Tactical time slot management under fulfillment considerations in e-grocery

Luís Miguel Vaz de Carvalho

Master's Dissertation

Supervisor: Prof. Flávia Barbosa

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia e Gestão Industrial

2019-07-01

Abstract

E-gorocery retail is one of the fastest growing businesses in Portugal and in the World. In a busy society, customers value the convenience of home-deliveries and the exemption in spending time commuting to the stores. Retailers recognize the growth of e-commerce in recent years and its long-term potential, being focused on building a strong market position while planning sustainable operations. They are challenged with the trade-off of providing attractive delivery windows to the customers while keeping fulfillment operations efficient. Therefore, an online retail store needs to decide not only the optimal combination of available slots but also the delivery fees to provide to its customers.

This project was motivated by the case of a food retailer, aiming to develop analytical support on decision-making concerning the e-commerce operations. The current planning process is empirical, with no risk-assessment and no long-term impact analysis. Therefore, the retailer cannot consider the impact on customer satisfaction and fulfillment operations. The scope of this thesis adds to the arduousness of the planning task, due to both consumer behavior and fulfillment operations being considered alongside.

The proposed approach recreates the consumer behavior and the fulfillment operations, using a simulation procedure, to test the impact of different delivery plans. Consumer behavior is approximated by the estimation of key parameters to characterize the costumers' profiles and the development of a machine learning model to estimate delivery preferences. The fulfillment operations are approximated by the estimation of the different phases on the delivery process, in order to assess the impact on operational costs. To assess the developed methodology and incite the advisory support of the project, differentiated pricing and slotting approaches are applied to generate new delivery scenarios.

The results generated from the application of the methodology developed show a significant improvement in reducing the uncertainty over the outcome of new delivery plans and improving the retailer's planning capacity. The suggested scenarios are expected to balance the usage of the available delivery options, while reducing the costs associated with the fulfillment operations, without impacting the number of deliveries.

Resumo

O retalho online é um dos negócios com maior crescimento em Portugal e no mundo. Em sociedades em que a gestão do tempo é fundamental, os clientes valorizam a conveniência de entregas em casa e a possibilidade de realizar as suas compras no conforto do lar. Os retalhistas reconhecem o crescimento dos serviços de *e-commerce* e o seu potencial de longo prazo, pelo que estão cada vez mais focados no desenvolvimento de uma forte posição no mercado e no planeamento de operações sustentáveis. São, também, confrontados com o *trade-off* entre apresentar janelas de entrega atrativas para os clientes, ou manter as operações de abastecimento eficientes. Portanto, um retalhista on-line precisa de decidir, não apenas a combinação ideal de janelas de entregas disponíveis, mas também as taxas de entrega a serem aplicadas aos seus clientes.

Este projeto foi motivado pelo caso de um retalhista alimentar português, com o objetivo de desenvolver o suporte analítico relativo à tomada de decisão das operações de e-commerce. O atual processo de planeamento é empírico, sem análise de risco e sem avaliação de impactos no longo prazo. No modelo atual, o retalhista não é capaz de considerar o impacto nas operações de entrega e na satisfação do cliente. O objeto de estudo desta dissertação apresenta uma dificuldade e relevância acrescida, por conciliar o comportamento do consumidor e as operações de entrega.

A abordagem proposta recria o comportamento do consumidor e as operações de entrega, através de uma ferramenta de simulação, sendo desta forma possível testar o impacto de diferentes planos de entrega. O comportamento do consumidor é estudado pela aproximação de parâmetros-chave, com o intuito de caracterizar o perfil dos clientes, e pelo desenvolvimento de um modelo de *machine leaning* para prever as preferências de entrega. As operações de entregas são analisadas pela aproximação das diferentes fases do processo de abastecimento, com o objetivo de avaliar o impacto nos custos operacionais. Para estudar a aplicabilidade da metodologia desenvolvida e promover o processo de consultoria do projeto, são aplicadas abordagens táticas de *pricing* e de definição de janelas de entregas para gerar novos planos entregas.

Os resultados obtidos com a implementação do projeto demonstram uma melhoria significativa na redução da incerteza sobre o a aplicação prática de novos planos de entrega e na melhoria da capacidade de planeamento do retalhista. Os cenários sugeridos equilibram o uso das opções de entrega disponíveis, reduzindo os custos associados às operações de abastecimento, sem impactar o número de entregas.

Acknowledgements

The philosopher São Tomás de Aquino defined three levels of gratitude in his text "Tratado de Gratidão". The first level is the most superficial level, where we recognize the others for their actions. The second level is the one related to acknowledgment, which is used to praise someone for their help. The third level reflects our commitment to the ones who have supported us.

The acknowledgments in languages, such as English or German, use the words "thank you" and "danke" that refer to the first level, while French and Spanish point to the second level with the words "merci" and "grazie". However, Portuguese is the only language that allows thanking in the deepest level with the word "obrigado". Bearing this in mind, I will proceed this chapter in Portuguese to thank everyone who helped me with such a powerful word: OBRIGADO.

Obrigado à LTPlabs, pela oportunidade de participar num projeto tão ambicioso e desafiante e de trabalhar com uma equipa repleta de mentes brilhantes. Foi sem dúvida um período essencial para o meu crescimento pessoal e profissional.

Obrigado ao Professor Pedro Amorim, por me ter desafiado a participar neste projeto. Toda a motivação, confiança e conhecimento foram essenciais ao desenvolvimento desta dissertação.

Obrigado aos Engenheiros Daniel Pereira e Paulo Sousa, pela forma cuidada e atenta com que reviram este trabalho, mas também pelos valiosos conhecimentos transmitidos e incansável disponibilidade em me orientar da melhor forma.

Obrigado à Professora Flávia Barbosa, pela disponibilidade e apoio constante no desenvolvimento da dissertação, mas também pelos valiosos conselhos.

Obrigado a toda a minha equipa de andebol do F.C. Gaia, por fazerem parte de algo tão importante na minha vida. Conquistar aquilo que conquistámos no mesmo ano em que desenvolvi esta dissertação acrescenta um sentimento e valor especial a todo este trabalho.

Obrigado a todos os meus amigos e colegas que me acompanharam durante o meu percurso académico. Partilhar bons e maus momentos rodeados daqules de quem gostamos torna o caminho mais fácil.

Obrigado ao meu Irmão, por ter sido uma fonte de inspiração em todas as fases do meu crescimento.

Obrigado à minha Mãe, por ser o meu porto-seguro e por ser o melhor abraço que posso receber.

Obrigado ao meu Pai, por ser o meu maior exemplo de vida, por ser o pai e homem que eu espero um dia vir a ser.

Obrigado à Maria, por todo o apoio, carinho e compreensão incondicional. Percorrer este caminho sem a tua presença não teria tido o mesmo significado.

"Não tenhamos pressa, mas não percamos tempo"

José Saramago

Contents

Acronyms and Symbols

ABM	Agent-Based Modelling
DEM	Discrete-Event Modelling
DT	Decision Trees
GAM	Generalized Attraction Model
GBM	Gradient Boosting Machines
IIA	Independence of Irrelevant Attribute
MAPE	Mean Absolute Percentage Error
MNL	Multinomial Logit Model
NL	Nested Logit Model
RF	Random Forests
SD	System Dynamics
SKU	Stock Keeping Unit

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In e-commerce, the e-grocery field has been growing 15% in the last two years and is expected to keep increasing (Nielsen, 2018). Thereby, many retail companies have been adopting this service to follow the market trend. The home delivery is attractive to the customers for various reasons, such as convenience over the delivery process, a busy lifestyle, exemption of transportation to the mall, or simply fast access to the online platforms. However, the delivery planning is a critical function to the retailer due to the slot appointment being the most important shopping feature (UPS, 2017). The delivery operations require the presence of the customer at home, the customers value the control over the delivery period and their willingness to wait depends on the weekdays, which leads to complex decisions for the retailer when planning the fulfillment operations.

The retailer needs to define which are the delivery windows available for the customers to book. The decision is influenced by the expected consequences on the fulfillment operations and by the service provided to the customers. Regarding the former, the main challenge is maintaining transportation costs under control. This efficiency is influenced by factors such as traveling distance, customer density, expected road traffic and available time to perform the deliveries to the customers. The sense of the quality of the e-commerce service is dependent on the delivery options available and the suitability to the consumers' needs. Thus, the retailer has to balance the trade-off between deliveries adjusted to customer preferences and efficient operations.

The process of setting the available delivery options should also encompass the definition of the price to be paid by the customer. This fee should reflect not only the difficulty of the retailer to perform the delivery operation but also the importance of different schedules to the customer. Therefore, the retailer should understand which are the factors that drive the customer's choice and how to combine them with the business goals. From the retailer's perspective, the fee paid for the delivery by the customer is not expected to generate profit to the business, but to offset the costs coming from the fulfillment operations.

Agatz et al. (2008b) defines the process of setting the delivery windows availability and price as time slot management. Regarding the complex decisions to be undertaken under the scope of this management problem, it is essential to have an analytical support behind the process of defining the available delivery windows and the fees to be assigned, in order to be more confident that the delivery offers are the most suitable to the customers' profile and the retailers' fulfillment operations.

1.2 Project description

This project was conducted in a Portuguese food retailer that is one of the main players in the e-grocery national market. This company has been having troubles defining the available delivery windows and the fees to charge for the delivery operation. The current process is conducted by a retailer's internal team and is empirical. There is a lack of analysis regarding the impact of different delivery plans since measurements and conclusions can only be withdrawn from real life scenarios. Consequently, the company realizes the urgency of developing methods that support the choice of the most efficient fulfillment plan, while increasing customer satisfaction.

To improve the current time slot management practices of the retailer, the project's main goal is the development of an analytical methodology to estimate the expected results on the fulfillment operations and on the consumer acceptance of different delivery windows and fees to be charged. Thereby, it is expected to help the decision-making on the tactical definition of the retailer's strategy, or, in other words, the plan for the delivery operations on the medium-term time horizon. It is important to clarify that all the slots' prices, present in this work, were transformed and operational costs used in relative values, due to confidentiality agreements.

The project is divided into five main stages, as described in Figure 1.1.

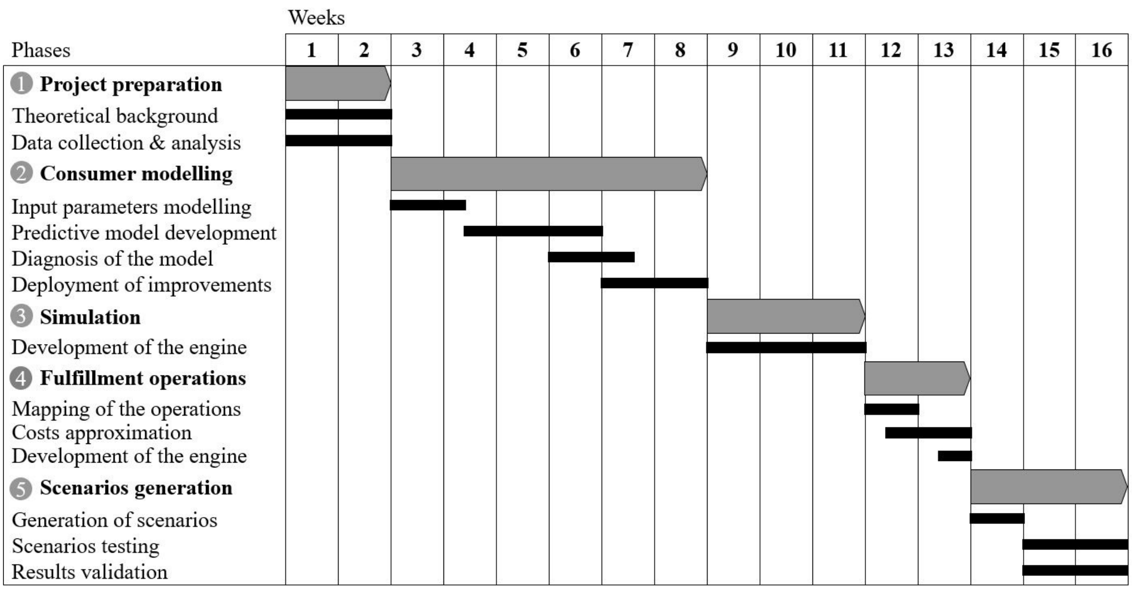

Figure 1.1: Project's timeline.

The first phase of the project encompassed the literature review and processing of the data available. These activities were used to set the foundations for the upcoming steps and had as main goal the strategy formulation.

In the second stage, the project had as goal the consumer behavior modeling, to recreate the decisions based on different delivery options and fees. Thereby, there was a need to define the parameters that characterize the customers and to develop a predictive model to foresee the consumers' choices.

The third stage was the development of a simulation procedure to recreate customer behavior virtually. The objective of this stage was to quickly reproduce the impact of different delivery options, reducing the uncertainty of a new option when deploying on real-environment.

When testing the impact of a delivery plan, it is not only important to assess the effect on consumers' choices but also on the efficiency of the delivery operations. Thus, in the fourth phase, a procedure to estimate the impact over the fulfillment operations was developed.

The final stage encompassed the development of scenarios to test the developed methodology and produce an overview of how the retailer could improve its e-commerce operations.

1.3 Dissertation structure

This thesis is divided into six chapters. It is organized with the following structure:

Chapter 1 introduces the project, where the motivation is contextualized, the problem described and the main objectives overviewed.

Chapter 2 provides a summary of the current state of the art for the most relevant themes to be developed throughout the thesis. First of all, the most relevant time slot management studies were described. Then, the current approaches to consumer behavior modeling were overviewed. To conclude, the forecasting topic is broadly covered

Chapter 3 describes in greater detail the current retailer's context, the limitations of the actual processes and the available data.

Chapter 4 depicts the proposed approach to tackle the time slot management problem. Firstly, it describes the process of modeling the consumers' choices and the integration of a simulation engine. Afterward, the estimation of fulfillment operations is summarised. Finally, the procedures to generate new delivery scenarios is overviewed.

Chapter 5 presents the results of the implementation of the developed methodology and the expected outcome of the generated scenarios.

Chapter 6 draws some conclusions of the project, alongside with future improvement to the work developed and some concerns over the applied methodology.

Chapter 2

Literature Review

This chapter has the goal of introducing to the reader into the theoretical background used as support for the development of this project, related not only with the time slot management but also other topics relevant for the methodology implemented. Section 2.1 introduces the most relevant approaches to tackle time slot management problems. In Section 2.2, the current practices to model consumer behavior are analyzed. In Section 2.3, the theoretical support used to develop the model to predict the consumers' delivery choices is presented. To conclude, in Section 2.4, simulation is presented as one way to recreate e-commerce operations virtually.

2.1 Time slot management

The time slot management problem is commonly faced by e-grocery retailers and consists of finding the optimal combination of available slots and delivery fees to provide to online customers. According to Agatz et al. (2008b), the problem can be seen as a traditional revenue management problem, that is defined as the utilization of analytics and consumer behavior to set the optimal combination of availability and price to maximize revenue. Generally, revenue management problems can be solved using quantity-based and price-based approaches. In the time slot management field, quantity-based approaches are related with the decision of setting which delivery time windows to offer while price-based approaches use delivery fees as a tool to manage demand. A retailer can apply both options at different moments of the process in a static or dynamic way. The static approaches are based on forecasts and manage demand tactically based on a set of conditions. The decisions are valid during the whole booking horizon and are not updated when new customers arrive at the system. Commonly, the retailers use these methods to establish their operations over a medium-term horizon. The dynamic problems manage demand on real-time, this is demand management policies are updated as soon as new information is available. Usually, dynamic approaches are suitable for short-term planning. Therefore, the combinations of quantity/price and static/dynamic strategies divide the time slot management into four different categories, as is summarized in the framework of Table 2.1 proposed by Agatz et al. (2008a). Each category will be characterized in the following subsections.

Table 2.1: Framework of Time Slot Management.

	Capacity allocation	**Pricing**
Static	Differentiated slotting	Differentiated pricing
Dynamic	Dynamic slotting	Dynamic pricing

2.1.1 Differentiated slotting

The differentiated slotting is a tactical problem of defining the delivery capacity available. The decision of setting the slots available to the customers encompass four relevant parameters: time slot length, time slot overlap, number of time slots offered and timing (Agatz et al., 2008b).

Time slot length is the difference between the maximum delivery hour and minimum hour promised by the retailer to deliver the purchase. It presents a dilemma between higher customer service or more operational costs to the retailer. Mintel (2016) found out that 32% of online grocery shoppers use this service due to adjusted delivery time windows. However, the improvement in customer service due to narrower delivery slots tends to increase delivery costs.

Time slot overlapping happens when two different time slots have a range of time in common. This parameter, alongside with the number of time slots, might increase the attractiveness of the service, as there are more choices available. However, the combination of different offers also impacts the efficiency of the delivery operations due to greater difficulty in managing more delivery options.

Timing is related to the time of the day that a slot is available and has an important role in the routing efficiency. Delivery vehicles have to visit different zip codes during a single time window. Thus, an area with more customers' density tends to have lower travel time between stops, which highlights the importance of clustering nearby zip codes in the same delivery slot, to minimize the distance between deliveries.

Agatz et al. (2008a) studied the tactical problem of selecting which time slots to offer in each zip code of a delivery region. The goal of their work was to maximize the service level while minimizing delivery costs. To develop this research, two crucial assumptions were established: the demand in each zip code is independent of the offered time slots and the demand is evenly distributed between the offered time slots. To estimate the costs of a time slot, the delivery routes were estimated using two approaches: the continuous approximation proposed by Daganzo (1987) and an integer programming approach using the seed-based scheme from Fisher and Jaikumar (1981). To select the best slots, the following schedules were tested: the outcome of the continuous approximation and integer programming approaches, the original schedule of a retailer, all time slots available and a single time slot that spans an entire shift. The computational studies unveiled a 25% increase in the delivery costs when two-hour slots were offered instead of slots that span an entire morning or period. Moreover, the geographical differentiation for the offered slots allowed up to 10% savings in delivery costs when compared with a full range of time slots offering.

Hernandez et al. (2017) also solved a tactical time slot problem that consisted of selecting a particular delivery shift schedule for each zone. The day was divided in full day, two or three shifts and the schedules were a combination of weekdays and day shifts. This work developed the methodology of Agatz et al. (2008a) by explicitly solving the vehicle routing problem using two different heuristics. The first one starts by solving the periodic vehicle routing problem to define shift schedules. Next, a greedy heuristic is used to create a schedule for the day, connecting the last zone of a schedule and the first one of the following one. To conclude, the routes of each individual day are optimized using the unified tabu search. The second heuristic solves the whole periodic vehicle routing problem with time windows, without decomposing it as the first heuristics.

Cleophas and Ehmke (2014) solved a problem more focused in managing the demand than Agatz et al. (2008a) and Hernandez et al. (2017), that were more focused in managing the fulfillment operations. The concepts of vehicle routing and demand fulfillment were used in an integrated way to maximize revenues by offering desirable time windows to highly profitable consumers, in a limited transport capacity environment. Initially, the transport capacity and cost-minimizing routes were computed to be used in the process of orders declining. This process was based on expected requests that were approximated by the exponential smoothing method. The number of requests to accept from each basket value range is estimated using the Expected Marginal Revenue heuristics, whose aim is to find protection levels for higher basket values and booking limits on lower basket values. The simulation used to evaluate the impact of value-based demand fulfillment found out that the acceptance limit based on demand fulfillment controls might increase the revenues up to 24%. However, this approach has the downside of only considering the short-term value of the order and not the future value of a customer.

2.1.2 Differentiated pricing

Pricing is used to discriminate and differentiate the different products and services. In time slot management, pricing is the assignment of a fee to a delivery time window. Consequently, the purpose of differentiated pricing is to define the price of a delivery time slot statically. Agatz et al. (2008a) define region-based, time-based, size based and service-based pricing as approaches to manage delivery fees.

Region-based is related with defining the price a region should pay for the service, based on the distance from the warehouse or the density of customers. However, Balkan (2008) alerts for the possibility of customers disappointment when targeted with a higher price over their area. Time-based pricing uses price to smooth demand across the periods, avoiding the most popular hours of becoming too crowded. Commonly, the main challenge lies in identifying which time slots should have a discount or a surplus. Service-based pricing expands time-based pricing not only by the differentiation between periods but also between slots with different lengths. Moreover, size-based pricing uses discounts or surpluses in the delivery fees as tools to promote the value of the sales. Currently, Amazon offers the delivery service when a given target price is exceeded, while Tesco charges an extra fee if a minimum basket value size is not achieved.

To the date of this thesis, Klein et al. (2017) is the only work to explore the differentiated pricing field and used the work of Agatz et al. (2008a) as a benchmark. The main goal was to tactically set the price for the available time windows in different zip codes. Even though the goal of the study was to maximize the profit, the purpose of the price was to influence customers to choose slots that minimize the delivery costs instead of generating extra profits. When making pricing decisions, the operational routing costs were considered by incorporating a model based approximation of the vehicle routing problem in the optimization model. Throughout the computational studies, best results were achieved for the approaches that follow a differentiated time slot pricing approach, due to its effectiveness to generate cost-effective schedules.

2.1.3 Dynamic slotting

Dynamic slotting is the decision-making of setting in real-time which slots are available for each customer. This decision can be influenced by capacity constraints, such as picking capacity in the warehouse, available driving time and physical fleet size (Campbell and Savelsbergh, 2005) or according to profitability concerns. Agatz et al. (2008b) pointed out that using first-come-first-served policies does not maximize the expected profit. Thus, it is needed to solve the dilemma of accepting orders based on its expected profit or declining them and wait for more profitable customers.

The paper from Campbell and Savelsbergh (2005) was one of the leading studies to develop work in the time slot management field. The scope of the work was to decide which deliveries to accept and to reject, as well as the time slot to offer to maximize expected profits. The decision was based on an insertion heuristics, that considers all the accepted customers and evaluates whether the delivery under evaluation can be inserted in one of the considered routes. If the algorithm found a feasible delivery route the order was accepted, otherwise it was rejected. Moreover, the future demand was approximated estimating the probability of a customer placing a request in the future. A simulation study where the supply and demand vary to test different scenarios was conducted to generate a stream of delivery requests and evaluate the applicability of the developed strategy. The main outcome of the study was that evaluating the feasibility of a delivery in a slot increases the profitability and reduces the risk of delivery failure, compared with only limiting the number of deliveries in a time slot.

Ehmke and Campbell (2014) extended the study of Campbell and Savelsbergh (2005) by considering time-dependent and stochastic travel times in congested areas. The goal of the work was to maximize the number of accepted requests while assuring feasible routes. The set of offered slots was known a prior, but bookings acceptance was performed in real-time. Thereby, the acceptance of orders until achieving the slot capacity was compared with a dynamic approach where the solution for a vehicle routing problem was estimated. Throughout the work, it was corroborated that acceptance mechanisms that consider travel time outperform non-time dependent mechanisms.

Mackert (2019) dynamically changed the slots available to influence customers' choices. The goal of the study was to maximize the expected profit of the delivery schedule, by offering a set of delivery option based on the requests' generated profit, expected consumer behavior and opportunity cost. The customer choice behavior was approximated using a Generalized Attraction Model. The opportunity cost of each order was computed by the estimation of future orders that might not be accepted due to accepting the one under analysis. The approximation of the opportunity cost jointly considered the impact on future available slots and delivery routes. A delivery booking was accepted when the generated profit before the delivery overcompensated the opportunity cost. This approach generated 7% more profit than myopic opportunity cost approximation and 4% more than an estimation based on insertions heuristics.

2.1.4 Dynamic pricing

Dynamic pricing is based on providing different delivery fees to different customers in real-time. This option provides a more powerful tool than just accepting or rejecting a booking. Pricing can incentive customers to select particular delivery options instead of just prevent them from choosing some options with dynamic slotting (Agatz et al., 2008b). This strategy might have the downside of customers perceiving differentiated pricing as unfair (Kimes and Wirtz, 2003). However, some researchers point out that the acceptance of this kind of policies is a matter of habituation (Wirtz and Kimes, 2007).

Campbell and Savelsbergh (2006) studied how to influence consumer behavior using price incentives while decreasing routing costs. The incentives were only offered to time slots with chances of being selected and could not be offered to all time slots since the increase on the choice probability of a slot was the result of the reduction in the probability of other slots. To test the incentives scheme, two optimization models were developed, one for one-hour length slots and others that considered wider time slots. The feasibility of serving an order in a period was computed using insertion heuristics, that tries to insert the request in the scheduled routes. Thus, it was concluded that it was enough to provide incentives at most to three delivery slots to reduce delivery costs and enhance profits. Moreover, it was stated that using incentive schemes to accept wider time slots is easier than to encourage the picking of specific time slots.

Asdemir et al. (2009) developed a Markov decision-process-based pricing model that adjusted delivery prices for different choices of the customers. Thereby, the developed model recognized the need for smoothing the demand over the delivery capacity and setting the prices so that less popular options should be filled earlier than popular options. The work recognized that the optimal combination of prices makes the e-grocer indifferent between delivering in any available time window. Moreover, it stated that in a dynamic environment declining a request from a customer can be misunderstood and affect the long term sustainability of the business since a lost delivery also represents a lost sale.

The work developed in Campbell and Savelsbergh (2006) was extended in Yang et al. (2014) by incorporating stochastic information about future requests. The expected delivery costs were estimated not only considering the up-to-date accepted orders but also the forecast of upcoming orders. When deciding the price and availability of each slot, consumer behavior was also considered through the multinomial logit model. It was found out that dynamic pricing without considering future demand can produce worse results than using static price policies. Even though this work sets the baseline for some of the most recent studies in this field, some downsides are verified. The schedules only considered the historical data not including already accepted orders over the period to produce updated forecasts. Thus, the already accepted orders and the upcoming orders were not directly linked.

Yang and Strauss (2017) developed a dynamic pricing approach to set the delivery fees per area. The delivery costs were estimated by clustering the problem in area-specific problems and used a continuous approximation of the total traveling distance. The main contribution to the existing literature lies on the incorporation of expected order displacement costs besides the delivery cost into the opportunity cost. The proposed method opposes the view of Yang et al. (2014) by stimulating the demand in low-demand areas. This strategy avoids the higher charges in these areas to promote demand growth.

The pricing policy developed in Yang et al. (2014) sets the baseline for the work of Klein et al. (2018). A mixed-integer linear programming approach was used to approximate the opportunity cost and the forecasted demand used the most recent available data. Moreover, the delivery costs updated the parameters of the seed-based scheme using the locations of the already accepted customers as input. The developed approach can consider the lost profit due to limit delivery capacity.

2.2 Consumer behavior

In time slot management, when testing different sets of available slots, it is essential to understand the impact on customers' choices. Thereby, it is needed to model consumer behavior to understand how is the decision-making driven. Throughout the existing literature, it is possible to identify some authors that completely ignore the customer choice behavior and others that use simple assumptions or more sophisticated models.

The works that use simpler approaches set the chances of a customer choosing each delivery option *a priori*, using standard assumptions about customer behavior. Campbell and Savelsbergh (2005) matched a profile to the customers and identified which time slots were acceptable. Cleophas and Ehmke (2014) used a similar approach, considering known probabilities of a slot being chosen depending on customers' population segment or delivery area. Campbell and Savelsbergh (2006) knew beforehand the probability of each customer choosing a delivery time slot. However, a dynamic approach was used to capture the effect of price incentives. The probability of slot being chosen increased by a given rate when an incentive was assigned and the probability of other slots decreased in equal proportion. ´

The researches that use more sophisticated models resemble each other by the utilization of discrete choice models. Asdemir et al. (2009), Yang et al. (2014), Yang and Strauss (2017), and Klein et al. (2018) used the Multinomial Logit Model. On the other hand, Mackert (2019) used a Generalized Attraction Model and Klein et al. (2017) used a General Non-Parametric Rank-based Choice Model. The aforementioned models will be addressed in the following section.

2.2.1 Discrete choice models

Discrete choice models explain people's choice among a set of alternatives mathematically and are based on the random utility theory. The random utility theory was developed based on the probabilistic choice theory (Bierlaire, 1998).

When people face repeated choices between two different alternatives, often they choose one option and in some instances, they choose the other one, as pointed out by Tversky (1969). Therefore, deterministic models are not suitable to represent the decision process, leading to the development of models whose output is the probability of different choice outcomes. To model the choice, it is defined a level of utility U_j for each alternative $j \in J$. The random utility theory assumes that the individual is rational and selects the alternative with the highest utility. However, the utilities are not known by the analyst and are treated as random variables. The most common representation of utility U_j can be formulated as a linear regression decomposed in a systematic component V_j, involving explanatory variables, and an unobserved component, that can be represented by the random variable ε_j:

$$U_j = V_j + \varepsilon_j \tag{2.1}$$

The Multinomial Logit Model (MNL) applies a transformation to the utilities of the alternatives so they can be interpreted as the probabilities of selecting each one:

$$U_1 = V_1 + \varepsilon_1 \iff P_1 = \frac{e^{U_1}}{e^{U_1} + e^{U_2} + e^{U_3}}; \tag{2.2}$$

$$U_2 = V_2 + \varepsilon_1; \iff P_2 = \frac{e^{U_2}}{e^{U_1} + e^{U_2} + e^{U_3}}; \tag{2.3}$$

$$U_3 = V_3 + \varepsilon_1 \iff P_3 = \frac{e^{U_3}}{e^{U_1} + e^{U_2} + e^{U_3}} \tag{2.4}$$

The MNL assumes that the errors are independent and identically distributed, following a Gumbel distribution. Moreover, the model relies on the strong assumption of the Independence of Irrelevant Attributes (IIA). This property can be stated as: "The ratio of the probabilities of any two alternatives is independent of the choice set" (Ben-Akiva and Bierlaire, 1999). Thereby, the choice probabilities ratio of two alternatives is independent of the presence of any other alternatives, or, in other words, the choice probabilities of options A and B do not depend on the set that contains A and B.

The limitation of the IIA assumption is exposed to situations with alternatives with correlated random utilities. To circumvent the IIA property, the Nested Logit Model (NL) can be used instead (Smith et al., 2018). The NL model differs from the MNL by grouping alternatives in subsets where the alternatives exhibit a higher degree of similarity between the set than alternatives in different groups (Ben-Akiva, 1973). Moreover, the NL has a hierarchical structure with as many levels as there are decisions in the developed model. Although this statistical representation uses more degrees to represent better the data, it does not represent the individual's decision process (Koppelman and Bhat, 2006).

The Generalized Attraction Model (GAM) proposed by Gallego et al. (2015) corrects the MNL that overestimates the choice probability of a slot by incorporating the probability of non-available slots (Mackert, 2019). The GAM is an extension of the Basic Attraction Model (BAM), that attaches a measure of attractiveness to the different alternatives to model the choice probabilities. However, the GAM also uses a measure of dissatisfaction if an alternative is not available. Therefore, a high level of dissatisfaction due to narrower delivery option or the nonexistence of the preferred slots might lead to the cancellation of the booking process. The BAM and the MNL are similar models, but the measures of attractiveness used in the BAM are embodied as exponential utilities in the MNL.

The General Non-Parametric Rank-based Choice Model used in Klein et al. (2017) is based on the assumption that customers choose a slot according to a preference list. Moreover, it is compatible with the different random utility models, such as the MNL. The preference list establishes a ranking based on the combination of slot and price. Moreover, if the combinations available are not suitable for customers' preferences, they may prefer to cancel the booking. It is expected that the customers choose the option with their highest level of preference.

2.3 Forecasting models

The previous section presented the procedures used in the current literature to model consumer behavior when selecting a delivery slot. This section explores other forecasting models that could be used instead to predict the choice of a consumer when facing a set of alternatives.

Forecasting involves the analysis of historical data to capture patterns and trends to predict future events. The ability to forecast is highly influenced by the level of understanding of the factors that affect the event to forecast, how much data is available and if the forecast is affected by its predictions (Hyndman and Athanasopoulos, 2018). Gonçalves (2000) breaks down forecasting methods in qualitative and quantitative methods.

Qualitative methods use subjective analysis and perceptions to conduct the forecast and may or may not take the historical data into account. Long-term forecasts or situations where available data does not allow the utilization of statistical methods or mathematical models are more prone to the utilization of this method. The expert's opinion, delphi method and market studies are examples of qualitative models.

Quantitative methods are used in situations where data has quality enough to extract insights and transform them into knowledge, to perform predictions. These methods are divided into time series methods and causal or regression models. Time series models rely entirely on historical data by identifying patterns, seasonality and trends verified in the past. On the other hand, causal methods express mathematically cause and effect relationships between explanatory variables and the event to predict. It is the most sophisticated kind of forecasting tool and may incorporate time series analysis (Van Ryzin and Vulcano, 2015).

Causal and regression models encompass a wide range of different methods from simpler models such as simple linear regressions to more sophisticated machine learning algorithms. Simple linear regressions are the simplest causal and regression models used in cases of strong linear relationships between two variables. However, this linear relationship can be extended to three or more variables resulting in multiple linear regression or non-linear regressions when the relationship is better explained through a non-linear mathematical function. In cases where the inclusion of causal information increases the prediction accuracy, supervised learning techniques are also suitable. Supervised learning techniques use different algorithms to perform a mapping of different inputs to an output.

The supervised learning algorithms are commonly divided into single models and ensembles. As suggested by the name, single models use an individual procedure that is trained to forecast the output, while ensembles combine several single models to produce a more powerful and flexible tool. The ensembles are classified as homogeneous and heterogeneous. The first ones use the same algorithm to produce the predictive models, while the latter uses models generated from different algorithms, as is the example of stacking. Homogeneous ensembles are subdivided in parallel ensembles, such as bagging, and sequential ensembles, like boosting.

Decision trees (DT) are an example of singles models and represent the decision-making process graphically through a tree. It can be used as a non-parametric effective machine learning model technique for regression and classification problems through the CART algorithm, developed by Breiman (2017). The decision tree can be divided into root nodes and leaf nodes. The first one represents an input variable and a split point on that variable, while the latter is an output variable used to make a prediction. Moreover, the selection of which inputs to use and the split points are chosen using greedy heuristics that minimizes a cost function. The tree construction ends using a predefined stopping criteria and the final result produces a set of mutually exclusive rules. The downsides of DT are the inability to capture relations that could have been captured with a continuous function, the risk of overfitting and capture noise from the data set. Overfitting is a modeling error that occurs when the model is too fitted to the training data but unable to predict future occurrences. It might be sorted out by limiting the length between the root and the leaf or increasing the minimum number of observations used to branch the data set. Besides that, Breiman (1996) circumvented this limitation through the utilization of bagging.

Bagging is a parallel ensemble learning that divides a standard data set in smaller and independent training sets to produce distinct predictive models. When presented with new data, the predicted score or category is assigned using majority or weighted voting. Random Forests (RF), introduced by Breiman (1996), is an ensemble learning method for classification and regression that creates a forest (of decision trees) in a random way trained by the bagging method. Moreover, in RF algorithms, each node is only allowed to split over a random subset of available predictors. Breiman (2001) proved that this randomness introduced by selecting the best partitioning decision from a random subset of features reduces the variance and bias of the bagging algorithm, even deteriorating the performance of the single model.

Boosting is an ensemble learning technique that uses the complete training set to train a series of single models sequentially, building models that correct the misclassifications of the previous ones. Each model has a voting right, according to its accuracy to score new observations, on the validation set. One of the most popular boosting methods is Gradient Boosting Machines (GBM) that is similar to RF. However, the trees in GBM are built in a gradual, additive and sequential manner, instead of in parallel as the case of RF.

To conclude, it is necessary to highlight that the models trained to predict the customers' choices in the current project are GBM and RF.

2.4 Simulation

To conclude the literature review chapter, this section intends to induce the reader on the topic of simulation, to be used on the proposed approach to recreate the e-commerce environment virtually. Simulation is described as the process of designing a model of a real system and conducting experiments, to evaluate the outcome of various strategies or understand the behavior of a system and its operations (Shannon, 1975). The development of this method and its growing utilization has emerged from the need to tackle complex real-world problems quickly and in an affordable way. Testing new solutions might be too expensive or changing the real-environment impossible, which leads to the necessity of developing a virtual model to find the right solutions (Grigoryev, 2015).

The development of a simulation model allows to test a significant amount of scenarios and analyze in high-detail the performance and impacts of each alternative (Swaminathan et al., 1998). Even though the power and advantages of a simulation model, a simulation project leads to various challenges and efforts. The deployment of a simulation tool is time-consuming, requires knowledge of the behavior to model and needs a proper validation and verification of the models (Hillier and Lieberman, 2001). Moreover, there is a risk of over-customization as a consequence of trying the perfect fit for the real-environment, which can imply low flexibility and no re-usability.

As every project, the deployment of a simulation model also requires a well-structured development plan and the alignment of all parties involved. Savory and Mackulak (1994) decomposed the process in five phases: model formulation and creation, coding of the simulation model, execution of the simulation model, summary of results, output analysis and interpretation.

The evolution of the simulation models and the emergence of new fields using this tool led to its diversification. Three main approaches were developed: system dynamics (SD), agent-based modeling (ABM) and discrete-event modeling (DEM). SD is more focused on long-term strategic decisions with higher abstraction level, while ABM and DEM are more used in tactical and operational decisions with medium to low abstraction level.

SD represents a general picture of a complex system ignoring finer details, such as events or specific characteristics. As described by Grigoryev (2015), this approach is widely used in situations where the properties of the whole cannot be found by connecting the individuals' properties. Moreover, it allows the analysis of systems' states and the rate at which states change.

ABM models follow a bottom-up approach, where the interaction between entities and their autonomous behavior is used to provide insights about the overall system (Bonabeau, 2002). This method allows the developer to gather insights and characterize the behavior of each entity but has a limited capacity of defining the system's states and flows. The simulation developed in this thesis is considered to be on the ABM scope.

Contrary to ABM, DEM follows a top-down approach where the collection of events impact the system instead of the entities individual behavior. The model can be represented as a flow of events and processes that trigger other operations. Therefore, it is mainly used in problems where the different states and changes are known beforehand.

Chapter 3

Problem description

In the e-commerce environment, there is a trade-off between customer satisfaction and efficient delivery operations. From the customer perspective, delivery slots adjusted to its schedules promote the shopping experience due to more convenience in the delivery process. However, from the e-tailer point of view, it represents more pressure over the fulfillment operations, with a higher risk of missing the promised delivery hour and higher delivery costs. Therefore, the retailer seeks to balance these downsides with an attractive offer to customers that prevents the loss of purchases.

The problem acquires higher importance when considering that e-commerce is a growing business and the retailers face it as an opportunity to assure their long-term sustainability. Thereby, building a strong market position since the emergence of the service is crucial to ensure wide market penetration in the future. However, the development of the e-commerce operations needs to be aligned with a proper design of the operations to assure long-term sustainability in order to avoid failure as the case of Webvan. Webvan was an e-grocery company that went bankrupt in 2011. The fast growth and aggressive strategy were pointed out as some of the reasons for the failure (Agatz et al., 2008b).

The retailer evaluates the different delivery designs and available slots on real-environment, which can have a prejudicial impact on the overall efficiency of operations and customer satisfaction. Therefore, it is important to establish a procedure to assess different scenarios and potential changes before putting them in practice.

This chapter intends to assess the current performance of the retailer. In the first section, the current online booking process is addressed. In Section 3.2, the actual process for setting the combination of available slots and price is described, as well as its limitations. Afterward, in Section 3.3 a preliminary analysis of the operations and consumer preferences is conducted based on the data available.

3.1 Booking process

The e-tailer considered in this study conducts its e-commerce transactions through its website. Firstly, the customers willing to make an order have to be registered and logged in the online platform. Secondly, they have to select the desired products from an assortment similar to the physical stores. After selecting all the products, the customers proceed to the checkout page. When in the checkout area, the customers have to choose the delivery option: home delivery or pick up in store. Focusing on the first option, the customers must have a delivery address assigned to their account. However, more than one address might be associated, so that customers have to decide the location where they wish to receive the package. Afterward, the customers should select a delivery time slot from the set of available options or leave the website and give up on the booking process. Currently, the available slots have lengths that range from 2 hours to 4.5 hours in a delivery date up to one month away. To conclude the request, the customers have to pay the order value and a delivery fee to receive the purchased items at home, which varies between slots and days. The customers might also decide to buy a quarterly or yearly subscription, making all the deliveries free, for requests that are worth more than an established value. Concerning the pick-up in store, the customers have to select a physical store of the retailer where to collect the order and an admissible time window over the opening hours.

The delivery slots available to the customers are constrained to three parameters: booking time, delivery area and delivery capacity.

Booking time Each delivery slot can only be selected before a previously defined hour. After that hour, known as cut-off time, it is not possible to book the slot, as exemplified in Figure 3.1.

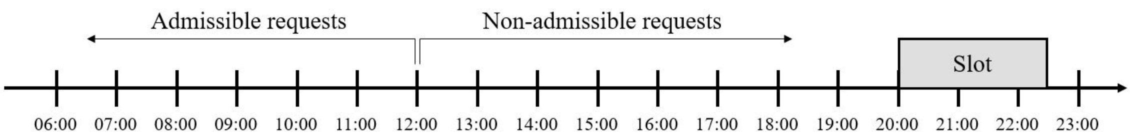

Figure 3.1: Illustrative example of cut-off hour for a slot.

Delivery area The retailer clusters zip codes in the same area. The available slots, length and delivery fees vary across those areas. Moreover, high demand zones have a wider set of options when compared with low demand areas. An illustrative example is that a customer living in a high-demand area might choose up to 10 different slots, but a customer based in a low-demand area only has a set of 4 available slots.

Delivery capacity Each slot in each delivery area has a fixed capacity. Thereby the number of orders that can be accepted in that slot is limited. When the limit on the number of orders is reached, the slot becomes unavailable and no further deliveries can be placed.

3.2 Retailer internal process

The current process of defining the combination of slots and price is conducted by an internal team of the retailer with three main goals: business growth, customer retention and learning.

For the company, e-commerce business has long-term potential and there is a great motivation in developing a solid structure to assure the success and sustainability of the operations. Thereby, the team is focused on offering attractive delivery choices to incentive the customers to try the service.

The business growth includes not only increasing the number of customers trying the service but also the number of customers that use it continuously. The number of active customers and the periodicity of their orders are key metrics to evaluate the success of the service. Therefore, the company tries to promote the connection of the customers to the service. The design of an attractive delivery policy and an effective subscription plan are some of the implemented strategies.

With the focus on business growth and customer retention, the retailer wants to learn as much as possible about the service and customers' preferences. Therefore, they analyze in real-environment the customers preferences regarding different delivery options. The variation of available slots and the respective customer behavior are expected to be fundamental to develop the background and analytical support necessary to understand the optimal set of slots and delivery fees. Such experiments are conducted in an *ad-hoc* basis, where no predefined methodology is followed to set the combination of slots and prices. The retailer captures which combinations are more effective and keeps them until feeling the need to revise them.

The retailer uses an internal platform where the internal team sets for each delivery area the available slots, the delivery fees, the maximum number of admissible orders and the time limit for order placement. Moreover, the starting time for the picking operations to the specified slot is also defined. Picking consists in collecting and preparing the items requested by each customer's orders.

The picking operations are carried out in fulfillment stores, which are specialized and strategically located for e-commerce operations. The operators collect the ordered products from the shelves like normal customers and redirect them to the shipment area to be collected by the transportation vans. Currently, the retailer conducts the operations through 12 fulfillment stores distributed over the country and all of them were considered in this project. The transportation service is outsourced. However, the retailer is responsible for solving the vehicle routing problem to set the order of the delivery addresses to visit.

3.2.1 Limitations

An analysis of the retailer's internal processes, previously described, disclosed some limitations of the current e-commerce operations:

No analytical support: Currently the process of defining the set of slots available and the prices is empiric, lacking analytical support to help the decision making. The decisions are based on the learning process of trial-error described and on the *know-how* of the stakeholders. Moreover, the retailer uses the operations of similar businesses as a benchmark.

No risk assessment: A change in the combination of slots available and the respective price might lead to substantial changes in the delivery operations. The impact of the different mixes available should be analyzed not only from the acquisition and retention of customers point of view but also based on the impact on the fulfillment operations. The retailer does not possess any tool or procedure to anticipate or simulate the impact of a change in the mix.

No consumer behavior modeling: The impact of a change on the consumer choices does not have analytical and predictive support. Thereby, the changes are only tested on real-environment. An alteration that was originally expected to have a great impact and acceptance on customers can easily disappoint the expectations. Thus, the effect of any change on customers should be analyzed and measured to avoid losing future purchases.

No long-term planning of operations: The retailer is focused on developing its e-commerce service and firm its position as the main brand in the Portuguese market. However, the operations are being developed to fulfill short-term needs. The retailer is focused on optimal solutions for short-term problems, which does not imply optimal solutions for long-term goals and needs. Therefore, to assure the sustainability of the business, the decision-making process should involve the analysis of the impact of the changes in the growth and future of the service.

No integration between commercial and operational planning: The current processes evidence a lack of integrated planning between customer experience and fulfillment operations. The retailer should plan both sides simultaneously, to maximize synergies and ensure the optimal decision for the overall e-commerce operations.

3.3 Retailer transactional data

The retailer captures information about customers, deliveries and orders as a result of customers' activity on the online platform. This information can be used to improve customer experience and fulfillment operations.

3.3.1 Available data

The data that was used in this work was collected between October of 2016 and October of 2017 and encompassed the following aspects of the retailer e-commerce business:

Geographical division: The retailer identifies the zip codes where the customers can use the e-commerce service and groups them by areas.

Fulfillment stores: The products to be delivered to customers' houses are prepared in the retailer's fulfillment stores and afterward shipped. Each store has a set of delivery regions and they become responsible for fulfilling all the requests.

Orders: The customers' orders are registered with a shipping number, a key that encodes the customer identity, the order date, the delivery zip code, the store responsible for preparing the shipping, the slot chosen by the customer, the delivery day and the price paid for the delivery. The retailer does not collect any information about customers that give up on the booking process when selecting a delivery option.

Ordered products: The products requested in each order are registered with the shipping number, the requested quantity, the unit price, the direct discount value, the value of the discount in the retailer customer card and the picking store.

Product information: The products are organized hierarchically in a structure composed by a commercial division, a business unit, a category, a sub-category, a base unit and a stock keeping unit (SKU) number.

Available slots The available slots in each delivery date present in the recorded dataset are stored with the minimum possible delivery hour, maximum possible delivery hour, the maximum number of deliveries, the area where the deliveries are admissible, the store that prepares the shipments to deliver in the respective slot and the cut-off time.

3.3.2 Descriptive analysis

The analysis of the delivery slots chosen by the customers encompasses two main characteristics: the length of the slots selected and the number of days between the order and the delivery. Figure 3.2 evidences the preference of the customers for narrower slots, as suggested in Mintel (2016). From the historical data available, 96% of the orders were delivered in slots with a length of 2 hours and a half or less. However, it is noteworthy that this result might be consequence not only of customers' preferences but also of the slots made available by the retailer. This Figure also suggests that the deliveries are distributed across the day. The majority of the slots only overlap for at most 30 minutes.

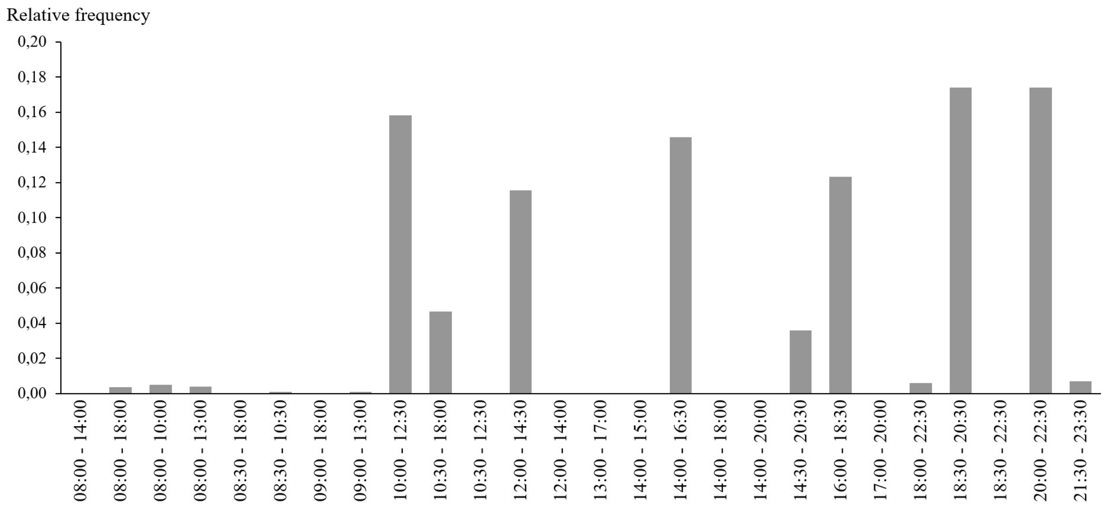

Figure 3.2: Distribution of orders per slot.

Figure 3.3 shows that customers are more prone to book deliveries within a week, and preferably for the next day. Moreover, it suggests that customers are not attracted to book the delivery to the day of the request. However, this phenomenon may be explained by the set of available slots. Firstly, most of the slots have a cut-off time that is before midday. Secondly, delivery capacity is limited. Therefore, the slots might not have enough capacity to accept same-day deliveries.

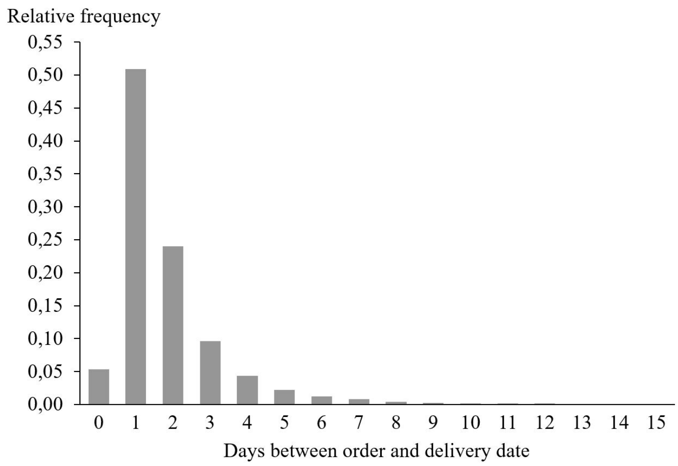

Figure 3.3: Distribution of number of days between order and delivery date.

The analysis of the historical dataset also discloses the existence of two big groups: the group of customers with free-deliveries and the group that paid for the deliveries. The first one represents 46% of the deliveries, while the latter 54%. The paid deliveries range from fees of 4.90 € to 18.90 €, as it is represented in Figure 3.4. However, it is possible to notice a predominance of orders that pay a fee from 8.90 € to 11.90 €. This phenomenon is explained by the set of the available slots having a price that belongs predominantly to this range. Moreover, the predisposition of a customer to pay higher or lower prices is influenced by factors that increase the perception of service quality, such as slot start hour or slot length.

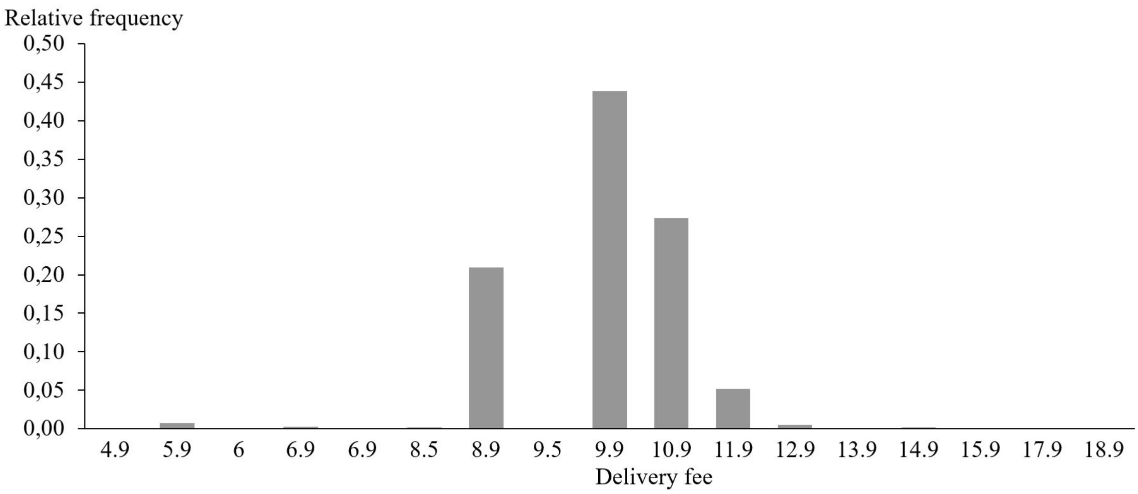

Figure 3.4: Distribution of price paid per delivery.

Chapter 4

Methodology

In this chapter, the methodology developed to overcome the time slot management problem is described. The proposed approach is overviewed in Section 4.1. After the summary presentation of the methodology developed, consumer behavior modeling is depicted in Section 4.2. Afterward, the way that the fulfillment operations are estimated is presented in Section 4.3. Finally, Section 4.4 summarises the procedure to generate scenarios to test the proposed approach.

4.1 The proposed approach

The current project develops knowledge in the time slot management field by using an approach that combines the simulation of consumer behavior and the estimation of fulfillment costs to recreate the retailer's e-commerce operations virtually, as is exemplified in Figure 4.1. Through the application of the developed approach, the retailer can assess the impact of different scenarios. Therefore, it is expected to reduce the uncertainty over the operational results when deploying a new combination of available slots or changes in the charged prices.

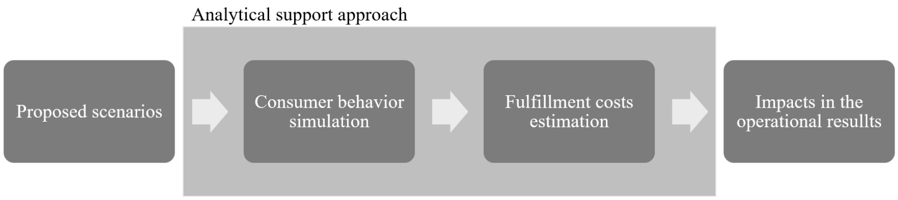

Figure 4.1: Proposed approach.

The first stage in consumer behavior modeling embraces data retrieval and processing to be used in the remaining project. Besides gathering the current combinations offered by the retailer and analyzing its performance, the modeling of the demand is also a crucial phase. This phase consists of modeling key parameters that represent the customer, such as order period or basket value to recreate the demand profile.

The second major step in consumer behavior modeling is the development of a machine learning model to predict consumers' choices. The consumer behavior is simulated using a predictive model that estimates the selected slot for each customer. The choices of the customers are used to create the expected service for each slot and fulfillment store. With a model trained to capture the effect of different combinations, it is possible to test the impact of changing the price and the available slots on the customers' behavior.

Both stages of consumer behavior modeling are used to develop a simulation procedure. The simulation is used to design the e-commerce operation virtually in order to assess the customers' preferences and choices to different available slots and charged prices.

After modeling the customers' behavior and estimating their delivery choices, the following stage encompasses the assessment of the proposed slots on the fulfillment operations. The final recommendation regarding the proposed combination is only conducted after the analysis of the operations performance. To evaluate the fulfillment operation, the transportation costs are estimated by the expected time spent on each stop, the traveling time between customers and the traveling time from and to the depot.

One of the main problems when designing new scenarios and alternatives is to select which are the best combinations to evaluate. The high number of possible prices and available slots that could be tested increases the dimensionality of the problem. Testing all of them would be computationally expensive and time-consuming. Therefore, the fourth and final stage proposes two methods to generate scenarios and test them preliminarily. Each approach uses a different tool to manage demand. The first method seeks to level the demand across the different slots, using the price as a driver. The second approach uses the slots' availability to manage demand.

4.2 Consumer behavior simulation

The current section explains each stage of consumer behavior modeling to be applied in the simulation procedure. Firstly, the process used to estimate the demand profile is presented and, afterward, the different stages of the development of the machine learning model are depicted. The objective of the simulation procedure was to understand the effect of potential changes on the delivery options by the customer point of view.

4.2.1 Input parameters' modelling

The main goal of this stage was to find out how to model each customer's profile. Each customer has its unique characteristics and the input parameters were used to recreate them. The customer was characterized by the weekday and period of the order, basket value, value spent in each product's category (food, frozen products and other products) and geographical location.

The first stage of the input parameters' modeling phase was the removal of outliers and defective information, followed by the fit of statistical distributions to the data. The Kolmogorov-Smirnov (K-S) test was used to evaluate the adjustment of the data points to the statistical distributions under evaluation.

4.2.1.1 Arrivals per day and period

The number of orders per day varies across different weekdays. Therefore, it was required to approximate the demand profile for each day of the week. The demand is approximated by a log-normal distribution fitted to the historical data of the correspondent day of the week. Cleophas and Ehmke (2014) approximated the expected number of requests using exponential smoothing. However, the development of a proper time series model would require a larger data set to capture effects such as the seasonality and trend between years.

The demand per period of the day is estimated using a statistical distribution that captures the proportion of customers booking in different periods, for different weekdays. The fitted distribution is a log-normal or a normal distribution, based on the distribution with better statistical results. The periods were divided considering the slots' cut-off hours and hours with similar demand across the day, as represented in Figure 4.2. The proportion of each period for each weekday was estimated using a fitted log-normal distribution.

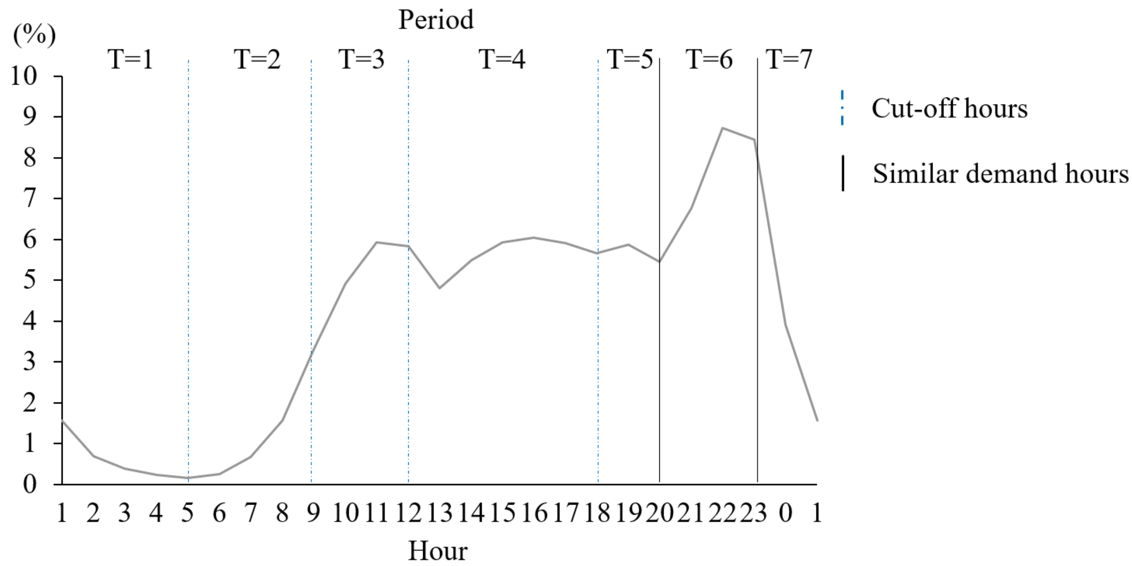

Figure 4.2: Distribution of orders per day period.

4.2.1.2 Demand characterization

The characterization of each customer was crucial to recreate the generated demand. It consisted of determining the basket value, the value spent on each products' category and the geographical location of each customer. The basket value is generated using a log-normal distribution fitted to the historical data, as shown in Figure 4.3. However, the other aforementioned characteristics are estimated using the real distributions, since no statistical test was found to be significant.

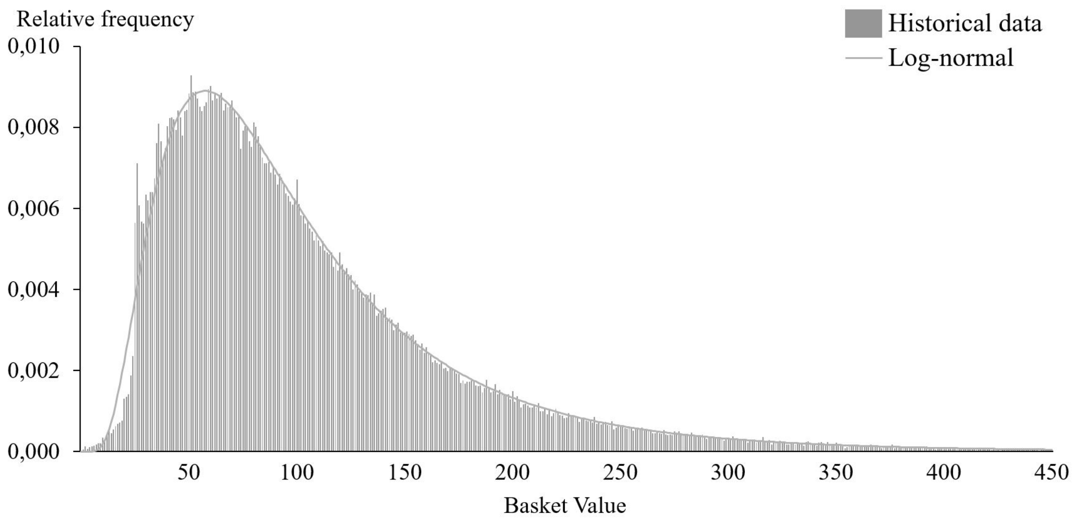

Figure 4.3: Distribution of customers' basket value.

4.2.2 Consumer behavior modeling

The developed model to predict the customers' choices assumes that the decision-making process is based on the comparison of different alternatives. Bettman and Zins (1979) propose that this evaluation can be alternative-based, where the customers evaluate different options one by one. The attributes of one option are assessed before the customer proceeds to the next alternative. Throughout the proposed predictive model, the evaluation of the available slots starts by analyzing the offers with more probability of being chosen and proceeds to the choices with less probability. If a customer finds an option to be suitable for his needs, it is assumed that the customer stops the evaluation process and selects that slot. However, if a customer does not find any suitable alternative, it is assumed that he gives up on the booking process.

4.2.2.1 Predictive procedure

The aforementioned procedure used to represent consumers' choices involves a set of binomial decisions. Those decisions are represented through a binomial model that estimates the outcome based on customer profile and slot characteristics, such as price, relative price, minimum delivery hour, slot length and number of days until delivery. Therefore, a machine learning algorithm was trained to estimate the probability of a customer selecting each option. After predicting the probability of each alternative to be selected, they are sorted by descending order. Finally, until finding the selected slot, a random number is generated. If the random number is lower or equal than the choice probability of the slot under analysis, the slot is selected. Otherwise, the procedure is repeated until finding a suitable slot or revisiting all the possible alternatives. The customer is assumed to give up on the booking process if no suitable alternative is found. The described approach is exemplified in Algorithm 1. The preparation of the binomial model is described in the following subsection.

for *each slot* **do**
    Run Binomial model
    Set Probability: p_i = P(slot s_i is chosen)
end
Order p_i by descending order
Selected slot = False
i = 1
while i < *Number of options* **&** *Selected slot* = *False* **do**
    Generate random number n_i
    if $n_i <= p_i$ **then**
        Selected slot = True
        Choice = s_i
    else
        $i = i + 1$
    end
end

Algorithm 1: Slot choice procedure.

4.2.2.2 Model preparation

The first step in the development of the predictive model involved the preparation of the available data to use it as input for training the model. Afterward, algorithms with different parameters were trained and tested to select the best option. Finally, the selected model was evaluated.

Data preparation

The first stage in the direction of accurately understanding the behavior of a customer when facing a set of potential choices was the identification of potential drivers on his choices. The behavior is expected to be affected not only by the customer's profile but also by the characteristics of the slots available. The variable identification process led to the necessity of gathering the information retrieved from different sources. This process of data gathering and consolidation was fundamental to ensure the quality of the model to be developed.

The information available did not exempt the data transformation and feature selection to improve predictive accuracy. This phase was mainly focused on improving the information about the available slots. Therefore, the following assumption was considered: a customer is always willing to pay a lower price than the price that he paid. Thus, when reconstructing the customer's set of possible choices, the observation of the chosen slot was replicated for the set of possible lower prices. Otherwise, the model could interpret the preference for a more expensive fee for the same slot, just because the slots available had that price assigned more often and would not represent rational behavior. The price was induced as a discrete variable to favor this transformation. Furthermore, this stage also embraced the identification of errors in the data and existing outliers. Those observations were not considered in the process of training the predictive model.

Training and feature selection

As aforementioned, the modeling object of the predictive model was the acceptance of an available slot. Thereby, the observations were mapped in the data set through a binary variable representing the customer's decision since the problem was handled as a binomial classification. However, a multinomial formulation could also be used, where the selected slot would be predicted directly. However, the no booking option could not be modeled since the available data set does not contain any information about the customers' withdrawal, which led to the option of developing a multinomial model being discarded.

Before the machine learning implementation, a decision to be made concerns the model's flexibility and the set of forecasts that the model will be asked to perform. Therefore, two starting options were available:

- Train a single model that could use the slot start hour and length as independent variables to predict the slot probability. Thereby, these two variables could be changed to test the impact on customers' decision.

- Train as many models as available slots in the historical data. However, the slot start hour and length could not be used as independent variables, because for the same slot, those two parameters are constant and therefore would be irrelevant for the probability prediction.

Since training a model instead of training one model for each slot is less computationally expensive, a preliminary assessment of both models was performed. Both approaches have been tested by training a GBM algorithm over the available data set. There was no evidence that training a model per slot had superior predictability accuracy over the single model. Moreover, training a model per slot reduced the flexibility of the scenarios to be tested because testing new slots would not possible. Therefore, it was decided to develop a single model.

Algorithm selection

In order to select a model that performs reliably and satisfyingly, it was needed to ensure that different options were trained to be able to select the best one. Moreover, it was necessary to define metrics to evaluate the quality of the model. Two different metrics were used:

- The accuracy of the slot choice estimated to each customer.

- The distribution of the choices over the available slots.

From the retailer point of view, the main focus was on having a model that can predict the workload of each slot, which may not support high accuracy on forecasting each customer's choices. In an extreme situation, it would be represented by failing the prediction for every customer but having a perfect estimation of how many times each slot was chosen. For testing purposes, the different models were preliminarily evaluated by the accuracy of predicting the selected slot, since the application of this metric speed up the evaluation process. Finally, to validate the chosen model, the slots' distribution was assessed.

To select the best model, two algorithms were tested: RF and GBM. These are ensemble learning methods recognized by their success on improving the prediction accuracy over single models, which motivated its application.

Both algorithms were implemented using the H2O's machine learning cluster (H2O, 2019), through the R studio, and were ubject to a hyper-parameter optimization, as exemplified in Figure 4.1. Hyper-parameters are variables of an algorithm that have to be set before applying a learning algorithm to the training data set. The GBM algorithm had their number of trees, maximum depth, learn rate and column sample rate allowed to vary. The RF algorithm also had their number of trees, maximum depth and column sample rate allowed to change. The number of trees is the number of trees built to produce the ensemble. The maximum depth is the maximum allowed tree size. Commonly, higher values make the model more complex and, therefore, more prone to overfitting. Learn rate specifies the rate at which GBM algorithm learns when training a model, which helps to avoid overfitting, and column sample rate is the rate of sampling the column dependent variable without replacement.

Table 4.1: Hyper-parameters considered for search.

Models Family	**Parameters tuned**
RF	number of trees, maximum depth and column sample
GBM	number of trees, maximum depth, learn rate and column sample

Model evaluation

The evaluation procedure is essential to ensure that the proposed model is reliable. Most of all, it is important to understand what is expected to obtain once the predictive model is applied in a real context. To that end, it was necessary to split the available data into a training data set to train the predictive model and a validation data set to evaluate the model's performance. This division in training and validation had to ensure the equal proportion of the different classes in both data sets.

The model to adopt during the project was selected with the accuracy of the slot choice and based on the hyper-parameter optimization previously described. The models present in Table 4.2 were the result of the hyper-optimization search and the top-performance models were selected according to the accuracy of predicting the slot choice. The predictive model developed from the GBM algorithm was found to be the most suitable model to represent reality. It is an ensemble of 100 trees, with a maximum depth of 20.

Table 4.2: Accuracy of the top-performance models.

	GBM	RF
1	25%	23%
2	19%	17%
3	17%	15%

Applying the predictive model, the accuracy of estimating the chosen slot by the customer was 25%. However, this accuracy represents a lift of 2.5, which implies a good improvement when compared to a random predictor. Lift is a measure of performance that compares the predictive model with a random choice targeting model.

The second stage of the validation of the proposed predictive model involved the comparison of the predicted distribution with the real distribution of the customers across the available slots. The comparison only encompassed the slots that the retailer had available during the development of this project. As exemplified, in Figure 4.4, the model captures the most popular slots and the less popular ones. The overall distribution has a mean absolute percentage error (MAPE) of 19%. Considering the complexity of the predictive problem, it was possible to assume that the model was an acceptable approximation of the real scenario. MAPE was computed by the following expression, where A_t represents the actual value and F_t denotes the forecast value:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

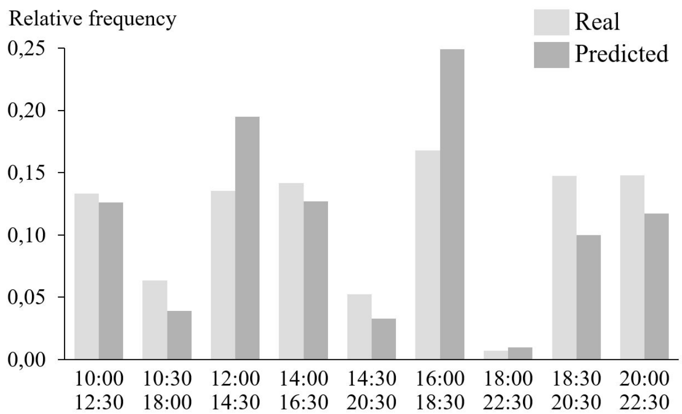

Figure 4.4: Distribution of customers between available slots.

4.2.3 Simulation

One of the benefits with the proposed approach is the integration of the machine learning model in a simulation system, thus recreating the e-commerce operations virtually. The simulation is run on Anylogic®, which is a software 100% coded in Java. This software was chosen due to its simple implementation of the machine learning model, that is also coded in Java, which speeds up the simulation process.

The simulation procedure developed encompasses the following stages: available slots setting, demand generation, customers' choice prediction, slots' availability update and final results report.

The first stage of the simulation process consists of setting the available slots. It is necessary to define for each fulfillment store and geographical area the open slots and the delivery fees. Moreover, the maximum number of admissible deliveries should also be established. This stage intends to recreate the process followed by the retailer's e-commerce internal team.

The second phase is the demand generation and it can be done with two different approaches. The first one uses the statistical distributions mentioned in the current section, while the second one allows the user to explicitly specify the characteristics of the customers with purchases over the weeks to be tested.

The third stage is the procedure of estimating the choice of each customer. This process occurs customer-by-customer since the slot selected for a customer can influence the choice of the following customer. The slots' availability is limited. Thus, it was necessary to model the system in a way that when the maximum capacity of a slot was reached, the option would become unavailable. Otherwise, if the customers' choices were foreseen at the same time, it could lead to overcapacity which would compromise the fit to the real-scenario.

Finally, the results of the simulated period are exported to allow the analysis of the scenarios to be tested. The following section will look in detail into the process of estimating the fulfillment costs as a result of the generated demand.

4.3 Fulfillment costs estimation

When analyzing the effect of different combinations of slots available, it is important to understand the impact not only on the consumers' choices but also on the fulfillment operations. The main costs are the delivery operations, which led to the analytical assessment of the changes in those costs.

To evaluate the delivery costs, some authors solve the vehicle routing problem explicitly to generate admissible routes and foresee the delivery costs, while others use approximations of the delivery costs. Throughout this project, the impact on fulfillment operations is estimated by considering three components: traveling time from fulfillment store to the first stop and last stop, traveling time between consecutive stops and time spent in the stops.

4.3.1 Travelling time between fulfillment store and first or last stop

The traveling time needed to arrive at the first stop or to return to the depot is dependent on the areas where the demand is generated. Regions, where the fulfillment store is outside the main counties or with a higher covered area, might have more unpredictable traveling times. However, a fitted distribution for all depots is used to estimate the traveling time, as shown in Figure 4.5.

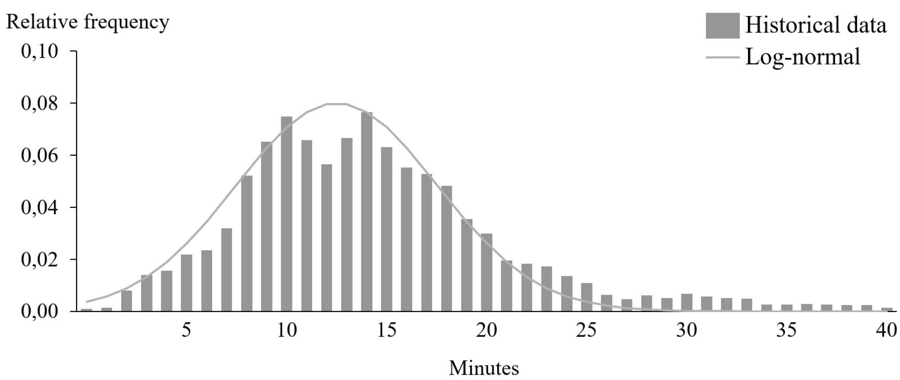

Figure 4.5: Travelling time between fulfillment store and first or last stop.

4.3.2 Travelling time between stops

The traveling time between consecutive customers is highly dependent on the density of the customers over the area where the operations are developed. Thus, areas with a high density of active customers are expected to have a low traveling time due to the low distance between them, while the opposite should be verified in low-density areas. Therefore, a regression where the traveling time is related to the density of customers with orders for the same slot was developed. The area considered was the area of the counties where the demand was generated.

The fitted distribution is represented in Figure 4.6. The R^2 of the fitted regression is 80%, which evidences a good correlation between the dependent and the independent variable. However, factors such as expected traffic or period of the day could be accounted for in the regression to capture the relation more effectively.

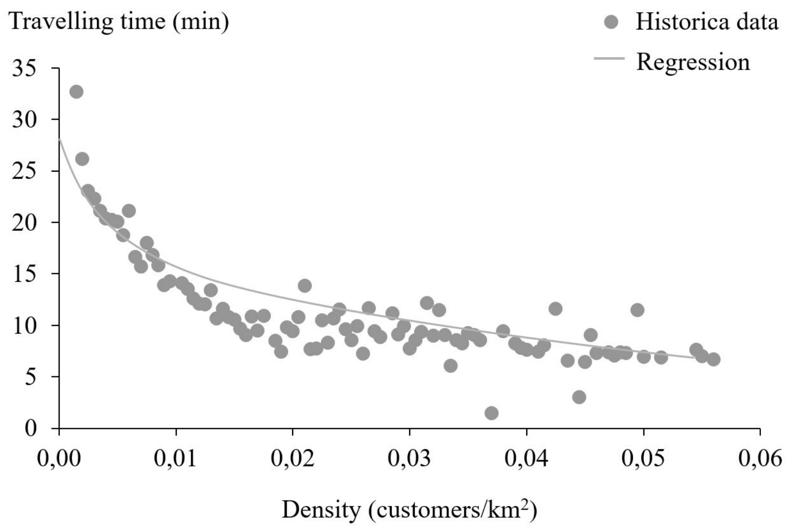

Figure 4.6: Travelling time between consecutive customers.

4.3.3 Time spent at the stops

The time spent at each stop can also impact the number of possible deliveries by slot. The delivery task can be affected by diverse factors, such as the non-attendance of the customer at home, the type of the building or house, the number of boxes and the type of products to deliver. However, considering all these events is a complex task, and a normal distribution can represent the time needed to complete all the tasks, as shown in Figure 4.7

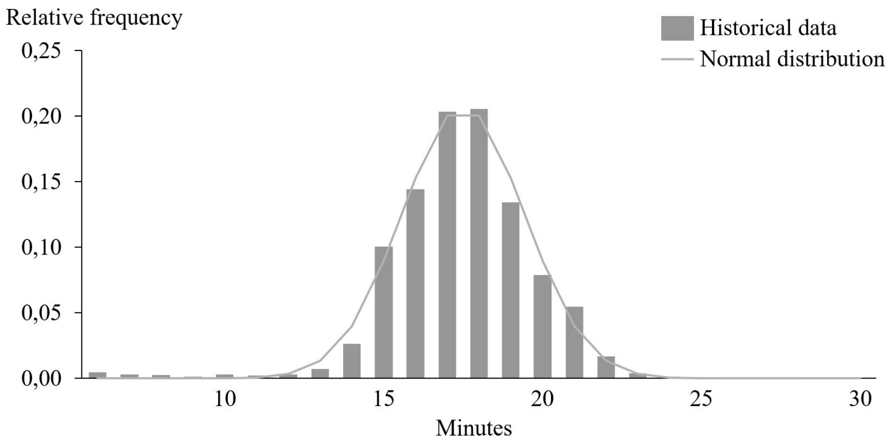

Figure 4.7: Time spent at each stop.

4.4 Prescription

Until this section, it has been described the approach to evaluate different scenarios. However, the procedure is unable to produce by itself recommendations to the retailer. Thus, this section intends to generate admissible scenarios to test the impact on the retailer's e-commerce operations.

4.4.1 Differentiated pricing approach

In time slot management, one of the main goals of the retailers is to balance the demand across the available slots. To achieve this objective, a differentiated pricing approach is utilized which is based on a time-based policy, as proposed by Agatz et al. (2008a). Through the utilization of discounts and surpluses, it is intended to balance the demand not only between day periods but also between slots with different lengths.

As aforementioned, one of the main challenges when testing a new scenario is to define how to set the combinations to test. Thereby, it was performed a sensitivity analysis to the predictive model to capture how the probability of a slot being chosen varies with its price. These probabilities were used in an optimization model to set which fee to assign to each slot.

From an optimization perspective, the problem emerges as a linear optimization problem. Consider:

Sets:

S: the set of available slots

F_s*:* the set of admissible fees for slot s

Parameters:

p_{fs}*:* probability of the slot $s \in$ S with the delivery fee $f \in F_s$ being chosen

a: the average price for the set of available slots

Variables:

$$X_{fs} = \begin{cases} 1 & \text{if delivery fee } f \text{ was assigned to slot } s. \\ 0 & \text{otherwise.} \end{cases}$$

M: auxiliary variable to balance the probability of different slots being chosen

Formulation:

$$\text{Minimize } M \tag{4.1}$$

$$\sum_{f \in F_s} p_{fs} \cdot X_{fs} - \frac{1}{|S|} \cdot \sum_{f \in F_s} \sum_{s \in S} p_{fs} \cdot X_{fs} \leq M \qquad , \forall s \in S \tag{4.2}$$

$$\sum_{f \in F_s} p_{fs} \cdot X_{fs} - \frac{1}{|S|} \cdot \sum_{f \in F_s} \sum_{s \in S} p_{fs} \cdot X_{fs} \geq -M \qquad , \forall s \in S \tag{4.3}$$

$$\sum_{f \in F_s} X_{fs} = 1 \qquad , \forall s \in S \tag{4.4}$$

$$\frac{1}{|S|} \cdot \sum_{f \in F_s} \sum_{s \in S} f \cdot X_{fs} = a \tag{4.5}$$

The objective of the optimization problem is to minimize the difference between the probability of each slot being chosen and the available slots' average probability. Therefore to ensure its linearity, the objective function is modeled by the utilization of the auxiliary variable M (constraint 4.1) and the constraints 4.2 and 4.3. The other set of constraints ensures the design of a feasible set of available slots: each available slot $s \in S$ only can have one delivery fee assigned (constraint 4.4) and the average fee of the available slots should satisfy the retailer's pre-defined target average fee a (constraint 4.5).

4.4.2 Differentiated slotting approach

In the previous subsection, it was intended to use the price as the main tool to manage the demand. However, as proposed by Agatz et al. (2008a), the available capacity can also be utilized to manage demand through quantity-base mechanisms. Thereby, it is proposed a differentiated slotting approach that is used to define which slots should be open in different delivery areas.

One of the main challenges in differentiated slotting is to define a set of available slots that do not prevent the customer acquisition in low-demand areas due to an insufficient set of delivery options while keeping the fulfillment operation efficient. Therefore, when analyzing a combination of available slots, the number of customers that give up on the booking process should be an accountable metric.

To define which slots should be available in each delivery area, it is proposed a heuristics that starts by setting all the slots open, regardless of the localization. Afterward, the simulation procedure is applied to capture the distribution of the customers over the slots. If the average number of customers scheduling their deliveries for a given slot is unable to fill a van, then, that slot is closed. This analysis is performed by delivery area. To promote the demand all over the regions, a set of slots is to be open regardless of the demand level. Through, the application of the algorithm 2, it is possible to define which slots should be open in each delivery area.

The developed algorithm is expected to cluster demand geographically, in order to avoid the dispersion of customers over the delivery slots in the same area. Moreover, the existence of a minimum number of slots open also assure the demand growth, regardless the insufficient number of customers in the past.

```
Set all slots open;
Generate demand;
Predict customers' choices;
for each delivery area a do
    for each slot s do
        if Average number of deliveries in the slot s at the delivery area a <= Number of
          deliveries to fill a van then
            Close slot s at delivery area a;
        else
            Keep slot s open at delivery area a;
        end
    end
    Print open slots at the delivery area a
end
```

Algorithm 2: Procedure to define the open slots

Chapter 5

Results

By the time this project was completed, the proposed approach was yet to be implemented by the retailer. Thereby, all presented results were obtained through the application of the methodology defined in Chapter 4. Section 5.1 shows the expected results of managing the demand through a differentiated pricing approach. Finally, Section 5.2 presents the expected effects of applying a differentiated slotting method.

5.1 Experimental results on differentiated pricing

As shown in Figure 3.2, the historical data unveils that the utilization of the slots is highly unbalanced. Therefore, the objective was to understand the best way to set the slots' price to assure a more balanced offer. This improvement on the demand balancing was expected to increase the efficiency of the fulfillment operations, by smoothing the workload over the delivery periods. The developed optimization model was used as input on the simulation procedure, by generating the scenarios that had to be assessed. Due to the stochastic nature of the input parameters of the developed simulation, each scenario was run 30 times.

5.1.1 Current average price

Using the current average price of the available slots as the baseline, the first analysis had two major outcomes: the diagnosis of the retailer's current performance and finding the best combination of prices to balance the demand over the slots. Thereby, the optimization model described in Subsection 4.4.1 was used to produce a new combination of prices for the available slots. The differences between the slots' prices and the average price for the current policy and the proposed combination are shown in Table 5.1.

Table 5.1: Slots' prices to be tested for the current average price.

Slot	Real scenario	Proposed scenario
10:00 - 12:30	+0.40 €	+0.55 €
10:30 - 18:00	-0.55 €	-0.75 €
12:00 - 14:30	-0.25 €	-0.10 €
14:00 - 16:30	+0.35 €	+0.45 €
14:30 - 20:30	-0.45 €	-0.75 €
16:00 - 18:30	-0.25 €	-0.15 €
18:00 - 22:30	-0.35 €	-0.75 €
18:30 - 20:30	+0.55 €	+0.75 €
20:00 - 22:30	+0.55 €	+0.75 €

The current combination of prices leads to under utilization of wider slots, such as the slots "10:30 - 18:00", "14:30 - 20:30" and "18:00 - 22:30". Through the optimization of the prices policy, it is possible to reduce the standard deviation of the utilization from 4.25% to 1.35%. When balancing the demand, improvement is also evidenced in the reduction of the difference between the utilization of the most popular slot and the less popular slot, from 10.7% to 4.5%. Moreover, the number of consumers using larger slots is expected to increase by 48%. The differences between customers' distribution in the current and proposed scenarios are highlighted in Figure 5.1.

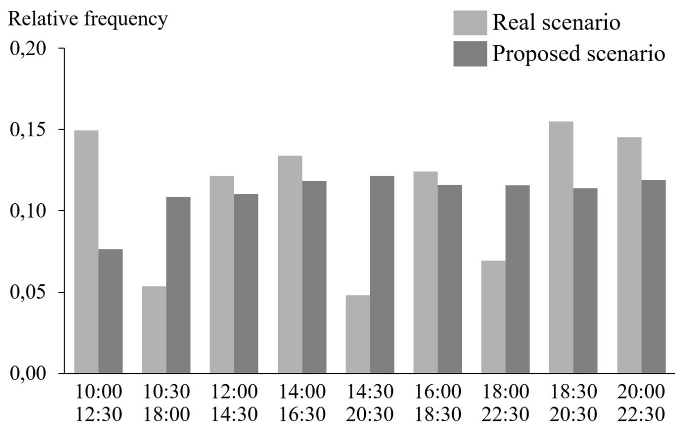

Figure 5.1: Slots' utilization for the current and proposed combinations to the current average price.

An important aspect to bear in mind is that the expected outcome of a new scenario should be analyzed not only from the demand balancing point of view but also from an operational perspective. Thereby, it was relevant to assess the impact on customers' withdrawal and the consequences on the fulfillment costs. The proposed scenario is expected to have no influence on the average number of deliveries per week. However, demand balancing is expected to decrease by 3.6% the delivery costs of the overall e-commerce operation.

5.1.2 New pricing policy

When analyzing the current prices policy and the cost per delivery, it is possible to observe that the fulfillment operations are unprofitable, by approximately 2.15 € per delivery. Motivated by this situation, it was evaluated the impact on increasing the slots' price to reduce the gap between the operational costs and the revenue on delivery fees. The goal of this assessment was not to have profit on the delivery operations but to offset the delivery costs.

To perform the aforementioned analysis, the real scenario and the proposed scenario in subsection 5.1.1 (scenario 1) are compared with two scenarios where the average price is increased by 0.35 € (scenario 2) and 0.85 € (scenario 3). The combination of prices to be tested relative to the scenario's average price were the result of the developed optimization model and are shown in Table 5.2.

Table 5.2: Slots' prices to be tested.

Slot	Current price	Scenario 1	Scenario 2	Scenario 3
10:00 - 12:30	+0.40 €	+0.55 €	+0.50 €	+0.60 €
10:30 - 18:00	-0.55 €	-0.75 €	-0.70 €	-0.70 €
12:00 - 14:30	-0.25 €	-0.10 €	-0.10 €	-0.10 €
14:00 - 16:30	+0.35 €	+0.45 €	+0.20 €	+0.20 €
14:30 - 20:30	-0.45 €	-0.75 €	-0.80 €	-0.70 €
16:00 - 18:30	-0.25 €	-0.15 €	0.00 €	-0.10 €
18:00 - 22:30	-0.35 €	-0.75 €	-0.50 €	-0.60 €
18:30 - 20:30	+0.55 €	+0.75 €	+0.70 €	+0.70 €
20:00 - 22:30	+0.55 €	+0.75 €	+0.70 €	+0.70 €
Average Price	Current	Current	+0.35 €	+0.85 €

Scenarios 2 and 3 present two situations of more balanced demand, compared with the real scenario. The differences are more prominent to the real scenario, as mentioned in the previous subsection, since scenario 1 is already successful when balancing the demand. The standard deviation of the utilization decreases to 1.00% in scenario 2 and 1.24% in scenario 3. The expected distribution for the scenarios 2 and 3 are presented in the Figure 5.2 and the expected business results are displayed in the Table 5.3.

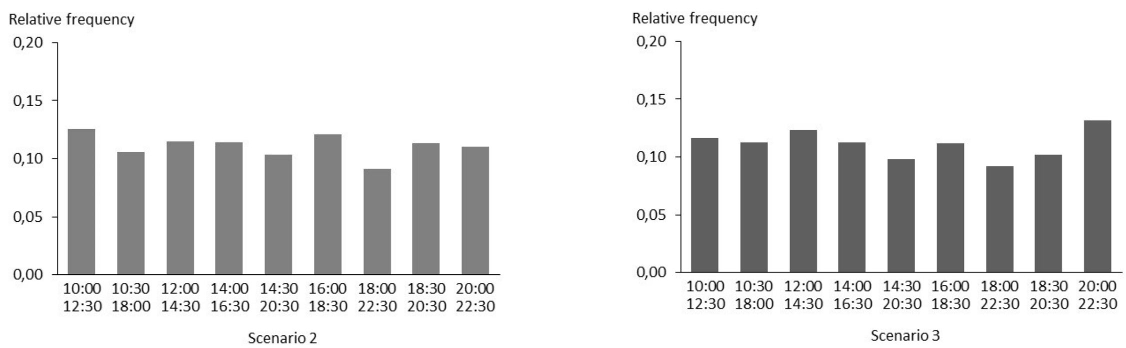

Figure 5.2: Expected utilization per slot for the generated scenarios.

Table 5.3: Results for the proposed scenarios with differentiated pricing.

	Real scenario	Scenario 1	Scenario 2	Scenario 3
Number of deliveries	Baseline	0.0%	-1.8%	-21.1%
Cost per delivery	Baseline	-3.6%	-7.8%	-4.6%

The second scenario is expected to decrease the number of deliveries per week, around 2% when compared to the real scenario. However, delivery costs are expected to be reduced by 7.8%. Despite the increase in customers' withdrawal, this is considered a scenario with a high potential, due to its capacity on balancing the demand and decreasing the operational costs. The impact on the number of deliveries is expected to be mitigated by subscription packages and expansion of the operations to uncovered areas.

When comparing scenarios 1 and 3, similar performances are verified on balancing the demand and on the operational costs per delivery. However, the decrease by 20% in the number of e-commerce customers would make the scenario 3 undesirable. The retailer's goal of growing the e-commerce business would be compromised by a scenario with such a significant impact on the number of customers.

To conclude, it is possible to corroborate that scenarios 1 and 2 are the most suitable ones for the current e-commerce business' goals. Therefore, both scenarios were used in the analysis discussed in the following section.

5.2 Experimental results on differentiated slotting

The previous section was focused on using the price as a tool to manage the demand. Now, it is possible to analyze the impact of changing the slot's availability on the overall operational performance. One of the problems in e-commerce deliveries is that customers are not equally divided across the country. There are areas with a high number of regular customers and others with sporadic deliveries, which leads to different operational implications to the retailer as exemplified in Figure 5.3. Therefore, it is relevant to concentrate the customers over the slots as much as possible to avoid high delivery costs to the retailer.

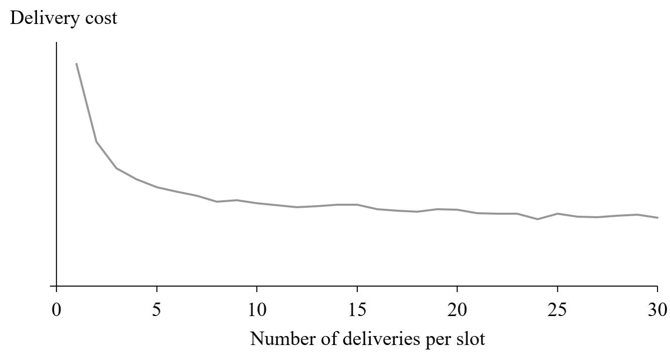

Figure 5.3: Relation of cost per delivery according to the number of deliveries per slot.

The results overviewed in this section compare the case where all the slots are open (baseline) with the following scenarios, in low-demand areas:

- Scenario 1 - Only the wider slots are open (10:30 - 18:00), (14:30 - 20:30) and (18:00 - 22:30);
- Scenario 2 - One wider slot (10:30 - 18:00) and one narrower night slot are open (20:00 - 22:30);
- Scenario 3 - Only one wider slot is open (10:30 - 18:00);

As mentioned in the previous section, the three scenarios were tested for two of the previously proposed prices' combinations: current average price (combination A) and increasing the average price by 0.35 € (combination B).

The utilization of the algorithm developed in Section 4.4.2 triggered at least one of the following outcomes: a decrease in the number of deliveries or of the cost per delivery. While the last effect is desirable to assure the sustainability of the operations, the first can have a prejudicial effect on business growth. The results presented in Table 5.4 summarise how the outcomes vary when the generated scenarios are applied for the two aforementioned prices combination.

Table 5.4: Results for the proposed scenarios with differentiated slotting.

Average price: current	Baseline	Scenario 1	Scenario 2	Scenario 3
Number of deliveries	0.0%	-0.1%	-0.9%	-3.5%
Cost per delivery	-3.6%	-11.9%	-8.1%	-14.4%
Average price: +0.35 €	Baseline	Scenario 1	Scenario 2	Scenario 3
Number of deliveries	-1.8%	-5.8%	-8.5%	-12.7%
Cost per delivery	-7.8%	-10.3%	-8.5%	-9.4%

For the prices' combination A, one of the most promising results is verified when the three wider slots are open, in low-demand areas (scenario 1). This scenario has an insignificant impact on the number of deliveries and delivery costs are expected to decrease more than 10%, comparing to the current scenario. The reduction of delivery costs is expected to benefit the goal of increasing the sustainability of e-commerce operations. The third scenario also discloses a promising reduction on the cost per delivery, but the high decrease in the number of deliveries makes the hypothesis of only opening one slot in low demand areas undesirable.

The expected outcome for the generated scenarios with combination B is worse than the overall performance of the scenarios with combination A. The higher delivery fees combined with a low offer decrease the number of deliveries for every scenario. Moreover, a low number of deliveries per slot is expected to increase the cost per delivery, as was previously suggested in Figure 5.3.

The retailer faces a trade-off between the number of deliveries and delivery efficiency. Increasing the delivery fees has the capacity of reducing the difference between the delivery price and its cost, but has a negative impact on the number of customers. On the other hand, finding the optimal combination for the current range of fees allows to sustain the number of deliveries, yet, the decrease on the difference between deliveries' costs and revenues is not that significant. Bearing this in mind, having a decrease in the number of customers is more harmful than not catching up breakeven on the delivery operations, for the current retailer's goals. Therefore, the described results led to the conclusion that the most suitable approach is the adoption of the optimized combination for the current average price, along the scenario where only wider slots are available, in low-demand areas. The proposed scenario is expected to generate yearly savings of 11.9% for the overall e-commerce operations.

Chapter 6

Conclusions and future work

The developed project focused on the time slot management practices of a Portuguese e-grocer. The objective was to propose and develop a methodology that provides analytical support to address the decision-making surrounding the availability of the delivery windows and respective fees. The analytical support had two main goals: the development of a tool to quickly access the impact of different delivery scenarios and the improvement of the current operations.

The project started with the analysis of the current operations, identifying the available time windows, the charged fees and the overall utilization of different delivery options. Subsequently, the two main fields of the e-commerce business had to be understood and represented: the consumer behavior and the fulfillment operations. The consumer behavior is recreated by the estimation of key parameters to characterize the demand and by the development of a machine learning model to estimate the consumers' choices when facing a set of delivery options. The fulfillment operations are approximated through the utilization of statistical distributions and a regression that considers both the number of deliveries and the respective covered area. A simulation model is used to foresee the customers' choices when facing a set of delivery alternatives and, after that, the fulfillment operations are estimated to understand the impact on operational costs. This procedure was used to assess the impact of different delivery scenarios, in order to evaluate the best alternative for the e-grocer to adopt.

The developed methodology enables the retailer to quickly assess the expected outcome of different delivery scenarios. Up to date, the process accountable for setting the available delivery windows and the fees to charge is of an empirical nature, with no procedure to foresee the expected outcome. Therefore, the generated scenarios were only evaluated on real-environment, which could have a prejudicial effect on the customers' satisfaction and operations sustainability. The existence of a tool to analyse the expected outcomes could prevent these undesirable effects. The simulation procedure enables the retailer to have a holistic and integrated view of the expected performance of the delivery operation for different delivery windows and regions. Moreover, it allows the retailer to test different scenarios in a short period of time, instead of waiting for various weeks to understand the impact of possible changes.

The utilization of a simulation tool to recreate the e-commerce operations provides greater comfort to the changes soon to be applied in a real scenario since many aspects of the delivery operations can be analyzed in greater depth. It is possible to assess the slots' expected utilization, the number of withdrawals and the effect of the price assigned to each slot. Consequently, the retailer is able to gather information to support his own decision-making, related to the time slot management. Moreover, both the stochastic nature of the input parameters and of the customers' choices increase the confidence of the final results, by assuring that the final outcome is not only based on a particular case but rather on the multitude of different possible events.

The proposed approach to test the impact of different scenarios is itself deprived of advisory capabilities. Also, the infinite combinations of delivery windows and possible prices to charge increase the dimensionality and complexity of the problem, which preventing the test of all the possible scenarios. Thus, the existence of methods that generate scenarios to be evaluated is of great importance. Throughout the project, an optimization model was developed to set the prices of the delivery windows and a heuristic procedure to set which are the delivery windows available in each geographical area.

According to the results of this thesis, a scenario with optimized charged fees, maintaining the current average price, and only making available wide delivery windows, in low-demand areas, is expected to produce better results than the current scenario. Even though the proposed scenario has not been implemented, the comparison with the performance of the *as-is* situation discloses a more balanced usage of the delivery windows and a reduction on the cost per delivery. The standard deviation between the delivery windows' utilization is expected to decrease from 4.25% to 1.37% and the delivery costs should decrease more than 10%. However, the proposed changes are not expected to have a significant impact on the number of customers.

The current state of the art in time slot management is not very vast and only a restrict set of papers have addressed the tactical field on this problem. Thus, the proposed approaches to generate scenarios extend the existing literature on the fields of tactical pricing and slotting. Moreover, among the available literature, it is possible to verify that all of the studies either make standard assumptions about consumer behavior or use regression models to approximate the behavior. Thereby, the utilization of a machine learning model to estimate the consumers' preferences is an upgrade to the current practices. Despite the improvement on the richness of the predictive model when compared to current approaches, the model has some shortcomings worth mentioning. The customers' information only embraces commercial data, such as basket value or booking time. A recommendation for future studies would be the usage of personal information, such as age, working situation and historical delivery preferences, to improve the predictive accuracy.

Despite the explicit improvement on the retailer's internal approach through the application of the proposed methodology and the expected gains for the generated scenarios, more remarks can be pointed out to the developed project. The proposed scenarios do not consider regional differentiation concerning the prices defined for the available delivery windows. Thus, a suggestion for future work would be to analyze how to set the fees not only for the available slots but also for the different delivery areas. Moreover, the experimental studies only embraced existing slots in the historical data set. The study of the expected outcome of new delivery slots should also be performed for further improvements on the e-commerce operations. The developed project only encompassed the tactical field of the time slot management problem. This study is crucial to set the retailer's operations in the medium-term and to build the foundations for more complex management methods. However, future work should also encompass the utilization of dynamic approaches, in order to update the delivery policies in real-time as soon as new information is available.

Bibliography

Agatz, N., Campbell, A., Fleischmann, M., and Savelsbergh, M. (2008a). Time slot management in attended home delivery. *Erasmus Research Institute of Management (ERIM)*, 45.

Agatz, N., Campbell, A., Fleischmann, M., van Nunen, J., and Savelsbergh, M. (2008b). Demand management opportunities in e-fulfillment: What internet retailers can learn from revenue management. *Erasmus Research Institute of Management (ERIM)*.

Asdemir, K., S. Jacob, V., and Krishnan, R. (2009). Dynamic pricing of multiple home delivery options. *European Journal of Operational Research*, 196:246–257.

Balkan, A. (2008). Can geographical pricing survive the information age. `https://2018.ar.al/1170/`. accessed: 10.05.2019.

Ben-Akiva, M. and Bierlaire, M. (1999). *Discrete Choice Methods and their Applications to Short Term Travel Decisions*, pages 5–33. Springer US, Boston, MA.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models.* PhD thesis, Massachusetts Institute of Technology.

Bettman, J. and Zins, M. (1979). Information format and choice task effects in decision making. *Journal of Consumer Research*, 6:141–53.

Bierlaire, M. (1998). *Discrete Choice Models*, pages 203–227. Springer, Berlin, Germany.

Bonabeau, E. (2002). Agent-based modeling: methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 3:7280–7287.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Campbell, A. M. and Savelsbergh, M. (2006). Incentive schemes for attended home delivery services. *Transportation Science*, 40(3):327–341.

Campbell, A. M. and Savelsbergh, M. W. P. (2005). Decision support for consumer direct grocery initiatives. *Transportation Science*, 39(3):313–327.

Cleophas, C. and Ehmke, J. (2014). When are deliveries profitable?: Considering order value and transport capacity in demand fulfillment for last-mile deliveries in metropolitan areas. *Business and Information Systems Engineering*, 6(3):153–163.

Daganzo, C. F. (1987). Modeling distribution problems with time windows: Part i. *Transportation Science*, 21(3):171–179.

Ehmke, J. and Campbell, A. (2014). Customer acceptance mechanisms for home deliveries in metropolitan areas. *European Journal of Operational Research*, 233:193–207.

Fisher, M. and Jaikumar, R. (1981). A generalized assignment heuristic for vehicle routing. *Networks*, 11:109 – 124.

Gallego, G., Ratliff, R., and Shebalov, S. (2015). A general attraction model and sales-based linear program for network revenue management under customer choice. *Operations Research*, 63(1):212–232.

Gonçalves, J. (2000). Gestão de aprovisionamentos: Stocks - previsão - compras. *Publindustria*.

Grigoryev, I. (2015). *AnyLogic 7 in Three Days*. CreateSpace Independent Publishing Platform, Scotts Valley, CA, EUA.

H2O (2019). H2o documentation site. `http://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html`. accessed: 01.03.2019.

Hernandez, F., Gendreau, M., and Potvin, J.-Y. (2017). Heuristics for tactical time slot management: a periodic vehicle routing problem view. *International Transactions in Operational Research*, 24(6):1233–1252.

Hillier, F. S. and Lieberman, G. J. (2001). *Introduction to Operations Research*. McGraw-Hill, New York, NY, USA, seven edition.

Hyndman, R. J. and Athanasopoulos, G. (2018). Forecasting: principles and practice. `OTexts.com/fpp2`. accessed: 01.05.2019.

Kimes, S. E. and Wirtz, J. (2003). Has revenue management become acceptable?: Findings from an international study on the perceived fairness of rate fences. *Journal of Service Research*, 6(2):125–135.

Klein, R., Mackert, J., Neugebauer, M., and Steinhardt, C. (2018). A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum*, 40(4):969–996.

Klein, R., Neugebauer, M., Ratkovitch, D., and Steinhardt, C. (2017). Differentiated time slot pricing under routing considerations in attended home delivery. *Transportation Science*, 53(1):236–255.

Koppelman, F. and Bhat, C. (2006). *A Self Instructing Course in Mode Choice Modelling: Multinomial and Nested Logit Models*. U.S. Department of Transportation Federal Transit Administration, Washington, DC USA.

Mackert, J. (2019). Choice-based dynamic time slot management in attended home delivery. *Computers & Industrial Engineering*, 129:333 – 345.

Mintel (2016). Online grocery retailing uk. *Technical report*.

Nielsen (2018). Global online grocery. `http://www.nielsen.com/eu/en/press-room/2018/global-online-grocery-purchasing-is-up-15-percent-in-last-two-years.html`. accessed: 20.05.2019.

Savory, P. and Mackulak, G. (1994). The science of simulation modeling. *Industrial and Management Systems Engineering Faculty Publications*, 26:1–56555.

Shannon, R. E. (1975). *Systems Simulation: the art and science*. Prentice-Hall, Upper Saddle River, NJ, EUA.

Smith, H., Currie, C., Chaiwuttisak, P., and Kyprianou, A. (2018). Patient choice modelling: how do patients choose their hospitals? *Health Care Management Science*, 21(2):259–268.

Swaminathan, J. M., Smith, S. F., and Sadeh, N. M. (1998). Modeling supply chain dynamics: A multiagent approach. *Decision Sciences*, 29(3):607–632.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76:31–48.

UPS (2017). Ups pulse of the online shopper - a customer experience study. *UPS*.

Van Ryzin, G. and Vulcano, G. (2015). A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science*, 61(2):281–300.

Wirtz, J. and Kimes, S. (2007). The moderating role of familiarity in fairness perceptions of revenue management pricing. *Journal of Service Research - J SERV RES*, 9:229–240.

Yang, X. and Strauss, A. (2017). An approximate dynamic programming approach to attended home delivery management. *European Journal of Operational Research*, 263:935–945.

Yang, X., Strauss, A., Currie, C., and Eglese, R. (2014). Choice-based demand management and vehicle routing in e-fulfillment. *Transportation Science*, 50:473–488.