Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Business Analytics & Quantitative Marketing

Unveiling Customer Motivations in Grocery Shopping: A Correlated Topic Model Approach with Variational Inference

Hugo Kroonen (471994)

Supervisor:	R. Paap
Second assessor:	D. Fok
Date final version:	9th October 2023

Abstract

In today's competitive grocery retail landscape, it is crucial to understand customer shopping behaviour. However, analysing these transactions presents challenges due to their large volume and dimensionality. Correlated topic modelling (CTM) emerges as a powerful tool to address these challenges by unveiling a set of underlying purchase motivations. Leveraging over 100,000 transactions from a prominent Dutch grocery retailer, we employ variational Bayesian inference to estimate the CTM model efficiently and scalably. The resulting purchase motivations are evaluated for quality and interpretability, measured by coherence and similarity.

Our approach extends beyond identifying latent motivations at the shopping trip level, as it also explores their interrelationships and estimates the effects of customer demographics and contextual factors. An innovative contribution to the existing literature lies in our ability to distinguish between promotion-driven and intrinsic motivations, providing valuable insights for developing more effective promotion strategies. Additionally, the identified motivations enable personalised product recommendations, combining novelty and explainability, even for new customers. With these insights, retailers can improve customer segmentation, personalised communication, tailored advertising, and recommendation systems, deepening their understanding of their customer base and fostering greater engagement.

Contents

Chapter 1

Introduction

In the dynamic landscape of Dutch supermarkets, millions of customers visit physical stores and online platforms daily in search of products that align with their preferences (Statista, 2022). To meet this demand, retailers are strategically investing in personalised communication strategies. These strategies include tailored product recommendations, targeted advertisements, and customised search features (Tyrała et al., 2022). These initiatives aim to establish deeper connections with customers, ultimately enhancing customer satisfaction and building brand loyalty, a crucial objective in the competitive supermarket industry (Tyrała et al., 2022).

Before sophisticated recommendation systems and tailored offers can become effective, companies must understand their customers' unique preferences (Christidis et al., 2010). In large retail settings, this task poses a substantial challenge due to data sparsity, given that many customers only purchase a small fraction of the extensive product range available (Carrasco et al., 2022). This challenge highlights an urgent need for an intelligent segmentation approach capable of handling transaction data's vast volume and dimensionality while yielding interpretable results.

To address this need, the central question of this thesis emerges: 'Can we develop a model that unveils the motivations driving grocery shopping behaviours, facilitating understanding, segmentation, and personalised product recommendations?'

Recent research has adapted topic modelling techniques, such as latent Dirichlet allocation (LDA), originally designed for extracting latent topics from text documents, to customer profiling and identifying purchasing patterns (Blanchard et al., 2017). When implemented in large retail settings, these modelling methods have demonstrated particular efficacy in addressing the data-sparsity problem (Reisenbichler & Reutterer, 2019). By treating product purchases as words and a customer's purchase history as a document, these methods can uncover topics indicative of customer's motivations in the retail context (Carrasco et al., 2022; Hornsby et al., 2020). To model the effect of customer demographics on initial motivation probabilities, Jacobs et al. (2016) extended LDA to LDA-X. However, traditional LDA models cannot capture correlations between motivations, leading to the introduction of the correlated topic model (CTM) by Blei & Lafferty (2007). Building upon CTM, Jacobs et al. (2021) incorporated interdependencies between motivations and the time dimension by modelling customer motivations at the shopping basket level. This approach yields superior predictive performance and a deeper understanding than traditional LDA models.

In this thesis, we present an innovative segmentation methodology based on the CTM approach that uncovers the underlying motivations driving customers' shopping behaviours. Specifically, we focus on its implementation within the grocery retail industry. As such, we contribute to the existing literature in two ways. First, we extend the CTM approach by incorporating promotion-related information to identify latent motivations that are more likely in discounted shopping baskets, which is valuable information for retail companies (Aguilar-Palacios et al., 2021). Second, since previous research has not explored regional differences in motivations, our research addresses this gap by investigating whether differences exist between the motivations of city and non-city customers.

We demonstrate our model's power using an extensive dataset from a large Dutch grocery retailer, encompassing over 120,000 transactions made by more than 1,100 customers. We use variational Bayesian inference to estimate our CTM model with customer-, shopping trip-, and promotion-related features, a fast and scalable estimation method suitable for large-scale models (Blei et al., 2017). We base our model's evaluation on the interpretability of motivations, using coherence and similarity measures and predictive performance.

To preview our findings, we have identified 30 interpretable motivations summarising grocery customers' shopping behaviour. These motivations are diverse, ranging from sustainability preferences to dietary choices and budget-conscious shopping habits. Furthermore, our research reveals motivations that respond considerably to discounts and motivations that remain relatively unaffected by discount offers. Additionally, we uncover significant effects of customer demographics and shopping trip contexts on motivation relevance, along with notable correlations between motivations. Equipped with this information, we can unveil customers' intricate shopping preferences over time and design a product recommendation algorithm distinguished by its predictive accuracy, novelty of recommendations, and proficiency in generating predictions for new customers.

The remainder of the thesis is organised as follows: In Chapter 2, we conduct a comprehensive review of the current literature, concentrating on the applications of topic modelling within the retail context. In Chapter 3, we describe the specifics of the dataset we have used to estimate our correlated topic model. Chapter 4 presents a detailed examination of our chosen methodology and the variational inference algorithm employed. Our model's outcomes and practical applications are presented in Chapter 5. Lastly, in Chapter 6, we discuss our results, drawing conclusions regarding their managerial implications and inherent limitations and suggesting potential avenues for further research.

Chapter 2

Literature

This chapter will provide a detailed discussion of previous research on applying topic modelling within the retail domain. Our primary objective is to outline how these studies have successfully uncovered the shopping motivations of customers. Importantly, we also examine how these approaches have tackled the data sparsity challenges described in Chapter 1.

2.1 Latent Dirichlet Allocation

In recent research, scholars have explored the application of topic modelling as a flexible framework for clustering extensive datasets within the retail sector. This exploration has arisen from the recognition that specific characteristics of customer behaviour in grocery retailing make conventional clustering techniques unsuitable. Unlike more standard applications, the preferences of grocery customers often span multiple segments simultaneously (Jacobs et al., 2016). Additionally, customer preferences within this domain are dynamic rather than static, as Jacobs et al. (2021) highlighted. Another critical challenge in this context is data sparsity, as some customers have limited shopping histories (Carrasco et al., 2022).

Topic modelling methods have proven their effectiveness in addressing this challenge of data sparsity in large retail settings (Reisenbichler & Reutterer, 2019). One of the most widely adopted Bayesian topic modelling techniques, latent Dirichlet allocation (LDA), has proven to be a flexible and scalable unsupervised machine learning technique with significant potential. Originally designed for uncovering latent topics within textual documents, LDA assumes that documents are composed of random mixtures over latent topics, where each topic is a distribution over words (Blei et al., 2003).

A different perspective is required to apply LDA effectively in the retail context. Here, customer purchase histories are treated as documents, and the purchased products are viewed as words. In this transformed context, LDA is a valuable tool for unveiling the underlying topics, referred to as 'motivations' by Jacobs et al. (2016), that drive a customer's purchasing behaviour. These motivations are distributions over the entire product assortment, highlighting products frequently purchased together to fulfil a particular purpose or motivation (Carrasco et al., 2022). For example, a barbecue motivation would result in buying barbecue-related products. As such, the frequency with which products are purchased together in a shopping basket determines the fit of LDA (Hornsby et al., 2020).

LDA belongs to the family of mixed membership models, assigning every product a probability of belonging to each shopping motivation. Specifically, LDA is a generalisation of a finite mixture model, as it assumes that observed purchasing behaviours come from multiple latent groups (i.e., motivations) that differ in proportions for each customer, all from the same distributional family (Reisenbichler & Reutterer, 2019). Therefore, selecting a specific product means doing multiple draws from this finite mixture. For a new customer, this process will also be a finite mixture with the same mixture of components but different proportions. LDA is comparable to exploratory or model-based factor analysis, as both aim to reduce dimensionality by identifying latent common factors from a large dataset (Reisenbichler & Reutterer, 2019).

In retail research, the identified motivations through LDA can reflect a customer's purchasing preferences (Reisenbichler & Reutterer, 2019). The methodology implies that customers differ in their tendency to be driven by each motivation and can be motivated by multiple motivations when shopping. By modelling purchase behaviour through the lens of customer motivations, specific similarities can be discovered between customers based on shared motivations (Jacobs et al., 2016). Furthermore, these motivations transcend product groups, meaning that multiple product groups can belong to one motivation, while a single product group can correspond to multiple motivations (Jacobs et al., 2016).

2.2 Extensions of LDA

Building upon the LDA framework, Jacobs et al. (2016) introduced their 'LDA-X' model, which proved to be a powerful tool for describing and predicting customer behaviour within the context of a medium-sized online retailer in the Netherlands. The model posits that specific motivations are associated with preferences for certain products in the retailer's assortment, such as baby products, toiletries, or weight care products. Purchases of specific products might also link to multiple motivations. For instance, a customer motivated by environmental care may prefer vegan toiletries, while another customer motivated by animal care may also purchase these products. To account for changes in customer types over time, the model also incorporates the time of adoption at the retailer as a variable. This variable is defined as the number of days between the initial operating day of the retailer and the customer's first order. However, it is important to note that this variable is limited in its ability to capture all relevant factors related to the dynamics of customer behaviour.

In addition to purchasing data, the model proposed by Jacobs et al. (2016) incorporates available customer characteristics. This approach addresses questions related to the impact of demographics on purchasing behaviour, such as how age affects specific motivations. Furthermore, including customer information can improve the model's predictive capabilities, particularly in cases where purchase data is limited, such as for a new customer (Jacobs et al., 2016).

The proposed model by Jacobs et al. (2016) is powerful since it can provide real-time predictions and recommendations, even when interacting with massive amounts of data. The model's strength is particularly notable when recommending a single product, and its recommendations include products in the long tail of a store's product assortment. Consequently, recommendations are more creative and are potentially valued more by customers. Moreover, the model demonstrates scalability, can deal with additional information that might be available, and performs better than traditional methods such as collaborative filtering. Collaborative filtering, a popular method for predicting a customer's future purchase, faces various challenges, such as data sparsity, the cold-start problem, and the difficulty of including customer characteristics (Gharahighehi et al., 2022; Joorabloo et al., 2020). It may also face difficulties in capturing purchase patterns driven by less frequently bought products (Jacobs et al., 2016).

Building upon their earlier work, Jacobs et al. (2021) introduce an extension to the LDA-X model to enhance the treatment of the time dimension and model customer behaviour more realistically. Instead of regarding a customer's entire purchase history as a single basket, the model treats each shopping trip separately. This approach accounts for differences in shopping motivations between weekdays and weekends and seasonal changes in behaviour. Furthermore, this approach may provide insights into sudden changes in customer behaviour, such as adopting a vegetarian diet.

2.3 Correlated Topic Modelling

One of the limitations of using LDA in the context of large-scale retail is its inability to model correlations between topics or motivations. This limitation stems from LDA's use of the Dirichlet distribution to model the variability in motivation proportions, which assumes independence (Blei et al., 2003). Blei & Lafferty (2007) introduced the correlated topic model (CTM) to address this shortcoming. The CTM model uses the logistic normal distribution to be more flexible than the Dirichlet, incorporating a covariance structure between motivations (Blei & Lafferty, 2007). Consequently, the CTM model has a higher predictive performance than LDA and can discover patterns between motivations that are useful for the retailer (Blei & Lafferty, 2007). The generative process assumed by CTM is nearly identical to that of LDA, with the primary distinction being that it draws customers' motivation proportions from a logistic normal distribution instead of a Dirichlet distribution (Blei & Lafferty, 2007).

Inspired by CTM, Jacobs et al. (2021) incorporated the correlation structure between motivations in their extension of the LDA-X model. Their findings highlight that motivations can have positive correlations, such as one associated with general cleaning and one related to floor cleaning. Retailers can leverage these correlations to identify cross-selling opportunities (Jacobs et al., 2021).

Jacobs et al. (2021) decided to use a set of 100 motivations. However, they did not provide specific performance or interpretability measures to substantiate this choice. The resulting motivations are specific and encompass different product categories. Furthermore, low-volume products often define specific motivations, indicating the model's ability to identify purchase patterns related to long-tail products. Jacobs et al. (2021) also found that demographics and other explanatory variables related to each shopping trip significantly impact motivations. Moreover, their obtained correlation matrix shows intuitive positive correlations, for instance, between two gardening-related motivations. Ultimately, the authors concluded that their model outperforms the LDA-X model proposed by Jacobs et al. (2016).

2.4 Evaluation of Topic Models

The studies by Carrasco et al. (2022) and Hornsby et al. (2020) extended their research scope beyond extracting motivations from purchase histories. These papers also studied the quality of resulting motivations, a dimension overlooked in the previously discussed literature, highlighting its importance in model selection.

Carrasco et al. (2022) focused on the inherent uncertainty in motivations, with resulting topics or motivations appearing and disappearing in different posterior samples. The authors introduced measures to assess topic coherence, distinctiveness, and credibility to address this challenge. A topic is coherent if the most probable products link to a single motivation, distinctive if it differs from other topics, and credible if it appears in many LDA posterior draws. Carrasco et al. (2022) also adopted a hierarchical clustering approach to summarise topics from multiple posterior distributions.

In a similar effort, Hornsby et al. (2020) conducted extensive studies to evaluate the quality of the motivations extracted using their model. The authors first asked participants to assign motivation labels to the ten most relevant products, ensuring that the assigned motivation labels were representative. Subsequently, they assessed the coherence of the resulting motivations by asking participants to identify any 'intruding' products that did not belong to a particular motivation. In a final assessment, Hornsby et al. (2020) tested whether the LDA probabilities could predict a customer's age, gender, and region. The tests produced significant results, indicating the effectiveness of the LDA model in extracting motivations from purchase histories.

Overall, research demonstrates that topic modelling offers a scalable statistical framework capable of handling large volumes of data while providing valuable insights into customer behaviours. Particularly promising in large-scale retail contexts is the integration of correlation structures within the correlated topic model. In the following chapter, we will introduce the specific dataset used to refine our approach to uncovering customer motivations.

Chapter 3

Data

For this research, we use data from a prominent grocery retailer in the Netherlands. The retailer possesses extensive transaction-level and customer-level data. The transaction data covers online and in-store transactions, although this study focuses on in-store purchases. Given the dynamic nature of customer motivations, our analysis will be on recent purchases spanning two months. Specifically, we will use transactions between March 1st, 2023, and May 1st, 2023. Leveraging loyalty cards and customer IDs, transaction records for this period can be associated with individual customers. However, the data has been fully anonymised to safeguard customers' privacy. We restrict our analysis to customers who have used personal loyalty cards, ensuring the possibility of tracking the same customer's purchasing patterns over time.

The raw transaction-level dataset provides information on in-store product purchases, encompassing approximately 20,000 distinct products across various categories such as food, pharmaceuticals, household goods and pet food. The retailer has detailed information on these products, including their taxonomy, category levels, and labels indicating vegan, vegetarian, or biological attributes. For this study, we will aggregate products based on their lowest-level taxonomy, representing the product's most specific subgroup. This lowest level provides sufficiently detailed information to obtain actionable and interpretable purchasing motivations. For example, the product 'dark brown gluten-free half bread' will be aggregated to 'gluten-free bread' and 'plant-based XL burger' becomes 'burger (vegetarian)'. This aggregation reduces the dataset to 2,965 products. Table 3.1 provides an overview of the ten most frequently purchased lowest-level taxonomies.

Product	**Basket %**	**Product**	**Basket %**
Bananas	12.98%	Free-range eggs	5.46%
Milk (fresh)	9.08%	Self-baking bread	4.66%
Cucumber (fresh)	7.16%	Energy drinks	3.07%
Rolls (fresh)	7.10%	Pilsner	2.90%
Fresh buns	6.34%	Fresh pastries	2.79%

Table 3.1: Top 10 Most Purchased Products (Lowest-Level Taxonomies) with Basket Percentages

Typically, a crucial pre-processing step in correlated topic models is removing stop words, frequently occurring words without significant meaning (Hornsby et al., 2020). However, in this study, we will maintain all stop words as we believe every product contributes to understanding different customer motivations.

Given the substantial volume of data, processing the entire dataset is impractical and inefficient. Therefore, we have chosen a more focused and manageable approach, balancing computational efficiency with the need for a sufficient number of observations to enable meaningful analysis. Specifically, we have extracted a random sample equivalent to 0.02% of the relevant customer population. This final dataset retains a representative subset of 122,645 transactions involving 1,131 unique customers across 11,798 shopping trips. This dataset reveals a diverse range of shopping behaviours and motivations.

In addition to transaction-level data, customer-level information is also available. This data contains customer location details and the ages of 49% of the sampled customers. These details allow us to examine potential differences in customer preferences and motivations across different demographic groups. Moreover, we integrate shopping trip-level information, providing insight into the context of each purchase. This data includes factors like the time of day, the day of the week, and the discounted proportion of the total basket price. Summary statistics regarding these customer and shopping trip variables are presented in Appendix B.1.

In the next chapter, we will delve into our methodology and introduce our proposed model, providing a comprehensive framework for quantitatively assessing how customer and trip attributes collaboratively determine purchasing motivations.

Chapter 4

Methodology

This chapter will present our model for uncovering customers' grocery shopping motivations. The chapter starts with an introduction to the core model and our extensions in Sections 4.1 and 4.2. Then, Section 4.3 will outline the details of our estimation methodology using variational Bayesian inference techniques. The following sections, 4.5, 4.6, and 4.7, outline our strategies for evaluating the impacts of explanatory variables, assessing motivation quality, and measuring the model's performance in product recommendations, respectively.

4.1 The CTM Model

In this research, we will employ the correlated topic model (CTM) due to its distinct advantages in uncovering the relations between motivations, distinguishing it from the LDA method as explained in Section 2.3. Drawing inspiration from the findings by Jacobs et al. (2021), which highlight the presence of significant correlations among motivations of retail customers, we aim to explore whether similar correlations exist within the context of grocery retail.

Moreover, in addition to modelling products and customers, Reisenbichler & Reutterer (2019) have highlighted the significant promise of incorporating time-dependent marketing variables in marketing research. Considering that customers' behaviours and preferences evolve and motivations vary across shopping trips, factors such as changing personal tastes and contextual influences (e.g. seasonal effects) are critical. For instance, customers may be motivated during summer by barbecue-related products, whereas in winter, traditional Dutch stew-related products may be more appealing. Unlike many previous studies that aggregate purchases over time and combine all shopping trips into a single basket, we recognise that ignoring the time dimension disregards significant information. Hence, following the approach by Jacobs et al. (2021), we adopt a dynamic modelling perspective and distinguish between individual shopping trips in our CTM model.

To uncover the hidden motivations, our Bayesian CTM approach aims to find the posterior distribution of the latent variables conditioned on all the observed product purchases. We denote the products in the retailer's assortment by $j = 1, \dots, J$ and the customers by $i = 1, \dots, I$. In the transaction dataset outlined in Chapter 3, each customer i engages in B_i shopping trips.

The total amount of N_{ib} products purchased by customer i in the b-th shopping trip are stored in the set $\boldsymbol{y}_{ib}$, where $y_{ibn} \in \{1, \ldots, J\}$ for $n = 1, \ldots, N_{ib}$.

The CTM model's generative process for a specific customer $i \in \{1, \ldots, I\}$ going to the store and buying a total of N_{ib} products in the b-th shopping trip is as follows. For purchasing the n-th product, the customer i selects a motivation m, denoted by $z_{ibn} \in \{1, \ldots, M\}$. Each activated motivation z_{ibn} depends on a customer-specific motivational probability vector $\boldsymbol{\theta}_{ib} = [\theta_{ib1}, \ldots, \theta_{ibM}]$. This motivational vector summarises the importance of each motivation for a specific customer during the b-th shopping trip. So, the probability that the activated motivation for purchase by customer i in shopping trip b was motivation m is θ_{ibm}:

$$P(z_{ibn} = m \,|\, \boldsymbol{\theta}_{ib}) = \theta_{ibm}, \tag{4.1}$$

where $\theta_{ibm} \geq 0$ and $\sum_m \theta_{ibm} = 1$. As described in Section 2.3, the CTM model samples these customer-specific motivation vectors $\boldsymbol{\theta}_{ib}$ from a logistic normal distribution to allow for correlations between motivations. We follow the suggestion by Jacobs et al. (2021) to draw $\boldsymbol{\theta}_{ib}$ from the softmax function of $\boldsymbol{\alpha}_{ib}$, which is an unrestricted, stochastic parameter vector that we will model in Section 4.2:

$$\boldsymbol{\theta}_{ib} \equiv \text{softmax}(\boldsymbol{\alpha}_{ib}) = \frac{\exp(\boldsymbol{\alpha}_{ib})}{\sum_{m=1}^{M} \exp(\alpha_{ibm})}. \tag{4.2}$$

Each $\boldsymbol{\alpha}_{ib} \in \mathbb{R}^M$ is an M-dimensional vector, in which scalar elements α_{ibm} represent the prior importance of motivation m for customer i in the b-th shopping trip. If a value α_{ibm} is large, there is a higher probability that a customer makes a purchase caused by motivation m. If a-priori, a motivation is not relevant for a customer, its weight will be close to zero.

Then, based on the activated motivation z_{ibn} that drives the purchase, a specific product j is bought. This specific product is sampled from the probability vector belonging to the activated motivation $\boldsymbol{\phi}_{z_{ibn}}$. Each motivation $m = 1, \ldots, M$ is defined by a vector $\boldsymbol{\phi}_m = [\phi_{m1}, \ldots, \phi_{mJ}]$. Each $\boldsymbol{\phi}_m$ is a J-dimensional probability vector over the entire product assortment. In this probability vector, each element $\phi_{mj} \geq 0$ represents the probability that product j will be bought if the customer activates motivation m, where $\sum_j \phi_{mj} = 1$:

$$P(y_{ibn} = j \,|\, z_{ibn} = m, \boldsymbol{\phi}) = \phi_{mj}. \tag{4.3}$$

We need to estimate the unknown motivation-specific $\boldsymbol{\phi}_m$ probability vectors from the data. The product co-purchases observed during shopping trips are crucial in uncovering these probability vectors. If a specific set of products is consistently bought together in many shopping trips, it implies a connection with a specific motivation (Jacobs et al., 2021).

The CTM model incorporates a Dirichlet prior on the $\boldsymbol{\phi}_m$ vectors, with prior parameters $\boldsymbol{\beta}_0$, indicating prior probabilities of products for each motivation. Usually, each element of $\boldsymbol{\beta}_0$ is set to a standard value of J^{-1}, where J is the total number of products. This choice ensures that the prior does not give preference to any particular products but treats them equally in terms of initial probabilities. Hence, we can express $\boldsymbol{\beta}_0$ as a vector of ones ($\mathbf{1}$) of length J scaled by J^{-1}:

$$\boldsymbol{\phi}_m \sim \text{Dirichlet}_J(\boldsymbol{\beta}_0 = \mathbf{1}_J J^{-1}). \tag{4.4}$$

To summarise, (4.5) defines the marginal probability that customer i will buy product j in the b-th shopping trip. In this equation, the likelihood that a customer has a specific motivation when visiting the store (θ_{ibm}) is multiplied by the probability that a particular product j is bought when a customer activates this specific motivation (ϕ_{mj}), summed over all motivations:

$$\begin{aligned}
&P(y_{ibn} = j \mid \{\boldsymbol{\phi}_l\}_{l=1}^{M}, \boldsymbol{\theta}_{ib}) \\
&= \sum_{m=1}^{M} P(y_{ibn} = j \mid z_{ibn} = m, \{\boldsymbol{\phi}_l\}_{l=1}^{M}) P(z_{ibn} = m \mid \boldsymbol{\theta}_{ib}) \\
&= \sum_{m=1}^{M} \phi_{mj} \cdot \theta_{ibm}.
\end{aligned} \tag{4.5}$$

4.2 Extending the CTM Model

In this section, we present our novel contributions to the existing body of literature regarding the application of CTM within the retail context. Our contributions include the incorporation of the effects of promotions and customer location into the inference of customers' motivations.

To begin, building upon the recommendations of Jacobs et al. (2021), who suggest incorporating various retailer-controlled factors, we have chosen to focus on integrating promotions into our CTM model. One of the primary goals for retailers is to optimise their promotional strategies to maximise profitability. Achieving this entails minimising discounts offered on products that customers would have purchased even without promotions, while strategically promoting products that have the potential to stimulate additional sales. Consequently, retailers must distinguish between customers' intrinsic motivations and those influenced by discounts.

To account for the influence of promotions, we will introduce a model feature that quantifies the promotional intensity of a customer's basket. Specifically, this feature represents the percentage discount applied to the total price of the customer's basket. By analysing the significance of the promotion-related feature in motivation assignments, we can indicate which motivations seem highly sensitive to discounts (exhibiting a high price elasticity of demand) and which seem to have relatively stable demand.

Besides this feature on the promotional intensity of a basket b of customer i, we will incorporate other trip-specific variables, including the time of day and the day of the week. These are denoted by the K_X-dimensional vector $\boldsymbol{x}_{ib}$, where K_X denotes the total amount of trip-specific variables. In addition to incorporating these trip-specific features, we will add to the literature by integrating location information for each customer i. This information will be part of the customer-specific demographic variables represented by the K_H-dimensional vector $\boldsymbol{h}_i$. Where K_H denotes the total amount of variables included to describe the age and location of each customer.

Following the methodology proposed by Jacobs et al. (2021), we assume that these customer- and trip-specific variables influence the prior likelihood of each customer's motivations in a specific shopping trip. This influence is reflected in $\boldsymbol{\alpha}_{ib} = [\alpha_{ib1}, \ldots, \alpha_{ibM}]$, where each element α_{ibm} corresponds to a specific motivation m. This parameter vector is the input for the softmax function in Equation (4.2), and the resulting output will be the prior probability distribution for the customer-specific motivation vectors $\boldsymbol{\theta}_{ib}$. Hence, we model α_{ibm} as follows:

$$\alpha_{ibm} = \mu_{ibm} + \epsilon_{ibm} = \kappa_{im} + \boldsymbol{x}_{ib}^{\intercal}\boldsymbol{\delta}_m + \boldsymbol{h}_i^{\intercal}\boldsymbol{\gamma}_m + \epsilon_{ibm}. \tag{4.6}$$

The model for α_{ibm} is linear, consisting of a predictable part μ_{ibm} and a stochastic part ϵ_{ibm}. The predictable part μ_{ibm} consists of model parameters whose posterior distributions will be estimated using our CTM model.

To elaborate, the customer-specific intercept κ_{im} will capture the baseline importance of motivation m for a customer i throughout all shopping trips. The vector $\boldsymbol{\kappa}_i = [\kappa_{i1}, \ldots, \kappa_{iM}]$ captures the M intercepts for each customer i. We define $\boldsymbol{\kappa}_i$ as a multivariate normal distribution:

$$\boldsymbol{\kappa}_i \sim \text{MVN}_M(\boldsymbol{\mu}_\kappa, \boldsymbol{\Sigma}_\kappa). \tag{4.7}$$

The parameter vector $\boldsymbol{\mu}_\kappa$ indicates the presence of each motivation across all customers, while the covariance matrix $\boldsymbol{\Sigma}_\kappa$ captures the relevant correlation structure between the motivations. The priors for both the mean and covariance (transformed to precision matrix $\boldsymbol{\Lambda}_\kappa \equiv \boldsymbol{\Sigma}_\kappa^{-1}$) of $\boldsymbol{\kappa}_i$, as suggested by Jacobs et al. (2021), are as follows:

$$\boldsymbol{\mu}_\kappa \sim \text{MVN}_M(\boldsymbol{\mu} = \mathbf{0}_M, \boldsymbol{\Sigma} = \mathbf{I}_M), \tag{4.8}$$

$$\boldsymbol{\Lambda}_\kappa \sim \text{Wishart}_M(n = 2M, \boldsymbol{V} = \boldsymbol{I}_M(2M)^{-1}). \tag{4.9}$$

It is important to note that we do not have strong prior beliefs regarding the correlations between motivations. Therefore, we adopt an empirical approach, allowing the observed data to inform us about these correlations. Our choice of non-informative priors reflects this approach.

The motivation-specific parameters $\boldsymbol{\delta}_m$ and $\boldsymbol{\gamma}_m$ represent the effects of the trip- and customer-specific factors on the importance of each motivation. We will estimate the posterior distributions of these parameters using our data, and they are assigned the following priors:

$$\boldsymbol{\delta}_m \sim \text{MVN}_{K_X}(\boldsymbol{\mu} = \mathbf{0}_{K_X}, \boldsymbol{\Sigma} = \boldsymbol{I}_{K_X}), \tag{4.10}$$

$$\boldsymbol{\gamma}_m \sim \text{MVN}_{K_H}(\boldsymbol{\mu} = \mathbf{0}_{K_H}, \boldsymbol{\Sigma} = \boldsymbol{I}_{K_H}). \tag{4.11}$$

Lastly, the stochastic part ϵ_{ibm} accounts for unexplained variation in motivational importance after accounting for the trip- and customer-specific factors. We will model this error term using a normal distribution:

$$\epsilon_{ibm} \sim N(0, \sigma^2_{\alpha_m}). \tag{4.12}$$

Here, the motivation-specific variance $\sigma^2_{\alpha_m}$ allows specific motivations to exhibit more variability than others. The prior specifications for these motivation-specific variances and further details regarding prior specifications of model parameters are provided in Appendix F. We assume that the error terms ϵ_{ibm} are independent across customers and that there is no auto-correlation between the error terms of different shopping trips or motivations for a given customer.

4.3 Variational Inference

The generative process of our CTM model using M motivations for a customer making B shopping trips and buying N items in each trip can be summarised by the joint distribution of latent and observable variables in the model:

$$p(\boldsymbol{z}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\sigma_\alpha}, \boldsymbol{\mu}_\kappa, \boldsymbol{\Sigma}_\kappa, \boldsymbol{y}|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\beta}_0) = \prod_{m=1}^{M} p(\boldsymbol{\phi}_m|\boldsymbol{\beta}_0) \prod_{b=1}^{B} p(\boldsymbol{\theta}_b|\boldsymbol{\alpha}_b) \left(\prod_{n=1}^{N} p(z_{b,n}|\boldsymbol{\theta}_b) p(y_{b,n}|\boldsymbol{\phi}_{z_{b,n}}) \right)$$
$$\times\, p(\boldsymbol{\kappa}|\boldsymbol{\mu}_\kappa, \boldsymbol{\Sigma}_\kappa) \left(\prod_{m=1}^{M} p(\boldsymbol{\delta}_m) \prod_{m=1}^{M} p(\gamma_m) \prod_{b=1}^{B} p(\boldsymbol{\epsilon}_b|\boldsymbol{\sigma}_\alpha) \right) \times p(\boldsymbol{\mu}_\kappa) p(\boldsymbol{\Sigma}_\kappa) p(\boldsymbol{\sigma}_\alpha). \tag{4.13}$$

To estimate the conditional density of the latent variables given the observed product purchases $\boldsymbol{y}$, it is necessary to compute the conditional posterior distribution of the topic structure, which is of interest. Following Bayes' theorem:

$$p(\boldsymbol{z}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\sigma_\alpha}, \boldsymbol{\mu}_\kappa, \boldsymbol{\Sigma}_\kappa|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\beta}_0) = \frac{p(\boldsymbol{z}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\sigma_\alpha}, \boldsymbol{\mu}_\kappa, \boldsymbol{\Sigma}_\kappa, \boldsymbol{y}|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\beta}_0)}{p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{\beta}_0)}. \tag{4.14}$$

However, this poses an estimation problem as calculating the denominator in Equation (4.14) is required, representing the observations' marginal probability, commonly known as the 'evidence'. Unfortunately, directly computing the evidence integral is infeasible due to the need to sum over all possible hidden topic structures, making it computationally intractable (Blei & Lafferty, 2007). Therefore, to obtain estimates for the model parameters and motivations, approximation methods must be employed to estimate the posterior distribution.

When selecting an estimation method, a difficulty arises from the CTM model. The CTM model replaces the convenient Dirichlet distribution, conjugate to the motivation assignment, with the logistic normal or softmax distribution in (4.2), which is not conjugate to the multinomial distribution. Consequently, posterior inference becomes more challenging (Blei & Lafferty, 2007). Traditional approximation methods like Gibbs sampling are no longer suitable (Reisenbichler & Reutterer, 2019). Moreover, the scale and dimensionality of the data make Metropolis-Hastings MCMC sampling impracticable (Blei & Lafferty, 2007). Additionally, due to the hierarchical nature of our model and the numerous customer- and trip-specific parameters, Hamiltonian

Monte Carlo (HMC) is also ineffective (Jacobs et al., 2021). As a result, recent research has turned to variational Bayes, also called variational inference (VI), as an approximate method for estimating the CTM model, improving speed and scalability (Hornsby et al., 2020; Jacobs et al., 2021). Therefore, our study will rely on variational inference to estimate our model.

4.4 Estimation Using VI

In Section 4.3, we presented our Bayesian CTM model. This model consists of latent variables and parameters, which we will denote by the set $\boldsymbol{\omega}$. Our primary objective is to approximate the intractable posterior distribution $p(\boldsymbol{\omega}|\boldsymbol{y})$ as defined in Equation (4.14). To achieve this, we employ the Bayesian technique of variational inference, which transforms the process of posterior inference into an optimisation problem.

The fundamental idea behind variational inference is to select a predetermined family of tractable joint probability distributions, denoted as Q, over the unknown model parameters $\boldsymbol{\omega}$. We refer to a specific distribution $q(\boldsymbol{\omega}) \in Q$ as a 'variational distribution' (Bishop, 2006). The goal of our optimisation process is to identify the variational distribution $q(\boldsymbol{\omega})$ within Q that minimises the Kullback-Leibler (KL) divergence from the exact conditional posterior distribution $p(\boldsymbol{\omega}|\boldsymbol{y})$:

$$q^*(\boldsymbol{\omega}) = \arg \min_{q(\boldsymbol{\omega}) \in Q} \mathrm{KL}(q(\boldsymbol{\omega}) \,\|\, p(\boldsymbol{\omega}|\boldsymbol{y})), \tag{4.15}$$

$$\mathrm{KL}(q(\boldsymbol{\omega}) \,\|\, p(\boldsymbol{\omega}|\boldsymbol{y})) = \mathbb{E}[\log q(\boldsymbol{\omega})] - \mathbb{E}[\log p(\boldsymbol{\omega}, \boldsymbol{y})] + \log p(\boldsymbol{y}). \tag{4.16}$$

This optimised candidate density $q^*(\boldsymbol{\omega})$ serves as an approximation for the exact conditional density (Blei et al., 2017). However, as mentioned in Section 4.3, directly computing the evidence $p(\boldsymbol{y})$ is computationally infeasible. To address this challenge, Blei et al. (2017) proposed an alternative objective that is equivalent to the KL divergence up to $\log p(\boldsymbol{y})$, which is a constant with respect to $q(\boldsymbol{\omega})$:

$$\mathrm{ELBO}(q) = \mathbb{E}[\log p(\boldsymbol{\omega}, \boldsymbol{y})] - \mathbb{E}[\log q(\boldsymbol{\omega})]. \tag{4.17}$$

This objective function is known as the Evidence Lower Bound (ELBO). The first term corresponds to the expectation of the joint density being approximated, while the second term represents the entropy of the variational density. Maximising this ELBO is equivalent to minimising the KL divergence and thus approaching the true posterior (Jordan et al., 1999). The complexity of this task depends on the family of distributions used for optimisation (Bishop, 2006). The restricted family of distributions Q should balance being flexible enough to approximate the density closely and simple enough to optimise efficiently.

Given that our model specification in Section 4.3 exclusively contains distributions belonging to the exponential family, we can leverage the variational inference algorithm proposed by Jacobs et al. (2021). This algorithm has demonstrated statistical and computational efficiency. Their approach employs mean-field variational inference, which assumes that the variational distribution $q(\boldsymbol{\omega})$ can be factorised over simpler distributions. Specifically, it factors the variational distribution over each unknown latent variable $\omega_v \in \boldsymbol{\omega}$ where $v = 1, \ldots, V$ and V is the total number of latent variables in the model. Consequently, each latent variable ω_v is governed by its corresponding variational factor, the variational density $q_v(\omega_v)$ (Bishop, 2006):

$$q(\boldsymbol{\omega}) = \prod_{v=1}^{V} q_v(\omega_v). \tag{4.18}$$

Using a Coordinate Ascent Variational Inference (CAVI) optimisation algorithm, we iteratively update these variational factors $q_v(\omega_v)$ to maximise the ELBO in Equation (4.17), until achieving convergence. Boyd & Vandenberghe (2004) have demonstrated that this approach guarantees at least a local optimum. The expressions for these updates of the variational distributions $q_v(\omega_v)$ corresponding to each parameter in $\boldsymbol{\omega}$ have been derived by Jacobs et al. (2021). Detailed explanations and solutions regarding these derivations are presented in Appendix G.2 and G.3, respectively.

The monitor the CAVI algorithm's convergence, we evaluate the ELBO after each iteration. A comprehensive outline of the algorithm and the initialisation procedures designed to ensure a consistent solution, as recommended by Jacobs et al. (2021), are provided in Appendix G.1.

4.5 Effects of Explanatory Variables

In the previous sections, we have provided an overview of our model, which includes customer-related information and specific shopping trip attributes. However, the relationship between these variables and customers' motivation-activation probabilities ($\boldsymbol{\theta}_{ib}$) is not straightforward due to the non-linear nature of the softmax function in Equation (4.2). This function depends on $\boldsymbol{\alpha}_{ib}$, which consists of $\boldsymbol{\mu}_{ib}$ and the error term $\boldsymbol{\epsilon}_{ib}$ (refer to Equation (4.6)). Therefore, understanding the effect of each variable on motivation importance requires evaluating odds ratios.

To evaluate odds ratios, we need to compare the motivation probabilities associated with specific values of each variable to a baseline scenario. For this baseline, denoted as $\boldsymbol{\mu}^B$, we use sample means for $\boldsymbol{x}_{ib}$ and $\boldsymbol{h}_i$, the population mean for $\boldsymbol{\kappa}_i$ ($\boldsymbol{\mu}_\kappa$), and the posterior means for the estimated parameters $\boldsymbol{\delta}_m$ and $\boldsymbol{\gamma}_m$, essentially representing the characteristics of an average shopping trip. Subsequently, we modify each variable separately while keeping the other variables constant with respect to the average shopping trip. For binary variables (dummies), we set one specific level to 1, and for the continuous variable basket discount percentage, we consider a shock of 50% compared to the average shopping trip. We use these altered values to calculate $\boldsymbol{\mu}^S$.

For the computation of odds ratios, we implement (4.19) as explained by Jacobs et al. (2021). In this equation, we evaluate the probability of a specific motivation m after applying a particular shift in variables (θ_m^S), relative to the baseline probability (θ_m^B) while integrating out the error term ($\boldsymbol{\epsilon}$). Here, $p(\boldsymbol{\epsilon})$ denotes the density of $\boldsymbol{\epsilon}$, as defined in Equation (4.12).

$$\mathbb{E}\left\{\frac{\theta_m^S}{\theta_m^B}\right\} = \int_\epsilon \frac{\exp(\mu_m^S + \epsilon_m)/\sum_{l=1}^M \exp(\mu_l^S + \epsilon_l)}{\exp(\mu_m^B + \epsilon_m)/\sum_{l=1}^M \exp(\mu_l^B + \epsilon_l)} * p(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon}. \tag{4.19}$$

This approach enables us to evaluate how particular motivations become either more or less likely as a result of various factors, such as a customer's location and the timing of their shopping trip.

4.6 Motivation Quality

The motivations discovered through our CTM model must be relevant but also interpretable. In business contexts, topic models often require thorough refinement and validation by company experts (Chuang et al., 2012). In this section, we will provide detailed explanations of the methods employed to assess the usefulness and robustness of the resulting motivations.

Although recent extensions of the LDA model, such as the utilisation of the hierarchical Dirichlet process, have enabled the estimation of the number of topics directly from the data, it remains common practice to predefine the number of motivations (Blei & Lafferty, 2007). In the case of large retail settings, previous studies have commonly employed a range of 25 to 100 motivations (Hornsby et al., 2020; Jacobs et al., 2021). Given the extensive range of products offered by large grocery retailers and the numerous customers, we expect substantial variability in individuals' purchasing behaviour. Consequently, selecting a relatively high number of motivations may be necessary to identify specific purchasing patterns. However, there is a potential issue of overfitting associated with this approach. Therefore, we must not set the number of motivations too high to ensure the relevance of our results. However, a too-small number of motivations may result in very broad motivations. In summary, there is a trade-off between the risk of overfitting and the generation of overly general motivations (Blair et al., 2020).

Typically, post-hoc methods are employed to evaluate the appropriateness of the chosen number of motivations. Recent research in this field has highlighted two aspects for assessing the quality of motivations: the 'coherence' of the associated products (1) and the 'distinctiveness' of the motivations (2) (Chuang et al., 2012).

To evaluate whether the resulting motivations represent a specific customer need, we utilise coherence as a first measure proposed by Carrasco et al. (2022). For example, products such as 'spaghetti', 'tomato sauce', and 'olive oil' can readily be linked to a motivation for preparing Italian pasta. In contrast, a motivation lacking coherence consists of products that do not appear to serve a specific customer need. Hence, topic coherence ensures that the resulting motivations are interpretable for human understanding.

Research has shown that measures of product co-occurrence, such as point-wise mutual information (PMI) and its normalised version (NPMI) demonstrate a high correlation with topic coherence as judged by human experts (Lau et al., 2014). The intuition behind these measures is that when products frequently appear together within shopping baskets, it suggests a level of semantic relatedness between them (Aletras & Stevenson, 2014). NPMI, as a standardisation of PMI quantifies the likelihood of seeing two products within the same basket compared to their individual probabilities (Carrasco et al., 2022). Given that NPMI demonstrates a higher correlation with topic coherence than PMI, we compute the average NPMI scores for the ten most probable products to assess the coherence of a specific motivation (Lau et al., 2014):

$$\text{PMI}(y_i, y_j) = \log\left(\frac{P(y_i, y_j)}{P(y_i)P(y_j)}\right); \quad i \neq j \quad 1 \leq i, j \leq 10. \tag{4.20}$$

$$\text{NPMI}(y_i, y_j) = \frac{\text{PMI}(y_i, y_j)}{-\log P(y_i, y_j)}; \quad i \neq j \quad 1 \leq i, j \leq 10. \tag{4.21}$$

In these equations, $P(y_i)$ represents the probability of observing a specific product y_i in a basket across the entire dataset, while $P(y_i, y_j)$ is the probability of both products y_i and y_j appearing together in the same basket. Focusing on the top ten products aligns with standard practices in measuring average NPMI and PMI (Aletras & Stevenson, 2014; Lau et al., 2014). These scores range between -1 and 1, with scores closer to 1 indicating higher coherence.

In addition to evaluating the coherence of the resulting motivations, we will also assess their distinctiveness to validate the chosen number of motivations. This measure will indicate whether motivations are sufficiently distinct or exhibit high similarity, thus failing to provide new insights (Carrasco et al., 2022). For example, two motivations characterised by alcoholic beverages are not very distinct, as they convey the same customer need. Xing et al. (2019) have demonstrated the superiority of the cosine distance as a measure of topic distinctiveness compared to other similarity measures. As such, we will calculate the minimum cosine distance (CD) between a specific motivation $\boldsymbol{\phi}_m$ and different motivations $\boldsymbol{\phi}_l$ for $l = 1, \dots, M$, where $l \neq m$, to indicate the distinctiveness of that motivation (Carrasco et al., 2022). The cosine distance is defined as:

$$CD(\boldsymbol{\phi}_m, \boldsymbol{\phi}_l) = 1 - \frac{\boldsymbol{\phi}_m \cdot \boldsymbol{\phi}_l}{\|\boldsymbol{\phi}_m\|\|\boldsymbol{\phi}_l\|}. \tag{4.22}$$

By considering both coherence and distinctiveness, we have two evaluation measures to assess the quality of the resulting motivations generated by our model. Although predictive performance and likelihood measures are also commonly used to evaluate model effectiveness, we choose not to incorporate these to select the number of motivations in this study. As explained by Jacobs et al. (2021) and Carrasco et al. (2022), these measures do not account for the interpretability of the resulting motivations, potentially leading to an excessive number of motivations that sacrifice the practicality of the model.

Following these metrics, Stevens et al. (2012) have discovered that topic coherence generally increases with the number of topics or motivations. This finding suggests that when increasing the number of motivations, they become more specific to customer needs (Aletras & Stevenson, 2014). However, Carrasco et al. (2022) demonstrated that an excessive increase in the number of motivations may sacrifice topic distinctiveness. In other words, as more motivations are introduced, some may become increasingly similar to others.

Once the model has been estimated using variational inference with the appropriate number of motivations, the resulting outputs will include the motivations, each with assigned probabilities for individual products, as well as the proportions of motivations attributed to each customer (Blei & Lafferty, 2007). To facilitate managers' understanding, providing motivations with meaningful labels is crucial. The process of labelling motivations is inherently subjective and relies on the expertise of the retailer's employees. Typically, researchers assign topic labels to each motivation based on the associated products with the highest probabilities, selecting approximately ten to fifteen products for this labelling task (Carrasco et al., 2022; Jacobs et al., 2021). In this research, we entrust the labelling process to employees from different roles within the large Dutch grocery retailer who will base their labels on the top ten products.

4.7 Product Recommendations

Once we have successfully obtained insights into each customer's relevant motivations and understood the effects of customer- and trip-specific variables, a promising opportunity emerges: the ability to generate product recommendations based on customers' motivations. Therefore, we will assess the effectiveness of our model in predicting customers' future purchases.

We are interested in constructing a predictive distribution encompassing a customer's potential selections from the assortment. This predictive distribution is composed of the conditional probabilities that a customer buys a specific product j, which we will estimate as follows, building upon the work by Griffiths & Steyvers (2004):

$$\begin{aligned} P[\tilde{y}_{in} = j | \boldsymbol{z}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_0, \boldsymbol{y}] &= \sum_{m=1}^{M} P[\tilde{y}_{in} = j | \tilde{z}_{in} = m, \boldsymbol{z}, \boldsymbol{\beta}_0, \boldsymbol{y}] P[\tilde{z}_{in} = m | \boldsymbol{z}_i, \boldsymbol{\alpha}_i] \\ &= \sum_{m=1}^{M} \mathbb{E}[\phi_{mj} | \boldsymbol{z}, \boldsymbol{\beta}_0, \boldsymbol{y}] \mathbb{E}[\theta_{im} | \boldsymbol{z}_i, \boldsymbol{\alpha}_i] \\ &= \sum_{m=1}^{M} \phi_{mj} \cdot \frac{\alpha_{im} + \pi_{im}}{\sum_{l=1}^{M} \alpha_{il} + \pi_{il}}. \end{aligned} \tag{4.23}$$

Here, we use the model parameters $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_0$, the latent purchase assignments $\boldsymbol{z}$, and observed purchases $\boldsymbol{y}$. We need to evaluate two expectations to derive the predictive distribution: $\mathbb{E}[\phi_{mj} | \boldsymbol{z}, \boldsymbol{\beta}_0, \boldsymbol{y}]$ and $\mathbb{E}[\theta_{im} | \boldsymbol{z}_i, \boldsymbol{\alpha}_i]$.

To calculate the first expectation of the probability of purchasing product j when motivation m is activated, we use the probabilities ϕ_{mj} from the estimated posterior probability distributions $\boldsymbol{\phi}_m$. Since these motivation-product probability vectors are defined at the customer-base level, we have substantial information to estimate these posterior probabilities in our CTM model.

The second expectation involves estimating the customer-specific relevance of each motivation. To efficiently generate predictions for many customers and update these customer-specific motivation relevances when new shopping trips occur, we adopt a practical and computationally feasible approach to estimate this expectation. This approach involves an approximation step in which we calculate π_{im}, indicating the proportion of purchases by customer i driven by motivation m. We calculate these proportions $\boldsymbol{\pi}_i$ by the normalised dot product of a customer's product purchase frequencies $\boldsymbol{f}_i = [f_{i1}, \ldots, f_{iJ}]$ and the $(J \times M)$ matrix concatenating all motivation-product (column) vectors $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_M]$. For a specific customer i and motivation m:

$$\pi_{im} = \frac{\boldsymbol{f}_i \cdot \boldsymbol{\Phi}_{:m}}{\sum_{l=1}^{M} (\boldsymbol{f}_i \cdot \boldsymbol{\Phi}_{:l})}, \tag{4.24}$$

where $\mathbf{\Phi}_{:m}$ denotes the m-th column vector of $\mathbf{\Phi}$, corresponding to $\boldsymbol{\phi}_m$. This approach is an approximation to obtain motivation-relevance vectors for each customer based on product purchases only. Combining this information with the prior likelihoods $\boldsymbol{\alpha}_i$ as shown in Equation (4.23), we estimate $\mathbb{E}[\theta_{im}|\boldsymbol{z}_i, \boldsymbol{\alpha}_i]$.

Extending this methodology, our model can incorporate real-time, trip-specific information to generate recommendations tailored to the specific shopping trip. Nonetheless, this enhancement comes at the expense of increased computational demands. Therefore, we have not incorporated real-time, trip-specific information into our product recommendation process. The estimated correlation structure among motivations will also not directly be used, as we do not anticipate its importance in enhancing predictive performance.

The final computed conditional probabilities are aggregated to construct the ultimate predictive distribution over all products for each customer. Subsequently, by sampling from this predictive distribution, we can generate predictions and offer tailored product recommendations. Our proposed methodology guarantees the efficient updating of customer preferences while also enabling the generation of predictions for new customers.

To evaluate the effectiveness of our model-generated product recommendations, we consider prediction sets of different sizes (S). Depending on context, customers might receive a single product recommendation or multiple product options. To quantitatively evaluate the performance of our model, we employ the (average) hit rate as a fundamental metric, following the suggestions of Jacobs et al. (2016). This metric measures the proportion of recommended products $\boldsymbol{r}_i$ for a given customer i that were actually purchased by the customer in their last transaction, as indicated by the hold-out samples H_i. To make the metric comparable across prediction sets of different sizes, we divide this proportion by the maximum number of potential hits, which is the minimum of the number of recommended products S and the unique number of products bought by customer i in their hold-out transaction u_i'. The hit rate is represented by the formula:

$$\text{Hit Rate}_i = \frac{1}{\min(S, u_i')} \sum_{s=1}^{S} \mathbb{I}[r_{is} \in H_i]. \tag{4.25}$$

Additionally, we introduce the concept of the (average) 'novelty hit rate' to assess the novelty of our recommendations. This metric evaluates the hit rate on products a customer has bought for the first time during his or her last transaction, as denoted by the set N_i, where:

$$\text{Novelty Hit Rate}_i = \frac{1}{\min(S, u_{new,i}')} \sum_{s=1}^{S} \mathbb{I}[r_{is} \in N_i]. \tag{4.26}$$

To summarise, this chapter has presented a comprehensive overview of the CTM model, our proposed model extension, the estimation approach, the evaluation of the resulting motivations, and the generation of product recommendations. Subsequently, in the following chapter, we will present the results obtained through applying this methodology.

Chapter 5

Results

This chapter provides an overview of the outcomes obtained by applying the correlated topic model (CTM) on the large Dutch grocery retailer's dataset. It starts with a summary of the purchase motivations identified by the CTM model in Section 5.1. Following this, we illustrate how these motivations can effectively describe a customer's shopping journey in Section 5.2. The subsequent Section 5.3 delves into the effects of trip- and customer-specific variables incorporated within the model. Section 5.4 is dedicated to exploring the relations between different motivations. Lastly, Section 5.5 concludes this chapter by evaluating our model-generated product recommendations. The findings presented in this section rely on the estimated variational posterior distributions of the model parameters.

5.1 Shopping Motivations

We apply the CTM model outlined in Chapter 4 to analyse the transactions and customer data of the Dutch grocery retailer, as detailed in Chapter 3. The objective is to uncover the underlying motivations driving customer shopping behaviour. As discussed in Section 4.6, one crucial step is determining the appropriate number of motivations (M) for our CTM model.

Therefore, we proceed to evaluate the performance of different models, each employing varying numbers of motivations. Specifically, we examine M values of 10, 20, 30, 40, 50, 70 and 90. We will only consider CTM models with less than 90 motivations to ensure our model summarises customer behaviour in an interpretable and concise manner. To estimate each model, we complete 2,000 iterations of the CAVI algorithm mentioned in Appendix G.1 and capture samples at every 500 iterations. We achieve convergence of the Evidence Lower Bound (ELBO) for all trained models.

The evaluation of these models focuses on two essential measures, as described in Chapter 4: motivation coherence and distinctiveness. To ensure robust and reliable estimations of the different models, we calculate the mean performance of each model on these two measures across ten different random samples, generated using the same procedure as described in Chapter 3. A summary of the evaluation results for these models is presented in Table 5.1.

Model	Topics	Performance (mean)	
		Coherence (NPMI)	Distinctiveness (CD)
CTM-10	10	-0.177	0.894
CTM-20	20	-0.038	0.954
CTM-30	30	0.005	0.953
CTM-40	40	-0.104	0.942
CTM-50	50	-0.167	0.941
CTM-70	70	-0.191	0.907
CTM-90	90	-0.207	0.893

Table 5.1: Mean Coherence and Distinctiveness Scores for CTM Models with Different Numbers of Topics

Upon comparing the outcomes of models with varying sizes of M, differences in the resulting motivations emerge. Confirming the findings of Carrasco et al. (2022), we have noted that topic distinctiveness declines as the number of motivations increases above a certain threshold (in this case, M=20/30). When analysing the resulting motivations in large models, we have identified new motivations that were not apparent in models with fewer motivations. However, we have also noticed that many of those new motivations were similar, thus failing to provide new insights and causing overall distinctiveness to decrease. Furthermore, we have observed a decline in motivation coherence when using larger models with more than 30 motivations. This finding contradicts the conclusions of Stevens et al. (2012), who suggested that topic coherence generally increases with an increasing number of topics. Consequently, our observation suggests that using more than 30 motivations in our model introduces more noise than valuable insights.

Considering motivation coherence and distinctiveness, the CTM model with 30 motivations is optimal for the Dutch grocery retailer to summarise customer behaviour. This model's output reveals a diverse range of shopping motivations. As described in Chapter 4, we have utilised a panel of employees representing diverse roles within the large Dutch grocery retailer to label each motivation based on the ten most probable products. The results of this labelling process can be found in Appendix A, showing the resulting motivations, including the product's importance probabilities captured by the corresponding $\boldsymbol{\phi}_m$ vectors.

These motivations identified by our CTM-30 model encompass a diverse range of customer behaviours, including customers' meal and product type preferences, lifestyles, allergies, activities, holidays, specific pet needs, and budgets. For instance, motivations 25, 14, 18, 28 and 15 highlight preferences for 'typical Dutch' products and motivations for cooking a 'Mexican meal', 'Italian meal', 'Asian meal', and 'ready-to-eat' meal, respectively. Motivations 8 and 27 correspond to 'protein-rich' and 'vegan/vegetarian' diets or lifestyles, respectively. Motivation 20 and 23 show necessities for 'housekeeping' and 'baking' activities, while motivation 13 contains holiday products related to 'Easter'. Motivations 4 and 29 include 'appetisers' and products for 'coffee/tea'. Additionally, motivations 10 and 26 are related to 'cat care' and 'dog care'. Motivation 7 shows a motivation related to higher-priced luxurious products, indicating a 'high budget' motivation. Moreover, various other motivations are identified, such as motivations related to households with a 'baby' (motivation 24), a 'fruits /vegetables' motivation (motivation 17), as well as more unhealthy motivations such as 'energy/soft drinks' (motivation 16) and 'sweet snacking' (motivation 5). Figure 5.1 shows the percentage of shopping trips in which each of the top ten most occurring motivations is most likely. For all motivations, this percentage ranges from 7.12% to 0.66%. This distribution highlights that some motivations are activated more than others, though all motivations appear relevant.

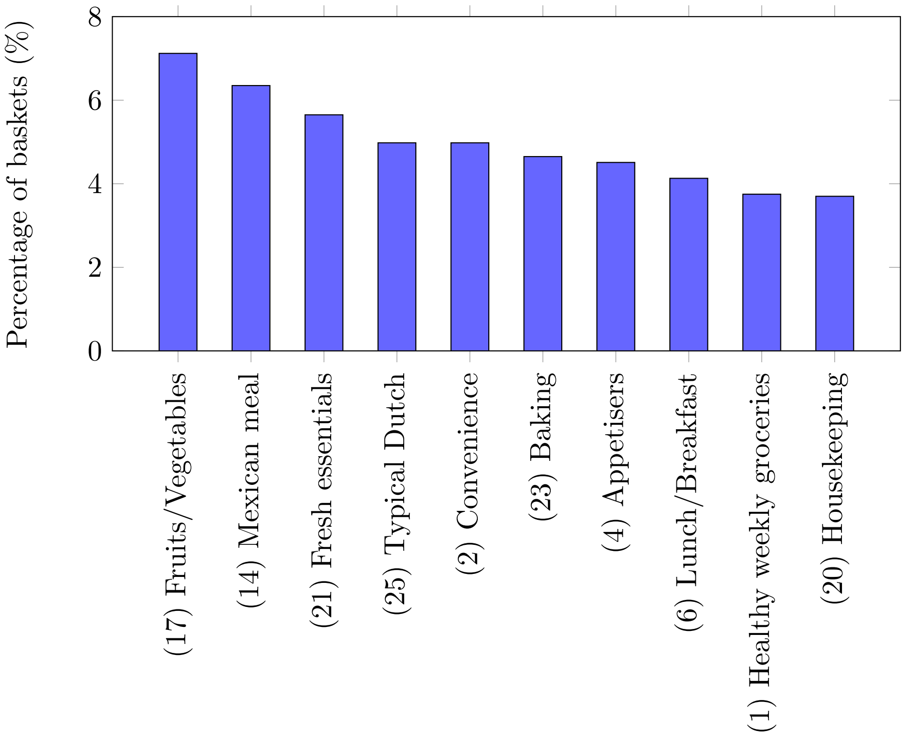


Figure 5.1: Top 10 Most Occurring Motivations with Basket Percentage Distribution

The majority of these motivations encompass products from different product categories. For example, motivation 18 contains products such as meat, cheese, vegetables, sauce, seasoning, pasta, and wine, all related to an 'Italian Meal'. Such motivations emphasise the importance of identifying customers' overarching motivations, as recognising frequently purchased combinations of different types of products can have significant commercial implications. Additionally, it is interesting to observe that certain products fulfil multiple motivations, such as milk associated with 'lunch/breakfast' and 'Easter' motivations, as well as the 'typical Dutch' motivation and 'baking' motivation.

The CTM-30 model captures broad motivations as well as more specific motivations. Some motivations, such as 'cat necessities' and 'energy/soft drinks' are highly specific, requiring only three products to encompass the majority ($\geq$ 50%) of the product purchases under those motivations. Notable, for the 'energy/soft drinks' motivation, 38% of the probability mass in its $\boldsymbol{\phi}_m$ vector is attributed to energy drinks only.

However, some motivations exhibit a more general nature, such as motivation 21, which requires more than 15 products to cover at least 50% of the purchases under this motivation (refer to Appendix A). The products associated with this motivation also do not have a clear and specific purpose or customer need, making it more challenging to label this motivation. Nevertheless, this motivation appears prevalent across many shopping trips (refer to Figure 5.1). It is the most likely motivation in 4.98% of the shopping trips, surpassing the average occurrence of 3.33%.

The motivation that prevails in most shopping trips (7.12%) is motivation 17, centred around 'fruits/vegetables'. For this motivation, at least 13 fruits and vegetables are required to amount to 50% of the probability mass in this vector. Labelling this motivation is more intuitive since all the products belong to the fruits and vegetables categories.

5.2 Customer Profile

Using the shopping motivations identified in Section 5.1, we can reveal the distinct shopping patterns of individual customers at the Dutch grocery retailer. As an example, we analyse the shopping journey of the customer with ID 29, which occurred between March 1st and April 23rd. We will focus on each shopping trip's three most probable motivations, as depicted in Figure 5.2 below.

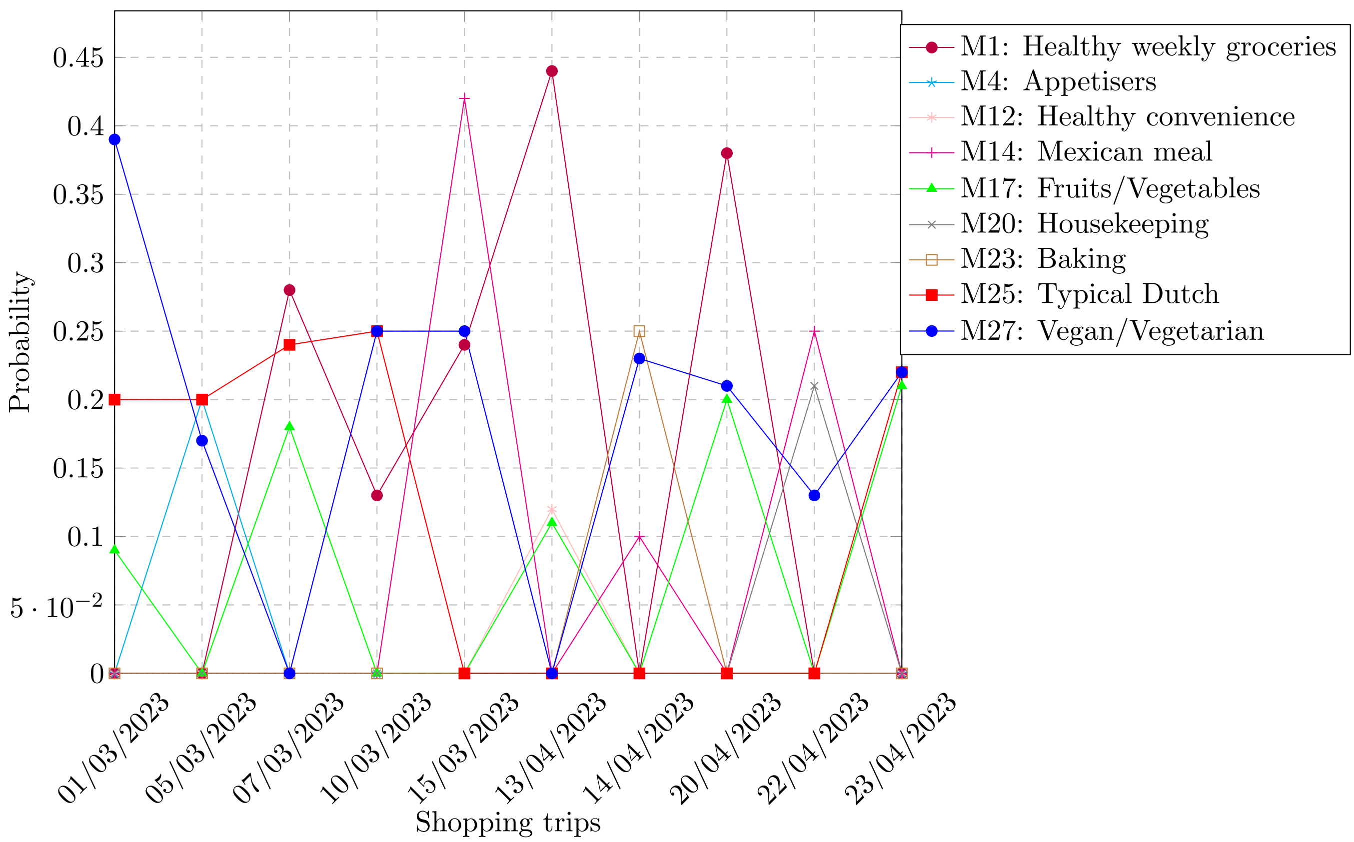


Figure 5.2: Relevant Motivations per Shopping Trip for Customer with ID 29

This analysis reveals that the customer seems to follow a vegan or vegetarian diet, as the motivation of 'vegan/vegetarian' is activated in almost every shopping trip. Additionally, motivations related to 'fruits/vegetables' and 'healthy weekly groceries' show importance in approximately half of the shopping trips. Moreover, some motivations reveal preferences regarding the dinner types of this customer, such as the 'typical Dutch' and 'Mexican meal' motivations.

Furthermore, our approach to modelling motivations at the shopping trip level enables the identification of incidental motivations. For instance, on April 14th, the customer's purchases suggest a baking day, while on April 22nd, we can identify a housekeeping motivation. Evidently, by using a trip-specific modelling strategy, we can understand the general shopping patterns of a customer, as well as more specific and intricate motivations.

5.3 Effects of Customer- and Trip-Specific Variables

The relevance of motivation m in the b-th shopping trip of customer i can be partially attributed to a customer's inherent preferences and characteristics and partially to contextual factors and marketing instruments. By modelling θ_{ibm} using the latent α_{ibm} as described in Chapter 4.2, we can capture the effects of customer characteristics through γ_m and the impact of contextual factors through β_m.

To measure the effect of these variables on the relevance of each motivation, we need to calculate the odds ratios as described in Chapter 4.5. The baseline for these odds ratios is the average shopping trip across all customers in the sample. The resulting odds ratios are summarised in Appendix D.

5.3.1 Effects of Customers' Age and Location

Our model included available information on customers' age and location, represented by age group dummies and a binary variable indicating whether a customer lives in a city (with a population of at least 200,000). Because of missing age data, a dedicated dummy variable indicates unknown age values. We take the levels 'Age: 45-55' and 'Location: No city' as baseline levels. The relatively large odds ratios in Appendix D indicate the significant impact of these customer demographics on the relevance of specific motivations.

When examining the effect of age on motivations, differences among age groups are evident. For instance, customers aged between 0 and 25 years and between 25 and 35 years show a 430% and 176% increase in the likelihood of having an 'energy/soft drinks' motivation (odds ratios 4.30 and 1.76). In contrast, customers in the age group 55-65 and 65+ demonstrate a 61% and 70% decrease in likelihood of this motivation, respectively. The youngest customers, between 0-25 years old, are less likely to be motivated by 'housekeeping' (0.43) and 'beer/self-care' (0.52). Customers between 25 and 35 years old are especially likelier to have a 'vegan/vegetarian' motivation and a 'fruits/vegetables' motivation (1.73 and 1.45). Customers aged 35-45 are also 52% more likely to be motivated by the 'fruits/vegetables' motivation, but 58% less likely to be motivated by a 'high budget'. The 'high budget' motivation is most likely for customers between

45-55 years old (1.51), who are also most interested in Asian and Italian meals (1.39 and 1.32). Customers between 55-65 are more likely to be motivated by the 'beer/ready-to-eat' motivation and 'appetisers' (1.38 and 1.33). Additionally, findings show that the 'coffee/tea' motivation and 'baking' motivation become 227% and 29% more likely for customers aged 65 and above, respectively. On the other hand, the 'high budget' motivation becomes 59% less likely for these customers. Notably, the relatively small odds ratios for the unknown age group indicate that customers who do not disclose their age do not differ much from those who do.

We also observe specific differences regarding the impact of a customer's location. Notably, when customers live in a city, the likelihood of being motivated by 'convenience' diminishes by 35%, and the motivation for 'typical Dutch' food becomes 28% less likely. In contrast, the probabilities of being motivated by 'housekeeping' and 'high budget' increase by 46% and 36%, respectively.

Our model setup also allows us to identify the demographic profile most inclined to shop with each specific motivation. For instance, considering the 'vegan/vegetarian' lifestyle motivation, the odds ratios reveal that this motivation is more likely to be relevant for young customers in the age groups of 25-35 (odds ratio 1.73) and 35-45 (1.39) while being less likely for older customers in the age groups of 55-65 (odds ratio 0.87) and 65+ (0.53) and young customers between 0-25 years (0.76). Furthermore, this motivation becomes significantly more likely for customers residing in larger cities (1.32). Figure 5.3 summarises the demographic profile corresponding to the 'vegan/vegetarian' motivation.

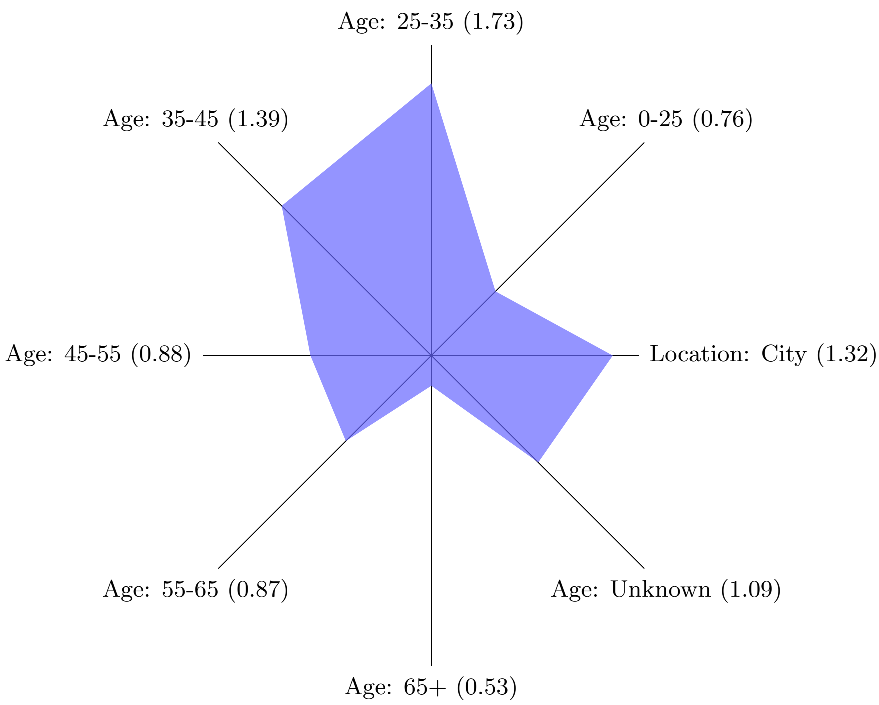


Figure 5.3: Customer Profile Corresponding to Vegan/Vegetarian Motivation, Using Odds Ratios

In contrast, the ‘typical Dutch’ motivation appeals to a different customer segment. This motivation, centred around traditional Dutch ingredients and products, becomes more likely for older customers in the age groups of 55-65 (odds ratio 1.246) and 65+ (1.47) while being less likely for younger customers aged 0-25, 25-35 and 35-45 (0.71, 0.81 and 0.84). Additionally, unlike the ‘vegan/vegetarian’ motivation, this motivation becomes 29% less likely for customers living in cities.

5.3.2 Effects of Timing of Shopping Trip

Our model also considers contextual factors of shopping trips. These are represented by dummy variables for the time of day (before or after 5 p.m.) and the day of the week (weekday or weekend) of the shopping trip. Within this specification, ‘Time: Before 5 p.m.’ and ‘Day: Weekday’ serve as the baseline levels for the respective dummy variables.

Our analysis reveals that both variables significantly influence the relevance of specific motivations. After 5 p.m., motivations related to breakfast and lunch, such as ‘sweet breakfast’ and ‘fresh buns’, become 61% and 54% less likely, respectively. However, motivations associated with unhealthy choices, such as ‘energy/soft drinks’ and ‘sweet snacking’, become 36% and 27% more likely, respectively. Additionally, most motivations centred around meals, such as the ‘beer/ready-to-eat’ motivation, become more likely after 5 p.m.

During weekends, specific motivations also become more relevant. The ‘fresh buns’, ‘appetisers’ and ‘convenience’ motivations become 30%, 23%, and 14% more likely, respectively. On the other hand, motivations such as ‘healthy convenience’, ‘Asian meal’, and ‘typical Dutch’ become less likely during the weekends by 17%, 18%, and 10%, respectively. Furthermore, we can determine when specific meals are most likely. The ‘beer/ready-to-eat’ and ‘Mexican meal’ motivations appear more likely during weekends, with 20% and 9%, respectively. In contrast, the ‘typical Dutch’ and ‘Asian meal’ motivations seem more likely during weekdays, with 10% and 16%. Interestingly, the Italian motivation appears to have similar likelihoods during weekends and weekdays, with an odds ratio of 0.98.

5.3.3 Effects of Discounts

To extend the existing literature on topic models for understanding customer behaviour, we have introduced a feature in our model regarding each basket’s discount percentage. The primary objective of this feature is to provide a quantitative indicator for distinguishing between intrinsic motivations and motivations particularly stimulated by discounts.

Analysing the odds ratios, we find that the motivation for ‘beer/self-care’ experiences the most notable increase in likelihood in discounted baskets (odds ratio 1.27). Additionally, several other motivations appear more likely when discounts are involved: ‘cat care’ increases in likelihood by 16%, ‘housekeeping’ by 17% and ‘vegan/vegetarian’ by 11%.

We can also identify motivations that maintain consistent likelihoods in discounted and non-discounted baskets. Motivations like ‘fruits/vegetables’, ‘convenience’, and ‘sweet snacking’ seem less sensitive to discounts (odds ratio 1.00). The same holds for the 'typical Dutch', 'Asian meal' and 'healthy weekly groceries' motivations (1.01, 0.98 and 0.99).

On the other hand, some motivations become less likely in discounted baskets, including the 'Mexican meal' motivation by 13%, 'energy/soft drinks' motivation by 11%, and 'baking' motivation by 12%. It is important to note that, however, we cannot conclude that this implies that customers shift toward motivations other than these when discounts are offered, as we do not have information on promotions of products that are not included in the customer's basket.

5.4 Relations Between Motivations

In terms of exploring the relationships between motivations, our correlated topic model, using the logistic normal distribution, enables us to capture the covariance structure among motivations through $\mathbf{\Sigma}_k$ (refer to Section 4.2).

We visualise the resulting correlation matrix in Appendix C. Within this correlation matrix, notable correlations between specific motivations are evident. Approximately 5% of the pairwise correlations are either above 0.5 or below −0.5, indicating significant associations. These specific 5% of correlations are particularly interesting and valuable for further investigation.

To illustrate, we consider a specific motivation and explore its strongest correlations with other motivations. Motivation 16, related to energy and soft drinks, exhibits noteworthy positive and negative correlations. These correlations are summarised in Table 5.2.

Table 5.2: Strongest Correlations with Motivation (16): Energy/Soft Drinks

Motivation	Correlation
(5) Sweet snacking	+0.61
(11) Savoury snacking	+0.55
(2) Convenience	+0.52
(9) Sweet breakfast	+0.47
(10) Cat care	+0.46
(21) Fresh essentials	−0.70
(17) Fruits/Vegetables	−0.57
(14) Mexican meal	−0.53
(27) Vegan/Vegetarian	−0.47
(8) Protein-rich	−0.41

Interestingly, the 'energy/soft drinks' motivation demonstrates significant positive correlations with other motivations associated with unhealthy choices. For instance, it exhibits a positive correlation with the 'sweet snacking' motivation (+0.61), the 'savoury snacking' motivation (+0.55) and the 'sweet breakfast' motivation (+0.47). An interesting finding is that the 'energy/soft drinks' motivation also positively correlates with the 'cat care' motivation (+0.46).

In line with intuition, these seemingly unhealthy motivations negatively correlate with more health-conscious motivations. For example, the ‘energy/soft drinks’ motivation exhibits a significant negative association with the ‘fruits/vegetables’ motivation (-0.57), the ‘vegan/vegetarian’ motivation (-0.47) and the ‘healthy breakfast’ motivation (-0.41). Also, more surprisingly, the ‘Mexican meal’ motivation appears to be negatively correlated with this motivation (-0.53).

Overall, we can observe noteworthy positive correlations among all ‘healthy’ motivations and all ‘unhealthy’ motivations. These relationships are depicted in Figure 5.4. Significant negative correlations were found between ‘healthy’ and ‘unhealthy’ motivations, such as between the ‘savoury snacking’ motivation and the ‘fruits/vegetables’ (-0.63) or ‘vegan/vegetarian’ motivation (-0.61).

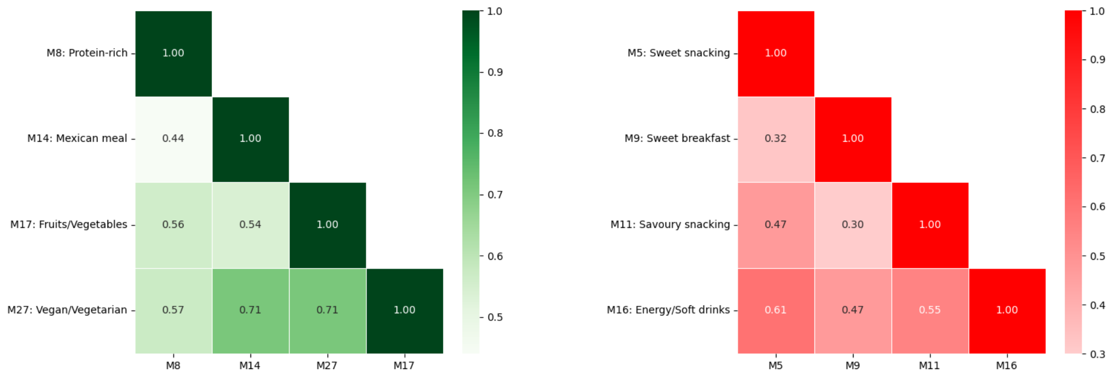


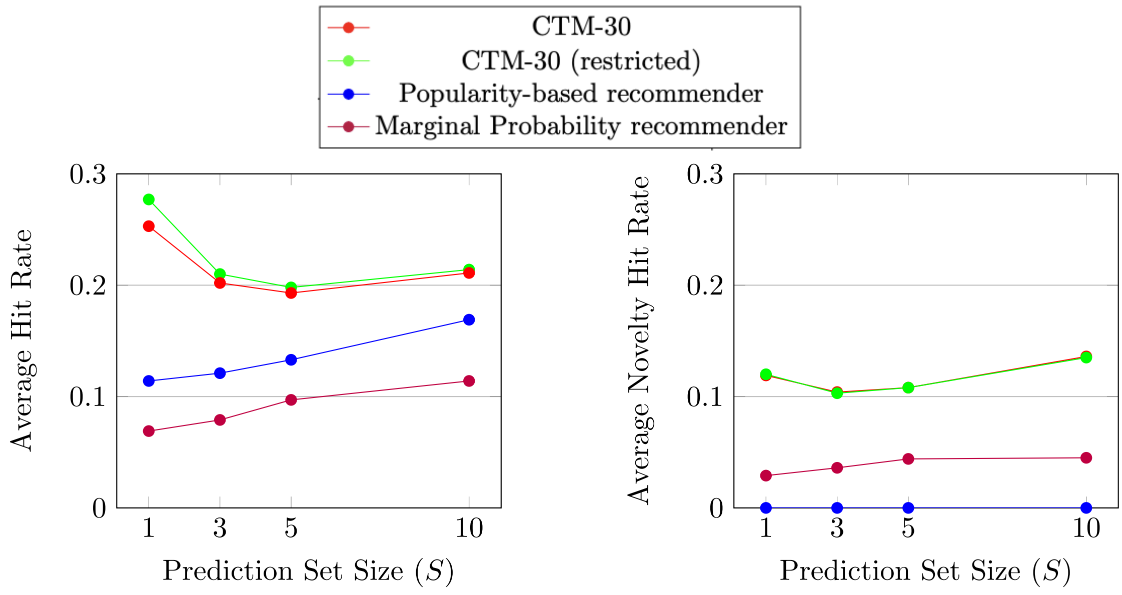
Figure 5.4: Correlations Between Seemingly Healthy and Unhealthy Motivations

These findings suggest that customers exhibit consistent shopping behaviour during shopping trips, favouring various health-conscious options or indulging in less healthy products. The negative correlations indicate a trade-off between those two groups of motivations during shopping trips, indicating that customers driven by specific health-conscious motivations tend to avoid unhealthy products and vice versa.

5.5 Product Recommendations

We use a systematic approach to evaluate the performance of our CTM-30 model-generated product recommendations, drawing ten distinct random samples following the procedure detailed in Chapter 3. This approach ensures we can reliably assess the average (novelty) hit rates, as outlined in Chapter 4.7. Each dataset undergoes division into training and test sets, with each customer’s last shopping trip allocated to the test set. Our prediction sets encompass the recommendations generated for individual customers, with varying sizes, including 1, 3, 5, and 10 products. These prediction sets are evaluated against the purchase data from our designated test sets to calculate the average (novelty) hit rates. The results are presented in Appendix E.

To provide a benchmark for comparison, we contrast the predictive performance of our model with three alternative benchmark models. Firstly, we consider a restricted CTM-30 model that excludes all customer-specific effects. The second benchmark involves a popularity-based recommender recommending each customer's most commonly purchased products. Given the repetitive nature of grocery shopping, we expect this approach to consistently achieve reasonable performance, reflecting customers' tendency to buy familiar products. The third benchmark uses marginal probabilities. Here, a product's likelihood is determined by its relative purchase frequencies across the entire dataset, offering a general view of product popularity without examining individual customer heterogeneity.



(a) Average Hit Rate Comparison

(b) Average Novelty Hit Rate Comparison

Figure 5.5: Comparison of Average Hit Rate and Novelty Hit Rate for Different Recommenders

Upon analysing the average hit rates depicted in Figure 5.5a, it becomes evident that the CTM-30-model, leveraging the uncovered motivations, can generate predictions surpassing the marginal probability recommender and the popularity-based recommender. Particularly noteworthy is the CTM-30 model's ability to excel in single-product recommendations. For about 28% of the customers, our model successfully suggests a product that was actually purchased during their last shopping trip. While the popularity-based recommender consistently demonstrates respectable performance, as anticipated, it faces a constraint in its inability to generate novel recommendations - products the customer has not previously purchased. This shortcoming is highlighted by evaluating the (average) novelty hit rate in Figure 5.5b.

This figure demonstrates that the CTM-30 model effectively generates innovative recommendations. Impressively, over 10% of CTM-30's recommended products are novel and accurate suggestions. These are product recommendations that customers actually purchase for the first time during their hold-out transactions. Surprisingly, in both evaluated scenarios, the restricted CTM-30 model, excluding customer-specific effects, performs similarly to the CTM-30 model while demonstrating faster computational speed. This observation suggests that customer-specific effects may not significantly enhance product recommendations.

To further investigate the CTM-30 model's predictive strength, we examine its performance when generating predictions for new customers who have only visited the store once before, resulting in limited purchase data. The outcomes, visually presented in Figure 5.6, indicate that the full CTM-30 model outperforms the restricted CTM-30 model in this context. Thus, including customer-specific effects in generating recommendations enhances the predictive capabilities of the model when interacting with new customers. The results of the CTM-30 model approach the performance of the popularity-based recommender while generating novel predictions as well as opposed to the popularity-based approach. Again, the marginal probability recommender is outperformed by the CTM-30 models, especially when recommending only one to five products. Compared to existing customers, the lower average hit rates for new customers highlight that existing customers' purchase histories provide valuable information for predicting future purchases.

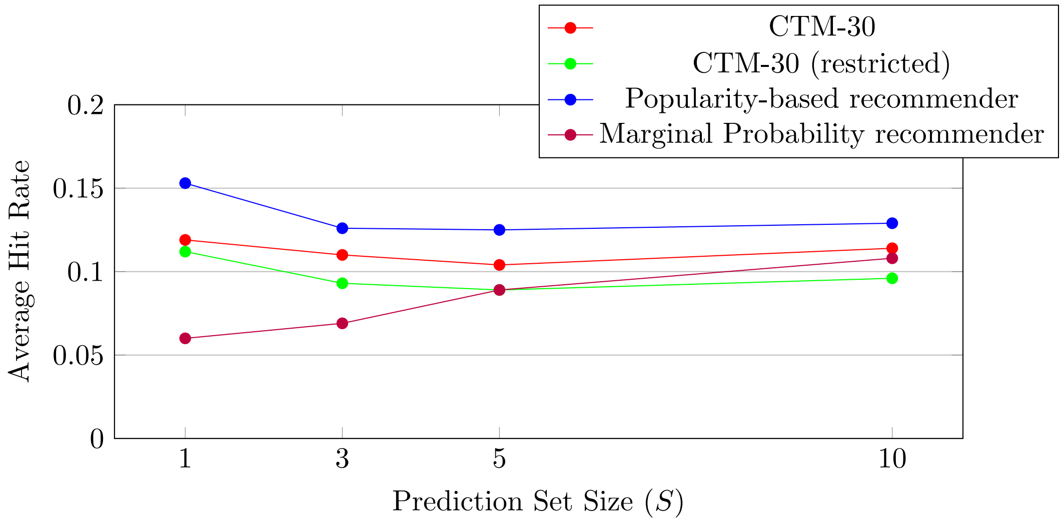


Figure 5.6: Average Hit Rate Comparison for New Customers

In summary, we thoroughly analysed the results produced by our CTM-30 model using data from a prominent Dutch grocery retailer. The analysis has provided valuable insights into customers' shopping motivations, the impact of customer- and trip-specific variables, the relationships between these motivations, and the effectiveness of our model in delivering product recommendations. In the next chapter, we will provide an in-depth discussion of these findings, highlighting the implications and significance within the broader context of our research.

Chapter 6

Discussion

6.1 Conclusion

In this research, we have leveraged the power of an efficiently estimated correlated topic model with variational inference to unveil the underlying motivations driving customers' grocery shopping behaviours. Our approach, as hypothesised, empowers retailers to gain a deeper understanding of their customers, enhance segmentation strategies, and generate personalised product recommendations that can enhance customer satisfaction and loyalty.

Our study analysed a dataset containing over 100,000 transactions from a prominent Dutch grocery retailer. We discovered that customer behaviour can be effectively summarised using a concise set of 30 purchasing motivations. This specific set of motivations was selected for its superior interpretability and assessed based on the similarity and coherence of these motivations. Consistent with the findings of Hornsby et al. (2020), our research revealed that grocery shoppers have thematic and goal-directed motivations, including lifestyles, meal preferences, dietary choices, activities, pet care, budgets, household circumstances, and more. Additionally, our findings align with the research by Jacobs et al. (2021), showcasing that motivations can identify highly specific customer behaviours within distinct subgroups of customers.

Each motivation is represented by a distribution across the entire product assortment, transcending traditional product categories. Understanding how products from different categories align with these motivations offers promising opportunities for cross-selling strategies, assortment optimisation, and web and store design improvements. Our findings emphasise that individuals associate grocery products with specific goals and themes, whereas supermarkets typically organise products taxonomically. Consequently, our research suggests supermarkets should consider revising their store layout to harmonise with customers' goals and needs.

By modelling customer motivations at the shopping trip level, we can effectively monitor how customers' motivations evolve. This capability enables retailers to tailor their communication and targeting strategies to align with each customer's current motivations. This personalised approach can potentially enhance the effectiveness and engagement of communication efforts by addressing the underlying motivations rather than superficial factors.

Moreover, our research explored the influence of customer demographics on their motivations. Our results uncovered distinct patterns among customers of varying age groups. Additionally, in contrast to previous studies, we found differences in motivations between customers living in cities and customers living outside of cities. Specifically, city customers appear to be more motivated by housekeeping and high budgets while showing less motivation for typical Dutch food and convenience. These insights have significant implications for targeted advertising efforts and offer valuable guidance to suppliers of retailers seeking to refine their products, tailor marketing strategies and introduce new products that align with the preferences of different customer segments.

Moreover, in agreement with Jacobs et al. (2021) and Hornsby et al. (2020), our research highlights that customer motivations can vary depending on the context of each shopping trip, influenced by the day of the week and the time of day. These findings highlight the potential for implementing real-time promotions and dynamic marketing strategies that adapt to these contextual variations.

In addition to these insights, our research has enriched the current body of customer motivation theory by introducing a novel distinction between intrinsic and promotion-related motivations. We have identified motivations discounts can particularly stimulate, with the 'vegan/vegetarian' motivation being a notable example. Building upon the findings by Carrasco et al. (2022) regarding reducing alcohol, fat, salt and sugar-related motivation consumption, our research offers a promising avenue for promoting sustainable consumption practices.

The correlated topic model provides a significant advantage in uncovering the connections between various motivations. Our specific model has successfully revealed that correlations between particular motivations are significant, aligning with the findings of Jacobs et al. (2021). In particular, we observed strong correlations among healthy motivations and those classified as more unhealthy. The identified correlations hold meaningful implications, especially in cross-selling and search optimisation. Retailers can leverage these correlations to suggest or display products that are positively related to each customer's underlying motivations.

Furthermore, in validating the final part of our hypothesis, our model has demonstrated its ability to efficiently generate product recommendations at scale. It consistently outperformed popularity-based recommendation systems and proved its effectiveness in providing predictions for new customers. Additionally, in line with the findings of Jacobs et al. (2016), our model excels in suggesting single products, including innovative recommendations with lower sales volumes situated at the tail end of the product assortment. Moreover, our model's recommendations are inherently intuitive and explainable, relying on customer motivations as guiding principles rather than popular product categories.

6.2 Limitations and Further Research

Our approach to topic modelling introduces a degree of subjectivity in determining the optimal number of motivations and assigning labels to each motivation. While our results demonstrate high face validity and are supported by findings of other studies, further measurements can be employed to enhance the validity of motivations. One potential approach involves conducting experiments in which respondents are tasked with labelling motivations or evaluating their quality, providing a means for rigorous testing and validating the motivations.

Furthermore, it is important to acknowledge that our sample exclusively consists of loyalty card customers. This approach may introduce a selection bias, as these customers' motivations might differ from those who do not engage in such loyalty programs. Additionally, our analytical timeframe is confined to a 2-month period, which could limit our ability to capture variations in motivations related to seasonal trends. Moreover, we have not incorporated online transaction data, which may reveal different customer motivations. To enhance our findings' generalisability and applicability, we should explore larger samples, including data from other grocery retailers and online purchases.

Moreover, during the iterative training of the correlated topic model, motivations exhibit variability by appearing and disappearing in different posterior samples, indicating uncertainty. To enhance the robustness of our findings, we could have employed a (hierarchical) clustering approach to cluster motivations from different samples. This method would have enabled us to identify recurring motivations that consistently drive the customer base, ensuring the reliability of our results.

Additionally, our analysis does not consider household sizes, as we do not know whether multiple individuals use the same loyalty cards for shopping. Investigating how household dynamics influence shopping motivations and behaviours in future research endeavours would be beneficial. Subsequent studies could also incorporate supplementary contextual data, such as product pricing, product attributes such as nutritional information or various product labels, and online web and app usage data. Additionally, the exploration of real-time recommendation systems that adapt to the specific context of each shopping trip represents an exciting avenue for further research.

Lastly, retailers can use our research to identify customers motivated towards healthier and more sustainable dietary choices. An intriguing avenue for further research involves incorporating discount information on products not included in a customer's shopping basket. This approach can provide a deeper insight into how customers switch between different motivations in response to discounts. As a result, retailers can develop effective promotion strategies to encourage more healthy and sustainable purchasing motivations among their customers.

References

Aguilar-Palacios, C., Muñoz-Romero, S. & Rojo-Álvarez, J. L. (2021). Causal quantification of cannibalization during promotional sales in grocery retail. *IEEE Access*, *9*, 34078-34089. doi: 10.1109/ACCESS.2021.3062222

Aletras, N. & Stevenson, M. (2014, 04). Measuring the similarity between automatically generated topics.. doi: 10.3115/v1/E14-4005

Bishop, C. M. (2006). Pattern recognition and machine learning. In (chap. 10). Springer-Verlag.

Blair, S. J., Bi, Y. & Mulvenna, M. D. (2020, 1). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, *50*, 138-156. doi: 10.1007/s10489-019-01438-z

Blanchard, S. J., Aloise, D. & Desarbo, W. S. (2017, 6). Extracting summary piles from sorting task data. *Journal of Marketing Research*, *54*, 398-414. doi: 10.1509/jmr.15.0388

Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. (2017, 4). *Variational inference: A review for statisticians* (Vol. 112). American Statistical Association. doi: 10.1080/01621459.2017.1285773

Blei, D. M. & Lafferty, J. D. (2007, 8). A correlated topic model of science. *The Annals of Applied Statistics*, *1*. doi: 10.1214/07-aoas114

Blei, D. M., Ng, A. Y. & Edu, J. B. (2003). *Latent dirichlet allocation michael i. jordan* (Vol. 3).

Boyd, S. & Vandenberghe, L. (2004). *Convex optimization.* Cambridge University Press.

Carrasco, M. V., Manolopoulou, I., O'Sullivan, J., Prior, R. & Musolesi, M. (2022, 6). Posterior summaries of grocery retail topic models: Evaluation, interpretability and credibility. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *71*, 562-588. doi: 10.1111/rssc.12546

Christidis, K., Apostolou, D. & Mentzas, G. (2010). Exploring customer preferences with probabilistic topics models. In (p. 12-24).

Chuang, J., Manning, C. D. & Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. In (p. 74-77). doi: 10.1145/2254556.2254572

Gharahighehi, A., Pliakos, K. & Vens, C. (2022). Addressing the cold-start problem in collaborative filtering through positive-unlabeled learning and multi-target prediction. *IEEE Access*, *10*, 117189-117198. doi: 10.1109/ACCESS.2022.3219071

Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5228-5235. doi: 10.1073/pnas.0307752101

Hornsby, A. N., Evans, T., Riefer, P. S., Prior, R. & Love, B. C. (2020, 6). Conceptual organization is revealed by consumer activity patterns. *Computational Brain and Behavior*, *3*, 162-173. doi: 10.1007/s42113-019-00064-9

Jacobs, B., Donkers, B. & Fok, D. (2016, 5). Model-based purchase predictions for large assortments. *Marketing Science*, *35*, 389-404. doi: 10.1287/mksc.2016.0985

Jacobs, B., Fok, D. & Donkers, B. (2021, 9). Understanding large-scale dynamic purchase behavior. *Marketing Science*, *40*, 844-870. doi: 10.1287/mksc.2020.1279

Joorabloo, N., Jalili, M. & Ren, Y. (2020, 01). Improved collaborative filtering recommendation through similarity prediction. *IEEE Access*, *8*, 202122-202132. doi: 10.1109/ACCESS.2020.3035703

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*, 183–233. doi: 10.1007/978-94-011-5014-9_5

Lau, J. H., Google, D. N. & Baldwin, T. (2014). *Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality.*

Pearl, J. (1988). Chapter 3 - markov and bayesian networks: Two graphical representations of probabilistic knowledge. In J. Pearl (Ed.), *Probabilistic reasoning in intelligent systems* (p. 77-141). San Francisco (CA): Morgan Kaufmann. doi: https://doi.org/10.1016/B978-0-08-051489-5.50009-6

Reisenbichler, M. & Reutterer, T. (2019, 4). Topic modeling in marketing: recent advances and research opportunities. *Journal of Business Economics*, *89*, 327-356. doi: 10.1007/s11573-018-0915-7

Statista. (2022). *Supermarket visit frequency in the netherlands 2022.* Retrieved 2023-06-20, from `https://www.statista.com/statistics/886964/supermarket-visit-frequency`

Stevens, K., Kegelmeyer, P., Andrzejewski, D. & Buttler, D. (2012, 07). Exploring topic coherence over many models and many topics..

Tyrała, S., Orwat, A. & Makowski, (2022). Trends and sales models in e-commerce: Examples of best practices. *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu*, *96*(1), 89–105. doi: 10.5604/01.3001.0016.0361

Xing, L., Paul, M. J. & Carenini, G. (2019). Evaluating topic quality with posterior variability. *ArXiv*, *abs/1909.03524*.

Appendix A

A.1 Motivations (CTM-30)

(1) Healthy weekly groceries

Product	P
Bananas	0.168
Multigrain bread	0.067
Fresh mandarins	0.060
Whole wheat bread	0.048
Fresh hamburger	0.042
Peanut butter	0.039
Fresh red grapes	0.039
Fresh white grapes	0.037
Semi-skimmed yogurt	0.034
Sliced mature cheese	0.031
Currant buns	0.031
Full-fat yogurt	0.028
Sweet pears	0.024
Breakfast crackers	0.024
Fresh potato wedges	0.023

(2) Convenience

Product	P
Buns (fresh)	0.118
Shaped Chips	0.055
Various Chips	0.046
Pizza	0.041
Whole Wheat Bread	0.035
Fruit Drink	0.032
Cola	0.030
French Fries	0.028
Free-Range Eggs	0.024
Tortilla Chips	0.023
Orange Soda	0.022
Frozen Bread	0.022
Young Cheese	0.019
Young Gouda	0.018
Ham	0.018

(3) Beer / Self-care

Product	P
Pilsner beer (can)	0.347
Dishwasher tablets	0.051
Men's deodorant	0.042
Pilsner beer (crate)	0.041
Women's deodorant	0.033
Sangiovese (Italy)	0.027
Toothpaste strong	0.027
Strong filter coffee	0.026
Pinot Grigio (Italy)	0.025
Diverse	0.023
Shower gel all skin	0.023
Shampoo all hair	0.021
Toothpaste white	0.021
Verdejo	0.020
Toothpaste fresh	0.019

(4) Appetisers

Product	P
Herbed cream cheese	0.076
Egg salads (spread)	0.048
Chicken salad (spread)	0.039
Fresh olives	0.036
White bread (oven)	0.033
Toasts	0.033
Cheese cubes	0.033
Salami, fuet	0.032
Meat appetisers	0.032
Meat aperitif snacks	0.031
Low-fat chips	0.029
Party peanuts	0.028
Fish salads (spread)	0.027
Pastry	0.027
Luxury toasts	0.023

(5) Sweet snacking

Product	P
Chocolate cookies	0.087
Paprika chips	0.073
Plain chips	0.053
Milk chocolate	0.052
Wine gums	0.049
Milk chocolate nuts	0.048
Chocolate candies	0.037
Assorted candy mix	0.034
Savory crackers	0.033
Regular cola cans	0.031
Sweet licorice (soft)	0.030
Popcorn	0.026
Flavored pretzels	0.025
Flavored chocolate	0.023
Mixed licorice	0.023

(6) Lunch / Breakfast

Product	P
Yoghurt drinks	0.0607
Brown bread	0.0602
Baked white rolls	0.0526
Cream cheese	0.0523
White bread	0.0487
Noodle soup	0.0407
Margarine spread	0.0350
Cucumber (fresh)	0.0309
Chocolate milk	0.0280
Toast crackers	0.0267
Sandwich sausage	0.0257
Everyday cookies	0.0229
Fruit biscuits	0.0209
Bapao buns	0.0194
Lemonade syrup	0.0162

(7) High budget

Product	P
Skyr - Icelandic yogurt	0.069
Luxurious chips	0.066
Assorted nuts	0.058
Fresh beef steak	0.054
Fresh shrimps	0.047
Cashew nuts (packet)	0.044
Nespresso (strong)	0.044
Fresh pizza	0.043
Herbal tea	0.040
Sauvignon Blanc	0.037
Nespresso (regular)	0.036
Moist toilet paper	0.035
Paracetamol	0.028
Crispbread	0.026
Granola bars	0.019

(8) Protein-rich

Product	P
Plain Greek yogurt	0.107
Low-fat quark	0.089
Chicken fillet	0.085
Corn cakes	0.064
Tuna in water	0.055
Cottage cheese	0.043
Flavored water bottle	0.037
Cherry tomatoes	0.035
Oats (regular)	0.033
Cucumber (snack)	0.032
Rice cakes	0.029
Laundry detergent	0.022
Bananas	0.019
Toast crackers	0.018
Light cream cheese	0.015

(9) Sweet breakfast

Product	P
Packaged bread	0.618
Fresh croissant	0.133
Chocolate bread rolls	0.079
Muffin	0.055
Focaccia	0.046
Goat's milk	0.009
Lactose-free fat milk	0.019
Iced tea	0.007
Goat's yogurt	0.006
White bread (fresh)	0.003
Yogurt snacks	0.003
Dairy snacks	0.003
Dates (assorted)	0.002
Baby wipes	0.002
Dates	0.002

(10) Cat care

Product	P
Cat food single pack	0.153
Cat snacks	0.141
Coffee pastry	0.138
Cat food multipack	0.119
Dry cat food	0.064
Premium ice cream	0.049
Vitamin water	0.048
Fruit-flavored drink	0.044
Concentrated softeners	0.040
Iced tea	0.030
Cat litter	0.027
Syrah, Shiraz	0.020
Chardonnay	0.011
Multi-grain buns	0.010
Cheese sauce	0.007

(11) Savoury snacking

Product	P
Frikandelbroodje	0.177
Hot dog sausages	0.105
Sausage roll	0.105
Regular coffee pads	0.061
Homemade soup	0.038
Berliner bol	0.037
Sanitary pads (day)	0.036
Canned pineapple	0.026
Sausage roll	0.022
Ice cream cones	0.022
Fresh strawberries	0.021
Frozen burgers	0.021
Yogurt, quark (1 pers.)	0.021
Lactose-free milk	0.019
Cheese rolls	0.018

(12) Healthy convenience

Product	P
Fruit Salad	0.082
Meal Salad	0.075
Young Cheese	0.061
Mango Pieces	0.058
Small Salads	0.052
Nut Bars	0.050
Yoghurt drinks	0.049
Pancakes	0.047
Steam Meals	0.040
Caffeine-Free Cola	0.036
Health Yoghurt	0.032
Dairy Snacks	0.030
Muesli Rolls	0.026
Yoghurt Drinks	0.025
Flavored Quark	0.020

(13) Easter / Water

Product	P
Long-life whole milk	0.205
Carbonated water	0.088
Still bottled water	0.076
Tonic water	0.055
Fresh tompoucen	0.050
Small bottled water	0.043
Milk Easter eggs	0.040
Almond pastries	0.039
Filled Easter eggs	0.038
Mixed Easter eggs	0.036
Oven bitterballen	0.031
Yogurt with muesli	0.027
Seasonal chocolates	0.020
Lasagna ready meals	0.018
From 12 months purees	0.017

(14) Mexican meal

Product	P
Chicken fillet - piece	0.058
Mozzarrella	0.042
Mushrooms (fresh)	0.040
Tortilla, wrap, burrito	0.036
Creme fraiche	0.034
Cherry tomatoes (fresh)	0.033
Tomato paste	0.030
Grated cheese	0.029
Fresh roma tomatoes	0.027
Yellow onions	0.026
Rocket	0.025
Garlic (fresh)	0.025
Peppers mix (fresh)	0.021
Pepper (fresh)	0.019
Red onions	0.018

(15) Beer / Ready-to-eat

Product	P
Pilsener - bottle	0.110
Asian ready-to-eat	0.086
Blond beer	0.061
Tripel beer	0.053
Chocolate milk (packs)	0.047
Italian ready-to-eat	0.046
Alcohol-free beer	0.042
Raw vegetable salad	0.040
Dark chocolate	0.029
World ready-to-eat	0.029
Coffee creamers	0.026
Instant coffee (regular)	0.024
Vegetable mix (canned)	0.024
India Pale Ale (IPA)	0.024
Wheat beer	0.023

(16) Energy / Soft drinks

Product	P
Energy drinks	0.379
Bottled water	0.087
Diet cola cans	0.068
Candy bars multipack	0.054
Sports drink bottles	0.047
Diet cola bottles	0.034
Orange soda (cans)	0.030
Ice cubes	0.027
Ice tea (cans)	0.017
Yogurt snacks	0.016
Surinamese soft drinks	0.012
Anchovies (canned)	0.012
Rice cakes	0.012
Egg cookies	0.010
Lemon, lime beverages	0.008

(17) Fruits / Vegetables

Product	P
Fresh cucumber	0.092
Avocado	0.060
Fresh zucchini	0.059
Fresh strawberries	0.041
Free-range eggs	0.039
Fresh blueberries	0.037
Red bell pepper	0.035
Fresh pasta	0.033
Fresh spinach	0.031
Fresh broccoli	0.028
Fresh lemons	0.024
Mixed salad greens	0.023
Carrots	0.020
Fresh eggplant	0.020
Fresh chicken fillet	0.018

(18) Italian meal

Product	P
Ground beef	0.151
Grated cheese	0.100
Stir-fry mix (fresh)	0.070
Minced meat	0.064
Pasta sauce	0.055
Italian seasoning mix	0.048
Sieved tomatoes	0.038
Cooking cream	0.028
Vegetable mix	0.027
Spaghetti	0.027
Ready-to-bake bread	0.023
Macaroni	0.022
Tomato paste	0.022
Italian meal package	0.017
Montepulciano (Italy)	0.016

(19) Refreshments

Product	P
Iced coffee	0.162
Flavored Quark (large)	0.115
Multifruit Juice	0.110
Iced tea	0.100
Cup Soup	0.080
Fresh Smoothie	0.057
Full-fat Milk	0.039
Chilled orange juice	0.035
Yoghurt Snacks	0.026
Fish Salads	0.024
Coffee Creamer	0.021
Breaded Fish	0.016
Multigrain Bread	0.016
Cheese rolls	0.012
Festive bread	0.012

(20) Housekeeping

Product	P
Laundry detergent	0.074
Toilet paper	0.070
Kitchen paper towels	0.067
Freshly baked cake	0.063
Dishwashing liquid	0.058
Toilet blocks	0.045
Bleach	0.043
Cabernet Sauvignon	0.040
Whipped cream	0.035
Chardonnay	0.034
Ground beef	0.034
Toothpaste sensitive	0.028
Moist cleaning wipes	0.024
Cleaning wipes	0.023
All-purpose cleaner	0.023

(21) Fresh essentials

Product	P
Buttermilk (fresh)	0.058
Fresh oranges	0.058
Whole leeks (fresh)	0.054
Belgian endive	0.037
Free-range eggs	0.032
Rusk	0.030
Crispbread	0.029
Skimmed yogurt	0.026
Full-fat quark	0.023
Pickles	0.022
Fresh green beans	0.021
Fresh ham	0.020
Large pudding	0.019
York ham (fresh)	0.019
Fresh green beans	0.018

(22) Treats / 1-person

Product	P
Appelflappen	0.133
Salad portion	0.098
Mousse (portion pack)	0.075
Quiches	0.073
Pudding (portion pack)	0.051
Cream Cheese	0.039
Refreshing Fruit Drinks	0.029
Water	0.028
White rolls	0.022
Fabric Softeners	0.022
Fresh Cake	0.021
Green Tea	0.020
Sauvignon Blanc	0.019
Soft Licorice	0.018
Whipped Cream	0.016

(23) Baking

Product	P
Fresh whole milk	0.147
Unsalted butter	0.147
Whipped cream	0.046
Dark chocolate	0.034
Salted butter	0.032
Wheat flour	0.031
Granulated sugar	0.021
Luxury chocolate	0.018
Frozen puff pastry	0.017
Fresh blue grapes	0.016
Strawberry jam	0.015
Salt (various)	0.015
Puffed grain	0.014
Gouda cheese	0.013
(Self-rising) flour	0.013

(24) Baby

Product	P
Fresh Cookies	0.143
Baby food (1 year)	0.071
Baby Face Wipes	0.040
Fruit snacks	0.037
Baby cereal	0.036
Flavored Water	0.033
Easter Bunny	0.028
Fruit yogurt	0.019
Candy Bars	0.014
Oriental Soup	0.013
Soup Croutons	0.013
Frozen Desserts	0.011
Tomato soup	0.011
Mushroom Soup	0.011
Fresh Strawberries	0.011

(25) Typical Dutch

Product	P
Semi-skimmed milk	0.186
Potatoes with skin	0.033
Vanilla custard	0.032
Vegetable juice (fresh)	0.031
Bacon strips	0.030
Baby potatoes - peeled	0.026
Sausage (rookworst)	0.023
Hagelslag	0.023
Schnitzel	0.022
Chicken schnitzel	0.022
Cauliflower	0.021
Applesauce (jar)	0.020
'Stamppot' potatoes	0.019
Spinach (frozen)	0.016
Baby potatoes	0.016

(26) Dog care

Product	P
Wet dog food	0.105
Dog snacks	0.083
Luxury desserts	0.071
Frozen dog food	0.047
Fresh cake	0.041
Skimmed milk	0.037
White wine	0.036
Semi-skimmed milk	0.031
Coffee creamer	0.028
Pudding	0.022
Porridge - large	0.021
Breaded fish	0.020
Flavored quark	0.019
Porridge	0.015
Gouda cheese	0.015

(27) Vegan / Vegetarian

Product	P
Vegetarian chicken	0.070
Oat drink	0.068
Vegetarian schnitzel	0.054
Flavoured vegan yogurt	0.051
Vegetarian burgers	0.047
Vegetarian meat	0.034
Long-life soy drink	0.033
Coconut milk	0.031
Flavored hummus	0.028
Vegetable spread	0.027
Plain hummus	0.027
Plain vegan yogurt	0.027
Naan bread	0.021
Herbal tea	0.020
Spelt bread (fresh)	0.019

(28) Asian meal

Product	P
Prawn crackers	0.089
Stiry-fry vegetables	0.083
Chicken fillet - cubes	0.072
Diet cola bottles	0.041
Satay sauce	0.038
Chicken fillet - pieces	0.038
Orange juice (fresh)	0.033
Indonesian seasoning	0.026
Lettuce	0.025
Noodles	0.021
Easter cookies	0.020
White rice	0.019
Chinese cabbage	0.018
Mixed vegetables	0.018
Jasmine, pandan rice	0.017

(29) Coffee / Tea

Product	P
Ground coffee (regular)	0.098
Black tea	0.069
Peanuts (small)	0.055
Coffee beans (strong)	0.049
Coffee cookies	0.043
Tempranillo (Spain)	0.038
Rooibos tea	0.030
Pre-packaged bacon	0.027
Various red wine	0.026
Party nuts (small)	0.026
Coffee creamer	0.026
Coffee filters	0.024
Laundry detergent	0.021
Dolce Gusto coffee	0.020
Roast beef	0.020

(30) Fresh buns

Product	P
Fresh buns	0.659
Donuts	0.213
Multigrain Bread	0.031
Heavy Beer	0.014
Roast Beef	0.009
Regular Cola	0.008
Fresh Cake	0.006
Milk Rolls	0.006
Canned Mushrooms	0.004
Buttermilk	0.004
Pastry	0.004
Filet Americain	0.003
Children's Yogurt	0.003
Carpaccio	0.003
Green Tea	0.003

Appendix B

B.1 Descriptive Statistics

Table B.1: Summary of Customer-Specific Variables

Variable	**Percentage**
Age	
Unknown	50.66%
0-25	6.45%
25-35	9.20%
35-45	8.05%
45-55	7.61%
55-65	7.34%
65+	10.69%
Location	
City	7.96%
No city	92.04%

Table B.2: Summary of Trip-Specific Variables

Variable	**Percentage**
Day of Week	
Weekday	74.30%
Weekend	25.70%
Time of Day	
Daytime	71.76%
Evening	28.24%
Promotion (average)	
Discount	11.13 %

Appendix C

C.1 Correlations Between Motivations

Correlations	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1,00	-0,32	-0,18	-0,23	-0,29	-0,10	0,05	0,28	-0,10	-0,23	-0,26	0,03	-0,17	0,15	-0,28	-0,39	0,26	-0,05	-0,09	-0,11	0,27	-0,18	0,10	0,00	0,07	0,02	0,33	0,13	-0,08	0,02
2	-0,32	1,00	-0,02	0,13	0,45	0,49	-0,32	-0,53	0,20	0,24	0,55	-0,15	0,14	-0,41	-0,01	0,52	-0,61	0,25	0,21	-0,08	-0,49	0,24	-0,35	-0,04	0,15	0,03	-0,58	-0,03	-0,08	0,15
3	-0,18	-0,02	1,00	0,10	-0,09	-0,15	0,17	-0,14	-0,12	0,22	-0,05	-0,20	-0,06	0,07	0,46	-0,04	0,05	0,01	-0,27	0,24	0,13	-0,03	0,02	-0,16	-0,09	-0,08	-0,08	-0,12	0,21	-0,14
4	-0,23	0,13	0,10	1,00	0,11	0,01	0,08	-0,17	-0,14	-0,08	0,06	-0,04	0,02	0,03	0,29	0,10	-0,07	0,06	-0,05	0,04	-0,09	0,01	-0,11	-0,12	-0,10	-0,08	-0,15	0,02	0,13	-0,21
5	-0,29	0,45	-0,09	0,11	1,00	0,28	-0,26	-0,45	0,32	0,35	0,47	-0,07	0,27	-0,46	0,05	0,61	-0,48	-0,08	0,27	-0,12	-0,53	0,33	-0,16	0,06	-0,03	-0,02	-0,41	-0,31	-0,07	0,20
6	-0,10	0,49	-0,15	0,01	0,28	1,00	-0,39	-0,38	0,10	0,11	0,39	-0,14	0,07	-0,41	-0,18	0,26	-0,51	0,28	0,15	-0,07	-0,30	0,20	-0,33	-0,01	0,30	0,24	-0,48	0,03	-0,04	0,10
7	0,05	-0,32	0,17	0,08	-0,26	-0,39	1,00	0,32	-0,11	-0,13	-0,42	-0,02	-0,13	0,41	0,21	-0,22	0,45	-0,16	-0,21	0,05	0,20	-0,25	0,19	-0,21	-0,35	-0,31	0,42	-0,04	0,04	-0,19
8	0,28	-0,53	-0,14	-0,17	-0,45	-0,38	0,32	1,00	-0,13	-0,32	-0,49	0,12	-0,15	0,44	-0,21	-0,41	0,56	-0,21	-0,17	-0,10	0,35	-0,31	0,30	-0,06	-0,22	-0,13	0,57	0,07	-0,05	-0,09
9	-0,10	0,20	-0,12	-0,14	0,32	0,10	-0,11	-0,13	1,00	0,38	0,30	0,11	0,17	-0,35	-0,19	0,47	-0,24	-0,27	0,40	-0,22	-0,46	0,32	-0,20	-0,06	-0,21	0,03	-0,08	-0,36	-0,37	0,63
10	-0,23	0,24	0,22	-0,08	0,35	0,11	-0,13	-0,32	0,38	1,00	0,35	-0,02	0,15	-0,34	0,12	0,46	-0,32	-0,11	0,16	-0,05	-0,38	0,29	-0,22	-0,01	-0,09	0,09	-0,32	-0,31	-0,17	0,26
11	-0,26	0,55	-0,05	0,06	0,47	0,39	-0,42	-0,49	0,30	0,35	1,00	-0,03	0,21	-0,54	-0,03	0,55	-0,63	0,09	0,35	-0,04	-0,50	0,34	-0,32	0,10	0,16	0,24	-0,61	-0,14	-0,10	0,28
12	0,03	-0,15	-0,20	-0,04	-0,07	-0,14	-0,02	0,12	0,11	-0,02	-0,03	1,00	0,16	-0,20	0,04	0,13	-0,09	-0,20	0,44	-0,14	-0,21	0,21	-0,17	0,29	-0,16	0,12	0,12	-0,07	-0,19	0,19
13	-0,17	0,14	-0,06	0,02	0,27	0,07	-0,13	-0,15	0,17	0,15	0,21	0,16	1,00	-0,28	0,07	0,37	-0,25	-0,21	0,26	-0,03	-0,34	0,23	-0,07	0,13	-0,15	0,08	-0,13	-0,26	-0,05	0,11
14	0,15	-0,41	0,07	0,03	-0,46	-0,41	0,41	0,44	-0,35	-0,34	-0,54	-0,20	-0,28	1,00	0,00	-0,53	0,71	0,05	-0,50	0,06	0,48	-0,53	0,38	-0,28	-0,16	-0,44	0,54	0,24	0,09	-0,36
15	-0,28	-0,01	0,46	0,29	0,05	-0,18	0,21	-0,21	-0,19	0,12	-0,03	0,04	0,07	0,00	1,00	0,10	-0,05	-0,03	-0,06	0,22	-0,04	0,05	-0,04	0,02	-0,23	-0,05	-0,10	-0,12	0,26	-0,24
16	-0,39	0,52	-0,04	0,10	0,61	0,26	-0,22	-0,41	0,47	0,46	0,55	0,13	0,37	-0,53	0,10	1,00	-0,57	-0,14	0,47	-0,16	-0,70	0,44	-0,32	0,04	-0,15	-0,01	-0,47	-0,33	-0,20	0,36
17	0,26	-0,61	0,05	-0,07	-0,48	-0,51	0,45	0,56	-0,24	-0,32	-0,63	-0,09	-0,25	0,71	-0,05	-0,57	1,00	-0,16	-0,45	0,01	0,52	-0,45	0,44	-0,21	-0,27	-0,37	0,71	0,11	0,01	-0,26
18	-0,05	0,25	0,01	0,06	-0,08	0,28	-0,16	-0,21	-0,27	-0,11	0,09	-0,20	-0,21	0,05	-0,03	-0,14	-0,16	1,00	-0,17	0,08	0,08	-0,13	-0,16	-0,08	0,38	0,03	-0,27	0,38	0,10	-0,22
19	-0,09	0,21	-0,27	-0,05	0,27	0,15	-0,21	-0,17	0,40	0,16	0,35	0,44	0,26	-0,50	-0,06	0,47	-0,45	-0,17	1,00	-0,23	-0,50	0,40	-0,31	0,25	-0,12	0,18	-0,24	-0,17	-0,27	0,46
20	-0,11	-0,08	0,24	0,04	-0,12	-0,07	0,05	-0,10	-0,22	-0,05	-0,04	-0,14	-0,03	0,06	0,22	-0,16	0,01	0,08	-0,23	1,00	0,17	-0,12	0,05	0,02	0,13	0,10	-0,08	0,03	0,31	-0,24
21	0,27	-0,49	0,13	-0,09	-0,53	-0,30	0,20	0,35	-0,46	-0,38	-0,50	-0,21	-0,34	0,48	-0,04	-0,70	0,52	0,08	-0,50	0,17	1,00	-0,36	0,36	-0,07	0,14	-0,05	0,34	0,27	0,26	-0,37
22	-0,18	0,24	-0,03	0,01	0,33	0,20	-0,25	-0,31	0,32	0,29	0,34	0,21	0,23	-0,53	0,05	0,44	-0,45	-0,13	0,40	-0,12	-0,36	1,00	-0,27	0,22	-0,04	0,22	-0,34	-0,27	-0,16	0,30
23	0,10	-0,35	0,02	-0,11	-0,16	-0,33	0,19	0,30	-0,20	-0,22	-0,32	-0,17	-0,07	0,38	-0,04	-0,32	0,44	-0,16	-0,31	0,05	0,36	-0,27	1,00	0,06	-0,15	-0,22	0,30	-0,05	0,15	-0,26
24	0,00	-0,04	-0,16	-0,12	0,06	-0,01	-0,21	-0,06	-0,06	-0,01	0,10	0,29	0,13	-0,28	0,02	0,04	-0,21	-0,08	0,25	0,02	-0,07	0,22	0,06	1,00	0,07	0,27	-0,13	-0,04	0,04	0,01
25	0,07	0,15	-0,09	-0,10	-0,03	0,30	-0,35	-0,22	-0,21	-0,09	0,16	-0,16	-0,15	-0,16	-0,23	-0,15	-0,27	0,38	-0,12	0,13	0,14	-0,04	-0,15	0,07	1,00	0,23	-0,32	0,30	0,11	-0,08
26	0,02	0,03	-0,08	-0,08	-0,02	0,24	-0,31	-0,13	0,03	0,09	0,24	0,12	0,08	-0,44	-0,05	-0,01	-0,37	0,03	0,18	0,10	-0,05	0,22	-0,22	0,27	0,23	1,00	-0,32	-0,02	0,05	0,03
27	0,33	-0,58	-0,08	-0,15	-0,41	-0,48	0,42	0,57	-0,08	-0,32	-0,61	0,12	-0,13	0,54	-0,10	-0,47	0,71	-0,27	-0,24	-0,08	0,34	-0,34	0,30	-0,13	-0,32	-0,32	1,00	0,02	-0,10	-0,06
28	0,13	-0,03	-0,12	0,02	-0,31	0,03	-0,04	0,07	-0,36	-0,31	-0,14	-0,07	-0,26	0,24	-0,12	-0,33	0,11	0,38	-0,17	0,03	0,27	-0,27	-0,05	-0,04	0,30	-0,02	0,02	1,00	0,06	-0,22
29	-0,08	-0,08	0,21	0,13	-0,07	-0,04	0,04	-0,05	-0,37	-0,17	-0,10	-0,19	-0,05	0,09	0,26	-0,20	0,01	0,10	-0,27	0,31	0,26	-0,16	0,15	0,04	0,11	0,05	-0,10	0,06	1,00	-0,35
30	0,02	0,15	-0,14	-0,21	0,20	0,10	-0,19	-0,09	0,63	0,26	0,28	0,19	0,11	-0,36	-0,24	0,36	-0,26	-0,22	0,46	-0,24	-0,37	0,30	-0,26	0,01	-0,08	0,03	-0,06	-0,22	-0,35	1,00

Figure C.1: Motivation Correlation Matrix

Appendix D

D.1 Odds Ratios

Table D.1: Summary of Odds Ratios for Trip-Specific Variables

Variable	Lowest odds ratios	Mean	Highest odds ratios
Discount percentage	M14: Mexican meal (0.87)	1.01	M3: Beer/Self-care (1.27)
	M23: Baking (0.88)		M20: Housekeeping (1.19)
Weekday	M30: Fresh buns (0.91)	1.00	M28: Asian meal (1.06)
	M4: Appetisers (0.92)		M12: Healthy convenience (1.06)
Weekend	M28: Asian meal (0.83)	1.01	M30: Fresh buns (1.30)
	M12: Healthy convenience (0.83)		M4: Appetisers (1.23)
Before 5 p.m.	M16: Energy/Soft drinks (0.88)	1.01	M9: Sweet breakfast (1.44)
	M15: Beer/Ready-to-eat (0.88)		M30: Fresh buns (1.35)
After 5 p.m.	M9: Sweet breakfast (0.39)	0.99	M16: Energy/Soft drinks (1.36)
	M30: Fresh buns (0.46)		M15: Beer/Ready-to-eat (1.34)

Table D.2: Summary of Odds Ratios for Customer-Specific Variables

Variable	Lowest odds ratios	Mean	Highest odds ratios
Age: Unknown	M28: Asian meal (0.78)	1.00	M17: Fruits/Vegetables (1.18)
	M7: High budget (0.82)		M3: Beer/Self-care (1.11)
Age: 0-25	M20 Housekeeping: 0.43	1.14	M16: Energy/Soft drinks (4.30)
	M26: Dog care (0.40)		M9: Sweet breakfast (2.65)
Age: 25-35	M3: Beer/Self-care (0.52)	0.99	M16: Energy/Soft drinks (1.76)
	M15: Beer/Ready-to-eat (0.61)		M27: Vegan/Vegetarian (1.73)
Age: 35-45	M7: High budget (0.42)	0.90	M17: Fruits/Vegetables (1.52)
	M15: Beer/Ready-to-eat (0.44)		M30: Fresh buns (1.48)
Age: 45-55	M17: Fruits/Vegetables (0.71)	1.04	M7: High budget (1.51)
	M27: Vegan/Vegetarian (0.88)		M28: Asian meal (1.39)
Age: 55-65	M16: Energy/Soft drinks (0.39)	0.92	M15: Beer/Ready-to-eat (1.38)
	M9: Sweet breakfast (0.53)		M4: Appetisers (1.33)
Age: 65+	M9: Sweet breakfast (0.21)	0.87	M29: Coffee/Tea (2.23)
	M16: Energy/Soft drinks (0.30)		M25: Typical Dutch (1.47)
Location: City	M2: Convenience (0.67)	1.02	M20: Housekeeping (1.46)
	M25: Typical Dutch (0.72)		M7: High budget (1.36)
Location: No City	M20: Housekeeping (0.97)	1.00	M2: Convenience (1.03)
	M7: High budget (0.97)		M25: Typical Dutch (1.03)

Appendix E

E.1 Average Hit Rate and Novelty Hit Rate Evaluation

Table E.1: Average Hit Rate and Novelty Hit Rate for Different Recommenders

Model	**Average Hit Rate**				**Average Novelty Hit Rate**			
	$S = 1$	$S = 3$	$S = 5$	$S = 10$	$S = 1$	$S = 3$	$S = 5$	$S = 10$
CTM-30	0.253	0.202	0.193	0.211	0.119	0.104	0.108	0.136
CTM-30 (no customer effects)	0.277	0.210	0.198	0.214	0.120	0.103	0.108	0.135
Popularity-based recommender	0.114	0.121	0.133	0.169	0	0	0	0
Marginal Probability Model	0.069	0.079	0.097	0.114	0.029	0.036	0.044	0.045

E.2 Average Hit Rate Evaluation for New Customers

Table E.2: Average Hit Rate Comparison for New Customers

Model	**Average Hit Rate**			
	$S = 1$	$S = 3$	$S = 5$	$S = 10$
CTM-30	0.119	0.110	0.104	0.114
CTM-30 (no customer effects)	0.112	0.093	0.089	0.096
Popularity-based recommender	0.153	0.126	0.125	0.129
Marginal Probability Model	0.060	0.069	0.089	0.108

Appendix F

F.1 Prior Specifications

This section provides an overview of the prior specifications for the population-level parameters, as suggested by Jacobs et al. (2021). We begin by transforming variance parameters into their corresponding precision parameters. Specifically, the variance of ϵ_{ibm}, denoted as $\sigma^2_{\alpha_m}$, becomes precision parameter $\tau_{\alpha_m} \equiv \sigma^{-2}_{\alpha_m}$. This transformation results in a precision vector $\boldsymbol{\tau_a} = [\tau_{\alpha_1}, ..., \tau_{\alpha_M}]$. Additionally, we transform the covariance matrix $\boldsymbol{\Sigma}_\kappa$ of the $\boldsymbol{\kappa}_i$ variables into a precision matrix denoted as $\boldsymbol{\Lambda}\kappa$ which is defined as the inverse of $\boldsymbol{\Sigma}\kappa$.
Furthermore, we assign priors to each population parameter in a way that ensures conjugacy with the corresponding parameter's full-conditional distribution. The prior distributions for the population-level parameters are chosen to be relatively uninformative. For $m = 1, ..., M$ we specify the following priors:

$\boldsymbol{\phi}_m \sim \text{Dirichlet}_J(\boldsymbol{\beta_0} = \mathbf{1}_J J^{-1})$,

$\boldsymbol{\delta}_m \sim \text{MVN}_{K_X}(\boldsymbol{\mu} = \mathbf{0}_{K_X}, \boldsymbol{\Sigma} = \boldsymbol{I}_{K_X})$,

$\boldsymbol{\gamma}_m \sim \text{MVN}_{K_H}(\boldsymbol{\mu} = \mathbf{0}_{K_H}, \boldsymbol{\Sigma} = \boldsymbol{I}_{K_H})$,

$\tau_{\alpha_m} \sim \text{Gamma}(\alpha = 1, \beta = 1)$,

$\boldsymbol{\mu}_\kappa \sim \text{MVN}_M(\boldsymbol{\mu} = \mathbf{0}_M, \boldsymbol{\Sigma} = \mathbf{I}_M)$,

$\boldsymbol{\Lambda}_\kappa \sim \text{Wishart}_M(n = 2M, \boldsymbol{V} = \boldsymbol{I}_M(2M)^{-1})$.

Appendix G

G.1 Coordinate Ascent Optimisation Algorithm

To maximise the ELBO of our model, as referenced in Equation (4.17), we use the Coordinate Ascent Optimisation Algorithm as suggested by Jacobs et al. (2021). The pseudocode of this algorithm is given in Algorithm 1. Following each update step within the algorithm, it is assured that the ELBO will not decrease. In each iteration and for each customer i, we conduct $L = 25$ subiterations to jointly optimise the variational distributions tailored to that specific customer. The optimisation process ends once the ELBO has reached convergence. For detailed insights into the steps involved in updating each parameter's variational distribution, refer to Appendix G.2.

Algorithm 1 VI Coordinate Ascent Optimisation Algorithm

Require: $q(\boldsymbol{\omega})$
while ELBO has not converged **do**
    for $i = 1$ to I **do**
        for $l = 1$ to L **do**
            for $b = 1$ to N_{ib} **do**
                Update $q(\boldsymbol{z}_{ib})$, $q(\boldsymbol{\alpha}_{ib})$
            end for
            Update $q(\boldsymbol{\kappa}_i)$
        end for
    end for
    Update $q(\boldsymbol{\phi})$, $q(\boldsymbol{\delta})$, $q(\boldsymbol{\gamma})$, $q(\boldsymbol{\tau}_\alpha)$, $q(\boldsymbol{\mu}_\kappa)$, $q(\boldsymbol{\Lambda}_\kappa)$
end while

To initiate the CAVI algorithm and ensure a consistent solution, the parameters of the variational distribution $q(\boldsymbol{\omega})$ need to be initialised. The choice of several initial parameters can impact the optimisation outcome, considering the updating order. The multivariate normal distributions are initialised with zero means, and $q(\boldsymbol{\Lambda}_\kappa)$ is initialised as shown in F.1. Regarding the initialisation of $q(z_{ibn})$, it does not affect the optimisation outcome. However, it is necessary to initialise $q(\boldsymbol{\phi}_m)$. For this purpose, we generate pseudocounts obtained from a collapsed Gibbs LDA algorithm using the available data, following the example of Jacobs et al. (2021). We employ 10,000 iterations, using a burn-in of 5,000 iterations. Then, we iterate through the parameters, replacing each with a revised estimate based on the variational updates in Appendix G.2. This revised estimate is evaluated using the current estimates of all the other parameters.

G.2 Derivation of the Variational Updates

To derive the variational updates, we leverage known results of Jacobs et al. (2021). In their VI approach, a partitioning function $F(\boldsymbol{\omega})$ is introduced that preserves all elements of each multivariate parameter in a distinct subset, denoted as w. Following the mean-field variational inference approach, each probability distribution in Q can be factorised over the unknown parameters $\boldsymbol{\omega}$. To find a solution, the CAVI algorithm iterates over the subsets of parameters w and updates the corresponding variational distribution $q(w)$ to minimise the ELBO.

When employing the mean-field assumption, Q no longer captures the true posterior distribution precisely. Instead, $q(w)$ serves as the variational approximation of the marginal posterior distribution of w. The accuracy of this approximation is influenced by the partitioning function $F(\boldsymbol{\omega})$. Finer partitioning, such as factorising multivariate parameters with only one parameter per w, sacrifices accuracy for computational simplicity. In our model, given the expected uncertainty in customer and trip-specific parameters, as many customers purchase only a few products per trip, we keep all elements of a multivariate parameter in a single subset w. This approach aligns with the methodology proposed by Jacobs et al. (2021).

By leveraging established findings from the literature on variational inference (VI), Jacobs et al. (2021) derive the optimal variational distribution $q^*(w)$ for each conditionally conjugate parameter in $\boldsymbol{\omega}$. For all variables except $\boldsymbol{\alpha}_{ib}$, we use the corresponding conjugate priors. By doing so, the complete conditional distribution remains in the same exponential family (Blei et al., 2017). All parameters except $\boldsymbol{\alpha_{ib}}$ are considered conditionally conjugate in our specific model.

When applying the mean-field assumption to distributions of the exponential family, as described in Bishop's work (2016), we can express the optimal variational distribution in a more general form. Since w is a conditionally conjugate parameter in our model and follows an exponential distribution, the optimal variational distribution for w can be expressed as follows, as shown by Jacobs et al. (2021):

$$q^*(w) \propto h(w) \exp\left(t(w)^\intercal \widetilde{E}_{\mathrm{MB}_w}\{\eta(\mathrm{MB}_w)\}\right). \tag{G.1}$$

To determine the optimal variational distribution of w, as described in Equation (G.1), it is essential to select appropriate functions h and t that match the functional form associated with the base measure and the sufficient statistic of the prior distribution for w (Jacobs et al., 2021). When inferring a random variable using a set of variables, considering only a subset of variables is often sufficient, commonly referred to as the Markov blanket (Pearl, 1988). The Markov blanket for w, denoted as MB_w, consists of a subset of variables that contain relevant information for inferring w.

The CAVI algorithm operates iteratively by setting each natural variational parameter of the optimal variation distribution $q^*(w)$ to the variational expectation $\widetilde{E}_{\mathrm{MB}_w}\{\eta(\mathrm{MB}_w)\}$ of its natural conditional parameter $\eta(\mathrm{MB}_w)$. This expectation is computed given all other parameters and observations. The natural parameter $\eta(\mathrm{MB}_w)$ corresponds to the full-conditional distribution $p(w|\mathrm{MB}_w)$.

To compute this expectation, it suffices to have access to the variational distributions of the parameters within the Markov blanket for w. Equation (G.1) allows for the derivation of updates for all parameters in the model that are conditionally conjugate. The closed-form solutions for these variational updates, derived by Jacobs et al. (2021), are presented in Appendix G.3. These updates will be used by the Coordinate Ascent Variational Inference (CAVI) optimisation algorithm to optimise the ELBO.

G.3 Solutions for the Variational Updates

Here, we present the update steps tailored to each parameter's variational distribution, as derived by Jacobs et al. (2021). To begin, we specify the variational expectations of ϵ_{ibm}, which are crucial for the variational updates of $\boldsymbol{\alpha}_{ib}$. We employ the specifications detailed in Equation (4.6):

$$\begin{aligned} &\widetilde{E}\{\epsilon_{ibm}\} \equiv \widetilde{E}\{\alpha_{ibm}\} - \widetilde{E}\{\mu_{ibm}\}, \\ &\widetilde{E}\{\epsilon^2_{ibm}\} \equiv \widetilde{E}\{\alpha^2_{ibm}\} + \widetilde{E}\{\mu^2_{ibm}\} - 2\widetilde{E}\{\alpha_{ibm}\}\widetilde{E}\{\mu_{ibm}\}. \end{aligned} \tag{G.2}$$

Following this, by using the expression for μ_{ibm}, we derive the following equation

$$\widetilde{E}\{\boldsymbol{\mu}_{ib}\} = \widetilde{E}\{\boldsymbol{\kappa}_i\} + \widetilde{E}\{\boldsymbol{D}\}\boldsymbol{x}_{ib} + \widetilde{E}\{\boldsymbol{G}\}\boldsymbol{h}_i, \tag{G.3}$$

where matrices $\boldsymbol{D}$ (M x K_x) and $\boldsymbol{G}$ (M x K_H) store all $\boldsymbol{\delta}_m$ and $\boldsymbol{\gamma}_m$ vectors, respectively. This implies that each row in these matrices corresponds to a specific motivation. Additionally, the vectors $\boldsymbol{x}_{ib}$, $\boldsymbol{h}_i$, $\boldsymbol{\epsilon}_{ib}$ and $\boldsymbol{\alpha}_{ib}$ are collected in matrices $\boldsymbol{X}$ ($\sum_{i=1}^{I} B_i$ x K_X), $\boldsymbol{H}$ ($\sum_{i=1}^{I} B_i$ x K_H) and $\boldsymbol{E}$ ($\sum_{i=1}^{I} B_i$ x M).

To facilitate understanding, we will introduce more relevant notation. Firstly, the notation $\langle \boldsymbol{\epsilon}_{ib} + \boldsymbol{\kappa}_i \rangle$, which signifies the cleaned effect of $\boldsymbol{\alpha}_{ib}$ of all variables except for $\boldsymbol{\kappa}_i$. Calculating the variational moments of such terms for $\boldsymbol{\alpha}_{ib}$ is often necessary, and notably, these moments are independent of $q(\boldsymbol{\kappa}_i)$. Furthermore, we use the values $\mathbf{0}$, $\mathbf{1}$ or the identity matrix $\boldsymbol{I}$ to represent the fixed values of prior parameters, as outlined in Appendix F. Additionally, the notation $[x]$ equals 1 if x is true and 0 otherwise. Moreover, the function $d(\boldsymbol{v})$ denotes the creation of a diagonal matrix from vector $\boldsymbol{v}$, while $\widetilde{E}\{\boldsymbol{\eta}\}$ represents the expectation of $\boldsymbol{\eta}$ under its variational distribution $q(\boldsymbol{\eta})$. In mathematical terms, this means $\widetilde{E}\{\boldsymbol{\eta}\} \equiv \mathbb{E}_{q(\boldsymbol{\eta})}\{\boldsymbol{\eta}\}$.

Finally, we use Equation (G.1) to derive the updates for all conditionally conjugate parameters in our model. For a more detailed discussion, Please refer to Appendix G.2 for a more detailed discussion.

Update $q(z_{ibn})$: M-dimensional Categorical with probability that $z_{ibn} = m$ given by

$$\widetilde{p}_{ibnm} = \frac{\exp(\widetilde{E}\{\alpha_{ibm}\}+\widetilde{E}\{\log\phi_{m,y_{ibn}}\})}{\sum_{l=1}^{M}\exp(\widetilde{E}\{\alpha_{ibl}\}+\widetilde{E}\{\log\phi_{l,y_{ibn}}\})}.$$

(G.4)

Update $q(\boldsymbol{\phi}_m)$: J-dimensional Dirichlet with parameter $\tilde{\boldsymbol{\beta}}_m$ for which the jth element is given by

$$\widetilde{\beta}_{mj} = J^{-1} + \sum_i \sum_{b=1}^{B_i} \sum_{n=1}^{N_{ib}} \widetilde{E}([z_{ibn} = m])[y_{ibn} = j].$$

(G.5)

Update $q(\tau_{\alpha_m})$: Gamma with parameters

$$\widetilde{a}_{\tau_{\alpha_m}} = 1 + \tfrac{1}{2}\sum_i B_i,$$

$$\widetilde{b}_{\tau_{\alpha_m}} = 1 + \tfrac{1}{2}\sum_i \sum_{b=1}^{B_i} \widetilde{E}\{\epsilon_{ibm}^2\}.$$

(G.6)

Update $q(\boldsymbol{\Lambda}_\kappa)$: M-dimensional Wishart with parameters

$$\widetilde{n}_{\Lambda_k} = 2M + I,$$

$$\widetilde{\boldsymbol{V}}_{\Lambda_k} = \left(\boldsymbol{I}_M(2M) + \sum_i \widetilde{E}\{\boldsymbol{\mu}_k\boldsymbol{\mu}_k^\intercal + \boldsymbol{\kappa}_i\boldsymbol{\kappa}_i^\intercal - \boldsymbol{\mu}_k\boldsymbol{\kappa}_i^\intercal - \boldsymbol{\kappa}_i\boldsymbol{\mu}_k^\intercal\}\right)^{-1}.$$

(G.7)

Update $q(\boldsymbol{\kappa}_i)$: M-dimensional multivariate Normal with parameters

$$\widetilde{\boldsymbol{\Sigma}}_i = \widetilde{E}\left\{\boldsymbol{\Lambda}_k + d(\boldsymbol{\tau}_\alpha)(B_i - 1)\right\}^{-1},$$

$$\widetilde{\boldsymbol{\mu}}_i = \widetilde{\boldsymbol{\Sigma}}_i\widetilde{E}\left\{\boldsymbol{\Lambda}_k\boldsymbol{\mu}_\kappa + d(\boldsymbol{\tau}_\alpha)\left(\sum_{b=2}^{B_i}\langle\boldsymbol{\epsilon}_{ib} + \boldsymbol{\kappa}_i\rangle\right)\right\}.$$

(G.8)

Update $q(\boldsymbol{\delta}_m)$: K_X-dimensional multivariate Normal with parameters

$$\widetilde{\boldsymbol{\Sigma}}_{\delta_m} = \widetilde{E}\left\{\boldsymbol{I}_{K_X} + (\boldsymbol{X}^\intercal \boldsymbol{X})\,\tau_{\alpha_m}\right\}^{-1},$$

$\widetilde{\boldsymbol{\mu}}_{\delta_m} = \widetilde{\boldsymbol{\Sigma}}_{\delta_m} \widetilde{E}\left\{\mathbf{0}_{K_X} + \tau_{\alpha_m}\left(\boldsymbol{X}^\intercal \langle \boldsymbol{e}_m + \boldsymbol{X}\boldsymbol{\delta}_m\rangle\right)\right\}.$

(G.9)

Update $q(\boldsymbol{\gamma}_m)$: K_H-dimensional multivariate Normal with parameters

$$\widetilde{\boldsymbol{\Sigma}}_{\gamma_m} = \widetilde{E}\left\{\boldsymbol{I}_{K_H} + (\boldsymbol{H}^\intercal \boldsymbol{H})\,\tau_{\alpha_m}\right\}^{-1},$$

$\widetilde{\boldsymbol{\mu}}_{\gamma_m} = \widetilde{\boldsymbol{\Sigma}}_{\gamma_m} \widetilde{E}\left\{\mathbf{0}_{K_H} + \tau_{\alpha_m}\left(\boldsymbol{H}^\intercal \langle \boldsymbol{e}_m + \boldsymbol{H}\boldsymbol{\gamma}_m\rangle\right)\right\}.$

(G.10)

Update $q(\boldsymbol{\mu}_\kappa)$: M-dimensional multivariate Normal with parameters

$$\widetilde{\boldsymbol{\Sigma}}_{\mu_\kappa} = \widetilde{E}\left\{\boldsymbol{I}_M + \boldsymbol{\Lambda}_{\boldsymbol{\kappa}} I\right\}^{-1},$$

$\widetilde{\boldsymbol{\mu}}_{\mu_\kappa} = \widetilde{\boldsymbol{\Sigma}}_{\mu_\kappa} \widetilde{E}\left\{\mathbf{0}_M + \boldsymbol{\Lambda}_\kappa \sum_i \boldsymbol{\kappa}_i\right\}.$

(G.11)

Since $\boldsymbol{\alpha}_{ib}$ is the only not conditionally conjugate parameter, we cannot calculate its optimal variational distribution analytically. Hence, we turn to the prior family for $\boldsymbol{\alpha}_{ib}$, relying on a set of independent normal distributions. Please refer to the paper by Jacobs et al. (2021) for the full details on the variational update for $q(\boldsymbol{\alpha}_{ib})$.