Waste reduction in e-groceries fulfillment centers

A case study at Picnic

J. Meijboom

TU Delft, Picnic B.V.

February 2019

Master thesis report

MSc Transport Engineering and Logistics

Thesis committee: Dr. ir. D.L. Schott, Dr. B. Atasoy & ir. F.B. Gorte

Delft University of Technology

FACULTY MECHANICAL, MARITIME AND MATERIALS ENGINEERING
Department Marine and Transport Technology

Mekelweg 2
2628 CD Delft
the Netherlands
Phone +31 (0)15-2782889
Fax +31 (0)15-2781397
www.mtt.tudelft.nl

Specialization:	Transport Engineering and Logistics
Report number:	2018.TEL.8303
Title:	**Waste reduction in e-groceries fulfilment centers: a case study at Picnic**
Author:	J. Meijboom

Title (in Dutch)	Vermindering van verspilling in online supermarkt fulfilmentcentra: een casestudy bij Picnic

Assignment:	Master's thesis
Confidential:	yes
Initiator (university):	Dr.ir. D.L. Schott
Initiator (company):	Ir. F. B. Gorte
Supervisor:	Dr. B. Atasoy
Date:	January 21, 2019

Preface

This research aims at identifying and confirming causes of food waste to ultimately reduce food waste. A subject that was brought up by Picnic, when we had decided that I could do my graduation research there. I knew little about the subject, but fortunately no-one else in the company really had an idea what was going on. It was interesting and educational to start on a totally new project within a company. The main intention for Picnic was to gain insights into the main drivers of their food waste, so they could effectively reduce waste. It turned out that the goals of a company do not always require a very scientific approach or method. The biggest challenge for me was to balance the practical desires of a company with the scientific requirements for my graduation. I have learned a lot during my graduation: working at a fast growing, young company, analyzing (big) data, balancing different interests (such as science & practice), quickly learning to use new software and improving my communication skills.

I would like to thank some people who helped me successfully complete the research or enriched my life in another way. First of all, I would like to thank my daily supervisor Bilge Atasoy, for our meetings, thinking with me in the process, showing me that *"statistics are cool"* and always having fun. Also I would like to thank Dingena Schott, for her direct and precise, yet kind feedback in the important meetings.

I would also like to thank the people from Picnic, a close, smart and friendly team. Special thanks to Frank Gorte, for formulating the assignment, helping me get on track and staying there, always with a smile. Also I would like to thank Frederik Nieuwenhuys, Thijs Bender, Peter Renting, Steven van de Ridder, Linda Rietveld and Floris Boekema for always making some time to help me with the research, software or waste reduction improvements at Picnic. I've had a great time graduating at Picnic, mostly thanks to the atmosphere and all the people.

With the completion of my thesis comes an end to my life as a student. I would like to take the opportunity to thank my dear family, for unconditionally supporting me during the full duration of my studies (and my life of course). Also a big thanks to the wonderful group of friends I've met during my time as a student. And last but not least I would like to thank my girlfriend, Bregje, for her love and support since the day I met her.

Jurriaan Meijboom
Delft, January 30, 2019

Abstract

Food waste is an increasing problem worldwide. Around one third of all produced food is estimated to end up as waste. This has big environmental and economical consequences. In literature, little is known about the causes of waste, its key performance indicators and how to reduce it. Some root causes were found, but none were quantitatively confirmed.

E-groceries are a relatively new type of supermarket, that operate on a big scale and have access to detailed data about their customers and their shopping behavior. This data enables detailed quantification of the waste. Picnic is one of these supermarkets, operating in the Netherlands and Germany. A case study focused on reduction of food waste was performed at their fulfilment centers. Also at Picnic, little was known about waste and its causes.

This research aims at identifying root causes in a quantitative way and using gained insights to reduce waste. The research question was:

What are the root causes of food waste at an online supermarket's fulfilment centres, and how can food waste be reduced?

The case study that was performed consisted of three parts:

- **An exploratory part**, in which data was collected and investigated. Chilled products caused 80 % of the waste, while only consisting of 20 % of the assortment.
- **A qualitative part**, in which KPIs and expected root causes were identified and hypotheses generated. Waste amount in units was used as a key performance indicator. Expected root causes were identified for four domains: assortment, supply chain, fulfilment center processes and other.
- **A quantitative part**, in which variables were constructed by means of separate analyses to measure these expected root causes. Twelve variables were constructed for assortment and supply chain related factors. The significance of the hypothesized root causes were tested with multivariate regression tools, focusing only on chilled products. Two data sets have been tested, both on granularity article x fulfilment center x financial period. The first containing the full data, including a majority of zero waste data points. The second containing only waste. All factors showed a significant relationship with waste, in at least one of the regression models, confirming that the identified factors are indeed root causes of food waste.

The results were used to improve waste reduction processes at Picnic. The gained insights in root causes ensured effective handling of waste cases: actions could be taken according the root cause of specific cases. This efficient continuous improvement has led to a waste reduction of 40 %, expressed in costs per item sold.

Samenvatting

Voedselverspilling is wereldwijd een toenemend probleem. Ongeveer een derde van alle geproduceerde levensmiddelen wordt geschat te eindigen als afval. Dit heeft grote gevolgen voor het milieu en de economie. In de literatuur was er weinig informatie beschikbaar over de oorzaken van afval, de belangrijkste indicatoren en hoe deze te verminderen. Er werden root causes gevonden, maar geen ervan waren kwantitatief bevestigd.

Online supermarkten kunnen een uitkomst bieden. Ze zijn een relatief nieuw type supermarkt, die op grote schaal werken en toegang hebben tot gedetailleerde gegevens over hun klanten en hun winkelgedrag. Deze data maakt de kwantificatie van waste mogelijk. Picnic is een van deze supermarkten, actief in Nederland en Duitsland. Een case study gericht op het verminderen van voedselverspilling werd uitgevoerd in hun fulfilment centers. Ook vanuit Picnic was er weinig bekend over afval en de oorzaken ervan.

Dit onderzoek is gericht op het identificeren van onderliggende oorzaken op een kwantitatieve manier en het gebruiken van opgedane inzichten om afval te verminderen. De onderzoeksvraag was:

Wat zijn de hoofdoorzaken van voedselverspilling in de fulfilmentcentra van een online supermarkt en hoe kan voedselverspilling worden verminderd?

De casestudy die werd uitgevoerd bestond uit drie delen:

- **Een exploratief deelt**, waarin gegevens werden verzameld en onderzocht. Gekoelde producten veroorzaakten 80% van de verspilling, terwijl ze slechts 20% van het assortiment vormden.
- **Een kwalitatief deel**, waarin key performance indicators en verwachte hoofdoorzaken werden geïdentificeerd en hypothesen werden gegenereerd. De hoeveelheid afval in eenheden/aantal stuks werd gebruikt als key performance indicator. Er zijn root causes geïdentificeerd voor vier domeinen: assortiment, supply chain, fulfillment center-processen en overig.
- **Een kwantitatief deel**, waarin variabelen werden geconstrueerd door middel van afzonderlijke analyses om deze verwachte hoofdoorzaken te meten. Twaalf variabelen werden geconstrueerd voor assortiment- en supply chain-gerelateerde oorzaken. De beschikbare hoofdoorzaken en hun hypotheses werden getest met multivariate regressieïnstrumenten, waarbij alleen gekoelde producten werden gebruikt. Twee datasets zijn getest, beiden op granulariteit artikel x fulfillment center x financiële periode. De eerste met de volledige data, inclusief een meerderheid van 'geen verspilling' datapunten. De tweede dataset bevatte alleen verspilling. Alle factoren toonden een significante relatie met voedselverspilling, in ten minste één van de regressiemodellen, wat bevestigde dat de geïdentificeerde factoren inderdaad oorzaken zijn.

De resultaten zijn gebruikt om afvalreductieprocessen bij Picnic te verbeteren. De nieuw verkregen inzichten in de oorzaken zorgden voor een effectieve afhandeling van de meest urgente verspillingsgevallen en concrete acties konden worden ondernomen op basis van de bevestigde oorzaken. Deze efficiënte continue verbetering heeft geleid tot een afvalvermindering van 40 %, uitgedrukt in verspillingskosten per verkocht artikel.

Contents

List of Figures

List of Tables

List of abbreviations

AOR Article order rate
BBD Best-before date
COGS Cost of goods sold
CoV AOR Coefficient of variation of the article order rate
CS Customer success
cts Eurocents
CU Consumer unit
D_f Forecasted demand
DB Dashboard
DM Dimension
DWH Data warehouse
EDA Exploratory data analysis
EPV Electric Picnic vehicle
FA Feestdag article
FAO Food and Agriculture organization
FC Fulfilment center
FEFO First-expired-first-out
FIFO First-in-first-out
FT Fact table
KPI Key performance indicator
MECE Mutually exclusive, collectively exhaustive
MOQ Minimum order quantity
OoA Out of assortment
PD Packaging date
PIM Product inventory management
POM Purchase order management
RQD Research question: design/case study
RQI Research question: Implementation
RQL Research question: literature
SCF Supply chain factors
SF Safety factor
SKU Stock keeping unit
TU Trade unit
WMS Warehouse management system
WRAP Waste and Resource Action Programme

1
Introduction

1.1. Food waste is caused by increasing global consumption

The world's population is growing faster than ever and all over the world countries are developing at a rapid pace. This brings along a growth in consumption that is unprecedented. The food industry is one of the biggest industries in the world that has to keep up with this pace.

A huge effort is made to produce enough food to be consumed by the world's rising demand. And not only demand, because with this rising development come rising standards. Consumers ask for a complete assortment of food that is available at all times. Because of extensive use of greenhouses to produce fresh fruit and vegetables combined with a fast and sophisticated supply chain, the consumers rely on the freshest products at the highest possible quality. In order to meet this demand of both high quantity and high quality, the food supply chain overproduces food and as a result, a part of the food becomes waste. The total yearly food waste is estimated to be around 1.55 billion tons (mass) [1].

1.2. Food waste has considerable environmental and economic impact

Environmental relevance
According to the Food and Agriculture Organization, globally, around 1/3 of all produced food is wasted (FAO 2011) [5]. This is an astonishing figure. It indicates that all food that is wasted, could easily feed twice the world's population that is in hunger [6].

In addition all this food that was produced but wasn't consumed, did cost soil, water, energy, fertilizer, transportation etc. This significantly contributes to the emission of greenhouse gases. FAO calculated that 8% of greenhouse gases are emitted in the process of the production and transportation of food that is wasted. This means that if "food waste were a country, it would be the third largest emitter of greenhouse gases, after the USA and China!" [7].

In order to meet the conditions of the Paris agreement [8], agriculture's environmental footprint should be drastically reduced, despite a required increase in global food production [9]. The united nations set up 17 sustainable development goals [10], of which the second goal directly concerns food waste and ending hunger [11]. For all these reasons, all parties within the supply chain responsible should do everything within their power to reduce food waste.

Economic relevance
It goes without saying that food waste brings along economic waste. All the resources to produce and transport food that is wasted have all cost money. Reducing food waste automatically improves economic situations; farmers have to produce less, less food has to be distributed, supermarkets have to buy less and allocate less space in warehouse and consumers have to buy less. Also less human resource in the form of processing (time) is needed along the whole value chain.

1.3. E-grocers can reduce food waste

The supply chain of food can be divided into five parts:

- Production
- Handling and storage
- Processing and packaging
- Distribution and retail
- Consumption

Across all these parts, waste is generated. Figure 1.1 shows the distribution in millions of tonnes of food wasted.

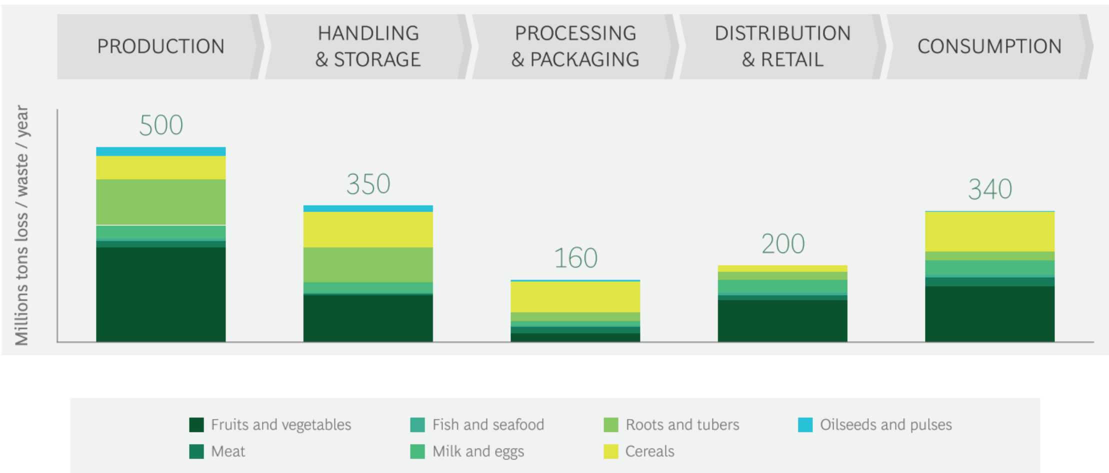


Figure 1.1: Food waste across supply chain. *Source: BCG 2018 [1]*

In the distribution & retail sections, e-grocers are relatively young. E-grocers are supermarkets, where people order their groceries online, through a website or an app. Picnic is one of these. They do not have any physical stores and orders are received through their app.

Supermarkets such as Picnic can accurately track consumers orders and shopping behavior. Therefore accurate forecasts can be made about demand. For a group of products, demand might already be known for the next day. If so, products can be delivered at the distribution centers just in time to ensure no waste or decrease unnecessary food production. This way, online supermarkets can be in a unique position to reduce food waste from the distribution and retail phase.

1.4. Research goal and scope

Little is known about the amount and the causes of food waste at supermarkets, and even less is known about waste drivers at online supermarkets. Therefore, apart from the already mentioned environmental and economic relevance in the reduction of food waste, this study is relevant for expansion of the knowledge of mapping food waste. Due to the extensive stock and warehouse management systems to store, pick and deliver the orders in online supermarkets a huge amount of data is available. E-grocers form an excellent environment to study generation of food waste. This study moves to a new field of research aiming at identifying accurate key performance indicators (KPIs) and root causes of food waste.

The goal of this research is to gain insight in root causes for food waste at e-groceries fulfilment centers (FCs). The e-grocer 'Picnic' forms the scope of the case study for qualitative and quantitative evaluation of food waste. Based on the outcome, recommendations can be given to reduce the amount of waste that is generated.
Wasted time and energy that are associated with this waste is not taken into account.

1.5. Research questions

The main research question, to gain proper understanding of food waste at an online supermarkets fulfilment center is:

What are the root causes of food waste at an online supermarket's fulfilment centres, and how can food waste be reduced?

To get to a complete answer in a structured way, the following sub questions have been formulated.

Literature

- RQL1: What are commonly used and applicable key performance indicators on food waste?
- RQL2: What is the current knowledge of root causes for food waste in the retail part of the food supply chain?
- RQL3: How do the root causes of food waste in e-groceries differ from traditional supermarkets?
- RQL4: What are state-of-the-art food waste reduction mechanisms?

Design/Case study

- RQD1: Which data should be used to measure waste performance at Picnic?
- RQD2: Which factors are root causes for food waste?
- RQD3: Can the root causes be confirmed, using a statistical model?

Implementation

- RQI: Can the gained information on root causes provide actionable insights to reduce waste?

1.6. Scientific contribution

We expect that the findings of this study will contribute to future research on food waste at the retail stage of the supply chain. A new field of research will be opened up: food waste reduction by means of quantified root cause analysis.
Different root causes will be identified and variables will be constructed to measure them. The root causes will be tested on their contribution to waste and their significance, using multivariate statistical analyses.
The identified root causes are expected to be universal to a certain extent, and both further scientific research and other retail stores in practice can benefit from the approach as undertaken in this research, as well as the results.

1.7. Outline of the report

Figure 1.2 schematically shows the outline of the body of the report. The body exists of a literature study and a case study. The outline is as follows:

- Literature
  - **Chapter 2) Food waste - literature:** definitions, KPIs and root causes of food waste are discussed. RQL1 and RQL2 are answered.
  - **Chapter 3) Traditional vs e-grocers:** E-grocers are introduced and compared to traditional supermarkets. RQL3 is answered.
  - **Chapter 4) Food waste - state of the art:** state of the art processes/technologies to reduce waste are discussed. RQL4 is answered
- Case study: the case study consists of three parts: an exploration part, a qualitative part and a quantitative part.
  - **Chapter 5) Picnic:** the company is introduced, fulfilment processes and waste are discussed.
  - **Chapter 6) Design of the case study:** methodology of the case study is elaborated. Not in Figure 1.2, since not part of the body.
  - **Chapter 7) Data extraction, KPIs:** this Chapter is concerned with the process of data collection and extraction, KPIs are discussed and global insights of food waste at Picnic is shown. RQD2 is answered.
  - **Chapter 8) Identification of root causes of waste:** this chapter consists of two parts. In the qualitative part factors, representing expected root causes are identified and hypotheses are formulated for these factors. Then in a qualitative part indicating variables are constructed for these factors.
  - **Chapter 9) Modeling:** The identified factors are tested on the relation to waste and their significance, using two statistical models.
  - **Chapter 10) Implementation:** gained insights are used for continuous improvement of waste performance at Picnic.

The report finishes with a chapter with a **conclusion & future research directions**.

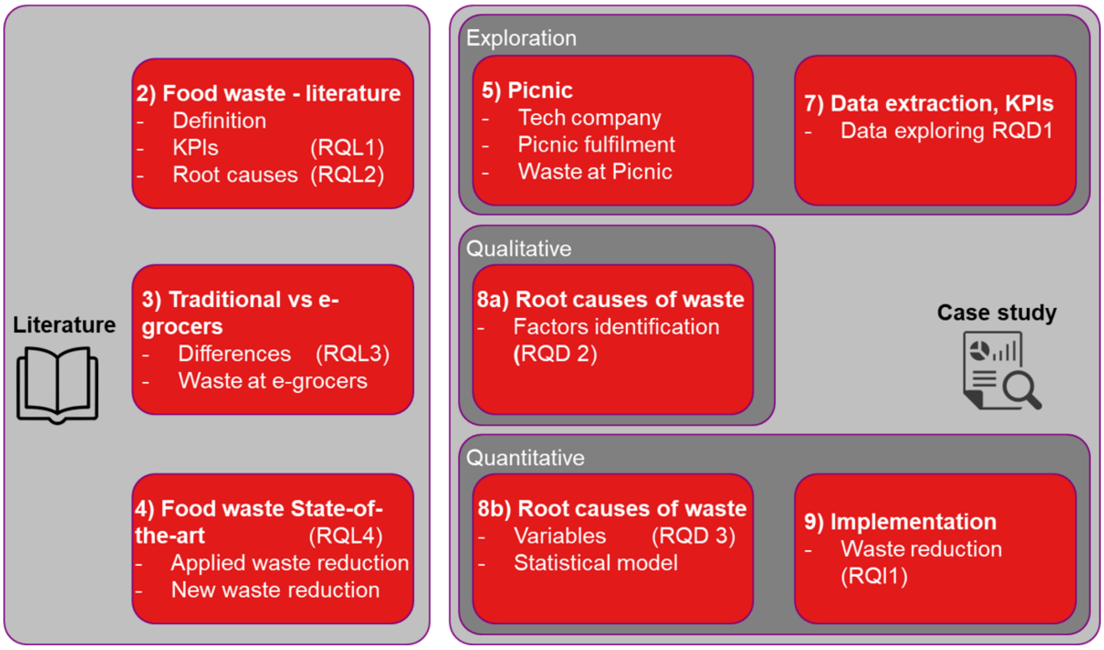


Figure 1.2: Outline of the report

I

Part 1: Literature and background

2

Food waste - a literature review

The first step in answering the main research question *"What are the root causes of food waste at an online supermarket's fulfilment centres, and how can food waste be reduced?"*, requires knowledge about previous studies on the topic of food waste and knowledge about online supermarkets. In this chapter the definition of food waste, its commonly used key performance indicators (KPIs) are given and root causes are discussed.

2.1. Definition of food waste

First of all, let us be clear on the definition of food waste. Not one all-round definition is used, since definitions vary throughout literature and waste can be expressed in different units. The most commonly used definition is described by the Food and Agriculture Organization of the united nations (FAO) as being a part of 'food loss' [12]. "Food waste can be conceived as the result of decisions made by consumers, supply chain actors or other stakeholders, and it represents a subset of the total food losses [13]." Food loss and food waste are differentiated in the following way (Figure 2.1):

- **Food loss** takes all the food that is lost in the total food supply chain in consideration; from producer to consumer. For the biggest part food loss is caused by the way the food supply chain works, including food production, supply system or legal framework.
- **Food waste** is a part of food loss. FAO considers food waste to be food loss that is or was at some point fit for human consumption, or which has expired mainly by economic behaviour, poor stock management or neglect [14].

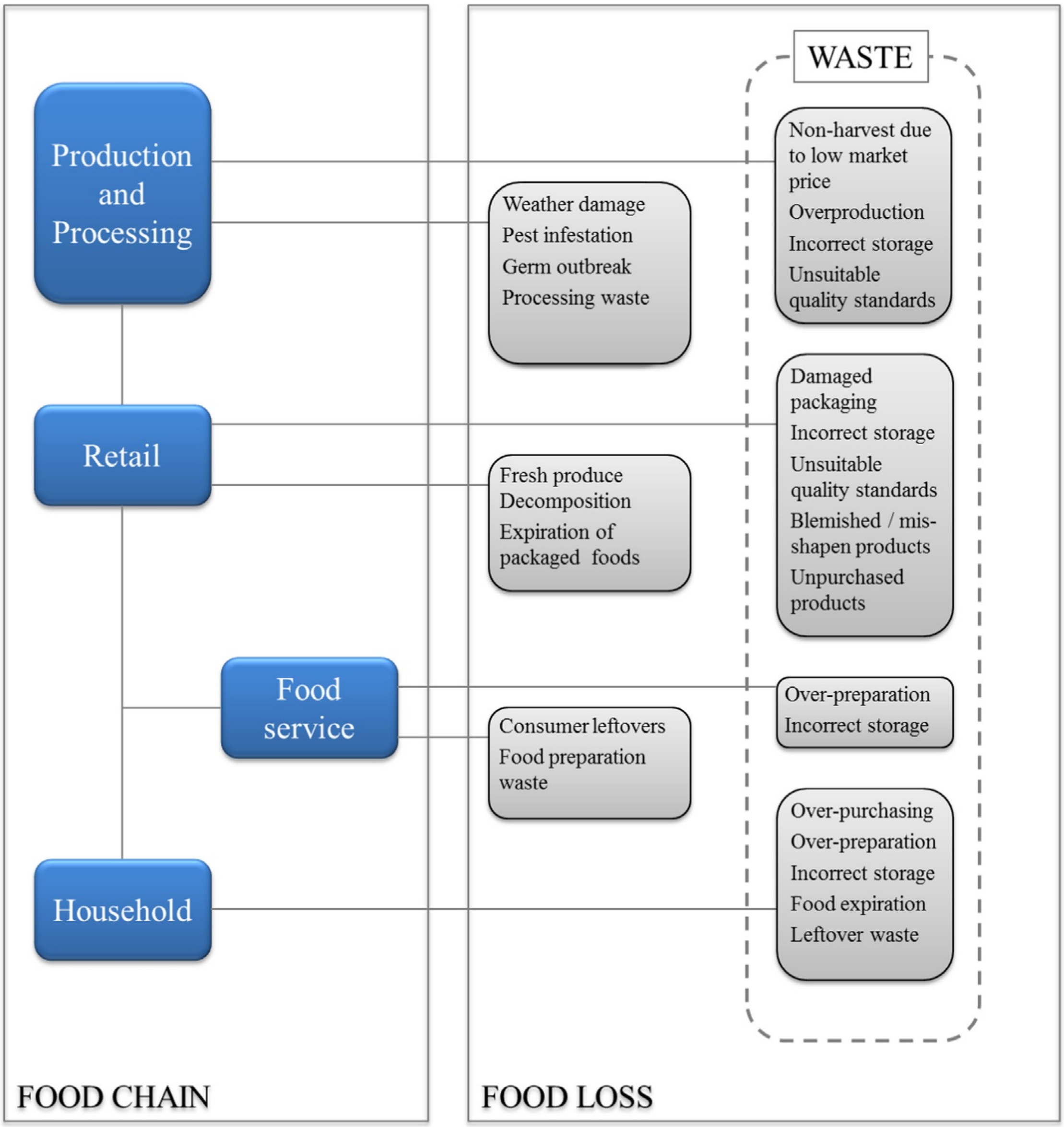


Figure 2.1: Food Loss and Food Waste, source; Cicatiello 2016 [2]

2.2. Waste key performance indicators (KPIs)

This section answers the first sub question of the literature part:

RQL1: What are commonly used and applicable key performance indicators on food waste?

2.2.1. Waste measured as an absolute value: mass

It is important to have a KPI that can accurately measure waste performance in order to create fact-based actionable insights. However, just like the definition, there is no generally used performance indicator to measure waste. The European Commission is developing a methodology to measure food waste and its relevant indicators, based on best practices. However, no results have been published so far [15].

In literature, several studies measure food waste on a large scale: nationally, internationally or even globally. For this, the most common unit of measure is mass as an absolute number (all studies mentioned in [3]),[16],[17]. This is done by either weighing the waste, or calculating the weight by multiplying the number of waste units by a known weight.

Figure 2.2 compares 9 different studies that all have reported food waste in the food supply, using kg/p/y as a unit of measurements. There are large differences in estimation of food waste between studies, ranging from around 125 to 300 kg/p/y. This difference is caused by the fact that there is no agreed upon definition and no concrete agreements on what is to be regarded as waste and what is not.

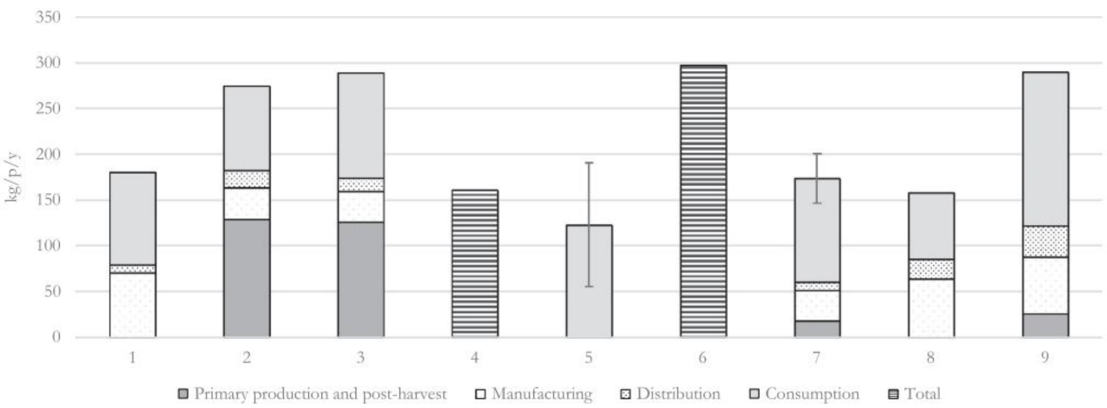


Figure 2.2: Comparing studies. *Source: Corrado 2018 [3]*

Fig. 3. European FW generation per capita and per year and uncertainty (minimum and maximum value for 5: 95% confidence interval for 7), when reported in the study. 1: Monier et al. (2010); 2: FAO (2011); 3: Bräutigam et al. (2014); 4: data collected by Eurostat (2017b); 5: Vanham et al. (2015); 6: Porter et al. (2016); 7: Stenmarck et al. (2016); 8: Tisserant et al. (2017); 9: Kemna et al. (2017). FAO (2011) includes only edible fraction of FW.

Figure 2.3: Comparing studies references. Source: Corrado 2018 [3]

Another disadvantage of using mass as a performance indicator is that mass can change; e.g. when a fresh product gets older (most fresh products have a moisture content of between 70% and 80% [18]), water evaporates and mass is 'lost.'

2.2.2. Alternative absolute waste indicators

The Waste and Resources Action Programme (WRAP) is a registered charity that aims to achieve a circular economy. Reducing waste is one of their main targets. It shows that next to mass and a percentage, food waste can also be measured in units, monetary value or carbon (tonnes) [19]. Some complications with these three is that units do not automatically say something about the amount of waste, especially when compared to mass. Monetary value is not an obvious KPI for measuring wast in a supply chain, since the value of a product is not constant during the different stages in the supply chain. Carbon tonnes are difficult to measure; many more aspects need to be measured.

2.2.3. Waste indicated by a percentage

The company 'Spoiler Alert' [20] focuses on educating on the topic of food waste. Its aim is reducing waste. Spoiler alert also concentrates on companies, government and provinces, therefore looking on a smaller scale compared to the studies mentioned in Section 2.2.1. As a KPI, it advises to use a shrink ratio, measuring units of waste, expressed as:

$$\textit{Shrink ratio} = \frac{\textit{Units unsold}}{\textit{Units purchased}} \tag{2.1}$$

A recovery ratio is used to measure the salvaged items:

$$\textit{Recovery ratio} = \frac{\textit{Units recovered}}{\textit{Units unsold}} \tag{2.2}$$

A study on food waste in six Swedish retail stores measures waste as mass of waste divided by the mass of delivered goods (Eriksson et al., 2012) [16].

$$\textit{Waste quotient} = \frac{\textit{Mass of waste}}{\textit{Mass delivered}} \tag{2.3}$$

2.2.4. KPIs conclusion

The waste and Resource Action programme says that whatever is measured, can be managed. What they mean by this is that it does not necessarily matter what KPI is measured, as long as that KPI is accurate and reliable and can be tracked. In literature the units mentioned in this Section are not frequently found.

2.3. Root causes of food waste

This section answers the question:

RQL2: What is the current knowledge of root causes for food waste in the retail part of the food supply chain?

Food waste cannot be caused by one similar aspect across all companies. They are different for every continent, country, region, store type, store size and company policy. It makes sense that not one single factor is causing all food waste. Food waste can be caused by a variety and oftentimes combinations of factors.

No concrete studies have been found that describe food waste at a retail level, based on quantified root causes. Nothing was found about root causes of food waste at an online supermarket's fulfilment centers. Some studies did show root causes, based on qualitative research.

Figure 2.4 shows the result of a root cause analysis of food waste by means of a case study in six Norwegian companies, focusing on the logistics and retail part of chilled food products chain with fixed shelf life. The root causes were divided into four areas of interest:

- Data utilization
- Planning decisions
- Product Damage
- Execution of plan

Various root causes, related to forecasting, inventory control policy, stock levels, product handling and picking are given. All shown root causes are expected to play a role in food waste generation at e-grocers. The root causes were not quantitatively validated and important waste influencing factors might be specifically related to the conditions in Norway such as long delivery distances and harsh winter conditions.

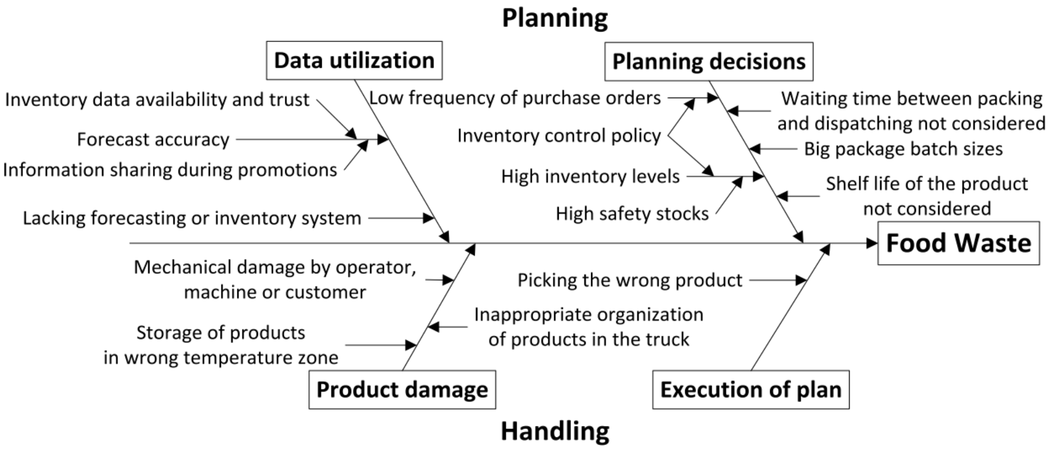


Figure 2.4: Root causes of food waste in logistics and distribution, *Source: Chabada et al. 2014, [4]*

Furthermore, root causes were found in two other studies: Teller et al. 2018 [21] and Mena et al. 2011 [22]. These were mainly based on interviews with store managers and experts in the field of food manufacturing and retail. The root causes identified in these two studies are compiled and summarized as follows:

1. **Mega-trends** in market

   (a) **Rise in demand of fresh products:** over the last few years, demand for fresh products has risen. This puts a strain on the food supply chain and can have waste as a result.
   (b) **Rise in demand for products that are out of season:** just like fresh products, also products that are out of season are expected to be in the assortment, the whole year round. These products might have to be imported, which as a consequence causes a more rigid supply flow.
   (c) **Rise in demand for products without preservatives:** another trend is that consumers want fresher products without preservatives. This causes fresh food to expire sooner, and become waste.

2. **Natural causes** related to products and processes

   (a) **Short shelf life.** Products that come into the shops with an expiry date that is too close to the date when delivered are likely to be wasted. Getting better products with a longer shelf life might be achieved by establishing a more efficient supply of products and will reduce waste.

(b) **Seasonality** of demand can cause waste. If for example a fresh product is ordered at the supplier the whole week in the same pattern, but the customers only order on Fridays, the rest of the week this product can cause waste.

(c) **Weather fluctuation** effects on demand are hard to estimate. BBQ products are a good example, when the sun is shining and the weather is sweet, the demand for BBQ products rise enormously. When it is warm but cloudy, demand will be less. Or when it is too hot, demand will also start to go down.

(d) **Long lead time for imported products:** mentioned under rise in demand for products that are out of season. Products with a long lead time are less adaptable to changing demand. When too much is ordered, the ordering company will be over stocked.

3. **Management root causes** on which management in a company can have a direct impact:

   (a) **Forecasting difficulties and poor ordering.** Humans are until some level predictable creatures, but there always is uncertainty. In order to have enough products in stock, demand usually needs to be forecasted. These forecasts rarely are perfect and can lead to either unavailable products or waste, depending on the desired service level. Most stores order more than they think they need, to be able to absorb some variation in demand.

   (b) **Service levels.** High service levels in terms of availability require more inventory. Apart from extra stock keeping costs, it also leads to waste. Smaller store format, think local convenience stores, tend to have lower service levels. They rather run out of stock than be stuck with too much perishable inventory. Therefore they order conservatively. On the other side of the spectrum there are large hypermarkets. These stores have a large assortment and attract customers by offering the highest service levels in terms of available products and quality. This requires more inventory and accurate forecasting. Hypermarkets tend to over order and have more waste relatively.

   (c) **Undesirable customer behavior.** When customers are shopping, they have a choice to either take the first item on the top of a stack/at the front of the shelves, or to look for the freshest products available. When everyone takes the freshest products, the least fresh products are more likely to become waste.

   (d) **Replenishment strategies.** A store can influence their customers shopping behavior (and therefore waste) by gradually replenishing shelves with fresher products. This can have a positive effect on waste, as customers don't have a choice anymore to take the freshest products of the shelf. Doing this requires more labor as shelves need to be filled multiple times. Also more space is required in the stockroom.

   (e) **Lack of information sharing** is crucial in the food supply chain and if not done properly can cause serious waste. An example is the bullwhip effect (Lee 1997) [23]. This effect is a change in demand at the retailer that causes bigger effects in earlier stages of the supply chain, ending in extreme over production at production level, cancelled orders and a lot of waste in the whole supply chain.

   (f) **Performance measurement and management.** It is important to have accurate KPIs and measurements of these for food waste. They need to be monitored and actions should be taken by management to reduce waste.

   (g) **Waste management responsibilities.** Someone needs to be responsible for the measuring, monitoring of waste, which lead to actions to reduce waste. If no-one has the responsibility, no-one will take action and more waste will result.

(h) **Cold chain management.** Cold chain is a mechanism to conserve fresh products, the occasional failure of cooling mechanisms can lead to waste of serious proportions. Although a breakdowns in Northern Europe are rare, they occur more often in warmer areas.

(i) **Training** Procedures for stacking, shelving, stock rotation need to be well understood and carried out by all employees. If not trained properly, personnel will cause waste. Waste caused by insufficient training will become more obvious in busy times, such as Christmas, when usually more temporary employees are hired.

(j) **Quality standards.** In Section 1.1 the rising physical quality standards were mentioned. Expecting perfect products causes waste, since a product that is not pretty enough will be taken out of the shelves and a store loses customers when they have stock-outs too often.

(k) **Promotions management.** Promotions cause different demand patterns, which are harder to predict. Waste can be a result of an imperfectly planned/forecasted promotion.

(l) **Packaging.** While more intensive types of packaging can improve shelf life, it might be harder to discard. A balance needs to be found between shelf life and type and amount of packaging needed.

2.4. Conclusion

So far it has gotten clear that in literature there are no set-in-stone definitions of food waste, KPIs or root causes. The most commonly used definition is given by the Food and Agriculture Organization [12], as: *"food loss that was at some point fit for human consumption, or which has expired mainly by economic behavior, poor stock management or neglect."*

The most commonly used KPI is mass, however mass is generally used at large scale researches such as regions or countries. Even when measuring in mass, outcomes are still vague and inconsistent among different studies. Percentages can be a clear indicator, but there is no general approach on basis of which terms the percentage needs to be calculated. Units as an absolute value, or a monetary value are not found in literature, but can be interesting KPIs when regarding just one stage of the supply chain.

Root causes for food waste can be split up in three categories. 1) Mega-trends in market, 2) Natural causes, and 3) Management root causes. The latter being the most interesting for this research, since influence from management decisions can be tracked. Trends in the market and natural causes can be seen as given.

3
Traditional supermarkets & e-grocers

In Chapter 2 traditional supermarkets have been discussed. This chapter shows food waste definition, KPIs and compares root causes found at traditional supermarkets with root causes at e-grocers.

This chapter answers the question:

RQL3: How do the root causes of food waste in e-groceries differ from traditional supermarkets?

3.1. E-grocers are different than traditional supermarkets and other online stores

First let's discuss the general differences between e-grocers and traditional supermarkets. In traditional supermarkets, customers go and walk around doing their shopping. At e-grocers, the customers order their groceries online and the e-grocer's employees pick their orders. Here, all inventory is stored in large scale fulfilment centers. After the orders are picked, they are distributed to the customers. How this is done, depends on the type of distribution model that is chosen by the e-grocer. At the traditional supermarkets the inventory is divided over a more complex network of distribution centers and various stores.

Another big difference between the two types of supermarkets is the fact that at e-grocers, data can be accurately stored with a much higher granularity. E-grocers exactly know the shopping behavior of different customers. They have all the individual data of all customers and all orders. This causes a very detailed view of customer shopping behavior, than can be found at traditional supermarkets. The traditional supermarkets have more of a black-box approach. They are able to track which products are moving through the distribution network, but they are not necessarily aware of the individual data of their customers. An example: when a product is out of stock, in a physical supermarket this is not necessarily noticed immediately. An e-grocer knows exactly when a product runs out, can estimate the actual demand for that day with more accuracy and adjust future orders to compensate for this availability.

When we look at other common non-food online stores, such as bol.com or coolblue, a big difference is the complexity of the distribution model. Online supermarkets have to do with various products in different temperature zones. The temperature zones are ambient, chilled and frozen. Among the whole range of products many fresh products are found. As the word implies, these products have limited freshness, so therefore they need to be delivered as soon as possible, while for a lot of the fresh products, also being cooled!

3.2. Waste at e-grocers

Although waste at e-grocers might be similar to traditional supermarkets in terms of types of products, the way waste emerges can be quite different. First of all, there is the aspect of customers in a store, versus order pickers in a warehouse. Both can cause waste, but it will most likely happen for different reasons. Customers in a traditional supermarket will probably choose the freshest products, thereby increasing the probability of older products to go to waste. Order pickers are not likely to pay any attention to the freshness dates, but they are more likely to cause waste by damaging products, since they are always under time pressure. Also, e-grocers usually give a freshness guarantee on perishable products. This means that when a customer receives fresh products, the products should at least have the guaranteed freshness left. The freshness guarantee puts the distribution chain under pressure, since products cannot be sold until the actual best-before date (BBD). Traditional supermarkets can reduce waste when products are on the verge of expiring, by discounting them. This is more difficult at e-grocers, since tracking of individual products needs to be done on a much bigger scale.
Note: BBD is often seen as expiry date, although a product might be fit for consumption after. Farmers and suppliers who apply these dates often take a conservative BBD, to minimize risk of people getting sick. Being too conservative with the BBD marking will increase waste.

Online supermarkets have high service levels. Especially concerning availability, customers need to be happy at all times. Unavailable products stand out more in an app than in a physical store. Because of the trade-off between unavailability and waste, usually there is a strong emphasis on availability of products and less on reducing waste. Another aspect might be that customers usually aren't aware of or concerned with the food waste of supermarkets, while they are very aware of product availability.

On the other hand, because of the higher availability of data, e-grocers should be able to forecast their demand more accurately. In short, there are different reasons and factors playing a role in waste generation at e-grocers and traditional supermarkets. Section 3.3 discusses the effect of different identified root causes of traditional supermarkets on waste at e-grocers.

3.3. Discussion root causes: expected effect at e-grocers

The root causes of food waste at traditional supermarkets can also generate waste at e-grocers. However, they might have a different effect because of the different size or distribution method. Table 3.1 was constructed to show the expected effects of management-related root causes as identified in Section 2.3 at e-grocers. The left column shows the management related causes. The mega-trends and natural causes are not taken into account, since these effects are expected to have a similar effect and cannot be influenced by management. The middle column shows the expected relative effect. An upwards arrow means more expected waste at e-grocers, a downwards arrow means less expected waste at e-grocers. An equal to sign means the expected waste is considered equal. The right column gives a short explanation about the expected effect. All root causes are discussed in more detail below Table 3.1

Food waste root causes of traditional supermarkets	Expected effect on waste at e-grocers	Explanation
Forecasting difficulties and poor ordering	↓	Better data availability and bigger scale forecasting
Service levels	↑	Generally higher service levels
Undesirable customer behavior	↓↓	No customers in warehouses
Replenishment strategies	↓	Replenishment does not influence picking
Lack of information sharing	↓	Less chains to share information
Performance measurement and management	=	KPIs and measuring is equally important
Waste management responsibilities	=	Monitoring and actions are equally important
Cold chain management	↓	More control of whole chain, until delivery at customers home
Training	=	Correct procedure execution equally important
Quality standards	↑↓	Depends on company standards
Promotions management	↓	More data available for forecasting
Packaging	=	Same characteristics for same products

Table 3.1: Food waste root causes of traditional supermarkets; expected effect on e-grocers

(a) **Forecasting difficulties and poor ordering** is assumed to cause less waste at e-grocers. Forecasting at e-grocers is done on scale of a fulfilment center, whereas forecasting at traditional supermarkets is usually done at the scale of one store.

(b) **Service levels** of e-grocers are generally higher than traditional supermarkets, because a missing item tends to be more noticeable in an app or website than in a physical shop. Hence more waste is expected.

(c) **Undesirable customer behavior** is not present at e-grocers. Orders are picked by dedicated order pickers. These pickers can have an influence on waste. Especially when shelves are not stocked in a first-expired-first-out (FEFO) way. Items can be damaged by personnel.

(d) **Replenishment strategies** also have a different effect. Replenishment strategies does not influence pickers. It can however contribute to waste, if the shelves are not filled in a FEFO way. In general, it is assumed that the customer behavior, combined with standard replenishment causes more waste than when an online supermarket is replenishing in a correct way.

(e) **Lack of information sharing** is problematic when more links are present in the information chain. With the bigger scale of e-grocers, less waste is expected.

(f) **Performance measurement and management.** The impact of performance measurement and management, when performed in the same way, depends on the managements decisions to reduce waste.

(g) **Waste management responsibilities.** Similar to the performance measurements, the responsibilities are similar, although fewer people can manage waste at fulfilment centers.

(h) **Cold chain** issues are expected to have a smaller effect. Online supermarkets rely on the fully functional cold chain more than traditional supermarkets. When a problem arises at an online supermarket, almost all of the orders will be affected. Another reason for a smaller expected effect is that e-grocers control the cooling process until the point of delivery at the customers home.

(i) **Training** is a crucial point for any food distributor. The impact is expected to be of equal size.

(j) **Quality standards.** Waste caused by quality standards is mainly impacted by the chosen quality standards of a company. When high quality standards are chosen, more waste will follow, and vice versa.

(k) **Promotions management** is usually based on accurate forecasting and rapid response to developing demand patterns during a promotion. Since we consider e-grocers to be able to forecast more accurately, less waste is expected as a result.

(l) **Packaging** influences the shelf life of products. Management can talk to suppliers to change packaging, but the relative waste between traditional supermarkets and e-grocers is assumed to be similar.

3.4. Conclusion

E-grocers are different from traditional supermarkets and common online shops.
They differ from traditional supermarkets since there are no physical stores and the order picking takes place in a fulfilment center. This requires a different distribution method. They differ from other (non-food) online shops because e-grocers handle many perishable products. Combining the different distribution method with perishable products that also have a freshness guarantee requires a highly efficient supply chain and might have waste as a result. On the other hand e-grocers have the advantage of a more detailed data availability and therefore better forecasting ability.

The effects of root causes for food waste that were found in literature for traditional supermarkets were discussed. Most root causes were assumed to be present in waste generation at e-grocers, although in most cases the effect was expected to be of a different magnitude.

4

Food waste: state-of-the-art

This chapter answers the question:

RQL4: What are some state-of-the-art food waste reduction mechanisms?

4.1. Applied waste reduction mechanisms

The easiest way to reduce waste is to make sure the food that cannot be sold by a supermarket, is still used for consumption. Donation of food is a frequent method to reduce waste. Especially at e-grocers, it can be more beneficial, since fresh products have a guaranteed shelf life remaining, so can generally still be eaten for a couple of days. However, it is best to reduce products that cannot be sold in the first place. A common technique for physical supermarkets is discounting [24]. When products are on the verge of expiring, and are therefore less likely to be sold than their fresher co-products, a discount can be applied to stimulate the sale of this product and with that prevent waste. Discounting works best at supermarkets with accurate inventory with registered best-before dates (BBDs) or on a smaller scale supermarket, at which employees can manually check products.

4.2. New waste reduction mechanisms

The most frequent reason why a product might be waste is because it expires. Uncertainty in perishability of products causes retailers to set conservative expiration dates. Ketzenberg et al. [25] conducted a research concerning uncertainty in perishability. There are three unwanted effects that can occur when perishability is unknown:

1. Receive new units of inventory that may spoil prior to units already in inventory from prior periods.
2. Sell product that is already spoiled.
3. Discard product that is still good for sale.

His goal was to optimize setting of expiration dates to reduce the sale of spoiled product and waste. Ketzenberg approached this by formulating a Markov Decision Process that balances risks and helps setting an accurate expiration date.

Another way to look at perishable products was researched by Buisman et al. [24]. He introduced dynamic shelf life (DSL) and is defined as *"a shelf life that can be adjusted to the actual quality of the product, eighter by adjusting the date or by indicating the quality of a product with a different technique, such as Time-Temperature Indicators"*. The shelf life can be determined based on a microbiological model that assesses food quality by bacterial growth.

4.3. State-of-the-art example: Ocado reports 0.02 % waste

Ocado, the leading purely online supermarket of the UK, says to produce almost zero waste (Ocado, 2018) [26]. They report to have 0.02 % of the products wasted. Some key points they mention that attribute to near-zero waste are:

- A short supply chain is aimed for. This means that they aim for as much locally sourced products as possible. The supply chain needs to be fast and efficient as well.
- Few locations: only three large automated fulfilment centers are serving the whole United Kingdom.
- Controlled temperature zones and a cold chain make sure products have the longest possible shelf life and minimizes waste.
- Tailor made software makes sure all automated hardware is always functioning optimally.
- Packaging. In consultation with suppliers Ocado makes sure products are packaged in a way that improves shelf life.
- Continuous optimization makes sure the company is constantly improving, with a high focus on waste reduction.

Note: some care needs to be taken when looking at these numbers; this is a self-proclaimed statement, with no added data. The annual report of Ocado 2017 mentions 0.7 % (GBP waste/GBP retail revenue) [27]. Apart from that, as mentioned before, there are no commonly accepted performance indicators or definitions. Ocado has donated over 2200 tons of food to charities. This research would call food that was not sold as planned waste, although it can and often is still usable and donated to charities.

4.4. Conclusion

Near zero waste can be achieved, when the supply chain is highly efficient and unsold products can be distributed further to charities or farmers to be consumed by humans or animals. Some other techniques are: data availability to monitor, forecast and improve demand and inventory, adequate packaging of materials, discounting products that are about to expire. Newer innovative ways of reducing waste are applying smarter expiration dates to the products.

II

Part 2: Case study

5
Picnic

The case study was performed at Picnic B.V. Picnic was founded in 2015 and is a rapidly growing online supermarket based in the Netherlands and Germany. This chapter gives an introduction of the company, its distribution model and its operations.

5.1. Picnic as a business: the modern milk man

Picnic attempts to recreate the experience of the milk man from back in the days, who was a friendly person and came by once a day to deliver a full range of products. Its motto is 'creating the best milk man on earth.' Of course Picnic has modernized the concept to the present and the future.
Picnic's distinctive features are:

- Orders via app only. The apps are made in-house by a team of developers, to have full control and adaptability.
- Lowest price, free delivery. Unlike some other online supermarkets, Picnic does not charge for delivery. Orders can be placed from a minimum of € 25. This is also lower than most competitors.
- Hub and spoke delivery systems. From Picnics fulfilment centers, trucks supply hubs with picked orders. From there Electronic Picnic Vehicles (EPVs, see Figure 5.1) deliver to the customers.
- Fixed time slots are used for delivering groceries. The EPVs drive fixed routes, efficiently delivering groceries to customers, similar to a bus route.
- The EPVs themselves are very efficient and green. They are electric vehicles, which means they do not emit any CO2 and make less noise than conventional trucks. They are also very narrow, with access to the groceries from the sides. This in contrary to their competitors bigger diesel vans that often block the road and have to be entered through a walk path in the middle to have access to the groceries.
- Groceries are delivered from ambient and cooled boxes (totes). To deliver the fresh products, different temperature zones are installed in the FCs, namely ambient, chilled and frozen. The groceries need to be delivered at their best temperature. For this ice packs are used, in insulated totes.

Figure 5.1: Electric Picnic Vehicle (EPV), ©Picnic B.V.

5.2. Picnic fulfilment processes

As this research focuses on food waste at fulfilment centers, a schematic flow of the products in a Picnic FC is shown in Figure 5.2. It represents the system boundaries.
All processes and actions in the FCs are driven and monitored by the WMS, using scanner devices. Picnic FC employees (shoppers) wear these on the display on their arm and the scanner on their finger, as can be seen in Figure 5.3.

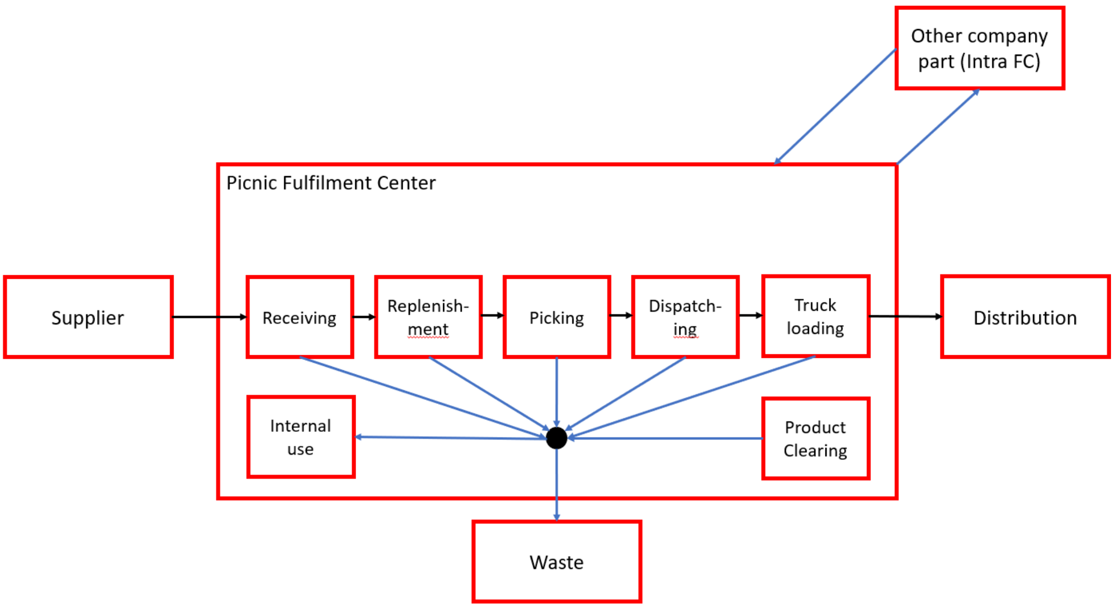


Figure 5.2: System boundaries

In the FCs, three temperature zones are present: ambient, chilled and frozen. Apart for some minor differences, the process in these zones is basically the same and will therefore be described generally.

The products go through the following processes in an FC: first they arrive in inbound trucks from a supplier and are unloaded, usually in rolling containers or on pallets. Most products need to be sorted to be sent to the right locations. Until this, the process is called receiving.

The next process is the replenishment process. This is similar to the process of stocking shelves in a physical supermarket, apart from the fact that the shelves are stocked from behind. This way the order pickers (shoppers) and replenishers can work simultaneously, without being in each others way. Also, FIFO (first in first out) stacking should be a more natural process.

In the picking process, shoppers pick the products that they need to pick to complete the orders. When the shoppers have finished their rounds, the totes are transferred to dispatch frames. These are taller frames that, when stacked, fit in the trucks and when unstacked in EPVs. Finally the dispatch frames are loaded into the outbound trucks.

Figure 5.3: Scanning in the fulfilment center, ©Picnic B.V.

Waste can come from all these processes. Important to mention is that not all products that are not sold, are waste by definition. Products can be used for internal use; products that are needed without ordering. An example of this can be forgotten products in the order for the canteen, or paracetamol/ibuprofen. Sometimes products are delivered at one FC, but have to be distributed to another, this is registered by the action intra-fc on the scanner.

5.3. Picnic is a data-driven tech company

Being an online supermarket that only lets customers order through an app, all Picnics customer actions and behavior can be registered. When an order is placed, this data, combined with all other order data is used to make a logical allocation for products in the right totes, to provide for an effective picking process.

All warehouse actions are data driven and (by means of the scanner actions) registered to keep track of stock levels, warehouse efficiency and performance. An efficient routing schedule for the EPVs to deliver the groceries is conceived by data-driven algorithms. Furthermore, historical data is used to analyze past demand and be able to accurately forecast demand. The forecast is used for ordering the right amount of products. In short, Picnic relies hugely on a vast amount of complex data.

5.4. Waste at Picnic is unknown

Even though all the scanning actions are stored, the exact waste at Picnic was still unknown. Primarily this was caused by the fact that the company is quite young and it had not been a priority. Now they are growing at such a fast pace, and want to convey the image of being a very green company, it is time to take action. In the scanner actions must lie the answer to waste, however it was unclear what actions needed to be regarded and which ones to be neglected. Also the difference between incoming and outgoing flows of products could be regarded. The difference between these two should be waste. Unfortunately it was not as easy as merely comparing the two. More about that later.

5.5. Conclusion

Picnic is a fast growing online supermarket that distinguishes itself by its efficient distribution model and smart use of data. It aspires to be as green as possible and therefore food waste needs to be reduced to a minimum. No serious analyses on food waste had been performed, so quantification and causes were unknown.

6
Design of the case study

Since no guidelines had been found how to study and predict the process of food waste generation, neither in literature nor at Picnic, I went into this research with a blank canvas and the research therefore was of the exploratory kind. Figure 6.1 shows the case study part the research once again. The case study exists of three parts: an exploration part, a qualitative part and a quantitative part.

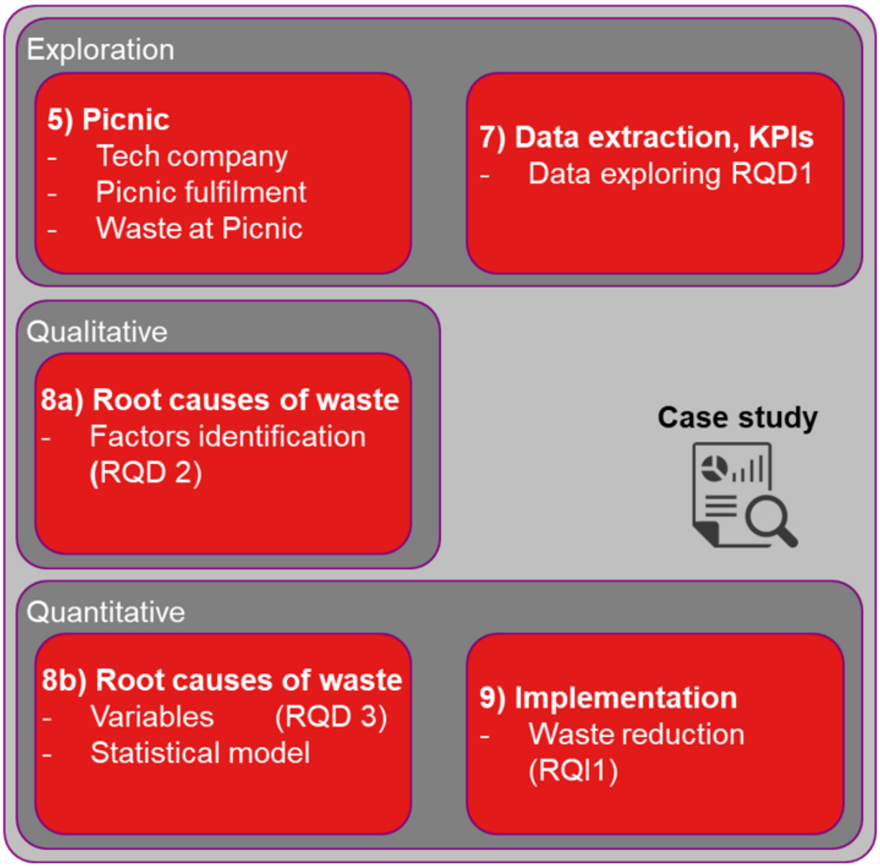


Figure 6.1: Outline of the report, case study

6.1. Exploration

The first part of the case study concerned itself with the question:

RQD1: Which data should be used to measure waste performance at Picnic?

The data that was going to be used, still needed to be accumulated and reviewed for relevance and reliability. From analyzing this data, combined with the indicated root causes that were found in literature, hypotheses about expected root causes were formed, divided in different categories:

- **Assortment**
- **Supply chain**
- **FC process**
- **Other**

Subsequently quantitative analysis of the found predictors was performed based on different KPIs and on various levels of the company. The formed hypotheses were tested by means of a linear regression model.

6.2. Qualitative analysis

The first step in the research was to gain an understanding of the data related to waste at Picnic. Exploratory data analysis (EDA) (Jebb, 2017) [28] was performed on the available waste data at Picnic. Picnic has various data that might be relevant to analyze waste stored in different locations. The data with the highest granularity that was quality assured by the company was found in the 'Stock mutations endpoints' (see Section 7.2). This data was extracted and filtered to only contain definitive waste.

From understanding the data and processes related to waste, four different domains were defined that drive waste: assortment, supply chain, FC processes and a rest category: other. This was at the basis of the research question:

"RQD2: Which factors are root causes of food waste?"

For each of these domains, expected root causes (factors) were set up, from the insights in literature, exploring the data and expert interviews.

6.3. Quantitative analysis: construction of variables

For as many identified factors as possible individual indicators were set up. This was done for all factors related to assortment and supply chain. No indicators were found for FC processes and other.

A data set was constructed to perform analyses on. The data set contains all identified factors and the KPI waste CU, as an outcome. This set has a data point for each combination of article (a constantly changing assortment of around [redacted] SKUs), FC [redacted], and financial period [redacted]. This comes down to around [redacted] data points. 80 % of waste comes from chilled products. The chilled areas contain fresh products. Further analyses have only focused on articles in the chilled areas, to gain the most important insights in waste.

Most analyses and data structuring have been done with Excel. Tableau software has been used for visualization and data blending.

6.4. Quantitative analysis: regression models

The goals of the research were to confirm that the quantified factors were indeed root causes, structuring the data to show insights in the process of food waste and to identify trends in the data. To confirm this, statistical models were applied to the constructed data set. This part answers the question:

"RQD3: Can the root causes be confirmed, using a statistical model?"

Since the research had mostly been exploratory and a lot of data had been collected, an accessible statistical tool was required to assess the significance of factors in a multivariate analysis. Waste was not expected to be perfectly predictable by making a regression model with the established expected root causes, so therefore a perfect fit was not the goal.

The first regression technique to be tested was multiple linear regression as it can show trends in the relation between multiple independent variables and waste.

Non-linearity turned out to play a role in the model, which caused violation of some linear regression assumptions. To improve the modeling and address the non-linearity in the data, two different regression models were considered. The dependent variable could be considered a count of a certain event: waste/no waste. Therefore two regression types were eligible: logistic regression and Poisson regression.

Logistic regression describes a model to predict if an event will occur or not. A binary dependent variable is required, so the preferred KPI needed to be decreased in granularity to waste/no waste. Logistic regression still had some of the same (violated) assumptions as linear regressions, but with a lower granularity output data. I experimented shortly with logistic regression, but due to the amount of zeroes in the data, the best predictive model was to predict zero with a very high probability. For these reasons logistic regression was discontinued and not included in the report.

The distribution of the dependent variable looked somewhat like a Poisson distribution. However, it was not exactly a Poisson distribution, but could be modeled with a negative binomial regression model. Negative binomial regression therefore has been performed on the same data sets to show that improvements can be achieved when nonlinearities are addressed and also to reinforce the conclusions about root causes.

SPSS statistical software and Excel have been used to run the regression models and process results.

7

Data extraction, KPI selection and first insights

This chapter answers the question:

RQD1: Which data should be used to measure waste performance at Picnic?

Picnic's data structure is explained, possible waste data is discussed. Then the chosen data is explained and corresponding KPIs are set up.

7.1. Large amounts of data from different sources

The vast amount of data that Picnic collects, cannot all be stored. It simply would be too much information, not to mention the work that would go into structuring all data. Instead, data engineers make sure that relevant data is stored in a structured way.

Figure 7.1 gives a simplified view of the most relevant data sources that were used during this research. Data that is generated or registered by the operational WMS can be extracted from so-called endpoints. These are small bundles of data regarding one specific topic. The endpoints are usually FC specific, and only small amounts can be extracted at once to not impact the performance of the WMS. The bulk of the data is stored in Picnic's data warehouse (DWH). This is the big data source that can be used for performance tracking and analysis. The data is constructed in structured tables and can be accessed from servers via Tableau software.

Data that is stored in the DWH can come from different sources. The first one being operational systems, such as the WMS. The second source is a variety of spreadsheets and flat files. E.g. lists for clearing products, planning schedules, productivity tracking. Picnic app events is the third source. This is all the information that is stored from the app, such as customer behavior, time spent on pages, searches. The final source is reference data, such as demographics and marketing data.

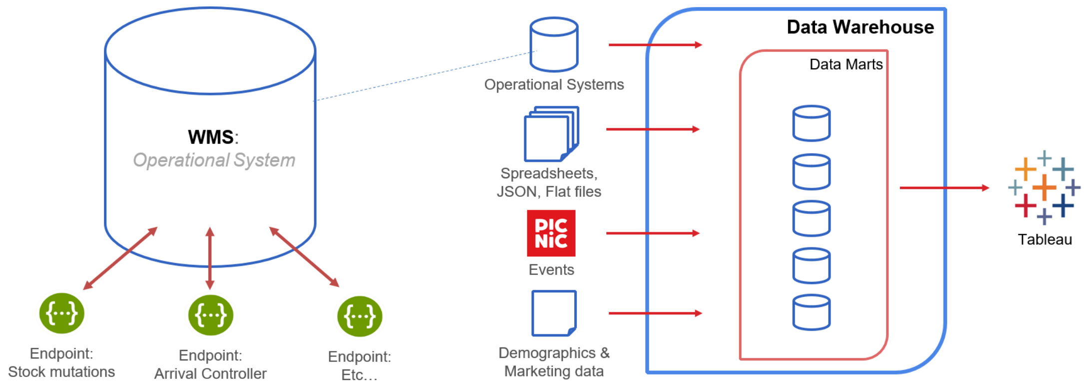


Figure 7.1: Picnic data structure extract, *Source: adjusted image of Picnic data team presentation*

7.2. Waste data from WMS

The waste data for this research came predominantly directly from the WMS, using the endpoint 'Stock Mutations.' The data in this source has the highest granularity, being actual scanning actions from the FC. The format is a table that contains (among other things) the article id, type of action, reason of this action, number of CUs and a time stamp. Data can be extracted per FC. An example data extract can be found in Appendix B.

The endpoint data has been tested extensively and is quality assured. It means that all scanning actions in the warehouse end up in the data. This is not to say however that all actions are correct! Humans can and will make errors and this will by definition also show in the data.

Since the data from the endpoint can display all stock mutations, not all data points are related to waste. Figure 7.2 shows which combinations of type and reason are considered waste. There are three types of actions that can be waste: Stock Clearing, Stock Adjustment and Stock Count.

Stock clearing is associated with the removal of products from a shelf and is usually done when products cannot be sold anymore because of freshness reasons. A 'blank' stock clearing is an actual clearing. This is considered waste and therefore used in the analyses. The reason 'Adjust for BBD clearing' is an automated action in the WMS, that adjusts the stock when different stock levels are detected. This is not waste and therefore not used for analysis.

A stock adjustment can be performed to register products that are taken out of the shelves or thrown away for another reason. The reasons 'Bad quality' and 'Damaged' are waste who speak for themselves. They are used for analysis. The reason 'Out of assortment' concerns a clearing because a product is no longer in the assortment: customers cannot order it anymore and the product is cleared. This definitely is waste and is used for analysis. The reason that 'Old production date and freshness guarantee removal' are in the stock adjustment type section, is section somewhat odd. They are deprecated reasons that concern a stock clearing for freshness reasons. In my opinion they should be either excluded from stock adjustments in general, or moved to the stock clearing section. For freshness clearing reasons in general, an actual stock clearing should be used, but when shoppers at some point have become used to using 'old production date' and 'freshness guarantee removal,' they might still use these.

A stock count is an adjustment to the stock, when the physical stock does not match the expected stock. The normal reason 'Count' is the hardest one to judge. When a person is not using the appropriate clearing or adjustment types to clear products, they might do it with a stock count. If so, it is waste. However, daily thousands of count actions are done to adjust stock and only extremely remarkable mistakes are noticed. In this study, We assumed that Picnic employees scan correctly. A stock count therefore is not considered waste.

Figures 7.3 and 7.4 represent a simplified version of the stock mutations, and the stock mutations considered for analysis.

WMS Registration type	Reason	Waste	Used?
Stock Clearing	Adjust for BBD clearing	No	No
	'blank'	Yes	Yes
Stock Adjustment	Bad quality	Yes	Yes
	Damaged	Yes	Yes
	Out of assortment	Yes	Yes
	Old production date	Yes	Yes
	Freshness guarantee removal	Yes	Yes
	Internal use	No	No
	Intra FC move	No	No
Stock Count	Count	Unclear	No
	Count from zero	No	No

Figure 7.2: WMS stock mutations

WMS Registration type	Reason
Stock Clearing	Freshness
Stock Adjustment	Quality
	Damaged
	Assortment changes
Stock Count	Adjustment of stock levels

Figure 7.3: WMS stock mutations simplified

WMS Registration type	Reason
Stock Clearing	Freshness
Stock Adjustment	Quality
	Damaged
	Assortment changes
Stock Count	Adjustment of stock levels

Figure 7.4: WMS stock mutations chosen for analysis

Figures 7.5 shows the distribution of waste, divided between stock clearings and stock adjustments. Around two thirds of the waste is cleared for freshness reasons and one third for other reasons.

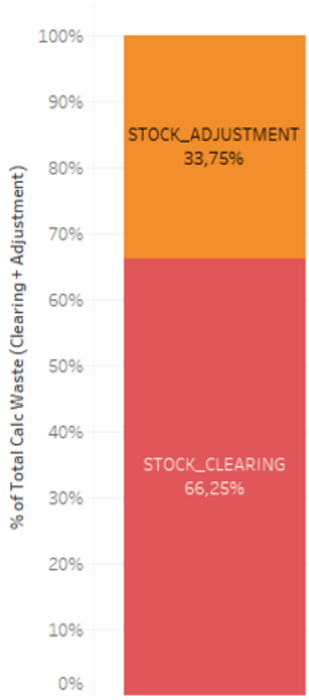


Figure 7.5: Waste composition

7.3. Waste changes over time

Waste is influenced by different factors, that can all change over time. Three identified main factors are: demand, assortment and procedures in the FCs. They are explained in more detail in Chapter 8.

To get an idea of how waste is influenced, a longer period of time needs to be considered.

Waste is regarded per financial period (P) of exactly 4 weeks: 28 days. A total period of 8 finacial periods have been considered: P1 up to and including P8. This is the period from 01/01/2018 to 12/08/2018. The length of a financial period has been chosen to be in accordance with the financial reporting of Picnic. Figure 7.6 shows the composition of waste in the first 8 periods, divided per temperature zone. The first insight is that by far the most waste is generated by the chilled products. This is striking, since of the roughly [redacted] SKUs that were considered, around [redacted] were ambient, [redacted] were chilled and [redacted] frozen. However, the effect was to be expected, since the chilled products are by definition fresh products. This means that they are likely to expire sooner than most products in the ambient temperature zone. There are fresh products in ambient, such as bananas, tomatoes and some other fruits and vegetables. This puts ambient in second position with regards to waste. Frozen not only has a relatively small range of SKUs, but also the products that are frozen usually have a very long shelf life.

The second insight is that the amount of waste per temperature zone is not steady. P1 and P2 have much higher ambient waste numbers than P3 to P6. The products that caused these higher numbers were mostly Christmas and new years products that were cleared, after they would not be sold anymore. An example is a Christmas stollen.

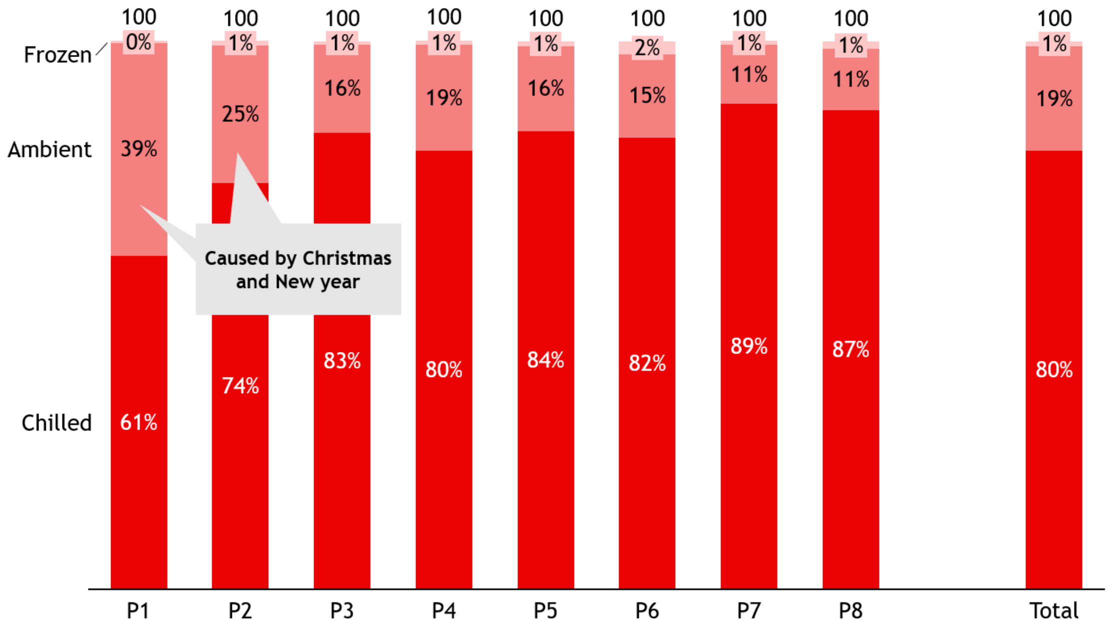


Figure 7.6: Waste per temperature zone per period

7.3.1. Waste - a black box approach, did not work

I have tried identifying waste at Picnic in a black box manner. Inbound products that came into the fulfilment center (were received) were compared with the outbound products the went out (were sold). Internal use and intra FC were adjusted. To do this, a set period of time needed to be regarded. Unfortunately for whichever length of period I chose, no realistic results were achieved. This was caused by three effects:

- The borders of an observed period were hard edges: overlapping effects between two periods were cut off and this blurred the data, since large quantities could be received, sold, or taken out of the assortment.
- Freshness or expiry dates (BBDs/PDs) were not always registered or obtainable. This way individual shipments or batches could not directly be traced.
- Receiving data is unreliable because of mistakes at the supplier and/or mistakes at receiving (that might be compensated with stock counts, but are therefore 'lost' in the data).

7.4. Other data sources used for waste analysis

In addition to the raw number of products that are wasted, we want to know which products, and mostly why they became waste and how this relates to the context. Information about the context can be found in different data sources.

- In the data warehouse, all incoming data is processed and divided in a similar way as the endpoints. The DWH is designed as a dimensional model (Kimball 2008) [29]. The small blocks into which the data is divided are called dimensions (DM) and are related to one subject, for example: a table with information about articles. Different dimensions can be attached to a central table called a 'fact' table (FT). For this research, the MART 'orderline' was used as a primary source from the DWH. The dimensions and fact that were used for this study are:
  - **DM Article:** information about each article: article id, name, product category, temperature zone
  - **DM Date delivery:** delivery date of orders
  - **DM delivery:** filter for actual order
  - **DM order:** order id, to calculate the number of orders and demand quantity.
  - **FT orderline:** financial data about orders: sales qty, cost of goods sold (COGS, see Section 7.4.1), net sales
- Another source that was often used was the product information management (PIM). This Excel file contains data concerning the SKUs, such as name, category, supplier, dimensions and weights, content, temperature zone, consumer units per trade units (CU/TU), freshness guarantee in days, etc.
- Picnic also uses a lot of google sheets. These sheets can have all sorts of functions, but mainly are automated processes that keep track of schedules, product move lists, clearing lists and productivity measurements. One google sheet that was used for the waste analyses was one that kept track of the order groups of products. This sheet was live, so always up to date with the current assortment.

7.4.1. Cost Calculation

The purchase price of products at Picnic is expressed as Cost Of Goods Sold (COGS). The stock of a product is evaluated based on the latest purchase price at a supplier. The COGS are registered for each day that products are sold. However, not every day, all products are sold. If a table is constructed, with on the Y-axis the products ([redacted] SKUs) and dates on the X-axis, only the days that this product has been sold, has values. There are a lot of empty fields. When a product was cleared on one of these days, no value could be found for the cost of the product. If no value was available, the closest previous value was chosen. It that also was not available, the closest next value was chosen.

7.5. KPI construction and selection

The most commonly used KPI for food waste, as found in literature, is mass. As mentioned in section and described in Section 2.2 this KPI is mainly useful in research based on a larger scale. At Picnic, measuring the mass of food waste is not particularly useful. Picnic as a company is interested in KPIs such as quantities, percentages costs and causes. These KPIs give more insights that can lead to concrete actions.

A more useful KPI reported in literature for measuring food waste at Picnic is the shrink rate [20]. This shows the units wasted relative to the amount of units that was purchased. Especially when split to different levels such as product, product type or supplier, it displays waste intensity. The shrink rate was defined as:

$$\textit{Shrink rate} = \frac{\textit{Units unsold}}{\textit{Units purchased}} \tag{7.1}$$

A problem with this measure is that the term 'units purchased' is found to be an unreliable data source at Picnic. The data can be found in two ways:

1. Units purchased from invoice data. The problem with this metric is that what is purchased is not always delivered to the warehouse. There is often a discrepancy between invoices and receivings.
2. Units received from WMS data. When receiving units, a scanner action is used by someone in the fulfilment center and is registered in the WMS. This action is prone to human error.

Therefore, new KPIs needed to be defined, that are based on the most reliable data. As mentioned in Section 2.2.2, it was mentioned in literature that it does not necessarily matter which KPI is chosen, as long as it comes from a reliable data sources and it is possible to monitor closely. For this research five KPIs were constructed. Figure 7.1 shows the amount of the described indicators, for period 1 up to and including 8, 2018.

1. The most convenient and least complex KPI is *'absolute waste in units (CU).'* This is the sum of the items that are considered waste, as scanned in the FC and explained in Section 7.2.
2. The second KPI is the *'Absolute cost of waste in Euros.'* It is the total number in units multiplied by the cost of goods sold (COGS) at the day of waste. The cost of goods sold is a term used at Picnic to relate to the purchase price of an article. The article and date need to be taken into account, since the COGS can vary per day.
3. The third KPI is *'Relative waste: CU waste/CU sold'.* This approaches the above mentioned shrink ratio the most. The number of items sold is a reliable data source at picnic, since all orders and customers depend on this number being correct. Percentages can show the intensity of waste for different articles, periods or FCs. Unfortunately, in some situations there is a problem with this type of KPI, described in section 7.5.1.
4. The fourth KPI, especially defined for Picnic, is *'Waste, expressed in cents/item sold.'* Picnic's fulfilment reporting expresses most costs of operation in cents per item sold. This way, different costs that Picnic has across various divisions can be compared and efficiency can be monitored. Costs are generally considered per financial period and are divided by the total number of sold products in that period.
5. A fifth KPI was also constructed specifically for Picnic to assess if products are worth wile having in the assortment, solely based on the waste a product generates. It is called *'Waste margin impact.'* The cost of waste (per item) is divided by the gross margin. Gross margin is defined as the net sales - COGS, so can be seen as profit before costs. If the cost of waste is bigger than the gross margin per item, there is a negative profitability of the product, the margin impact is bigger than 100 % and Picnic should seriously consider removing this product from the assortment.

[redacted]

Table 7.1: KPIs with quantities, P1 - P8, 2018

For further analysis, Abs waste CU and Relative waste CU were the most convenient KPIs, since they came from the most reliable sources and were least distorted by other data sources. They are most similar to the shrink ratio [20], measuring units of waste. They are discussed in the following two Sections.

7.5.1. KPI 'Relative waste CU'; a product goes out of assortment

When a product goes out of assortment, stock is left that is usually wasted. A problem with the KPI 'relative waste (CU wasted/CU sold)' is that when no sales have taken place in a period of time and waste is registered, very high or unrealistically low waste percentages are registered as a result. Figure 7.7 shows this effect.

Case 1 describes the situation where a product is taken out of the assortment at the beginning of period 1. There is a lot of waste, on this product, represented by the red bar. Since there are no more sales after the product is taken out of the assortment, an extremely high percentage of waste is possible (highest measured was [redacted] %).

Case 2 describes a situation when a product is taken out of the assortment at the end of a period. Only a small part of the waste is registered in the period itself, which causes an very low percentage of waste. The rest of the waste might be cleared in the next period. Then sales are nill and waste is considerable. A percentage cannot be calculated.

This problem increases when analysis is done on higher granularity data. For example, if one product goes out of assortment as described in case 1, and therefore causes extreme relative waste. When comparing this to the relative of the whole FC, the effect will be much less, due to the bigger scale: the total numbers of sales will always be much higher than numbers of waste.

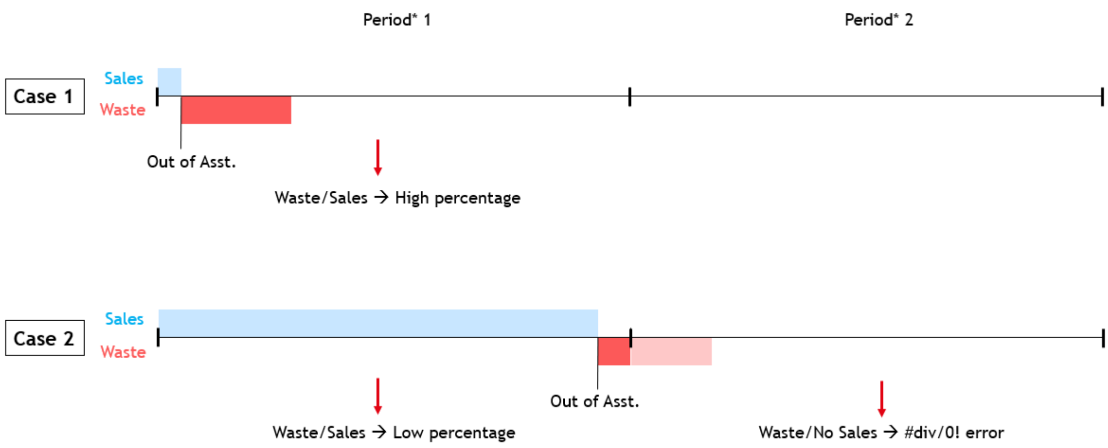


Figure 7.7: Out of assortment waste; KPI problems

7.6. Chosen KPIs: CU waste for analysis

Different KPIs of waste might be useful for different purposes.
Relative waste is an appealing KPI, since it can display intensity of waste. However it can give a wrong impression for certain cases, as previously mentioned.

Absolute cost of waste, waste in cts/item and waste as a margin impact are all useful measurements for Picnic, and mostly in line with their reporting. These KPIs are mainly used in the implementation to assess which actions to take to improve waste at Picnic.

For further analysis purposes, absolute waste CU has been chosen as the main KPI. It is the most reliable and complete data for waste at Picnic at this moment. The output is an absolute number of products wasted. The indicator is not distorted by any other data sources.

7.7. On overview of waste at Picnic

7.7.1. High waste for young FCs

Even though relative waste does not always give a realistic view for detailed analysis, on the granularity of /week /FC it works sufficiently adequate. An overview of the relative waste per FC over time is given in Figure 7.8. Blue represents all stock adjustments, so fore example broken items, or out of assortment waste. Orange represents all stock clearings, based on freshness reasons.

FC0 and FC1 are the more mature FCs. FC2 opened in week 5 and FC3 opened in week 21. We can clearly see the difference in mature FCs vs young FCs. Peaks are seen at the opening of an FC that quickly decreases in the following months. The average of all FCs is [redacted] %. The FC that is considered most efficient; FC1, scores best with an average percentage of [redacted] %. The smaller blue peaks in FC0 and FC1 are mostly leftover stock from Christmas, that is cleared as 'out of assortment,' as described earlier.

[redacted]

Figure 7.8: Waste per FC over time

7.7.2. Waste decreases when orders increase

Figure 7.9 shows the waste percentage of units sold versus the number of orders of an FC. It shows that the more orders an FC fulfills, the less waste is caused. This is beneficial for the future of Picnic. It can expect waste numbers to drop significantly when more FCs perform according to their capacities. FC0 has hit its capacity of around [redacted] orders per week. FC1 fulfilled the most orders ranging from [redacted] to [redacted]. FC2 has seen significant growth and performed well on waste. FC3 has seen a progressive drop in waste, from the moment it started.

A trend line with an $R^2 = 0.50$ and $P - value < 0.0001$ can be fitted with the formula:

[redacted] (7.2)

[redacted]

Figure 7.9: Waste vs. number of orders

7.8. Data inconsistencies

Every now and then, there were actions showing up in the waste data, that were not actual waste. An example:
In the data analyses, every now and then huge waste items came up, with articles that had 'voedselbank (food bank)' in the name. This requires some explanation: Picnic customers can order food packages to go to the Dutch food banks. If they do, this is registered in their order. Picnic wants to thank the customer on behalf of the food bank with a flyer. An order picker would scan the flyer and put it with the other groceries in the allocated tote.

After such an action has ended, there might be flyers left. When the flyers are scanned to take them 'out of assortment,' they appear in the waste data. And with the way the cost of waste is calculated, there can be very high cost of waste for these flyers, while not actually being physical products that are wasted.
These items which were definitely not waste, have been deleted from the data.

7.9. Conclusion

The most reliable, quality assured, high granularity data comes directly from the WMS, by means of the endpoint: "stock mutations." The data is combined with other sources from the

DWH, google sheets and lists to give sufficient information about waste at picnic.

KPIs as found in literature proved not applicable for this research. Either input data for a KPI was not accurately measured (mass) or data was not the most reliable data regarding waste (received items). Therefore five new KPIs have been set up that could be used. The most appealing KPI is relative waste, expressed as a percentage of sold products. This KPI however cannot be used for in-depth analysis, because it is not applicable to all products on all levels of granularity. This metric gives extreme non-representative, outlying values.

The KPI waste CU has been used for analyses and regression. The KPIs cost of waste, relative waste, waste in cts/item and waste as an impact on the gross margin were mainly used for Picnic's purposes and reporting.

Waste changes over time, influenced by different so-far unidentified factors. We have seen that around 80 % comes from the chilled temperature zone, around 19 % from ambient and 1 % from frozen. This indicates that waste is directly linked to fresh products.
Figure 7.8 and 7.9 show that new FCS have more waste and the more an FC is working towards its capacity, the less waste arises.
For the full waste analysis, data has been structured and gathered per financial period of 28 days. This length was chosen to be in line with the operations reporting, as performed by Picnic.

8 Identification of root causes of waste

This chapter answers the question:

RQD2: Which factors are root causes of food waste?

In this chapter, expected root causes at Picnic are discussed that were identified from literature research, exploring the data and expert interviews. Hypotheses about the effect of the expected root causes were generated. The identification of expected root causes was a qualitative process. By means of quantified analyses, indicating variables related to these expected root causes have been constructed, for this case study specifically.

8.1. Qualitative effects of factors on food waste

During the literature research and the exploratory data analysis phase, hypotheses about root causes were created. This was done by analyzing literature, identifying the biggest waste cases at Picnic, talking to experts, inspecting properties of products, demand and processes and hereby identifying possible problems. For all these properties, data was gathered and analyzed. From different waste cases, it soon became clear that there was not one single factor causing most of the waste. It was likely to be a combination of different factors that were causing waste.

An example was that I found large amounts (in absolute Euros) of waste on a certain type of instant meals. All meals from this type had a lot of waste. When inspecting various different properties of this product, it turned out that demand was low, but very volatile, the articles were ordered with a high safety factor (explained in Section 8.2.2), the products were relatively expensive, and the delivery schedule was irregular, with a large maximum time between deliveries. These possible reasons were included as possible identified root causes. On top of that, data was collected for each of these possible reasons for all data points.

Resulting from expert interviews, four domains were expected to cause waste. Each domain is expected to contain various different root causes of waste. Figure 8.1 shows an overview of the four domains: assortment, supply chain, FC process and other, and their identified factors that are assumed to attribute to the waste generation at Picnic. This can be considered a driver tree. All the branches have a possible root cause for waste. Note that this is not a mutually exclusive and collectively exhaustive (MECE) tree. Waste can be caused by a combination of factors that are shown in this tree. Most of the expected root causes were also identified in the literature study. They were arranged to their respective domains.

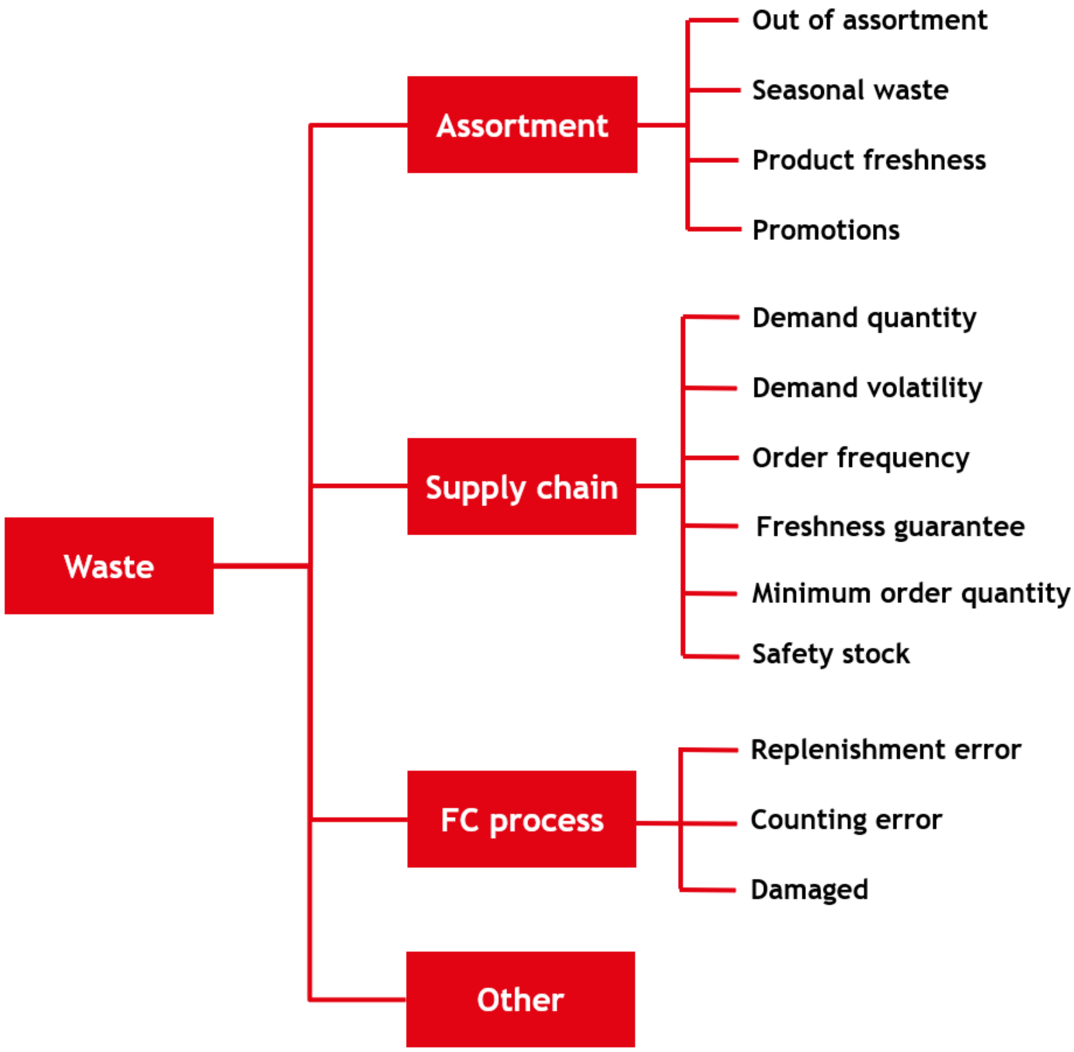


Figure 8.1: Picnic's waste domains and expected root causes

8.1.1. Factors: Assortment

Waste can be caused by different aspects of the articles in the assortment. Freshness reasons, changes in the assortment and irregular products can be causes of waste. Picnic focuses on having a compact but complete, effective assortment that pleases most customers. The department Customer Success (CS) collects requests from customers for products to be included in the assortment. When demand is high for certain products, these products can be included in the assortment. This leads to a constantly changing assortment.

Assortment related waste issues can be divided into more specific assumed root causes. I made the following subdivision: the firs two factors that are described, are physical aspects of the products. The second two are related to assortment choices of management.

The first factor is **seasonal.** This factor came up in an exploratory analysis, when it became clear that there had been more waste on products with some sort of seasonal labeling. This means that they are not always in the assortment. They can be in the assortment for a longer period of time. Seasonal products usually have different demand and/or ordering patterns than the regular assortment. Some examples are: chocolate letters, Christmas luxury meats, Easter chocolate eggs, BBQ meats/packages, Asparagus. Especially the event related products, such as Sinterklaas and Christmas articles are ordered in another way than regular products. These products have to be ordered a long time ahead and are therefore not dependent on a short-term forecast, but more on the growth of the company. Especially for event articles, such as Christmas products, availability is crucial for customer satisfaction. Because of the fact that forecasting demand is more complicated, the hypothesis about seasonal products is that they cause more waste than regular products.

The second factor related to the assortment is **product freshness.** This is a straightforward factor, and was described in literature.
Different fresh products can have different shelf lives. Also, not every delivery has the same shelf life remaining for the same products. This can vary over time. When a product has lower shelf life remaining, it needs to be sold sooner in order to not become waste. The hypothesis about freshness in terms of shelf life is that the shorter the shelf life, the more waste will be caused.

The third factor that is identified concerns the period when a product is taken **out of the assortment**. This was a factor that was identified by exploring the data at Picnic and had not been found in literature. Picnic's assortment is not only expanding: sometimes products are taken out of the assortment. This can have various reasons. A product might not be profitable, agreements with a supplier might not work, there might be no demand for the product anymore, etc.. When a product is taken out of the assortment, the item disappears in the app. It can happen that picnic has not sold all of the products, and has leftover stock. This stock usually is cleared and considered waste. The hypothesis is that when a product is taken out of the assortment, it is likely to cause waste.

The fourth and final factor related to the assortment is considered to be **promotions**. The factor was identified in literature as a possible cause of waste. At Picnic promotions are discussed with suppliers, a few weeks before the starting date of the promotion. Usually promotions have a duration of one week. Forecasting demand during a promotion is done by learning from promotions in the past. Considering a forecast for a bigger period of time, with less stable forecasting parameters and different demand patterns, the hypothesis is that promotions cause waste.

8.1.2. Factors: Supply Chain

Supply chain factors, such as demand patterns of customers, agreements with suppliers and chosen order strategies can all influence waste. The supply chain probably is at the basis of most of the food waste. Many links and countless variables can affect food waste in the supply chain. In this section supply chain is restricted to Picnic's part in it. This can be Picnic's interactions and agreements with customers and suppliers, but also managerial decisions that have been taken. For this study, six factors have been identified that attribute to supply chain waste.

Demand quantity was the first factor that came forward from literature [4]. When regarded by itself it is a straightforward variable. When a product is in high demand, the order quantities are bigger and will likely be causing more waste, since a smaller deviation in the forecast will lead to a bigger waste number. The hypothesis is that products with a higher demand cause more waste.

A second factor that is associated with demand and is expected to be at least as important, is its volatility. The **volatility of demand** says something about its predictability. This factor can be regarded as the root cause found in literature, described as forecasting difficulties [22], [21]. When unpredictable products are ordered on the basis of a forecast, it is very likely to generate a lot of waste. The hypothesis for the volatility of demand is; the more volatile a product, the more waste.

Order frequency is the next factor that was identified. It was also indicated in literature. Order frequency is mostly dependent on the supplier or the agreements that have been made with a supplier with regards to the days that he/she delivers. Different suppliers can have different delivery schedules. Most suppliers come on an agreed set of days. Some suppliers have a bi-weekly delivery schedule. When a supplier delivers often, there are more moments to adjust the stock and a shorter period of time needs to be forecasted. When there are less deliveries, a longer period of time needs to be forecasted and less moments are available to adjust. The hypothesis is: a higher order frequency leads to less waste.

Freshness guarantee (FG) is expected to have an impact on waste. This factor is extra sensitive for e-grocers and therefore came forward in data analyses.
It is the number of days Picnic guarantees a product to remain fresh and fit for consumption. It is a parameter that can be changed easily. If products tend to have high waste due to their freshness guarantee, it can be reduced almost instantly. The hypothesis about freshness guarantee is that a (too) high freshness guarantee causes waste.

Minimum order quantity (MOQ) is the minimum amount of products that can be ordered in one delivery. This factor was taken into account because it was mentioned in expert interviews. The minimum order quantity often has to do with the supplier or the way the products are packaged. Especially on lower orders, a multiple of the MOQ can cause differences between the forecast and the ordered quantity. The hypothesis is: large minimum order quantities generate more waste.

Safety stock is an extra stock on top of (predicted) demand that is used to guarantee availability of products. It is a common phenomenon in retail/ supply chain and was identified as a root cause in literature [4].
The hypothesis is that a bigger safety stock causes more waste. This is mainly the case when the waste is considered from a relative point of view. Picnic's order algorithm is formulated in a way that slow-movers receive a relatively higher safety stock than fast movers. More is explained in the quantitative part of safety stock.

8.1.3. Factors: FC processes

In the FC there can be various reasons for products to become waste. One can think of replenishment strategies, physical layout of an FC, number of times a product is handled and human errors in the scanning actions. Some of these processes that are recorded with scanning actions are: receiving, moving stock, replenishing, picking and dispatching. The factors that are listed here mostly came forward from talking to employees and during my onboarding time, when I worked in the FC for a couple of days to experience the processes first hand.

The first important cause presumed to be related to waste is **replenishment.** The replenishment process happens manually, from the back of the shelves. It is important that the shelves are stacked in a first-expired-first-out (FEFO) way. This means that products that expire first, need to be placed at the front of the shelves and therefore are picked first. It can happen that a product that expires sooner, is delivered at a later point in time or in a different batch than products that are already on the shelves. Shoppers would need to check all expiry dates of the products on the shelves and compare them with the expiry dates of the newly received products. It is not likely that shoppers will do this consciously, because in general they are under time pressure and they have no incentive to check all products. The hypothesis is that non-FEFO replenishment leads to waste.

The second cause is related to human errors in **waste registration.** All waste accounting counts on data that is generated by shoppers through the scanner. When an error is made, whether this is a typing error or a counting error, this will show up in the data. Incorrect registration can work two ways: there can be a registration of waste while there is no waste, or there can be no registration of waste while there is waste. Both reasons will influence the assumed stock levels and, if not corrected in time, the order algorithm. When the WMS thinks there is no more stock, while in reality there still is, more stock will be ordered. This means that the product will be over stocked and chances for waste increases. If no physical stock is present, but the WMS thinks there still is, this will lead to a lower order and increase the chance of unavailability. Both are undesired effects.

The third cause is when products are **damaged.** As mentioned before, products need to be handled in different processes. During all of these handlings, there is a chance of breaking or damaging a product. Damaged products cannot be sold anymore, and are considered waste.

Another factor that can be considered waste is **theft.** Being a shopper can be tough, it is hard physical work and can be monotonous. This can make people hungry or thirsty. Then being in an environment with plenty of food and drinks around, it could stimulate people to take one of the products on the shelves. Empty cans of soda and used chewing gum packages are notorious to show up in the shelves/bins every now and then. These products cannot be sold anymore and therefore are waste. Theft is expected to have a very small impact and is not considered for further analysis.

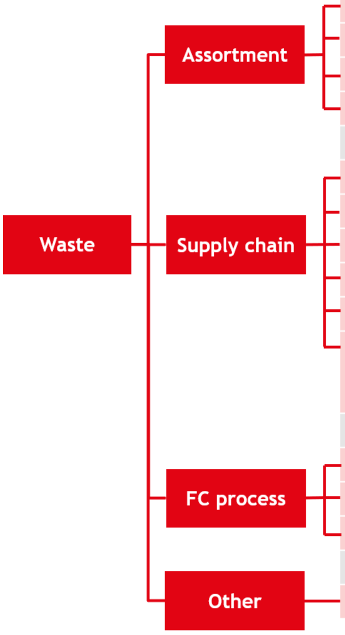
8.1.4. Factors: other

Not all root causes for waste have been listed above. One could still think of more reasons. However, the above mentioned factors were considered the most important and relevant in the research. The bigger effects that have been described in the literature part of this study have not been considered, such as market mega-trends with its changing demand patterns and natural causes such as weather fluctuations that make everyone want to go BBQ'ing. Also, hard to measure aspects such as training of shoppers quality have been left out of the analysis.

8.2. Quantitative indicators for factors

Having identified the expected factors associated with food waste at an online supermarket's fulfilment centers, we are interested in measuring them.
From exploring the data for previous processes, I had a good feeling with the data and was able to distinguish which factors could be quantified, and which could not. Most variables that were identified needed separate analysis in order to collect and obtain the correct data for all considered data points. This task of identification of the correct data and the variables for measuring the factors is one of the most important contributions of this thesis. After establishing the variables a data set was created that had granularity level: FC/Article/Period.

For the factors associated with assortment and supply chain, variables have been found for each factor. No accurate variables were identified for the FC processes.
Figure 8.2 shows the factors with their variables, as well as type and range of values, that came from the analyses. They will each be discussed.



Factor	Variable name	Type	Range
Out of assortment	Goes out	Nominal	[0 / 1]
Seasonal	Season	Nominal	[0 / 1]
Product freshness	FC period (days)	Discrete	-3 - 344
Promotions	Promo pressure	Continuous	0% - 100%
Demand qty	Article Order Rate (AOR)	Continuous	0 - 0.50
Demand volatility	Coefficient of Variation AOR	Continuous	0 - 5
Order frequency	Max T betw. Deliveries	Discrete	1 - 7
Freshness guarantee	FG (days)	Discrete	0 - 21
Minimum order qty	CU/TU	Discrete	1 - 92
Safety stock	Safety Factor (SF), High SF, High SF * CU/TU	Continuous, nominal, interaction	0.2 - 0.5, [0 / 1], 0 - 92
Replenishment error	?	-	-
Counting error	?	-	-
Damaged	?	-	-
Unidentified	?	-	-

Figure 8.2: expected root causes and indicating variables

8.2.1. Quantitative indicators: Assortment

- Out of assortment: **'Goes out'** is a binary value, that was set up using data from the DWH; MART orderline. Articles either were in the assortment or out. Analysis was done to check if an article in a FC in a period was in or out of assortment. The first period that a product went out of assortment this the variable was assigned a 1, meaning that the product goes out in this period. For analysis purposes, the period after a product went out of assortment was also considered 'Goes out', in case taking the product of assortment happened at the end of a period, or waste was present for this product in the next period.
- Seasonal: **'Season'** is also a binary value, applied to products that in article databases such as PIM and salesforce had the label, 'season, bbq, sint or kerst.' Articles that have the label seasonal, are considered seasonal independent of time.

- Product freshness: **'FC period'** was defined. This also required an extensive analysis. The term product freshness was introduced to indicate the maximum amount of time in days that products could be held in the FC. One can see it as a shelf life, within the FC. The analysis that was done calculated the last delivery date minus day of receiving, taking into account all receiving data with freshness dates for all products for the full period and all freshness guarantees of these products. The most occurring freshness (mode) that products were delivered with was used. The freshness guarantee was then subtracted from this, to determine the FC period.
- Promotions: **promo pressure** is a metric that can be used to see how much of the total quantity sold was sold in a promotion. This measure was analyzed, using the MART orderline, assessing the amount of sales in promotion and the total amount of sales, on FC/article/period granularity. It is defined as:

$$Promo\ pressure\ = \frac{Promotion\ sales\ quantity}{Original\ sales\ quantity} \tag{8.1}$$

8.2.2. Quantitative indicators: Supply Chain

- Demand quantity: **Article order rate** is a metric mainly used by Picnic's Purchase Order Management (POM). It is used to forecast demand. It displays the demand as a ratio of the amount of orders and is defined by:

$$AOR = \frac{Original\ sales\ quantity}{Order\ quantity} \tag{8.2}$$

  The AOR can be looked at from different perspectives. It can be regarded, for the whole of picnic, per FC, per day, week, month or year. etc. For this data set, an AOR was calculated per FC, Article, date. An average AOR was then calculated for the 28 days in one period. Note: although the AOR compensates for fluctuating demand through the week, by dividing the sales qty by number of orders, meaning that the AOR can be constant through the week even though sales quantities vary. On the other hand, taking an average of the AORs takes away daily seasonality. An example: if one product is ordered much more on Fridays, such as beer: the AOR will be higher on Friday. In the average article order rate this can not be seen.

- Demand volatility: **Coefficient of variation of the article order rate (CoV AOR).** This is the variable that is assumed to have the biggest impact on waste. Volatility can be expressed as the coefficient of variation; a ratio between the standard deviation and the mean of a series. As an example: when a product has a coefficient of variation of 1 with a mean demand of 100, this means that the next day demand might be 100 again, but might just as well be 0 or 200! Basically products with a high coefficient of variation are unpredictable. In this case the article order rate is used as the demand variable. So volatility of demand is defined as coefficient of variation of the article order rate:

$$CoV\,AOR = \frac{\sigma_{AOR}}{\mu_{AOR}} \tag{8.3}$$

  Both σ (standard deviation) and μ (mean) are calculated in a separate analysis, using the 28 AORs per period. Picnic's forecast takes the day of the week into account, so compensates for daily seasonality. Still it is expected that with this definition of periodical coefficient of variation of the article order rate, the influence on waste can be measured.

- Order frequency. A metric that can indicate waste, caused by the order frequency, was identified to be the **maximum time between deliveries**. The parameter that is expected to influence waste most, is the longest period between two deliveries, since this is the longest period of uncertainty. This metric came into view when an analysis was done on a high waste item. The delivery schedule was inspected and this turned out to be bi-weekly. Most deliveries were 2 days apart, but the maximum time between deliveries was 5 days! This was assumed to cause the waste. All suppliers delivery schedules have been analyzed. This metric therefore is supplier specific, including a range of articles. It is independent of time.

- Minimum order quantity (MOQ). Picnic has no definite registration of a minimum order quantity. Another measure needed to be composed.
  The products we know as consumers are called consumer units (CU). This can be for example a pack of rice. At the supplier these are packed in a bigger carton, called a trade unit (TU). **CU/TU** often can be seen as the minimum order quantity. This metric is used and the data came from an extract of the product inventory management. It is independent of time or FC.
  An example for further explanation: the previously mentioned packets of rice are in a TU with 20 CUs and the MOQ is 1 TU, or 20 CUs. When forecasted demand is 22 CUs; 2 x 20 = 40 CUs will be ordered, meaning 18 more than forecasted. Rice fortunately does not expire so soon, so in this case it is no problem. However, fresh products will suffer more from this phenomenon.

- Safety stock: as mentioned in the qualitative part of safety stock, a higher safety stock is expected to relatively cause more waste.
  The way the safety stock is generated, is not very straightforward. First, a safety factor is determined per article in an order. Small orders receive a higher SF than a bigger order. This is a factor that is ordered on top of the forecasted order to compensate for missed deliveries. Over predicting with potentially waste as a result is considered less bad than under predicting, with potentially unavailability as a result. For products with a daily forecasted demand (D_f) of less than 10 units, a safety factor of [redacted] is applied. For products with a daily predicted demand of more than 100 units, a safety factor of [redacted]. Between 10 and 100 the SF is determined by [redacted], expressed as Equation 8.5.

$$SF = \text{[redacted]},\ for\ D_f \leq 10 \tag{8.4}$$

$$SF = \text{[redacted]}\ for\ 10 \leq D_f \leq 100 \tag{8.5}$$

$$SF = \text{[redacted]},\ for\ D_f \geq 100 \tag{8.6}$$

  Safety stock now needs to be measured. There are three ways that this can be done.

  1. **SF.** The continuous variable Safety Factor could be used. A High safety factor is expected to cause high relative waste. The question is if this is visible in an absolute KPI such as CU waste, since faster moving products (therefore having a lower safety factor) are expected to cause more waste.
  2. **High SF.** This is a binary variable, and receives a 1 for the highest safety factor of [redacted] and a 0 for a SF < [redacted]. This should make a clearer distinction between high and low safety factors.

3. **High SF x MOQ.** This is the minimum order quantity for slow moving products: an expected risk group for waste. Again it is expected to cause high relative waste, and is remain to be seen how well this is visible with an absolute waste KPI.

- FG is registered in the product inventory management. The freshness guarantee is expressed in number of days and is independent of FC and period.

8.3. Illustration of different types of waste

To clarify different possibilities of waste, Figure 8.3 shows three simplified possible examples. On the X-axis an undefined timeline, starting in 2018. The purple lines represent stock levels, the red arrows indicate waste.

The top example is one that is related to the assortment. Christmas stollen is a product that was ordered in large quantities in December. Picnic did not sell all products; some were left in the FC. They were stacked on a rolling container, and nobody consciously paid attention to it anymore. After a while someone noticed that this pallet was still standing in a corner. Picnic obviously was not going to sell it anymore, so all breads had to be cleared.

The second example is more complicated. Certain instant microwave meals show high waste numbers, especially expressed in Euros. When inspecting the data it struck me that the delivery schedule was very irregular. The supplier had a two-weekly delivery schedule. In this schedule, the longest time between two deliveries was 5 days. These microwave meals usually were delivered with a remaining shelf life of 6 days. However, Picnic offers a freshness guarantee of 3 days. This means that Picnic only has 3 days to sell the batch of products. This is shorter than the maximum time between deliveries! Therefore, all products that had not been sold, had to be cleared before the next batch was delivered. On top of that, the demand for these products was quite volatile (a CoV of around 0.6), so the products were hard to forecast. Even though it was a slow moving product, this was one of the top waste articles.

To conclude, an example from the FC. Bananas are the most frequently ordered products at Picnic (followed by milk and cucumbers). Bananas are ordered per pallet. Imagine two pallets of bananas have been standing at the unloading dock and therefore have been overlooked in the FC. A shopper might notice there are no more bananas at the banana stock location and adjusts the stock to 0. Before this is noticed in the data, POM already ordered two extra pallets of bananas, to be ordered the next day. When this shipment arrives, the other pallets are 'found,' and Picnic is stuck with a ...load of bananas. This is represented by the dotted line. The demand for bananas does not rise, so there is a big chance that (a part of) these bananas will go to waste.

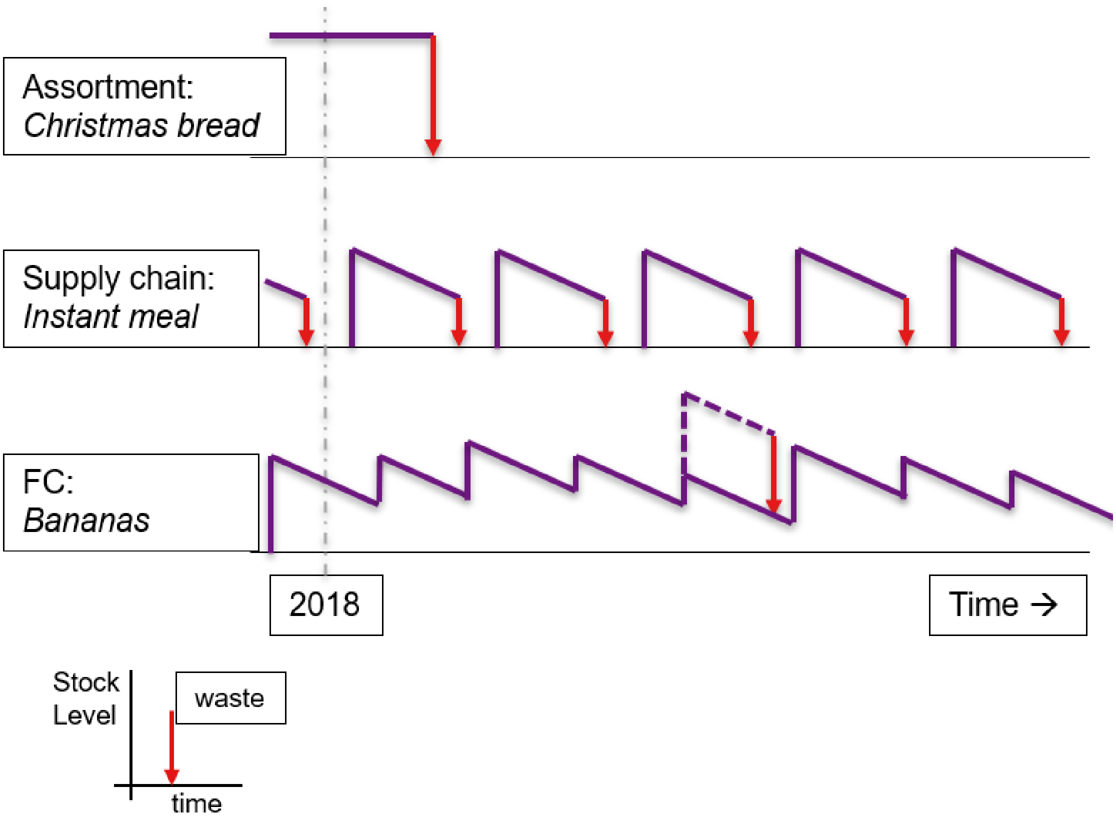


Figure 8.3: Root cause examples

8.4. Conclusion

Expected root causes of waste were identified from data exploration at Picnic, discussed and quantified where possible. The root causes were divided into three main domains: assortment, supply chain and FC processes. A fourth domain (other) was set up to cover all unidentified causes of waste.

For the assortment and supply chain related root causes, variables were constructed and a data set built. No quantified data has been found to test FC processes.

9 Modeling

In order to assess to what extent the expected root causes are indeed playing a role in the generation of waste, multivariate statistical models have been used to test hypotheses and th e factors on significance and the hypotheses. Since 80 % of the waste generated at Picnic comes from the chilled temperature zones, while these sections only contain around 23% of the SKUs, performing a regression analysis on only this temperature zone would most likely be an efficient way to gain insights in the most important root causes of waste.

> This section answers the question:
>
> *RQM1: Can the root causes be confirmed, using a statistical model?*

9.1. Input data cleaning

Before the data could be used for statistical models such as regression, the data needed to be processed, cleaned and some data points excluded when relevant.

The full data set for all available waste data concerning the chilled temperature zone consists of 47k (46881) rows. The measured KPI is CU waste and all data points have granularity FC/Article/Period. In this full data set there are some inconsistencies, such as cells with unavailable values or errors. Rows with incomplete or incorrect cells were excluded. 44k (44137) rows remained. From these rows, there were a handful of extremely high waste numbers, that were considered to be errors instead of waste. Six values with values above 1000 CU waste were removed from the data. The final data set consisted of 44131 data points, with 13 columns: 1 column for the dependent variable CU waste, and 12 for the factors.

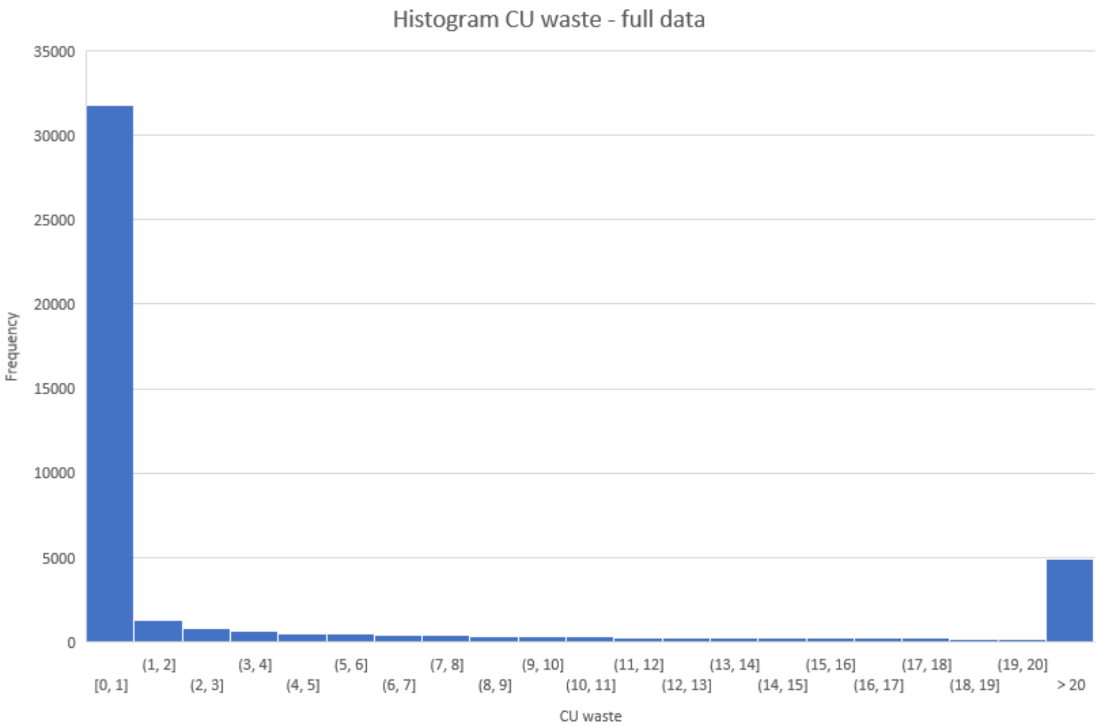


Figure 9.1: Histogram CU Waste - full data

Figure 9.1 shows the distribution of the dependent variable. This is an extremely skewed distribution, mainly because the biggest part of data consists of zeroes. On top of that, the tail of the distribution is very long. At first, we are interested in the most important effects that generate waste. Analysis on the full data gives insight in significance of root causes, distinguishing waste and zero waste.

On the other hand, waste reduction is the ultimate goal. Therefore it is interesting to see the cases that are causing the biggest waste. A second analysis focuses only on the cases where waste has been identified, excluding the zero waste data points. This gives a view of the extent to which the factors play a role in increasing waste. Excluding rows with zero waste leaves around one third of the original data set: 15177 rows. Figure 9.2 shows the distribution of the data set without zeroes. This still is a very skewed distribution, but judging by the relatively higher overflow bucket (waste > 20 CU) more influence of bigger waste numbers is to be expected.

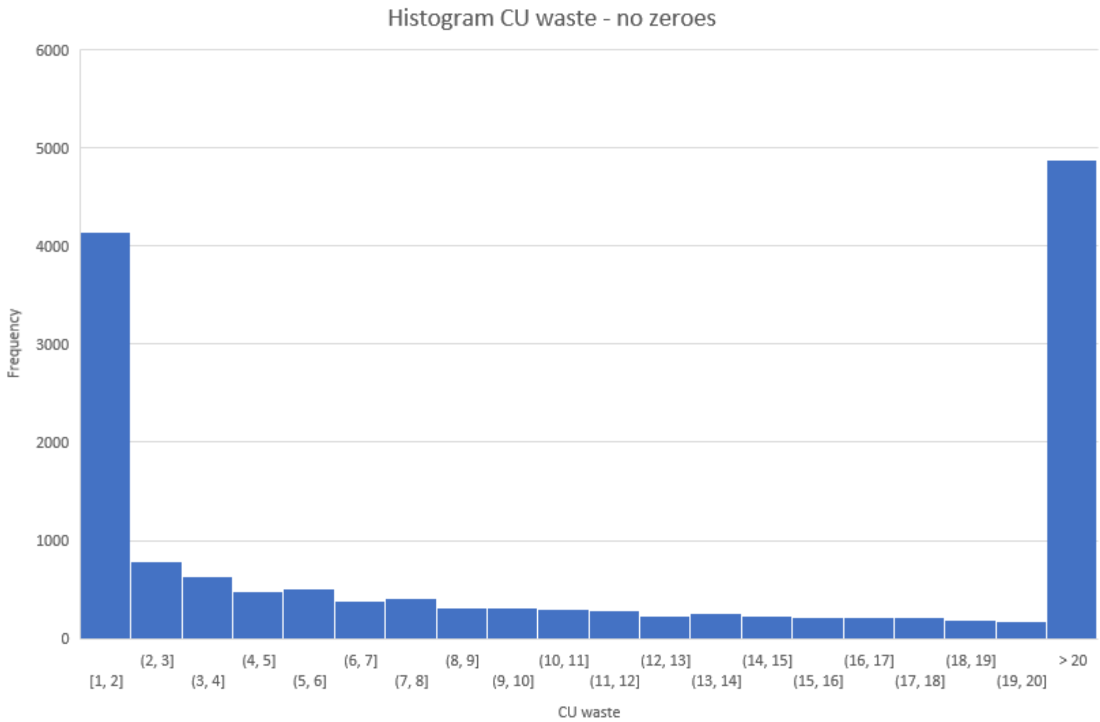


Figure 9.2: Histogram CU waste - no zeroes

9.2. Regression types

The goal of modeling the data set with statistical models is to examine if the 12 identified factors as described in 8.2 are significant contributors to waste.
Different statistical models are available to test hypotheses about data sets. Because 12 factors have been identified, a multivariate model is required. Regression models are a commonly used method. Which regression model to use depends mostly on the properties of the dependent variable and which model provides the best fit. The dependent variable CU waste can be regarded as two types of variable: a continuous variable or a count variable.

When considering CU waste a continuous variable; a multivariate data set is usually modeled with a **multiple linear regression** model [30].

CU waste can also be regarded as a count of waste instances. Count variables often follow a Poisson distribution and can be modeled by a Poisson regression. This data set does not follow an exact Poisson regression (explained in Section 9.4), but can be modeled in a similar way, as a **negative binomial regression** model.

Section 9.3 describes the multiple linear regression analysis and Section 9.4 describes the negative binomial regression analysis.

9.3. Multiple linear regression model

The model that the regression is trying to fit, is a linear model, as indicated by equation 9.1, with a dependent variable CU waste y_i, coefficients β_i (with β_0 the intercept), factors x_i and i the index for each observation.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_n x_{ni} \tag{9.1}$$

In this case, the dependent variable y is CU waste, and β_1 to β_{12} are the coefficients corresponding to factors x_1 to x_{12}.

A linear regression model shows different output parameters. The most relevant parameters are discussed.
Factor parameters:

- *Coefficients* show what relation the concerned factor has on the model. A positive value indicates that Waste rises when the factor increases. A negative value indicates waste is decreasing when the factor increases.
- The *standard error* is a dispersion coefficient. The smaller the standard error, the more accurate the factor is.
- The *t-statistic* expresses the probability that the coefficient of the parameter is not zero. It is calculated by dividing the factors coefficient by its standard error. The larger the t-statistic, the less chance of the parameter to be zero.
- The *P-value* shows the probability that the coefficient of the parameter is zero. A p-value smaller than 0.05 is considered significant and thereby rejects the null hypothesis.

Model parameters:

- R^2 is the coefficient of determination. This coefficient explains which part of the variance in the dependent variable (CU waste) can be explained by the model. It shows how well a model is able to predict. Adjusted R^2 is the R^2, adjusted for the number of predictors in the model.
- The *standard error* represents the average distance between the results and the regression line in units of the dependent variable. It represents the precision of the model.
- The *F-statistic* and '*significance F*' test the null hypothesis that all the model's regression coefficients are zero.

9.3.1. Linear regression results

The results for both the full data and the results for the data without zeroes are shown in Figure 9.1 and Figure 9.2 respectively

SUMMARY OUTPUT **CHILLED, FULL DATA**

Regression Statistics	
Multiple R	0.2619
R Square	0.0686
Adjusted R Square	0.0683
Standard Error	27.856
Observations	44131

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	2520537	210045	270.7	0
Residual	44118	34234527	776		
Total	44130	36755064			

	Std. Coefficient	*Std. SE*	*Coefficient*	*SE*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-	-	15.897	1.585	10.03	1.16E-23	12.79	19.00
Goes Out	0.031	0.0046	8.025	1.200	6.69	2.27E-11	5.67	10.38
Season	0.032	0.0047	6.100	0.900	6.77	1.26E-11	4.34	7.87
FC period	-0.089	0.0054	-0.057	0.003	-16.54	2.68E-61	-0.06	-0.05
promo	0.046	0.0049	8.229	0.868	9.48	2.56E-21	6.53	9.93
AOR	0.082	0.0071	132.212	11.420	11.58	5.92E-31	109.83	154.59
CoV AOR	0.135	0.0049	12.720	0.462	27.51	3.04E-165	11.81	13.63
T max <> del.	0.043	0.0048	1.234	0.138	8.91	5.06E-19	0.96	1.51
FG (days)	-0.052	0.0058	-0.382	0.042	-9.06	1.32E-19	-0.46	-0.30
CU/TU	-0.014	0.0081	-0.067	0.038	-1.75	8.00E-02	-0.14	0.01
Safety factor	-0.063	0.0089	-25.616	3.638	-7.04	1.94E-12	-32.75	-18.48
High SF	-0.024	0.0089	-1.542	0.572	-2.70	7.03E-03	-2.66	-0.42
High SF * CU/TU	0.017	0.0088	0.088	0.046	1.91	5.66E-02	0.00	0.18

Table 9.1: Linear regression results: full data

SUMMARY OUTPUT **CHILLED, NO ZEROES**

Regression Statistics	
Multiple R	0.2414
R Square	0.0583
Adjusted R Square	0.0575
Standard Error	44.128
Observations	15177

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	1827357	152280	78.2	2.7217E-187
Residual	15164	29528510	1947		
Total	15176	31355867			

	Std. Coefficient	*Std. SE*	*Coefficient*	*SE*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-	-	35.060	3.380	10.37	4.06E-25	28.43	41.69
Goes Out	0.003	0.0080	0.712	2.257	0.32	7.53E-01	-3.71	5.14
Season	0.072	0.0081	24.377	2.725	8.95	4.14E-19	19.04	29.72
FC period	-0.099	0.0096	-0.149	0.015	-10.25	1.39E-24	-0.18	-0.12
promo	0.096	0.0093	23.944	2.314	10.35	5.21E-25	19.41	28.48
AOR	0.019	0.0125	32.897	21.896	1.50	1.33E-01	-10.02	75.82
CoV AOR	0.050	0.0098	9.174	1.810	5.07	4.07E-07	5.63	12.72
T max <> del.	0.050	0.0083	2.344	0.388	6.03	1.63E-09	1.58	3.11
FG (days)	-0.085	0.0106	-1.154	0.144	-8.03	1.06E-15	-1.44	-0.87
CU/TU	0.039	0.0118	0.285	0.086	3.30	9.86E-04	0.12	0.45
Safety factor	-0.069	0.0149	-36.192	7.831	-4.62	3.84E-06	-51.54	-20.84
High SF	-0.064	0.0152	-5.938	1.420	-4.18	2.89E-05	-8.72	-3.16
High SF * CU/TU	0.043	0.0125	0.491	0.144	3.42	6.35E-04	0.21	0.77

Table 9.2: Linear regression results: no zeroes

9.3.2. Linear regression discussion

In general both models are significant: both the high values for F as well as well as the significance of 0 and 2.7 E -187 show that there is no chance that all the regression coefficients are zero. In short, the null hypothesis is rejected for both models.

However, the model is not an accurate predictor of the exact amount of waste. An Adjusted R Square value of 0.068 for the full data and 0.058 for the data without zeroes shows that only 6.8 % and 5.8 % respectively of the variation of the dependent variable can be explained by this combination of independent variables. Also the standard errors are quite large, especially considering this skewed distribution, in which most values are 0 or 1. Because the model is not expected to have high predictive power, no regression equation is given.

Looking at the significance of individual factors, we see some extremely significant values. This means that it can be said with confidence that the identified factors have some sort of relation with the amount of waste.
In Table 9.3 the coefficients, t-Stat and P-value for both models are summarized. The following sections discuss the hypotheses and meaning of factors, classified per domain. Only the direction of coefficients (positive or negative) is discussed, since the models have only low predictive power. The t-statistic and P-value can be discussed in more detail, since they indicate the importance of the factor in the model rather than the quantity.

		Linear regressions comparison					
		FULL DATA			NO ZEROES: WASTE ONLY		
	Factor	*Coefficients*	*t Stat*	*P-value*	*Coefficients*	*t Stat*	*P-value*
	Intercept	15.9	10.0	1.16E-23	35.1	10.4	4.06E-25
Assortment	Goes Out	8.0	6.7	2.27E-11	0.7	0.3	7.53E-01
	Season	6.1	6.8	1.26E-11	24.4	8.9	4.14E-19
	FC period	-0.1	-16.5	2.68E-61	-0.1	-10.3	1.39E-24
	promo	8.2	9.5	2.56E-21	23.9	10.3	5.21E-25
Supply Chain	AOR	132.2	11.6	5.92E-31	32.9	1.5	1.33E-01
	CoV AOR	12.7	27.5	3.04E-165	9.2	5.1	4.07E-07
	t max <> deliveries	1.2	8.9	5.06E-19	2.3	6.0	1.63E-09
	FG (days)	-0.4	-9.1	1.32E-19	-1.2	-8.0	1.06E-15
	CU/TU	-0.1	-1.8	8.00E-02	0.3	3.3	9.86E-04
	Safety factor	-25.6	-7.0	1.94E-12	-36.2	-4.6	3.84E-06
	High SF	-1.5	-2.7	7.03E-03	-5.9	-4.2	2.89E-05
	High SF * CU/TU	0.1	1.9	5.66E-02	0.5	3.4	6.35E-04

Table 9.3: Linear regression comparison

Assortment

Out of assortment - goes out
The coefficient of the variable 'goes out' matches the hypothesis: when products are taken out of the assortment, chances of generating waste rise. However, the coefficient is not significant for the model without zeroes. This means that the factor does play a significant role in differentiating between waste or no waste. If waste is given, it does not play a significant role any longer.

Seasonal - Season
The positive coefficients of both regressions indicate that the hypothesis of seasonal products causing waste can be verified. Both models show positive coefficients and high t-stats and significance.

Freshness - FC period
Freshness, with its indicator FC period, behaves as expected: a longer FC period, so more shelf life, causes less waste. Both coefficients are in the top of the significant factors. For both models it has the second highest significance.

Promotion - promo
The hypothesis of promotions causing waste can also be validated. A positive coefficient shows that the higher the promo pressure, so the more products are sold in a promotion, the more waste is to be expected. In the model with only waste it has the highest significance.

Supply Chain

Demand quantity - AOR
The article order rate plays different roles in the two models. In the full data, the coefficient is very significant, while in the model with no zeroes the significance is only 0.13. This means it is not significant by the $\alpha = 0.05$ standard, but still quite significant. In both models AOR has a positive relation with absolute waste. It can be interesting to test the AOR against relative waste, since this will give a better view of the impact of the waste.

Demand volatility - CoV AOR
This is the variable that was expected to be strongly related to waste. The volatility in demand is having the biggest effect on waste in the full data model: the higher the coefficient of variation, the higher the waste. T stat is the highest of all and therefore it confirms the hypothesis. In the no zeroes model it is by far not the most significant parameter anymore, although still very significant.

Order frequency - Maximum time between deliveries
A low order frequency, indicated by a higher maximum time between deliveries causes waste. This is consistent with the hypothesis. It might have an interaction effect with other variables, such as FC period, but this has not been tested in these models.

FG days
The negative coefficient of the freshness guarantee demonstrates that a the longer the guarantee, the less waste. This is counter intuitive. It can be explained by the fact that Picnic is setting their freshness guarantees too conservatively.

Minimum order quantity - CU/TU
Minimum order quantity, expressed as CU/TU has a reasonably low t-stat in both models. In the full data model the value is 0.08, so not significant by the $\alpha = 0.05$ assumption, but pretty close. Interestingly enough in the full model the coefficient is negative, indicating that a higher minimum order quantity causes less waste. In the only waste model, a positive coefficient shows that given waste, a higher minimum order quantity does increase waste. This effect can be explained by non-linearity in the data: there are many articles with a high minimum order quantity that never caused waste, but for the products that are sensitive to waste, the minimum order quantity does increase waste.

Safety stock - SF, high SF, high SF x CU/TU
The hypothesis is that a higher safety stock causes more waste. In Figure **??** safety factor has a negative coefficient and does not concur with the hypothesis about safety stock. This looks like the higher a safety factor, the more waste is generated. This is not necessarily true. A

high safety factor inherently means an article that has not very high demand and is therefore inversely correlated with AOR. Since Demand (AOR) has a positive coefficient with waste, this effect is explained.
A high SF (categorical variable), was composed to presumably cause a more extreme representation. It does not make a difference for the coefficients. The absolute KPI CU waste might not be the best indicator to reveal waste for slow movers.
On the other hand a CU/TU given a high safety factor, the variable that was created to identify waste due to safety stock in slow movers, does show a positive trend. Both factors can be considered (almost) significant.

9.3.3. Strongest factors

Table 9.4 shows the factors, ranked by absolute t-stat. The results show us that these two models are in fact quite different processes.

Different models were to be expected, since there are so many zeroes in the full data. For all data points in which the dependent variable are zeroes, the 12 variables do have values; they just don't attribute to waste in that period. This way, the model with the full data likely distinguishes between waste and no waste, whereas the model with only waste distinguishes between a lot and a little waste, since waste is a given.

While having the same input variables, that are mostly significant in both models, the order of significant factors varies quite a bit.
The strongest factor in the full data model is the coefficient of variation of the article order rate. Strangely enough in the waste only model it comes in at the sixth place. This indicates that a high CoV AOR does increase the probability of a product going to waste, but when waste is given, does not influence the amount by as much. Promotion takes the first place in the no zeroes model. It is likely to think that when promotions cause waste, they cause higher numbers of waste. This can explain the first place in the no zeroes model. FC period is very important in both models: fresher products cause less waste. In the full data model there is an interesting mix of supply chain and assortment related factors in the top. In the waste only model, the top three factors are assortment related causes.

	Linear regression									
	FULL DATA					NO ZEROES: WASTE ONLY				
Rank	*Factor*	*Coefficients*	*t Stat*	*P-value*	*Domain*	*Factor*	*Coefficients*	*t Stat*	*P-value*	*Domain*
1	CoV AOR	12.7	27.5	3.04E-165	SC	promo	23.9	10.3	5.21E-25	Asst
2	FC period	-0.1	-16.5	2.68E-61	Asst	FC period	-0.1	-10.3	1.39E-24	Asst
3	AOR	132.2	11.6	5.92E-31	SC	Season	24.4	8.9	4.14E-19	Asst
4	promo	8.2	9.5	2.56E-21	Asst	FG (days)	-1.2	-8.0	1.06E-15	SC
5	FG (days)	-0.4	-9.1	1.32E-19	SC	t max <> deliveries	2.3	6.0	1.63E-09	SC
6	t max <> deliveries	1.2	8.9	5.06E-19	SC	CoV AOR	9.2	5.1	4.07E-07	SC
7	Safety factor	-25.6	-7.0	1.94E-12	SC	Safety factor	-36.2	-4.6	3.84E-06	SC
8	Season	6.1	6.8	1.26E-11	Asst	High SF	-5.9	-4.2	2.89E-05	SC
9	Goes Out	8.0	6.7	2.27E-11	Asst	High SF * CU/TU	0.5	3.4	6.35E-04	SC
10	High SF	-1.5	-2.7	7.03E-03	SC	CU/TU	0.3	3.3	9.86E-04	SC
11	High SF * CU/TU	0.1	1.9	5.66E-02	SC	AOR	32.9	1.5	1.33E-01	SC
12	CU/TU	-0.1	-1.8	8.00E-02	SC	Goes Out	0.7	0.3	7.53E-01	Asst

Table 9.4: Factors with strongest significance in linear regression

9.3.4. Linear Regression assumptions

A linear regression model makes several assumptions. These are checked to validate the model. Table 9.5 shows which assumptions are made for a linear regression and if they were met by this model. Explanation for each assumption follows below.

	Regression assumptions	Justified? Full data	Justified? No zeroes
1	Additivity and linearity	×	×
2	Independent errors	✓	✓
3	Homoscedasticity	×	×
4	Normally distributed errors	✓	✓
5	Variable types	✓	✓
6	No multicollinearity	✓	✓
7	Non-zero variance	✓	✓

Table 9.5: Linear regression assumptions check

1. Additivity and linearity:
   The relation between the dependent variable and any continuous independent variable is assumed to be linear. The relations of all non-binary variables are shown in Figure 9.3 for the full data set and 9.4 for the data set without zeroes. Also some discrete variables are presented, to give an idea of the relation. As one can see, most variables cannot necessarily be called linearly related to the dependent variable. The sheer amount of data points makes some graphs hard to read and stochastic behavior in the data is expected to cause the non-linearity.

2. Independent errors:
   The residual terms of the model should be uncorrelated. This can be tested with a Durbin-Watson test. This test is defined as the sum of squared difference of residuals divided by the sum of squared residuals:

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})}{\sum_{i=1}^{n}(e_i^n)} \tag{9.2}$$

   This value can vary between 0 and 4. A value of 2 means that residuals are uncorrelated (Durbin Watson, 1953) [31]. The calculated Durbin Watson coefficient for the full data is 1.830 and for the no zeroes data 1.853. This assumption can be validated.

3. Homoscedasticity
   The residuals at each level of the predicted value should have the same variance. Figure 9.5 and 9.6 show the standardized residuals versus the predicted values. A clear slightly tilted bottom can be seen (0 waste and 1 waste, respectively) with a non parallel cloud of data above it. The data is not homoscedastic.
4. Normally distributed errors:
   This assumption is related to the distribution of the residuals. They are assumed to be normally distributed with a mean of 0. Figure 9.7 and 9.8 shows the histogram of standardized residuals. Both distributions seem to be normally distributed.
5. Variable types:
   All independent variables must be of the scale or binary type. This is confirmed.

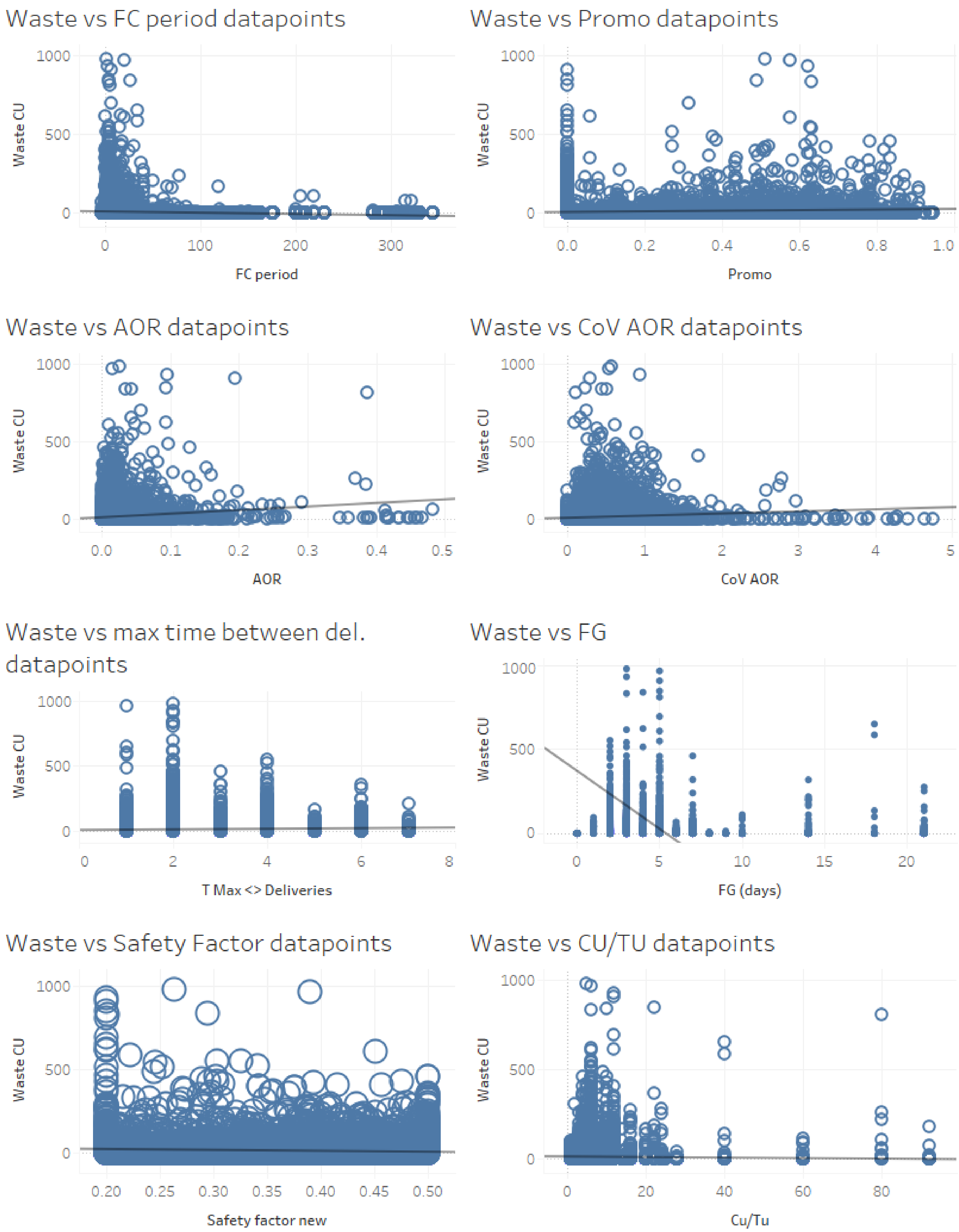


Figure 9.3: Relation plots of CU waste vs IVs, for full data set

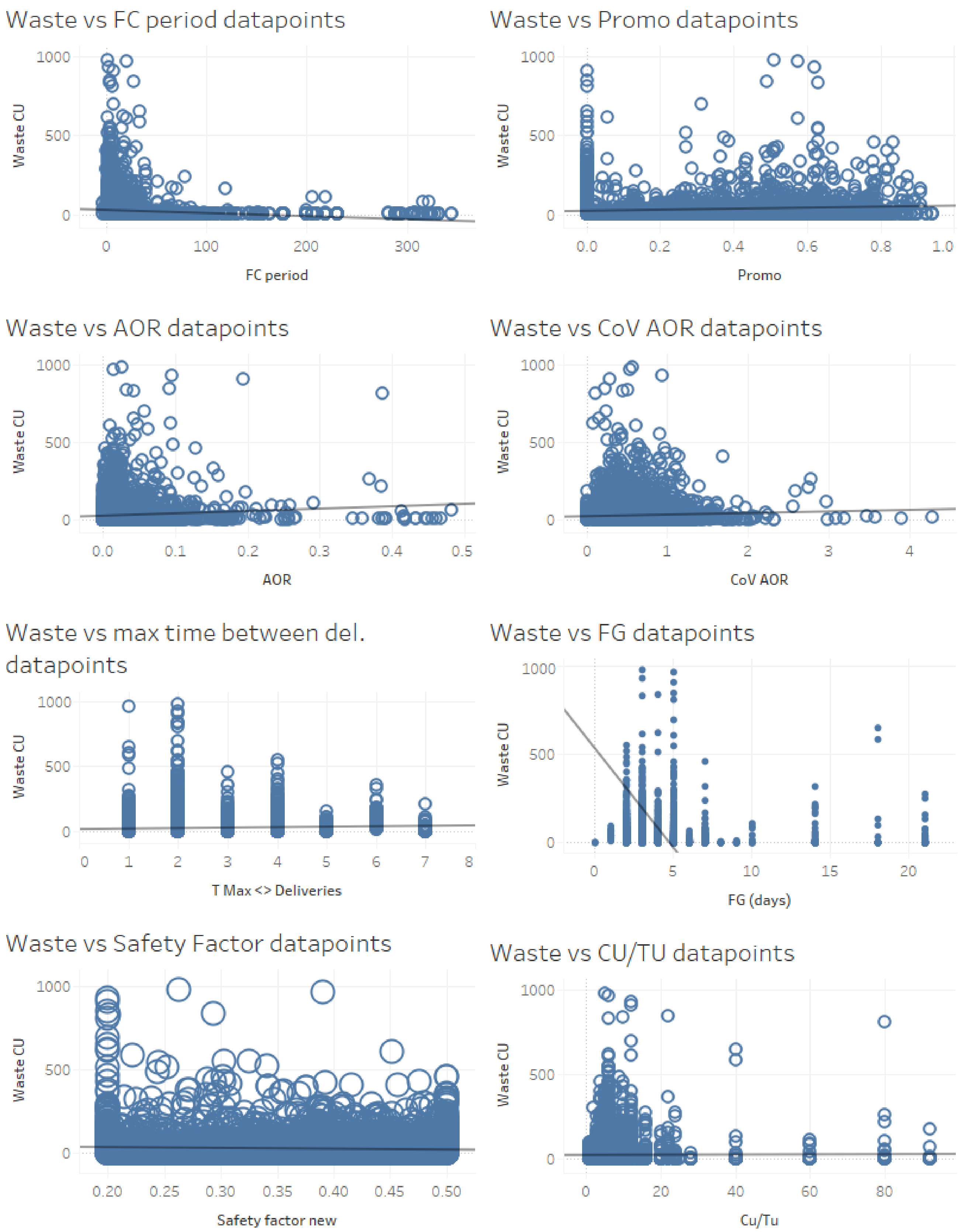


Figure 9.4: Relation plots of CU waste vs IV, for no zeroes data

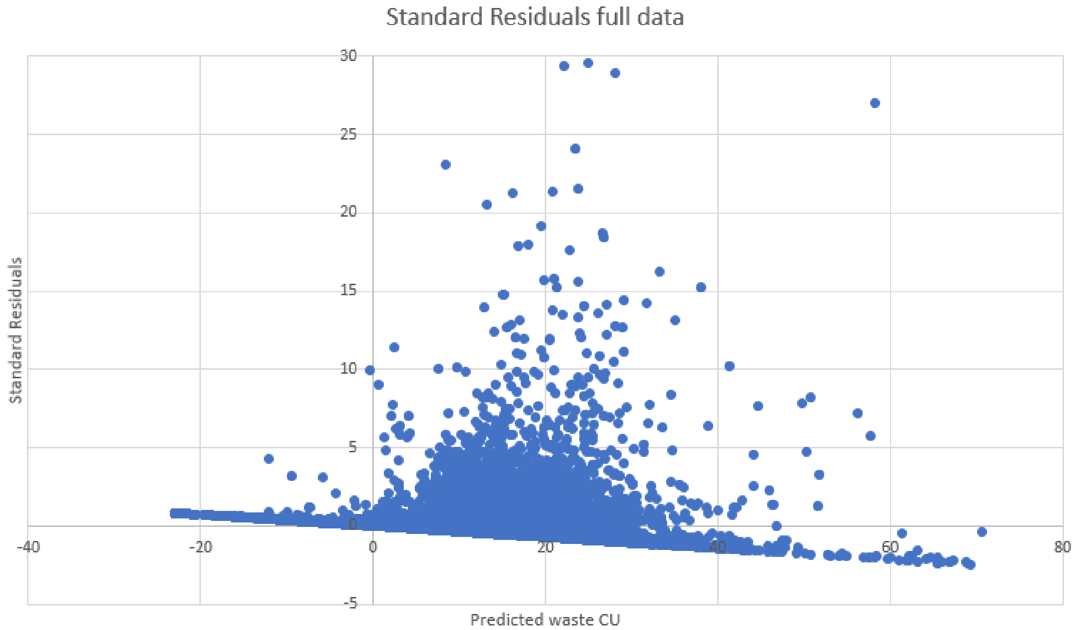


Figure 9.5: Standard residuals plot full data

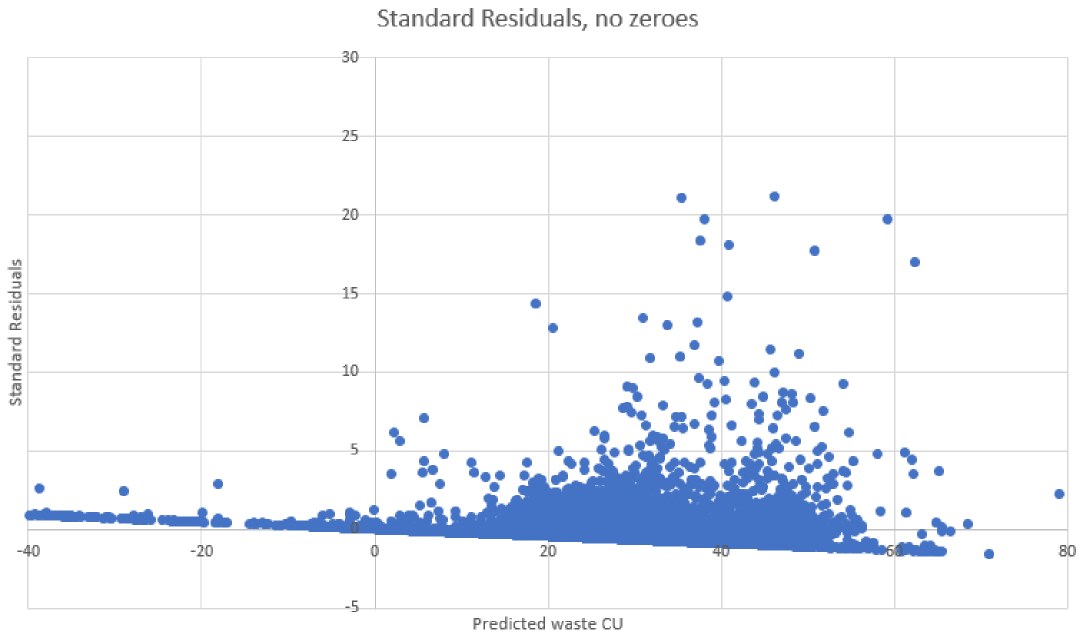


Figure 9.6: Standard residuals plot no zeroes

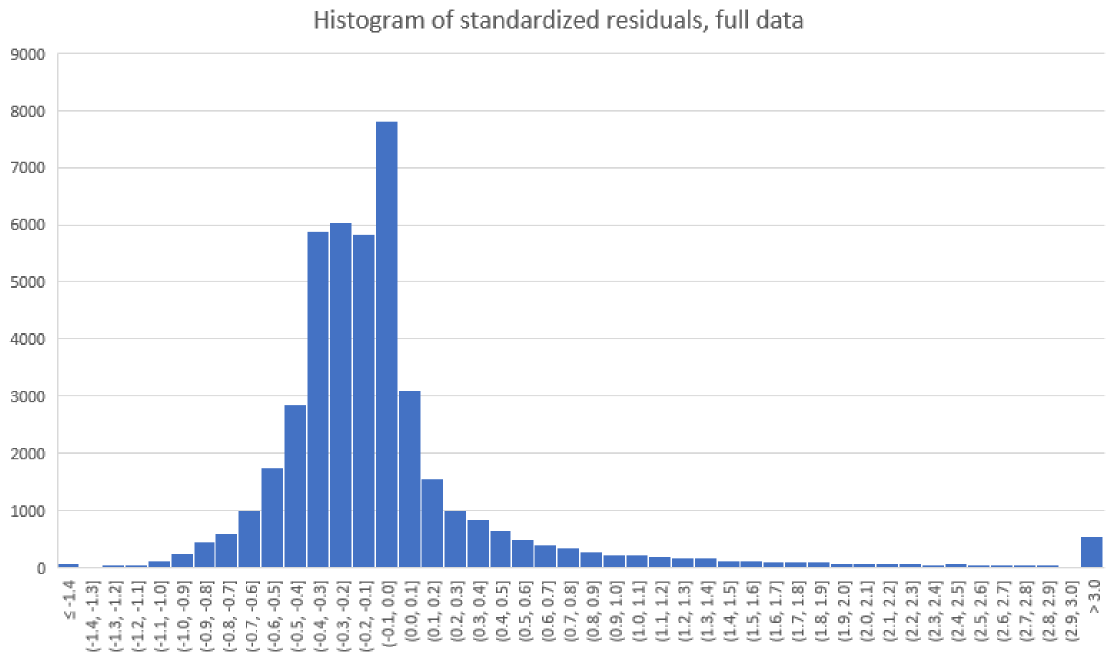


Figure 9.7: Histogram of standard residuals full data

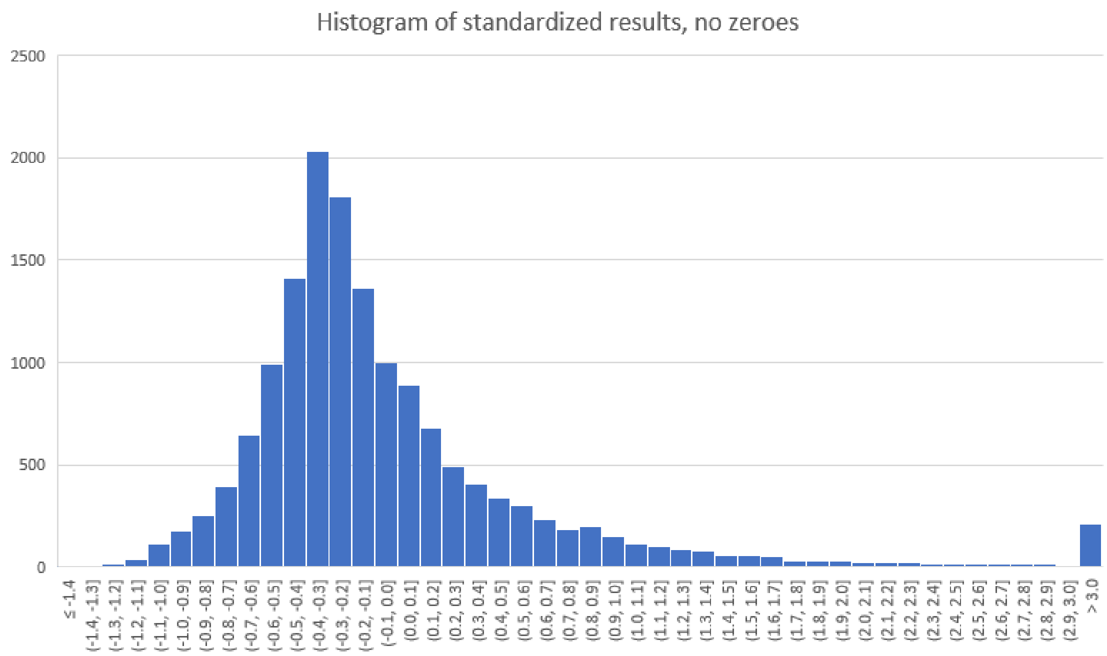


Figure 9.8: Histogram of standard residuals no zeroes

6. Multicollinearity:
   Different independent variables cannot be correlated. A correlation factor of 0.8 or higher assumes correlated factors. This causes the model to be influenced by dependency between variables. Tables and shows that no relations are bigger than 0.8. Hereby this assumption is validated.
7. Non-zero variance:
   The predictors should have some variation in value: they cannot all be zero. This can be confirmed.

CORRELATION ANALYSIS FULL DATA

	Waste CU	FC period	T max <> del.	Goes Out	Season	promo	AOR	CoV AOR	Safety factor	High SF	CU/TU	High SF * CU/TU	FG (days)
Waste CU	1												
FC period	-0.12	1											
T max <> del.	0.08	-0.21	1										
Goes Out	0.05	-0.03	0.01	1									
Season	0.03	-0.07	0.01	0.10	1								
promo	0.11	0.01	0.11	-0.02	-0.05	1							
AOR	0.14	-0.04	-0.02	0.02	-0.05	0.10	1						
CoV AOR	0.15	0.06	0.04	0.05	-0.06	0.29	0.05	1					
Safety factor	-0.13	0.03	0.03	0.01	0.06	-0.14	-0.71	0.02	1				
High SF	-0.11	0.02	0.06	0.01	0.07	-0.15	-0.49	-0.08	0.76	1			
CU/TU	-0.03	0.32	-0.13	-0.01	-0.07	0.00	0.28	-0.02	-0.17	-0.13	1		
High SF * CU/TU	-0.10	0.24	-0.04	-0.02	-0.02	-0.09	-0.26	-0.05	0.42	0.56	0.55	1	
FG (days)	-0.11	0.49	-0.25	-0.04	-0.09	-0.02	-0.01	0.05	-0.01	-0.04	0.45	0.30	1

Table 9.6: Correlations full data

CORRELATION ANALYSIS, NO ZEROES

	Waste CU	FC period	T max <> del.	Goes Out	Season	promo	AOR	CoV AOR	Safety factor	High SF	CU/TU	High SF * CU/TU	FG (days)
Waste CU	1												
FC period	-0.13	1											
T max <> del.	0.08	-0.18	1										
Goes Out	0.01	-0.02	-0.01	1									
Season	0.07	-0.05	-0.02	0.18	1								
promo	0.13	0.04	0.09	-0.04	-0.05	1							
AOR	0.09	0.00	-0.06	-0.02	-0.03	0.08	1						
CoV AOR	0.06	-0.01	0.09	0.01	0.06	0.40	-0.24	1					
Safety factor	-0.09	-0.06	0.12	0.05	0.05	-0.14	-0.67	0.36	1				
High SF	-0.07	-0.08	0.16	0.05	0.05	-0.17	-0.42	0.31	0.74	1			
CU/TU	0.01	0.32	-0.14	-0.01	-0.04	0.01	0.47	-0.12	-0.29	-0.24	1		
High SF * CU/TU	-0.05	0.11	0.07	0.02	0.00	-0.10	-0.27	0.20	0.49	0.66	0.20	1	
FG (days)	-0.12	0.56	-0.24	-0.04	-0.06	-0.01	0.03	-0.03	-0.08	-0.12	0.46	0.17	1

Table 9.7: Correlations no zeroes

9.4. Negative binomial regression model

As mentioned in Section 9.2, the dependent variable can also be regarded as a count: CU waste can be seen as a number of times that an event of waste has taken place for an article, in an FC in a period. There are no negative values, and there is no natural upper bound. This count data is typically modeled with a Poisson regression.

The Poisson regression has 5 assumptions:

1. Dependent variable consists of count data
2. The independent variables are continuous, ordinal or nominal variables
3. The observations should be independent.
4. The distribution of counts follows a Poisson distribution
5. The mean and the variance of the model are identical

The fifth assumption forms a problem: the mean and the variance are not equal in both models as can be seen in Tables 9.8 and 9.9. The mean in both data sources is a lot smaller than the standard deviation, which is the square root of the variance. The variability is greater than can be explained by the assumed statistical model. This is called an overdispersed model and cannot give the desired output, when modeled with a Poisson regression.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
WasteCU	44131	0	980	8.01	28.860
Valid N (listwise)	44131				

Table 9.8: Descriptives of CU waste - full data set

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
WasteCU	15177	1	980	23.29	45.455
Valid N (listwise)	15177				

Table 9.9: Descriptives of CU waste - data set with no zeroes

There is an alternative: **negative binomial regression**. Negative binomial regression is a generalization of Poisson regression which loosens the restrictive assumption that the variance is equal to the mean, as stated by the Poisson model [32],[33]. In this type of regression, the dependent variable follows a negative binomial distribution. For this case, assumptions 1, 2 and 3 still stand, but 4 and 5 are no longer required. Since we have just explained that the dependent variable can be seen as count data, the independent variables are not changing and we have already shown that the observations are independent, we can validate the assumptions and are safe to proceed with this model.

The equation the regression model is trying to fit, is a generalized form of the linear model, shown by Equation 9.3. It takes the (natural) logarithm of the dependent variable:

$$ln(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_n x_{ni} \tag{9.3}$$

This implies:

$$y_i = e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+...+\beta_n x_{ni}} \tag{9.4}$$

The output of the negative binomial model is different than the output of the linear regression model. The most important factors that the negative binomial regression model displays are:

- *Likelihood Ratio Chi-Square* is used to test if the variables are dependent. A low value indicates that the results may be misleading.
- *df* is the degrees of freedom of the model. Both model have 12 degrees of freedom (13 columns, -1).
- Sig: stands for *significance*, similar to the linear regression model. In general a value of lower than $\alpha = 0.05$ is considered significant.
- B stands for the *coefficient* of a factor. B is often also expressed as exp(B), since it is a logarithmic model. B can be best compared to the coefficient of the linear regression model.
- *Wald Chi-Square* determines the extent of significance. It can be compared to the t-stat of linear regression.

9.4.1. Negative binomial regression results and discussion

Regression analyses have been run with the same data sets as the linear regression model: one with the full data set and one with the data set without zeroes. Full regression results can be found in appendix F. The significance of the models are shown in Tables 9.10 and 9.11. These models are very significant, with a value of 0.000, indicating that the null hypothesis for the full models is rejected. Table 9.12 shows the factors, ranked on their significance.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
7391.434	12	.000

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG [a]

a. Compares the fitted model against the intercept-only model.

Table 9.10: Omnibus test of negative binomial regression with full data

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
2294.524	12	.000

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG [a]

a. Compares the fitted model against the intercept-only model.

Table 9.11: Omnibus test of negative binomial regression with data with no zeroes

	Negative binomial regression									
	FULL DATA					WASTE ONLY: NO ZEROES				
rank	*Factor*	*B*	*Wald Chi-Square*	*Sig.*	*Domain*	*Factor*	*B*	*Wald Chi-Square*	*Sig.*	*Domain*
1	CoV_AOR	4.66	3989.8	0.000	SC	FC_period	-0.01	556.2	0.000	Asst
2	FC_period	-0.01	1000.9	0.000	Asst	FG	-0.06	214.3	0.000	SC
3	T_max_del	0.23	215.0	0.000	SC	Promo	0.85	174.3	0.000	Asst
4	FG	-0.06	179.5	0.000	SC	T_max_del	0.12	117.3	0.000	SC
5	Promo	-1.13	155.2	0.000	Asst	Season	0.76	105.1	0.000	Asst
6	High_SF	-0.59	112.6	0.000	SC	SF	-1.43	41.2	0.000	SC
7	AOR	17.91	74.5	0.000	SC	CU_TU	0.02	38.1	0.000	SC
8	Goes_out	1.03	69.1	0.000	Asst	CoV_AOR	0.28	30.3	0.000	SC
9	SF	-3.49	61.9	0.000	SC	High_SF	-0.19	21.2	0.000	SC
10	CU_TU	0.02	15.3	0.000	SC	High_SFxCU_TU	0.02	12.2	0.000	SC
11	High_SFxCU_TU	-0.01	4.7	0.029	SC	AOR	0.13	0.0	0.842	SC
12	Season	0.10	1.2	0.269	Asst	Goes_out	0.76	0.0	0.865	Asst

Table 9.12: Factors with strongest significance in negative binomial regression

Inspecting the results of the negative binomial regressions, we can see that almost all factors are significant. This confirms that correct root causes have been identified, and that it is worth the effort to improve statistical methods to describe root causes of waste.

Comparing the full data regression and the no zeroes regression, the directions of the B coefficients are mostly the same, except for promotion and High SF x CU/TU. Both have negative coefficients in the full data analysis and positive coefficients in the no zeroes analysis. It can indicate that there is still some non-linearity for this case. Promo was expected to have a positive coefficient, which can be seen in the no zeroes regression. I do not have an explanation for the coefficients of high SF x CU/TU, apart from that it might only have a negative coefficient in this model, together with the other factors. The unexpected coefficient might actually be caused more by the other parameters of the model, than a one on one relation with the dependent variable.

9.5. Comparison of regression models

Comparing the results of the linear and negative binomial regressions there are some striking resemblances, such as the coefficient of variation of the article order rate being the strongest factor in the full data, and decreasing in importance in the no zeroes data regression. The freshness variable 'FC period' is an important factor in all results. Order frequency also plays an important role. Significance wise, it is interesting to note that the same two variables (AOR and Goes out) are insignificant in both no zeroes regressions.

There are also some differences. Mostly the order of most important factors, and differences in significant factors. The factor season is not significant in the full data negative binomial regression. It is significant in the linear regression, where all parameters are (almost) significant.

The two models, being of different regression types, are not commonly compared on goodness-of-fit. An indicator can be the sum of squared residuals. Table 9.13 shows a comparison of the sum of the squared residuals, for the two regression models and the two data sets. The sum of squared residuals (SSR) for the two models using the no zeroes data set are very similar. The results for the full data set however are extremely different. Where the linear regression models SSR are somewhat higher for the full data than the no zeroes data, the negative binomial regression model gets an extremely high value for the full data regression. This indicates that the negative binomial is slightly better at predicting waste for the data set with only waste, but is much worse at predicting the full data set. It might be explained by the fact that the negative binomial regression model cannot predict 0 waste, since the predicted value $y_i = e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+...+\beta_n x_{ni}}$ can never be 0. For the most part, the full data set exists of zeroes.

	Sum of squared residuals	
	Linear	Negative binomial
Full data	34.2 M	1.74E+19
No zeroes	29.5 M	29.3 M

Table 9.13: Sum of squared residuals compared

9.6. Conclusion

The idea of using regression models was to test if the identified expected factors for generating waste were indeed root causes. Two data sets had been created: one with the full data and one with only waste. Two regression types have modeled these two data sets.

The linear regressions have been an accessible tool to find relations between the different factors and the dependent variable: CU waste. The linear regression models showed highly significant results, meaning that the expected factors indeed are strongly related to waste. Not all linear regression assumptions were validated. Especially the skewed distribution of the dependent variable, and its non-linear relations with the independent variables ensured that this model could not be trusted to be fully accurate and therefore be considered a valid model.

The negative binomial regression model was used, with CU waste regarded as a count variable. All negative binomial regression assumptions could be validated. Similar to linear regression, factors had the same type of effect and highly significant results were obtained.

The high significance of factors in all models and the coefficients of parameters indicate that the 12 identified factors are indeed root causes. Also it was shown that the two models (full data versus no zeroes) were not exactly the same processes. Different factors were significant and a different order in the most important parameters was present.

Concluding this section, we've learnt four things about modeling waste:

1. The identified factors are significantly related with waste.
2. It is hard to predict waste with regression analyses, because of low goodness-of-fit of the regression models.
3. The linear regression model gave good insights in the main goal: identifying the root causes, even though not all regression assumptions could be validated.
4. The negative binomial regression model showed similar results and thereby confirmed the results of the linear regression. Because all assumptions could be validated, this is a valid model.
5. It can be beneficial to improve modeling to further approach and indicate waste with more complex models.

10
Implementation

Since waste reduction is the ultimate goal for Picnic, this chapter is devoted to the practical implementation and further development of waste reduction.

10.1. Implementation at Picnic

The case study at Picnic mostly was an interaction between exploring, hypothesizing, testing and using gained insights to take actions. Exploring consisted mostly of talking to Picnic employees and collecting and inspecting data. Hypotheses rose from this, that were tested with statistical models. When the hypotheses were confirmed and categorized, the correct actions could be taken as a logical next step for specific cases of waste.

This chapter answers the question:
RQM2: Can the regression model provide actionable insights to reduce waste?

10.1.1. Good results at Picnic by continuously improving

The continuous improvement of identifying a problem and tackling it has been an important part of the research process. In the most cases the costs of waste or the relative waste of all products in all FCs were analyzed and sorted. For the top bleeders; the products with the highest absolute or relative waste, a deep dive into the data was done. This was done in parallel with the statistical analyses. Specific actions would come forward from communicating the findings in the data as well as the statistical models to the responsible people. An example of these actions is: negotiating with suppliers to have an extra delivery moment or smaller minimum order quantity. An outcome can also be to decide to stop having a product in the assortment, or decrease the safety factor for unpredictable items. As a consequence Picnic chooses to accept a higher risk of unavailability.

As mentioned in Section 7.5, the average waste in cts per sold item was [redacted]. I am proud to say that the average waste of the last 8 weeks has decreased to an average of [redacted] cts/item, with the lowest values dipping below 1 cts/item. A comparison between this decrease in cts/item and a scenario in which the cts/item theoretically remained the same is shown in Table 10.1, assuming a linear decrease of waste cts/item over the last four periods.

[redacted]

Table 10.1: Rough calculation difference in waste expenses for Picnic

Sections 10.1.2 and 10.1.3 give an overview of the actions or recommendations, related to the identified and validated root causes. Some of the actions have already been used in practice.

10.1.2. Actions & recommendations: assortment

Out of assortment - goes out
When products are (planned to be) taken out of assortment, adjust ordering. Currently most products are ordered automatically, until they are taken out. An option can be to set a negative safety factor towards the end of the products life at Picnic. This way Picnic can sell all the products before a product becomes unorderable in the app. Currently this action is being done at some of the Christmas/new years products

Seasonal - Season
Seasonal products are only a very small part of the total waste, but can involve high numbers of waste. Seasonal products should be considered on a product level. Analyzing and learning from previous years on a product level is probably the most useful method. Seasonal products are at some point taken out of the assortment by definition, so in the type of waste there will be overlap with out of assortment waste and the previously mentioned actions should also be considered.
Furthermore; BBQ seasonal products might be coupled to a weather forecast. This can already be analyzed with historical waste and weather data.

Freshness - FC period
To receive products with the highest possible freshness is never a bad thing. It is always good to negotiate with suppliers how to get the best possible freshness on products. Packaging can also play a role. For example, meat can be vacuum packed, which can improve shelf life.

Promotion - promo
Promotions in some cases had an opposite effect than expected. Promotions do play a signif-

icant role in the process of waste generation, but need further research to confirm the exact type of effect. When an extreme positive coefficient was present for all cases, it would show that promotions are always over ordered and causing waste. This result tells Picnic that they do a good job on forecasting promotions.

10.1.3. Actions & recommendations: supply chain

Demand quantity - AOR
A high AOR is causing waste, however not always in a significant way. This needs to be researched with a relative waste parameter. The absolute value of waste compared to an absolute demand quantity turned out to be not a very expressive variable.

Demand volatility - CoV AOR
Products that have a highly volatile demand, are more likely to cause waste. It was the strongest variable in the full data regressions. It is not wise to order an unpredictable product by a forecast. An assessment will have to be made for products with a high coefficient of variation based on the trade-off between waste and unavailability. These unpredictable products are likely to have more of both, but ordering in a different way can influence this. For example, unpredictable products with a low safety factor (and therefore AOR) can get an unpredictability correction on its order quantity or safety factor.

Order frequency - Maximum time between deliveries
For freshness reasons, it is always favorable to have the highest possible order frequency. Although for some products this is not necessary and it would not make sense to drive a supplier deliver toilet paper twice a day. A balance needs to be found for the products for which the order frequency does influence waste.

Minimum order quantity - CU/TU
The minimum order quantity was also a significant factor that contributes to waste. Here also a balance needs to be found for a reasonable way of packing CUs in a TU. From a waste perspective, preferably with the least CUS. For some slow moving meats, new agreements have been formed with the supplier to supply less CU/TU.

Safety stock - SF, high SF, high SF x CU/TU
The safety stock was a hard one to decipher. Currently the purchase order management is working on adjusting the way of formulating the safety factor, so I would advise to await this and see what effect this has. As previously mentioned, seasonality of products and volatility of demand should be taken into account.

FG days
The freshness is an important parameter for Picnic. Picnic wants to be able to guarantee the most freshness, but increasing the freshness guarantee will also increase waste. This can also be an optimization study.

10.1.4. Waste dashboards

There was a demand from Picnic for an (automated) process, keeping track of waste and showing insights in waste performance. For this reason, I have made 4 dashboards, using tableau software. These dashboards are refreshed weekly. For now data has to be gathered, joined and processed manually, but a specific data MART in the data warehouse is being engineered to automate this process. This waste MART will contain the parameters that I identified and quantified.

The most important dashboard is shown in Figure 10.1. It shows an overview of the waste data of the last 8 weeks. In the top left corner, the absolute waste in Euros is shown. Each FC has its own color. We can see that in general the waste drops, especially when we look at FC0 to FC3. FC4 was opened in week 42 and causes relatively a lot of waste. This can be seen in the line graph at the bottom left. This is the waste expressed in cts per item. FC1 is considered the role model FC for Picnic and generally performs the best on waste. FC2 and FC3 are close to each other and also perform well. FC3 is quite young, but rapidly increasing waste performance and FC4 just started, which we knew generates more waste, but also sees rapid decline.

In the top right corner, the order groups are shown, sorted on the ones generating the most waste. In this case it is Nijland, a supplier of chicken products. The table shows the amount of Euros waste per week, per FC. Totals are shown for the order groups, weeks and FCs.

The bottom center and right tables show the top waste articles. The center table is sorted on margin impact, so the products that are the least profitable because of their waste. The right table is sorted on cost of waste and shows the biggest waste items of the last 8 weeks. In the order group table, the viewer can click different aspects, for example Order group Nijland, FC3. All figures and tables will filter by this data and only show information about Nijland FC3 for the last 8 weeks. This ensures that the users of this dashboard can easily find problematic waste causes.

Every week someone from purchase order management inspects these dashboards. The top waste items are given a closer look. Because of the identified root causes by the regression model, she (or he) can have a better understanding of the factors that could have influenced this waste. When the reasons are found, this will be communicated to the responsible people, mostly category management for assortment issues, POM and/or category management for SC issues, and FC leads/captains for FC related problems.

The other three dashboards are found in Appendix D and display:

- DB0: all KPIs (cost of waste, waste EU/net sales, margin impact, relative waste qty (also split per broken and quality issues and waste cts/item) per FC and total for the last 8 weeks.
- DB2: a breakdown of waste product for quality and breakage reasons. This dashboard can be filtered on FC, Article and Purchase order group.
- DB3: shows the top broken and quality items for last week

10.1.5. Waste was labeled with MECE waste tree

An approach that was tested around halfway through the research was a preliminary labeling method to label all waste data points with their root cause. This way a distribution of waste could be obtained. It showed the percentages of waste per proposed root cause. This was done in a 'mutually exclusive and collectively exhaustive (MECE)' way [34]. A tree as shown in Figure 10.2 was designed, to be able to label each data point in certain 'buckets' of waste. The driver tree (which consisted then mainly of expected root causes) was used as a guideline for the labeling. A data point could only end up in one of the labeled waste buckets. Thresh-

[redacted]

Figure 10.1: Waste Dashboard 1, week 47

olds were chosen to assess data points on whether they belonged in the bucket or not. The outcomes of these models were very sensitive to the thresholds and the hierarchy of the tree. The thresholds and hierarchy were merely based on my view and experience from data exploration so far. However, with the knowledge that was created during this study, the hierarchy and thresholds of the MECE waste tree can be tuned to construct a more accurate model that labels data points with their correct root causes.

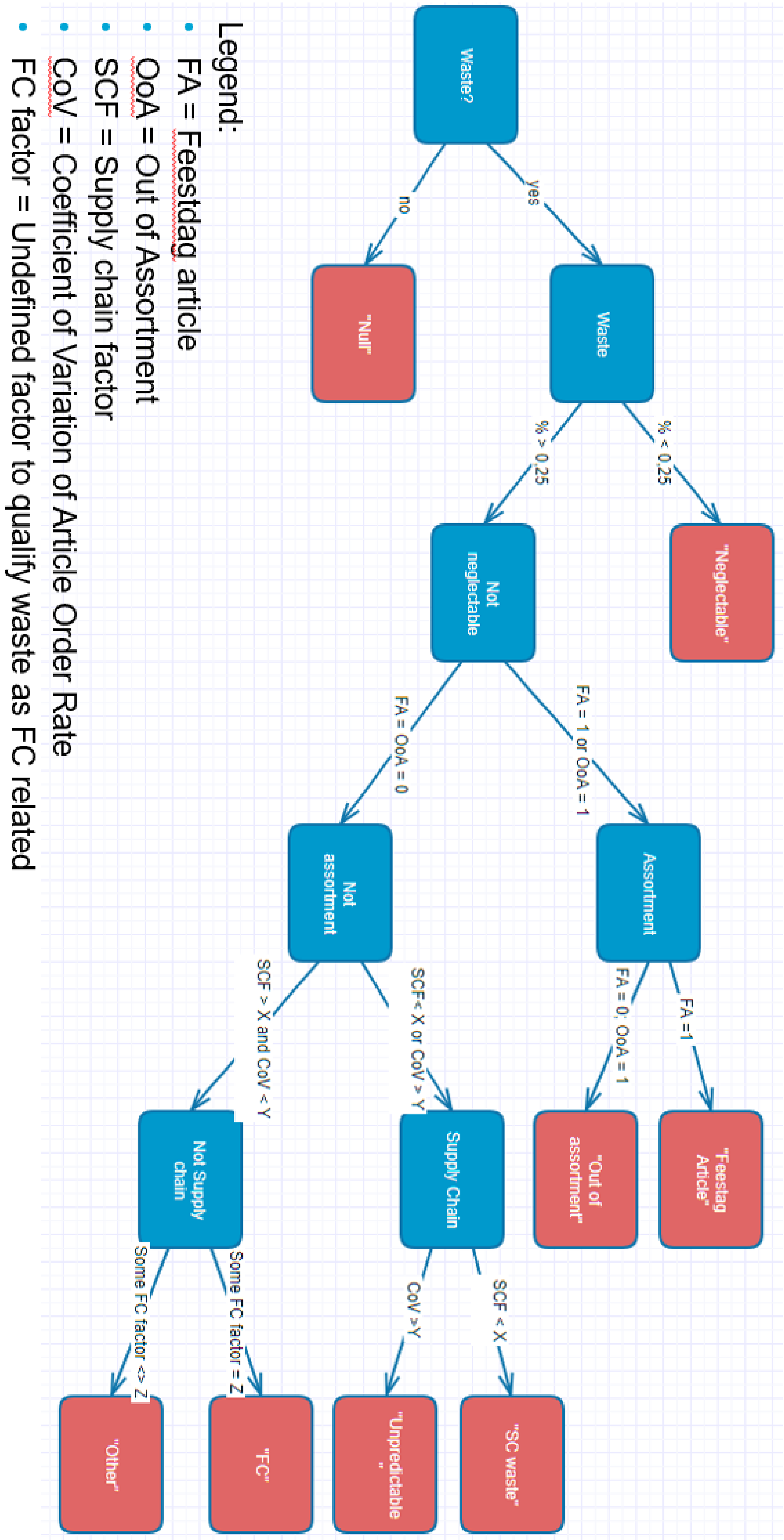


Figure 10.2: MECE tree root causes

10.2. Recommendations for Picnic

10.2.1. Update WMS scanning actions

The first problem I ran into was that it was not clear which data was waste and which data was not. Scanner actions had not been adjusted although procedures were not supposed to be used anymore. A first step is to improve the possible actions. They have to be clear. To be totally clear, a new way of registration can be implemented, where 'waste' is a scanning action. Realizing this is a major operation and also other processes rely on the data structure as is used now, I suggest some minor changes in the WMS registration. They are found in Figure 10.3. The waste MART can then easily filter the right data for waste and be able to label the freshness, quality, damaged and out of assortment waste.

The stock counts data is assumed to contain some waste data, because shoppers use this action a lot and a stock count is an easy way to adjust stock levels, also when it should actually be registered as waste. Picnic would benefit from having more certainty about the stock counts. This would lead to a clearer picture of waste and a higher stock certainty.

Current WMS registration: actions with reasons	**Proposed WMS registration: actions with reasons**
Stock Clearing • Blank (BBD Clearing) • Adjust for BBD clearing (automatic WMS stock adjustment; not actual clearing; used to allow clearing action. Always in combination with clearing action)	Stock ***Freshness*** Clearing • BBD Clearing • PD Clearing
Stock Adjustment • Bad Quality • Damaged • Freshness Guarantee Removal • Internal Use • Intra FC Move • Old Production Date • Out of assortment	Stock Adjustment • Bad Quality • Damaged • Internal Use • Intra FC Move • Out of assortment
Stock Count • Count • Count from zero	Stock Count • Count • Count from zero • Adjust for freshness clearing

Figure 10.3: Suggestion to improve stock mutations

10.2.2. Validate waste data

The next problem is that the data needs to be validated. This is important for the reliability of the dashboards and its interpretations, but also for waste reporting. Currently this waste report is not linked to the operational and financial waste reporting. Operational reporting gets is data from an endpoint that is much more aggregated and therefore cannot see the individual waste actions. The waste data from my analyses and the financial data showed similar patterns, but were never exactly matching. This needs to be researched, so the reporting can be more accurate.

10.2.3. Improve weekly waste dashboards

An improvement for the weekly waste dashboards is to include the day of the week. Deep dives on certain waste cases showed that often waste was cleared on a specific day of the week. This might have to do with the delivery schedule for example, and cannot be seen in a weekly view. Also Sundays are much calmer days, an FC might even be closed. This will affect waste and this effect might become visible in a daily view.

Also, the dashboards can be expanded by including the specific root causes and concrete actions that need to happen in order to reduce waste. Include the suggested MECE tree logic for labeling in dashboards.

III

Part 3: Conclusion & recommendations

11

Conclusion & future research directions

The main research question was:

What are the root causes of food waste at an online supermarkets' fulfilment centres, and how can food waste be reduced?

In literature little information was found about root causes for food waste at a retail level. No generally accepted KPIs for measuring and comparing food waste were found that were applicable to the retail level of the supply chain. Some root causes were described in literature for traditional supermarkets, but no quantified data was available to support these expectations.

Comparing these root causes to the online supermarkets root causes, most effects were expected to be present although having a different effect or magnitude. Causes related to in-store customer behavior could be neglected at online supermarkets, but causes related to customer demand forecasting, the freshness of products and an efficient supply chain are expected to have a bigger effect.

Some (online) supermarkets are able to achieve near zero waste, by organizing a highly effective supply chain, optimizing internal processes, and have food salvation mechanisms, such as donations to charity.

A case study was done at Picnic; a fast growing online supermarket in the Netherlands and Germany. Picnic aspires to be a green company, and therefore needs to focus on their waste habits and performance. Before this research started, little was known about the waste patterns and drivers.

By means of an exploratory research, data was gathered and five KPIs were defined: wasted units (CU waste), cost of waste, relative waste, waste cts/CU sold and waste margin impact. The first three showed waste in consumer units, waste as a monetary value and waste as a percentage of units sold respectively. The last two were constructed for Picnic specifically and are in line with their operational reporting. They indicate waste in euro cents per item sold and waste as an impact on the gross margin. The most practical and reliable KPI for analysis purposes was waste measured in consumer units (CU waste).

The first insights into waste showed that around 2/3 of the waste was freshness related. 1/3 of the waste had causes related to quality, damaged items, or assortment changes. The pattern of waste was not constant over time. It varied per period and per FC. New FCs showed higher relative waste numbers than more mature FCs. Furthermore, around 80 % of the waste came from chilled products, while chilled products form around 20 % of the total assortment. Frozen products caused almost no waste.

Expected root causes of waste were identified for three main domains: assortment, supply chain and FC processes. The factors were:

- Assortment
  - Out of assortment
  - Seasonal waste
  - Product freshness
  - Promotions
- Supply chain
  - Demand quantity
  - Demand volatility
  - Order frequency
  - Freshness guarantee
  - Minimum order quantity
  - Safety stock
- FC processes
  - Replenishment error
  - Counting error
  - Damaged

The factors relating to the assortment and the supply chain were expected to have a big impact on waste, and indicating variables could be constructed and assessed in a clear way. A data set with granularity FC/Article/Period was constructed for these factors, for chilled products only. The factors related to FC processes were not quantified in this study.

Two multivariate regression model types tested the relations and significance of the factors to the dependent variable CU waste. For each regression model, two data sets were tested: one full data set, including a majority of data points where no waste had taken place and one data set with only waste. The dependent variable in all four models was CU waste.

Linear regression models were used to indicate the most important parameters. All factors showed highly significant relations to waste in the multivariate models. The predictive power of the models was not strong. This was caused mainly by a skewed distribution of the dependent variable and its relations to the independent variables. Therefore not all regression assumptions could be validated.

Negative binomial regression models were used to indicate significant factors, while addressing some of the non-linearity in the data sets. Factors were again highly significantly correlated to the output variable; CU waste. The results in significance of the model and of the factors were comparable to the linear regression model. All assumptions could be validated for this model.
Taking both regression models and both data sets into considerations, the five most important causes were:

- High **demand volatility** causes waste. Especially in the full data regressions, the volatility is the strongest factor. It demonstrates that unpredictable products cause waste.

- **Product freshness** is one of the most important factors in all regressions. Products with longer freshness cause less waste.
- **Promotions** are particularly important in the no-zeroes regressions, indicating that given waste, promotions play a big role.
- **Freshness guarantee** is again an important factor in all regressions. Since this variable is a chosen 'setting' for a product, it can easily be influenced.
- **Order frequency** also plays an important role in generating waste. The least frequent delivery schedules caused the most waste.

11.1. Contributions

11.1.1. Scientific Contribution

This study contributed to understanding and quantifying food waste in the retail part of the supply chain. To my knowledge, it is the first time that expected root causes for food waste have been quantitatively tested on significance in the waste generation process, using a multivariate analysis.

The study opens up a new field of research on reducing waste, using quantified data. An approach was shown to gain insights in root causes of waste, using exploratory data analysis, qualitative analysis on identifying factors, quantitative analysis on measuring these factors. Various variables to measure the expected root causes have been constructed, specifically for this research. These variables can be used by other researches or companies to gain insight in food waste performance and ultimately, reduce food waste.

The root causes are assumed to be mostly universal and are expected to have similar effects at other online supermarkets fulfilment centers. Moreover, most factors are expected to cause waste in any food handling company in the retail sector. The root causes are related to issues or processes that are to some extent also present in most other retail stores. Demand of products, promotions, product freshness and safety stocks are aspects of any food handling company and the identified root causes in this research are closely related to all of these.

11.1.2. Contribution to Picnic

The most important causes of waste have been identified, specifically for Picnic. Waste issues at Picnic can now be inspected with accurate insights in the root causes. Weekly updated waste dashboards have been constructed and show waste performance of the last 8 weeks. By a process of continuously improving on waste performance as well as gaining knowledge, the average waste expressed in cents per sold item has dropped from an average for the first 8 periods of 2018 of [redacted] to an average for the last 8 weeks (d.d. 11/2018) of [redacted]. The trend in weekly cost of waste is still decreasing. When Picnic continues focusing and improving their waste performance, Picnic might also be able to achieve near zero waste!

11.2. Limitations of the study

Both the linear and the negative binomial regression models that were used, do not show complete predictability of waste, while also relating all factors that influence waste. This was mostly caused by non-linearity in the model.

Also, this research was mostly exploratory; there was a strong emphasis on understanding the data and researching which factors were contributing to waste. In order to obtain the general influences, the following limitations have been applied:

- Only CU waste was regarded as a dependent variable for the model. This had as a result that some factor effects could not be evaluated optimally, such as the AOR. Ideally some a relative waste KPI could be used.
- Since 80 % of waste came from chilled products, only these were considered for regression analysis. The waste patterns for ambient and frozen can be different, but we do not know yet.
- The full data availability of Picnic has not been used. Using the detailed data of customers, customer types etc. different waste patterns might become visible.

11.3. Future research directions

Some recommendations are a direct result of the limitations. These recommendations are:

- Include a reliable relative waste KPI in the data. For some factors, such as AOR and SF, the effects can be explained in more detail.
- Analyze waste at the other two temperature zones. Do they have the same root causes?
- utilizing the full potential of the data availability at e-grocers. Detailed information about individual customers and orders can be used to improve waste prediction or identify waste patterns.

Quantification could be extended by improving the statistical models or using different, more complex models.

- Aim for linearity between variables. See if transforming data can achieve linearity for all relations between dependent and independent variables
- Interaction of two or more variables might be able to improve the model. Several promising interaction effects are expected. For example the interaction between volatility of demand and order frequency, or freshness of products combined with freshness guarantee.
- Other mathematical/statistical models, such as a zero-inflated regression model, or machine learning to improve the predictive power of models. Since there is an abundance of zeroes in the data set, a zero-inflated regression model can be considered. This model distinguishes two kinds of zeroes: the first kind is caused by a certain parameter, the second is not.

A final recommendation is to find a relation that describes the trade-off between waste and unavailability. When the influencing parameters are known and modeled, a company can make accurate decisions about how much waste or unavailability they allow.

Bibliography

[1] Uwe Hoering. Loss and Waste. pages 1–10, 2012. URL http://www.dandc.eu/en/article/lots-food-wasted-worldwide-aid-agency-efforts-improve-matters-hardly-help-poor.

[2] Clara Cicatiello, Silvio Franco, Barbara Pancino, and Emanuele Blasi. The value of food waste: An exploratory study on retailing. *Journal of Retailing and Consumer Services*, 30: 96 – 104, 2016. ISSN 09696989. doi: 10.1016/j.jretconser.2016.01.004.

[3] Sara Corrado and Serenella Sala. Food waste accounting along global and European food supply chains: State of the art and outlook. *Waste Management*, 79:120–131, 2018. ISSN 18792456. doi: 10.1016/j.wasman.2018.07.032. URL https://doi.org/10.1016/j.wasman.2018.07.032.

[4] Lukas Chabada, Cecilie Damgaard, Heidi Dreyer, Hans Hvolby, Iskra Dukovska-Popovska, Iskra Dukovska-Popovska Logistical, Bernard Grabot, Bruno Vallespir, Samuel Gomes, Abdelaziz Bouras, and Dimitris Kiritsis. Logistical Causes of Food Waste: A Case Study of a Norwegian Distribution Chain of Chilled Food Products. pages 273–280, 2014. doi: 10.1007/978-3-662-44739-0_34. URL https://hal.inria.fr/hal-01388261.

[5] FAO. Key facts on food loss and waste you should know! | SAVE FOOD: Global Initiative on Food Loss and Waste Reduction | Food and Agriculture Organization of the United Nations. URL http://www.fao.org/save-food/resources/keyfindings/en/.

[6] How 'Ugly' Fruits and Vegetables Can Help Solve World Hunger. URL https://www.nationalgeographic.com/magazine/2016/03/global-food-waste-statistics/.

[7] FAO. Global food loss and waste. *Food wastage footprint & Climate Change*, (1):1–4, 2013.

[8] Unfccc. ADOPTION OF THE PARIS AGREEMENT - Paris Agreement text English. Technical report. URL https://unfccc.int/sites/default/files/english{_}paris{_}agreement.pdf.

[9] Jonathan A. Foley, Navin Ramankutty, Kate A. Brauman, Emily S. Cassidy, James S. Gerber, Matt Johnston, Nathaniel D. Mueller, Christine O'Connell, Deepak K. Ray, Paul C. West, Christian Balzer, Elena M. Bennett, Stephen R. Carpenter, Jason Hill, Chad Monfreda, Stephen Polasky, Johan Rockström, John Sheehan, Stefan Siebert, David Tilman, and David P.M. Zaks. Solutions for a cultivated planet. *Nature*, 478(7369):337–342, 2011. ISSN 00280836. doi: 10.1038/nature10452.

[10] About the Sustainable Development Goals - United Nations Sustainable Development, . URL https://www.un.org/sustainabledevelopment/sustainable-development-goals/.

[11] Goal 2: Zero hunger | UNDP, . URL http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-2-zero-hunger.html.

[12] Fao. SAVE FOOD: Global Initiative on Food Loss and Waste Reduction Definitional framework of food loss. 2014.

[13] Jean C Buzby and Jeffrey Hyman. Total and per capita value of food loss in the United States q. *JOURNAL OF FOOD POLICY*, 37:561–570, 2012. doi: 10.1016/j.foodpol.2012.06.002. URL http://dx.doi.org/10.1016/j.foodpol.2012.06.002.

[14] Matteo Vittuari, Paolo Azzuro, Silvia Gaiani, Manuela Gheoldus, Stephanie Burgos, Lusine Aramyan, Natalia Valeeva, David Rogers, Karin Östergren, Toine Timmermans, and Hilke Bos-Brouwers. Recommendations and guidelines for a common European food waste policy framework. Technical report, 2016.

[15] European Commision. An EU action plan for the circular economy. Technical report, 2015. URL https://eur-lex.europa.eu/resource.html?uri=cellar:8a8ef5e8-99a0-11e5-b3b7-01aa75ed71a1.0012.02/DOC{_}1{&}format=PDF.

[16] Mattias Eriksson, Ingrid Strid, and Per Anders Hansson. Food losses in six Swedish retail stores: Wastage of fruit and vegetables in relation to quantities delivered. *Resources, Conservation and Recycling*, 68:14 – 20, 2012. ISSN 09213449. doi: 10.1016/j.resconrec.2012.08.001.

[17] Peter Willis. Waste arisings in the supply of food and drink to households in the UK. Technical report, LeeWillis2010. URL www.wrap.org.uk.

[18] Mathava Kumar, Yan Liang Ou, and Jih Gaw Lin. Co-composting of green waste and food waste at low C/N ratio. *Waste Management*, 30(4):602–609, 2010. ISSN 0956053X. doi: 10.1016/j.wasman.2009.11.023. URL https://ir.nctu.edu.tw/bitstream/11536/5572/1/000275766700009.pdf.

[19] WRAP supply chain waste presentation. Technical report. URL http://www.wrap.org.uk/sites/files/wrap/2WRAPManageandMeasureSCWasteWEBFINAL.pdf.

[20] Nell Fry. A Strategic Guide for Using Data to Drive Food Loss and Waste Reductions. Technical report, 2017. URL http://info.spoileralert.com/hubfs/AStrategicGuideforUsingDatatoDriveFoodLossandWasteReductions{%}5B09-2017{%}5D.pdf?t=1538666292581{&}utm{_}source=hs{_}automation{&}utm{_}medium=email{&}utm{_}content=56100293{&}{_}hsenc=p2ANqtz-{_}zoFu8Egn.

[21] Christoph Teller, Christina Holweg, Gerald Reiner, and Herbert Kotzab. Retail store operations and food waste. *Journal of Cleaner Production*, 185:981 – 997, 2018. ISSN 09596526. doi: 10.1016/j.jclepro.2018.02.280.

[22] Carlos Mena, B. Adenso-Diaz, and Oznur Yurt. The causes of food waste in the supplier-retailer interface: Evidences from the UK and Spain. *Resources, Conservation and Recycling*, 55(6):648–658, 2011. ISSN 09213449. doi: 10.1016/j.resconrec.2010.09.006. URL http://dx.doi.org/10.1016/j.resconrec.2010.09.006.

[23] Sequoia Capital and Engineering Management. Chains. pages 93–103, 1997.

[24] M. E. Buisman, R. Haijema, and J. M. Bloemhof-Ruwaard. Discounting and dynamic shelf life to reduce fresh food waste at retailers. *International Journal of Production Economics*, xxx:1 – 11, 2017. ISSN 09255273. doi: 10.1016/j.ijpe.2017.07.016.

[25] Michael Ketzenberg, Gary Gaukler, and Victoria Salin. Expiration dates and order quantities for perishables. *European Journal of Operational Research*, 266:569 584, 2018. ISSN 03772217. doi: 10.1016/j.ejor.2017.10.005.

[26] Ocado reveals near zero food waste levels – Ocado Group, . URL http://www.ocadogroup.com/news-and-media/news-centre/2018/nationalfoodwaste.aspx.

[27] Ocado Group plc Annual Report and Accounts Ocado Brands and Partners World Shops OCADO RETAIL OCADO SOLUTIONS DRIVING GROWTH. Technical report, . URL www.ocadogroup.com.

[28] Andrew T. Jebb, Scott Parrigon, and Sang Eun Woo. Exploratory data analysis as a foundation of inductive research. *Human Resource Management Review*, 27(2):265–276, 2017. ISSN 10534822. doi: 10.1016/j.hrmr.2016.08.003. URL http://dx.doi.org/10.1016/j.hrmr.2016.08.003.

[29] Ralph Kimball, Margy Ross, Warren Thorntwaite, and Joy Mundy. *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses (Second ed.)*. 2008. ISBN ISBN 978-0-470-14977-5.

[30] Choosing the Correct Type of Regression Analysis - Statistics By Jim. URL http://statisticsbyjim.com/regression/choosing-regression-analysis/.

[31] J Durbin and G S Watson. Biometrika Trust Testing for Serial Correlation in Least Squares Regression. III. Technical Report 1, 1971. URL https://about.jstor.org/terms.

[32] NCSS and LLC. NCSS Statistical Software Negative Binomial Regression. Technical report. URL https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative{_}Binomial{_}Regression.pdf.

[33] Shengping Yang and Gilbert Berdine. Negative Binomial Regression StatiSticS column. *The Southwest Respiratory and Critical Care Chronicles*, 3(10), 2015. doi: 10.12746/swrccc2015.0310.135. URL http://pulmonarychronicles.com/index.php/pulmonarychronicles/article/viewFile/200/480.

[34] Barbara Minto. *The pyramid principle*. Pearson Eductaion Limited, 2008. ISBN 9780273710516.

Appendices

A
Research paper

Identification and confirmation of root causes of food waste in the retail part of the supply chain

J. Meijboom[*], B. Atasoy[*], F.B. Gorte[**] and D.L. Schott[*]

[*]Section Transport Engineering and Logistics, Department Maritime & Transport Technology, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, The Netherlands
[**]Picnic B.V., Amsterdam, The Netherlands

Keywords: food waste, root causes, exploratory data analysis (EDA), e-groceries, online, food supply chain, retail, chilled

Abstract: Food waste is a significant problem worldwide and serious actions need to be taken to reduce it, both on the research side and in practice by companies. To achieve this, accurate insights in the root causes of food waste are required. In literature root causes have been described only in expert interviews, not in quantified analyses. The aim of this study is to identify and quantitatively confirm root causes of food waste in the retail part of the supply chain. The ultimate goal is to show an approach to gain insights to reduce food waste as well as laying a foundation for future research.

Exploratory data analysis was performed in a case study at online supermarket 'Picnic B.V.' in the Netherlands. High data availability at the online supermarket made this research possible. Expected root causes were identified, hypotheses were generated and variables were constructed to measure them.

Four waste domains were identified: assortment, supply chain, fulfilment center processes and other, of which we focused on the first two. In these, twelve expected root causes were identified. Negative binomial regression models have been used to statistically confirm the root causes. The five most important root causes to come forward from the regression models are: demand volatility, product freshness, promotions, freshness guarantee and order frequency.

This study has shown an approach to quantitatively identify and confirm root causes of food waste. Root causes are expected to be universal and this approach can be applied in other food handling companies in retail.

1 INTRODUCTION

Parallel to the gradual expansion of food production and the demand for high quality products, the amount of food waste has increased to astonishing levels, estimated to 30 % of produced food [1]. This indicates that in theory, all food that is wasted, could easily feed twice the world's population that is in hunger [2]. Waste has important environmental and economical consequences. In terms of emissions; if "food waste were a country, it would be the third largest emitter of greenhouse gases, after the USA and China." [3].

In order to meet the conditions of the Paris agreement [4], agriculture's environmental footprint should be drastically reduced, despite a required increase in global food production [5]. Furthermore, all the resources to produce and transport food that is wasted have all cost money. Reducing food waste automatically improves economic situations; farmers have to produce less, less food has to be distributed, supermarkets have to buy less and allocate less space in warehouse and consumers have to buy less.

Literature, analyzing the causes of food waste in retail is scarce. No clear definitions [6] and key performance indicators [7] have been created to measure food waste at a retail level. Some root causes have been identified [8], [9], [10]. They are mostly related to product freshness, demand patterns, delivery schedules, data availability, incorrectly executed processes and management decisions. However, none of these were quantitatively confirmed.

E-grocers can accurately track consumers orders and shopping behavior. This generates large size high resolution data. Therefore, accurate forecasts can be made about demand. This forecasting accuracy and data availability gives e-grocers specifically a unique opportunity to measure, quantify and possibly reduce waste.

This study aims at identifying and confirming expected root causes of food waste production at an online supermarket in a quantitative way. The approach to identify root causes will have a universal aspect and can open up a new field of research on quantitative analysis of causes of food waste at the retail stage of the supply chain. Ultimately this research contributes to the knowledge about reducing food waste; an urgent societal issue.

2 METHODOLOGICAL FRAMEWORK

The study was designed as an exploratory case study at an online supermarket operating in the Netherlands, to identify and confirm the root causes of food waste. The methodology is designed in three phases of analysis:

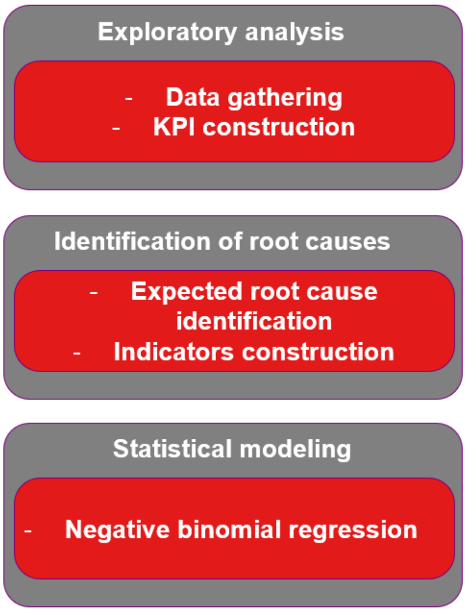


Figure 1: Schematic methodology

- **An exploratory analysis**: data was collected and investigated. Key performance indicators (KPIs) were constructed and an area of focus was determined.
- **Identification of root causes**: expected root causes were identified and hypotheses generated in a qualitative way. Identification of expected root causes occurred mostly on the basis of inspecting the available data and local expert interviews. Quantitative indicators were constructed to measure the expected root causes. These were created by means of separate analyses for all factors. A data set was constructed for chilled articles, containing 45k rows and 13 columns: 1 dependent variable and 12 independent variables.
- **Statistical modeling**: the significance of the hypothesized root causes were tested with a negative binomial regression model.

3 ANALYSIS AND RESULTS

3.1 Exploratory analysis: KPI and waste focus

As mentioned in the introduction, there was no generally accepted KPI for food waste. In literature the most common unit of measure is mass [7], [11], [12]. The studies that used mass as a KPI usually had a much bigger scope than only retail of one company. Some other KPIs have been identified, such as percentage of units purchased [13] or percentage of mass delivered [11]. At Picnic, most cost KPIs are expressed in euro's, or in euro cents per item sold. The KPI to be used for this research was waste in absolute number of consumer units: CU waste. This indicator is a combination of different stock mutations in the warehouses, that are accurately registered by means of scanning actions. The source is a reliable, quality assured and most complete data source with detailed information about waste, that is currently available at Picnic. The indicator is not distorted by other data sources.

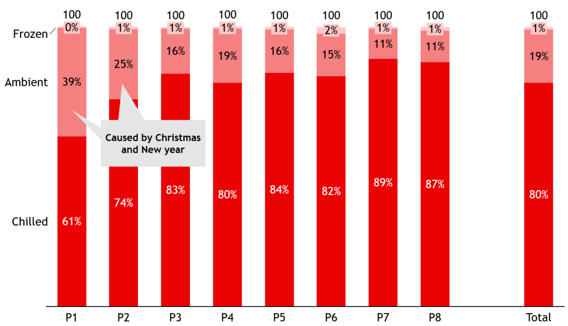


Figure 2: Waste per temperature zone, P1 - P8, 2018

Figure 2 shows the distribution of waste over the course of eight periods per temperature zone (a period consists of 28 days). On average, chilled products caused 80 % of the waste, while consisting of only 20 % of the assortment. Therefore, for analysis

Waste	Factor	Variable name	Type	Range
Assortment	Out of assortment	Goes out	Binary	[0 / 1]
	Seasonal	Season	Binary	[0 / 1]
	Product freshness	FC period (days)	Discrete	-3 - 344
	Promotions	Promo pressure	Continuous	0% - 100%
Supply chain	Demand qty	Article Order Rate (AOR)	Continuous	0 - 0.50
	Demand volatility	Coefficient of Variation AOR	Continuous	0 - 5
	Order frequency	Max T betw. Deliveries	Discrete	1 - 7
	Freshness guarantee	FG (days)	Discrete	0 - 21
	Minimum order qty	CU/TU	Discrete	1 - 92
	Safety stock	Safety Factor (SF), High SF, High SF * CU/TU	Continuous, binary, interaction	0.2 - 0.5, [0 / 1], 0 - 92
FC process	Replenishment error	?	-	-
	Counting error	?	-	-
	Damaged	?	-	-
Other	Unidentified	?	-	-

Table 1: Root causes (factors) and quantitative indicators (variables)

purposes, only waste from the chilled areas has been considered.

3.2 Identification of root causes: factors, hypotheses and quantification

Expected root causes were identified for four domains of the online supermarket; assortment, supply chain, fulfilment center (FC) processes and other. Table 1 shows the domains, the expected root causes (factors), quantitative indicators (variables) and their types and range.

For the assortment and supply chain related factors twelve quantitative indicators (variables in Table 1) were constructed. In the FC domain also various causes for products to become waste were identified, mostly related to human errors. For those no adequate quantitative measures could be constructed, and therefore they were not included in this study.

The following sections describe the factors and their hypotheses about generating waste.

3.2.1 Assortment

Assortment related waste is associated with physical properties of products or changes in the assortment.

- **Out of assortment** indicates if a product is discontinued. It is expected to cause waste, since often stock is left in an FC when not everything is sold. Out of assortment is indicated by the binary variable 'Goes out.' This factor was identified for this research specifically; it was not found in literature.
- **Seasonal.** There are different types of seasonality, such as BBQ meats, Christmas products, but also asparagus. They are all considered seasonal and are indicated by the binary variable 'Season.' Due to their irregular demand and ordering patterns seasonal products are expected to cause waste.
- **Product Freshness** is expected to play a role in waste, since products with a longer shelf life are expected to cause less waste. The indicating variable is called 'FC period' and was also constructed specifically for this research. FC period shows the maximum number of days that a product can be in an FC. It is the outcome of an analysis that takes the mode of the remaining shelf life and subtracts the freshness guarantee.
- **Promotions** was an identified factor in literature.

Similar to seasonal products, promotions can have irregular demand and order patterns. They are therefore harder to forecast and are expected to cause waste. Promotions are indicated by 'Promo pressure;' the percentage of products that was sold in a promotion.

3.2.2 Supply Chain

The supply chain factors are related to the company's role in the supply chain. This can be interactions and agreements with customers and suppliers, but also managerial decisions that have to be taken.

- **Demand quantity** is expected to cause waste. Products with higher demand are likely to cause more waste. This variable was also found in literature. The variable to measure demand quantity for this research is called the Article Order Rate (AOR), and is defined by the amount of orders in a day divided by the amount of orders of that same day. For periods longer than one day, the average AOR is taken.
- **Demand volatility** indicates the dispersion of demand, and thereby its predictability. Most products are ordered on the basis of a forecast, and therefore unpredictable products are expected to cause waste. The constructed variable for the demand volatility is expressed as the coefficient of variation of the article order rate ('CoV AOR'), defined by the standard deviation of the AOR divided by its mean.
- **Order frequency** is based on the delivery schedule that is agreed upon with suppliers. A high order frequency ensures that lower quantities have to be ordered in one delivery and more stock adjustment moments are available. A high order frequency is expected to cause less waste than a low order frequency. The constructed indicator for the order frequency is the maximum time between deliveries (Max T betw. del.).
- **Freshness guarantee** is a factor that specifically applies to e-grocers. After delivery, a freshness guarantee shows the minimum amount of days that a product should still be fresh. This puts the supply chain under pressure and a high freshness guarantee (on fresh products) is expected to cause waste. The indicating variable is called 'FG' and describes the guarantee in days.
- **Minimum order quantity (MOQ)** is the minimum amount of products that can be ordered in one delivery. This factor came forward from expert interviews. Picnic did not have a specific MOQ, but it can be related to the packaging size, expressed as the variable 'consumer unit per trade unit' (CU/TU). The hypothesis is that large MOQ generates more waste.
- **Safety stock** is an extra stock that is kept on top of predicted demand to guarantee availability of products. The hypothesis is that a large safety stock causes more waste. Three variables were developed for the safety stock in this study:
  - A safety factor (SF), the factor that is ordered on top of the predicted order, ranging between 0.2 for products that have a high daily demand and 0.5 for products that have a low daily demand.
  - A high safety factor: a binary variable for the safety factor, that only is 1 for slow movers.
  - the high SF multiplied by the MOQ, indicated by SF x CU/TU an expected indicator for problematic slow movers.

3.3 Statistical Modeling

For the twelve available independent variables and the dependent variable 'CU waste', two data sets were constructed, with a data point for each product, FC and financial period which after cleaning resulted in 44131 data points. The first data set contained all data points. Figure 3 shows the distribution of the dependent variable, including a majority of zero waste data points. The second data set contained only waste data points (Figure 4), excluding zero waste. The distinction between the two data sets was made because the hypothesis was that the factors causing the event of waste (waste/no waste) are different from the factors causing increasing waste, given that waste is present.

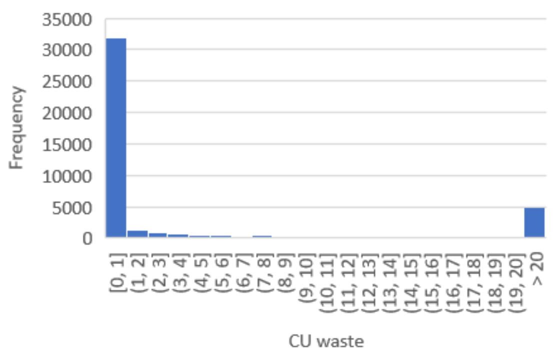


Figure 3: Histogram CU Waste - full data

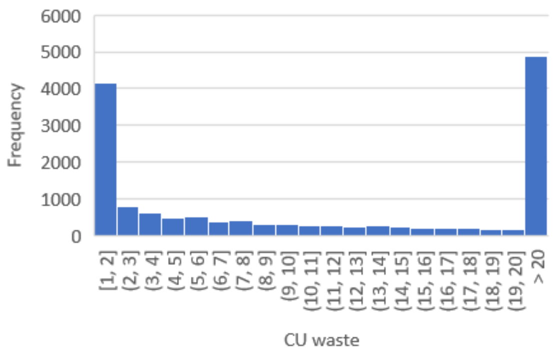


Figure 4: Histogram CU waste - no zeroes

Negative binomial regression models have been applied to this full data set and a data set without zeroes, to test the hypotheses and significance of factors. A negative binomial regression model is a generalized multivariate linear model in which the dependent variable Y_i is a count variable. The dependent variable in this analysis can be seen as a count of waste instances. The negative binomial regression model is given by Equation (1). This model takes the natural logarithm of the dependent variable to find coefficients (β) for the independent variables (x_i), i.e. the impact of the independent variables on the waste [14].

$$ln(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_n x_{ni} \qquad (1)$$

The significance of the models is shown in Table 2 and indicates if these models provide a better fit than the intercept-only model (when only β_0 is present). Both values are 0.000, which tells us that this is true and the null hypothesis is rejected very strongly in both cases.

	Omnibus test	
	Full data	No zeroes
Likelihood Ratio Chi-Square	7391	2294
Degrees of freedom	12	12
Significance	0.000	0.000

Table 2: Omnibus test negative binomial regression

The regression results of the two models with coefficients (β, indicated by B) Wald Chi-Square values, significance and the relative domain of the root cause are shown in Table 3. All factors showed a highly significant relationship with waste in at least one of the regression models, confirming that these are indeed root causes of food waste.
Comparing the two models, the coefficients of the same factors are mostly in the same direction, so the effects on waste of most factors is clear in this multivariate analysis. Only promo and high SFxCU/TU have opposite coefficients. This indicates that these factors have opposite effects in the model with and without waste.
Furthermore, the order of most important root causes in terms of significance vary between the two models. This indicates that the model describing the full data set is in fact describing a different process than the model describing the 'only waste' scenario.

4 CONCLUSION AND FUTURE RESEARCH DIRECTIONS

The hypothesized root causes of food waste in a Dutch e-grocer could be studied in detail by quantitative indicators, made possible by the high data availability. Expected root causes of waste were discussed and quantified where possible, expressed as factors and variables. The factors relating to the assortment and the supply chain were expected to have a big impact on waste, and could be measured with variables, constructed for this research. By means of negative binomial regression models the identified factors were confirmed as root causes of food waste.

Not all domains could be investigated for causes of food waste due to the lack of numeric data in the causes in the FC domain, and the unknown domain of 'other'.

The five most important factors to come forward from the regression analyses were:

- High **demand volatility** (CoV AOR) causes waste. Especially in the full data regressions, the volatility is the strongest factor. It demonstrates that unpredictable demand for products causes waste.
- **Product freshness** (FC period) is one of the most important factors in all regressions. Products with longer freshness cause less waste.
- **Promotions** (Promo pressure) are particularly important in the no-zeroes regressions, indicating that given waste, promotions play a big role.
- **Freshness guarantee** (FG) is again an important factor in all regressions. Since this variable is a chosen 'setting' for a product, it can easily be influenced.
- **Order frequency** (Max T betw. del.) also plays an important role in generating waste. The least frequent delivery schedules caused the most waste.

	Negative binomial regression									
	FULL DATA					WASTE ONLY: NO ZEROES				
rank	Factor	B	Wald Chi-Square	Sig.	Domain	Factor	B	Wald Chi-Square	Sig.	Domain
1	CoV AOR	4.66	3989.8	0.000	SC	FC period	-0.01	556.2	0.000	Asst
2	FC period	-0.01	1000.9	0.000	Asst	FG	-0.06	214.3	0.000	SC
3	Max T betw. del.	0.23	215.0	0.000	SC	Promo pressure	0.85	174.3	0.000	Asst
4	FG	-0.06	179.5	0.000	SC	Max T betw. del.	0.12	117.3	0.000	SC
5	Promo pressure	-1.13	155.2	0.000	Asst	Season	0.76	105.1	0.000	Asst
6	High SF	-0.59	112.6	0.000	SC	SF	-1.43	41.2	0.000	SC
7	AOR	17.91	74.5	0.000	SC	CU/TU	0.02	38.1	0.000	SC
8	Goes out	1.03	69.1	0.000	Asst	CoV AOR	0.28	30.3	0.000	SC
9	SF	-3.49	61.9	0.000	SC	High SF	-0.19	21.2	0.000	SC
10	CU/TU	0.02	15.3	0.000	SC	High SF x CU/TU	0.02	12.2	0.000	SC
11	High SF x CU/TU	-0.01	4.7	0.029	SC	AOR	0.13	0.0	0.842	SC
12	Season	0.10	1.2	0.269	Asst	Goes out	0.76	0.0	0.865	Asst

Table 3: Factor significance, negative binomial regression

This study has opened up a new field of research, by showing a new approach to quantify waste and identify the most important factors. Quantitative statistical analysis tested hypotheses and significance of identified causes. To the authors' knowledge, this has not been published before.

The case study has taken place at an e-grocer; large amounts of data were available and made the construction of KPIs and factors possible. The identified root causes are also expected to be relevant for other food handling companies in the retail part of the food chain, such as traditional supermarkets. The identified and confirmed root causes of food waste can aid in efficient reduction of waste at retail companies. Also, companies can use this study's approach to identify their specific root causes.
Future research can improve the modeling of waste, focusing on:

- different KPIs, such as a relative waste KPI, expressing waste as a percentage of purchased products
- different temperature zones. Do they have similar root causes as chilled products?
- interaction effects between factors. Several promising interaction effects are expected. For example the interaction between volatility of demand and order frequency, or freshness of products combined with freshness guarantee.
- utilizing the full potential of the data availability at e-grocers. Detailed information about individual customers and orders can be used to improve waste prediction or identify waste patterns.
- other mathematical/statistical models, such as a zero-inflated regression model, or machine learning to improve the predictive power of models.
- finding a relation that describes the trade-off between waste and unavailability. When the influencing parameters are known and modeled, a company can make accurate decisions about how much waste or unavailability they allow.

REFERENCES

[1] Fao. SAVE FOOD: Global Initiative on Food Loss and Waste Reduction Definitional framework of food loss. 2014.

[2] How 'Ugly' Fruits and Vegetables Can Help Solve World Hunger.

[3] FAO. Global food loss and waste. *Food wastage footprint & Climate Change*, (1):1–4, 2013.

[4] Unfccc. ADOPTION OF THE PARIS AGREEMENT - Paris Agreement text English. Technical report.

[5] Jonathan A. Foley, Navin Ramankutty, Kate A. Brauman, Emily S. Cassidy, James S. Gerber, Matt Johnston, Nathaniel D. Mueller, Christine O'Connell, Deepak K. Ray, Paul C. West, Christian Balzer, Elena M. Bennett, Stephen R. Carpenter, Jason Hill, Chad Monfreda, Stephen Polasky, Johan Rockström, John Sheehan, Stefan Siebert, David Tilman, and David P.M. Zaks. Solutions for a cultivated planet. *Nature*, 478(7369):337–342, 2011.

[6] C Caldeira, C Sara, and S Serenella. Food waste accounting. Technical report, 2017.

[7] Sara Corrado and Serenella Sala. Food waste accounting along global and European food supply chains: State of the art and outlook. *Waste Management*, 79:120–131, 2018.

[8] Christoph Teller, Christina Holweg, Gerald Reiner, and Herbert Kotzab. Retail store operations and food waste. *Journal of Cleaner Production*, 185:981 – 997, 2018.

[9] Carlos Mena, B. Adenso-Diaz, and Oznur Yurt. The causes of food waste in the supplier-retailer interface: Evidences from the UK and Spain. *Resources, Conservation and Recycling*, 55(6):648–658, 2011.

[10] Lukas Chabada, Cecilie Damgaard, Heidi Dreyer, Hans Hvolby, Iskra Dukovska-Popovska, Iskra Dukovska-Popovska Logistical, Bernard Grabot, Bruno Vallespir, Samuel Gomes, Abdelaziz Bouras, and Dimitris Kiritsis. Logistical Causes of Food Waste: A Case Study of a Norwegian Distribution Chain of Chilled Food Products. pages 273–280, 2014.

[11] Mattias Eriksson, Ingrid Strid, and Per Anders Hansson. Food losses in six Swedish retail stores: Wastage of fruit and vegetables in relation to quantities delivered. *Resources, Conservation and Recycling*, 68:14 – 20, 2012.

[12] Peter Willis. Waste arisings in the supply of food and drink to households in the UK. Technical report, LeeWillis2010.

[13] Nell Fry. A Strategic Guide for Using Data to Drive Food Loss and Waste Reductions. Technical report, 2017.

[14] Michael Zwilling. Negative Binomial Regression. *The Mathematica Journal*, 15, 2013.

B
Extract endpoint data

Figure B.1: Endpoint extract data

C
Extract regression data

Waste CU	Goes Out	Season	FC period	promo	AOR	CoV AOR	nax <> deliveri	FG (days)	CU/TU	Safety factor	High SF	High SF * CU/TU
0	0	0	63	0	0.006	0.423	1	21	16	0.5	1	16
0	0	0	14	0	0.012	0.347	2	5	6	0.427888721	0	0
0	0	0	7	0	0.008	0.419	2	5	6	0.480182129	0	0
0	0	0	17	0	0.004	0.489	2	7	6	0.5	1	6
0	0	0	39	0	0.005	0.416	1	5	12	0.5	1	12
0	0	0	14	0	0.009	0.485	2	5	6	0.494532554	0	0
0	0	0	18	0	0.013	0.305	2	4	6	0.426029325	0	0
37	0	0	4	0	0.022	0.179	2	2	8	0.359527459	0	0
0	0	0	6	0.4449	0.05	0.717	2	3	9	0.254136819	0	0
11	0	0	6	0	0.036	0.456	2	5	11	0.304554292	0	0
20	0	0	3	0	0.006	0.373	2	3	4	0.5	1	4
72	0	0	2	0.6397	0.009	1.051	2	3	8	0.478593227	0	0
0	0	0	21	0	0.02	0.379	1	5	6	0.368200192	0	0
7	0	0	21	0	0.006	0.477	1	5	6	0.5	1	6
0	0	0	116	0	0.001	0.614	2	7	6	0.5	1	6
15	0	0	21	0.0071	0.004	0.464	1	5	6	0.5	1	6
8	0	0	6	0	0.005	0.362	2	3	4	0.5	1	4
0	0	0	7	0	0.012	0.272	2	3	5	0.449047307	0	0
1	0	0	20	0	0.005	0.488	2	3	4	0.5	1	4
38	0	1	4	0	0.006	0.51	2	4	4	0.5	1	4
0	0	1	7	0	0	0	2	3	2	0.5	1	2
2	0	0	6	0.3048	0.09	0.275	2	3	8	0.2	0	0
10	0	0	13	0	0.002	0.512	2	3	6	0.5	1	6
5	0	0	12	0	0.002	0.54	2	3	1	0.5	1	1
37	1	1	3	0	0.001	0	2	3	4	0.5	1	4
25	1	1	3	0	0.001	0.693	2	3	4	0.5	1	4
0	0	0	3	0	0.009	0.319	2	3	6	0.469814998	0	0
0	0	0	6	0	0.167	0.208	2	3	12	0.2	0	0
3	0	0	30	0	0	0	2	3	8	0.5	1	8
0	0	0	1	0.0562	0.03	0.252	2	3	12	0.323638625	0	0
57	0	0	3	0	0.006	0.517	2	3	3	0.5	1	3
45	0	0	3	0	0.007	0.412	2	3	3	0.5	1	3
29	0	0	3	0	0.005	0.505	2	3	3	0.5	1	3
1	0	0	8	0	0.112	0.163	2	4	16	0.261012835	0	0
1	0	0	8	0	0.002	0.471	2	5	4	0.5	1	4
18	0	0	7	0	0.033	0.191	2	3	10	0.306858486	0	0
0	0	0	8	0	0.017	0.269	2	3	10	0.388877799	0	0
12	0	0	5	0	0.06	0.144	2	3	10	0.228724484	0	0
0	0	0	6	0	0.04	0.174	2	3	10	0.281493865	0	0
0	0	0	6	0.0249	0.034	0.246	2	3	8	0.304346826	0	0
0	0	0	30	0	0	0	1	3	9	0.5	1	9
0	0	0	4	0	0.072	0.214	2	5	22	0.206977134	0	0
69	0	0	30	0	0.104	0.232	2	3	92	0.2	0	0
0	0	0	2	0	0	0	2	3	2	0.5	1	2
78	0	0	4	0	0.014	0.546	2	3	4	0.435037132	0	0
20	0	0	4	0	0.009	0.392	2	3	4	0.46474759	0	0
22	0	0	4	0	0.012	0.568	2	3	4	0.456161589	0	0
0	0	0	152	0	0	0	2	3	12	0.5	1	12
30	0	0	5	0	0.009	0.318	2	5	6	0.479783084	0	0

Figure C.1: Extract regression data

D
Waste dashboards

[redacted]

Figure D.1: DB0; waste KPIs last 8 weeks

[redacted]

Figure D.2: DB2; Broken and quality waste quantities

[redacted]

Figure D.3: DB3; top waste items last week: broken and damaged

E

Linear regression results

```
 GET
  FILE='C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Wa
ste - JM\Afstuderen\Regression\RegressionSPSS\full_data.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
GET
  FILE='C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Wa
ste - JM\Afstuderen\Regression\RegressionSPSS\data_no_zeroes.sav'.
DATASET NAME DataSet2 WINDOW=FRONT.
DATASET ACTIVATE DataSet1.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT WasteCU
  /METHOD=ENTER FC_period T_max_del Goes_out Season Promo AOR CoV_AOR SF Hi
gh_SF CU_TU
    High_SFxCU_TU FG
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /SAVE PRED ZPRED COOK RESID ZRESID.


 REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT WasteCU
  /METHOD=ENTER FC_period T_max_del Goes_out Season Promo AOR CoV_AOR SF Hi
gh_SF CU_TU
    High_SFxCU_TU FG
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /SAVE PRED ZPRED COOK RESID ZRESID.
```

Regression

Notes

Output Created		18-DEC-2018 15:43:29
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\full_data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	44131
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any variable used.
Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE ZPP /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT WasteCU /METHOD=ENTER FC_period T_max_del Goes_out Season Promo AOR CoV_AOR SF High_SF CU_TU High_SFxCU_TU FG /SCATTERPLOT=(*ZRESID ,*ZPRED) /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID) /SAVE PRED ZPRED COOK RESID ZRESID.
Resources	Processor Time	00:00:06.45
	Elapsed Time	00:00:05.29

Notes

	Memory Required	11664 bytes
	Additional Memory Required for Residual Plots	504 bytes
Variables Created or Modified	PRE_2	Unstandardized Predicted Value
	RES_2	Unstandardized Residual
	ZPR_2	Standardized Predicted Value
	ZRE_2	Standardized Residual
	COO_2	Cook's Distance

Descriptive Statistics

	Mean	Std. Deviation	N
WasteCU	8.01	28.860	44131
FC_period	28.74	45.156	44131
T_max_del	2.05	1.005	44131
Goes_out	.01	.111	44131
Season	.02	.149	44131
Promo	.05307	.162645	44131
AOR	.00799	.017922	44131
CoV_AOR	.36335	.306733	44131
SF	.4671	.07047	44131
High_SF	.72	.447	44131
CU_TU	7.33	6.092	44131
High_SFxCU_TU	4.95	5.502	44131
FG	5.26	3.956	44131

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	FG, SF, Goes_out, CoV_AOR, Season, T_max_del, Promo, CU_TU, FC_period, AOR, High_SFxCU_TU, High_SF[b]	.	Enter

a. Dependent Variable: WasteCU

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics	
					R Square Change	F Change
1	.262[a]	.069	.068	27.856	.069	270.684

Model Summary[b]

Model	Change Statistics			Durbin-Watson
	df1	df2	Sig. F Change	
1	12	44118	.000	1.729

a. Predictors: (Constant), FG, SF, Goes_out, CoV_AOR, Season, T_max_del, Promo, CU_TU, FC_period, AOR, High_SFxCU_TU, High_SF

b. Dependent Variable: WasteCU

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2520537.048	12	210044.754	270.684	.000[b]
	Residual	34234527.03	44118	775.976		
	Total	36755064.08	44130			

a. Dependent Variable: WasteCU

b. Predictors: (Constant), FG, SF, Goes_out, CoV_AOR, Season, T_max_del, Promo, CU_TU, FC_period, AOR, High_SFxCU_TU, High_SF

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	15.897	1.585		10.032	.000
	FC_period	-.057	.003	-.089	-16.545	.000
	T_max_del	1.234	.138	.043	8.915	.000
	Goes_out	8.025	1.200	.031	6.689	.000
	Season	6.100	.900	.032	6.775	.000
	Promo	8.229	.868	.046	9.484	.000
	AOR	132.212	11.420	.082	11.578	.000
	CoV_AOR	12.720	.462	.135	27.513	.000
	SF	-25.616	3.638	-.063	-7.040	.000
	High_SF	-1.542	.572	-.024	-2.695	.007
	CU_TU	-.067	.038	-.014	-1.751	.080
	High_SFxCU_TU	.088	.046	.017	1.907	.057
	FG	-.382	.042	-.052	-9.063	.000

Coefficients[a]

Model		Correlations Zero-order	Correlations Partial	Correlations Part
1	(Constant)			
	FC_period	-.124	-.079	-.076
	T_max_del	.083	.042	.041
	Goes_out	.045	.032	.031
	Season	.026	.032	.031
	Promo	.107	.045	.044
	AOR	.144	.055	.053
	CoV_AOR	.146	.130	.126
	SF	-.132	-.034	-.032
	High_SF	-.112	-.013	-.012
	CU_TU	-.031	-.008	-.008
	High_SFxCU_TU	-.103	.009	.009
	FG	-.105	-.043	-.042

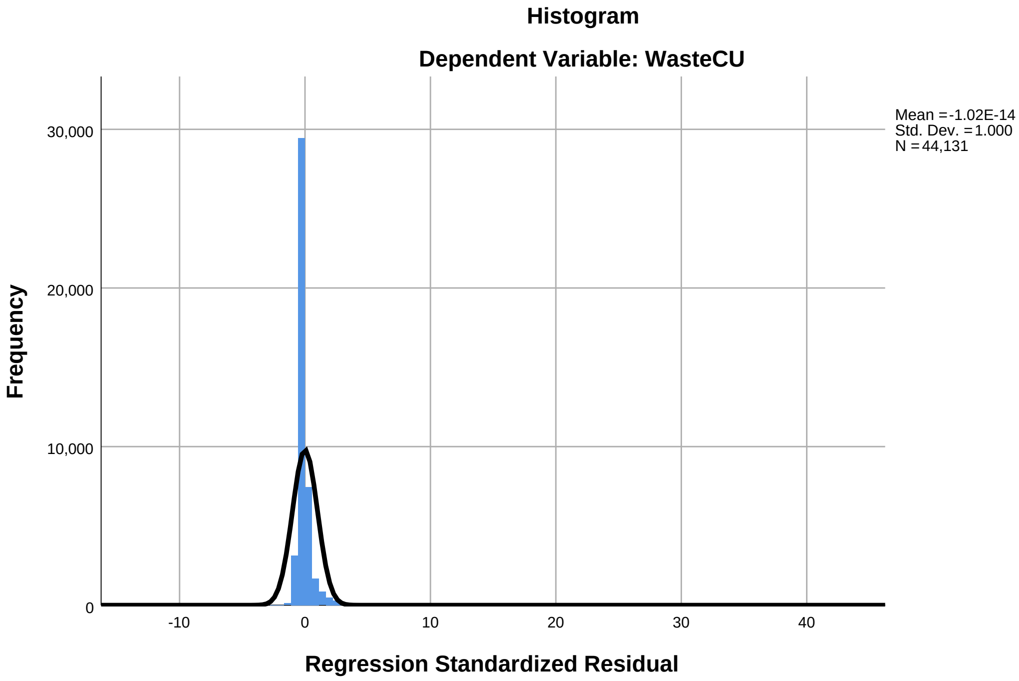
a. Dependent Variable: WasteCU

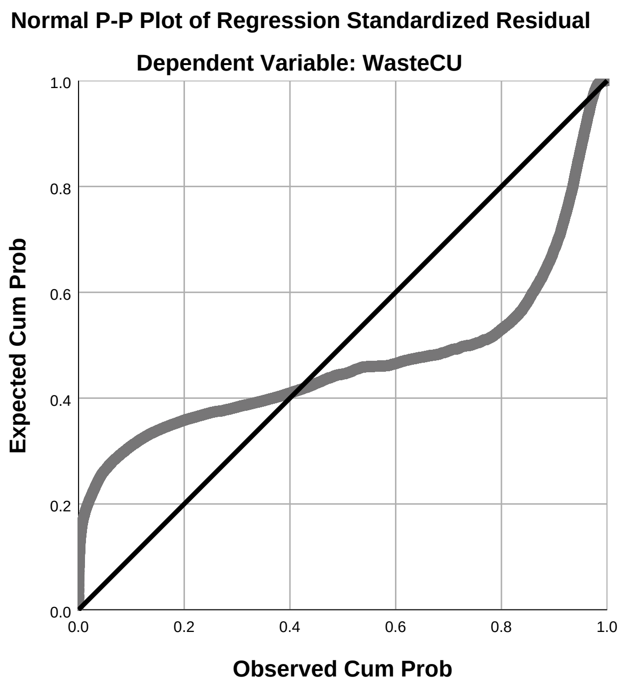
Residuals Statistics[a]

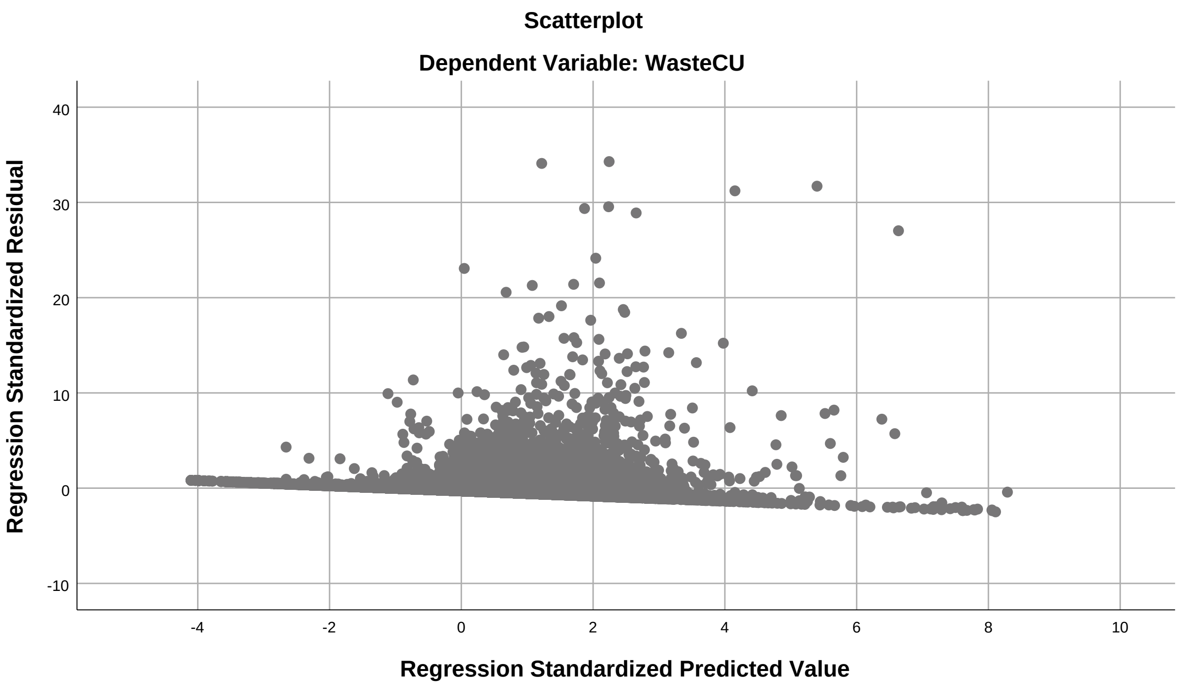
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-22.99	70.65	8.01	7.558	44131
Std. Predicted Value	-4.102	8.289	.000	1.000	44131
Standard Error of Predicted Value	.165	4.616	.403	.257	44131
Adjusted Predicted Value	-23.02	70.98	8.01	7.560	44131
Residual	-69.288	955.039	.000	27.853	44131
Std. Residual	-2.487	34.284	.000	1.000	44131
Stud. Residual	-2.493	34.292	.000	1.000	44131
Deleted Residual	-69.619	955.477	.000	27.880	44131
Stud. Deleted Residual	-2.493	34.758	.000	1.003	44131
Mahal. Distance	.545	1210.636	12.000	31.722	44131
Cook's Distance	.000	1.059	.000	.005	44131
Centered Leverage Value	.000	.027	.000	.001	44131

a. Dependent Variable: WasteCU

Charts








```
 DATASET ACTIVATE DataSet2.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
```

```
  /DEPENDENT WasteCU
  /METHOD=ENTER FC_period T_max_del Goes_out Season Promo AOR CoV_AOR SF Hi
gh_SF CU_TU
    High_SFxCU_TU FG
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /SAVE PRED ZPRED COOK RESID ZRESID.
```

Regression

Notes

Output Created		18-DEC-2018 15:45:29
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\data_no _zeroes.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	15177
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any variable used.

Notes

Syntax		REGRESSION /DESCRIPTIVES MEAN STDDEV CORR SIG N /MISSING LISTWISE /STATISTICS COEFF OUTS CI(95) R ANOVA CHANGE /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT WasteCU /METHOD=ENTER FC_period T_max_del Goes_out Season Promo AOR CoV_AOR SF High_SF CU_TU High_SFxCU_TU FG /SCATTERPLOT=(*ZRESID ,*ZPRED) /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID) /SAVE PRED ZPRED COOK RESID ZRESID.
Resources	Processor Time	00:00:02.00
	Elapsed Time	00:00:01.71
	Memory Required	11472 bytes
	Additional Memory Required for Residual Plots	504 bytes
Variables Created or Modified	PRE_1	Unstandardized Predicted Value
	RES_1	Unstandardized Residual
	ZPR_1	Standardized Predicted Value
	ZRE_1	Standardized Residual
	COO_1	Cook's Distance

```
[DataSet2] C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80
 Waste - JM\Afstuderen\Regression\RegressionSPSS\data_no_zeroes.sav
```

Descriptive Statistics

	Mean	Std. Deviation	N
WasteCU	23.29	45.455	15177
FC_period	16.10	30.077	15177
T_max_del	2.20	.968	15177
Goes_out	.03	.162	15177
Season	.02	.135	15177
Promo	.06874	.182633	15177
AOR	.01286	.025951	15177
CoV_AOR	.49869	.245378	15177
SF	.4491	.08674	15177
High_SF	.61	.487	15177
CU_TU	6.56	6.216	15177
High_SFxCU_TU	3.3050	3.96900	15177
FG	4.5439	3.34650	15177

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	FG, Promo, Goes_out, AOR, Season, T_max_del, High_SFxCU_TU, CoV_AOR, FC_period, CU_TU, SF, High_SF[b]	.	Enter

a. Dependent Variable: WasteCU

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics	
					R Square Change	F Change
1	.241[a]	.058	.058	44.128	.058	78.201

Model Summary[b]

Model	Change Statistics			Durbin-Watson
	df1	df2	Sig. F Change	
1	12	15164	.000	1.760

a. Predictors: (Constant), FG, Promo, Goes_out, AOR, Season, T_max_del, High_SFxCU_TU, CoV_AOR, FC_period, CU_TU, SF, High_SF

b. Dependent Variable: WasteCU

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1827356.888	12	152279.741	78.201	.000[b]
	Residual	29528510.34	15164	1947.277		
	Total	31355867.22	15176			

a. Dependent Variable: WasteCU

b. Predictors: (Constant), FG, Promo, Goes_out, AOR, Season, T_max_del, High_SFxCU_TU, CoV_AOR, FC_period, CU_TU, SF, High_SF

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	35.060	3.380		10.371	.000
	FC_period	-.149	.015	-.099	-10.252	.000
	T_max_del	2.344	.388	.050	6.035	.000
	Goes_out	.712	2.257	.003	.315	.753
	Season	24.377	2.725	.072	8.945	.000
	Promo	23.944	2.314	.096	10.347	.000
	AOR	32.897	21.896	.019	1.502	.133
	CoV_AOR	9.174	1.810	.050	5.068	.000
	SF	-36.192	7.831	-.069	-4.622	.000
	High_SF	-5.938	1.420	-.064	-4.183	.000
	CU_TU	.285	.086	.039	3.295	.001
	High_SFxCU_TU	.491	.144	.043	3.417	.001
	FG	-1.154	.144	-.085	-8.028	.000

Coefficients[a]

Model		95.0% Confidence Interval for B Lower Bound	Upper Bound
1	(Constant)	28.434	41.686
	FC_period	-.178	-.121
	T_max_del	1.583	3.106
	Goes_out	-3.712	5.135
	Season	19.035	29.719
	Promo	19.408	28.480
	AOR	-10.022	75.816
	CoV_AOR	5.626	12.723
	SF	-51.542	-20.842
	High_SF	-8.721	-3.156
	CU_TU	.115	.454
	High_SFxCU_TU	.209	.772
	FG	-1.435	-.872

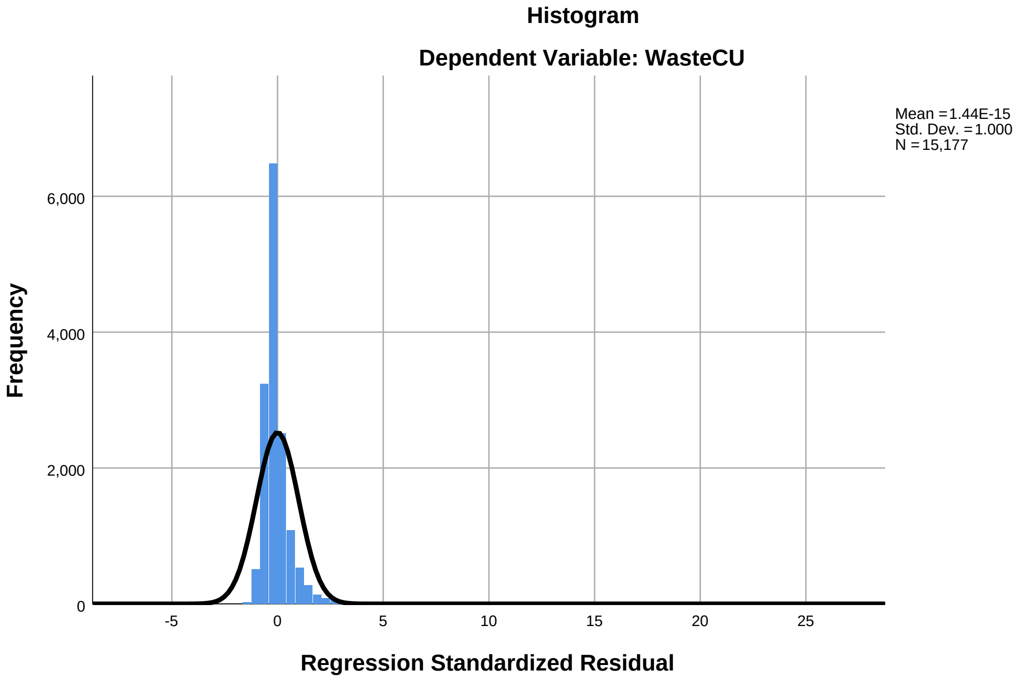
a. Dependent Variable: WasteCU

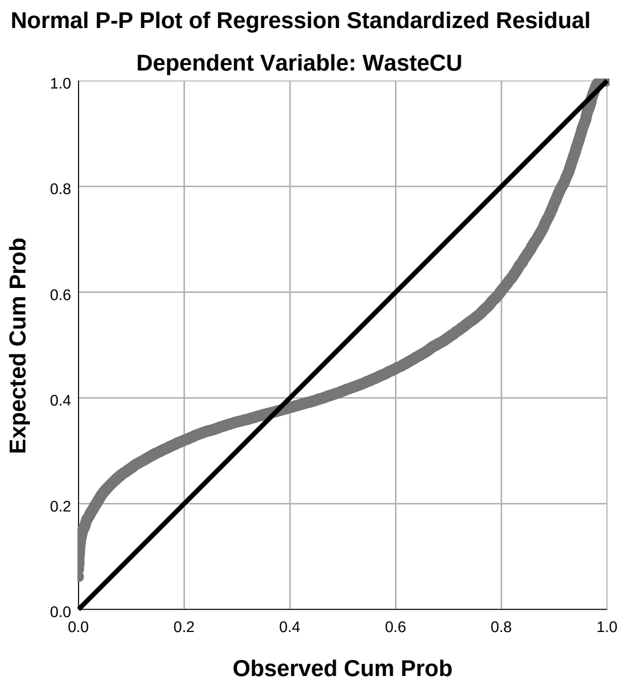
Residuals Statistics[a]

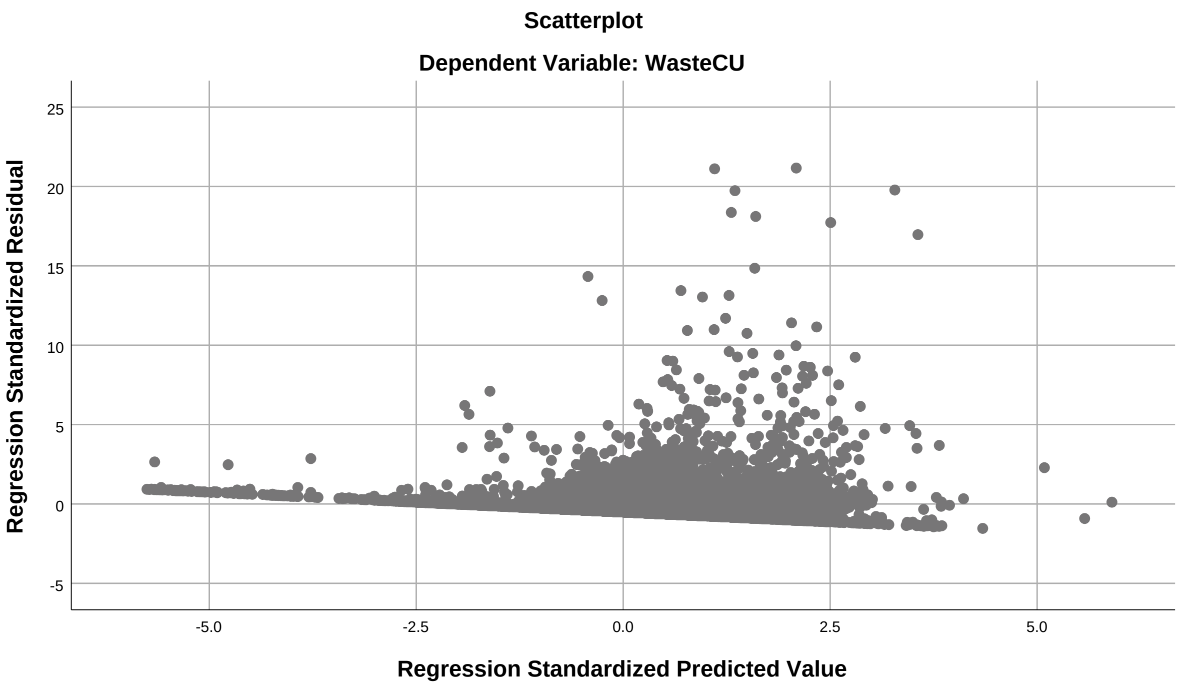
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-39.84	88.07	23.29	10.973	15177
Std. Predicted Value	-5.753	5.903	.000	1.000	15177
Standard Error of Predicted Value	.495	11.301	1.083	.704	15177
Adjusted Predicted Value	-40.12	87.93	23.28	10.985	15177
Residual	-67.938	933.775	.000	44.111	15177
Std. Residual	-1.540	21.161	.000	1.000	15177
Stud. Residual	-1.557	21.170	.000	1.001	15177
Deleted Residual	-69.448	934.595	.001	44.194	15177
Stud. Deleted Residual	-1.557	21.489	.000	1.004	15177
Mahal. Distance	.913	994.258	11.999	28.962	15177
Cook's Distance	.000	.522	.000	.005	15177
Centered Leverage Value	.000	.066	.001	.002	15177

a. Dependent Variable: WasteCU

Charts


Histogram
Dependent Variable: WasteCU
Mean =1.44E-15
Std. Dev. =1.000
N =15,177
Frequency
6,000
4,000
2,000
0
-5
0
5
10
15
20
25
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: WasteCU
Expected Cum Prob
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob




```
 DATASET ACTIVATE DataSet1.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT WasteCU
  /METHOD=ENTER FC_period T_max_del Goes_out Season Promo AOR CoV_AOR SF Hi
gh_SF CU_TU
    High_SFxCU_TU FG
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*ZRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /SAVE PRED ZPRED COOK RESID ZRESID.


 * Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=FC_period WasteCU MISSING=LIS
TWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE
  /FITLINE TOTAL=YES.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
```

```
  DATA: FC_period=col(source(s), name("FC_period"))
  DATA: WasteCU=col(source(s), name("WasteCU"), unit.category())
  GUIDE: axis(dim(1), label("FC_period"))
  GUIDE: axis(dim(2), label("WasteCU"))
  GUIDE: text.title(label("Simple Scatter with Fit Line of WasteCU by FC_pe
riod"))
  ELEMENT: point(position(FC_period*WasteCU))
END GPL.
```

GGraph

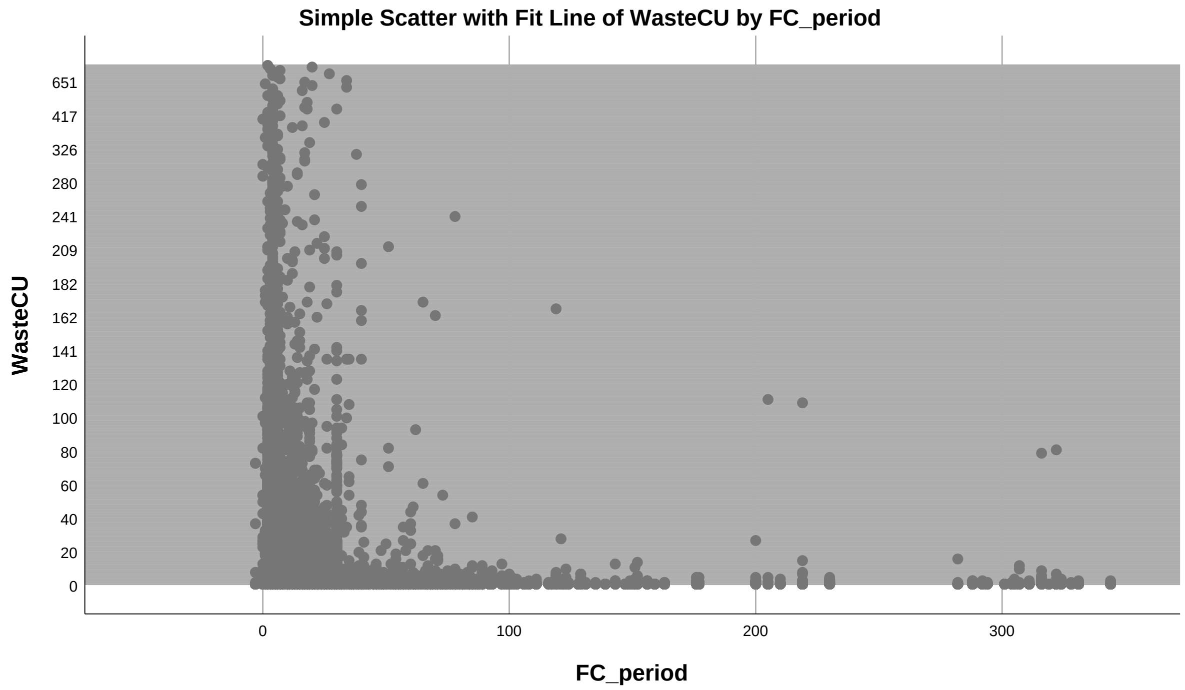
Notes

Output Created		19-DEC-2018 20:03:29
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\full_data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	44131

Notes

Syntax		GGRAPH
		  /GRAPHDATASET NAME="graphdataset" VARIABLES=FC_period WasteCU MISSING=LISTWISE REPORTMISSING=NO
		  /GRAPHSPEC SOURCE=INLINE
		  /FITLINE TOTAL=YES.
		BEGIN GPL
		  SOURCE: s=userSource(id("graphdataset"))
		  DATA: FC_period=col(source(s), name("FC_period"))
		  DATA: WasteCU=col(source(s), name("WasteCU"), unit.category())
		  GUIDE: axis(dim(1), label("FC_period"))
		  GUIDE: axis(dim(2), label("WasteCU"))
		  GUIDE: text.title(label("Simple Scatter with Fit Line of WasteCU by FC_period"))
		  ELEMENT: point(position(FC_period*WasteCU))
		END GPL.

Resources	Processor Time	00:00:00.66
	Elapsed Time	00:00:00.40




```
DESCRIPTIVES VARIABLES=WasteCU
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptives

Notes

Output Created		19-DEC-2018 21:36:24
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\full_data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	44131
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	All non-missing data are used.
Syntax		DESCRIPTIVES VARIABLES=WasteCU /STATISTICS=MEAN STDDEV MIN MAX.
Resources	Processor Time	00:00:00.05
	Elapsed Time	00:00:00.05

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
WasteCU	44131	0	980	8.01	28.860
Valid N (listwise)	44131				

```
DATASET ACTIVATE DataSet2.
FREQUENCIES VARIABLES=WasteCU
  /ORDER=ANALYSIS.
```

Frequencies

Notes

Output Created		19-DEC-2018 21:36:42
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\data_no _zeroes.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	15177
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data.
Syntax		FREQUENCIES VARIABLES=WasteCU /ORDER=ANALYSIS.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.03

```
[DataSet2] C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80
 Waste - JM\Afstuderen\Regression\RegressionSPSS\data_no_zeroes.sav
```

```
DESCRIPTIVES VARIABLES=WasteCU
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptives

Notes

Output Created		19-DEC-2018 21:36:55
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\data_no _zeroes.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	15177
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	All non-missing data are used.
Syntax		DESCRIPTIVES VARIABLES=WasteCU /STATISTICS=MEAN STDDEV MIN MAX.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.02

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
WasteCU	15177	1	980	23.29	45.455
Valid N (listwise)	15177				

```
 DATASET ACTIVATE DataSet1.
DATASET ACTIVATE DataSet1.

SAVE OUTFILE='C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)
\80 Waste - '+
    'JM\Afstuderen\Regression\RegressionSPSS\full_data.sav'
  /COMPRESSED.
DATASET ACTIVATE DataSet2.
DATASET CLOSE DataSet1.
```

F
Binomial negative regression results

```
 DATASET ACTIVATE DataSet1.
* Generalized Linear Models.
GENLIN WasteCU BY Goes_out Season High_SF (ORDER=ASCENDING) WITH FC_period
T_max_del Promo AOR
    CoV_AOR SF CU_TU High_SFxCU_TU FG
  /MODEL Goes_out Season High_SF FC_period T_max_del Promo AOR CoV_AOR SF C
U_TU High_SFxCU_TU FG
    INTERCEPT=YES
 DISTRIBUTION=NEGBIN(MLE) LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100 MAXSTEPHA
LVING=5
    PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD) CILEVEL
=95 CITYPE=WALD
    LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION (EXPONENTIATED)
  /SAVE MEANPRED XBPRED COOK RESID DEVIANCERESID.
```

Generalized Linear Models

Notes

Output Created		18-DEC-2018 15:48:16
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\full_data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	44131
Missing Value Handling	Definition of Missing	User-defined missing values for factor, subject and within-subject variables are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables in the model.
Weight Handling		not applicable

Notes

Syntax		GENLIN WasteCU BY Goes_out Season High_SF (ORDER=ASCENDING) WITH FC_period T_max_del Promo AOR CoV_AOR SF CU_TU High_SFxCU_TU FG /MODEL Goes_out Season High_SF FC_period T_max_del Promo AOR CoV_AOR SF CU_TU High_SFxCU_TU FG INTERCEPT=YES DISTRIBUTION=NEGBIN (MLE) LINK=LOG /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100 MAXSTEPHALVING=5 PCONVERGE=1E-006 (ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3 (WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FULL /MISSING CLASSMISSING=EXCLUDE /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION (EXPONENTIATED) /SAVE MEANPRED XBPRED COOK RESID DEVIANCERESID.
Resources	Processor Time	00:00:08.39
	Elapsed Time	00:00:08.61
Variables Created or Modified	Predicted Value of the Linear Predictor	XBPredicted
	Predicted Value of the Mean of the Response	MeanPredicted
	Raw Residual	Residual
	Deviance Residual	DevianceResidual
	Cook's Distance	CooksDistance

```
[DataSet1] C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80
 Waste - JM\Afstuderen\Regression\RegressionSPSS\full_data.sav
```

Model Information

Dependent Variable	WasteCU
Probability Distribution	Negative binomial (MLE)
Link Function	Log

Case Processing Summary

	N	Percent
Included	44131	100.0%
Excluded	0	0.0%
Total	44131	100.0%

Categorical Variable Information

			N	Percent
Factor	Goes_out	0	43576	98.7%
		1	555	1.3%
		Total	44131	100.0%
	Season	0	43123	97.7%
		1	1008	2.3%
		Total	44131	100.0%
	High_SF	0	12184	27.6%
		1	31947	72.4%
		Total	44131	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean
Dependent Variable	WasteCU	44131	0	980	8.01
Covariate	FC_period	44131	-3	344	28.74
	T_max_del	44131	1	7	2.05
	Promo	44131	.000	.944	.05307
	AOR	44131	.000	.482	.00799
	CoV_AOR	44131	.000	4.743	.36335
	SF	44131	.20	.50	.4671
	CU_TU	44131	1	92	7.33
	High_SFxCU_TU	44131	0	92	4.95
	FG	44131	0	21	5.26

Continuous Variable Information

		Std. Deviation
Dependent Variable	WasteCU	28.860
Covariate	FC_period	45.156
	T_max_del	1.005
	Promo	.162645
	AOR	.017922
	CoV_AOR	.306733
	SF	.07047
	CU_TU	6.092
	High_SFxCU_TU	5.502
	FG	3.956

Goodness of Fit[a]

	Value	df	Value/df
Deviance	30434.697	44117	.690
Scaled Deviance	30434.697	44117	
Pearson Chi-Square	461044.626	44117	10.450
Scaled Pearson Chi-Square	461044.626	44117	
Log Likelihood[b]	-83545.166		
Akaike's Information Criterion (AIC)	167118.332		
Finite Sample Corrected AIC (AICC)	167118.341		
Bayesian Information Criterion (BIC)	167240.061		
Consistent AIC (CAIC)	167254.061		

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG[a]

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
7391.434	12	.000

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG[a]

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	70.629	1	.000
Goes_out	69.146	1	.000
Season	1.223	1	.269
High_SF	112.572	1	.000
FC_period	1000.917	1	.000
T_max_del	214.996	1	.000
Promo	155.237	1	.000
AOR	74.464	1	.000
CoV_AOR	3989.784	1	.000
SF	61.943	1	.000
CU_TU	15.254	1	.000
High_SFxCU_TU	4.745	1	.029
FG	179.471	1	.000

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test
			Lower	Upper	Wald Chi-Square
(Intercept)	2.187	.2668	1.664	2.710	67.211
[Goes_out=0]	-1.029	.1237	-1.271	-.786	69.146
[Goes_out=1]	0[a]	.	.	.	.
[Season=0]	-.102	.0924	-.283	.079	1.223
[Season=1]	0[a]	.	.	.	.
[High_SF=0]	.591	.0557	.482	.700	112.572
[High_SF=1]	0[a]	.	.	.	.
FC_period	-.012	.0004	-.013	-.011	1000.917
T_max_del	.235	.0160	.203	.266	214.996
Promo	-1.125	.0903	-1.302	-.948	155.237
AOR	17.911	2.0756	13.843	21.979	74.464
CoV_AOR	4.655	.0737	4.511	4.800	3989.784
SF	-3.492	.4437	-4.362	-2.622	61.943
CU_TU	.018	.0047	.009	.027	15.254
High_SFxCU_TU	-.011	.0049	-.020	-.001	4.745
FG	-.062	.0046	-.071	-.053	179.471
(Scale)	1[b]				
(Negative binomial)	7.498	.0746	7.353	7.646	

Parameter Estimates

Parameter	Hypothesis Test			95% Wald Confidence Interval for Exp(B)	
	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	1	.000	8.911	5.282	15.033
[Goes_out=0]	1	.000	.357	.280	.455
[Goes_out=1]	.	.	1	.	.
[Season=0]	1	.269	.903	.753	1.082
[Season=1]	.	.	1	.	.
[High_SF=0]	1	.000	1.806	1.619	2.014
[High_SF=1]	.	.	1	.	.
FC_period	1	.000	.988	.987	.989
T_max_del	1	.000	1.264	1.225	1.305
Promo	1	.000	.325	.272	.387
AOR	1	.000	60062018.48	1027646.820	3510394809
CoV_AOR	1	.000	105.118	90.980	121.453
SF	1	.000	.030	.013	.073
CU_TU	1	.000	1.018	1.009	1.028
High_SFxCU_TU	1	.029	.989	.980	.999
FG	1	.000	.940	.932	.949
(Scale)					
(Negative binomial)					

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

```
 DATASET ACTIVATE DataSet2.
* Generalized Linear Models.
GENLIN WasteCU BY Goes_out Season High_SF (ORDER=ASCENDING) WITH FC_period
T_max_del Promo AOR
    CoV_AOR SF CU_TU High_SFxCU_TU FG
  /MODEL Goes_out Season High_SF FC_period T_max_del Promo AOR CoV_AOR SF C
U_TU High_SFxCU_TU FG
    INTERCEPT=YES
 DISTRIBUTION=NEGBIN(MLE) LINK=LOG
  /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100 MAXSTEPHA
LVING=5
    PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD) CILEVEL
```

```
=95 CITYPE=WALD
    LIKELIHOOD=FULL
  /MISSING CLASSMISSING=EXCLUDE
  /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION (EXPONENTIATED)
  /SAVE MEANPRED XBPRED COOK RESID DEVIANCERESID.
```

Generalized Linear Models

Notes

Output Created		18-DEC-2018 15:50:46
Comments		
Input	Data	C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80 Waste - JM\Afstuderen\Regression \RegressionSPSS\data_no _zeroes.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	15177
Missing Value Handling	Definition of Missing	User-defined missing values for factor, subject and within-subject variables are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables in the model.
Weight Handling		not applicable

Notes

Syntax		GENLIN WasteCU BY Goes_out Season High_SF (ORDER=ASCENDING) WITH FC_period T_max_del Promo AOR CoV_AOR SF CU_TU High_SFxCU_TU FG /MODEL Goes_out Season High_SF FC_period T_max_del Promo AOR CoV_AOR SF CU_TU High_SFxCU_TU FG INTERCEPT=YES DISTRIBUTION=NEGBIN (MLE) LINK=LOG /CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100 MAXSTEPHALVING=5 PCONVERGE=1E-006 (ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3 (WALD) CILEVEL=95 CITYPE=WALD LIKELIHOOD=FULL /MISSING CLASSMISSING=EXCLUDE /PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION (EXPONENTIATED) /SAVE MEANPRED XBPRED COOK RESID DEVIANCERESID.
Resources	Processor Time	00:00:02.11
	Elapsed Time	00:00:02.17
Variables Created or Modified	Predicted Value of the Linear Predictor	XBPredicted
	Predicted Value of the Mean of the Response	MeanPredicted
	Raw Residual	Residual
	Deviance Residual	DevianceResidual
	Cook's Distance	CooksDistance

```
[DataSet2] C:\Users\Jurriaan\Dropbox (Picnic)\100 fc intern projects (1)\80
 Waste - JM\Afstuderen\Regression\RegressionSPSS\data_no_zeroes.sav
```

Model Information

Dependent Variable	WasteCU
Probability Distribution	Negative binomial (MLE)
Link Function	Log

Case Processing Summary

	N	Percent
Included	15177	100.0%
Excluded	0	0.0%
Total	15177	100.0%

Categorical Variable Information

			N	Percent
Factor	Goes_out	0	14769	97.3%
		1	408	2.7%
		Total	15177	100.0%
	Season	0	14897	98.2%
		1	280	1.8%
		Total	15177	100.0%
	High_SF	0	5860	38.6%
		1	9317	61.4%
		Total	15177	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean
Dependent Variable	WasteCU	15177	1	980	23.29
Covariate	FC_period	15177	-3	344	16.10
	T_max_del	15177	1	7	2.20
	Promo	15177	.000	.942	.06874
	AOR	15177	.000	.482	.01286
	CoV_AOR	15177	.000	4.290	.49869
	SF	15177	.20	.50	.4491
	CU_TU	15177	1	92	6.56
	High_SFxCU_TU	15177	.00	92.00	3.3050
	FG	15177	.00	21.00	4.5439

Continuous Variable Information

		Std. Deviation
Dependent Variable	WasteCU	45.455
Covariate	FC_period	30.077
	T_max_del	.968
	Promo	.182633
	AOR	.025951
	CoV_AOR	.245378
	SF	.08674
	CU_TU	6.216
	High_SFxCU_TU	3.96900
	FG	3.34650

Goodness of Fit[a]

	Value	df	Value/df
Deviance	17408.320	15163	1.148
Scaled Deviance	17408.320	15163	
Pearson Chi-Square	36912.788	15163	2.434
Scaled Pearson Chi-Square	36912.788	15163	
Log Likelihood[b]	-60871.662		
Akaike's Information Criterion (AIC)	121771.325		
Finite Sample Corrected AIC (AICC)	121771.353		
Bayesian Information Criterion (BIC)	121878.110		
Consistent AIC (CAIC)	121892.110		

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG[a]

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
2294.524	12	.000

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG[a]

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	1102.868	1	.000
Goes_out	.029	1	.865
Season	105.085	1	.000
High_SF	21.160	1	.000
FC_period	556.214	1	.000
T_max_del	117.347	1	.000
Promo	174.269	1	.000
AOR	.040	1	.842
CoV_AOR	30.311	1	.000
SF	41.150	1	.000
CU_TU	38.116	1	.000
High_SFxCU_TU	12.154	1	.000
FG	214.346	1	.000

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test
			Lower	Upper	Wald Chi-Square
(Intercept)	4.108	.1466	3.821	4.396	784.959
[Goes_out=0]	.011	.0623	-.111	.133	.029
[Goes_out=1]	0[a]	.	.	.	.
[Season=0]	-.764	.0745	-.910	-.618	105.085
[Season=1]	0[a]	.	.	.	.
[High_SF=0]	.185	.0403	.106	.264	21.160
[High_SF=1]	0[a]	.	.	.	.
FC_period	-.008	.0004	-.009	-.008	556.214
T_max_del	.122	.0112	.100	.144	117.347
Promo	.847	.0642	.721	.973	174.269
AOR	.128	.6457	-1.137	1.394	.040
CoV_AOR	.280	.0509	.181	.380	30.311
SF	-1.425	.2222	-1.861	-.990	41.150
CU_TU	.018	.0029	.012	.023	38.116
High_SFxCU_TU	.016	.0045	.007	.024	12.154
FG	-.061	.0042	-.069	-.053	214.346
(Scale)	1[b]				
(Negative binomial)	1.445	.0153	1.415	1.475	

Parameter Estimates

Parameter	Hypothesis Test df	Sig.	Exp(B)	95% Wald Confidence Interval for Exp(B) Lower	Upper
(Intercept)	1	.000	60.833	45.638	81.087
[Goes_out=0]	1	.865	1.011	.895	1.142
[Goes_out=1]	.	.	1	.	.
[Season=0]	1	.000	.466	.403	.539
[Season=1]	.	.	1	.	.
[High_SF=0]	1	.000	1.204	1.112	1.302
[High_SF=1]	.	.	1	.	.
FC_period	1	.000	.992	.991	.992
T_max_del	1	.000	1.129	1.105	1.155
Promo	1	.000	2.332	2.057	2.645
AOR	1	.842	1.137	.321	4.031
CoV_AOR	1	.000	1.324	1.198	1.463
SF	1	.000	.240	.156	.372
CU_TU	1	.000	1.018	1.012	1.024
High_SFxCU_TU	1	.000	1.016	1.007	1.025
FG	1	.000	.941	.933	.949
(Scale)					
(Negative binomial)					

Dependent Variable: WasteCU
Model: (Intercept), Goes_out, Season, High_SF, FC_period, T_max_del, Promo, AOR, CoV_AOR, SF, CU_TU, High_SFxCU_TU, FG

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.