, how someone reacts to other people, someone can be ego oriented or group oriented. The second dimension, on the y-axis, distinguishes psychological elements, which are extrovert or introvert people (van Hattum, 2010).

An example of the BSR model can be found in Figure 3. A short description of the four lifestyles (SmartAgent, 2012):

» The yellow lifestyle represents harmony. People with a yellow lifestyle are spontaneous and very social, they are extravert and group-oriented. It is easy for them to make new friends. Besides a big social network, family and the neighbourhood are a central point in their lives.
» The green lifestyle means security. People with a green lifestyle are group-oriented but also introvert. Their family, neighbourhood and their privacy are all important for them. They don't like changes and want everything as normal as can be.
» The blue lifestyle means control over life. People with a blue lifestyle are introvert and ego-oriented. They want to have a high social status and a luxury lifestyle. They make strategic decisions about their career to be successful.
» The red lifestyle means vitality. People with a red lifestyle are ego-oriented and extravert. They are free-spirits and want a lot of freedom in their lives. Enjoying life, besides work and family, is for them the most important thing.

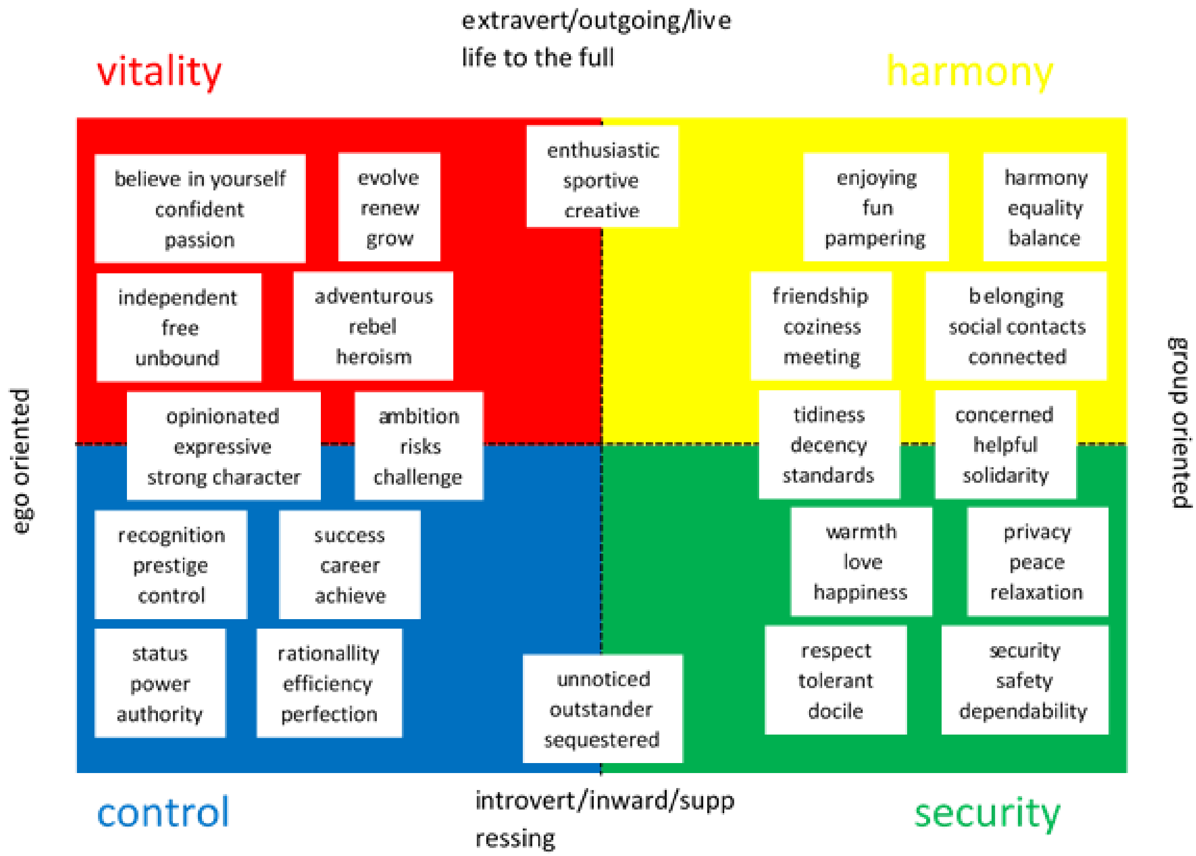

Figure 3: BSR-model (van Hattum, 2010)

If you want to know in which lifestyle category you would be, you can take a look at the website of SmartAgent (Smartagent, 2013).

2.2.3. Multinomial logistic regression model

The multinomial logistic regression model (or multinomial logit model) is a regression model that can handle more than two categories of the dependent or outcome variable. The multinomial logistic regression model is used because it does not make assumptions about linearity, normality or homoscedasticity of the variables (Starkweather & Moske, 2011). This is needed because the used variables are nominal, not ordinal.

The model cannot assume linearity because the first or last group of a variable does not indicate the first or last lifestyle (yellow or red). For example family phase, the first group is young and alone, which indicates a red lifestyle. The last groups, older, indicate the green lifestyle, not yellow (as expected when variable would be linear). The middle group, family, indicates the blue and yellow lifestyles.

The model also cannot make assumptions about normality. If the data would be normal distributed, it would mean that every lifestyle and every value of a variable would be distributed uniformly over the population, this is not the case.

The model cannot make assumptions about homoscedasticity because the dataset does not have an equal variance of all the data. Sometimes, one value of a variable predicts lifestyles much better than other values, some values do not have a predictive application at all. There are also great differences between variables.

Also, the model assumes non-perfect separation of the outcome variable (Starkweather & Moske, 2011). This is needed because people never belong for 100% in one lifestyle, but have always traits of different lifestyles. Also, people with similar household indicators can have different lifestyles. One value of a variable does not indicate a lifestyle but only gives a higher probability for a lifestyle.

A multinomial logistic regression model is a model that predicts the odds between probabilities of a reference category and other categories of the dependent variable. The dependent variable is the variable that is predicted, in this case the BSR-lifestyles. The odds of the reference variable are always 1. This way the predicted probabilities can be calculated easily. See the following example, where the person get assigned the blue lifestyle (Table 1):

Lifestyles		**Odds**	**Probability**
Yellow		1	1/(1+0,5+2+1,5) = 0,2
Yellow	Green	1 : 0,5	0,1
Yellow	Blue	1 : 2	0,4
Yellow	Red	1 : 1,5	0,3

Table 1: Odds example for an imaginary case, person would have blue lifestyle

The odds of a respondent are calculated with corresponding Betas. The odds are then: $O(Y=c \mid X=x) = EXP(\beta^{(c)}_0 + x\beta^{(c)}_1 + x\beta^{(c)}_2 + \ldots + x\beta^{(c)}_p)$. Where Y is the dependent variable with category c. X is the observation, with the set of values of variables, p is the number of variables. The first Beta, $\beta^{(c)}_0$, corresponds only with the category, not with the variables.

The Betas are calculated by SPSS by using an optimization method, where the Log Likelihood and rho-square are optimized. Also, the standard error and the corresponding 95% confidence interval of the β and EXP(β) are calculated.

Rho square is a measurement for how well a regression model can predict the variance of data, or how well the regression model fits the data. A perfect model will have a rho-square value of 1. But the data in this dataset has a very high variance and a lot of outliers. That is why the rho-square cannot become high, but it can still be optimized.

The Log Likelihood (LL) is a measurement of the difference between the real probabilities and the predicted probabilities for a certain outcome. The LL should be as close to zero as possible, but can never be completely zero. There can also be made a comparison between the LL of the null-model and the LL of the final model. The difference needs to be as high as possible.

Also the Chi-square and significance explain something about the difference between the final model and the null-model. The Chi-square gets bigger when there is a bigger difference between the performance of the null-model and the performance of the final model. So the Chi-square needs to be as big as possible. The significance is calculated with Chi-square as input, where the significance needs to be as low as possible. To be 95% sure that the final model is different than the null model, the significance needs to be 0,05 or lower. With this number, there cannot be made assumptions about how significant the model itself is. It only tells that the model is significantly different from the null model.

2.3. Approach

The construction of the multinomial regression model consists of a few phases. First, the different datasets are merged and the variables are converted, so that they can be used and are consistent with the OViN dataset. After that, variables are analysed to make a selection of variables to be tested for inclusion into the final model. The selected variables are used in the final model. The last phase is the creation of the multinomial logistic regression model itself.

2.3.1. Merging of datasets

In the first phase, datasets from different sources (DGW, CBS and OViN) are merged and made consistent. The datasets from DGW 2008 and DGW 2012 are merged, also data from CBS is added. The variables are made consistent with the OViN dataset. That way, the created model can be applied to the OViN dataset and other datasets with mobility indicators. Information about these datasets can be found in paragraph 2.1. Most difficult challenges in merging the datasets are the consistency of variables. The datasets from DGW2008 and DGW2012 do not use the same variables. Variables from both datasets are converted to variables that are also present in OViN. CBS data is added using the zip codes or neighbourhood codes of respondents. This data includes the degree of urbanism and distances to services.

2.3.2. Method for selection of variables

Variables are analysed to make a selection for the final model. This is needed to prevent overfitting of the model and to limit the number of possible combinations to test. The variables that have no meaning are filtered out. This makes the model more reliable and better applicable at other datasets.

The variables are tested in two ways. The significance of the variables are tested and the multinomial logistic regression models of the individual variables are tested on their performance. More information about the acceptable performance of variables can be found in chapter 3.

2.3.3. Multinomial logistic regression analysis

The multinomial logistic regression model is constructed using SPSS software. This software calculates β's for every variable for every lifestyle. These β's indicate how the variable attached to this beta is related to the probability of a person having a certain lifestyle. The multinomial logistic regression model compares three lifestyles to one control lifestyle, normally the yellow lifestyle. For the model, it makes no difference which lifestyle is chosen to be the control lifestyle. A β below zero for a value for a certain lifestyle indicates that the likelihood of that lifestyle, for a person with the value, is lower than the control lifestyle. A β above 0 indicates that that probability is higher.

Another thing to consider is the interaction between variables. The different variables used cannot be seen completely apart from each other. They influence each other and can have different meanings depending on the values of other variables. But taking the nature of a multinomial logistic regression model and the calculation times into account, a full factorial analysis cannot be performed.

In this study, there will be an acceptable performance rate of 46% for the final model. The next paragraph explains more about the performance rate.

2.3.4. Acceptable performance rate of the multinomial logistic regression model

This paragraph elaborates on the acceptance rate of the performance of the final model.

The BSR model divides people into four groups; Yellow 'Harmony', Green 'Security', Blue 'Control' and Red 'Vitality'. The lifestyle of a person is determined by their answers to questions like: 'Which character traits fit you best?', 'Which occupations, regardless of your own work situation, fit you best?' and 'Which values fit you best?' This results in a lifestyle that fits that person best. But that does not mean that that person doesn't have any traits of other lifestyles, in fact, the person does have traits of (nearly) all lifestyles. A visualization of the model can be found in Figure 4.The dominant lifestyle is on average 60% present in the lifestyle distribution of a person (van Hattum, 2010). Only the dominant lifestyle is known in the original DGW dataset. It would be better to use the distribution of lifestyles and predict this distribution. The model only predicts the dominant lifestyle. This means that a model that predicts these lifestyles based on other variables can never have a better performance than 60%.

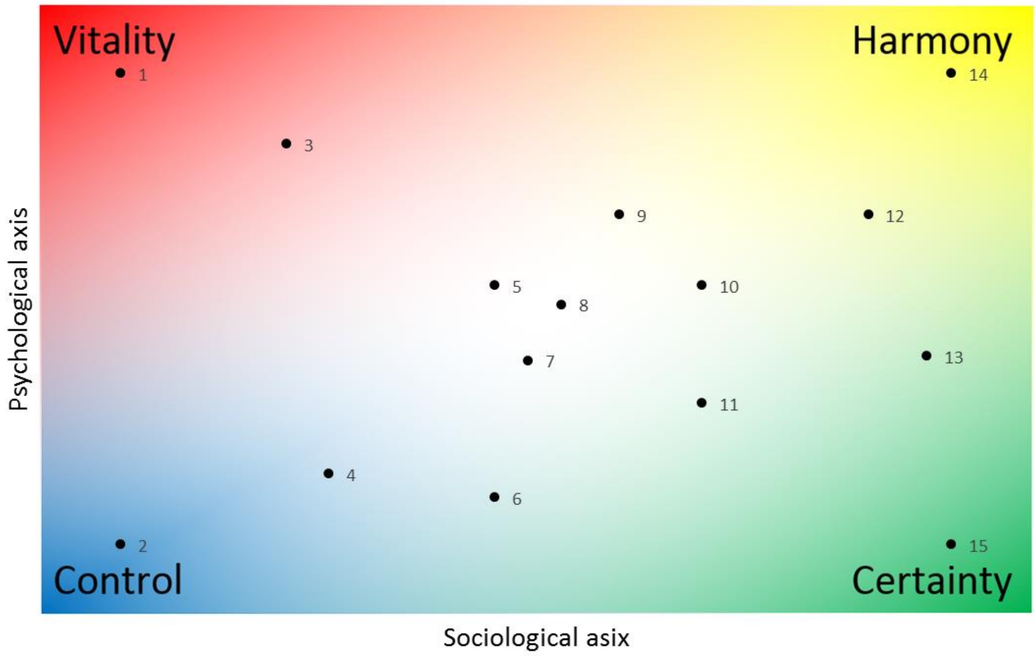

Figure 4: Visualization of the BSR model. The four lifestyles cannot be visualized as four separate clusters (happens normally), but the can be seen as lifestyle gradients. Only from respondents 1, 2, 14 and 15 can the lifestyle be determined clearly. Also from respondents 3, 4, 12 and 13 the lifestyles can be determined. The lifestyles from the other respondents, especially 5, 7 and 8, cannot be determined.

Also, lifestyles are based on a psychographic classification. In this research, there is developed a model that describes the relation between measurable household variables and the classification. A high percentage of correct predictions would detract this psychographic classification.

The two arguments above conclude that the model conducted in this research has a lower probability than 60%. To decide whether the prediction rate is acceptable, there is made a comparison with a null-hypothesis. The null-hypothesis is a uniform distribution of the lifestyles, where the distribution of lifestyles in the Rotterdam is taken into account (Yellow: 31,5%; Green: 29,0%; Blue: 20,7%; Red: 18,8%). Because of that, the null-hypothesis is a model that predicts 26% of the cases correct. Considering all these arguments, a performance of 45% for the final model seems correct, this model would predict nearly twice as much lifestyles correct than the null-model.

The SmartAgent Company itself uses a comparable methodology for data fusion. They use measurable household variables (which are derived from CBS) to predict lifestyles in neighbourhoods. They agree with a performance of 45% (van Hattum, 2010). That makes the minimum performance rate of 46% that is used in this research acceptable.

Besides the performance rate, it also needs to be significantly different from the Null-model and needs to have a lower Log Likelihood than the initial model.

3. Measurable household variables and the BSR lifestyle model

In this chapter, all variables are presented, then measurements of variables are enclosed. The last evaluation consists of the presentation of individual variable models (which are multinomial logistic regression models). Lastly, the final model that describes the relation between BSR lifestyles and mobility behaviour is presented.

3.1. Variable overview

In this paragraph, all the variables in the dataset that will be used to make the multinomial regression model will be summarized. There will also be made a choice regarding the variables that will be included and excluded in further analysis.

Variable	Explanation	Data source
Gender	Man / Woman	DGW 2008/2012
Age_Recoded	Age divided into groups.	DGW 2008/2012
Summarized family phase	Young/old, alone/together or family	DGW 2008/2012
Extensive family phase	Age group, alone/together/family	DGW 2008/2012, merged information from other variables
Household income	Measured in the degree of modality for the corresponding year.	DGW 2008/2012, present in different forms, this nominal scale is used
Type of housing	Different types of housing	DGW 2008/2012
House owning	Buying or renting the house	DGW 2008/2012
Education	From primary school to university	DGW 2008/2012
Employment situation	Includes unemployment/retirement/ student/homemaker	DGW 2008/2012
Land of birth	Land of birth of respondent	DGW 2008/2012
GMcode	Code of the municipality from the respondent. Codes are used to merge data from CBS and 'De Grote Woontest'.	DGW 2008/2012
WKcode	Code of the district from the respondent. Codes are used to merge data from CBS and 'De Grote Woontest'.	DGW 2008/2012
BUcode	Code of the neighbourhood from the respondent. Codes are used to merge data from CBS and 'De Grote Woontest'	DGW 2008/2012
BSR LIFESTYLE	Lifestyle of the respondent	DGW 2008/2012
Car	Number of cars present in the household of the respondent	DGW 2008
Distance Main Road	Mean distance in neighbourhood of respondent to main road.	CBS
Distance train station	Mean distance in neighbourhood of respondent to train station.	CBS
Population Density	Mean population density in neighbourhood of respondent.	CBS
Degree of urbanism	Mean address density in the neighbourhood of the respondent.	CBS

Table 2: Variables for multinomial logistic regression

Because of time constraints and the presence of variables in the OVIN datasets (mobility datasets), not all of the variables can be evaluated. Choices regarding the variables that are included and excluded are made based upon their relevance and the presence in OVIN. The following variables will be excluded because of lack of presence in OVIN datasets:

» Type of housing
» Education partner
» Employment partner
» Land of birth father
» Land of birth mother

The following variables will be excluded because of lack of relevance, this means that the variables are already present in another variable, that variables are too much coherent with other variables that are already present, or that there is a lack of coherence with the BSR lifestyle model.

» Age (as a number)
» People per household
» Origin
» Generation
» Distance Market
» Distance School
» Distance swimming pool
» Population Density

The following variables will be included in the used dataset but excluded from evaluation. These variables are used to merge data from CBS and the DGW.

» GMcode
» WKcode
» BUcode

Confirming to the excluded variables, the next variables are included in the dataset and evaluation:

» BSR lifestyle → this is the depended variable in the analysis.
» Gender
» Age_Recoded
» Summarized family phase
» Extensive family phase
» Income
» Employment
» Education
» Car ownership
» Land of birth
» Degree of urbanism
» Distance to main road
» Distance to train station

3.2. Minimal performance of individual variables

The variables are tested in two ways. The significance of the variables is tested and the multinomial logistic regression models of the individual variables are tested on their performance.

The significance of a variable is measured in the likelihood ratio test, all variables are included in this test. The variable is compared to a null model, which indicates no relationship between the variable and BSR-lifestyles. When the significance is below 0,05, there is a significant relation between the variable and BSR-lifestyles, with a 95% probability level. When there is no significant relation between the variables and BSR-lifestyles, the variable won't be included in the final model. A significance test is performed for every individual variable. The variances are measured in percentages per value per lifestyle (so that the percentage of one lifestyle of all the values is 100%) the significant differences between lifestyles are found with confidence intervals. these tables can be found in the appendix. The histograms with distributions of variables over lifestyles are presented in the next paragraph.

The performance of the individual variable models will be evaluated in two ways. Firstly the model fit, where the chi-square is calculated, when the p-value is beneath 0.01, the model is fitted. Secondly, the classification table of the model is evaluated, where the observed and predicted frequencies are summarized. It also contains the overall correct predicted percentage. If a variable is not predictive, its model will predict that all respondents are in the yellow lifestyle, because the most respondents of DGW are in that lifestyle. This model would have an overall correct predicted percentage of 31,4%. A model where the respondents are equally distributed over the four lifestyles, where their presence is taken into account, has a correct predicted percentage of 26,1%. The minimum for a variable to be included in the final model is set at 32%. This percentage correctly predicted observations is the most important performance indicator.

3.3. Variance and numbers of values in variables

In this paragraph, the variances of variables and lifestyles are summarized. All bar graphs with descriptions can be found in appendix A (page 33). Cross tables, where graphs are based upon, can be found in appendix C (page 40). In this chapter, two variables are illustrated as examples. After that, there is given a summary of all evaluated variables.

3.3.1. Age

The mean age of the respondents is between 35 and 55 years. The effect of age on the lifestyles can mainly be seen in the red and the green lifestyles. The age of the green lifestyle is significantly older and the red lifestyle is significantly younger than the other lifestyles (Figure 5).

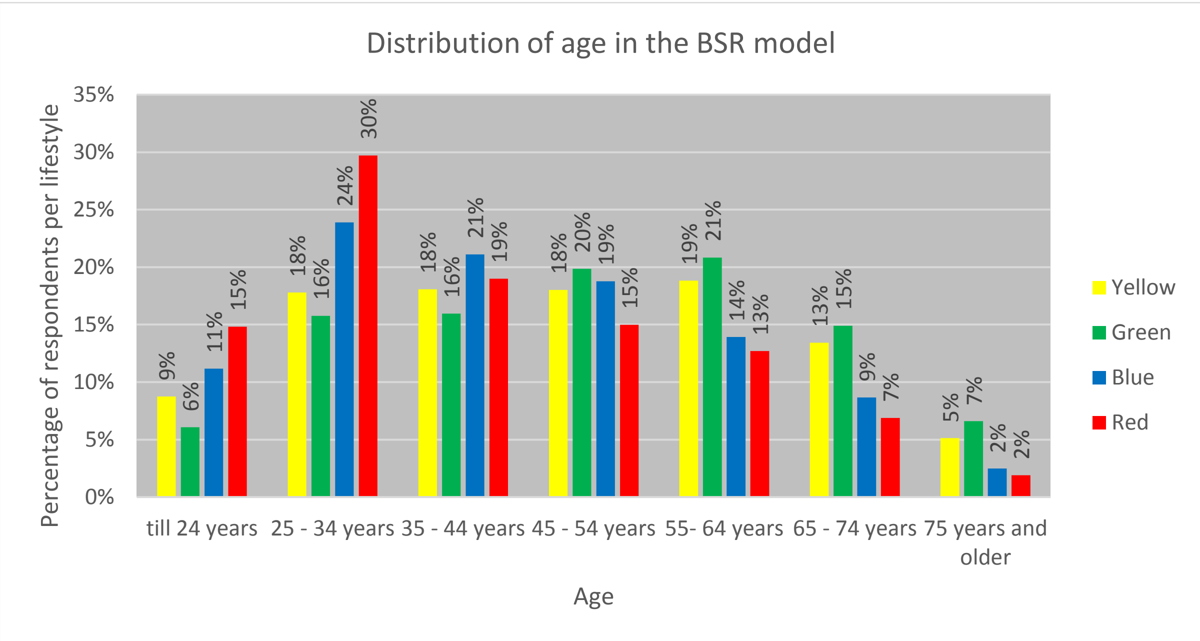

Figure 5: Distribution of age over the BSR lifestyles

3.3.2. Income

The income of the population has mainly effect on the yellow and blue lifestyles. The yellow lifestyle has less income and the blue lifestyle has more income than the other lifestyles. Most significant is the difference at twice the modal income. Another significant difference can be found in the red lifestyle. A group of people does not want to share their income, in the red lifestyle, this group is the smallest (Figure 6).

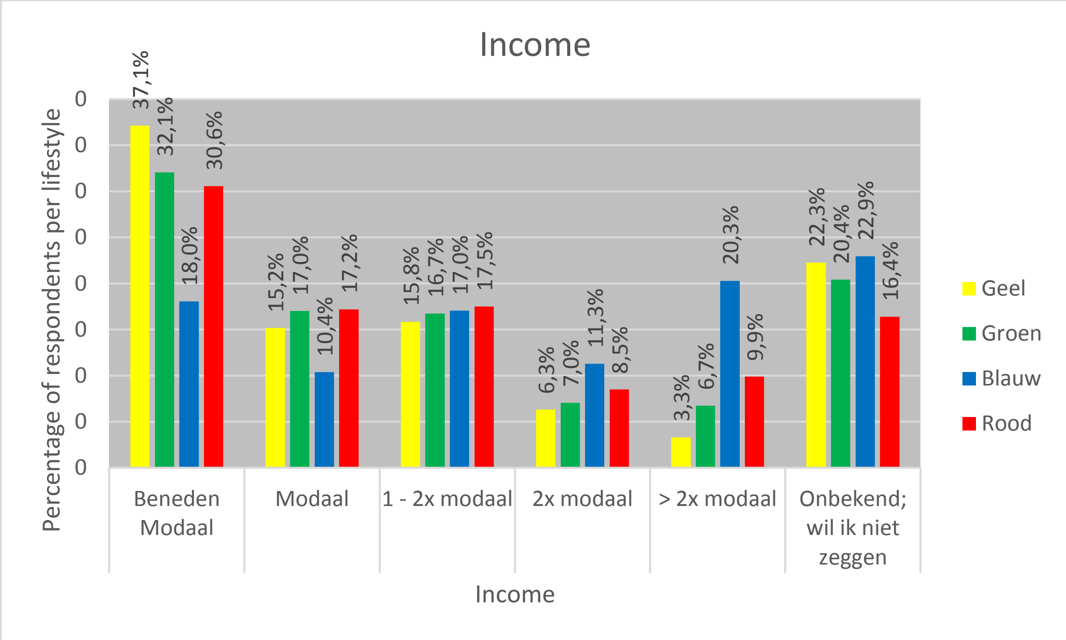

Figure 6: Distribution of income over the BSR lifestyles

3.3.3. Summary of all evaluated variables

Above, there can be seen that distributions of age and income are different for different lifestyles. Especially the higher and lower incomes and ages show significant differences between lifestyles. In all variables, there can be found significant differences between lifestyles.

There was expected that the distribution of genders over the lifestyles would be equal. But the yellow lifestyle consists of clearly more woman than man. In both family phases, there can be seen that the red lifestyle is definitely younger than other lifestyles. They are also more often alone and without a family. The yellow and blue lifestyles have a lot more families than other lifestyles.

In the employment status can be seen that people with a yellow lifestyle are more often housewives and less often entrepreneurs. A green lifestyle is more likely to be retired, a blue lifestyle is less likely to have no job and a red lifestyle is more likely to be a student.

The education of a respondent has also a big influence on its lifestyle. The yellow lifestyle is the most significantly different from the other lifestyles and is less educated. The red and blue lifestyles are more educated than the other lifestyles, where red has the highest difference with other lifestyles.

Measurements for car ownership are only made in the ‘De Grote Woontest’ survey of 2008. That is why this variable is less reliable than other variables. Households with a red lifestyle have fewer cars than other lifestyles. Households with a blue lifestyle have more cars than the average household.

For the land of birth, there cannot be found a significant difference between lifestyles. Also for distances to train stations and to main roads, there cannot be found significant differences. The degree of urbanism has a significant difference for the red lifestyle, this lifestyle live much more often in a high-density area (like the city centre) than other lifestyles.

3.4. Performance of single variable models

Several things are taken into account when choosing variables for the definitive Multinomial Logistic Regression model. Firstly, there has to be a significant relationship between the evaluated variable and the dependent variable (BSR lifestyle). Secondly, the performance of individual variable models consisting of only one variable and the dependent variable will be measured. The variables will be evaluated with the method described in paragraph 3.2.

All bar graphs of evaluated variables with descriptions can be found in appendix B (40). The tables where these graphs are based upon, can be found in appendix C (46). In this chapter, two variables are illustrated below as an example. After that, there is given a summary of all evaluated variables.

3.4.1. Age

The model made with only the variable age performs for **32,4%.** This performance is low compared to other variables. But this model has a few strong points; this model makes a difference between the yellow and green lifestyles. It also includes the red lifestyle in the prediction of lifestyles. The blue lifestyle is excluded because there the blue lifestyle does not have an age group where it is significantly more present (as seen above).

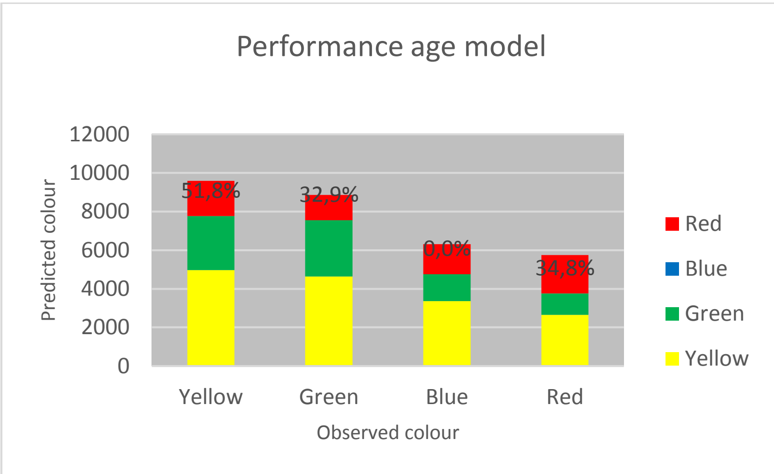

Figure 7: Performance of age model

3.4.2. Income

The model that is conducted with this variable performs for **35,6%**, this is average compared to the other variables. However, relates to the yellow and blue lifestyles. This shows that the other lifestyles don't have a significant high or low income.

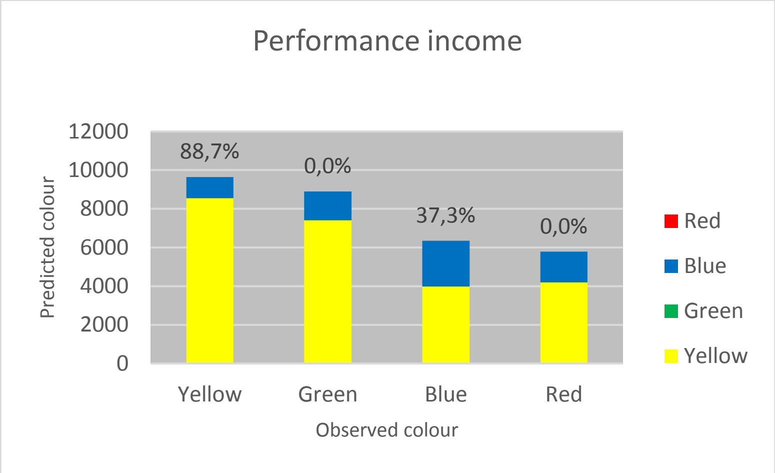

Figure 8: Performance income

3.4.3. Summary of all evaluated variables

Most single variable models can make a distinction between two or three lifestyles. The employment situation can even make a distinction between all four lifestyles. The performance of the single variable models of land of birth and distance to main road or train station perform very badly. They only predict the yellow lifestyle for all respondents. This is due to the fact that there are no significant differences between lifestyles. The single variable model of education performs best, with 38,2%.

3.5. Selection of variables for final model

Selection criteria are used to select variables for the final model. The fist selection criterion is the performance of the individual model of every variable. A null-hypothesis model predicts that every respondent is in the yellow lifestyle, that makes the correct predicted percentage 31,7% (if all respondents are taken into account). The minimum correct predicted percentage for a variable to be included in the final model is 32%. That way, the following variables are excluded:

- Land of birth
- Mean distance to main road
- Mean distance to train station

Also, extended family phase is excluded from the model. This is due to the similarity with summarized family phase, extended family phase does not have a better performance. Lastly, also car ownership is excluded in the final model. This because this variable is only included in half of the dataset (only data from 2008).

4. Multinomial Logistic Regression Analysis, final model

4.1. Conducting the final model

The selected variables are added to a multinomial logistic regression model one by one, until the performance indicators are optimized and do not significantly change anymore. Also, the variables education and employment status were transformed slightly to fit better and to be more easily applicable to the OViN dataset. Parameter estimates of the model can be found in Appendix D (page 59), there will be elaborated on them in the next paragraph. Model fitting information can be found in Appendix E (page 62). The model fitting information shows that the model is significantly different from the null model. Because of the big dataset, the Log-likelihood and Chi-square values are very high. The rho-squared does not show an optimal result, but cannot be optimized further with the current variables. This is due to the high variance and the high number of variables in the multinomial logistic regression model.

The final model, made with the selected variables has a performance of 46,7%. The biggest predicted value for every category is the correct, observed category. The figure below shows significant differences between the correct predicted values and incorrect predicted values. The only uncertainty can be found between the Green 'Security' lifestyle and the Yellow 'Harmony' lifestyle. This is due to the similarities between the green and yellow lifestyles and due to the fact that the green lifestyle is most difficult to define and looks most like the yellow lifestyle.

The performance of the final model could be improved by adding interactions between variables, recoding of variables, merging of alternatives or merging of classes of variables. But these additions will not radically improve the performance of this model. A few exercises showed a maximum improvement of just a few percent. Also, merging of alternatives (lifestyles) will detract the usability of the BSR model. There is chosen to not do this improvement because it will be a lot of effort for small improvements. Also, not all variables can be recoded into perfect functioning variables, because they still need to be applicable to OViN and CBS datasets.

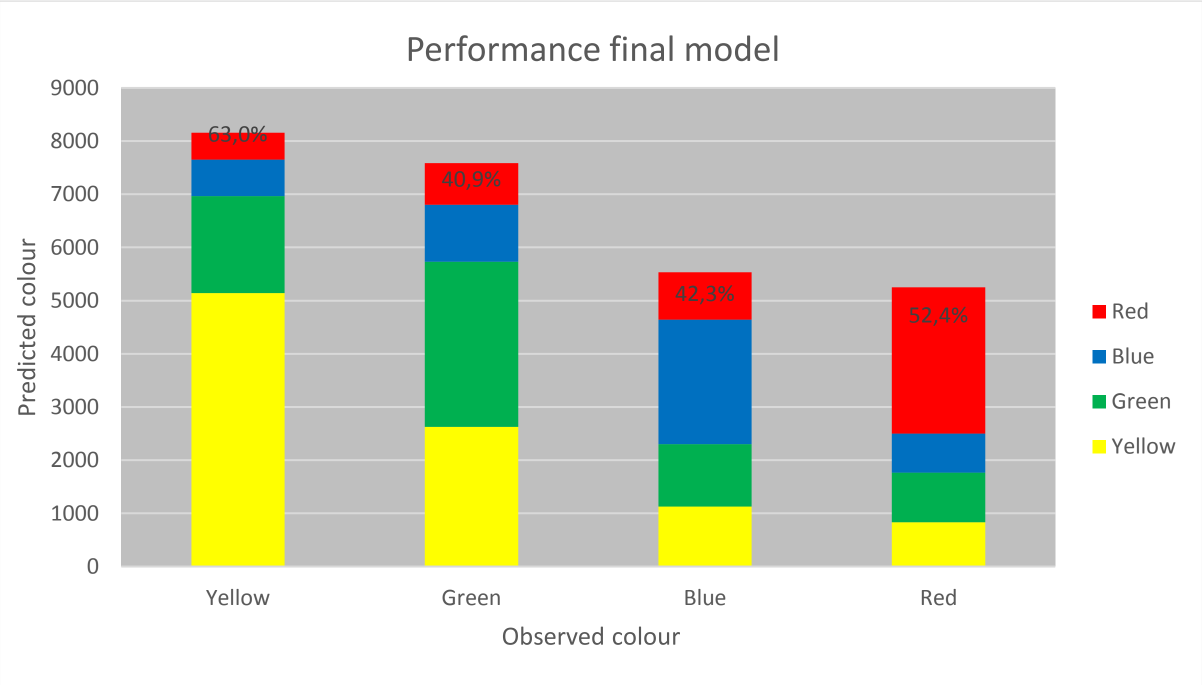

Figure 9: Performance final model

4.2. Parameter estimates of the final model

The full table with parameter estimates of the final model can be found in Appendix D (page 59). In this paragraph, a reduced table is shown. In Table 3, only the significant values of parameters are shown. Also, all redundant values are excluded (last category of every variable).

		Yellow "Harmony"		Green "Safety"		Blue "Control"	
	BSR_LIFESTYLE	B	Sig.	B	Sig.	B	Sig.
	Intercept	0,629	0,065	0,094	0,789	-0,518	0,182
Gender	[Man]	-0,718	0,000	-0,097	0,019	0,239	0,000
Age	[Till 24 years]			-1,144	0,000		
	[25 till 34 years]			-0,779	0,000		
	[35 till 44 years]	-0,369	0,027	-0,630	0,000		
	[45 till 54 years]	-0,373	0,015	-0,409	0,006		
	[55 till 64 years]	-0,288	0,043	-0,470	0,001		
	[65 till 74 years]			-0,472	0,000		
Income	[TILL 0,75X MODAL]	0,200	0,005			-0,521	0,000
	[0,75 - 1,1X MODAL]					-0,612	0,000
	[1,1 - 1,7X MODAL]	-0,170	0,010	-0,177	0,006	-0,446	0,000
	[1,7 - 2,1X MODAL]	-0,482	0,000	-0,356	0,000	-0,346	0,000
	[MORE THAN 2,1X MODAL]	-1,040	0,000	-0,617	0,000		
Family phase	[Young, alone]	-0,511	0,000				
	[Young, together]			0,325	0,009		
	[Family]	0,810	0,000	0,439	0,000	0,586	0,000
	[Older, alone]			0,312	0,014		
	[Older, together]	0,428	0,000	0,531	0,000	0,313	0,014
Employment status	[Working]	-0,632	0,000	-0,382	0,038		
	[Scholar/Student]	-0,846	0,000				
Education	[PRIMARY SCHOOL]	0,821	0,004	0,722	0,015		
	[HBO, WO]	-1,441	0,000	-0,781	0,003		
Degree of urbanism	[Degree of urbanism=1] (<500 addresses)	2,469	0,000	2,341	0,000	2,254	0,000
	[Degree of urbanism=2]	2,084	0,000	1,947	0,000	2,005	0,000
	[Degree of urbanism=3]	1,978	0,000	1,864	0,000	1,828	0,000
	[Degree of urbanism=4]	1,731	0,000	1,694	0,000	1,596	0,000
	[Degree of urbanism=5]	1,724	0,000	1,781	0,000	1,588	0,000
	[Degree of urbanism=6]	1,759	0,000	1,741	0,000	1,462	0,000
	[Degree of urbanism=7]	1,058	0,000	1,136	0,000	0,878	0,000
	[Degree of urbanism=8]	1,505	0,000	1,375	0,000	1,012	0,000
	[Degree of urbanism=9]	0,535	0,007	0,633	0,000	0,826	0,000
	[Degree of urbanism=10]	0,568	0,000	0,512	0,000	0,459	0,000
	[Degree of urbanism=11]	0,634	0,000	0,645	0,000	0,452	0,000
	[Degree of urbanism=12]	0,683	0,000	0,642	0,000	0,732	0,000
	[Degree of urbanism=13]	0,397	0,000	0,215	0,031		

Table 3: Reduced table of parameter estimates of the final model

In this table, the β and significance are most interesting parameters to evaluate. The red lifestyle is used as reference category and is thereby not included in the table. For the application of the model, it is not important which lifestyle is used as reference category. A beta below zero for a variable in a lifestyle means that that value indicates a lower probability at the certain lifestyle than the red lifestyle. A beta above zero indicates higher probability. The significance of a variable needs to be below 0,05 to be a significant factor in the model. The intercept is the β_0 of the lifestyle. There can be seen that some variables are significant for one colour, but not for another one. More information about multinomial logistic regression models can be found in paragraph 2.2.3.

In the table can be found that age is a significant variable for the green and yellow lifestyles, not for the blue lifestyle. This means that people with a blue lifestyle are equally distributed over the age categories, which can also be found in the variance analysis of chapter 3. For the green lifestyle can be seen that an older age group indicates a higher beta. This means that an older person has a higher probability of having a green lifestyle.

In contrast, the variable income shows a high significance for all of the lifestyles, especially the blue lifestyle. The betas of this variable are interesting and show clear parallels with the previous analyses. The betas of the yellow lifestyle range from high to low (higher to lower probabilities to be a yellow lifestyle). This is due to the fact that people with a yellow lifestyle have more often a lower income. It also shows the difference between the red and yellow lifestyles. Yellow lifestyle consists of more people with a lower income than the red lifestyle, the red lifestyle consists of more people with a higher income. The blue lifestyle shows the opposite of the yellow lifestyle. This is because people with a blue lifestyle have a significantly higher income.

The variable family phase shows that people with a family are likely to have a yellow or blue lifestyle. Older people who are together are more likely to have a green lifestyle people who are younger and alone are less likely to have a green lifestyle. Younger people, especially people who are alone are more likely to have a red lifestyle.

The variables employment status and education show a smaller number of significant variables. Most outstanding is that the workers and students have a lower probability of having a yellow than a red lifestyle. People who did HBO or WO have also a lower probability of having a yellow or green lifestyle. But people who only did primary school have a higher probability of having a yellow or green lifestyle.

The degree of urbanism shows that people who live in a rural area have a lower probability of having a red lifestyle, and a higher probability of having a yellow, green or blue lifestyle. To probability for having a red lifestyle get higher for people who live in urban areas. Probabilities for the other lifestyles are about equal. The only thing that stands out is that the probability for a blue lifestyle in urban areas faster declines than for the yellow and green lifestyles.

An overall thing that stands out in the parameter estimates is the fact that parameters of the yellow and green lifestyles are comparable. This can be seen in the performance of the model. The distinction between the yellow and green lifestyle is hard to predict. The lifestyles have a lot of similarities.

4.3. First implementation of the final model in OViN 2015

An exploratory implementation of the final model in the OViN 2015 dataset is made to give a first impression of the application of the BSR lifestyle model in mobility datasets. The outcomes of values of different variables are plausible. But there has to be kept in mind that uncertainty in the model is caused by the fact that only dominant lifestyles are known, not distributions of lifestyles as explained in paragraph 2.3.4. This uncertainty creates images of lifestyles that are less clear than they would be if the distribution was known. The bar graph beneath shows an example of the outcome of the application.

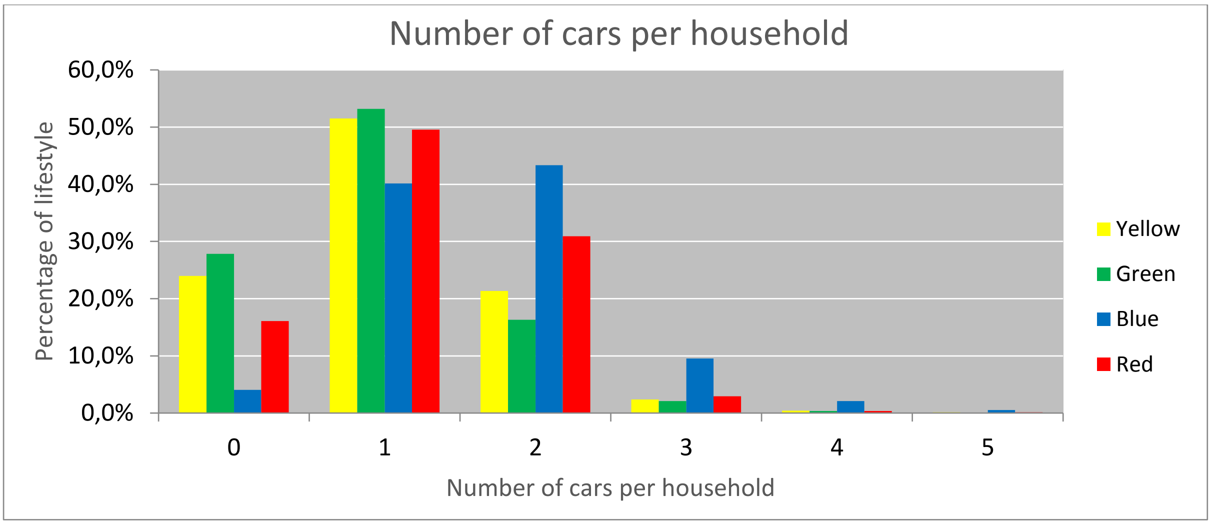

Figure 10: Number of cars per household regarding the implementation of the final model in OViN dataset

The fact that the blue lifestyle has significantly more cars than other lifestyles is most outstanding in this graph. It was expected that this lifestyle would have most cars. The blue 'control' lifestyle likes to live a luxury life and wants to show their wealth. Another thing that stands out is the fact that the yellow and especially green lifestyles have fewer cars than the average. This is due to the fact that their neighbourhood is very important to them, they can walk or cycle a lot and do not need a car. For the green lifestyle also applies that they are often elderly people who do not want to drive anymore.

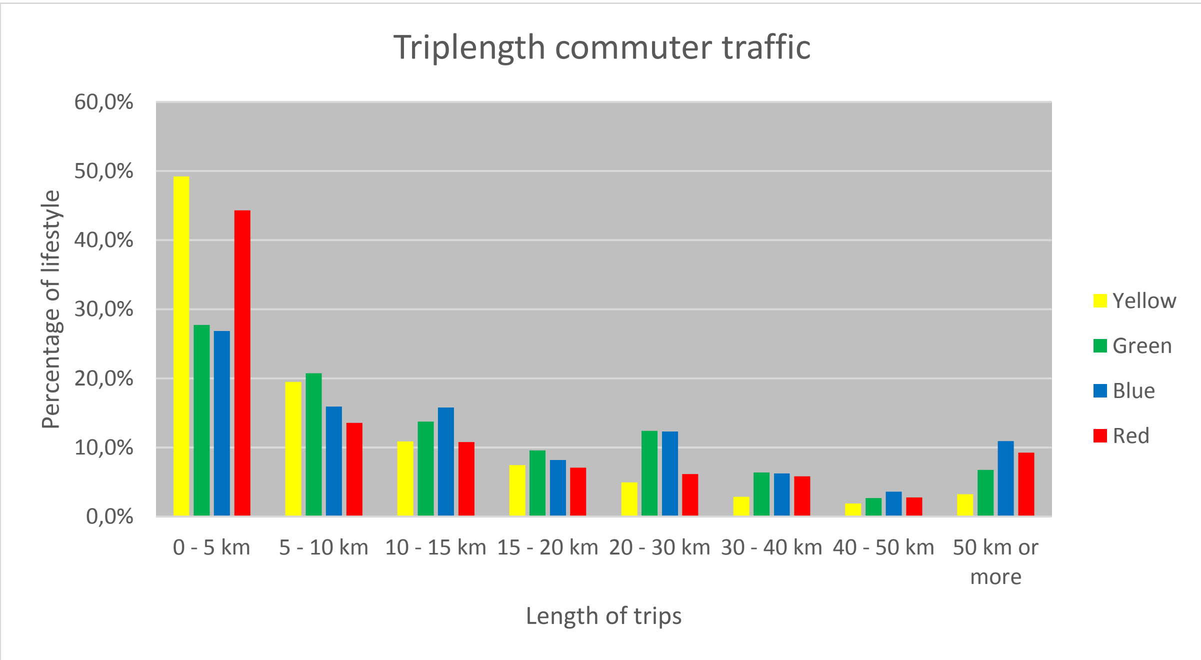

Figure 11: Length of trips for commuter traffic

In the graph above, trip lengths for commuter traffic are show. Most outstanding are the high number of short trips for the red and yellow lifestyles. However, the red lifestyle also makes relatively much long trips. The blue and green lifestyles make most average trips, between 5 and 40 kilometres.

These results are corresponding to results from a research of the municipality of Rotterdam to the relation between lifestyles and transportation (de Graaf, 2011). In that research, there was stated that "blue" people have more cars and also use them more. The values of car ownership, as a variable present in DGW dataset 2008, is comparable to the number of cars presented in this first implementation.

There are other results that show similar explainable outcomes. They are not shown in this paper because this was not the aim of the research and there is too much uncertainty about the results.

5. Conclusion and Recommendations

In this research, the relation between measurable household indicators and lifestyles that is found is described with a multinomial logistic regression model. This model is chosen because it does not make any assumptions about linearity, normality and homoscedasticity. First, the individual variables have been evaluated. After that, the significant variables were selected and combined into the final model.

Firstly, the research questions will be discussed beneath.

1. *What are the characteristics of the different lifestyles in Rotterdam?*

This research question was not difficult to answer. There was enough literature about this subject, mainly from the SmartAgent Company. The BSR model is used to describe the different lifestyles. This is explained in chapter 2.2.2.

The yellow lifestyle represents harmony. People with a yellow lifestyle are very social, they are extravert and group-oriented. They have a big social network, family and the neighbourhood are important in their lives.

The green lifestyle means security. These people are group-oriented but are also introverts. Their family, neighbourhood and privacy are important to them. They don't like changes and want to live a normal life.

The blue lifestyles represents control over life. These people are introvert and ego-oriented. They want to have a high social status and a luxury lifestyle. They make strategic decisions about their career to be successful.

The red lifestyle means vitality. People with a red lifestyle are ego-oriented and extravert. They are free-spirits and want a lot of freedom in their lives. Enjoying and experience life is very important to them.

2. *How can, according to literature, lifestyles of households be related to quantitative household indicators?*

This question relates strongly to the first question, but besides the lifestyles, literature and databases are further analysed. There is decided that the Multinomial Logistic Regression model is used. There is also decided which method is used to select variables, every variable is evaluated individually in two ways. This question provides the theoretical foundation for this research, exact information can be found in chapter 2.

3. ***Is there a relationship between lifestyles and their quantitative household indicators?***

This research question is most important and also took most the time to answer. The Multinomial Logistic Regression model is used. That way, the recommendation from question 2 is followed. First, individual variables are analysed, to investigate the relations between the variables and lifestyles. After that, significant variables are combined to specify the final model. That way, the final model is conducted and the relationship between lifestyles and their quantitative household indicators is described. The model performs for 46,7%, that is sufficient to apply to other datasets with mobility behaviour. A few exercises showed that interactions between variables or merging of values can slight improve the performance of the model. But this will be a lot of effort for a small improvement. It will not be interesting to investigate this in future research. More information can be found in chapters 3 and 4.

4. *Is there a relationship between lifestyles and mobility behaviour?*

This question is asked to give an early prediction of the relationship between lifestyles and mobility behaviour. Before starting the research, there was known that there was a big possibility that due to time constrains, this question could not be answered. That is also the case. However, as illustrated in chapter 4, there is done a first exploration of this research question. The results plausible and are explainable with the definitions of lifestyles.

But the lifestyles cannot be used as an explanatory variable in the current form. There is too much uncertainty attached to the predictions and distributions of lifestyles. Most important factor in the uncertainty is that the distributions of lifestyles are not known, therefore the uncertainty of the lifestyle model cannot be determined. Only when there is a dataset available with the distributions of the different lifestyles, the real uncertainty can be calculated.

5. *Which information is needed to complete the datasets that are going to be used in the next part of the research?*

This question was asked to give a good start to the next part of the research, it is answered in this chapter. The answer did not turned out to be a recommendation for completing a dataset. It is a recommendation about what to do in further research and how to start this.

5.1. Recommendations

The model that is conducted in this research predicts 46,7% of the dominant lifestyles. This seems low, but it is good enough to apply to datasets like OViN. It can be used to map place of living for lifestyles. Using OViN, the model can also be used to globally say something about how and where lifestyles want to travel and about their attitudes toward travel modes. This is mainly on a strategic level. It is not expected that the BSR model can be used to generate the number of trips or exact distributions. Only rough distributions regarding the whole population or the total number of trips can be generated with the model.

It is interesting to investigate the relations between mobility characteristics and lifestyles further. The relationship between lifestyles and mobility behaviour can be broader investigated than the municipality of Rotterdam did in their report *Leefstijlen, verkeer en vervoer* (de Graaf, 2011). Different topics can be studied extensively, also new topics, like trip length, can be studied. The significance of these correlations will be difficult to investigate. Not only because there is uncertainty in predicting the dominant lifestyles itself, but also because the distribution of lifestyles in the original dataset is unknown. That is why the extent to which predictions are correct or incorrect is unknown.

Besides investigating this relationship it would also be important to investigate the possibilities for applying the BSR lifestyles in current traffic models. It would be interesting to make a comparison between models with and without the BSR lifestyles in them. Before starting new research, it is important to consider the added value of BSR lifestyles. In my opinion, the BSR-lifestyles will not extra value as an explanatory variable to the existing traffic models, due to the high level of uncertainty in the model. However, it would be interesting to add the relevant variables from this research to existing mobility models.

However, it would be relevant to label respondents in mobility datasets with a lifestyle, which is not explanatory. These lifestyle labels can be used for communication purposes. For the municipality of Rotterdam, this communication can be internal and external communication. The municipality of Rotterdam uses the BSR model in different fields. It can be useful to use BSR lifestyles to indicate group behaviours. Also, there can be added typical lifestyles to results of transportation models like trips or emissions. That way, policies can be adapted to the behaviour and worldviews of the different lifestyles and people of different lifestyles can be approached in an appropriate and appealing way, similar to the marketing strategies for which the BSR model was originally intended.

6. Conclusie en aanbevelingen

Het verband dat tussen meetbare huishoudindicatoren en levensstijlen is gevonden is in dit onderzoek beschreven met het multinomiaal logistische regressie model. Deze is gekozen omdat het geen aannamen maakt over lineaire verdeeldheid, normaliteit en homoscedasticiteit. De individuele variabelen zijn geëvalueerd, waarna de significante variabelen zijn geselecteerd, die zijn samengevoegd tot een model.

Hieronder worden eerst de onderzoeksvragen besproken.

1. *Wat zijn de eigenschappen van de verschillende levensstijlen in Rotterdam*

Deze onderzoeksvraag was niet moeilijk te beantwoorden. Er was veel literatuur beschikbaar, voornamelijk vanuit de SmartAgent Company. Het BSR model is gebruikt om de verschillende levensstijlen te beschrijven. De vraag is beantwoord in hoofdstuk 2.2.2.

De gele levensstijl staat voor harmonie. Mensen met een gele levensstijl zijn erg sociaal, ze zijn extravert en groep georiënteerd. Ze hebben een groot sociaal netwerk, waar familie en de buurt centraal staan.

De groene levensstijl staat voor zekerheid. Deze mensen zijn groep georiënteerd, maar zijn ook introvert. Hun familie, buren en privacy zijn erg belangrijk voor hun. Ze houden niet van veranderingen en willen een zo normaal mogelijk leven leiden.

De blauwe levensstijl staat voor controle over het leven. Deze mensen zijn introvert en ego georiënteerd. Ze willen een hoge sociale status en veel luxe in het leven. Ze maken strategische beslissingen over hun carrière om succesvol te zijn.

De rode levensstijl staat voor vitaliteit. Mensen met een rode levensstijl zijn ego georiënteerd en extravert. Ze zijn vrije geesten en willen veel vrijheid in hun leven. Plezier hebben in het leven en veel dingen meemaken zijn voor hun belangrijk.

2. *Hoe kunnen, volgens de literatuur, levensstijlen van huishoudens zijn verbonden met hun kwantitatieve huishoud indicatoren?*

Deze vraag relateert sterk met de eerste vraag, maar niet alleen de levensstijlen, maar ook de literatuur en datasets zijn verder geanalyseerd. Er is besloten om een Multinomiaal Logistisch Regressie model te gebruiken. Ook is de methode die zal worden gebruikt om variabelen te selecteren bepaald, elke variabele is geëvalueerd op twee manieren. De verdere theoretische onderbouwing kan worden gevonden in hoofdstuk 2.

3. ***Is er een verband tussen levensstijlen en hun kwantitatieve huishoud indicatoren?***

Deze onderzoeksvraag is het belangrijkst en kostte daarom ook het meeste tijd om te beantwoorden. Er is niet alleen gezocht naar of er een verband bestaat, al snel bleek dat dat zo is, maar er is ook onderzocht hoe dat verband eruit ziet. Het multinomiaal logistisch regressie model is gebruikt, zoals in vraag 2 was aanbevolen. Eerst zijn de individuele variabelen geanalyseerd, waarna de significante variabelen zijn gecombineerd tot een definitief model. Op die manier beschrijft het definitieve model zo goed mogelijk de relatie tussen huishoudindicatoren en levensstijlen. Het model voorspelt 46,7% van de levensstijlen juist. Dit is voldoende om toe te passen op datasets met mobiliteitsgedrag. Een aantal vingeroefeningen lieten zien dat interacties tussen variabelen of samenvoegen van variabelen de prestatie van het model met een paar procent kunnen verbeteren. Maar dit is erg veel moeite voor een kleine verbetering. Ook omdat er een model ontwikkeld is die prima werkt, zal het niet interessant zijn om hier in vervolgonderzoek naar te kijken. Meer informatie is te vinden in hoofdstuk 3 en 4.

4. *Is er een verband tussen levensstijlen en mobiliteitsgedrag?*

Deze vraag is gesteld om een eerste voorspelling te kunnen geven over de relatie tussen mobiliteitsgedrag en levensstijlen. Voordat het onderzoek werd gestart was er al bekend dat er door tijdsdruk een grote kans was dat deze vraag niet zou worden beantwoord. Dit is dan ook het geval. Hoewel er wel een eerste observatie gedaan is. Hieruit blijkt dat er logische verbanden kunnen worden gevonden tussen levensstijlen en mobiliteitsgedrag. Meer resultaten kunnen worden gevonden in hoofdstuk 4.

Belangrijk om te onthouden is dat de levensstijlen in de huidige vorm niet kunnen worden gebruikt als directe verklarende variabelen in modellen. Er is te veel onzekerheid verbonden met de voorspellingen en verdelingen van levensstijlen. Het belangrijkst hierin is de onzekerheid over de verdeling van levensstijlen van personen, deze is onbekend in de huidige dataset. Hierdoor kan de werkelijke onzekerheid van het model niet worden berekend. Alleen als er een dataset beschikbaar komt waar deze verdelingen wél bekend zijn, kan de onzekerheid berekend worden.

5. *Welke informatie is nodig om de datasets compleet te maken, zodat ze kunnen worden gebruikt in het volgende deel van het onderzoek?*

Deze vraag is gesteld om de start van het eventuele vervolgonderzoek soepel te laten verlopen. Het resultaat bleek geen aanbeveling te zijn voor het complementeren van een dataset, maar een aanbeveling voor het vervolgonderzoek.

6.1. Aanbevelingen

Het model dat in dit onderzoek is gemaakt werkt voor 46,7%. Dit is goed genoeg om resultaten te verkrijgen en toe te passen op een dataset van bijvoorbeeld OViN. Het model kan worden gebruikt om een algemene verdeling te maken van levensstijlen over de ruimte. Door gebruik van het OViN kan het model worden gebruikt om globaal iets te zeggen over het aantal ritten of houding tegenover vervoerswijzen van levensstijlen. Dit blijft op strategisch en globaal niveau. Het valt niet te verwachten dat het BSR model kan worden gebruikt om exacte aantallen ritten of distributies te genereren. Alleen ruwe verdelingen ten opzichte van een hele populatie of totaal aantal ritten kunnen worden gebruikt.

In een vervolgonderzoek zouden deze relaties tussen levensstijlen en mobiliteitskenmerken verder moeten worden uitgediept. Het zou meer of relevantere relaties tussen levensstijlen en mobiliteit kunnen opleveren dan de gemeente Rotterdam in hun rapport *Leefstijlen, verkeer en vervoer* (de Graaf, 2011) heeft onderzocht. Verschillende onderwerpen kunnen uitgebreider worden bestudeerd, ook nieuwe onderwerpen, zoals ritlengte, kunnen worden onderzocht. De significantie van deze correlaties zal lastig zijn om te onderzoeken. Niet alleen omdat er onzekerheid is in het voorspellen van de dominante levensstijlen zelf, maar ook omdat de distributie van levensstijlen niet aanwezig is in de originele dataset. Hierdoor is de mate waarin voorspellingen correct of incorrect zijn niet bekend.

Naast het uitdiepen van deze relatie is het ook belangrijk om de mogelijkheden voor het toepassen van het BSR model in huidige verkeersmodellen. Het zou interessant zijn om een vergelijking te maken tussen een model met en een model zonder BSR levensstijlen. Voordat er een vervolgonderzoek wordt gestart, is het belangrijk de toegevoegde waarde van het BSR model te overwegen. In mijn opzicht zouden BSR levensstijlen als verklarende variabelen niets toevoegen aan de huidige modellen, omdat er veel onzekerheid in het voorspellen van levensstijlen. Maar het zou wel relevant zijn om de variabelen die in dit onderzoek zijn gebruikt, toe te voegen aan de huidige verkeersvoorspellingen, zonder de tussenkomst van het BSR model.

Het zou wel relevant zijn om de BSR levensstijlen als een manier van communiceren te gebruiken. Respondenten kunnen worden gelabeld met een levensstijl, die niet wordt gebruikt als verklarende variabele, maar na berekeningen wordt gebruikt als een manier om te communiceren. Deze communicatie kan intern zijn bij de gemeente Rotterdam, waar het BSR model al wordt gebruikt in verschillende disciplines. Maar de communicatie kan ook extern zijn. Het BSR model kan worden gebruikt om gedragingen van de levensstijlen te identificeren. Hiermee kan beleid worden afgesteld op de gedragingen en wereldbeelden van de verschillende levensstijlen. Ook kunnen mensen op een passende manier worden aangesproken, op een manier die lijkt op de marketing strategieën waarvoor het BSR model oorspronkelijk voor is ontwikkeld.

7. Bibliography

Bamberg, S., Hunecke, M., & Blöbaum, A. (2007). Social context, personal norms and the use of public transportation: Two field studies. *Journal of Environmental Psychology*, 190-203.

Benes, F. (2015). *Mobility at Old Age: The predictive power of perceived car-dependence on the mobing plans of eldery Swedes.* Stockholm: Stockholm University.

de Graaf, P. (2011). *Leefstijlen, verkeer en vervoer.* Rotterdam: Centrum voor Onderzoek en Statistiek.

Goudappel Coffeng. (2012). *Lifestyle-clustering en mobiliteitsgedrag Mogelijkheden voor toepassing van BSR clustering op mobiliteitsgedrag en verkeersmodellen.* Deventer: Goudappel Coffeng.

Hammouch, M. B. (2007). *De toepasbaarheid van leefstijlen.* Delft: Technische Universiteit Delft.

Hattum, P. v., & Hoijtink, H. (2008). *The Proof of the Pudding is in the Eating Data Fusion: An Application in Marketing.* Utrecht: Utrecht University.

Hildebrand, E. D. (2003). Dimensions in elderly travel behaviour: A simplified activity-based model using lifestyle clusters. *Transportation*, 285-306.

Lamme, A. W. (2003/2010). *SmartBSR Belevingsonderzoek; Achtergronden bij een metataal.* Amersfoort: The SmartAgent Company.

Reitsma, D. (2012). *Bestuurders: "Begrijpen wij elkaar?"; De burger op zoek naar de bestuurder en de bestuurdre zoekt de burger. Het toepassen van leefstijlen in het openbaar domein.* Amersfoort: The SmartAgent Company.

Scheiner, J., & Holz-Rau, C. (2007). Travel mode choice: affected by objectivee or subjective determinants? *Transportation*, 487-511.

SmartAgent. (2008). *Woonbeleving Regio Rotterdam 2008.* Leusden: SmartAgent.

Smartagent. (2011). *Utrechts Autodelen: perceptie en praktijk.* Amersfoort: SmartAgent. Opgehaald van http://www.crow.nl/vakgebieden/verkeer-en-vervoer/bibliotheek/kennisdocumenten/utrechts-autodelen-perceptie-en-praktijk

SmartAgent. (2011). *Utrechts Autodelen: perceptie en praktijk.* Amersfoort: SmartAgent. Opgehaald van http://www.crow.nl/vakgebieden/verkeer-en-vervoer/bibliotheek/kennisdocumenten/utrechts-autodelen-perceptie-en-praktijk

SmartAgent. (2012). *Woonservice Gebieden Rotterdam.* Amersfoort: SmartAgent.

Smartagent. (2013, November 7). *BSR-English*. Opgehaald van Smartagent: http://www.hoedenktu.nl/ennis/surveys/BSREngels/

Smartagent. (2013, November 7). *BSR-Nederlands*. Opgehaald van Smartagent: http://www.hoedenktu.nl/ennis/surveys/bsrwebsite/

Smartagent. (2013). *De Grote Woontest 2012 in de Regio Rotterdam.* Amersfoort: Smartagent.

Starkweather, J., & Moske, A. (2011, August). *Multinomial logistic regression.* Opgeroepen op December 2016, van http://it.unt.edu/sites/default/files/mlr_jds_aug2011.pdf van Hattum, P. (2010). *Brand Strategy Research; Van theorie naar statistiek.* Amersfoort: The SmartAgent Company.

Van Hattum, P., & Hoijtink, H. (2009). Market Segmentation Using Brand Strategy Research: Bayesian Inference with Respect to Mixtures of Log-Linear Models. *Journal of Classification*, 297-328.

8. Appendices

8.1. Appendix A. Variance and numbers of values in variables

In this chapter, bar graphs of evaluated variables can be found with short descriptions.

8.1.1. Gender

It was expected that lifestyles would have an almost equal distribution of both genders. For the green, blue and red lifestyle (Starkweather & Moske, 2011), this is true with a distribution between 45% and 55% percent of the two genders. But people with a yellow lifestyle are more likely to be a woman, 71% of people in this lifestyle are woman (Figure 12). This is a significant difference. More tables of this lifestyle can be found in appendix C, chapter 8.3.1 at page 46.

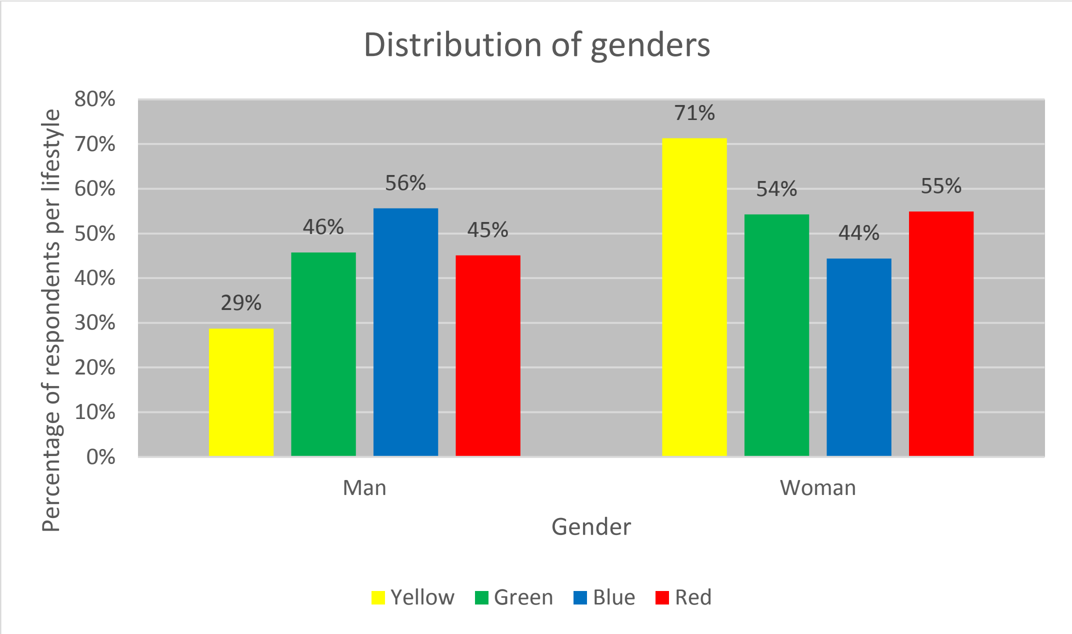

Figure 12: Distribution of genders over the BSR lifestyles

8.1.2. Age

The mean age of the respondents is between 35 and 55 years. The effect of age on the lifestyles can mainly be seen in the red and the green lifestyles. The age of the green lifestyle is significantly older and the red lifestyle is significantly younger than the other lifestyles (Figure 13). More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.2 at page 47.

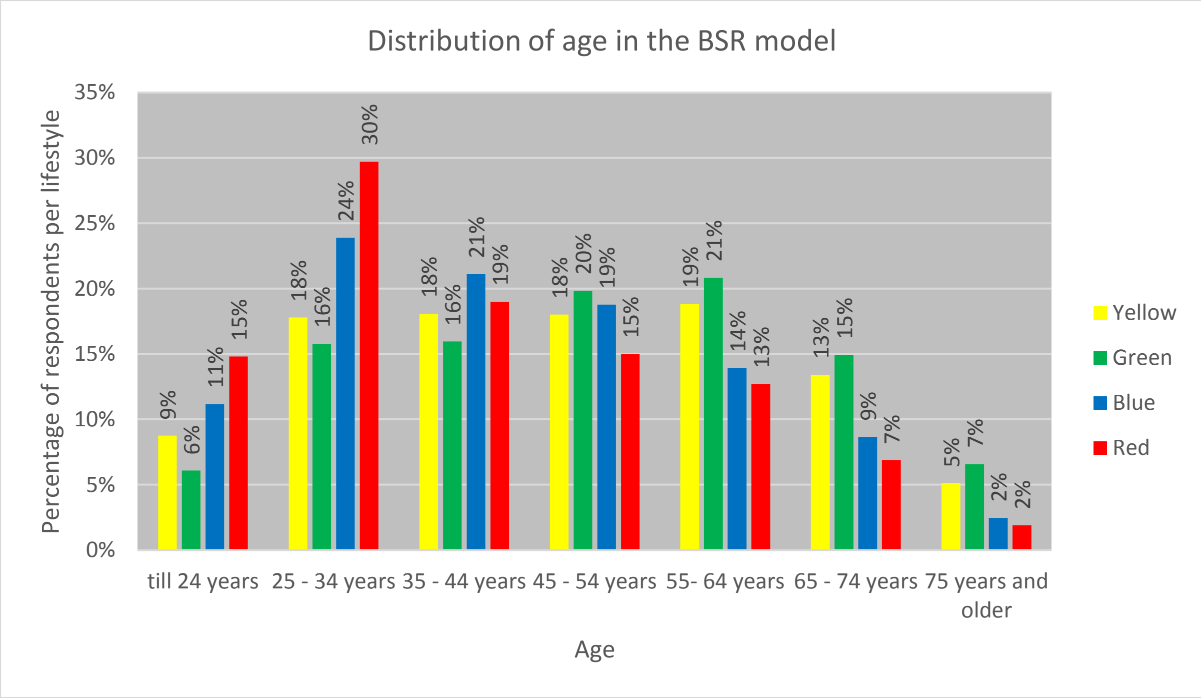

Figure 13: Distribution of age over the BSR lifestyles

8.1.3. Family phase summarized

The different phases that are being distinguished in this indicator are: *1-person young, 2-person young, family, 1-peron old, 2-person old* and *different*. Young 1-person households are less likely to have a yellow lifestyle and more likely to have a red lifestyle. Green lifestyles are more likely to be old. Blue and red lifestyles are less likely to be old, where blue is less likely to be alone and red is less likely to be together. Yellow and Blue lifestyles are most likely to have a family, this can be found in Figure 14. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.3 at page 48.

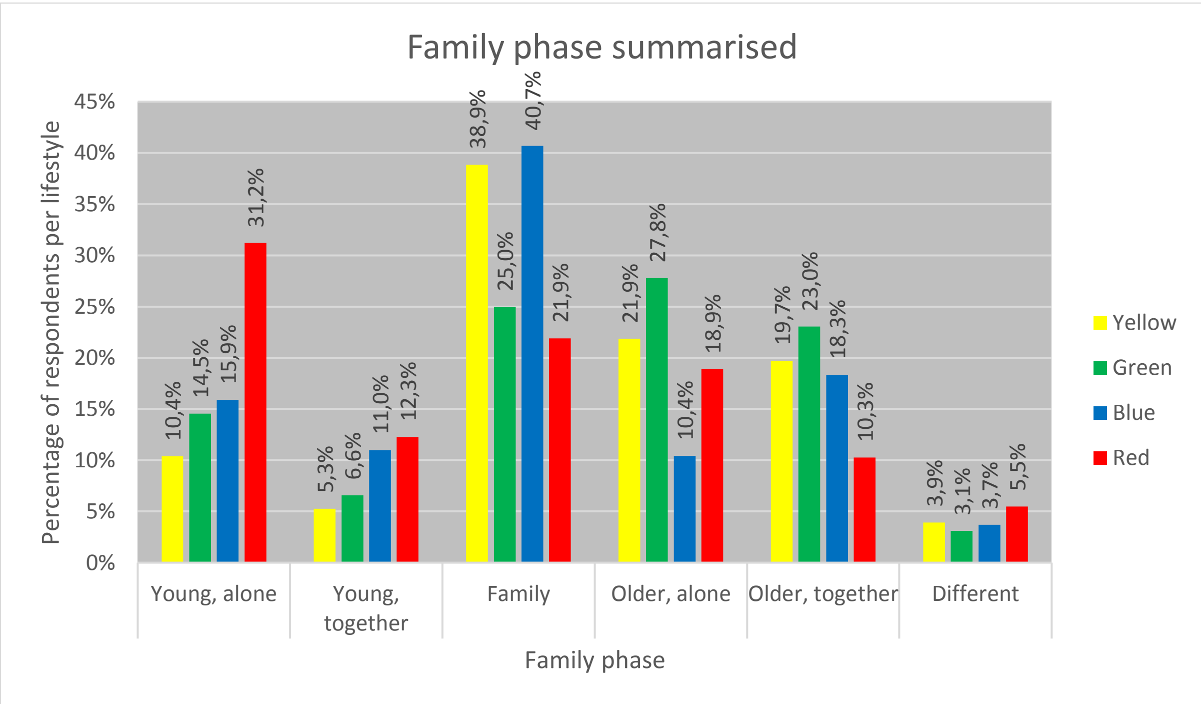

Figure 14: Distribution of the summarized family phase over the BSR lifestyles

8.1.4. Family phase extended

A combination of the distribution from the variables age and summarized family phase can be found in this variable. There are a few lifestyles that significantly differ from the mean. There is again found that the red lifestyle is relatively young and do not want to bond to a family. The green lifestyle is relatively old and also does bond less to a family than average. The yellow and blue lifestyles like to bond to a family, where the yellow lifestyle is a bit older than the blue lifestyle (Figure 15). More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.4 at page 49.

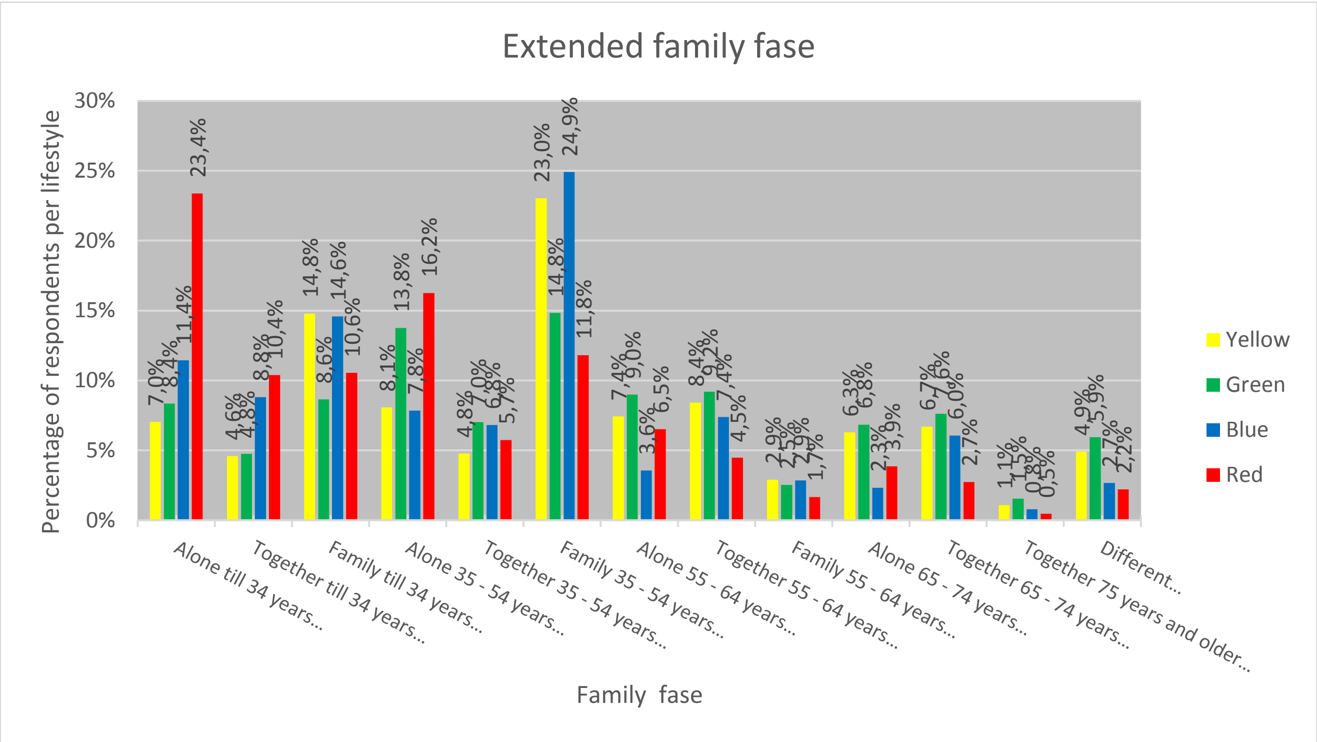

Figure 15: Distribution of the extended family phase over the BSR lifestyles

8.1.5. Income

The income of the population has mainly effect on the yellow and blue lifestyles. The yellow lifestyle has less income and the blue lifestyle has more income than the other lifestyles. Most significant is the difference at twice the modal income. Another significant difference can be found in the red lifestyle. A group of people does not want to share their income, in the red lifestyle, this group is the smallest (Figure 16). More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.5 at page 50.

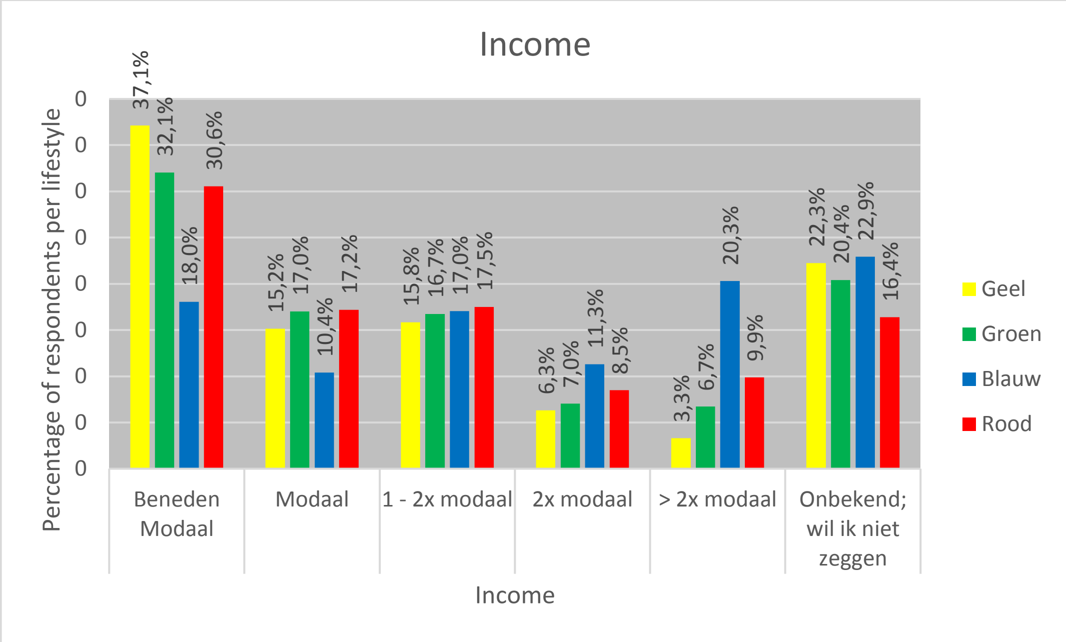

Figure 16: Distribution of income over the BSR lifestyles

8.1.6. Employment status

The employment situation can indicate which lifestyle a person has. People with a yellow lifestyle are less likely to be an entrepreneur and more likely to be a housewife. A green lifestyle is more likely to be retired, a blue lifestyle is less likely to have no job and a red lifestyle is more likely to be a student (Figure 17). More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.6 at page 51.

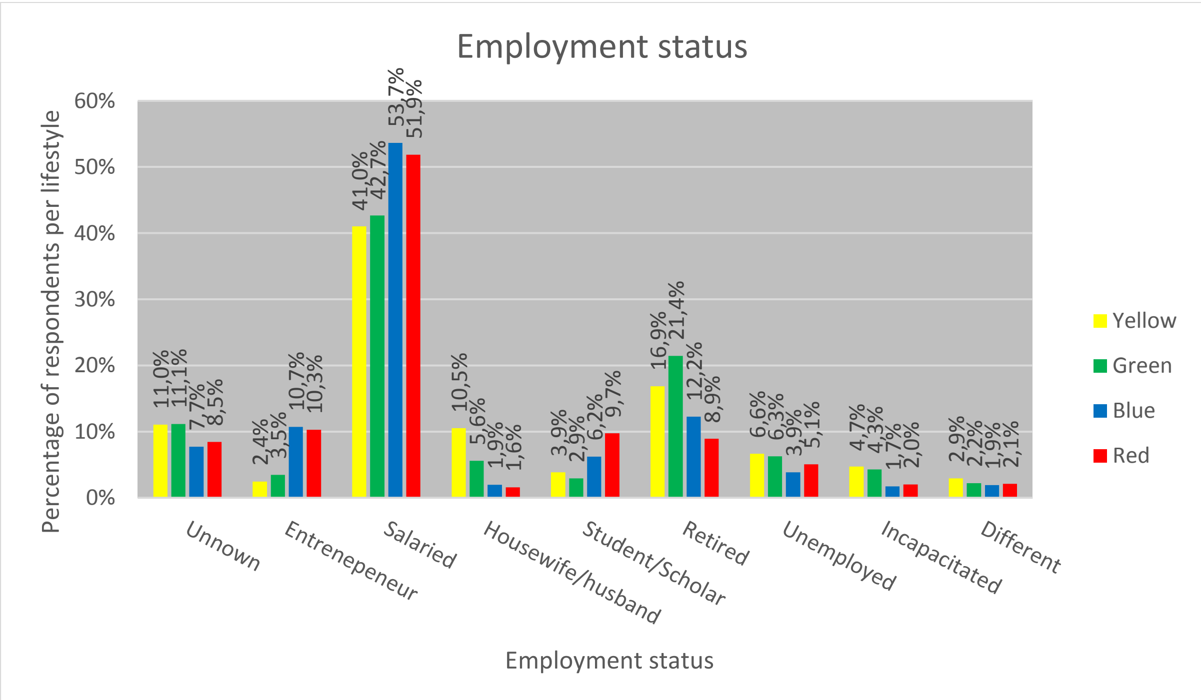

Figure 17: Distribution of employment statuses over the BSR lifestyles

8.1.7. Education

The education of a respondent has a big influence on its lifestyle. The yellow lifestyle is the most significantly different form the other lifestyles and is less educated. The red and blue lifestyles are more educated than the other lifestyles, where red has the highest difference with other lifestyles. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.7 at page 52.

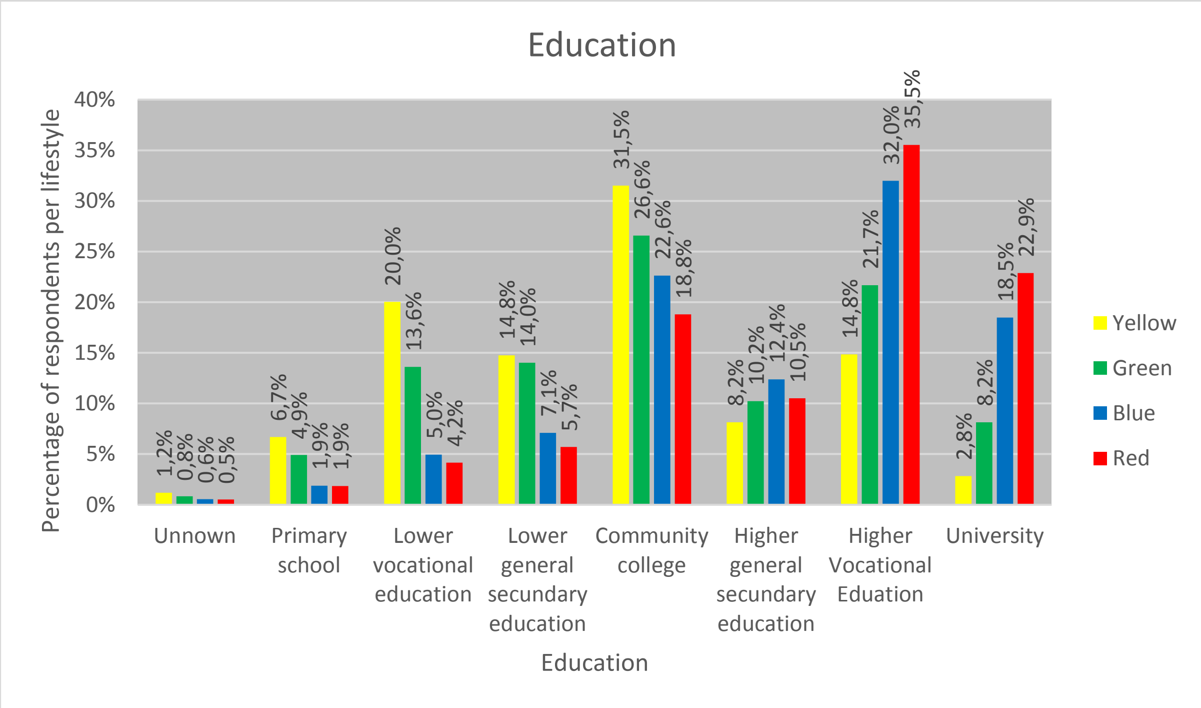

Figure 18: Distribution of education over the BSR lifestyles

8.1.8. Car ownership

Measurements for car ownership are only made in the 'De Grote Woontest' survey of 2008. That is why this variable is less reliable than other variables. Car ownership can indicate the lifestyle of a household. Households with a red lifestyle have less cars than other lifestyles. Households with a blue lifestyle have more cars than the average household (Figure 19). More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.8 at page 53.

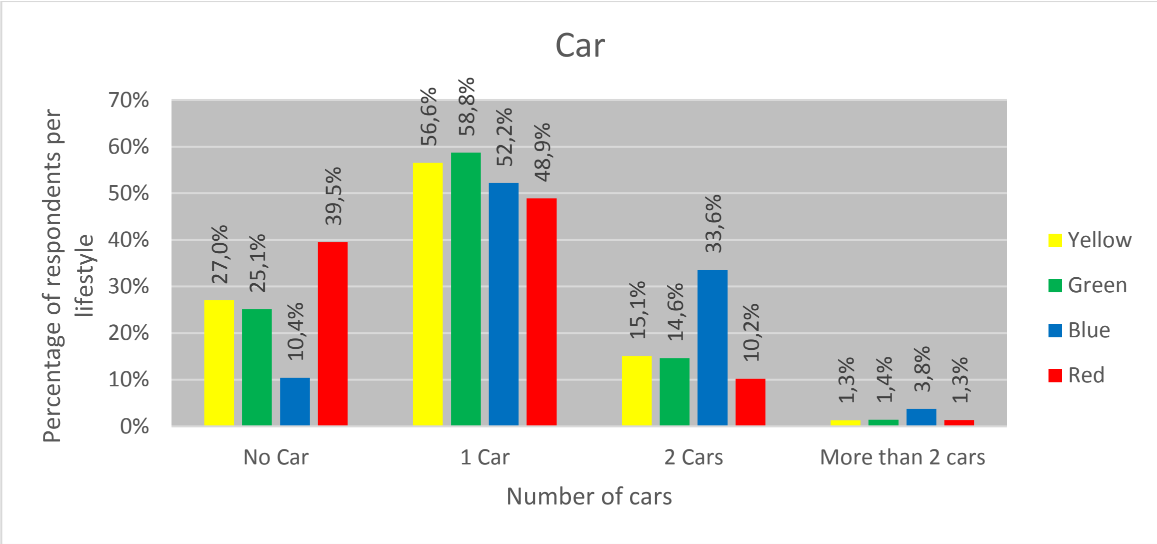

Figure 19: Distribution of car ownership and BSR lifestyles

8.1.9. Land of birth respondent

The hypothesis was that land of birth from a person would influence their lifestyle. However, there cannot be found a significant difference between lifestyles and a land of birth. That is why the histogram is not presented here. Most respondents are born in the Netherlands. The other percentages of birth lands are too low to draw conclusions out of them. Tables and figures of this lifestyle can be found in appendix C, chapter 8.3.9 at page 54.

8.1.10. Degree of urbanism

Most outstanding is this variable is that the red lifestyle does like to live in cities or places with a high degree of urbanism. The other lifestyles are more equal and live a more quit area in a city or village (Figure 20). Tables and figures of this lifestyle can be found in appendix C, chapter 8.3.10 at page 55.

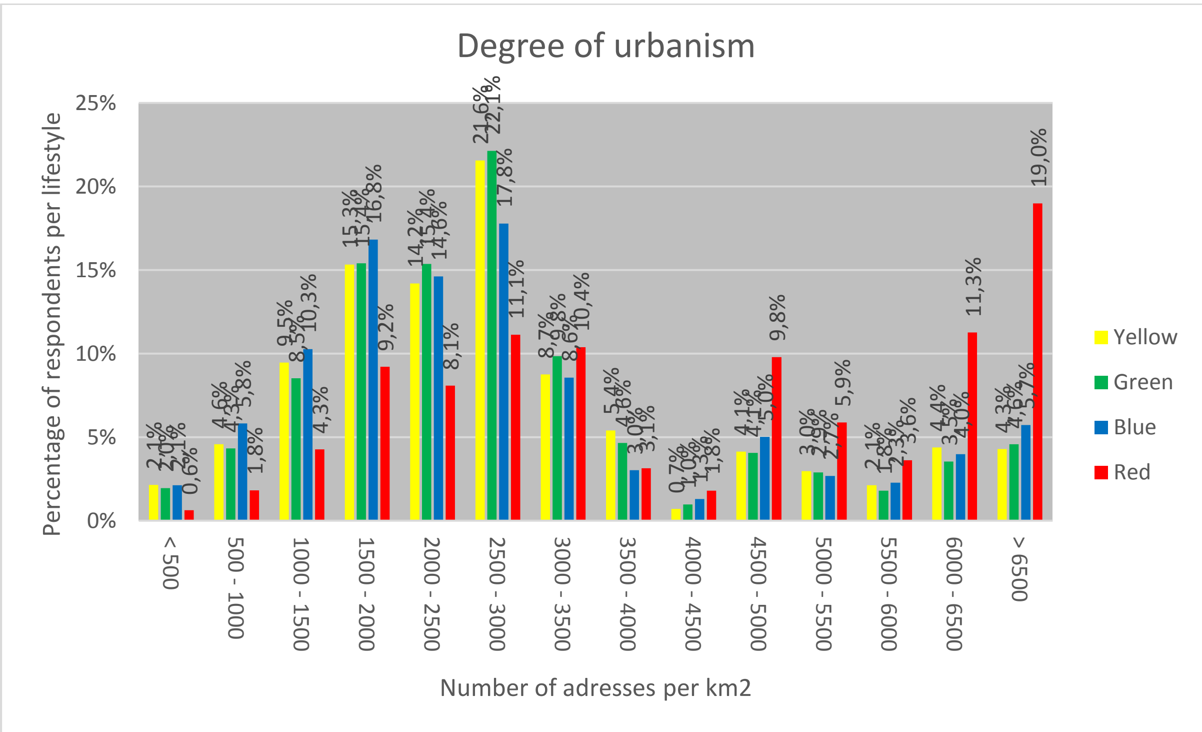

Figure 20: Distribution of degree of urbanism and BSR lifestyles

8.1.11. Mean distance to a main road

This variable shows a small number of differences between the different lifestyles. The only significant difference is found in the fact that the red lifestyle lives further away from main roads. This is due to the high number of people in red lifestyles that are living in cities (Figure 21). More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.11 at page 57.

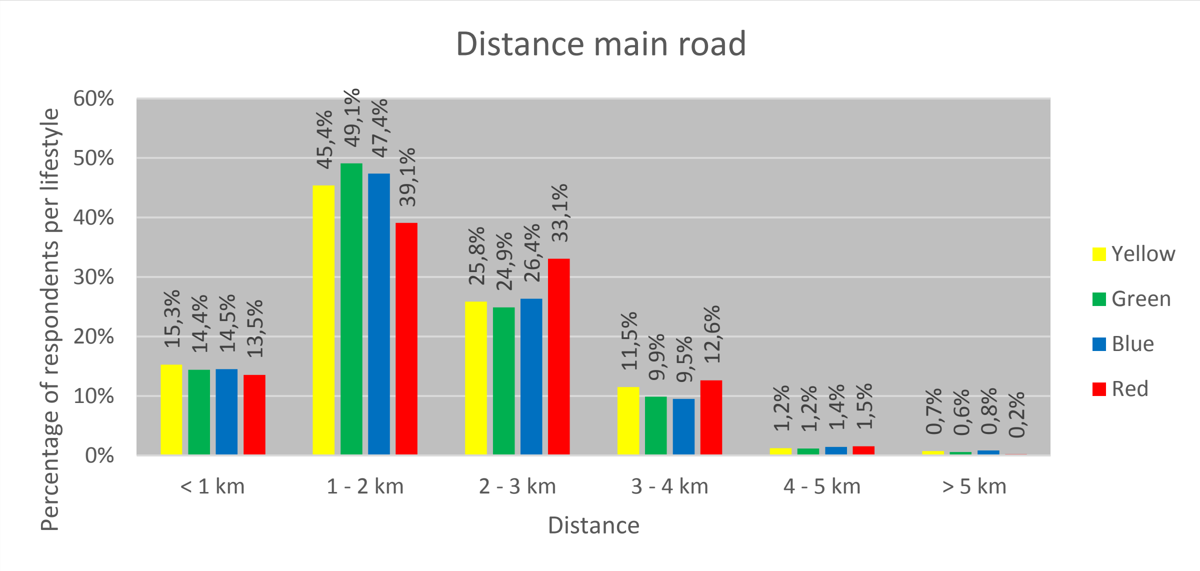

Figure 21: Distribution of distances to main road and BSR lifestyles

8.1.12. Mean distance from a train station

Again, there are no big differences between the lifestyles. The only difference can be found in the fact that the red lifestyle lives closer to a train station than other lifestyles. This is also due to the fact that red lifestyles mostly live in cities, where there is a higher number of train stations (Figure 22). More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.12 at page 58.

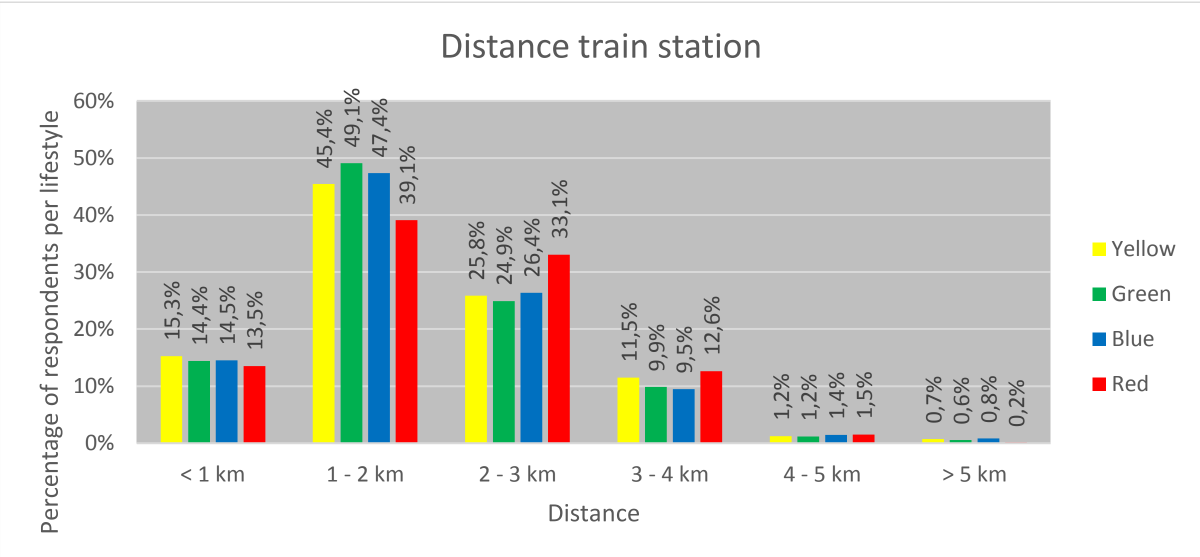

Figure 22: Distribution of distances to train station and BSR lifestyles

8.2. Appendix B. Performances of single variable models

In this chapter, a summary of the performances of all single variable models is given.

8.2.1. Gender

The model that is made with the variable gender performs for **35,2%** for the mean of the whole model. In comparison to the other models it is high, but there are only values for the yellow and green lifestyles. This is due to the fact that there are only 2 categories in this variable. The model distinguishes only two colours. More tables of this lifestyle can be found in appendix C, chapter 8.3.1 at page 46.

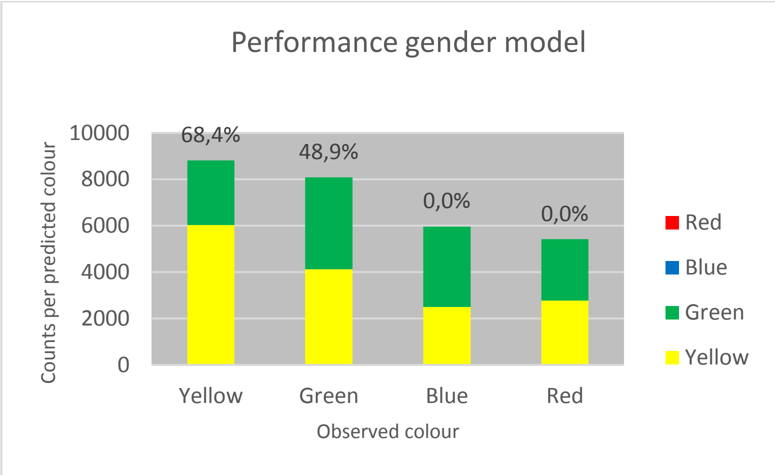

Figure 23: Performance of gender model

8.2.2. Age

The model made with only the variable age performs for **32,4%.** This performance is low compared to other variables. But this model has a few strong points; this model makes a difference between the yellow and green lifestyles. It also includes the red lifestyle in the prediction of lifestyles. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.2 at page 47.

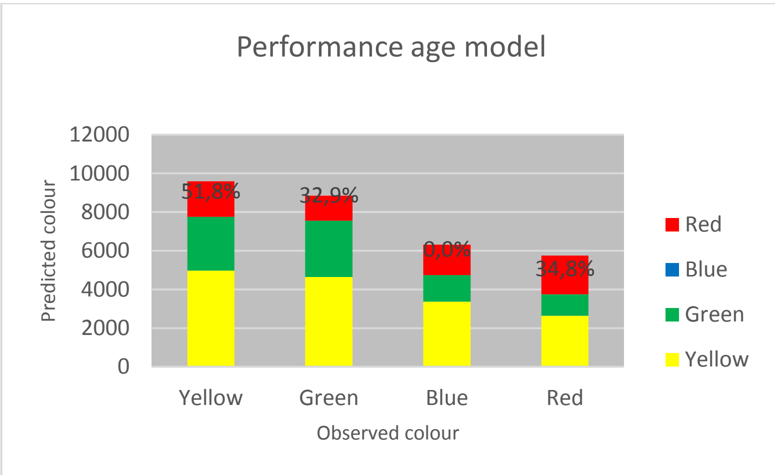

Figure 24: Performance of age model

8.2.3. Family phase summarized

With **36,5%**, the model performs relatively well. It makes a distinction between the yellow, green and red lifestyles, but is not an explanatory variable for the blue lifestyle. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.3 at page 48.

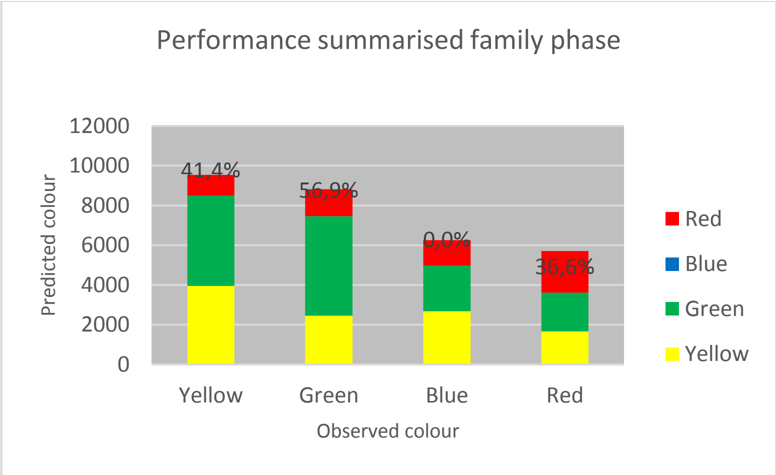

Figure 25: Performance of summarized family phase model

8.2.4. Family phase extended

The model performs well for **36,4%**. That is just as good as the summarized model. It does not have any advantage to include the extended model over the summarized model. In further models, there will be included the summarized family phase model to improve calculation speed. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.4 at page 49.

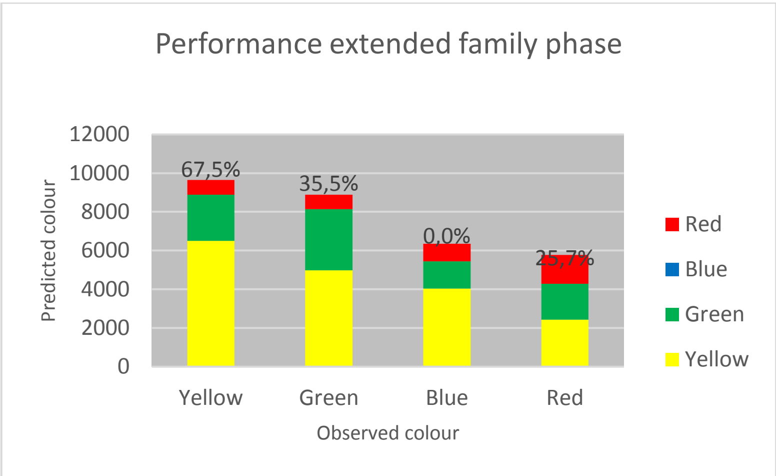

Figure 26: Performance extended family phase model

8.2.5. Income

The model that is conducted with this variable performs for **35,6%**, this is average compared to the other variables. However, it only relates to the yellow and blue lifestyles. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.5 at page 50.

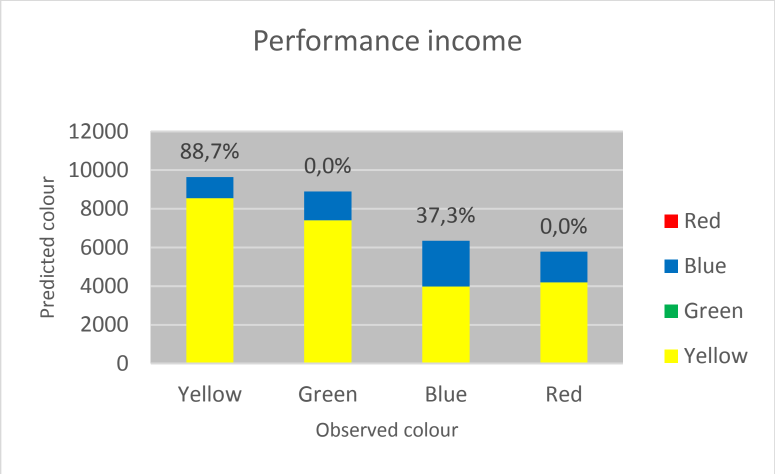

Figure 27: Performance income

8.2.6. Employment status

The model made with this variable performs for **34,2%**. Compared to the other variables, this is an average score. The fact that is distinguishes all four lifestyles in the classification table is better than some other variables like income or land of birth. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.6 at page 51.

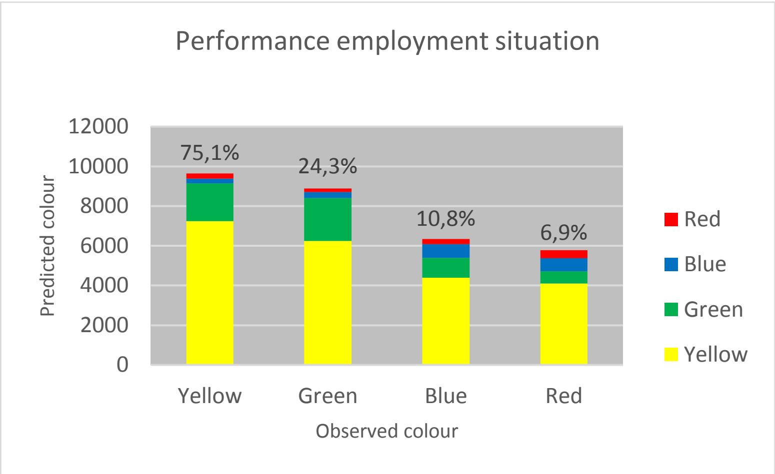

Figure 28: Performance employment status

8.2.7. Education

The model conducted with only this variable performs with **38,2%** the best of all models from one variable. But it does not have a relation with the blue lifestyle. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.7 at page 52.

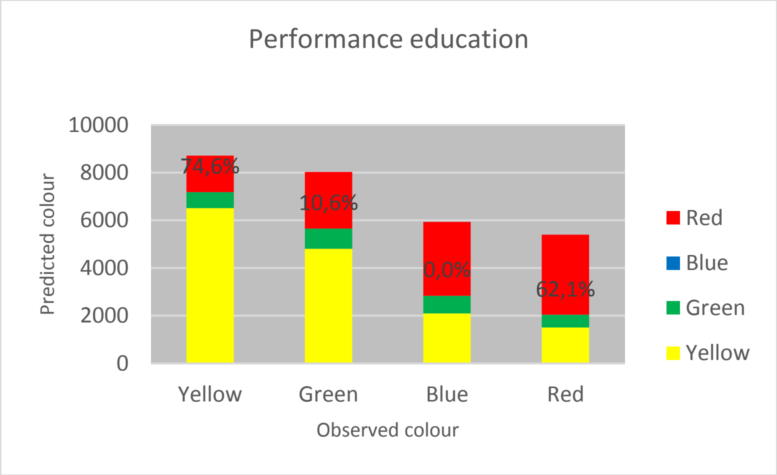

Figure 29: Performance education

8.2.8. Car ownership

With **33,9%** the model conducted with only this variable performs relatively low. It only makes a distinction between the yellow and blue lifestyles. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.8 at page 53.

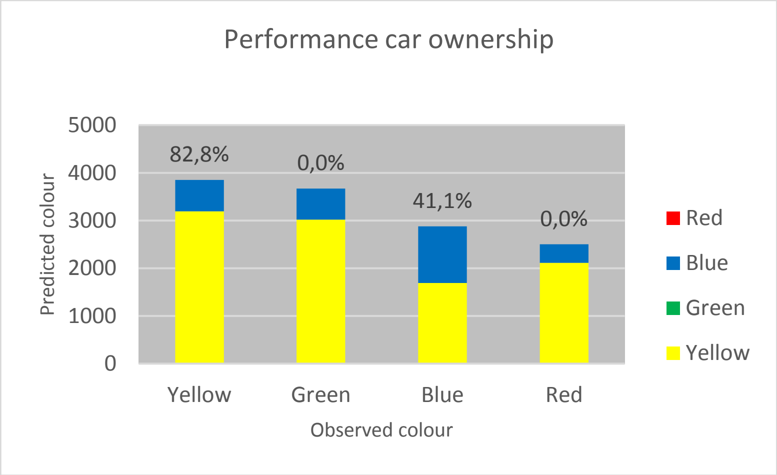

Figure 30: Performance car ownership

8.2.9. Land of birth

The model conducted with only this variable performs relatively bad, **31,4%**. This is due to the fact that more than 80% of the respondents is Dutch. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.9 at page 54

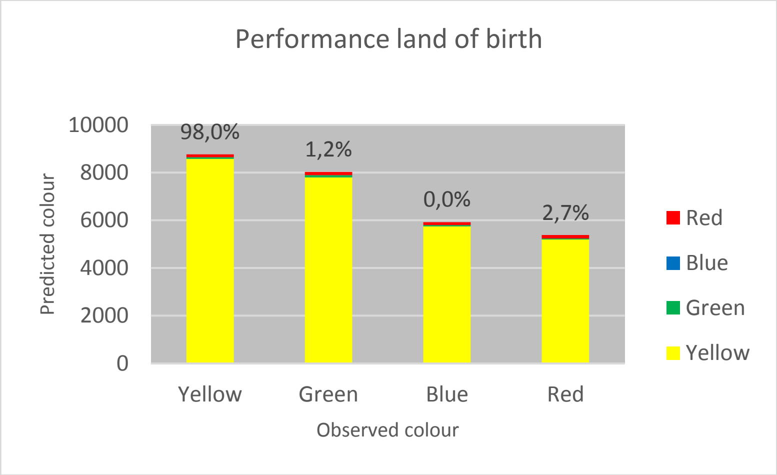

Figure 31: Performance land of birth respondent

8.2.10. Degree of urbanism

The model conducted with only this variable performs for **35,5%**, compared to other variables this is average. It is a big benefit that this model makes a distinction between the yellow, green and red lifestyles. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.10 at page 55.

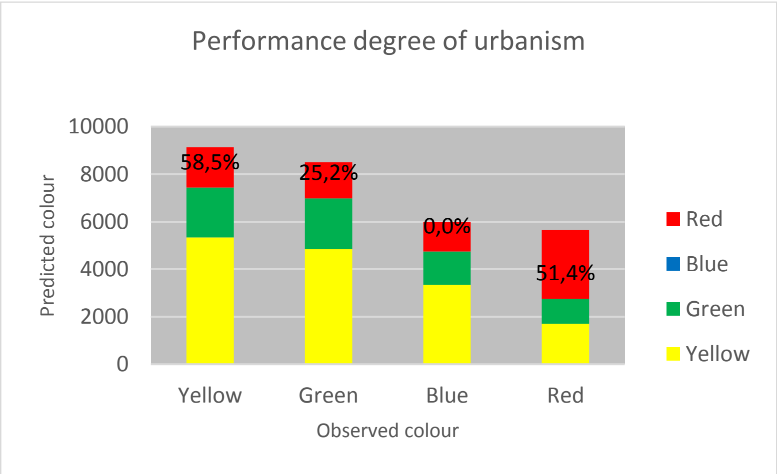

Figure 32: Performance degree of urbanism

8.2.11. Distance to main road

The model conducted with only this variable performs badly with **31,4%** and no distinguishing between different lifestyles. The model predicts only the yellow lifestyle. This is due to the fact that in Rotterdam, there is not enough variation between the different lifestyles. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.11 at page 57.

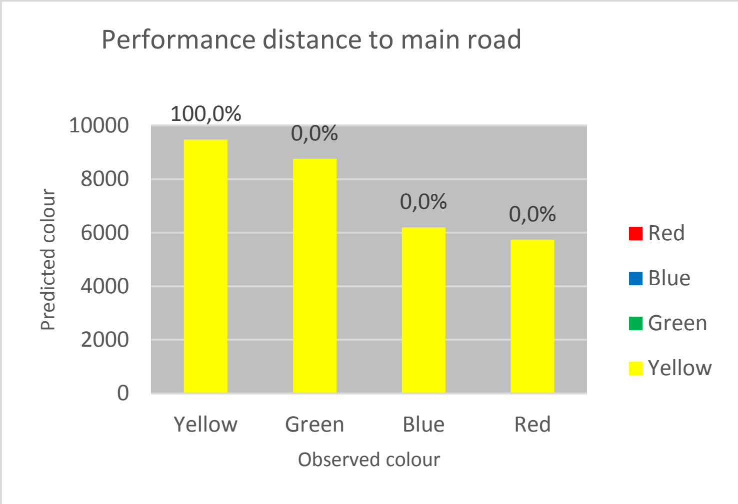

Figure 33: Performance mean distance to main road

8.2.12. Distance to train station

With **31,9%** the model performs under average. It only predicts the yellow and red lifestyles. More tables and figures of this lifestyle can be found in appendix C, chapter 8.3.12 at page 58.

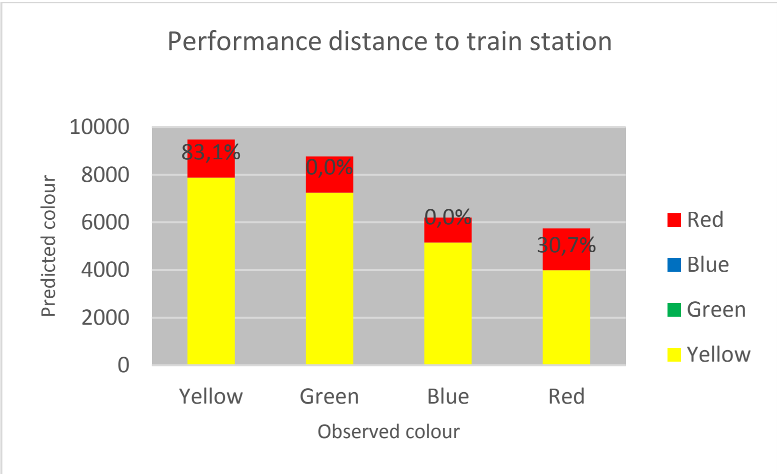

Figure 34: Performance mean distance to train station

8.3. Appendix C. Cross tables and performances of single variables

The cross tables show the number of people with a certain lifestyle and a certain type of household indicator. This can be age group, income group, housing type etc. Only the cross tables with the most significant differences are being discussed.

The count means how much respondents there are counted in a group. The expected count is the count that is expected when all the lifestyles would be equally distributed over the different types of the household indicator. The % within BSR_LIFESTYLE is the percentage of people from one lifestyle that has the certain type of household indicator. These values are used to make calculations.

A green or red bold value with indicates a positive or negative significant difference.

Significant differences are different when the 95% confidence intervals do not interfere. The confidence intervals are calculated with the following formulas: $U=p-2\sqrt{(p(1-p))/n}$ and $V=p+2\sqrt{(p(1-p))/n}$. Where U is the lower and V is the upper bound. p is the sampling fraction counts/total counts, n is the total counts of one lifestyle.

8.3.1. Gender

Cross table

Gender * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	Total
Gender	Man	Count	2517	3614	3245	2384	11760
		Expected Count	3707,9	3345,3	2469,0	2237,9	11760,0
		% within BSR_LIFESTYLE	**28,7%**	45,7%	55,6%	45,1%	42,3%
	Vrouw	Count	6246	4292	2590	2905	16033
		Expected Count	5055,1	4560,7	3366,0	3051,1	16033,0
		% within BSR_LIFESTYLE	**71,3%**	54,3%	44,4%	54,9%	57,7%
Total		Count	8763	7906	5835	5289	27793
		Expected Count	8763,0	7906,0	5835,0	5289,0	27793,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

*Table 4: Cross table Gender * BSR lifestyles*

Model Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	6032	2783	0	0	68,4%
Groen	4130	3955	0	0	48,9%
Blauw	2495	3464	0	0	0,0%
Rood	2767	2651	0	0	0,0%
Overall Percentage	54,5%	45,5%	0,0%	0,0%	**35,3%**

Table 5: Classification table performance gender model

8.3.2. Age

Cross Table

Age_Recoded * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	Total
Age_Recoded	t/m 24 jaar	Count	834	528	691	839	2892
		Expected Count	916,4	835,1	595,4	545,1	2892,0
		% within BSR_LIFESTYLE	8,8%	**6,1%**	11,2%	**14,8%**	9,6%
	25 t/m 34 jaar	Count	1694	1369	1478	1683	6224
		Expected Count	1972,1	1797,3	1281,4	1173,1	6224,0
		% within BSR_LIFESTYLE	17,8%	**15,8%**	23,9%	**29,7%**	20,7%
	35 t/m 44 jaar	Count	1721	1385	1306	1076	5488
		Expected Count	1738,9	1584,8	1129,9	1034,4	5488,0
		% within BSR_LIFESTYLE	18,1%	**16,0%**	21,1%	19,0%	18,3%
	45 t/m 54 jaar	Count	1715	1722	1162	849	5448
		Expected Count	1726,3	1573,3	1121,7	1026,8	5448,0
		% within BSR_LIFESTYLE	18,0%	19,8%	18,8%	**15,0%**	18,1%
	55 t/m 64 jaar	Count	1792	1808	861	719	5180
		Expected Count	1641,3	1495,9	1066,5	976,3	5180,0
		% within BSR_LIFESTYLE	18,8%	**20,8%**	13,9%	12,7%	17,2%
	65 t/m 74 jaar	Count	1278	1294	536	390	3498
		Expected Count	1108,4	1010,1	720,2	659,3	3498,0
		% within BSR_LIFESTYLE	13,4%	14,9%	8,7%	**6,9%**	11,6%
	75 jaar en ouder	Count	488	572	153	108	1321
		Expected Count	418,6	381,5	272,0	249,0	1321,0
		% within BSR_LIFESTYLE	5,1%	**6,6%**	2,5%	1,9%	4,4%
Total		Count	9522	8678	6187	5664	30051
		Expected Count	9522,0	8678,0	6187,0	5664,0	30051,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

*Table 6: Cross table Age * BSR lifestyle*

Performance

	Predicted				
Observed	Geel	Groen	Blauw	Rood	Percent Correct
Geel	4971	2789	0	1834	51,8%
Groen	4636	2916	0	1302	32,9%
Blauw	3360	1387	0	1560	0,0%
Rood	2644	1109	0	2000	34,8%
Overall Percentage	51,2%	26,9%	0,0%	21,9%	**32,4%**

Table 7: Classification table performance age model

8.3.3. Summarized family phase

Cross table

Gezinsfase_Kort*BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	
Gezinsfase_Kort	Jong, alleen	Count	984	1256	974	1757	4971
		Expected Count	1575,3	1437,3	1021,0	937,4	4971,0
		% within BSR_LIFESTYLE	**10,4%**	14,5%	15,9%	**31,2%**	16,6%
	Jong, samen	Count	499	567	674	691	2431
		Expected Count	770,4	702,9	499,3	458,4	2431,0
		% within BSR_LIFESTYLE	5,3%	6,6%	11,0%	12,3%	8,1%
	Gezin	Count	3677	2156	2495	1233	9561
		Expected Count	3029,8	2764,4	1963,7	1803,0	9561,0
		% within BSR_LIFESTYLE	38,9%	25,0%	40,7%	**21,9%**	32,0%
	Ouder, alleen	Count	2069	2399	640	1064	6172
		Expected Count	1955,9	1784,5	1267,7	1163,9	6172,0
		% within BSR_LIFESTYLE	21,9%	**27,8%**	**10,4%**	18,9%	20,7%
	Ouder, Samen	Count	1865	1990	1125	578	5558
		Expected Count	1761,3	1607,0	1141,6	1048,1	5558,0
		% within BSR_LIFESTYLE	19,7%	**23,0%**	18,3%	**10,3%**	18,6%
	Anders	Count	370	267	226	309	1172
		Expected Count	371,4	338,9	240,7	221,0	1172,0
		% within BSR_LIFESTYLE	3,9%	3,1%	3,7%	5,5%	3,9%
Total		Count	9464	8635	6134	5632	29865
		Expected Count	9464,0	8635,0	6134,0	5632,0	29865,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

*Table 8: Cross table summarized family phase * BSR lifestyle*

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	3945	4542	0	1050	41,4%
Groen	2451	5012	0	1344	56,9%
Blauw	2687	2282	0	1285	0,0%
Rood	1662	1958	0	2093	36,6%
Overall Percentage	35,4%	45,5%	0,0%	19,0%	**36,5%**

Table 9: Classification table performance summarized family phase model

8.3.4. Extended family phase

Cross table

Gezinsfase_Lang * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	Total
Gezinsfase_Lang	Alleen t/m 34 jaar	Count	673	728	713	1330	3444
		Expected Count	1090,9	993,7	710,8	648,7	3444,0
		% within BSR_LIFESTYLE	7,0%	8,4%	11,4%	**23,4%**	11,4%
	Samen t/m 34 jaar	Count	438	414	549	590	1991
		Expected Count	630,6	574,5	410,9	375,0	1991,0
		% within BSR_LIFESTYLE	4,6%	4,8%	8,8%	**10,4%**	6,6%
	Gezin t/m 34 jaar	Count	1415	753	908	600	3676
		Expected Count	1164,3	1060,6	758,7	692,3	3676,0
		% within BSR_LIFESTYLE	14,8%	**8,6%**	14,6%	10,6%	12,2%
	Alleen 35 t/m 54 jaar	Count	773	1198	489	924	3384
		Expected Count	1071,9	976,4	698,4	637,4	3384,0
		% within BSR_LIFESTYLE	8,1%	13,8%	7,8%	**16,2%**	11,2%
	Samen 35 t/m 54 jaar	Count	456	612	424	326	1818
		Expected Count	575,8	524,5	375,2	342,4	1818,0
		% within BSR_LIFESTYLE	4,8%	7,0%	6,8%	5,7%	6,0%
	Gezin 35 t/m 54 jaar	Count	2204	1292	1552	672	5720
		Expected Count	1811,8	1650,4	1180,6	1077,3	5720,0
		% within BSR_LIFESTYLE	23,0%	14,8%	24,9%	**11,8%**	18,9%
	Alleen 55 t/m 64 jaar	Count	711	784	222	371	2088
		Expected Count	661,4	602,4	430,9	393,3	2088,0
		% within BSR_LIFESTYLE	7,4%	**9,0%**	**3,6%**	6,5%	6,9%
	Samen 55 t/m 64 jaar	Count	805	801	460	254	2320
		Expected Count	734,8	669,4	478,8	437,0	2320,0
		% within BSR_LIFESTYLE	8,4%	9,2%	7,4%	**4,5%**	7,7%
	Gezin 55 t/m 64 jaar	Count	277	220	178	94	769
		Expected Count	243,6	221,9	158,7	144,8	769,0
		% within BSR_LIFESTYLE	2,9%	2,5%	2,9%	**1,7%**	2,5%
	Alleen 65 t/m 74 jaar	Count	600	596	145	219	1560
		Expected Count	494,1	450,1	322,0	293,8	1560,0
		% within BSR_LIFESTYLE	6,3%	6,8%	2,3%	3,9%	5,2%
	Samen 65 t/m 74 jaar	Count	640	663	377	155	1835
			581,2	529,4	378,7	345,6	1835,0
		% within BSR_LIFESTYLE	6,7%	7,6%	6,0%	**2,7%**	6,1%
	Samen 75 jaar en ouder	Count	103	134	49	27	313
		Expected Count	99,1	90,3	64,6	59,0	313,0
		% within BSR_LIFESTYLE	1,1%	1,5%	0,8%	0,5%	1,0%
	Anders	Count	469	517	166	125	1277

	Expected Count	404,5	368,4	263,6	240,5	1277,0
	% within BSR_LIFESTYLE	4,9%	5,9%	2,7%	2,2%	4,2%
Total	Count	9564	8712	6232	5687	30195
	Expected Count	9564,0	8712,0	6232,0	5687,0	30195,0
	% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 10: Cross table extended family phase

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	6507	2370	0	765	67,5%
Groen	4974	3158	0	759	35,5%
Blauw	4030	1404	0	912	0,0%
Rood	2427	1864	0	1485	25,7%
Overall Percentage	58,5%	28,7%	0,0%	12,8%	**36,4%**

Figure 35: Classification table Performance extended family phase model

8.3.5. Income

Cross table

Inkomen * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	Total
Inkomen	Beneden modaal	Count	3550	2792	1124	1739	9205
		Expected Count	2915,8	2655,5	1899,2	1734,5	9205,0
		% within BSR_LIFESTYLE	37,1%	32,1%	18,0%	30,6%	30,5%
	Modaal	Count	1450	1482	647	978	4557
		Expected Count	1443,5	1314,6	940,2	858,7	4557,0
		% within BSR_LIFESTYLE	15,2%	17,0%	10,4%	17,2%	15,1%
	1-2x modaal	Count	1516	1459	1062	996	5033
		Expected Count	1594,3	1451,9	1038,4	948,4	5033,0
		% within BSR_LIFESTYLE	15,8%	16,7%	17,0%	17,5%	16,7%
	2x modaal	Count	605	613	704	483	2405
		Expected Count	761,8	693,8	496,2	453,2	2405,0
		% within BSR_LIFESTYLE	6,3%	7,0%	11,3%	8,5%	8,0%
	>2x modaal	Count	315	586	1264	562	2727
		Expected Count	863,8	786,7	562,6	513,9	2727,0
		% within BSR_LIFESTYLE	3,3%	6,7%	20,3%	9,9%	9,0%
	Onbekend; Wil ik niet zeggen	Count	2129	1779	1429	932	6269
		Expected Count	1985,8	1808,5	1293,4	1181,3	6269,0
		% within BSR_LIFESTYLE	22,3%	20,4%	22,9%	16,4%	20,8%
Total		Count	9565	8711	6230	5690	30196
		Expected Count	9565,0	8711,0	6230,0	5690,0	30196,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 11: Cross table income

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	8553	0	1089	0	88,7%
Groen	7413	0	1478	0	0,0%
Blauw	3977	0	2369	0	37,3%
Rood	4188	0	1588	0	0,0%
Overall Percentage	78,7%	0,0%	21,3%	0,0%	**35,6%**

Table 12: Classification table performance income

8.3.6. Employment status

Cross table

Arbeidssitatie_Respondent * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				Total
			Geel	Groen	Blauw	Rood	
Arbeidssitatie_Respondent	Onbekend;	Count	1056	971	482	481	2990
		Expected Count	947,2	862,5	616,9	563,4	2990,0
		% within BSR_LIFESTYLE	11,0%	11,1%	7,7%	8,5%	9,9%
	Zelfstandig	Count	234	301	666	584	1785
		Expected Count	565,5	514,9	368,3	336,3	1785,0
		% within BSR_LIFESTYLE	2,4%	3,5%	10,7%	10,3%	5,9%
	Loondienst	Count	3923	3718	3343	2950	13934
		Expected Count	4414,2	4019,6	2874,7	2625,5	13934,0
		% within BSR_LIFESTYLE	41,0%	42,7%	53,7%	51,9%	46,1%
	Huisman/vrouw	Count	1006	485	121	90	1702
		Expected Count	539,2	491,0	351,1	320,7	1702,0
		% within BSR_LIFESTYLE	10,5%	5,6%	1,9%	1,6%	5,6%
	Student/Scholier	Count	370	254	388	554	1566
		Expected Count	496,1	451,8	323,1	295,1	1566,0
		% within BSR_LIFESTYLE	3,9%	2,9%	6,2%	9,7%	5,2%
	Gepensioneerd; AOW; VUT; Wachttijdgeld	Count	1612	1867	763	507	4749
		Expected Count	1504,5	1370,0	979,7	894,8	4749,0
		% within BSR_LIFESTYLE	16,9%	21,4%	12,2%	8,9%	15,7%
	werkloos; bijstand; werkzoekend	Count	636	546	240	288	1710
		Expected Count	541,7	493,3	352,8	322,2	1710,0
		% within BSR_LIFESTYLE	6,6%	6,3%	3,9%	5,1%	5,7%
	Arbeidsongeschikt	Count	450	375	107	114	1046
		Expected Count	331,4	301,7	215,8	197,1	1046,0
		% within BSR_LIFESTYLE	4,7%	4,3%	1,7%	2,0%	3,5%
	Anders	Count	278	193	119	121	711
		Expected Count	225,2	205,1	146,7	134,0	711,0
		% within BSR_LIFESTYLE	2,9%	2,2%	1,9%	2,1%	2,4%
Total		Count	9565	8710	6229	5689	30193
		Expected Count	9565,0	8710,0	6229,0	5689,0	30193,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 13: Cross table employment status

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	7243	1897	251	251	75,1%
Groen	6245	2160	309	177	24,3%
Blauw	4392	1000	687	267	10,8%
Rood	4103	606	669	398	6,9%
Overall Percentage	71,7%	18,5%	6,3%	3,6%	**34,2%**

Table 14: Classification table performance employment status

8.3.7. Education

Cross table

Opleiding_Responden * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	Total
Opleiding_Responden	Onbekend	Count	102	65	32	28	227
		Expected Count	71,4	64,5	47,8	43,4	227,0
		% within BSR_LIFESTYLE	1,2%	0,8%	0,6%	0,5%	0,8%
	Lagere school; basisonderwijs	Count	580	385	110	98	1173
		Expected Count	368,7	333,2	247,0	224,1	1173,0
		% within BSR_LIFESTYLE	**6,7%**	4,9%	1,9%	1,9%	4,3%
	Lager beroepsonderwijs	Count	1738	1066	288	219	3311
		Expected Count	1040,7	940,5	697,2	632,5	3311,0
		% within BSR_LIFESTYLE	**20,0%**	13,6%	5,0%	4,2%	12,0%
	MAVO; MMS; MULO; ULO	Count	1279	1098	413	301	3091
		Expected Count	971,6	878,0	650,9	590,5	3091,0
		% within BSR_LIFESTYLE	14,8%	14,0%	7,1%	5,7%	11,2%
	MBO	Count	2733	2082	1315	991	7121
		Expected Count	2238,3	2022,8	1499,5	1360,4	7121,0
		% within BSR_LIFESTYLE	**31,5%**	26,6%	22,6%	**18,8%**	25,8%
	HAVO; VWO	Count	707	802	720	554	2783
		Expected Count	874,8	790,5	586,0	531,7	2783,0
		% within BSR_LIFESTYLE	8,2%	10,2%	**12,4%**	10,5%	10,1%
	HBO	Count	1286	1699	1858	1873	6716
		Expected Count	2111,0	1907,7	1414,2	1283,0	6716,0
		% within BSR_LIFESTYLE	**14,8%**	21,7%	32,0%	**35,5%**	24,3%
	WO	Count	246	639	1073	1206	3164
		Expected Count	994,5	898,8	666,3	604,4	3164,0
		% within BSR_LIFESTYLE	**2,8%**	8,2%	18,5%	**22,9%**	11,5%
Total		Count	8671	7836	5809	5270	27586
		Expected Count	8671,0	7836,0	5809,0	5270,0	27586,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 15: Cross table education

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	6508	682	0	1529	74,6%
Groen	4810	849	0	2364	10,6%
Blauw	2094	749	0	3087	0,0%
Rood	1507	536	0	3354	62,1%
Overall Percentage	53,2%	10,0%	0,0%	36,8%	**38,2%**

Table 16: Classification table performance education

8.3.8. Car ownership

Cross table

Car * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				Total
			Geel	Groen	Blauw	Rood	
Car	Geen auto	Count	1044	891	286	918	3139
		Expected Count	972,4	892,1	689,1	585,4	3139,0
		% within BSR_LIFESTYLE	27,0%	25,1%	**10,4%**	**39,5%**	25,2%
	1 Auto	Count	2184	2082	1429	1138	6833
		Expected Count	2116,7	1941,9	1500,1	1274,3	6833,0
		% within BSR_LIFESTYLE	56,6%	58,8%	52,2%	48,9%	54,8%
	2 Auto's	Count	583	519	919	238	2259
		Expected Count	699,8	642,0	495,9	421,3	2259,0
		% within BSR_LIFESTYLE	15,1%	14,6%	**33,6%**	**10,2%**	18,1%
	Meer dan 2 auto's	Count	51	51	103	31	236
		Expected Count	73,1	67,1	51,8	44,0	236,0
		% within BSR_LIFESTYLE	1,3%	1,4%	**3,8%**	1,3%	1,9%
Total		Count	3862	3543	2737	2325	12467
		Expected Count	3862,0	3543,0	2737,0	2325,0	12467,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 17: Cross table car ownership

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	3191	0	661	0	82,8%
Groen	3017	0	653	0	0,0%
Blauw	1693	0	1182	0	41,1%
Rood	2116	0	384	0	0,0%
Overall Percentage	77,7%	0,0%	22,3%	0,0%	**33,9%**

Table 18: Classification table performance car ownership

8.3.9. Land of birth

Cross table

Gebland_Respondent * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				Total
			Geel	Groen	Blauw	Rood	
Gebland_Respondent	Nederland	Count	7657	7144	5105	4789	24695
		Expected Count	7701,8	7056,7	5199,8	4736,7	24695
		% within BSR_LIFESTYLE	87,40%	89,00%	86,30%	88,80%	87,90%
	Turkije	Count	177	77	87	45	386
		Expected Count	120,4	110,3	81,3	74	386
		% within BSR_LIFESTYLE	2,00%	1,00%	1,50%	0,80%	1,40%
	Marokko	Count	141	71	70	32	314
		Expected Count	97,9	89,7	66,1	60,2	314
		% within BSR_LIFESTYLE	1,60%	0,90%	1,20%	0,60%	1,10%
	Suriname	Count	211	188	160	106	665
		Expected Count	207,4	190	140	127,6	665
		% within BSR_LIFESTYLE	2,40%	2,30%	2,70%	2,00%	2,40%
	Nederlandse Antillen; Aruba	Count	118	103	85	58	364
		Expected Count	113,5	104	76,6	69,8	364
		% within BSR_LIFESTYLE	1,30%	1,30%	1,40%	1,10%	1,30%
	Indonesië	Count	59	96	53	47	255
		Expected Count	79,5	72,9	53,7	48,9	255
		% within BSR_LIFESTYLE	0,70%	1,20%	0,90%	0,90%	0,90%
	Japan	Count	2	0	1	1	4
		Expected Count	1,2	1,1	0,8	0,8	4
		% within BSR_LIFESTYLE	0,00%	0,00%	0,00%	0,00%	0,00%
	Ander land in Europa, Noord-Amerika of Oceanië	Count	119	135	119	148	521
		Expected Count	162,5	148,9	109,7	99,9	521
		% within BSR_LIFESTYLE	1,40%	1,70%	2,00%	2,70%	1,90%
	Ander land	Count	278	216	237	163	894
		Expected Count	278,8	255,5	188,2	171,5	894
		% within BSR_LIFESTYLE	3,20%	2,70%	4,00%	3,00%	3,20%
Total		Count	8762	8030	5917	5390	28101
		Expected Count	8762	8030	5917	5390	28101
		% within BSR_LIFESTYLE	100,00%	100,00%	100,00%	100,00%	100,00%

Table 19: Cross table land of birth respondents

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	8586	59	0	119	98,0%
Groen	7799	96	0	135	1,2%
Blauw	5745	53	0	119	0,0%
Rood	5195	47	0	148	2,7%
Overall Percentage	97,2%	0,9%	0,0%	1,9%	**31,4%**

Table 20: Classification table performance land of birth respondents

8.3.10. Degree of urbanism

Cross table

Stedelijkheid * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	Total
Omgevingsadressendichtheid	< 500 adressen	Count	195	166	127	35	523
		Expected Count	163,1	151,7	107,1	101,0	523,0
		% within BSR_LIFESTYLE	2,1%	2,0%	2,1%	**0,6%**	1,8%
	500 - 1000 adressen	Count	418	367	349	103	1237
		Expected Count	385,8	358,9	253,3	239,0	1237,0
		% within BSR_LIFESTYLE	4,6%	4,3%	**5,8%**	**1,8%**	4,2%
	1000 - 1500 adressen	Count	865	724	616	242	2447
		Expected Count	763,3	709,9	501,1	472,8	2447,0
		% within BSR_LIFESTYLE	9,5%	8,5%	10,3%	**4,3%**	8,4%
	1500 - 2000 adressen	Count	1401	1309	1009	521	4240
		Expected Count	1322,5	1230,0	868,3	819,2	4240,0
		% within BSR_LIFESTYLE	15,3%	15,4%	16,8%	**9,2%**	14,5%
	2000 - 2500 adressen	Count	1297	1305	877	457	3936
		Expected Count	1227,7	1141,8	806,0	760,5	3936,0
		% within BSR_LIFESTYLE	14,2%	15,4%	14,6%	**8,1%**	13,4%
	2500- 3000 adressen	Count	1969	1880	1067	630	5546
		Expected Count	1729,9	1608,9	1135,7	1071,5	5546,0
		% within BSR_LIFESTYLE	21,6%	22,1%	17,8%	**11,1%**	18,9%
	3000 - 3500 adressen	Count	799	836	514	587	2736
		Expected Count	853,4	793,7	560,3	528,6	2736,0
		% within BSR_LIFESTYLE	8,7%	9,8%	8,6%	10,4%	9,3%
	3500 - 4000 adressen	Count	494	395	181	178	1248
		Expected Count	389,3	362,0	255,6	241,1	1248,0
		% within BSR_LIFESTYLE	5,4%	4,6%	3,0%	3,1%	4,3%
	4000 - 4500 adressen	Count	64	83	78	102	327
		Expected Count	102,0	94,9	67,0	63,2	327,0

		% within BSR_LIFESTYLE	0,7%	1,0%	1,3%	1,8%	1,1%
	4500 - 5000 adressen	Count	378	345	301	554	1578
		Expected Count	492,2	457,8	323,1	304,9	1578,0
		% within BSR_LIFESTYLE	4,1%	4,1%	5,0%	9,8%	5,4%
	5000 - 5500 adressen	Count	270	245	161	333	1009
		Expected Count	314,7	292,7	206,6	194,9	1009,0
		% within BSR_LIFESTYLE	3,0%	2,9%	2,7%	5,9%	3,4%
	5500 - 6000 adressen	Count	193	153	136	205	687
		Expected Count	214,3	199,3	140,7	132,7	687,0
		% within BSR_LIFESTYLE	2,1%	1,8%	2,3%	3,6%	2,3%
	6000 - 6500 adressen	Count	401	300	239	637	1577
		Expected Count	491,9	457,5	322,9	304,7	1577,0
		% within BSR_LIFESTYLE	4,4%	3,5%	4,0%	11,3%	5,4%
	> 6500 adressen	Count	392	389	343	1075	2199
		Expected Count	685,9	637,9	450,3	424,9	2199,0
		% within BSR_LIFESTYLE	4,3%	4,6%	5,7%	19,0%	7,5%
Total		Count	9136	8497	5998	5659	29290
		Expected Count	9136,0	8497,0	5998,0	5659,0	29290,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 21: Cross table degree of urbanism

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	5342	2096	0	1698	58,5%
Groen	4841	2141	0	1515	25,2%
Blauw	3349	1391	0	1258	0,0%
Rood	1709	1044	0	2906	51,4%
Overall Percentage	52,0%	22,8%	0,0%	25,2%	**35,5%**

Table 22: Classification table performance degree of urbanism

8.3.11. Mean distance main road

Cross table

Afst_Hoofdweg * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				
			Geel	Groen	Blauw	Rood	Total
Afst_Hoofdweg	< 1 km	Count	1447	1261	900	776	4384
		Expected Count	1377,0	1272,6	900,4	833,9	4384,0
		% within BSR_LIFESTYLE	15,3%	14,4%	14,5%	13,5%	14,5%
	1 - 2 km	Count	4305	4302	2936	2243	13786
		Expected Count	4330,3	4001,8	2831,6	2622,4	13786,0
		% within BSR_LIFESTYLE	45,4%	49,1%	47,4%	39,1%	45,7%
	2 - 3 km	Count	2450	2181	1634	1898	8163
		Expected Count	2564,0	2369,6	1676,6	1552,8	8163,0
		% within BSR_LIFESTYLE	25,8%	24,9%	26,4%	33,1%	27,0%
	3 - 4 km	Count	1090	866	589	726	3271
		Expected Count	1027,4	949,5	671,8	622,2	3271,0
		% within BSR_LIFESTYLE	11,5%	9,9%	9,5%	12,6%	10,8%
	4 - 5 km	Count	117	101	89	88	395
		Expected Count	124,1	114,7	81,1	75,1	395,0
		% within BSR_LIFESTYLE	1,2%	1,2%	1,4%	1,5%	1,3%
	> 5 km	Count	71	50	51	10	182
		Expected Count	57,2	52,8	37,4	34,6	182,0
		% within BSR_LIFESTYLE	0,7%	0,6%	0,8%	0,2%	0,6%
Total		Count	9480	8761	6199	5741	30181
		Expected Count	9480,0	8761,0	6199,0	5741,0	30181,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 23: Cross table mean distance to main road

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	9480	0	0	0	100,0%
Groen	8761	0	0	0	0,0%
Blauw	6199	0	0	0	0,0%
Rood	5741	0	0	0	0,0%
Overall Percentage	100,0%	0,0%	0,0%	0,0%	31,4%

Table 24: Classification table performance mean distance to main road

8.3.12. Mean distance train station

Cross table

Gebland_Respondent * BSR_LIFESTYLE Crosstabulation

			BSR_LIFESTYLE				Total
			Geel	Groen	Blauw	Rood	
Distance_Train_Station	< 1,5 km (walking distance)	Count	1606	1513	1051	1762	5932
		Expected Count	1863,3	1722,0	1218,4	1128,4	5932,0
		% within BSR_LIFESTYLE	16,9%	17,3%	17,0%	30,7%	19,7%
	1,5 - 3 km	Count	3178	3000	2188	2285	10651
		Expected Count	3345,5	3091,8	2187,7	2026,0	10651,0
		% within BSR_LIFESTYLE	33,5%	34,2%	35,3%	39,8%	35,3%
	3 - 6 km	Count	2298	2154	1588	1080	7120
		Expected Count	2236,4	2066,8	1462,4	1354,4	7120,0
		% within BSR_LIFESTYLE	24,2%	24,6%	25,6%	18,8%	23,6%
	6 - 9 km	Count	686	601	431	175	1893
		Expected Count	594,6	549,5	388,8	360,1	1893,0
		% within BSR_LIFESTYLE	7,2%	6,9%	7,0%	3,0%	6,3%
	9 - 12 km	Count	546	501	295	152	1494
		Expected Count	469,3	433,7	306,9	284,2	1494,0
		% within BSR_LIFESTYLE	5,8%	5,7%	4,8%	2,6%	5,0%
	12 - 15 km	Count	653	565	335	163	1716
		Expected Count	539,0	498,1	352,5	326,4	1716,0
		% within BSR_LIFESTYLE	6,9%	6,4%	5,4%	2,8%	5,7%
	15 - 18 km	Count	395	314	222	90	1021
		Expected Count	320,7	296,4	209,7	194,2	1021,0
		% within BSR_LIFESTYLE	4,2%	3,6%	3,6%	1,6%	3,4%
	> 18 km	Count	118	113	89	34	354
		Expected Count	111,2	102,8	72,7	67,3	354,0
		% within BSR_LIFESTYLE	1,2%	1,3%	1,4%	0,6%	1,2%
Total		Count	9480	8761	6199	5741	30181
		Expected Count	9480,0	8761,0	6199,0	5741,0	30181,0
		% within BSR_LIFESTYLE	100,0%	100,0%	100,0%	100,0%	100,0%

Table 25: Cross table mean distance to train station

Performance

Observed	Predicted				
	Geel	Groen	Blauw	Rood	Percent Correct
Geel	7874	0	0	1606	83,1%
Groen	7248	0	0	1513	0,0%
Blauw	5148	0	0	1051	0,0%
Rood	3979	0	0	1762	30,7%
Overall Percentage	80,3%	0,0%	0,0%	19,7%	**31,9%**

Table 26: Classification table performance mean distance to train station

8.4. Appendix D. Parameter estimates of the final model

The Red lifestyles is used as reference category

BSR_LIFESTYLE		β	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Yellow 'Harmony'	**Intercept**	0,629	0,341	3,394	1	0,065			
Gender	**[Man]**	-0,718	0,043	277,044	1	0,000	0,488	0,448	0,531
	[Woman]	0b			0				
Age	**[Till 24 years]**	-0,353	0,186	3,596	1	0,058	0,702	0,488	1,012
	[25 till 34 years]	-0,208	0,171	1,483	1	0,223	0,812	0,581	1,135
	[35 till 44 years]	-0,369	0,167	4,872	1	0,027	0,691	0,498	0,959
	[45 till 54 years]	-0,373	0,154	5,887	1	0,015	0,688	0,509	0,931
	[55 till 64 years]	-0,288	0,142	4,094	1	0,043	0,750	0,568	0,991
	[65 till 74 years]	-0,227	0,134	2,876	1	0,090	0,797	0,614	1,036
	[74 years and older]	0b			0				
Income	**[TILL 0,75X MODAL]**	0,200	0,071	7,942	1	0,005	1,222	1,063	1,405
	[0,75 - 1,1X MODAL]	0,048	0,065	0,554	1	0,457	1,049	0,924	1,191
	[1,1 - 1,7X MODAL]	-0,170	0,066	6,662	1	0,010	0,843	0,741	0,960
	[1,7 - 2,1X MODAL]	-0,482	0,081	35,251	1	0,000	0,617	0,526	0,724
	[MORE THAN 2,1X MODAL]	-1,040	0,084	153,732	1	0,000	0,353	0,300	0,417
	[UNNOWN; DON'T WANT TO SAY]	0b			0				
Family phase	**[Young, alone]**	-0,511	0,120	18,022	1	0,000	0,600	0,474	0,760
	[Young, together]	-0,028	0,122	0,052	1	0,820	0,973	0,765	1,236
	[Family]	0,810	0,108	56,543	1	0,000	2,247	1,820	2,776
	[Older, alone]	-0,119	0,125	0,899	1	0,343	0,888	0,695	1,135
	[Older, together]	0,428	0,123	12,131	1	0,000	1,534	1,206	1,952
	[Different]	0b			0				
Employment status	**[Working]**	-0,632	0,180	12,388	1	0,000	0,531	0,374	0,756
	[Housewife]	0,303	0,207	2,142	1	0,143	1,354	0,902	2,033
	[Scholar/Student]	-0,846	0,210	16,299	1	0,000	0,429	0,285	0,647
	[Unemployed]	-0,267	0,200	1,796	1	0,180	0,765	0,518	1,132
	[Incapacitated]	0,198	0,211	0,880	1	0,348	1,219	0,806	1,845
	[Retired]	-0,172	0,188	0,838	1	0,360	0,842	0,583	1,217
	[Different]	-0,532	0,216	6,052	1	0,014	0,587	0,384	0,897
	[Unknown]	0b			0				
Education	**[PRIMARY SCHOOL]**	0,821	0,286	8,260	1	0,004	2,272	1,298	3,976
	[MAVO, VMBO, VGLO, LAVO, MULO]	0,448	0,258	3,032	1	0,082	1,566	0,945	2,594
	[MBO, HAVO, VWO]	-0,323	0,255	1,600	1	0,206	0,724	0,439	1,194
	[HBO, WO]	-1,441	0,256	31,757	1	0,000	0,237	0,143	0,391
	[UNNOWN]	0b			0				
Degree of urbanism	**[Degree of urbanism=1] (<500 addresses)**	2,469	0,220	126,412	1	0,000	11,807	7,678	18,157
	[Degree of urbanism=2]	2,084	0,139	224,375	1	0,000	8,039	6,120	10,559

	[Degree of urbanism=3]	1,978	0,109	330,555	1	0,000	7,231	5,842	8,950
	[Degree of urbanism=4]	1,731	0,091	358,540	1	0,000	5,649	4,722	6,758
	[Degree of urbanism=5]	1,724	0,094	337,793	1	0,000	5,607	4,665	6,738
	[Degree of urbanism=6]	1,759	0,087	409,997	1	0,000	5,807	4,898	6,885
	[Degree of urbanism=7]	1,058	0,093	128,091	1	0,000	2,880	2,398	3,459
	[Degree of urbanism=8]	1,505	0,120	158,169	1	0,000	4,505	3,563	5,696
	[Degree of urbanism=9]	0,535	0,198	7,272	1	0,007	1,707	1,157	2,518
	[Degree of urbanism=10]	0,568	0,104	29,811	1	0,000	1,764	1,439	2,163
	[Degree of urbanism=11]	0,634	0,118	28,760	1	0,000	1,886	1,496	2,378
	[Degree of urbanism=12]	0,683	0,136	25,335	1	0,000	1,980	1,517	2,583
	[Degree of urbanism=13]	0,397	0,102	15,129	1	0,000	1,488	1,218	1,817
	[Degree of urbanism=14] (>6500 addresses)	0b			0				
Green 'Safety'	**Intercept**	0,094	0,350	0,072	1	0,789			
Gender	**[Man]**	-0,097	0,041	5,491	1	0,019	0,908	0,838	0,984
	[Woman]	0b			0				
Age	**[Till 24 years]**	-1,144	0,187	37,309	1	0,000	0,319	0,221	0,460
	[25 till 34 years]	-0,779	0,167	21,687	1	0,000	0,459	0,331	0,637
	[35 till 44 years]	-0,630	0,163	14,974	1	0,000	0,532	0,387	0,733
	[45 till 54 years]	-0,409	0,148	7,662	1	0,006	0,664	0,497	0,887
	[55 till 64 years]	-0,470	0,137	11,860	1	0,001	0,625	0,478	0,817
	[65 till 74 years]	-0,472	0,128	13,513	1	0,000	0,624	0,485	0,802
	[74 years and older]	0b			0				
Income	[TILL 0,75X MODAL]	0,039	0,071	0,301	1	0,583	1,040	0,904	1,196
	[0,75 - 1,1X MODAL]	-0,013	0,064	0,042	1	0,838	0,987	0,871	1,119
	[1,1 - 1,7X MODAL]	-0,177	0,065	7,446	1	0,006	0,838	0,738	0,951
	[1,7 - 2,1X MODAL]	-0,356	0,078	20,963	1	0,000	0,700	0,601	0,816
	[MORE THAN 2,1X MODAL]	-0,617	0,074	68,845	1	0,000	0,539	0,466	0,624
	[UNNOWN; DON'T WANT TO SAY]	0b			0				
Family phase	[Young, alone]	0,070	0,122	0,333	1	0,564	1,073	0,845	1,361
	[Young, together]	0,325	0,124	6,825	1	0,009	1,384	1,084	1,765
	[Family]	0,439	0,112	15,224	1	0,000	1,551	1,244	1,933
	[Older, alone]	0,312	0,127	6,060	1	0,014	1,366	1,066	1,751
	[Older, together]	0,531	0,125	18,003	1	0,000	1,701	1,331	2,174
	[Different]	0b			0				
Employment status	**[Working]**	-0,382	0,184	4,315	1	0,038	0,683	0,476	0,979
	[Housewife]	0,356	0,213	2,799	1	0,094	1,428	0,941	2,167
	[Scholar/Student]	-0,392	0,216	3,277	1	0,070	0,676	0,442	1,033
	[Unemployed]	0,088	0,203	0,191	1	0,662	1,093	0,734	1,626
	[Incapacitated]	0,271	0,215	1,594	1	0,207	1,311	0,861	1,998
	[Retired]	0,016	0,191	0,007	1	0,931	1,017	0,699	1,478
	[Different]	-0,403	0,222	3,298	1	0,069	0,668	0,433	1,032
	[Unknown]	0b			0				
Education	**[PRIMARY SCHOOL]**	0,722	0,296	5,942	1	0,015	2,058	1,152	3,676

	[MAVO, VMBO, VGLO, LAVO, MULO]	0,410	0,268	2,348	1	0,125	1,507	0,892	2,548
	[MBO, HAVO, VWO]	-0,104	0,265	0,153	1	0,695	0,901	0,536	1,516
	[HBO, WO]	-0,781	0,265	8,647	1	0,003	0,458	0,272	0,771
	[UNNOWN]	0b			0				
Degree of urbanism	[Degree of urbanism=1] (<500 addresses)	2,341	0,217	116,294	1	0,000	10,387	6,788	15,894
	[Degree of urbanism=2]	1,947	0,136	203,970	1	0,000	7,006	5,363	9,151
	[Degree of urbanism=3]	1,864	0,106	311,035	1	0,000	6,448	5,242	7,932
	[Degree of urbanism=4]	1,694	0,087	377,805	1	0,000	5,439	4,585	6,452
	[Degree of urbanism=5]	1,781	0,089	397,645	1	0,000	5,936	4,983	7,071
	[Degree of urbanism=6]	1,741	0,083	443,014	1	0,000	5,701	4,848	6,704
	[Degree of urbanism=7]	1,136	0,088	167,814	1	0,000	3,115	2,623	3,700
	[Degree of urbanism=8]	1,375	0,117	137,276	1	0,000	3,954	3,141	4,976
	[Degree of urbanism=9]	0,633	0,178	12,592	1	0,000	1,884	1,328	2,672
	[Degree of urbanism=10]	0,512	0,099	26,909	1	0,000	1,668	1,375	2,025
	[Degree of urbanism=11]	0,645	0,113	32,655	1	0,000	1,906	1,528	2,378
	[Degree of urbanism=12]	0,642	0,132	23,669	1	0,000	1,900	1,467	2,462
	[Degree of urbanism=13]	0,215	0,100	4,653	1	0,031	1,240	1,020	1,508
	[Degree of urbanism=14] (>6500 addresses)	0b			0				
Blue 'Control'	**Intercept**	-0,518	0,388	1,784	1	0,182			
Gender	[Man]	0,239	0,043	31,159	1	0,000	1,270	1,168	1,382
	[Woman]	0b			0				
Age	[Till 24 years]	-0,114	0,199	0,325	1	0,568	0,892	0,604	1,319
	[25 till 34 years]	-0,057	0,183	0,096	1	0,756	0,945	0,659	1,353
	[35 till 44 years]	-0,143	0,180	0,630	1	0,427	0,867	0,609	1,234
	[45 till 54 years]	-0,153	0,168	0,830	1	0,362	0,858	0,617	1,193
	[55 till 64 years]	-0,247	0,156	2,508	1	0,113	0,781	0,575	1,061
	[65 till 74 years]	-0,104	0,147	0,497	1	0,481	0,901	0,675	1,203
	[74 years and older]	0b			0				
Income	[TILL 0,75X MODAL]	-0,521	0,079	43,308	1	0,000	0,594	0,509	0,694
	[0,75 - 1,1X MODAL]	-0,612	0,070	76,141	1	0,000	0,542	0,473	0,622
	[1,1 - 1,7X MODAL]	-0,446	0,067	43,749	1	0,000	0,640	0,561	0,731
	[1,7 - 2,1X MODAL]	-0,346	0,077	20,141	1	0,000	0,707	0,608	0,823
	[MORE THAN 2,1X MODAL]	0,048	0,068	0,487	1	0,485	1,049	0,918	1,199
	[UNNOWN; DON'T WANT TO SAY]	0b			0				
Family phase	[Young, alone]	-0,184	0,121	2,301	1	0,129	0,832	0,656	1,055
	[Young, together]	0,056	0,121	0,213	1	0,645	1,057	0,835	1,339
	[Family]	0,586	0,112	27,348	1	0,000	1,796	1,442	2,237
	[Older, alone]	-0,232	0,134	3,019	1	0,082	0,793	0,610	1,030
	[Older, together]	0,313	0,128	5,999	1	0,014	1,368	1,065	1,757
	[Different]	0b			0				
Employment status	[Working]	-0,318	0,207	2,353	1	0,125	0,728	0,485	1,092
	[Housewife]	-0,304	0,245	1,535	1	0,215	0,738	0,457	1,193
	[Scholar/Student]	-0,243	0,231	1,110	1	0,292	0,784	0,499	1,233

	[Unemployed]	-0,167	0,231	0,521	1	0,470	0,847	0,539	1,330
	[Incapacitated]	-0,265	0,251	1,118	1	0,290	0,767	0,469	1,254
	[Retired]	-0,170	0,216	0,618	1	0,432	0,844	0,552	1,289
	[Different]	-0,287	0,247	1,357	1	0,244	0,750	0,463	1,217
	[Unknown]	0b			0				
Education	[PRIMARY SCHOOL]	0,209	0,339	0,379	1	0,538	1,232	0,634	2,395
	[MAVO, VMBO, VGLO, LAVO, MULO]	0,123	0,300	0,168	1	0,682	1,131	0,628	2,034
	[MBO, HAVO, VWO]	0,038	0,296	0,016	1	0,899	1,038	0,581	1,855
	[HBO, WO]	-0,249	0,296	0,709	1	0,400	0,780	0,437	1,392
	[UNNOWN]	0b			0				
Degree of urbanism	**[Degree of urbanism=1] (<500 addresses)**	2,254	0,219	105,837	1	0,000	9,529	6,202	14,641
	[Degree of urbanism=2]	2,005	0,135	220,004	1	0,000	7,428	5,699	9,682
	[Degree of urbanism=3]	1,828	0,106	296,187	1	0,000	6,221	5,052	7,661
	[Degree of urbanism=4]	1,596	0,088	327,235	1	0,000	4,935	4,151	5,867
	[Degree of urbanism=5]	1,588	0,092	300,781	1	0,000	4,896	4,091	5,858
	[Degree of urbanism=6]	1,462	0,086	291,842	1	0,000	4,316	3,650	5,105
	[Degree of urbanism=7]	0,878	0,091	92,307	1	0,000	2,406	2,012	2,878
	[Degree of urbanism=8]	1,012	0,129	61,519	1	0,000	2,752	2,137	3,544
	[Degree of urbanism=9]	0,826	0,172	23,148	1	0,000	2,283	1,631	3,196
	[Degree of urbanism=10]	0,459	0,100	21,225	1	0,000	1,583	1,302	1,925
	[Degree of urbanism=11]	0,452	0,121	13,995	1	0,000	1,571	1,240	1,990
	[Degree of urbanism=12]	0,732	0,132	30,546	1	0,000	2,079	1,604	2,694
	[Degree of urbanism=13]	0,116	0,102	1,271	1	0,260	1,122	0,918	1,372
	[Degree of urbanism=14] (>6500 addresses)	0b			0				

8.5. Appendix E. Final model fitting

Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	49491,361	49515,919	49485,361			
Final	40049,716	41081,161	39797,716	9687,644	123	0,000

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	33794,851	32265	0,000
Deviance	29657,967	32265	1,000

Pseudo R-Square

Cox and Snell	0,306
Nagelkerke	0,327
McFadden	0,133