Marketing mix modeling algorithms for FMCG industry

Symeon Tsinaslanidis

SID: 3308180023

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

JULY 2022

THESSALONIKI – GREECE

Marketing mix modeling algorithms for FMCG industry

Symeon Tsinaslanidis

SID: 3308180023

Supervisor:	Dr. Agamemnon Baltagiannis
Supervising Committee Members:	Dr. Christos Berberidis
	Dr. Dimitrios Baltatzis

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

JULY 2022

THESSALONIKI – GREECE

Abstract

This dissertation was written as a part of the MSc in Data Science at the International Hellenic University.

Perhaps the most challenging, cruelest and disciplined markets are the world of fast-moving consumer goods: pure science and exceptional thinking, investment in consumer and competition study for a fully oriented business orientation, merit quality and constancy in marketing information decide market share, viability and sustainability. During each point of the model, a range of iterations of the Industry Life Cycle model are used to guide the focus of the marketing activities. Launch Engineering allows FMCG companies to be more profitable, strengthen branding, extend marketing messaging, monitor ad agencies and refine category management. A simpler, quicker road to trial and brand acceptance is included in FMCG data. Relevant patented (pre-launch) new product pre-launch review method almost removes the risk of not preparing a product launch; sophisticated market segmentation strategies give you a strategic 'edge'. Improved advertising returns, trade investment (sometimes referred to as promotional budget), sales marketing & public relations (pr & advertising) also account for FMCG consulting fees.

The exploration of three linear regression techniques, in order to explain the volume of sales, highlights strengths and weaknesses, of each one of them. Two of those techniques are applied in practice and their results are compared

I would like to express my gratitude to my supervisor, Prof. Baltagiannis, for his support, during this challenging task, and the opportunity he gave me to acquire knowledge from this exiting industry of FMCG. Of course, I owe a BIG ‘Thank you!’ to my family and especially my mother Sophia, my father Savvas and my sister Paraskevi who were always very supportive.

Symeon Tsinaslanidis

07/07/2022

Table of Contents

1 Introduction

Normally, when a new product is launched into a market, it undergoes a sequence of market stages; these steps are growth initiation, maturity, and eventually the decline period. These phases chronologically follow each other and are hence referred to as the PLC of the product life cycle. In the business setting, the PLC sequence or series is closely related to the trends and has resulting implications on the marketing combination of goods and marketing techniques. The product lifecycle graph is a graph that usually maps sales against the phases of the product.

In the launch process of the PLC, businesses usually seek to create brand knowledge of the product with the use of a marketing blend. Initial steps entail the development of couples with intellectual property rights, such as trademarks, for consistency and branding. If there are already existing players, the price policy may be low to ease penetration into the market, although it may be high if there are no rivals and this helps the initial expense to be easily recovered If the marketing targets early adapters and innovators at the development stage of the PLC, distribution is typically selective, the companies seek to maximize their share market by providing additional features to the product quality while retaining costs. Distribution is improved to satisfy the increased demand, while when your customer says your brand is to be de-listed, the campaign reaches a wider audience. The brand may be described as merchandise, services, or both; in other words, it is something that meets the needs of the consumer. - product has its own limited life; however, it shares the same feature, and we describe the "Product life cycle" period the product goes through.

2 Literature review

As the commodity demand and rivals change overtime, the FMCG identifying and positioning approach shifts. Most curves of the Commodity Life Cycle are represented as bell-shaped. Usually, this curve is broken into four stages: presentation. Development, maturity, and fall (Day,[6]). To say four items is a commodity with a life cycle. Products have a short life; product sales go through various phases, each providing the seller with different obstacles, opportunities and problems; revenues grow and fall at different stages of the PLC; and products at each point of their life cycle need different selling, environmental, production, buying and human resource techniques (Tibben-Lembke, [27]). According to Johansson [10], the commodity Life Cycle resembles the S curve in the traditional marketing illustration, with the development duration corresponding to when the S has the steepest ascent. This is why a new product is frequently launched to reap first-mover opportunities in international markets. The PLC is a conceptual instrument that offers a way to explain the pricing trends of goods over their time on the market, whether they are products or service products (Meldrum and Mc Donald, [14]). Researchers have established PLC trends from six to seventeen distinct. A growth-slump maturity pattern often typical of small kitchen appliances, the cycle-recycle pattern defines the sales of new drugs and the scalloped pattern, specifically for new product features, uses or consumers, are three common alternative PLC patterns. For the analysis of a product group, a commodity forms a product or a brand, the PLC definition may be used.

To interpret commodity and consumer conditions, the PLC term is better used. The PLC definition lets managers recognize the key marketing problems at each point of the life of a product as a planning strategy and create important alternate marketing tactics. Wasson [30] suggests that fashions end when they reflect a purchasing compromise and customers continue to search for missing attributes. In introducing new products, one of four techniques can be adopted by a company: fast skimming, slow skimming, rapid penetration and slow penetration. A rapid revenue rise could be seen in the development stage. Early adopters enjoy the item and additional buyers begin to purchase it. New rivals are moving in, drawn by the prospects. They add new features for the commodity and broaden distribution. Normally, the maturity period lasts longer than the preceding stages and provides brand management with daunting obstacles. The stage of maturity divides development, stable, and decaying maturity into three stages. Any businesses abandon weaker goods and new products at this point.

Finally, as in the case of oatmeal, the decline might be gradual or quick, as in the case of the Edsel engine. Sales can sink below zero, or at a low stage, they can petrify. Unfortunately, most firms have not built a well-designed policy to handle their ageing goods (Alexander

[1]). This also means that any particular good or service is expected to go from birth to death through multiple stages of its life. According to Tanner and Chonko [26], products have been related to live beings. They are introduced to the market or are born. Then they mature (n sales) and die out at some point.

3 Product life cycle model

The life cycle of the product typically consists of five main stages or phases: product creation, product launch, product growth, product maturity and final decline of the product. These stages exist and are common to all goods or facilities ranging from a certain car manufacture to a multimillion-dollar lithography instrument to a one-cent capacitor. Depending on the product, these stages can be separated into smaller ones and must be addressed as a new product is to be launched into a market since they determine the sales success of the product. The definition also extends to utilities, but the form can vary markedly. Product types and the industry will also be protected. It should be remembered that a strong 'predictor' of product behavior is not inherently the product lifecycle. Instead, it will assist the marketer to understand the market. A product may have gone through a period of exponential growth, for instance, and sales may have started to level off. This does not actually mean that the product is maturing; it may merely be a transient slowing that culminates in the rapid growth of product sales. For living beings, it is possible to get a very shrewd sense of where they are in their life cycle, how long they are going to survive, and as a result of the kind of complications that will arise at any given moment, it is more difficult to obtain this degree of product comprehension. The accompanying map illustrates how sales vary as a product moves through the different phases of its lifecycle.

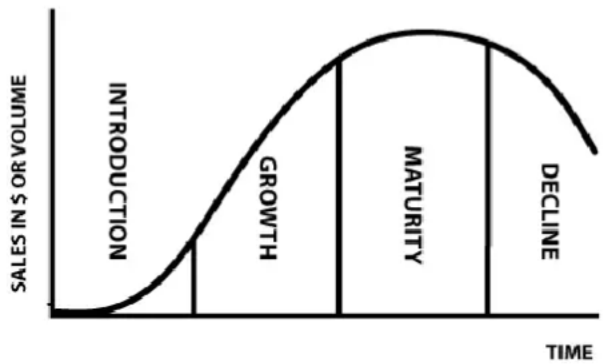

Figure 1 The product lifecycle

A new product takes time to begin selling in volume. There could be production or distribution challenges to deal with. The marketplace may be unfamiliar with the commodity and it takes time to build visibility. As a result, during the launch process, product sales saw sluggish growth. The FMCG changes the price, location (where the commodity is sold) and advertising to satisfy its marketing goals. In wide markets with a high potential for competition, for example, it would make sense to spend heavily in promotion and to start at low rates. This technique would also refer to a commodity for which, with economies of scale, manufacturing costs would decrease rapidly. The FMCG penetrates easily with this approach before rivals have the ability to launch rival products.

A fast rise in sales volume characterizes the growth space. This is generated by increased demand for goods. The FMCG and logistics challenges are likely to be tackled, and the organization is much more conscious of the commodity. Since economies of scale have begun to take place, it should be possible for advertisers to raise advertising efforts. Competition would start to stiffen at the same time and hence the marketer can make appropriate changes to the 4 Ps of marketing. For instance, when adding additional features, it might be necessary to tweak the goods. Competition can be fended off in this manner. In order to draw in more price-sensitive customers, it could even make sense to reduce costs a bit.

The maturity period is defined by the leveling off of market volumes. Competition is high at this stage and margins can continue to suffer. Signs of reaching this stage are that rivals can begin to advertise more strongly or use other promotional means to improve sales.

Finally, product prices are starting to decline, and some serious campaign decisions need to be taken at this stage. By modifying any of the product characteristics, repositioning it or packaging it with other items, it might be possible to prolong the life of a product. On the other hand, eliminating the commodity from your portfolio might make sense.

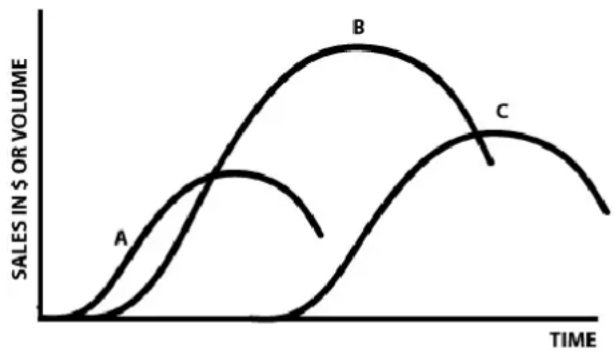

Figure 2 Simple product portfolio for products A, B and C

The philosophy behind product portfolio management is that products will gradually achieve maturity and degrade, ultimately. Although extending their lives and certain products have an incredibly long product life cycle could be necessary, it makes sense to manage a program of constantly adding fresh products. In this way, other product revenues grow as other product revenues level out or decrease. This is seen in Example 2.

3.1 Product development phase

When a company discovers and introduces a new product concept, the FMCG Product Creation process continues. This includes converting numerous pieces of data and transforming them into a new product. Before being exposed to target consumers across test markets, a product typically undergoes many modifications requiring a lot of money and time during production. Those goods that pass the test market are then put in the actual marketplace and the process of product launch begins. Sales are negligible during the product production process and sales are negative. It is a period of investment with almost no return (Gehin, Zwolinski & Brissaud, [8]).

3.2 Introduction phase

The release stage of an FMCG product involves the launch of the product and its specifications to have it released in such a manner that at the time of selling it would have optimum impact. The launch of 'Windows XP' by Microsoft Corporation is a clear example of such a launch. In relation to a product's maturity process, this time can be defined as a money sinkhole. Significant publicity and promotional spending are widespread, and fast yet expensive service standards are added. The FMCG must be willing to expend a lot of money to then get a small amount of it back. In this step, FMCG delivery arrangements are applied. It is really important to get the commodity in both counters and is considered an impossible task. By employing external consultants or contracting the entire logistics arrangement, certain firms escape this stress. This has the advantage of evaluating a major marketing instrument, such as outsourcing. Pricing is something else that a business should remember at this point (Pourhejazy, Sarkis & Zhu, [18]).

Product pricing typically embraces one or two methods that are well-structured. Early buyers will pay a lot for something different, and this will help to mitigate the previously described sinkhole a little. Later, so that the commodity can become competitive, the price strategy can be more stringent. Another solution is a pre-set offer that is assumed to be the best one to boost sales. However, this needs a very clear view of the demand and of what a consumer is likely to pay for a newly launched commodity. An effective stage of product introduction can also benefit from measures taken by the company before the product is released on the market. In the formulation of the campaign plan, these actions are used. Through the use of market analysis, this is done during product growth. To produce a product design, consumer specifications on design, costing, maintenance, and packaging are invaluable. A consumer will tell a business what characteristics are desirable to the product and what characteristics the product does not have. He would illustrate the ways in which the commodity will become useful and beneficial. In this way, a business would know what to expect from consumers and rivals before its product is launched to the market. In the launch process, a brand blend will also help identify the intended demographic during product promotion and advertisement (Sharma, [23]).

3.3 Growth phase

The FMCG's development process provides the pleasure of watching the take-off of the commodity in the industry. In order to concentrate on growing market share, this is the appropriate timing. If the commodity (introduction into a 'virgin' market or an established market) was first launched into the market, so it is reasonably straightforward to obtain market share. A modern rising market alarms attention to the competition. The business must explain all the goods sold and strive to separate them from those of the opponent. A periodic product adjustment process is an efficient strategy to prevent rivals from copying or selling similar products from gaining market share. Licenses and copyrights, design sophistication and poor product part supply are other obstacles.

Promotion and advertisement continue, but not to the degree that it is in the initial stage and is based on market leading tasks and not on raising awareness of the commodity. The use of foreign advertising consultants is a successful strategy. This duration is the time to build efficiencies and increase the availability and service of goods. In winning consumer confidence, cost effectiveness and time-to-market and price and discount strategies are major factors. Throughout the development cycle, sufficient visibility in all markets is a worthy priority. It's necessary to control the growth stage. Companies also bring a lot of time into the manufacturing process, overestimating their marketplace. In predicting consumer demands, reliable forecasts can provide critical insight into the production planning process. After having arranged for relative output capacity, it is futile to raise client demands and commodity demand. The error of over committing must not be made by a firm. This will result in the lack of consumers who will not discover the commodity on their own.

3.4 Maturity phase

The maturity period arrives as the market becomes flooded with versions of the basic product and all rivals are described in terms of an alternate product. In this process, increase in market share is at the detriment of the company of someone else, rather than the growth of the market itself. This period is the period during which the commodity gets the best returns. The most profitable time is for a business that has reached its market share target, while a company that falls below its market share goal would rethink its marketing positioning in the marketplace. New brands are launched during this time even though they clash with the main product of the business and model updates are more regular (product, brand, and model). This is the moment for the life of the commodity to be extended. In regard to competitive policy, pricing and promotion policies are also modified, i.e., pricing goes up and down according to the rivals and discounts and coupons are implemented in the case of consumer goods. In terms of consistency and reliability, marketing and advertisement change from the perspective of getting new consumers to the scope of product differentiation. The distribution battle persists with multi-distribution networks (Moss & Schuiling, [16]).

A successful maturity period of the commodity is extended beyond the timely aspirations of everyone. "Tide" washing powder, which has grown old, is a prime example of this and it is still rising. The decision to pull a product appears to be a difficult process, because before agreeing to take it out of the market, there are a lot of problems to be addressed. Dilemmas such as servicing, supply of replacement parts, response to service competitions in filling the demand void are some of the concerns that raise the difficulty of the decision process to remove a product from the market. For diminishing goods, firms also maintain a high pricing strategy that raises the profit margin and eventually discourages the "few" faithful remaining consumers from purchasing it. Telegraph submission over facsimile or email is such a case. It is often difficult for an enterprise to conceptualize a product's decline signals. A product downturn is usually followed by a drop in retail revenue. It is often difficult to identify its recognition, as marketing teams are typically too enthusiastic due to the great popularity of the product emerging from the maturity level. This is the moment to begin eliminating commodity combinations from the market that are lacking in their market position (Moss & Schuiling, [16]). This must be performed cautiously as it is not always obvious which difference in the commodity brings revenue. Prices must be kept affordable and marketing can be pulled down to a degree that makes the presence of the commodity noticeable and keeps the "loyal" buyer at the same time. It narrows delivery. The basic channel should be kept successful, but it is important to abandon alternate channels. A 0800-telephone line with distribution by a reputable delivery firm, paid by the customer, for example, is worth holding.

4 Market Mix Modeling (MMM)

4.1 Introduction

There was a period when we viewed conventional marketing techniques as an art form, and the victories or shortcomings they create. Marketing efforts lacked clarity for elusive, untraceable effects and were generally perceived as being born out of the imaginative skills of star marketing practitioners.

Big data's arrival changed things; in the 1980s, massive investment in data gathering and research saw a transition in thinking of marketing as at least partially, if not entirely, empirical. The move toward a full-blown science approach has been daunting for many advertisers, however. New technology, difficult algorithms and mathematical applications frequently leave us struggling to keep up with each other.

Just 8-10% of sales revenue goes to marketing operation in the retail sector, leaving CMOs and marketing managers to confront the question of how and where to spend their small marketing budgets. The greatest issue, understandably, is how best to distribute this budget to a wide array of marketing events. In the context of Marketing Mix Modeling (MMM), methodical, systematic planning will help you solve this issue by identifying the optimum mix of marketing variables and proving the return on investment (ROI) that your carefully studied marketing approach offers (Chan & Perry, [5]).

4.2 What is Marketing Mix Modelling (MMM)

The 4 Ps of the marketing mix are familiar to all of us: commodity, price, location, promotion. It is a fundamental part of the philosophy of marketing that considers what variables are needed to succeed in an organization. Well, marketing mix modeling is closely linked to the 4Ps in that it attempts to evaluate how much achievement each element has produced and to predict what potential performance will be achieved by adjusting and optimizing the marketing mix.

Modelling the marketing mix is a predictive approach to assess the efficacy of marketing strategies by breaking down composite data and discriminating between marketing techniques and advertising events inputs and other uncontrollable performance factors. The effects of your marketing mix model study will then notify the composition of potential marketing activities with a degree of confidence, i.e. adjusting input 'a' will impact output 'b' (Banfi et al., [2]).

Benefits of Marketing Mix Modelling

- Enables marketers to prove the ROI of their efforts.
- Returns insights that allow for effective budget allocation.
- Facilitates superior sales trend forecasting.

Limitations of Marketing Mix Modelling

- Lacks the convenience of real-time modern data analytics.
- Critics argue that modern attribution methods are more effective as they consider 1 to 1, individual data.
- Marketing Mix Modelling does not analyse customer experience (CX)

4.2.1 Establish goals

Note, in the first place, the entire justification for MMM is to learn systemic information that can boost your marketing efforts and refine your budget allocation. Yet the corporate priorities need to be transparent and realistic above this.

Describe the main questions that you want to address by the modeling of the marketing mix. Any examples of places to explore and to ask yourself questions can include:

- ***Budget***

Which marketing tactics have the best median return on investments (MROI)?

- ***Media***

Would increasing the TV advertising budget by 15% increase our incremental sales?

- ***Pricing***

What is the impact of a price change on sales and profits?

- ***Competitive***

What promotional strategies for rivals have the most important effect on sales?

During the planning process, the questions you ask the company will then direct the scale and range of your MMM review to help you understand what detail is required to carry out your plans.

4.2.2 Align your organization to understand the data

Modelling the campaign mix allows you to gather a vast volume of data within your organization from many various fields. You would need to involve the gatekeepers of each data set to do this, define obligations, and create a data collection timeline.

Any number of these figures will likely need to be engaged:

- CMO (Chief Marketing Officer)
- TV and media agency partners
- Marketing agency partners
- CRM manager
- Marketing executives

4.2.3 Identify relevant data

Your business would have internal archives of data; this is where it is convenient to view and evaluate your business and consumer data for analysis purposes. A critical aspect for MMM is the consistency of your information; data that is reliable, tidy, and logically processed can save you time and effort when it comes to repurposing your analysis. You should enlist the assistance of any colleague responsible for handling corporate data repositories and resources at this level. (Banfi et al., [2]).

4.2.4 Understand your access to data, including any limitations

Build a comprehensive list of the information you possess and intend to participate in the analysis; attempt to obtain as much data as possible, including any fees or subscriptions necessary to view data from third parties. You may still need to pay for any disruptions involved with offline data usage.

4.3 Nielsen points to 4 stages of the MMM process

4.3.1 Collect

Econometric techniques are used to predict product demand generated by marketing tactics within the compilation stage of marketing mix modeling by dividing product revenue into 2 categories of sales drivers:

1. Incremental drivers

There are the controllable components that the marketing staff introduces. Incremental drivers are running on a short-term basis; information is gathered on week-to-week transactions that differ based on the sales.:

- Above-the-line operation in the newspapers (TV, print ads, digital ads, promotions, and discounts, etc.)
- Factors beyond the line (temporary sales rates, sales incentives, coupons, social media, campaigns for direct mail marketing, in-store marketing, meetings, and conferences).)

2. Base drivers

Sales that are produced in the absence of any incremental marketing effort are the essential outcome for a business. Basic findings are also the result of brand value and credibility, such as customer satisfaction, which has been established over many years (Cain, [3]).

Base-drivers are the following components:

- Price: The price of a product is an important fundamental driver of a marketing mix since the price dictates both the customer category at which a product is marketed and the advertisements that are offered to the selected demographic to promote the product.

- Distribution: The number of outlets in the shop, the product inside those locations, and the shelf life of that stock are all known as the marketing mix's simple drivers. Store sites and the inventory inside them are static and can be identified without any marketing intervention by consumers.
- Seasonality: In a market year, such fluctuations arise frequently and can thus be counted on to drive revenue with a degree of predictability. For the business, seasonal sales, such as the winter holiday time, are big factors. The eCommerce business, for example, expanded 16.7 percent in 2018, reaching up to $123.90 billion thanks to the holiday shopping spree.
- Macroeconomic variables: The analysis of how the overall economy and markets work is macroeconomics. The effects of problems such as inflation, gross domestic product (GDP), unemployment, etc. was considered. Macroeconomic conditions may have a huge effect on base sales when it comes to MMM, such as a rise in unemployment, rates will lower the purchasing power of consumers and, thus, sales will decrease.

4.3.2 Model

P.M Cain [3] calls the "logical choice" for marketing combination modeling ventures the Time Series analysis (regression model). If you're scratching your head right now, don't hesitate. From MathWorks, here is a short definition:

Time series regression is a predictive approach focused on the reaction history (known as autoregressive dynamics) and the conversion of dynamics from appropriate predictors to predict a potential response. Regression of time series will help you understand and forecast the action from experimental or empirical evidence of complex systems. Regression of time series is widely used to model and predict economic, financial, and biological processes. (Usta et al., [28]).

1. Analyze

The outputs of your selected model will be analyzed in the research stage; these outputs will come in the form of sales decomposition, which with each modelled strategy breaks down the results into volume.

When evaluating the decomposition of revenue, there are 3 important metrics.:

- Effectiveness
- Efficiency
- Median Return of Investment (MROI)

2. Optimize

In essence, this final stage of MMM lets you transform your outputs into inputs - full circle style; suggesting, to refine the content mix for future promotions, you use the findings of your research.

A "what if" simulation will provide part of the optimization. Your marketing model's findings are calculations that indicate the relationship between marketing practices and sales outcomes. With these calculations, if adjustments are made to the marketing mix, you will foresee what will happen.

Marketing Mix Modeling (MMM) has been an important technique in recent decades to help enterprises maximize the distribution of budgets to different forms of advertising, such as internet networks, television, etc. (Chan & Perry, [5]).

MMM is a tool that aims to measure the effect on revenue or market share of many marketing sources. The goal of using MMM is to consider how much each marketing input contributes to revenue and how much each marketing input needs to be spent on. It is often used to refine the budget for expenditures on these multiple marketing inputs.

In terms of Return on Investment, MMM helps to determine the efficiency of each marketing input. In other words, a higher investment return (ROI) marketing input is more efficient as a tool than a lower ROI marketing input.

MMM leverages the methodology of Regression and then uses the research done to derive key knowledge and observations. Let's speak about the different principles and functional implementations related to MMMM comprehension (Chan & Perry, [5]).

4.4 Multi-Linear Regression

MMM uses the Multi-Linear Regression concept, as described above. The dependent variable may be revenue or market share, while price, TV spending, delivery, discounts, and coupons etc. are the independent variables widely used. Nowadays, to raise brand recognition, the new medium is extensively used. Inputs such as digital spending, website visits (amount of clicks, number of impressions, number of product page views, etc.) may also be used as inputs for MMMM if available (Milani & Lourenço, [15]).

Between the dependent variables and predictors, an equation is formed. Depending on the relationship between the dependent variable and varied advertisement inputs, this equation could be linear or non-linear. There are some factors which have a non-linear relationship with revenue, such as TV ads. This means that the rise in TV GRP is not directly commensurate with the increase in revenue.

TV GRP is called a non-linear variable since an advertisement can only produce interest among consumers to a degree, according to advertisers. Increased exposure to ads will not produce any more incremental recognition among clients after a certain stage, since they are already aware of the brand.

The betas created from the study of Regression assist in quantifying the effect of each data. The beta basically indicates that one unit increase in the input value will increase Beta units' sales/profit, leaving the other marketing inputs stable.

4.5 Linear and Non-linear Impact of predictors

Any predictors display a linear relationship with revenue, i.e., sales will continue to grow as we increase these predictors. But predictors such as Gross Rating Point (GRP) for TV do not have a linear sales effect. Only to a certain degree will an increase in TV GRPs increase sales.

Each incremental unit of GRP will have less effect on revenue until the saturation point is reached. Therefore, on such non-linear variables, certain transformations are made to use them in linear models.

The proportion of the audience targeted by ads is raised by increasing the volume of advertising, thereby increasing demand, but a linear rise in advertising penetration does not have a direct linear impact on demand. Usually, each incremental volume of advertisement allows the rise of demand to have an increasingly lower effect (Kumar, Shankar & Aljohani, [13]).

5 Data science in retail

5.1 Introduction

As of late as there has been a paradigm shift over to consumers, the shopping experience has shifted dramatically. Customers can locate and examine products from a range of devices without much of a stretch, often when strolling around a shop. Via online networking, they will share their insights on retailers and brands and affect potential prospective consumers. Retailers need to introduce modern and innovative methodologies to attract and maintain customers in order to succeed in this current multi-channel environment. Although this means researching customer preferences, it also involves evolving demands for the individual product being sold.

Input from a retail company about how it treats its clients and/or how the consumer responds to its service represents the role of the company in cultivating its customer base. However, historical evidence is not only an indication of how poor or well the company is doing, it can also offer a very valuable perspective into what can be done, based on customer behavior, to increase its profits. The studies summarizing overall behavior may not have the requisite valuable insights to assess how particular clients are likely to behave because patterns in general behavior are just too large. It is also important to integrate even more nuanced methods of labelling to determine which category to cater to (Sakoda, G., Takayasu & Takayasu, [22]).

It needs more than summary reports in order for retailers to build a substantive dialogue with consumers that respects the desired level and mode of interaction of the shopper, which is why consumer insight and predictive analytics have the potential to transform the retail marketing industry dramatically. Therefore, customer intelligence modeling is a simulative discipline in cognitive science and data science and is thus a process of evaluating and offering data-driven insights into the past and thus forecasting market behavior with the assistance of the organization concerned (Sheth & Kellstadt, [24]).

The most important consumer intelligence reflects on how and when clients are going to make future decisions and needs quantitative and qualitative analysis to be combined. A mixture of exchange-driven data and human-derived perspectives, which can only come from a rich conversation with clients, offers the most accurate and actionable image of the client. A simplistic vision of the consumers could be accessible to decision-makers who focus on quantitative data points alone. Yet decision-makers who have a rich dialogue with their clients are well prepared to consider and make informed decisions about transaction data. Customer analytics must also incorporate both raw transactional and behavioural data in order to produce deviated steps from the current insights of the organization. Therefore, the mechanism designed to provide additional insight into the purpose of the customer is not only about a vast volume of collected data, but also about the data produced (Kim, [11]).

In a way that changes market operations today, the position of data science and artificial intelligence is growing its footprint. In addition to other sectors, such as BFSI (Banking, Financial Services and Insurance) and others, retail and CPG (Consumer Packaged Goods) are perhaps one of the most active industries in adopting these technical advances early in the game. Most market experts point out that over three-fourths of companies expect to implement data science and AI in retail by 2021. From the point of view of retailers, the emphasis is obviously on a) optimizing consumer service by personalization and so-called hyper personalization, b) increasing productivity and optimization to minimize costs and c) maximizing the value added as much as possible through automation. These are, of course, guided by the "mindset", "attitude" and "ability" to go in detail to achieve the strategic path.

5.2 Key challenges today

There could be lots of different capacities in which Data Science could play a part in driving success if we start looking at retailers and jot down their issue statements or market goals. For starters, if we look at some questions that customers need for a smarter answer, it is as follows: how to boost customer experience; how to better learn the behaviour of customers; how to optimize costs, maximize productivity in most operations that are doing well and that are not doing well; how to provide better value to customers today relative to what they have earlier received; how do we innovate on products and services so as to increase the acceptance, visibility and sales and so on.

According to analyst reports, approximately one-fifth of organizations have deployed data science and AI technologies in manufacturing, indicating that there is around 70 - 80 percent opportunity where there can be scope for improvisation of current solutions in design phases, there can be scope for new solutions to be applied, there can be scope for new techniques to be found and prepared for path towards a robust solution (Provost & Fawcett, [19]).

7 Major application themes: Some of the top usage of Data and AI in Retail in next few quarters could continue to be the following:

1. 1. Personalization Hyper (Product personalization for users, Service personalization specific to a user, a set of users etc.)
2. 2. Customer reviews (e.g. NBOs — Next Best Deals etc.)
3. 3. Automated warehouse-picking machines or robots (Automation in tasks for picking and movement of goods)
4. 4. Visual search using image recognition (making it easy for clients to access what they are searching for) and online search monitoring.
5. 5. Detection of Fraud (always simple to think but complex to handle)
6. 6. Effective pricing policy (e.g. seasonal pricing, incentive pricing or event-driven pricing etc).
7. Virtual reality or augmented reality

In specific, in merchandising, brand strategy, CRM, etc., consumer insights and marketing analytics will continue to play an important role. These can concentrate on real-time hyper-personalization of buyers, financial and assortment-based merchandise preparation, inventory optimization, market forecasting, advertising effectiveness and its effects on sales, etc (Wedel & Kannan, [31]).

Focus for Business Impact: Most strategies and actions around these strategies create impact by:

- Improving the service of customers,
- To respond quickly on decisions based on the monitoring of consumer actions,
- Increase of traffic on websites, department store traffic or footfall,
- Improving reaction time for promotions,
- Optimizing store inventory by estimating criteria,
- Improving the quality of staff production by putting automation into operation for very particular tasks.

Approach: The "POET" literary approach will apply poetry to science in order to help push performance. Here's how data science's position would probably become crucial for retailers:

P: Practicality and Productionization: viability, data approach, how and how to implement the system given data quality and data usefulness, etc.

O: Operationalization: surveillance, recording, carrying out several tests to decide and test which experiment would be the one to use for the solution to the problem for which we are solving it.

E: Expandability-to offer clarity, very necessary. It is important to clarify what is happening from step 1 to step N and it can be communicated to relevant stakeholders. The more we log and explain measures in EDA, feature engineering, tests, end user assessment metrics, the easier it is.

T: Clarity, confidence, what is being done and how it is being done, reveal the criteria produce an effect on the prediction target, which observations are expressed, which methodology is considered, and why, depending on what experiment and indicators are around it. To a greater degree, the recording of risk, compliance, data sensitivity aspects will benefit end users.

5.3 Consumer Behavior Modelling

Consider a supplier in a certifiable situation who would wish to better alert high-esteemed, steadfast clients who send off an appearance of withdrawing from the brand. With seven days, a prescient model worked through historical data will know the clients are responsible for shopping again, encouraging the store to give them a chance to be the dedicated consumers they actually are. If such consumers are unrealistic to order within seven days but have a strong usual request esteem, the prescient model will also emerge (Sibalija & Majstorovic, [25]). The retailer should provide a driving power for these consumers to take the customers back to the brand. In any case, it is important to predict what consumers are inclined to do to see how best to finish the conversation with them. Retailers should use promotional options leveraging bits of information gained from consumer experience and prescient examination to move forward. The knowledge community of a retailer must procure components from all areas of the organization, including retail experts, data nerds and data scientists. When we press deeper into the age of big data, these main elements will set retailers up for achievement. Consequently, the key is collaboration with domain experts. A three-legged stool resembles this organized effort. Each leg is important to staying stable and fulfilling the expected purpose for the stool. The three legs of the stool are retail experts, coders, and prescient modelers or computer scientists with respect to gaining consumer knowledge (Roberts & Lilien, [21]).

5.3.1 The Retail Experts

Retail specialists have sufficient domain knowledge, and the problem customer insight is intended to be explained best. They propose established characteristics that offer consistency to both the brand and the promotional activity of the company (Casas et al., [4]).

5.3.2 Coders

Knowledge nerds are required to configure these thoughts and store them in a suitable archive that can frequently prompt the merchant with exceedingly extended information management needs. If the data must be used to make plans or decide on key promotional choices in the event that it is properly packed away and got to, be it as it can. Data out of control means worthless information and a squandered chance.

5.3.3 The Data Scientists

The processed information is then supposed to be used by prescient modelers and data scientists to create models that meet certain market objectives initially set by the retail master. Prescient models identify relations between noteworthy data and consequent outcomes such that it is possible to predict close-term and long-haul consumer conduct. This leg of the stool is meant to solve concerns, such as the likelihood of whether a customer will make their next purchase and what the estimate of the purchase will be. These relations are complicated now and again, to the extent where machine learning algorithms alone can uncover them (Kim et al., [12]).

Data Science aims to enhance the understanding of consumers. Cognitive modeling of patterns in consumer behaviour leads to customer input and industry analyses in the sentiment analysis of social media streams, call center records, product ratings, etc. It also helps to come up with personalised purchase reference reviews. These predictive analytics help to develop a strategic framework to boost consumer service.

Merchandizing-Data science, using heat sensors and image processing to detect behavioural trends, aims to create enhanced layouts, advertising shows and product placements. With the aid of video data processing, some of the other merchandizing approaches include recognizing retail habits, comparing trends and cross-selling opportunities. From a blend of internal and external data (e.g. economic predictions, weather and traffic statistics, holiday and seasonal trends), it also allows to collect higher everyday income and quicker business increases through comprehensive market basket research.

Marketing - Deploy real-time pricing using "second by second" analytics to help location-based and customized deals on mobile devices (e.g. supply chain and inventory data, competitor pricing, market and consumer behavior data). Driven campaigns can also be carried out using algorithms to identify customers and determine the most suitable platforms.

Customer category research also does this. Online behavioral profiling and web analytics will help build personalized deals that can definitely target a certain group of individuals.

Supply Chain Logistics- Using GPS-enabled Big Data Telematics, data science can help to refine routes and make transport more efficient. Based on in-store data, it may also aid in more successful supplier agreements. With the help of structured and unstructured data, demand-driven forecasts can be made (Pattnaik & Behera, [17]).

5.3.4 The adoption strategy for big data

The adoption strategy for big data consists of the following steps.

Strategy - Many retailers find the need to assess their scalability in regard to their own knowledge situation, when the capacities of big data are blooming. A consistent plan is therefore needed to identify a path map for the implementation of big data in the enterprise. It will help to quantify and create a business case for the uncertainties.

Pilot- To define small business divisions or divisions that can be used to do a pilot test, retail businesses should be made. The team in charge should define solutions, priorities, figure out main success metrics (KPIs) that can be conveniently assessed in relation to the enterprise and have simple, understandable ideas.

Adoption of Large Scale- After the pilot test gives acceptable results, the organization will continue to use big data analytics through the company's multiple avenues. Using a top-down approach to build understanding of the importance of big data analytics is more useful in order to guide corporate policies.

Managing Development- It is crucial to keep a good check that the execution of the solution is well handled such that the gains can be spread across all divisions of business.

Big data roadmap and adoption was initially motivated by cost savings, but it helps to take a more strategic view that seeks to provide the enterprise topline with benefit.

Control of large data volumes from heterogeneous sources.

5.3.5 Poor information quality

Poor information quality effects each of these ranges in an unexpected way.

Deliverability

Advertisers work to increase deliverability in order to generate emails displaying rates of improvement. Organizations typically attribute 32 percent of sales to the email platform. Be that as it may, in the last 12 months, 66 percent of advertisers have faced problems with email deliverability, mostly attributable to off-base details. Basically, advertisers do not communicate their painstakingly made messages to consumers without exact details (Purcarea, [20]).

Single client perspective

For advertisers, keeping a solitary consumer view was more critical than any other division in the report. A crucial move in retaining transparent data that can be used for information-driven understanding is to integrate heterogeneous data into a lone client perspective. A solitary consumer viewpoint is ultimately used to gain market knowledge and emphasis on highlighting deals, as the study indicates, something that many advertisers expect to affect in an inexorably digital world.

Upgrading understanding and dedication

For a clear end target, advertisers work to develop awareness or comprehension to improve consumer experience and sales for growth. As information occupations broaden and advertisers turn out to be more modern with information management, it is common for this field to begin to become more pervasive.

Educating senior executives and ensuring that intelligence programs are guided by well-defined market needs and the best possible integration of quantitative and qualitative methods. The retail sector is facing a big opportunity to consider its clients and satisfy them. To allow productive targeted marketing policies, product development and supply chain planning, they need to coordinate the huge variety and volume of unstructured customer-oriented data. Big data uses smart analytics to identify consumer expectations and monitor patterns in a cost-effective way.

6. Dataset and data exploration

The exploratory data analysis was conducted in R (version 4.1.2) using the *dplyr* (version 1.0.7), *plotly* (version 4.10.0) and *corrplot* (version 0.92) libraries.

6.1 Dataset

The given dataset consists of twelve (12) variables:

1. Week: number of week that corresponds to the values of the other variables
2. Volume of sales: the volume of sales, for the given week, of our product in national level
3. Actual price: the average national price that our product was sold, after any promotion applied, for the given week
4. Base price: the average national price of our product, before any promotion applied, for the given week
5. Competitor 1 price: the average national price that the first competitor product was sold, for the given week
6. Competitor 2 price: the average national price that the second competitor product was sold, for the given week
7. Competitor 3 price: the average national price that the third competitor product was sold, for the given week
8. Competitor 4 price: the average national price that the fourth competitor product was sold, for the given week
9. Competitor 5 price: the average national price that the fifth competitor product was sold, for the given week
10. Competitor 6 price: the average national price that the sixth competitor product was sold, for the given week
11. Competitor 7 price: the average national price that the seventh competitor product was sold, for the given week
12. Promo %: the average national percentage of discount, for our product, that was applied on the base price in form of promotion, for the given week

The dataset consists of 146 rows, each one representing the observed data of one week, starting from 1 to 146.

In table 1, using the *summary* function from R, there is a depiction of the variables' minimum value, maximum value, mean, median, first quantile and third quantile.

Table 1 Key statistics of variables

	Min	**1st Qu**	**Median**	**Mean**	**3rd Qu**	**Max**
Volume of sales	36840	60561	72678	90317	117132	219624
Actual price	0.758	0.907	0.9915	0.9967	1.0953	1.222
Base price	0.995	1.09	1.129	1.12	1.154	1.222

Promo %	0	3	11	13.47	22	41
Competitor 1 price	0.539	0.644	0.67	0.6716	0.7005	0.755
Competitor 2 price	0	0	0	0.1715	0.5015	0.721
Competitor 3 price	0	0	0	0.5123	1.2265	1.414
Competitor 4 price	0	0	0	0.2389	0.52	0.618
Competitor 5 price	0	0	0	0.3524	0.7498	0.964
Competitor 6 price	0	0	0	0.09649	0	0.705
Competitor 7 price	0	0	0	0.1208	0	0.89

Based on the output of summary, we can observe the presence of zero values for the variables of Competitor 2 price, Competitor 3 price, Competitor 4 price, Competitor 5 price, Competitor 6 price and Competitor 7 price. More specifically, based on the median value the Competitor 2 price, Competitor 3 price, Competitor 4 price and Competitor 5 price variables have at least 50 % of their observations with zero value and based on the third quantile value the Competitor 6 price and Competitor 7 price variables have at least 75 % of their observations with zero value.

The Figures 3 - 13 of the variables verifies the above.

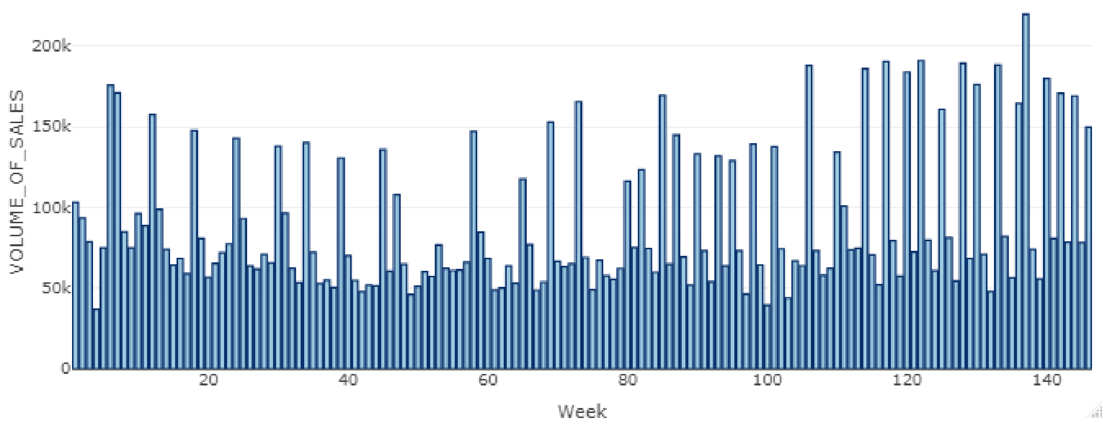

Figure 3 Volume of sales per week

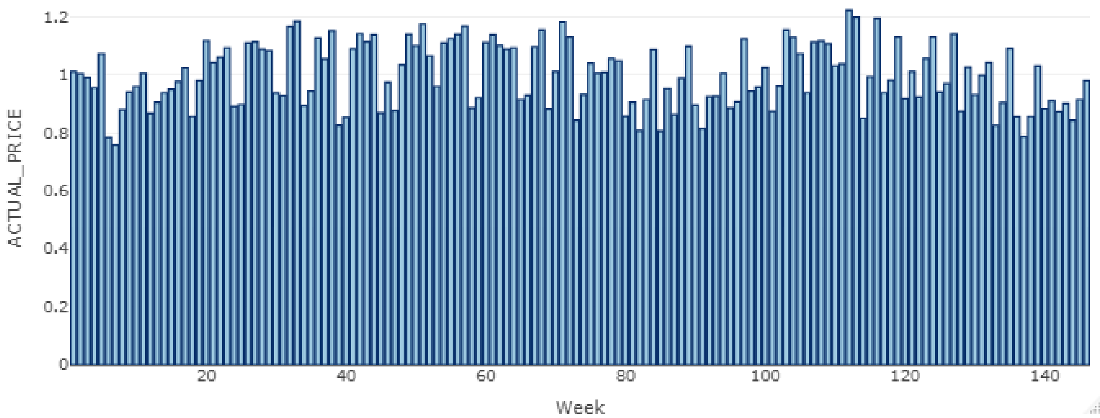

Figure 4 Actual price of our product per week

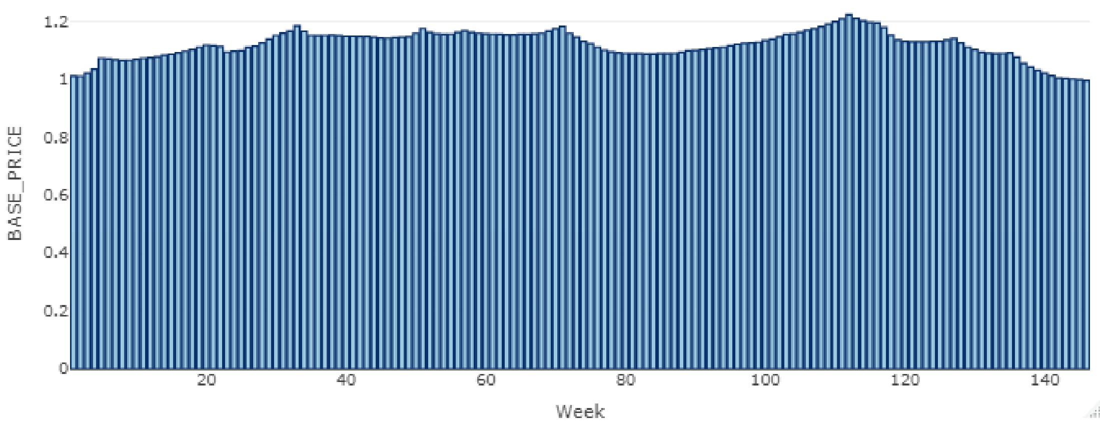

Figure 5 Base price of our product per week

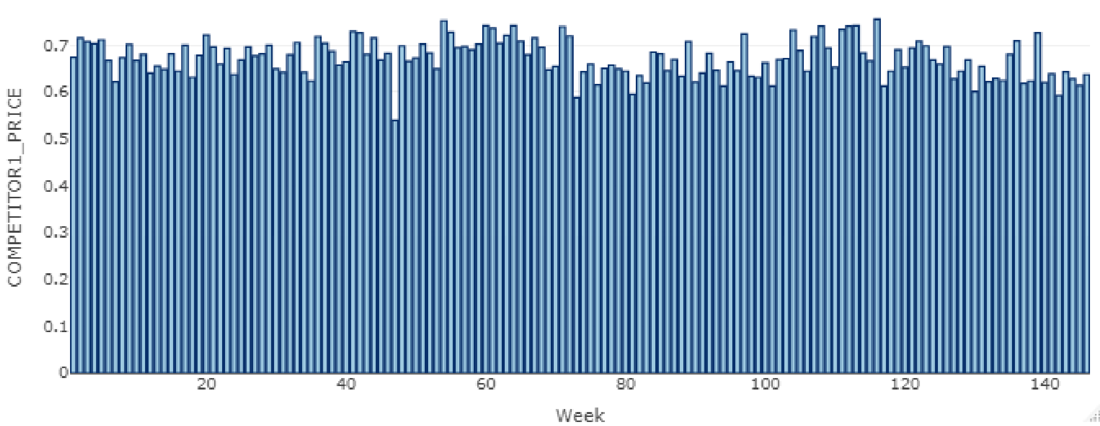

Figure 6 Price of the first competitor product per week

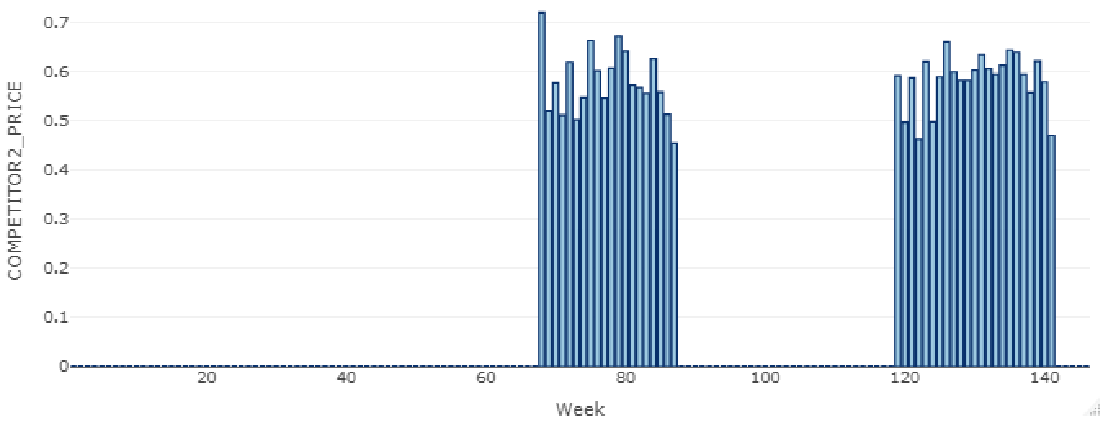

Figure 7 Price of the second competitor product per week

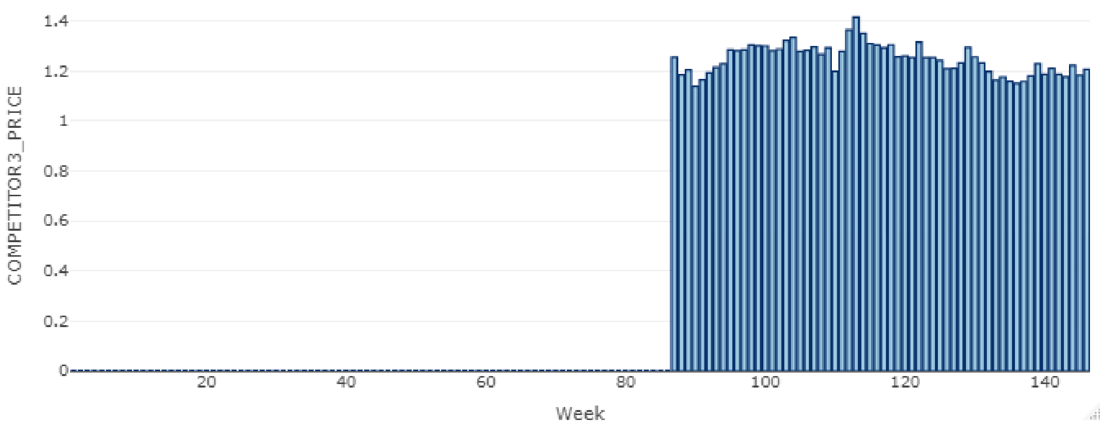

Figure 8 Price of the third competitor product per week

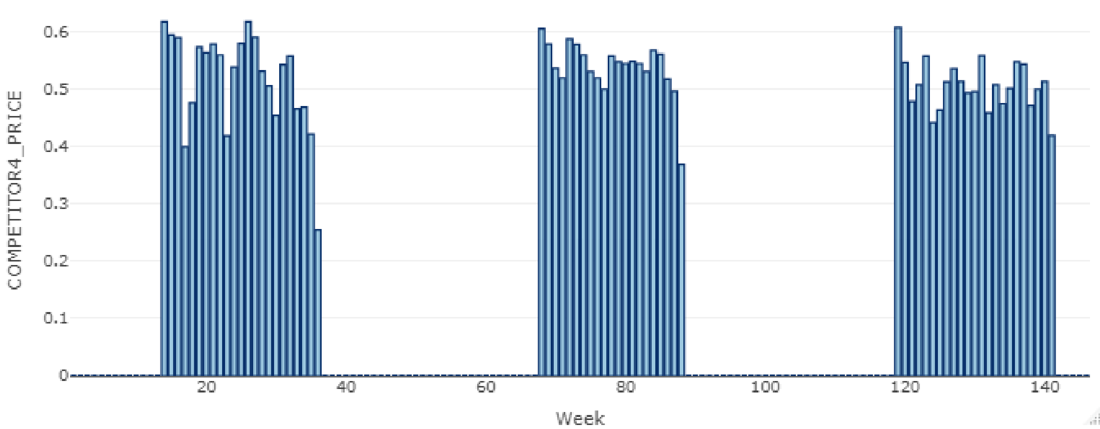

Figure 9 Price of the fourth competitor product per week

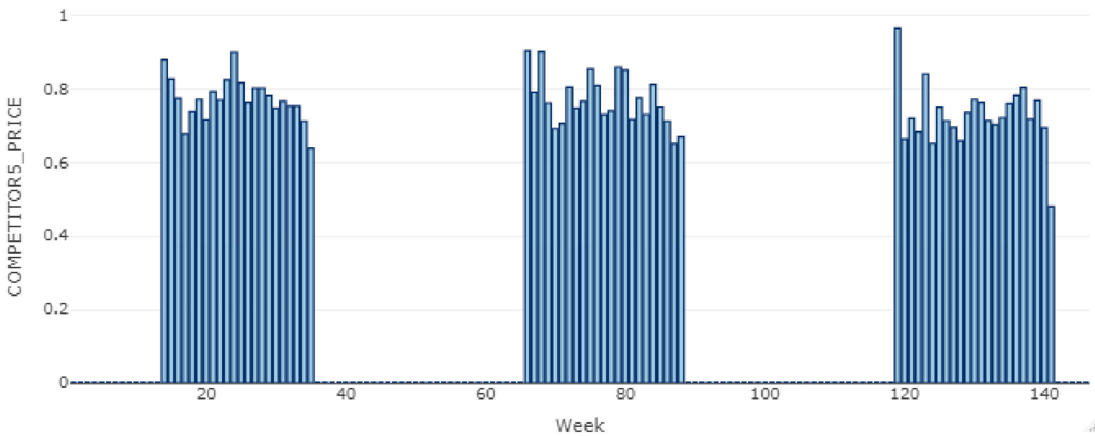

Figure 10 Price of the fifth competitor product per week

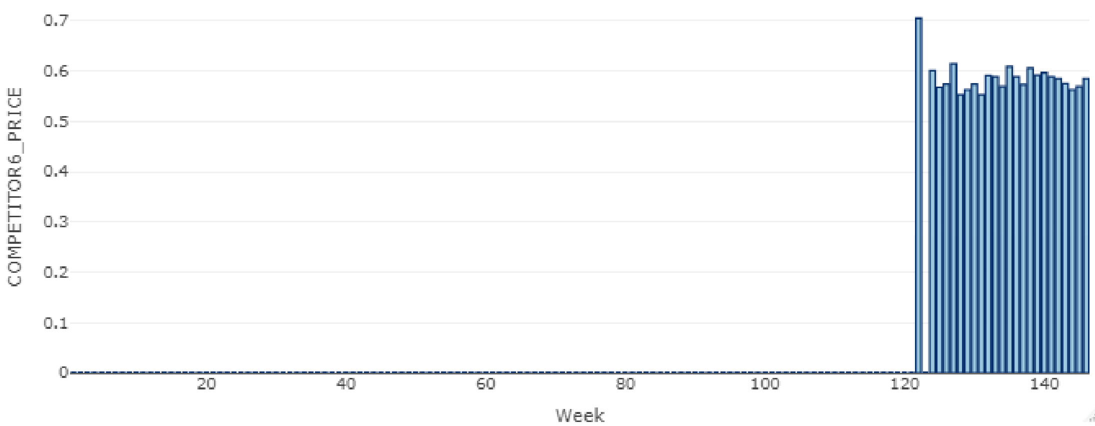

Figure 11 Price of the sixth competitor product per week

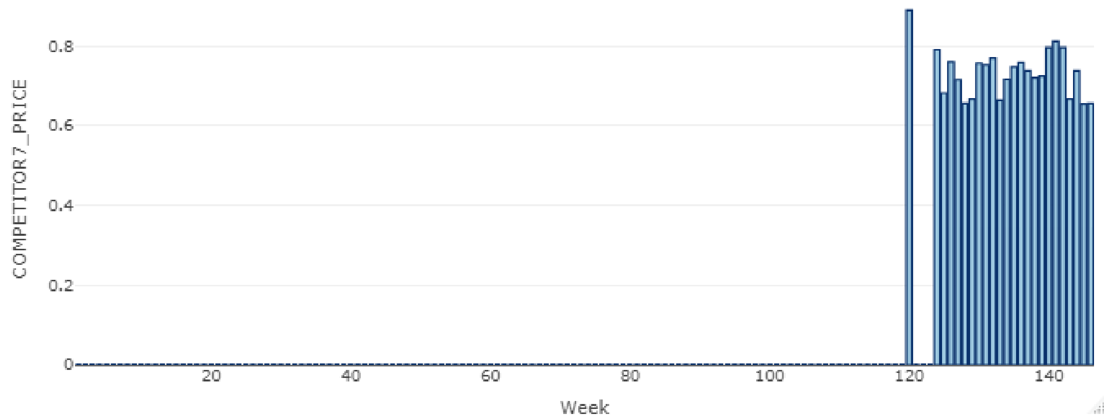

Figure 12 Price of the seventh competitor product per week

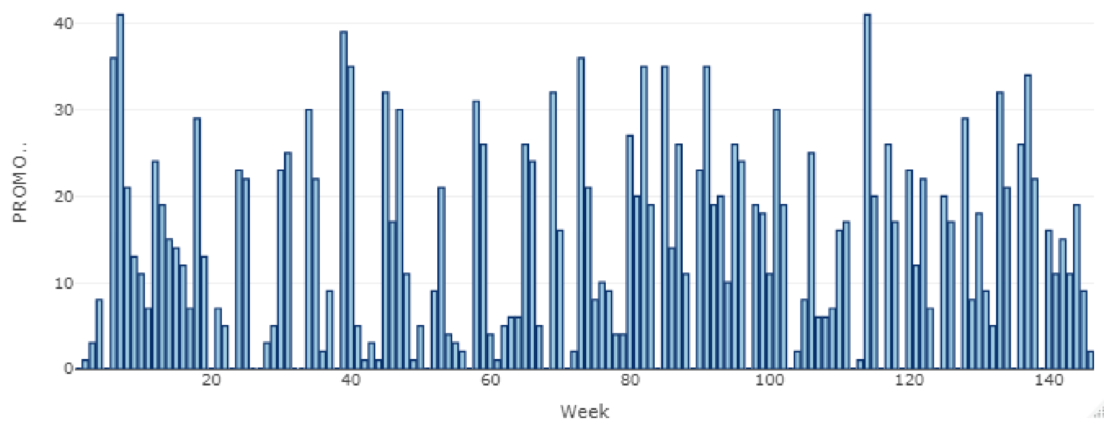

Figure 13 Percentage of base price discount applied in our product per week

6.2 Feature Engineering

At this point we should explore possible new variables that could enhance the models' fitting and predictability and transformations of already existing predictors.

6.2.1 After promo variable

Additionally to the twelve features available, we should also consider including an extra, "after promo", variable. This is a dummy variable which marks weeks with no active promotion that comes right after a week with promotion. So, this dummy indicates a non-promo week right after a promo week with 1 and 0 otherwise. The expected benefit by this variable's addition is to capture a negative effect of promotions, through the drop of sales in the after-promo period.

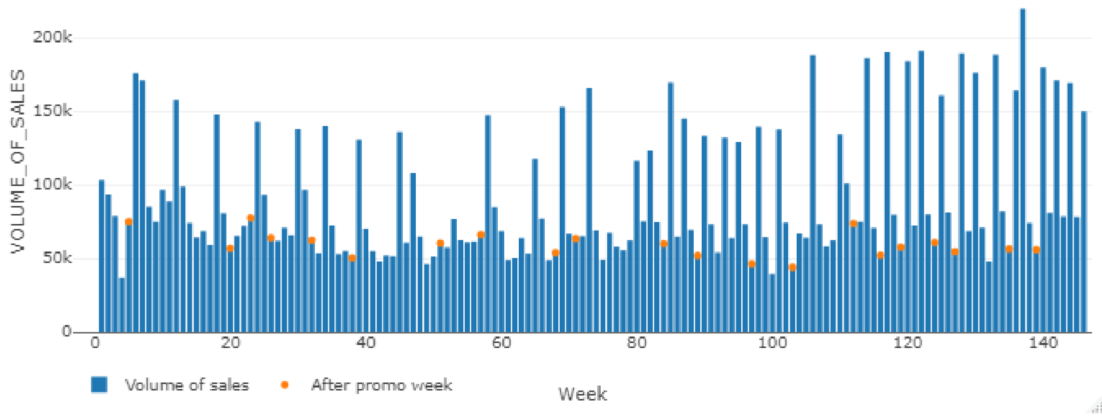

Figure 14 Construction and depiction of after promo week variable

6.2.2 Bank holidays

Marking a week as a bank holiday, if it includes within its days a bank holiday, could explain a gain in sales, due to this condition. Although a helpful variable, it cannot be constructed in this dataset, because the origin country of these data is unknown (different bank holidays in different countries) and the variable representing the weeks is not in date format (e.g., 2019-11-14) but in numeric format (e.g., 1, 2…), indicating only the order.

6.2.3 Price variables

Since, in the dataset, both actual and base price are present, as well as the promotion percentage variable, for the examined product's sales, a logical question is whether to use all of them or some of them. As a first step, a new variable, PRICE_BY_BASE_AND_PROMO, is constructed through base price and promo.

$$x_i = BasePrice_i - BasePrice_i * \frac{Promo_i}{100}$$

Secondly, a correlation between actual price and the newly constructed variable needs to be calculated. A linear relationship between variables could explain the choice of Pearson's coefficient whereas the Spearman's correlation works fine with a monotonic relationship as well. To examine the relationship between these variables a scatter plot is required.

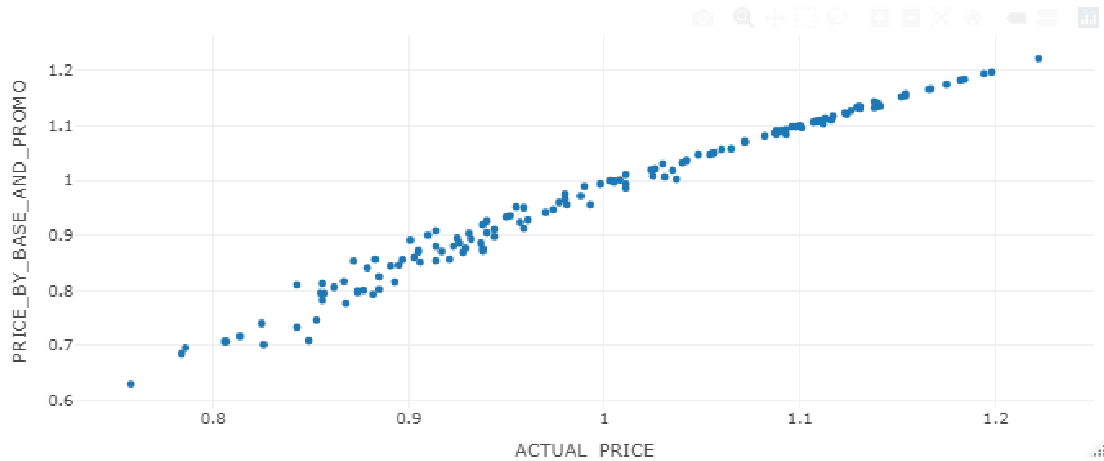

Figure 15 Relationship between actual price and PRICE_BY_BASE_AND_PROMO

Figure 15 indicates a linear relationship, so, a correlation matrix, among those variables, will be constructed, using the *cor* function for calculation, with the Pearson's coefficient as the chosen method and *corrplot* library for visualization.

Figure 16 Correlation among price related variables

In Figure 16, on the upper diagonal half there are the calculated correlation values, while on the bottom diagonal half there is a depiction of those values, based on intensity and direction (bigger circle indicates higher correlation coefficient and blue indicates positive correlation while red indicates negative). Based on Graph 14, all those variables should not be included in one model simultaneously, although, combinations of those could be examined in different models.

6.2.4 Volume of sales

Typically, normally distributed variables tend to have better behavior in a linear regression model. The examination of the volume of sales, which will be used as the depended variable, is very important. A Shapiro – Wilk test of normality can identify if a sample is normally distributed. The null hypothesis of this test is that the population follows a normal distribution. Using the R function *shapiro.test*, the p-value is 7.852e-12. So, for a confidence level of 95% (a = 0.05), the null hypothesis of normality is rejected (7.852e-12 < 0.05).

A histogram of the volume of sales could give us an indication of the distribution of this variable. Using the *hist* function we can get the histogram of volume of sales.

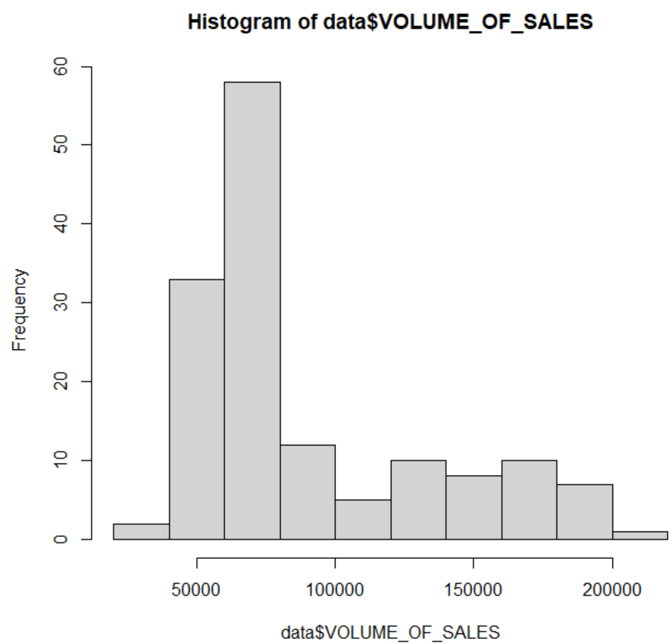

Figure 17 Histogram of volume of sales

Based on Figure 17, the distribution of the sales' volume could be a log-normal. Thus, a logarithmic transformation is a reasonable one for this variable and should be examined.

We should also check if the lagged values of the volume of sales are important. To do so, the *pacf* function will be used on the data that will be used to train the models. The training set consists of 104 weeks e.g. 2 years (~ 70% of the whole time series).

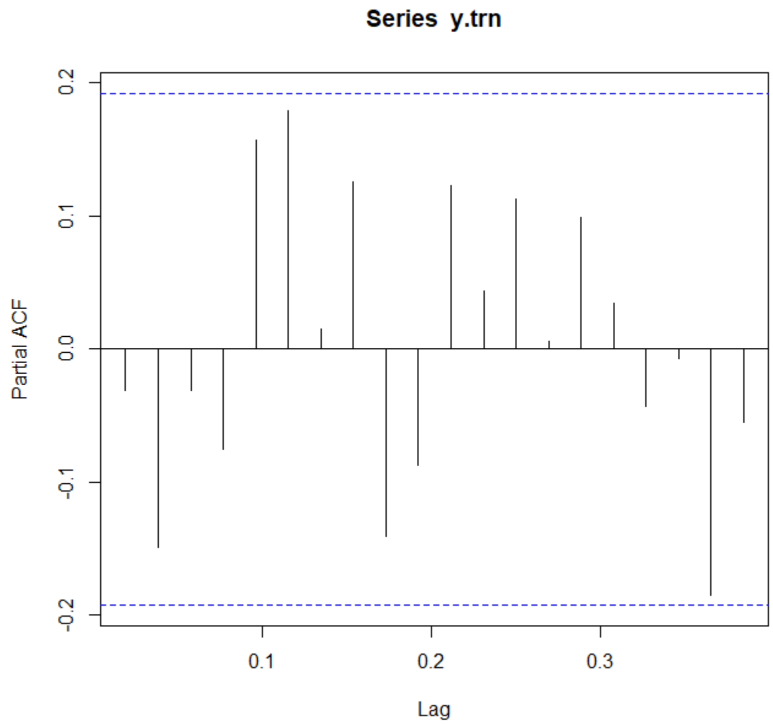

Figure 18 Important lag values of volume of sales

Figure 18 informs us that no lags are significant.

6.2.5 Handle zero-valued observations

Another issue that needs to be addressed is the presence of zero values in the second, third, fourth, fifth, sixth and seventh competitor price variables. One way to handle this is through a simulation of those prices.

Suppose that $X^{(i)}$ and $Y^{(i)}$ are the *price* and *demand* of a product at time i, $X_{var}^{(i)}$ is the variability of price at time i around a constant average price $\underline{X}$, and β is the price elasticity of demand.

For a level-level demand price model, we have:

$$X^{(i)} = \underline{X} + X_{var}^{(i)}, \qquad X_{var}^{(i)} \in \{\underline{X}_{min}, \underline{X}_{max}\} \qquad [1]$$

$$Y^{(i)} = \beta X^{(i)} + e, \qquad e \sim N(0, \sigma_e^2) \qquad [2]$$

where σ_e^2 is the constant variance of the residuals e.

Combining expressions [1] and [2] we get:

$[2, 1] \rightarrow$

$$X^{(i)} = \frac{Y^{(i)} - e}{\beta}$$

$$\underline{X} + X_{var}^{(i)} = \frac{Y^{(i)} - e}{\beta}$$

$$X_{var}^{(i)} = \frac{Y^{(i)} - e}{\beta} - \underline{X},$$

which leads up to the final simulation model for $X_{var}^{(i)}$:

$$X_{var}^{(i)} = \frac{Y^{(i)} - N(0, \sigma_e^2)}{\beta} - \underline{X}$$

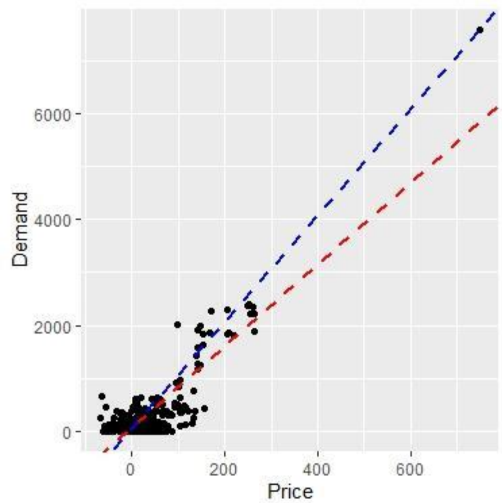

Figure 19 Sales vs. simulated prices with regression slope (red) and theoretical slope (blue)

Although a highly sophisticated way to fill the zero values of the price variables, in this dataset we are missing the information of sales.

An alternative way to fill these zero values would be based on their distribution. A Shapiro – Wilk test on those variables can indicate if a normal distribution is followed, for an a = 0.05.

Table 2 Shapiro – Wilk test on variables with zero values

Variable	Shapiro - Wilk p-value	Null hypothesis of normality
Competitor 2 price	0.5238	not rejected
Competitor 3 price	0.2247	not rejected
Competitor 4 price	0.0002266	rejected
Competitor 5 price	0.04327	rejected
Competitor 6 price	4.88E-05	rejected
Competitor 7 price	0.1504	not rejected

Based on Table 2, Competitor 2 price, Competitor 3 price and Competitor 7 price variables are not rejected for the hypothesis of normality, thus an alternative way to replace the zero values would be by filling these observations with values produced by a normal distribution with mean and standard deviation the means and standard deviations of these variables, using the function *rnorm*.

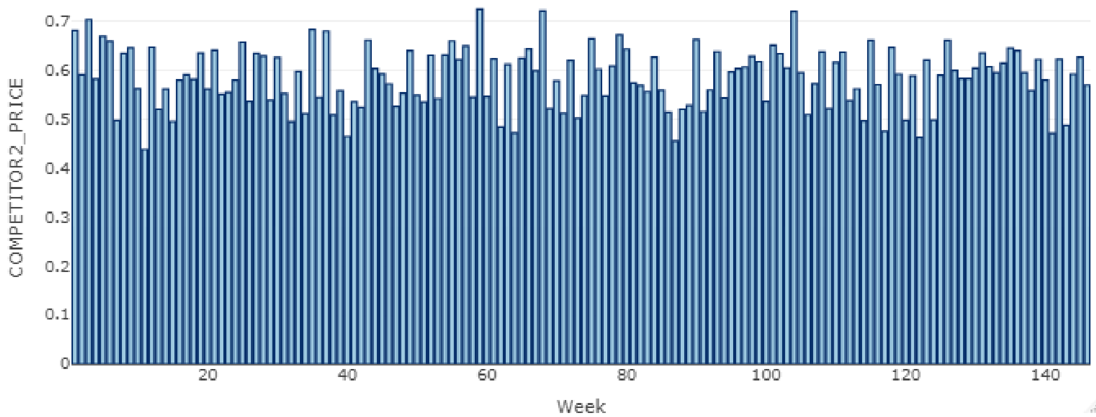

Figure 20 Competitor 2 price after filling the zero-valued observations

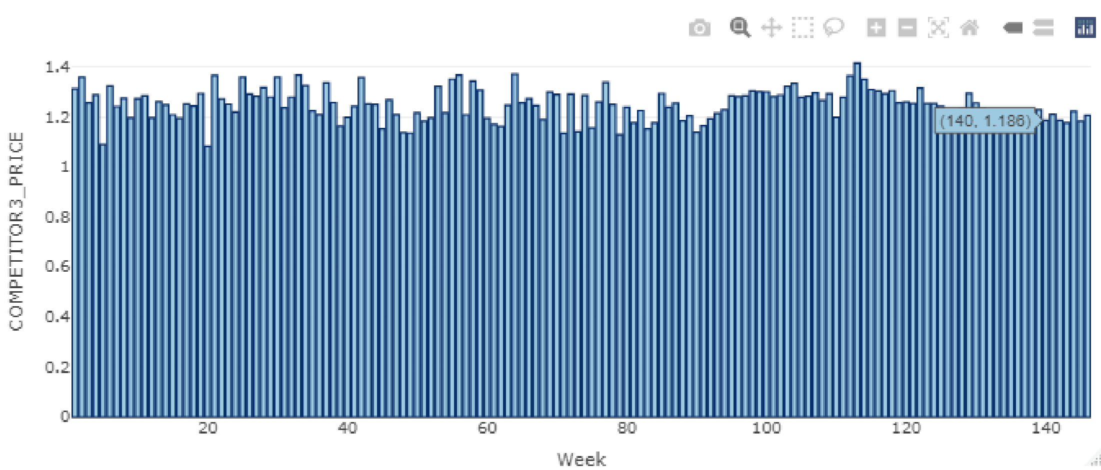

Figure 21 Competitor 3 price after filling the zero-valued observations

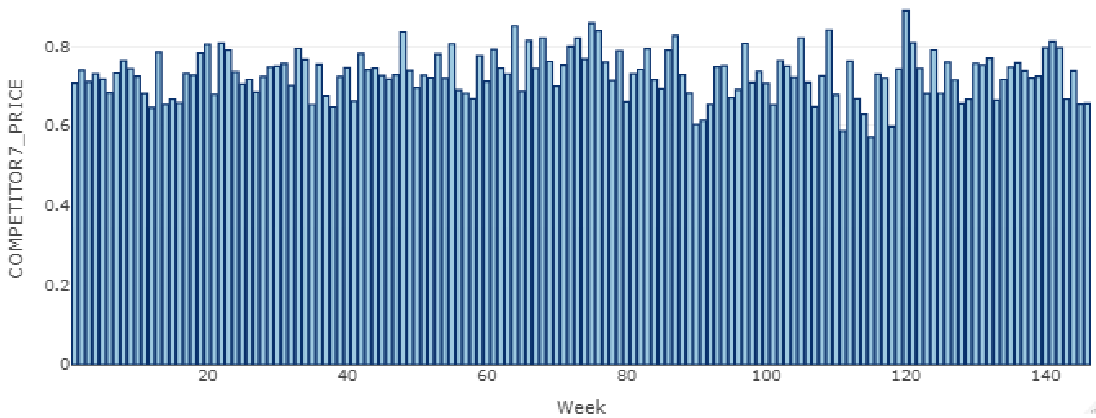

Figure 22 Competitor 7 price after filling the zero-valued observations

In Figures 20 – 22 there is a depiction of those variables, after filling the zeros. When fitting the different models these variables will be included in their original, due to the fact that it is unknown to us if these are missing values, or transformed forms or will be excluded. A combination of all competitors' product prices in one variable, by taking the average value of all prices at the same week, could also be tested. The total market sales could give an insight if these products where out of market for specific periods, although this is unavailable to us. We will treat these zeros in two ways, assuming:

- They are missing values, so we have filled these zeros in the way we have previously discussed
- They are periods were the products were out of market, so we will use the original variables by replacing the zeros with nulls

7. Modelling and evaluation

In this chapter there will be discussed three types of linear regression models, a simple multiple linear regression, a time series linear regression and a state space linear regression. The evaluation part contains the statistical evaluation tests for all models produced and an evaluation of models' predictability and finally, the modeling part and the selection of best model.

7.1 Multiple linear regression model

A linear regression model is of the following form

$$y = b_0 + b_1 x_1 + \cdots + b_n x_n + e \ (1)$$

Equation (1) represents a multiple linear regression model, where y is the depended variable, or the variable that we will predict and $x_1, \dots, x_n$ are the n independent variables. The $b_1, \dots, b_n$ coefficients measure the *marginal effects* of the independent variables.

7.2 Time series linear model

In the equation (1) now, the element of time is added.

$$y_t = b_0 + b_1 x_{1,t} + \cdots + b_n x_{n,t} + e_t \ (2)$$

A time series linear model has the ability to incorporate time series components as predictors, and specifically trend and seasonality. An example of such a model in predicting sales is the following.

Let $x_1 \in \{0,1\}$ be the variable representing *current product promos* and $x_2 \in \{0,1\}$ be the variable representing corresponding *companion / competitor product promos*. The encoded value of $x_j \in \{0,1\}$ indicates whether promotion j was activated at time t.

Therefore, we can express the predicted sales for product i, at time t, and for promo predictor $j = \{1, 2\}$, in the form of the following regression model:

$$Sales_{i,t} = b_{0,i} + b_{1,i}x_{1,i} + b_{2,i}x_{2,i} + \ldots + [Prices + Time + \ldots] \quad (2)$$

By using our historical data on sales, promotions, and prices, we can estimate model coefficients and use them to interpret relative promotion effect or forecast sales at time $t + k$.

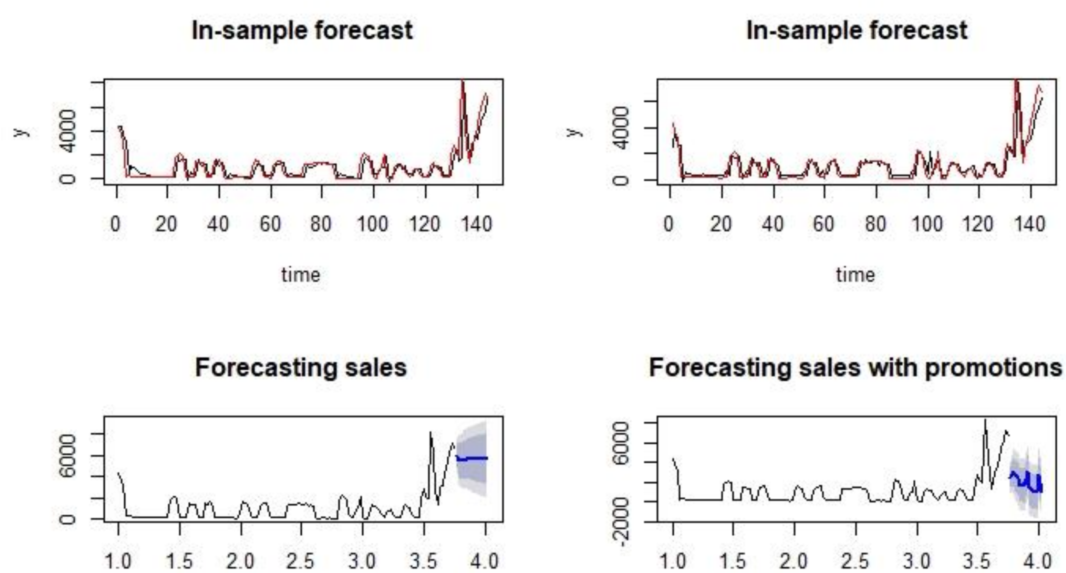

Figure 23 Example of time series linear regression model in forecasts

7.3 State space linear regression

The State Space Time Series Regression modelling, (Koopman and Shephard (2004)) [34] applies a moving base (local level) estimation approach that accounts for the explanatory factors that have influenced sales but there are no accurate data available to track them. Other, more rudimentary techniques often have the effects of unknown parameters misattributed to various other factors like media, promotion, cross effects etc skewing their effectiveness estimate. The state space (unobserved components method) avoids these shortcomings.

State Space Regression, (Koopman, Shephard and Doornik (1998)) [33], (Koopman and Shephard (2004)) [34] is based on a Gaussian State Space form consisting of a transition equation and a measurement equation; a typical formulation is:

$$a_t + 1 \;=\; d_t \;+\; T_t\, a_t \;+\; H_t\, e_t, \qquad a_1 \sim N(a, P), \qquad\qquad t = 1, \dots, n$$

$$\theta_t \;=\; c_t \;+\; Z_t\, a_t \; (4)$$

$$y_t \;=\; \theta_t \;+\; G_t\, a_t, \qquad e_t \sim NID(0, I)$$

The deterministic matrices T_t, Z_t, H_t and G_t are referred to as system matrices and they usually are sparse selection matrices. The vectors d_t and c_t are fixed and can be useful to incorporate known effects or known patterns into the model (i.e., regression effects), otherwise they are zero. When the system matrices are constant over time, we drop the time-indices to obtain the matrices T, Z, H and G. In the general case the omega matrix is defined as $\Omega =< H_t H_t^T \,, H_t G_t^T \,| G_t H_t^T \,, G_t G_t^t \; >$. This is a $(m + N)\; x\; (m + N)$ matrix with N being the number of observations available at time t and m being the number of observations of the state vector a_t. In the case of multiple linear regression in State Space the system matrices are set to $T_t = I_k$, $Z_t = X_t$, G_t = σ_ξ^2 , ε'1, where ε'1 = (1,0, 0…) and $H_t = 0$.

Some advantages to this approach are:

1. It accounts for the parameters that have influenced sales that we are not aware of, or we do not have an accurate way to track them – because many times the effects of these unknown parameters are misattributed to various other factors like media, or promotions, and thus skewing their effectiveness estimate. An implication of this advantage is that the base estimate changes with time – this gives a direct result on the sales level in absence of any promotion.
2. Apart from the advantages on business sense, the State Space approach offers more robust results in a statistics sense. For example, the regression model in State Space leads to the so-called marginal or modified-profile likelihood function, which is known to have better small-sample behavior than the standard concentrated likelihood.
3. Moreover, the beta coefficient estimates result from a maximum likelihood procedure rather than the standard ordinary least square (OLS) lines allowing for more robust inference and insights to be drawn as this approach returns the result which is more likely to happen.

7.4 Models' evaluation

The testing of a model's validity and the comparison among models will be conducted based on the tests and metrics in Table 3.

Table 3 Evaluation of a linear regression

Criteria	Value	Comments
R-squared adjusted	[0, 1]	The highest the value, the highest the explained variation of the dependent
Durbin-Watson p-value		Since the p-value is greater than a given threshold, we cannot reject the H0 of no correlation among the residuals
Jarque-Bera p-value		Since the p-value is greater than a given threshold, we cannot reject the H0 of normality of the residuals
Breusch-Pagan p-value		Since the p-value is greater than a given threshold, we cannot reject the H0 of homoscedasticity of the residuals
AIC	> 0	Comparative among models. The best model fit is indicated by the lowest AIC
VIF	≥ 1	1 indicates no correlation between any two predictors. 1 to 5 indicates moderate correlation. > 5 indicates potentially severe correlation

7.4.1 Statistical evaluation of the models

The systematic statistical testing of models contains the Durbin-Watson test for autocorrelation, the Jarque-Bera statistic for the normality of the residuals, the Breusch-Pagan test for heteroscedasticity and the Variance Inflation Factor (VIF) to detect multicollinearity.

- Durbin-Watson Test

The presence of correlation between the residuals consists a violation of the assumptions in a linear regression. To detect the presence of autocorrelation the Durbin – Watson test can be applied.

$$DW = \frac{\Sigma_{t=2}^{T}(\upsilon_t - \upsilon_{t-1})^2}{\Sigma_{t=1}^{T}\upsilon_t^2}$$

Where, u_t is the residual and T is the number of observations

The test uses the following hypotheses:

H0 (null hypothesis): There is no correlation among the residuals.

H1 (alternative hypothesis): The residuals are autocorrelated.

A DW statistic under 1 or greater than 3 should cause a concern (Field (2004))[32].

- Jarque-Bera Test

The normality test for the residuals of the model. If the residuals are not normally distributed, the residuals should not be used in Z tests or in any other tests derived from the normal distribution, such as t tests, F tests and chi-square tests. If the residuals are not normally distributed, then the dependent variable or at least one explanatory variable may have the wrong functional form, or important variables may be missing, etc. Correcting one or more of these systematic errors may produce residuals that are normally distributed. The presence of outliers should also be investigated in such a case. The test used here for normality is the Jarque-Bera statistic which is a goodness-of-fit measure of departure from normality, based on the sample kurtosis and skewness. The literature suggests that a reasonable approximation of the distribution of this statistic is the chi-squared with 2 degrees of freedom. Note that in case where this test fails the model should not necessarily be rejected but further investigation, perhaps by looking in the auxiliary residuals is required. The Jarque-Bera test is defined by:

$$JB = \frac{n}{6}s^2 + \frac{1}{4}k^2$$

Where,

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n} \left(x_i - \underline{x}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n} \left(x_i - \underline{x}\right)^2\right)^2} - 3$$

and

$$s = \frac{\frac{1}{n}\sum_{i=1}^{n} \left(x_i - \underline{x}\right)^3}{\left(\frac{1}{n}\sum_{i=1}^{n} \left(x_i - \underline{x}\right)^3\right)^{\frac{3}{2}}}$$

- Breusch-Pagan Test

It is used to determine if heteroscedasticity is present in a linear regression model.
H0 (null hypothesis): Homoscedasticity is present (the residuals are distributed with equal variance)
H1 (alternative hypothesis): Heteroscedasticity is present (the residuals are not distributed with equal variance)
If the H0 is rejected, the standard errors of the regression may be unreliable.

- VIF

VIF is used to detect multicollinearity by measuring the correlation and its strength among the independent variables of a model.

7.4.2 Selection of a good model

To compare and select between two valid models the Akaike Information Criterion (AIC) and the R-squared adjusted will be used. The selection of the variables for the regression will be made through a stepwise regression process.

- R-squared adjusted

  The R-squared adjusted is a measurement of how much of the variation of the dependent variable is explained by the independent variables. Unlike R-squared, this measure, which incorporates R-squared in its calculation, penalizes the use of extra input variables. So, just by adding a useless extra variable the R-squared adjusted will not rise. The calculation of R-squared is given by

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

  Where, RSS is the Residuals Sum of Squares and TSS is the Total Sum of Squares.

  The R-squared adjusted is given by

  $$Adjusted\ R^2 = 1 - \frac{(1 - \mathrm{R}^2)(\mathrm{n} - 1)}{\mathrm{n} - \mathrm{k} - 1}$$

  Where, n is the number of observations, k is the number of independent variables and R^2 is the R-squared values determined by the model

- AIC

  The Akaike information criterion is used to compare fit among regression models.

  $$\mathrm{AIC} = 2n - 2\log(L)$$

  Where, n is the model's predictors (for 2 predictors n = 4) and L is the log-likelihood of the model
  The lower the value of the AIC, the better the fit of the model.

7.4.3 Stepwise regression

For a large number of independent variables e.g., 16, the total number of models needs to be calculated is ~65,000 (2^{16}). A stepwise regression could be applied in this case, to avoid the calculation of all possible models. A forward stepwise starts with only the intercept. One by one the predictors are added to the model and only remains the one that improves the most the measure of predictive accuracy (AIC). This continues until no more improvement. A backwards stepwise starts with all the independent variables. Then removes variables one by one and keeps the model if this has an improved AIC, until no farther improvement.

Using the function *step,* we can apply a stepwise regression in R.

7.5 Modeling

Even though a state space linear model seems to be a better choice than a linear model with time elements or a multiple linear regression model, in this dissertation will be examined the multiple linear regression and the time series linear regression approaches. In the modeling phase, many possible models will be fitted.

In the time series linear regression fits, the volume of sales will be treated as a time series with a frequency of 52 (weeks), and a multiple linear regression will be applied with different combinations of predictors, through the stepwise regression, in original and transformed form, alongside with time related variables (trend and seasonality). A more robust approach would be to fit all possible model.

The train and test split will be 70 – 30 %, so, the training set will include 104 weeks (~70% of time series) and the test set will include 42 weeks (~30% of time series).

In the multiple linear regression, the volume of sales will be the independent variable.

7.5.1 Sales decomposition for time series linear modeling

A time series consists of four components, level, trend, seasonality and noise, out of which two are the main ones, trend and seasonality. By using the *cmav* function we will apply a centered moving average to the data.

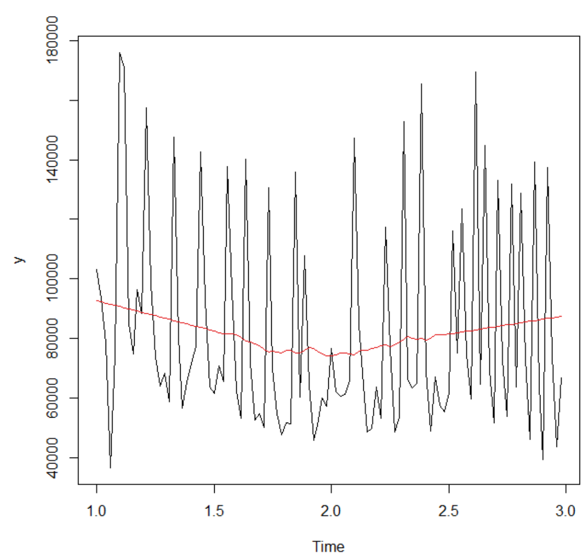

Figure 24 Trend plot of volume of sales

In Figure 24, the sales' trend shows a general downward pattern until around the end of first year and a general upward pattern later. This could be an indication of a Piecewise linear trend and should be farther explored.

We will check for seasonality by using the *seasplot*. This will detrend the series, if needed, and produce the seasonal plot. It will also provide a p-value for the presence of seasonality.

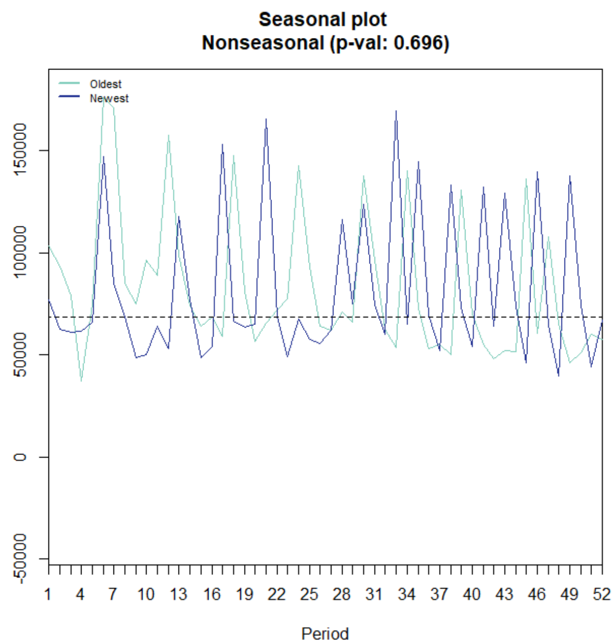

Figure 25 Seasonal plot of volume of sales

In Figure 25, based on the two years of data in the training set we cannot detect any seasonal pattern.

7.5.2 Time series linear modeling procedure

The modeling will be conducted, in the training set, with the *tslm* model using a forward stepwise regression. We have identified the presence of trend, so this will be included in the list of predictors. The full predictors' list is in Table 4.

Table 4 Number of predictors

Predictor's name	**Comments**
trend	The trend of the volume of sales time series
ACTUAL_PRICE	The actual price of the product
BASE_PRICE	The base price of the product
PROMO	The promotion percentage of the product at this week

AFTER_PROMO	Dummy variable that marks the week after a promotion has ended
PRICE_BY_BASE_AND_PROMO	Variable that was constructed by the base price and the promotion percentage
COMPETITOR1_PRICE	The price of the first competitor product
COMPETITOR2_PRICE	The price of the second competitor product
COMPETITOR3_PRICE	The price of the third competitor product
COMPETITOR4_PRICE	The price of the fourth competitor product
COMPETITOR5_PRICE	The price of the fifth competitor product
COMPETITOR6_PRICE	The price of the sixth competitor product
COMPETITOR7_PRICE	The price of the seventh competitor product
COMPETITOR2_PRICE_FILLED	The price of the second competitor product, with zero values filled
COMPETITOR3_PRICE_FILLED	The price of the third competitor product, with zero values filled
COMPETITOR7_PRICE_FILLED	The price of the seventh competitor product, with zero values filled
COMPETITOR_AVG_PRICE	The average price of all competitors' products original prices

The independent variable will be the volume of sales or the logarithmic transformation of them. Consequently, we can create four groups of testing models.

- Group 1 has the sales as the independent variable and trend, actual price, base price, promotions, after promo dummy, the price produced by the base price and the promotions, average competitor price and all original competitor price variables as predictors.
- Group 2 has the sales as the independent variable and trend, actual price, base price, promotions, after promo dummy, the price produced by the base price and the promotions, competitor 1 price and all the filled competitor price variables as predictors.
- Group 3 has the logarithmic transformation of sales as the independent variable and trend, actual price, base price, promotions, after promo dummy, the price produced by the base price and the promotions, average competitor price and all original competitor price variables as predictors.
- Group 4 has the logarithmic transformation of sales as the independent variable and trend, actual price, base price, promotions, after promo dummy, the price produced by the base price and the promotions, competitor 1 price and all the filled competitor price variables as predictors.

The forward stepwise time series linear regression will be conducted once in each one of these groups. The final outcome of the stepwise proposes the models in Table 5.

Table 5 Stepwise regression outcome

Group	Selected Independent Variables	AIC
1	Trend, Actual Price, Base Price, Promotion, After promo	2073.3
2	Trend, Actual Price, Base Price, Promotion, Competitor 2 price filled, Competitor 3 price filled, Competitor 7 price filled	2068.8
3	Trend, Actual Price, Base Price, Promotion	-300.86
4	Trend, Actual Price, Base Price, Promotion, Competitor 2 price filled, Competitor 7 price filled	-304.29

In Table 5 it appears that based on AIC the preferred model to continue would be the selected model from Group 4.

$\log(SalesVolume_t) = 12.004 - trend + 3.94ActualPrice_t - 5.05BasePrice_t + 0.06Promo_t - 0.59Competitor2PriceFilled_t + 0.86Competitor7PriceFilled\_t$ (1)

Equation (1) is the selected model based on AIC, from those that where the final suggestions from the stepwise regression. There is no certainty that this model is the best possible, but is a good model (James, Witten, Hastie, & Tibshirani (2014)) [35]. The statistical evaluation of the four models will show if they are reliable or not.

Table 6 Statistical evaluation of proposed model

Model	Durbin - Watson statistic (p-value)	Jarque-Bera p-value	Breusch-Pagan p-value	VIF value
Group 1	2.09 (0.644)	0.01015	0.0003975	2808.74
Group 2	2.08 (0.492)	0.3693	0.001094	1900.659
Group 3	1.94 (0.874)	0.02639	0.03198	1535.273
Group 4	1.98 (0.896)	0.6522	0.02803	1720.082

Based on Table 6, all four models have no correlation among their residuals (Durbin – Watson test). For a significance level of 95% the null hypothesis of normally distributed residuals (Jarque – Bera test) is rejected in the models of Group1 and Group 3 and is not rejected in the models of Group 2 and Group 4. For the same significance level (95%) the null hypothesis of the Breusch – Pagan test (no heteroscedasticity in residuals) is rejected. So, there is strong evidence that, in all models, the residuals do not have constant variance. The VIF value of all models is way above 5. This strongly indicates high correlation between predictors. Indeed, the presence of actual price, base price and promotions simultaneously in the same model could cause that.

The stepwise regression, as it was executed, did not help to find a valid time series linear regression model and, consequently, we will not procced with the testing. The safest approach would be to execute all possible models.

7.5.3 Multiple linear regression procedure

In the multiple linear regression models the stepwise regression technique will not be applied. The presence of null values in the competitors' products price cannot be handled by the *step* function. We will test some models that make a business sense. The R function *lm* will be the one that will fit our multiple linear regression.

The first limitation in the predictors is that the examination of actual price and promotion percentage will not be conducted in the same model with base price. This combination creates problems of correlation among predictors. Secondly, the filled competitor prices will not be tested in the same model with the original competitor prices. In the latter, the zeros have been replaced with nulls. We will also test the volume of sales and the logarithmic transformation of them.

Table 7 Multiple linear regression models

Model	Variables	AIC	R-squared adj.
1	log(VOLUME_OF_SALES) = BASE_PRICE + PROMO + AFTER_PROMO + COMPETITOR2_PRICE + COMPETITOR3_PRICE	11.437	0.7428
2	log(VOLUME_OF_SALES) = BASE_PRICE + PROMO + AFTER_PROMO + COMPETITOR4_PRICE	9.116	0.6785
3	log(VOLUME_OF_SALES) = ACTUAL_PRICE + PROMO + AFTER_PROMO + COMPETITOR2_PRICE + COMPETITOR3_PRICE	11.236	0.7449
4	log(VOLUME_OF_SALES) = ACTUAL_PRICE + PROMO + AFTER_PROMO + COMPETITOR1_PRICE + COMPETITOR4_PRICE	8.458	0.6859
5	VOLUME_OF_SALES = BASE_PRICE + PROMO + AFTER_PROMO + COMPETITOR1_PRICE + COMPETITOR2_PRICE + COMPETITOR3_PRICE + COMPETITOR7_PRICE	456.84	0.7053
6	VOLUME_OF_SALES = BASE_PRICE + PROMO + AFTER_PROMO + COMPETITOR3_PRICE_FILLED + COMPETITOR4_PRICE + COMPETITOR6_PRICE + COMPETITOR7_PRICE_FILLED	458.66	0.6904
7	VOLUME_OF_SALES = ACTUAL_PRICE + PROMO + AFTER_PROMO + COMPETITOR2_PRICE + COMPETITOR3_PRICE + COMPETITOR7_PRICE	417.59	0.6878

8	VOLUME_OF_SALES = PROMO + COMPETITOR3_PRICE_FILLED + COMPETITOR4_PRICE + COMPETITOR6_PRICE	457.06	0.6836

In Table 7 there is a model for each unique combination based on the limitations defined above. Based on R-squared adjusted the model that explains most of the variation of the dependent variable is model 3. On the other hand, based on AIC, the prevailing model is the model 4. In both of those models the volume of sales was given in the *lm* function in its logarithmic transformed form. Now will have a look in the statistical evaluation for the validity of the models.

Table 8 Multiple linear regression models' statistical evaluation

Model	Durbin - Watson statistic (p-value)	Jarque-Bera p-value	Breusch-Pagan p-value	VIF value
1	2.238 (0.768)	0.6886	0.2907	< 3.1
2	1.753 (0.26)	0.1071	0.09143	< 1.6
3	2.254 (0.728)	0.7223	0.2435	ACTUAL_PRICE > 13, PROMO > 9
4	1.847 (0.454)	0.00876	0.0724	ACTUAL_PRICE > 10, PROMO > 9
5	2.28 (0.966)	0.7194	0.7609	< 3.7
6	2.72 (0.288)	0.954	0.8297	< 3.8
7	2.58 (0.556)	0.9435	0.6626	ACTUAL_PRICE > 11, PROMO > 8
8	2.58 (0.312)	0.9522	0.8903	< 1.2

Based on table 8, models 3 ,4 and 7 appear to have very high VIF values for the actual price and promotion predictors, which does not come as a surprise. The actual price variable actually incorporates any promotion made in this week. So, models 3, 4 and 7 should not be considered as candidates. Among the remaining models, model 1 has a higher R-squared adjusted but model 2 has a lower AIC. We should farther analyze those two.

- Model 1

$$\log(VolumeSales) = 7.239 - 1.644BasePrice + 0.053Promo + 0.295AfterPromo + 0.669Competitor2Price + 3.9599Competitor3Price$$

- Model 2

$$\log(VolumeSales) = 11.3128 - 0.376BasePrice + 0.035Promo + 0.165AfterPromo - 0.1079Competitor4Price$$

By comparing the equations of the two models, model 2 appears to have an unexpected coefficient for the competitor 4 price variable. Let's transform the coefficient of the Competitor4Price predictor so it represents a percentage change in the volume of sales.

$$PrcChange_i = (e^{b_i} - 1) * 100 \text{ (4)}$$

Where i represents the i^{th} predictor

So, applying equation (4) in the Competitor4Price coefficient we have a -10.23%, meaning that, by one unit of increase for Competitor4Price, the volume of sales is decreasing by 10.23%. This is contradictory because we should expect that an increase in the price of a competitor product, should also increase the volume of sales of our targeted product.

Considering this, model 1 is the model that more logically explains the volume of sales.

8. Conclusion and discussion

Even though this dataset had time sequence in its structure (weekly data) a suitable model found through a more simplistic approach, a multiple linear regression. The state space linear approach could not be tested due to its theoretical and practical complexity, but could have resulted in a better solution, that also explains variation from unknown factors.

In this dissertation, the forward stepwise regression was used to find a good (not the best) model in the time series analysis and an intuitively approach, due to limitations in *step* function to handle missing values, to find a good model in the multiple linear regression analysis. An approach that would fitted all possible models, in both analyses, could have resulted in better final model.

The main scope was to find a model that explains the volume of sales, for the given dataset, and not the predictability of the model. So, we mainly focused on the statistical evaluation of the models and not on their ability to predict the volume of sales.

Bibliography

1. Alexander, C. (1964). *Notes on the Synthesis of Form* (Vol. 5). Harvard University Press.
2. Banfi, F., Bhandari, R., Gordon, J., & Umblijs, A. (2012). Marketing Mix Modeling. In Perrey & Spillecke (eds), *Retail Marketing and Branding: A Definitive Guide to Maximizing ROI,* (pp 161-177). U.S.: Wiley.
3. Cain, P. M. (2014). *Dynamic marketing mix modelling and digital attribution*. US: Market Science Consulting.
4. Casas, P., Chapman, J., Hunt, R., Kozyreff, G., Lacey, A., Larrieu, E., ... & Staron, L. (2004). *Models of consumer behaviour*. Study Group Report, ESGI.
5. Chan, D., & Perry, M. (2017). Challenges and opportunities in media mix modeling.
6. Day, G. S. (1981). The product life cycle: analysis and applications issues. *Journal of marketing, 45*(4), 60-67.
7. Deliya, M. (2012). Consumer behavior towards the new packaging of FMCG products. *National Monthly Refereed Journal of Research in Commerce and Management*, *1*(11), 199-211.
8. Gehin, A., Zwolinski, P., & Brissaud, D. (2008). A tool to implement sustainable end-of-life strategies in the product development phase. *Journal of Cleaner Production, 16*(5), 566-576.
9. Hernandez, T., & Bennison, D. (2000). The art and science of retail location decisions. *International Journal of Retail & Distribution Management, 28*(8), 357-367.
10. Johansson, G. (2002). Success factors for integration of ecodesign in product development. *Environmental management and health, 13*(1), 98-107.
11. Kim, K. Y. (2014). Business intelligence and marketing insights in an era of big data: The q-sorting approach. *KSII Transactions on Internet and Information Systems (TIIS), 8*(2), 567-582.

12. Kim, M., Zimmermann, T., DeLine, R., & Begel, A. (2017). Data scientists in software teams: State of the art and challenges. *IEEE Transactions on Software Engineering, 44*(11), 1024-1038.

13. 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.

14. Meldrum, M., & McDonald, M. (1995). *Key marketing concepts*. Basingstoke: Macmillan.

15. Milani, G., & Lourenço, P. B. (2010). Monte Carlo homogenized limit analysis model for randomly assembled blocks in-plane loaded. *Computational Mechanics*, *46*(6), 827-849.

16. Moss, G., & Schuiling, I. (2004). A brand logic for pharma?: A possible strategy based on FMCG experience. *Journal of Medical Marketing*, *4*(1), 55-62.

17. Pattnaik, M., & Behera, M. K. (2016). *How Predictive Analytics is Changing the Retail Industry*. Paper presented at the International Conference on Management and Information Systems. Chitkara University Bangkok, September, 23-24.

18. Pourhejazy, P., Sarkis, J., & Zhu, Q. (2019). A fuzzy-based decision aid method for product deletion of fast moving consumer goods. *Expert Systems with Applications, 119*, 272-288.

19. Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, *1*(1), 51-59.

20. Purcarea, T. (2014). A Review of the Different Marketing Opinions on Marketers' Maturity and Challenges in the Second Half of 2014. *Holistic Marketing Management Journal, 4*(4), 05-14.

21. Roberts, J. H., & Lilien, G. L. (1993). Explanatory and predictive models of consumer behavior. *Handbooks in operations research and management science, 5*, 27-82.

22. Sakoda, G., Takayasu, H., & Takayasu, M. (2019). Data science solutions for retail strategy to reduce waste keeping high profit. *Sustainability, 11*(13), 3589.

23. Sharma, N. (2013). Marketing strategy on different stages PLC and its marketing implications on FMCG products. *International Journal of Marketing, Financial Services & Management Research, 2*(3), 121-136.

24. Sheth, J., & Kellstadt, C. H. (2021). Next frontiers of research in data driven marketing: Will techniques keep up with data tsunami? *Journal of Business Research*, *125*, 780-784.
25. Sibalija, T. V., & Majstorovic, V. D. (2012). An integrated approach to optimise parameter design of multi-response processes based on Taguchi method and artificial intelligence. *Journal of Intelligent manufacturing*, *23*(5), 1511-1528.
26. Tanner Jr, J. F., & Chonko, L. B. (2002). Using trade shows throughout the product life cycle. *Journal of Promotion Management*, *8*(1), 109-125.
27. Tibben-Lembke, R. S. (2002). Life after death: reverse logistics and the product life cycle. *International Journal of Physical Distribution & Logistics Management*, *32*(3), 223-244.
28. Usta, M., Darabi, H. R., Ryan, J., & Data, G. (2017). *On optimal budget allocation using marketing mix models*. Paper presented at the Marketing Analytics Conference, Austin, Texas, June, 5-6.
29. van Elzakker, M. A., Zondervan, E., Raikar, N. B., Grossmann, I. E., & Bongers, P. M. (2012). Scheduling in the FMCG industry: An industrial case study. *Industrial & engineering chemistry research*, *51*(22), 7800-7815.
30. Wasson, C. (1974). *Dynamic Competitive Strategy and Product Life Cycles*. St. Charles: Challenge Books.
31. Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97-121.
32. Field, A., (2009), Discovering statistics using SPSS, third edition, Sage Publications.
33. Koopman, S.J.M. & Shephard, N. & Doornik, J.A., (1998), Statistical Algorithms for Models in State Space Using SsfPack 2.2.
34. Koopman, S.J.M and Shephard, N., (2004), State Space and Unobserved Component Models, Cambridge University Press.
35. James, G., Witten, D., Hastie, T., & Tibshirani, R., (2014), An introduction to statistical learning: With applications in R, New York: Springer.

Appendix

Table of Figures

Table of Tables