ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis [Data science and marketing analytics]

Conjoint analysis of the soft drinks market

Name student: Jeroen Schreuder

Student ID number: 452973

Supervisor: Fok, D.

Second assessor: Bellet, C.S.

Date final version: 25-11-2020

Table of contents

Abstract

The soft drink consumption patterns have recently been changing by an increased availability of different types of drinks in the last few years. This shift in consumption patterns is a problem and opportunity for the soft drinks industry. If the soft drink producers do not react, they might miss the boat and lose profit. Therefore, the goal of this research is to create a marketing strategy with regards to specific product attributes of soft drinks for different consumer segments. This has been done with the use of choice-based conjoint analysis based on the multinomial logit. The segmentation has been done based on age, the consumption frequency and the consumption moment. The results showed that the general preference consists out of a soft drink in a glass bottle without sugar, branded as top brand and the product should be recyclable. There appeared to be some small differences between the preferences of the different groups, but these differences were not large enough to advice soft drink produces to perform segmentation based on this research.

1. Introduction

1.1 Background

Soft drinks are beverages composed of water and other ingredients that provide specific sensorial attributes. The soft drink consumption patterns have recently been changing by an increased availability of different types of drinks in the last few years. The consumers' preference has changed partly due to the specific attributes of the soft drinks. A research on the consumption patterns of soft drinks showed that American adults changed their soft drinks consumption pattern in 2002 compared to 1977 (Redondo, Gomez-Martinez & Marcos, 2014). Soft drink consumption rose with 20% in this period. The Healthy Lifestyle in Europe by Nutrition in Adolescence Cross Sectional Study (HELENA-CSS) evaluates the number of beverages consumed across urban areas from 10 European cities. They found that sugar-sweetened beverages are the second most consumed beverage (430 ml per day) after water (821 ml per day) (Redondo, Gomez-Martinez & Marcos, 2014). The soft drink consumption in the Netherlands was above 100 liters per capita 2009. In 1970 the Dutch consumed 55.3 liters per capita. However, when considering the last years, the liters of soft drinks per capita in the Netherlands is slightly decreasing and the liters of alternatives for soft drinks per capita is increasing (CBS, 2010; FWS, 2018). The Dutch consumed 99.6 liters per capita in 2014. This amount is decreasing every year. In 2018, the Dutch consumed 93.7 liters soft drinks per capita (Table 1). On the contrary, alternatives (flavoured waters and non-processed waters for example) for soft drinks are gaining terrain and people consume more of them every year (Table 1). Thus, when considering the information presented above, there seems to be a shift in the consumption patterns in the soft drinks market. This shift is a problem and an opportunity for the producers in the soft drinks industry. Think about Coca-Cola, Fanta, Schweppes or Hero for example. It is important for the soft drink industry to react and not miss the boat. If they ignore this shift, this will probably lead to losing customers and profit. However, reacting on time (reacting before the competitions does) might lead to an increase in profit and market share. Thus, the changing consumption patterns are an opportunity as well. Therefore, it would be interesting to investigate how the soft drink industry should react to this shift in consumption in terms of creating an appropriate marketing strategy. The marketing strategy is the plan that should be made for the industry as reaction to this shift. This will be elaborated in section 1.3.

Some research has been done about consumer preference with regards to soft drinks. Jansen & Verkooijen (2014) determined the underlying attributes behind soft drink consumption. They

made a distinction between different socioeconomic groups in their research for targeting purposes. However, they used a regular questionnaire to gain insight into consumer preference. This is a questionnaire where the researcher asks the respondents preference directly. This is not most reliable, because consumers might not reveal their real preferences in this type of survey. Redondo, Gomez-Martinez & Marcos (2014) looked at the sensory attributes of soft drinks and one of their findings was that health, calorie content and price are important attributes for preferring a certain type of soft drink. Another research done by Enneking, Neumann & Henneberg (2007) looked into consumer preference in the soft drink market. For future research, they suggested to look into the pricing of soft drinks. They used a conjoint analysis to examine what factors influence a consumer's purchasing decision based on three different soft drinks. The participants had to taste the soft drinks. However, taste is a factor which can only be investigated in a real setting where participants have to taste the drinks. A survey will be used in my research, thus it is not possible to include taste. It is only possible to describe the taste, but this does not come close to the real tasting experience. Therefore, it has been decided that taste will not be included in this research. Nonetheless, it should be mentioned that taste is an important attribute in having a preference for one soft drink over another (Enneking, Neumann & Henneberg, 2007).

1.2 Goal and explanation of this research

The goal of the research is to investigate the consumer preference towards soft drinks in the Netherlands. I would like to examine what type of soft drink they prefer, and which factors determine this preference. Market research on soft drinks consumption is usually done based on conventional surveys. The drawback of this approach is that participants might not reveal their real preferences. Besides, it does not come close to the real choice task for preferring one soft drink over another. Hence, a better approach is choice based conjoint analysis (CBC). With the use of this method, I will examine the consumer preference towards soft drinks. The benefit of CBC compared to a regular survey where preferences of participants are being asked is that with CBC it is possible to calculate the WTP for different combinations of attribute levels. Besides, CBC comes closer to the real choice task compared to a regular survey (a survey where the respondents' preference is asked directly). Furthermore, I will make segments across the data based on variables that have to be determined in the theory part of this research. With the use of these variables, it will be possible to make segments in the data. To execute CBC, I need to

identify the most important attributes of soft drinks. These attributes will be based on literature research.

1.3 Problem description and research questions

This research is scoped on examining the consumer preference towards soft drinks and to create a marketing strategy around this preference. The marketing strategy will be the solution to the changing consumer behaviour discussed in the introduction. If the soft drinks industry does not respond to this change, their profits might decrease. On the contrary, reacting on time (reacting before the competitions does) might lead to an increase in profit and market share. Thus, creating this strategy as response to the changing consumer behaviour is actually about gaining knowledge about the consumer preference towards soft drinks, right now it is crucial to get a better understanding of the consumers' needs because of the changing consumer behaviour earlier discussed. The marketing strategy will consist out of the most preferred product for different consumer segments aimed at selling in the supermarkets. These variables will be discussed in the theory part of this research. Examples of these segmentation variables are age or how often someone consumes soft drink. To create the marketing strategy for the soft drinks industry aimed at selling in the supermarkets, the following research question has been formulated:

"How should the marketing strategy be made for soft drinks?"

To answer the main-research question, the following sub-questions have been made:

1. *Which attributes are important for soft drinks?*

   This question will provide an answer to the first step in the CBC, namely identifying the most important attributes of soft drinks. The answer to this question will be based on previous research done in the soft drinks area.

2. *In what way should consumers be segmented in the soft drinks market?*

   It is important to determine in what way consumers should be segmented for marketing purposes in the soft drink industry. The answer will be provided with existing literature about segmentation.

3. *What are the most important attributes that determine the consumers preference?*

   The answer to the question above will provide the most important attributes that determine the consumer preference. The most important attributes will be used to create a marketing strategy.

4. How does the Willingness to pay(WTP) differ across different attributes and segments?

The WTP is an important part of the analysis. It gives the amount of money that a consumer is willing to pay given a specific combination of attributes. This is needed to determine the overall marketing strategy because the pricing is part of a marketing strategy as well.

1.4 Outline

In the next section, the theoretical section covers important underlying concepts for this research. Furthermore, literature will be discussed to make the foundation for the conjoint research. In this section will be investigated which attributes are the most important for in the soft drinks market. Next, the method section will cover information about the methods used for the analysis used in this research. The section after this will be about the data in this research. In this section, important aspects about the data will be discussed. Following, the results section will cover the most important findings in this research. Finally, the conclusion provides an overall concluding answer to the main-research question.

2. Theory

This research combines different concepts to solve the marketing strategy problem. Because of this academic works which are necessary for the foundation of this research will be discussed. First, a section about the types of beverages on the market will be presented. Second, some theoretical concepts which are needed to understand some concepts in this research will be discussed. Third, the types of consumers in the market will be discussed in terms of lifestyles. Fourth, segmentation variables will be determined and discussed.

2.1 The soft drinks market

Table 1

Dutch beverage consumption in Liters

	2014	2015	2016	2017	2018
Soft drinks*	99.6	95.9	94.4	93.8	93.7
Non-processed waters	22.5	23.8	25.8	27.2	28.8
Flavoured waters	2.6	3.3	4.2	5.2	5.9
Syrups**	29	29.6	29.4	28.9	29.1
Juices	15.2	14.4	13.5	12.1	11
Nectars***	11.2	11	10.2	9.4	8.8

* Carbonates, still drinks, iced tea and sports and energy drinks
** Ready to drink
*** Juices contain 100% juice, nectars contain 25%-99% juice

As can be seen in Table 1, the average soft drinks consumption per capita in liters in the Netherlands has been slowly decreasing in the last years from 99.6 to 93.7. This means that a transition is taking place in the consumption of soft drinks. The national Frisdrankonderzoek (2017) concluded that the Dutch are becoming more aware of their intake of soft drinks. Cola regular is still the most popular soft drink but soft drink producers have to cope with this change in consumption. Consumers are choosing lower calorie and or low sugar soft drinks. This makes the soft drinks isle in the supermarket one of the most innovative categories within the supermarket. This can sometimes be confusing for the consumers. It appears to be the case that a substantial part of the consumers does not have adequate product knowledge about soft drinks. Many of them overestimate the amount of calories and or sugar in cola zero for example. Besides, they think that artificial sweeteners have a negative effect on health. However, the European food authority concluded that they are safe to consume within set boundaries (FWS, 2017).

2.2 Willingness to pay (WTP)

Willingness to pay (WTP) is specified as the maximum price that people would pay for a product. With regards to CBC, the WTP is often calculated for a change of an attribute level. For example, how much does the WTP change when someone prefers a recyclable bottle over a non-recyclable bottle. Knowledge about the WTP for a product on behalf of the customer plays a crucial role in many areas of marketing management with regards to pricing decisions or new product development. Therefore, it is an important part for the creation of a marketing strategy in the soft drinks industry. Despite advances in both academic and applied pricing research, many companies do still make their pricing decisions without a profound understanding of the likely response of their buyers and competitors to alternative pricing strategies. This results in failing to pursue an adequate pricing strategy that is custom made to their marketing environment. As a result, they are at risk of ignoring valuable sources to increase the profitability of their products offered to the customer. This type of failing often leads to "intuitive pricing" (Beidert, Hahsler & Reutterer, 2006). Several approaches have been developed to measure WTP. The primary distinctions between several approaches are whether they measure the WTP in a direct or indirect way. Some marketeers prefer to use the direct approach by asking the respondents to state their WTP for a specific product in an open-ended question format. This might not lead to the optimal result because it is possible that consumers do not reveal their real WTP. Hence, they might not know what they are exactly willing to pay for a specific product. Because of this, others prefer choice-based conjoint (CBC) analysis where the WTP is calculated based on the consumers' choices among several product alternatives (Milner, Hofstetter, Krohmer & Zhang, 2011). With respect to this study, the WTP will be calculated for the most important combinations of attributes with regards to the preferred type of soft drink. For this research, only the base price of a product will be used. This means that promotions will not be taken into account. They will not be taken into account, because CBC will be used in this research and with this method it is not feasible to use many different prices. Therefore, promotions are left out in this research.

2.3 Consumer decision making

Szmigin, & Piacentini (2018) refer to high involvement consumer decision making which is the process by which (1) consumers identify their needs, (2) collect information, (3) evaluate alternatives, and (4) make the purchase decision. These actions are determined by psychological and economic factors, and are influenced by environmental factors such as cultural, group, and

social values. They also refer to low involvement decision making. The main difference between both decision-making processes is that in high involvement decision making a choice is made following a process of search and evaluation, where some kind of evaluative criteria will be applied. In the low involvement decisions making the choice is made without these intervening steps. This means that according to high involvement decision making, consumers identify a need in stage one, with regards to soft drinks this could be the need of drinking a cola. The second stage is about collecting information where the consumer will gain knowledge about the cola drink. In the third stage, possible alternatives are evaluated. This set contains a few other alternatives such as Pepsi and other drinks. The drinks within this set are being called the "evoked set". This set refers to the alternatives that person knows. In the last stage, stage four, the purchase decision is being made. In my example, this means that the customer will decide whether to buy the coke or not. In low involvement decision making, the consumer would not search for alternatives. High involvement decision making is more likely to happen within the soft drinks market, because there are many types of soft drinks and the consumer can search for alternatives when he or she does not like one particular soft drink.

The book Thinking fast and slow by Kahneman (2011) refers to two different systems when consumers are making a decision. With system one he refers to the part that performs routine operations, with little effort and no feeling of voluntary being in control. He considers this system as being the routine mode. On the contrary he speaks about system two which he considers to be the reflective mode. This mode refers to effortful attention to a mental activity and is associated with choice and concentration. Both are distinct systems however they are complementary systems. System one assesses the current situation and delivers updates and system two seeks new/missing information to make decisions. Both systems are important regarding decision making to soft drinks for example. System one is the automatic system which is involved when the customer goes to the beverage isle in the supermarket and grabs a particular drink. System two is involved when the customer stops to think about the beverage choices. When everything goes smooth, system two adopts the suggestions made by system one (Kahneman, 2011). Why is this important for the decision-making process and this research? Well, it has implications for market research. When the participants in this research are answering questions, it is important that their system two is active in the decision-making process because if it is not, it is possible to have superfluous answers with little depth. However, if it is

possible to engage system 2 in this research, then the answers and choices become more meaningful. Therefore, choice-based conjoint will be used. This method requires more thinking and engagement from the participant compared to an open question format or just asking the participants preferences.

2.4 Consumer preference

To create a marketing strategy for the soft drinks industry, it is highly important to understand what factors are the driving source behind consumers' choices. Therefore, it is important to understand the concept of consumer preference. Consumer preference refers to the process of selecting an option which has the greatest anticipated value among a number of options by the consumer in order to satisfy needs or desires. The preference is the result of the consumer decision making process described in section 2.3. Consumer preferences are defined as subjective individual tastes of various bundles of goods measured by utility (Ubeja & Patel, 2014).

This research will be based on CBC. This method measures utility of the respondents by giving the participants multiple choice sets and asking which option they prefer. This type of experiment can be referred to as a discrete choice experiment (DCE). Lancaster (1966) made a new approach towards consumer preference. He states that a good on itself does not give utility to the consumer within a DCE. However, it possesses characteristics that give rise to utility (Stigler, 1945). A consumer good will possess more than one characteristic and many of these characteristics will be shared by more than one good. Within the DCE, this are the product attributes which yet have to be determined for this research. Examples of these attributes are the product shape, product size, product color, product package and the product brand. Voicu (2013) assumes that utility is made out of attributes of a particular product as well. When doing research to identify consumers' preference the consumers may express their needs and desires and still may act in a totally opposing way. It is even possible that they are not even aware of the true motivations behind their own buying behaviour or that they could possibly react to factors that determined last minute changes with regards to their buying behaviour. Therefore, Voicu (2013) concludes that the psycho-physiological process behind the consumer is difficult to take into account. He assumes that preferences are revealed through choices, this is also being referred to as "revealed preferences". The "revealed preferences" cannot be investigated based on a survey, because surveys do not replicate the real choice task. However, CBC comes closer to the real choice task

compared to asking consumers what they prefer. It seems to be the case that consumers' reveal their preferences during the choice process, but how is this decision being made? To give an answer to this question, we have to assume that the participants of choice experiments are rational and want to maximize their utility while having stabile preferences at the same time. However, consumers are very much constrained in their choices (Ubeja & Patel, 2014). These constraints are defined by the consumers' income and the prices of the consumer pays for the good. Nonetheless, the consumer will choose their most preferred bundle of goods. This means that choosing one bundle of goods given the constraint can be seen as an optimization problem. The consumer will choose the consumption bundle that will maximize its own utility with respect to the budget constraint (Ryan, Gerard & Amaya-Amaya, 2008). An example in the soft drinks market is someone with a budget of €10.00 who has to choose between cola (€2.50) and Fanta orange (€1.00). Assuming that this person does want to divide the budget over a combination of drinks and having a certain amount of money left, he has to choose between different combinations of drinks. A person shall choose a combination of the drinks and amount of money to keep that will lead to the highest utility for this person. The person also derives utility out of the outside good, which is the money he or she has left. Therefore, a combination of goods and money left can be chosen as well. For example, this person can buy four cola's and zero Fanta orange or five Fanta orange and two cola or 1 cola and 1 Fanta and keep €6.50 or many more different combinations. It has been demonstrated that consumers' preference can be measured effectively and that a study on consumer preference can provide a more thorough understanding on the choice's consumers make (Voicu, 2013). Therefore, CBC is a useful analysis to give insight in consumers' choice.

2.5 Types of soft drinks

It is important to get a better understanding of the available types of soft drinks on the market, because the types of soft drinks have to be determined for the CBC. When asking a participant to choose a preferred soft drink. It is important to have different types of soft drinks to be able to have a preference of that consumer. A good overview of the types of soft drinks can be found at the Albert Heijn[1] and jumbo[2] on their online website. When viewing their website, the following brands are present: Coca-Cola, Fanta, Pepsi, Ah private label, Red Bull, Lipton, Schweppes and some other brands which are not as commonly known as these brands. Most of

[1] https://www.ah.nl/producten/frisdrank-sappen-koffie-thee/frisdrank?page=2
[2] https://www.jumbo.com/

these brands do have multiple variants of their drinks in the form of light versions and or zero sugar versions. AH and Jumbo have been chosen to look at because it they have the largest market shares in the Netherlands with 34.9% and 21% respectively (Schelfaut, 2020). When looking at Table 1, it can be noticed that there are other types of drinks as well on the beverage market. Nonetheless, according to FWS (2018), the soft drinks category contains carbonates, still drinks, iced teas and sports and energy drinks.

2.6 Types of consumers

In this research, different types of consumers should be considered. Consumers have different lifestyles and these lifestyles do probably affect their consumer behaviour. Kvaavik, Andersen & Klepp (2006) investigated the intake of carbonated soft drinks and the association between this intake and lifestyle factors and bodyweight of people aging from 15 to 33 years old. The study showed the people show stability from early adulthood into later adulthood. More importantly, they found that a high intake of sugar-sweetened soft drinks was associated with reduced levels of leisure-time physical activity among and with increased smoking prevalence and increased energy intake among men compared to people with having a low long-term intake of soft drinks. According to this study, the intake is an important factor in the lifestyle of a consumer because they tend to have a different lifestyle when consuming more soft drinks. Nevertheless, they did not find any significant relationship between the soft drink intake and bodyweight. Another study done by Kumar and Ray (2018) examined the consumption patterns and attitudes towards soft drinks among Indian youth. They used cluster analysis to obtain a better understanding among the attitudes of young consumers towards soft drinks. They found two clusters of consumers with different consuming frequency of soft drinks consumption. This is interesting for my research, because this implies that differences in consumption frequency possibly leads to different choice patterns among the participants. The group who consumed more frequently, drank more different types of soft drinks compared to the group that did consume the most infrequent.

Some studies found a positive relationship between soft drink consumptions and overweight in children and adolescents (Ludwig, Peterson & Gortmaker, 2001; Giammattei, Blix, Marshak & Wollitzer, 2003) while others did not (Forshee, Anderson & Storey, 2008). However, it is not possible and or ethical responsible to target people based on their weight. Therefore, this will not be investigated in this research. When taking the studies above into account. It can be concluded that there is a relationship between soft drink consumption and lifestyle. This finding

might have an effect on the choices that consumers make regarding soft drinks. Then the following question arises: how should this be taken into account for this research? According to Weinstein (2004), companies target their customers based on their consumption frequency in order to increase their consumption and to "*move them up the usage ladder".* When looking at lifestyle from that perspective it becomes easier to categorize consumers into groups. Previously discussed literature discussed unhealthy lifestyles. To be able to take a part of consumer lifestyle into account for this research, the participants will be asked how often they drink soft drinks. When asking this, it can be examined whether differences in consumption frequency lead to different soft drink preferences.

2.7 Consumption moment

The consumption frequency has been discussed in section 2.6. Not only, the consumption frequency seems to be important, but it might be the case that the consumption moment is an important segmentation variable as well. Unfortunately, there is little literature available about the consumption moment of soft drinks. Enneking, Neumann and Henneberg (2008) examined intrinsic and extrinsic product attributes by means of a choice-based conjoint experiment. They included a few exogenous variables in their research. One of the questions was: "When do you drink carbonated soft drinks?". The participants could choose out of the following options: "on the way" or "during work out". They found relatively high coefficients for both variables compared to other variables. This means that both variables have an impact on the choice's consumers' making with regards to soft drinks. However, a day has more moments than "during workout" and "on the way". Therefore, I will include "during dinner", "during work", "in the evening", "in spare time" and "during lunch" as well.

2.8 Generation marketing

Each generation has unique expectations, experiences, generational history, lifestyles, values, and demographics that influence their buying behaviours. Because of this, many companies use a multi-generational framework to try to understand and gain the attention of these diverse buyers. Multi-generational marketing refers to the practice of appealing to the unique needs and behaviour of individuals within more than one specific generational group, with a generation being a group of individuals born and living about the same time (Smola & Smutton, 2002). Each generation needs an appropriate marketing strategy that fits that group. Smola and Smutton (2002) and William and page (2011) do both divide the age groups in the following groups: Baby

boomers(1946-1964), generation X(1965-1977), generation Y(1978-1994) and generation Z(1995-2010).

2.9 Product attributes

In this part of the research, the most important attributes regarding soft drinks will be selected based on literature. Sorenson and Bogue (2006) identified extrinsic and intrinsic attributes of a range of novel stimulant beverages. They used the following attributes: brand, flavour, carbonation level, added ingredients and type of packaging. Where the packaging had the following options: glass bottle, aluminum can and plastic bottle. For the flavour attribute they used the following attribute levels: Blend of orange and spring water, blend of apple juice and spring water or lemon and lime-flavoured spring water. The most important attributes were the flavour, type of packaging and added ingredients. The attribute levels for the type of packaging were aluminum can, plastic bottle and glass bottle. For the brand they decided to give the participant two options: "familiar brand" and "new brand". Both attribute levels had a small effect on the utility. In my own research I will use the *packaging attribute* as well with the same attribute levels as in the research of Sorenson and Bogue (2006), but a top brand option will be added. The study done by Enneking, Neumann and Henneberg (2007) used three different branding options: national brand, regional brand and private label, because their goal was to assist soft drinks manufacturers and not one specific brand. Furthermore, they used the price, labelling and taste as attributes. Price, brand and the taste (amount of sugar) were the most important attributes. They found high negative effects for all the used sweetener compared to the option with 100% sugar. The sweetener systems were 100% sugar, 67% sugar, conventional sweetener, or diet sweet up. Because the sweetener systems were important as well, I will use them in my own research as well. However, using them in the same way as being done by Enneking, Neumann and Henneberg (2007) is the approach which I will use as well in my research but not in the exact same way. Coca-Cola uses "regular", "light" and "zero" for example. Other brands do this as well. However, the association between a brand and this type of referring to "regular", "light" and "zero" might be too strong. Therefore, it would be better to refer to drinks with and without sugar. Drinks without sugar can contain other sweeteners. These sweeteners can be divided in artificial and natural sweeteners (Mooradian, Smith & Tokuda, 2017). To include the *type of drink* in this research, it has been decided to use the following attribute levels: "with sugar", "without sugar", "artificial sweetener" and "natural sweetener".

Another study done by Myers (2003) looked at the impact of different attributes on soft drink preference. He went for another approach compared to previous discussed research with regards to the brand attribute. He used the top 9 brands in the United-States as attribute levels while other researchers did not use brand names as attribute levels (Sorensen & Bogue,2006; Enneking, Neumann & Henneberg, 2007). Furthermore, he used the following attributes: calorie content, sugar content, sodium content, caffeine content, carbonation, taste, sweetness and price. Myers (2003) concluded that the brand has a greater importance compared to the overall preference for a specific brand meaning that brand is highly important in the choice process. In this study, I do not want to use real brands, because the study aims to find a marketing strategy for the soft drinks industry. When using real brands, the study cannot be aimed to the entire soft drink industry but only to those brands. Therefore, I choose to use the approach used by Sorensen & Bogue (2006) and Enneking, Neumann & Henneberg (2007). This means that the *brand attribute* will have the following levels: "New brand", "top brand", "private label".

The study by Redondo, Gomez-Martinez and Marcos (2014) investigated the sensory attributes of soft drinks and their influence on consumers' preference. They claim that consumer's preferences to choose one drink over another is not merely influenced by genetic, psychological and environmental factors, but also by the beverage's specific attributes. There are several sensorial systems involved in the brain when tasting a particular type of drink. The taste and mouthfeel sensations are the main factors influencing a consumer to choose a beverage over another one. This preference already occurs when being a child and is more or less the same when later being in adulthood. These attributes can be used to influence the consuming patterns. This research is in line with other research who claimed that taste is an important factor (Sorenson & Bogue, 2006). However, measuring taste without really tasting a soft drink is not possible. The research done by Sorenson & Bogue (2006) used descriptions of taste as attribute levels. I will not use this in my research, because as said earlier, taste is only measurable when the tasting process really takes place. Therefore, from a conjoint perspective with a questionnaire, describing taste does not replicate the taste to a certain level. As a result of this, taste will not be used in this research.

UNESDA is a European Association. It represents the interests of the European soft drinks industry. UNESDA's general director Nicholas Hodac said the following: "Creating a sustainable future is our generation's biggest challenge and the realization of a circular economy and a carbon-neutral Europe are top priorities for both governments and businesses." (Hodac, 2019). Meaning

one of the goals of the soft drinks industry is to become more sustainable. For example, production is now closer to markets to minimize transport and energy emissions. The goal of UNESDA is to have 100% recyclable bottles in the industry by the end of 2025. For the industry it is important to know how highly the customers value sustainability. Furthermore, sustainability is becoming a more important topic in society, sustainable marketing becomes more and more important. It will even be a moral obligation in the future (Kumar, Rahman, Kazmi & Goyal, 2012). Therefore, including an attribute with regards to sustainability in the research is beneficial. Previous research did not use an attribute linked to sustainability (Sorensen & Bogue, 2006; Enneking, Neumann & Henneberg, 2007; Redondo, Gomez-Martinez & Marcos, 2014; Meyers, 2003). My research will have an attribute about *recyclability* with the following attribute levels: recyclable and not recyclable.

Lastly, the *price* should be an attribute for the conjoint analysis. This attribute is needed because otherwise it will not be possible to derive the WTP across different attributes and segments. In the studies of Sorensen and Bogue (2006) and Enneking, Neumann and Henneberg (2007), the price is an attribute with a high coefficient compared to the other attributes, meaning that the price is important for choosing one product over another. The pricing attributes will have the following levels in this research per 330ml: €0.20, €0.40, €0.60, €0.80. The prices are set per 330ml, because the packaging attributes contains: "glass bottle", "aluminum can" and "plastic bottle". Plastic and glass bottles are also available in larger sizes than 330ml. However, aluminum cans are usually not available in larger sizes than 330ml. Therefore, I have chosen to set all the prices per 330ml. The prices should not be too high, because a high price will probably not be chosen. On the other hand, the prices should not be too low to prevent one price to be always chosen.

The following question (sub-question one) can be answered at this point in the research: "*Which attributes are important for soft drinks?"*, the following attributes will be used for the conjoint analysis including the attribute levels. First, packaging will be used with the following levels: aluminum can, plastic bottle and glass bottle. Second, the type of soft drink will be used with the following levels: "with sugar", "without sugar", "artificial sweetener" and "natural sweetener". Third, the brand will be used with the following levels: "new brand", "top brand" and "private label". Fourth, an attribute about recyclability will be implemented with the following levels: "recyclable" and "not recyclable". Lastly, price per 330ml will be included with the following levels: €0.20, €0.40, €0.60, €0.80.

At this point in the research, the answer to the following question will be given (sub-question two): *"In what way should consumers be segmented in the soft drinks market?"*, the answer to this question was unknown in the problem description in section 1.3. The age variable will be used where the groups are divided into the 4 following subgroups: baby boomers, generation X, generation Y and generation Z. Furthermore, their consumption frequency will be asked with the following options: "never", "once a month", "weekly", "daily" and "multiple times per day" . Lastly, the moment of consumption will be asked with the following options: "breakfast", "lunch", "dinner", "during sports", "on the way", "during work", "spare time" and "not".

3. Methods

In this chapter, the techniques being incorporated for this research will be discussed. First, choice-based conjoint analysis will be discussed which is the foundation of the research. Then, the setup for the conjoint analysis will be discussed which is the generation of the choice design. Following, the concept of willingness-to-pay and the likelihood-ratio test will be explained. Finally, the ROC (Receiving Operator Curve) and variable selection will be discussed.

3.1 Conjoint analysis

Typically, consumer preferences are estimated by means of conjoint analysis using online or paper-and-pencil surveys (Decker & Trusov, 2010). In conjoint analysis, participants have to rate or rank a product or choose their most preferred option out of a set of products. Conjoint analysis has received considerable academic and industry attention as a major set of techniques for measuring buyers' tradeoffs among multi-attributed products and services (Green & Srinivasan, 1990; Fader & Hardie, 1996). It is mostly used to gain insight into consumers' preference. Choice-based conjoint analysis is unique in the way that the participants make choices instead of giving a rating or ranking to a product. Choice data does not fit well into linear models. Hence, marketers have developed choice models, which are well suited to understand the relationship between the attributes of products and customers' choices among sets of products (Chapman & Mcdonell, 2015).

3.1.1 Types of conjoint analysis.

Researchers have developed different types of conjoint analysis (rating, ranking and choice based) as well as different techniques to estimate parameters for conjoint models (Natter & Feurstein, 2002). Rating- and ranking based conjoint analysis consist out of pairwise comparisons where the respondent needs to rank or rate a product, while choice based conjoint analysis is used to ask respondents to choose the most preferred product out of a set of alternatives. Ranking, rating and choice based are methods used in conjoint analysis, with all three methods it is possible to decompose products into attribute levels and to estimate the part-worth utilities of these attribute levels. They differ in the estimation method being used. In choice based conjoint, the utility is based on the choice option that respondents prefer while in ranking based conjoint analysis, it is based on rankings for a certain product and in rating based conjoint it is based on a rating, for instance, on a scale from 1- to 10 (Louviere & Woodworth, 1983). One of the benefits of choice-

based conjoint over other choice-modelling methods is that it is closer to true decision making in real life compared to a regular questionnaire.

3.2 The choice model

The standard random utility model will be used for this research, it is needed to understand how the total utility of an alternative is derived (Huber & Zwerina, 1996). With Conjoint-analysis it is possible to derive the part-worth utilities of all the attribute levels. Part-worth utility is the utility that one attribute level contributes to the total utility. Louviere and Woodworth (1983) specified how to derive the utility of an alternative in conjoint analysis. First one has to calculate the consumers' utility for alternative k with the following formula:

$$v_a = \sum_k \beta_{ka} * x_k \quad (1)$$

$$u_a = v_a + \varepsilon_a \quad (2)$$

Each individual is being faced with a set of choice alternatives. The set of all relevant choices is α, containing J members. Subsets of α are denoted by capital letters and single members by lowercase letters. For instance, A is a set of choices within α and a is an alternative within choice set A . They assume that value, v_a is associated with each alternative a in choice set A. Each of the members A is described by some combination of levels of K attributes by vector x' = (x_1, x_2, … x_k), x_k refers to the observed value of the attribute (Louviere & Woodworth, 1983). The utility u_a denotes how much each alternative is liked or valued and is a function of the vector of attributes (x) of the alternatives v_a where β_{ka} is the weight of attribute k in the valuation of alternative a. For every respondent, the utility for a choice consists out of a deterministic component (equation 1), that represents the latent preference structure at the aggregate level, and a random component. This random component is an error term which can be seen in equation 2. The random component is caused by fluctuations in perceptions, attitudes, or other unmeasured factors (McFadden, 1986). As being discussed earlier, multiple techniques have been developed to estimate parameters within conjoint analysis. For this research the model discussed below will be used.

3.2.1 The model.

A linear regression would not be ideal for choice data, this would result in chances bigger and or smaller than zero. Therefore, the multinomial logit is the solution (equation 3). The multinomial logit model computes the probability of choosing an alternative as a function of the

attributes of all the alternatives available. Various authors have employed it in marketing (Guadagni & Little, 1983).

Consider a choice experiment with N choice sets referred to as C_n, indexed by n=1,…,N. Each choice set is characterized by a set of alternatives $C_n = \{X_{1n,\ldots,}X_{Jnn}\}$. Then the probability that a consumer will choose alternative i from choice set C_n, under the assumption that the ε_a's from equation 2 are independently, identically Gumbel-Distributed equals:

$$P_{jn} = \frac{e^{x_{in}\beta}}{\sum_{j=1}^{J_n} e^{x_{in}\beta}} \qquad (3)$$

The multinomial logit model in this research will have a choice variable as dependent variable. This variable takes value 1 if an alternative within a choice set is being chosen and 0 otherwise. The independent variables are the variables presented in Table 4. The formula used for this research has the following form based on equation 3 where x_k are the attributes used in this research:

$$U_a = \beta_1 * Packaging_k + \beta_2 * TypeOfSweetener_k + \beta_3 * Brand_k + \beta_4 * Recyclability_k + \beta_5 * Price_k + \varepsilon_a$$

The parameter estimates will be obtained with R, which is a software program used in statistics. The "mlogit package" will be used, this package uses the equation discussed below (equation 4).To estimate the β's, the following formula will be maximized, which is being referred to as the log-likelihood function:

$$L(Y \mid X,\beta) = \sum_{n=1}^{N}\sum_{j=1}^{J_n} y_{jn} \ln(P_{jn}(X_n,\beta)) \qquad (4)$$

Here, Y is a matrix of choices with elements y_{jn}, where y_{jn} equals 1 if alternative i is chosen in C_n and 0 otherwise. $P_{jn}(X_n,\beta)$ refers to equation 3, this is the choice probability of alternative j in set C_n. The β's are estimated using the derivative of equation 4 with respect to each parameter.

3.3 Design generation

In conjoint analysis, not only a model for analysis should be selected. A few steps before this are of major importance. The researcher has to select an appropriate choice design with respect to the attributes and attribute levels. This research will have 5 attributes with 2- to 4 attribute levels per attribute (Table 4). To conduct the research, it should be determined how many choice-sets and

how many alternatives will be presented per question. The used choice sets in a survey are called the "choice design". To make the design, a full factorial design could be used. This is the combination of all attribute levels in all possible combinations. For this research, a full factorial design will consist out of 288 alternatives (multiply the attribute levels 3 * 4 * 3 * 2 * 4). The total number of choice sets when 4 products per choice set are used is equal to $\frac{288!}{4!284!}$. However, it is not possible to use that many choice sets in a survey due to the fact that respondents are not able to process that much choice sets in a survey. Therefore, a subset should be selected. This subset is called a fractional factorial design where the researcher takes a subset out of the entire choice sets. A subset is called orthogonal when the joint occurrence of any of the two levels of two different attributes appear in profiles with equal frequencies to the product of their marginal frequencies. In this research, this would be satisfied when "glass bottle" and "with sugar" appear in one-twelfth of the cases, because "glass bottle" should appear in one-third of the cases and "with sugar" in one-fourth of the cases, their joint occurrence equals one-twelfth (Huber & Zwerina, 1996). An orthogonal subset improves the choice design, because it ensures that some attribute levels are not overrepresented neither or underrepresented, this leads to better estimation of the coefficients.

3.3.1 Efficient choice designs.

Several authors have proposed designs that decrease the number of respondents or choices per respondent needed to achieve an expected level of accuracy (Huber & Zwerina, 1996). This is done by making the choice design efficient. Out of this rises the following question: "what makes choice designs efficient and how to measure efficiency?". Huber and Zwerina (1996) discuss the properties that make a choice design efficient. They measure efficiency in terms of the D_p error which is a probability-centered estimate of error (Huber & Zwerina, 1996). The goal is to minimize this error around the estimated parameters. The first property that makes a choice design efficient is *level balance*. Level balance is the requirement that the levels of an attribute occur with equal frequencies. For instance, within this research, the attribute level "glass bottle" should occur in precisely one-third of the cases, because it is part of the attribute "packaging" which has three attribute levels. The second criterion is *orthogonality*. In an ideal world, level balance and orthogonality could both be satisfied. However, many designs do not allow to satisfy both criteria. Consider, for example, a design with a three-level attribute and one four-level attribute. When the total number of choice sets used in a survey is not a common product of 3 and 4, level balance and orthogonality cannot both be satisfied. A third criterion is *minimal overlap*. The first two criteria

are independent with respect to the choice sets, they define which profiles will be used. Whereas minimal overlap is important within a choice set. Minimal overlap means that the probability that an attribute can repeat itself within a choice set should be as small as possible. An example of the violation of this criterion would be the case where the levels of one attribute are the same across all alternatives within a choice set. This leads to a choice set that provides no information of the attribute's value. When assuming that every consumer is indifferent, meaning that this person derives equal utility out of all the attribute levels within an attribute, an efficient utility-neutral design can be generated where the three criteria mentioned above are a consequence of this, but they need to be satisfied to call a design a utility-neutral design (Huber & Zwerina, 1996). The most efficient design will be the design with the lowest D_p error (Huber & Zwerina, 1996).

3.3.2 Utility balance.

The downside of utility-neutral designs is that it is possible to have choice sets that are uninformative when the prior parameters are zero, because of dominant alternatives within a choice set. This can be solved by improving the efficiency of the designs by balancing the utilities of alternatives in each choice set. This can be achieved by altering the prior parameter estimates to a non-zero value, this means that the researcher makes an estimation about the expected values of the parameters in order to rank attribute levels instead of having no preference order within an attribute. Huber and Zwerina (1996), first make a utility-neutral design where they assume that the prior parameter estimates are 0, meaning that $\beta = 0$. Changing these values of β lead to different probabilities of an alternative being chosen by the consumer. When setting the β's to a nonzero value, the probability that a consumer will choose alternative i from choice set C_n increases when $x > 0$ as β increases and decreases when β decreases and vice versa when $x < 0$ (based on equation 3). To illustrate the impact of nonzero β's, Huber and Zwerina (1996) space the partworths for a three-level attribute evenly between -1 and 1. Hereby, the part worth -1 corresponds to attribute level 1, 0 to attribute level 2 and 1 to level 3. This should be done in the order that reflects the true expected part-worths. When alternatives in a choice set have more or less similar utilities, this is referred to as *utility balance*. This avoids having dominant alternatives.

Huber and Zwerina (1996) showed that using prior values of β in the right direction lead to more efficient choice designs compared to using designs where $\beta = 0$. Partworths can only be roughly estimated before the experiment. Generating the most efficient design requires prior knowledge of one or more parameters. Prior knowledge of a certain parameter is only required

when it is reasonable to expect that there is a clear preference within an attribute beforehand. The prior knowledge is needed to state a preference within an attribute. For example, a price of €0.20 is in general preferred over a price of €0.40 when two products are identical. To generate the most efficient design, it is important to assign negative utility to the higher prices in order to avoid uninformative choice sets.

Huber and Zwerina (1996) investigated how the gains from utility balance change if the assumed prior values are incorrect. They concluded that there are losses in the gain from utility balance. However, as bad as the losses may be, they are still substantially smaller than the efficiency losses from disregarding priors altogether. Thus, when the parameters are in the right direction, the analyst is better of being wrong about prior values of β than disregarding them completely. When the prior estimates of β have a nonzero value, the first two criteria for efficient choice designs are no longer valid. They are no longer properties of an efficient choice design. Hence, the combination of choice sets will be different compared to the state where β is equal to 0, because when β is equal to zero, the probability that an alternative will be chosen is equal for all alternatives. When β has a nonzero value, level balance and orthogonality will probably not be satisfied anymore. However, it will be attempted to generate a choice design that will come close to both criteria.

3.3.3 Prior values of price parameter.

In this research, the design could be improved with regards to the attribute *price*. Lower prices are in general preferred over higher price (Huber & Zwerina, 1996). When having two alternatives where all attribute levels are the same except the prices, the one with the lowest price is probably chosen most of the times. Thus, this would be an uninformative choice set. To solve this, *utility balance* will be used as being discussed above. To create informative choice sets, the prior values of β for the price variable should be determined. Previous research done by Enneking, Neumann and Henneberg (2007) about soft drinks concluded that higher prices result in a decrease in utility compared to lower prices. Sorenson and Bogue (2006) also modelled consumer preference for soft drinks. They used a rating based conjoint analysis with a logit model. They concluded that lower prices are preferred over higher prices as well. Basing the prior values of β on the research from Sorenson and Bogue (2006) is not possible due to the fact that they used a different type of conjoint analysis and different prices. Basing the prior values of β on the study of Enneking, Neumann and Henneberg (2007) is not optimal as well, because they used different prices in their

research compared to my research. Therefore, the parameter estimates of their study are probably not the same that they will be in my research. It can be concluded that estimating prior values of β is a difficult task, because the β values are unknown prior to the research. The partworths can only be roughly estimated before the experiment (Huber & Zwerina, 1996).

Previous research was not comparable enough to base the values of β on. Therefore, another method should be used to base the prior β values on. One way of addressing this problem is by estimating the probabilities for a choice experiment and use the β's from that experiment as prior values of β. This will be done with the use of equation 3. With this formula it is possible to calculate the chances that a participant will choose an alternative within a choice set. When these probabilities are not too extreme, the proposed values of β will be used. The following prior values will be used: €0.20 will have value 0 which is the reference level, €0.40 will have a prior value of $-\frac{1}{3}$, €0.60 will have a prior value of $-\frac{2}{3}$ and €0.80 a prior value of -1. Assigning these values mean that the prices of €0.80, €0.60 and €0.40 all have negative utility compared to the reference level where €0.80 is the least preferred and €0.20 the most preferred attribute level. Now, imagine a choice set with four different choices where all the attributes have the same level except the prices differ. An example of such a choice set can be found in the table below (Table 2).

Table 2

Example choice set

Alternative	**Packaging**	**Type of sweetener**	**Brand**	**Recyclability**	**Price**
1	Aluminum can	With Sugar	New brand	Recyclable	€0.20
2	Aluminum can	With Sugar	New brand	Recyclable	€0.40
3	Aluminum can	With Sugar	New brand	Recyclable	€0.60
4	Aluminum can	With Sugar	New brand	Recyclable	€0.80

When examining Table 2, it becomes clear that a respondent would always choose alternative 1. This alternative is the same as the other alternatives, but it has the lowest price. Hence, this is a dominant alternative within this choice set. A dominant alternative is an alternative that will always be chosen within a choice set by the participants. This will be solved by using prior values of β which has been discussed earlier. To determine if the earlier proposed values of β are appropriate,

the probabilities that an alternative is being chosen in this choice set for all the alternatives according to equation 3 will be calculated. The results of the calculations can be found in Table 3.

Table 3

Alternative probabilities

Alternative	**Calculation**	**Result**
1	$\frac{e^{0}}{e^{0} + e^{-\frac{1}{3}} + e^{-\frac{2}{3}} + e^{-1}}$	0.3849
2	$\frac{e^{-\frac{1}{3}}}{e^{0} + e^{-\frac{1}{3}} + e^{-\frac{2}{3}} + e^{-1}}$	0.2758
3	$\frac{e^{-\frac{2}{3}}}{e^{0} + e^{-\frac{1}{3}} + e^{-\frac{2}{3}} + e^{-1}}$	0.1976
4	$\frac{e^{-1}}{e^{0} + e^{-\frac{1}{3}} + e^{-\frac{2}{3}} + e^{-1}}$	0.1416

The results indicate that there are no extreme probabilities when using the proposed values of β. It can be concluded that the chances of an alternative being chosen in the example decreases when the value of β becomes more negative. However, specifying prior values of β will lead to more efficient choice designs compared to having al β's equal to 0. As being stated earlier, the analyst is better of being wrong about prior (except $\beta > 0$ is worse) values of β than disregarding them completely (Huber and Zwerina, 1996).

3.3.4 Choice design in this research.

The choice set used for this research can be found in Appendix 1. The design has been generated with JMP which is a statistical program. With JMP it is possible to generate choice designs for a choice-based conjoint analysis. JMP makes the choice design according to the research of Huber and Zwerina (1996) which has been discussed earlier in the method section. To make the design, JMP optimizes the D_p error (Huber & Zwerina, 1996). Furthermore, the proposed prior values of β for the price attribute discussed in section 3.3.3 have been used to solve the problem of uninformative choice sets. The values of β will be equal to zero for all the other

attributes used in this research, because there was no expectancy for the preference order within these attributes beforehand.

3.4 Willingness-to-pay

The willingness to pay is the maximum amount that a consumer is willing to pay for a product or service. It will be calculated for all the attributes in the models that will be used in this research. The WTP is necessary to answer sub-question 4. The WTP is often used instead of the part-worth estimates in conjoint analysis, because it is a measure of relative importance of the attributes. The average WTP can be derived for a particular level of an attribute by dividing the coefficient for that level by the numerical price coefficient (Chapman & McDonnell, 2015).

3.5 Likelihood-ratio test

This research will contain multiple multinomial logit models. The first model is the base model with the attributes only. The second model is the base model with interactions between the attributes and age groups. The third model is the base model with the interaction between attributes and how often someone consumes soft drinks. The last model is the base model with the interactions between attributes and when someone consumes soft drinks. The likelihood-ratio test will be used to test whether or not the age groups, how often someone consumes soft drinks and when someone consumes soft drinks have a significant effect on the choices people make. The likelihood-ratio test compares two models where one model has a subset of the parameters of the other model (Chapman & McDonnell, 2015). The tests will be performed between model 1 and 2, 1 and 3, 1 and 4.

3.6 ROC curve

The four models that will be created can all be used to predict whether or not someone will choose an option or not. For this research, the data will be randomly divided into 70% training data and 30% test data. The training and test data will only be used for prediction, the entire dataset will be used to create the final models. The prediction performance can be assessed with the use of the Receiver Operator Characteristic (ROC) and the AUC (area under the curve). The ROC curve is a popular graphic to simultaneously display the true positive and false positive rates. This research has a ‘choose 1 from 4’ setting. Therefore, an adaption has been made to be able to use the ROC curve for this research. The choice variable in this research equals 1 if an alternative within a choice set is chosen and 0 otherwise. All observations in the test set were predicted with the training data. Out of this, the probability that an observation would belong to class 1 or 0 was extracted. Thus,

the multiclass classification was transformed to binary classification in order to be able to use the ROC curves. The classification went as follows. For example, one could choose a threshold of 50% and classify all predicted values lower than 50% belonging to class 0 and all observations with a predicted value higher than 50% belonging to class 1. After doing this, the number of true positives and false positives can be calculated. The same can be done for another threshold, for example a threshold of 10%, meaning that the researcher wants more certainty of an observation belonging to class 1. When using this threshold, all observations with predicted values higher than 10% will be in class 1 and all observations with a predicted value lower than 10% will belong to class 0. Again, the true positives and false positive rate can be calculated. This process can be repeated for every value between 0 and 100. When this is being done, all the true positive and false positive rates can be plotted against each other. This plot is being referred to as the ROC curve.

Often, the AUC is used as a simple metric to define how an algorithm performs over the whole space (Davis & Goadrich, 2006). An ideal ROC curve will hug the top left corner, so the larger the AUC the better the classifier (Chapman & McDonnell, 2015). An AUC of 0.5 means that the model is not better than a random guess. It should be interpreted in this way, an AUC of 0.50 means that the model classifies 50% of all observations correctly irrespective of the threshold being chosen. AUC values below 0.5 are even worse than a random guess. A perfect model would have an AUC equal to 1 (Kwartler, 2017; Chapman & McDonnell, 2015; Davis & Goadrich, 2006).

3.7 Variable selection

This research will contain four different models. One base model with the attributes only. The base model including the interactions with the different generations. A base model including the interaction with the consumption frequency and a base model including the interaction with the consumption moment. The last three models will be large compared to the sample size. In order to prevent overfitting, a variable selection method is required. For this research, backward selection based on the p-value will be used. To arrive at a model with fewer parameters, the full model with p parameters will be estimated first. Next, the variable with the largest p-value will be removed if this p-value is larger than 0.10. Next, the new model with p-1 parameters will be estimated. Again, the parameter with the largest p-value will be removed if the value is larger than 0.10 and the model with p-2 parameters will be estimated. This process repeats itself until all variables are significant at the 10% level. First, the variable selection will be done for all the models based on the training data in order to assess their prediction performance with the use of the ROC curves. The final

models will be based on the entire dataset and the variable selection will be done again for these models.

4. Data

4.1 Attributes and choice sets

The data being used in this research has been collected through a survey. The survey contained 10 choice sets with four alternatives per choice set (Appendix 1). The choice sets have been made with JMP which is a statistical software program. The used attributes and attribute levels are presented in Table 4. The respondents had to choose their preferred alternative out of the presented choice sets which were presented in a random order. At the end of the survey, some demographic questions were asked to gain knowledge about the sample. Furthermore, questions about how often someone consumed soft drinks and when someone consumed soft drinks were asked. These questions will be used to make segments across the data.

Table 4

Attributes and attribute levels

Packaging	**Type of Sweetener**	**Brand**	**Recyclability**	**Price**
Aluminum can	With Sugar	New brand	Recyclable	€0.20
Plastic bottle	Without Sugar	Top brand	Not recyclable	€0.40
Glass bottle	Artificial sweetener	Private label		€0.60
	Natural sweetener			€0.80

4.2 Sample

The used data for this research is obtained in the period from 13 July till 24 august 2020. The survey was distributed amongst friends, family and fellow students. Furthermore, social media has been used to distribute the survey as well. This led to a total number of 376 responses. After deleting incomplete responses, 350 responses were suitable for this research. The sample consists out of 248 women and 102 men. For this research, it is important to obtain participants from varying ages, to be able to have meaningful results about choices made by different generations. The sample consists out of 110 participants born between 1946 and 1964, 116 participants born between 1965-1977, 68 participants born between 1978 and 1994 and 56 participants were born between 1995 and now. When looking at the participant's education, 11 participants did VMBO, 32 participants did HAVO/VWO, 62 participants did MBO and 245 participants did HBO/Universitair. The data

from the sample will be used to conduct the conjoint analysis as described in the methodology. The statistical software program “R” will be used to do this.

5. Results

This chapter presents the results of the CBC in order to give answer to the following sub-questions (sub-question 3 and 4): "What are the most important attributes that determine the consumers preference?" and "How does the Willingness to pay (WTP) differ across different attributes and segments?". Both questions help to give answer to the main research question which will be answered in the conclusion. This section will first provide all the four different models. Following, the ROC curves and likelihood-ratio tests will be presented. Finally, both sub-questions will be answered.

5.1 Base model

The collected data has been analyzed with the use of the multinomial logit. The parameters are all estimated with respect to the reference level of an attribute. First, the parameters and the WTP for the base model have been estimated. The base model is the model with the attributes only as parameters (Table 5). From the results it is clear that all parameters are significant at the 0.001 significance level. The results from the base model show that participants derive the most utility out of a soft drink in a glass bottle without sugar, branded as top brand and the product needs to be recyclable. Besides, the price parameter shows a negative relation between the price and the utility that a person derives out of a soft drink, meaning that higher prices make the soft drink less attractive.

To be able to calculate the WTP, the price variable is treated as a numerical variable. This makes it possible to draw conclusions beyond the used price range used in this research (€0.20-€0.80). This is reasonable, because the distances between the prices are equal, meaning that a linear relationship between prices can be assumed. The WTP indicates that "With Sugar", "Artificial Sweetener" and "Recyclable" have the most effect on the participants' utility compared to the other attribute levels. Furthermore, it should be noted that attribute levels with a negative WTP lead to a lower WTP. The WTP should be interpreted with respect to the reference level of an attribute. Consider, for example, a person who is willing to pay x euro for a soft drink with an aluminum can. This person will on average pay x + €0.462 for a soft drink out of a plastic bottle compared to the aluminum can soft drink. From the WTP results, it is clear that some WTP's are relatively large compared to other ones. For example, a person is willing to pay 1.949 euro less for a soft drink per 330ml with sugar compared to a soft drink without sugar. The WTP's with the largest effect are compared to the used prices in this research not unrealistic. However, paying 1.949 euro less for a

soft drink without sugar compared to a soft drink with sugar per 330 ml is a large amount when considering that some soft drinks cost 0.50 euro per 330 ml. Stated differently, a consumer is indifferent between a soft drink with sugar and a product without sugar priced 1.949 euros cheaper.

Table 5

Estimation results and WTP base model

	Estimate	Std.error	WTP
Glass bottle	0.357***	0.06	0.462
Plastic bottle	0.212***	0.051	0.274
Aluminum can	Reference	-	-
With Sugar	-1.506***	0.06	-1.949
Natural Sweetener	-0.529***	0.049	-0.684
Artificial Sweetener	-1.431***	0.064	-1.852
Without Sugar	Reference	-	-
Private label	-0.567***	0.065	-0.734
New brand	-0.341***	0.048	-0.441
Top brand	Reference	-	-
Recyclable	1.312***	0.047	1.700
Not Recyclable	Reference	-	-
Price	-0.773***	0.164	-

., significant at $p < 0.10$
*, significant at $p < 0.05$
**, significant at $p < 0.01$
***, significant at $p < 0.001$

5.2 Base model including age

The second model that has been estimated with the multinomial logit is the base model including interactions with the age groups (Table 6). This research uses four different age groups in order to be able to segment customers based on the generation. To arrive at this model, backward selection based on the p-value ($p<0.10$) has been used. The full model contained 36 variables including all interactions between the attributes from the base model and the age groups. After the variable selection, the model contains 18 variables, meaning that the variable selection method reduced the number of parameters with 18. The estimation results in Table 6 show that most of the parameters are significant at the 0.001 level.

In Table 7, the WTP per generation is displayed for each variable. In general, "Recyclable", "Artificial Sweetener" and "With Sugar" have the largest WTP. Hence, these attributes are the most important attributes compared to the other attributes. Again, the WTP's for these attributes are relatively large compared to the prices for soft drinks per 330 ml. For example, a person is

willing to pay 2.19 euro more for a recyclable soft drink compared to a non-recyclable soft drink. Table 7 shows some differences in the WTP between the four different generations. The generation born between 1946 and 1964 has the highest WTP for a Glass bottle compared to an aluminum can. This generation experiences the largest decrease in WTP when a soft drink contains sugar compared to a soft drink without sugar. Moreover, they have the highest WTP for recyclable soft drinks compared to a non-recyclable soft drink. This generation prefers a glass bottle without sugar, branded as top brand and the product should be recyclable.

Table 7 shows no remarkable differences between the generation born between 1965 and 1977 and the other generations. This generation has a strong negative WTP for all sort of sweeteners compared to the reference level (Without Sugar), but they have the largest decrease in WTP and utility (Table 6) for a soft drink with sugar compared to a soft drink without sugar. Furthermore, a private label and new brand both lead to a decrease in utility and WTP compared to a top brand. Lastly, this generation highly values recyclability. Thus, this generation prefers a glass bottle without sugar, branded as top brand and the product should be recyclable.

The third generation, born between 1978 and 1994 display more or less the same WTP's as the previous generations. They have negative WTP for all the sweeteners compared to a product without sugar. They assign the most utility and highest WTP to a top brand. The utility and WTP is high compared to other utilities and WTP's, meaning that recyclability is important for this generation. In the end, they have exactly the same preference as the previous generations, meaning that they prefer a glass bottle without sugar, branded as top brand and the soft drink must be recyclable.

The last generation used in this research, the generation born between 1995 and now display other preferences compared to the other generations. Opposing to the other generations, they have negative WTP for a glass bottle and prefer a plastic bottle. Furthermore, all sweeteners have a negative WTP compared to without sugar. In line with the other generations, the utility and WTP for a recyclable product are high compared to the other attribute levels. Additionally, the estimation results from Table 6 display a negative interaction between the price and the generation born between 1995 and now. The other generations have a price coefficient of -0.572, meaning that if the price increases with €1, the utility decreases with 0.572. The generation born between 1995 and now have a total negative utility effect of -1.707, meaning that if the price increases with €1 and everything else remains the same, the utility decreases with 1.707. This is almost three times larger

compared to the other groups. This means that this generation is more price sensitive in comparison with the other generations.

Appendix 2. Table 1 shows an overview of the preferences for all the generations. In general, the preferences remain stable over the different generations, meaning that they do not differ much. The generation born between 1995 and now prefers another type of packaging while the other generations prefer the same packaging. Moreover, there appear to be differences in the WTP between the generations for a glass bottle, but the differences are small. Some generations have a stronger preference for a soft drink without sugar compared to other generations, however they all prefer a soft drink without sugar. Lastly, the generation born between 1946 and 1964 value recyclability more compared to other generations.

Table 6

Estimation results and WTP base model including age

	Estimate	Std.Error
Glass bottle	0.292***	0.077
Glass bottle:1946-1964	0.462**	0.117
Glass bottle:1995-now	-0.314*	0.14
Plastic bottle	0.211***	0.052
Aluminum can	Reference	-
With Sugar	-1.334***	0.112
With Sugar:1946-1964	-0.758***	0.152
With Sugar:1965-1977	-0.309*	0.139
With Sugar:1995-now	0.644***	0.172
Natural Sweetener	-0.584***	0.054
Natural Sweetener:1995-now	0.284*	0.134
Artificial Sweetener	-1.611***	0.073
Artificial Sweetener:1995-now	0.884***	0.163
Without Sugar	Reference	-
Private label	-0.56***	0.065
New brand	-0.357***	0.048
Top brand	Reference	-
Recyclable	1.254***	0.056
Recyclable:1946-1964	0.282**	0.099
Not Recyclable	Reference	-
Price	-0.572**	0.184
Price:1995-now	-1.135**	0.397

., significant at $p < 0.10$
*, significant at $p < 0.05$
**, significant at $p < 0.01$
***, significant at $p < 0.001$

Table 7

Willingness to pay per generation

	1946-1964	1965-1977	1978-1994	1995-now
Glass bottle	1.319	0.511	0.511	-0.013
Plastic bottle	0.369	0.369	0.369	0.124
Aluminum can	Reference	Reference	Reference	Reference
With Sugar	-3.66	-2.87	-2.33	-0.404
Natural Sweetener	-1.020	-1.020	-1.020	-0.176
Artificial Sweetener	-2.820	-2.820	-2.820	-0.455
Without Sugar	Reference	Reference	Reference	Reference
Private label	-0.979	-0.979	-0.979	-0.328
New brand	-0.624	-0.624	-0.624	-0.209
Top brand	Reference	Reference	Reference	Reference
Recyclable	2.682	2.190	2.190	0.900
Not Recyclable	Reference	Reference	Reference	Reference

5.3 Base model including consumption frequency

The third model which has been estimated is the base model including interactions from the consumption frequency. Again, backward selection based on the p-value ($p<0.10$) is used. The full model contained 45 variables. After the variable selection, the model contains 24 variables. The variable selection led to the removal of the main effects of the bottle types, therefore the effect of glass bottle and plastic bottle is equal to zero for the reference category (daily). The interaction between the packaging and consuming multiple times per day has been removed by the variable selection as well, consequently this effect is estimated to be equal to zero.

Table 9 displays the WTP results categorized by the consumption frequency. The variables with the largest effect on the utility and WTP are “With Sugar”, “Artificial Sweetener” and “Recyclable”. These WTP’s are large compared to the prices of soft drinks per 330 ml. The other WTP’s are reasonable, in general. For example, a consumer born between 1965 and 1977 is willing to pay 0.51 euro more for a glass bottle compared to an aluminum can. This is reasonable , because it is in line with the used prices in this research.

The consumers who never consume soft drinks value a plastic bottle more compared to the other generations (Table 9). They have a lower WTP for a soft drink with sugar or an artificial sweetener compared to the other generations. They are the only group with a positive utility for “Private label” and “New brand”. They have the highest WTP for a soft drink branded as private label. This is remarkable, because all the other groups prefer the top brand. This group of consumers values recyclability more than other generations, because they have the highest WTP and utility

for a recyclable product in comparison with the other generations. Overall, this group prefers a glass bottle without sugar, branded as private label and the soft drink should be recyclable.

The second group is the group that consume soft drinks once a month. It is directly noticeable that this group has the highest negative WTP for a natural sweetener compared to the other groups. Overall, they do not prefer sweeteners in their soft drink, just as the other generations. Furthermore, the utility and WTP are both negative for private label and new brand, meaning that this group prefers a top brand. Regarding recyclability, this group prefers a recyclable soft drink. All in all, this group prefers a glass bottle without sugar, branded as top brand and the product should be recyclable.

The consumers who consume on a weekly basis have a higher WTP for a plastic bottle compared to the other packaging types. This group has another preference with regards to packaging. In accordance with the other groups, they assign negative utility and WTP to all sort of sweeteners. Furthermore, they have a negative WTP for the private label and new brand, meaning that they prefer a top brand. This group highly values recyclability just as the other groups. These consumers prefer a plastic bottle without sugar, branded as top brand and the product should be recyclable.

The consumers who consume on a daily basis have negative utility and WTP for all sort of sweeteners and they highly value recyclability. For this group, there is no information regarding their preference around the packaging type, due to the fact that the variable selection led to the removal of the coefficients for this group. However, other coefficients of this group are available and the results show that this group prefers a soft drink without sugar, branded as top brand and the product should be recyclable.

The last group is the group who consume multiple times per day. For this group, there is no information about the packaging as well, due to the same reason as with the precious group. The preference of this group is in line with the previous group and does not show any extraordinary results.

For an overview of the preferences per group, see Appendix 2. Table 2 The table shows that in general the preferences remain the same over all the groups, except for the consumers who never consume soft drinks and the consumers who consume on a weekly basis. Lastly, the price coefficient is negative for all the generations, meaning that higher prices lead to a decrease in utility.

Table 8

Estimation results base model including consumption frequency

	Estimate	Std.Error
Glass bottle	0	-
Glass bottle:Never	0.906***	0.207
Glass bottle:OnceAMonth	0.417***	0.093
Glass bottle:Weekly	0.357***	0.101
Plastic bottle	0	-
Plastic bottle:Weekly	0.375***	0.087
Aluminum can	Reference	-
With Sugar	-1.131***	0.104
With Sugar:Never	-0.845***	0.23
With Sugar:OnceAMonth	-0.626***	0.146
With Sugar:Weekly	-0.383**	0.133
Natural Sweetener	-0.443***	0.06
Natural Sweetener:OnceAMonth	-0.256*	0.103
Artificial Sweetener	-0.95***	0.125
Artificial Sweetener:Never	-1.359***	0.308
Artificial Sweetener:OnceAMonth	-0.798***	0.174
Artificial Sweetener:Weekly	-0.624***	0.16
Artificial Sweetener:MultipleTimesPerDay	0.574*	0.232
Without Sugar	Reference	-
Private label	-0.613***	0.066
Private label:Never	0.741*	0.295
New brand	-0.568***	0.085
New brand:Never	0.589**	0.193
New brand:OnceAMonth	0.341**	0.112
New brand:Weekly	0.205.	0.108
Top Brand	Reference	-
Recyclable	1.305***	0.049
Recyclable:Never	0.664**	0.203
Not Recyclable	Reference	-
Price	-0.831***	0.165

., significant at $p < 0.10$
*, significant at $p < 0.05$
**, significant at $p < 0.01$
***, significant at $p < 0.001$

Table 9

Willingness to pay segmented by consumption frequency

	Never	Once a month	Weekly	Daily	Multiple times per da
Glass bottle	1.090	0.502	0.429	-	-
Plastic bottle	-	-	0.451	-	-
Aluminum can	Reference	Reference	Reference	Reference	Reference
With Sugar	-2.38	-2.113	-1.821	-1.360	-1.360
Natural Sweetener	-0.533	-0.841	-0.533	-0.533	-0.533
Artificial Sweetener	-2.78	-2.1	-1.891	-1.140	-0.449
Without Sugar	Reference	Reference	Reference	Reference	Reference
Private label	0.154	-0.738	-0.738	-0.738	-0.738
New brand	0.026	-0.273	-0.436	-0.683	-0.683
Top brand	Reference	Reference	Reference	Reference	Reference
Recyclable	2.369	1.570	1.570	1.570	1.570
Not Recyclable	Reference	Reference	Reference	Reference	Reference

5.4 Base model including consumption moment

The fourth model which has been created is the base model including interactions between the attributes and the moment of consumption. At first, the full model was made and contained 72 variables. To decrease the number of parameters, backward selection based on the p-value is used ($p<0.10$). The final model contains 42 variables. Table 11 shows an overview of the WTP per consumption level. It can be seen that some WTP's are relatively large compared to the other ones and even large compared to the WTP's in the previous models. This model contains relatively extreme WTP's compared to the other models. For example, a person is willing to pay 8.98 euro less for a soft drink with sugar compared to a soft drink without sugar during breakfast. Stated differently, a consumer is indifferent between a soft drink with sugar and a product without sugar priced 8.98 euros cheaper. Obviously, this is not a realistic WTP, a consumer would probably never pay that amount more for a soft drink of 330 ml. However, this value indicates that some consumers just dislike soft drinks with sugar. Consumers who never consume soft drinks are willing to pay 2.32 euro for a recyclable soft drink compared to a non-recyclable soft drink. These consumers are

indifferent between a soft drink that is recyclable and a soft drink that is not recyclable priced 2.32 euros cheaper. This amount is not realistic as well, because it is relatively large compared to the used prices in this research (0.20, 0.40, 0.60, 0.80).

Table 11 shows that consumers who consume during breakfast have the highest WTP for a glass bottle. Furthermore, it is noticeable that the WTP's for with sugar and artificial sweetener are large compared to other ones, meaning that a soft drink with sugar or an artificial sweetener during breakfast are highly disliked. The other sweetener systems are not feasible as well, their WTP's are large too.

The breakfast consumers have a high utility and WTP decrease compared to the other groups when the soft drink is branded as private label. Recyclability is less important compared to other attributes, because the absolute value of the WTP is small compared to the other ones. Overall, these consumers have a preference for a recyclable soft drink. They prefer a glass bottle without sugar, branded as top brand and the product should be recyclable.

The second group are the consumers who consume soft drinks during lunch. They prefer a glass bottle slightly more than an aluminum can. Their preferences are less strong compared to the previous group. However, all the sweeteners have a negative WTP. These consumers prefer a soft drink without sugar. Regarding the branding, they have smaller WTP's compared to the previous group. Nonetheless, they do not prefer a private label or new brand. This group prefers a glass bottle without sugar branded as top brand and the product should be recyclable.

The people who consume during lunch prefer a glass bottle without sugar, branded as top brand and the product should be recyclable. The consumers who consume during dinner have a slightly different preference compared to the previous groups. First of all, they have the highest WTP for a plastic bottle. They do not prefer any type of sweetener. They assign more negative utility to a new brand compared to a private label which is opposed in comparison with the groups before. However, they have the same preference, namely a top brand. This group has almost the highest WTP for a recyclable product compared to the other groups. This group prefers a plastic bottle without sugar, branded as top brand and the product should be recyclable.

The consumers who consume during sports prefer a glass bottle above all other packaging types. In line with the previous groups, they do not want any type of sweetener. Overall, they prefer a glass bottle without sugar, branded as top brand and the product should be recyclable.

Next, the consumers who are on the way have a strong preference and high WTP for a glass bottle. This preference is stronger compared to the other groups. It is remarkable that these consumers have relatively low WTP's for the private label and new brand. They are almost indifferent regarding the brand. In contrast to the other groups, they prefer a new brand. They also have a relatively high WTP for a recyclable product compared to the other groups. In general, this group prefers a glass bottle without sugar, branded as private label and the product should be recyclable.

Next, the consumers who consume during work have a different preference regarding the packaging. They are the only group with negative WTP's for glass bottle and plastic bottle both, meaning that they prefer the aluminum can. Additionally, they have negative WTP's for all sweeteners and they dislike the private label and new brand more compared to other groups. This group prefers an aluminum can without sugar, branded as top brand and the product should be recyclable.

The consumers who consume in their spare time have positive utility for the glass bottle and plastic bottle. However, the utility and WTP is higher for the glass bottle. Furthermore, they prefer a product without sugar and they dislike other brands than the top brand. They prefer a glass bottle without sugar branded as top brand and the product should be recyclable.

The last group are the consumers who do not consume soft drinks. It might sound strange to have preferences about a group who do not consume soft drinks. In fact, knowing the preferences of a group of potential buyers can be interesting. This group prefers a glass bottle and dislikes the plastic bottle more compared to the other groups. They also highly dislike soft drinks with sugar. It is interesting to notice that this group has the highest WTP for recyclability compared to the other groups. In general, the preferences of this group are more or less the same as of the other groups. They prefer a plastic bottle without sugar, branded as top brand and the product should be recyclable. In general, Appendix2. Table 3 shows that the preferences remain the same with exceptions for consumers who consume during work or during dinner. Besides, as being discussed, some WTP's are larger than other ones.

Table 10

Estimation results base model including consumption moment

	Estimate	Std.Error
Glass bottle	0.493***	0.071
Glass bottle:Lunch	-0.456*	0.212

Glass bottle:On the way	0.37*	0.172
Glass bottle:During work	-0.711***	0.175
Plastic bottle	0.291***	0.065
Plastic bottle:Lunch	-0.515**	0.187
Plastic bottle:Dinner	0.263*	0.126
Plastic bottle:During sports	-0.422**	0.163
Plastic bottle:On the way	0.439**	0.145
Plastic bottle:During work	-0.375*	0.149
Plastic bottle:Not	-0.685***	0.188
Aluminum can	Reference	-
With Sugar	-1.729***	0.074
With Sugar:Breakfast	-5.817**	2.109
With Sugar:Dinner	0.809***	0.135
With Sugar:During work	0.718**	0.165
With Sugar:Not	-0.502*	0.208
Natural Sweetener	-0.576***	0.054
Natural Sweetener:Breakfast	-1.508*	0.737
Natural Sweetener:During work	0.276.	0.151
Artificial Sweetener	-1.61***	0.079
Artificial Sweetener:Breakfast	-2.792.	1.466
Artificial Sweetener:Dinner	0.312.	0.165
Artificial Sweetener:During work	0.923***	0.179
ArtificialSweetener:Not	-0.864**	0.287
Without Sugar	Reference	-
Private label	-0.52***	0.078
Private label:Breakfast	-3.817*	1.56
Private label:On the way	0.488**	0.177
Private label:During work	-0.463**	0.168
New brand	-0.256***	0.057
New brand:Breakfast	-1.62.	0.914
New brand:Dinner	-0.268*	0.119
New brand:On the way	0.287*	0.131
New brand:During work	-0.54***	0.139
Top brand	Reference	-
Recyclable	1.336***	0.06
Recyclable:Lunch	-0.382*	0.174
Recyclable:Dinner	0.307*	0.133
Recyclable:On the way	0.235.	0.133
Recyclable:During work	-0.484***	0.141
Recyclable:Not	0.609***	0.176
Not Recyclable	Reference	-
Price	-0.841***	0.176
Price:Breakfast	8.357.	4.289

Price:During sports	-0.989.	0.531

., significant at $p < 0.10$
*, significant at $p < 0.05$
**, significant at $p < 0.01$
***, significant at $p < 0.001$

Table 11

Willingness to pay segmented by consumption moment

	Breakfast	Lunch	Dinner	During sports	On the way	During work	Spare time	Not
Glass bottle	0.587	0.044	0.587	0.587	1.027	-0.258	0.587	0.587
Plastic bottle	0.346	-0.267	0.659	-0.156	0.868	-0.1	0.346	-0.469
Aluminum can	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
With Sugar	-8.98	-2.060	-1.098	-2.060	-2.060	-1.206	-2.060	-2.657
Natural Sweetener	-2.475	-0.685	-0.685	-0.685	-0.685	-0.357	-0.685	-0.685
ArtificialSweetener	-5.24	-1.920	-1.548	-1.920	-1.920	-0.82	-1.920	-1.03
Without Sugar	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Private label	-5.158	-0.618	-0.618	-0.618	-0.037	-1.168	-0.618	-0.618
New brand	-2.234	-0.304	-0.622	-0.304	0.037	-0.947	-0.304	-0.304
Top brand	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Recyclable	1.590	1.135	1.955	1.590	1.869	1.014	1.590	2.315
Not Recyclable	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference

5.5 Prediction performance

To assess the prediction performance of all four models, the data has been randomly divided into a train set (70%) and test set (30%). First, models were made with the train set. Next, the variable selection has been performed on these models. Then, the models were used to predict the test data. The models presented in the results are based on the whole dataset. The variable selection has been done again for these models. To assess the performance, ROC curves have been made for each of the four models based on the training data and the AUC measures were used to get an understanding of their performances. All four ROC curves are displayed in figure x. The AUC of the base model equals 0.813. This means that the model predicts 81.3% of the test data correctly irrespective of what classification threshold is chosen. The base model including interactions with age has an AUC of 0.816. Meaning that the model predicts 81.6% of the test data correctly irrespective of what classification threshold is chosen. The base model including the interactions with consumption frequency has an AUC of 0.813. Meaning that the model predicts 81.3% of the test data correctly irrespective of what classification threshold is chosen. The last model, the base

model including the consumption moment has an AUC of 0.819. Meaning that this model predicts 81.9% of the test data correctly irrespective of what classification threshold is chosen. All these values are reasonably good, because it is better than the scenario where the AUC would be 0.50 meaning that the model is not performing better than a tossing a coin. The graphs and the AUC indicate that the models all perform similarly with some minor differences. The differences between the ROC curves are small because the predicted probabilities do not differ much between all four models.

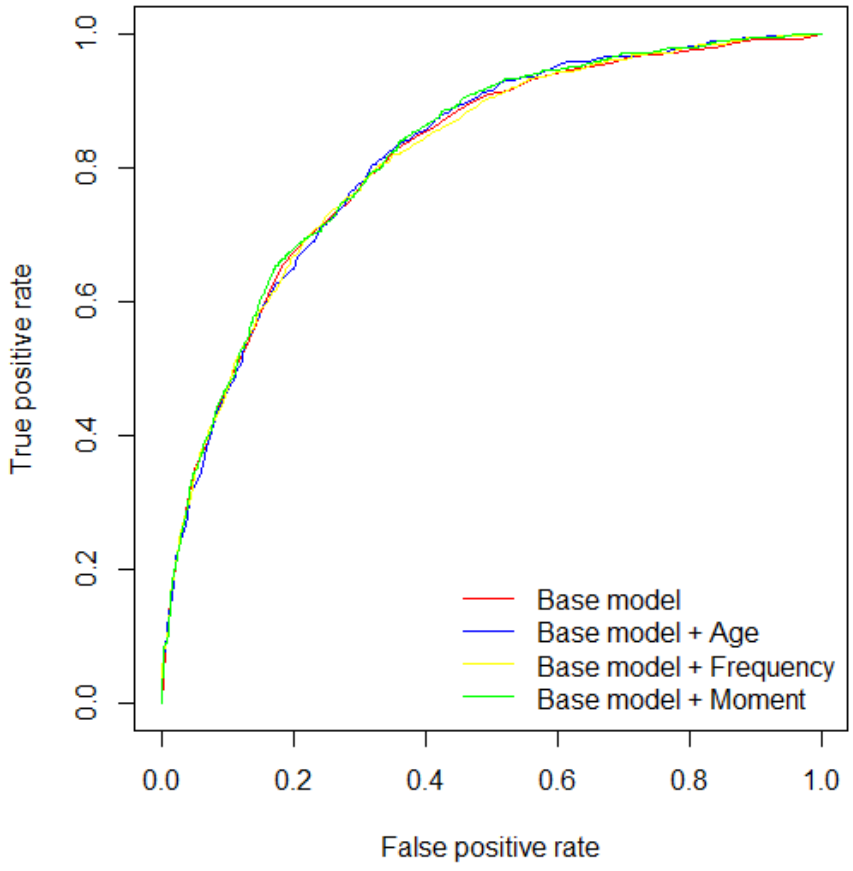

Figure 1. ROC curves

5.6 Likelihood-ratio tests

To test whether or not the base model including age, base model including consumption frequency and the base model including consumption moment are significantly different from the base model, three likelihood ratio-tests were performed as being described in the methodology. All the tests showed that the models were significantly different compared to the base model, meaning that the models including the segmentation variables fit the data better.

5.7 Answers to sub-questions 3 and 4

At this point in the research, sub-questions 3 and 4 can be answered. First sub-question 3 will be answered and finally sub-question 4.

5.7.1 What are the most important attributes that determine the consumers preference?

The answer to this question will be based on the WTP's for all the four models. In general, the attribute levels with higher WTP's compared to the other WTP's are deemed to be the most important attribute levels. According to the base model without any form of segmentation, the most important attributes are "With Sugar", "Artificial Sweetener" and "Recyclable". When segmenting the data per generation, the same variables remain the most important. Segmentation based on the consumption frequency leads to the same answer. Lastly, the model including the consumption moment shows the same variables as the most important ones. In all four models, the WTP's for "With Sugar" and "Artificial Sweetener" are negative compared to the base level (Without Sugar). "Recyclable" has a positive WTP in all four models. Hence, the most important attributes are "With Sugar", "Artificial Sweetener" and "Recyclable".

5.7.2 How does the Willingness to pay (WTP) differ across different attributes and segments?

In general, the highest WTP is achieved when the soft drink is made out of a glass bottle without sugar, branded as top brand and the product should be recyclable. When taking consumers ages into account, the general preference and direction of the WTP's remains the same. However, the generation born between 1995 and now prefers a plastic bottle, meaning that this attribute level has the highest WTP compared to the aluminum can and glass bottle. Moreover, it is noticeable (Table 7) that the generation born between 1995 and now have a lower negative WTP for all three sweetener systems compared to without sugar. This means that the aversion to the sweeteners is

less strong compared to the other generations. Furthermore, people born between 1946 and 1964 are willing to pay more for a recyclable soft drink compared to the other generations.

When taking consumption frequency into account, the general preference and thus the direction of the WTP's remains the same. Except for consumers who never consume soft drinks or once a month. They have a higher WTP for a glass bottle compared to the other packaging types. The people who consume on a weekly basis have the highest WTP for a plastic bottle. Furthermore, the people who never consume soft drinks have a higher WTP for the private label compared to a top brand. The consumers with other consumption frequency prefer a top brand. Furthermore, it is noticeable that people who consume less often, have a stronger negative WTP for all sort of sweeteners. Lastly, the consumers who never consume are willing to pay the most for a recyclable soft drink compared to the other generations.

Segmentation based on the consumption moment leads in general to the same general preference and highest WTP as being stated earlier, namely a glass bottle without sugar branded as top brand and the product should be recyclable. However, there are some differences between the different consumption moments. Consumers who consume during work of on the way prefer an aluminum can and plastic bottle respectively. Meaning that these consumers have higher WTP's for these preferences compared to an aluminum can. Furthermore, the people who consume during breakfast have the WTP for any sort of sweetener compared to a product without sugar. Meaning that soft drinks during breakfast are highly undesirable. Consumers who consume on the way have the highest WTP for a new brand in contrast to the other groups.

It can be concluded that the differences between the WTP's are present, however, the direction of the WTP is in most cases pointed in the same way. This means that the most preferred soft drink remains almost the same when examining this across different segments.

6. Conclusion

The soft drink consumption patterns have recently been changing by an increased availability of different types of drinks in the last few years (Redondo, Gomez-Martinez & Marcos, 2014). Soft drink producers have to react to this shift or otherwise this could lead to losing profit. The goal of this research was to investigate consumer preference towards soft drinks in the Netherlands in order to address the changing consumer patterns. This goal was achieved in the form of a marketing strategy for different segments. This marketing strategy is an advice for the soft drink industry aimed at selling in the supermarkets. In order to achieve the goal of this research, the following main-research question was made: "*How should the marketing strategy be made for soft drinks?"*. To answer this question, the research was divided into four sub-questions. The answers will be repeated below to be able to answer the main-research question in the end. The first two sub-questions were answered according to literature. The last two sub-questions were answered according to the results of this research.

The first sub-question was: *"Which attributes are important for soft drinks?"*. The following attributes were deemed to be important according to the literature: packaging, type of drink, brand, recyclability and price.

The second sub-question was needed in order to identify how consumers should be segmented in the soft drink market. The question was: *"In what way should consumers be segmented in the soft drinks market?"*. The first segmentation used in this research was based on generations. The following 4 different generations were used: 1946-1964, 1965-1977, 1978-1994 and 1995-now. Secondly, segmentation based on consumption frequency was used with the following levels: "never", "once a month", "weekly", "daily" and "multiple times per day". The last segmentation was based on the moment of consumption with the following levels: "breakfast", "lunch", "dinner", "during sports", "on the way", "during work", "spare time", "not".

The third sub-question was: *"What are the most important attribute levels that determine the consumers' preference?"*. This question was answered with the use of four different models which were made with the multinomial logit. The WTP's were computed for all attribute levels in all the models. It can be concluded that "With Sugar", "Artificial Sweetener" and "Recyclable" are the most important attribute levels for the consumers' preference. This is in line with the trends discussed earlier in the research. The national Frisdrankonderzoek (2017) concluded that the Dutch are becoming more aware of their intake of soft drinks. Consumers are choosing lower calorie and

or low sugar soft drinks. Thus, this trend is in line with the results from this research. Furthermore, sustainability is becoming more important (Kumar, Rahman, Kazmi & Goyal, 2012), therefore, it is not surprising that consumers have a high WTP for a recyclable soft drink. The goal of the industry is to become more sustainable, and according to this research, the consumers prefer a recyclable product is in line with the sustainability trend.

The fourth and final sub-question was: *"How does the Willingness to pay differ across different attributes and segments?"*. In general, the WTP is in most cases pointed in the same direction, meaning that the preference is in most cases the same. As being discussed in the results, there are some small differences in preferences. However, these differences are too small to create different strategies for different types of consumers. Therefore, the advice for the soft drink producers is to use one type of preferred drink as general preference according to this research. Based on the variables used in this research, segmentation is not recommended. Using other variables could lead to having more differences between different consumers.

With the use of the results in the previously answered research questions, it is now possible to give an answer to the following main-research question: :*"How should the marketing strategy be made for soft drinks?"*. Advise for a marketing strategy can be made with regards to the packaging, type of drink, brand, recyclability and price. As being stated earlier, the preferences remain almost the same across different segments. This holds for the segmentation by generation, consumption frequency and consumption moment. The expectation was that the segmentation techniques would lead to more difference in preferences. It could be the case that the three most important variables overruled the other variables and therefore, participants looked mostly at these attributes when choosing an option in the survey. The general preference is a soft drink in a glass bottle that does not contain sugar. The soft drink should be a top brand and the product should be recyclable. The advice for soft drink producers is to focus on this preference. The outcome that consumers prefer a glass bottle is somewhat remarkable. The research of Sorenson and Bogue (2006) showed that consumers in general prefer a plastic bottle which is not in line with the preference revealed in this research, the glass bottle. It might be the case that consumers first look at other attributes. It could be the case that the three most important attributes in this research("With Sugar", "Artificial Sweetener" and "Recyclable") overrule the other attributes. Hence, this could be the explanation why glass bottle appears to be the preferred packaging above the other packaging options.

This research revealed the most important attributes as well. These are "With Sugar", "Artificial Sweetener" and "Recyclable". In practice, this means that producers should focus on these attributes. This means that it is important to focus on products without sugar because a product with sugar leads to a substantial decrease in WTP. The same reasoning retains for the attribute level "Artificial Sweetener". Furthermore, it is important to make every soft drink recyclable because the WTP increases in all 4 different models when a product is recyclable. Therefore, producers should not overlook this aspect and make every soft drink recyclable. Producers should be careful with the pricing strategies. All four different models showed a relatively strong negative relation between the price and the utility compared to other utilities, meaning that an increase in price leads to a decrease in utility. This relation was even stronger for consumers born between 1995 and now, meaning that these consumers are even more price sensitive compared to other generations.

To conclude, the answer to the problem stated can be formulated as follows: the soft drink should be in a glass bottle and not contain sugar, branded as top brand and it needs to be recyclable according to this research. Furthermore, segmentation is not advised based on this research, due to the result that the differences in preference between different consumers are too small. Additionally, the focus for the soft drink producers should be on the most important attributes which are "With Sugar", "Artificial Sweetener" and "Recyclable", because these attributes lead to the highest changes in WTP. In practice, this means that they should focus on making products without sugar or artificial sweeteners and that the product needs to be recyclable.

6.1 Limitations and future research

This research has some limitations which will be discussed below. First of all, one major limitation lies in the fact that it is not possible to make exact pricing strategies with the used methodology in this research. Some WTP's are realistic but some are not. For example, it is not realistic that a person would pay 8.98 euros less for a soft drink with sugar compared to a soft drink without sugar during breakfast. The WTP's are in many cases large compared to the used prices in this research and therefore not realistic. For future research, I recommend using another method to estimate the WTP's. A research based on a real experiment and real sales data could lead to better insights.

Another limitation of this research is the used variable selection technique. The technique led to the removal of some main effects. Because of this, there was no information regarding the

packaging for all consumption moments. For future research I recommend using another variable selection method, because this might lead to other results. For example, a multinomial logit with an elastic net.

Another shortcoming of this research is the conjoint methodology. Conjoint analysis does not replicate the real choice task in the supermarket. In conjoint analysis, the product attributes are stated explicitly, whereas this is not always the case with products in the supermarket. Therefore, it is advised to create an environment which comes closer to the real choice task for future research.

This research used specific attributes and attribute levels. Other choices regarding the attributes and their levels might lead to different results. Therefore, it is recommended to use other attributes and or attribute levels in order to investigate this. Furthermore, the used method in this research was the multinomial logit where the parameters were estimated according to the maximum-likelihood. For future research, it is suggested to use another method to know if this leads to other results.

The models made in this research were all tested to be significantly different from the base model with the likelihood-ratio test. All models were significantly different compared to the base model. Then the prediction performance of all four models were tested. This resulted in similar prediction performances. This is a contradicting result. The models showed similar prediction performance but the models including interactions were significantly different from the base model. The results show in general similar preferences across all four models. This is in line with the fact that the prediction performances are almost similar. Yet there is no explanation why the results from the likelihood-ratio tests are in contradiction with the results from the prediction performances. Therefore, it is advised for future research to use another methodology which might lead to results that are better explainable.

Regardless of all shortcomings, this research is a contribution to the already existing literature about soft drinks. Because the methodology with the specific attributes used in this research was not used before in the same way. Furthermore, this research investigated the effect of recyclability in relation to the consumer preference, this has not been done in previous research (Sorensen & Bogue, 2006; Enneking, Neumann & Henneberg, 2007; Redondo, Gomez-Martinez & Marcos, 2014; Meyers, 2003).

References

Breidert, C., Hahsler, M., & Reutterer, T. (2006). A review of methods for measuring willingness-to-pay. *Innovative Marketing*, *2*(4), 8-32.

CBS; Centraal Bureau voor Statistiek; Statistics Netherlands (2010). Voedings- en genotmiddelen; consumptie per Nederlander. Retrieved on 11-03-2020 from http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=37154&D1=a&D2=0- 1,11,21,31,41,51,61,71,81,86,91,96,101-110,l&HD=120530-1657&HDR=G1&STB=T.

Chapman, C. and McDonnell Feit, E. (2015). R for Marketing Research and Analytics. 2nd ed. Boston: Springer, pp.374 - 377.

Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240).

Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. International Journal of Research in Marketing, 27(4), 293-307.

Enneking, U., Neumann, C., & Henneberg, S. (2007). How important intrinsic and extrinsic product attributes affect purchase decision. *Food Quality and Preference*, *18*(1), 133-138.

Fader, P. S., & Hardie, B. G. (1996). Modeling consumer choice among SKUs. Journal of marketing Research, 33(4), 442-452.

Forshee, R. A., Anderson, P. A., & Storey, M. L. (2008). Sugar-sweetened beverages and body mass index in children and adolescents: a meta-analysis. The American journal of clinical nutrition, 87(6), 1662-1671.

FWS; Nederlandse Vereniging Frisdrank, Waters, Sappen; Dutch Association for Soft Drinks, Waters and Juices (2018). Drankenconsumptie in Nederland. Inzicht in de Nederlandse drankenconsumptie op basis van de Voedselconsumptiepeiling 2014-2018. The Hague, The Netherlands: Dutch Association for Soft Drinks, Waters and Juices.

FWS. (2017, August 22). Retrieved from fws.nl: https://www.fws.nl/nieuws/groei-verfrissende-drankjes-met-minder-tot-geen-suiker-en-calorie%C3%ABn

Giammattei, J., Blix, G., Marshak, H. H., Wollitzer, A. O., & Pettitt, D. J. (2003). Television watching and soft drink consumption: associations with obesity in 11-to 13-year-old schoolchildren. Archives of pediatrics & adolescent medicine, 157(9), 882-886.

Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. Journal of marketing, 54(4), 3-19.

Guadagni, P. M., & Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. Marketing science, 2(3), 203-238.

Hodac.(2019, December 2). theparliamentmagazine. Retrieved from https://www.theparliamentmagazine.eu/articles/partner_article/unesda-union-european-soft-drinks-associations/how-soft-drinks-industry-0

Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. Journal of Marketing research, 33(3), 307-317.

Jansen & Verkooijen, 2014. Determinants of soft drink consumption. Differences by socioeconomic status. *Wageningen Universiteit.*

Kahneman, D. (2011). Thinking Fast and Slow. United-States: Farrar, Straus and Giroux.

Kuhfeld, W. F., Tobias, R. D., & Garratt, M. (1994). Efficient experimental design with marketing research applications. Journal of Marketing Research, 31(4), 545-557.

Kumar, N., & Ray, S. (2018). Attitude towards soft drinks and its consumption pattern: a study of Gen Y consumers of India. British Food Journal.

Kumar, V., Rahman, Z., Kazmi, A. A., & Goyal, P. (2012). Evolution of sustainability as marketing strategy: Beginning of new era. *Procedia-Social and Behavioral Sciences*, *37*, 482-489.

Kvaavik, E., Andersen, L. F., & Klepp, K. I. (2005). The stability of soft drinks intake from adolescence to adult age and the association between long-term consumption of soft drinks and lifestyle factors and body weight. Public health nutrition, 8(2), 149-157.

Kwartler. (2017). *Text Mining in Practice with R.* Chennai: Wiley.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, *74*(2), 132-157.

Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. Journal of marketing research, 20(4), 350-367.

Ludwig, D. S., Peterson, K. E., & Gortmaker, S. L. (2001). Relation between consumption of sugar-sweetened drinks and childhood obesity: a prospective, observational analysis. The lancet, 357(9255), 505-508.

McFadden, D. (1986). The choice theory approach to market research. Marketing science, 5(4), 275-297.

Miller, K. M., Hofstetter, R., Krohmer, H., & Zhang, Z. J. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, *48*(1), 172-184.

Mooradian, A. D., Smith, M., & Tokuda, M. (2017). The role of artificial and natural sweeteners in reducing the consumption of table sugar: A narrative review. Clinical nutrition eSPen, 18, 1-8.

Myers, C. A. (2003). Managing brand equity: a look at the impact of attributes. Journal of product & brand management.

Natter, M., & Feurstein, M. (2002). Real world performance of choice-based conjoint models. European Journal of Operational Research, 137(2), 448-458.

Redondo, N., Gomez-Martinez, S., & Marcos, A. (2014). Sensory attributes of soft drinks and their influence on consumers' preferences. *Food & function*, *5*(8), 1686-1694.

Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008). Discrete choice experiments in a nutshell. In Using discrete choice experiments to value health and health care (pp. 13-46). Springer, Dordrecht.

Schelfaut. (2020, January 22). AD. Retrieved from https://www.ad.nl/economie/jumbo-groeit-snoeihard-door-ah-stabiel-en-lidl-en-aldi-verliezen-marktaandeel~a916a3e8/?referrer=https://www.google.com/

Silayoi, P., & Speece, M. (2007). The importance of packaging attributes: a conjoint analysis approach. European journal of marketing.Szmigin, & Piacentini. (2018). Consumer Behaviour . Oxford: Oxford University Press.

Sorenson, D., & Bogue, J. (2006). Modelling soft drink purchasers' preferences for stimulant beverages. International journal of food science & technology, 41(6), 704-711.

Stigler, G. J. (1945). Agricultural & Applied Economics Association. Journal of Farm Economics, 27(2), 303-314.

Ubeja, S., & Patel, R. (2014). Consumer preference towards soft drinks: a perceptual study. *Public Business Review International*, *6*(9), 80-86.

Voicu, M. C. (2007). Characteristics of the consumer preferences research process. Marketing Research, 44, 347-356.

WEINSTEIN, A. (2004). Handbook of Market Segmentation: strategic targeting for business and technology firms. New York: Haworth. ISBN 0-7890-2156-0.

Wey Smola, K., & Sutton, C. D. (2002). Generational differences: Revisiting generational work values for the new millennium. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *23*(4), 363-382.

Williams, K. C., & Page, R. A. (2011). Marketing to the generations. *Journal of Behavioral Studies in Business*, *3*, 1.

Appendix

Appendix 1

Choice sets used in this research

	Choice Set	Choice ID	Verpakking	Soort zoetstof	Merk	Recyclebaar	Prijs
1	1	1	Plastic fles	Zonder suiker	A-merk	Recyclebaar	€0.20
2	1	2	Aluminium blikje	Natuurlijke zoetst...	Huismerk	Niet-recyclebaar	€0.20
3	1	3	Glazen fles	Met suiker	Nieuw merk	Recyclebaar	€0.20
4	1	4	Aluminium blikje	Natuurlijke zoetst...	Nieuw merk	Niet-recyclebaar	€0.40
5	2	1	Aluminium blikje	Zonder suiker	Huismerk	Niet-recyclebaar	€0.80
6	2	2	Aluminium blikje	Kunstmatige zoet...	Huismerk	Recyclebaar	€0.40
7	2	3	Plastic fles	Met suiker	A-merk	Recyclebaar	€0.60
8	2	4	Plastic fles	Natuurlijke zoetst...	Nieuw merk	Recyclebaar	€0.80
9	3	1	Plastic fles	Kunstmatige zoet...	A-merk	Niet-recyclebaar	€0.80
10	3	2	Glazen fles	Zonder suiker	Huismerk	Recyclebaar	€0.80
11	3	3	Aluminium blikje	Kunstmatige zoet...	Nieuw merk	Niet-recyclebaar	€0.60
12	3	4	Glazen fles	Natuurlijke zoetst...	Huismerk	Niet-recyclebaar	€0.40
13	4	1	Aluminium blikje	Zonder suiker	A-merk	Niet-recyclebaar	€0.40
14	4	2	Plastic fles	Natuurlijke zoetst...	Nieuw merk	Niet-recyclebaar	€0.60
15	4	3	Plastic fles	Kunstmatige zoet...	Nieuw merk	Niet-recyclebaar	€0.80
16	4	4	Aluminium blikje	Met suiker	A-merk	Recyclebaar	€0.80
17	5	1	Aluminium blikje	Met suiker	Huismerk	Recyclebaar	€0.60
18	5	2	Plastic fles	Natuurlijke zoetst...	A-merk	Recyclebaar	€0.40
19	5	3	Glazen fles	Zonder suiker	Nieuw merk	Niet-recyclebaar	€0.60
20	5	4	Aluminium blikje	Met suiker	Nieuw merk	Niet-recyclebaar	€0.80
21	6	1	Plastic fles	Met suiker	Nieuw merk	Recyclebaar	€0.20
22	6	2	Glazen fles	Natuurlijke zoetst...	A-merk	Recyclebaar	€0.20
23	6	3	Aluminium blikje	Zonder suiker	Nieuw merk	Recyclebaar	€0.20
24	6	4	Glazen fles	Kunstmatige zoet...	Huismerk	Niet-recyclebaar	€0.40
25	7	1	Glazen fles	Kunstmatige zoet...	Huismerk	Niet-recyclebaar	€0.60
26	7	2	Plastic fles	Zonder suiker	Huismerk	Niet-recyclebaar	€0.20
27	7	3	Glazen fles	Kunstmatige zoet...	Nieuw merk	Recyclebaar	€0.40
28	7	4	Aluminium blikje	Met suiker	A-merk	Niet-recyclebaar	€0.40
29	8	1	Plastic fles	Zonder suiker	Nieuw merk	Niet-recyclebaar	€0.40
30	8	2	Aluminium blikje	Kunstmatige zoet...	A-merk	Niet-recyclebaar	€0.60
31	8	3	Plastic fles	Natuurlijke zoetst...	Nieuw merk	Recyclebaar	€0.60
32	8	4	Glazen fles	Met suiker	A-merk	Niet-recyclebaar	€0.80
33	9	1	Glazen fles	Met suiker	Nieuw merk	Recyclebaar	€0.40
34	9	2	Plastic fles	Kunstmatige zoet...	Huismerk	Recyclebaar	€0.20
35	9	3	Plastic fles	Kunstmatige zoet...	A-merk	Niet-recyclebaar	€0.20
36	9	4	Aluminium blikje	Natuurlijke zoetst...	Nieuw merk	Recyclebaar	€0.20
37	10	1	Glazen fles	Met suiker	Nieuw merk	Niet-recyclebaar	€0.20
38	10	2	Plastic fles	Met suiker	Huismerk	Niet-recyclebaar	€0.60
39	10	3	Glazen fles	Zonder suiker	A-merk	Recyclebaar	€0.60
40	10	4	Aluminium blikje	Natuurlijke zoetst...	Huismerk	Niet-recyclebaar	€0.80

Appendix 2. Tables

Appendix 2. Table 1

Preferences segmented by generation

Generation	Packaging	Type of Sweetener	Brand	Recyclability	Price
1946-1964	Glass bottle	Without Sugar	Top Brand	Recyclable	Negative utility
1965-1977	Glass bottle	Without Sugar	Top Brand	Recyclable	Negative utility
1978-1994	Glass bottle	Without Sugar	Top Brand	Recyclable	Negative utility
1995-now	Plastic bottle	Without Sugar	Top Brand	Recyclable	More negative utility

Appendix 2. Table 2

Preferences segmented by consumption frequency

Consumption frequency	Packaging	Type of Sweetener	Brand	Recyclability	Price
Never	Glass bottle	Without Sugar	Private label	Recyclable	Negative utility
Once a month	Glass bottle	Without Sugar	Top Brand	Recyclable	Negative utility
Weekly	Plastic	Without Sugar	Top Brand	Recyclable	Negative utility
Daily	-	Without Sugar	Top Brand	Recyclable	Negative utility
Multiple times per day	-	Without Sugar	Top Brand	Recyclable	Negative utility

Appendix 3. Table 3

Preferences segmented by consumption moment

Consumption moment	Packaging	Type of Sweetener	Brand	Recyclability	Price
Not	Glass bottle	Without Sugar	Top brand	Recyclable	Negative utility
Spare time	Glass bottle	Without Sugar	Top Brand	Recyclable	Negative utility
During work	Aluminum can	Without Sugar	Top Brand	Recyclable	Negative utility
On the way	Glass bottle	Without Sugar	New Brand	Recyclable	Negative utility
During sports	Glass bottle	Without Sugar	Top Brand	Recyclable	More negative utility
Dinner	Plastic bottle	Without Sugar	Top brand	Recyclable	Negative utility
Lunch	Glass bottle	Without Sugar	Top Brand	Recyclable	Negative utility
Breakfast	Glass bottle	Without Sugar	Top Brand	Recyclable	Positive utility