ERASMUS UNIVERSITEIT ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Master Thesis Data Science and Marketing Analytics

Internet marketing, websites functionalities and customer loyalty.

Student name: Jimena Leon Testa

Student ID number: 541569

Supervisor: Dr. S.L. Malek

Second assessor: Dr. Ph.H.B.F. Franses

Abstract

In a globalized economy, marketing strategy is shifting towards building longtime customer relationships, instead of centering mainly on increasing sales, making concepts like Relationship Marketing and Customer Loyalty fundamental. Moreover, globalization and access to information and technology are resulting in an important expansion of internet sales. Thus, researches are trying to understand the particularities of the former concepts in an online environment. In this sense, several studies address issues related to relationship quality and customer loyalty in an online environment, from a customer satisfaction and trust point of view, by surveying customers or interviewing people inside the industry. In this thesis we try to fill the literature gap using a different approach. Specifically, we try to find the main drivers of internet customer loyalty by analyzing internet marketing strategies and website functionalities that successful internet retailers are applying. We do this by modeling the *Returning Shoppers' rate* of each website, first with a traditional statistical model, Lasso Linear Regression, and, second, comparing its prediction performance with Boosting Regression Trees, a more complex machine learning model. We find that Relationship Marketing also plays a big role in the online setting and is fundamental to increase customer loyalty. Additionally, we identify that online consumers mostly value 4 types of website functionalities which we classify as: "Convenience features", "Easy to use features", "Personalized features" and "Promotions features". Finally, although we do not achieve a high prediction performance, our study sets the precedents to keep developing and applying machine learning models in Relationship Marketing and Customer Loyalty's literature, as there is still plenty of potential to keep improving these models, as more data becomes available every day.

Keywords: customer loyalty, relationship marketing, internet marketing, website functionalities, lasso regression, boosting, machine learning.

Table of contents

1. Introduction

In a globalized world economy, where new technologies are developing incredibly quickly, consumers and companies can access information about products and competitors that years ago was unimaginable to reach. As a consequence of this, traditional marketing strategies like differentiation, market niches and cost leadership are becoming less effective (Bauer et. al, 2002), as competitors can rapidly respond and outperform other company's strategies. To adapt to this new context, marketing strategy is shifting more and more towards building longtime customer relationships, instead of centering mainly on increasing sales (Bauer et. al, 2002; Reinartz et. al, 2004; Rafiq et. al, 2013). As Reinartz et. al (2004, p.293) state "organizations are, in essence, moving away from product- or brand-centric marketing towards a customer-centric approach." In this sense, several studies suggest that gaining new customers is, on average, five times more expensive than retaining already existing customers (Bauer et. al, 2002; Athanasopoulou, 2009; Christodoulides and Michaelidou, 2010). Thus, retaining customers appears to be the most efficient and profitable path for companies.

Moreover, globalization and access to information and technology are resulting in an important expansion of internet sales. For example, as we can see in *Figure 1*, USA e-commerce retail sales represented just 5% of the total retail sales in 2007 (in millions of dollars), while this figure rose to 16% in 2019.
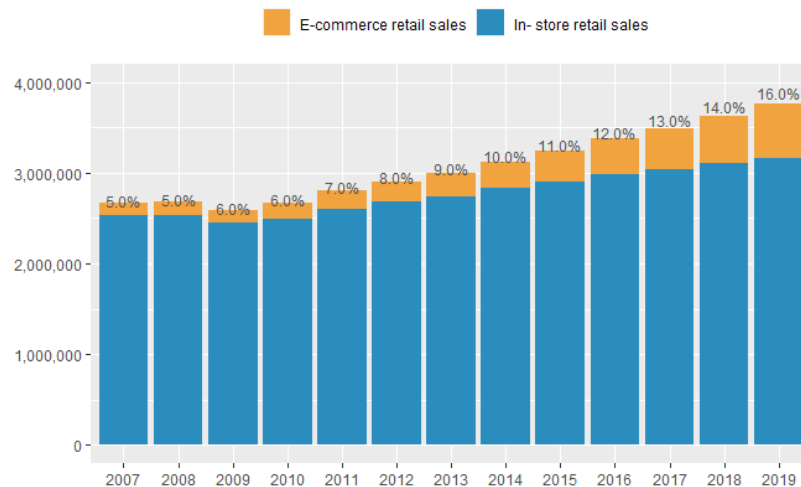

Figure 1: E-commerce share in total retail sales (in millions of dollars) in the USA. Source: Digital Commerce 360.

What is more, in recent months, with the COVID-19 pandemic, the e-commerce channel became essential for retail, as restrictions to go out lead to most of the companies having to close their physical stores. As a consequence, those companies that did not have online channels to connect with customers, had to adapt as quickly as possible in order to stay in business.

Therefore, to fit in this new internet-information era, it is important for marketers to understand what customers value when shopping online. Concepts like online relationship marketing and internet customer loyalty are becoming more relevant, as companies are trying to understand not only what makes people visit their website, but how to make them return to the website (Ilfeld and Winer, 2002). In this sense, Danaher et. al (2002) find that big brands show higher customer loyalty rates when bought online than in offline environments. Additionally, according to Reichheld and Schefter (2000, p.106) "Most of today's on-line customers exhibit a clear proclivity towards loyalty, and Web technologies, used correctly, reinforce that inherent loyalty". What is more, Zineldin (2000, p.15) finds that "effective use of relationship based on IT encourages the establishment of long-term relationship marketing with customers, suppliers, competitors, and others in the organization's external environment".

Thus, retailers are investing heavily in learning more about these matters. For example, in the late 2000's the UK pharmacy chain Boots began testing digital strategies with "clinical-style" methodologies. Furthermore, around the same period the giant American retailer Macy's launched a program to attract the best technology experts in the industry, to help them boost their online channel (Rigby, 2011).

In this context, several studies address issues related to relationship quality and customer loyalty in an online environment, from a customer satisfaction and trust point of view, by surveying customers or interviewing people inside the industry. For example, Novak, et. al (2000) find that consumers having a compelling online experience in general value shopping features that make shopping "smooth" (like easy ordering, easy to contact, easy to cancel, easy returns and quick delivery), customer support, variety, and quality information. Similarly, Gommans et.al (2001) find that features like fast page loads, easy to navigate, personalization, designed for target audience, language options, effective search options and quick

shopping check out process are fundamental drivers of customer loyalty. Additionally, concerning customer service, the authors indentify that features like fast response to customer inquiries (or frequently asked questions sections), easy to contact, free online applications, easy payment methods, fast delivery, delivery options and customer reward system are important.

To the best of my knowledge, there is still little literature about how specific internet marketing strategies and website functionalities can influence the company's relationship with consumers, analyzing them directly (not having to rely in surveys to customers on interviews), and about how this results in customer loyalty.

Thus, this thesis intends to fill that gap by answering the following research question:

How can internet marketing and website functionalities increase customer loyalty?

To answer this question, we use data from the "top 500 online retailers report" of US database for 2016, web scraped from Digital Commerce 360. The report contains a wide range of variables, including internet marketing activity indicators, website and sales performance indicators, website functionalities indicators and customer demographics (see Data section). Thus, we analyze specific data about website's features and internet marketing of successful companies.

According to Anderson and Srinivasan (2003, p.125), e-loyalty is defined as "the customer's favorable attitude towards an electronic business resulting in repeat buying behavior". Following their definition, in this thesis we measure customer loyalty with the returning shoppers' rate, assuming that websites that have higher returning rates, have higher customer loyalty rates.

Additionally, we aim not only to find the main drivers of shoppers returning (customer loyalty), but to create a machine learning model that can predict more accurately what the returning rate of my website will be, given those drivers. This, also contributes to enhancing marketing literature, as to the best of my knowledge, there is still no research on this topic that implements machine learning.

Therefore, this thesis helps marketing managers to allocate their budget more efficiently, as they can focus on the main drivers found to increase customer loyalty. What is more, it helps companies that do not have a web store yet or that want to improve it, estimate the success of implementing certain functionalities on the website.

To find the main drivers of the *Returning Shoppers' rate* of the websites, we perform a regularized linear regression. Particularly, we implement a Lasso liner regression. Lasso regression allows us to avoid overfitting the model on the training sample, which is particularly important in our case, as we have a limited amount of observations and a high number of variables (see Data section). This way, we can then generalize results and predict better other website's returning rate. Additionally, as the aim of this thesis is not only to find the main drivers that explain customer loyalty, but to help marketers predict how their strategy will perform, we then contrast the linear regression with the predictions of a more complex Machine Learning model. Specifically, we perform Boosted Regression Trees.

This report is structured as follows. First, a literature review about the academic papers already written on relationship marketing and customer loyalty on the internet and how they relate to our study. After that, a data section explaining our data set and containing descriptive statistics. Thereafter, a methodology section containing the characterization of the methods used to perform our analysis and the key performance measures implemented to evaluate them. Next, we present the analysis of the models and results. Finally, we end the report with the conclusion, including the academic and managerial implications and the limitations of our study.

2. Literature Review

The topic of Customer Loyalty has been widely researched in the marketing literature. Specifically, the concept emerged in the 1950s, when it was mainly defined as a repeating purchasing behavior (Srinivasan et.al, 2002). One of the first researches to go deep into the topic was George H.Brown, who classifies customer loyalty in 4 categories: "undivided loyalty", "divided loyalty", "unstable loyalty", and "no

loyalty" (Brown, 1952), depending on the purchase frequency. Later, in the 1960's other authors like John U. Farley focused on defining the concept using the economics of information framework. According to the researcher, loyalty depends on the market research and information each household does or has. Consequently, if one considers that most brands are good substitutes between each other, households that have more information tend to be less loyal (Farley, 1964). Nevertheless, nowadays, in a highly globalized economy, this theory seems inadequate, as customers have easy access to much more information, especially in an online setting. Furthermore, in the late 1960's some researchers suggested that the behavioral definition of customer loyalty was not enough and added the attitudinal aspect to the definition, which is based on the premise that a customer is loyal not only by repeatedly buying a brand, but by having a positive attitude towards it (Jacoby and Chestnut,1978). Since then, much of the research conducted on Customer Loyalty, has been on finding different customer loyalty measures combining the attitudinal and behavioral approaches (Schultz and Bailey, 2000).

The topic of Relationship Marketing first emerged in the academic literature, in the early 80's in the context of Services Marketing, when Leonard Berry defined it as "attracting, maintaining and in -multi service organizations- enhancing customer relationships" (Berry, 1995, p.236). It became more popular later in the 90's, when authors started emphasizing on the importance of the relationship to be long lasting and how it can increase customer loyalty. For example, Evans and Laskin (1994) find that relationship marketing can help companies to differentiate from their competitors, building a relationship with customers that will lead to customer loyalty.

In the late 90s, with the rise of internet, relationship marketing and customer loyalty gained more relevance in an online setting, as companies needed to start learning how to connect with the customers on this environment. As a matter of fact, technology and internet are driving us to something similar to what economist call "perfect market", what is increasing competition between companies, making customer loyalty crucial for companies to succeed (Srinivasan et. al, 2002). According to Kozlenkova et.al (2017, p.21) "as online sales grow and customers gain e-commerce experience, online shopping also is evolving from primarily a transactional exchange to a more relational-based exchange, similar to traditional retail

interactions". In this sense Bauer et.al (2000) find that internet has the potential to boost customer trust and commitment in relationship marketing. Moreover, Zineldin (2000) states that in order for relationship marketing to be an efficient strategy tool, technology must be effectively implemented.

Considering all these, we see that even though Relationship Marketing and Customer Loyalty have been present in the marketing literature for a long time, the analysis of the topics in an online environment is still relatively recent. Consequently, still little is known about the particularities of these key marketing concepts in an online setting.

Below we present a more detailed analysis of this concepts and how we understand them in this thesis.

2.1 Relationship Marketing

There exist various definitions of relationship marketing, which highlight different characteristics of the concept, but most of the authors agree that it should be a long-lasting relationship. For example, Jackson (1985, p.2) defines it as "marketing oriented toward strong, lasting relationships with individual accounts". Similar to them, Morgan and Hunt (1994, p.34) define relationship marketing as "all marketing activities directed towards establishing, developing, and maintaining successful relational exchanges" and Evans and Laskin (1994, p.440) state that "Relationship marketing is the process whereby a firm builds long-term alliances with both prospective and current customers so that both seller and buyer work toward a common set of specified goals". Furthermore, Zineldin (2000) emphasizes that developing good communication channels is fundamental for marketing relationships and summarizes the key characteristics of relationship marketing established by Adrian Payne in 1995, as focus on customer retention, orientation to customer value, long time- scale, high customer service emphasis, high customer contact and quality concern.

More recently, researchers emphasize that online relationships are different from offline relationships because we cannot interact directly with other people (Yadav and Pavlou, 2014; Verma et.al, 2016, Kozlenkova et al.,2017; Steinhoff et. al, 2019). Moreover, according to KPMG (2017) online relationships are considerably different than offline relationships. Specifically, in their "2017 Global Consumer Report" the firm states that the "path to purchase" differs significantly between offline and online transactions. This is, because, even though the consumers still go through the same four stages "awareness", "consideration",

"conversion" and "evaluation", in the online journey consumers have access to more information and can go back and forth from stage to stage. In this sense, Kozlenkova et.al 2017 also establishes that there are some characteristics of the online channel that make offline marketing strategy inadequate. For example, "anonymity", which is forcing companies to develop new techniques to try to understand the customer.

As we can see, researchers and marketers are still trying to figure out the differences between online and offline Relationship Marketing, and how to generate a long-lasting relationships with consumers in the online environments, where consumers have much more access to information (Steinhoff et. al, 2019). For example, by interviewing website designers to understand how, among others, Relationship Marketing influences web design, Geissler (2001) finds that online consumers are more impatient. Thus, response time is fundamental in an online environment. Also, the author finds that interacting with consumers with policies like gathering customer information, encouraging feedback, answering questions and providing information in the ordering process are key elements marketers should take into account.

2.2 Customer Loyalty

Relationship Marketing is crucial to achieve Customer (or brand) Loyalty, which can also have different characteristics in an online context, as customers tend to value other aspects when shopping online. Thus, researchers often refer to online customer loyalty as "e-loyalty".

As mentioned before, early definitions of customer loyalty, such as Brown's (1952) definition, focused mainly on the behavioral aspects of the concept (characterizing as a repetitive behavior). Another example is Cunningham's (1966, p.118) definition, who measures brand loyalty as "the proportion of total purchases represented by the largest brand used". Later, the attitudinal aspect was added to the concept, arguing that not only the repetitive purchase mattered, but the consumer has to have a positive attitude towards the brand (Jacoby and Chestnut,1978). Considering all these, Jacoby and Kyner (1973,p.2) define brand loyalty as "the biased (i.e., nonrandom), behavioral response (i.e., purchase), expressed over time, by some decision-making unit, with respect to one or more alternative brands out of a set of such brands, and is a function of psychological (decision- making, evaluative) processes". Moreover, Keller (1993, p.8) states that customer

loyalty "occurs when favorable beliefs and attitudes for the brand are manifested in repeat buying behavior".

Additionally, Oliver (1999, p.34) adds the notion of commitment to the concept and defines it as "a deeply held commitment to rebuy or re patronize a preferred product/service consistently in the future, thereby causing repetitive same-brand or same brand-set purchasing, despite situational influences and marketing efforts having the potential to cause switching behavior."

Relating to our subject of study, according to Gommans et. al (2001) customer loyalty is more difficult to achieve in an online context, as customers have access to much more information than they have when they buy offline. Additionally, analysts often have trouble finding the correct measure for e-loyalty, as not only repurchasing, but also repeated visits to the website and time spent on the website can matter (Smith, 2000). Thus, taking into consideration the previous research on customer loyalty and trying to adapt it to an online setting, Anderson and Srinivasan (2003, p.125), define e-loyalty as the "customer's favorable attitude towards an electronic business resulting in repeat buying behavior".

2.3 Previous Research

Several studies have focused on finding what elements drive customers to keep returning and buying on a website. Novak, et. al (2000) find that consumers having a compelling online experience in general value shopping features that make shopping "smooth" (like easy ordering, easy to contact, easy to cancel, easy returns and quick delivery), customer support, variety, and quality information. Similarly, Chen and Chang (2003) find factors that influence online shopping experience positively are associated with interactivity, transaction and fulfillment. Furthermore, Wolk and Theysohn (2017) find 16 factors that affect website traffic. These factors are: quality, uniqueness, relevance, personalization, branding, price level, price discrimination, business model, payment system, interactivity, website organization, navigation, accessibility, actuality, credibility and visibility of the website.

Other authors focus on trust as a major loyalty driver. Even though trust is always a key determinant when it comes to customer loyalty, it is even more important in an online environment, where customers do not have real contact with the sellers (Gommans et al, 2001). In this sense, Bauer et. al (2002, p.159) find that

"customers who trust a corporation feel more committed to it". Additionally, among other factors, the authors find that efficient information transfer can also increase commitment and that the possibility to shop online increases trust among customers. Verma et. al (2016) find that mainly, trust, relationship quality and relationship satisfaction contribute to customers to be more loyal on online retailing. What is more, Rafiq et. al (2013) concludes that trust has no direct effect on customer loyalty on the e-retail context, but that the main driver of loyalty is relationship satisfaction. According to the authors, the effect of trust is transmitted through relationship satisfaction.

Meanwhile, other studies focus on more specific drivers of e-loyalty. Among other determinants, like brand building, trust and security and value prepositions, Gommans et.al (2001) find that Website and Technology and Customer Service are key determinants of customer loyalty online. Specifically, within Website and Technology they find that features like fast page loads, easy to navigate, personalization, designed for target audience, language options, effective search options and quick shopping check out process are fundamental. Additionally, concerning customer service, they find that features like fast response to customer inquiries (or frequently asked questions sections), easy to contact, free online applications, easy payment methods, fast delivery, delivery options and customer reward system are important.

In the same vein, Reichheld and Schefter (2000) describe how Dell created a customer loyalty council and discovered three main drivers of e-loyalty, namely, order fulfillment, product performance and post-sale service and support. Also, by studying leading internet companies' websites and surveying their customers, the authors find five main determinants of customer loyalty, which are, quality customer support, on-time delivery, compelling product presentations, convenient and reasonably priced shipping and clear privacy policies. Furthermore, Srinivasan et. al (2002) identify seven factors that significantly impact e-customer loyalty, which are, customization, contact interactivity, care, community, cultivation, choice and character.

2.4 Applying the theoretical framework to our case

As we have seen above, researches have written a lot about online relationship marketing and customer e-loyalty, and about the key aspects that marketers should take into account when defining their strategies.

13

Researchers find that the two most important loyalty drivers are trust and satisfaction. In this sense, there are some key elements on the online environment that they find contribute to boost them. Specifically, in what relates to website features and functionalities most studies agree that features related to security, delivery fulfillment, personalization, quickly answering customers' requests, helping customers to search quicker, helping customers to understand better, easing checkout process, easy and secure payment methods and post-sale services are important.

In this report we try to connect all these features already found, to the internet marketing policies and website functionalities we know successful retail websites are already implementing. We will measure the success of these policies by looking at the *Returning Shoppers' rate*, following Anderson and Srinivasan definition of e-loyalty. Additionally, we will be doing this by implementing machine learning techniques to optimize the prediction and create a framework that marketers can then use to predict the effectiveness of their strategy.

3. Data

The data set consists of web scraped data from Digital Commerce 360's "top 500 online retailers in the US" 2016 report. The original data contains 280 variables and 500 e-retailers. The data includes features that indicate whether several marketing activities are performed inside the company or outsourced to another company (e.g., customer service software, marketplace management, affiliate marketing, e-mail marketing, Online advertising, etc.), as well as variables that indicate if the website has specific features (e.g., preview search, product customization, frequently asked questions' section, etc.). Furthermore, it has variables that indicate if the website offers certain customer services (e.g., private label credit card, international delivery, next day delivery, free shipping, live chat, multiple languages, etc.). Finally, we can also find other variables like number of social media followers, payment methods, merchant type, sales, conversion rate, monthly visits, percentage of traffic from different channels, average ticket, conversion rate and shoppers' demographics.

Taking into consideration our theoretical framework and removing variables that even though useful had a high number of missing values, we end up with a data set of 67 variables and 370 companies. A complete list of the variables included with their descriptive summary statistics can be found in Appendix A - *Table 1*. Note that the web traffic variables are not included in the analysis, as we understand that the loyalty rate influences them and not the other way round. In other words, websites that have higher loyalty rates, also have higher direct traffic rates, as customers go directly to buy what they want.

As explained in the previous sections, our goal is to predict customer loyalty, thus, our dependent variable is the *Returning Shoppers' rate 2015*. The variable indicates the percentage of returning shoppers out of the total website shoppers during the year 2015. As we can see in *Figure 2* the variable is quite evenly distributed, with a mean value of 38% and a median value of 37%.
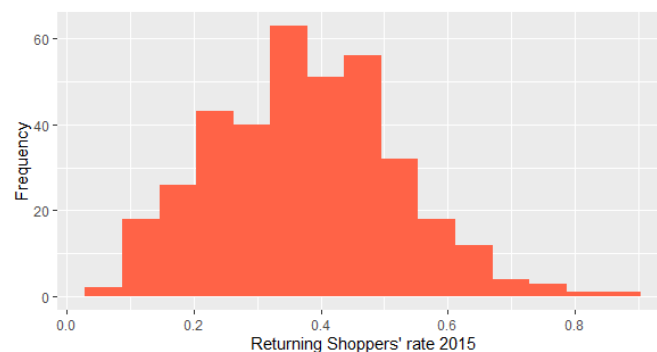

Figure 2: Histogram depicting Return Shoppers' rate distribution.

Other continuous independent variables included in the analysis (*Monthly 2015 Email Campaigns*, *Monthly Average Paid Search Spending 2015, Response time in seconds and Site Availability*) are transformed using the logarithm function to get a less skewed distribution, particularly for *Monthly Email Campaigns* (see Appendix A – *Figure 1*). Additionally, as we can see in Appendix A – *Figure 2* no strong correlations are found between the continuous independent variables, and the strongest correlation between these variables and *Returning Shoppers' rate* appears to be with *Monthly Email Campaigns* (0.4). As we explain in the methods section, the continuous variables are standardized in order to perform lasso regression, so that the

tuning parameter λ is just to all variables. Lastly, is important to mention that we express the *Monthly Average Paid Search Spending 2015* relative to sale, so that we can compare it between companies.

To get a first impression of how the chosen features may impact the *Returning Shoppers' rate* we plot exploratory graphs for a selected group of variables (*Figure 3*). We can observe associations with *Monthly Email Campaigns*, *Monthly Average Paid Search Spending* and *Response time in seconds*. Additionally, if we look at the categorical variables, we can see that the presence of most website functionalities or internet marketing strategies, seem to affect positively the Returning Shoppers' rate, except for *Google Wallet* and *Next Day Delivery*.

In order to perform our analysis, we divide the data in two subsets. The training dataset, where we train the model and tune the parameters and the test dataset (or validation set), where we measure the real performance of our models (out of sample performance). In this report we choose to use 80% of the observations on the training data set and 20% of the observations on the test data set. We randomly assigned the observations to each dataset, and checked that the mean Returning Shoppers' rate was similar in both subsets.
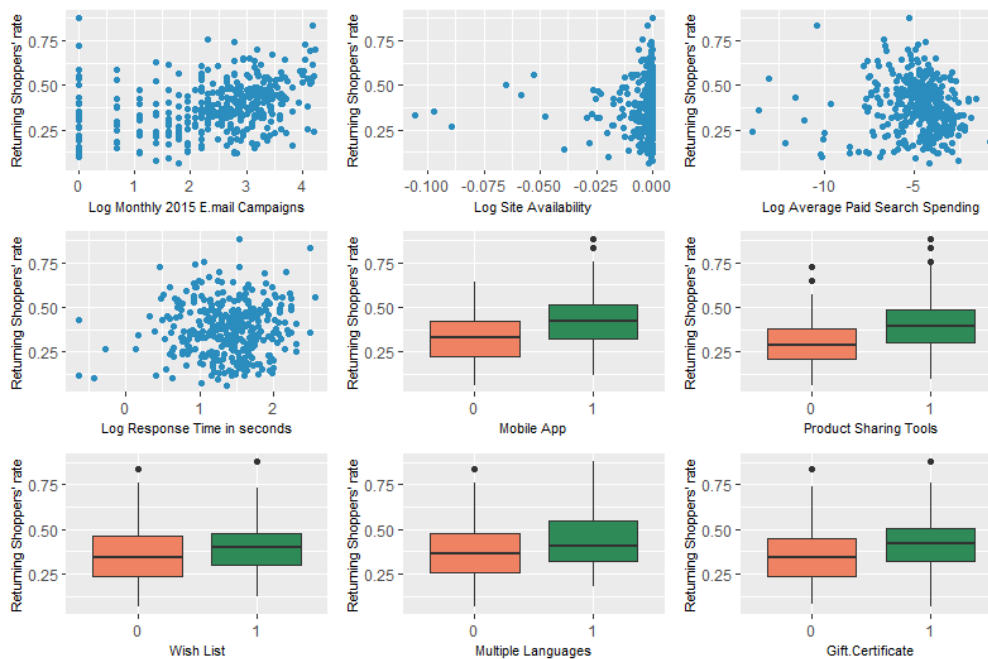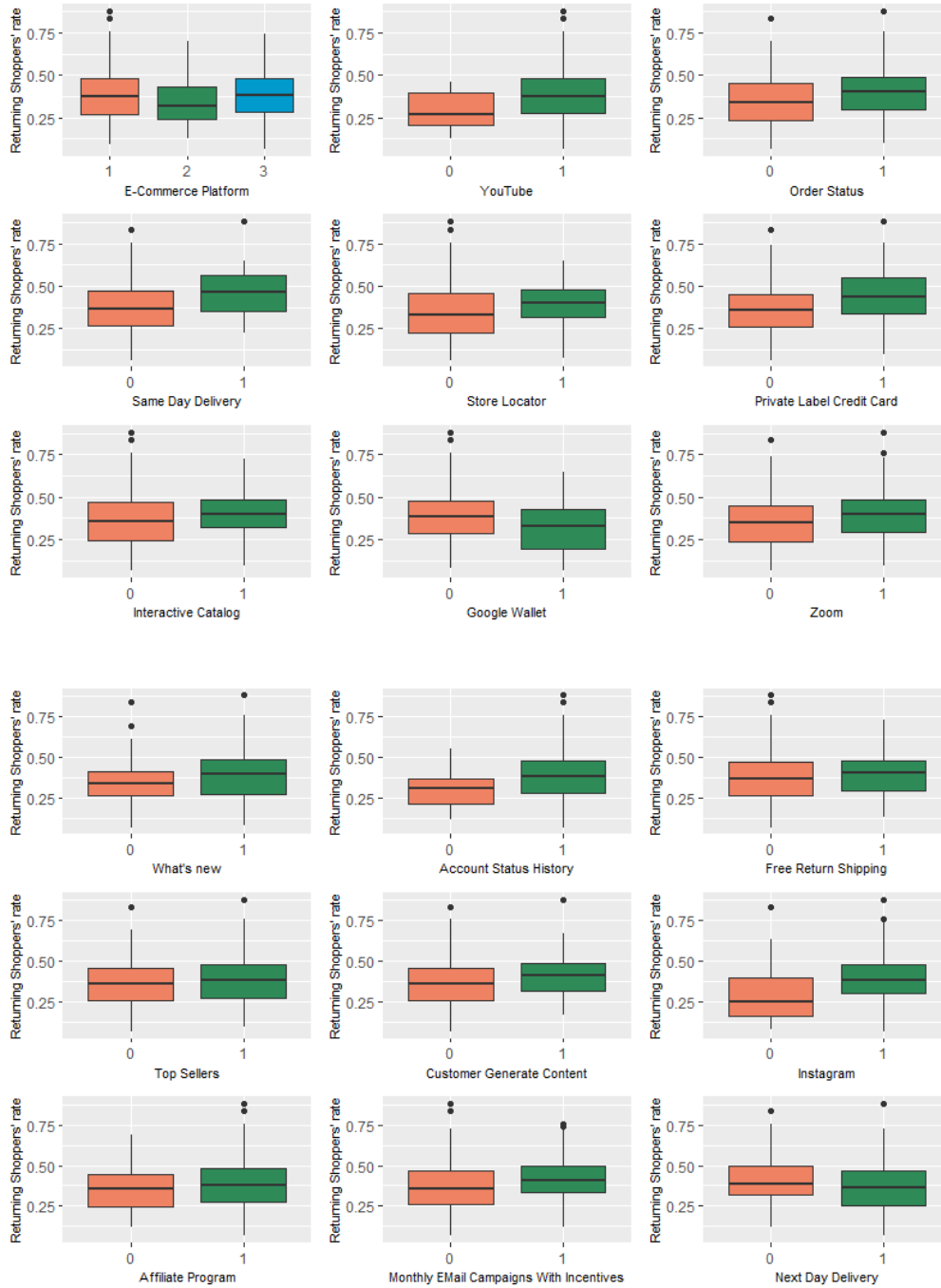
Figure 3: Scatter plots and box plots depicting the Return Shoppers' rate relation with the explanatory variables.

4. Methods

4.1 Regression models

There are two types of machine learning algorithms supervised and unsupervised. In supervised learning the aim is to predict an outcome based on a set of features (the predictors), while in unsupervised learning the aim is to find patterns within the data to get a better understanding of it. Additionally, within the supervised methods we have two types of problems: regression problems and classification problems. In regression problems the goal is to predict a quantitative outcome, whereas in classification problems the goal is to predict a categorical outcome (Boehmke and Greenwell, 2020). As the purpose of this report is to predict the returning shoppers' rate (a numerical variable), we solve a regression problem. As previously mentioned, we will use two different techniques to predict the returning shoppers' rate: Lasso Linear Regression and Boosted Regression Trees. Lasso Linear Regression is a classical statistical method, while Boosted Regression Trees is a more complex machine learning technique. The two of them are supervised methods that can be used to predict continuous outcomes. Boosting can also be adapted for classification problems, but that is outside the scope of this thesis.

Before explaining the models, we introduce the *Bias-variance trade-off*, which is a trade-off we run into every time we try to make predictions. Thus, understanding it will help us to get a better comprehension of the models.

The variance of a model represents how sensitive the fit of the model is to changes in the training data set. A method has high variance when different training data sets lead to important differences in the estimated models. The bias of a model is the error that we introduce when we fit the data approximating it to a rather simple mathematical form, when, in reality, the form is more complex. As we introduce more predictors (increase the flexibility) into a model, the fit will be better, contributing to reducing the bias of the model. However, we are simultaneously increasing the variance of the model, as we are improving the fit on that specific training sample. This is what is called the bias-variance trade-off. As our goal is to predict, we want

an specification of the model that works well on out of sample observations as well, and thus, we are willing to allow an increase on the bias of the model, in exchange for a reduction on the variance (*see Figure 4*).
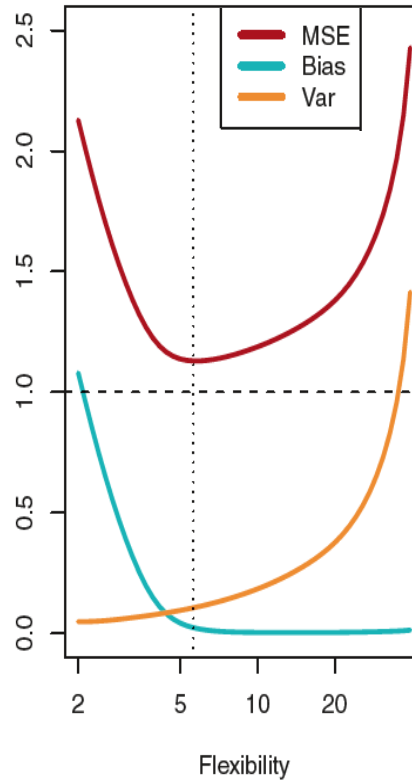

Figure 4: Representation of the Bias-variance trade off. Image taken from James et.al (2013, p.36).

4.1.1 Lasso Linear Regression

Introduction to Multiple Linear Regression

Multiple Linear Regression is a widely implemented statistical method used for predicting a quantitative outcome (the dependent or response variable) with a set of predictors (the independent variables), which assumes that there is a linear relationship between the response and the predictors. Using similar notation to James et. al (2013, p.71) the model is defined by:

$$(1) \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \epsilon,$$

where Y is the response variable, X_j are the predictor variables and β_j the weights or coefficients. The β_j quantify the effect of a one unit change in X_j on the response variable, given all the other variables stay constant (p is the total number of predictors). β_0 is the intercept, and it represents the expected value of Y when all the X_j are zero. ϵ is a random error term assumed to be normally distributed with mean zero and constant variance. The predictors X_j as well as the errors ϵ are assumed to be independently distributed.

The goal is to find the β_j such as that the linear model fits best to the data. There are several approaches for doing this. The most commonly used is the Ordinary Least Squares (OLS) approach, which finds the β_j by minimizing the Residual Sum of Squares (RSS), which is the squared sum of the difference between the real values and the estimated values. Using similar notation to James et. al (2013, p.72) RSS is defined as:

$$(2) \quad RSS = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 = \sum_{i=1}^{n}(y_i - \widehat{\beta_i}x_{i1} - \widehat{\beta_i}x_{i2} - \ldots - \widehat{\beta_i}x_{ip})^2,$$

where y_i represents the observed values of my dependent variable, $\widehat{y_i}$ the estimated values of y_i , $\hat{\beta}$ the estimated values for the coefficients β, n the number of observations and p the number of predictors.

Lasso Linear Regression

In this thesis, we will apply Lasso restriction to the Multiple Linear Regression. When the number of predictors is high related to the number of observations, OLS can have two major problems. On the one hand, the estimations have high variance and, consequently, the method might not find a unique solution for the minimization problem. On the other hand, because of the high variance we can also end up overfitting the model on the training sample. Therefore, as our goal is prediction, we want a model that performs well on unseen data.

Lasso is a regularization method which helps us fitting the model on a high dimensional space, by shrinking certain coefficients to zero, and thus, reducing the variance of the model and performing variable selection.

The minimization problem with Lasso restriction is very similar to OLS, but we add an additional term to the equation, the penalty term. This way, using similar notation to James et.al (2013, p.219) we have,

$$(3) \quad \sum_{i=1}^{n}(y_i - \beta_0 \sum_{j=1}^{p} \beta_j \, x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| = RSS + \lambda \sum_{j=1}^{p} |\beta_j|.$$

The first term is exactly the same as equation (2), while the second term is the Lasso restriction and it is called the shrinkage penalty. λ is a non-negative tuning parameter ($\lambda \geq 0$). If $\lambda = 0$ then we have the OLS problem, while as λ increases, the higher the penalty and the more coefficients are reduced to zero. Note that the restriction does not apply to β_0, as like we mentioned previously, the intercept represents the expected average value of the response variables when all the predictors (X_j) are zero and, if we applied the penalty to it, we will lose interpretation. In summary, by shrinking the coefficients towards zero (or exactly to zero), Lasso regression can still perform well, even in a high dimensional space, trading a little increase on the bias of the estimators for a reduction of the variance.

We choose the value λ using cross-validation, choosing λ such as that the prediction error is minimized. Specifically, in this report we choose it using 10-fold cross validation. In 10-fold cross validation we divide the training sample in 10 equally sized sub-samples, treating the first sub-sample as the validation set and training the model in the other 9, we then do the same with the second one and so one. To measure the performance, we use The Mean Squared Error (MSE), choosing the value of λ that yields the lowest MSE. As we train the model 10 times, the MSE is an average of all the 10 estimates of the test error. Following the same notation as James et.al (2013, p.29) the MSE is defined as:

$$(4) \quad MSE = 1/n \sum_{i=1}^{n}(y_i - \hat{y_i})^2.$$

There are two common approaches to choose λ, λ_{min} or λ_{1se}. λ_{min} is the value of λ that yields the lowest cross validated average MSE, while λ_{1se} is the λ one standard deviation away from this value and it is considered to be a more conservative option as it will shrink more parameters towards zero (and, thus

the probability of overfitting the model is smaller). In this thesis we choose to work with or λ_{min}, as we would like to get more insights on which variables might affect customer loyalty.

It is worth mentioning that unlike OLS, where the scale of the predictors is not an issue, as if we multiple the level of the variable by a constant C for example, then the resulting coefficient of that variable will be scaled by a factor of $1/C$(James et.al, 2013), it is an issue for Lasso. This is because when we add the Lasso penalty term, all the coefficients β_j are equally penalized. Thus, in order to perform the method, we must standardize the variables. Otherwise, variables that have a higher variance will also have higher coefficients. In this report we turn the variables into z-scores, by subtracting its mean and dividing it by its standard deviation (see data section). By doing this we lose interpretability, and we can no longer interpret the β_j as the effect of an increase of one unit of X_j on the response variable (now our effects are in terms of standard deviations). Nevertheless, as all the variables are in the same scale, we can still compare β sizes, to see which variables have the greatest influence.

4.1.2 Boosted Regression Trees

Ensemble methods aggregate the predictions of base learners (or weak learners) in order to improve the accuracy of the models. According to Hastie, et al (2009, p.605) "ensemble learning can be broken down into two tasks: developing a population of base learners from the training data, and then combining them to form the composite predictor".

Boosting is an ensemble method which works by training models subsequently, doing this in a way in which every new model learns from the mistakes of the previous model. There are different boosting algorithms, one of the most popular ones is the AdaBoost algorithm (Adaptive Boosting) formulated by Yoav Freund and Robert Schapire in 1996. Later, in 1999 Jerome Friedman formulated the Gradient Boosting algorithm, which is more flexible than the AdaBoost approach and it is the one we use in this thesis. Boosting can be applied to a wide range of statistical and machine learning techniques, to solve both regression and classification problems. One of the most common applications of the method, and the one

we use in this report is to boost Decision Trees. Thus, before explaining how exactly it is that Gradient

Boosting works, we will start with an introduction to Decision Trees.

Decision Trees

Decision Trees can also be used to address both, regression and classification problems. In this thesis we

will focus on how to implement them to solve regression problems, and thus, we call them Regression

Trees. The Regression Trees' algorithm is based on dividing the data in different non-overlapping sub-

spaces called "regions", what is called "Recursive Partitioning". The algorithm does this by choosing

different cut points based on the values of the predictor variables and usually performing binary splits.

Then, depending on which region my observation falls into, the prediction the tree will yield. The name of

the method comes from the fact that we can then visualize the different decision rules and regions in the

form of a 'tree". We call the points where the regions split "internal nodes", the segments of the tree that

connect the different regions the "branches" and the final points, where the prediction is shown, the "leaves

of the tree" or "terminal nodes". *Figure 5* depicts an example of a Regression Tree with 4 terminal nodes.

Taking it to mathematical terms and using similar notation to James et. al (2013, p.314), Regression Trees

can be expressed as:

$$(5) \quad f(x) = \sum_{m=1}^{M} c_m \, I(x \in R_m),$$

where R_m represents the different M regions the data is split into, and c_m represents the predicted value for

the observations that fall into the mth region (R_m).

There are different Recursive Partitioning algorithms that we can use to find the optimal solution for the

decision trees. The most commonly used is the CART algorithm, which chooses the best split by

minimizing the RSS at each step. Nevertheless, this approach has two major problems. On the one hand it

can lead to overfitting, as we tend to grow large trees. This can be solved by pruning the trees, so that we

reduce the flexibility and, thus, the variance. On the other hand, CART trees suffer from selection bias, tending to select variables with more splitting points, such as continuous variables or categorical variables with many categories. While selection bias may not be a problem for prediction, it is for interpretability, as features that have no relation to the response variable might be selected (Hothorn et. al, 2006).
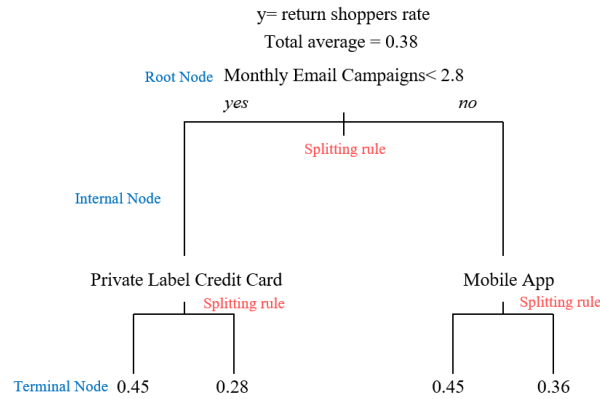

Figure 5: Regression Tree Example.

Thus, as most of the variables we have are categorical, in this thesis we will apply Gradient Boosting with an "Unbiased Recursive Partitioning algorithm". Many studies conclude that one of the main causes of the selection bias is that algorithms like CART combine the variable selection and split selection in the same step (Hothorn et. al, 2006). Thus, in Unbiased Recursive Partitioning we separate the process into two steps. Additionally, the selection of a variable is based on its statistical significance, and not just in minimizing a certain loss function.

According to Hothorn et. al (2006) a generic Recursive Partitioning Algorithm includes the following steps:

1) Test the partial null hypothesis of independence between the independent variables and the response variable ($H_0: D(Y/X) = D(Y)$). If we reject the hypothesis, we choose the variable with the strongest association to the dependent variable to make the split.

2) For the selected variable implement a split.

3) Repeat steps 1) and 2) for the next splits until we can no longer reject H_0.

In this thesis, we specifically work with Conditional Inference Trees, which were introduced by Hothorn, Hornik, and Zeileis in 2006. As the distribution of the explanatory variables is unknown, Conditional Inference Trees select variables calculating its significance by performing permutation tests, which consist of fixing the values of the explanatory variables and conditioning them on the permuted values of the response variable. The authors give different specifications of statistics we can use to do this. In this report, we choose to use the max function as test statistic and the test type t-statistic. By using the t-statistic we just compute the regular p-values and by using the max function we choose the maximum standardized value of the chosen statistic. Additionally, we work with a min criterion of 95%, which means that in order to reject H_0 (independence between the explanatory variables and the response variable) the p-value has to be 5% or lower.

Unlike minimizing a loss function, as the CART algorithm does (the RSS), after selecting the variable we select the best split by performing permutation tests again and selecting the split that yields the maximum standardized value of our t-statistic.

The algorithms will stop making splits when the variables are no longer significant at the level of significance we have chosen, or at the stop criterion we have set; for example, a minimum number of observations per terminal node or a maximum number of terminal nodes (max depth).

Finally, it is important to notice, that unlike linear regression, where we had to apply a Lasso restriction to perform variable selection, Decision Trees already do this when they choose the best split at each point. Therefore, variables that do not contribute to improve the fit of the model will automatically be left out. Additionally, Regression Trees work well with mixed data (categorical and continuous variables) and can capture non-linear relationships between the independent variables and the response variable, even identifying interactions and complex relationships, that other simple methods like linear regression cannot capture.

But Decision Trees have a major problem, which is that they suffer from high variance, meaning that its output is highly dependent on the specific training sample. That is why Decision Trees are often improved using ensemble methods. Even though, we lose interpretability, ensemble methods can highly improve the accuracy of the models.

Gradient Boosting

As aforementioned Boosting is an ensemble method. Boosting works by subsequently training models and adding their results, what is known as "Stage wise additive modeling". The main idea is that each new model learns from the mistakes of the others. There are different ways of doing this. Gradient Boosting works by training the first model on the whole data and then training the subsequent models on the residuals of the previous models. According to Friedman (2002, p.367), "Gradient boosting constructs additive regression models by sequentially fitting a simple parameterized function (base learner) to current "pseudo"-residuals by least squares at each iteration". Thus, unlike other ensemble methods like bagging or random forest that mainly focus on reducing the variance of the model, boosting mainly reduces the bias, as it is constantly improving the fit. That is why, as we mention later, boosting usually trains shallow trees, which have low variance. This way, the algorithm usually ends up reducing both, the bias and the variance of the model.

Before going deep into how exactly the gradient boosting algorithm works, it is important to explain the intuition of the gradient descent algorithm, which is the approach boosting uses to find the best solution. The gradient descent algorithm is a widely used approach that allows us to find the parameters that minimize a specific loss function. We can use gradient descent in any loss function that is differentiable. The algorithm begins by assigning a random value to the parameters we want to estimate. We then calculate the gradient (slope of the curve) of the loss function for that specific value. The value of the slope determines the size of the step the algorithm will take towards minimizing the loss function. Thus, the gradient descent will take bigger steps when it is far from the optimum and smaller steps when it is closer (as the slope in that point is zero). Additionally, a tuning parameter called the learning rate (which varies between 0 and 1)

is added to the equation, so that we are be able to scale the size of the steps. This is, because we do not want the algorithm to move too fast, as it may skip the minimum and end up on the other side of the function, for example. Thus, we multiply the slope by the learning rate, so that the algorithm takes smaller steps (Gerón, 2017). *Figure 6* depicts an example of the gradient descent algorithm with a concave loss function.
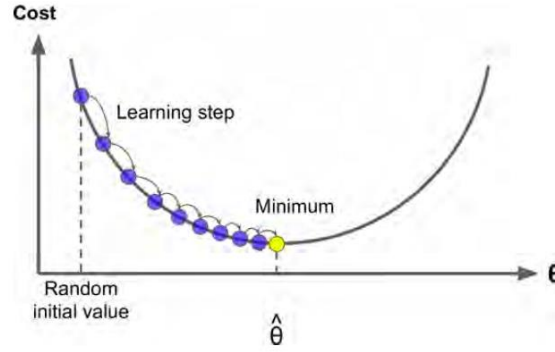

Figure 6: Representation of the Gradient descent algorithm. Image taken from Gerón (2017, p.111).

According to Friedman (2002) Gradient boosting uses gradient descent to estimate the pseudo residuals. "The pseudo-residuals are the gradient of the loss functional being minimized, with respect to the model values at each training data point evaluated at the current step" (Friedman, 2002, p.367).

This way, using similar notation to Hastie et. al (2009, p.361), the gradient boosting algorithm can be represented by the following steps:

1. Initialize $f_0(x) = argmin_\gamma \sum_{i=1}^{N} L\,(y_i, \gamma)$

2. For $m = 1$ to M:

    (a) For $i = 1, 2, \ldots, N$ compute:

    $$r_{im} = -[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}]_{f=f_{m-1}}$$

    (b) Fit a regression tree to the targets r_{im} giving terminal regions $R_{jm}, j = 1, 2, \ldots, J_m$.

    (c) For $j = 1, 2, \ldots, J_m$ compute:

$$\gamma_{jm} = argmin_\gamma \sum_{x_i \in R_{jm}} L\left(y_i, f_{m-1}(xi) + \gamma\right)$$

(d) Update $f_m(x) = f_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$

3. Output $\hat{f}(x) = f_M(x)$

In order to understand it better we will walk step through each step of the algorithm. In step 1 we initialize the problem with a constant value that we will the try to optimize using the gradient. L is the loss function we want to minimize, which depends on the values of the response variable y_i and on the predicted values γ. There are different loss functions that we can use for regression problems, the most commonly used, and the one we use in this thesis, is the squared error loss function: $L = \frac{1}{2}[y_i - f(x_i)]^2$, which's gradient is $y_i - f(x_i)$. With this loss function, the optimal constant model to initialize the problem is a tree with just one terminal node that yields the average value.

Step 2 is divided in 4 parts. M is the number of trees we fit and N the number of observations we have. In part a), we compute the pseudo residuals r_{im} by minimizing the loss function respect to $f(x)$ our prediction, using the gradient descent approach. We do this for each mth tree and ith observation, evaluating the results at the current step, using the value of the previous estimation f_{m-1}. In part b) we train a new regression tree on the pseudo residuals found in a). In part c) we determine the output values for each terminal node γ_{jm}, adding them to previous prediction f_{m-1}. Given that we choose the squared error loss function, the optimal value yield each terminal node will always be the average of the observations that fall into each leaf. Finally, we update the final prediction, that is the previous prediction plus the prediction of the tree (note that the second part of equation (d) is just the equation of a regression tree presented in (5)). Every tree's prediction is weighted (or scaled) using the parameter v. v is a regularization parameter, which varies between 0 and 1, and we use to control the learning rate of the boosting algorithm. M and v help us avoid overfitting the model on the training sample.

Thus, we have three important parameters we need to optimize: the number of trees M, the learning rate v and the number of splits per tree J. Choosing the correct number of trees is important, as we are fitting every new tree on the residuals of the previous and that can lead us to overfitting. The learning rate also helps us to avoid overfitting, as it is a shrinkage parameter that slows down the optimization problem, scaling the contribution of every new tree. Thus, small values are usually preferred. Nevertheless, a learning rate that is too small can result on never reaching the optimum or to being too computationally expensive. Finally, the depth of the tree is important as it controls the variance of our model, as well as the order of interaction. If we do boosting just with stumps (a tree that has only one terminal node) we will have an additive model with no interactions. Generally, rather shallow trees work well (as they have low variance), so unlike other machine learning methods we do not grow complex trees or perform pruning. We can also set the minimum number of observations per terminal node instead of choosing the number of terminal nodes.

4.2 Performance measures

In order to compare the performance of the models we use the Mean Square Error (MSE), defined in equation (4) computed in the test sample. The model that has the lowest out of sample MSE has the best performance. Additionally, even though we cannot compare it between models (unless we have the same amount of predictors or adjust it) we will also look at the R^2 of the models, to have an idea of the proportion of variance they are explaining. Using similar notation to James et. al (2013) we define R^2 as:

$$(9) \quad RSS = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \ ,$$

where RSS is the residual sum of squares presented in (2) and $TSS = \sum(y_1 - \overline{y})^2$ represents the Total Sum of Squares. $\overline{y}$ is the mean of the response variable. Thus, TSS represents the variance Y already has before fitting the model. Consequently, we can interpret the R^2 as the amount of the variance explained by the model.

Additionally, as we are trying to gain business insights about which variables affect customer loyalty, it is important for us not only to see the prediction accuracy of the models, but also which variables influence it the most. In Lasso regression this is pretty straight forward, as we can directly observe the size and direction of the coefficients. As the variables are standardized (all of them are in the same level), we can directly compare the absolute values of the coefficients and determine which variables are more important for our analysis.

Decision Trees are also easy to interpret, as we can directly observe the variables used to make the splits. However, when we apply boosting we lose interpretability, as we no longer have a single tree, but a sum of trees that we do not observe. Boosting is what is usually called a "black box" model, as we cannot directly see what the algorithm does. Thus, in this report we use other methods that help us interpret the boosted trees results and gain additional insights. Specifically, we will look at the "Variable importance" and at the "Partial Dependency Plots". Both are model agnostic methods, which means they can be applied to any model, independently of how the model is formulated.

4.2.1 Variable importance

There are two common ways of computing the variable importance in black box models. The first one is leaving out features from the model one at a time and measuring the change in performance. The second one, and the one we use in this thesis, is by doing permutation. As previously mentioned, in this report we use the MSE to measure performance. In the leaving out features approach, we train new models excluding the predictors one at a time and then measure the increase in the MSE when we take out each predictor. Predictors that when taken out of the model result in higher increases of the MSE are more important. In the permutation approach, we permute the values of a certain predictor (permute the rows within that feature) and leave all the other variables unchanged. This way, we are breaking the relationship the predictor has with the response variable. We then estimate the model and calculate the increase in the MSE. If after permuting the values the MSE does not show significant changes, it means that variable is actually not important for prediction, as changing its values did not have an effect on the outcome. Same as the leaving

out approach, the bigger the increase in the MSE when we permute the values, the more important the variable is. Using similar notation to Molnar (2019, p.154) the permutation algorithm consists of:

Input: Trained model f, feature matrix X and Loss function $L(y, f)$

1. Compute the original model error e_{orig} using $L(y, f)$. In this case we use MSE.

2. For each predictor $j = 1, 2, ., p$ do:

   (a) Generate feature matrix X_{perm} by permuting predictor j in the data X.

   (b) Estimate the new error e_{perm} using X_{perm} to predict y

   (c) Compute permutation feature importance $FI_j = e_{perm}/e_{orig}$ or $FI_j = e_{perm} - e_{orig}$

3. Sort the features in descending order using FI

But by computing the variable importance we still don't know in what direction the predictor affects the outcome, we only know whether it influences the response variable or not. Thus, we also apply Partial Dependence Plots (PDP), which help us estimate the marginal effect of the predictors on the response variable.

4.2.2 Partial Dependency Plots

Partial dependence plots (PDPs) help us understand the relationship between the predictor and the response variable. The plots can depict linear, monotonous or even more complex associations (Molnar, 2019). Using similar notation to Molnar (2019), the partial dependency function for regression can be represented by the following equation:

$$(10) \quad \hat{f}_{x_s}(x_s) = E_{x_c}[\hat{f}(x_s, x_c)] = \int \hat{f}(x_s, x_c) d\mathbb{P}(x_c),$$

where x_s represents the variables which's effect on the prediction we want to estimate (maximum 2 by plot) and x_c all the other predictor variables. $\hat{f}$ is our machine learning model, in this case the Boosted Decision

Trees. The partial function $\hat{f}_{x_s}$ tells us the marginal effect on the prediction, given a fixed value for x_s.

Thus, we can have a better idea of how that feature relates to the response variable.

We estimate $\hat{f}_{x_s}$ by fixing a value for x_s, and calculating the new predictions, leaving the original values

for all the other x_c. We then average these new predictions and get the function that we plot. Using similar

to Molnar (2019) we have that:

$$(10) \quad \hat{f}_{xs}(x_s) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(x_s, x_c^{(i)}),$$

where n is the number of observations.

But how do we choose the fixed values for x_s? The approach Friedman (1999) uses to do this, is splitting

the variable we want to analyze into J different, equally-spaced values. According to Boehmke and

Greenwell (2020) the algorithm then works as follows:

For a selected predictor x_s:

1. Build a grid of j equally spaced values $x_1, x_2, ., x_j$ across x_s distribution.

2. For each of the J values selected for x_s:

(a) Copy the training data set and replace the values of x_s for each fixed value of x_s.

(b) Apply the machine learning model with the fixed value of x_s.

(c) Average all the predictions for each fixed value of x_s.

3. Plot the average predictions against $x_1, x_2, \ldots, x_j$.

In this report, we use the DALEX package to compute the partial dependency measures, which has a default number of J=100.

PDP can also be computed for categorical variables, just by replacing all the instances of x_s for a specific category, instead of making a grid with different values.

It is important to notice, that even though PDPs are efficient and easy to interpret, they are assuming independence between the x_s and the other x_c. This can sometimes be a problem, as features can be correlated with each other, and thus, that assumption could lead to unrealistic predictions. Additionally, as we compute the marginal effect as the average of all predictions, heterogeneous effects between observations might be hidden. These effects can cancel each other, leading us to conclude that x_s has no effect on the outcome, when it actually does.

5. Results

5.1 Lasso Regression

We start by performing the Lasso linear regression. To do this we first select the best value λ, by performing 10-fold cross validation. As explained in the methodology, in this report we choose to work with λ_{min} which is the value of λ that yields the lowest average cross validated MSE. As we can see in Appendix B - *Figure 1* we clearly have an overfitting problem, as as we increase λ, and thus, reduce the amount of explanatory variables, the average MSE drops. After choosing λ_{min}, we end up with a subset of 23 variables (plus the intercept) out of the initial 67 explanatory variables. Appendix B - *Figure 2* shows how coefficients are shrunken as the value of λ increases.

As we can see in Table 1, the model has a MSE of 0.01325 and of R^2 of almost 40% on the training sample, while this figures rise to 0.01641 and drop to 20% on the test sample respectively.

As the variables are standardized, we cannot interpret the coefficients directly, but we can compare the sizes and determine which variables have the greatest influence on the *Returning shoppers' rate,* as well as the direction of the effect.

 As we see in *Figure 7*, *Private Label Credit Card* appears to be the most important variable when explaining customer loyalty and has a positive coefficient. This is quite intuitive, as private label credit cards are often associated to loyalty programs, which give points or benefits to buyers, creating a closer relationship with them. The second biggest and positive coefficient is the one associated to *Mobile App*, indicating that offering a mobile app increases the level of engagement of the shoppers. Moreover, having *Product sharing tools* (e.g., "share this" bottom in Facebook or WhatsApp), as well as the number of *Monthly Email Campaigns* implemented during the year, also appear to impact positively on the returning rate of the customers,.

Additionally, even though they have a lower impact, other features related to convenience and shipping like *Order Status*, *Free Return Shipping*, *Pre-order* and *Same Day Delivery* appear to contribute positively to the customer re-purchasing on the E-commerce shop. Also, variables related to personalized offers or easy to click features like *Alternative Views*, *Customer Generated Content* and *Top sellers* have a positive effect on the *Returning shoppers' rate*. Finally, other features related to internet marketing and social media contribute slightly to increase customer loyalty (see *Figure 7*).

On the other hand, being on *Google Shopping* negatively affects the *Returning Shoppers' rate*. This is probably because customers shop in that platform instead of on my website. *E-commerce Platform 2* (which means that the platform is managed by the company and by an external company) makes customer's loyalty drop, meaning that is better that the own company manages the platform or directly outsourced it to an external expert than to manage it jointly. Furthermore, even though contrary to what we find in our Literature Review, offering *Next Day Delivery* also affects the *Returning Shoppers' rate* negatively. This could be because companies that have this service do not fulfil it on time, making the customer lose trust. Finally, the *Average Paid Search Spending* is also found to discourage shoppers to come back to the

website, suggesting that companies that spend more in advertising do this to attract new customers, as people are not coming back or that they are still looking for ways to strengthen the relationship with consumers.

Lasso Linear Regression with first order interactions was also performed but the performance was slightly worse than the Regular Lasso model. A table with the relevant coefficients is presented in Appendix B – *Table 2.*
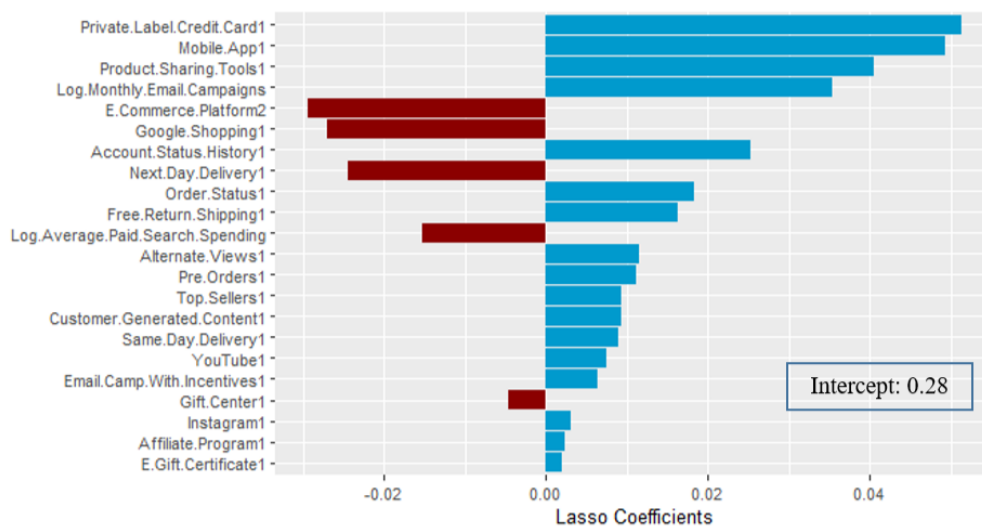

Figure 7: Lasso linear regression coefficients.

Table 1: Model Performance.

	MSE		R^2	
	Train	Test	Train	Test
Lasso Linear Regression	0.01325	0.01641	39.40%	19.47%
Boosted Regression Trees	0.00350	0.01620	89.63%	22.24%

5.2 Boosted Regression Tree

To perform the Boosted Regression tree, we tune the parameters using 10-fold cross validation. We choose a fixed learning rate of 0.01 and tune the number of trees and the max depths of the tree. We end up

performing boosting with 300 trees and a max depth of 6 nodes. As we can see in Table 1, the boosting algorithm clearly outperforms Lasso Linear Regression with a MSE of 0.0035 vs 0.01325 (and a much better R^2: 90%) in the training sample. Nevertheless, when we look at the performance in the test set (out of sample) both models show a quite poor and similar performance, with boosting performing just slightly better.

Taking a look at the variable importance (*Figure 8*), we can see that the model selects almost the same variables as Lasso Linear Regression. Boosting adds *Search Engine Marketing* as an important variable and *Email to a Friend*. If we analyze the partial dependency plots in Appendix B – *Figure 3*, *Search Engine Marketing* 2 and 3 appear to contribute positively to the prediction, meaning that is better to outsource this activity or hire an expert to help you than to perform it in-house. *Email to a friend* appears to have negative effect, which suggests that people might be annoyed by this marketing strategy. The algorithm also adds the *Response Time in Seconds*, but the effect seems counter-intuitive, as higher values contribute to a higher *Returning Shoppers' rate*.
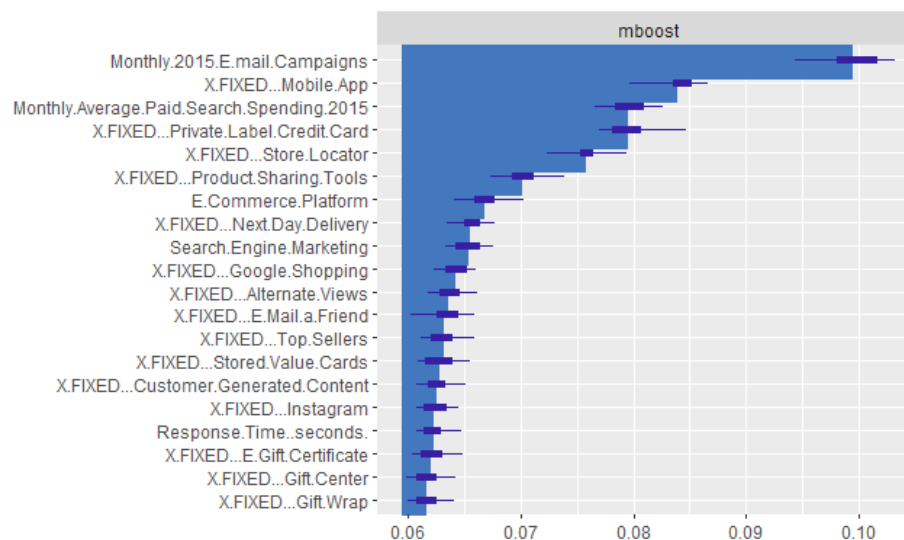

Figure 8: Variable Importance for Boosted Regression Tree – Based on Permutation Test.

6. Conclusion

6.1 Main findings

We performed Lasso linear regression and Boosting Regression Trees on internet marketing and website functionality data for 370 US E-retailers in order to predict the *Returning Shoppers' rate*. Our aim was to find the main drivers of customer e-loyalty using the first method and to optimize the prediction with the later one. Both methods yield similar conclusions, but given the limited amount of observations we have, we did not achieve a good predicting performance. As a matter of fact, Boosting appears to reduce significantly the bias of our model, but did not manage to effectively solve our variance problem. Nevertheless, we have to consider that measuring customer loyalty is a difficult task, as as we have seen in the Literature Review, there are many variables (mostly behavioral) that a company cannot control. Thus, even a model that explains a little amount of variance is useful for managers to understand how to influence it better.

Nevertheless, by using a different approach than previous studies, with information that is more accessible for companies, we managed to come to similar conclusions to the ones we discuss in the theoretical framework and to get some additional insights. Thus, coming back to our research question: How can internet marketing and website functionalities increase customer loyalty?

We conclude that internet customers like features related to "convenience", as most of the fast and easy delivery and returning options contribute to increasing the *Returning Shoppers' rate*. At the same time, internet shoppers also appear to like features that "personalize their experience" as variables like *Alternative Views* and *Customer Generated Content* have positive effects on our response variable.

Additionally, we find that customers value what we call "Easy to use features or functionalities". For example, having a *Mobile App* appears to be one of the main drivers of engaging consumers to shop again in your website. In the same vein, features like *Product Sharing Tools* and *Top Sellers section* and a *Store locator option* also contribute positively.

Moreover, marketing strategies like *Monthly Email Marketing Campaigns* and *Private Label Credit Cards* are also found to highly contribute to customer loyalty. Thus, we conclude that the internet consumer values "promotions" as both strategies are associated to this. Furthermore, both strategies are closely related to relationship marketing, confirming that building a relationship with the online customer is crucial to achieve success.

Furthermore, we also conclude that presence on social media is fundamental as strategies such as having an *Instagram page* or a *YouTube channel* are found to contribute positively, suggesting that customer probably frequent these platforms.

On the other hand, spending on paid search does not seem to be increasing the label of engagement of the customers. Moreover, being on other sale platforms, like google shopping, also discourages customers to buy on the company's website, as they probably buy directly through those channels.

6.2 Academic Implications

Coming back to our literature review, our results support that Relationship Marketing significantly contributes to building customer loyalty in an online environment as well, as we find that marketing strategies like email campaigns and private label credit cards greatly influence the *Returning Shoppers' rate*.

Additionally, our study confirms that features that Novak et.al (2000) and Cheng and Chang (2000) identify that contribute to customers having a satisfactory online experience, which are mostly associated to delivery and fulfilment, effectively contribute to increasing customer loyalty.

Moreover, it also supports Gommans et.al (2001) study, by finding that personalization and effective searching features increase the *Returning Shoppers' rate*.

However, we did not find elements to affirm that features that contribute to "building trust", like *Buy Online, Pick up at Store*, *Shipping Tracking* and *Shipping Cost Calculator* play a role in increasing customer loyalty, as Gommans et.al (2001), Bauer et al. (2002) and Verma et.al (2016) suggest. Furthermore, we did

not find evidence that a smaller *Response time in seconds* of the website is relevant to explain the *Returning Shoppers' rate,* as Geissler (2001) or Gommans et.al (2001) identify.

Taking all these into consideration, this study sets the precedents that analyzing website functionalities or marketing strategies and, thus, information that companies already have in their data bases, can give us valuable insights about what customers value and how to engage them. What is more, we have proven that by using the correct tools, machine learning methods can also give us very relevant insights. This is very important as data availability is increasing and analysts have to deal with extremely large datasets with multiple predictors that simpler statistical methods cannot handle. Thus, introducing this kind of methods to marketing literature can be extremely useful. As a matter of fact, even though companies are already using them to understand their customers it is still relatively new in the marketing academic research.

Thus, we believe that there is still much more potential to keep researching about this topic and applying machine learning methods to companies' data (instead of surveying customers). Our data set was relatively small and by gaining access to more data prediction could be significantly improved.

6.3 Managerial Implications

Considering all of the above mentioned, we suggest that companies keep investing in strategies that contribute to generate a closer relationship with the online consumer. Particularly, Email Marketing Campaigns and Loyalty Programs are proven to be effective.

Furthermore, investing in personalization and on features that make the online experience smoother appears to be fundamental as well. Developing a Mobile app is highly recommended.

Moreover, having efficient delivery options and focusing on order fulfillment is also fundamental if the e-retailer wants to increase the engagement of customers.

On the other hand, companies must evaluate whether the google shopping channel is an efficient alternative, or if they prefer to work on driving traffic directly to their webpage. Being on this type of platforms might be a good marketing strategy in the beginning and enables people to find your brand easily, but the company

must evaluate how profitable it is, or if it is more efficient to invest in other type of campaigns that help them build a close relationship with customers, like we mentioned above. Additionally, they might also want to review their payed search spending, as it does not appear to be effective.

6.4 Limitations of our study

Our study has various limitations. The most relevant one is the sample size, as after cleaning our data we end up with a sample of 370 and 67 variables. This is partially compensated if we consider that it is a representative sample, as we already know that the this are successful e-commerce companies. Nevertheless, this is a limitation regarding the algorithms we can apply. For example, more complex algorithms like Neural Networks cannot be performed in a sample this size, with this amount of predictors, as we have too many parameters to estimate. It will be interesting if in the future we could gather more updated information and combine the data sets, to see if we get a better out of sample performance.

Additionally, we have to consider that this are already well know companies and the results might be different if we consider startups for example, as people are just starting to get acquainted with them and might not trust them yet. Furthermore, even though most of this companies are multinational, the study is limited to the US market, and thus, our study mostly represents the preferences of the North American consumer.

Moreover, although they are all online retailers, they focus on different markets and products. Therefore, if we had enough information, conducting a study for specific industries might help us achieve a better prediction accuracy.

Finally, even though we were not able to achieve a high prediction performance, we were able to get very useful insights from the website's and internet marketing data. Additionally, we were able to implement a machine learning method, which has not been used yet for this kind of research. Improvements can be made in the future if we gather more data, but we can see that even though Boosting is a black box model, the

performance measures applied allow us to understand the marketing problem equally well than a more traditional method like Lasso Regression.

References:

Anderson, R., & Srinivasan, S. S. (2003). E-satisfaction and e-loyalty: A contingency framework. *Psychology and Marketing*, 20(2), 123–138.

Athanasopoulou, P. (2009). Relationship quality: A critical literature review and research agenda. *European Journal of Marketing*, 43(5), 583-610.

Bauer, H.H., Grether, M. and Leach, M. (2002). Building Customer Relations over the Internet. *Industrial Marketing Management*, Vol. 31, No. 2, pp. 155-163.

Bellizzi, A. J. (2000). Drawing Prospects to E-Commerce Websites. *Journal of Advertising Research* 41, 1/2: 43–53.

Berry, Leonard. L, (1995). Relationship Marketing of Services - Growing Interest, Emerging Perspectives. *Journal of the Academy of Marketing Science*, 23(4), pp. 236-245.

Boehmke, B., Greenwell, B. (2020). Hands-On Machine Learning with R. New York: Chapman and Hall/CRC.

Brown, George H. (1952). Brand loyalty – fact or fiction? *Advertising Age*,23 (June 9), 53–55.

Chen, S. & Chang, T. (2003). A descriptive model of online shopping process: some empirical results. *International Journal of Service Industry Management*, Vol. 14 No. 5, pp. 556-569.

Cunningham, S. M. (1966). Brand Loyalty--What, Where, How Much? *Harvard Business Review* 34 (January-February): 116-128.

Danaher, P. J., Wilson, I. W., & Davis, R. A. (2003). A comparison of online and offline consumer brand loyalty. *Marketing Science*, 22(4), 461-476.

Evans, J.R. & Laskin, R.L. (1994). The Relationship Marketing Process: A Conceptualization and Application". *Industrial Marketing Management*, 23, pp. 439-452.

Farley, J. (1964). Brand Loyalty and the Economics of Information. *The Journal of Business*, 37(4), 370-381.

Friedman, J. H., (2002). Stochastic Gradient Boosting. Computational Statistics and Data Analysis, 38 (4), 367–378.

Friedman, J.H., (1999). Greedy function approximation: a gradient boosting machine. Technical Report,Department of Statistics, Stanford University.

Geissler, G.L. (2001). Building customer relationships online: The Web site designers' perspective. *Journal of Consumer Marketing*, Vol.18No.6, pp.488-502.

Gerón A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly

Gommans, M., Krishnan, K.S., & Scheffold, K.B. (2001). From brand loyalty to e-loyalty: A conceptual framework. *Journal of Economic and Social Research*, 3(1), 43–58.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer.

Hothorn, T., Hornik, K.t & Zeileis A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*,15:3, 651-674.

Ilfeld, J.S. and Winer, R.S. (2002). Generating Website Traffic. *Journal of Advertising Research*, Vol. 42, No. 5, pp. 49-61

Jackson, B. (1985). Winning and Keeping Industrial Customers. Lexington, KY: Lexington Books.

Jacoby, J., & Kyner, D. (1973). Brand Loyalty vs. Repeat Purchasing Behavior. *Journal of Marketing Research*, 10(1), 1-9.

Jacoby J. and Chestnut R. W. (1978). Brand Loyalty Measurement and Management. New York: Wiley.

James G., Witten D., Hastie T., and Tibshirani R (2013). An Introduction to Statistical Learning. Springer-Verlag, New York.

Keller, K. (1993). Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *Journal of Marketing*, 57(1), 1-22.

Kozlenkova, I. V., Palmatier, R. W., Fang, E., Xiao, B., & Huang, M. (2017). Online relationship formation. *Journal of Marketing*, 81(3), 21–40.

KPMG (2017). The truth about online consumers: 2017 global online consumer report. Retrieved 26 March 2018 from https://assets.kpmg/content/dam/kpmg/xx/pdf/2017/01/the-truth-about-online-consumers.pdf

Molnar, Christoph (2019). Interpretable machine learning. A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 53(3), 20–38.

Novak, T.P., Hoffman, D.L., & Yung, Y.-F. (2000). Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. *Marketing Science*, 19(Winter, 1), 22–42.

Oliver, R. (1999). Whence Consumer Loyalty? *Journal of Marketing*, 63, 33-44.

Rafiq, M., Fulford, H., & Lu, X. (2013). Building Customer Loyalty in on Online Retailing: The Role of Relationship Quality. Journal of Marketing *Management*, 29, 494-517.

Reichheld, F. & Schefter, P. (2000). E-Loyalty: Your Secret Weapon on the Web. *Harvard Business Review*, 78(4): 105-114.

Reinartz, W., Kraft, M. and Hoyer, W.D. (2004), The Customer Relationship Management Process: Its Measurement and Impact on Performance, *Journal of Marketing Research*, Vol.41, No. 3, pp. 293-305.

Rigby, D. (2011). "The Future of Shopping." *Harvard Business Review* 89 (12): 64–75.

Schultz D. E., Bailey S. E (2000). Customer/Brand Loyalty in an Interactive Marketplace. *Journal of Advertising Research*, 40 (3), 41-52.

Smith, E. R. (2000) E-Loyalty: How to keep customers coming back to your website. New York: Harper Collins.

Srinivasan, S.S., Anderson, R., & Ponnavolu, K. (2002). Customer loyalty in e-commerce: An exploration of its antecedents and consequences. *Journal of Retailing*, 78(1), 41–50.

Steinhoff, L., Arli, D. Weaven, S. and Kozlenkova, I. (2019). Online relationship marketing. *Journal of the Academy of Marketing Science*, 47, issue 3, p. 369-393

Verma, V., Sharma, D., & Seth, J. (2016). Does relationship marketing matter in online retailing? A meta-analytic approach. *Journal of the Academy of Marketing Science*, 44(2), 206–217.

Wolk A. and Theysohn S. (2007) Factors influencing website traffic in the paid content market, *Journal of Marketing Management*, 23:7-8, 769-796.

Yadav, M. S., & Pavlou, P. A. (2014). Marketing in computer-mediated environments: research synthesis and new directions. *Journal of Marketing*, 78(1), 20–40.

Zineldin, M. (2000). Beyond relationship marketing: Technologicalship marketing. *Journal of Marketing Intelligence & Planning*, /S(l), 9-23.

Appendixes

Appendix A

Table 1: Variable description and main statistics.

| Variable	Description	Mean	SD	Median	Min	Max	% Category = 1
Return Shoppers 2015	**% of returning shoppers out of the total shoppers in 2015**	**0.38**	**0.15**	**0.37**	**0.06**	**0.88**	-
Monthly 2015 E-mail Campaigns	Total number of Email campaigns conducted in 2015.	17.43	14.76	14	0	69	-
Annual Average Paid Search Spending 2015	Average paid search spending in 2015 as a porcentage of sales	0.02	0.03	0.01	0	0.43	-
Response Time (seconds)	Average website page load time in seconds	4.36	1.97	4.1	0.53	13.04	-
Site Availability	% of times the website is available.	1	0.01	1	0.9	1	-
Account Status History	Binary variable indicating whether the website has Account Satus/History: 0=No 1=Yes.	-	-	-	-	-	90%
Affiliate Program	Binary variable indicating whether the E-commerce shop has an Affiliate Program: 0=No 1=Yes.	-	-	-	-	-	67%
Alternate Views	Binary variable indicating whether the website has alternative views: 0=No 1=Yes.	-	-	-	-	-	78%
American Express	Binary variable indicating whether the E-commerce shop accepts American Express Credit Card: 0=No 1=Yes.	-	-	-	-	-	96%
Bill Me Later	Binary variable indicating whether the website has the option to bill me later: 0=No 1=Yes.	-	-	-	-	-	24%
Blog	Binary variable indicating whether the E-commerce shop has a blog: 0=No 1=Yes.	-	-	-	-	-	66%
Buy Online Pickup In Store	Binary variable indicating whether the E-commerce shop has the option to buy online and pick up at store: 0=No 1=Yes.	-	-	-	-	-	20%
Color Change	Binary variable indicating whether the website changes its color: 0=No 1=Yes.	-	-	-	-	-	31%
Customer Generated Content	Binary variable indicating whether the website has Customer Generated content: 0=No 1=Yes.	-	-	-	-	-	21%
Deferred Payment	Binary variable indicating whether the website has Deferred Payment: 0=No 1=Yes.	-	-	-	-	-	16%
E Gift Certificate	Categorical variable indicating who is in charge of the E-Commerce Platform: 1= In-house, 2= Mixed , 3=Outsourced.	-	-	-	-	-	42%
E Mail a Friend	Binary variable indicating whether the website has the option to by a E-Gift certificate: 0=No 1=Yes.	-	-	-	-	-	71%
E Mail Pop Up Box	Binary variable indicating whether the website has the option to by a E-Gift certificate: 0=No 1=Yes.	-	-	-	-	-	39%
E-Commerce Platform	Binary variable indicating whether the website has an E-mail pop up box: 0=No 1=Yes.	-	-	-	-	-	38%
Estimated Delivery Date	Binary variable indicating whether the website shows you the estimated delivery date: 0=No 1=Yes.	-	-	-	-	-	43%
Facebook	Binary variable indicating whether the E-commerce shop has a facebook: 0=No 1=Yes.	-	-	-	-	-	99%
Free Return Shipping	Binary variable indicating whether the E-commerce shop offers free return shiping: 0=No 1=Yes.	-	-	-	-	-	10%
Free Shipping	Binary variable indicating whether the E-commerce shop offers free shiping: 0=No 1=Yes.	-	-	-	-	-	65%
Frequently Asked Questions	Binary variable indicating whether the website has a frequently asked questions section: 0=No 1=Yes.	-	-	-	-	-	91%
Gift Center	Binary variable indicating whether the website has a gift center: 0=No 1=Yes.	-	-	-	-	-	29%
Gift Message	Binary variable indicating whether the website offers the option to put a gift message: 0=No 1=Yes.	-	-	-	-	-	33%
Gift Wrap	Binary variable indicating whether the website offers the option to wrap gift: 0=No 1=Yes.	-	-	-	-	-	28%
Google Shopping	Binary variable indicating whether the E-commerce shop is also in google shopping: 0=No 1=Yes.	-	-	-	-	-	76%
Google Wallet	Binary variable indicating whether you can pay with google wallet: 0=No 1=Yes.	-	-	-	-	-	16%
Guided Navigation	Binary variable indicating whether the website has guided navigation: 0=No 1=Yes.	-	-	-	-	-	75%
In Store Returns	Binary variable indicating whether the E-commerce shop has an instagram: 0=No 1=Yes.	-	-	-	-	-	33%
Instagram	Binary variable indicating whether the E-commerce shop offers to option to return your purchase in store: 0=No 1=Yes.	-	-	-	-	-	84%
Interactive Catalog	Binary variable indicating whether the website has an interactive catalog: 0=No 1=Yes.	-	-	-	-	-	36%
International Shipping	Binary variable indicating whether the website offers international shipping: 0=No 1=Yes.	-	-	-	-	-	60%
Live Chat	Binary variable indicating whether the website has a live chat: 0=No 1=Yes.	-	-	-	-	-	36%
Mobile App	Binary variable indicating whether the E-commerce shop also has a mobile app: 0=No 1=Yes.	-	-	-	-	-	52%
Mobile Optimized Site	Binary variable indicating whether the website is mobile optimized: 0=No 1=Yes.	-	-	-	-	-	83%
Monthly E Mail Campaigns With Incentives	Binary variable indicating whether the E-commerce shop implements monthly email campaigns with incentives: 0=No 1=Yes.	-	-	-	-	-	22%
Multiple Languages	Binary variable indicating whether the website has the option to change the language: 0=No 1=Yes.	-	-	-	-	-	10%
Next Day Delivery	Binary variable indicating whether the E-commerce shop offers next day delivery: 0=No 1=Yes.	-	-	-	-	-	72%
Order Status	Binary variable indicating whether the website shows your order status: 0=No 1=Yes.	-	-	-	-	-	52%
Outlet Center	Binary variable indicating whether the E-commerce shop has an outlet center: 0=No 1=Yes.	-	-	-	-	-	49%
PayPal	Binary variable indicating whether the website offers you the option to pay with Paypal: 0=No 1=Yes.	-	-	-	-	-	50%
PayPal Express Checkout	Binary variable indicating whether the website has Paypal express check out: 0=No 1=Yes.	-	-	-	-	-	24%
Pinterest	Binary variable indicating whether the E-commerce shop is on Pinterest: 0=No 1=Yes.	-	-	-	-	-	87%
Pre Orders	Binary variable indicating whether the website has the option to pre-order: 0=No 1=Yes.	-	-	-	-	-	32%
Private Label Credit Card	Binary variable indicating whether the E-commerce shop has a private label card (usually associated to a loyalty program): 0=No 1=Yes.	-	-	-	-	-	24%
Product Comparisons	Binary variable indicating whether the website gives you the option to compare products: 0=No 1=Yes.	-	-	-	-	-	31%
Product Customization	Binary variable indicating whether the website allows you to customize the product: 0=No 1=Yes.	-	-	-	-	-	35%
Product Recommendations	Binary variable indicating whether the website recomends you products: 0=No 1=Yes.	-	-	-	-	-	76%
Product Sharing Tools	Binary variable indicating whether the website lets you share products: 0=No 1=Yes.	-	-	-	-	-	79%
Quick View	Binary variable indicating whether the website has a quick view option: 0=No 1=Yes.	-	-	-	-	-	32%
Real Time Inventory	Binary variable indicating whether the E-commerce shop keeps real time inventory: 0=No 1=Yes.	-	-	-	-	-	59%
Same Day Delivery	Binary variable indicating whether the E-commerce shop offers same day delivery: 0=No 1=Yes.	-	-	-	-	-	6%
Search Engine Marketing	Categorical variable indicating who is in charge of the Search Engine Marketing: 1= In-house, 2= Mixed , 3=Outsourced.	-	-	-	-	-	30%
Ship from Store	Binary variable indicating whether the E-commerce shop offers ship from store: 0=No 1=Yes.	-	-	-	-	-	7%
Shipment Tracking	Binary variable indicating whether the website allows you to track your shipment: 0=No 1=Yes.	-	-	-	-	-	72%
Shipping Cost Calculator	Binary variable indicating whether the website has a shipping cost calculator: 0=No 1=Yes.	-	-	-	-	-	27%
Shopping Cart Abandonment E Mail Program	Binary variable indicating whether the website sends an automatic email when you abandon your cart: 0=No 1=Yes.	-	-	-	-	-	35%
Store Locator	Binary variable indicating whether the website has a store locator: 0=No 1=Yes.	-	-	-	-	-	56%
Stored Value Cards	Binary variable indicating whether the website allows you to store value cards: 0=No 1=Yes.	-	-	-	-	-	35%
Top Sellers	Binary variable indicating whether the website shows you the top sellers: 0=No 1=Yes.	-	-	-	-	-	65%
Trending Products	Binary variable indicating whether the website has a trending products section: 0=No 1=Yes.	-	-	-	-	-	10%
What s New	Binary variable indicating whether the website has a what's new section: 0=No 1=Yes.	-	-	-	-	-	72%
Wish List	Binary variable indicating whether the website lets you build a wish list: 0=No 1=Yes.	-	-	-	-	-	58%
YouTube	Binary variable indicating whether the E-commerce shop uses Youtube: 0=No 1=Yes.	-	-	-	-	-	94%
Zoom	Binary variable indicating whether the E-commerce shop uses Zoom: 0=No 1=Yes.	-	-	-	-	-	60%
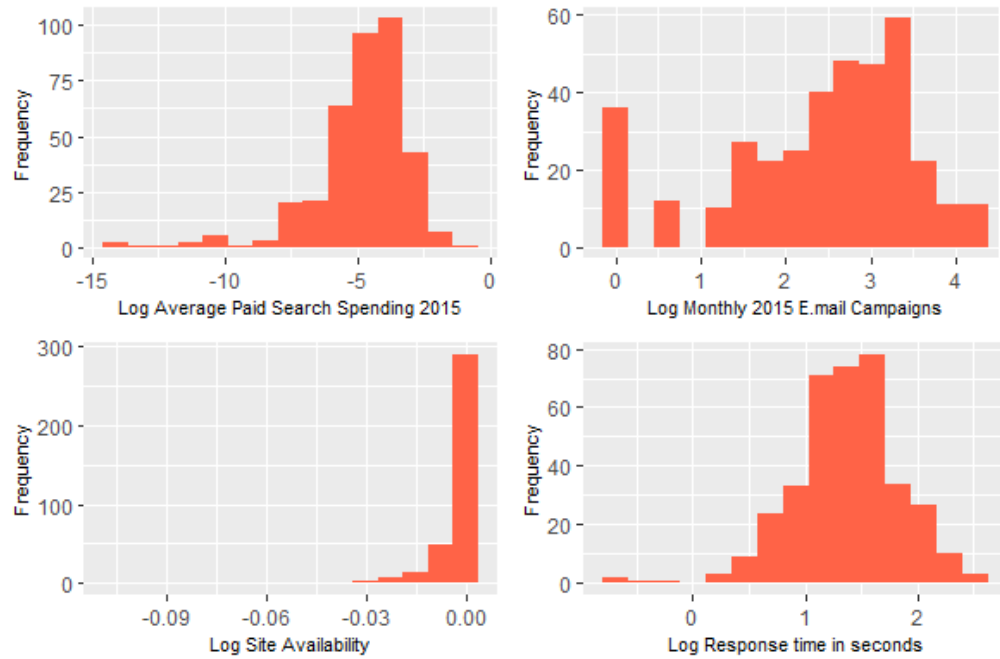
Figure 1: Histograms depicting the distribution of the continuous variables.
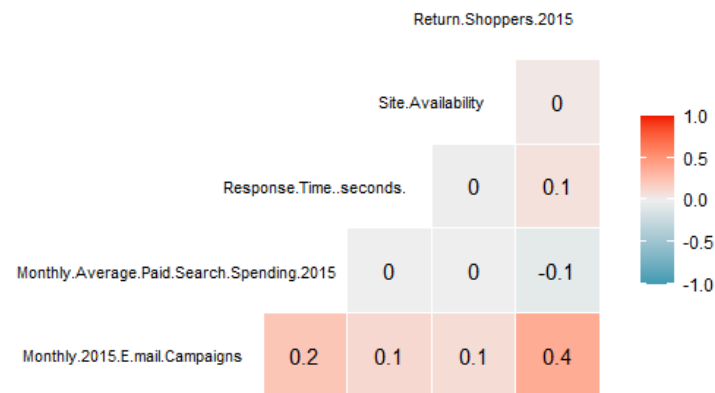

Figure 2: Correlation plot depicting the linear association between the continuous variables.
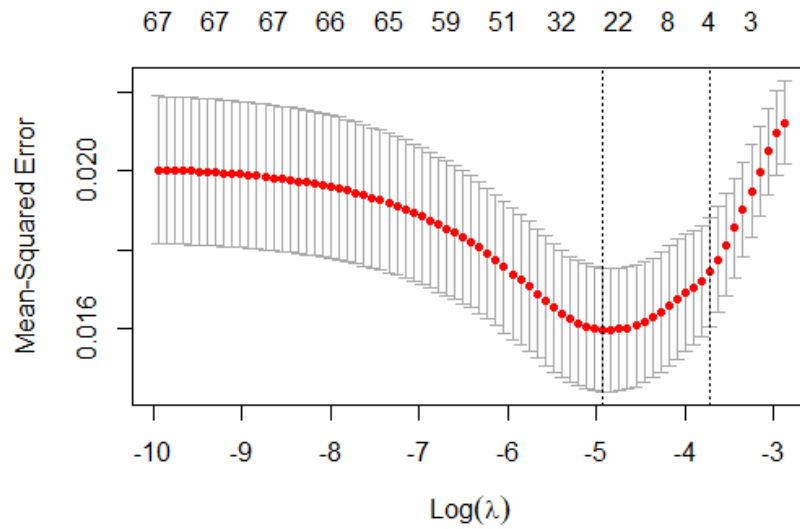
Appendix B

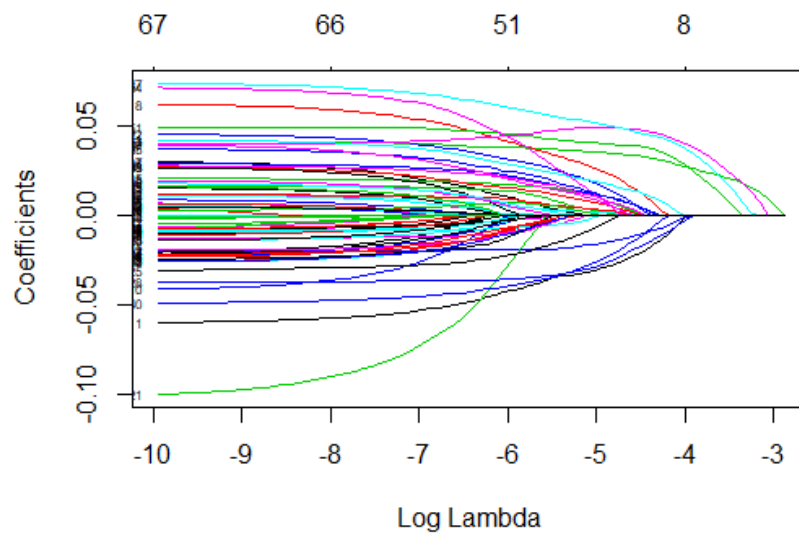*Figure 1: Cross-Validated Mean Square Error for different levels of lambda in Lasso Linear Regression*

Figure 2: Coefficient profile plot for different values of lambda in Lasso Linear Regression.

Table 1: Lasso Linear Regression with interactions.

| Variables	Coefficients
(Intercept)	0.310
Same.Day.Delivery1:Customer.Generated.Content1	0.154
Private.Label.Credit.Card1:Top.Sellers1	0.073
Private.Label.Credit.Card1:Same.Day.Delivery1	0.052
Product.Sharing.Tools1:YouTube1	0.043
Product.Sharing.Tools1:Monthly.Average.Paid.Search.Spending.2015	-0.025
Product.Sharing.Tools1:E.Commerce.Platform2	-0.024
Mobile.App1:Free.Return.Shipping1	0.024
Mobile.App1:YouTube1	0.023
Monthly.2015.E.mail.Campaigns:E.Commerce.Platform3	0.020
Monthly.2015.E.mail.Campaigns	0.020
Account.Status...History1:Alternate.Views1	0.019
Private.Label.Credit.Card1:Affiliate.Program1	0.019
Mobile.App1:Account.Status...History1	0.019
Order.Status1:Pre.Orders1	0.017
Private.Label.Credit.Card1:E.Commerce.Platform2	-0.017
Order.Status1:Free.Return.Shipping1	0.017
YouTube1:Monthly.E.Mail.Campaigns.With.Incentives1	0.015
Google.Shopping1	-0.015
Free.Return.Shipping1:Monthly.Average.Paid.Search.Spending.2015	-0.014
Next.Day.Delivery1:Gift.Center1	-0.012
Next.Day.Delivery1:YouTube1	-0.009
Monthly.Average.Paid.Search.Spending.2015:E.Gift.Certificate1	-0.008
Order.Status1:YouTube1	0.008
Monthly.2015.E.mail.Campaigns:Alternate.Views1	0.008
Mobile.App1:Affiliate.Program1	0.007
Mobile.App1:Monthly.Average.Paid.Search.Spending.2015	-0.005
Product.Sharing.Tools1:Alternate.Views1	0.005
Monthly.E.Mail.Campaigns.With.Incentives1:Affiliate.Program1	0.004
Mobile.App1:Customer.Generated.Content1	0.004
Google.Shopping1:YouTube1	-0.004
Monthly.Average.Paid.Search.Spending.2015:Same.Day.Delivery1	-0.004
Order.Status1:Monthly.Average.Paid.Search.Spending.2015	-0.004
Private.Label.Credit.Card1:Order.Status1	0.003
Product.Sharing.Tools1:Affiliate.Program1	0.002
Order.Status1:Top.Sellers1	0.002
Monthly.2015.E.mail.Campaigns:Affiliate.Program1	0.002
Private.Label.Credit.Card1:Monthly.2015.E.mail.Campaigns	-0.002
Product.Sharing.Tools1:Gift.Center1	-0.002
YouTube1:Instagram1	0.002
Alternate.Views1:Top.Sellers1	0.001
Order.Status1:Same.Day.Delivery1	0.001
Alternate.Views1:Affiliate.Program1	0.001

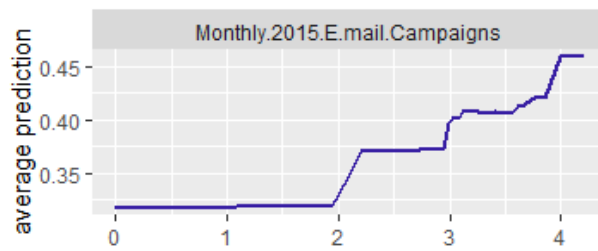## Log Monthly E.mail Campaigns
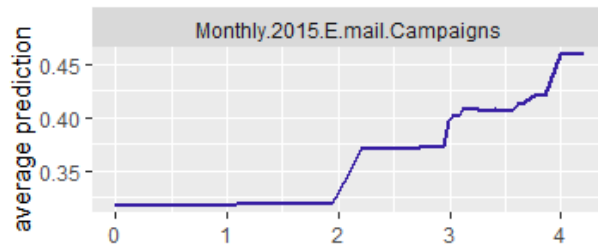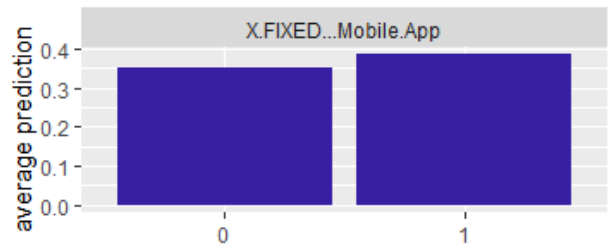
Mobile App

Log Average Paid Search Spending
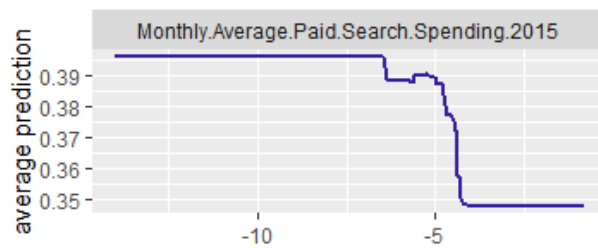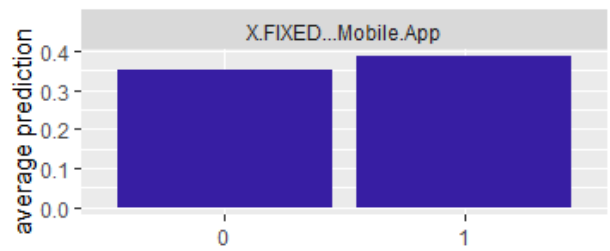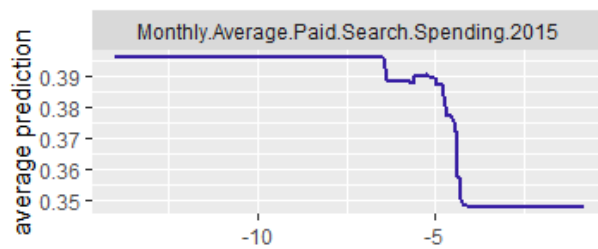

Private Label Credit Card
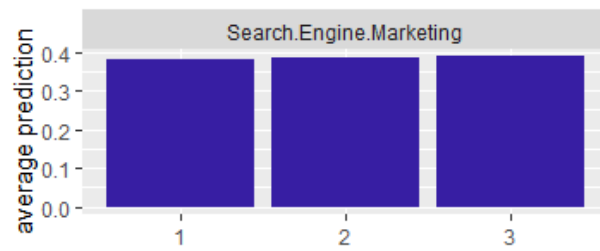

Log Monthly E.mail Campaigns
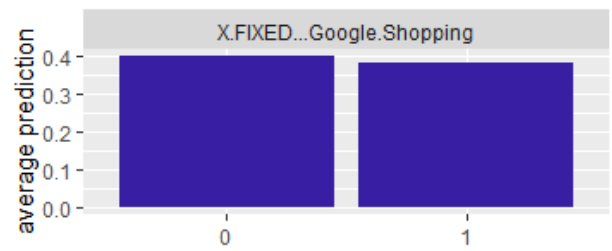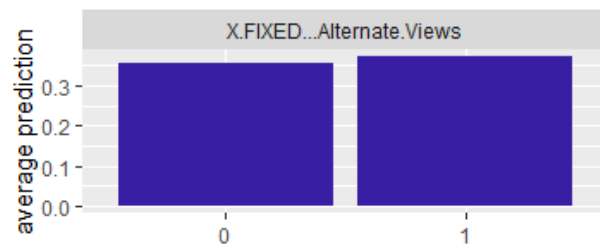

Mobile App

Log Average Paid Search Spending
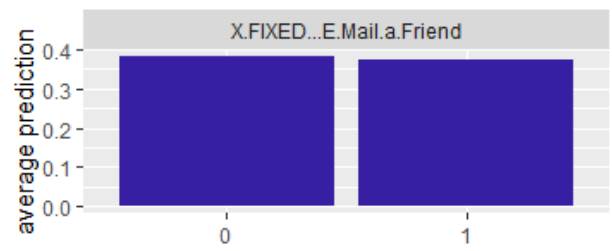

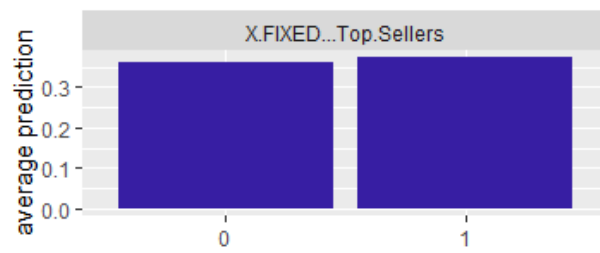## Private Label Credit Card

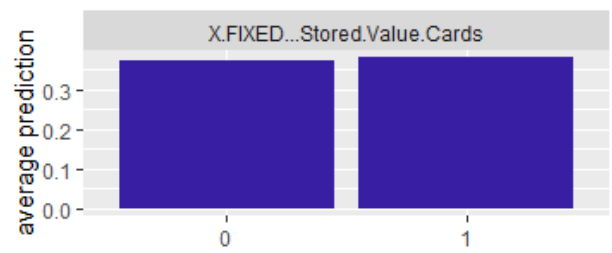## Search Engine Marketing

Google Shopping

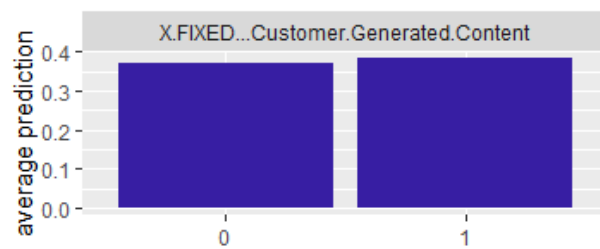## Alternate Views

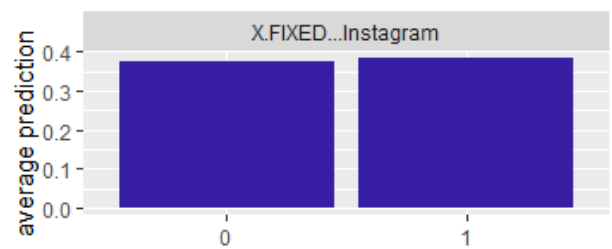## Email a friend

Top Sellers

Stored.Value.Cards

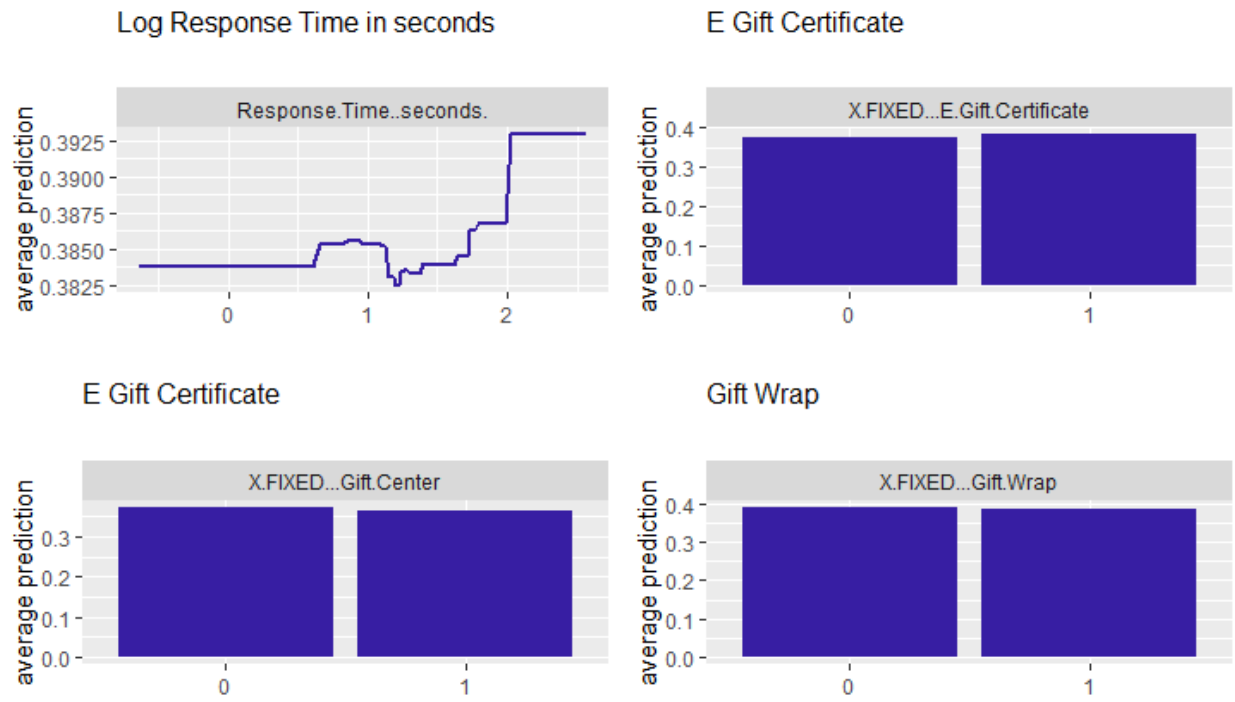## Customer Generated Content

Instagram

Figure 3: Partial Dependency Plots for the main predictors in Boosted Regression Trees.