MASTER

Improving retail order forecasts by predicting competitive promotions

van Doorn, G.

Award date:
2019

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Industrial Engineering and Innovation Sciences
Operations Planning Accounting and Control Research Group

Improving Retail Order Forecasts by Predicting Competitive Promotions

Master Thesis

Gwen van Doorn

Supervisors:
dr. A.E. Akçay, TU/e
dr. K.H. van Donselaar, TU/e
dr. M. Firat, TU/e
Y.B.N. Suurmeijer MSc, Bright Cape
P.L.P. Meeuws MSc, Swinkels Family Brewers

In partial fulfillment of the requirements for the degree of
Master of Science in Operations Management and Logistics

Eindhoven, October, 2019

Eindhoven University of Technology
School of Industrial Engineering
Series Master Theses Operations Management and Logistics

Keywords: promotion, competitor, promotion prediction, manufacturer, sales forecasting, machine learning, multi-label classification

Abstract

Previous literature showed that incorporating promotions of related products is beneficial to the forecasting performance. This information is, however, only known to retailers and not to manufacturers. Therefore, this research aims to give insight into the predictability of promotions of competitors and addresses the practical value of predicted promotions by incorporating them into the retail order forecast of a manufacturer in the beer sector. In this research a competitive promotion prediction model is created. The input variables of the model are based on insights derived from a visualization analysis performed on historic promotion data. The competitive promotion prediction model is approached as a multi-label classification problem. A decision tree applied with the label power-set method is able to predict 43% of the promotions correctly and 46% of the predicted promotions is also a promotion in reality. The predictions from the competitive promotion prediction model are incorporated in a retail order forecast model which is modeled with a decision tree. This model does not give substantiation about the practical value of predicted promotions but shows excellent results compared to the current forecasting method of the manufacturer.

Executive summary

More accurate sales forecasts have the potential to decrease safety stock and thereby reducing costs while maintaining the level of service. Previous literature showed that incorporating promotions of related products is beneficial to the forecasting accuracy (Kuo, 2001; Divakar, Ratchford & Shankar, 2005; Ali, Sayın, Van Woensel & Fransoo, 2009; Huang, Fildes & Soopramanien, 2014; Ma, Fildes & Huang, 2016; Ma & Fildes, 2017). The promotions of competitors are, however, only known to retailers and not to manufacturers. Therefore, this research aimed to improve retail order forecast accuracy of Swinkels Family Brewers by gaining insight into the predictability of promotions of competitors. This aim was reached by answering the following research question:

Are the promotions of competitors predictable and do predicted promotions of competitors improve retail order forecast accuracy?

Research design
To answer the research question, first the promotion behavior of competitors was examined. This was done by analyzing historic promotion data and extracting patterns from that. Thereafter, competitive promotions were predicted by means of insights from the competitive promotion analysis. Finally, these predictions were incorporated in the retail order forecasts. This approach is visualized in Figure 1.

Figure 1: General research approach

Competitive promotion analysis
The competitive promotion analysis was performed on five product groups: pilsner crates, pilsner other, 0.0%, fresh and characterful. The analysis was supported by various visualization techniques and the results are shown in Figure 2. For all groups hold that certain brands are often on promotion in the same week. This especially applies to brand which are brewed by the same company. For all but the 0.0% group hold that promotions are related to national holidays. Furthermore, a certain degree of cyclicity is present in the promotions. In the fresh group and characterful group also seasonality plays a role. These findings apply most strongly to the biggest brands and the biggest product groups. These products and brands generate the highest revenue for a retailer and therefore have more to say in the supermarket's promotion calendar.

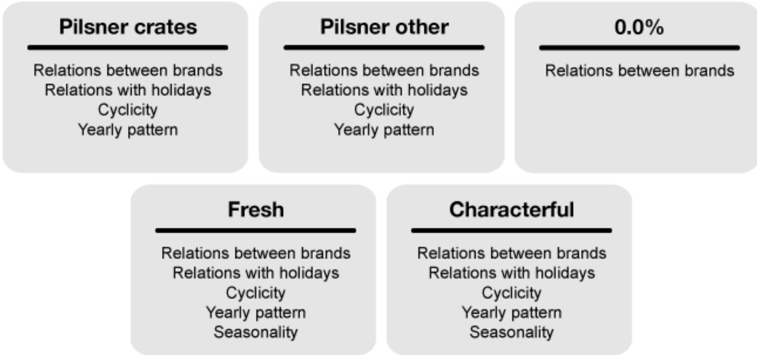

Figure 2: Insights competitive promotion analysis

Competitive promotion prediction model

The competitive promotion prediction model aims to predict promotions of competitive brands. The model was approached as a multi-label classification problem in which for every week it is predicted whether product groups of competitive brands have a promotion or not. The product groups of competitive brands are referred to as 'competitive brand-groups' in the remainder of this summary. The input variables for the competitive promotion prediction model were derived from the insights of the competitive promotion analysis. The input variables include a variable indicating the week number, a variable indicating the number of the week in the data set, a variable indicating whether there is a holiday in the predicted week, variables indicating the number of weeks since the last promotion of competitive brands and variables indicating whether Bavaria is on promotion since then it is less likely that another brand is on promotion as well.

Several methods which are able to deal with multi-label classification problems were applied to a decision tree and a random forest. These methods are the binary relevance method, the label power-set method, the ensemble of classifier chains an the algorithm adaptation. Furthermore, two benchmark models were created to compare the performance with. One benchmark is a random prediction which takes into account the amount of promotions of a competitive brand-group. The other benchmark is a prediction which is based on the probability distribution of the number of weeks between promotions of a competitive brand-group.

The label power-set approach together with a decision tree results in a model which is particularly suitable for predicting competitive promotions of the pilsner groups at Albert Heijn. The created model outperforms both benchmarks on four out of five performance measures. Hereby the model is able to predict 43% of the promotions correctly and 46% of the predicted promotions is also a promotion in reality.

Retail order forecast model

The aim of the retail order forecast model is to forecast the orders of retailers as precise as possible. Furthermore, the retail order forecast model should address whether the performance of the competitive promotion prediction model is good enough to provide practical value. That is, is the performance of the competitive promotion prediction model sufficient to improve retail order forecast accuracy. The retail order forecast model only provides forecasts for the weeks around a promotions. In the other weeks (i.e. the baseline weeks) the forecasts from the current method are used.

The retail order forecast was modeled with a decision tree and with three different sets of input variables. The first set consist only of promotional data of the forecasted product such as the promotion mechanism, whether the product is placed on second placement and whether the promotion is during a holiday. The second set consists of promotional data of the forecasted product and actual promotions of competitors. The third set consists of promotional data of the forecasted product and predicted promotions of competitors provided by the competitive promotion prediction model. The current forecasting method is used as a benchmark.

The retail order forecast model based only on data of the forecasted product showed a significant improvement compared to the current forecasting method on pilsner products at Albert Heijn. The performance improvement amounts to an improvement of 20.5% overall and 58.7% in the case of a promotion. The incorporation of the predicted competitive promotions into the retail order forecast model did not result in a greater improvement. However, incorporating actual competitive promotions has not led to a greater improvement either. The practical value of the predictions can therefore not be assessed. An explanation as to why the forecast performance of the retail order forecast model does not improve by incorporating promotions of competitors is the limited amount of available data. The model with the best performing set of input variables was also applied to other products and other supermarkets and also here it outperformed the current forecasting method which shows the practical potential of the model.

Conclusions and recommendations

Despite the fact that there is no exclusion about the practical value of predicted competitive promotions, the retail order forecast model showed excellent results compared to the current forecasting method. Storing and using historic data is thus a major added value in forecasting. Besides the practical importance, this study also contributes to literature. While existing studies on forecasting with competitive promotions are researched from the retailer's point of view this study is the first to study that from a manufacturer's point of view. Therefore, this study showed how to take into account competitive promotions despite the fact that this information is unknown to a manufacturer. The created competitive promotion prediction model enables manufacturers to generate predictions of promotions of competitors and incorporate them in the retail order forecast model.

Due to the limited amount of available data the practical value of the competitive promotion prediction is unexplored. More research is desired in this field. Furthermore, due to the innovative nature of this research, no benchmark from practice or literature was available for comparison. Therefore, future research is challenged to improve the competitive promotion prediction model by using this study's results as a benchmark. As soon as there is a definite and positive answer about the practical value of the competitive promotion predictions, Swinkels Family Brewers can take steps to implement the competitive promotion prediction model in practice.

On the other hand, Swinkels Family Brewers is recommended to instantaneously implement the retail order forecast model based only on data of the forecasted product. High data quality is a condition for using the retail order forecast model in practice. To ensure data quality it is recommended to Swinkels Family Brewers to minimize data processing efforts by using fixed data formatting and storing all data from the same product under the same product number. Even with renewed packaging, the product should be stored under the original product number. Furthermore, Swinkels Family brewers is recommended to collect more data to improve the retail order forecast model. More specifically, record the percentage discount given to the end consumer next to the promotion mechanism.

Preface

This thesis is written in order finalize the master Operations Management and Logistics at the Eindhoven University of Technology. This also means that my student time has come to an end. I would like to take this opportunity to thank the people who have supported, motivated and inspired me in these years.

First of all, I would like to thank my supervisor Alp Akçay for the guidance during this project. Your critical view has helped me to never loose sight of the goal and your questions assured me in making well-motivated decisions. Additionally, I would like to thank Karel van Donselaar as second supervisor. Your experience and knowledge of this field of research were of great value.

Next, I want to thank my supervisors from Bright Cape and Swinkels Family Brewers for offering me the opportunity to conduct this research. Youp, thank you for giving me freedom and confidence at the same time in finding my way through the project. Especially your knowledge on the more technical aspects of the project was very helpful. You were always willing to give feedback, even from across the ocean. Pieter, thank you for your business insights and getting me acquainted in the company. I greatly appreciate your involvement and commitment, without your efforts I would not have been able to gather all the required data. Furthermore, I would like to thank my colleagues for creating a fun and pleasant working atmosphere.

As I said before, this project also marks the end of my student time. A time in which I was able to develop myself personally and in which I experienced great memories and experiences. My study would definitely not have been as fun without the girls from Formidabel. Thanks to my fellow board members from UniPartners Nederland for the amazing year in which we achieved a lot together. And my field hockey team, de Assepushters, thank you for all the good times. I would not have missed it for the world!

Finally, my family deserves a special thanks. To my parents and brother, thank you for your unconditional support and being there for me during every step of the way. And last but certainly not least, Fin, thank you for always offering a listening ear even though you probably did not always have a clue what I was talking about. You cheered me up in stressful situations and encouraged me to stay positive. I deeply appreciate your support and love in anything I do.

Gwen van Doorn

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Often manufacturers forecast retail orders using historic order data (Williams & Waller, 2010). However, the increase of point-of-sales (POS) sharing between a retailer and its suppliers (Bourland, Powell & Pyke, 1996; Gavirneni, Kapuscinski & Tayur, 1999; Cachon & Fisher, 2000; Lee, So & Tang, 2000; Raghunathan, 2001; Kulp, Lee & Ofek, 2004) enables more and more suppliers and manufacturers to use POS data for forecasting. This increase in POS sharing has raised a new question in the forecasting domain, namely, does the use of POS data in addition to historic order data increase the forecast accuracy (Jiang, Zhong & Klein, 2000; Williams & Waller, 2010; Choi, Hui, Ng & Yu, 2012).

In addition to the consideration between POS and order data, a major difficulty in forecasting is the domination of price promotions on the retail shop floor. A price promotion is a temporary reduction in the price offered to customers. A price promotion can be induced directly by a retailer or indirectly by a manufacturer. An important driver of the sales increase during a promotion is the brand switching behavior of customers as a result of the temporary price reduction (Gupta, 1988). Van Heerde, Gupta and Wittink (2003) showed that when you account for an increase in category volume due to a promotion, which is also partly beneficial to the non-promoted brands, 33% of the sales elasticity is due to brand switching. In other words, if a promoted brand gains 100 units, the non-promoted brands together lose 33 units. Price promotions do not only cause brand switching, but also forward buying behavior and those two together make sales forecasting during a promotion a complex task (Ailawadi, Harlam, Cesar & Trounce, 2006). Another complexity is that promotions of competing products influence the sales, which increases the difficulty of sales forecasting even more (Walters, 1991; Struse III, 1987; Walters, 1991; Kamakura & Kang, 2007).

This report presents a research concerning the improvement of retail order forecasts of manufacturers. The remainder of this chapter is organized as follows. In section 1.1 related literature is reviewed. Section 1.2 states the research objective and research questions. Section 1.3 sets the scope of the research and in section 1.4 the methodology of the research is described.

1.1 Literature review

In this section literature about forecasting in a manufacturing company is reviewed. The literature is divided into three topics. The first topic is about whether a manufacturer should use POS data or order data in forecasting. The second topic is about the value of incorporating promotions from competitive products in the forecast. Lastly, literature about competitive promotion behavior is reviewed.

1.1.1 Forecasting based on point of sales data or order data

Preliminary research about the best data source for sales forecasting suggests, without empirical evidence, that POS data is the best source at all stages of the supply chain, since it captures independent consumer demand.(Kiely, 1998; Lapide, 2005; Holmström, Främling, Kaipia & Saranen, 2002). For the past decade, research has focused on empirically substantiating this. Williams and Waller (2010) empirically investigated whether retail suppliers should use historic order data or POS data to forecast orders from retailers. They use data from three different product categories to compare simple statistical models using historic order data and POS data. One of the product categories has a highly seasonal demand and one has a short shelf-life. They found that, for a four week forecasting horizon, at least one of the two error measures decreases for all product categories by using POS data rather than order data. When the forecast horizon is extended the seasonal and short shelf-life products showed no statistical results in favor of POS data anymore. They argue that this may be attributed to product characteristics such as seasonality. But they conclude that using POS data, in general, outperforms order data.

Building on these results, Williams, Waller, Ahire and Ferrier (2014) propose a new method to reduce the error in the forecast of retail orders. This new method uses both POS data and historic order data and it accounts for the correction which a retailer makes for an earlier forecast error. Due to this earlier forecast error the retailer will adjust the amount of the next order to compensate for the imbalance in the inventory. Only using POS data or only using order data does not take into account this behavior in the order forecast. Williams et al. (2014) compare the performance of using both POS and order data to the performance of using only one of them. The authors showed that a combination of both data sources outperforms a single data source. Furthermore, using only POS data gives a lower forecast error compared to only using order data.

1.1.2 Incorporating competitive promotions in forecasting

The effect of promotions on sales is a widely researched topic in literature (Chakraborty, Mehrotra, Mohan & Ranka, 1992; Lachtermacher & Fuller, 1995; Kumar, Rao & Soni, 1995; Agrawal & Schorling, 1996). According to Kuo (2001) that effect was simplified in previous research, since the effect of promotions is not so straight forward but instead more fuzzy. Therefore, his study proposed a fuzzy neural network which is initiated by a genetic algorithm (GFNN). In addition, the fuzzy promotion of related products is taken into account. Their results showed that their model had a higher accuracy than a statistical method not incorporating the promotions of related products. However, it was only taken into account whether a related product is on promotion.

On the contrary, Divakar et al. (2005) took, among others, competitor's promotion prices and variables into account in the sales forecast. They performed their research on the brand level and at the SKU (Stock Keeping Unit) level intra-brand competition can exist. Therefore, Ali et al. (2009) researched sales forecasting at the SKU level. Ali et al. (2009) explored the trade-off between model complexity and the corresponding data preparation costs and forecasting accuracy in forecasting retail sales in the presence of promotions. This was done by varying the forecasting technique from simple time series models to sophisticated machine learning models such as support vector machine and regression trees and by varying the input variables from historical sales and promotion data of the concerned product to combinations of historical sales and promotions data of the concerned and related products. Their results showed that in periods with promotion, incorporating promotional data about related products is beneficial for the model accuracy.

Whereas previous mentioned research only considered competitive information in the promoted periods, Huang et al. (2014) also considered this in non-promoted periods. They researched forecasting sales at the SKU level for retailers in the presence of promotions, incorporating competitive information. Three types of Autoregressive Distributed Lag (ADL) models were applied of which two incorporate competitive information and one does not. The ADL models were compared to two benchmark methods: the simple exponential smoothing (SES) model and the base-times-lift approach in which baseline forecast are produced and adjustments are made in case of a promotion. In forecasting, a distinction was made between a promoted period and non-promoted period. During all periods, the base-time-lift approach performed better than the SES model. All ADL models on their turn performed better than both benchmark models, but the ADL models incorporating competitive information had better performance than the ADL model which did not incorporate that information. The same conclusions hold when considering only the promoted periods. Forecasts for the non-promoted periods, however, yielded contrary results. Again the SES method performed as worst, but the base-time-lift approach outperformed the ADL model without competitive information. The ADL models with competitive information, again, had the best performance.

In addition to previous research, Ma et al. (2016) did not only take into account intra-category promotional information, but also inter-category promotional information. They showed that incorporating more information about related products improve accuracy compared to only incorporating information about the forecasted product. However, the largest part of the accuracy improvement is due to incorporating intra-category information, only a small part is due to inter-category information.

Most recently Ma and Fildes (2017) extended the research about cross-product promotional information with incorporation of cross-period promotional information. Namely, they considered the promotional effects on the forecasted product in the current period, the promotional effects on the forecasted product in successive periods (cross-period) and the promotional effects on other products in the same category in the current period (cross-product). They modeled sales as a function of promotional effects, long term trend and seasonality with a Autogressive Distributed Lag (ADL) model. The ADL model incorporating cross-product promotional information was compared to an ADL model incorporating only promotional information of the forecasted product. The results showed that, in general, the ADL model incorporating cross-product information performed better than the alternative models and they conclude that incorporating cross-product and cross-period promotional information can improve the accuracy of SKU demand forecasting.

1.1.3 Future competitive promotion behavior

The importance of competitive promotional information is stressed in the previous section, but in practice this information is often only available to the retailer and not to the manufacturer. This is confirmed by Yang, Goh, Jiang, Zhang and Akcan (2015), who argued that previously proposed forecasting methods which incorporate promotion information of competitive products (e.g. Huang et al., 2014) are not applicable at the manufacturer-level. They used forecasts incorporating competitive promotional information at the retailer-level and reconciled this to an aggregated forecast on manufacturer-level. They argued that in this way retailers do not have to share potentially sensitive business information (like a promotion calendar). However, they assume that retailers' forecasts incorporate promotional information and that those forecasts are shared with manufacturers. Since this is often not the case in practice, in this chapter literature about competitive promotion behavior is reviewed.

Traditionally, competitive structures were researched by measuring brand switching behavior of consumers. For example, Grover and Srinivasan (1987) classified competing brands within customer segments by applying structure analysis. However, those types of researches only describe competitive structures and do not account for competitive behavior or actions of firms.

Another traditional method for researching competitive structures, which considers behavior, is structural estimation analysis. Structural estimation research which does address promotion behavior of competitors in the retail sector is focused on reaction functions. Vilcassim, Kadiyali and Chintagunta (1999) developed a game-theoretic formulation to determine how a firm can react optimally, both in terms of direction and magnitude, to the marketing actions of another firm. Furthermore, their formulation is able to reveal the structure which acts as the foundation of those competitive reactions. The drawback of structural estimation techniques is that results are driven by restrictive assumptions instead of data (Dubé et al., 2002). Therefore, Ailawadi, Kopalle and Neslin (2005) use a game theoretic point of view which is combined with an empirical analysis for predicting the competitive response to a change in promotion policy. They conclude that the promotion strategy of manufacturers is random and they do not respond to each other on a week-by-week basis. Competitor reactions are also researched by Steenkamp, Nijs, Hanssens and Dekimpe (2005) by means of empirical analysis. Specifically, they researched how competitors react to price promotions and advertising, both on the short and long term. They showed that the main response to price promotions and advertising is no response at all. They also took into account that a promotion induced by a manufacturer has a longer response time compared to a promotion induced by a retailer (Leeflang & Wittink, 1992, 1996). These researches on competitor reactions confirm the demonstration of Rao, Arjunji and Murthi (1995) that retail promotions are independent across competitors. They explained this by the fact that the competitors strategy is unknown to a firm and therefore its promotion decision must be independent of the competitors promotion action.

The aforementioned literature about competitive promotion behavior mainly describes historic behavior. In forecasting future demand, it is essential to foresee future promotional behavior. Divakar et al. (2005) predicted the competitor's price and feature and display activity with regression models and they concluded that that information is generally well predictable. However, they predicted this on brand-level and they did not consider price promotion timing. Groot and Musters (2005) predicted optimal promotion timings, i.e. when the number of competitive brands on promotion is the lowest. They also found that the competitor promotional planning is predictable to some extent, namely they predicted the market with a success rate of 56% whereas a random guess would have a success rate of 50%.

1.1.4 Gap in existing literature

The reviewed literature indicates that a manufacturer should use POS data over order data (Kiely, 1998; Lapide, 2005; Holmström et al., 2002; Williams & Waller, 2010). Additionally, one of the most recent papers on this topic suggests that both order data and POS data should be used since that benefits the forecast accuracy the most. The benefit of this approach is that it accounts for how a retailer corrects for earlier forecast errors in future orders (Williams et al., 2014).

A review on the incorporation of promotions of related products showed that incorporating such information is beneficial to the forecasting performance (Kuo, 2001; Divakar et al., 2005; Ali et al., 2009; Huang et al., 2014; Ma et al., 2016; Ma & Fildes, 2017). The biggest flaw in most of the research in that area is that promotional information of competitors is only known to retailers and not to manufacturers. There is little research to suggest that the promotion planning is predictable. However, the research that has been done only focused on the amount of products on promotion, the timing of promotions was not considered. (Divakar et al., 2005;

Groot & Musters, 2005)

A gap in existing literature is thus identified on the prediction of price promotions of competitors. Also Abraham and Lodish (1987) give the forecasting of future competitive promotions as a possible direction for future research. Despite this direction for future research, no one to the best of our knowledge has studied this.

1.2 Research Questions

The reviewed literature in the previous section showed that incorporating promotions of competitors improves forecast accuracy. However, promotions of competitors are unknown to manufacturers. Therefore, this research aims to assess the predictability of promotions of competitors. Furthermore, it will be assessed whether the incorporation of predicted promotions of competitors improves retail order forecast accuracy. The research objective is formulated as follows:

Improve retail order forecast accuracy by gaining insight into the predictability of promotions of competitors.

To reach this research objective, the following main research question is formulated:

Are the promotions of competitors predictable and do predicted promotions of competitors improve retail order forecast accuracy?

In order to answer the main research question, the following sub-questions are defined:

1. *Which promotion planning behavior of competitors can be extracted from historic promotion data?*
2. *How should the competitive promotion prediction model be designed?*
   (a) *Which features should be chosen or created to predict a promotion event?*
   (b) *Which method should be used to predict a promotion event?*
3. *How should the retail order forecast model be designed?*
   (a) *Which features should be chosen or created to forecast retail orders?*
   (b) *Which method should be used to forecast retail orders?*
4. *Does incorporating predicted promotions of competitive brands improve retail order forecast accuracy?*

1.3 Scope

This research is limited to forecasting from a manufacturer's point of view. Particularly, the forecasting of retail orders in which the sales to the end consumer is subjected to promotions is of interest. The manufacturer should have competitors in the sector in which it performs and the future promotion planning of competitors should be unknown. Furthermore, the manufacturer should have access to data about historic promotions (including competitor's promotions) and to both historic order data and POS data.

1.4 Methodology

This research is structured according to the reflective and regulative cycle developed by Van Aken (1994) and van Strien (1997) as shown in figure 1.1. The reflective cycle aims to develop a design knowledge which is generally applicable. The design knowledge is acquired from the regulative cycle in which a specific case is solved.

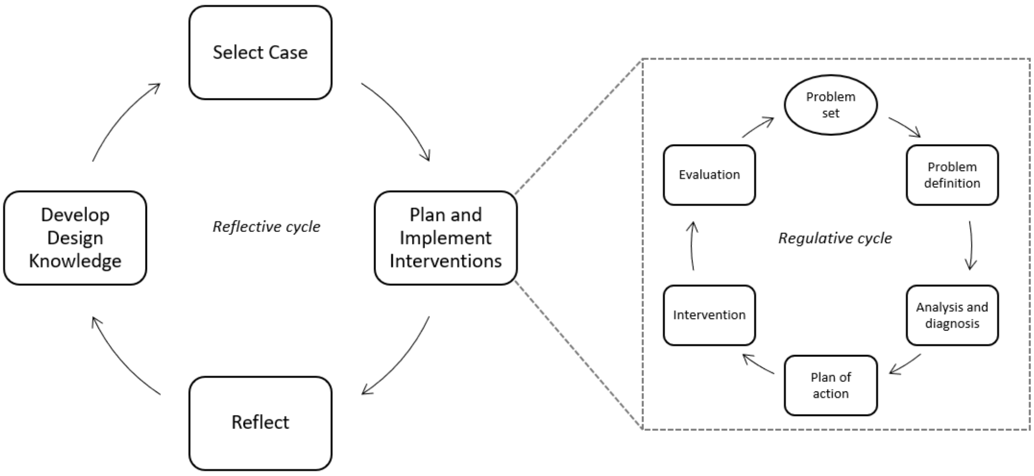

Figure 1.1: Reflective and regulative cycle
(Van Aken, 1994; van Strien, 1997)

First, the current situation of the selected case will be analyzed which corresponds to the analysis and diagnosis step of the regulative cycle. Sub-question 1, sub-question 2 and sub-question 3 address the plan of action and intervention. Sub-question 4 evaluates the intervention and with that the regulative cycle is finalized. Finally, a reflection on the case will be executed to develop design knowledge and with that also the reflective cycle is completed. Chapter 2 describes the selected case.

1.4.1 General approach

The plan of action for solving the case consists of building two models. One model aims to predict competitive promotions and the other model aims to forecast retail orders. The predictions of the competitive promotions will serve as an input for the retail order forecast model. In order to be able to predict competitive promotions, the promotion behavior of competitors needs to be analyzed first. This will be done by analyzing historic promotion data and extracting patterns from that. This general approach is visualized in figure 1.2.

Figure 1.2: General research approach

The literature is consulted to form a range of methods which could be used for both the competitive promotion prediction model and the retail order forecast model and out of this range an appropriate method is chosen. As described in the literature review (section 1.1), a gap in existing literature is identified on the prediction of promotions of competitors. Therefore, methods which were applied in a similar situation (i.e. event prediction) are consulted. Customer churn prediction is a widely discussed topic in literature and can also be seen as event prediction. Customer churn occurs when a customer stops doing business with a company. To predict a churn event or the time till the next churn event occurs various classification techniques have been used. Especially neural networks, decision trees and random forests are widely used (Miguéis, Van den Poel, Camanho & e Cunha, 2012) and therefore those will be considered as the range of methods for predicting competitive promotions. As to the methods for retail order forecasting, Ali et al. (2009) researched the relative performance of sales forecasting techniques ranging from more traditional time series models like linear regression to sophisticated machine learning models of support vector machine and regression trees. They found that simple time series technique perform at least as good as machine learning techniques in periods without promotions. In contrast to that, all machine learning techniques perform better in periods with promotions, especially when promotions of related products are incorporated. The support vector machine and regression tree has a particularly outstanding performance. Therefore, the support vector machine and regression tree will be considered as the range of methods for forecasting retail orders. The evaluation of the models will be based on quantitative methods.

1.5 Outline

The remainder of this report is structured as follows. Chapter 2 gives the description of the selected case and problem. Chapter 3 describes the analysis of competitive promotion behavior. In Chapter 4 the competitive promotion prediction model is described and Chapter 5 describes the retail order forecast model. Finally, Chapter 6 provides the conclusions of the research.

Chapter 2

Case and problem description

This chapter presents the company at which the research is applied and the available data at the company. Furthermore it is described how the performance of the forecasting method is evaluated and the evaluation of the current performance is presented.

2.1 Company background

The research is applied at Swinkels Family Brewers. Swinkels Family Brewers is a Dutch family business which operates in the beer, soft drink and malt sector. Swinkels Family Brewers was founded in 1719. They have six brewery locations and two malt houses in the Netherlands, Belgium and Ethiopia, their portfolio consists of 26 brands and their products are consumed in more than 130 countries.

Swinkels Family Brewers sells their products through different channels. This research is focused on beer sold in the retail sector. Retailers place an order at Swinkels Family Brewers and Swinkels Family Brewers attempts to predict those orders as well as possible. Accurate forecasts are important since they are a resolution for the trade-off between service and costs. This trade-off arises due to safety stock which avoids stock-outs but reduces profit. More accurate forecasts have the potential to decrease safety stock and thereby reducing costs while maintaining the level of service.

There are many brands active in the beer market in the Netherlands. However, only few of the brands are brewed independently. Most of the brands are owned by a company which owns several other brands. Also Swinkels Family Brewers, as stated before, owns several brands. In the remainder of this paper, competitive brands mean brands which are not owned by Swinkels Family Brewers.

2.2 Promotions

52.7% of all beer sold in Dutch retail in 2018 was sold on promotion. When zooming in on the pilsner subcategory, 54.4% was sold on promotion. And when only considering pilsner sold in crates, this amount was even 69.3%. This indicates a high promotion pressure in the beer category. In 2017 48.7% of all beer sold in Dutch retail was sold on promotion. This amount was 47.8% in 2016, 47.0% in 2015 and 42.7% in 2014. This indicates that the promotion pressure has only increased since 2014, which demonstrates the relevance of incorporating promotions in forecasting.

Account managers of Swinkels Family Brewers are responsible for the retail promotion planning. The account manager makes, in agreement with the retailer, decisions on the planning of the promotions on the retail shop floor. The account manager creates an initial promotion planning for every brand and presents this to the retailer. Since the retailer deals with account managers from different brands and firms, it is their job to align all requested promotions from the account managers. The result of these negotiations is a promotion planning, which is certain at least four weeks upfront. This promotion planning consists only of Swinkels Family Brewers' own brands. During the negotiations, the promotion planning of competitors is also brought up. However, in practice it is hard to draw conclusion from the negotiations about the competition since the retailer is also negotiating with those competitors. Therefore, in this research it is assumed that the account manager is unaware of promotions of competitors.

To cover for the price reduction on the retailer shop floor, Swinkels Family Brewers gives a discount on the wholesale price. In this way, a retailer is able to purchase products with a discount from Swinkels Family Brewers. The period in which the retailer can purchase products with a discount is different per retailer but this is generally one or two weeks before the promotion. These weeks are defined as the loading weeks.

2.3 Forecasting process

Swinkels Family Brewers uses a forecast method for forecasting retail orders which can be considered as a base-time-lift approach. A tool provides a monthly initial baseline forecast per retailer based on exponential smoothing, which is divided proportionally over the weeks in that month. Adjustments to the baseline forecasts are made by account managers in case of a promotion. These adjustments are based on the lift effect of the last like promotion and by their own judgement. In the end, the demand planner is responsible for the forecasts and can make judgmental adjustments as desired. The final forecasts are weekly forecasts for every product at every supermarket provided four weeks upfront. The purpose of the final forecast is two-fold. On the one hand, forecasts on retailer-level serve as grip for the account manager for a better negotiation position. On the other hand, forecasts aggregated to all retailers together serve as an input for the production planning. The forecasting process is displayed visually in figure 2.1.

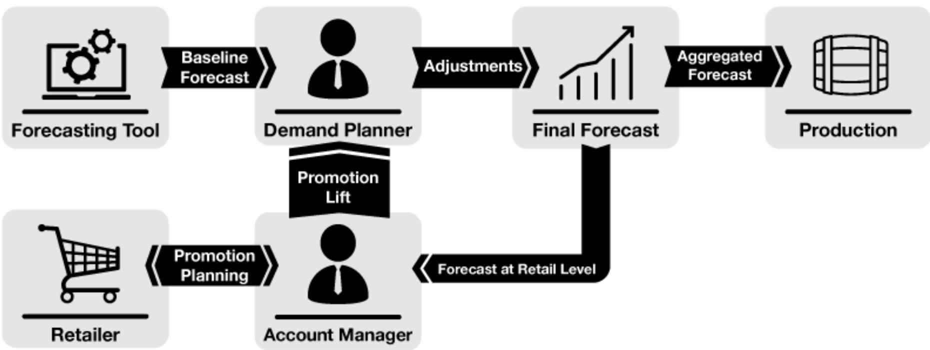

Figure 2.1: Forecasting process

Currently, the promotions of competitive products are not taken into account in the forecast since those are unknown. Furthermore, the forecast method in the period of a promotion can be considered as subjective. This is because the account managers, who provide the initial forecast in the case of a promotion, are responsible for the contact with the retailers and therefore the account managers are inclined to overestimate the lift of a promotion. In this way the account managers can guarantee that there is enough stock available and thereby the retailer will not be surprised by a stock-out. The production department, on the other hand, is primarily driven by costs and desires forecasts that are as precise as possible. The demand planner finds himself in the middle trying to find a balance. The performance of the current forecasting method is described in section 2.5.2.

2.4 Available data

The available data sets consist of historic order data, POS data and promotion data collected from week 1 of 2015 till week 12 of 2019. The order data consists of weekly orders from retailers, the POS data consists of weekly sales to consumers and the promotion data consists of price promotions in the supermarket. The historic order data contains only Swinkels Family Brewers' brands, while the POS data and promotion data also contains brands of competitors. Before analyzing the data, it is essential to ensure that the three data sets contain the same retailers and same products of Swinkels Family Brewers. This is not the case for all retailers and therefore it is chosen to only analyze the retailers which are present in all data sets. The number of remaining retailers in the data set is nine. All nine retailers are supermarkets. Furthermore, Swinkels Family Brewers' products which have changed from package, either temporally or permanently, received a new product number while the product itself has not changed. In order to deal with this, it is chosen to adapt the data in such a way that data from the different product numbers of one product is merged and only the original product number is kept. It is also decided to remove products which were introduced to the market after 2015 (i.e. new product introduction) or withdrawn from the market between 2015 and 2019 (i.e. end of life). In this way only products which were sold throughout the entire period between 2015 and 2019 remain in the data set. An overview of the retailers (i.e. supermarkets) and products in the data set can be found in Appendix A.

The products in the data sets are categorized into five different groups, namely pilsner crates, pilsner other, 0.0%, fresh and characterful. An explanation of every group can be found in table 2.1.

Group	**Explanation**
Pilsner crates	Pilsner beer sold in crates belong to this group.
Pilsner beer other	Pilsner beer sold not in crates belong to this group (e.g. pilsner sold in cans).
0.0%	All types of non-alcoholic beers belong to this group.
Fresh	White beers, radlers, ciders and Mexican style beers belong to this group.
Characterful	Special beers and heavy beers belong to this group.

Table 2.1: Product groups

2.4.1 Order & POS data

The order and POS data consists of weekly orders and weekly sales to consumers, respectively. The final order and POS data set consist both of data from 30 SKUs (Stock Keeping Unit). The SKUs are from five different brands of Swinkels Family Brewers. How the SKUs are divided over the five groups is visible in figure 2.2. To give a complete picture, the product volumes in the data set divided over the product groups is also made visible (figure 2.3). As one can see, most of the products in the data set are characterful beers (40%), followed by 0.0% beers (23%)

and pilsner other (17%). However, the pilsner crates group is the group with the highest volume sold to the retail (39%), shortly followed by the pilsner other group (36%). This explains why the decision was made to divide pilsner into subgroups of crates and non-crates (other).

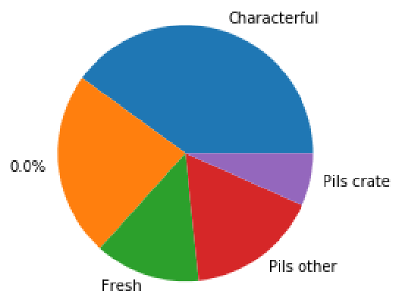

Figure 2.2: Number of products divided over groups

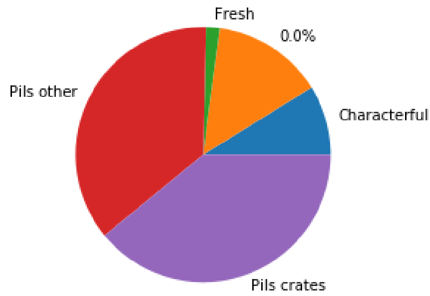

Figure 2.3: Product volumes divided over groups

In total, there are 147 product-retailer combinations in both the order data set and POS data set. Not all retailers sell all products, which explains why the number of product-retailer combinations is not equal to the number of SKUs multiplied by the number of retailers.

2.4.2 Promotion data

The promotion data consists of price promotions in the supermarket. The promotion data is two fold. On the one hand, the promotion data consists of the promotions of all brands. These promotions are on product group level. This data is purchased from an external party. On the other hand, the promotion data consists of the internal data about promotions of Swinkels Family Brands. These promotions are on product level and is more detailed than the external promotion data. For example, the internal promotional information also contains data about second placements. The external promotion data is available from 2015 till now, while the internal promotion data is only collected from 2017.

There are a more than 140 brands in the external promotion data, therefore it is chosen to include only the largest competitor brands which, together with Swinkels Family Brewer's brands, account for a market share of at least 80% per group. Private label brands (i.e. brands owned by retailers) are excluded from the analysis, since there is no data available about the promotions of private label brands. The selected competitive brands and their market shares can be found in Appendix B.

The external promotion data set consists of data about promotions from the promotion brochures, but also from television commercials or online newsletter. It is decided to only focus on promotions from the brochures, since the brochure always consists of the full promotion program. In this way, the data is not contaminated by promotions which are mentioned in a TV commercial as well as in the brochure. Furthermore, only promotions in which a price reduction is the case are selected. Promotions in which customers get a free premium instead of price reduction are out of scope in this research. Furthermore, when the data states that 'all variants' of a brand are on promotion it is assumed that all groups in which the brand is active are on promotion except for the pilsner crates group. Imagine a brand which is active in the pilsner crate group, the pilsner other group and the fresh group. If the promotions states that all variants of that brand are on promotion, the promotion accounts for both the pilsner other and the fresh group. The pilsner crates group is not covered by 'all variants' since the data shows that the pilsner crates always have a separate mention in the brochure due to its importance and size.

The final external promotion data set consists of 8217 promotions of 22 competitor brands in 9 supermarkets. Those 22 competitor brands are owned by 9 companies. Appendix B also shows which brands belong to which companies. The final internal promotion data set consists of 1707 promotions of Swinkels Family Brewer's brands in 9 supermarkets.

2.5 Current forecasting performance

In this section the performance of the current forecasting method is evaluated. First, the performance measures which are used for the evaluation is described. Accordingly, the current performance of the retail order forecasts is determined.

2.5.1 Performance measures

The performance of the retail order forecast model is evaluated using the mean absolute error and the Mean Absolute Promotion Volume Deviation.

The **mean absolute error (MAE)** calculates the deviation between the forecast and the actual value for every observation and has been widely used in performance evaluation of forecasting methods (Huang et al., 2014). The MAE is a scale dependent measure which means that it cannot be used across data sets of different units. In this research only one data series (the retail orders) will be evaluated, so the scale dependency is not a problem. Another property of a scale dependent error measure is that it considers the weight of individual SKUs. The average order amount per product ranges a lot and in that case a relative error measure such as the mean absolute percentage error (MAPE) would not give extra weight to the accuracy of the large volume SKUs. Furthermore, in the periods when the actual order is 0 the MAPE would be undefined as you cannot divide by 0. For these reasons, the MAE is chosen as one of the performance measures for the retail order forecast model. Since the MAE corresponds to an error, the lower the MAE the better the better the forecast method performs. The MAE is calculated as follows:

$$MAE = \frac{1}{k}\frac{1}{n}\sum_{i=1}^{k}\sum_{t=1}^{n}|Y_{i,t} - \hat{Y}_{i,t}| \tag{2.1}$$

Where k is the number of product-retailer combinations, n is the total number of observations per product-retailer combination, $Y_{i,t}$ is the actual order for product-retailer combination i at time t and $\hat{Y}_{i,t}$ is the forecast order for product-retailer combination i at time t.

The **Mean Absolute Promotion Volume Deviation (MAPVD)** measures the deviation between the forecast and the actual value only for observations which encounter a promotion. This is a self-invented performance measure and it is created since the current forecast method of Swinkels Family Brewers is considered as subjective in periods with a promotion. This measure reflects the performance specifically in a promotion period, whereas the MAE reflects the overall performance regardless of whether there is a promotion. A promotion period is defined as the promotion week itself plus the two loading weeks before the promotion. The MAPDV is cacluated as follows:

$$MAPVD = \frac{1}{k}\frac{1}{P}\sum_{i=1}^{k}\sum_{p=1}^{P}|Y_{i,p} - \hat{Y}_{i,p}| \tag{2.2}$$

Where k is the number of product-retailer combinations, P is the total number of promotions per product-retailer combination, $Y_{i,p}$ is the actual order for product-retailer combination i at promotion period p and $\hat{Y}_{i,p}$ is the forecast order for product-retailer combination i at promotion period p.

2.5.2 Current performance

The performance of the current forecasting method is evaluated according to the performance measures described in the previous section. The order data set is used for the evaluation. The performance of the current forecasting method is calculated over the last 25 weeks of the data, which corresponds to week 40 of 2018 up until week 12 of 2019. The performance is presented in table 2.2. The table also shows the average order per week per product and the average promotion volume per promotion per product. The performance can be interpreted as follows. The average absolute deviation between the forecast order and actual order is 28.76 hectoliters for any product-retailer observation at any week during the last year of the data. In other words, on average the forecast deviates every week approximately 9627 beer bottles from the actual order. For every promotion, the forecasted orders deviate on average 131.29 hectoliters from the actual orders.

Method	**MAE (HL)**	**Average weekly order (HL)**	**MAPVD (HL)**	**Average promotion volume (HL)**
Current method	28.76	53.92	131.29	397.42

Table 2.2: Performance current forecasting method

Figure 2.4 shows the actual and forecast orders for the best selling product of Swinkels Family Brewers at one retailer from week 40 2018 to week 12 2019. The baseline weeks are indicated with a dot, promotions weeks are indicated with a square and the one or two weeks before a promotions are indicated with a triangle. The triangle indicates thus the weeks in which the retailer can buy products against a discounted wholesale price.

As the figure indicates, the deviations between actual and forecast orders are the biggest in the periods with a promotion. It was already stated earlier that the promotional forecasts can be considered as subjective. The figure also shows that the lift estimation in a promotion is more based on judgements and not of the lift effect of the most recent promotion. This pattern is also visible in the forecasts of other products. One may thus conclude that judgemental adjustments lead to systematic biases which is not beneficial for the forecasting accuracy. This is also supported by literature ((Cooper, Baron, Levy, Swisher & Gogos, 1999; Fildes, Nikolopoulos, Crone & Syntetos, 2008; Fildes, Goodwin, Lawrence & Nikolopoulos, 2009; Franses & Legerstee, 2010)).

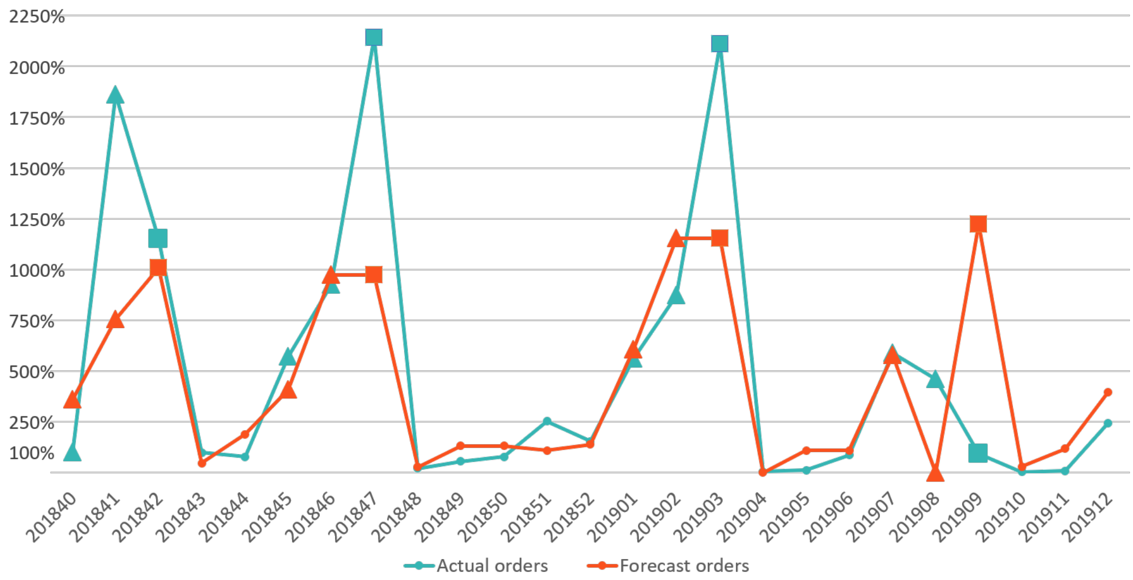

Figure 2.4: Actual and forecast orders of one SKU at one retailer of the current forecasting method. A dot stands for a baseline week, a square stands for a promotion week and a triangle stands for one or two weeks before a promotion. 100% corresponds to the average baseline volume.

Furthermore, one can see that in the week before the promotion and in the promotion week the retailer orders are forecasted to be approximately the same whereas the actual order is often much higher in the promotion week. However, this is not always the case as can be seen in the first promotion (2018 week 42). This figure, together with the high MAPVD suggest that the biggest opportunity for improvement lies in the periods around the promotions.

Chapter 3

Analysis competitive promotion behavior

This chapter provides an analysis on the promotion behavior of competitors in the beer market. Expert interviews are the basis for this analysis. Additionally, historic promotion data is examined. The aim of the competitive promotion behavior analysis is to detect input variables for the competitive promotion prediction model which is described in Chapter 4.

3.1 Findings from expert interviews

In order to investigate the promotion behavior of competitors, the promotion behavior of Swinkels Family Brewers is examined by interviewing account managers. According to the account mangers, there are a few important drivers for their promotion behavior. The most important drivers concerning the timing of promotions are the spread over the year and holidays. Swinkels Family Brewers prefers their promotions to be roughly equally distributed over the year and they use the promotions from the previous year as a starting point for this. The account managers also indicate that they distinguish between the type beer in the promotion planning. Furthermore, they take into account national holidays when planning their promotions. Carnaval, in particular, is an important holiday for them since Swinkels Family Brewers is established and located in the south of the Netherlands. Also other holidays play a role, especially the holidays which create an extended weekend.

3.2 Promotion data analysis

From the expert analysis a few drivers for the promotion planning have emerged. Those are the distribution of promotions over the year, the promotion planning of the previous year, holidays and the type of beer. Visualization techniques on the promotion data are used to test these drivers for competing brands. Also other drivers for promotion planning behavior, which are not mentioned in the expert interviews, may come to light through the visualizations. In the visualization, only the promotions in 2015, 2016, 2017 and 2018 are considered. The promotions in 2019 are not considered because only a part of that year is present in the data which could create a distorted image. Since there are a lot of brands and retailers in the data, we chose to start the analysis with promotions of one retailer. Figure 3.1 and Figure 3.2 indicate that the Albert Heijn is the largest supermarket in terms of number of promotions and order amount of Swinkels Family Brewers' products. Therefore, Albert Heijn is the chosen retailer to start the analysis with. The promotions of Albert Heijn always start on Monday and end on Sunday.

Besides, Figure 3.2 shows that the order amount is decreasing over the years. This is due to the decreasing pilsner consumption and increasing consumption of other beer types. As described before, 75% of the order volume in the data set is from pilsner beers. Swinkels Family Brewers is therefore also expanding its portfolio. For example, in 2016 Swinkels Family Brewers took over specialty beer brewery Palm.

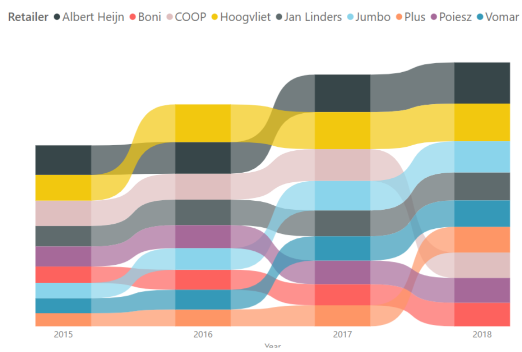

Figure 3.1: The number of promotions per retailer per year

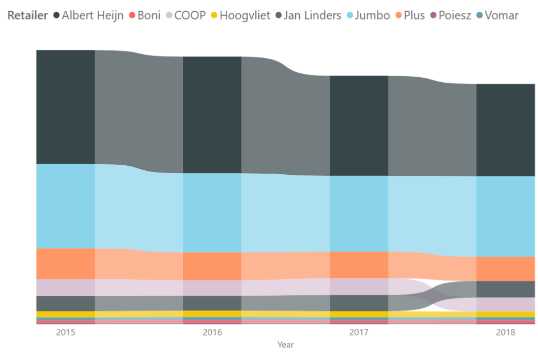

Figure 3.2: Order amount (in HL) of Swinkels Family Brands' products per retailer per year

3.2.1 Analysis promotions at Albert Heijn

The first aspect which is analyzed is the timing of promotions of different types of beer. The five groups (i.e. pilsner crates, pilsner other, 0.0%, fresh and characterful) are used for this.

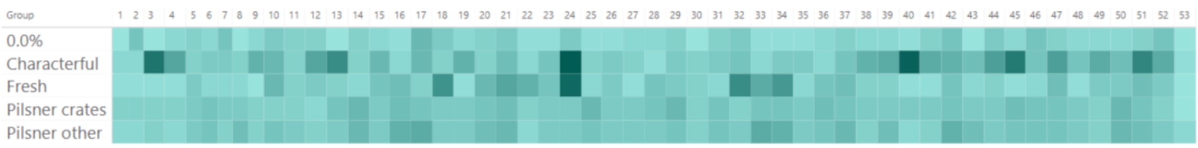

Figure 3.3: Heatmap of promotions at Albert Heijn in 2015-2018

Figure 3.3 shows a heatmap of the promotions per week per group at Albert Heijn in 2015, 2016, 2017 and 2018. The darker the color, the more promotions there were in that week. The figure shows that the promotions of the two pilsner groups are divided equally over the year, while the promotions of the fresh and characterful group show a slightly more seasonal pattern. The promotions of fresh beers seem to occur more in summer, for example week 24, 32, 33 and 34. The promotions of the characterful group seem to happen mostly in autumn or winter, for example between week 40 and 51 and week 3, 4, 12 and 13. Promotions of the characterful group also show a peak in week 24, which can be promotions of blond beers which belong to the characterful group but can have a fresh flavour. Furthermore, the number of promotions in the 0.0% group seems a little lower than the number of promotions in the other groups, indicated by the lighter color. All in all, Figure 3.3 indicates differences between the timing of promotions of different groups. Therefore, the analysis is performed on the groups separately.

3.2.1.1 Pilsner crates

The relevant brands in the pilsner crates group are Heineken, Hertog Jan, Grolsch, Amstel, Bavaria and Jupiler. Figure 3.4 visualizes the promotions of pilsner crates at Albert Heijn. The heatmap in the figure indicates that there is a promotion of beer crates almost every week. Often there is one brand on promotion, but promotions of two or more brands occur as well. The fact that there are almost no weeks without promotion (in white) could suggest a strategy in which the competitors are trying to keep another competitor out of the market (Lal, 1990). This could also be initiated by the retailer instead of the competitors themselves, since the retailer is ultimately responsible for the promotion planning.

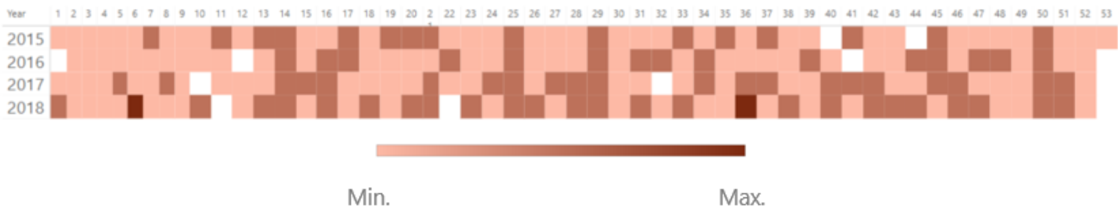

Figure 3.4: Heatmap of pilsner crates promotions at Albert Heijn

The graph in Figure 3.5 shows the pilsner crate promotions at Albert Heijn. This graph allows one to compare the promotions of the brands of this group for every year. Interestingly, the promotions of Hertog Jan and Jupiler beer crates are often in the same week. As Appendix B points out, Hertog Jan and Jupiler are owned by the same company. Furthermore, Heineken is often the only brand on promotion, but when it is sharing its promotion week it is always - in 2015 to 2018 - with Amstel. Also here it applies that Heineken en Amstel are owned by the same company. Lastly, Bavaria's pilsner crate promotions are regularly in the same week as Grolsch' pilsner crate promotions. However, Bavaria and Grolsch are owned by a different brewer. To find out if this is on purpose, verifications are made with account managers from Swinkels Family Brewers. According to them, Bavaria and Grolsch do not cannibalize each other. In other words, Bavaria and Grolsch do not take away customers from each other. Of course a retailer also benefits from a promotion of brands who do not cannibalize each other. This could therefore also hold for why Heineken and Amstel on the one hand and Hertog Jan and Jupiler on the other hand are on promotion in the same week. Another reason could be of a more commercial nature. Namely, the cost contribution to be in the promotion folder per brand could be lower when multiple brands of one brewer are on promotion compared to when the brands are on promotion in different weeks.

Figure 3.5 also allows one to compare the yearly promotions of a brand in the pilsner crates group. Also, the weeks with a holiday are visualized with a black line in order to test whether brands are often on promotion during holiday weeks. The holidays are national holidays and include Easter, Carnaval, Kingsday, Ascension Day, Pentecost, Christmas and New Year's Eve. Despite Carnaval being a regional holiday and not a national holiday it is included as holiday since account managers from Swinkels Family Brewers' indicated it is an important holiday for them in promotion planning. Figure 3.5 indeed confirms that Bavaria is always on promotion during the Carnaval holiday. Also Jupiler is often on promotion during Carnaval. Heineken is always on promotion during Christmas, as only brand. For other holidays it holds that Amstel has often promotions during the week of Easter Monday, Grolsch has often promotions during the week of Whit Monday, Heineken during the weeks of Whit Sunday and Easter Monday and Hertog Jan and Jupiler during Easter Sunday. Furthermore, one can see in Figure 3.5 that some brands have promotions during the same weeks every year. See for example Amstel's promotions in week 16, Grolsch' promotions in week 39 or Heineken's promotions in week 8.

Finally, Figure 3.5 suggests a certain degree of cyclicity, i.e. the promotions occur often after the same time has elapsed. To examine this further, histograms of the weeks between the promotions of the brands are shown in Figure 3.6. The figure shows that promotions of Grolsch, Heineken and Hertog Jan, in particular, show a strong degree of cyclicity. For those brands one can say

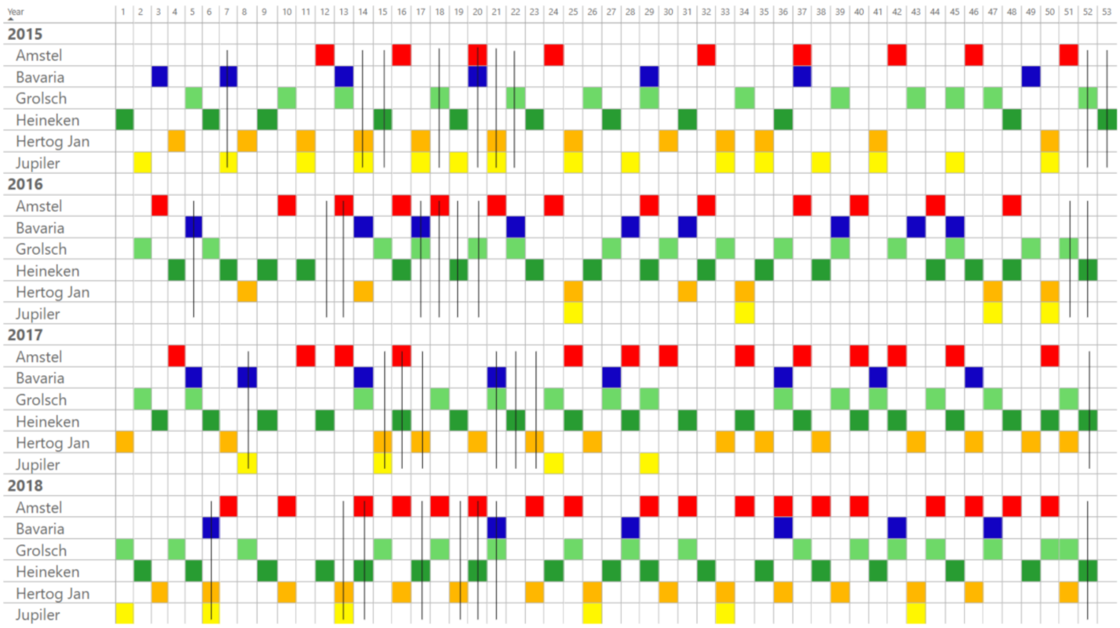

Figure 3.5: Pilsner crates promotions at Albert Heijn. The dark lines indicate the weeks with a holiday and the colors are unique for every brand.

that pilsner crates promotions occur approximately every three weeks. According to Appendix B Grolsch, Heineken and Hertog Jan have the biggest market share in the pilsner crates group at the nine supermarkets in the data set. To test whether this is also the case at Albert Heijn, Figure 3.7 shows the market share of brands in the pilsner crates group at Albert Heijn. The figure confirms that Heineken, Grolsch and Hertog Jan have the biggest market share of all pilsner crates brands at Albert Heijn. This suggests that there is a relation between the size of the brand and the cyclicity of the promotions. Furthermore, the number of promotions of these brands is higher than the number of promotions of other brands. This together indicates that the size of the brand influences the say of a brand in the promotion planning at Albert Heijn.

In conclusion, the analysis of the pilsner crates group has led to insights which could be helpful in predicting promotions of that group. Those insights are that particular brands are often on promotion in the same week. This seems to be related to the companies which own the brands. Vice versa, particular brands are often on promotion as only brand. Furthermore, the analysis of the promotions suggests that national holidays play an important role in the promotion planning of pilsner crates. Also important are the cyclicity of the promotions, indicating that often promotions occur after a fixed time. Finally, sometimes pilsner crates promotions occur in the same week every year.

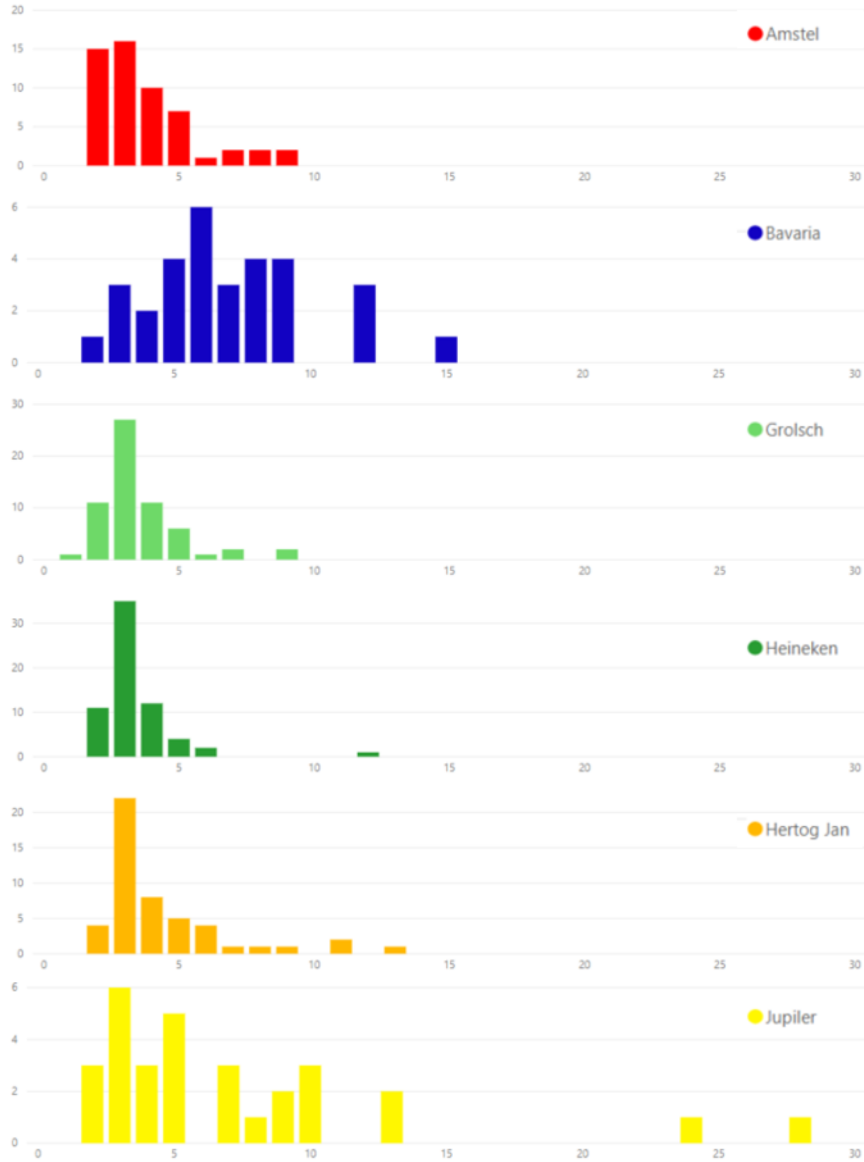

Figure 3.6: Histogram of weeks between pilsner crates promotions.

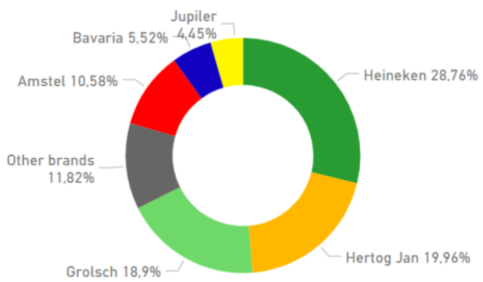

Figure 3.7: Market share of brands in the pilsner crates group at Albert Heijn (2015-2018)

3.2.1.2 Pilsner other

The relevant brands in the pilsner other group are Heineken, Amstel, Bavaria, Grolsch, Hertog Jan, De Klok and Jupiler. However, De Klok did not have any promotions in the pilsner other group at Albert Heijn in the past years. Therefore, the analysis of pilsner other promotions is only focused on Heineken, Amstel, Bavaria, Grolsch, Hertog Jan and Jupiler. A heatmap of the promotions in the pilsner other group is shown in Figure 3.8. The heatmap shows that the promotions are divided equally over the year and that there is a promotion almost every week. This might again indicate a strategy in which the brands (or the retailer) are trying to keep another competitor out of the market (Lal, 1990).

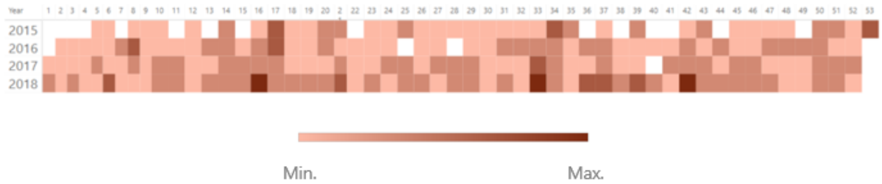

Figure 3.8: Heatmap of pilsner other promotions at Albert Heijn

To examine the relationships between the brands in the pilsner other group, Figure 3.9 is shown. Almost the same conclusions as in the pilsner crate group hold. Hertog Jan's and Jupiler's pilsner other, which are owned by the same company, are often in the same week on promotion. Bavaria's and Grolsch's pilsner other are often in the same week on promotion although they are not owned by the same company. Different from the pilsner crates group, Heineken is often alone on promotion, or with Amstel or Grolsch. Heineken and Amstel are owned by the same company, but Grolsch is owned by a different company. Also different from the pilsner crates group, in the pilsner other group it happens more often that there are three or more brands on promotion at the same time.

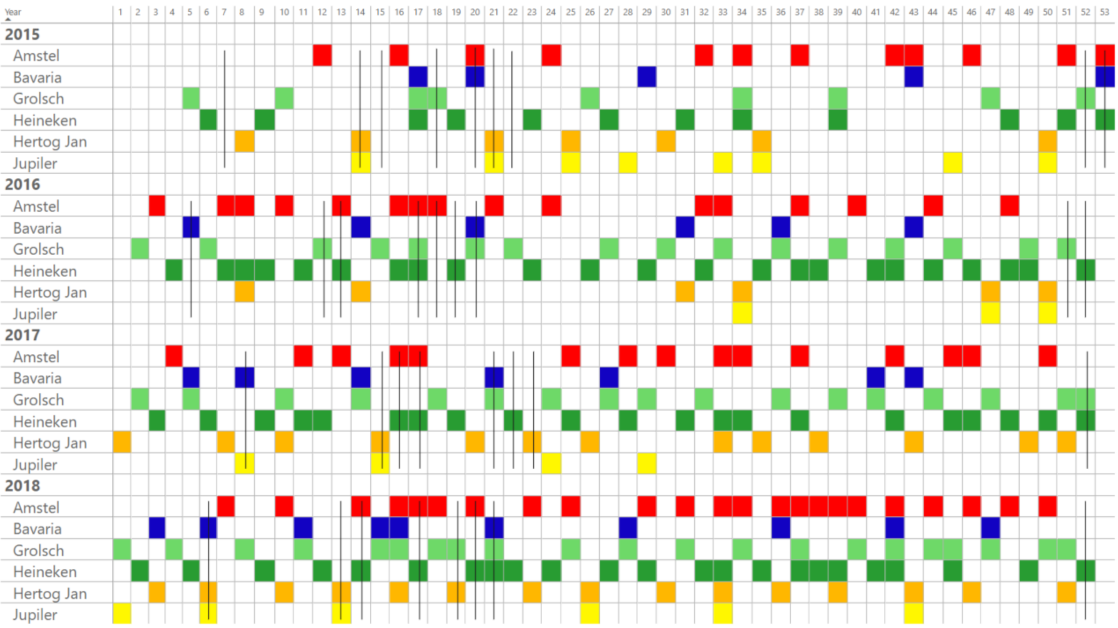

Figure 3.9: Pilsner other promotions at Albert Heijn. The dark lines indicate the weeks with a holiday and the colors are unique for every brand.

Figure 3.9 also presents the holiday weeks and shows that some brands are often on promotion during holidays. Namely, Amstel is often on promotion in the week of Kingsday and Easter Monday, Bavaria is often on promotion during Carnaval, Heineken is often on promotion during Easter Monday and always on promotion during Christmas, Hertog Jan is often on promotion in the week of Easter Sunday and Jupiler is often on promotion during Easter Sunday. Another thing that stands out in the figure is that some of the brands in the pilsner other group have often promotions in the same week every year (e.g. Amstel in week 37, Grolsch in week 39 or Heineken in week 9). Third, in the pilsner other group it occurs that promotions of one brand are in consecutive weeks. Lastly, the figure indicates some cyclicity. This will again be tested in histograms.

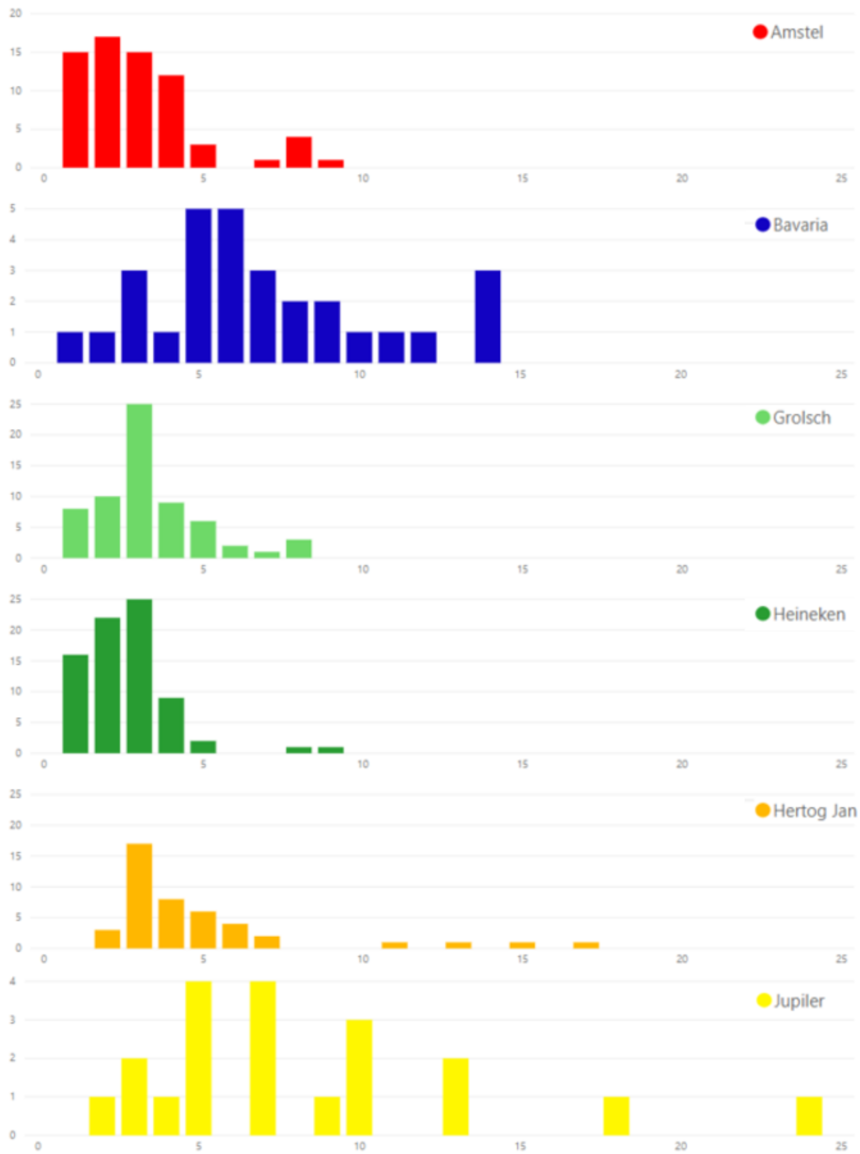

Figure 3.10: Histogram of weeks between pilsner other promotions.

Figure 3.10 presents the histograms of the weeks between the pilsner other promotions at Albert Heijn. The degree of cyclicity in pilsner other promotions seems less than the degree of cyclicity in pilsner crates promotions. In the pilsner other group the promotions of Hertog Jan and Grolsch, particularly, occur after a fixed time. Also Heineken shows cyclicity in their promotions. Those are the same brands which show cyclicity as in the pilsner crates group. However, in the pilsner other group those brands belong to the biggest brands but they are not necessarily the three biggest brands as indicated by Figure 3.11. A side note to Figure 3.11: the other brands have a big market share in the pilsner other group which is, among others, due to private label brands which are not considered in this analysis.

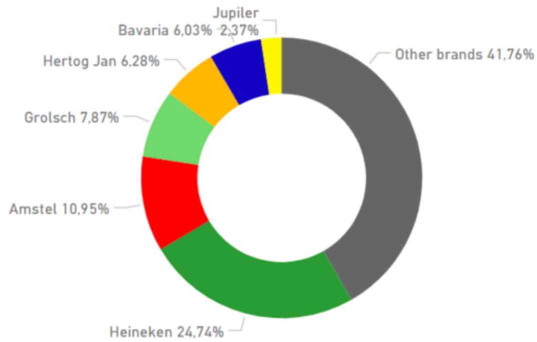

Figure 3.11: Market share of brands in the pilsner other group at Albert Heijn (2015-2018)

To conclude, the visualization of promotions in the pilsner other groups has led to the following insights. There are relations between the promotions of the different brands in the pilsner other group. Additionally, the promotions are related to national holidays. Which brand is related to which holiday is brand dependent. Last, the cyclicity of promotions and the week number is also an important factor in the promotion planning.

3.2.1.3 0.0%

The relevant brands in the 0.0% group are Amstel and Bavaria. The alcohol free beer segment is recently emerging. Amstel and Bavaria are the only 0.0% brands which are sold in the entire period between 2015 and 2019. For the 0.0% group the same analysis as for the pilsner crates and pilsner other groups has been carried out. There are only small insights from analyzing promotions in the 0.0% group. Namely, the promotions of Bavaria and Amstel are often not in the same week. The small number of insights is mostly because there are only few brands in this group. The graphs supporting these results can be found in Appendix C.

3.2.1.4 Fresh

The relevant brands in the fresh group are Amstel, Bavaria, Corona, Desperados, Erdinger, Franziskaner, Grolsch, Hertog Jan, Hoegaarden and Wieckse. Jillz is also part of the fresh group, but Jillz was not on promotion at Albert Heijn during the past years. Therefore, this analysis is focused on the aforementioned fresh brands. The analysis of fresh promotions is again performed with graphs which are presented in Appendix D. All in all, a few insights are derived from the analysis of the promotions in the fresh group. The insights are especially in the area of brand dependencies. Namely, brands owned by the same company are often in the same week on promotion. For example, Hoegaarden, Corona and Hertog Jan or a combination of two of those brands are often on promotion in the same week and the same holds for Amstel, Desperados and Wieckse. Furthermore, all brands in the fresh group or a large part of them are regularly all on promotion in the same week. This is the result of promotions in which for example all summer beers, all radlers or all white beers are on promotion. In terms of brand specific insights, cyclicity is especially found in promotions of Grolsch and Hertog Jan. Also only a few relations between promotions and holidays are discovered. Namely, Grolsch's fresh beers are often on promotion during Christmas. Lastly, the promotions of fresh beer bands occur most often in spring and summer.

3.2.1.5 Characterful

The relevant brands in the characterful group are Affligem, Bavaria, Brand, Duvel, Gulpener, Grimbergen, Grolsch, Heineken, Hertog Jan, La Chouffe, La Trappe, Leffe, Palm and Westmalle. There are patterns discovered in the promotions in the characterful group, which can be found in the graphs in appendix E. Specifically, the promotions of characterful beers happen mostly in the beginning and end of the year. This could be due to the autumn bok beers which belong to the characterful group. One can also imagine that special beers are drunk more in winter, which also explains the high amount of promotions in the beginning of the year. Furthermore, the promotions of certain brands occur often in the same week. Namely, Heineken, Brand, Leffe and Affligem which are brewed by the same brewer. Duvel and La Chouffe are also on promotion in the same week regularly and are owned by the same company. Finally, Bavaria and Grolsch is, again, a frequent combination. The promotions of some brands are related to national holidays. Finally, the promotions of some brands show cyclicity or a yearly pattern in which promotions occur often in the same week every year. Brand, Grolsch, Heineken and Hertog Jan show some degree of cyclicity, while the promotions of the other brands in the characterful group show hardly any cyclicity.

3.3 Summary

In the previous sections the analysis of the promotion at Albert Heijn is presented. The analysis is broken down into five groups of beer; pilsner crates, pilsner other, 0.0%, fresh and characterful beers since the timing of promotions differs among the groups. The analysis has provided insights in all groups. For all groups hold that particular brands are on promotion in the same weeks or never on promotion in the same weeks. For all but the 0.0% group hold that national holidays play an important role in the promotion planning. Furthermore, there are often the same amount of weeks between the promotions and sometimes promotions are in the same week every year. In the fresh group and characterful group also seasonality plays a role. An overview of the insights is presented in Figure 3.12.

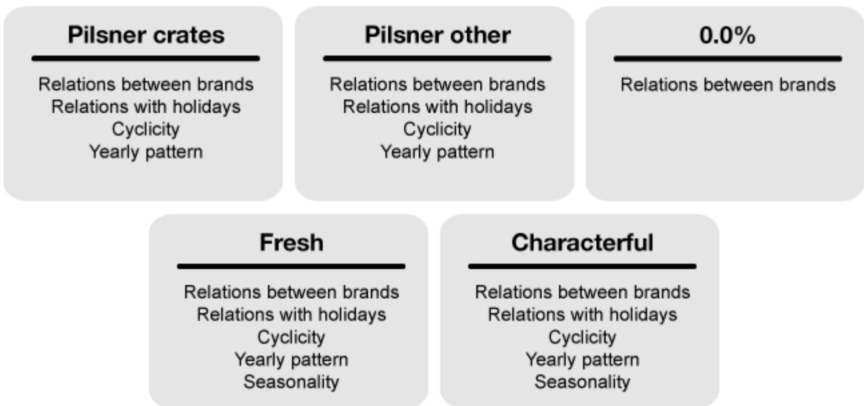

Figure 3.12: Insights promotions at Albert Heijn

Interestingly, the five groups share insights. This could suggest a relation between the groups. To get a better understanding of the relations between the groups, Figure 3.14 shows the promotions at Albert Heijn in 2017 of the brands which have products in more than one group. The figure indicates that there are a lot of similarities between the timing of promotions of different groups of the same brand. More specifically, often a promotion of a brand concerns products from multiple groups. This is particularly the case for the pilsner crates and pilsner other group.

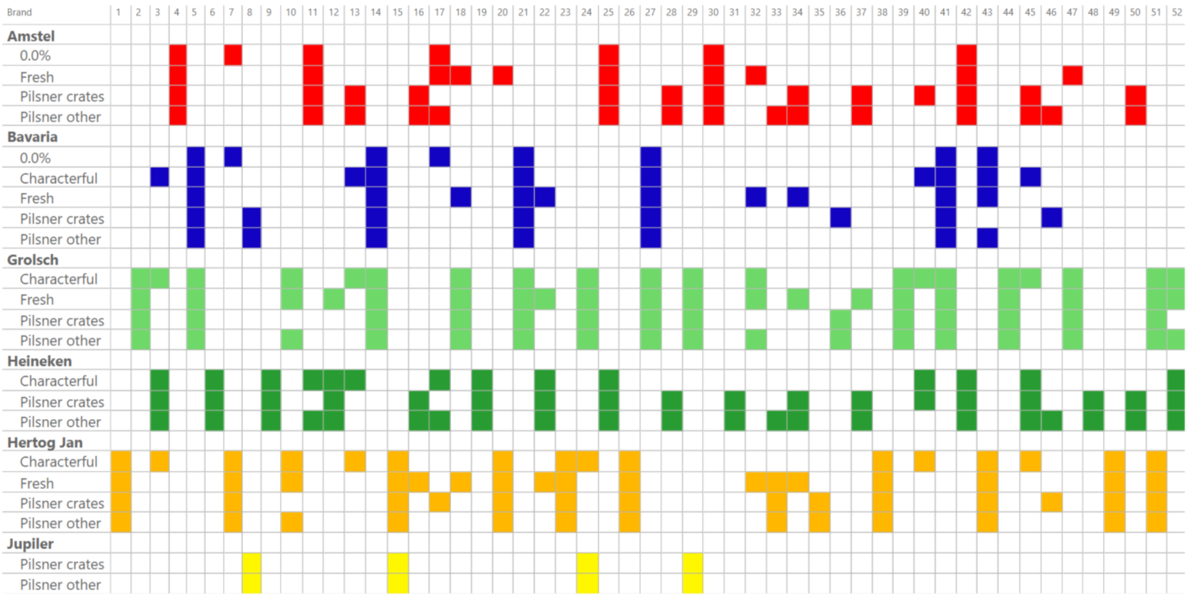

Figure 3.13: Promotions of multiple groups of brands at Albert Heijn in 2017

Another interesting insight is the cyclicity of the promotions. The cyclicity is especially the case for promotions of Heineken, Grolsch and Hertog Jan although they are not the biggest brands in all groups. They are, however, the three biggest brands in the pilsner crates group. As figure 3.3 indicates, the pilsner crates group has the highest market share compared to other groups at Albert Heijn. Therefore, it could be that since Heineken, Grolsch and Hertog Jan are the biggest brands in the pilsner crates group and the pilsner crates group is the biggest group at Albert Heijn, they have a big say in the promotion calendar. Accordingly, the big say could result in a geater degree of cyclicity.

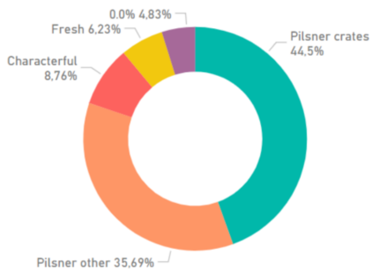

Figure 3.14: Market share of groups at Albert Heijn (2015-2018)

In conclusion, the comparison of promotions of different groups of one brand has led to a new insight. Namely, the promotions products in different groups of one brand are often in the same week. This is especially the case for the pilsner crates and pilsner other group. All in all, patterns in the promotions behavior at Albert Heijn are found in all groups. However, the analysis on promotion behavior has only been carried out for one retailer (Albert Heijn).

3.3.1 Other supermarkets

Previous section presented the analysis of competitive promotions at one retailer. In this section a small analysis is made of how promotions at Albert Heijn relate to other supermarkets. The duration of promotions is different for the supermarkets. For example, the duration of Albert Heijn's promotions is from Monday to Sunday while the duration of promotions at Jumbo is from Wednesday to Tuesday. To be able to make a comparison of promotions with different durations, the duration of the promotions is converted to the week (Monday - Sunday) to which the biggest part of the promotion corresponds. Imagine a promotion at Jumbo which takes place from Wednesday week 11 to Tuesday week 12. That promotion is then converted to week 11 since five of the seven days of the promotion are in week 11.

Figure 3.15 shows all promotions from all groups at nine retailers in 2017. The figure shows that Boni, Jan Linders, Plus, Poiesz and Vomar have significant less promotions than Albert Heijn. Coop and Hoogvliet also seem to have less promotions but the difference is smaller. The most notable promotions are those at Jumbo. This notable pattern is caused by the seasonal offers of Jumbo in which promotions endure for an entire season. Moreover, Jumbo offers yearly promotions and a few brands in the pilsner crates group had a yearly promotion in 2017. Obviously, this figure does not show particular promotion strategies. However, a totally different strategy such as Jumbo's is visible. An analysis of the promotions should be performed at every supermarket separately to find specific promotion behavior but that is left out of scope for this research due to time restrictions. The promotions at Jumbo are also left out of scope in the remainder of this research since there is little uncertainty in seasonal and yearly promotions compared to weekly promotions which makes a promotion prediction less relevant.

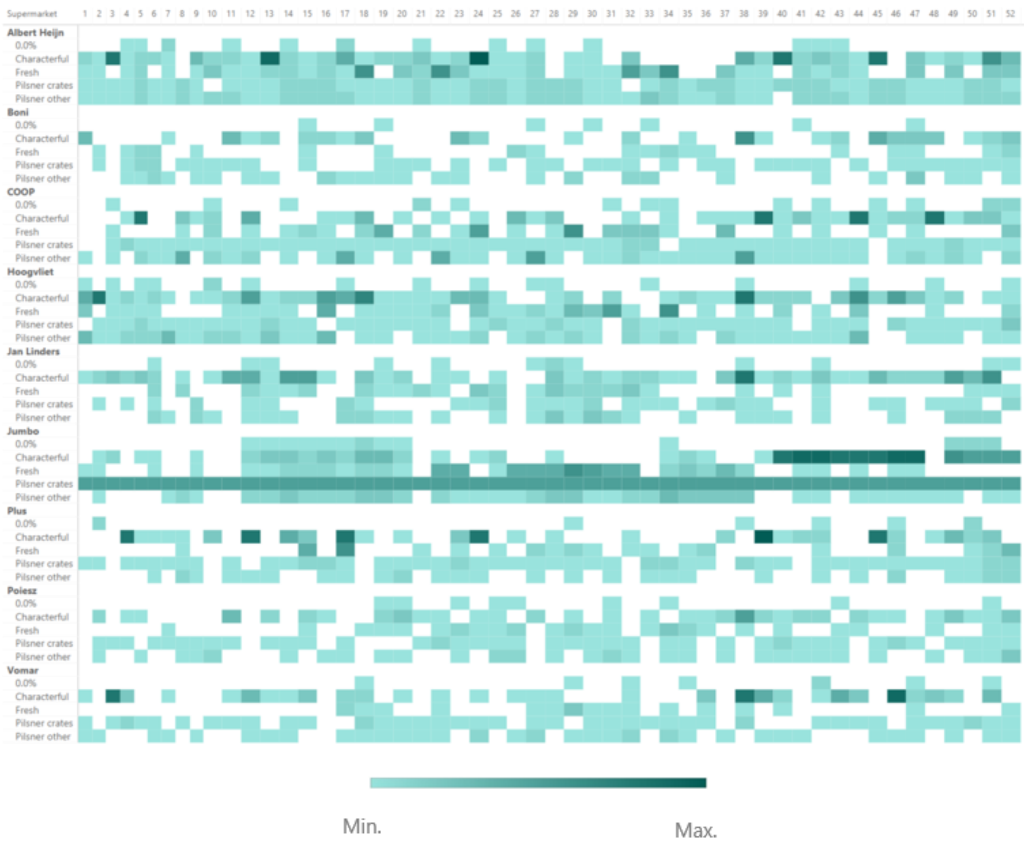

Figure 3.15: Heatmap of promotions at different supermarkets in 2017

Chapter 4

Competitive promotion prediction model

This chapter describes the competitive promotion prediction model which is based on the insights derived from the analysis of competitive promotions. It is described how the model is structured, how the performance of the model is tested and how the experiment is set up. Finally, the results of the competitive promotion prediction model are presented. The foundation for the competitive promotion prediction model is based on insights derived from the analysis of competitive promotions in Chapter 3. The aim of the competitive promotion prediction model is to predict the promotions of competitive brands. These predictions serve as one of the input variables of the retail order forecast model as described in Chapter 5.

4.1 Model description

The analysis in Chapter 3 showed that there is cyclicity, seasonality and a yearly pattern in the promotions. Furthermore, promotions are related to holidays and the number of promotions has increased over the years. These insights are translated into input variables for the competitive promotion prediction model. Additionally, there are relations between the promotions of brands and relations between the promotions of groups of one brand. Therefore, the model should also account for correlations between brands and groups. Furthermore, the competitive promotions need to be predicted four weeks ahead since the retail order forecast are required four weeks upfront due to production lead times.

The discovered promotion behavior does not apply to all researched brands. Therefore, the brands for which the behavior does not apply are left out of consideration in the competitive promotion prediction model. Table 4.1 shows the brands which remain in scope. Furthermore, the promotions of Swinkels Family Brewers' brands are not displayed in Table 4.1 as they are not being predicted since this information is obviously known to them. Promotions of Swinkels Family Brewers' brands should be, however, taken into account some way because their promotions are related to promotions of competitive brands. Namely, if a brand from Swinkels Family Brewers is on promotion it is less likely that a competitor is on promotion as well. Therefore, the promotions of Swinkels Family Brewers' brands are included as input variable.

Pilsner crates	Pilsner other	0.0%	Fresh	Characterful
Amstel	Amstel	Amstel	Amstel	Affligem
Grolsch	Grolsch		Corona	Brand
Heineken	Heineken		Desperados	Grolsch
Hertog Jan	Hertog Jan		Grolsch	Heineken
Jupiler	Jupiler		Hertog Jan	Hertog Jan
			Hoegaarden	Leffe
			Wieckse	

Table 4.1: Considered competitive brands in promotion prediction model

4.1.1 Input variables

The input variables for the competitive promotion prediction model are derived from the insights of the promotion behavior analysis. The following input variables (or features) are determined: $Week$, $Yearweek$, $Season$, $Holiday$, the different *Last promotion* variables and the different *Promotion Bavaria* variables. An overview of the input variables is provided in Table 4.2. A more detailed description of the input variables can be found in Appendix G.

Variable	Description	Variable type
$Week$	Week number	Discrete
$Yearweek$	A number adding up for every week in the data	Discrete
$Season$	The season	Categorical
$Holiday$	The type of holiday	Categorical
$Last\ promotion_1$	The number of weeks since the last promotion of brand-group 1 at a specific supermarket	Discrete
...	...	...
$Last\ promotion_X$	The number of weeks since the last promotion of brand-group X at a specific supermarket	Discrete
$Promotion\ Bavaria_{pilsnercrates}$	A promotion of Bavaria pilsner crates	Binary
$Promotion\ Bavaria_{pilsnerother}$	A promotion of Bavaria pilsner other	Binary
$Promotion\ Bavaria_{0.0\%}$	A promotion of Bavaria 0.0%	Binary
$Promotion\ Bavaria_{fresh}$	A promotion of Bavaria fresh	Binary
$Promotion\ Bavaria_{characterful}$	A promotion of Bavaria characterful	Binary

Table 4.2: Input variables of competitive promotion prediction model

$Week$ is a metric variable which indicates the week number. $Yearweek$ is a metric variable which indicates the number of the week in the data set. This number adds up for every week in the data starting at the first week of the data. $Yearweek$ takes a value of 54 for week 1 of 2016, while $Week$ takes a value of 1 for that week. $Yearweek$ hereby captures the increasing number of promotions during the years. $Season$ is a non-metric variable which represents one of the four seasons (winter, spring, summer or autumn). When a season change takes place in a week, $Season$ is assigned as the coming season. $Holiday$ is a non-metric variable which indicates whether a holiday takes place in that week. The holidays are Carnaval, Easter Sunday, Easter Monday, Ascension Day, Whit Sunday, Whit Monday, Kingsday, Christmas and New Years Eve. Easter Sunday and Easter Monday, for example, are taken into account separately since the analysis of competitive promotions indicated that some brands are often on promotion during one of those holidays. If Christmas and New Years Eve are in the same weeknumber, the holiday of that week is assigned as Christmas since this is a more important holiday for the promotion planning. The *Last promotion* variables are non-metric variables which stand for the number of weeks since the last promotion of a specific brand-group combination. This means that *Last promotion* is not one feature, but instead one feature for each considered brand-group combination. Lastly, promotions of Swinkels Family Brewers' own products are known in advance and the analysis of competitive promotions revealed relations between competitive promotions and promotions of Bavaria, which is why promotions of Bavaria are taken into account by the *Promotion Bavaria* variables as non-metric input variables. Only the Bavaria brand is taken into account as a Swinkels Family Brewers brand since this is their only brand in the pilsner crates and pilsner other group which showed relations to other brands in the competitive promotion analysis.

4.1.2 Dependent variable

The dependent variable for the competitive promotion prediction model is whether there is a promotion in a specific week for the brand-group combinations. To emphasize, the dependent variable is a promotion of a group and not a promotion of a specific product. The competitive promotion prediction problem is a classification problem since it has to be determined if a brand-group-week combination belongs to the category of promotion or the category of no promotion. There are several approaches to a classification task as displayed in Figure 4.1. The classification approach determines the form of the dependent variable.

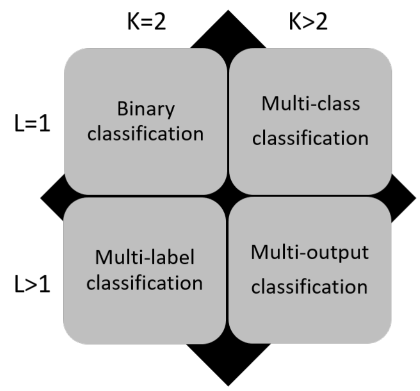

Figure 4.1: Classification approaches: L is the number of labels and K is the number of classes (source: (Read, Martino, Olmos & Luengo, 2015))

The multi-label classification is chosen as the approach for the competitive promotion prediction model since in this approach an observation can be classified into multiple classes whereas in the single-label approaches (binary classification and multi-class classification) the classes are mutually exclusive. Multi-label classification is a common approach in text categorization (Tsoumakas & Katakis, 2007). In text categorization the text categories would be the labels and the two classes for every label would be whether the text belongs to that category or not. This way, a text can be categorized in multiple categories. A multi-label classification approach is used for the competitive promotion prediction model, since multiple brands can be on promotion in the same week. The labels are the brand-groups combinations and the two classes are whether that brand-group combination is on promotion or not. This results in a multi-label classification model with the number of labels equal to the number of considered brand-group combinations with two classes each; one class for a promotion and one class for no promotion. A single observation corresponds to one week in which all labels are classified into one of the two classes.

4.1.3 Classification method

As described in Section 1.4.1, the literature is consulted to form a range of methods which could be used for the competitive promotion prediction model. For this reserach we choose to focus on neural networks, decision trees and random forests. Often raised in literature is that neural networks are hard to interpret (Lou, Caruana & Gehrke, 2012). Therefore, it is decided to stick to decision trees and random forests for the competitive promotion prediction model. As the name suggest, a decision tree has the structure of a tree and splits the data into subsets based on if-then-else decision rules. The root node represents the entire sample and is at the top of the decision tree. The sample in the root node is split on all available variables and the split which results in the most homogeneous sub-samples is selected. This way, the root node is divided into two or more sub-nodes. The new sub-node is called a decision node if and only if the decision node splits further into sub-nodes. If the new sub-node does not split further it is called a leaf node or terminal node. The leaf nodes represents a single class in the case of a single-label classification problem. The structure of a decision tree is represented in Figure 4.2. One of the main disadvantages of decision tree is overfitting. This can be overcome by the random forest algorithm. A random forest can be seen as a forest of decision trees. More specifically, a random forests consists of a number of decision trees. The final class is determined by the mode of the classes of the individual trees.

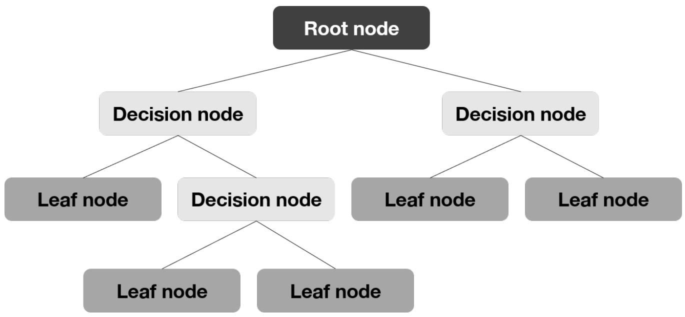

Figure 4.2: Decision tree representation

Existing, traditional algorithms only deal with single-label classification problems and not with multi-label classification problems. This is also the case with decision trees and random forests since the leaf nodes are only able to represent the classes of a single label and not of multiple labels. There are methods to handle multi-label classification problems and they can be grouped in two categories: problem transformation methods and algorithm adaptation methods. In problem transformation methods, the multi-label classification problem is transformed to one or more single-label classification problems. Algorithm adaptation methods adapt existing algorithms in order to be able to handle multi-label problems directly. (Tsoumakas & Katakis, 2007).

The most used problem transformation method is the binary relevance method. This method transforms the multi-label problem into one binary classification problem for each label. However, since the binary problems are classified separately, correlations between labels are neglected. Another common problem transformation method is the label power-set method. The label power-set method transforms the multi-label problem into a multi-class problem. For every label combination in the data, a class is created. In contrast to the binary relevance method, the label power-set method is able to model label correlations. The label power-set method is, however, criticized by its computational complexity and tendency to overfit as it can only model label combinations which exist in the training data. The third problem transformation is the classifier chains method. The classifier chain method is based on the binary relevance method but overcomes the drawbacks of the binary relevance method. Namely, the classifier chains method deals with label correlations. The classifier chain method transforms the multi-label problem into multiple binary problems. The binary classifiers to classify these binary problems are chained one after the other where each classifier takes into account the predictions of the previous classifier. Obviously, the performance of the classifier chain is heavily dependent on the order of the chain. To solve this problem ensemble of classifier chains can be used. In ensemble of classifier chains several classifier chains with a random order are trained. The final prediction for each label is equal to the majority of the predictions of the individual chains, similar to the principle of random forests. (Read et al., 2015)

4.2 Performance measure

The competitive promotion prediction model is a multi-label classification problem. The performance of multi-label classification problems can be assessed as fully correct, partially correct or fully incorrect. To give a complete picture of the performance of the model, the performance measures account for both fully correctness and partially correctness. The selected performance measures are: the Exact Match Ratio, Hamming Loss, Precision, Recall and F1-score.

The Exact Match Ratio (MR) is the fraction of observations that have all their labels correctly classified. Thereby it does not distinguish between partially correct and fully incorrect. The Exact Match Ratio is defined as follows: (Sorower, 2010)

$$MR = \frac{1}{n}\sum_{t=1}^{n} I(Y_t = \hat{Y}_t) \tag{4.1}$$

Where n is the number of observations, Y_t is the actual set of labels at time t, $\hat{Y}_t$ is the predicted set of labels at time t and I is the indicator function.

Hamming Loss (HL) is the fraction of the total number of labels of all observations which is incorrect and therefore it accounts for partial correctness. As the word itself suggests it is a loss function, which means that the optimal value is 0. Hamming Loss is calculated as follows: (Schapire & Singer, 2000)

$$HL = \frac{1}{nL}\sum_{t=1}^{n}\sum_{j=1}^{L} XOR(Y_{t,j}, \hat{Y}_{t,j}) \tag{4.2}$$

Where n is the number of observations, L is the number of labels, $Y_{t,j}$ is the actual class of label j at time t and $\hat{Y}_{t,j}$ is the predicted class of label j at time t.

It happens more often that a brand is not on promotion than it is on promotion, therefore the promotion data set is an unbalanced data set. In other words, a big part of the labels belongs to one class, namely the class of not being on promotion. In unbalanced data sets, measures such as MR and HL can reflect the underlying class distribution which is misleading. Therefore, Precision, Recall and F1-score are also used as performance measures. All three measures are micro-averaged, which means that they are calculated globally over all instances and all labels. Hereby it takes into account possible label imbalance, meaning that a label with more promotions receives more weight in the calculation.

Precision (P) is the fraction of predicted positive labels which is also a positive label in actuality. It is calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{4.3}$$

Where TP is the number of true positives and FP is the number of false positives.

Recall (R) is the fraction of actual positive labels which is also classified as a positive label. It is defined as follows:

$$R = \frac{TP}{TP + FN} \tag{4.4}$$

Where TP is the number of true positives and FN is the number of false negatives.

The F1-score (F1) is the harmonic mean of Precision and Recall. Both contribute equally to the F1-score. The F1-score is calculated as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4.5}$$

In this research, the F1-score is the most important performance measure of the above mentioned performance measures, since this measure is able to deal with unbalanced data-sets and reflects both the Precision and Recall. Therefore, the F1-score is the most prominent measure in evaluating the competitive promotion prediction model.

4.3 Experimental setup

Section 4.1 described the design of the competitive promotion prediction model. Namely, the input variables are $Week$, $Yearweek$, $Season$, $Holiday$, time since last promotion for every brand-group combination (*Last promotion*) and whether Bavaria's groups are on promotion ($Promotion\ Bavaria$). The output of the model is a binary prediction for every brand-group combination whether there is a promotion in a specific week or not. A binary prediction is chosen since this is as similar as possible to the actual promotions. This facilitates practical use of the model, especially for the use in the retail order forecasts. The model is approached as a multi-label classification model, in which every brand-group combination is a separate label.

In order to test which method works best for the prediction of competitive promotions several models are created and tested. Table 4.5 gives an overview of the models. Since there is no current prediction model at Swinkels Family Brewers or no comparable model in literature, two benchmarks models are custom made. These benchmark models are used to compare the competitive promotion prediction model with. Model 1a is a benchmark which provides a stratified random prediction which means that it takes into account the class distribution of a label. For example, suppose label X has one promotion in every 10 weeks. Then model 1a flips a coin for every week where there is a 10% probability on a promotion and a 90% probability on no promotion. Because one random prediction can happen to be much better than the other, the random stratified prediction is repeated a thousand times and the final performance is averaged over those thousand times. This is also a commonly used benchmark in customer churn prediction (Oentaryo, Lim, Lo, Zhu & Prasetyo, 2012). Model 1a takes, however, not into account the time between promotions. Therefore another benchmark is created, model 1b. Model 1b determines for every label the probability distribution on the number of weeks between promotions. The probability distribution is then used to generate the probability on a promotion in a specific week given the current number of weeks since last promotion. The probability on a promotion is obviously higher as more time has passed since the last promotion. If that probability on a promotion is greater than 0.5, a promotion is predicted for that label. Appendix H provides a more detailed explanation and the calculation for the generation of the promotion probability for model 1b. The benchmark models are chosen in a way to test whether complex machine learning models add value to the prediction compared to the simple and easily interpretable benchmark models. Models 2 and 3 approach the multi-label problem with the binary relevance solution and use a decision tree and random forest as classification method, respectively. In models 4 and 5 the label-power set approach and a decision tree and random forecast, respectively, is applied. Model 6 and 7 use ensemble of 10 classifier chains with random orders and a decision tree and random forest. Lastly, models 8 and 9 approach the multi-label problem with an adapted decision tree and an adapted random forest. The models are chosen in such a way that the best classification method and the best approach to the multi-label problem can be tested interchangeably.

Model	**Classification method**	**Multi-label approach**
Model 1a (Benchmark)	Random stratified	
Model 1b (Benchmark)	Distribution time between promotions	
Model 2 (DT-BR)	Decision Tree	Binary Relevance
Model 3 (RF-BR)	Random Forest	Binary Relevance
Model 4 (DT-LP)	Decision Tree	Label Power-set
Model 5 (RF-LP)	Random Forest	Label Power-set
Model 6 (DT-ECC)	Decision Tree	Ensemble of Classifier Chains
Model 7 (RF-ECC)	Random Forest	Ensemble of Classifier Chains
Model 8 (DT-AA)	Decision Tree	Algorithm Adaptation
Model 9 (RF-AA)	Random Forest	Algorithm Adaptation

Table 4.3: Competitive promotion prediction models

For every model the experimental setup as displayed in Figure 4.3 is executed. The setup contains an outer loop for model evaluation and an inner loop for hyperparameter optimization. The outer loop splits the data in a training set and a test set, while the inner loop splits the training set further into a training subset and a validation set. This is done to be able to make a fair overall model evaluation, since the test set is kept apart from the training of the model. In this way, the test set can be considered as 'new data'. The division of data into a training, validation and test set is also conventional in research. As Figure 4.3 points out, the evaluation of the hyperparameter optimization and the evaluation of the final model are performed based on the rolling origin recalibration technique as described by Tashman (2000). They found that the rolling origin recalibration technique improves the reliability of out-of-sample tests for single time series data. This is because when new data has become available, the model is retrained with the most up-to-date information. Changes in promotion planning behavior, which are not uncommon since it is a manual task, can hereby be detected quickly. However, the rolling origin recalibration technique requires automated data collection, which decreases practical feasibility. Therefore, a sensitivity analysis is performed on the rolling origin recalibration and the results of this are shown in Section 4.4. The rolling origin recalibration technique for the inner loop and outer loop is displayed in Figure 4.4 and 4.5, respectively.

The rolling origin recalibration technique for the inner loop is only performed on the training set of 169 weeks (i.e. 2015 week 1 - 2018 week 12). The training set is divided into a training subset of 117 weeks (2015 week 1 - 2017 week 12) and a validation set of 52 weeks (2017 week 13 - 2018 week 12). In every iteration, the model is re-estimated on the training subset and tested on one week of the validation set. The first iteration uses the complete training subset minus

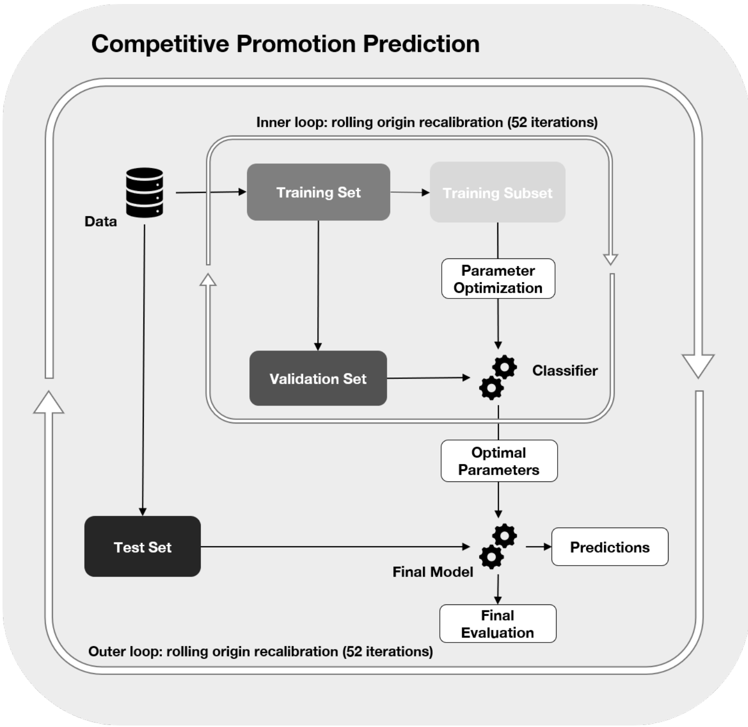

Figure 4.3: Experimental setup of competitive promotion prediction model

three weeks (because of a forecast horizon of 4 weeks) for estimating the model and the first week of the validation for prediction. In the next iteration, the first week from the validation set is added to the training subset and the model is re-estimated. The re-estimated model and the second week of the validation set are used for prediction. This process is repeated in every iteration. In this way, data from the validation set that has become available is used for recalibration of the model. This process stops when there is no more data to add to the training subset, resulting in 52 iterations. After performing the 52 iterations, predictions for the entire validation set are obtainedproduced after which the performance can be evaluated and the optimal hyperparameters can be determined based on this performance.

The rolling origin evaluation for the outer loop is performed on the entire data set of 221 weeks. The data is split in a training set (2015 week 1 - 2018 week 12) and a test set (2018 week 13 - 2019 week 12). Again, in every iteration the week from the test set which became available is added to the training set and the model is re-estimated. This results also in 52 iterations. Predictions for the entire test set are obtained in these iterations after which the overall model performance can be determined. This overall model performance is determined only on the test set and is compared to the overall performance of the other models in order to find the best performing model.

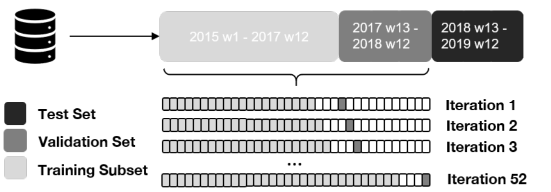

Figure 4.4: Rolling origin recalibration inner loop

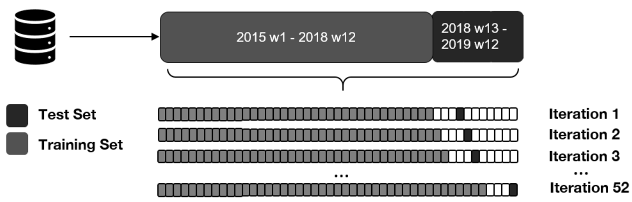

Figure 4.5: Rolling origin recalibration outer loop

4.3.1 Hyperparameter optimization

The hyperparameter optimization in the inner loop employs a grid search to test different combinations of values of hyperparameters of the decision tree and random forest. The combination of hyperparameter values with the highest performance according to the $F1$ (Equation 4.5) is considered as optimal hyperparameters. Only the $F1$ is considered in the hyperparameter optimization so there is no need to consider between the different performance measures. $F1$ is chosen since it deals well with unbalanced data sets. The overall model evaluation in the outer loop is performed according to all performance measures described in Section 4.2.

In the decision trees, the following hyperparameters are tuned. The maximum depth of the tree and the minimum number of samples for a split. The maximum depth of the tree indicates how deep the tree can be. A deeper tree can capture more information about the data, but also has a bigger tendency to overfit. The maximum depth hyperparameter is ranging from 5 to 25 with steps of 10 and also 'None' is tested. The minimum number of samples for a split represents the number of samples which is required for splitting a node. When this number is higher, the generalizability of the tree increases which can lead to underfitting. The minimum number of samples for split is varying between 2 samples and 20%, 40% and 60% of the samples.

In the random forests the same hyperparameters are tuned as in the decision tree. There is one additional hyperparameter which only applies to random forests and not to a single decision tree. Namely, the number of estimators which indicates the number of trees in the forest. The number of estimators ranges from 10 to 110 with steps of 50.

The drawback of using a gridsearch in hyperparameter optimization is that it can be computationally expensive and therefore time consuming. To deal with this initial experiments are carried out on the hyperparameter values and the final grid is based on these experiments.

4.4 Results

The models for predicting competitive promotions are created in Python using the scikit-learn (Pedregosa et al., 2011) and scikit-multilearn (Szymański & Kajdanowicz, 2017) library.

The analysis of competitive promotions already pointed out that the discovered insights in promotion behavior apply more strongly to certain groups. Therefore, the models as described in Table 4.5 are first applied to the groups separately at Albert Heijn. In this way, it can be researched if the models work better for some groups or if the model performs equally well on the separate groups. The Albert Heijn is again chosen since it is the most relevant retailer in terms of order volume. Figure 4.6 shows the F1-scores of the best performing machine learning model, the random benchmark and the benchmark based on distribution of time between promotions for every group. The machine learning model with the best performance is shown in between brackets for every group. As is visible, for every group the machine learning models perform better than the random prediction. Not all machine learning models do, however, outperform the benchmark based on distribution. This is especially the case for the the fresh group and characterful group. These groups carry the largest number of brands which causes a higher model complexity. A higher model complexity on its turn probably causes the machine learning models to perform worse. The benchmark based on distribution is not affected by model complexity and therefore performs better for these groups. The machine learning model and distribution benchmark for pilsner other group have the most outstanding performance and the machine learning model and distribution benchmark for the pilsner crates group have the second best performance. This is not surprisingly since the insights from the promotion analysis also applies most strongly to those groups. Also, the greatest amount of volume is in the two pilsner groups which may indicate that these groups are the most critical groups in promotion planning. With the same reasoning it could be argued that the other groups are less critical in promotion planning which makes them harder to predict.

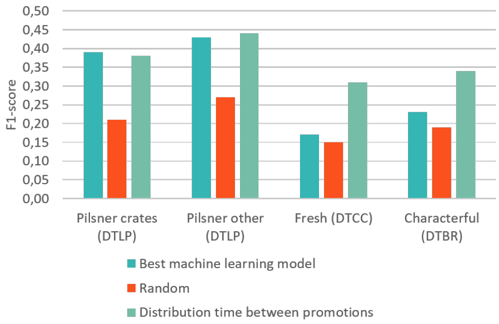

Figure 4.6: Results of Competitive Promotion Model for separate groups at Albert Heijn

The analysis of promotions showed that there are relations between the promotions of groups at Albert Heijn. The groups are combined to see if information from other groups adds value to the predictions. By this it is meant that one model is used for the prediction of both groups at once instead of multiple models for the multiple separate groups. This way, information about the time since last promotions of the one group can be used to predict promotions of the other group and vice versa. Since the two pilsner groups are performing best, those are combined to 'all pilsner'. Furthermore, the two pilsner groups contain the same brands so another combination with all the brand-groups from those brands is tested called 'pilsner brand all groups'. Finally, a combination of all brands and groups ('all groups') in one model is tested. The results are visualized in Figure 4.7. The first bars represent the separate pilsner crates group and the separate pilsner other group, which are included to enable easy comparison. The figure shows that for the three group combinations the benchmark based on the distribution of the time between promotions is performing very well. For the pilsner brands all groups combination and the all groups combination the distribution benchmark is even performing much better than the machine learning model. This is probably also due to the higher model complexity for these groups. For the all pilsner combination the distribution benchmark is performing almost equally well as the best machine learning model. Combining the pilsner crates group with the pilsner other group is considered valuable since the performance has increased compared to pilsner crates and remained approximately the same as the pilsner other group. Also the other combinations are promising.

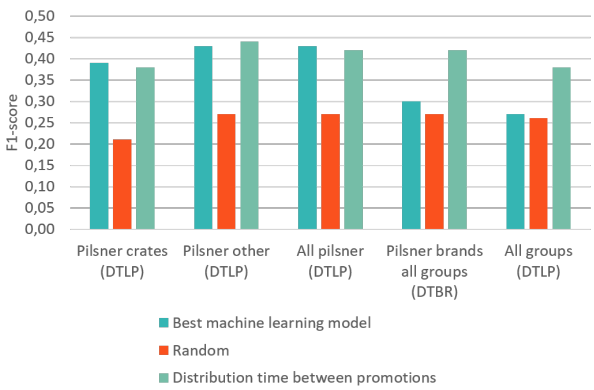

Figure 4.7: Results of Competitive Promotion Model for combined groups at Albert Heijn

Up to now the best results in terms of F1-score have been achieved on DTLP applied to all pilsner combination (F1 = 0.43), the benchmark model 1b applied to all pilsner combination

(F1 = 0.42) and the benchmark model 1b applied to the pilsner brands all groups combination (F1 = 0.42). The other performance measures of these are displayed in Table 4.4. When only taking into account the F1-score the DTLP applied to all pilsner performs best. However, it can be argued that the benchmark model 1b is easier to interpret than the DTLP. The DTLP has also a higher MR, a lower HL and a higher Precision than the benchmark 1b applied to all pilsner and pilsner brands all groups. Therefore, despite its poorer interpretability, the DTLP applied to all pilsner is selected as the best performing model at Albert Heijn so far.

Applied to	**Model**	**HL**	**MR**	**Precision**	**Recall**	**F1**
All pilsner	Benchmark 1b	0.45	0.00	0.35	0.54	0.42
	DTLP	0.35	0.19	0.46	0.43	0.43
Pilsner brands all groups	Benchmark 1b	0.44	0.00	0.36	0.52	0.42

Table 4.4: Models with highest F1-score

To provide a compact summary, Table 4.5 shows the performance of the other models applied to all pilsner at Albert Heijn. The optimal hyperparameters are also displayed in brackets, in which the first number corresponds to the maximum depth of the tree, the second number corresponds to the minimum number of samples for a split and the third number, if applicable, corresponds to the number of estimators. The numbers in bold are the best scoring numbers on that particular performance measure. Table 4.5 shows that the best performing model to predict promotions in all pilsner groups is the decision tree with label powerset (DTLP). It outperforms the F1-score of all benchmarks and other machine learning models. The Recall of benchmark 1b is, however, not outperformed by the DLTP. As the precision of benchmark 1b is a lot lower, it can be concluded that benchmark 1b often predicts a promotion that is not correct. The DTLP is thus better at correctly predicting when there is no promotion which is also valuable information. Additionally, the DTLP has a higher MR than all other models. The MR is 0.19 which means that 19% of the weeks is classified completely correct. This confirms that there are indeed correlations between the promotions of both pilsner groups which are captured by the label power-set. This makes the binary relevance approach less suitable which is also apparent from the results. The ensemble of classifier chains, on the other hand, is suitable in the case of label correlation, however it cannot be guaranteed that the order of the individual chains is optimal, assuming there is an optimal order. This could therefore be the reason why it is not performing well. The algorithm adaption does also take into account label correlation but is apparently not suitable to this problem. A drawback of the label powerset method is that it is only able to predict label combinations which are observed in the training set. That the label powerset performs well could mean that the training set in this situation contains all possible label combinations or that all occurring label combinations in the training set are also represented in the test set. The latter is more likely which would mean that the test set is a good representation of the data. Additionally, the table shows that almost all random forests perform worse than the decision trees with the same approach to the multi-label problem. For example, the RFBR has a lower F1-score than the DTBR. This can be explained by the fact that a single decision tree can happen to give a good result, while an ensemble of trees gives a more stable and thus a more reliable result. In the case of the label powerset approach, the random forest and decision tree perform equally well. This indicates that the predictions of the trees are stable and that all trees in the ensemble return the same prediction. The DTLP is chosen as the best performing model since the running time is substantially shorter than the running time of the RFBR. In essence, the RFBR would also have been a good choice as best performing model.

Forceast horizon	Models	Optimal Hyperparameters	HL	MR	Precision	Recall	F1
4	Benchmark 1a		0.39	0.01	0.32	0.27	0.27
	Benchmark 1b		0.45	0.00	0.35	**0.54**	0.42
	2 DTBR	[25, 0.2]	0.42	0.02	0.31	0.28	0.28
	3 RFBR	[25, 0.2, 10]	0.34	0.00	0.22	0.12	0.14
	4 DTLP	[25,0.2]	0.35	**0.19**	**0.46**	0.43	**0.43**
	5 RFLP	[25, 0.2, 60]	0.34	0.15	**0.46**	0.42	0.42
	6 DTECC	[5, 0,2]	0.41	0.00	0.31	0.29	0.28
	7 RFECC	[25, 0.2, 110]	**0.31**	0.00	0.09	0.05	0.06
	8 DTAA	[25,0.2]	0.34	0.02	0.21	0.12	0.15
	9 RFAA	[25, 0.2, 10]	0.32	0.00	0.08	0.03	0.04

Table 4.5: Results competitive promotion prediction 'all pilsner' at Albert Heijn

4.4.1 Sensitivity analysis

A sensitivity analysis is performed on the best performing model: DTLP for all pilsner at Albert Heijn. The analysis is focused on two important components of the model setup, namely the forecast horizon and the number of iterations in the rolling origin recalibration. The forecast horizon is a requirement of the business at which the research is applied, but other businesses might require a different horizon. The rolling origin recalibration has been included in the sensitivity analysis since this affects the practical feasibility of the model.

First, it is tested whether the prediction performance improves when the forecast horizon is reduced. The horizon is reduced from 4 to 3, 2 and 1 weeks and the results are shown in Table 4.6. All performance measures remain approximately the same so it can be concluded that the prediction performance does not improve nor does it strongly decrease. Apparently, reducing the forecast horizon decreases the uncertainty barely. The input variable which depends most on the forecast horizon is the *Last promotion* variables. The other variables ($week$, $Yearweek$, $Holiday$ and the $Promotion\ Bavaria$ variables) are known regardless of the forecast horizon. To put it differently, when predicting a promotion 4 weeks ahead it is already known which weeknumber it is and whether there is a holiday or promotion of Bavaria in that week. However, the time since last promotion is more uncertain since it could be that there is a promotion in the weeks between now and the week which is forecasted. This uncertainty increases as the forecast horizon increases. This uncertainty is, however, taken into account in the model by training the model also on the time since last promotion since the forecast horizon. In case of a forecast horizon of 4 weeks, the model is trained on the time since last promotion of 4 weeks ago instead of the time since the last recent promotion (which could be between now and the forecasted week). This could explain why the performance of the model is not very sensitive to the forecast horizon.

Forceast horizon	HL	MR	Precision	Recall	F1
4	0.35	0.19	0.46	0.43	0.43
3	0.32	0.17	0.48	0.42	0.43
2	0.36	0.10	0.44	0.36	0.38
1	0.33	0.10	0.47	0.42	0.41

Table 4.6: Sensitivity analysis on the forecast horizon

Lastly, it is investigated how the number of iterations in the rolling origin recalibration method influences the prediction performance. The number of iterations is reduced from 52 to 13, 4 and 1 iterations. In case of 52 iterations, the model is recalibrated every week, 13 iterations means the model is recalibrated every four weeks, 4 iterations means the model is recalibrated every quarter of a year and 1 iteration means the model is recalibrated every year. Table 4.7 shows the results of the sensitivity analysis on the number of iterations. Again, the performance barely changes which means the model is not sensitive to the number of iterations in the recalibration method. This implies that the recalibration period can be increased which increases the practical feasibility while the prediction performance is maintained.

Number of iterations	HL	MR	Precision	Recall	F1
52	0.35	0.19	0.46	0.43	0.43
13	0.35	0.19	0.43	0.39	0.41
4	0.32	0.19	0.47	0.39	0.42
1	0.33	0.19	0.46	0.43	0.44

Table 4.7: Sensitivity analysis on the number of iterations in the rolling origin recalibration method

4.4.2 Other supermarkets

Before, the analysis and prediction of competitive promotions focused mainly on Albert Heijn since it is Swinkels Family Brewers' biggest customer and since Albert Heijn has the highest number of promotions. The group combination which works best for prediction at Albert Heijn is also tested separately at other supermarkets. Other supermarkets include Boni, COOP, Hoogvliet, Jan Linders, Plus, Poiesz and Vomar. As discussed before, the promotions of Jumbo are out of scope. Figure 4.8 shows the results of the promotion prediction for all pilsner at the other supermarkets. Almost all models perform better than random which indicates that the features carry not only relevant information for the Albert Heijn. At some supermarkets, Hoogvliet (F1 = 0.23) and Vomar (F1 = 0.24) for example, the models perform better than at other supermarkets. However, not a single model at one of the supermarkets approaches the prediction performance at Albert Heijn.

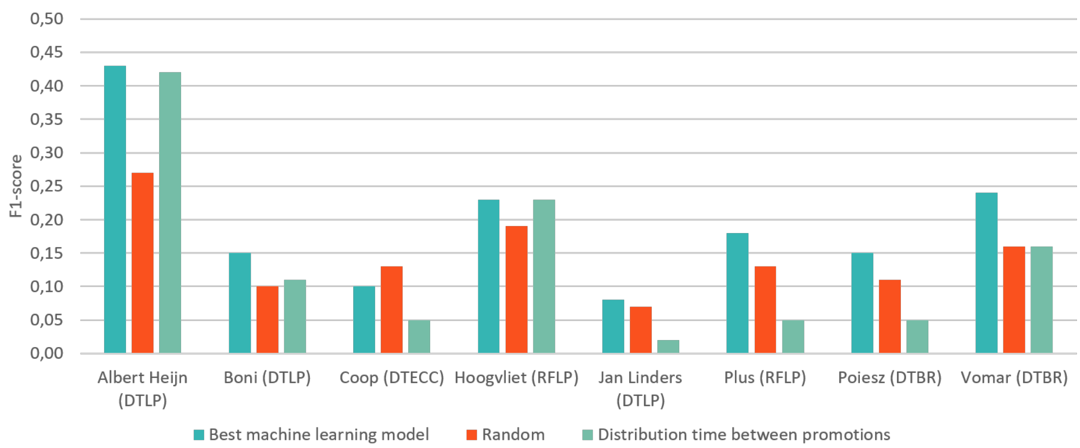

Figure 4.8: Results of Competitive Promotion Model for 'all pilsner' at all supermarkets

Another interesting question is whether the promotions at other supermarkets carry information which could help to predict promotions at Albert Heijn. It could be the case that brands plan the promotions of their products not in the same week at multiple supermarkets, for example. To test whether this is the case, the best performing model so far (the model for all pilsner) is extended to other supermarkets. This extension is two-fold. Namely, the information about historic promotions at other supermarkets (i.e. the time since last promotion of brand X at supermarket Y) could be included for predicting only promotions at Albert Heijn. On the other hand, the information about historic promotions of all supermarkets could serve to predict the promotions at all supermarket at once. The results of these extensions are shown in Figure 4.9. The performance of the two benchmark models do not change when all supermarkets are included as input and the promotions at Albert Heijn are still the output variables. This can be explained by the fact that the benchmark models are only based on the input from the predicted labels. More specifically, the benchmark models are in this case only based on promotions at Albert Heijn and are not affected by the promotions of other supermarkets. The performance of the machine learning model is lower when all supermarkets are included as input and promotions at Albert Heijn are the output. The pilsner promotions at other supermarkets, evidently, do not carry enough information which could help to predict promotions at Albert Heijn. Albert Heijn is the biggest supermarket formula in the Netherlands which could explain why the promotions at Albert Heijn do not depend on promotions at other supermarkets. The other supermarkets in the data set are substantially smaller than Albert Heijn. The second biggest supermarket chain (Jumbo), also significantly bigger than the other supermarkets, is out of scope for the competitive promotion prediction model as discussed before. It could be the case that promotions at Jumbo, in contrast to promotions at the other supermarkets, do carry information about Albert Heijn's promotion, which is an interesting question for further research. When promotions at all supermarkets are included as both input and output, i.e. promotions at all supermarkets are predicted, the random model has the highest F1-score indicating that the promotions at all supermarkets are not predictable with the current model design. An explanation for why incorporating all supermarkets in one model does not give satisfying results is that other supermarkets could employ different promotion strategies than Albert Heijn causing the chosen features to be less relevant for those supermarkets.

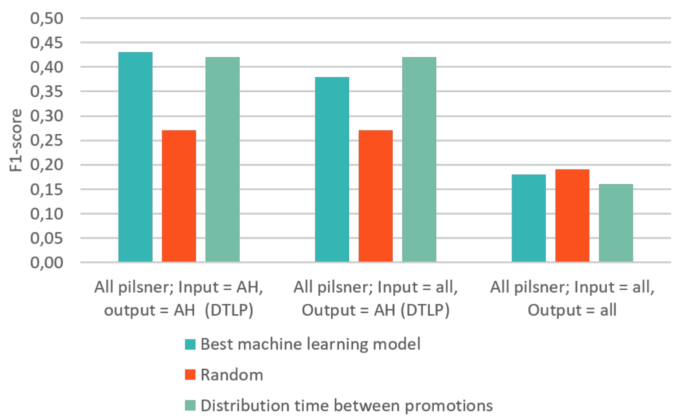

Figure 4.9: Results of Competitive Promotion Model for incorporating other supermarkets in 'all pilsner' at Albert Heijn

4.5 Summary

This chapter described the competitive promotion prediction model. Several experiments are executed and almost all models in the experiments perform better than a random model and also better than a model which is based on a simple rule. This means that the invented models make sense and machine learning does add value to the prediction. The best predictions are obtained for promotions at the Albert Heijn, on which also the analysis is performed where the features are derived from. More specifically, the best predictions are obtained by predicting promotions in the pilsner crates and pilsner other group at Albert Heijn. The performance is, however, hard to judge since there is neither a model in practice nor a model in literature to compare the performance with. Therefore, the next chapter provides more insights into the practical value of the competitive promotion prediction model.

Chapter 5

Retail order forecast model

This chapter describes the retail order forecast model. First, the model design is described. Second, it is described how the experiment is setup and finally the results of the retail order forecast model are presented. The aim of the retail order forecast model is to forecast the orders of retailers as precise as possible. Furthermore, the retail order forecast model should address whether the performance of the competitive promotion prediction model as described in Chapter 4 is good enough to provide practical value. That is, is the performance of the competitive promotion prediction model sufficient to improve retail order forecast accuracy.

5.1 Model description

The retail order forecast model should forecast retail orders as precise as possible, since accurate forecasts are a resolution for the trade-off between services and costs. The retail order forecast model as described in this chapter only provides forecasts for the weeks around a promotion: the two weeks prior to a promotion week, the promotion week itself and the week following a promotion week. These weeks are defined as the promotion weeks. In the weeks not in vicinity of a promotion week (i.e. the baseline weeks), the current forecast method is used. This is decided since the analysis of the current forecasting performance (Section 2.5.2) showed that the biggest opportunity for improvement is in the weeks around a promotion. Besides, experiments pointed out that only forecasting weeks around a promotion with the retail order forecast model and using the current forecast method in the baseline weeks improves the accuracy compared to forecasting all weeks with the retail order forecast model.

The forecasts serve as guidelines for the production planning and therefore the forecasts are required four weeks upfront, i.e. the forecast horizon is four weeks due to production lead times. The forecasts should be provided on product-retailer-week level, which means that a SKU sold to a specific retailer in a specific week is treated as a single observation. For every product-retailer a separate model is built. Experiments have also been carried out with all product-retailer combinations in one model in which the product-retailer combination can learn from related combinations, but this does not benefit the model performance.

5.1.1 Dependent variable

The dependent variable is the weekly order amount of a specific product-retailer combination in hectoliters. The dependent variable is a continuous variable. As described in Section 2.5.1 the forecast will be assessed with the Mean Absolute Error (MAE) and the Mean Absolute Promotion Volume Deviation (MAPVD). The MAE measures the average absolute deviation per week between the forecast order and the actual order for every product-retailer combination. The MAPVD measures the average absolute deviation per promotion between the forecast promotion volume and the actual promotion volume for every product-retailer combination. Promotion volume is defined as the sum of orders in the promotion week and the two weeks before the promotion week. As described before, a separate model is built for every product-retailer combination. The MAE and MAPVD are calculated for every product-retailer combination and thereafter this is averaged over all product-retailer combinations. The MAE and MAPDV give thus an average deviation per product-retailer combination per week (MAE) or per promotion period (MAPDV).

5.1.2 Input variables

The input variables for the retail order forecast model are based on literature, expert interviews and outcome of the competitive promotion prediction model. Frequently used explanatory variables for sales forecasting in literature are price and promotional information of the concerned product, seasonality, holidays and events, and orders in the previous weeks (Cooper et al., 1999; Divakar et al., 2005; Ali et al., 2009). Literature also suggests that a manufacturer should use POS data instead of order data in forecasting (Kiely, 1998; Lapide, 2005; Holmström et al., 2002; Williams & Waller, 2010). Therefore the POS of the previous weeks are also a possibility as input variables in addition to orders in the previous weeks. Additionally, stockpiling behavior during promotions and holidays is important to incorporate (Ailawadi et al., 2006). As a results from stockpiling, the demand in the week after a promotion is often lower compared to a week which is not following a promotion. Besides, experts from Swinkels Family Brewers pointed out that retailers can purchase products with a discount one or two weeks before the promotion week in the supermarket. These weeks are called the loading weeks. Also temperature is often used as independent variable in sales forecasting. However, since the forecast horizon is four weeks and weather predictions are rather unreliable four weeks upfront it is decided to not take into account temperature.

Summarizing, the initial input variables derived from literature, expert interviews and the competitive promotion prediction model are: the weeknumber, the season, a variable indicating whether the forecasted week is promotion week, a variable indicating whether the forecasted week is following a promotion week, a variable indicating whether the forecasted week is prior to a promotion week, a variable indicating whether the forecasted week is two weeks prior to a promotion week, a variable indicating whether there is a second placement in the case of a promotion, a variable indicating the promotion mechanism in the case of a promotion, a variable indicating whether there is a holiday in the forecasted week, a variable indicating whether the forecasted week is prior to a week with a holiday, historic POS for the last four weeks, historic orders for the last four weeks and (predicted) promotions of competitors. Previous work showed that also the percentage discount offered to consumers in a promotion is proved to be a good predictor. However, this information is not available at Swinkels Family Brewers and therefore this is not included as initial input variable. That these initial input variables work well in previous literature does, however, not mean that they also work well for the products at Swinkels Family Brewers. By means of a correlation matrix it was checked which input variables correlate highly with the dependent variable. The input variables which correlate lower with the dependent variable are removed one by one while keeping an eye on the MAE and MAPVD. This only leaves variables in the model which have a positive effect on the forecast performance. Furthermore, experiments have also been carried out on predicting the promotional POS and then distributing this volume over the promotion week and loading weeks, since Swinkels Family Brewers is still interested in the orders per week instead of only the promotional volume. The distribution over the promotion weeks and the loading weeks was done with a fixed percentage based on historic data, but hereby the model was not able to handle the varying purchasing behavior of a retailer. The variety of this behavior is demonstrated with the standard deviations of the percentages of the division of the order amount over the loading weeks and the promotion week in Appendix F. Therefore, it was also tested how the model would behave if a forecast of POS would be included as a input variable. How to cope with the varying division of promotion volume over the loading weeks and promotion week is, however, a fruitful avenue for future research.

The resulting and final input variables are shown in Table 5.1. As one can see, the variables concerning historic orders, historic POS and forecasted POS have not made it to the final input variables. The retail orders around a promotion week are apparently not affected by the retail orders and sales in the supermarkets in the weeks before. Morever, the forecast horizon is four weeks which means that the historic orders and historic POS are at least four weeks away from the forecasted week. Also the forecasted POS does not seem to be a good predictor of the retail orders. However, this can also be the result of inadequate POS forecasts. This could be investigated further but that is left out of scope for this research and the forecasted POS is removed as an input variable. Also the weeknumber and season are removed from the input variables as they also increased the forecast error.

Variable	**Description**	**Variable type**
$Promotion$	A promotion in the forecasted week	Binary
$Promotion_{following}$	A promotion in the week before the forecasted week	Binary
$Promotion_{prior}$	A promotion in the week after the forecasted week	Binary
$Promotion_{prior2}$	A promotion in two weeks after the forecasted week	Binary
$Mechanism$	The mechanism (i.e. type of discount) of the promotion in case of a $promotion$, $promotion_{following}$, $promotion_{prior}$ or $promotion_{prior2}$	Categorical
$Secondplacement$	Whether the product is placed on second placement in case of a $promotion$, $promotion_{following}$, $promotion_{prior}$ or $promotion_{prior2}$	Binary
$Holiday$	The type of holiday in the forecasted week	Categorical
$Holiday_{prior}$	The type of holiday in the week after the forecasted week	Categorical
$CompetitivePromotion_1$	A (predicted) promotion of competitive brand-group 1	Binary
...	...	...
$CompetitivePromotion_X$	A (predicted) promotion of competitive brand-group X	Binary

Table 5.1: Input variables of retail order forecast model

Important to mention is that the final input variables $Mechanism$ and $Secondplacement$ do not only concern the promotion week itself but also the weeks prior to and following a promotion week. Namely, a different promotion mechanism affects not only the orders in the promotion week but also the orders in the loading weeks and the same holds for if there is a second placement during a promotion. The $CompetitivePromotion_{brand-group}$ input variables are based on the predictions of the competitive promotion prediction model of Chapter 4. The $CompetitivePromotion_{brand-group}$ are binary variables indicating a promotion (1) or no promotion (0) of a competitive brand. The predicted competitive promotions carry a certain degree of uncertainty, which is not captured by a binary variable. Therefore, also continuous variables indicating the probability that a competitive brand is on promotions are tried. This was, however, less effective than the binary variables and therefore the definitive input variables include binary variables for the promotions of competitors. The promotions of competitive brands are on brand-group level and only concern promotions from the pilsner crates and pilsner other groups at Albert Heijn, as the competitive promotion prediction model performs best on these groups. The promotions of competitive brand-groups are only included for products from the same group which are sold at the same supermarket. For example, for the forecasts of the Bavaria's pilsner crate Albert Heijn combination only the promotions of brands from the pilsner group are included, whereas for forecasts of the Bavaria's pilsner crate Jumbo combination no competitive promotions are included. Both predicted and actual competitive promotions are tested as input variables in order to compare the performance of predicted and actual competitive promotions. A more extensive description of the input variables of the retail order forecast model is presented in Appendix I.

5.1.3 Regression method

The dependent variable of the retail order forecast model is a continuous variable and thereby the forecasting method should be able to predict a continuous variable. This is in contrast to the competitive promotion prediction model which predicts a categorical variable. Methods for predicting continuous variables are called regression methods. Section 1.4.1 already described that the range of methods for the retail order forecast model consists of the support vector machine and regression tree. The regression tree is chosen as the regression method for the retail order forecast model since it has an outstanding performance when the number of input variables is high which is the case when taking into account competitive promotions. Moreover, the hyperparameter choice of the kernel is a big issue in using the support vector machine and it is the kernel that makes the interpretation of the model very difficult. Also the training time of a support vector machine can be very high, which decreases the practical feasibility.

5.2 Experimental setup

Section 5.1 described the design of the retail order forecast model. Namely, the input variables concern a dummy variable for whether the forecasted week is a promotion week, a dummy variable for whether the week is following a promotion week, a dummy variable for whether the week is prior to a promotion week, a categorical variable indicating the promotion mechanism, a dummy variable for whether there is a second placement during the promotion, a categorical variable for the holiday in the forecasted week and a categorical variable for the holiday in the week prior to the holiday and dummy variables for the promotions of competitive brand-groups. The output of the model is the order in HL for a product-retailer combination in a specific week.

Model	Regression method	Competitive promotions?
Model 1 (benchmark)	Current forecasting method	No
Model 2 (DT)	Regression Tree	No
Model 3 (DT-ACP)	Regression Tree	Yes, actual competitive promotions
Model 4 (DT-PCP)	Regression Tree	Yes, predicted competitive promotions

Table 5.2: Retail Order Forecast models

The main subject of this part of the research is to examine the added value of predicted promotions. Therefore a number of models as presented in Table 5.2 are created and tested. The first model is the benchmark model consisting of the forecasts made with the current forecasting method as explained in Section 2.3. This method does not incorporate competitive promotions. In contrast to model 1, model 2 does incorporate historic promotions of the forecasted product and learns from them. Model 2 uses all input variables as listed in Table 5.1 except for the promotions of competitive brand-groups. Model 3 and model 4 use the same model as model 2 but also incorporate actual competitive promotions and predicted competitive promotions, respectively. Model 4 can thus show whether the predicted promotions by the competitive promotion prediction model improve the retail order forecast accuracy.

As described before, a separate model is built for every product-retailer combination and for every product-retailer combination model 1 till 4 is executed according to the experimental setup in Figure 5.1.

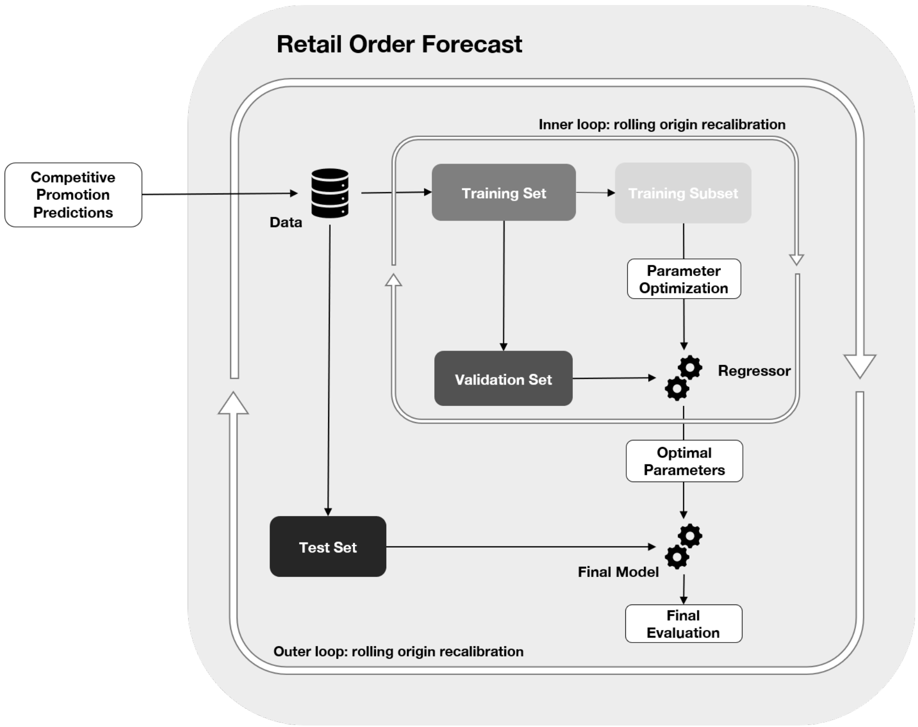

Figure 5.1: Experimental setup of retail order forecast model

First, the predicted competitive promotions for the two pilsner groups at Albert Heijn are added to the data. These are predicted by the decision tree with label power-set (DTLP) with optimal parameters which is trained on data from 2015 and 2016 and with the rolling origin recalibration method predictions are provided from 2017 on wards. These predictions are added to the input data for the retail order forecast model which is from week 1 of 2017 to week 12 of 2019 since the internal promotion data is only starting at 2017. In total there are thus 116 weeks in the data, which is a representative amount as validated with forecasting literature (Ali et al., 2009). Second, the data is split into a training, validation and test set. Again, the outer loop consists of the model evaluation and splits the data into a training set (2017 week 1 - 2018 week 39) and test set (2018 week 40 - 2019 week 12) while the inner loop consists of the hyperparameter optimization and splits the training set further into a training subset (2017 week 1 - 2018 week 14) and a validation set (2018 week 15 - 2018 week 39). The test set is only used for assessing the performance and not for model training.

Both the outer and inner loop exploit the rolling origin recalibration, which means that new available data is added to the training (sub)set and the model is retrained with the updated data. The number of iterations in the rolling origin recalibration varies per product-retailer combination since the number of promotions differs for all product-retailers. For every product-retailer combination holds that as soon as a promotion is over and the information and order amount of a promotion is known, this is added to the training data. The model only provides forecasts for the weeks around a promotion, so in order to perform the final evaluation the forecasts from the current forecast method for the baseline weeks are added to the promotional forecasts. This way, the final evaluation is performed on the same test size of 25 weeks for every product-retailer combination regardless of the number of promotions in the test set. It is hereby assumed that the number of promotions in the test set is in proportion with the total number of promotions in the data set.

5.2.1 Hyper parameter optimization

The inner loop consists of an optimization of the hyperparameters. Different combinations of hyperparameters are tested in a gridsearch to find the optimal combination. The optimal combination is the combination with the lowest MAE. The following hyperparameters are tuned in the retail order forecast model; the maximum depth of the tree and the minimum samples for a split. Those are the same hyperparameters as which are tuned in the competitive promotion prediction model. The maximum depth hyperparameter is again ranging from 5 to 25 with steps of 10 and also 'None' is tested. The minimum number of samples for a split is again varying between 2 samples and 20%, 40%, and 60% of the samples. Again, the grid is based on experiments with the hyperparameter values.

5.3 Results

The retail order forecast models as described in Table 5.1 are created in Python using the scikit-learn library (Pedregosa et al., 2011) and the results of the models are presented in this section.

The competitive promotion prediction model performs best on the two pilsner groups at Albert Heijn and therefore the results on Swinkels Family Brewer's products belonging to this group and retailer are presented first. More specifically, first the results are presented on the forecasts for products belonging to the pilsner crates and pilsner other group at Albert Heijn. This selection includes six products from two brands of Swinkels Family Brewers of which two products belong to the pilsner crates group and four belong to the pilsner other group. Table 5.3 and Figure 5.2 show the results of the retail order forecast model for the pilsner products at Albert Heijn. The optimal parameters are not shown in this table since that varies per product-retailer. The table also shows the average order per week per product and the average promotion volume per promotion per product.

Model	**MAE (HL)**	**Average weekly order (HL)**	**MAPVD (HL)**	**Average promotion volume (HL)**
Model 1 (benchmark)	162.97	381.90	517.48	1571.83
Model 2 (DT)	129.53	381.90	213.65	1571.83
Model 3 (DT-ACP)	141.32	381.90	361.05	1571.83
Model 4 (DT-PCP)	156.92	381.90	542.10	1571.83

Table 5.3: Results retail order forecast model pilsner products at Albert Heijn

Model 2 performs a lot better than model 1. The MAE decreases by 33.44 HL and the MAPVD decreases by 303.83 HL. The MAE of model 3 increases compared to model 2, but is still lower than the MAE of model 1. The same holds for the MAPVD of model 3. The MAE of model 4

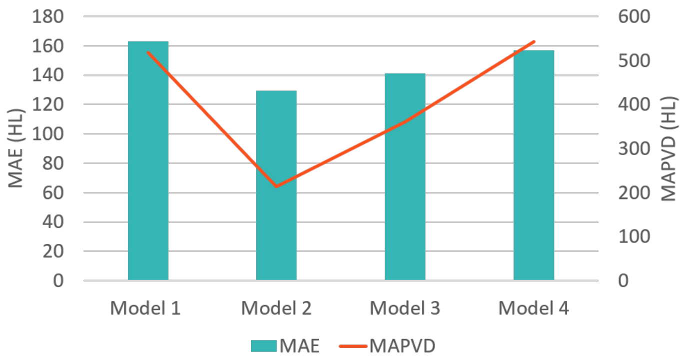

Figure 5.2: Results retail order forecast model pilsner product at Albert Heijn

increases again compared to the MAE of model 3 and is still lower than the MAE of model 1. The MAPVD of model 4 is, however, the highest of all models.

These results show that using data of all historic promotions (model 2) improves the current forecast accuracy substantially. Using all historic promotion data improves the MAE by 20.5% and the MAPVD by even 58.7%. The improvement can mainly be attributed to the bias created by judgemental adjustments in the current forecasting method. Also a big part is due to taking into account holidays and information of all historic promotions in model 2 instead of only the order volume of last promotion. This is tested with a method which can be considered the same as the current method but without judgemental adjustments. More specifically, in this experiment the promotion volume is only based on the most recent promotion and no manual adjustments are made. This method has a MAE of 149.05 HL and a MAPVD of 387.35 HL. These errors are lower than the errors in the current method but higher than the errors in model 2, which confirms that the judgemental adjustments decrease the performance and that taking into account more data on the promotions is very beneficial. The percentage difference in MAPVD between the current method and model 2 is higher than the percentage difference in MAE because it could be that the total promotion volume is forecasted well but the distribution of the promotion volume over the promotion week and loading weeks is incorrectly. A good example of this is visible in Figure 5.3, especially in the weeks around the promotion in week 42 of 2018. The total volume in the promotion week (week 42) and the loading week before the promotion week (week 41) seems to be forecasted well, but it is not distributed correctly over the weeks. Namely, the actual orders are lower in the promotion week than in the week before the promotion, whereas the forecasted orders are higher in the promotion week than in the loading week.

Furthermore, the results show that including the data on the promotions of competitors does not increase forecast accuracy. This is contrary to expectations. It would be expected that the promotion volume could be forecasted more accurately when the promotions of competitors are known. The MAE also decreases when only including the promotions of brand-groups with the highest correlation with the retail orders. This could be due to the low number of samples in which Bavaria is on promotion with other brands. For example, Bavaria's pilsner crate was on promotion 18 times in the data set, of which 11 times with only Grolsch' crate, twice with Amstel's and Grolsch' crates, twice with Jupiler's and Hertog Jan's crate, once with only Hertog Jan's crate, once with Amstel's and Hertog Jan's crate and once with no other brands. Some brands were thus on promotion together only once or twice in the dataset causing the model to have too few observations to learn from. Another explanation is that the total number of observations in the dataset is low. When the number of observations in a dataset is low, a higher number of input variables increases the model's tendency to overfit on the trainingset. In other words, when the number of observations in a dataset is low, a less complex model is

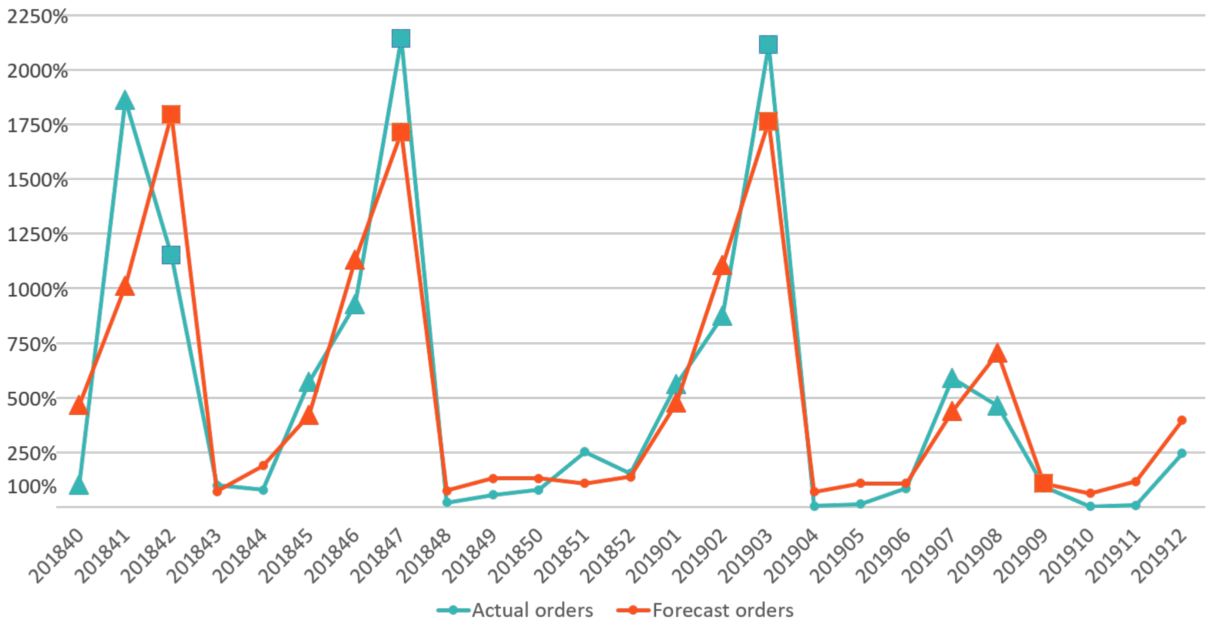

Figure 5.3: Actual orders and forecast order by model 2 of one SKU at Albert Heijn. A dot stands for a baseline week, a square stands for a promotion week and a triangle stands for a loading week. 100% corresponds to the average baseline volume.

often preferred over a more complex model. Model 2 can be considered as a less complex model compared to model 3 since it has less input variables and this could could be the reason that model 2 performs better than model 3.

Finally, the results show that the model which includes predicted competitive promotions does not perform as good as the model which incorporates the actual competitive promotions. This is obvious since the predicted promotions are not totally correct. That model 4 does not outperform model 2 is also legitimate since model 3 does not outperfom model 2 either. To put it differently, the actual competitive promotions do not improve forecast accuracy and neither do the predicted competitive promotions.

5.3.1 Other supermarkets

Before, the retail order forecasts were only focused on pilsner products sold at Albert Heijn. This section shows the results of the retail order forecast model when it is applied to the 147 product-retailer combinations in the data set. Table 5.4 shows the results. The MAE of model 2 is 1.29 HL lower than the MAE of model 1, which is a decrease of 4.49%. The MAPVD decreased by 37.5% to 82.04 HL. The percentage decrease on both measures is lower than the model applied to only pilsner products at Albert Heijn since the proportion of products which have no or few promotions is bigger when including all product-retailer combinations and the retail order forecast model only provides a promotional forecast. Also, input variables are optimized for the buying behavior of Albert Heijn. For another retailer it might be more optimal to remove the variable which indicates whether the forecasted week is two weeks before a promotion week because that retailer only receives a discount one week before the promotion. Because there is a separate model for every product-retailer this can be tested and optimized for every product-retailer in the future. Nevertheless, model 2 still outperforms the current forecasting method on all 147 product-retailer combinations which shows the potential of the model.

Model	**MAE (HL)**	**Average weekly order (HL)**	**MAPVD (HL)**	**Average promotion volume (HL)**
Model 1 (benchmark)	28.76	53.92	131.29	379.42
Model 2 (DT)	27.47	53.92	82.04	379.42

Table 5.4: Results retail order forecast model all product-retailer combinations

5.4 Summary

This chapter described the retail order forecast model. Model 2, the model which is based on all historic promotion data but does not incorporate competitive promotions, is the best performing model. The average absolute deviation per product per week between the actual order and forecast error in model 2 decreases by 20.5% compared to the current forecasting method when only considering pilsner products at Albert Heijn. For all pilsner products at Albert Heijn this is an improvement of approximately 200 HL in total per week, which amounts to more than 66,000 bottles of beer. When only considering the deviation of promotion volume, this forecast deviation decreases by 58.7% compared to the current forecasting method. This amounts to approximately 102,280 bottles of beer on average per promotion. Model 2 does not only perform excellent on the pilsner products at Albert Heijn, it also outperforms the current forecast method when considering all 147 product-retailers. Incorporating competitive promotions does, however, not improve forecast accuracy what could be attributed to the small number of observations in the data set which causes a less complex model to be more preferable over a more complex model.

Chapter 6

Conclusions and future research

This research tried to answer the question: 'Are the promotions of competitors predictable and do predicted promotions of competitors improve retail order forecast accuracy?' Therefore an analysis on the promotion planning of competitors was carried out, multiple models were designed and tested to predict promotions of competitors and the predictions of the best model were integrated in retail order forecasts.

The promotions of competitors were analyzed by means of various visualization techniques. The results showed that the found promotion planning behavior most strongly applies to the promotions of the biggest brands and biggest product groups. This can be explained by the fact that these brands and these products generate the highest revenue for a retailer and therefore have more to say in the supermarket's promotion calendar. The insights of the competitive promotion analysis were translated into input variables for the competitive promotion prediction model. This model was approached as a multi-label classification problem. The label power-set approach together with a decision tree resulted in a model which is particularly suitable for predicting competitive promotions of the pilsner groups. The performance of this model was better than that of a random model and a model based on a simple rule, proving that the developed model is able to learn from experience. The model is able to predict 43% of the promotions correctly and 46% of the predicted promotions is also a promotion in reality. The predictions of the competitive promotion prediction model were incorporated in the retail order forecast model in order to test the practical value of predicted competitive promotions. The retail order forecast model based on data of all historic promotions of only the forecasted product showed a significant improvement compared to the current forecasting method. More specifically, the performance improvement corresponds to an improvement of 20.5% overall and 58.7% in the weeks around a promotion. This improvement is probably due to the elimination of systematic bias in the current forecasting method created by judgemental adjustments. Additionally, the created model incorporates all available information of all historic promotions. The incorporation of the predicted competitive promotion into the retail order forecast model did not result in a greater improvement. However, incorporating actual competitive promotions did not result in a greater improvement either.

This researched aimed to assess the predictability of promotions of competitors. Furthermore it was assessed whether the incorporation of predicted promotions of competitors improves retail order forecast accuracy. The competitive promotion prediction model showed that the promotions of competitors are predictable to a certain extent. The predicted promotions of competitors did, however, not improve retail order forecast. Therefore, the practical value of the predictions could not be assessed, nor can be concluded that the promotions of competitors do not affect retail orders. Previous literature has namely already shown that the incorporation of competitive promotions is beneficial to the forecasting performance (e.g. Divakar et al., 2005; Ali et al., 2009; Huang et al., 2014). An explanation as to why the forecast performance of the retail order forecast model does not improve by incorporating promotion of competitors is that the total number of observations in the data set is too small. With a low number of observations a model with less input variables, i.e. a lower model complexity, is generally preferred in order to prevent overfitting. All in all, it can be concluded that the competitive promotions are predictable to some extent, but it is not proven if the competitive promotion prediction model provides practical value.

6.1 Contributions to practice and recommendations

Despite the fact that there is no substantiation about the practical value of predicted competitive promotions, the retail order forecast model showed excellent results compared to the current forecasting method. Storing and using historic (promotion) data is thus a major added value in forecasting. To maintain forecasting performance and perhaps even improve it in the future it is essential that the data quality is high. To ensure data quality, the following is recommended to Swinkels Family Brewers.

First, retain original product numbers in order to store all data from the same product under the same product number. Currently, products have changed from product number due to renewed packaging. As a result, data stored under different product numbers needs to be connected which leads to higher data processing efforts especially when multiple data sets are associated. Furthermore, for consistency reasons it is recommended to use fixed data formatting in which a data set always consists of the same variables. In this way, data processing efforts are also minimized and tools do not need to be adjusted for new unrecognizable variables.

Second, collect more data to improve the retail order forecast model. More specifically, record the percentage discount given to the end consumer. Previous literature showed that the forecast accuracy increases when more promotional variables are included (Ramanathan, 2012). Furthermore, the promotion mechanism is often not among the most crucial explaining variables in promotional forecasting (Divakar et al., 2005; Ali et al., 2009; Huang et al., 2014; Van Donselaar, Peters, de Jong & Broekmeulen, 2016; Ma et al., 2016; Ma & Fildes, 2017). This could indicate that the promotion mechanism on itself does not have a substantial contribution to the forecasting performance. Therefore, Swinkels Family Brewers is recommended to record the percentage discount as well as the promotion mechanism of a promotion. It is not recommended to record the absolute discount since research showed that percentage discount is a better predictor of the promotion volume than absolute discount (Derks, 2015; Nabuurs, 2017; van der Poel, 2010). This research showed that a less complex model, i.e. a model with a fewer amount of input variables, is favorable and therefore Swinkels Family Brewers is at the same time recommended to be careful with collecting more input data for the retail order forecast model.

6.2 Contributions to literature

Besides the practical importance, this study also contributes to the literature. While existing studies on forecasting with competitive promotions are researched from the retailer's point of view this study is the first to study that from a manufacturer's point of view. The difference between the two is that competitive promotions are known to a retailer but unknown to a manufacturer. A retailer is thus able to incorporate competitive promotions in the forecasts and previous research has shown that incorporating this information is beneficial for the forecasting performance (Kuo, 2001; Divakar et al., 2005; Ali et al., 2009; Huang et al., 2014; de Sousa, Oliveira & Ramos, 2016; Ma et al., 2016; Ma & Fildes, 2017). In practice, it is difficult for a manufacturer to obtain data on the planned promotions of competitors since retailers do not share this information. Therefore, this study showed how to take into account competitive information despite the fact that that information is not known. This is done by analyzing promotions of competitors and using these insights for creating the competitive promotion prediction model. The competitive promotion prediction model enables manufacturers to generate predictions of promotions of competitors and incorporate them in the retail order forecast. Hereby, this research filled in the gap in literature on the prediction of price promotions of competitors.

6.3 Limitations and future research

This study faces a number of limitations which could serve as areas for future research. First, the available data was limited which was probably the reason that the actual competitive promotions did not improve forecast accuracy. As a result, the question of whether the predicted promotions would improve the retail order forecast accuracy could not be answered. To put it differently, the practical value of the competitive promotion predictions is unexplored. The research question remains therefore unanswered and more research is desired in this field to better understand the implications of this research. Future research on this topic is recommended to first test the added value of the actual competitive promotions and hereby not to encounter surprises later.

Second, due to the innovative nature of this research, no benchmark was available to compare the results of the competitive promotion prediction model with. The model was dependent on created benchmarks rather than existing benchmarks from practice or literature. This made it difficult to interpret how the model performed. Future research is, therefore, challenged to improve the designed competitive promotion prediction model by using this study's results as a benchmark. Opportunities for improvements lie particularly in the area of the input variables, the multi-label approach and the classification method. For example, deep learning methods, such as neural networks, might be selected as classification method. These methods give little insights in the classification process but may be able to better capture the humane promotion planning behavior. The biggest advantage of deep learning models is that they can find patterns in a large amount of data without specifying explicit features. This eliminates the chance of an inappropriate model design. The model can also be expanded by predicting other promotion characteristics such as the price discount instead of only the timing of a promotion.

Third, in this research we tried to predict the promotional POS and then distribute this volume over the promotion week and loading weeks with a fixed percentage. However, we were not able to cope with the varying purchasing behavior of a retailer in this approach. More research is required on the promotional purchasing behavior of retailers and how this can be modeled. Coping with the purchasing behavior of retailers can possibly help to improve manufacturer's forecasts.

Last, the findings of this research are limited to one product category, namely the beer category. Furthermore, this research was mainly focused on one retailer, i.e. the biggest retailer of the Netherlands. Small experiments on other retailers have been executed but these have not been optimized. It would be interesting to experiment further with other retailers and also other product categories. These experiments could help to validate our results and could shed new light on diversity in the retail domain.

References

Abraham, M. M. & Lodish, L. M. (1987). Promoter: An automated promotion evaluation system. *Marketing Science*, *6*(2), 101–123.

Agrawal, D. & Schorling, C. (1996). Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model. *Journal of Retailing*, *72*(4), 383–407.

Ailawadi, K. L., Harlam, B. A., Cesar, J. & Trounce, D. (2006). Promotion profitability for a retailer: the role of promotion, brand, category, and store characteristics. *Journal of Marketing Research*, *43*(4), 518–535.

Ailawadi, K. L., Kopalle, P. K. & Neslin, S. A. (2005). Predicting competitive response to a major policy change: Combining game-theoretic and empirical analyses. *Marketing Science*, *24*(1), 12–24.

Ali, Ö. G., Sayın, S., Van Woensel, T. & Fransoo, J. (2009). Sku demand forecasting in the presence of promotions. *Expert Systems with Applications*, *36*(10), 12340–12348.

Bourland, K. E., Powell, S. G. & Pyke, D. F. (1996). Exploiting timely demand information to reduce inventories. *European journal of operational research*, *92*(2), 239–253.

Cachon, G. P. & Fisher, M. (2000). Supply chain inventory management and the value of shared information. *Management science*, *46*(8), 1032–1048.

Chakraborty, K., Mehrotra, K., Mohan, C. K. & Ranka, S. (1992). Forecasting the behavior of multivariate time series using neural networks. *Neural networks*, *5*(6), 961–970.

Choi, T.-M., Hui, C.-L., Ng, S.-F. & Yu, Y. (2012). Color trend forecasting of fashionable products with very few historical data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(6), 1003–1010.

Cooper, L. G., Baron, P., Levy, W., Swisher, M. & Gogos, P. (1999). Promocast: A new forecasting method for promotion planning. *Marketing Science*, *18*(3), 301–316.

Derks, L. A. D. (2015). *Improving promotion forecasts for the dutch fmcg market: a research executed at royal frieslandcampina n.v.* (Unpublished master's thesis). Eindhoven University of Technology.

de Sousa, J. P., Oliveira, J. M. & Ramos, P. (2016). Sales forecasting in retail industry based on dynamic regression models.

Divakar, S., Ratchford, B. T. & Shankar, V. (2005). Practice prize articlechan4cast: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. *Marketing Science*, *24*(3), 334–350.

Dubé, J.-P., Chintagunta, P., Petrin, A., Bronnenberg, B., Goettler, R., Seetharaman, P., ... Zhao, Y. (2002). Structural applications of the discrete choice model. *Marketing Letters*, *13*(3), 207–220.

Fildes, R., Goodwin, P., Lawrence, M. & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International journal of forecasting*, *25*(1), 3–23.

Fildes, R., Nikolopoulos, K., Crone, S. F. & Syntetos, A. A. (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society*, *59*(9), 1150–1172.

Franses, P. H. & Legerstee, R. (2010). Do experts' adjustments on model-based sku-level forecasts improve forecast quality? *Journal of Forecasting*, *29*(3), 331–340.

Gavirneni, S., Kapuscinski, R. & Tayur, S. (1999). Value of information in capacitated supply chains. *Management science*, *45*(1), 16–24.

Groot, R. D. & Musters, P. A. (2005). Minority game of price promotions in fast moving consumer goods markets. *Physica A: Statistical Mechanics and its Applications*, *350*(2-4), 533–547.

Grover, R. & Srinivasan, V. (1987). A simultaneous approach to market segmentation and market structuring. *Journal of Marketing Research*, *24*(2), 139–153.

Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing research*, *25*(4), 342–355.

Holmström, J., Främling, K., Kaipia, R. & Saranen, J. (2002). Collaborative planning forecasting and replenishment: new solutions needed for mass collaboration. *Supply chain management: An international journal*, *7*(3), 136–145.

Huang, T., Fildes, R. & Soopramanien, D. (2014). The value of competitive information in forecasting fmcg retail product sales and the variable selection problem. *European Journal of Operational Research*, *237*(2), 738–748.

Jiang, J. J., Zhong, M. & Klein, G. (2000). Marketing category forecasting: An alternative of bvar-artificial neural networks. *Decision Sciences*, *31*(4), 789–812.

Kamakura, W. A. & Kang, W. (2007). Chain-wide and store-level analysis for cross-category management. *Journal of Retailing*, *83*(2), 159–170.

Kiely, D. A. (1998). Synchronizing supply chain operations with consumer demand using customer data. *The Journal of Business Forecasting*, *17*(4), 3.

Kulp, S. C., Lee, H. L. & Ofek, E. (2004). Manufacturer benefits from information integration with retail customers. *Management science*, *50*(4), 431–444.

Kumar, A., Rao, V. R. & Soni, H. (1995). An empirical comparison of neural network and logistic regression models. *Marketing Letters*, *6*(4), 251–263.

Kuo, R. (2001). A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. *European Journal of Operational Research*, *129*(3), 496–517.

Lachtermacher, G. & Fuller, J. D. (1995). Back propagation in time-series forecasting. *Journal of forecasting*, *14*(4), 381–393.

Lal, R. (1990). Price promotions: Limiting competitive encroachment. *Marketing science*, *9*(3), 247–262.

Lapide, L. (2005). Account level forecasting needs downstream data. *JOURNAL OF BUSINESS FORECASTING METHODS AND SYSTEMS*, *24*(2), 17.

Lee, H. L., So, K. C. & Tang, C. S. (2000). The value of information sharing in a two-level supply chain. *Management science*, *46*(5), 626–643.

Leeflang, P. S. & Wittink, D. R. (1992). Diagnosing competitive reactions using (aggregated) scanner data. *International Journal of Research in Marketing*, *9*(1), 39–57.

Leeflang, P. S. & Wittink, D. R. (1996). Competitive reaction versus consumer response: Do managers overreact? *International Journal of Research in Marketing*, *13*(2), 103–119.

Lou, Y., Caruana, R. & Gehrke, J. (2012). Intelligible models for classification and regression. In *Proceedings of the 18th acm sigkdd international conference on knowledge discovery and data mining* (pp. 150–158).

Ma, S. & Fildes, R. (2017). A retail store sku promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research*, *260*(2), 680–692.

Ma, S., Fildes, R. & Huang, T. (2016). Demand forecasting with high dimensional data: The case of sku retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, *249*(1), 245–257.

Miguéis, V. L., Van den Poel, D., Camanho, A. S. & e Cunha, J. F. (2012). Predicting partial customer churn using markov for discrimination for modeling first purchase sequences. *Advances in Data Analysis and Classification*, *6*(4), 337–353.

Nabuurs, F. P. A. (2017). *Forecasting and replenishing promotions in a dutch grocery retail supply chain* (Unpublished master's thesis). Eindhoven University of Technology.

Oentaryo, R. J., Lim, E.-P., Lo, D., Zhu, F. & Prasetyo, P. K. (2012). Collective churn prediction in social network. In *2012 ieee/acm international conference on advances in social networks analysis and mining* (pp. 210–214).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*(Oct), 2825–2830.

Raghunathan, S. (2001). Information sharing in a supply chain: A note on its value when demand is nonstationary. *Management science*, *47*(4), 605–610.

Ramanathan, U. (2012). Supply chain collaboration for improved forecast accuracy of promotional sales. *International Journal of Operations & Production Management*, *32*(6), 676–695.

Rao, R. C., Arjunji, R. V. & Murthi, B. (1995). Game theory and empirical generalizations concerning competitive promotions. *Marketing Science*, *14*(3_supplement), G89–G100.

Read, J., Martino, L., Olmos, P. M. & Luengo, D. (2015). Scalable multi-output label prediction: From classifier chains to classifier trellises. *Pattern Recognition*, *48*(6), 2096–2109.

Schapire, R. E. & Singer, Y. (2000). Boostexter: A boosting-based system for text categorization. *Machine learning*, *39*(2-3), 135–168.

Sorower, M. S. (2010). A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis*, *18*, 1–25.

Steenkamp, J.-B. E., Nijs, V. R., Hanssens, D. M. & Dekimpe, M. G. (2005). Competitive reactions to advertising and promotion attacks. *Marketing science*, *24*(1), 35–54.

Struse III, R. W. (1987). Commentary approaches to promotion evaluation: A practitioner's viewpoint. *Marketing Science (1986-1998)*, *6*(2), 150.

Szymański, P. & Kajdanowicz, T. (2017, February). A scikit-based Python environment for performing multi-label classification. *ArXiv e-prints*.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, *16*(4), 437–450.

Tsoumakas, G. & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, *3*(3), 1–13.

Van Aken, J. E. (1994). De bedrijfskunde als ontwerpwetenschap: De regulatieve en de reflectieve cyclus. *Business engineering as a design science. The regulatory and the reflectory cycle), Bedrijfskunde*, *66*, 16–26.

van der Poel, M. J. (2010). *Improving the promotion forecasting accuracy at unilever netherlands* (Unpublished master's thesis). Eindhoven University of Technology.

Van Donselaar, K. H., Peters, J., de Jong, A. & Broekmeulen, R. A. (2016). Analysis and forecasting of demand during promotions for perishable items. *International Journal of Production Economics*, *172*, 65–75.

Van Heerde, H. J., Gupta, S. & Wittink, D. R. (2003). Is 75% of the sales promotion bump due to brand switching? no, only 33% is. *Journal of Marketing Research*, *40*(4), 481–491.

van Strien, P. J. (1997). Towards a methodology of psychological practice: The regulative cycle. *Theory & Psychology*, *7*(5), 683–700.

Vilcassim, N. J., Kadiyali, V. & Chintagunta, P. K. (1999). Investigating dynamic multifirm market interactions in price and advertising. *Management Science*, *45*(4), 499–518.

Walters, R. G. (1991). Assessing the impact of retail price promotions on product substitution, complementary purchase, and interstore sales displacement. *Journal of marketing*, *55*(2), 17–28.

Williams, B. D. & Waller, M. A. (2010). Creating order forecasts: point-of-sale or order history? *Journal of Business Logistics*, *31*(2), 231–251.

Williams, B. D., Waller, M. A., Ahire, S. & Ferrier, G. D. (2014). Predicting retailer orders with pos and order data: The inventory balance effect. *European Journal of Operational Research*, *232*(3), 593–600.

Yang, D., Goh, G. S., Jiang, S., Zhang, A. N. & Akcan, O. (2015). Forecast upc-level fmcg demand, part ii: Hierarchical reconciliation. In *2015 ieee international conference on big data (big data)* (pp. 2113–2121).

Appendix A

Retailers and products in data set

Retailers
Albert Heijn
Boni
Coop
Hoogvliet
Jan Linders
Jumbo
Plus
Poiesz
Vomar

Table A.1: Retailers in the data set

Brand	**Product**	**Packaging**	Group
Bavaria	Pilsner	Crate bottle 12x30 CL	Pilsner crates
Bavaria	Pilsner crate	Crate bottle 24x30 CL	Pilsner crates
Bavaria	Pilsner	Can 6x33 CL	Pilsner other
Bavaria	Pilsner	Can 12x33 CL	Pilsner other
Bavaria	Pilsner	Can 24x50 CL	Pilsner other
Bavaria	Pilsner cooled	Can 6x33 CL	Pilsner other
Bavaria	0.0% Original	Bottle 6x30 CL	0.0%
Bavaria	0.0% Original	Can 6x33 CL	0.0%
Bavaria	0.0% Original	Can 12x50 CL	0.0%
Bavaria	0.0% Wit	Can 6x33 CL	0.0%
Bavaria	0.0% Radler Lemon	Bottle 6x30 CL	0.0%
Bavaria	0.0% Radler Lemon	Can 6x33 CL	0.0%
Bavaria	0.0% Fruity Rose	Can 6x33 CL	0.0%
Bavaria	Radler Lemon	Bottle 6x30 CL	Fresh
Bavaria	Radler Lemon	Can 6x33 CL	Fresh
Bavaria	Radler Lemon	Can 12x50 CL	Fresh
Bavaria	Old Brown	Bottle 6x30 CL	Characterful
Bavaria	Bockbeer	Bottle 6x30 CL	Characterful
Bavaria	8.6 Original	Can 12x50 CL	Characterful
Bavaria	8.6 Extreme	Can 12x50 CL	Characterful
Claro	Tequila Flavoured	Bottle 3x33 CL	Fresh
La Trappe	Blond	Bottle 75 CL	Characterful
La Trappe	Double	Bottle 6x30 CL	Characterful
La Trappe	Triple	Bottle 6x30 CL	Characterful
La Trappe	Isid'or	Bottle 6x30 CL	Characterful
La Trappe	Isid'or	Bottel 75 CL	Characterful
La Trappe	Quadrupel	Bottle 6x30 CL	Characterful
La Trappe	Quadrupel	Bottel 75 CL	Characterful
La Trappe	White Trappist	Bottle 6x30 CL	Characterful
Landerbrau	Pilsner	Can 24x50 CL	Pilsner other

Table A.2: Swinkels Family Brewers' products in the data set

Appendix B

Competitive brands

Pilsner crates	Pilsner other	0.0%	Fresh	Characterful
Heineken (25%)	Heineken (21%)	Amstel (38%)	Amstel (29%)	Grolsch (13%)
Hertog Jan (18%)	Amstel (12%)	Bavaria (25%)	Grolsch (15%)	Leffe (12%)
Grolsch (17%)	PL AH basic (9%)	Grolsch (16%)	Desperados (10%)	Gulpener (10%)
Amstel (13%)	PL Pitt (8%)	Heineken (8%)	Corona (6%)	Brand (6%)
Bavaria (8%)	Bavaria (8%)		Jillz (5%)	Palm (6%)
Jupiler (5%)	Grolsch (6%)		Hoegaarden (4%)	Hertog Jan (6%)
	Hertog Jan (5%)		Bavaria (3%)	Bavaria (6%)
	PL Export (6%)		Wieckse (3%)	Duvel (5%)
	De Klok (5%)		Franziskaner (2%)	La Trappe (4%)
	Jupiler (2%)		Hertog Jan (2%)	La Chouffe (4%)
			Erdinger (2%)	Westmalle (3%)
				Grimbergen (2%)
				Affligem (2%)
				Heineken (2%)

Table B.1: Brands with biggest market share per group (PL = private label)

Company	**Brand**
Alken Maes	Grimbergen
Asahi	Grolsch
Duvel	Duvel La Chouffe
Erdinger	Erdinger
Franziskaner	Franziskaner
Gulpener	Gulpener
Heineken	Affligem Amstel Brand Desperados Heineken Jillz Wieckse
Inbev	Corona Hertog Jan Hoegaarden Jupiler Leffe
Swinkels Family Brewers	Bavaria La Trappe Palm
Westmalle	Westmalle

Table B.2: Brands owned by companies

Appendix C

Analysis 0.0% promotions at Albert Heijn

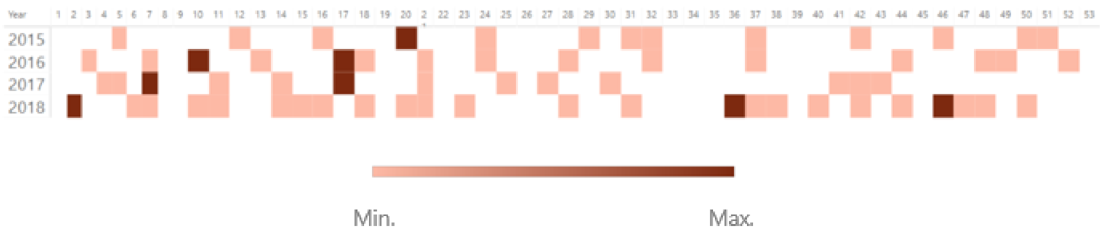

Figure C.1: Heatmap of 0.0% promotions at Albert Heijn

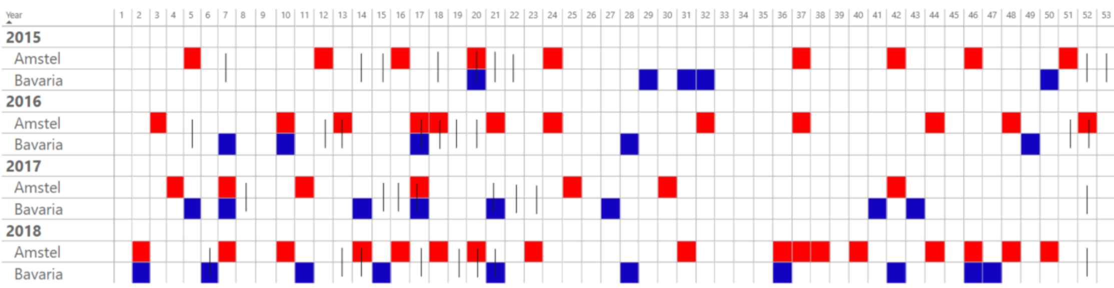

Figure C.2: 0.0% promotions at Albert Heijn. The dark lines indicate the weeks with a holiday and the colors are unique for every brand.

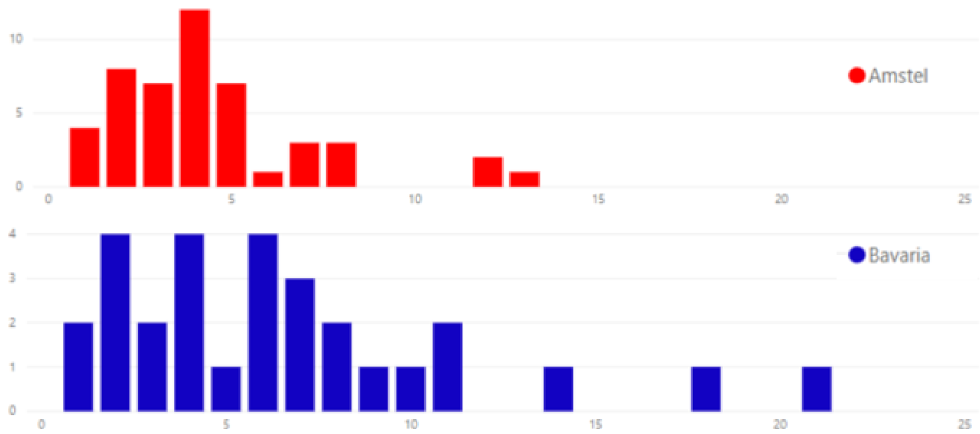

Figure C.3: Histogram of weeks between 0.0% promotions.

Appendix D

Analysis fresh promotions at Albert Heijn

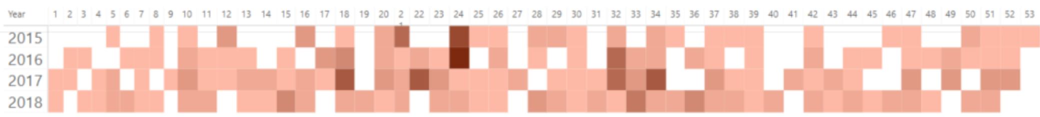

Figure D.1: Heatmap of fresh promotions at Albert Heijn

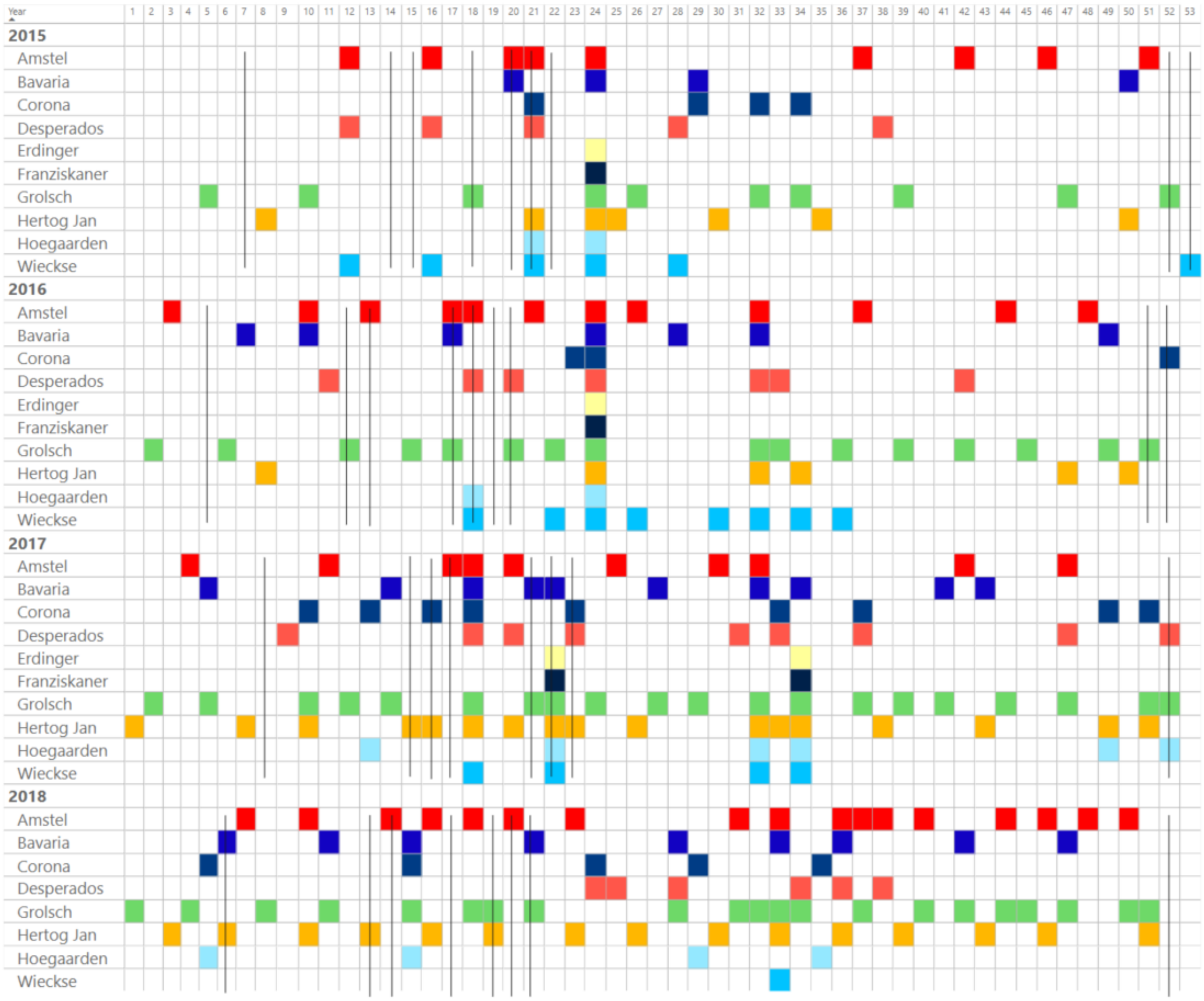

Figure D.2: Fresh promotions at Albert Heijn. The dark lines indicate the weeks with a holiday and the colors are unique for every brand.

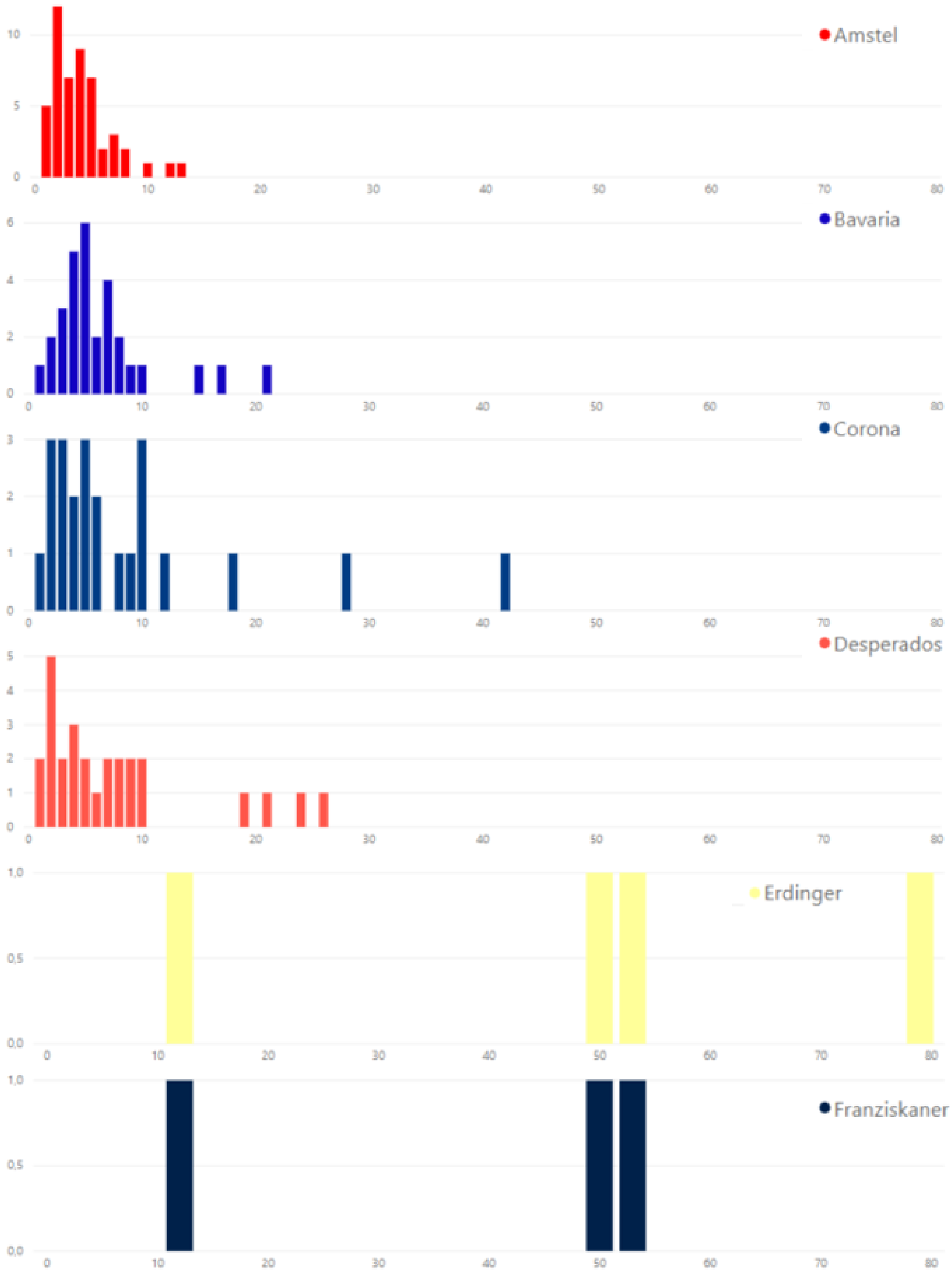

Figure D.3: Histogram of weeks between fresh promotions.

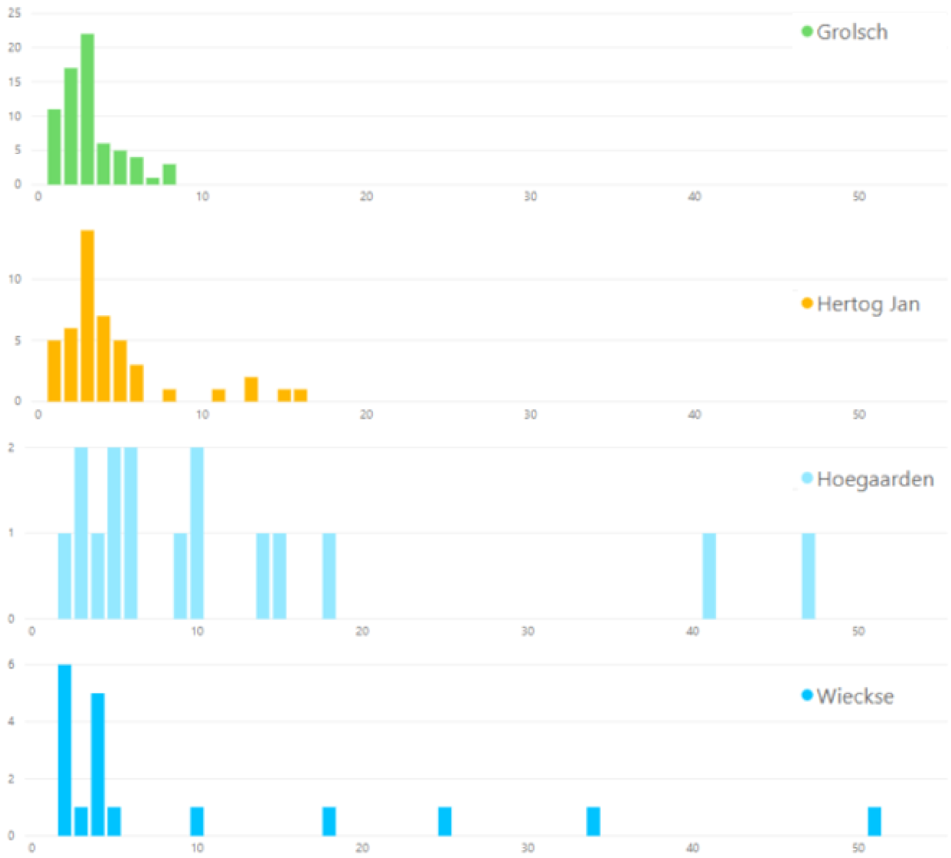

Figure D.4: Histogram of weeks between fresh promotions.

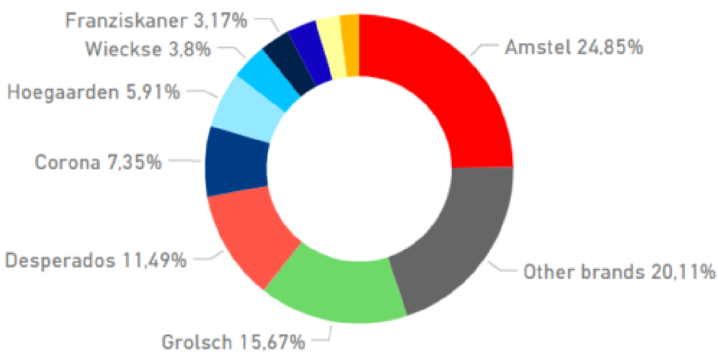

Figure D.5: Market share of brands in the fresh group at Albert Heijn (2015-2018)

Appendix E

Analysis characterful promotions at Albert Heijn

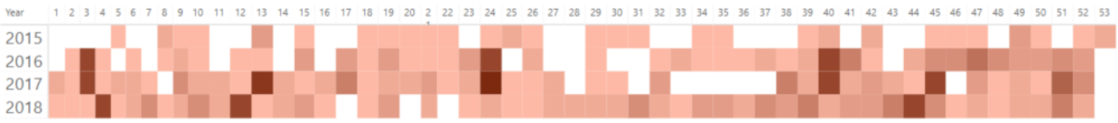

Figure E.1: Heatmap of characterful promotions at Albert Heijn

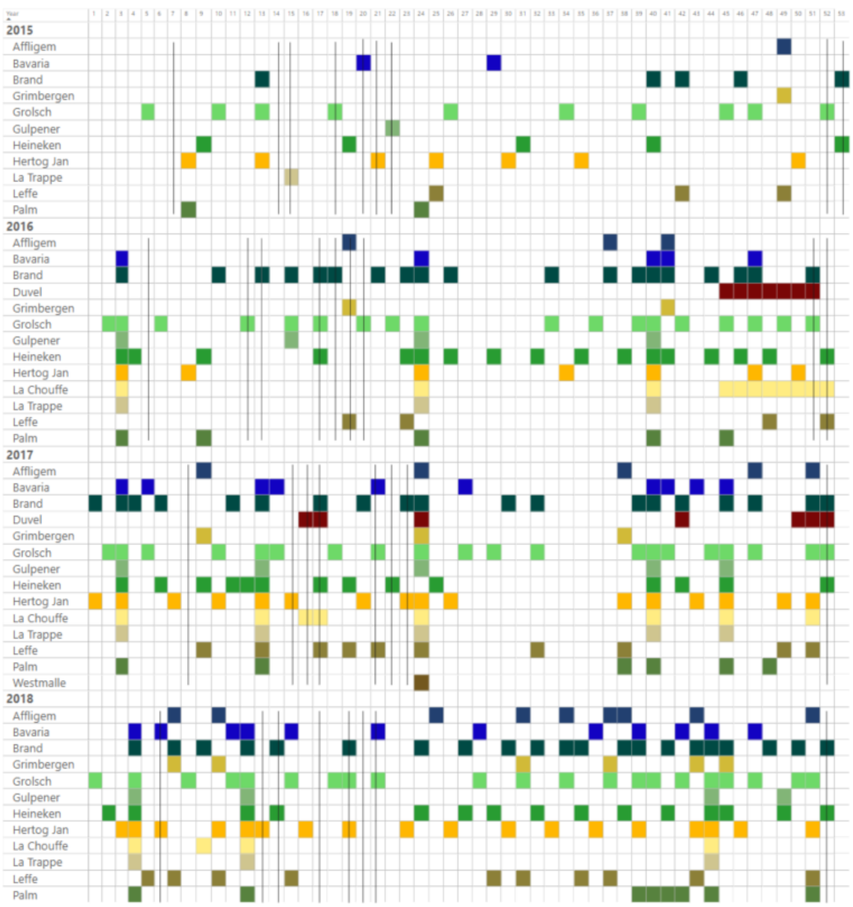

Figure E.2: Characterful promotions at Albert Heijn. The dark lines indicate the weeks with a holiday and the colors are unique for every brand.

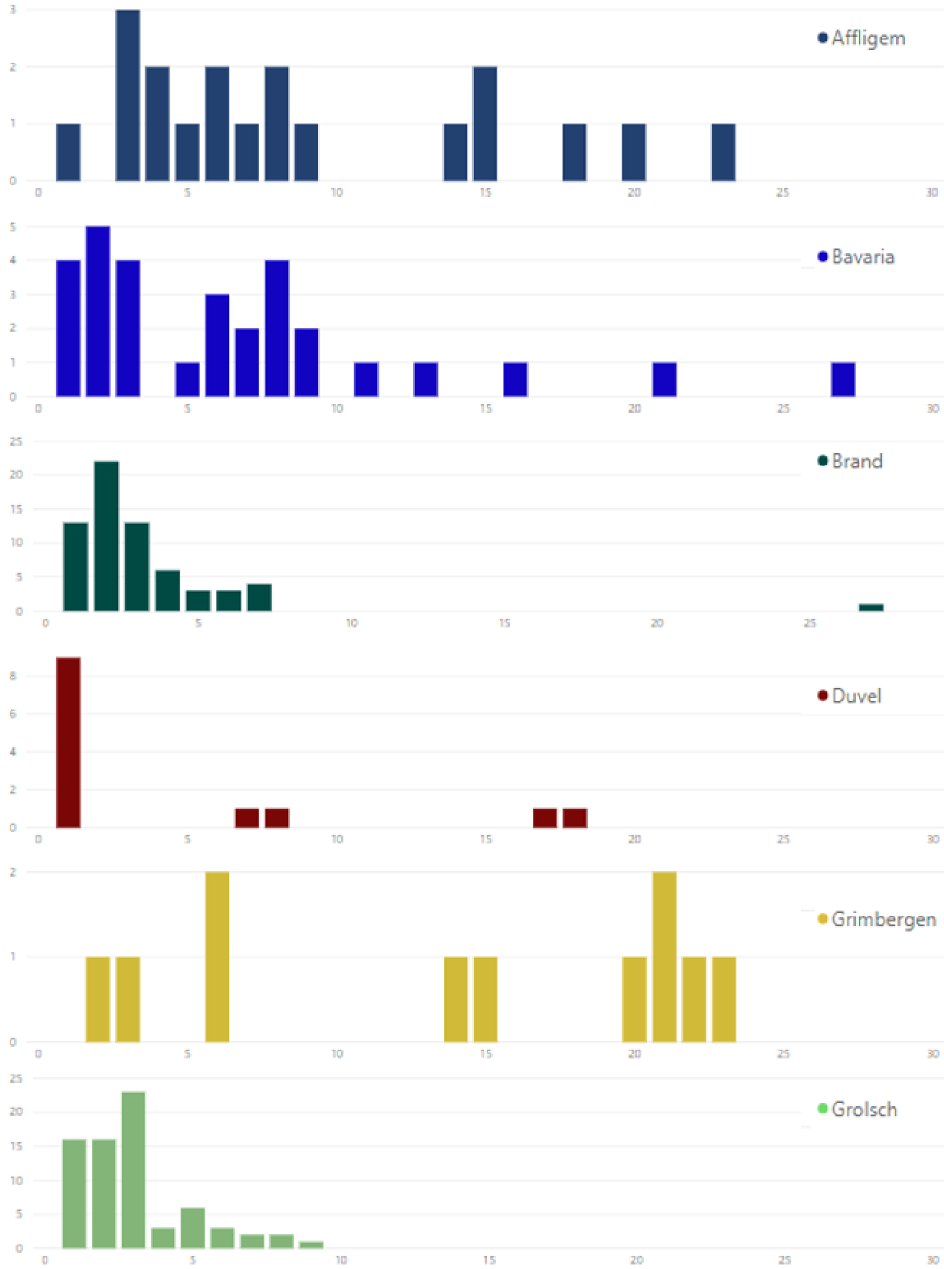

Figure E.3: Histogram of weeks between characterful promotions.

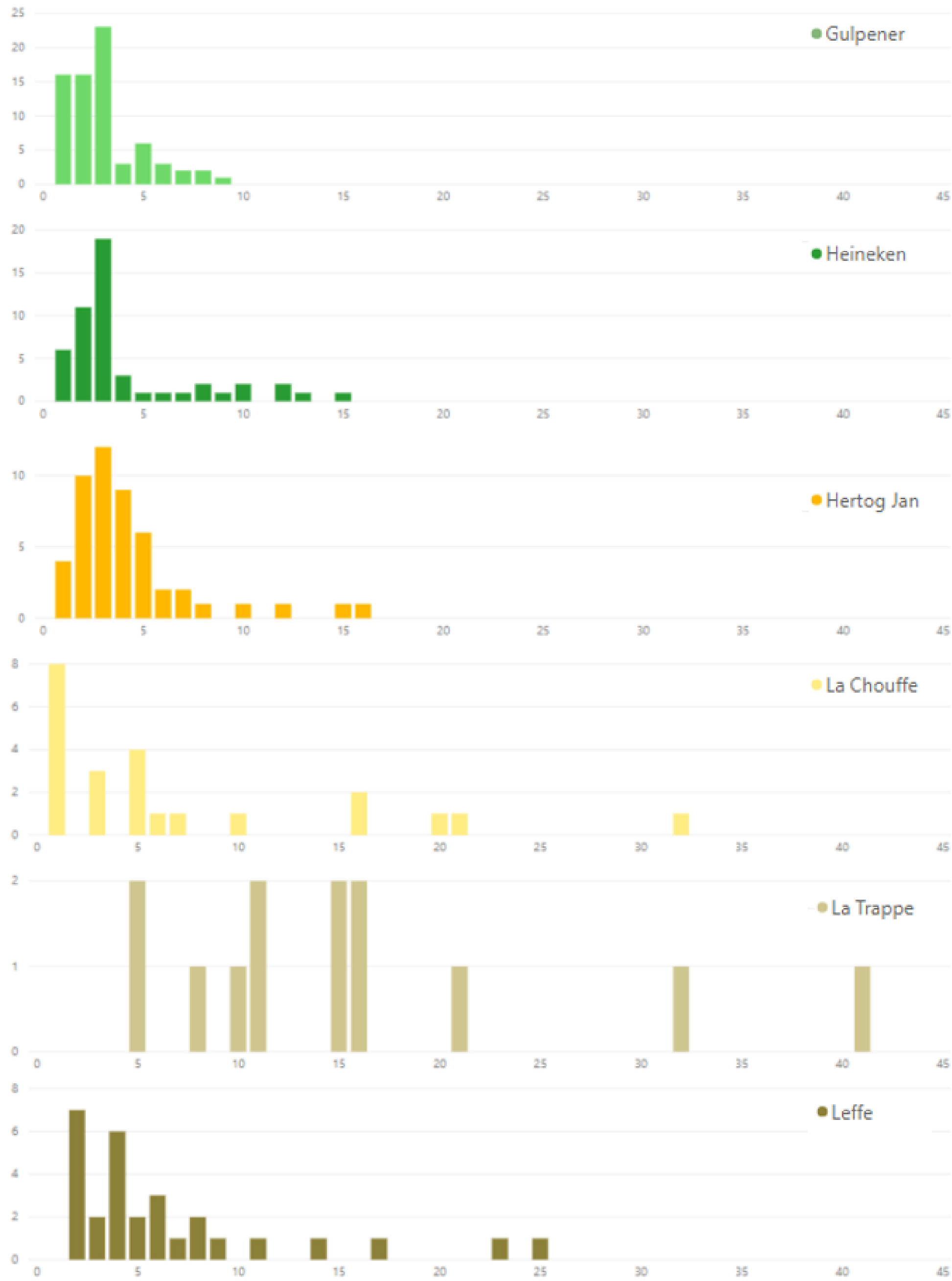

Figure E.4: Histogram of weeks between characterful promotions.

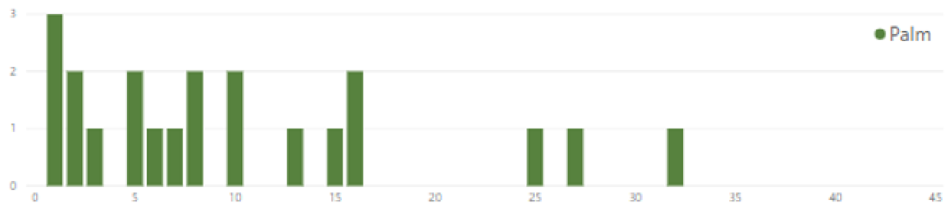

Figure E.5: Histogram of weeks between characterful promotions.

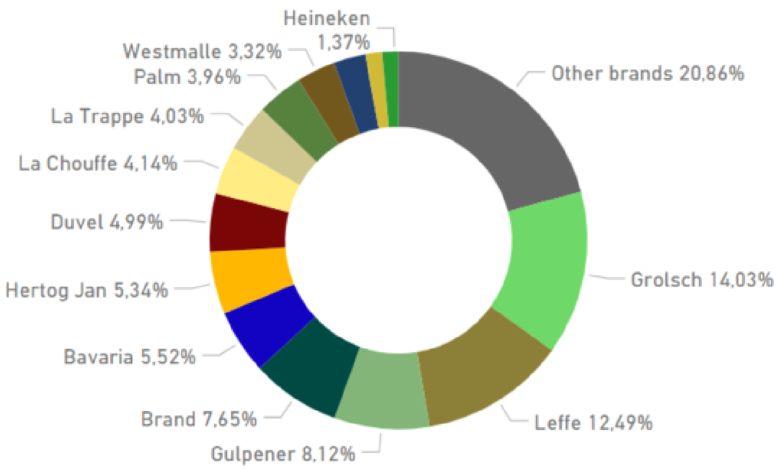

Figure E.6: Market share of brands in the characterful group at Albert Heijn (2015-2018)

Appendix F

Varying purchasing behavior of Albert Heijn

Week	**Standard deviation**
Loading week 1	0.35
Loading week 2	0.55
Promotion week	0.42

Table F.1: Standard deviation of division of promotion volume over loading weeks and promotion week

Appendix G

Input variables competitive promotion prediction model

Variable	Description	Variable type
$Week$	Week number	Discrete
$Yearweek$	A number adding up for every week in the data	Discrete
$Season$	The season (1: winter, 2: spring, 3: summer, 4: autumn)	Categorical
$Holiday$	The type of holiday (0: no holiday, 1: Carnaval, 2: Easter Sunday, 3: Easter Monday, 4: Kingsday, 5: Ascension day, 6: Whit Sunday, 7: Whit Monday, 8: Christmas, 9: New Years Eve)	Categorical
$Last\ promotion_{brand-group1}$	The number of weeks since the last promotion of brand-group 1 at a specific supermarket	Discrete
$Last\ promotion_{brand-group2}$	The number of weeks since the last promotion of brand-group 2 at a specific supermarket	Discrete
...	...	...
$Last\ promotion_{brand-groupX}$	The number of weeks since the last promotion of brand-group X at a specific supermarket	Discrete
$Promotion\ Bavaria_{pilsnercrates}$	A promotion of Bavaria pilsner crates (0: no promotion, 1: promotion)	Binary
$Promotion\ Bavaria_{pilsnerother}$	A promotion of Bavaria pilsner other (0: no promotion, 1: promotion)	Binary
$Promotion\ Bavaria_{0.0\%}$	A promotion of Bavaria 0.0% (0: no promotion, 1: promotion)	Binary
$Promotion\ Bavaria_{fresh}$	A promotion of Bavaria fresh (0: no promotion, 1: promotion)	Binary
$Promotion\ Bavaria_{characterful}$	A promotion of Bavaria characterful (0: no promotion, 1: promotion)	Binary

Table G.1: Input variables of competitive promotion prediction model

Appendix H

Calculation competitive promotion prediction benchmark model 1b

Model 1b of the competitive promotion prediction model fits a continuous distribution on the time between promotions for every label. The tested distributions are the beta distribution, the erlang distribution, the exponential distribution, the gamma distribution, the lognormal distribution, the normal distribution and the uniform distribution. Every distribution is fitted on the number of weeks between promotions and the sum of squared errors (SSE) is determined. The fitted distribution with the smallest SSE is selected as the final distribution for that label. Thereafter, for every week the probability on a promotion is determined with the final distribution by taking into account the time since the last promotion at this moment. The difficulty here is that the prediction horizon is four weeks and that there are eight different scenarios in the weeks between now and the predicted week. These scenarios are visually displayed in Figure H.1.

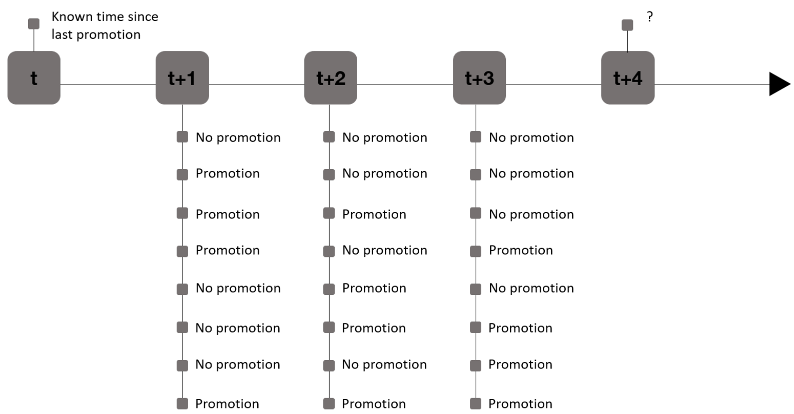

Figure H.1: Scenarios in weeks between now (t) and the predicted week (t+4)

For every scenario the probability on a promotion is calculated, which is based on the known time since last promotion at time t. The total probability in week t+4 is the sum of the probabilities of the eight scenarios. An example is given on the first scenario. In the first scenario there is no promotion in week t+1, no promotion in week t+2 and no promotion in week t+3. Suppose that the time since last promotion at week t is a. Then the probability on a promotion in week t+4 is:

$$P(X = a + 4 \mid X > a + 3) \tag{H.1}$$

However, the distribution is continuous and therefore the probability is:

$$P(a + 3.4 < X < a + 4.5 \mid X > a + 3) \tag{H.2}$$

Summing all scenarios leads to the following probability equation for a promotion in week t+4:

$$\begin{gathered}
P(Promotion_{t+4}) = \\
P(a + 3.4 < X < a + 4.5 \mid X > a + 3) + \\
P(a + 0.4 < X < a + 1.5 \mid X > a) \times P(2.4 < X < 3.5 \mid X > 2) + \\
P(a + 0.4 < X < a + 1.5 \mid X > a) \times P(0.4 < X < a + 1.5 \mid X > 0) \times P(1.4 < X < 2.5 \mid X > 1) + \\
P(a + 0.4 < X < a + 1.5 \mid X > a) \times P(1.4 < X < 2.5 \mid X > 1) \times P(0.4 < X < 1.5 \mid X > 0) + \\
P(a + 1.4 < X < a + 2.5 \mid X > a + 1) \times P(1.4 < X < 2.5 \mid X > 1) + \\
P(a + 1.4 < X < a + 2.5 \mid X > a + 1) \times P(0.4 < X < 1.5 \mid X > 0) \times P(0.4 < X < 1.5 \mid X > 0) + \\
P(a + 2.4 < X < a + 3.5 \mid X > a + 2) \times P(0.4 < X < 1.5 \mid X > 0) + \\
P(a + 0.4 < X < a + 1.5 \mid X > a) \times P(0.4 < X < 1.5 \mid X > 0) \times P(0.4 < X < 1.5 \mid X > 0)
\end{gathered}$$

When $P(Promotion_{t+4})$ is greater than 0.5, it will be classified as a promotion. If $1 - P(Promotion_{t+4})$ is greater than 0.5, i.e. $P(Promotion_{t+4})$ is smaller than 0.5, it will be classified as no promotion.

Appendix I

Input variables retail order forecast model

Variable	Description	Variable type
$Promotion$	A promotion in the forecasted week (0: no promotion, 1: promotion)	Binary
$Promotion_{following}$	A promotion in the week before the forecasted week (0: no promotion in the week before, 1: promotion in the week before)	Binary
$Promotion_{prior}$	A promotion in the week after the forecasted week (0: no promotion in the week after, 1: promotion in the week after)	Binary
$Mechanism$	The mechanisms (i.e. type of discount) of the promotion in case of a $promotion$, $promotion_{following}$, $promotion_{prior}$, $promotion_{prior2}$ (1: 1+1 free, 2: 20% discount, 3: 1 voor X, etc.)	Categorical
$Holiday$	The type of holiday in the forecasted week (0: no holiday, 1: Carnaval, 2: Easter, 3: Kingsday, 4: Ascension Day, 5: Pentecost, 6: Christmas, 7: New Years Eve)	Categorical
$Holiday_{prior}$	The type of holiday in the forecasted week (0: no holiday, 1: Carnaval, 2: Easter, 3: Kingsday, 4: Ascension Day, 5: Pentecost, 6: Christmas, 7: New Years Eve)	Categorical
$CompetitivePromotion_{brand-group1}$	A promotion of brand-group 1 (0: no promotion, 1: promotion)/probability	Binary
...	...	..
$CopmetitivePromotion_{brand-groupX}$	A promotion of brand-group X (0: no promotion, 1: promotion)/probability	Binary

Table I.1: Input variables of retail order forecast model