Erasmus School of Economics

ERASMUS SCHOOL OF ECONOMICS

Master Thesis - Data Science and Marketing Analytics Program

A comparative study on machine learning based marketing mix modelling: Comparing the performance and applications of regression methods and Random Forest

Name student	Alireza Mohaghegh
Student ID	538063
Date of final version	16.01.2022
Supervisor	Dr. Radek Karpienko
Second assessor	Dr. Kathrin Gruber

Abstract:

Marketing mix modelling refers to estimating the relationship of marketing efforts behind a product or service, to a dependant variable such as sales or market share. The main objective of this study is to compare the performance of machine learning techniques (namely Random-Forest) and simple linear regression (OLS) in marketing mix modelling applications.

The initial model developed is OLS linear regression, which creates a baseline for both prediction and interpretation for more complex models. Subsequently, Random Forest model is deployed and more complex nonlinear relationships with the response variable is captured. Also, this model has acquired higher prediction performance however, the improved accuracy in such models has a constant trade-off with interpretability which is addressed in the study.

This research will bring insight to organizations which intend to invest in new Machine Learning techniques in their marketing efforts and will compare the outcome of these models versus traditional statistical tools. The data utilized in the research is provide by FrieslandCampina, N.V. (A Dutch multinational dairy corporation which is based in Amersfoort) and consists of real-world market sales data.

Table of Contents

1. Introduction 7

2. Literature Review 10

2.1. What is Marketing Mix Modelling? 10

2.2. Modeling principles of MMM 11

2.3. History and Implementation of Marketing Mix 12

2.4. Challenges with MMM 12

2.5. Machine Learning Methods 15

2.6. Advantages Machine Learning 16

2.7. Disadvantages Machine Learning 18

3. Methodology: 19

3.1. Conceptual Modelling 19

3.2. MMM using linear models 20

3.3. Linear regression model and Least Squares 20

3.4. Assessing the accuracy of the model and $R2$ Statistic: 22

3.5. Prediction accuracy and overfitting in linear regression: 23

3.6. Machine learning methods for regression 23

3.7. Regression Trees 24

3.8. Bagging 25

3.9. Random forest 25

3.10. Partial Dependence Plot (PDP) 26

4. Data: 27

4.1. Description of the Dataset 27

4.2. Variable description **28**

4.3. Descriptive Statistics **29**

5. Analysis ***35***

5.1. Data Preparation: **35**

5.2. Marketing mix modelling using Linear Regression Model (OLS): **36**

5.3. Feature contribution (importance) in linear regression model **39**

5.4. Marketing mix modelling at country level: **40**

5.5. Partial Dependence analysis for linear regression **44**

5.6. Linear model performance on brand level **45**

5.7. Marketing mix modelling using Random Forest **47**

5.8. Feature importance in Random Forest **49**

5.9. Partial dependence analysis **50**

5.1. Random forest model performance on brand level **51**

6. Conclusion: ***53***

6.1. Linear regression: **53**

6.2. Random forest **55**

7. Bibliography ***57***

8. Appendix ***60***

8.1. Response Curves in marketing mix modelling **60**

8.2. Global segregate model: **62**

Table of Figures:

Figure 1- Graphical representation of OLS line in a three-dimensional setting with two predictors and one response. Under this setting the regression line becomes a plane. 22

Table 1- Variable name and description of different features included in the model: 28

Figure 2- Total sales volume of different brands per week in three-year time span of the data collection 30

Figure 3- Average market price for brand “NOYNOY” over years 31

Figure 4- Average sales volume for brand “NOYNOY” over years 32

Figure 5- Offline media spent of brand "NOYNOY" over years 33

Figure 6- Online media spent of brand "NOYNOY" over years 33

Figure 7- Weighted distribution (product availability) for brand "NOYNOY" 34

Figure 8- Correlation plot of numerical features in the dataset 35

Table 2- OLS Regression coefficient results for marketing mix modelling 37

Figure 10- Prediction versus Actual results (sales volume) for regression analysis 38

Figure 11- OLS Regression model's contribution plot. This plot is graphical representation for contribution of each feature in the model. Also, predicted versus actual results are demonstrated with brown and purple lines respectively. 40

Table 3- Measures of fit for country level marketing mix modelling using linear regression model 41

Table 4- predictive performance of the model on test data set 42

Table 5- OLS Regression coefficient results for marketing mix modelling 43

Figure 12- partial dependence plot of average price. This plot is a representation of marginal effect of price on models predictionsTable 5- OLS Regression coefficient results for marketing mix modelling 43

Table 6- OLS Regression coefficient results for marketing mix modelling 43

Figure 13- partial dependence plot of offline media investment. This plot is a representation of marginal effect of offline marketing investment on model’s predictions 45

Figure 14- Prediction versus Actual results (sales volume) for linear regression model analysis for brand “MILNER” 46

Figure 15- Prediction versus Actual results (sales volume) for linear regression model analysis for brand "NOYNOY" 46
Table 7- Predictive performance of OLS regression model at brand level 47
Table 8- Hyperparameter tunning result for Random Forest model: 48
Table 9- Predictive performance of Random Forest model: 48
Figure 16- Feature importance plot for Random Forest model 49
Figure 17- partial dependence plot of weighted distribution. This plot is a representation of marginal effect of product availability(wd) on model's predictions 50
Figure 18- partial dependence plot of product price. This plot is a representation of marginal effect of selling price on model's predictions 51
Table 10- Predictive performance of Random Forest at brand level on the test data set 52
Figure 19- Online and Offline advertising response curves, an output of marketing mix modeling 61
Figure 20- Representation of global segregate model 62

1. Introduction

Marketing mix modeling, which refers to "a statistical approach where quantified marketing activities over time are mathematically linked to a dependent variable, such as sales or revenues" (Wolfe Sr & Crotts, 2011). This tool is often deployed in the selection of marketing concepts and allows companies to statically model and evaluate the outcomes of their marketing efforts in different channels. In MMM models, most commonly desired response variable is sales of a specific product or brand (Tellis G. J., 2006).

Powered by increasing computational abilities and ever-growing data accessibility marketing mix decisions are increasingly rotted in quantitative rather than qualitative reasoning. Specifically, new models and methods are used to approach the marketing mix concepts, such as Machine Learning (ML) which has and will continue to dramatically transform marketing and marketing mix modelling (Huang & Rust, 2021).

Machine learning (ML) methods have deeply penetrated quantitative rooted marketing mix modelling as it displays a multitude of advantages over common statistical methods (Kumar, Shankar, & Aljohani, 2020; Kumar, Shankar, & Aljohani, 2020). The fact that, machine learning models can handle unstructured and hybrid data, great volumes of data and allows of model flexibility and increased prediction accuracy. One the other hand, one of the main disadvantages of ML are that they often lack interpretability (Rai, 2020).

Organisations aim to optimize their marketing mix spending by investing into concepts which display closest predicted results, leading to raised organisational interested in both informative and accurate models. Considering the pros and cons of the new ML techniques available for marketing mix modelling, raises the question that how organisations should balance machine learning models versus conventional statistical techniques?

The trade-off between interpretability and accuracy is a constant topic in data science department of organisations. Although literature on both ML techniques and MMM is abundant, yet there is more to be found out on how effective and profitable it is to invest in data driven marketing practices. This brings me to form the following primary research questions:

$RQ_{1.1}$: Investigation of the utilisation cost of machine learning methods, specifically in marketing mix modelling for FMCG business.

$RQ_{1.2}$: How beneficial would it be for the organisations to transform their quantitative Marketing research to a Machine Learning based structure? In other words, this research will look into how these new techniques will benefit companies and what are the new insights that can be extracted upon deployment of machine learning techniques.

Another important criterion in marketing departments of companies is how to measure and quantify the effectiveness of marketing activities. Marketing managers seamlessly are seeking quantifiable proof on how their decisions are contributing to companies' success. Considering high levels of investments that companies put into their marketing activities extends the need to understand the effectiveness and ROI of these spends. Marketers are mainly seeking answers to a) how to quantify the impact of specific marketing plan and b) how does current Brand Marketing Plan (BMP) will impact future sales levels (Pandey, 2021)

In MMM practices it is common to develop a model on aggregated sales data for specific brand. Considering the nature of this data it is usual to have weeks as unit of measurement which overtime (years) will not lead to a significant high number of data points. Marketeers, trying to use as many historical data point possible, often build models on aggregated multi-brand data. Considering that every brand has its own mix elements it is critical to test and compare the outcome of these models for every. Moreover, at the end of the day every company and business will try to use these models for predictions on own products rather that whole market. Therefore, it is important to compare prediction performance of different MMM methods at brand level to see if and how new machine learning methods will help companies in their future predictions. This research will aim to address the secondary research question which is formulated below:

$RQ_{2:}$ How Different are brand level insights driven from conventional marketing mix modelling methods, namely Linear Regression, in comparison to Machine Learning methods specifically, Random Forest?

Another crux that organisations are facing is to understand whether it is more beneficial of have a better prediction of future sales? Or to have a clearer interpretations of how current marketing campaigns are affecting sales levels. This matter, in the world of machine learning and statistical methods, is addressed by the interpretability vs accuracy trade-off. This brings me to form the following additional research questions:

RQ_3: How to address the trade-off between interpretability and accuracy in machine learning models, specifically in Marketing Mix Modelling?

The objective of this research is to investigate and compare different methods in modelling marketing mix and examine the effects and outcomes of these models. Specifically, this study will compare a traditional approach to marketing mix modelling, namely Linear Regression, and Random Forest as a relatively new approach to this problem. This study intends, to determine which model is preferable in the context of this research, with regards to the quality of fit, the uncertainty of the model and its interpretation. The following study is done in collaboration with FrieslandCampina, N.V. is a Dutch multinational dairy cooperative which is based in Amersfoort.

2. Literature Review

In the following, the relevant streams of existing literature regarding Marketing-mix-modelling (MMM), Machine learning (ML), and overlap of these streams will be discussed.

2.1. What is Marketing Mix Modelling?

The marketing mix consists of analysis of variables that can be manipulated by marketers and can be controlled to influence a brand's KPIs like sales or market share (Tellis G. J., 2006). Traditionally these variables, "4Ps" of marketing, is consisted of four main parts Product, Price, Place and Promotion (Kumar, Shankar, & Aljohani, 2020). ***Product*** incapsulate the company's product variables such as portfolio, differentiation from competitors, quality, etc. ***Price*** refers to the products' price and monetary discounts such as bulk shopping or deals. ***Place*** includes the information of distribution of product, availability, shelf position, etc. And lastly, ***Promotion*** refers to marketing advertising, feature sales promotions and advertising displays (Tellis G. J., 2006).

MMM can be extended so that it encapsulates not only the direct effect of marketing efforts but also, it can include the indirect marketing effects on the sales of the researched brands. Mix variables are chosen such that they are controllable and employed to influence sales or market position. Marketing mix modeling therefore is the act of tuning the investments in these four variables to optimize the manager's goal (Kumar, Shankar, & Aljohani, 2020).

Marketing Mix Modelling is employed as a decision-making tool in companies of varying sizes. The main goal of MMM is to gain insights about how marketing activities impact business metrics, such as sales. Furthermore, MMM measures the quantitative effect of ROI of various marketing actions (Kumar, Shankar, & Aljohani, 2020). MMM is necessary to determine the importance of marketing channels based on the relationship between sales and investment. Therefore, employing MMM helps the managers to optimize investments into the various marketing channels and additionally allows them to gain knowledge of the relationships of the channels to the sales. Especially as the marketing mix concerns variables which are controllable by the marketing manager and which impact brands sales and or market share (Tellis, 2006).

2.2. Modeling principles of MMM

Over the course of past decades, researchers and marketers have been constantly trying to find the ultimate solution to identify and measure the efficacy, effectiveness, and sensitivity of every unit of investment, spent on the business specific KPIs. In this regard, a variety of statistical, market-variable response to marketing mix models have been developed. Amongst these models most of them are focused on market response to advertising and pricing (Sethuraman, 1991). A plausible reason for this is that optimizing expenses related to changes in levels of aforementioned factors are vital for businesses and mainly these factors are indicative of success for business teams and mainly marketeers (Pandey, 2021).

Another underlying hypothesis in this approach is that past data on consumers and market-response variable contains valuable information in such way that one wants to gather as many as past data point as possible. Also, using historical data one might be able to predict how consumers and market will respond in future and this will enable better planning of marketing variables (Tellis G. J., 1995; Pandey, 2021).

2.3. History and Implementation of Marketing Mix

In prior literature it is mentioned that, first-using the term "marketing mix" is claimed by Borden (1965) and it was a description given by Culliton (1948) which suggested this to him . Culliton describes a business executive as "mixer of ingredients". An executive is "a mixer of ingredients, who sometimes follows a recipe as he goes along, sometimes adapts a recipe to the ingredients immediately available, and sometimes experiments with or invents ingredients no one else has tried" (Culliton, 1948; Goi, 2009).

The set of variables introduced by Borden included 12 different elements: pricing, channels of distribution, promotions, advertising, packaging, servicing, branding, personal selling, physical handling; and fact finding and analysis (Borden, 1965). This idea was further developed by McCarthy (1964) and the notion of marketing mix was defined as group of factors which are in control of marketeers to satisfy target. In McCarthy's version of marketing mix, the 12 aforementioned elements were grouped to 4Ps which include: product, price, promotion and place.

Over the years several additional "Ps" are introduced by researchers to be added into this concept. People, personnel, participants, physical evidence and process and process management, and physical facilities are a few of the additions that were made to original marketing mix framework (Goi, 2009).

2.4. Challenges with MMM

Majority of preliminary marketing literature suggests that all elements of 4P framework of the marketing mix are equally important and lagging behind in anyone those will translate into failure (Kellerman, 1995). However, a review of recent literature demonstrates that in business context the executives do not view the initial elements of 4Ps equally important but consider price and product components to be the most critical factors of success (Goi, 2009).

Moreover, the exercise of actually implementing the ultimate mix in a company is tasked by various departments and persons within the organization, and companies take great efforts in order fully integrate their marketing activities. On the other hand, consumers mainly experience the individual effects of elements in different occasions, times and places (Constantinides, 2002; Wang, 2005)

The intrinsic notion of 4Ps in marketing mix is often viewed as ignorant from a customer's perspective (Popovic, 2006). This rather product-oriented definition is so called a "marketing management perspective" and is addressed by redefining each element in the mix. The transformation and addressed by Lauterborn can be attained by "converting product into customer solution, price into cost to the customer, place into convenience, and promotion into communication, or the 4C's" (Lauterborn, 1990; Goi, 2009).

Moreover, other than being internally oriented and lacking consideration for customer behaviour, the challenges that marketing mix framework faces is not to allow interactions with, or initiated from customers and is regarded as customer-passive (Möller, 2006). Also, this definition is not conclusive to specific elements of service marketing and ignores the fact that product is not singular and is not offered in isolation. Companies usually sell products lines, or brands, in an interconnected ecosystem (Constantinides, 2002; Wang, 2005)

From a practical perspective, in marketing mix modeling for a certain brand or product, there are numerous variables which do have an impact on sales therefore, regardless of a model that one chooses it might be prone to high noise (Wigren, 2019). Moreover, marketing mix modeling includes building and training a model on historical data aggregated on weekly or monthly basis. One major limitation of this approach is the number of data points collected in any specific time span. As an example, weekly data collected on yearly basis yields no more than 52 data points which generally would not yield to a reliable model development. Also, one might decide to aggregate historical data for an elongated period which also might lead to misleading conclusions. Mentioned phenomena commonly known as *Lucas critique* which states that determining effect of change in economic policy, would lead to highly misleading conclusions, in case it is solely based on highly aggregated historical data (Lucas, 1976).

From a modelling perspective there are mainly two challenges with marketing mix modelling using conventional techniques. First and foremost, nonlinearity of marketing expenditures is one aspect which needs to be addressed. As defined by (Tellis G. J., 2006) the shape effect is ”change in sales in response to increasing intensity of advertising in the same time period”. This means by increasing the level of investment in any advertising channel the effect of respected channel on the response variable, sales, will be diminishing. Traditional statistical modelling methods such as regression have an underlying assumption on linearity of the features which is not usually the case in real world.

In other words, there might exist certain thresholds after which a saturation effect on the advertising channel happens. These aforementioned phenomena could be addressed by ad stock and decay curves incorporated in modelling the marketing mix.

Moving on, another challenge with marketing mix modeling using traditional methods is that they fail to predict big data sets if they do not adhere to specific assumptions e.g. linearity, distribution and samples size, limiting the applicability of traditional methods. Also in current complex marketing framework, traditional demand forecasting models based on simple analytical techniques are only capable to sense the demand signals with liner and exponential trends, cyclic behavior and seasonality" (Kumar, Shankar, & Aljohani, 2020).

2.5. Machine Learning Methods

Today consumer journeys are becoming increasingly complex and variables influencing consumers' behaviour is constantly evolving. Machine learning has enabled marketers to increase their models adaptation speed in this volatile marketing framework and has become the new normal in the marketing field. Moreover, using these methods contribution of each channel or campaign can be evaluated which is a long-standing goal for marketers. Incorporating Machine learning in everyday business decision has been a common practice for corporations (Nanayakkara, 2020). Machine learning enables transformation of marketing mix elements into a more customer centric framework. In more recent takes marketing mix the "Ps" are substitute with consumer, cost, convenience, and communication (Londhe, 2014).

Machine learning is a subfield of artificial intelligence as it can learn and improve without being explicitly programmed. There are multiple fields of machine learning with supervised and unsupervised being the most popular. Supervised machine learning algorithms functions with data, which is unclassified, they can discover hidden patterns in the given data set. Unsupervised machine uses past or current information to predict future outcomes. Supervised machine learning uses train and validation data sets to determine the predictive strength of the model.

By Integrating ML models into the traditional statistical methods, the managers were able to gain added dividends with increased precision. According to a research, conducted on an Asian internet campaign, using ML models and social network analysis led to significantly higher conversion rate opposed to using best-practice marketing (Sundsøy, 2014).

Modelling the marketing mix with the aid of machine learning methods poses a potentially fruitful solution. As research by Kumar, A., Shankar, R., & Aljohani, N. R. (2020) indicates that in a 360 marketing framework where, marketers are dealing with multiple promotional campaigns and channels, suggested machine learning methods, specifically random forest outperformed traditional statistical models in maximising revenues and ROI.

Machine learning enables companies to have dynamic pricing for products and services (Campbell, 2019) such practises are often seen in hospitality industry. With the ability to track buying trends machine learning can track patterns and suggest competitive pricing solutions. From a promotional standpoint machine learning will facilitate promotion strategy by identifying ROI on each channel and provide optimisation in both content and level of promotion.

Machine learning methods comes with many advantages and also has its down sides. To name a few, one of the main advantages is that machine learning algorithms can promptly handle unstructured data such as text, images, audio, and hybrid forms. Moreover, they can process huge amounts of data and are highly flexible in terms of dealing with non-linearities. The latter or so called “feature engineering” aspect is frequently cited as the key success factor in winning entries of data science contest. the following sections, these advantageous aspects of machine learning will be addressed in detail and some caveats of them will be pointed all as well.

2.6. Advantages Machine Learning

Traditional statistical analysis has been found to be insufficient to grasp all the necessary data points to sufficiently carry out MMM. It has been mentioned in prior research that, using traditional methods based on historical demand data points, predictions of the future customers demand behavior are no longer accurate and effective enough as they ignore critical factors (Kumar, Shankar, & Aljohani, 2020). Regression models are the fundamental basis of traditional marketing methods, nevertheless they often report short comings in capturing the quantitative effect of the researched marketing channels, specifically their long-term impact (Kumar, Shankar, & Aljohani, 2020).

Machine learning models can handle unstructured data which enables an extended extraction of useful insights. Application of these techniques on network or tracking data with complex structures combined with the possibility of processing images and audio, is a competitive advantages of machine learning techniques that have tremendously impacted popularity of these models (Ma, 2020).

The ability of handling large volumes of data is another advantage of Machine Learning models. Unlike economical models in which a typical rage of predictors and feature are introduced with in the models, Machine learning techniques can be trained on huge data sets and many variables. As a matter in fact certain machine learning models such as random forest and neural networks will generate substantially better results which data points exceed a few hundred thousand (Ma, 2020).

In such data settings where we are dealing with a handful of features in the data, traditional statistical methods provides us with credible results yet, in a more complex structure where possible associations in data increases statistical inferences become less accurate and the competitive advantage of machine learning techniques become more apparent (IJ, 2018).

Another interesting aspect to machine learning models is their degree of flexibility which enables users to capture non-linear relation between variables. Set of predictors in machine learning models can be present in the model with many forms of transformations. They can have higher order forms or transform into a binary variable. Actually, Machine Learning, encourages extensive upfront efforts on creating and transforming input variables. It has been mentioned in the literature that " feature engineering" plays a key role in success and popularity of machine learning methods (Romov, 2015). Feature engineering and non-linear model structure substantially assists the model in capturing the connection of predictor and response variable (Ma, 2020).

Machine learning and statistical methods both are used for prediction, yet the accuracy of machine learning methods is proved to outperform the latter in most cases (Sabbeh, 2018). Machine learning techniques unlock patterns in the data and perform prediction using general-purpose learning algorithms, and come as handy when one is dealing with rich and unwieldy data. On cases where dealing with "wide data" where, number of features exceeds the data points, machine learning techniques have substantial advantage. These methods make minimal assumptions about the relationships between features and also can be useful where data are gathered without a carefully controlled environment (IJ, 2018).

In a nutshell most of the aforementioned advantages of machine learning techniques put-together, can be summarized in their higher prediction accuracy. Mostly, statistical methods try to so casual relationships whilst, machine learning techniques focus of prediction and model performance (Ma, 2020).

2.7. Disadvantages Machine Learning

In machine learning model there is often no clear model structure and a clearly define linkage between response variable and features in the model. As mentioned earlier machine learning models to a great degree depend on feature engineering a flexible model structure which substantially decreases the interpretability of the model; In return, these models often benefit from a higher predictive accuracy which makes them referred to as black-boxes (Ma, 2020). In a marketing mix modeling setting interpretations and extracting insights are generally more valuable for marketeers therefore this may be considered as the main disadvantage for these techniques.

On the other hand, there are some post-hoc analysis which can be applied to machine learning methods and are independent of the model which may lead to some insights yet, these methods cannot fully account for the interpretability which in some cases is required from the models. The relationships between features and response in machine learning models are often considered correlational rather than casual (Ma, 2020). The aforementioned phenomena, makes it difficult perform certain a cause-and-effect analysis, which is essential for designing and evaluating marketing mix or other important decisions such as seg-mentation and engagement. Which in turn limits the use of these models in a core capacity in marketing (Ma, 2020).

Lastly, in certain marketing area such as capturing consumer differences in an individual level, consumer heterogeneity, machine learning methods are not proven to outperform conventional statistical techniques (Ma, 2020). Traditionally, dynamic data are generally analyzed using time series models or other analysis such as HMM (Hidden Markov Model) , which can be considered techniques in the field of statistics, rather than machine learning.

3. Methodology:

3.1. Conceptual Modelling

Demand estimating models have gained significant popularity over past years. They are certain form of marketing mix models, where market response variable is sales or market demand (Sabbeh, 2018). These models are especially used by companies in industries such as health care, retail, and internet industries or and other firms that has access to data and data collection techniques (Bajari, 2015). Right metric selection upon marketing mix practices is a crucial task for managers as they rarely have a shortage of metrics that they can include in the mix (Ofer Mintz, 2021).

Trying to meet the needs of different stakeholders in a business the elements included in the mix could differ (Mintz, 2013), but for this study the following model is proposed. If K products each with observable and measurable characteristics X_j, the demand of these K products in market m and time t is shown below:

$$\ln Q_{Kmt} = f\ (P_{mt}, A_{mt}, X_{mt}, D_{mt}, \epsilon_{Kmt})$$

In the equation above A is a representation of advertising and promotional activities, D a vector containing demographical information and P is the price vector. In the model above idiosyncratic shock is added with ϵ.

This study will estimate the relationship between the right-hand side of the equation and sales value (demand). The first model that is explored is linear model. Moving on a regression analysis with variable transformations will be deployed and later on a tree-based model, Random Forest, will be introduced which provide higher prediction accuracy with the price of limited interpretability (Hastie, 2009).

3.2. MMM using linear models

In linear models a continuous response variable is explained as a function of independent predictors. The following linear model captures the effect of basic marketing mix elements such as Promotion (advertising), price, place, and product (Pandey, 2021):

$$Y_t = \alpha + \beta_1 A_t + \beta_2 P_t + \beta_3 X_t + \beta_4 D_t$$

In this model Y_t is the response variable or sales of specific product, α is some sort of base for the dependant variable, and coefficients β_j represent the effect of j^{th} independent variable. Other mix elements in line with what was under in conceptual modelling [section 3.1] are as follows: *A*(Advertising), *P* (Price), *X* (Product Characteristics) and *D* (Demographic/Place).

In the model above *t* is representation of time for each of the variables. In this study unit of measurement will be considered weeks. This is a common practice in marketing mix modelling as the unit of measurement for aggregated sales data in this field is usually considered to be weeks.

3.3. Linear regression model and Least Squares

Considering $X^T = (X_1, X_2, \dots, X_k)$ as an input vector the linear model has the general form which is represented below. In this model the function $f(X)$ will predict a real- valued output Y. (Hastie, 2009)

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$$

Regardless of source of Variables "X_j's", this model is linear in parameters. In marketing mix applications in usual there is a training data set on which the linear regression model is fitted. The most popular estimation method includes calculating coefficients matrix $\beta = (\beta_0, \beta_1, \dots, \beta_{k)}$ in such way that residual sum of squares is minimized. The equation of residual sum of squares on OLS method are as follows:

$$RSS(\beta) = \sum_{i=1}^{N} \left(y_i - f(X_i)\right)^2 = \sum_{i=1}^{N} (y_i - \beta\, 0 \quad - \sum_{j=1}^{N} x_{ij}\beta_j)^2$$

If the training set includes random observations from the population this criterion would be reasonable. In case that observations are not random the criterion is still valid as long as $y_i's$ are conditionally independent given the inputs (Hastie, 2009; Gareth James, 2013). In the special case of two predictors figure 1 is a graphical representation of OLS method.

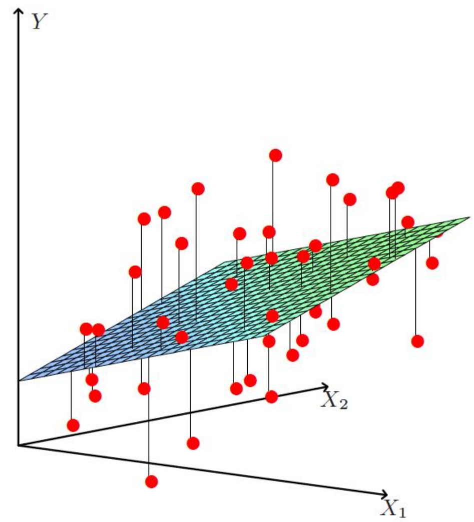


Figure 1- Graphical representation of OLS line in a three-dimensional setting with two predictors and one response. Under this setting the regression line becomes a plane.

3.4. Assessing the accuracy of the model and R^2 Statistic:

For any marketer, it is natural to assess the accuracy of the obtained model. In linear regression setting usually the quality of fit is measured by two related metrics: Residual Standard Error(RSE) and R^2 statistic (Hastie, 2009; Gareth James, 2013).As mentioned, it is assumed in the linear regression model that a relationship exists between X and Y which takes the following form: $Y = f(x) + \epsilon$ in which f, is an unknown function and ϵ is a mean-zero random error term. Due this error term even the true regression line (with a known β_o and $\beta_{1,\ldots}$) would not correctly predict all the responses. RSE is a representation of the variance of the error terms and is defined as follws:

$$RSE = \sqrt{(\frac{1}{n-2}RSS)}$$

RSE is considered a metric which defines the lack of fit in the model meaning if predictions are very close to the true values RSE will be very small and vice versa.

The problem often associated with RSE is that since it is measured in units of output it does not come with a clear guideline on what is a good RSE. On the other hand, R^2 statistic provides an alternative measure of fit indicating the proportion of explained variance. It takes a value in between zero and 1 and is calculated as follows:

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - RSS/TSS$$

In the above formula TSS is total sum of squares and is calculated as $TSS = \sum(y_i - y^{\bar{}})^{\wedge}2$. R^2 is a measure of variability in response variable that can be explained using the predictors. (Gareth James, 2013).

3.5. Prediction accuracy and overfitting in linear regression:

Linear regression in complex model settings, is prone to overfitting. This refers to a condition where statistical captures random errors and not the true relationship between features of the model. This phenomenon will result in a low MSE and high R^2 on train data yet, the prediction accuracy of the model on future data will be greatly affected

3.6. Machine learning methods for regression

A common Machine Learning technique for estimation and classification is to use so called “Tree based models”. This approach mainly includes stratification and segmentation of the predictor space. In these models the mean or mode value of the training observations belonging to the specific region is used as model prediction (Hastie, 2009; Gareth James, 2013). The main advantage in these models is their high interpretability which is due to their tree-based nature. On the other hand, these methods do not have a competitive prediction accuracy with other machine learning methods, therefore, a common approach is to combine multiple tree-based models to increase predictions accuracy. These techniques mainly include bagging and Random Forest which will be further discussed.

3.7. Regression Trees

The process of building a regression tree can be roughly summarized into two steps as explained in (Hastie, 2009):

1. First the predictor space containing all the possible values for $X_1, X_2, \dots, X_p$ will be divided into *K* non-overlapping and distinct regions $T_1, \dots, T_k$
2. For every observation that falls into any of *K* regions (assume T_i), the same prediction will be made which is the mean of the response values for the training observation in T_i

From a theory perspective these regions could have any shape or form, however for sake of simplicity and ease of interpretation the objective is to divide the predictor space into high-dimensional rectangles or boxes. Subsequently the end goal is to find these regions in such way that RSS which is defined below is minimized.

$$RSS = \sum_{j=1}^{K} \sum_{i \in T_j} (y_i - \hat{y}_{T_j})$$

In the equation above $\hat{y}_{T_j}$ is the mean of the response variable in the j-th box for the training set.

Due to computational limitations, it would be impossible to get every possible set of partitions in the predictor space, therefore a *top-down, greedy* approach is utilized (Gareth James, 2013). In this setting the splitting starts at the top of the tree and successively divides the predictor space. Also, at each step of the process the best split is made for that particular step not for the tree as a whole.

3.8. Bagging

Generally, decision trees suffer from high variance. In other words, if the training set is divided into multiple subsets and different decision trees are built based on the subsets the resulting trees might be quite different (Gareth James, 2013). Bootstrap aggregation is defined as a procedure in order to reduce the variance of a statistical method. For any given set of *k* independent observations with a variance of $\boldsymbol{\sigma}^2$ the variance of the mean of these observation is given by $\boldsymbol{\sigma}^2 \ / \ k$. In other words averaging set of observations will reduce the variance. A possible extension of this technique is its application in improving the variance of statistical methods, in such way that upon building a separate prediction model on many training sets taken from the population and averaging the resulting prediction which will lead to a low variance statistical learning model. The downside of this approach is that usually we do not have access to many training set therefore a possible solution is to take many samples from training set that we have and finally averaging the results. In this approach K different bootstrapped training set are taken and from training the model of each of those K samples (*assume k-th sample)* $f^{\wedge^{*k}}(x)$ are predicted. Subsequently averaging all the predictions will lead to a predictor model with lower variance (Gareth James, 2013). This process is called bagging for which the mathematical representation is provided below:

$$f_{bag}^{\wedge} = 1/K \sum_{k=1}^{K} f^{\wedge^{*k}}(x)$$

Bagging can improve performance of many regression methods, yet it is mainly used for decision trees.

3.9. Random forest

Random forest is an extension on bagging and is based on averaging a large collection of de-corelated trees (Hastie, 2009). As explained upon averaging high variance (noisy) but low biased models we can substantially improve performance of the utilized statistical method. Trees are preferred candidates for bagging as they are able to grasp complex interactions in the data, also if grown deep enough they would have relatively low bias (Hastie, 2009). As mentioned in this improved version of bagging the tweak is to have uncorrelated trees in our bag of trees (Gareth James, 2013). Upon building these decision trees at each split a random sample of *m* predictors are chosen as split candidates from the all *P* predictors. As expected only one of the chosen predictors is going to be used and afterwards a fresh sample of m predictors are going to be chosen again (Gareth James, 2013). As a rule of thumb usually, $m = \sqrt{p}$, in other words the number of candidates for predictors at each split is set equal to the square root of total number of predictors (Gareth James, 2013). Following aforementioned algorithm upon building the forests at each split a large number of predictors are not even allowed to be considered at the splits this rational will result in having uncorrelated trees in cases such as having a very strong predictors and many moderate predictors. In other words, the main difference between bagging and random forest is due to the number of predictors chosen at each split in such way if m would be chosen equal to p then both methods would be leading to same predictions. The algorithm of regression random forest are briefly represented below:

1. For b=1 to B draw a bootstrap sample of the predefined size from the training data.
2. Grow a random forest tree T_b, following the recursively repeating steps mentioned below for each node until the minimum node size is reached:
    a. m variables will be selected at random from the fill set of P predictors
    b. The best variable to split is selected and the node will be constructed
3. Output the collection of trees $\{T_b\}_1^B$
4. To make a prediction for a new point *x*: $f^{\wedge}{}_{rf}{}^{B}(x)$: $1/B \sum_1^B T_b(x)$

3.10. Partial Dependence Plot (PDP)

Partial dependence plots show the type of the relationship between, the dependant feature of the model (response variable – Sales in this study) and other independent features. Generally, these relations can be linear, monotonic, and even more complex. In case of a linear regression model

PDP indicates a linear relation. Partial dependence for regression is defined below (Friedman, 2001).

$$\hat{f}_S(x_S) = E_{x_C}\left[\hat{f}_S(x_S, X_C)\right] = \int \hat{f}_S(x_S, X_C)\, d\, P\,(X_C)$$

The x_S are the features for which the partial dependence function should be plotted and X_C are the other features used in the machine learning model $\hat{f}$. By calculating averages in the training set, known as Monte Carco method $\hat{f}$, or the partial function, is estimated. Partial dependence works by marginalizing the machine learning model output over the distribution of the features in set C, so that the function shows the relationship between the features in set S we are interested in and the predicted outcome.

4. Data:

4.1. Description of the Dataset

The Data-set utilized for the analysis, consists of real word marketing mix data for Greece cheese market. Specifically, data-set includes Nielsen weekly reported data for a period over two years from 02-01-2017 to 30-12-2019. Data-set includes sales and marketing spending data for 12 biggest cheese brands in Greece which are reported on daily basis. Additionally, other data sets such as weather data and event calendar data (indication certain days which might have influence sales data) for the abovementioned period of time are added to the dataset which will be further explained. In total the data set consists of 1872 data points and 17 features which will be explained.

It's worth mentioning that in specific the resources of Friesland-Campina (Royal Friesland-Campina N.V. is a Dutch multinational dairy cooperative which is based in Amersfoort, Netherlands.) is used in the analysis. FC produces 3 cheese brands (NONNOY, MILNER, FINA) in Greek market therefore the data available on these 3 brands is extended to online spending as well. This additional data will be used for an extension on the analysis, introduced in the this study.

4.2. Variable description

Table 1 presents the name and description of variables included in the dataset:

Table 1- Variable name and description of different features included in the model:

Variable name	Description	Type	Rational
TOT_vol_sum	Total Sales volume in number of cheese packages sold	Numeric	Market Response
TOT_price_avg	Sales price of specific brand in certain week	Numeric	Marketing Mix
TOT_price_index _avg	This value indicates level of promotion price for specific brand in a week. The formula for this calculating is as follows: non-promotion price/ average price	Numeric	Marketing Mix
Media_offline	Offline media spending (own brand’s) in that week for mentioned brand	Numeric	Marketing Mix
Media_online	Online media spending(own brand’s) in that week for mentioned brand	Numeric	Marketing Mix
COMP_TOTAL_ mediaspent	Offline media spend of the competition (other cheese brands) during that week	Numeric	Marketing Mix
date	Weekly values ranging from January 2017 to December 2019	Date format	Control
month	Indication on month of value reported	Numeric	Control
year	Indication on year of value reported	Numeric	Control
brand	This value indicates brand name which consists of 12 Biggest cheese brands in the Greece cheese market	Character	Control
quarter	Indication of quarter of the year	Numeric	Control
TOT_wd_max	Weighted distribution of specific brand in that week. This number	Numeric	

	indicates percentage score for the availability of the product across the stores that specific week		
event_New_Years _Day_max	New years day	Binary	Control
event_25th_of_M arch_national_holi day_max	25th of March Holiday	Binary	Control
event_Orthodox_ Easter_Monday_ max	Easter Monday	Binary	Control
event_Orthodox_ Easter_Sunday_m ax	Easter Sunday	Binary	Control

4.3. Descriptive Statistics

In this section, for the most important variables in the model, an overview will be given and, and a model free analysis will be provided .

- Total sales volume *(TOT_VOL_sum)* for all brands:

The data consist of aggregated weekly sales volume for 157 weeks (~ 3 years) for various brands. As mentioned, 12 brands are included in this analysis. Figure 2 is related to sales volume of these brands over the period of 3 years. As it is demonstrated in the figure the biggest brands in the market are 2 FC cheese brands (NOYNOY and MILNER) along with a competitor brand ADORO. In sales data usually the spikes are an indication of a promotional sales which is followed by a huge dip. This pattern is an indication of promotional shopping behaviour, in other words consumers buying products during the promotion would not buy the products after promotion therefore, there will be a huge dip in the sales data.

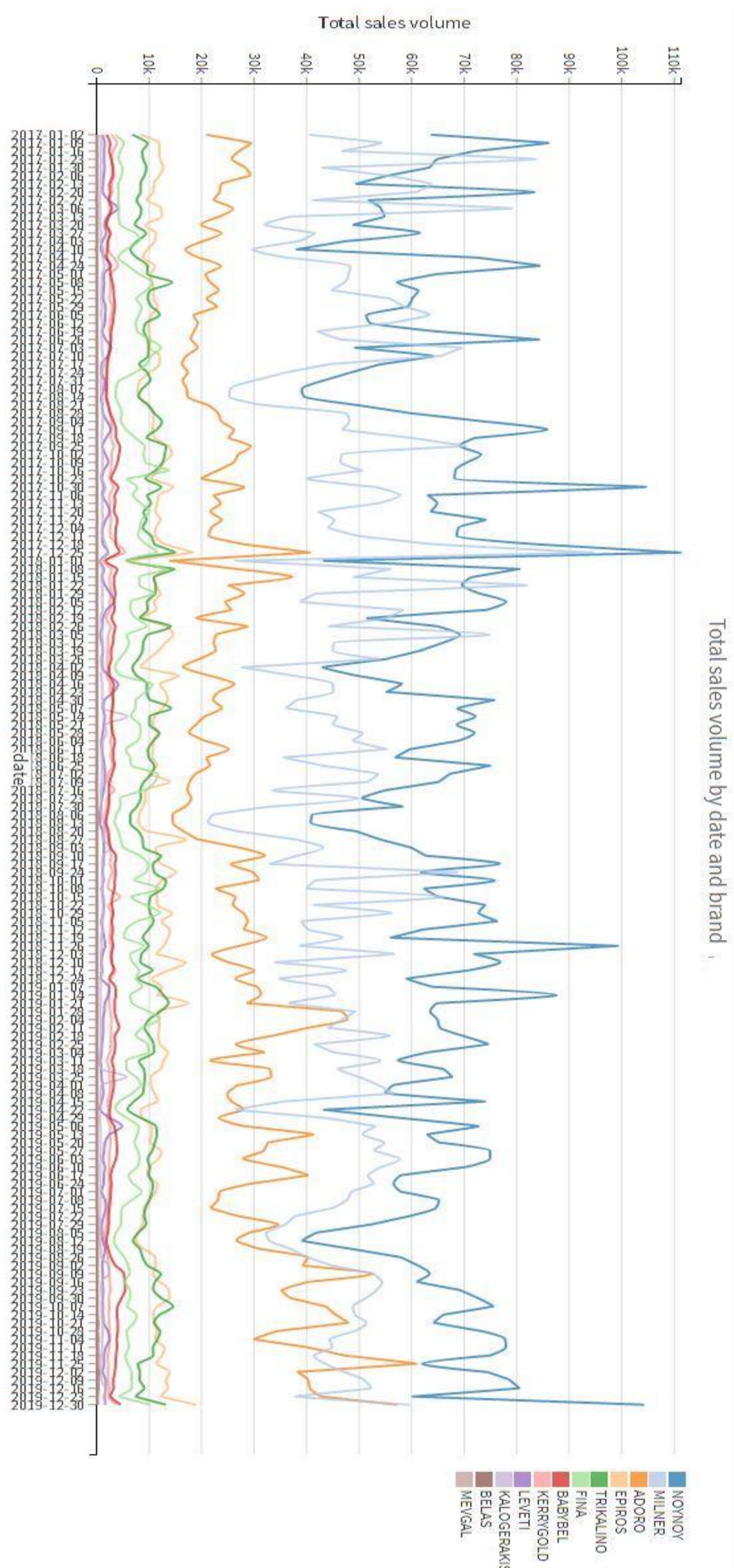


Figure 2- Total sales volume of different brands per week in three-year time span of the data collection

- Selling market price *(TOT_price_avg)*:

Looking into brand level features of the data-set, as an example, the biggest market player in Greek cheese market , NOYNOY, is selected for demonstration. Sales volume and price changes of the brand over time has been plotted below (Figure 3-4).

The General price pattern shows an increase from 2018 onward which is subsequently corresponding to a decrease in vol in that period. Also Volume shows a huge variation (spikes in the times series) in second half of 2018 and early 2019 which can be further investigated. In general pattern in the data shows the dips in the volume are corresponding to the spikes in the price which builds the foundation of the analysis and marketing mix model building.

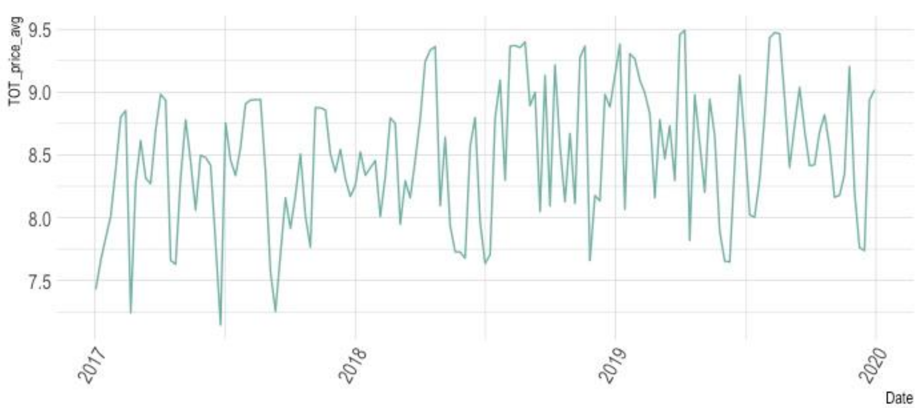


Figure 3- Average market price for brand "NOYNOY" over years

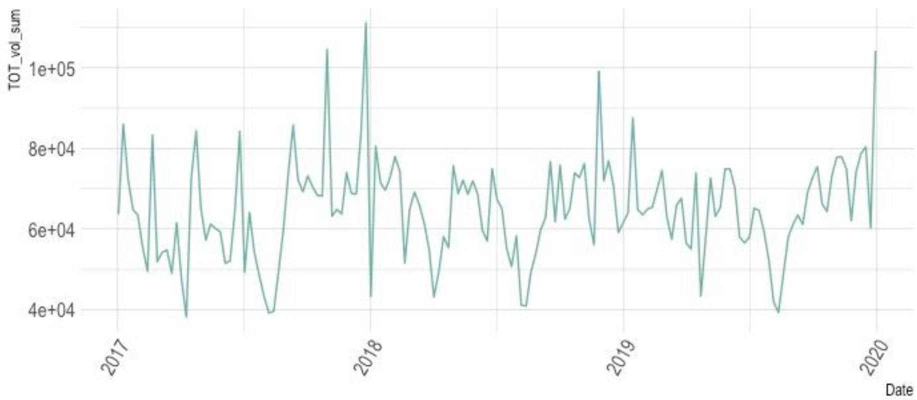


Figure 4- Average sales volume for brand "NOYNOY" over years

- Online and offline marketing advertising investment *(Media_offline and Meida_online):*

Another important feature in the data which worth investigation in the media investment level in both offline and online mediums. The general pattern of the data as demonstrated below indicates a higher level of investment in offline media, yet the investment cycles in offline setting are also considered to be longer. In an online platform due to the variety of assets and platforms more campaigns take place during the year. Online advertisement has higher reach and therefor the investment levels are generally considered to be lower. Another aspect of media advertisement is the Ad-stock. In other words, advertisement effect is not necessarily limited to the investment period and could last for several periods after advertising. This criterion should be considered upon investigating the effects of advertising on sales in marketing mix modelling. Offline and Online media spent for brand " NOYNOY" is plotted below(Figure 5-6):

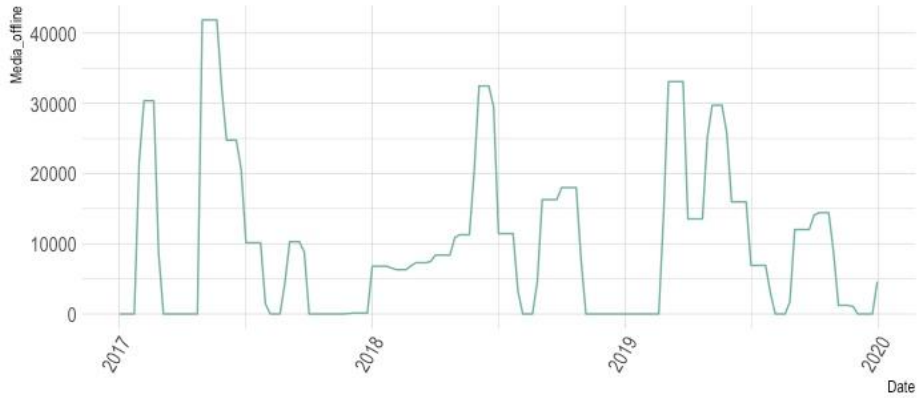


Figure 5- Offline media spent of brand "NOYNOY" over years

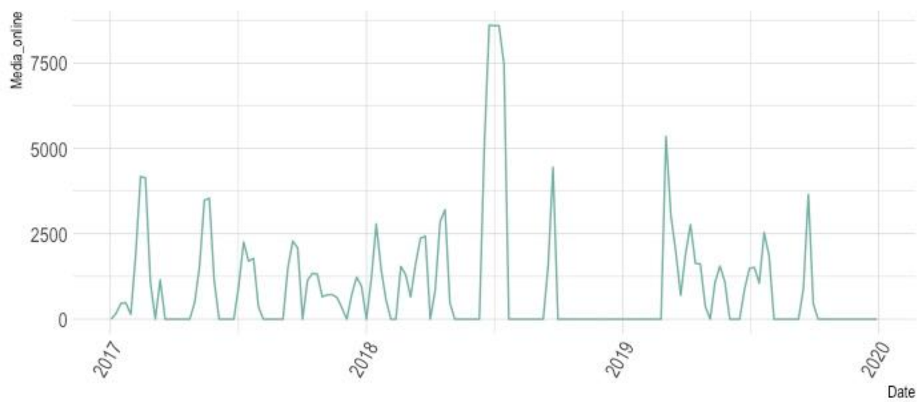


Figure 6- Online media spent of brand "NOYNOY" over years

- Product availably / distribution *(TOT_wd_max)*:

  Next feature that is going to be investigated is the weighted distribution of NOYNOY. As it is shown in the Figure 7, throughout the year the *wd* does not have a lot of variations. This would be important at a brand level however for different brands across the data this variable could be considered important as product availability directly affects the sales levels.

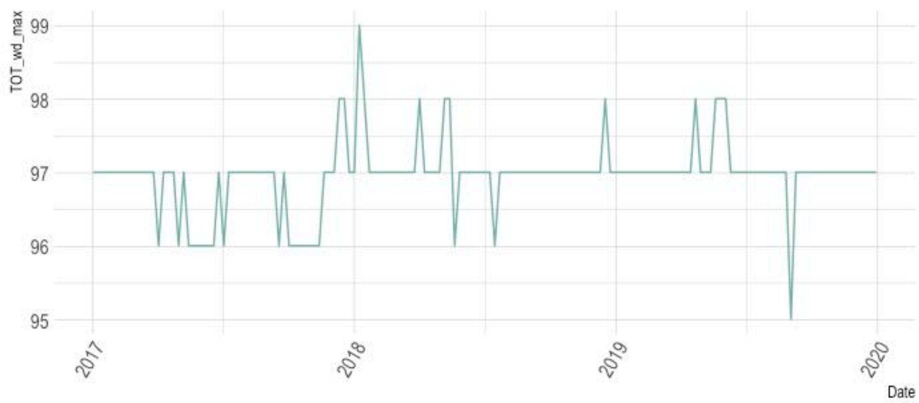


Figure 7- Weighted distribution (product availability) for brand "NOYNOY"

After exploring the data and getting a sense of variables and features of our data set the next step is to start the analysis. Next section in going to build up the analysis performed in this study with a stepwise approach. In this setting we are going to compare the performance and results of different methods and techniques in marketing mix modelling for individual brands and subsequently the analysis will be extended to whole Greek market.

As for the next step the correlation of different numerical features of the dataset is calculated. This will help to construct a general idea of how features are potentially correlated and also, will inform us of the potential multicollinearity of the features is important in certain methods. Figure 8 is graphical demonstration on correlations of most important features of the dataset. As expected, product availability and media investment have positive correlation with sales volume and price is negatively corelated with volume which is in-line with the initial assumptions. Moreover, the correlation values between predictive feature is no indication of multicollinearity between variables which builds the foundations of further analysis.

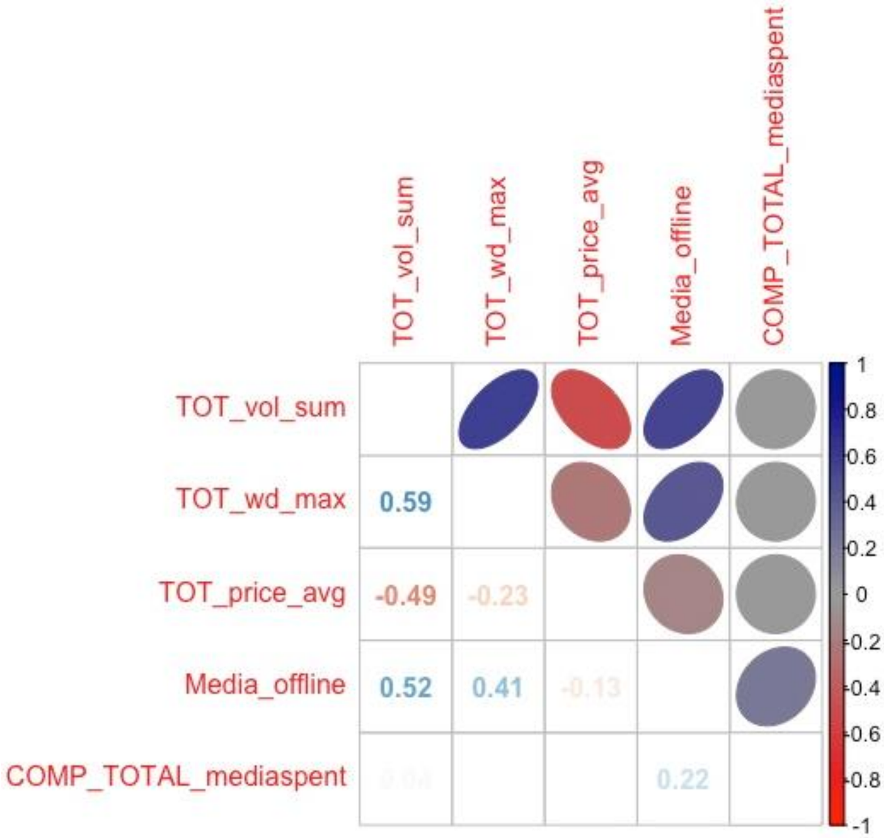


Figure 8- Correlation plot of numerical features in the dataset

5. Analysis

5.1. Data Preparation:

The main objective of this analysis is to compare and contrast the performance of machine learning techniques (namely Random-Forest) and simple linear regression (OLS) in marketing mix modelling applications. In this setting generally the objective is to make accurate predictions on target variable (in this study total sale volume) and try to explain the effect of different features of the model on target variable.

After checking for missing and invalid values in the data, the distributions of the features are investigated. A suitable transformation, accounting for such as right tailed price distribution, which is present in the data, is logarithmic transformation. Therefore, required transformations are applied to certain features in the data such as: Price, Weighted distribution, and Media spent of all brands in the market.

5.2. Marketing mix modelling using Linear Regression Model (OLS):

with the simplest model for singular brand and subsequently extent the model to multiple brands and a conclusive marketing mix model at a market level. A simple approach to marketing mix modelling is to fit a regression model to the data. The first step in the regression analysis will start by developing a marketing mix model for an individual brand in the data (eg. MILNER). The data points on this brand consists of weekly sales volume collection adding up to 157 weeks in between a three-year period. Afterwards the analysis will be extended to develop a market inclusive model for which the data for all brands will be incorporated. In this model along with features present in the data set, also the interactions of different variable such as brand, price, and media spent are added to the model.

Equation below is representation of the fitted model to the data. It is worth mentioning for earlier mentioned reasons variables: price, online investment, and offline investment are transformed into their logarithmic form as an example variable *MILNER_TOT_price_avg - (market price)* is replaced by Log (MILNER_TOT_price_avg+0.001). The value 0.001 is added prior to transformation to account for 0 values (missing prices in certain weeks).

$$\begin{aligned}
&\log(\text{Tot_vol_sum}) = \beta_0 + \beta_1 \log(\text{Milner_Tot_price_avg} + 0.001) + \\
&\beta_2 \text{Log}(\text{fc_gr_medium_brand_offline_MILNER_mecetsp} + 0.001) + \\
&\beta_3 \text{ Log}(\text{fc_gr_medium_brand_online_MILNER_mecetsp} + 0.001) + \beta_4 \text{ summer_hol_dummy} + \\
&\beta_5 \text{temperatureHight_avg} + \beta_6 \text{ event_new_Years_Day_max} + \beta_7 \text{event_25th_of_March_national_max} + \\
&\beta_8 \text{ event_Orthodox_Easter_Monday_max} + \beta_9 \text{event_Orhodox_Easter_Sunday_max}
\end{aligned}$$

In the model above β_0 is the intercept and β_1 to β_9 are the coefficients in the model. Below the results of regression model on brand MILNER data is presented. Also, following the coefficients, the explainability and measurements of the fit for the model are listed in table 2:

Table 2- OLS Regression coefficient results for marketing mix modelling

Variable Name	Coef	Std err	t	p>|t|	[0.25	0.975]
Intercept	15.1327	0258	58.554	0.000	14.622	15.643
Log (MILNER_TOT_price_avg+0.001)	-1.9731	0.120	-16.406	0.000	-2.211	-1.735
Log (fc_gr_medium_brand_offline_MILNER_mecetsp + 0.001)	0.0036	0.002	2.155	0.033	0.000	0.007
Log (fc_gr_medium_brand_online_MILNER_mecetsp + 0.001)	0.0021	0.001	1.482	0.141	-0.001	0.005
summer_hol_dummy	-0.1828	0.051	-3.567	0.000	-0.284	-0.082
temperatureHight_avg	-0.0075	0.002	-4.662	0.000	-0.011	-0.004
event_new_Years_Day_max	-0.1536	0.086	-1.781	0.077	-0.324	0.017
event_25th_of_March_national_max	-0.1421	0.068	-2.094	0.038	-0.276	-0.008
event_Orthodox_Easter_Monday_max	-0.1733	0.068	-2.560	0.011	-0.307	-0.039
event_Orhodox_Easter_Sunday_max	-0.4133	0.068	-6.076	0.000	-0.548	-0.279
R-squared	0.795					
Adj R-squared	0.783					
F-statistic	62.97					
Prob (F-statistic)	7.84e-46					

Relatively high value for R-squared is a demonstration of high explainability of the model. Looking at the coefficients obtained results are mainly significant and could be argued for causality relationships. In the above log-log model the Coefficient of price can be interpreted as the percentage of volume decrease that would be predicted by the model. The obtained relationships for media and price variables are in-line with the expectations. In other words, a negative relationship between volume and price and a positive relationship for media spent and predicated sales volume are demonstrated in the results. Also, temperature and Dummy variables have

A common approach includes controlling the Variance Inflation Factor (VIF-score) of an independent variable, which represents how well that variable is explained by other independent variables. Considering relatively low VIF scores of variable sin the model (1< <3.84) the potential multicollinearity of the features in the model is dismissed.

Figure 9 is representation of prediction vs actual values in the model. The red line is actual sales value and the blue line in the prediction of the model (regression-line) for response variable. This indicates the results are representative of market developments and are well estimated with this model.

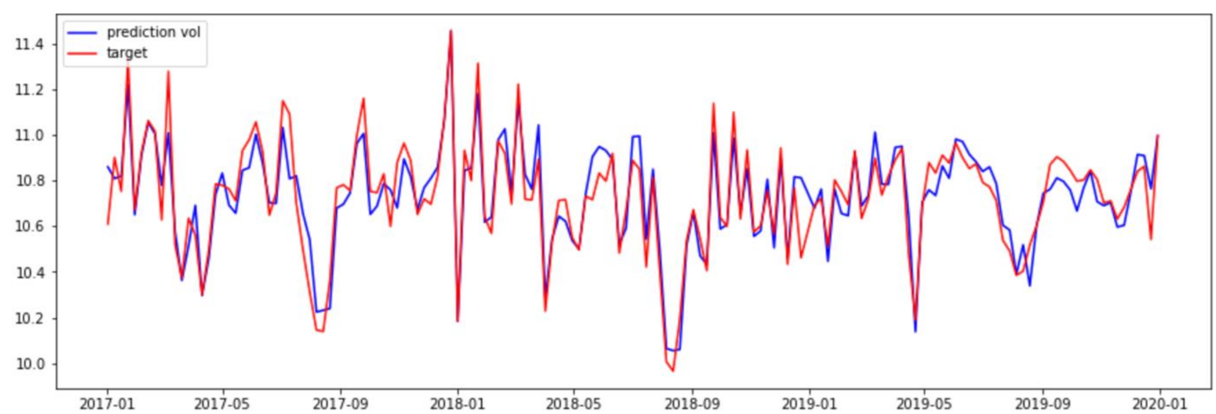


Figure 9- Prediction versus Actual results (sales volume) for regression analysis

5.3. Feature contribution (importance) in linear regression model

One of the preferred interpretations of in both statistical and machine learning models is often to comparing the effect of certain features of in model. In linear regression setting contribution plot (which shows to what extend each of the coefficients are contributing to the prediction) is a common representation of feature importance. In this simple model for visually demonstration, plot below is presented. For each data point: the actual sales value, predicted sales value, price, media spent (including online and offline) the base (coefficient value for controls in the model such as temperature and holiday dummies) have been plotted in Figure 10.

To better clarify with an example in 1st week of May,2018 the price of Milner multiplied by the obtained coefficient in the regression will be calculated and plotted in Orange.

Different colors give a visual sense of contribution of each of these features to the actual sales value. As in line with the regression coefficients the sensitivity of price in this case in much higher than media investments. This can be explained by the nature of the product (cheese) and the advertising conversion rate in the industry. In FMCG sector considering a usual market scenario, where there has not been a new product development, and where products have high levels of brand awareness the price sensitivity is much higher than other elements of the marketing mix. Consumers often have no preference in consuming certain products in similar category and the main indicative factor in their consumer journey would be the price. in such situation the sales volume are substantially affected by changes on price.

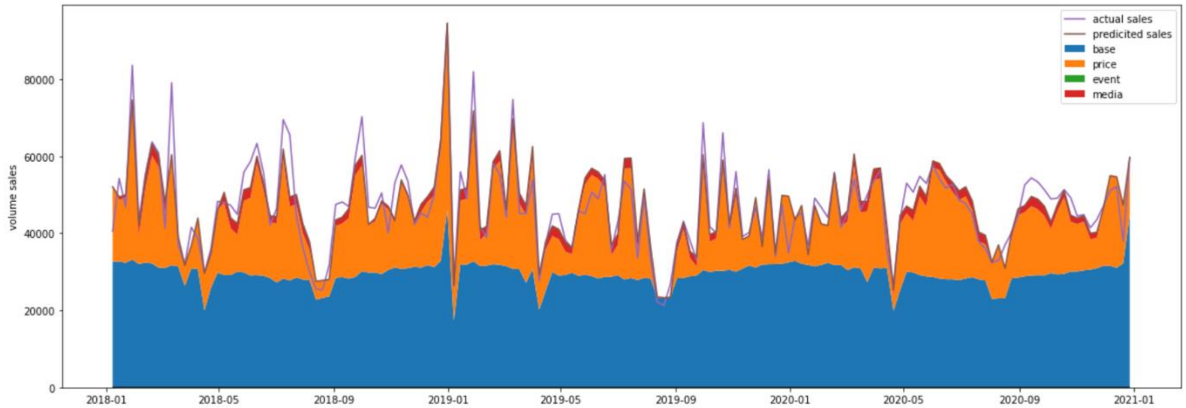


Figure 11- OLS Regression model's contribution plot. This plot is graphical representation for contribution of each feature in the model. Also, predicted versus actual results are demonstrated with brown and purple lines respectively.

5.4. Marketing mix modelling at country level:

The next step of the analysis includes the extension of regression analysis on whole data set. As mentioned data-set contains information on 12 biggest cheese brands in the Greek market, therefore, a categorical variable " brand" , is added to the model. Also certain features such as interaction between price, media spent and brand is added to the model. Following the original intention behind the study, in order to better compare the models, the data set is splitted into test and train sets and all models are trained and their performance are tested on an identical data set. In this analysis 30/70 rule for test/train set is implemented. Also considering this random selection an additionally added criterion makes sure that the class balance of different brands in test and train set are proportionate. This strategy makes sure that the test set is a good representative of training data in brand level analysis.

From a company perspective, marketing mix modelling would be beneficial for demand forecasting upon potential changes in price and promotional factors. Therefore, the developed model is intended to be utilized for a specific brand. In this analysis an additional step is taken, to test the model in the aforementioned concept. After using the developed model for predictions on test set the results are grouped by brands and the fit of the model for each brand in calculated in the same manner as the original model. These results will be interpreted following the analysis. Below is provided the mathematical expression of the linear regression model fitted on the data:

Model : log(TOT_vol_sum) = β_0 + β_1 * log(TOT_price_avg + 0.001) + β_2 * Log(TOT_price_index_avg + 0.001) + β_3 * log(Media_offline + 0.001) + β_4 * log(COMP_TOTAL_mediaspent + 0.001) + β_5 * log(Media_offline + 0.001):C(brand) + β_6 * C(brand) + β_7 * C(quarter) + β_8 * summer_hol_dummy + β_9 * temperatureHigh_avg + β_{10} event_New_Years_Day_max + β_{11} * event_25th_of_March_national_holiday_max + β_{12} * event_Orthodox_Easter_Monday_max + β_{13}* event_Orthodox_Easter_Sunday_max

In the model above β_0 is the intercept and β_1 to β_{13} are the coefficients in the model. Also, the explainability and measurements of the fit for the model are listed in Table 3:

Table 3- Measures of fit for country level marketing mix modelling using linear regression model

Dep. Variable	Log (TOT_vol_sum + 1)	R-squared:	0.989
Method	Least Squares	Adj. R-squared:	0.989
No. Observations:	1872	F-statistic:	4752.
		Prob (F-statistic):	0.00
		Log-Likehood:	72.452
		AIC:	-72.90
		BIC:	126.3

The extreme high R-squared of the model in a potential sign for overfitting. In marketing mix modelling problems, the market level data on different brands will increase the data points which will benefit the model in complicated ML models yet fitting linier regression model in the data is prone to overfitting. After perming predictions on the test set the other performance metrics of the model are listed in Table 4:

Table 4- predictive performance of the model on test data set

Model	MAE	RMSE	R2 Score
Ordinary Least Squares	0.09	0.15	0.96

In-line with our initial model the regression model not only can make predictions of the demand but also the model can provide a meaningful interpretation with the obtained coefficients. The results of the regression model is provided in the Table 5-6:

Variable Name	Coef	Std err	t	p>|t|	[0.25	0.975]
Intercept	12.7355	0.238	53.575	0.000	12.269	13.202
C(brand). BABYBEL	-0.8206	0.101	-8.108	0.000	-1.019	-0.622
C(brand). BELAS	-5.1469	0.051	-101.044	0.000	-5.247	-5.047
C(brand). EPIROS	0.0390	0.067	0.586	0.558	-0.092	0.170
C(brand). FINA	-1.0201	0.044	-23.201	0.000	-1.106	-0.934
C(brand). KALOGERAKIS	-2.2855	0.066	-34.458	0.000	-2.416	-2.155
C(brand). KEYYYGOLD	-1.6700	0.059	-28.241	0.000	-1.786	-1.554
C(brand). LEVETI	-2.2327	0.058	-38.717	0.000	-2.346	-2.120
C(brand). MEVEGAL	-5.2174	0.089	-58.314	0.000	-5.393	-5.042
C(brand). MILNER	0.8412	0.046	18.177	0.000	0.750	0.932
C(brand). NOYNOY	1.1384	0.186	6.129	0.000	0.774	1.503
C(brand). TRIKALINO	0.0245	0.076	0.323	0.746	-0.124	0.173
C(quarter). T.2	0.0493	0.025	1.978	0.048	0.000	0.098

Table 5- OLS Regression coefficient results for marketing mix modelling

C(quarter). T.3	0.1107	0.030	3.710	0.000	0.052	0.169
C(quarter). T.4	0.1115	0.018	6.279	0.000	0.077	0.146
Log(TOT_price_avg+0.001)	-1.5102	0.093	-16.178	0.000	-1.693	-1.327
Log(TOT_price_index_avg+0.001)	-0.6227	0.136	-4.592	0.000	-0.889	-0.357
Log(TOT_wd_max+0.001)	3.6063	0.474	7.603	0.000	2.676	4.537
Log(Media_offline+0.001)	0.0057	0.004	1.600	0.110	-0.001	0.013
Log(Media_offline+0.001).C(brand). BABYBEL	0.0133	0.006	2.314	0.021	0.002	0.024
Log(Media_offline+0.001).C(brand). BELAS	0.0860	0.008	10.331	0.000	0.070	0.102
Log(Media_offline+0.001).C(brand). EPIROS	-0.0043	0.005	-0.900	0.368	-0.014	0.005
Log(Media_offline+0.001).C(brand). FINA	0.0055	0.005	1.207	0.228	-0.003	0.015
Log(Media_offline+0.001).C(brand). KALOGERAKIS	0.0002	0.005	0.036	0.971	-0.009	0.010
Log(Media_offline+0.001).C(brand). KEYYYGOLD	0.0009	0.005	0.192	0.847	-0.008	0.010
Log(Media_offline+0.001).C(brand). LEVETI	-0.0032	0.005	-0.706	0.480	-0.012	0.006
Log(Media_offline+0.001).C(brand). MEVEGAL	0.0120	0.005	2.513	0.012	0.003	0.021
Log(Media_offline+0.001).C(brand). MILNER	-0.0001	0.005	-0.022	0.983	-0.009	0.009
Log(Media_offline+0.001).C(brand). NOYNOY	0.0033	0.018	0.179	0.858	-0.032	0.039
Log(Media_offline+0.001).C(brand). TRIKALINO	-0.0053	0.005	-1.110	0.267	-0.015	0.004
Log(Comp_TOTAL_mediaspent+0.001)	0.0330	0.012	2.653	0.008	0.009	0.057
Summer_hol_dummy	-0.0839	0.034	-2.492	0.013	-0.150	-0.018
temperatureHight_avg	-0.0055	0.002	-3.670	0.000	-0.009	-0.003
Event_New_Years_Day_max	-0.0620	0.050	-1.243	0.214	-0.160	0.036
Event_25th_of_March_national_holiday_max	-0.0601	0.042	-1.444	0.149	-0.142	0.022
Event_Orthodox__Easter_Monday_max	-0.1434	0.042	-3.445	0.001	-0.225	-0.062
Event_Orthodox__Easter_Sunday_max	-0.2745	0.042	-6.563	0.000	-0.356	-0.192

Table 6- OLS Regression coefficient results for marketing mix modelling

As expected, Product price is negatively affecting sales volume, and media spent is positively correlated with the target. Based on the results obtained variable price index, indicating the level of promotion negatively affects demand volume. In other words, the higher the promotion level (meaning lower price index) will result in more demand. Quarter dummy capturing potential seasonality of the market is significant and capture the variation of the target variable.

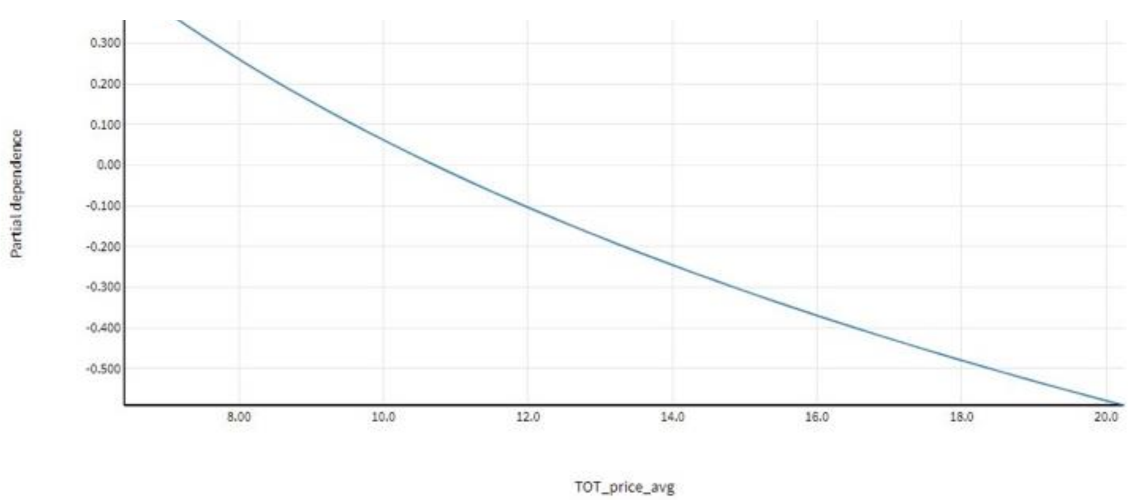
5.5. Partial Dependence analysis for linear regression

As discussed, partial dependence plots, indicate how model predictions will be affected on average upon changes in the selected feature. Following modelling the mix with regression model partial dependence analysis on the test set is performed. Following figures (Figure 11-12) show the changes in the predicted sales volume (market response) based on changes in the selected feature of the model. Liner model such as OLS will predict a constant and liner effect of these feature specifically, price on the volume. In other words, as one would expect as the price of the product will be increased the sales volume will be negatively affected.

In practice it is often the case that prices effect on sales does not follow a linear response. this phenomenon can be one of the downsides of linear regression models. In subsequent model development stages, the nonlinear price elasticity with more complicated models will be analysed.

Also, for media investments after a certain threshold a minimal effect is obtained. This result is in-line with the assumption that media investments will be saturated after a certain level and the effects of the media on volume is dependant of this response curves.

Figure 12- partial dependence plot of average price. This plot is a representation of marginal effect of price on models predictions



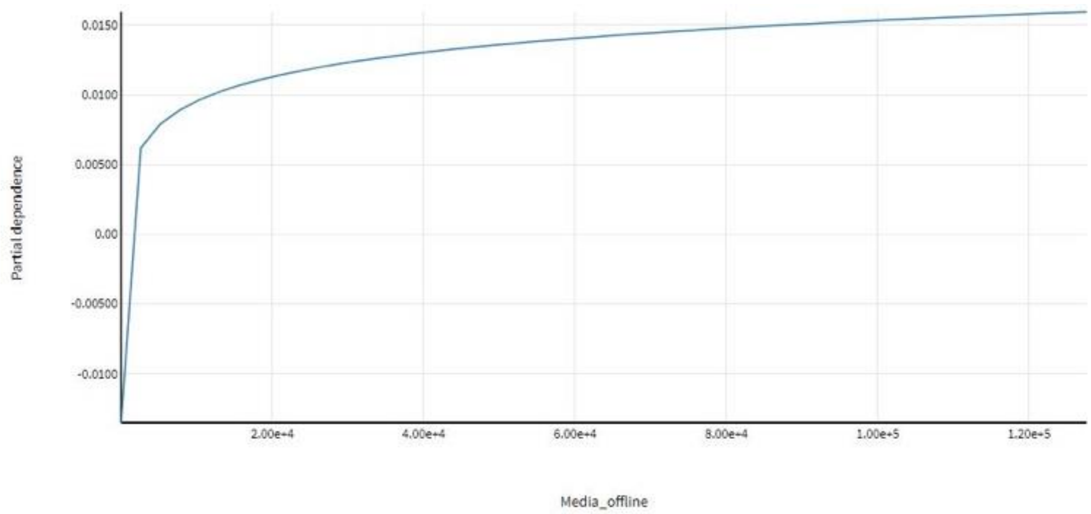


Figure 13- partial dependence plot of offline media investment. This plot is a representation of marginal effect of offline marketing investment on model's predictions

5.6. Linear model performance on brand level

To investigate the behaviour of the model across the certain feature and to see whether it performs identically a subpopulation analysis will be performed. The selected feature from the data set is product brand. As explained in the initial phase of the analysis the performance of marketing mix modelling should be consistent at brand level, in other words the fit of the model across different brands in the test data set will be calculated and it will be compared. Also, from a practical perspective at the end of the day the marketing mix model will be deployed to investigate ad predict the sale volume of a singular brand. In the figures (Figure 13-14), model prediction versus actual volume is plotted for two sample brands MILNER and NOYNOY (the two brands as stated are Friesland Campina's own brands hence, the brands as example) .

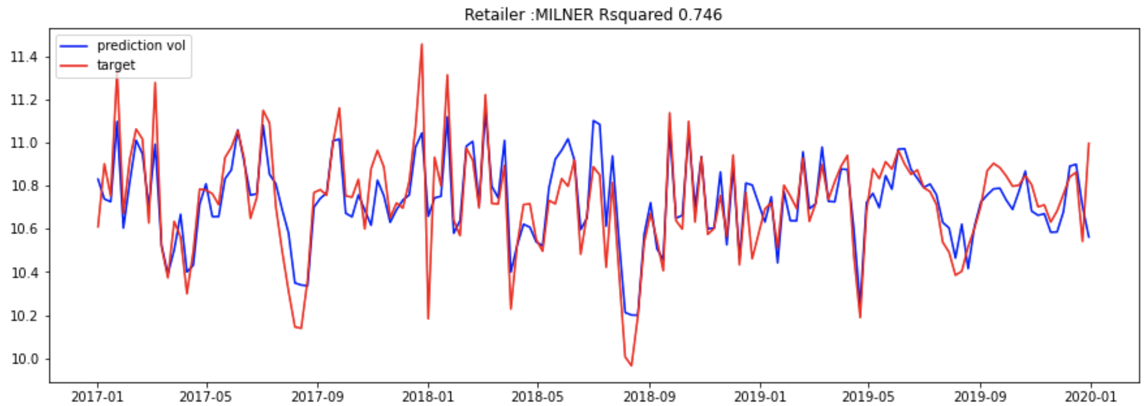


Figure 14- Prediction versus Actual results (sales volume) for linear regression model analysis for brand "MILNER"

"

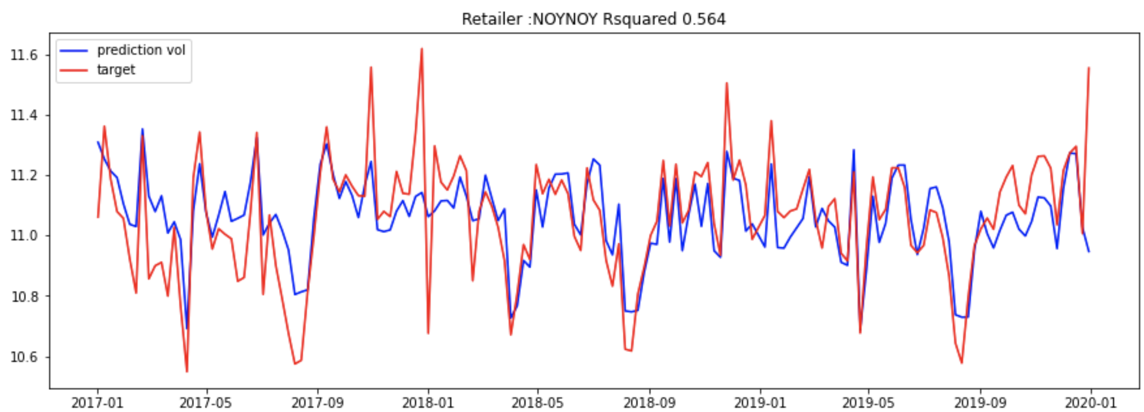


Figure 15- Prediction versus Actual results (sales volume) for linear regression model analysis for brand "NOYNOY"

The explainability of the model at a brand level is lower and inconsistent across different brands. It is interesting that, the brands with a higher total sale in the market (Market Leaders) have score substantially higher that small players. This phenomenon can be explained by lower volume variation of these smaller brands during the period. Also due to lower investments in these brands the correlation target variable and features in the model is considered to be lower than other competitors. Among the results for two specific brands ADORO and MEVGAL the model has demonstrated a very poor performance which is a sign of an omitted variable for this brands which is not captured with in the model. For specific brands effects such as product availability and out of stock issues and substantive promotional effects can cause predictions to deviate from target value. Predictive performance of model across different brand is listed in below Table 6 .

Brand	MODALITY (%)	Metric: RMSE	R2 Score
MILNER	8%	0.10	0.74
NOYNOY	8%	0.11	0.56
FINA	8%	0.07	0.71
ADORO	8%	0.21	0.12
BYBEL	8%	0.13	0.08
BELAS	8%	0.11	0.32
EPIROS	8%	0.10	0.50
KALOGERAKIS	8%	0.12	0.29
KERRYGOLD	8%	0.09	0.51
LEVETI	8%	0.12	0.53
MEVGAL	8%	0.12	0.43
TRIKALINO	8%	0.09	0.59

Table 20- Predictive performance of OLS regression model at brand level

5.7. Marketing mix modelling using Random Forest

Moving on from linear regression models for the next step in the analysis, a random forest model is trained on the training set . In this model the predictions of multiple decision trees are combined which will assist the model to improve the prediction accuracy. In order to better compare the results obtained with this model and OLS model another regression model is trained (fitted) on the same train dataset, as random forest, and using this model predictions on volume for a test data set is made. Subsequently the result of both model is compared and interpretations are made.

One of the important step training random forest model is hyperparameter selection. Changing these features in the model directly affects complexity and accuracy of the model. In this study a search grid has been defined for hyperparameter selections, also a 5 fold cross-validation approach is utilized in order to find the best model parameters. Based on the minimum OOB - MSE the best fit for the model is chosen. Table 7 contains the best tuned hyper parameters for the model.

Table 21- Hyperparameter tunning result for Random Forest model:

Parameter	Description	Value
ntree	Number of trees in the forest	200
mtry	Number of variables which is randomly selected at each split	15

Subsequently after training the model the predictions on the test set is made and the model performance, is obtained:

Table 22- Predictive performance of Random Forest model:

Model	MAE	RMSE	R2 Score
Random forest	0.08	0.13	0.98

5.8. Feature importance in Random Forest

As explained in the methods, one of the most beneficial post hoc analysis that would greatly help to interpret the outcomes of the model is feature importance plots. The results indicate product availability is the most important feature in the model. Product price, followed by total media spending in the competition are the second and third important variables in the model which are demonstrated in the figure below. In comparison with regression model, the influential features of random forest are more or less comparable with the exception of competition media spend which in regression model was not determined as a very effective feature, considering the value of its coefficient. A downside of random forest model is that the direct effect of each feature in the model is not explainable as easy as OLS regression. One possible solution to understand the effect of certain feature is to look into partial dependences obtained after predictions, which will be discussed in the following. Feature importance plot for Random Forest model is provided in Figure 15.

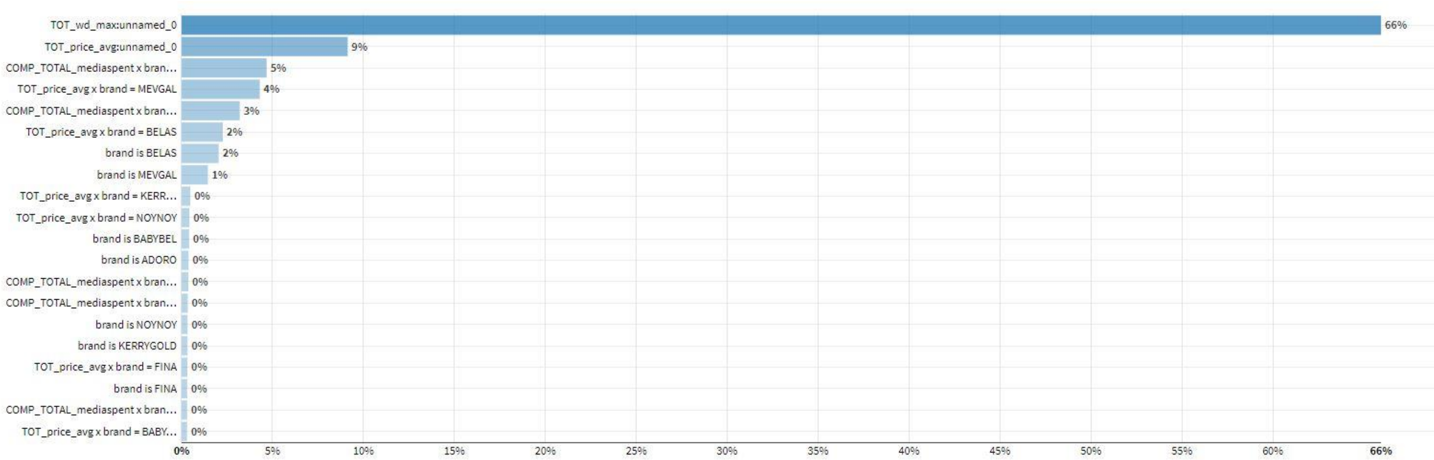


Figure 16- Feature importance plot for Random Forest model

5.9. Partial dependence analysis

Random forest algorithm, make it possible for the model to capture non-linear effects in the feature of the model. In the figures below to most important features, product distribution and price are demonstrated for further analysis. As shown in Figure 16, the distribution of the product follows somehow a bracket like effect meaning, there are some jumps is the prediction of the model when the distribution changes around certain thresholds(e.g. 50% to 60%). Moreover, it is interesting that the distribution effect slows down above 60% and somehow follows a horizontal line increasing to 90%. This result can be interpreted as the effect of product availability for the main competitors of the market being minimal as they have a relatively higher availability versus other market competitors.

The second plot (Figure 17), showing the changes in model's prediction in different price ranges demonstrates a more intense reduction for price increase than the results obtained from OLS model. Also changes due to price increases does not follow a linear effect which was previously obtained from linear regression model.

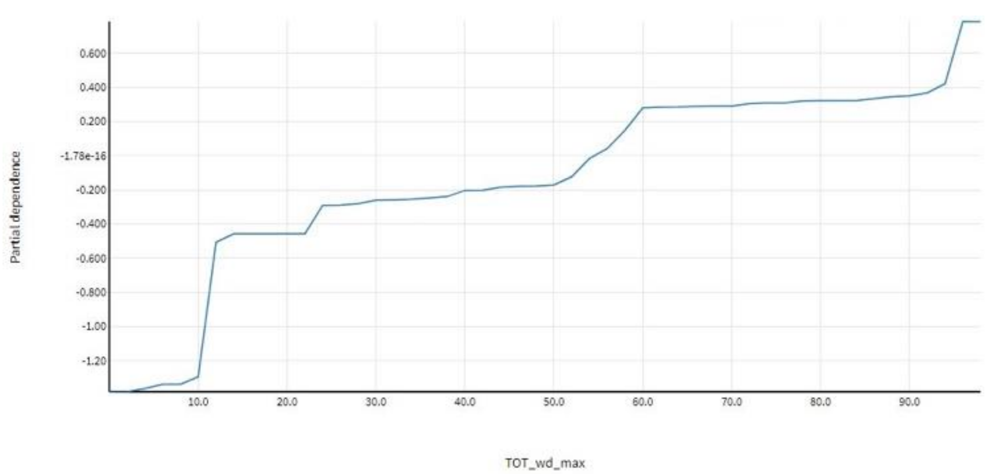


Figure 17- partial dependence plot of weighted distribution. This plot is a representation of marginal effect of product availability(wd) on model's predictions

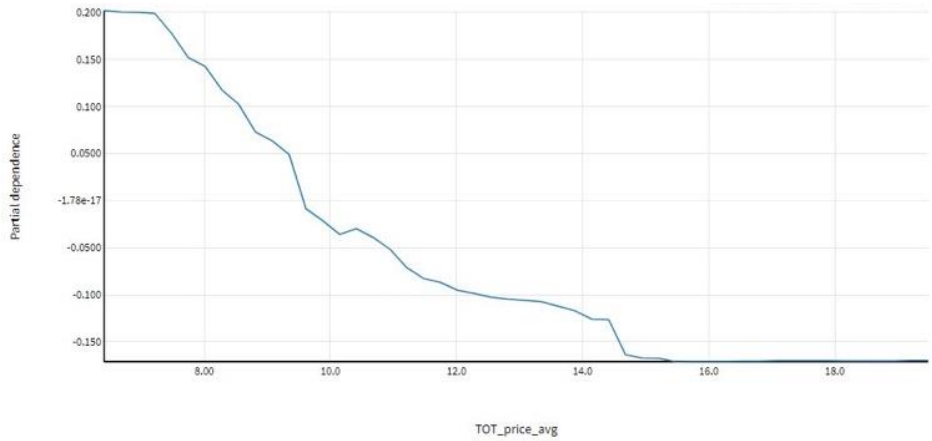


Figure 18- partial dependence plot of product price. This plot is a representation of marginal effect of selling price on model's predictions

5.1. Random forest model performance on brand level

As it is mentioned earlier form a company's point of view the true advantage of marketing mix models is to be able to give predictions and make interpretations and a brand level, therefore it does makes sense to review models behaviour with regard to different values of certain features. In this model the accuracy and fit of the model is reviewed for different brands and the results are demonstrated in the table below. In order to compare the predictions of OLS regression model and random forest a new linear regression line is fitted with in the test data set and using the results of this model predictions has been made on the test set. Subsequently, models performance metrics, are grouped by different values for brands and are compared with the outcome of random forest model. Table 9 represents an overview of the aforementioned comparison in between two models. As expected, random forest model is performing substantially better in explaining the variance and also has lower error value in almost all brands. Random forest model is shown to effectively capture the variation in the outcome variable and specifically in brands with lower market share and volatile sales patterns. The results show, for mid players in cheese market, using linear regression model forecasts may not be reliable as these brands usually have huge shifts in marketing budgets and their promotion plans. On the other hand random forest is substantially performing better for the same data set.

		Random Forest		Linear Regression	
Brand	MODALITY(%)	Metric: RMSE	R2 Score	Metric: RMSE	R2 Score
MILNER	8%	0.05	0.81	0.09	0.79
NOYNOY	8%	0.10	0.63	0.11	0.53
FINA	8%	0.06	0.78	0.07	0.74
ADORO	8%	0.15	0.32	0.21	0.21
BYBEL	8%	0.09	0.68	0.13	0.52
BELAS	8%	0.07	0.49	0.11	0.05
EPIROS	8%	0.07	0.54	0.10	0.42
KALOGERAKIS	8%	0.06	0.73	0.12	0.48
KERRYGOLD	8%	0.07	0.71	0.09	0.50
LEVETI	8%	0.08	0.41	0.12	0.31
MEVGAL	8%	0.10	0.48	0.12	0.19
TRIKALINO	8%	0.07	0.71	0.09	0.50

Table 23- Predictive performance of Random Forest at brand level on the test data set

6. Conclusion:

6.1. Linear regression:

As in regression modelling, one attempts to model the sales as closely as possible, incorporating as many significant variables as possible. Also, amongst other metrics, quality of fit and interpretability are considered main indicators of success in MMM development. (Wolfe Sr & Crotts, 2011). The assumption of a linear regression model is that the regression function $E(Y|X)$ is linear in the inputs $X_1, X_2, \dots, X_k$). These simple models provide ample interpretability and, in some cases, outperform more complicated non-linear (Hastie, 2009) models.

The first model deployed in the analysis, OLS linear regression, creates a baseline for both prediction and interpretation of marketing mix models. The values obtained based on model fitted in this data set are 0.96 for R-squared and 0.15 for MSE both of which are considered high scores in the industry. On the other hand, as expected, the predictive performance of OLS, translated into such as MAE and RMSE, is lower than Random Forest method used in the analysis.

A point which worth highlighting is that generally marketing mix modelling are constructed on aggregated monthly or weekly market level data which does not generate that many data points over a relatively long period of time (Tellis G. J., 2006). As an example, for a specific brand in this research data set is an aggregated retail level data set for a time span of over three years which in total generates less than 200 data points. Therefore, one of the benefits on OLS is, its applicability in relatively smaller data sets.

Secondly, the output of OLS regression provides a great degree of interpretability specially in marketing mix models. Moving on from the sales prediction performance the other objective of marketing mix modelling is to provide interpretations such as promotion or channel contribution for the product. Regression models provide such interpretations relatively easier and at the same time more user friendly to the stakeholders of the companies. Due to their simple nature and direct interpretability linear models are of importance in the industry, also they provide informative insights to marketers in order to get a better investment decision and increase ROI.

Comparing the performance of linear regression and random forest there is a trade-off between accuracy and interpretability which should be addressed depending on the application case of the data. In this study the interpretational heavy nature of marketing mix models is the main argument in favour of such models.

Marginal effect of features in the model of the response variable is another important outcome from marketing mix modelling. Specifically, companies are interested in understanding how certain price changes will affect sales of their brands. Upon perfuming such analysis OLS predicts a linear behaviour of the feature in the predictor space. This can be considered the main shortcoming of OLS in context of this study.

Based on OLS model price increase will negatively affect sales levels and diffract sensitivities in price brackets are ignored in this model. However, media investment plots have successfully depicted real life picture of the market. As one would expect investment would lead to an increase in sales level up to a certain threshold. This media saturation is caused due to the fact that advertising effects have diminishing returns on scale.

Moving on, at a brand level data, upon performance comparison for different brands in certain cases, OLS regression substantially underperforms other models. These instances are related to relatively newer brands in the market that either have not built a substantive past data or are undergoing significant price changes or promotions in their growth cycle. As explained linier in regression method is method noise is assumed to be independent of the predictors therefore, certain correlations in the data might lead to variance in the results.

Considering the trade of between interpretability and accuracy in the case of this study and possible data collection limitations for marketing mix models at a brand level OLS regression method provides comparative results and therefore it is still the main method for marketing mix modelling for the company.

One other main aspect which needs to be considered in the context of marketing mix modelling is how the trained models are going to be utilized from a company standpoint. These models after being trained on a market level data will be trying to predict the sales for different brands. What is of paramount importance is that the saturated markets such as FMCG products the price bracket of brands is greatly determinant of their sales. Subsequently, OLS regression has a lower performance at certain brands on the test sets. A more complicated prediction techniques which are able to capture high dimensional non linearities in predictor space will outperform this simple model are preferred.

6.2. Random forest

As mentioned earlier Random Forest model is based on the aggregation of the ensembled trees. Averaging the predictions of these trees will generate a lower variance. Random Forests are expansions of the notion around using a collection of predictors by introducing randomness into the variable set which are considered at the splitting node of the tress (Bajari, 2015).

One main advantage of random forest method is the increased prediction performance but at a relative low cost of lost interpretability. Amongst popular ML models popular in practice, Random Forest models are rather explainable to the stake holders therefor, it provides them a huge popularity in the practice. Also upon utilizing model agnostic methods the interpretability power of these models will be assisted and therefore such models are gaining popularity in practice.

In this study Random Forest model obtained values of 0.98 and 0.13 for R-squared and MSE respectively and outperformed OLS regression. Since Random Forest enables the model to capture complicated nonlinear relationships with the response variable, a higher prediction accuracy was resulted in the data. Also, in certain brands such as: Belas, Kerrygold, and Megval, with a more complex promotion pattern and a more volatile pricing strategy Random Forest was able to still result in acceptable prediction performance.

Feature importance analysis in Random Forest indicates that the effect of product distribution in the market on sales exceeds that of media investment which is a practical insight for companies. Also, the nonlinear marginal effect of price is perfectly captured with in the model. This indicates that after hitting price points 10 and 14 a drastic decrease in the sales volume will be seen which previously, based on linear regression was indicated to be a single rate reduction pattern.

Another benefit of machine learning techniques is the possibility of including multiple features in the data set and deploying these models correlated environment. Companies nowadays do have access to a substantive set of features and variables and have a tendency to generate complex insights on their target market. Random Forest provide the opportunity to use an extensive set of predictors and increase the prediction power of the model.

Another side to this equation is the complexity cost of such models which is directly in trade of with the resulted increase in accuracy. Advance machine learning methods such as Random Forest can provide substantial improvements in the performance of the model yet in cases such as marketing mix modelling the process behind the prediction is as important as the result itself. Machine Learning methods or so-called black-boxes provide an accurate prediction yet the exact process behind this prediction is rather unclear (Rai, 2020). In such cases that marketeers are after the effect of components of the marketing mix on the projected sales it would be more effective to have an interpretable model with less accuracy also a model which fits the ongoing business model but is less preferable from a solely statistical point of view can be preferable in such cases.

7. Bibliography

Bajari, P. N. (2015). Machine Learning Methods for Demand Estimation. *The American Economic Review*, 481-485.

Borden, N. H. (1965). The concept of the marketing mix. *Journal of advertising research*, 4(2), 2-7.

Campbell, C. S.-Y. (2019). From data to action: How marketers can leverage AI. . *Business Horizons*, 63(2), 227-243.

Constantinides, E. (2002). The 4S Web-Marketing Mix Model. *Electronic Commerce Research and Applications*, 1(1), 57-76.

Culliton, J. (1984). The Management of Marketing Costs. *Division of Research, Graduate School of Business Administration, Boston, MA: Harvard University.*

Gareth James, D. W. (2013). *An introduction to statistical learning : with applications in R.* Newyork: Springer.

Goi, C. L. (2009). A Review of Marketing Mix: 4Ps or More? *International Journal of Marketing Studies*, 1(1), 2-15.

Hastie, T. T. (2009). *The elements of statistical learning: data mining, inference, and prediction.* New York: Springer.

Huang, M. H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research*, 24(1), 30-41.

IJ, H. (2018). Statistics versus machine learning. . *Nature methods*, 15(4), 233.

Joy, J. (2006, May). Understanding Advertising Adstock Transformations. *Munich Personal RePEc Archive*.

Kellerman, B. J. (1995). Product and Pricing Courses are Underrepresented in Undergraduate Marketing Curricula. *Journal of Product & Brand Management*,, 4(1), 18-25.

Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial marketing management*, 90, 493-507.

Lauterborn, B. (1990). New Marketing Litany: Four Ps Passé: C-Words Take Over. . *Advertising Age*,, 61(41), 26.

Londhe, D. B. (2014). Marketing Mix for Next Generation Marketing. *Symbiosis Institute of Management Studies Annual Research Conference,* (SIMSARC13), 335-340.

Lucas, R. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 19-46.

Ma, L. a. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. *International Journal of Research in Marketing*, 481-504.

McCarthy, E. J. (1964). Basic Marketing, IL: Richard D. Irwin.

Mintz, O. &. (2013). What drives managerial use of marketing and financial metrics and does metric use affect performance of marketing-mix activities? *Journal of Marketing*, 77(2), 17–40.

Möller, K. (2006). The Marketing Mix Revisited: Towards the 21st Century Marketing by E. Constantinides. *Journal of Marketing Management*, 22(3), 439-450.

Nanayakkara, N. W. (2020). Application of Artificial Intelligence in MarketingMix: A Conceptual Review. *The Conference Proceedings of 11th International Conferenceon Business & Information ICBI, University of Kelaniya, Sri Lanka*, ISSN 2465-6399, (pp.530-542).

Ofer Mintz, T. J. (2021). The right metrics for marketing-mix decisions. *International Journal of Research in Marketing*, Volume 38, Issue 1,Pages 32-49.

Pandey, S. &. (2021). Marketing Mix Modeling (MMM) -Concepts and Model Interpretation. . *International Journal of Engineering and Technical Research. 10. 784. 10.17577/IJERTV10IS060396.*

Popovic, D. (2006). Modelling the Marketing of High-Tech Start-Ups. *Journal of Targeting, Measurement and Analysis*, 14(3), 260-276.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137-141.

Romov, P. a. (2015). Recsys challenge 2015: ensemble learning with categorical features. *Proceedings of the 2015 International ACM Recommender Systems Challenge*, 1-4.

Sabbeh, S. F. (2018). Machine-Learning Techniques for Customer Retention: A Comparative Study. *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 9.

Sethuraman, R. &. (1991). An analysis of the tradeoff between advertising and pricing. *Journal of Marketing Research*, 31, 160–174.

Tellis, G. J. (1995). Cracking the retailer's decision problem: Which brand to discount, how much, when and why? . *Marketing Science*, 14(3), 271–299.

Tellis, G. J. (2006). Modeling marketing mix. Handbook of marketing research.

Wang, K. L. (2005). A Comparative Study on Marketing Mix Models for Digital Products. *Proceedings of the First International Workshop on Internet and Network Economics (WINE'05), Hong Kong, China*,, 15-17, 660-669. .

Wigren, R. &. (2019). *Marketing Mix Modelling: A comparative study of statistical models.*

Wolfe Sr, M. J., & Crotts, J. C. (2011). Marketing mix modeling for the tourism industry: A best practices approach. . *International Journal of Tourism Sciences,* 11(1), 1-15.

Yuxue Jin, Y. W. (2017). *Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects.*

8. Appendix

8.1. Response Curves in marketing mix modelling

Marketing mix modelling is an embodiment of sales response function which demonstrates the effects of marketing activities on sales. The functional form of the model indicates the relationship of the dependant variable (sales) and marketing mix elements. In our initial model this relationship is considered to be linear which translates into an increase in advertising efforts will cause an increase in sales. In real world practices this might not always be the case and a phenomenon called "advertising saturation" occurs after certain level of investments in the media channels. In other words, after certain level of investment the effect of advertising on sales will be diminishing to scale (Yuxue Jin, 2017).

An extension of the basic linear model is so called multiplicative model in which the independent variables of MMM are multiplied together (Pandey, 2021). A general representation of this model is shown below:

$$Y_t = \exp(a) * \beta_1 B_t * \beta_2 C_t * \beta_3 D_t * \varepsilon_t$$

Where Y representing the market response and B,C, and D are Marketing mix elements in the model. A simple logarithmic transformation will linearize this model which will allow us to estimate regression coefficients using our OLS method. One of the main advantages of this model is that it allows sales (Y) to take variety of shapes depending on the value of the coefficients. As for the next step of the analysis using a python package response curves for online and offline media are constructed.

In the resulting sale response function the effect of different elements in the marketing mix will be captured. Functional form of the relationship between sales and the feature of the model reflects nature of the marketing activity. As an extension of the linear regression model the multiplicative mix model is fitted on the data and the resulting response curve for online and offline media activities are plotted below. Figure below, is a demonstration of saturation level for online and offline mediums. As expected, due to high level of investment on offline media this medium has a relatively higher saturation level, also this result shown investments above certain thresholds (8K€ for online - 42K € for online) will not result in an increased sales.

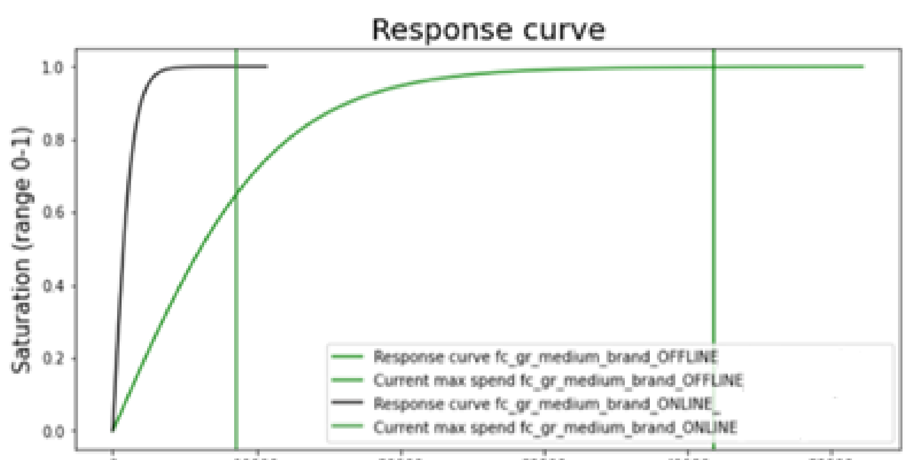


Figure 19- Online and Offline advertising response curves, an output of marketing mix modeling

8.2. Global segregate model:

Another way to assist the interpretability of black box ML models such as random forest is to replace the model with a more explainable models such as decision tree. Figure below is a demonstration of segregate model for which a maximum depth of 4 is selected. In this explainable model product availability and product price are main predictors of first nodes of the tree and other features will appear as the more in depth the model is.

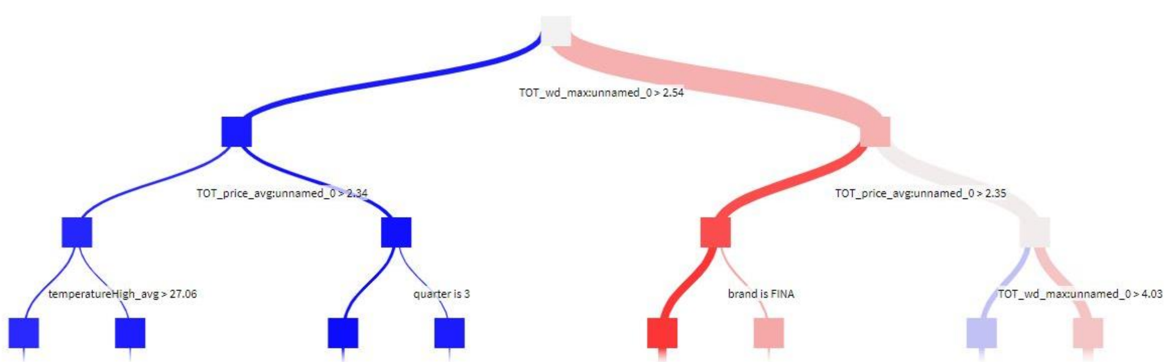


Figure 20- Representation of global segregate model