Dynamic Delivery Pricing Model Applied to Two Hypotheses Testing

Matilde Proença Aires

Dissertação de Mestrado

Orientador na FEUP: Prof. Alcibiades Soares Guedes

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia e Gestão Industrial

2019-07-01

“Não é o mais forte, nem sequer o mais inteligente da espécie que sobrevive. O que sobrevive é o que se mostra mais adaptável à mudança”

Leon C. Megginson, 1963

Resumo

Nos últimos anos o e-commerce – retalho online – sofreu um crescimento acelerado. Com este crescimento surgiram problemas até então inexistentes. No retalho tradicional todos os custos de transporte associados a um produto, desde a loja até à casa de um consumidor são inconscientemente suportados por esse mesmo cliente. No retalho online os mesmos custos são vistos como taxas adicionais sendo muitas vezes consideradas injustas e apenas uma forma do retalhista obter lucros extra. Deste modo empresas de retalho online são forçadas a encontrar o equilíbrio entre cobrar taxas de envio aos consumidores, ao mesmo tempo que os retêm e adquirem novos. Só é possível estabelecer este equilíbrio através de um processo iterativo em que são testados valores diferentes para essas taxas. A forma mais eficiente de comparar dois valores para uma mesma taxa online é através de testes de duas hipóteses também conhecidos como *A/B tests*.

Esta tese propõe a construção de um modelo de classificação para prever em que sessões é mais provável haver conversão – passar para a página seguinte no processo de compra num website / app. O algoritmo vai ser posteriormente integrado numa plataforma de *A/B testing* com o objetivo de melhorar o desempenho desta.

Para a construção do algoritmo dois *datasets* diferentes foram construídos. Um para cada uma das páginas (do website da empresa em estudo) em que o consumidor é confrontado com os preços de envio. Adicionalmente três algoritmos de classificação foram analisados – *Naive Bayes, Logistic Regression* e *Extreme Gradient Boosting.*

Após avaliar o desempenho de cada um dos modelos para cada um dos *datasets* a combinação que mostrou melhores resultados foi escolhida. Esta escolha recaiu sobre o modelo Extreme *Gradient Boosting* e o *dataset* relacionado com a página do website onde o consumidor é confrontado pela ultima vez pelas taxas de envio.

Em último lugar é demonstrado como o algoritmo seria integrado na atual plataforma de *A/B tests* e como a melhoraria.

A integração do algoritmo com a plataforma de *A/B tests* vai permitir a implementação de testes mais rápidos e com menos consumo de recursos.

Abstract

In the past years, e-commerce experienced rapid growth. Alongside this growth came a new problem. Shipping and handling fees that in traditional retail are unconsciously absorbed by customers, in e-commerce are consciously and unwillingly supported by consumers. Therefore, e-commerce companies struggle to find the perfect balance between charging those fees to their consumers, acquiring new ones, and maintaining the already existing clients. Finding this balance is only possible through an iterative process in which new values of shipping fees are tested. The best way to compare the impact of two different values for the same fee is through two hypotheses testing, also known as A/B testing.

This thesis proposes the construction of a machine learning classification algorithm to predict in which sessions a conversion – go to the following page on a website/app – is likely to occur. This algorithm is later going to be integrated on a shipping fee A/B testing platform to enhance its performance.

To construct the algorithm, two different datasets were collected - one for each of the two different pages in which shipping fees are conveyed to the consumer on the case study company website. Furthermore, three different classification algorithms were studied and optimized– Naïve Bayes, Logistic Regression, and Extreme Gradient Boosting.

After assessing the performance of all the three models combined with the two datasets, the association that showed the best performance was chosen. This combination was the Extreme Gradient Boosting model along with the dataset related to the website/app page where the consumer faces the shipping price for the last time before making a purchase.

Finally, a description of how the algorithm would be used to improve the current A/B testing platform is given.

The integration of the chosen algorithm with the A/B testing platform will allow for less costly and time-consuming A/B tests to be performed.

Acknowledgments

I would like to thank Farfetch for allowing me to develop my master thesis. Especially to the delivery development team members, that received me with open arms and were always there to help me. In particular to Tomás Palhinhas who acted as my mentor and guided me through every step of this project, to Ivo Nogueira who offered me crucial insights regarding Big Data and Machine Learning, to Lisandra Rocha that assisted me in the creation of all the datasets used throughout the development of this dissertation, to Ana Rita Moura for accompanying me and helping with my integration and finally to my team leader, Pedro Bastos.

To my thesis supervisor, Prof. Alcibiades Soares Guedes, whose help was indispensable for the completion of this document, a special thank you. I would also like to thank Prof. Vera Migueis for the support given with Machine Learning algorithms. To all the Industrial Engineering and Management teachers, thank you for being there throughout the path that led me to the conclusion of my master's degree and this thesis. I would also like to give a very special thank you to FEUP that, for the past five years has been like a second home to me.

Finally, I would like to thank my family, my friends and my boyfriend not only for the endless support during the past four and a half months but also for helping me every day to be the best version of myself.

Contents

Acronyms and symbols

3PL	Third Party Logistics
AOV	Average Order Value
AUC	Area Under Curve
AWB	Air Way Bill
BDA	Big Data Analytics
BO	Boutique Order
CM	Confusion Matrix
CRISP-DM	Cross Industry Standard for Data Mining
CV	Cross Validation
EGB	Extreme Gradient Boosting
FN	False Negatives
FP	False Positives
FR	Flat Rate
FS	Free-Shipping
GTV	Gross Transaction Value
IPO	Initial Public Offering
KDD	Knowledge Discovery in Databases
LR	Logistic Regression
ML	Machine Learning
NB	Naive Bayes
NFL	No Free Lunch
P&L	Profit and Loss
PDP	Product Description Page
PLP	Product List Page
PO	Portal Order
PP	Partitioned Pricing
PPC	Pay Per Click
ROC	Receiver Operating Characteristic
S&H	Shipping and Handling
SBU	Strategic Business Unit
TBGB	Tree Based Gradient Boosting
TN	True Negatives
TP	True Positives

List of figures

Listo of tables

1 Introduction

The undeniable exponential growth of e-tail, sale of products on the internet (Cambridge 2008), in the past decade has led to the reformulation of the business model of numerous retail companies, and the luxury sector was no exception. Alongside this reformation, a boom of new fashion-related technology start-ups occurred. This revolution, together with external factors allowed for a growth in the luxury sector of 5.1% CAGR (167bn€ to 262bn€) between 2008 and 2017. Moreover, the personal luxury goods market is predicted to grow to 290bn€ by 2020, as reported by RetailX (2019).

To reinforce the need to become digital, about 54% of BoF-McKinsey State of Fashion Survey respondents said that "omnichannel integration, investing in e-commerce and digital marketing is their number one priority for 2019" (Amed et al. 2018) which is in line with Statista prediction that 17% of the total revenue in the luxury goods sector will be generated through online sales by 2022 (Brinckmann 2018), a considerable increase from the 2017's 9% (D'Arpizio et al. 2017). All of these indicate the market's definite potential.

However, the luxury online sales sector also faces some adversities - one of these being the impact in customer satisfaction caused by the systematic differences that exist between online and offline shopping environments (Cao and Li 2015). Due to these adversities, the need to increase customer satisfaction and therefore, customer retention arises. According to Gounaris and Dimitriadis (2003), one of the three primary quality dimensions that affect customer satisfaction is "customer protection and risk decreasing comprising secure online payment and shipping aspects in terms of time, costs and options, and communication". Bamfield (2013) also claims that customer satisfaction is directly connected to how customers perceive prices and their fairness. Taking both these perspectives into account, e-tail companies should focus on how they calculate the shipping price of its products and how its clients perceive it.

As Nisar and Prabhakar (2017) observed, e-tail websites are a "gold mine of analytic tools to assist in effective retailing methods" making it easier for e-tail companies to alter shipping dynamically, hopefully allowing for better customer satisfaction and retention.

In view of the aforementioned statements and within the scope of FEUP's Industrial Engineering and Management course the following master thesis will focus on the creation of a machine learning tool to analyze the buying intention of Farfetch clients. This tool will then be used to alter the way shipping price alterations are being tested. This thesis also presents the analysis that should precede the creation of such a tool. The analysis will consider all the possible variables that may be relevant for the consumers' purchase intention with a special focus on shipping fees.

1.1 Farfetch

Farfetch is a rapid growing luxury e-commerce platform launched in October 2008, selling products from 25 boutiques in 5 countries (Farfetch 2019). Today it connects key consumers to more than 800 boutiques and almost 400 brand owners (Linnane 2018) with over 1000 brands available (Farfetch 2019). In September 2018 Farfetch launched its activity in the stock market. Its IPO (Initial Public Offering) valued the company at 5.12bn€ with its shares trading as much as 39% above its initial public offering (White and Fares 2018).

Farfetch has offices in 13 sites - Porto, Lisbon, Guimarães, Braga, São Paulo, Hong Kong, Moscow, Los Angeles, New York, London, Tokyo, Shanghai and Dubai – and currently employs a workforce of over 3000 employees. Also, in the later year, Farfetch has made some strategic partnerships. Of those, the partnerships with both JD.com and Harvey Nichols, two giants in the e-commerce and luxury segments respectively, should be highlighted.

Despite all this success, it is clear, as shown in Figure 1, that Farfetch benefited from the first-mover status as a luxury marketplace in the first part of the decade. In most recent years, there is an irrefutable trend shift towards luxury brand owners leading luxury e-commerce (Andersson and Shaw 2019).

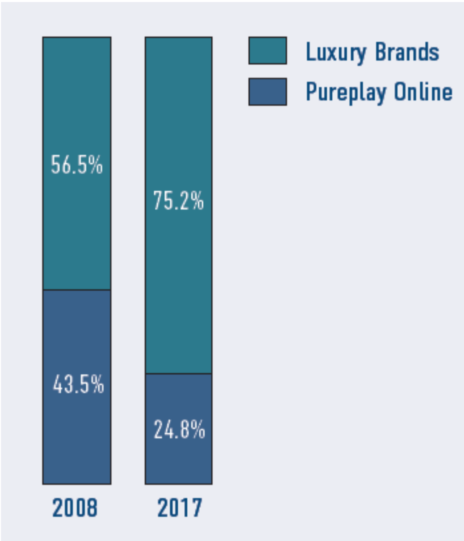

Figure 1 - Share of online revenue in the luxury sector assigned to pureplayers vs. luxury brand owners for 2008 and 2017, in (Andersson and Shaw 2019)

However, as aforementioned, it is indisputable that Farfetch is still growing and that this growth is in part supported by Farfetch's business model. This model is based upon a commission per sale, and unlike its competitors, Farfetch does not own any of the stock that is available for sale in its marketplace. This business model has both advantages – the savings on inventory expenses and the vast amount of product availability – and disadvantages as the increased likelihood of stockout and the augmented complexity of the delivery process.

As a result of this complexity, Farfetch uses a drop-shipping model. A drop-shipping model implies that the transport of goods between boutiques and consumers is established using third party logistic (3PL) partners.

Given that Farfetch ships to over 190 (Linnane 2018) countries and offers 4 different types of services – Standard, Express, Same Day Delivery and 90 Minutes Delivery – the 3PL partners perform more than 5000 different routes. This results in a complex distribution, implying extremely high shipping costs.

Price perception is one of the most important criteria to consider when trying to increase and maintain customer retention (Lewis 2006). On the other hand, shipping costs are Farfetch's biggest cost pool. Hence the shipping price charged to a Farfetch client does not always reflect the actual shipping cost in which Farfetch incurred to send a parcel.

Although Farfetch imputes a reduced shipping price to its clients, when compared to its competitors, Farfecth is practicing higher prices for either the same service or for a worse one. However, Farfetch cannot alter its prices without analyzing what those alterations may imply for the company's performance. Testing those alterations in the quickest possible manner is, therefore, a necessity. This necessity highlights the need for a project like the one exposed in this thesis.

1.2 Project objectives and methodology

As José Neves, Farfetch's founder and CEO, said in his interview to Expresso (Marques 2019), a company is only as strong as its values. Striving to stay relevant in an always changing market, Farfetch takes this vision very seriously. Being the most recent of the six Farfetch values "Amaze Customers" it is not surprising that, as abovementioned, the shipping price is lower than the actual shipping costs. As a matter of fact, for the year of 2018, about 31% of the shipping costs were supported by Farfetch.

Taking this into account, the objective of this dissertation project is to reformulate the way shipping price alterations are tested. To do so, a new tool will be introduced. A series of analyses on the relevant variables that may affect conversion - the number of purchases divided by the total number of visits to the website - alongside the shipping price will be performed.

The methodology used to structure the problem considered the creation of a series of steps as a way to deconstruct a complex problem into simpler ones. The steps were as follows:

- **Case study analysis** – fully understand the context of the problem, how the shipping prices are being calculated today and how any possible changes to those prices are being tested;
- **Variables analysis and solutions proposal** – understand which variables have the most significant impact on conversion, giving a special focus to shipping fees. Construct different models that based on those variables are able to predict if a consumer is going to convert or not;
- **Model selection** – propose, based on the conclusions derived from the previous steps, the most indicated model for the described problem;
- **Hypothesis testing** – create a new tool to test shipping prices' alterations based on the selected model. This tool should be able to test if the alterations made have a positive effect on Farfetch's gross transaction value (GTV);

To better plan the work needed to accomplish the objectives described in this section, a Gantt chart was developed as seen in Figure 2.

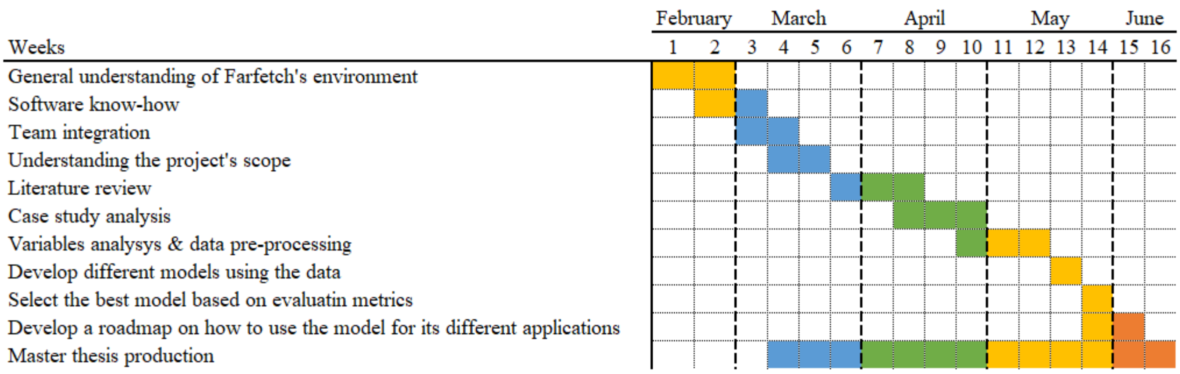

Figure 2 - Gantt chart with the different steps defined

1.3 Thesis structure

The remainder of this thesis is structured to give a complete scope of shipping fees two hypothesis test test and their proposed alteration. Chapter 2 is the literature review related to luxury e-commerce, shipping fees calculation methods, and machine learning algorithms. Chapter 3 describes the proposed methodology to be followed throughout the development of chapters 4 and 5. Chapter 4 encompasses the understanding of the Farfetch ecosystem with a special focus on the shipping fees calculation method, revenue structure, and two hypothesis testing. It corresponds to the first step of the proposed methodology. As for the following steps, they are all comprised in chapter 5. Chapter 5 focuses on the creation of a machine learning classification algorithm and in its integration with the current A/B testing platform creating a new A/B testing tool. Finally, chapter 6 reflects on the results obtained in the previous chapters and on future work to complement the studies comprised in this dissertation.

2 Literature Review

2.1 Luxury e-commerce

To fully comprehend how the luxury e-commerce functions, it is necessary to understand the scope of the luxury fashion industry, its core characteristics, and its consumers.

As stated by Cornell (2002), “luxury is particularly slippery to define” but its key components are the “strong element of human involvement and the very limited supply”. Furthermore, Kapferer (1997) claims that luxury is a way to define beauty, “it is art applied to functional items. Like light, luxury is enlightening. Luxury items provide extra pleasure and flatter all senses at once”.

A more objective view of what the key components of luxury are is given by Okonkwo (2009), stating that “these elements speak more to passion than reason” and are originality, creativity, craftsmanship, precision, exclusivity, high quality, and premium pricing.

Being luxury so ambiguous to define, Wiedmann et al. (2009) created a framework proposing that the luxury consumer decision-making process could be translated into four different dimensions: financial, functional, individual and social, classified as first order latent variables. These variables should then be divided into nine antecedent constructs related to the previously mentioned dimensions and to each other as shown in Figure 3.

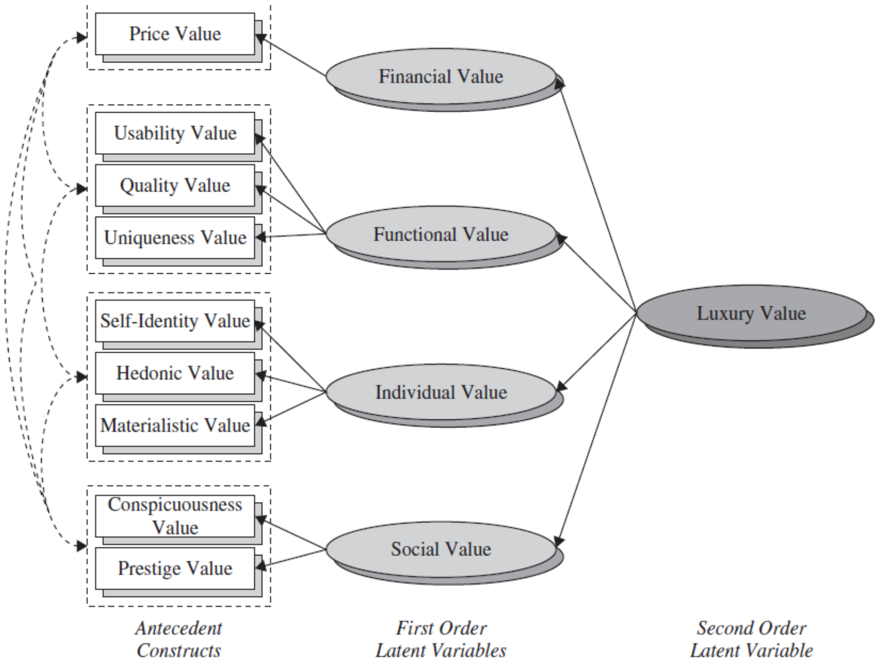

Figure 3 - Consumer decision making process as thought by Wiedmann et al. (2009)

An interesting discovery their research brought to light was the fact that all the antecedent constructs had a positive impact on the luxury value of an item, except for the usability value. This means that the higher the usability value of an item the lower is its luxury value. The usability value of an item can be described as both how well a product's proprieties suit a consumer's needs, and how easy it is to use while performing the tasks it was conceived to perform (Cambridge 2008).

In their research, Wiedmann et al. (2009) also affirmed that the different dimensions and antecedent constructs had different impacts on different consumers, thus the need to understand who the fashion luxury consumers are has arisen.

In the interest of this thesis, the fashion luxury consumers were divided into two main categories: online luxury buyers and in-store luxury buyers. In their paper about luxury shoppers, Xia Liu et al. (2013) discovered that these two segments of consumers were influenced by different motivational factors, which is in line with the study by Wiedmann et al. (2009). According to their paper, online shoppers tend to be more price-conscious, are pleased with the vast number of products available online, and find comfort in evaluating a product's value through reading online customer reviews. On the other hand, in-store shoppers are more risk-averse towards the security provided by online shopping, feel that seeing the product is a necessity before purchasing it and believe that the shopping experience and interaction in a luxury store add value to the product. Despite their differences, both these consumers feel the fundamental needs "to be admired, recognised, appreciated and respected" and one way of satisfying these needs is through their possessions (Okonkwo 2009).

After understanding the scope of the luxury fashion industry, its core characteristics, and its consumers, it is essential to have a high-level understanding of e-commerce. E-commerce is characterized by usually following a pull marketing approach – customers are drawn to information and purchases – by a low switching cost between sellers, by a weaker sales power and by being available to a mass consumer base (Okonkwo 2009).

Having comprehended both the fashion luxury industry and e-commerce it becomes easy to grasp why Dauriz et al. (2013) affirm that conventional wisdom says that selling online and especially on multibrand retail websites is for the lower and middle ranged products. However, they also claim that technology and e-commerce are revolutionizing the way people shop for luxury and that luxury brands cannot afford to ignore this transformation. Therefore, it is imperative that brands find the right balance between an online presence and sustaining their luxury image. To do so, it is necessary to design a strategy that best fits their needs.

On their article, Dauriz et al. (2013) identified three different digital archetypes for luxury brands and retailers, defined as follows:

- The "Plugged-in pro" – diversified retail strategy, for mono or multibrand stores, and complete use of the digital world (from social media to an online store);
- The "Selective e-tailer" – for monobarands stores that use the digital world only for new customers (digital marketing and online store for entry-level products);
- The "Hesitant holdout" – for small monobrand stores that use the digital world only as a complementary showroom.

For each one of these archetypes, a different online strategy should be implemented. Nevertheless, a good e-strategy is often supported by the same principles.

Understanding these principles means that first, it is necessary to acknowledge that luxury companies are structured and function in a way that does not readily accept radical changes, representing a problem for online integration. To face this issue, it is essential to re-think the structure and working method of all business aspects without compromising the normal functioning of the company. This implies incorporating a Strategic Business Unit (SBU) fully dedicated to the Internet, comprised of qualified personnel who would liaise with annex departments. This SBU needs to be led by skilled managers capable of making the best strategic decisions to ensure the appropriate positioning of the brand in the cyberspace. To do so, it is vital to avoid "the widespread practice of internal competition between departments for sales revenue and clients"(Okonkwo 2009).

Once all the principles are implemented, the company must find the most appropriate methods to monitor its online progress. An interesting metric to do so is to closely follow how consumers' online attitudes are reflected in their attitudes towards the brand offline. Besides this metric several new ones should be developed with one main objective as the goal, which is creating a luxury shopping experience in a mass market environment.

2.2 Shipping fees

Bearing in mind the strategic notions aforementioned, one thing that all managers dedicated to exposing luxury brands online should consider is the "online cart abandonment" phenomenon. This phenomenon is defined as "consumers' placement of item(s) in their online shopping cart without making a purchase of any item(s) during that online shopping session". Approximately 88% of online shoppers have abandoned their cart in the past, and every time a customer places an item in the cart there is about a 25% chance that the cart will be abandoned (Kukar-Kinney and Close 2010), making this a widespread and almost alarming phenomenon.

In their paper, Kukar-Kinney and Close (2010) identified the main drivers for online cart abandonment and explained that some of the reasons for this phenomenon are unavoidable. Notably, many consumers consider the online shopping cart either as entertainment or a tool to organize their shopping research, inducing them to only make a purchase either on a later session or via another channel. When this is not the case, online shoppers compare the products most relevant attributes as perceived by them between several retailers or between their idea of what that attributes should be like.

With the entertainment and organizational values of the shopping cart being the two most important factors for cart abandonment, the third most important driver was identified as being the concern for overall costs. Interestingly, this is also the main driver to buy from a physical store instead of using the virtual channel (Kukar-Kinney and Close 2010). This third factor, unlike the first two, is more actable upon. To better understand how to minimize this driver, it is necessary to comprehend its composition.

Overall costs are composed of the product's price, the shipping and handling fees (S&H) and taxes and duties. Regarding the product´s price, its value is not susceptible to change until the sale seasons. As for the S&H fees and taxes and duties, research suggests that online shoppers are less sensitive to the taxes and duties charges than to the S&H fees (Frischmann et al. 2012). This research is also supported by the surveys conducted by several institutions where 60% of the respondents claimed to have abandoned their cart when the shipping fees were added, and 50% of shoppers affirmed that the main drawback of online shopping where the shipping fees. These results are probably an effect of the fact that in traditional retailing, the S&H fees are unconsciously absorbed by the consumer (Lewis et al. 2006). Therefore, it is reasonable to assume that the S&H fees are the most relevant component of the overall cost related to the cart abandonment phenomenon.

Before exploring ways to minimize the impact of the S&H fees, it is crucial to grasp how the alterations made to these fees will affect the consumer's behavior. Especially given the element of discrete choice that these fees introduce into the customer's buying decision

(Lewis et al. 2006). Research found that the four main impacts S&H fees have are in customer retention and acquisition, purchase incidence (number of orders placed) and expenditure decisions with “multiple and potentially conflicting effects “ (Lewis 2006).

Regarding customer acquisition and retention, conventional wisdom is that acquiring a customer is more expensive than retaining one. However, shipping fees are introduced in a similar way to both new and returning customers. Furthermore, due to the mandatory nature of these fees, they may prevent the purchase by both current and potential customers (Lewis 2006). According to Morwitz et al. (1998), because S&H fees are only introduced to the consumer once they have already chosen to place an item in the shopping bag, prospective customers are more likely to under weigh these fees than returning customers who are already acquainted with the value that these fees will take. However, more recent research by Lewis (2006) found that due to the major relevance that is being given to the shipping fees (Lewis et al. 2006), these are more likely to affect the purchase intention of a new customer than the one of returning clients. Moreover, Lewis (2006) found that new customers are more sensitive to shipping fees that tend to penalize the basket size (in both number of items and total value) and contrariwise returning customers are more sensitive to the base shipping fee level (shipping paid before any number of items or value threshold is crossed).

As for the purchase incidence and expenditure decisions, some variables are beyond any company’s control. It was proved that customer expenditure tends to increase with the increase of the purchases made to a specific retailer. However, Lewis et al. (2006) were able to identify how shipping fees influence purchase incidence and expenditure empirically. Firstly, the higher the shipping fee per item, the lower the order volume will be. Secondly, when shipping fees have order size incentives (the higher the value paid for the items, the lower the shipping fee), the expenditure will be higher, and the purchase incidence will be lower. The reverse condition – “free shipping”- has the opposite consequences.

As a result of the different impacts that altering the S&H fees have - the high costs that shipping items to the consumer represents to a company (average loss of $4 to $16 per order shipped (Frischmann et al. 2012)) - and the constantly changing market conditions, each company needs to find the right S&H fees strategy.

In the literature, there are two main types of strategies, the partitioned pricing (PP) strategy, and the “free shipping” (FS) strategy. Both of these strategies try to take advantage of the consumers’ biased perceptions of S&H fees. For example, there is a segment of consumers who consider S&H fees as being unfair referred to as shipping-charge skeptics (Schindler et al. 2005).

When using a PP strategy, it is usual to only reveal the S&H fees at the end of the purchase as a way to benefit from the consumers’ high lock-in costs. Consequently, the PP strategy is more commonly used in markets where customers spend little cognitive effort when comparing prices, there are few shipping-charge skeptics and the products sold are of high volume (Mehmet et al. 2013). An advantage of the PP strategy is that the item price recalled by the consumer is lower than what he actually paid. When using a PP strategy, retailers often charge high gross product prices by increasing the S&H fees more than they decrease the product’s net price, as shown in Figure 4.

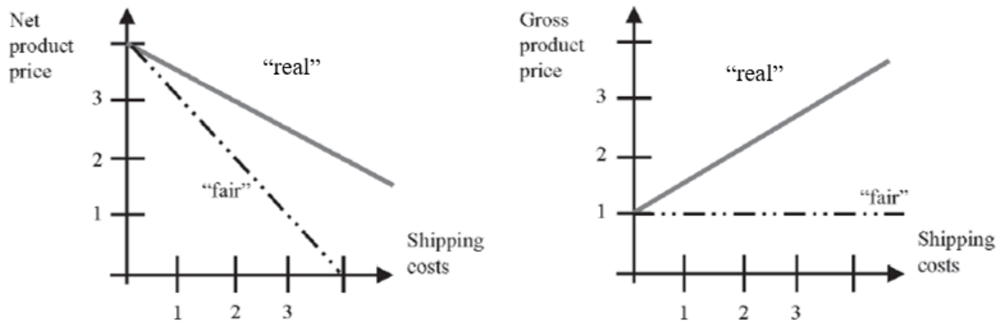

Figure 4 - Difference between the real prices charged by PP retailers Vs. the fair price, adapted from (Frischmann et al. 2012)

As for the FS strategy, it is more commonly used in markets where the number of shipping-charge skeptics is high, and the volume of the items shipped is small (Mehmet et al. 2013). FS offers are seen as being better than moderate S&H fees due to the zero price effect – " decrease in price from a non-zero value to zero, say from $1 to 0, increases demand more than the same decrease in the positive price range, say from $2 to $1" (Shampanieret al. 2007). By reason of this effect, companies that apply the FS strategy will most likely charge high gross product prices by increasing the product price more than the necessary value to cover the shipping costs (Frischmann et al. 2012).

An interesting finding by Mehmet et al. (2013) is that companies using the FS strategy tend to alter their prices 1.5 times more than those using the PP strategy. They also suggest that given the high volume of shipments that popular retailers perform they should offer "free-shipping" taking advantage of the shipping economies of scale.

Another less explored strategy in the literature is charging customized S&H fees, which takes advantage of both the heterogeneity that exists between consumers and the purchase records that companies keep.

2.3 Machine learning on big data

In the past years, the growing interest in big data has led to many attempts in defining what big data is. However, many still use only one dimension, size, to characterize it (Gandomi and Haider 2015). Nevertheless, a complex definition when talking about big data is beginning to be used more often. TechAmerica Foundation (2012) labels big data as being "a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information." Similarly, Gartner, Inc (2013) defines big data as being "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.". These two definitions are in line with what has been described as the 3 "Vs" of big data - volume, variety, and velocity. The "Vs" are used to minimize the incorrect characterization of big data, and each "V" embraces a spectrum of different measures for each dimension (Kiron et al. 2014). The three "Vs" and their ramifications are depicted in Figure 5.

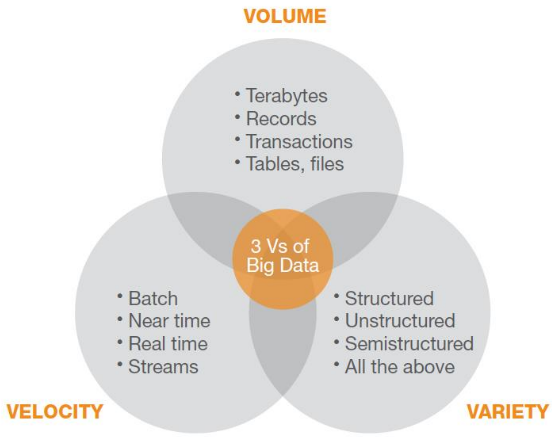

Figure 5 - The three Vs of big data, in (Kiron et al. 2014)

As the understanding of big data spreads, so does its use in analytics – big data analytics (BDA) – to improve business metrics. About 91% of Fortune 1000 organizations are allocating resources to BDA (Kiron et al. 2014). This allocation of resources has proven to be successful given that in their study McAfee and Brynjolfsson (2012) discovered that companies who use big data to make more informed decisions regarding their business are on average "5% more productive and 6% more profitable than their competitors".

A common form of BDA is machine learning (ML). ML is a concept that has been around for a long time, as Michie (1968) said in his 1968 paper: "Attempts to computerize learning processes date back little more than 10 yr.". An up to date definition of what ML is can be found on Mohri et al. (2013) book that states that "Machine learning can be broadly defined as computational methods using experience to improve performance or to make accurate predictions.". In their definition, the term "experience" is related to all the previously collected data that is available for analysis and therefore available for the computational methods to learn from. ML techniques have several practical applications. For example, credit card companies use it to identify fraud, Netflix and other streaming companies use it to make specific recommendations to their users, and the financial system uses it to deal with billions of trades (Kuhn and Johnson 2013). All these applications correspond to a particular category of learning problems. Three of the most popular techniques and also the ones lectured during the Industrial Engineering and Management course are classification, regression, and clustering that can be defined as follows (Mohri et al. 2013):

- Classification models are defined as algorithms that assign a category to each analyzed object (for example, classifying a customer as being fraudulent or not).
- Clustering models divide items into homogeneous groups and are usually used to analyze extensive data sets (for example, in customer segmentation to identify possible groups of consumers).
- Regression models are used to predict the real value of an item. For these types of models, there is a penalty for every incorrect prediction that depends on the magnitude of the difference between the real and the predicted values. These models are often used for, for example, predicting the stock values.

Even though there are numerous ML techniques, most fall under either supervised learning or unsupervised learning techniques. Unsupervised learning techniques are characterized by the fact that the data used to create a model is unlabelled, and the model attempts to make predictions for all unseen points. In supervised learning, the data already has predefined labels that are used to build a model of those labels distribution in terms of the predictor (independent) variables (Mohri et al. 2013). An example of unsupervised learning is clustering models. As for supervised learning, the most common example are classification models.

The quality of the data used to model an algorithm is crucial to determine the model's performance. A common issue regarding data quality lies in imbalanced datasets. There are 2 typical methods to balance data:

- Undersampling – randomly eliminate observations from the class with the highest number of observations;
- Oversampling – re-sample entries from the class with the lowest number of observations.

Figure 6 portrays the described methods of data balancing for a 2-class classification model.

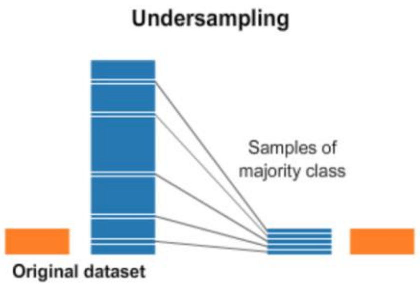

Figure 6 - Dataset balancing methods - undersampling and oversampling, in (Badr 2019)

Once the data balancing, if necessary, is finished, the next step to construct a model is data splitting. Data is usually split into a training dataset and a testing dataset. The training set is used to train the model – understand the relationship between the independent variables selected and the output variable - and the test set is used to evaluate the model's performance. The test set should never be used to train the model so that it provides unbiased values of the model performance (Kuhn and Johnson 2013).

There are two different, commonly used ways to split a dataset. The first one is the hold-out method in which the data is partitioned into two mutually exclusive datasets – the training and test datasets previously described. Usually, the training set has more data entries (about 2/3 of all the data) than the test set (Mohri et al. 2013). The other method is the k-fold cross-validation (CV). For this method, the data is randomly partitioned into k equal sized sets. A model is trained using k-1 sets and is tested on the held-out set. This process is repeated until all the k subsets were used as test sets. The performance of the algorithm is evaluated using an average of the k models' performance (Kuhn and Johnson 2013). A simplified version of this method is depicted in Figure 7 with k=3.

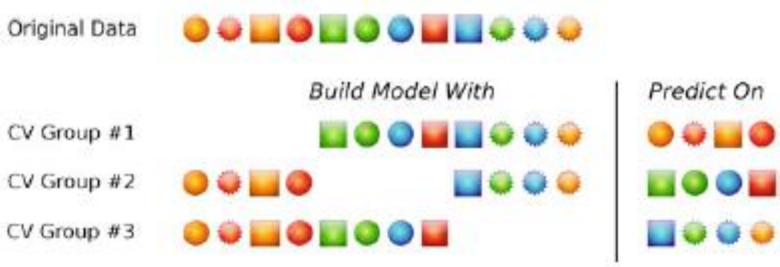

Figure 7 - Division of a dataset using a three-fold cross-validation technique, in (Han et al. 2012)

Comparing these two methods Hawkins et al. (2003) affirmed that the hold-out methods using "samples of tolerable size do not match the cross-validation itself for reliability in assessing model fit and are hard to motivate".

Having understood the concepts of BDA and ML, it is crucial to highlight that BDA, and consequently ML, alone has only a marginal effect on the environment it is being applied in. Every ML model created is only a piece of a business process. Therefore, several methodologies have been developed in the literature to align a BDA / ML projects with the business processes and environment of a company. The two most popular methodologies are the CRISP-DM methodology which stands for Cross Industry Standard for Data Mining and the Knowledge Discovery in Databases methodology (KDD).

CRISP-DM breaks down a project into 6 different phases (business understanding, data understanding, data preparation, modeling, evaluation, and deployment) as depicted in Figure 8.

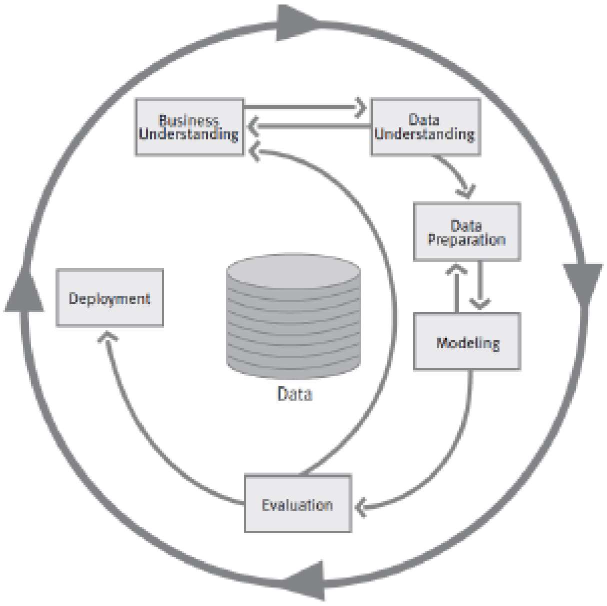

Figure 8 - Simplified view of the CRISP-DM methodology, in (Wirth and Hipp 2000)

It is essential to understand that the order by which each phase is performed is not strict. The arrows merely indicate the "most important and frequent dependencies between phases". The outer circle represents the continuous learning attained from every concluded project, from which future ventures may benefit from (Wirth and Hipp 2000) .

While KDD is less focused on the understanding of the business as a whole, it prioritizes the data preparation and modeling phases. Depending on the literature, the steps that compose the KDD methodology may vary. In their textbook, Han et al. (2012) identified seven different steps that one should follow when building a ML algorithm. The data selection phase may be incorporated into the data understanding process of the CRISP-DM methodology. The following three steps belong to the data preparation phase and are as follows: data cleaning, data integration, and data transformation. As for the remaining three steps – date mining, pattern evaluation, and knowledge presentation – they are equivalent to the modeling, evaluation, and deployment phases.

3 Methodology

The following chapter aims to better define the methodology that will be followed throughout the development of a solution for the problem exposed in the Introduction chapter – improve the two-hypothesis shipping fee testing method. Choosing a methodology to follow is extremely important because it makes specialized processes easier to finish. A methodology may be compared to a checklist of tasks to be performed in order to achieve an end result. Although a methodology simplifies complex problems, finding the right level of granularity to achieve the best results with the least possible effort may be a difficult task.

To make an informed decision on which methodology to use to address the problem in hand, two main methodologies were reviewed, CRISP-DM and KDD. Even though they are different in their approach to ML problems, it can be argued that they complement each other. Therefore, the methodology thought to develop this thesis was a combination of both those methodologies.

The first step of the thought methodology is in line with CRISP-DM's business understanding phase. This initial phase consists of understanding the business and the problem's objectives. With the objectives defined, it is important to identify where possible causes for the given problem reside. A crucial element of this phase consists of analyzing in further detail the identified causes and how they affect the company. This effect is measured using metrics, previously established by the company.

Data understanding happens simultaneously with the business understanding step once it is required to have at least some understanding of the data in order to define a problem. This is also in line with CRISP-DM's usual second phase. First, the data is collected in accordance with the established problem. Secondly, a quality assessment of the data is made in order to understand which variables will prejudice the model's construction. Then unambiguous and uncorrelated variables are selected – data selection phase from the KDD methodology – and finally, an exploratory analysis is performed.

The third step is data preparation. Parallel to what was described in the Literature Review chapter, this step may be divided into three tasks as defined by the KDD methodology. Data cleaning is a result of the previously performed exploratory analysis and consists of the removal of unwanted outliers and of other data points that may introduce noise into the dataset. Data transformation consists of altering the data in order to make it compatible with the chosen ML algorithm. As for data integration, it only happens when the data was collected from multiple data sources and needs to be combined. These tasks may be performed in any order and more than once if deemed necessary.

For the modeling phase, the ML model type (for example, clustering or classification) that best suits the problem is chosen. After, several different modeling techniques comprised in the model type chosen are used to build different models. Some of the different techniques require specific data formats. These requirements show the link that exists between data preparation and modeling. To correctly assess which model should be used as a solution for the ML problem, the models' parameters should be optimized – tuned – so that the values returned by the models are the best possible.

The next phase lies in evaluating the built models and choosing the one with the best performance. To do this, it is necessary to define which metrics should be considered and by which order of importance. The model that shows the best results for the selected metrics is chosen. Once a model has been opted for, it is once again assessed, this time using unseen data. The usage of unseen data allows for a better perception of what the chosen metrics would look like if the model was applied in new data.

The final step of the methodology is the deployment of the model. The attained knowledge needs to be explicitly represented so that the end user is able to understand it. For the development of this thesis, deployment was considered to be the creation of a usage roadmap. The real implementation of the model will require the conception of a repeatable data mining process within Farfetch's platform infrastructure. Like in many other ML problems, the deployment of the model is rarely carried out by the analyst who conceived it. For example, in Farfetch's case, the Business Intelligence team would be the one responsible for the deployment steps.

It is important to understand that just like in the CRISP-DM methodology the order by each phase is performed is not strict. For example, one can build a model, understand that a variable may be wrongly collected and then go back to the data selection task (part of the data understanding step).

The proposed methodology is depicted in Figure 9.

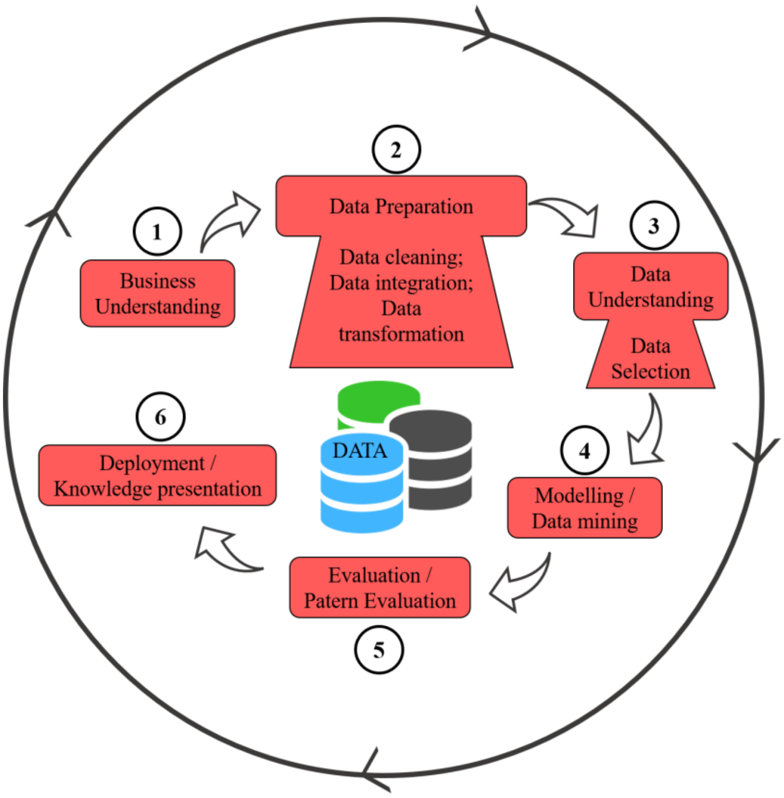

Figure 9 - Proposed methodology based on the KDD and the CRIP-DM methodologies

As it was proposed, this methodology combines CRISP-DM and KDD. Similar to CRISP-DM, the inner arrows and the numbers represent only the most frequent order in which each phase should be executed. Analogous to CRISP-DM the external arrows represent the fact that with every ML project concluded there is knowledge, that may be passed to the next ML project.

4 The case study

4.1 Farfetch ecosystem

Up until 2014, Farfetch had a single focus, which was its marketplace. From 2014 onwards, the idea of Platform thinking was embedded in the company's culture. As Sangeet Choudary (2016) wrote: "A platform thinking approach to building a business involves figuring out ways by which an external ecosystem of developers and users can be leveraged to create value". This new notion shifted Farfetch's focus. In Farfetch's and many other companies' cases, this ecosystem is composed of platform infrastructures (services, data & API), applications, a community of participants, and the transactions that occur between them. For Farfetch, these applications are, for example, Black & White websites, Store of the Future products, the Fashion Concierge app as well as the Farfetch Marketplace.

Black & White websites offer a white labeled website solution for luxury brands. Store of the Future has the objective of bringing together the in-store luxury experience with the knowledge that online acquired data provides. The Fashion Concierge app allows for clients with the highest annual spending to purchase exclusive and difficult to find products from non-Farfetch sellers. As for the Farfetch marketplace, it connects creators, curators, and consumers from all over the globe (Farfetch 2019). These applications are all enabled by the Farfetch platform infrastructure, as shown, in a simplified way, in Figure 10.

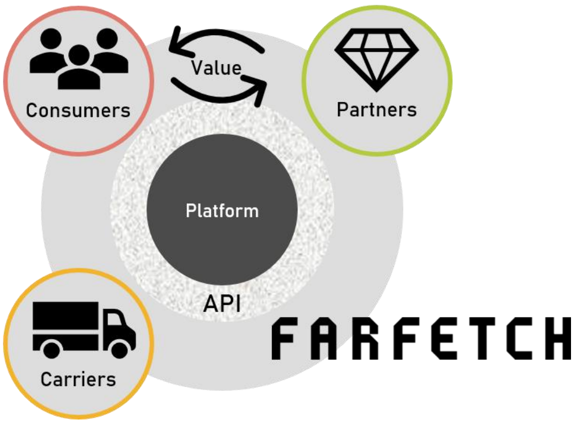

Figure 10 - Simplified view of Farfetch's platform infrastructure, adapted from (Walter 2017)

In the interest of this thesis, only the Frafetch Marketplace application will be studied.

4.2 Farfetch marketplace

As previously mentioned in the Introduction chapter, Farfetch does not own any of the stock it sells. However, it manages to have over 370k different products available online. This is only possible due to the partnerships it establishes. Throughout the years, Farfetch partnered up with boutiques – high-end luxury venues – department stores and most recently with brands.

Farfetch main strength is its stock breadth – product variety – accomplished by partnering up with over 800 boutiques all over the world. One of Farfetch's most recent acquisitions, Stadium Goods, also increases this breadth. As stated in the introduction, Farfetch grew very rapidly in the past 10 years. Alongside this growth came the need to increase Farfetch's stock depth – product quantity. To do so, Farfetch partnered up with both brands and department stores. Although these partnerships increased Farfetch's stock depth, it is still falling behind in this field when compared to its direct competitors. In May 2015 Farfetch acquired Browns, a London based boutique and today it represents about 8,41% of all Farfecth's online sales. This acquisition not only allowed to increase Farfetch's stock depth, but it also made it possible to better understand the in-store luxury consumer behaviour. Browns, being a 1st party business, also acts as a safeguard from the fact that Farfetch is an all third party business- financial/ reputational, legal, regulatory and operational risks (Scott and Spitse n.d.) may be minimized.

Being the Farfetch marketplace a part of the Farfetch platform, it is an omnichannel service (present in more than 5 different channels), accepting 19 different payment methods (data referring to 2018) and offering 4 different delivery methods.

Due to its worldwide presence, lack of stock and different delivery methods, Farfetch faces an augmented complexity of the delivery process. To tackle this issue, Farfetch uses a drop-shipping method with its tailor-made ordering process.

4.3 Ordering process

The tailor-made ordering process plays a crucial role in the shipping price calculation. To understand its impact, one must first fully understand the ordering process. This process was developed by Farfetch's CEO and founder, José Neves, during the early stages of the company, and it is one of the operational baselines around which the company revolves. This process starts when a consumer places an order, defined as a Portal Order (PO) – commonly referred to as the consumer's basket. As a result of Farfetch being a multi-partner marketplace, a PO may be divided into one or more Boutique Orders (BO) – set of products that are sold by the same stock point – which may, in turn, be comprised of one or more items as depicted in Figure 11.

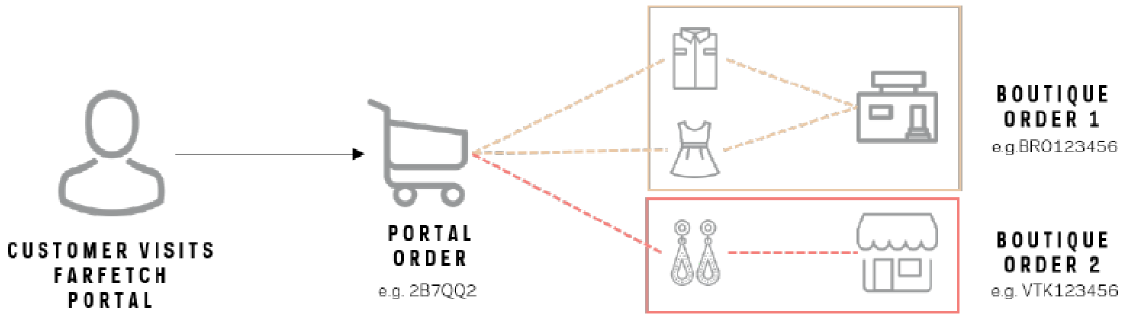

Figure 11 - Composition of a portal order

Once an order is placed it must go through 6 different steps before it is delivered to the client. The steps are as follow:

- Step 1 – Partner checks whether there is stock;
- Step 2 – Farfetch's fraud team approves or not the payment (happens simultaneously with step 1);
- Step 3 – Partner decides the packaging;
- Step 4 – Partner creates shipping label – air waybill (AWB);
- Step 5 – Partner sends a parcel through a carrier that also has a partnership with Farfetch;

- Step 6 - The parcel is in transit until it is delivered.

There are 3 auxiliary steps that are not directly related to the ordering process. The first one happens before an item is placed online in which partners send packages (slots) with up to 50 items they would like to sell on Farfetch's marketplace to a Farfetch production center – a Farfetch facility where the items that are sold online are photographed in a homogeneous way to guarantee a consistent image across catalogs and classified. The other 2 steps happen when a customer wishes to make a return. First, the client sends the item back to Farfetch, and secondly, Farfetch accepts or refuses the return, which is translated into refunding the client or not. The entire process along with the teams in it involved is represented in Figure 12.

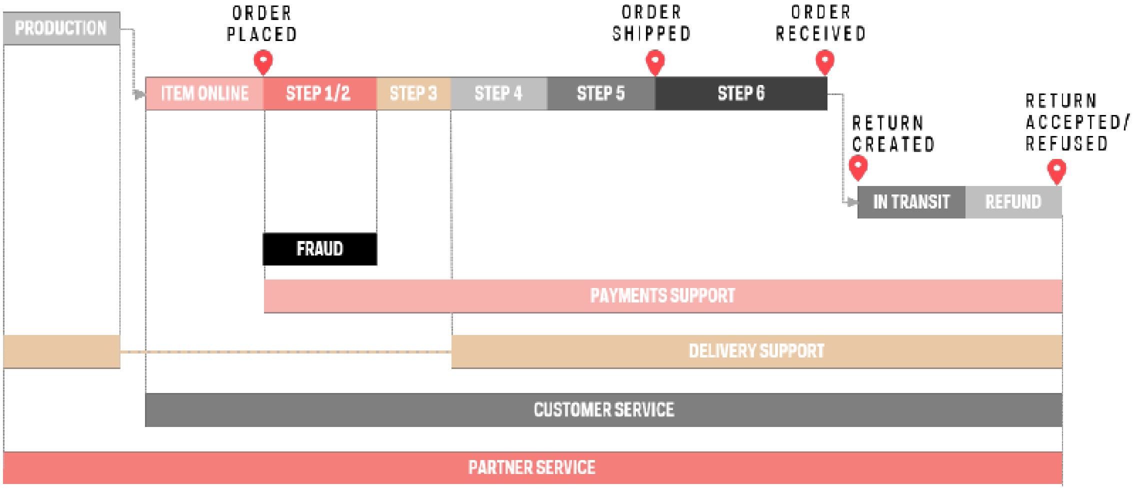

Figure 12 - Entire process since an order is placed, auxiliary steps and teams involved

4.4 Farfetch customers

One of the most important goals of the multidimensional operational process referred to in the previous chapter is in line with one of the already mentioned Farfetch value: "Amaze Customers". This goal is also related to both reducing costs and increasing revenue. This objective is very similar to the main goals of many other e-luxury companies.

To attain this goal, firstly, it is necessary to understand who Farfetch clients are. Therefore, Farfetch conducts regular surveys related to its consumers. Based on the most recent survey, the average Farfetch client is around 36 years old, with 54% of Farfetch customers being either millennials or from generation Z. They are usually female, with only 33% of male consumers. About 50% of the consumers are either married or sharing a home with a partner with an average household income of $120,312 and do not have children (only 38% of the respondents affirmed to have kids). About 73% of consumers own a house rather than renting one, and on 8% of the cases, Farfetch clients are homemakers. Regarding their occupation, almost 75% of the clients are employed, and 13% are still studying.

Although it is crucial to comprehend who the average Farfetch client is, it is also essential to recognize that not all customers are the same. Farfetch understands this, so it created a customer loyalty program, the Access Program, like many of its competitors. Even though each client has numerous specific characteristics, the one they all have in common and also the easiest one to account for is the money each one of them spends on the Farfetch website. Therefore, the Access Program divides consumers into tiers dependant on their annual spending. There are five tiers – from Tier 4 to VIP, being the Tier 4 clients the ones with lower annual spending and VIP clients the ones with greater spending (over $12,000 per year) - each one of them with different benefits. In the interest of this thesis, the most relevant benefit to highlight is that VIP clients are always given free-shipping regardless of their basket value and country. An interesting insight this program helped realizing was that even though VIP clients only represent 0.8% of all customers, they are responsible for 19.9% of all of Farfetch's GTV.

Another critical element in understanding e-tail company clients is to map their journey from the moment they enter the website up until they either convert – buy an item – or leave the website without having made any purchase. For most e-tail websites a simplified view of his journey starts on the landing page that often matches the "Homepage". Then follows the "Product List Page" (PLP) – page in which all the resulting items of an internal search are shown – and the "Product Description Page" (PDP) – page showing the selected item in more detail. The "Shopping Bag" page is the next page on the customer journey, which is where the consumer sees what he/she has decided to buy so far and the total value of his basket. The "Checkout" page is the final step where the buyer decides on whether to make a purchase or not. For Farfetch the "Checkout" page is referred to as the "Review" page, where, as the name suggests, the consumer reviews his/her purchase and decides on which shipping method better suits him/her.

As happens in many other e-commerce companies, a client that enters an e-commerce website does not always convert. For example, luxury e-tailers average conversion rate – the number of purchases divided by the total number of visits to the website – can be as low as 0.5% - 3%. On certain occasions, however, the conversion rate may not always be measured as whether a client bought or not an item but as whether a client moved to the following page or not – number of visits that advanced to the next page divided by the total number of visits of that page. This notion of conversion rate is helpful when it comes to understanding ways of either reducing costs or increasing revenue, the second part of the aforementioned goal.

To tackle conversion issues in a more objective and efficient way, it is essential to understand where the problems lie. By analyzing the conversion rate in each webpage separately it becomes simpler to identify where customers leave the website. Consequently, identifying where the main issues concerning customer experience and revenue loss reside also becomes more straightforward. The conversion rate in each step of the customer journey on Farfetch's website is portrayed in Figure 13. To simplify the analysis made, the PLP and the PDP were analyzed together.

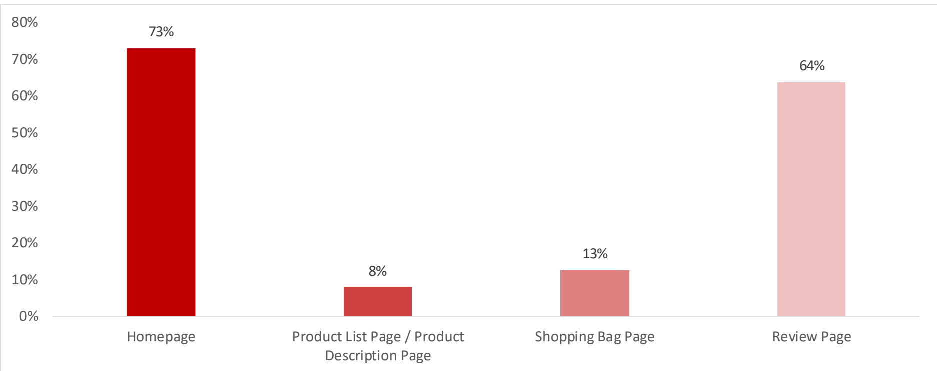

Figure 13 - Conversion rate by page

By analyzing Figure 13, interesting conclusions may be drawn. The most critical webpages regarding the conversion rate are the PLP and the PDP and the Shopping Bag page. From all the people that reach the PLP or the PDP during their customer journey only 7.80% go to the shopping bag, and of those only 12.52 % do not abandon their cart. Regarding the review page, even though the conversion rate assumes a value over 60%, this is the final touchpoint between Farfetch and its clients before a purchase is made. Therefore, constant efforts are being made to improve this rate to higher values. As for the Homepage, 73% of all consumers move to either a PLP or a PDP.

The Shopping Bag page and the Review page are closely connected having numerous elements in common. By tackling an issue in one of the pages, the same issue is being tackled on the other page. This makes the Shopping Bag page and the Review page more attractive when it comes to implementing changes. One element that is under Farfetch's control and only present on those pages is the shipping price, so the necessity of this thesis becomes clear.

4.5 Shipping fees

As mentioned in chapter 2.2 not all clients are the same. However, what all e-commerce companies' clients have in common is that every time they decide to make an online purchase, they will always buy or be offered (in case of free-shipping) a shipping service. Also, the shipping price is an important element of the previously mentioned Shopping Bag page and Review page. In both of those pages, the shipping price is presented to the consumer as shown in Figure 14.

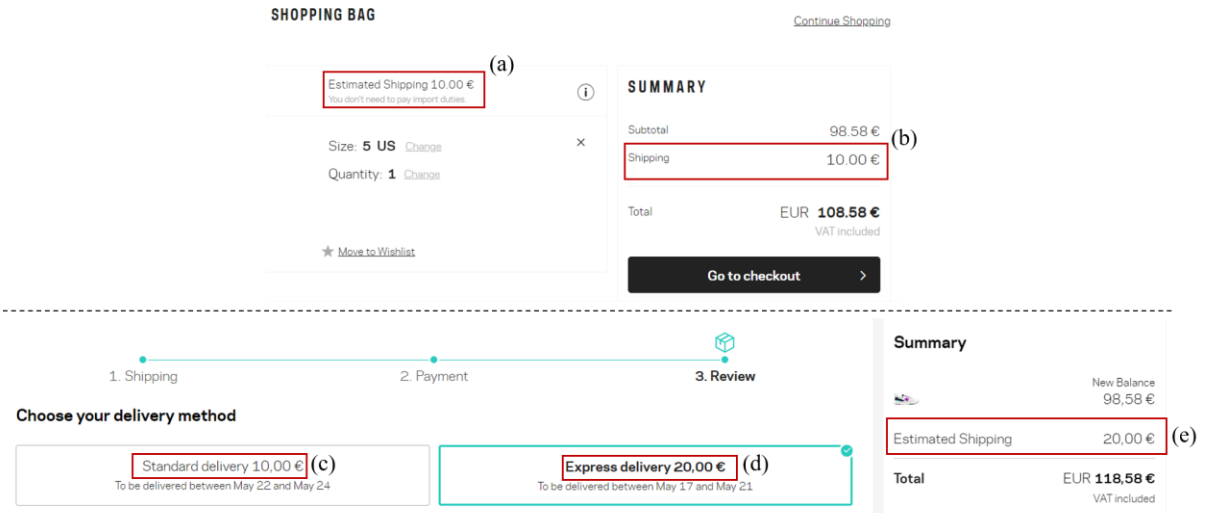

Figure 14 - Shipping information on the Shopping Bag page (upper Figure) and on the Review page (lower Figure)

The shipping price is conveyed to the consumer using the highlighted information in red.

By observing these figures, some problems are quickly identified such as the fact that on the same page the shipping fee is shown as both an estimation (a) /(e) and as a fixed price (b) /(c) /(d). This happens in both the Shopping Bag page and the Review page. The mismatched information may result in a lack of trust from the final consumer that is already prone to see shipping fees as being unfair and wrongly charged as referred to in the Literature Review chapter.

Also, Farfetch belongs to the large group of e-commerce companies described in the literature that do not take advantage of both the consumer's related data and of their customers' heterogeneity. To minimize customer distrust while increasing revenue, two main actions may be taken on the Shopping Bag page and the Review page. First, one should utilize consumers' related data to alter shipping fees in the most cost-effective and faster way. Secondly, what in every webpage is causing the consumers' mistrust should be identified and acted upon.

Because all companies' websites are different, identifying what may cause consumer mistrust on the Farfetch website will not result in interesting findings for the luxury e-commerce retailers. However, understanding consumer's price sensitivity to shipping fees on the e-luxury world, and using that information to alter delivery prices easily may result in interesting conclusions applicable not only to Farfetch but to almost every e-luxury company. Before studying price sensitivity and suggesting ways on how to apply those findings to alter the way shipping fees are being tested, it is necessary to understand both how shipping fees are being obtained in the case study company, Farfetch, and understanding how the evaluation of changes to those fees is being made.

4.5.1 Shipping fees calculation

Farfetch is a marketplace, and because it has partners all over the globe the number of different routes performed between partners and clients is higher than if it had its stock stored in a warehouse, like many of its competitors. Therefore, the shipping costs incurred by Farfetch are higher than the ones incurred by its competitors. Due to Farfetch's condition as a marketplace every time a PO is a multi-partner order (composed by more than one BO), there will be as many shipments made to the consumer as the number of BOs. This may be a critical situation given that each BO only represents, on average, 1.09 items. Therefore, each item on the consumer's basket will most likely represent a shipment. In the competitors' case, regardless of the number of items ordered, only one shipment will be necessary, assuming all items come from the same warehouse. Confirming that the logistics and costs involved in this process are higher than usual.

To minimize the effects of the high shipping costs while striving for competitiveness, Farfetch developed a personalized method of calculating shipping fees. This method takes into consideration both direct and indirect variables. Direct variables are explicitly explained to the consumer and are: the customer country, the customer basket value, and the delivery method chosen. Indirect variables are neither explained to the consumer nor directly related to the shipping price charged. However, as stated in the Introduction chapter, Farfetch covered 31% of its shipping costs in 2018 and has an objective of maintaining this ratio constant or decreasing it like it did in the past. Any increase in this ratio should be explained by an increase in Farfetch's net income. The ratio evolution is represented in Figure 15.

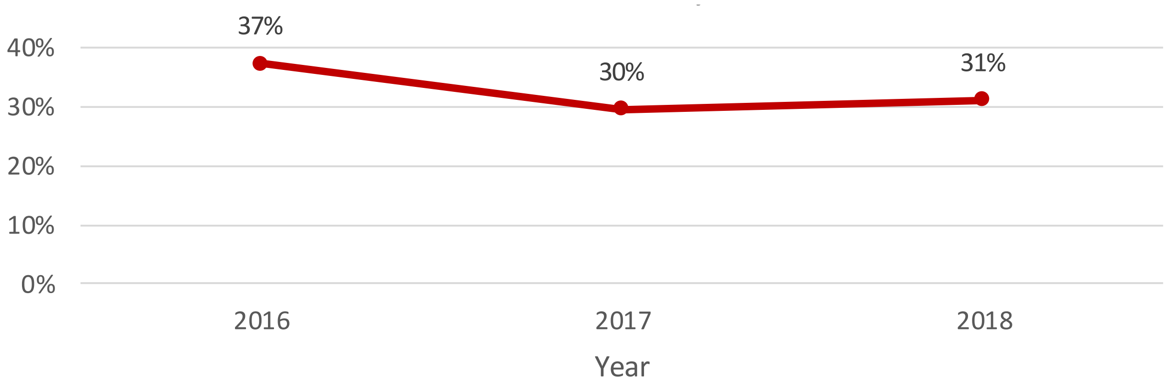

Figure 15 - Percentage of shipping costs suported by Farfetch

The method considers a basket value threshold, up until which the consumer pays a shipping fee for each BO in his/her basket. The threshold is constant for each country and independent of the shipping method chosen. Above the threshold, the customer pays a fixed fee – Flat Rate (FR) – regardless of the number of BOs in the order he/she placed. Some countries (United States, Russia, China, and Hong Kong) have free shipping above another basket value threshold. Because this method is not a common practice among other companies, the consumer is most likely not used to it, possibly adding an extra layer of mistrust between Farfecth and its customers. The shipping fee calculation method is presented to the consumer, as depicted in Figure 16.

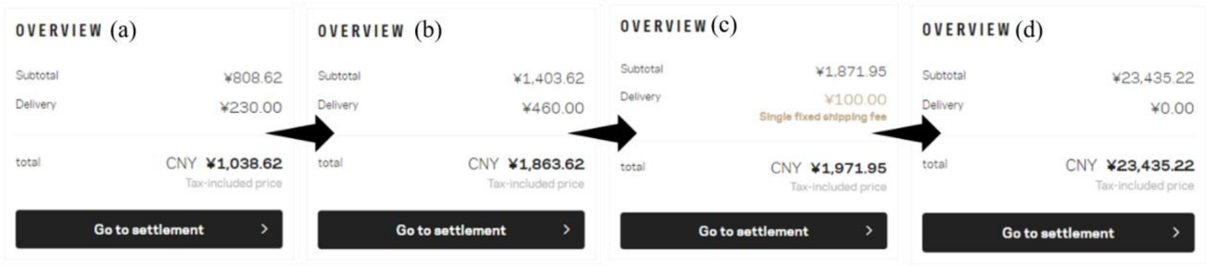

Figure 16 - Shipping fee calculation method as conveyed to the consumer

Suppose a consumer enters the Farfetch website and adds a product to his shopping bag amounting to a total basket value lower than the FR threshold. This scenario would result in him/her paying one shipping fee (sub-Figure (a) – ¥230) for the shipment of the order. In sub-Figure (b) the consumer has 2 products in his/her basket that still do not amount to a total value higher than the FR, therefore he/she pays a fixed shipping fee times the number of BOs in his basket (¥230*2=¥460). Once the consumer adds the third product to his/her basket, the basket value will be higher than the FR threshold but still lower than the free-shipping (FS) threshold – sub-Figure (c). This means that the shipping fee charged will be a fixed fee (¥100). In the last scenario represented by sub-Figure (d), the basket value is greater than both the FR threshold and the FS threshold. Therefore, the consumer will be offered a free shipping discount. The shipping fees could have been different if the shipment method chosen was different. However, the thresholds would have remained the same.

Farfetch has a free returns policy in which consumers may return any item in good conditions up until 14 days after receiving it, free of any charges and with the guarantee of a full refund (inclusive of duties and taxes if applicable) excluding the shipping fee initially paid. Due to this policy, the question has been raised of whether people are buying more products than the ones they want, to have a basket value higher than a given threshold. However, Farfetch's return rate (about 20% for 2018) is lower than other e-tailers return rate. Nevertheless, this doubt may be supported by the orders distribution regarding the basket value depicted in Figure 17 for the year 2018.

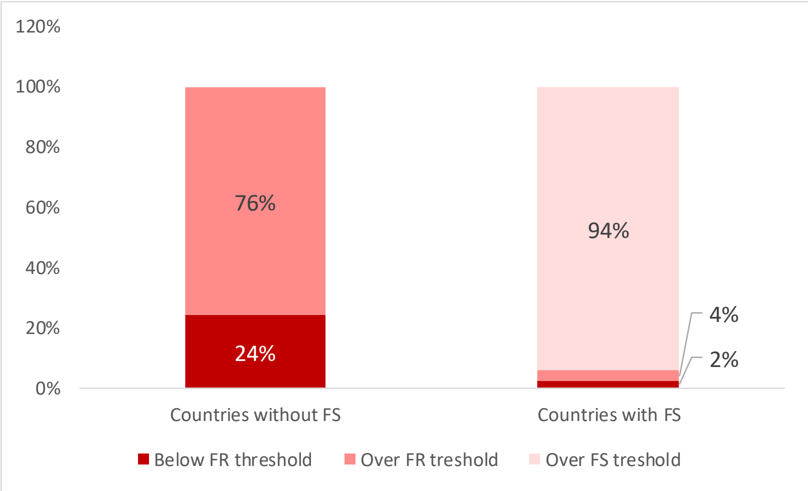

Figure 17 - Percentage of orders over and below the described thresholds for countries with and without free-shipping

The high amount of orders over the thresholds may in part be explained by the fact that Farfetch's average order value (AOV) is, for all countries, higher than the FR and the FS thresholds but also by the previously described phenomenon.

Despite the order distribution, Farfetch is still a better case study company than its competitors to understand shipping fees price sensitivity. As a result of its shipping fee calculation model, Farfetch has data on how the same customer reacted to different shipping prices for a given set of conditions making the price sensitivity study more reliable.

4.5.2 Shipping costs and revenues structure

To better understand the possible impacts that altering the shipping fees may represent for Farfetch, it is necessary to comprehend the shipping Profit and Loss statement (P&L). Foremost, it is essential to understand that any alteration made to the way shipping fees are being calculated will either have a positive or a negative impact on the shipping P&L – the shipping fees paid by the consumers represent part of the revenues of the shipping P&L. It is also vital to notice that a negative impact on the shipping P&L does not necessarily imply a negative impact on Farfetch's overall P&L. For example, a decrease in shipping fees – a decrease in the net value of the shipping P&L – may result in higher conversion rates, representing an increase in Farfetch's GTV, having a positive impact on its P&L.

Three main elements compose the shipping P&L: the value charged by the carriers, the shipping fees, the shipping subsidy, and the shipping fees charged to the partners.

The value charged by the carrier is a fixed value per route, weight, and service (for example, express service) that is subject to alterations when prices are negotiated with the different carriers. An increase in this value harms the P&L and vice-versa, therefore representing the costs on the P&L.

Shipping fees are paid by the consumer and are perceived as an extra cost paid for a shipping service. The consumer unconsciously pays the shipping subsidy. It represents a certain percentage of the value of the product and is already included in the item's final price. This value is always charged independently of the total basket value. The shipping subsidy is used to cover free shipping expenses. The shipping fees charged to the partners are used to support the shipping cost associated with the free returns policy. Both the shipping fees charged to the consumers and the partners and the shipping subsidy represent the revenues on the P&L, meaning that an increase of those values has a positive impact on the final shipping P&L and vice-versa.

The final value of the shipping P&L statement – net income – is a result of the costs mentioned above and revenues (net income = revenues – costs). The revenue distribution is as depicted in Figure 18.

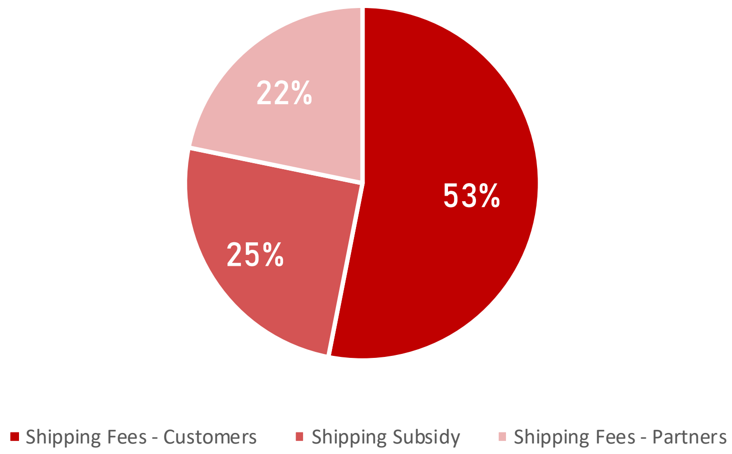

Figure 18 - Average shipping revenues distribution for the year of 2018

When analyzing Figure 18, it becomes clear that the shipping fees represent the majority of the shipping income. Altering those fees will have an impactful repercussion on the shipping P&L, possibly causing the increase of the ratio between shipping costs and shipping income described in chapter 2.2. To minimize possible negative impacts of altering shipping fees, Farfetch performs tests before implementing any change.

4.6 Current shipping fee testing method

The current method used by Farfetch, to test whether an alteration to a shipping fee is profitable or not is the A/B test method. An A/B test consists of comparing two versions of the same variable on a webpage or app and determining which one performs better for a given objective. This can be applied to testing alterations on an already existing variable or the creation of a new variable. For this thesis purpose, the variable to be altered is the value of the shipping fee. To do this, consumers are randomly chosen to belong to either a control or a test group. The consumers in the control group will be exposed to the current value/ aspect of the variable that is being tested, and the consumers in the test group will be shown the new value for that variable. The test and control group are similar in size. Once the test returns significant results, it is stopped, and the choice of altering or not the value of that variable becomes a more informed decision.

Farfetch's and many other luxury e-commerce companies' objective is to increase profit. The best way to do it is by increasing revenues by increasing sales, commonly known as top-line growth (Murphy 2019). In Farfetch's case, one way of doing so is by decreasing the value of shipping fees, which may, in turn, result in a higher conversion rate for both the shopping bag page and the review page. This increase will most likely lead to an increase in revenue and therefore, in profit. Another method is increasing the shipping fees value, which will increase the shipping income and therefore, the final profit of the company. The two previously described scenarios may not always be true, so the need to perform A/B tests arise. To divide the population to be tested, Farfetch randomly chooses which customers belong to the test and the control groups. To test if the change on the value of a shipping fee was successful, equation 4.1 is taken into consideration.

$$PP*GTV_{Extra}-Costs_{Extra} \tag{4.1}$$

Where:

PP, is the GTV percentage of profit
GTV_{Extra} is the difference between the GTV on the control group and the GTV on the test group
$Cost_{Extra}$ is the difference between the costs on the control group and the costs on the test group

If the value of the equation 4.1 is greater than zero then the A/B test is considered successful and the change made to the shipping fee is applied for all the tested population. Otherwise, the A/B test is considered unsuccessful and the changes are not applied. Implying one of the following:

- When the shipping fees are decreased the extra profit generated needs to, at least, cover those costs;
- When the shipping fees are increased, the loss in profit (PP*GVTExtra will assume a negative value) needs to be equal or lower than the extra shipping income (-CostsExtra assumes a positive value).

4.6.1 Problem and motivation

As aforementioned, Farfetch randomly chooses which consumers are to enter the A/B tests. This implies that both consumers who are going to change their attitude towards converting on an individual webpage and those who are not, are considered. Taking into account both types of consumers will cause the A/B test to return significant results in a longer time. It will also imply the loss of the shipping revenue of those consumers who would have converted anyway (when the alteration made to the shipping fees is a decrease in their value).

The online environment is known for being forever changing and growing, and online marketplaces are no exception. Along with this growth comes the appearance of several competitors for e-tailers. This fast-paced environment requires that any alteration made should be as fast as possible, in order to avoid any possible losses to competitors.

The objective of this thesis is to build a model that classifies consumers into possible converters - customers who are likely to convert with the current shipping price - or not. This will allow for Farfetch to make a more appropriate selection of which consumers should be considered or not in an A/B test:

- For a decrease in shipping prices, only those consumers who are unlikely to convert with the current shipping price should be considered;
- For an increase in shipping prices, only those consumers who are likely to convert with the current shipping price should be considered.

This will lead to faster results and to fewer costs incurred by Farfetch when implementing an A/B test making any alteration made to shipping fees more competitive.

5 Proposed solution

To follow the methodology described, it is not only necessary to understand a given business. It is also essential to comprehend what problems the business has and to define what path is going to be taken to tackle those problems. Throughout the development of this thesis, the path considered had the main objective of creating a model to understand how likely a person is to convert on both the Shopping Bag page and the Review page. One of the variables under Farfetch's control for both the studied pages is the shipping price. The shipping price is assumed to have a significant impact on conversion. This variable is also one of the most easily exchangeable variables and therefore, one of the most interesting to study the impacts of changing it. Thus, even though the ML algorithm will study conversion specifically, a special focus was given to shipping fees.

The output of the model to be created is a dependent variable assuming two values – did convert and did not convert – i.e., a series of inputs will be classified. To create the model, the input of whether someone converted or not will be given. Hence, the technique used will be a supervised ML technique. The ML technique that best suits these features is binary classification models. The final classification algorithm has two main objectives:

- Determine which variables are most relevant when it comes to conversion;
- Implement more efficient shipping fee A/B tests.

5.1 Data understanding

5.1.1 Data collection

Before building a classification model, it is necessary to construct a dataset in line with the model's purpose. The model's objective is related to understanding if, during a session, a user will convert on a particular page or not. For the year of 2018, an average of 909,801.8 sessions per day was registered. To process such high amounts of data in a reasonable time, Google BigQuery was used to retrieve the dataset.

When constructing the dataset, the first decision to be made is to decide what each row will represent. The model to be created is a classification algorithm that divides user sessions into sessions with conversion or not for a given page. Consequently, it was logical that each row should represent a session. Because two different pages were being studied – where the consumer interacts with shipping prices – two different datasets were created: one related to the Review page, and one related to the Shopping Bag page. For both datasets, each column represented a session attribute, and the variable ConvertReview / ConvertShopBag represented the dependent variable. A two-level factor variable that assumed the values of 1 or 0 depending on whether a conversion happened (0) or not (1).

Given that this thesis' primary objective is understanding how the shipping price influences conversion, only sessions in which something was added to the shopping bag, and therefore, a shipping price could be calculated were considered. Due to the database limitations, only sessions in which something was added to the bag on the product page were considered. Hence, customers who added something to their basket on the checkout page were disregarded. VIP clients were not considered because as referred in chapter 4.4 they are always given free shipping regardless of their session characteristics, meaning that their shipping price sensitivity could not be studied nor modeled. Sessions classified as bots, which are "software applications that perform automated tasks over the Internet" (Cambridge 2008) were also disregarded alongside sessions with a duration of 0 minutes. From the remaining sessions, some contained more than one shipping offer – meaning that two or more different shipping prices were presented to the user. Because it is impossible to infer what shipping price the customer would have chosen in case of non-conversion and because the vast majority of clients are only presented one shipping method – Express – only those 73.77% of clients were considered. For each of the created datasets, only sessions in which the consumer got to either the Shopping bag page or the Review page were considered. Despite all of these limitations, both the Shopping Bag and the Review datasets still contained 2,522,979 and 889,670 entries, respectively. All the initial variables considered (51) are represented in Table 1 and Table 2 for both datasets.

Table 1 - Datasets variables (1)

Name	Type	Description
Month	Categorical	Months 1 to 12
VisitorType	Categorical	New or Returning depending if it is the first time on the website or not
CustomerType	Categorical	Customer or Prospect depending on if a purchase as already been done or not
DeviceType	Categorical	Tablet, Mobile or Desktop depending on where the website / app is being used
DestCountry	Categorical	A list of 57 distinct countries and a level "Others"
Channel	Categorical	A list of 9 distinct ways of entering the Farfetch website / app
Gender	Categorical	Men, Women or Unknown
InteractWithPromoCodes	Categorical	Yes (1) or no (0)
AddToWishList	Categorical	Yes (1) or no (0)
InteractWithPhotos	Categorical	Yes (1) or no (0)
TimeSessionMinutes	Numerical	Length of a session
TotalClicks	Numerical	Total clicks during a session
DaysSinceLastVisit	Numerical	Days since the customer last came to the website/app
BasketValue	Numerical	Sum of the value of all the items in the basket (£)
NProducts	Numerical	Number of products in the basket
AvgPrice	Numerical	Average price of the items in the basket (£)

Table 2 - Datasets variables (2)

Name	Type	Description
MoreIncomeBrands_1 …	Categorical	Brands that bring the most income; From 1 to 5 being 5 the ones that generate more revenue
MoreProductsBrands_1 …	Categorical	Brands that sell more products; From 1 to 5 being 5 the ones that are most sold
MostSelledProducts_1 …	Categorical	Products that are sold in more quantities; From 1 to 5 being 5 the ones that are most sold
Gender_K …	Categorical	Gender of the products in the basket (K-kids, W-Women, M-Men)
Animal	Categorical	If the basket contains an animal product or not
Vintage	Categorical	If the basket contains a vintage product or not
Customizable	Categorical	If the basket contains a customizable product or not
ProductCategory_Accessories …	Categorical	Category of the products in the basket - 8 distinct categories and a label “Others”
Week	Categorical	Weeks from 1 to 52
ShippingFee	Numerical	Shipping fee to be paid by the consumer (£)
ConvertReview / ConvertShopBag	Categorical	1 or 0 depending on if the consumer converted (0) that page or not (1)

Classifying whether there was a conversion on the Shopping Bag page and on the Review page is fairly simple. To convert on both pages, the customer needs to go to the following page instead of either leaving the website/app or going to a previous page. The classification process is represented in Figure 19.

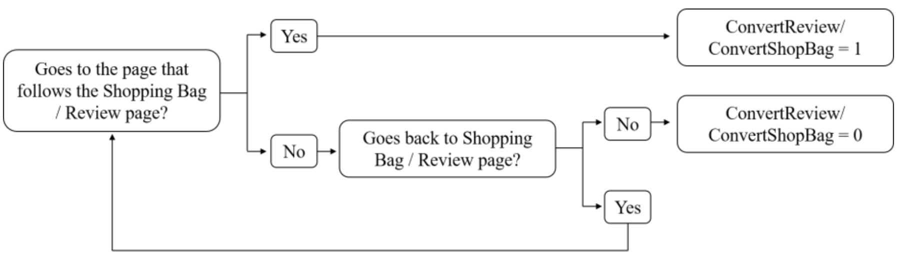

Figure 19 - Criteria for defining if a conversion happened or not

5.1.2 Datasets analysis

After building the datasets, it is essential to understand if the variables chosen to construct the model are the most indicated ones or not. To do this, first, an exploratory analysis on the entire datasets was performed. As a result of this analysis, some variables were removed from the initial datasets. Afterward, an exploratory analysis on both the Shopping Bag dataset and the Review dataset was performed to better understand how each variable would impact the conversion rate on both those pages. Some entries were excluded from both datasets as a result of the exploratory analysis. All the data handling from this chapter onwards was performed with the programming language R, version 3.6.0, through the free version of RStudio interface available at https://www.rstudio.com/products/rstudio/download/

Data assessment and selection

"TimeSessionMinutes" was a variable of interest because it was initially assumed that when a person makes a purchase – converts on the Review page – they also conclude their session. To test this assumption, it was presumed that both the length of a session until something is bought and the full length of a session, followed a normal distribution. After, two hypotheses were tested using a one-sided t-test statistic with a confidence level of 95%, where:

H0: µFull session in minutes = µSession until purchase in minutes

H1: µFull session in minutes > µSession until purchase in minutes

The variances of both the full session time and the session time until purchase were estimated using the Welch-Satterthwaite approximation to the degrees of freedom. The p-value obtained for a subsample of 500 entries was lower than 0.05 for the Review dataset, thus rejecting the null hypothesis. Due to these findings, when the variable "ConvertReview" is equal to 0, the variable "TimeSessionMinutes" was altered to only account for the time spent on the website/ app before the purchase was made. Due to the limitation imposed by the data collection, it is not possible to correctly assess the exact moment when someone converted on the Shopping Bag page. Therefore, the variable "TimeSessionMinutes" may introduce a certain degree of error in the final shopping bag model.

Throughout the year of 2018, due to a technical error Farfetch website was displaying products whose real value was of over $500 and was selling them for less than $1. All sessions in which this was verified were removed from both datasets.

Upon studying and altering the variable "TimeSessionMinutes" and removing unviable sessions, a correlation study was made to test if all the variables contained in the datasets were relevant. First, a correlation test using the Pearson coefficient and the Pearson coefficient as an approximation of the Phi coefficient (Gendy and Phys 2006) was made between binary and continuous variables. The results of the test are represented in Appendix A for the Shopping Bag page in Appendix B for the Review page.

According to Schober and Schwarte (2018), only variables with values of correlation over 0.9 are considered to have a "very strong relationship". Variables with a value of the correlation between 0.1 and 0.9 should be further analyzed to decide whether they should belong to the dataset or not. Taking a conservative approach to these values, variables with a correlation coefficient higher than 0.8 were removed, and all the others were kept. Table 3 shows the removed variables and their correlation coefficients with the kept variables for both datasets.

Table 3 - Kept and removed binary / numeric variables and their correlation coefficient, for both datasets

Removed Variable	Kept Variable	Correlation Coefficient: Review	Correlation Coefficient: Shopping Bag
MoreIncomeBrands_5	MoreProductsBrands_5	1.0000	1
Month	Week	0.9864	0.9826
ProductCategory_Child	Gender_k	0.8496	0.8504
MoreIncomeBrands_1	MoreProductsBrands_1	0.8246	0.8172

Regarding the 6 categorical variables that were not translated into numbers, a $\chi 2$ test with a confidence level of 95% was performed to evaluate the correlation between those variables. Due to the high amount of data collected, to obtain a significant p-value, a subsampling of 500 entries was performed to run the test. Because the factors "DestCountry" and "Channel" have a high number of levels, no significant p-value was obtained. Therefore, those variables were kept. Similarly to Table 3, Table 4 shows the removed and the kept variables, and the p-value of the $\chi 2$ test performed for both datasets.

Table 4 - Kept and removed non-numeric categorical variables and their p-value, for both datasets

Removed Variable	Kept Variable	P-Value: Review	P-Value: Shopping Bag
VisitorType	DeviceType	0.000075	-
VisitorType	CustomerType	0.004640	-
CustomerType	DeviceType	-	0.005643

Exploratory analysis

Having finished removing correlated variables and altering some wrongly collected data to get it closer to Farfetch's reality, the resulting datasets were composed of 889,515 entries – Review dataset – and 2,522,736 entries – Shopping Bag dataset. After the initial study, an exploratory analysis of the resulting datasets was performed. The purpose of this analysis was to understand which variables would have the most impact in predicting if a consumer would convert on a certain page or not.

A data point was considered an outlier if it was ± 3 times the standard deviation of the average value of all that variable data points. The following subsections are dedicated to the study of specific variables and their relationship with the dependent variable.

ConvertReview / ConvertShopBag

The most relevant variable to be studied is the dependent variable. For each dataset, Figure 20 represents this variable distribution.

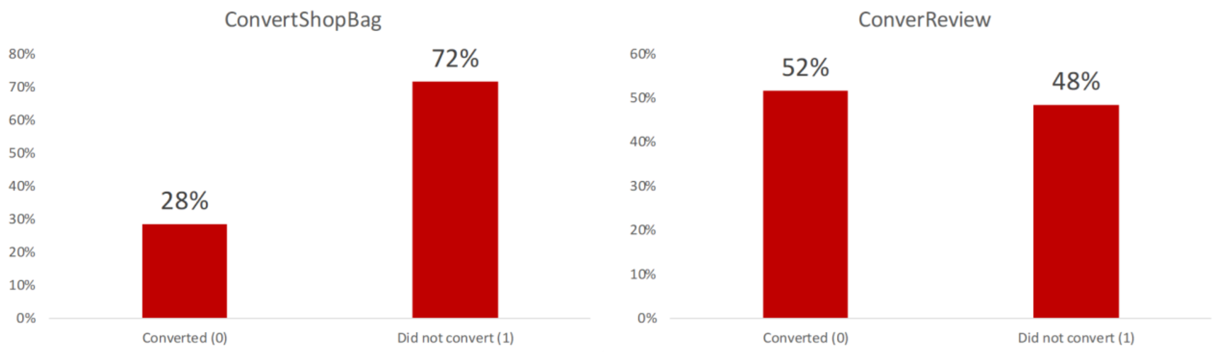

Figure 20 - Relative frequency plot for the dependent variable in both datasets

The datasets conversion distribution is fairly similar to the ones described in chapter 4. On the Review page, the percentage of conversion is 12 percentual points higher than the one described, and on the Shopping Bag page is 15 percentual points lower. The detected differences may be a result of the data collection limitations. To minimize the impact of these differences, both datasets were balanced.

TimeSessionMinutes and TotalClicks

As mentioned before, “TimeSessionMinutes” is the total length of the session in minutes, and “TotalClicks” is the total number of clicks a customer performs until the end of his / her session. Both these variables reflect information that can only be obtained at the end of a session and not when someone converts on a certain page. Even though “TimeSessionMinutes” was revised for the Review page the same could not be done for the “Shopping Bag page” nor for the “TotalClicks” variable. Therefore, both these variables may introduce a degree of error to the models. These variables distribution according to whether someone converted on a given page (0) or not (1) are plotted in Figure 21 and Figure 22.

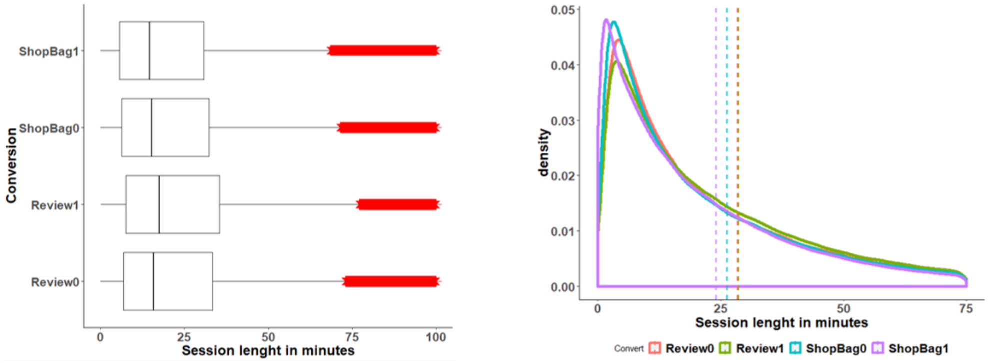

Figure 21 - Session length boxplots and density function, for both datasets

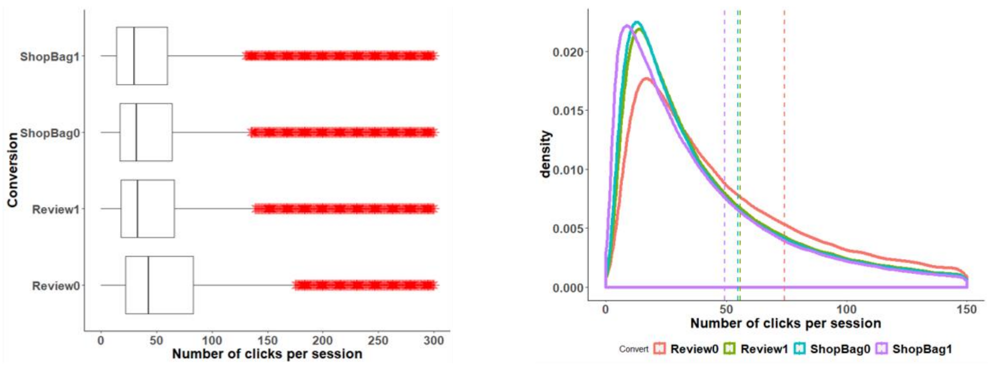

Figure 22 - Number of clicks per session boxplots and density function, for both datasets

From the density plots in Figure 21 and Figure 22 it is obvious that, on average, consumers who get to the Review page have longer sessions and perform more clicks than consumers who get to the Shopping Bag page – who may or may not get to the review page. An interesting finding is that people who convert on the review page spend less time on their session than those who do not. However, they perform more clicks. This may be explained by the fact that the “TotalClicks” variable could not be rectified and by the fact that, to convert you need to perform more clicks.

Both the total number of clicks and the session length have a great number of outliers. For both datasets, for both levels of the “Convert” variable, and for both the total number of clicks and the session length, more than 90% of the variable’s values interval is represented by outliers (7% for the session length and 12% for the number of clicks). These outliers are not represented on the boxplots, but the greatest value for the length of a session is 1400min, and the maximum number of clicks is 4923. Some consumers leave the Farfetch website or app open while not using it. This may explain the number of outliers found.

DaysSinceLastVisit

As the name suggests, this variable registers the days that passed since a consumer last visited the Farfetch website or app. This is an interesting variable because it will allow the understanding of whether Farfetch consumers behave in a similar way to the ones described in the Literature Review chapter. Consumers are said to use the shopping cart on a first session as a way to store items they would like to purchase. Only on a later session do consumers possibly buy the stored items from that first session. Figure 23 displays this variable distribution in accordance with whether a conversion happened or not for a given page.

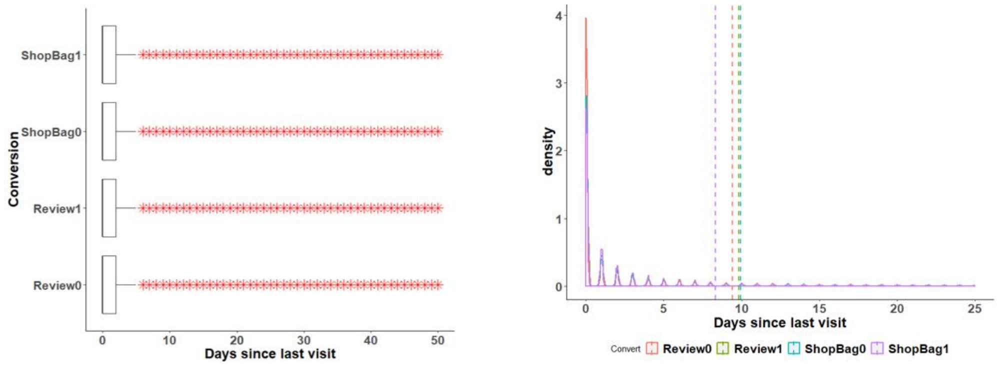

Figure 23 - Days since last visit boxplots and density function, for both datasets

Similarly to what happened with the previously analyzed variables, this variable outliers also assume greater values than the average.

Given that Farfetch was created over 10 years ago, the most extreme value this variable assumes is of a little more than 9 years. As shown in Figure 23 density plot, customers who did not convert on the Review page had not visited the Farfetch website/app in a longer time than the ones who did. This behavior was already expected, as discussed in the Literature Review chapter. The wavy format of the distribution is a result of this variable assuming discrete values instead of continuous ones.

AvgPrice, BasketValue, NProducts, ShippingFee

These variables are engineered features. Engineered features result of the combination of raw features and are created to attain more significant variables to train a model with. The “BasketValue” variable is just the sum of all the products value a consumer added to his/her shopping bag during a session. Likewise, “NProducts” is the count of the products the same consumer added to his bag during that session. The AvgPrice is the last two variables divided one by the other, as depicted in equation 5.1.

$$AvgPrice = \frac{BasketValue}{NProducts} \tag{5.1}$$

AvgPrice may be an interesting variable to analyze because a client may have numerous items on his/her shopping bag with a low average product price. In this case, there is a high likelihood of the total basket value being lower than the flat rate threshold. Which may induce customers not to convert. These three variables distributions are depicted in Figure 24, in Figure 25 and Figure 26.

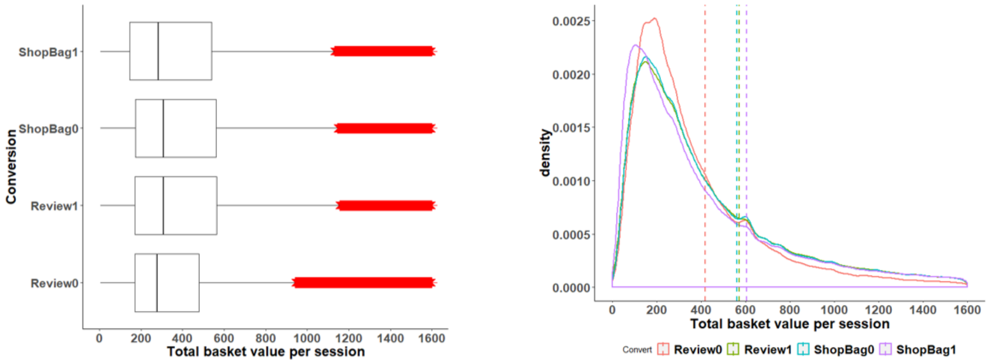

Figure 24 - Total basket value per session boxplots and density function, for both datasets

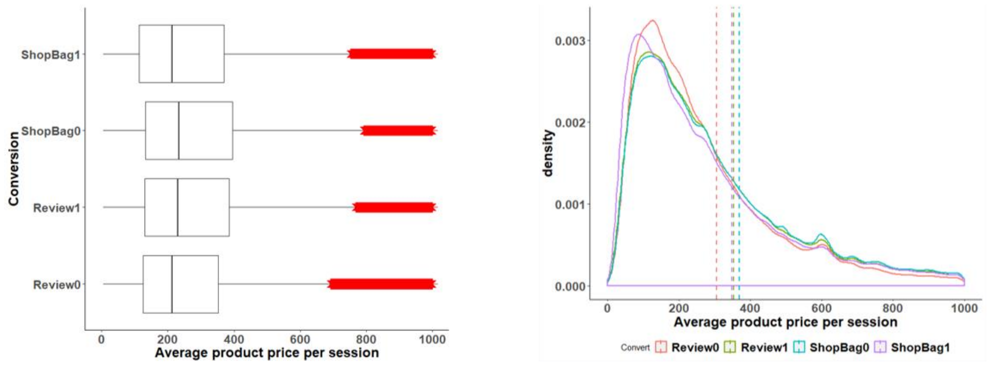

Figure 25 - Average products price added to the shopping bag per session boxplots and density function, for both datasets

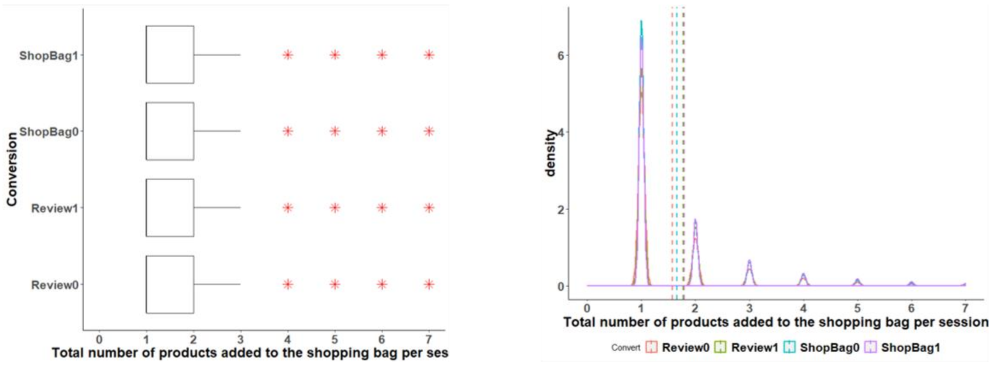

Figure 26 - Total number of products added to the shopping bag per session boxplots and density function, for both datasets

The average product price and the total basket value per session follow a similar distribution.

Nevertheless, the lowest basket value averages are related to converting in both the shopping bag page and the review page. The same does not happen for the averages of the variable AvgPrice. The lowest average for the average product price is related to converting in the

Review page (in line with the basket value distribution), but the highest average is related with converting on the Shopping Bag page. This difference may be explained by the fact that when there are commercial campaigns (for example decreasing prices by 20%) the decrease in prices is only observable after the shopping bag page. It may induce consumers to convert on that page but not converting on the review page. Homogeneously to what happened in the "DaysSinceLastVisit" distribution, the wavy format of "NProducts" is justified by the fact that the variable only assumes discrete values. All the average values for the number of products per session are lower than 2, which is an expected value given that in 2018 each purchase was on average composed of 1,65 products. Similarly to what happened in all the previously analyzed variables, outliers assume values +90% higher than the average values. This phenomenon may be explained by the fact that Farfetch has consumers who make purchases of extremely high values when compared to the norm.

All products prices are stored in pounds (£ - GBP) while the flat rate and free shipping thresholds were in the country currency. To overcome this issue, these values were converted to GBP using the session date to determine the currency conversion rate. For all the sessions analyzed, the shipping price that the consumer was seeing was calculated as explained in chapter 4.5.1. For this variable, its distribution is depicted in Figure 27.

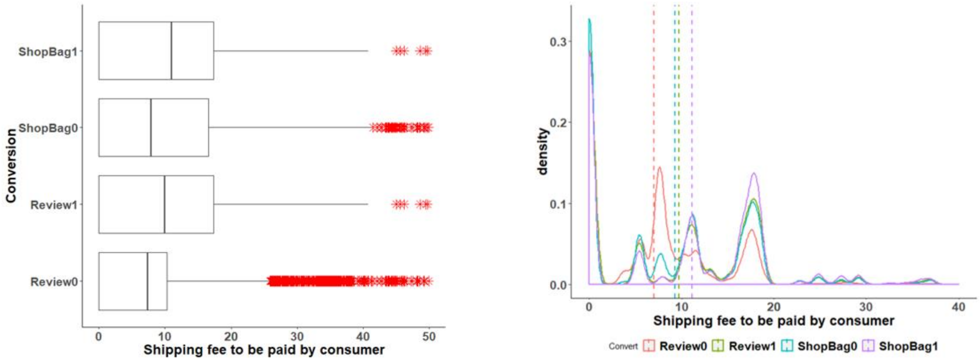

Figure 27 - Shipping fee to be paid by the consumer per session boxplots and density function, for both datasets

In the Shipping fees calculation chapter, it is said that most orders have an AOV over either the free shipping threshold or over the flat rate threshold. Therefore, the fact that most sessions have an associated shipping fee of 0 is easily explained. Also, two of the lower shipping fee averages are related with conversion on the Review page and on the Shopping Bag page. This may imply that shipping fees have a great impact on conversion. Regarding the outliers, they can be explained by those cases in which consumers add several items to their shopping bag that do not amount to a total value higher than the defined thresholds.

<u>InteractWithPromocodes and InteractWithPhotos</u>

There is no measure to define whether a consumer has a high purchase intention – the willingness of a customer to buy a certain product – or not. As a way to materialize this intention "InteractWithPromocodes" and "InteractWithPhotos" were analyzed. This analysis was based on the intuition that those who are interested in buying an item are more likely to interact with the photos of that item than those who are not. On account of promo codes (promotional codes to access discount campaigns) reducing the total price of an item, the interaction with them was also considered to have a positive impact on consumer's willingness to buy. Because one can only interact with promo codes after converting on the Shopping Bag page, this variable was not studied for that page and was removed from that dataset.

Figure 28 shows that those who interact with photos are more likely to convert on the Review page (purchasing an item) than those who do not. However, that is not the case for the Shopping Bag page. Similarly to what happened with the "DaysSinceLastVisit" variable, this difference between the Review page and the Shopping Bag page may be explained by what was reviewed in the literature. Consumers may add something to their shopping bag on a given session and making a purchase only on a later session.

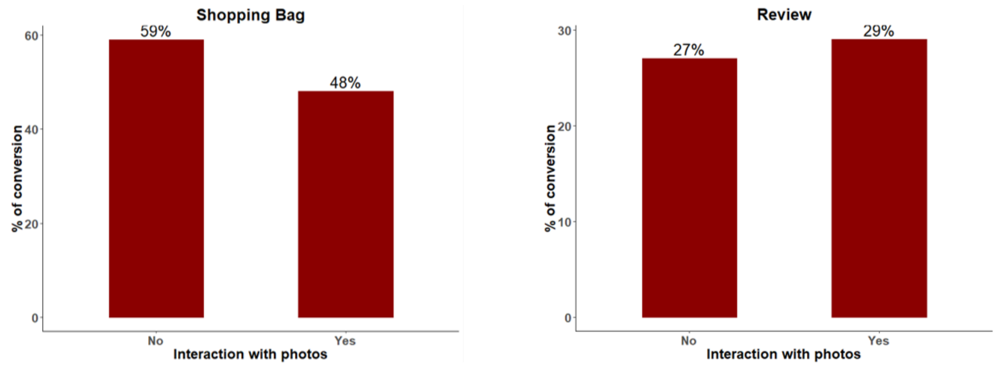

Figure 28 - Percentage of conversion every time a consumer interacted with photos before ending a session for both datasets

Because one may only interact with promo codes after the Shopping Bag page, it is unsurprising that the percentage of people who convert on the Review page and interact with promo codes is lower than the ones who do not interact with promo codes as plotted in Figure 29. When a promo code campaign goes live, a great number of clients goes to the review page only to see what the discount implies. These results contrast with the initial assumption that promo codes have a positive effect on conversion.

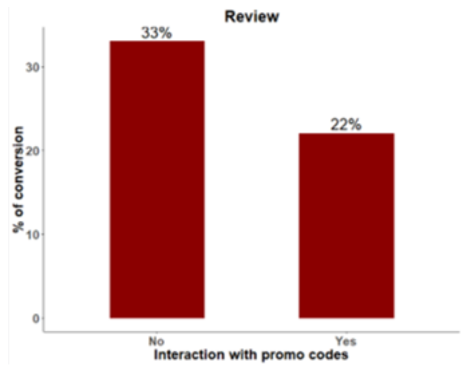

Figure 29 - Percentage of conversion every time a consumer interacted with promocodes before ending a session for the Review dataset

<u>MoreIncomeBrands, MoreProductsBrands, MostSelledProducts</u>

Farfetch sells more than 1000 brands and 37000 products online. The product and the brand chosen by the consumer were assumed to have an impact on his purchase decision. Because it was not viable to analyze each brand and product separately, the following binary variables were created. MoreIncomeBrands divides brands into five groups according to how much income they are generating. MoreProductsBrands divides brands into five groups according to how much products they are selling. As for MostSelledProducts it divides products into five categories according to how much items of that product have been sold. For all the variables, group 1 is composed of the brands and products that sell the least and group 5 by the ones that sell the most. All the groups for each variable have the same width meaning that they do not have the same number of elements. The relationship of these variables with conversion is depicted in Figure 30, Figure 31, and Figure 32.

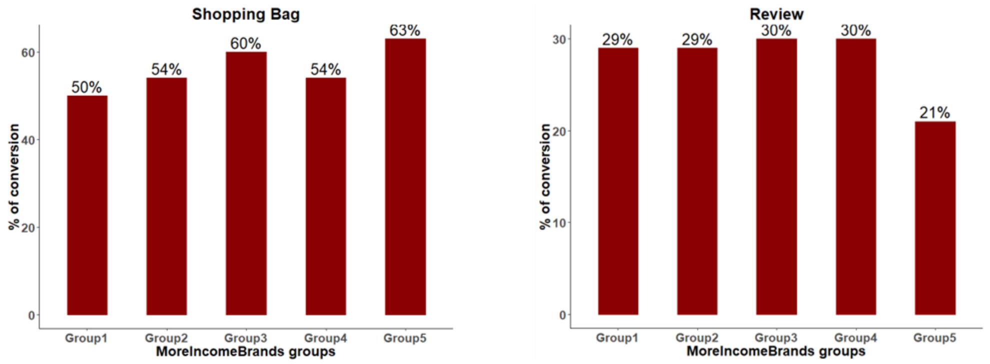

Figure 30 - Percentage of conversion every time a consumer added a brand belonging to a certain level of generated income to his shopping bag before ending a session, for both datasets

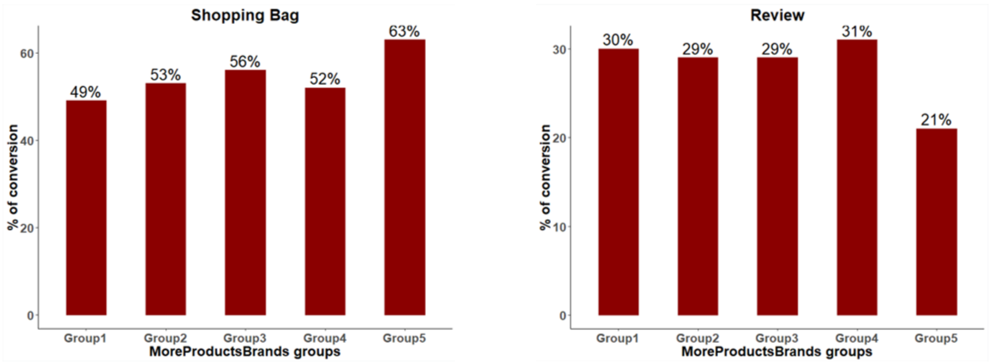

Figure 31 - Percentage of conversion every time a consumer added a brand belonging to a certain level of the number of products sold to his shopping bag before ending a session, for both datasets

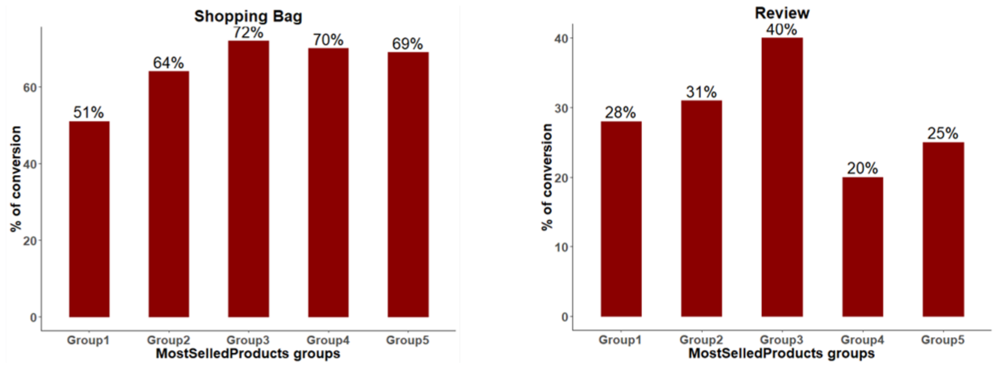

Figure 32 - Percentage of conversion every time a consumer added a product belonging to a certain level of the number of items sold to his shopping bag before ending a session, for both datasets

Regarding the brand-related variables, it is visible that group 1 and group 5 for both the Review and the Shopping Bag datasets assume a similar or equal value for those variables. Proving the identified correlation that exists between them. An interesting observation is that group 5 is the one that when present on the consumer's shopping bag results in more conversion on the Shopping Bag page. However, the same does not happen on the Review page. This may be because popular brands make consumers more interested in only deciding whether they are going to buy the product or not later in the session.

A similar pattern regarding the MostSelledProducts can be identified for group 5. A highlight of this variable is the behavior of group 3 for the Review page. When a consumer has a group 3 product in his/her basket and gets to the review page, the likelihood of him/her converting on that page increases when compared to other group's products. Due to this last variable behavior, it is expected that it will have a significant impact on conversion.

<u>DeviceType and Channel</u>

DeviceType holds information regarding from which device a consumer is accessing either the Farfetch website or app. There are some limitations with this variable, an example being that when someone is using their phone, it is not specified whether the operating system is Android, iOs or Windows. It is assumed that all consumers behave equally when on their phones, which may not be true.

Channel discriminates how consumers got to the website/app. There are seven different ways for a consumer to reach Farfetch's marketplace, representing seven of the nine levels of this variable:

- Direct – when a consumer types the site URL to get to the website;
- Organic – when a consumer searches for the website on a search engine and gets to the website;
- Affiliates – when a consumer reaches the website because a personality redirected them to the website. Every purchase by that consumer implies benefits for the redirector;
- Referral – when a consumer makes a purchase because another consumer redirected him there;
- Display – when a consumer is influenced to visit a website due to advertisements that appear on other websites. The company that is adverting needs to pay for that advertisement regardless of whether the consumer reached their website or not;
- Pay-Per-Click (PPC) – when a consumer gets to a company's website because of an add that is only being paid for when clicked;
- E-mail – when a consumer receives a promotional e-mail and decides to visit the website of the sender.

This variable has two other levels: "Not Defined" – when the path the consumer took to get to the website is unclear – and "Others" – when the channel used to get to the website is none of the above mentioned.

These two variables may be interesting to study because the rationale behind the reasons to go to a website, were thought of as being different for every channel and every device type used. The variables distribution, according to whether someone converted or not, is depicted in Figure 33 and Figure 34.

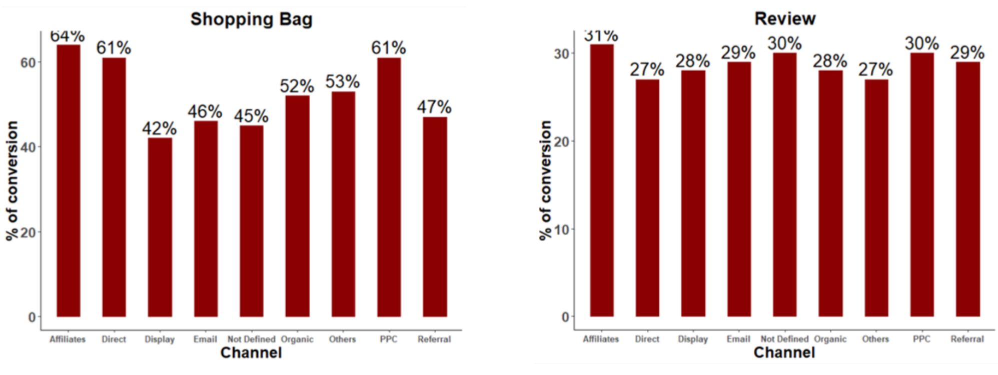

Figure 33 - Percentage of conversion depending on the channel used to get to Farfetch website/ app, for both datasets

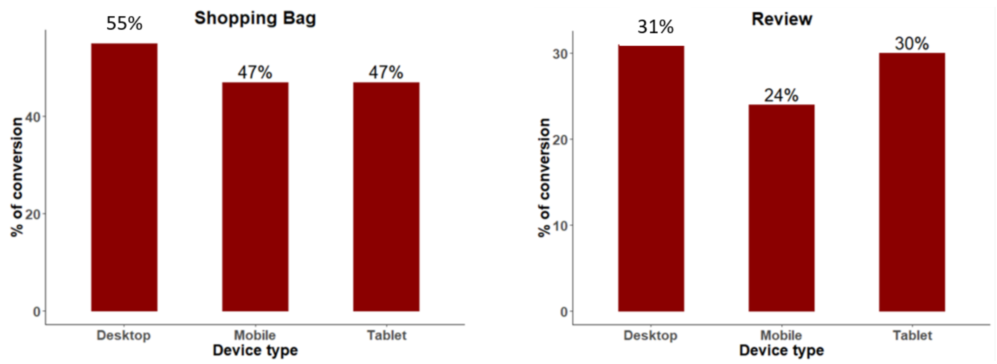

Figure 34 - Percentage of conversion depending on which device the Farfetch website/ app was being accessed from, for both datasets

By analyzing these figures, it is clear that the device type is more relevant when analyzing conversion on both the Review and the Shopping Bag pages than the Channel. This is because a more significant conversion related discrepancy is visible in the device type levels than in the channel levels. However, on the Shopping Bag page, the channel also assumes a significant relevance, especially on the levels: Affiliates, Direct and PPC. Direct and PPC may be justified by the intuition that when a consumer is interested in a product, he/she will directly search for the website and therefore convert on the shopping bag page (in Farfetch's case PPC ads appear when someone makes a Farfetch related search on a search engine). As for the Affiliates level, it may be justified by the strong influence some personalities have on consumers. In conclusion, these variables may play an important role when predicting conversion on both pages as it was initially assumed.

DestCountry

Different countries have distinct cultures, which may result in different conversion behaviors. To test this hypothesis, the variable "DestCountry" was created. It represents the destination country chosen by the consumer. It may not correctly represent the consumers' nationality, but it is the closest approximation possible with the available data. The relationship between conversion behavior and the destination country is depicted in Figure 35 for both pages under analyze.

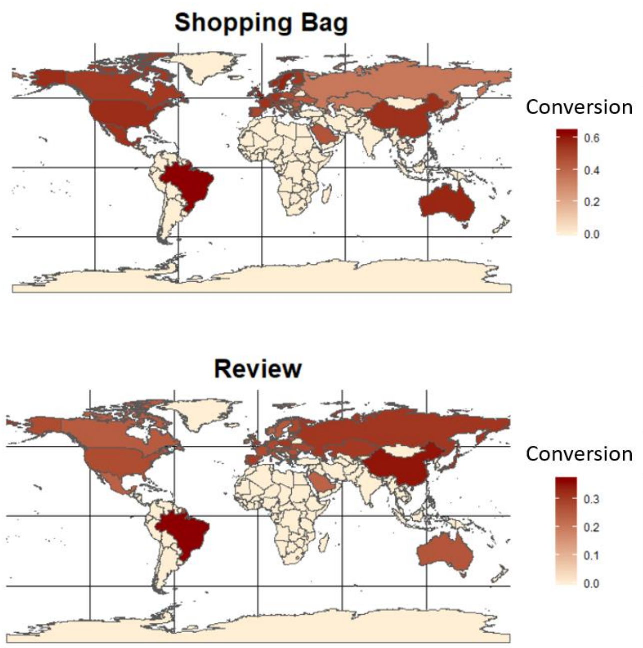

Figure 35 - Percentage of conversion for each destination country, for both datasets

For both pages represented in Figure 35, darker colors are related to higher conversion and vice-versa. It is interesting to notice that the United States convert relatively more on the shopping bag page than on the review page. The opposite happens for Russia. By observing these maps, it becomes clear that the destination country may have a high impact on predicting conversion.

5.2 Data preparation and algorithms selection

To prepare the data, one has to decide which ML techniques are going to be used. Being classification one of the most popular and significant areas of ML, it is unsurprising that ML researchers have proposed numerous classification algorithms. Due to this ever increasing selection of classification algorithms, the question of which algorithm better suits a given dataset and problem arises. As an answer to this question Wolpert and Macready (1997) developed the No Free Lunch (NFL) theorem: "If algorithm A outperforms algorithm B on some cost functions, then loosely speaking there must exist exactly as many other functions where B outperforms A". Bearing this in mind, the model selection must consider what evaluation metrics best suit a given problem.

To better comprehend the model developed in this thesis, Naive Bayes, Logistic Regression, and Extreme Gradient Boosting algorithms were analyzed in further detail.

5.2.1 Studied algorithms

Naive Bayes

Naive Bayes (NB) is a classification algorithm based on the Bayes' theorem. It is built under the assumption that all variables used to predict the outcome of a model (value of the dependent variable) are independent. This assumption is the root of the "naive" classification of this algorithm. This may be a disadvantage given that the independence assumption is rarely true in most real-world applications. However, one of the advantages of NB is its simplicity to build, which makes it useful for large datasets (Zhang 2004).

NB uses the Bayes' theorem to calculate the posteriori probability – "updated probability of an event occurring after taking into consideration new information" (Hayes 2019) – as explained in equation 5.2.

$$P(c|x) = \frac{P(x|c) * P(c)}{P(x)} \tag{5.2}$$

Where:

$P(c|x)$, posterior probability of class c (dependent variable) given predictor x (independent variable);

$P(x|c)$, probability of predictor x given class c

$P(c)$, the prior probability of class c

$P(x)$, the prior probability of attribute x

This equation is used to calculate the probability of a given set of attributes to belong to a certain class. The class with the higher posterior probability is the outcome value for the dependent variable (Zhang 2004).

Logistic regression

Regression is a method used to find statistical relationships between variables. There are 2 popular types of regression for classification models, the linear regression, and the logistic regression (LR). Both are used to represent the relationship between independent variables and dependent ones. The main difference between the two is that in linear regression, the relationship between the independent variables and the dependent ones is required to be linear (Lobo et al. 2010). Figure 36 represents the output for a binary classification problem using the two regressions in which the previous condition is not verified.

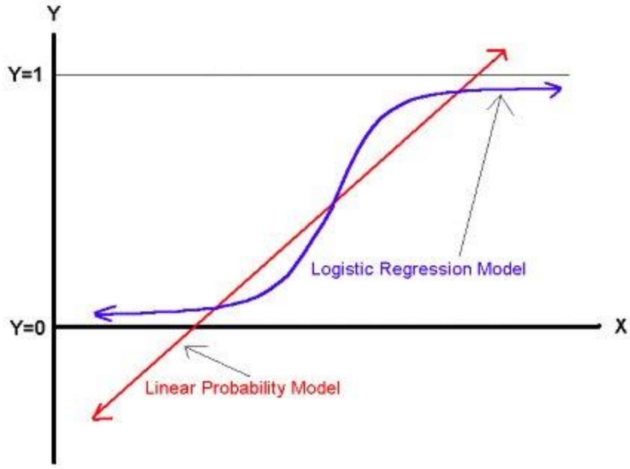

Figure 36 - Differences between a logistic regression model and a linear regression model, in (Lobo et al. 2010)

Binary LR is a model used for classification algorithms in which the outcome variable is a factor with only two levels (ex.: yes/no; 1/0). The output of a LR is the sum of each independent variable times its coefficient. This output may be described as how likely a given data point is to belong to a given class (Lobo et al. 2010).

Extreme gradient boosting

Gradient boosting is a ML technique mostly used for tree-based classification algorithms. Tree-based gradient boosting (TBGB) is very similar to the random forest algorithm. However, there are some differences. The most significant one being that in TBGB an ensemble of shallow and weak trees – results are only slightly better than random guessing – is built where each created tree learns and improves with the previous one. On the other hand, in the random forest algorithm, an ensemble of deep independent trees is built (Friedman 2002).

TBGB algorithm starts by building a first weak tree where all points are given the same importance – weight. Then a second tree is built where data points correctly classified by the first tree are given lower importance and vice-versa. This tree focusses on the higher weight data points. Some of these points are correctly classified. However, there are still misclassified points. This process continues for multiple iterations until a stopping condition is reached. The final model is a combination of all the created models - all the trees are given a score depending on their accuracy, and a weighted outcome is generated (Friedman 2002). This process is depicted in Figure 37, where the blue shade represents the values predicted as "+" and the pink shade the values predicted as "-".

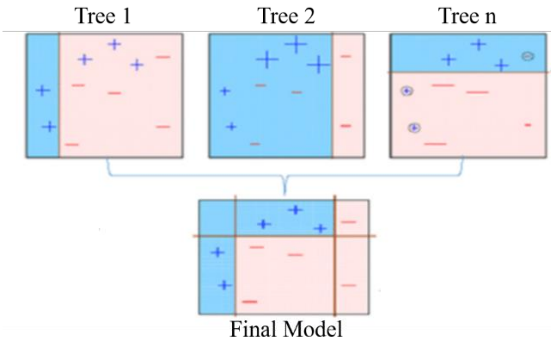

Figure 37 - Tree based gradient boosting simplification, in (Saraswat 2016)

TBGB is a "black-box" type of model – highly complex model – meaning that the interpretability of the results may be lowered. Also, it is very computationally expensive – both time and memory exhaustive. To tackle this issue, the Extreme Gradient Boosting (EGB) algorithm was created by Tianqi Chen in 2014 with its first stable release in 2017. It allows parallel computation, and it is generally 10 times faster than the normal gradient boosting. It also allows an incorporated hyperparameter – a priori set parameters – tuning. For a decision tree based EGB, the most commonly tuned parameters are (Chen et al. 2016):

- Nrounds – the total number of trees created. The higher the number of trees, the higher will the performance be. Nevertheless, the computational power demanded will also increase. A balance between performance and computational power must be found;
- Max_depth – the lower this value is, the higher is the depth of the tree. The higher the depth of the tree the higher the likelihood of overfitting – "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may, therefore, fail to fit additional data or predict future observations reliably" (Cambridge 2008);
- Col_sample_bytree – the number of features in a tree. Like the "Nrounds" hyperparameter, generally, the higher the number of features, the higher the performance and the higher the computational power needed.

5.2.2 Data preparation

Having understood the different classification algorithms to be used, only one transformation technique was needed.

The EGB algorithm only allows for numeric variables. Because both datasets had 6 non-numeric variables – "VisitorType", "CustomerType", "DeviceType", "DestCountry", "Channel" and "Gender" – the one-hot-encoding technique was applied. One-hot-encoding is a process by which categorical variables are converted into numeric binary ones. This process is described in Figure 38.

Session	Channel
1	Channel1
2	Channel2
3	Channel3
4	Channel4
5	Channel5

Session	Channel1	Channel2	Channel3	Channel4	Channel5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1

Figure 38 - One-hot-encoding explained for the "Channel" variable

5.3 Model selection and evaluation

After preparing the data, the next step of the methodology is building a model. Because it is not clear which model best suits the defined problem, three were constructed – Naïve Bayes, Logistic Regression, and Extreme Gradient Boosting. However, before constructing a model, it is necessary to split and balance the datasets as previously reviewed.

Because a very large number of entries composed the original datasets, the balancing method used was undersampling, performed as described in the Literature Review chapter. The resulting datasets were smaller with the review dataset being composed of 853,934 entries and the shopping bag dataset of 1,412,732 entries. An advantage of using smaller datasets is that they require less computational effort when building a model.

To overcome the lack of reliability described in the Literature Review chapter of the hold-out method, Max Kuhn et al. (2013) suggested using a cross-validation technique on the training set. This allows for the selection of the model that shows the best values for the evaluation metrics. After this selection, the actual performance of the chosen model may be evaluated using the testing set. Consequently, both balanced datasets were split into a training set composed of 70% of the entries and a test set composed by the rest. All the constructed models used a 10-fold cross-validation technique on the training set.

To select the best possible model, the evaluation metrics chosen to assess the models' performance need to be decided. To do so, first, it is necessary to understand the existing metrics in the literature for binary classification problems.

5.3.1 Evaluation metrics

As the NFL theorem argues, the best metrics for a given problem must be found in order to select the model that best suits it. For a binary classification model as the one proposed in this thesis, the possible evaluation metrics are the ones detailed in the following section.

Performance metrics

For classification problems, the concept of a confusion matrix (CM) is the base for several performance metrics. A CM shows the values predicted by a model and compares them to the actual values of the dependent variable. It assumes a size of n x n, where n is the number of the different levels of the dependent variable (Visa et al. 2011). An example of such a matrix with n=2 is displayed in Figure 39.

	Predicted Negative	Predicted Positive
Actual Negative	a	b
Actual Positive	c	d

Figure 39 - Example of a confusion matrix, in (Visa et al. 2011)

In this matrix, "**a**" represents the number of correct negative predictions also known as true negatives (TN), and b represents the number of incorrect positive predictions, otherwise known as false positives (FP). The sum of "**a+b**" is the total number of actual negative values. As for **c** and **d**, they follow a homologous logic with **c** being the number of false negatives (FN) and **d** being the number of true positives (TP) (Visa et al. 2011). Their sum amounts to the total number of actual positive values. In statistics, a FP is considered a type I error and a FN a type II error.

The most used performance metric based on the CM is accuracy, which can be defined by equation 5.3:

$$accuracy = \frac{a + d}{a + b + c + d} \tag{5.3}$$

Nonetheless, in many classification problems, accuracy alone is not the best metric to evaluate a classifier. Especially when the dataset used is imbalanced – if 80% of the data is positive, then an accuracy of 80% only means that the model predicted everything as positive. Hence several other performance metrics can be built using the CM in order to have a more accurate picture of the model's capabilities (Visa et al. 2011). These other metrics are:

- Sensitivity/ recall – measures the positive instances accuracy and it is defined by equation 5.4:

$$sensitivity = \frac{d}{d + c} \tag{5.4}$$

- Specificity – measures the negative instances accuracy and can be defined by equation 5.5:

$$specificity = \frac{a}{a + b} \tag{5.5}$$

- Precision – measures the model ability to not label an entry as positive when it is negative and is calculated as shown in equation 5.6:

$$precision = \frac{d}{b + d} \tag{5.6}$$

Given that the result of all the analyzed models is a probability of an entry to belong to a certain class or not, an operating point is defined as being the threshold from which that entry is either considered positive or negative. This is relevant because all the previously described metrics are only valid for a single operating point. To evaluate a model as a whole, regardless of the chosen threshold, the Receiver Operating Characteristic (ROC) curve has long been used.

The ROC curve shows the relationship between sensitivity and specificity Two ROC curves from different models created using the same dataset are depicted in Figure 40. The diagonal represents the performance of a random classifier.

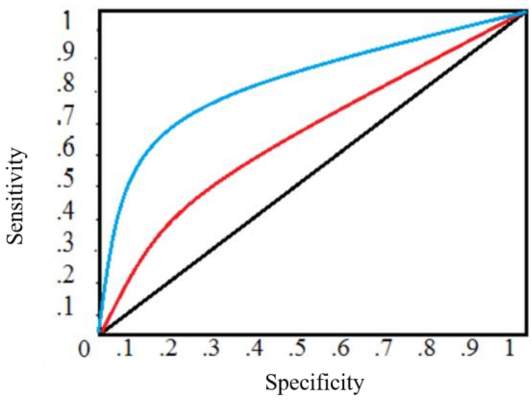

Figure 40 - ROC curves for two different models and random guessing line, adapted from (Han et al. 2012)

The closer a ROC curve is to the random classifier diagonal, the worse the model performance is. Point (0,1) represents a perfect classifier – all the positive predictions are correct – whereas point (1,0) represent the worst possible model. The ROC curve is helpful in determining what the most suitable operating point for a given problem is. However, it is not easy to use when comparing a wide variety of classification models. Therefore, the need for a new metric arose. The area under the ROC curve (AUC) became widely used to evaluate a model's performance. The higher the AUC, the better the model is at predicting positive values as positive and negative values as negative (Han et al. 2012).

Speed metrics

Regarding the speed of the model, there are two important metrics to consider. The time it takes to construct/train the model – training time – and the time necessary to use the model – prediction time. Both the speed metrics and the performance metrics should be considered when choosing the model that best suits a given problem.

5.3.2 Model testing

To select the best model, the metrics chosen were accuracy, AUC-ROC, and precision. Speed metrics were not consider given that none of the constructed models took more than a day to be built. To understand which model best suited the given problem, all the most common tuneable hyperparameters were optimized.

EGB is the only algorithm with tuneable hyperparameters. The hyperparameters tuned and their values were:

- Nrounds – assuming the values of 100 (default value) and 200 (higher value than the default number that should increase performance but low enough to not significantly alter the model's building speed);
- Max_depth – assuming the values of 3, 6 (default), 8, 10 and 15. The best value for this variable is usually between 3 and 10;
- Col_smaple_Bytree – assuming the values of 0,5 to 0,9 with an increment of 0,1. These values represent the percentage of the total features that are included in the model

The values of accuracy for each dataset, for each EGB model created, are depicted in Appendix C. For the Shopping Bag dataset, the best performance results were obtained when Nrounds was 200, Max_depth was 10 and Col_sample_Bytree was 0.5. The best values of the hyperparameters for the Review dataset were the same as the ones for the Shopping Bag dataset except for the Col_sample_Bytree which assumed a value of 0.9.

After choosing the best models' hyperparameters – only for EGB – it was necessary to test the models using unseen data. This necessity came from the fact that when the CV technique is used on the training data, the performance estimators assume optimistic values. To have a more realistic view of those estimators, the testing set initially separated from the training set is used. The best performance metrics results for all the three models for the testing set are represented in Table 5 for the Review dataset and in Table 6 for the Shopping Bag dataset.

Table 5 - Performance metrics results for the three models studied for the review dataset

Review	EGB	LR	NB
Accuaracy	79.73%	65.16%	62.83%
AUC-ROC	91.60%	73.36%	70.44%
Precision	92.55%	67.82%	72.97%

Table 6 - Performance metrics results for the three models studied for the shopping bag dataset

Shopping Bag	EGB	LR	NB
Accuaracy	68.93%	65.87%	60.76%
AUC-ROC	76.70%	71.95%	68.27%
Precision	71.28%	65.19%	68.57%

For both datasets, the model that showed the best results for all the considered metrics was the EGB. The ROC curves for that model for both datasets are depicted in Figure 41.

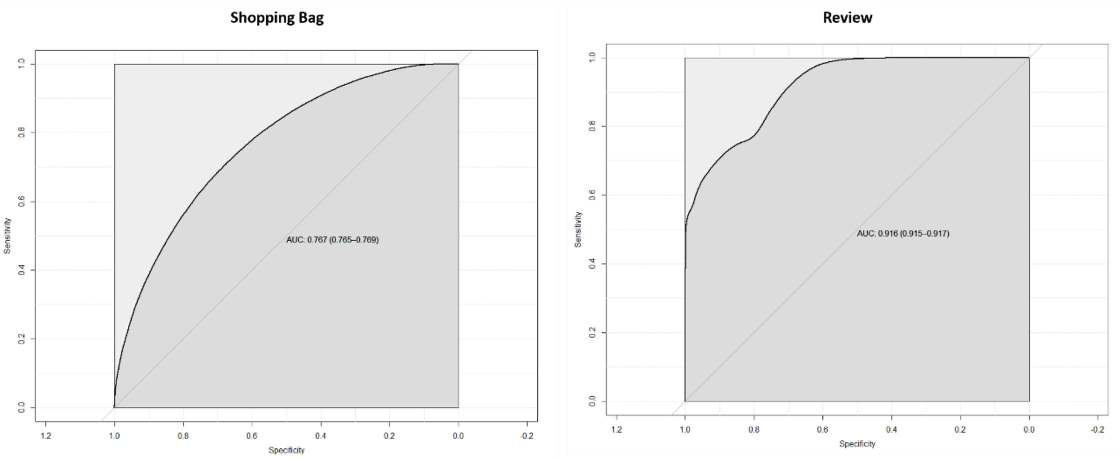

Figure 41 - ROC curves for the EGB model, for both datasets

To select which model was going to be used for the new shipping fee A/B testing tool, a variable importance study was performed. The study was conducted for the EGB models related to the Shopping Bag dataset and with the Review dataset. The results of this study for both pages are depicted in Figure 42.

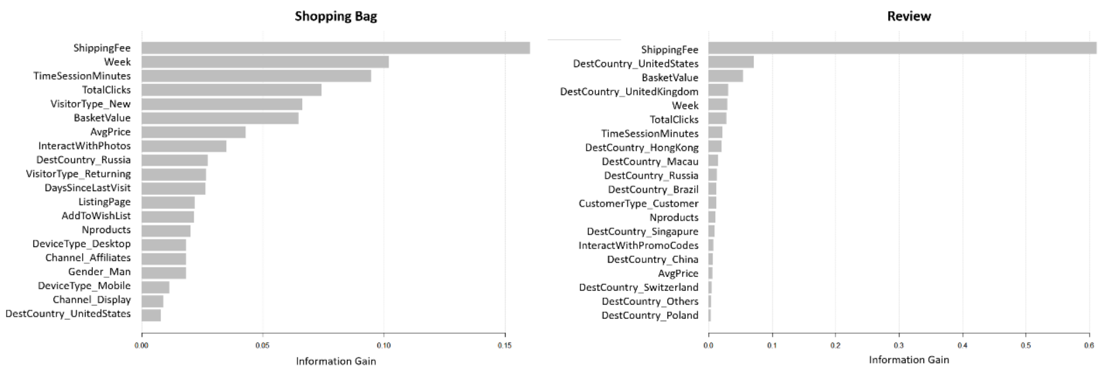

Figure 42 - Information gain with each variable for the EGB model, for both datasets

By observing Figure 42, it becomes clear that the shipping fees have a greater influence on the review page (+0.6 information gain) than on the Shopping Bag page (+0.15 information gain). Therefore, the Review model was the one considered for the creation of a new A/B testing tool.

5.4 Application

The final step of the methodology chosen is the deployment of the model. It was not possible to implement the model in real-time during the 4.5 month-long internship at Farfetch. Hence an explanation on how the model would be applied for the creation of a new A/B testing tool follows.

For the creation of the new tool to be possible, the constructed algorithm would need to be integrated with the already existent A/B testing platform. The algorithm would be constantly updated (for example, once a week) to consider only the previous moving year given that fashion is an ever-changing market.

Before using either the current or the to be implemented A/B testing platform, it is necessary to identify where conversion problems lie. Because a variables analysis was made, identifying where problems may reside is made simpler. Like the algorithm, these analyses would also be constantly updated to give important insights on what may be done to tackle conversion problems (for example, Australia has a relatively high conversion rate for the Shopping Bag page, but the same does not happen on the Review page). Once a possible action is defined on how to increase or decrease the shipping prices (following the previous example, altering the shipping fee for Australia), the A/B test is performed to evaluate whether the change is profitable or not for Farfetch.

The main difference between the old A/B testing platform and the proposed one is the people who are selected to participate in the A/B test. In the proposed platform, to execute an A/B test for a decrease in shipping prices, only sessions classified as sessions with no conversion by the EGB algorithm – for the current shipping fee – would enter the A/B test. If the objective of the A/B test were to assess an increase in shipping fees, then only consumers who would convert with the current shipping price would be considered. This pre-selection eliminates the noise caused by consumers who would most likely not change their behavior when presented with the new price.

Half of the selected sessions would still be shown the current price – control group – and the other half would be shown the new price – test group. The A/B test would only be considering those people in which the price change might alter their attitude towards conversion. Hence, the time to obtain significant results is expected to decrease. To test if the alteration in shipping fees will bring extra profit for Farfetch a new take on equation 4.1 is carried out.

To calculate the extra GTV equation 5.7 is considered.

$$GTVExtra = GTVTest * Precision \tag{5.7}$$

Where:

GTVTest is the GTV generated by the consumers on the test group

As for the extra costs equation 5.8 is considered.

$$CostsExtra = CostsTest * Precision \tag{5.8}$$

Where:

CostsTest is the GTV generated by the consumers on the test group

Having calculated these two variables, it is important to notice that it is not in Farfetch's best interest to charge different shipping prices to different clients in the same market. This is because Farfetch is a luxury marketplace. Therefore it does not want to be compared to non-luxury marketplaces like Amazon. Hence, for an A/B test to be considered successful, two outcomes need to be simultaneously true:

- The test group has to show better results than the control group;
- The extra profit generated by the test group needs to cover: the extra costs with it associated and the costs related with applying the new shipping fee to all Farfetch consumers of the under-analysis market.

Thus, a new variable needs to be created, absorption. This variable may be translated by equation 5.9. It represents the percentage of sessions considered in the test group when compared to all the test group equivalent sessions. This variable only needs to be considered for a decrease in shipping prices.

$$\text{Absorption} = \frac{ConsumersTest * Precision * X * Consumers}{\text{EquivalentTestConsumers} + ConsumersTest * \text{Precision}} \quad (5.9)$$

Where:

ConsumersTest is the number of consumers on the test group;

Consumers is the number of consumers whose behaviour was not predicted to change with the shipping price alteration;

EquivalentTestConsumers is the number of consumers that did not convert and were not on the test group (for a decrease in shipping fee):

X is the difference between the old shipping price and the new.

There are two possible outcomes for a shipping fee A/B test:

- Conversion for the control group is equal or better than conversion for the test group;
- Conversion is better for the test group than on the control group.

The possible outcomes of an A/B test when the test group outperforms the control group in terms of conversion are depicted in Figure 43 for the new A/B testing tool. This Figure also explains the new take on equation 4.1.

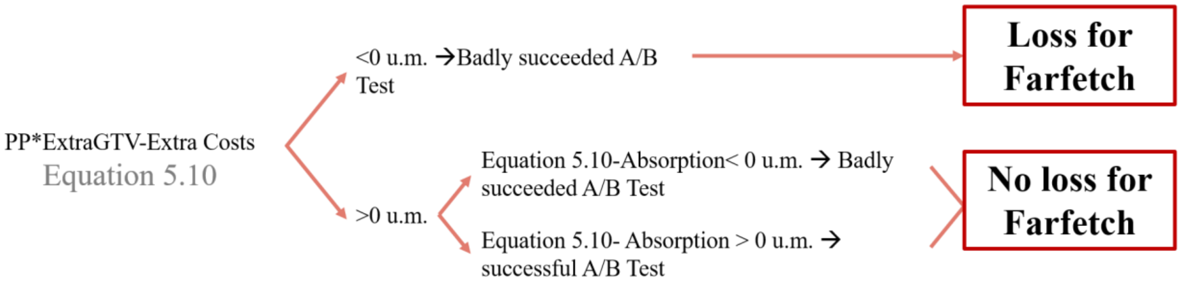

Figure 43 - Possible outcomes of the new shipping fee A/B test tool when the test group outperforms the control group

From the analysis of Figure 43, it becomes clear that the risks of performing an A/B test decrease – even when an A/B test is not successful it may not imply a loss for Farfetch (when decreasing the shipping prices). This is an extra scenario that the new A/B testing tool allows that is not possible with the current platform. Also, as previously stated, because the A/B test focus will be more specific, the time necessary to obtain significant results will decrease. Consequently, the costs undertaken during an A/B test implementation will also decrease.

5.5 Results limitations

The objective of the solution created was to build a classifier that could predict if a session would result in a conversion or not. This classifier should include shipping fees as a variable. As it was later expectedly confirmed, shipping fees are one of the most important features when a consumer is deciding to convert or not on both the Review page and on the Shopping Bag page. Also, it is essential to highlight that shipping fees play a more important role on the Review page than on the Shopping Bag page.

Although the model's accuracy was of almost 80% and its precision of over 90% there are some limitations to the model and some possible loss of information with the new A/B testing tool.

Regarding the model's limitation, it can only be applied in 73.77% of the cases. These cases exclude Intra-European Union transactions and domestic shipments in the United States. Also, there is some missing information that may be considered relevant (the type of operating system on the phone, the customer country instead of the destination country). Another important limitation is the fact that all shipping fees were calculated using the information on the products that were added to the shopping bag. However, if the consumer already had something in his bag from a previous session, those items were not accounted for. It was also not possible to classify an item as being on sale or not. Finally, it was assumed that when someone did not convert on the Review page, it was the last thing they did, which may not be true. There was no available data to correct these issues.

As for the A/B testing tool, even though it will reduce the time needed to obtain results and the total investment necessary to implement an A/B test, there will be some information loss. As described in chapter 4.5.1 there is the belief that some consumers buy unwanted items only to enjoy a reduced shipping fee. With the old A/B testing method, it was possible to understand how reducing the shipping fees would affect the return rate. This information will be lost while using the new platform. This is due to the fact that the new A/B tests will only consider those consumers in which a shipping price reduction would possibly alter the conversion behavior. Deciding whether to use or not the new platform must be a weighted decision on what is more relevant for the company:

- Understanding if a shipping fee change would be profitable for the company assuming an unaltered return rate, with faster and less costly results;
- Understanding if a shipping fee change would be profitable for the company assuming a variable return rate, with slower and more costly results.

Because fashion is an industry of pervasive change, with its consumers' behavior changing accordingly, the intuition is that most companies working in the fashion industry would prefer the first scenario.

6 Conclusions and future work

All e-commerce business, including luxury e-tailers, are continually striving for a better understanding of its consumers. Nowadays, numerous tracking applications allow for e-tailers to follow their consumers' actions on their websites. Although some ethical concerns regarding costumers privacy may be raised, the information obtained through those methods is considered of extreme relevance. What one does with such information may be vital for a company. Thus this thesis proposes the creation of a ML algorithm that uses that information to improve the performance of the current shipping price two hypothesis testing platform in the case study company, Farfetch. Along with this improvement, some interesting insights regarding the relationship between consumers' characteristics and conversion were obtained.

The objective for this thesis was demanding not only due to the dataset collection limitations but also, in part, because shipping is the only service every e-commerce consumer either buys or is offered. Thus, consumers have much information to decide on whether they find a specific shipping fee fair or not. This aspect is of extreme importance, given that many consumers claim that the greatest disadvantage of online shopping is the shipping fees. These are often seen as being biased and a way for retailers making an extra profit. However, shipping prices may not be altered without before assessing the profitability of that alteration. Therefore, a fast and intelligent way of testing modifications on the shipping prices is crucial.

Farfetch is a data-driven fast-growing company. Consequently, it collects data regarding their customer's behavior on their website. With this data, it is possible to understand the consumer's reactions to certain variables such as shipping fees. Due to its shipping price calculation method, Farfetch may charge different shipping prices to the same client in a different set of conditions. This makes the data available more interesting for constructing a machine learning model to predict conversion taking shipping fees into account. Understanding conversion is the first step to create a faster A/B testing tool.

To better comprehend conversion and due to the large amount of data made available by Farfetch, a machine learning algorithm was thought of and built. Because the shipping prices are only presented to the consumer on the Shopping Bag page and on the Review page (checkout) of Farfetch website/app, two models were created - one related with the Shopping Bag page and the other with the Review page. After pondering which model best suited the described problem the Extreme Gradient Boosting algorithm applied to the Review page was proved to yield the best results. However, the model created had some limitations, such as not being applicable to Intra-European Union shipments and domestic shipments in the United States or not considering whether an item was on sale or not. Despite its limitations, it was still used as a foundation for the new shipping A/B testing tool.

The current A/B testing tool considers all consumers when creating its control and testing groups. This implies that when an alteration to the shipping fees is being tested those consumers who would convert (decrease in shipping price) / not convert (increase in shipping price) regardless of the alteration made to the shipping fee are also being tested. These consumers add noise to the results, resulting in more time and resource consuming A/B tests.

The proposed A/B testing platform integrates the created algorithm to predict conversion in order to select the A/B test participants. Only those consumers who would possibly change their behavior towards conversion due to the alteration in shipping fees are considered. This pre-selection eliminates noise. Therefore, the time and resources necessary to achieve significant results are reduced, improving the A/B test application.

Although the project is concluded from a curricular point of view, it should not be considered finished. All the data collected to construct the model was retrieved from a soon to be discontinued data gathering platform – “Clickstream”. This platform is being replaced by the “Omnitracking” platform. “Omnitracking” will allow for a more complete data collecting process regarding shipping fees, number of products and consumer characteristics such as country of origin. The collection of session related variables will also be improved, especially, the length of the session until the review page is reached and the total number of clicks. “Omnitracking” will also monitor the consumer’s mouse actions. This will allow the identification of which shipping method the consumer would have chosen in the 26.23% of the cases where the consumer is shown more than one shipping fee. The improved data collection process will allow for a more in-depth, accurate, and precise model.

An interesting outcome from the A/B tests is the augmented information regarding consumers’ behavior toward conversion for different shipping fees. An interesting possible future work would be using this information to create a dynamic shipping price calculation tool. Much like Amazon, this tool would assign a different shipping price to each client in order to increase sales while increasing the shipping income. Once Farfetch had such a tool at its disposal, it would be a strategic decision whether to use it or not. Farfetch would have to ponder keeping its reputation as a luxury marketplace with a transparent shipping price calculation method or being compared to non-luxury marketplaces and therefore losing luxury clients but increasing its sales while increasing its shipping income.

Bibliography

Amed, Imran, Anita Balchandani, Marco Beltrami, Achim Berg, Saskia Hedrich, and Felix Rölkins. 2018. "The State of Fashion 2019." *McKinsey&Company*. https://doi.org/10.1163/156853010X510807.

Andersson, Thomas, and Martin Shaw. 2019. "RETAILX SECTOR ANALYST REPORT LUXURY 2019."

Badr, Will. 2019. "Having an Imbalanced Dataset? Here Is How You Can Fix It." 2019. https://towardsdatascience.com/having-an-imbalanced-dataset-here-is-how-you-can-solve-it-1640568947eb.

Bamfield, Joshua A. N. 2013. "Retail Futures 2018: Shop Numbers, Online and The High Street A Guide to Retailing in 2018."

Brinckmann, Madeleine. 2018. "Luxury Report 2018 – Luxury Fashion," no. August: 1–13. https://www.statista.com/study/56842/luxury-report---luxury-fashion/.

Cambridge. 2008. *Cambridge Advanced Dictionary*. Edited by Colin McIntosh. 4th Revise. Cambridge, United Kingdom: CAMBRIDGE UNIVERSITY PRESS. https://dictionary.cambridge.org/dictionary/english/e-tailing?fbclid=IwAR1SuZ2bUpdzMzMm4Md7DZuNdg_BnwscNK0BuIWN-Hd8ag-oX-pKIJR2HSQ.

Cao, Lanlan, and Li Li. 2015. "The Impact of Cross-Channel Integration on Retailers' Sales Growth." *Journal of Retailing*. https://doi.org/10.1016/j.jretai.2014.12.005.

Chen, Tianqi, Tong He, and Michael Benesty. 2016. "Extreme Gradient Boosting." *ArXiv*, 1–4. https://github.com/dmlc/xgboost.

Choudary, Sangeet. 2016. "A Platform Thinking Approach to Building a Business | Platform Strategy – by Sangeet Paul Choudary." 2016. https://platformed.info/a-platform-thinking-approach-business-engine-scale/.

Cornell, A. 2002. "Cult of Luxury: The New Opiate of the Masses." *Australian Financial Review*, 2002.

D'Arpizio, Claudia, Federica Levato, Marc-André Kamel, and Joelle de Mongolfier. 2017. "Luxury Goods Worldwide Market Study, Fall-Winter 2017: The New Luxury Consumer: Why Responding to the Millennial Mindset Will Be Key." *Bain & Company*.

Dauriz, Linda, Natalie Remy, and Nicola Sandri. 2013. "Luxury Shopping in the Digital Age." *McKinsey*.

Farfetch. 2019. "What Is Farfetch? - About Farfetch." 2019. https://aboutfarfetch.com/about/farfetch/.

Friedman, Jerome H. 2002. "Stochastic Gradient Boosting." *Computational Statistics and Data Analysis*. https://doi.org/10.1016/S0167-9473(01)00065-2.

Frischmann, Tanja, Oliver Hinz, and Bernd Skiera. 2012. "Retailers' Use of Shipping Cost Strategies: Free Shipping or Partitioned Prices?" *International Journal of Electronic Commerce*. https://doi.org/10.2753/jec1086-4415160303.

Gandomi, Amir, and Murtaza Haider. 2015. "Beyond the Hype: Big Data Concepts, Methods, and Analytics." *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

Gartner, I T. 2013. "Gartner IT Glossary." Technology Research. 2013.

Gendy, M S El, and Can J Phys. 2006. "Phi-Coefficient" 79 (7): 2001–6.

Gounaris, Spiros, and Sergios Dimitriadis. 2003. "Assessing Service Quality on the Web: Evidence from Business-to-Consumer Portals." *Journal of Services Marketing* 17 (5): 529–48. https://doi.org/10.1108/08876040310486302.

Han, Jiawei, Micheline Kamber, and Jian Pei. 2012. "Introduction." In *Data Mining*, 1–38. https://doi.org/10.1016/B978-0-12-381479-1.00001-0.

Hawkins, Douglas M., Subhash C. Basak, and Denise Mills. 2003. "Assessing Model Fit by Cross-Validation." In *Journal of Chemical Information and Computer Sciences*. https://doi.org/10.1021/ci025626i.

Hayes, adam. 2019. "Posterior Probability Definition." Investopedia. 2019. https://www.investopedia.com/terms/p/posterior-probability.asp.

Kapferer, Jean-Noël. 1997. "Managing Luxury Brands." *Journal of Brand Management*. https://doi.org/10.1057/bm.1997.4.

Kiron, David, Pamela Kirk Prentice, and Renee Boucher Ferguson. 2014. "The Analytics Mandate." *MIT Sloan Management Review*.

Kuhn, Max, and Kjell Johnson. 2013. *Applied Predictive Modeling [Hardcover]*. *Springer*. https://doi.org/10.1007/978-1-4614-6849-3.

Kukar-Kinney, Monika, and Angeline G. Close. 2010. "The Determinants of Consumers' Online Shopping Cart Abandonment." *Journal of the Academy of Marketing Science*. https://doi.org/10.1007/s11747-009-0141-5.

Lewis, Michael. 2006. "The Effect of Shipping Fees on Customer Acquisition, Customer Retention, and Purchase Quantities." *Journal of Retailing*. https://doi.org/10.1016/j.jretai.2005.11.005.

Lewis, Michael, Vishal Singh, and Scott Fay. 2006. "An Empirical Study of the Impact of Nonlinear Shipping and Handling Fees on Purchase Incidence and Expenditure Decisions." *Marketing Science*. https://doi.org/10.1287/mksc.1050.0150.

Linnane, Ciara. 2018. "Farfetch IPO: 6 Things You Should Know about This Luxury-Goods Marketplace - MarketWatch." 2018. https://www.marketwatch.com/story/farfetch-has-filed-for-an-ipo-six-things-to-know-about-the-luxury-goods-marketplace-2018-08-21?fbclid=IwAR2A1CP5m-odmaRXQ6b7yra9llvI6yvVqTR1dKRL78QWOhlo5y4P0LnKTg0.

Liu, Xia, Alvin C. Burns, and Yingjian Hou. 2013. "Comparing Online and In-Store Shopping Behavior towards Luxury Goods." *International Journal of Retail and Distribution Management*. https://doi.org/10.1108/IJRDM-01-2013-0018.

Lobo, Desmond, Paul Watters, and Xin Wen Wu. 2010. "A New Procedure to Help System/Network Administrators Identify Multiple Rootkit Infections." *2nd International Conference on Communication Software and Networks, ICCSN 2010*, 124–28. https://doi.org/10.1109/ICCSN.2010.14.

Marques, Nelson. 2019. “Especial Moda - Entrevista José Neves.” *Expresso*, March 16, 2019.

McAfee, Andrew, and Erik Brynjolfsson. 2012. “Big Data: The Management Revolution.” *Harvard Business Review*.

Mehmet, Gümüş, Li Shanling, Oh Wonseok, and Ray Saibal. 2013. “Shipping Fees or Shipping Free? A Tale of Two Price Partitioning Strategies in Online Retailing.” *Production and Operations Management*. https://doi.org/10.1111/j.1937-5956.2012.01391.x.

Michie, Donald. 1968. “‘Memo’ Functions and Machine Learning.”

Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. 2013. *Foundations of Machine Learning*. *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004.

Morwitz, Vicki G., Eric A. Greenleaf, and Eric J. Johnson. 1998. “Divide and Prosper: Consumers’ Reactions to Partitioned Prices.” *Journal of Marketing Research*. https://doi.org/10.2307/3152164.

Murphy, Chris B. 2019. “Understanding Bottom-Line Growth vs. Top-Line Growth.” Investopedia. 2019. https://www.investopedia.com/ask/answers/difference-between-bottom-line-and-top-line-growth/.

Nisar, Tahir M., and Guru Prabhakar. 2017. “What Factors Determine E-Satisfaction and Consumer Spending in e-Commerce Retailing?” *Journal of Retailing and Consumer Services*. https://doi.org/10.1016/j.jretconser.2017.07.010.

Okonkwo, Uché. 2009. “Sustaining the Luxury Brand on the Internet.” *Journal of Brand Management*. https://doi.org/10.1057/bm.2009.2.

Saraswat, Manish. 2016. “Beginners Tutorial on XGBoost and Parameter Tuning in R.” *Hackerearth*. https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/beginners-tutorial-on-xgboost-parameter-tuning-r/tutorial/.

Schindler, Robert M., Maureen Morrin, and Nada Nasr Bechwati. 2005. “Shipping Charges and Shipping-Charge Skepticism: Implications for Direct Marketers’ Pricing Formats.” *Journal of Interactive Marketing*. https://doi.org/10.1002/dir.20030.

Schober, Patrick, and Lothar A. Schwarte. 2018. “Correlation Coefficients: Appropriate Use and Interpretation.” *Anesthesia and Analgesia* 126 (5): 1763–68. https://doi.org/10.1213/ANE.0000000000002864.

Scott, Tim, and Nathan Spitse. n.d. “Third-Party Risk Is Becoming a First Priority Challenge | Deloitte Canada.” Deloitte. Accessed April 18, 2019. https://www2.deloitte.com/ca/en/pages/risk/articles/reduce-your-third-party-risk.html#.

Shampanier, Kristina, Nina Mazar, and Dan Ariely. 2007. “Zero as a Special Price: The True Value of Free Products.” *Marketing Science*. https://doi.org/10.1287/mksc.1060.0254.

TechAmerica Foundation. 2012. “Demystifying Big Data: A Practical Guide to Transforming the Business of Government.” *TechAmerica Foundation’s Federal Big Data Commission*. https://doi.org/10.1109/MCSE.2011.99.

Visa, Sofia, Brian Ramsay, Anca Ralescu, and Esther Van Der Knaap. 2011. “Confusion Matrix-Based Feature Selection.” In *CEUR Workshop Proceedings*.

Walter, Matthias. 2017. “How Platform Thinking Transforms Business, Society and Work.” 2017. https://medium.com/platform-innovation-kit/how-platform-thinking-transforms-business-society-and-work-ecbf988703c8.

White, Sarah, and Melissa Fares. 2018. “Farfetch Tops Price Range in IPO in Boon to Luxury

Market.” Reuters. 2018. https://www.reuters.com/article/us-farfetch-ipo-luxury/farfetch-tops-price-range-in-ipo-in-boon-to-luxury-market-idUSKCN1M11BQ?fbclid=IwAR0_M_bVyP9QHmV2J4BxxGfz2j0r0SE9Ry511SGxVE44zcoiuZIut6AhPZc.

Wiedmann, Klaus Peter, Nadine Hennigs, and Astrid Siebels. 2009. “Value-Based Segmentation of Luxury Consumption Behavior.” *Psychology and Marketing*. https://doi.org/10.1002/mar.20292.

Wirth, Rüdiger, and Jochen Hipp. 2000. “CRISP-DM : Towards a Standard Process Model for Data Mining.” In *Proceedings of the 4th International Conference on the Practical Application of Knowledge Discovery and Data Mining*. https://doi.org/10.1.1.198.5133.

Wolpert, David H., and William G. Macready. 1997. “No Free Lunch Theorems for Optimization.” *IEEE Transactions on Evolutionary Computation*. https://doi.org/10.1109/4235.585893.

Zhang, Harry. 2004. “The Optimality of Naive Bayes.” *AAAI*.

Appendix A: Upper correlation matrix for the Shopping Bag

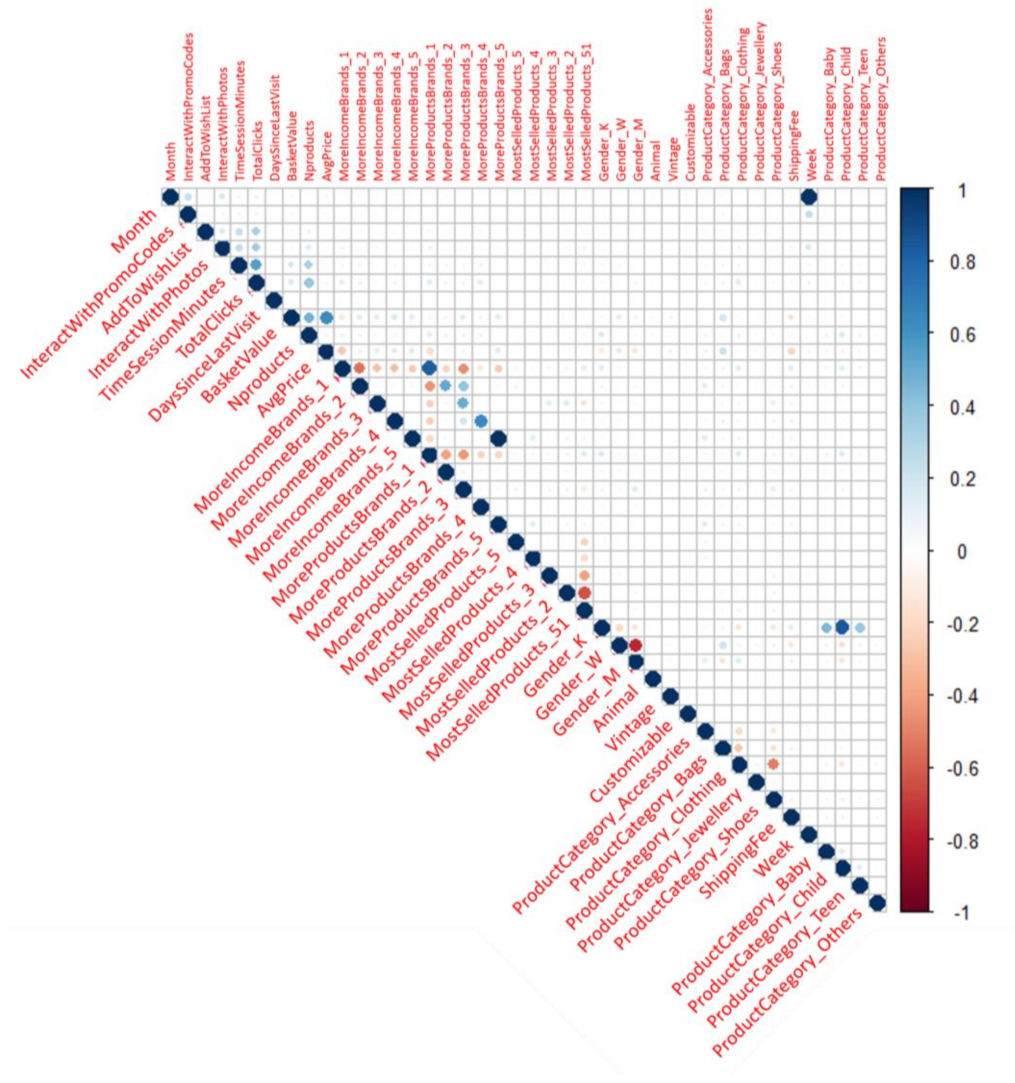

Figure A.1 - Upper correlation matrix for the Shopping Bag

Appendix B: Upper correlation matrix for the Review dataset

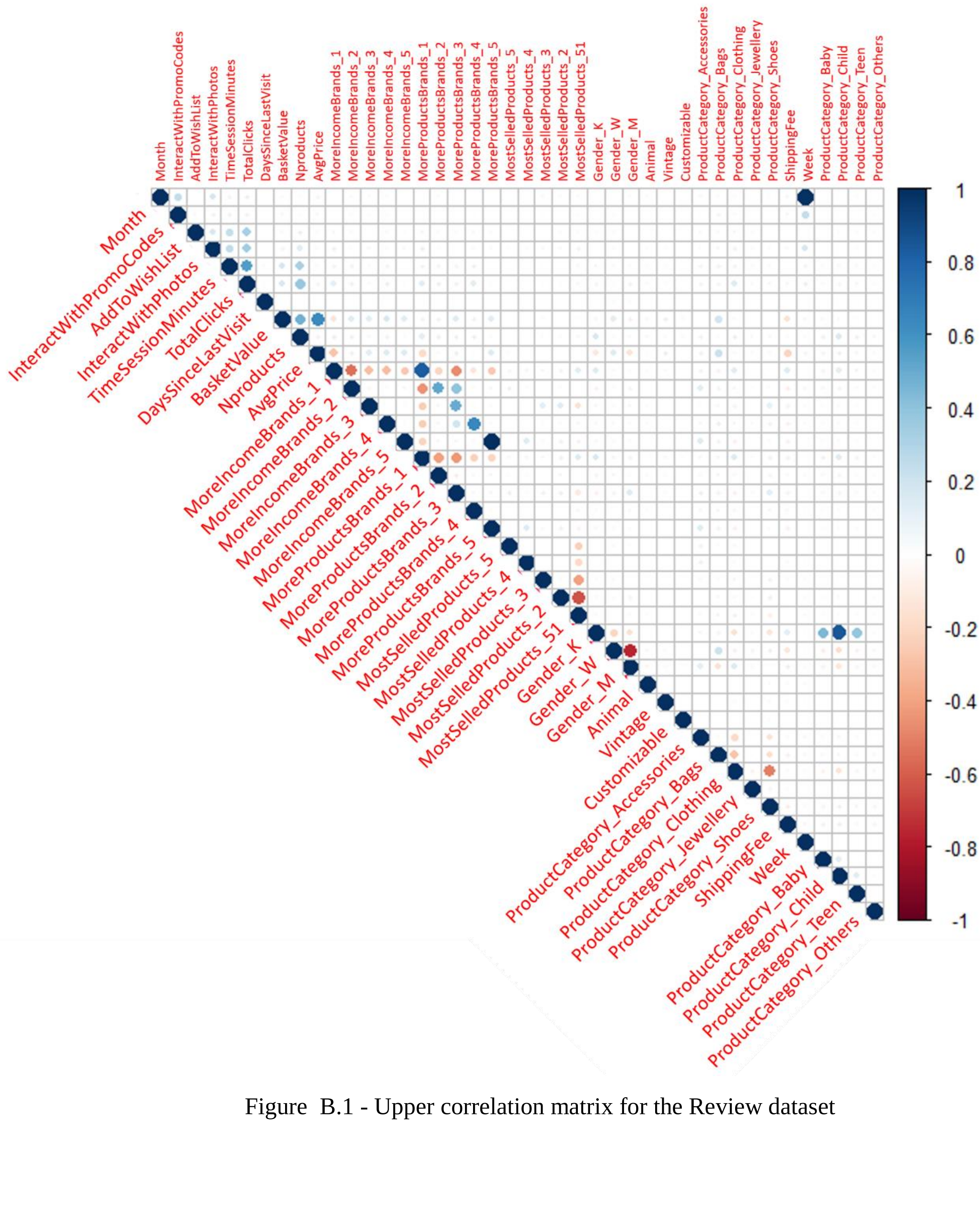

Figure B.1 - Upper correlation matrix for the Review dataset

Appendix C: Accuracy for the different values of the Extreme Gradient Boosting hyperparameters

Table C.1 - Accuracy for the different values of the Extreme Gradient Boosting hyperparameters

Max_depth	Colsample_Bytree	Nrounds	Accuracy_ShopBag	Accuracy_Review
3	0,5	100	68%	86%
3	0,6	100	68%	87%
3	0,7	100	68%	87%
3	0,8	100	68%	87%
3	0,9	100	68%	87%
6	0,5	100	70%	91%
6	0,6	100	70%	91%
6	0,7	100	70%	91%
6	0,8	100	70%	91%
6	0,9	100	70%	91%
8	0,5	100	71%	91%
8	0,6	100	71%	91%
8	0,7	100	71%	91%
8	0,8	100	71%	91%
8	0,9	100	71%	91%
10	0,5	100	71%	91%
10	0,6	100	71%	91%
10	0,7	100	71%	91%
10	0,8	100	71%	91%
10	0,9	100	71%	91%
15	0,5	100	71%	91%
15	0,6	100	71%	91%
15	0,7	100	71%	91%
15	0,8	100	71%	91%
15	0,9	100	71%	91%
3	0,5	200	69%	89%

3	0,6	200	69%	90%
3	0,7	200	69%	90%
3	0,8	200	69%	90%
3	0,9	200	69%	90%
6	0,5	200	71%	91%
6	0,6	200	71%	91%
6	0,7	200	71%	91%
6	0,8	200	71%	91%
6	0,9	200	71%	91%
8	0,5	200	72%	91%
8	0,6	200	72%	91%
8	0,7	200	72%	91%
8	0,8	200	72%	91%
8	0,9	200	72%	91%
10	0,5	200	72%	91%
10	0,6	200	72%	91%
10	0,7	200	72%	91%
10	0,8	200	72%	91%
10	0,9	200	72%	91%
15	0,5	200	72%	91%
15	0,6	200	72%	91%
15	0,7	200	71%	91%
15	0,8	200	71%	91%
15	0,9	200	71%	91%