FORECASTING PROMOTIONAL SALES IN COSMETIC INDUSTRY USING MACHINE LEARNING METHODS

ZAHİDE TUĞBA KARAKURT

MAYIS 2022

FORECASTING PROMOTIONAL SALES IN COSMETIC INDUSTRY WITH MACHINE LEARNING

BAHÇEŞEHİR ÜNİVERSİTESİ

LİSANSÜSTÜ FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ZAHİDE TUĞBA KARAKURT

BÜYÜK VERİ ANALİTİĞİ VE YÖNETİMİ DALINDA

YÜKSEK LİSANS DERECESİ İÇİN GEREKLİ ÇALIŞMALAR YERİNE GETİRİLMİŞTİR

MAYIS 2022

TEZ ONAY FORMU BURAYA GELECEK..

Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.

Ad, Soyad : Zahide Tuğba Karakurt

İmza :

ÖZET

KOZMETİK SEKTÖRÜNDE MAKİNE ÖĞRENMESİ
YÖNTEMLERİ İLE PROMOSYONEL SATIŞ TAHMİNİ

Zahide Tuğba Karakurt
Büyük Veri Analitiği Ve Yönetimi
Tez Danışmanı: Doç .Dr.Tevfik Aytekin
Mayıs 2022, 40 sayfa

Bu çalışmanın amacı, bir perakende kozmetik şirketinde satış indirim kampanyalarının dönemsel satış talep tahminine etkisini inceleyerek, aylık talep tahminlerini makine öğrenimi teknikleriyle geliştirme konusunda etkin bir çalışma geliştirmektir. Satış promosyonlarının tümü veya mağaza içi kampanyalar hedef müşteriye, sadık müşteriye ve potansiyeli olan ancak henüz firmanın müşterisi olmayanlara ulaşmayı sağlar. Pazarlama stratejisi olarak kampanyalar, satınalımı arttırmayı, daha fazla ciro üretilmesini böylece finansal olarak daha fazla kar elde edilmesini amaçlamaktadır.

Pazarlama içerisinde satış kampanyaları hayati önem taşıyan bir araçtır ve uygulamaları yıllar içinde rekabetçi perakende sektöründe kayda değer bir şekilde artış göstermektedir. İdeal fiyat konumlandırmaları ve indirim seviyeleri de bir firma için temel göstergelerden biridir. Diğer önemli bir nokta ise, bu tarz pazarlama stratejileri uygulanıyor iken aynı zamanda marka sadakati ve güvenirliliğinin de korunabilmesidir. Dolayısıyla bu iş gereklilikleri çok önemlidir. Çoğu şirket daha iyi satış tahminleri yapmak, harcamalarını kontrol etmek ve yeni müşteriler kazanmak için mücadele etmekteler. Yöneticiler ve analistler farklı değişkenlere ve deneyimlerine dayanarak satış tahminleri yapmaya çabalarlar. Ürün kampanyaları, raf yerleşimleri ve mevsimsel etkiler yönetici ve analistleri bir sonraki kampanya tahminlerinde basit lineer regresyon kullanmaya teşvik etmekteler. Bu geçmişten gelen yaklaşımlar insan beyni ile sonuçları analiz etme odaklı yaklaşımlardır.

Perakende şirketleri arasında sürekli büyüyen rekabet ve müşteri beklentilerinin sürekli değişimi, şirketlerin üzerinde yeni kampanya kurguları yaratmak için artan

satışlar ve pozitif kar elde etmek amacıyla büyük bir yük oluşturuyor. E-ticaret'in yükselişi müşterilerin almak istedikleri ürünlerin fiyat ve özelliklerini kolayca kıyaslama yapabildiği bir alan olması nedeniyle duruma yeni bir boyut kazandırıyor. Buna bağlı olarak kişiselleştirilmiş kampanyalar müşterilere fayda sağlayacak önemli bir araç olmaya başlıyor.

Yeni dönem tahminleme çalışmalarında makine öğrenimi teknikleri fark edilmektedirler. Makine öğrenimi teknikleri uygulanan bir model inşa ederek, her kategori ve ürün için doğruluk oranı daha yüksek tahminlemelere ulaşılabilmektedir. Firmalar, kendi sektör ve endüstrileri içerisinde yer alan rekabetçi dinamikleri, makine öğrenmesi tekniklerinin gücünü kullanarak kayda değer bir şekilde etkilemekte ve değiştirmekteler. Son zamanlarda makine öğrenmesi tekniklerinin uygulamaları, daha geniş bir bilgi kaynağına ve bilişim gücüne ulaşılabilirliği nedeniyle daha pratik uygulamalar olma yolundadır.

Bu çalışmamızda satış tahmini yapabilmek için lineer regresyon, rastgele orman, karar ağaçları ve XGBoost methodları gibi makine öğrenim teknikleri uygulayacağız. Modellerimizde indirim kurgusu ve mevsimsellik gibi değişkenlerin satış tahminine etkilerini inceleyeceğiz. Bu çalışma, indirim ve kampanyaların müşteri davranışları üzerindeki etkileri ve tahminleme algoritmalarımızın başarısı hakkında fikir ve bilgi sağlayacaktır.

Keywords: Satış kampanyaları, Müşteri Davranışları, Makine Öğrenimi, Rastgele Orman, Karar Ağaçları.

ABSTRACT

FORECASTING PROMOTIONAL SALES
IN COSMETIC INDUSTRY WITH MACHINE LEARNING

Zahide Tuğba Karakurt
Big Data Analytics and Management
Thesis Supervisor: Assoc. Prof. Tevfik Aytekin
Mayıs 2022, 40 pages

This study aims to observe and improve a deep work about periodical sale forecasting using machine learning techniques by analysing sales promotions' effect on them in a cosmetic retail company. Whole kinds of sale promotions or in-store promotions enable reaching the targeted customers, the loyal customers, and the potential ones who are not a regular customer of the company yet. As a marketing strategy, promotions are aimed at increasing purchases, generating more revenue hence, more profit for the company financially.

Sales promotion is a vital tool in marketing and its applications are remarkably increasing over the years in this competitive retail industry. Ideal positioning of price and discount levels is also a major indicator for the company. Another important point is that brand loyalty and reliability must be preserved when these marketing strategies are applied. Therefore, balancing these business needs is crucial. Most companies are struggling to make better sales predictions to control their spending and gain new customers. Managers and analysts try to predict sales based on different variables and their experiences. The promotions of products, the shelf placement, and seasonality effects prompt them to use basic linear regressions on their forecast of the next campaign results. This old approach depends on the human brain to analyze the results.

Ever-growing competition between retailers and changing consumers' expectations place a burden on companies for designing new promotions with growing sales and positive margins. The rising of e-commerce adds a new dimension to this field also, in which customers can more easily compare the price and features of products they

wanted to buy. Correspondingly customized promotions are becoming a more crucial tool to affect customers in favor of them.

In the new world of forecasting finds out the requirement of machine learning techniques. Building a model using machine learning, we have more accurate predictions for every category or product. With the power of machine-learning techniques companies significantly influence competitive dynamics in their industries and sectors. Recently, these techniques have become more practical because of the availability of computing power and a wider variety of data sources. We'll apply various machine learning methods such as lineer regression, random forest, decision trees, and XGBoost to predict sales.

In our models, we analyze the effect of important variables such as the level of the discount, promotion elasticity and seasonality to forecasting. These results enable managers and researchers to have an idea about promotion impacts on consumer behaviors and our predicting algorithms' success on forecasting.

Keywords: Sale Promotion, Customer Behaviors, Machine Learning, Random Forest, Decision Trees.

Anneme ve babama

TEŞEKKÜR

Bu tez çalışmasının planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen, engin bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren sayın hocam Doç. Dr. Tevfik Aytekin'e sonsuz teşekkürlerimi sunarım.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

ML	:	Machine Learning
K-NN	:	K- Nearest Neighbour
RF	:	Random Forest
XGBooST	:	Extreme Gradient Boosting
AR	:	Auto-Regressive
RMSE	:	Root Mean Square Error
MAPE	:	Mean Absolute Percentage Error
MAD	:	Mean Absolute Deviation
MSE	:	Mean Squared Error
M5P	:	M5 Base Model Tree
RSS	:	Residual Sum of Squares
CRM	:	Customer Relationship Management

1. INTRODUCTION

Sale promotions are immensely powerful marketing strategies applied by retail companies to stimulate the demand for products as we mentioned before. Additionally, to be more profitable, promotions help to reduce overstock problems, providing store traffic, meeting new products and, dominate the market between competitors [1]–[3]. Generally, retailers use numerous approaches to promote products, for example, price discounts, special deals (buy-two-get-one-free, thirty percent discount), vouchers, and special offer displays [4]. Meanwhile, there are lots of external components that determine the success of sale promotions, such as macro & microeconomic transformations, sensitivity to market trends, sales in complementary categories, seasonal effects and, other competitors [5].

From the vision of the companies, the forecasting of sales promotions is vital for inventory stock levels and production process planning for the supply chain. In retail, ideal, and accurate forecast means optimized and fulfilling stocked products on the shelves and to live a good shopping experience that communicates well about products and their usages with the store staff for the customers.

After arising the e-commerce shopping, customers changed their buying behaviors and their demands on products, shopping experience nearly for the past decade. Consumers are more informed about the product offerings and actual campaigns nowadays. They are reaching them more proactively fast to their expectation and demands. With the social media effects, they can easily access all information about offers before making a purchasing decision. Sales promotion is one of the key indicators that companies use for struggling rivals of the market. Companies have to try to get it out in front with their high discount effect and goods of their services.

1.1 PROBLEM STATEMENTS & OBJECTIVES

Some sales promotions have value promotions, price promotions, or extra gift offers. Value promotions include free-gift options for specified calculated customers considering loyalty data, container, or multi-bin promotions. Price promotion includes amount or discount level coupons, diverse discount offers, buy one get the second at half price, buy one get one free, and extra sample packs. There are many forms of sales promotion in retail to stimulate customers we will mention them in the following chapters in detail but one of the problems is getting hard in this issue which sales promotion pattern would be most effective and which one is attracting the customers to push purchasing it.

The retail companies are chasing one question and they try to overcome it by using data and machine learning technics power. The question is which sales promotion tools and applications such as coupons, price discounts, point of sale displays, free samples, threshold gifts affect the purchasing decision of the customers? Thus, the importance of observing and detecting which sales promotional method and technic should be chosen from the pool of machine learning methods is becoming vital when they plan to apply these kinds of strategies. We will be observing also the major factors that influence customers to choose the promotional pattern while making their purchasing decision in the details of products and category.

The process of designing promotional planning can be complicated and challenging for dynamic competitive markets, especially for the retail industry. Retail companies hire forecasters and analysts, who are devoted to predicting the future sales of products on the next offers. If they have complex and a vast number of promotions, these companies apply to forecast decision support systems. Some of the modern and innovative companies also applied statistical and machine learning approaches [6]. These methods count on data from historical actual sales to forecast upcoming future sales. Besides, a few of them can interpret these figures applied the methods. Technics such as Decision Trees, Support Vector Machines, or Neural Networks could provide more accurate estimations to managers but behind the results, current situations and updated information can be ignored [7].

An ideal forecasting algorithm must be consistent, interpretable, accurate, and have stability for promotion designing. If the algorithm is sufficiently adaptable, it performs well on various categories or product types. If the algorithm is stable, it means that the accuracy does not depend on time and doesn't affect seasonality. This minimizes the time-consuming processes which required re-training, re-testing, and update the data.

1.2 SCOPE AND LIMITATION OF RESEARCH

This research is focused on a monthly campaign period of product sales data from a cosmetic company in Turkey. We have a wide variety of products and lots of sub-categories in our data. Their trends and performances will be different in every season and every special time duration of the campaign. Observing and studying only public promotions such as discount offers as 30 percent discount, buy one get the second at half price, buy one get one free doesn't give the precise results all the time. Due to there are different types of sales promotions in the company which affects purchasing behavior of customers such as coupons, free samples, customized threshold gifts, birthday gifts also, we struggle to detect the 100 percent the main reason for sale increase.

In some outlier days and periods which applied extra marketing operations such as TV advertisement on specific products and outdoor visual activities, we encounter unusual boost effect and out of trend issue for some products or categories. We try to explain them sufficiently, that will never change our insights and results.

The Covid-19 pandemic started in the March of 2020, all stores of the company had to close after the second half of March until June of 2020. Therefore, in this period generally the website channel was available for shopping. Sales figures were not stable and sales decreased in this period because of both pandemic effects and closed stores. In worldwide also most companies especially in the retail industry, were affected too badly in this term because of stopping the shopping from stores and turning to online shopping. We normalize and ignore these periods in our study for required points, we will mention these points in the section process of data preparation.

2. LITERATURE REVIEW

In first section, pieces of knowledge about some of the related literature researches are given. There are too many studies on sales forecasting, some of them are specific to customized customers and their behaviors, some of them are focused on the effect of historical data for specific sales time series when a new product or new store is launched [8]. One of them provides an insight to us about working on a kNN method that modifies predictions which are automated weighted k-nearest neighbors where the distances are calculated based on a feature selection process that focuses on the resemblance of promotional sales [9]. We observe these two studies in detail in the Related Studies part.

This literature review part consists of the following sections: Related studies, types of sale promotions, price discount, the significance of sale promotions in retail, and effects on decision making.

2.1 RELATED AND PREVIOUS STUDIES

Improving accuracy is essential for studies that have uncertainty. In [10], it is mentioned that combining forecasts produced by different algorithms, will be increasing the possibility of forecasting accuracy. This work also mentioned time-series datasets that classical algorithms and machine-learning-based algorithms can be equally used.

The [8] study focused on the effect of historical data for specific sales time series when a new product or new store is launched. For [8] study there are some issues, for instance, not having historical data for a long period for a new product. To overcome this problem, they create a pattern from a similar product and expected that the new product will have a similar sales performance. Sales prediction is a regression problem rather than a time series problem. Implementations of regression approaches may sometimes give better results in comparison to time series methods. Because machine-learning algorithms make it easier to find patterns in the time series [8], in this study

we can talk about complicated patterns in the sales dynamics, they use supervised machine-learning methods.
Some of the most popular and well-known algorithms are tree-based machine-learning ones [11] like Random Forest, Gradient Boosting. We will use tree-based algorithms in our case also in the following chapter.

In [8], researchers try linear models, machine learning, probabilistic models. They use in [8] and study many approaches for time series forecasting and logistic regression with highly imbalanced data. They observe several types of patterns and effects in their data for example seasonality, trend, autocorrelation. They also analyze too much noise in the sales and outliers. They need to take into account noise and extreme values because of risk assessments. In the work, we can see that the main assumption of regression methods is the patterns in the historic data may be repeated in the future. The accuracy of the validation set is an essential indicator to choose an optimal number of iterations of machine-learning algorithms. The influence of machine-learning generalization consists in the fact of capturing the patterns in the whole set of data. This affects the make sales prediction usage if there is a small number of historical data for special time sales for a new product is launched.

The other study observed one provides an insight to us about working on a kNN method that modifies predictions which are automated weighted k-nearest neighbors where the distances are calculated based on a feature selection process that focuses on the resemblance of promotional sales. [9] This method learns online, and the algorithm is compared with an ensemble of regression trees and the forecast provided by the company and it performs over not only by the mean absolute error but also by the error deviations used in business. The recommended method in [9] significantly increases the accuracy of the forecast in many kinds of categories and geographical locations.

In [9], four implementations are applied. The first one is collinearity for retail data mostly a common issue, which can be complicated. The second one is feature selection for instance promotional events. The third implementation is the tuning of hyperparameter on a real dataset. The last one is the algorithm on a large real-market dataset generated from lots of promotions from many countries and stores. We can see

in [9] study is compared with Least Squares (LS), an ensemble of regression trees, and the employee' forecast. The prediction is in [9] work, generated at SKU level meaning that the data always fits in memory. In this model also researchers show us how to expand the algorithm by combining single learners and weak forecasters to increase accuracy.

The features and the weights of the campaigns use for calculating the forecast. For most promotions, the cost-sensitive or discount rate-sensitive ones are set by analysts. But in [9] study, the categorical data in datasets have low cardinality, in the situation that having large cardinality categorical data, requires application of dimensionality reduction.

2.2 TYPES OF SALE PROMOTIONS

For this company in our study, the campaign process is managed by two approaches which had to be considered different dynamics managing them. One of them is customized campaigns for every customer which consist of birthday campaigns, welcome packs, threshold gifts, samplings, specific churn-focused ones. We will mention this kind of campaign called provoke campaigns. These campaigns mainly focused on data-oriented of every specific customer behavior and purchasing. These promotions are always aimed to push more purchasing of the loyal customers and gain new customers. In this way, all recommendations and offers of a product can reach potential ones by constantly reminding them who needs them. Over time while the purchasing data is flowing and histories of customers accumulating, these customized campaign algorithms work better.

Another application on campaign management is in-store public offers which we will call spontaneous campaigns. These campaigns contain whole year campaign periods thematically which will play on store windows and e-commerce. These discount-based offers take place at all the points of the sale. Some of one might be categorically priced discount, at the same time some of one might be like starting from a specific price for a star product or product type.

There is a wide range of tools to use when promoting sales as we mentioned before such as samplings, birthday campaigns, coupons, welcoming packs, threshold gifts, and spontaneous public offers. Let's mention the types of campaigns in detail.

Sampling is a small size trial offered to the consumer of the product. This is considered to be the most effective to stimulate consumers for buying but it is also costly high. Because samples should be free to try out but depending on the product, sometimes an extra offer can apply like a threshold purchasing to have this sample for free. Especially if the product size and ml's are close to the main product, the company should cover the costs. The sample aims to experience and test a new product or an already existing product that has been in the product range for a certain period. Specifically, in cosmetic products, some allergy issues can happen, and consumers want to test the product on their skin before buying it.

Birthday campaigns aim that bring the customers again to store who have already a customer of this company. And keep continuity the loyal of customers, giving them a gift for their birthday. This kind of subtlety affects the customers and enables sympathy to the company.

Coupons aim that offers the customers a discount on specified products or a certain amount of cash discount. They can be used in different customer types or different points of sales for example in specific stores where support by mall management. Exclusively some mall managements apply these kinds of activation for all stores where they locate in their mall and integrate it. The advantage of Coupons can be gaining new customers and pushing loyal customers coming to the stores.

Welcoming packs are applied when prices are reduced on certain specific products directly on their package. An example of these packs can be two for one, or three for two kinds of deals, or enable an advantaged discount for purchasing the whole pack rather than buying one by one. Price packs are considered highly effective and important in cross-selling products which should use together.

Threshold gifts aim to grow the basket size, average selling price, and products in the basket. This offer presents a gift in the condition that if the purchase reaches a certain amount. This stimulates customers to buy one or two more products aim of having this gift.

Spontane public promotions, often also known as just POP's, and placement of the products in the forms of displays which are taking place in stores and e-commerce channels. Usually, this can be for example 30% discount on body care products or another specific product group, it can be also a specific product starting from a price offer. All these offers can take place in the same campaign period, it depends on the company's campaign strategy. Buy one get one free or buy one get the second 50% discount offers are also placed in spontaneous campaigns.

Integration of provoking and spontaneous campaigns between each other is crucial. Because as of financial, the company doesn't want too many offers at the same time or at the wrong time. Offers and campaigns mean reducing the margin and gain less money for companies. So, these all activations might devaluate the product's worth or brand attitude. The most important indicator when talking about a sales promotion is the actual discount rate after all these activations. This indicator makes necessary the integration between all these campaign tools.

2.3 IMPORTANCE IN PRODUCT LIFE CYCLE

When a product is born and launch, it passes through a line, a cycle, we called the product life cycle. Typically, this cycle has stages and generally, every product behaves similarly in all their lives. This cycle is divided into four-part. The first part is the launch period, the development, the maturity, and the setback or decline part. The product life cycle starts with when the product is developed. In this part, the product hasn't yet ready for the place. The product is created with its unique specialties to gain demand in the market. This means that it must be attractive for a potential customer to want to buy it.

The growth part is succeeded by the first period, where the product launches the market. The product still stays unknown for many customers in the situation of fulfilling its growth. Sometimes there will be some issues in the product in the produced process, which will be able to be fixed later. When a product launch, sold at a high price, not discounted. In this part, the product is starting to place its position on the market between its competitors. Its sales and demands will start to increase in the market becoming more aware of it. In the development part, the product needs steady advertising. The price of the product needs to be adapted to stay competitive. Distribution and visibility of the product are much more important to make the product available to buy on both channels' stores and e-commerce and any point of sale. And the staff of stores should be trained about wide knowledge of product features.

The maturity part will be the most important part for all promotion activities for the product. Promotion and brand loyalty comes from customers are vital in this part due to this indicates the duration of the cycle. A product should be focused on promotional activities. The company wants to extend this part because all the spendings should achieve the product goal and should turn into income. How long this stage will be long, turning into the last death part of the product, gain a time before declining.

In the last setback part, it is expected the sales of the product drop, the company must reduce the price and give a discount because the rest inventory of the product must be sold. If the product has an expiration date this part should be fast and quick. But this part comes after all the marketing strategies applied. So, the product sells at cheap prices in the market and the only aim will sell the inventory rather than becoming waste and unavailable for buying.

The product life cycle model is crucial in retail because this indicator can be used for business forecasting, to become a strategic decision on a product based on how well succeeded in the market between competitors. Some kinds of products can even fail in the first part of the introduction and sometimes never make it to maturity and development part. That's why the time of promotion is more important if the promotion activities make before its sufficient time. The product can devaluate and can sell lower than its value, companies don't want to face to face this issue, and also

being late to apply a discount on a product can be crucial and a big mistake too. Timing and integration of promotions are stand out in retail, especially in the cosmetic area.

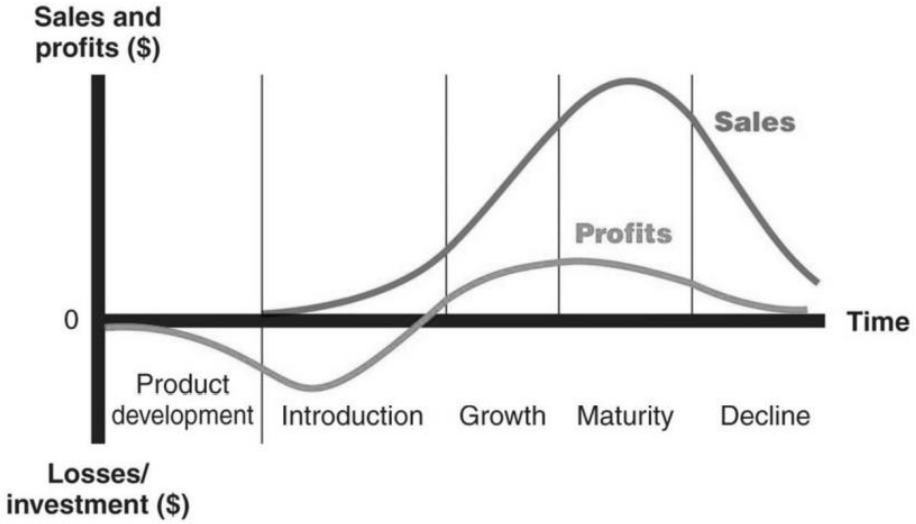

Figure 2.3. The Product Lifecycle_ was taken from [22]

2.4 SIGNIFICANCE OF SALE PROMOTIONS IN RETAIL

The significance of sale promotions increasing over time in this competitive retail industry globally. Promotion strategies are the changing marketing environment and dynamics with the new and creative ideas in the future expansion of the sales. It is becoming the survival part of the marketing strategy facing many obstacles for brands.

There are external and internal reasons we will explain in detail. Internal reasons for the company are that the promotions are acceptable due to it is effective for increasing the sales quantity and volume. The marketing and purchasing managers are responsible for the increase and achievement of the product range and sales promotion to reach the objective which was given and specified before. Reaching the targeted revenue, margin, and volume increase campaign management should be applied very strategically on time.

Externally the promotions are applied in a certain term of advertising for brand attitude and persuading loyal and potential customers. The market is incredibly competitive and flows with different varieties of products with similar competitor brands.

Customers steadily are stimulated on social media by other competitor products' advertising. Brands want to take advantage to get ahead of their competitors, changing the decision-making of customers. Choice of customers doesn't change very quickly, notably, if the customers are loyal to the brand so to gain a customer before becoming a loyal one for the competitor brand is vital and particularly important for every brand strategy.

2.5 EFFECTS ON DECISION MAKING

All steps of promotions aim to push and stimulate the customer for buying more and gain new customers. Repeated purchases are even important to keep the loyalty of customers and not lose them. To introduce a new customer requires sales promotions effects or highly advertising viral effect. bringing forth a completely new brand to a completely new customer. To create niche positioning for the brand about a specific product or product type, the company should stand out their differences and promotional power between the market in other competitive brands.

The actual results of a sale promotion can be focused on the long term or short term. It can also be affected by the store or channel choice, product category level choice, discount rate, stocking, merchandising, and new product situation of consumers. Short-term promotion effects happen during a certain time of the campaign. Long-term campaigns aim to build brand awareness, recognition, repeated purchase, influence sales, expand the targeted market and enable an opportunity to be a market leader in the competitor market.

Sales volume of products can increase during the period of the campaign, reaching the goal of switching brands and influence them to buy from the promoted category rather than other brands' products. But at this point a sales promotion may create a negative effect on the brand attitude, especially concerning advertising. Reason for gaining customers with discount effect, that sales promotion comes attractively with high promotion to consumers because of price savings. These price advantages mean that consumers may switch brands to receive another special offer from a competitor.

That’s why the promotion tools and timing are more vital for the brands and products’ target. Brands don’t want to face to face devaluating the new products which they invested a huge amount of money for it. Analyzing the customer's behaviors on time, observing competitors and the similar products’ pricing analysis, 360 degrees of marketing strategy integration are stand out in retail in the cosmetic market.

3. DATA EXPLORATION AND PREPARATION

3.1 DATASET DESCRIPTION

3.1.1 Data Overview

Our dataset consists of 12 features, 12 columns, 19897 rows and 1359 unique product sale details. We have six main category which are make-up, face care, body care, perfume, hair care and hygiene products. We indicate them on “family” coloumn in the data. The company’s operation and strategies are surrounded by these items. Every category has different dynamics in itself because of seasonality tendency, competitor price positioning and addressing different customer types. Each one of them’ price decision is made taking into account their competitors in the market and their product strength. On the other hand, every item’s cost can be unstable due to their ingredients. Some products’ cost might decrease and increase almost every month. This is the important decision maker on price positioning or price increase decision if needed.

In the figure 3.1 we can check how many instances every main category has. 51% of all comes from Make-up category because in make-up there are many color and tone options and every option are a new SKU, new instance.
Body and Hair have instances lower than the others, because of they have less variety on items, they have no color distinction or size distinction.

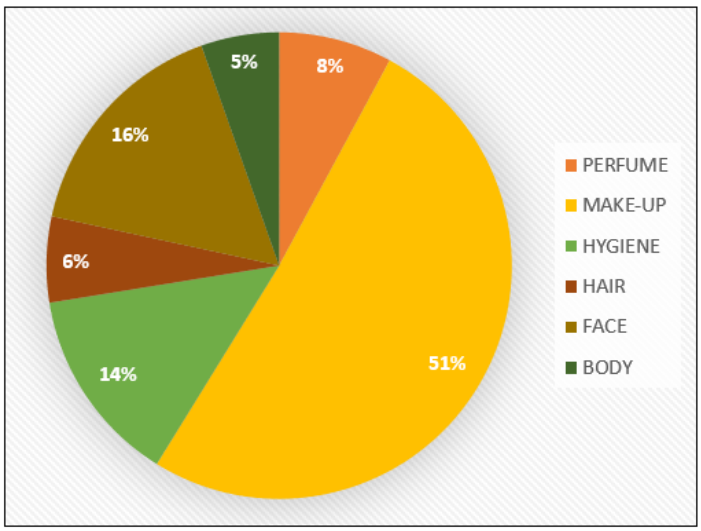

Figure 3.1. Distribution of instances based on main category

In the 3.2 figure as bellow, we can review the distribution of sale volume in our data. We saw the higher instances comes from Make-up category in figure 3.1 but in 3.2 only 11% of volume comes from Make-up group. So, there is no significant relation between number of SKU and sales volume.

Face category has the most sale volume compared by others, 29% of sale volume comes from only face category. Then hair and hygiene products dominate and sell a lot because of frequency of usage we'll mention it later.

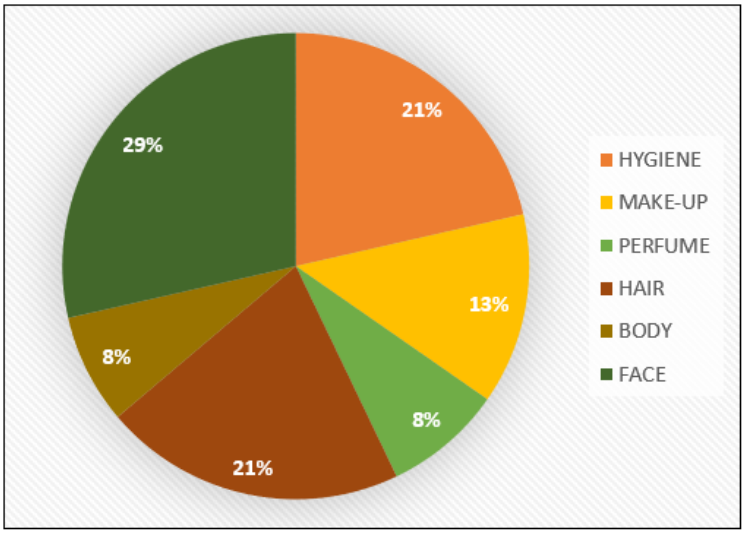

Figure 3.2. Distribution of sale volume based on main category

These six categories have twenty-one subcategories. We indicate them on "sub-family" coloumn in the data. For instance, which one of the main categories is Hair has four sub-category that are shampoos, conditioners, expert hair care and hair accessories.

Another column is product type, these are subgroups of sub-families. We have 58 types of product type for six total categories. For instance, hair styling is a product type of expert hair care.We have also sub-product type which is indicated on product sub-type coloumn in the data. We have 147 types of product sub-type for six main families. For instance, hair spray is a sub-product type of hair styling.

We have “Serie” and “Sub-serie” named column. Serie is like a range, as a specific line. For example, nutrition shampoo, nutrition conditioner and nutrition hair spray are called one seri which name is nutrition. But as we see these three products can have different subcategory also different main category. The only common ground is one of their ingredients is the same and we recommended by using these three products together as a marketing strategy mostly. Because their complementary products for each other. Sub-serie is a sub-group of series also. For example, Nutrition serie might have 2 sub-serie, names Nutrition 2019 and Nutrition 2021. This is informing us which year this serie launched. A serie might re-launch 2 or 3 years later because of re-formulation or re-packaging.

To make a clear these sub-groups and category perspective and relation, we can review in figure 3.1 as bellowed. This scheme indicates a part of hair product structure. As we see some sub-family like Shampoo might not have a product type, directly the lower level is a product name. But in Expert Care; this Sub family have a product type and also product sub-type. A shampoo and a hair spray can be in a same seri like Volume example in the below figure.

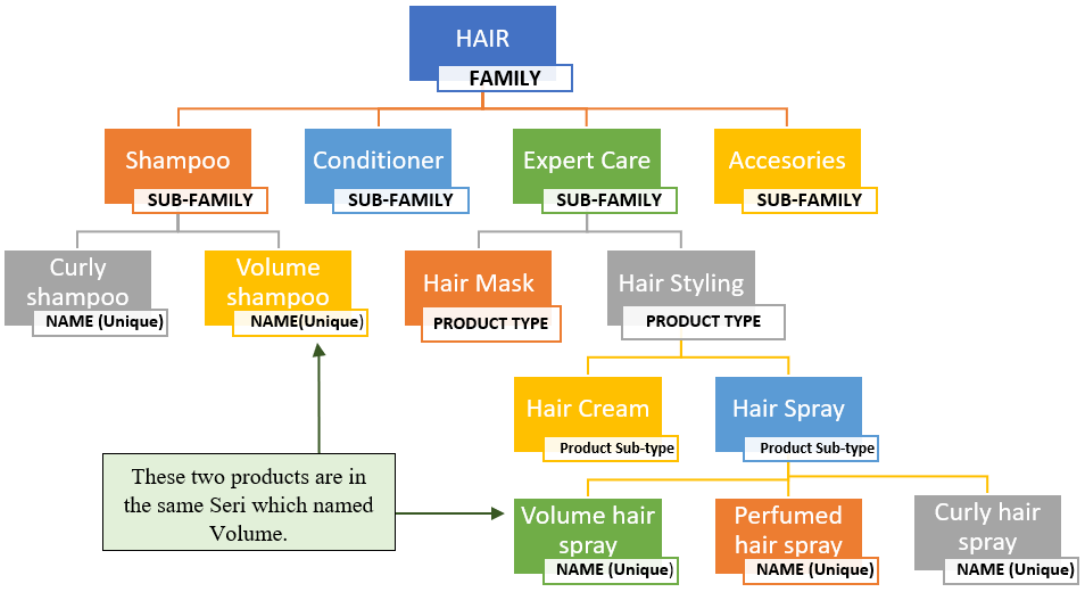

Figure 3.3. A part of hair product tree

We mentioned 7 features as above, all of them are nominal categoric values. We have 5 attributes more. Their column names are “Month, price, sales unit, revenue and discount.” I take into account the year of 2019 and 2021 12 monthly data into my

dataset. The reason not to take the year of 2020 is Covid period. The pandemic occurred in the year of 2020, because of the fact that stores had to be closed the doors for closure for many days. Price, sales unit, revenue and discount; these 4 features are numeric values, we'll detail them with dynamics.

"Price" named feature is represent a product's firt price without a promotion or discount. Price levels can vary depending on its main category. For instance, perfume items' average prices higher than face items and face items also more expensive than hair products. To get a general idea about six main category's first prices,we can check the figure 3.1.2 as bellow.

X matrices represent distribution of products' first price levels and Y matrix is sales unit of products. As we see higher than 400 , only Perfume products are seen in the table which are purple.

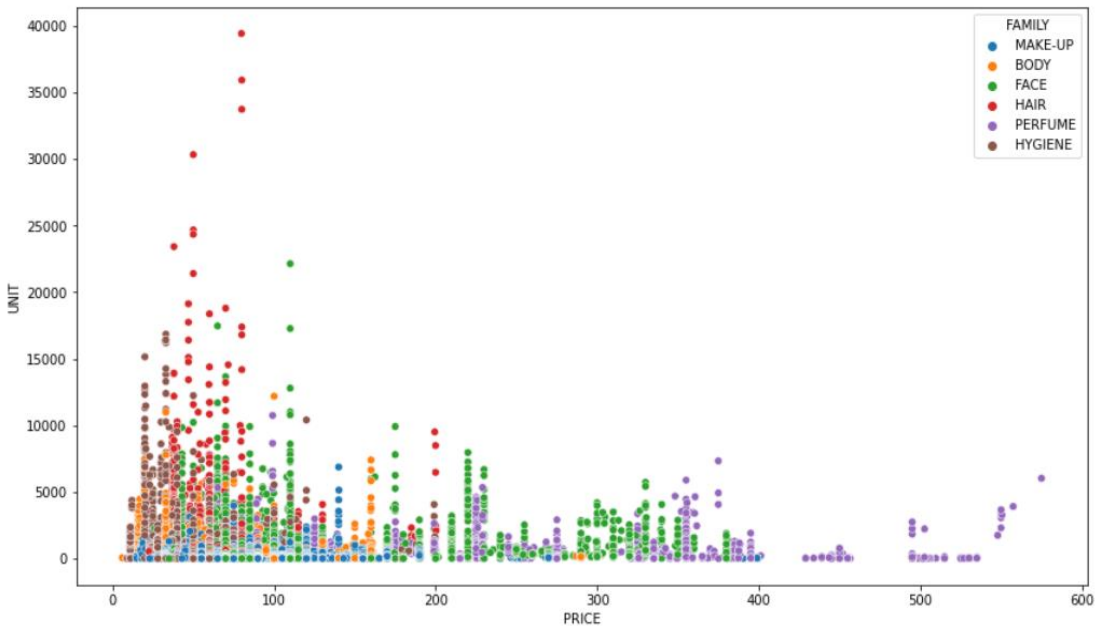

Figure 3.4. Scatter plot distribution of first price main category detailed

Between 0-100 TL we can see density of products. These products are hygiene, hair, and body products mostly and their sales are much more than the others. Consume of these products sell more often than a perfume or face care items. For example, a shampoo or a shower gel rebuy in two- or three-months period so that is the reason of mostly hygiene and hair products' sale quantity much more than rest ones.

Most important feature is “Sales unit” which means that sale volume, quantity. We’ll be predicting the volume and every detail and dynamics on volume is critical and had to observe cautiously.

In the figure of 3.5 as below, in some months and season different main category peaked and react differently routine. There is some seasonality effect on specific categories for instance, in December in both 2019 and 2021, Hygiene products boosted more than the other months. The reason is Christmas period , gift season affect especially Hygiene items.

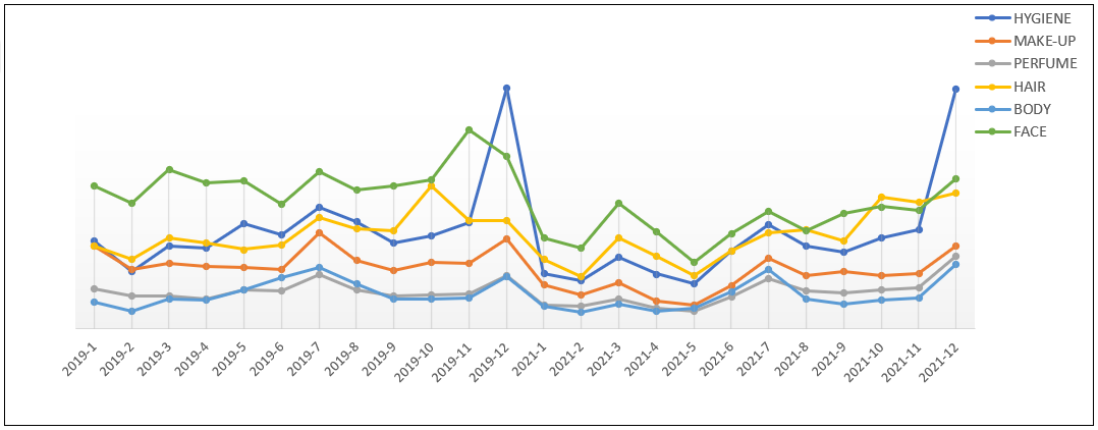

Figure 3.5 Monthly sales volume based on category

In Christmas period that is December for Turkey, people tent to buy gifts their relatives and their loved ones. This month is being very voluminous to increase the sales for retail as like all of the world.

In December Hygiene category launches new Christmas edition products also, they have some special packaging and trendy items that is also helping to boost the sales and help reaching out the targets.

Apart from seasonality in some specific terms, the company have tv commercials which supported by store windows and e-commerce and social media also. These marketing strategies and dynamics might increase the volume a lot outside the norm. For face category we see some peaks in November 2019 and in March 2021. These peaks come from tv commercial and specific campaign terms which focused on some

products in face. Store windows and social media effect is powerful for retail especially in cosmetic area. Sometimes an influencer shares a product of company in her/his link, although the company don't have a collaboration with them, in the word of mouth of product effect suddenly it can blow up the sales. Windows of stores also enormously powerful because the company has almost 250 stores and the location of these stores are strategic and attracts people and customers.

Other numeric feature is "Revenue" which means sales turnover. In the the figure of 3.6 as below, we can see the most revenue comes from Face and Perfume category, although we realize Perfume's volume was not so high. 60% of sale revenue comes from face and perfume. The only reason is the first price of these two main categories is as we mentioned before, higher than others and even though they don't sell too much, their prices enable reaching out the top for sale turnover.

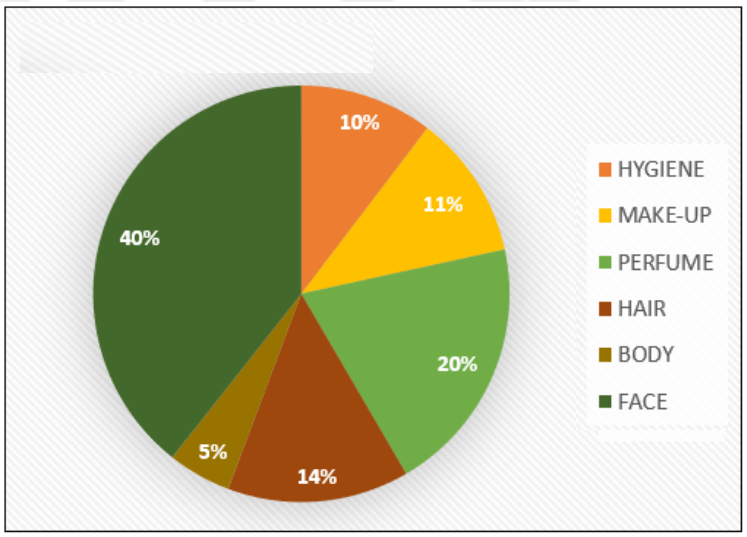

Figure 3.6 Sale revenue mix based on main category

Another thing we need to make a point is that although hygiene, bath and hair have lot of sale quantity, their mix is only 24% of sale revenue. Because of their price level is lower than face and perfume, their effect on margin and margin rate is fewer. For margin development and management, face, and perfume categories' margin – cost balance is more important than others since their weights in total.

Our last and one of the most important features is “discount” which means actual discount rate that were applied indicated month. Discount rate and sale quantity have positive correlation, we’ll mention this point following chapter.

Significantly, we think about from customer part if a company or brand gives offer higher discount on their product, this stimulate purchasing behavior of people and to have an advantage on purchasing and in consequence the sales will increase.

In the figure 3.7 as below, we can review that discount rate applied between which range mostly in monthly detail. As we see there are a few products which has no discount, discount rate 0% items in every month. Generally applied discount rate is aggregated between 20%-40%. There are a few products which they have more than 60% discount. These high discounted products might be destockage items which means old stocks, or which have limited expiration date, so they are consumed and bought promptly. Giving a high discount on these kinds of items is one of the highest decision and strategy to sell them faster .

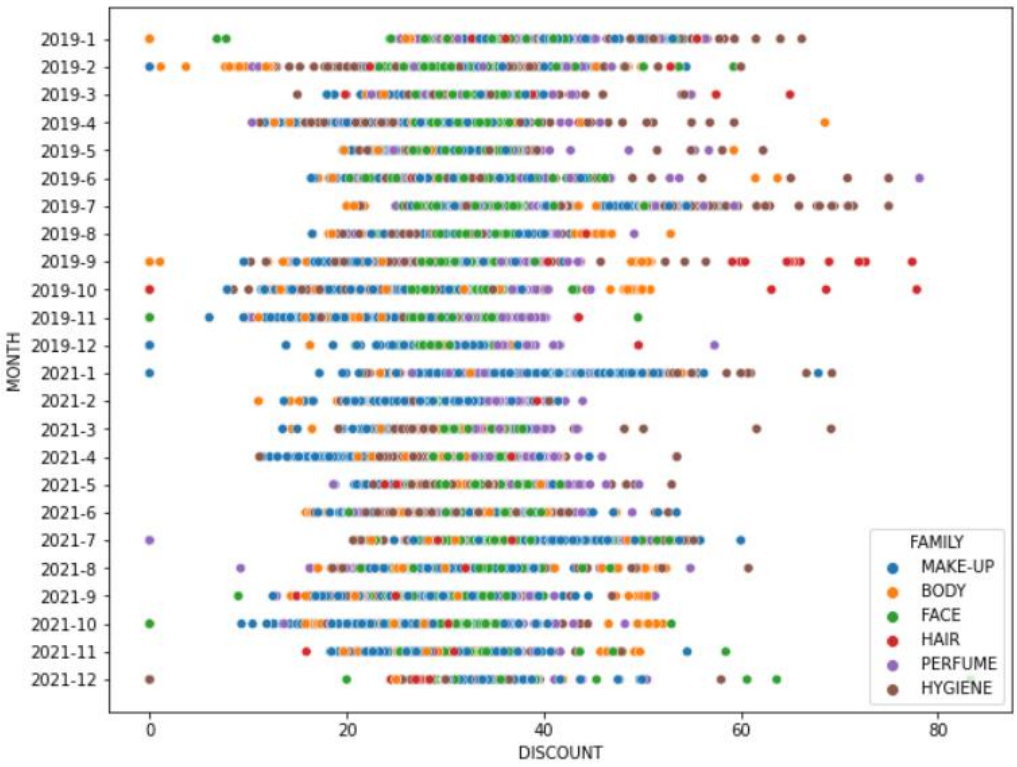

Figure 3.7. Scatter plot distribution of discount rate monthly detail

I want to check category basis discount rate distribution so the box plot in figure 3.8 gives an idea for this question. There are significant outlier items the reason is as I mentioned previous paragraph, especially there are many outliers in Make-up and Hygiene category, these categories have many SKUs. These two categories have lots of products which have many colors and scent variety and to consume old SKUs, high discount rates are given.

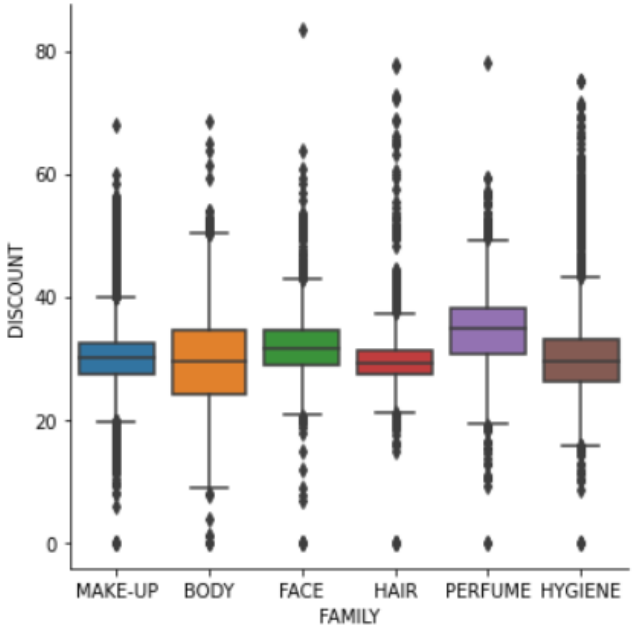

Figure 3.8. Box plot category discount rate distribution and outliers

Mostly in category detail, median values of all categories are close to each other. The median values have 30-35% discount rate for every category. Perfume is differentiating, it seems its discount rate close to 40% the reason is that there is a CRM birthday campaign strategy. Customers must buy 2 get one free campaign during their birthday month and mostly people prefer to buy perfumes when they use the campaign because of the highest price of perfume, they wanted to reach maximum advantage. So, a large part of perfume discount come to close 50% with using birthday campaign because perfume median discount level close to 40%.

For body category lower and upper limit have long widths as we can see the figure 3.8. The reason is that sun care products are in body category, most of the year especially in summer season they have 50% discount and sun care weight in body is not least, therefore upper limit close to 50%.

3.1.2 Features Used in Model Development

There are multiple types of factors that can make the model of Machine Learning more effective on many given tasks. One of the mostly used methods of feature selection is data correlation. Data correlation is a main impact on a model's performance. This will reduce the stress and pressure on the Machine Learning model during preprocessing and cleansing the data. Chosen data attributes for training the Machine Learning model would have an important impact on the efficiency of the model. Because of the irrelevant features that are presented, the model output will be decreased. Feature selection method enables an efficient way to remove data redundancy and irrelevant data that helps to reduce calculation time, improve, and develop accuracy, and build up understanding of your model [12].

Figure 3.9 Heatmap

The selection of features is critical in classification. Feature selection techniques are designed to improve classification efficiency by selecting the crucial features from the data sets.

Data correlation is a method that helps to predict one attribute from another attribute and is used as a basic quantity in many modeling techniques. If one feature increases, the correlation will become positive, so the other feature increases as well and negative if one feature increases there will be a decrease in another. If there is no relation

between any two attributes, then it is said to be no correlation[13]. If there is a linear relationship between the constant variables, then the Pearson correlation coefficient is used. If there is a non-linear relation between the constant variables, then the Spearman correlation coefficient is used. Since the considered data set is linear so the Pearson correlation coefficient is used for the selection of features in this study. This correlation for all the attributes will be shown in the one of the next parts.

It is a statistic measuring the linear correlation of two variables X and Y. It has a value between +1 and 1, where 1 is a linear positive correlation, 0 is not a linear correlation and 1 is a linear negative correlation[14]. The motivation for considering the correlation is when people know a score on one measure, they can make a prediction of another measure that is highly related to it more accurate. The more accurate the prediction, the stronger the relationship between the variables [15].

Table 3.1

Data set and feature details

Feature Name	**Description**	**Type**
Product Name	Product's special name	Categorical
Family	Main category of the product	Categorical
Sub Family	Sub category of the product	Categorical
Serie	Serie of the product. More than one products' serie is the same which are used for only one purpose	Categorical
Sub Serie	Means the launch the serie of the product.	Categorical
Product Type	Lower level of sub category of the product	Categorical
Product Sub Type	Lower level of product sub type of the product	Categorical
Month	The product' actual figures occur in the specific period	Numeric
Price	First price of the product, before the discounts	Numeric
Sales Unit	The product's volume that sold by in the period	Numeric
Revenue	The product's turnover that sold by in the period	Numeric
Discount Ratio	The product's reduce price ratio in the period	Numeric

In the table of 3.1 we want to expound our feature’s explanations. As we can see, we have categorical and numeric values, mostly categorical values we have.

Our numeric attributes' statistical details and summary indicated in table 3.2 as below. Mean value of discount rate calculated 31.4%, as we mentioned before mostly products aggregated close to 30% and some products were not given a discount that we can see the min discount rate value is 0%, and max value is calculated 83% which comes from some outliers items. The mean value of sales unit calculated 633, min value is 1. We ignored the products which are have no sale, max value calculated 39.419.

Table 3.2

Statistical summary of numeric columns

	PRICE	SALES UNIT	REVENUE	DISCOUNT
Count	19,897	19,897	19,897	19,897
Mean	90.17	633	37,357	31.39
Std	80.19	1,441	96,343	7.31
Min	5.89	1	28	0.00
25%	42.90	42	1,859	27.5
50%	64.90	156	6,951	30.33
75%	103.43	615	30,201	33.55
Max	575.00	39,419	2,246,115	83.3

3.2 DATASET PREPARATION

3.2.1 Data Cleaning

Data cleaning is the process of editing, correcting, and structuring data within a data set so that it's generally uniform and prepared for analysis. This includes removing corrupt or irrelevant data and formatting it into a language that computers can understand for optimal analysis. There is an often repeated saying in data analysis: "Garbage in, garbage out," which means that, if you start with bad data (garbage), you'll only get "garbage" results. Data cleaning is often a tedious process, but it's absolutely essential to get top results and powerful insights from your data. [16]

In machine learning, data scientists agree that better data is even more important than the most powerful algorithms. This is because machine learning models only perform as well as the data they're trained on. If you're training your models with bad data, the

end analysis results will not only be generally untrustworthy, but will often be completely harmful to your organization. [16]

Effective data cleaning process enable saving time and money and make the companies' organization more efficient. And also it allow using the same data sets for multiple analyses and functions.

3.2.1.1 Remove irrelevant data. I needed to figure out firstly what analyses I'll be running what are the relevant instances and what I'll not need. I filtered out data and extracted some of items which are not significant for my analyze. For example there were packs, pocket and bags in my sales dataset but I want to predict these items because of they are consumable materials and some of them are not selling with money like a gift. I extracted these kind of product types from my dataset.

3.2.1.2 Deduplicate the data. In the case of collecting data from multiple sources, one the problem that analysts face to face is duplicates. This is causing slow down your analysis. More to the point if we'll train a machine learning model on duplicated dataset the model will mostly give more weight to the duplicates, depending on how many times they were duplicated. So we need to remove them from our dataset for good-balanced results.

My problem for duplicates is that one unique product might have more than once sometimes because the name of products are changing when they relaunch. For example in next month due to the name of the product is changing, in analysis algorithms behave it, like a new product I could not singularize these products because of there is no significant link between old and new names of these products. But if the names are same or closer to each other I can singularize them at least most of them are behave single SKU.

3.2.1.3 Missing data. When I check my dataset, noticed we have some missing data on especially some categoric columns, for example in sub-product type or sub family detail. Because some products don't need to detailed third or forth level. Then I entered the cells manually in excell to use a formula and we get non-null cells for every columns.

3.2.1.4 Outliers. I have many outliers in our data as we saw in the box plot that I mentioned and indicated previous chapter. Depending on campaign period and playing window visuals or tv commercials the star product's sale might boost and we have 16-17 campaign terms in a year so the situation is faced many times in a year for different products.

Ignoring these outliers are not significant because I want them into my dataset to reach the effective prediction, so I keep on my dataset. But I ignore some rows which have low sales because of they are delist or stock-out items. Which means that these products will not proper to sell in the coming time so there is no need any forecasting for these items. The production of these is stopped mostly and there will be no demand for them.

3.2.2 Encoding categorical values

Categorical data contains label values that are considered nominal values. Each value has categories of different types. Besides, a few of the groups have a normal relationship with each other is known as natural ordering. The categorical data can be converted into numerical data to improve the efficiency of the Machine Learning model [17].

One hot encoding is the method where the data is represented in binary format and included as a feature. It is one of the most common methods, comparing each level of the numerical variable with a fixed starting point. In this thesis for the data set that had taken, one hot encoding is used to represent categorical variables as binary vectors [17].

We have seven attributes are categorical and need to convert binary values so I use one hot encoding. We can review in the figure of 3.3 and 3.4, the results and values when before and after applying encoding to family feature which is main category.

Table 3.3

Before one hot encoding

AILE	ALT_AILE	SERI	ALT_SERI	URUN_TIPI	ALT_URUN_TIPI
MAKYAJ	Ten Makyajı	Couleurs Nature	Pure Light	Pudra	Toz Pudra
VUCUT	Dudak Bakımı	Baume Levres Soin	Baume Levres Soin	Dudak Bakım	Dudak Balsamı
MAKYAJ	Ten Makyajı	Couleurs Nature	Pure Light	Pudra	Toz Pudra
VUCUT	Ayak Bakımı	Beaute des Pieds	Beaute des Pieds	Ayak Bakım	Ayak Peelingi
VUCUT	Dudak Bakımı	Baume Levres Soin	Baume Levres Soin	Dudak Bakım	Dudak Balsamı

Table 3.4

After one hot encoding

AILE_HİJYEN	AILE_MAKYAJ	AILE_PARFUM	AILE_SAÇ	AILE_VUCUT	AILE_YUZ
0	1	0	0	0	0
0	0	0	0	1	0
0	1	0	0	0	0
0	0	0	0	1	0
0	0	0	0	1	0

4. METHODS AND ANALYSIS

4.1 METHODS

4.1.1 Linear Regression

Linear regression is divided into simple linear regression and multiple linear regression. Simple linear regression (Formula 1) is an approach for predicting the response variable Y using a single predictor X. The predictors are called the independent variables and the response variable is called the dependent variable because it depends on the input. In contrary to simple linear regression, Multiple Linear Regression (Formula 2) uses multiple predictors to estimate the response variable Y [18].

$$Y = \beta 0 + \beta 1X + \epsilon \ (1)$$

$$Y = \beta 0 + \beta 1X1 + \beta 1X2 + \cdots + \beta pXp + \epsilon \ (2)$$

The coefficients of each attributes are not known. We want to need data to estimate these coefficients. Linear regression's aim is to find good estimator for each independent variable, the best prediction is obtained for the dependent variable. The objective of linear regression is minimazing the residual sum of squares. Besides, this method assumes that there is a linear relationship between the response and predictors [19].

The advantage of linear regression, algorithm is not complexible. There are also some disadvantages and potential problems in linear regression. Linear regression does not have a good performance if the variables have a non-linear relationship, are correlated, do not have a constant variance in the error or if an observation in the dataset is an outlier. Other fitting procedures may result in better prediction accuracy and model interpretability. Linear regression finds a coefficient for all predictors that are given as input. However, it often occurs that many of the input predictors do not have a relation with the response variable. This leads to unnecessary complex models and sometimes a lower performance. Model selection and regularization methods can reduce the number of variables of a linear model [19].

4.1.2 Decision Trees

Tree-based methods are simple to create and useful for interpretation [19]. Decision trees can be applied to regression and classification problems. A decision tree consists of a series of splitting rules. Once an observation followed a sequence of splits, a prediction is made for that observation. All observations from the data that followed the same sequence have the same prediction. A decision tree 32 consists of internal nodes and terminal nodes. An internal node represents a splitting criterion, and the terminal node consists of a prediction.

In the context of this research, an internal node could be the retailer or price, and a terminal node is the forecast of the promotional volume. The splitting criterion is determined in such a way that the Residual Sum of Squares (RSS) is minimized. It is often computational infeasible to consider all possible splits in the tree. Therefore, a greedy approach is used to construct a tree. The split is created at the predictors that have the highest impact. The next splits are based on the remaining set of predictors. This is a greedy approach because it only evaluates the possibilities at that moment, and it does not include future steps. The variables with the highest impact on the prediction are at the beginning of the tree. This process produces good estimates for the training data, but it often results in lower test performance due to overfitting (high variance). The decision tree is fitted well for the input data, but these relations do not have to hold for new datasets. Therefore, a decision tree results often in high variance due to overfitting. Other regression tree methods with a lower variance are random forests, bagging, and boosting.

4.1.3 Random Forest & Bagging

The Random Forest, which is a method based on the decision tree [20]. A Random Forest is simply building many decision trees, and it has two important characteristics. The number of trees that are built is called ntree. The first characteristic is that a Random Forest does not use all data as input, but it samples a new subset from the original dataset with replacement. This means that one observation from the original dataset might occur multiple times in the subset (called bootstrapping).

The advantage of the Random Forest is all data can be used for model building. The second characteristic of a Random Forest is that a subset of variables is chosen at each split. Because a subset of predictors is chosen, the individual trees are less correlated resulting in a better overall performance. The number of variables that are chosen at each split is called mtry. If a Random Forest is used for regression, then an often-used value for mtry is equal to the number of predictors divided by three. A Random Forest that uses all variables as input for each split is called bagging [19].

Other advantages of a Random Forest are the easy interpretability, it can be solved for classification and regression problems, it can handle large datasets with high dimensionality, it can identify the most significant variable, and it outputs the importance of each variable. A disadvantage of a Random Forest is that it is a black-box model. This means that we know the input and the output of the Random Forest, but we do not know how the model derives the output from the input [18].

4.1.4 XGBoost

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now [21].

The answer of why does XGBoost perform well is that GBoost and Gradient Boosting Machines are both ensemble tree methods that apply the principle of boosting weak learners using the gradient descent architecture. However, XGBoost improves upon the base GBM framework through systems optimization and algorithmic enhancements [21].

4.2. BASELINE CALCULATION

This study aims to develop a model for promotional forecasting per product. To calculate the total volume, we need baseline sales, seasonality and commercial effects of products.

Baseline sales are calculating by some kind of manuel simple mind. To predict 2021 monthly forecast, we consider 2019 monthly values with a formulated way of same season. We don't take into account 2020 monthly values because of covid pandemic. In 2020 untill June, in most of the months stores had to close the door and there was closure globally. In these terms e-commerce channels were active but people were not tend to shop from e-commerce compared to stores so sales were dropped dramatically. After June, stores' sales were not normal again, due to the boosting sales in the stores. People bought and shopped from the stores because they had postponed and waited their needs.

For 2021, forecasting monthly demand was also complicated because till the June there was partially closure for the stores of the brands as well. And specificly for the company we observed October, November and December were also abnormal months. The TV commercials and investment began in October and proceeded till December.In December for Turkey the currency crisis happened and mostly companies increased their prices because of boosting their costs. But the company we observed kept the prices same so in December people and customers bought much more than they need and they stockpile the products. One of the reason boosting the sales in December was Christmas period. In December 2021, also the sales increased abnormally because of gifting season.

We currently calculate September in 2021 sales, taking into account July in 2021 and August in 2021 values and the data of July- August and September in 2019 monthly sales. We take the seasonality effect to consider the actual sales of three months in 2019. We calculated how is the July in 2021 differs compared to 2019 July and how is the August in 2021 differs compared to 2019 August then we reach a coefficient to use for forecast the September in 2021.

Our baseline formula details are belowed. We reached out 44.9% for MAPE with our baseline method.

Sales unit_2019_avg = (Sales unit _2019_June+ Sales unit _2019_July+ Sales unit _2019_August)/3
Sales unit _2021_avg = (Sales unit _2021_ June + Sales unit _2021_ July + Sales unit _2021_August)/3

Prediction = (Sales unit _2021_avg/ Sales unit _2019_avg)* Sales unit _201

MAPE : (actual – pred)/actual -1

5.EXPERIMENTAL RESULTS

5.1. Random Forest

A random forest does not build one decision tree, but it builds many decision trees based on a bootstrapped dataset. We use 10-fold cross-validation to validate the results of each experiment.

The previous sales units and first prices are very important for determining the volume of a product. Therefore, the best results are obtained when these predictors are always included. Because all predictors are considered at each split, there is a high probability that each time the same predictors are chosen at the same split. This has as a consequence that the value for ntree does not have a large impact since the trees are highly correlated. This is not a big problem, because each decision tree is based on a new bootstrapped dataset. A Random Forest with all predictors as input is called a bagged tree [21].

We have 410 instances for the predicting september sales unit. Average percentage error of the products is 30.7%. For the baseline method this value was 44.9% as we mentioned and detailed in the previous section. So we improve and reduce the percent error almost 15% percent which seems very effective. Touching some keypoints and reducing a few outliers we'll reduce the error more then 30.7%. For all algorithms sklearn's default parameter values are used, no hyper parameter tuning is performed, we can see the MAPE values in table 5.1 as below.

Table 5.1
MAPE results for algorithms

	MAPE
Linear Regression	58.8%
Decision Tree	38.1%
XGBoost	35.2%
Random Forest	30.7%

In Table 5.2 we observe random twenty-five products results and how the model predict well. As we see some of our prediction is perfect almost the percentage error calculated 0.2%, some of them are reaching out 75% and more. On the basis of main group, how is the percentage error differs we can see in the Table 5.2. For body family percentage error higher than others, it is almost 69%. We'll mention the reason of this issue in the following chapter.

Table 5.2

Sample results applied Random Forest

FAMILY_	Actual	Pred	Percent_Error	PRICE
BODY	980	1,721	75.6%	160.46
MAKE-UP	21	36	71.0%	89.90
MAKE-UP	20	34	68.3%	164.97
MAKE-UP	140	192	37.2%	111.61
MAKE-UP	77	105	36.8%	84.90
FACE	1,007	647	35.8%	59.98
FACE	218	286	31.3%	85.10
HAIR	2,718	3,453	27.0%	60.34
FACE	6,214	4,711	24.2%	85.76
MAKE-UP	56	70	24.2%	125.47
MAKE-UP	277	344	24.1%	100.16
HYGIENE	1,096	1,344	22.6%	50.04
MAKE-UP	154	182	17.9%	120.55
HAIR	3,226	3,795	17.6%	79.76
MAKE-UP	218	253	16.1%	90.14
MAKE-UP	79	67	15.5%	85.97
MAKE-UP	74	85	14.5%	122.34
MAKE-UP	26	22	14.3%	89.90
MAKE-UP	30	27	11.7%	90.73
FACE	1,037	935	9.8%	99.15
BODY	2,253	2,074	7.9%	99.95
FACE	79	74	5.8%	74.90
PERFUME	110	114	3.8%	69.90
HYGIENE	1,029	1,061	3.1%	27.95
MAKE-UP	400	388	2.9%	64.95
HAIR	4,337	4,344	0.2%	59.91

Table 5.3

Main group basis percentage error

FAMILY	Average of Percent_Error	SKU
HYGIENE	26.9%	31
MAKE-UP	28.8%	241
PERFUME	22.3%	23
HAIR	20.5%	28
BODY	68.7%	24
FACE	33.0%	63
Grand Total	**30.7%**	**410**

We observed having a few outliers thet we couldn't reject and eliminate in data cleaning step. As bellowed five products errors are highest ones and we want to reduce them and check the results after reducing. The percentage error decrease 26.9% from 30.7% after reducing these five items which means almost 4 points improvement.

In Figure 5.2. we can observe which group error decrease after the step and how was the results. The error of body group dropped to 30.1% from 68.7% which means decreasing half-way. The error of face group dropped to 25.7% from 33.0% this is also important because most important group to avoid prediction deviation is face family.

Table 5.4

Reduced outliers

FAMILY_	Actual	Pred	Percent_Error	PRICE	Reducing Reason
FACE	497	2,909	485.3%	70.06	Delist Item
BODY	511	2,882	464.1%	161.00	Demand Transition
BODY	116	462	298.2%	131.02	Demand Transition
BODY	15	53	255.4%	129.90	Delist Item
MAKE-UP	34	104	205.9%	154.90	Demand Transition

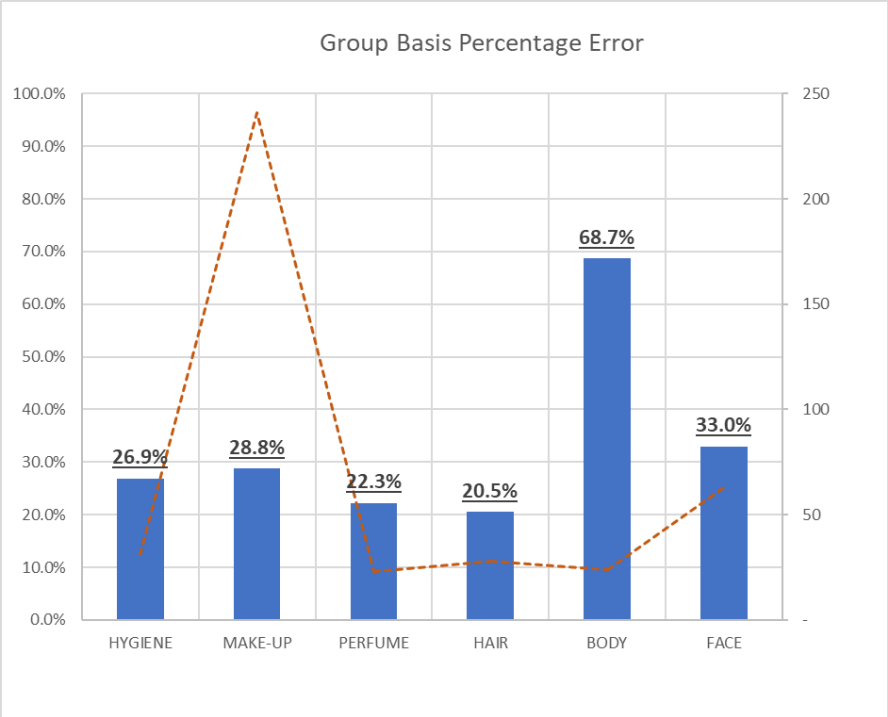

Figure 5.1. Graph of main group's error

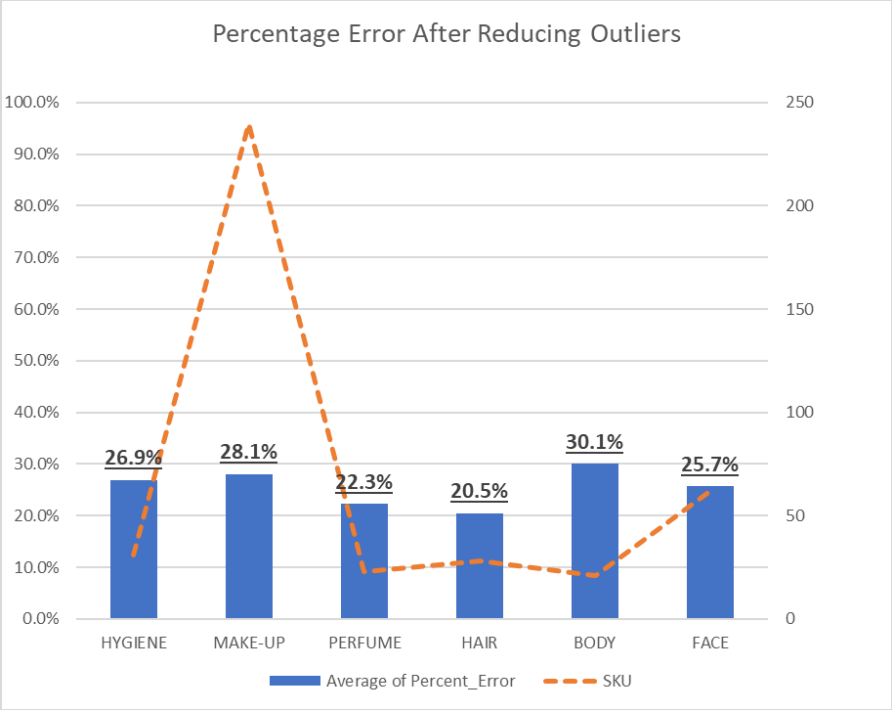

Figure 5.2. Main group's error after reducing outliers

In the Figure 5.3 as bellow although some predictions have no big errors their importance in demand is critical because of their volüme and values are too high. Every deviation of predicting on these items caused by overstock problem. So the demand planners need attention more than on them compared to the rest of. We can see them on the right side and below the line.

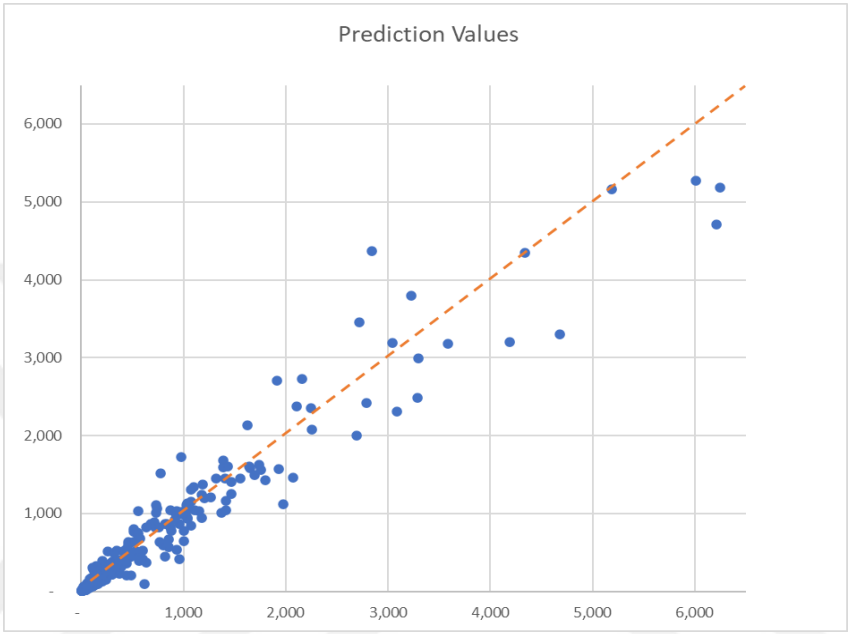

Figure 5.3. Scatter plot of predictions

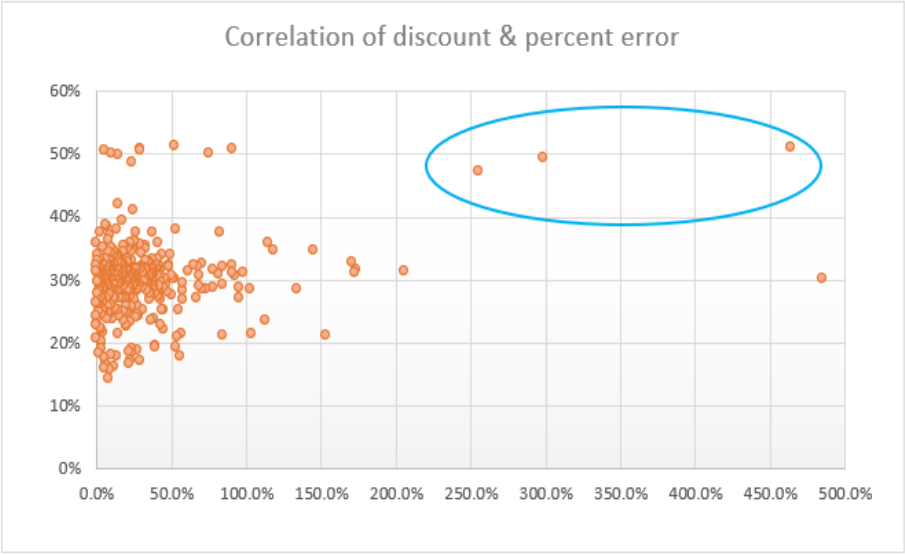

Figure 5.4. Distribution of errors in detail of discount

In the Figure 5.4, I want to check on if is there any effects of discounts on outliers. Because if applied too high discount on an item than unusual, the item can react and boost its sales. As we can see, three of them which circled with blue line, have discount nearly 50%. But the reason of these products have big error isn't, they have high discount. Because of the items are delisted which means they produced no more, as a strategy it is given more discount for their few stocks to consume.

In the figure of Figure 5.5 there is an item indicated in Face group which has shortage problem. The importance of realizing these kind of items are crucial. Because we couln't catch the items taking into account their errors because their errors calculated not so high. But for example the highlighted one predict as 6.768 but actually sold 3.803 units. Although the percentage error of this item is not so much 78%, the units are significantly important. The difference prediction and actual value is the shortage. Shortage means that because of there is no sufficient products in the store, loss of sales are consist of. In this case we can't talk about prediction error, we should calculate shortage and should try to minimize it.

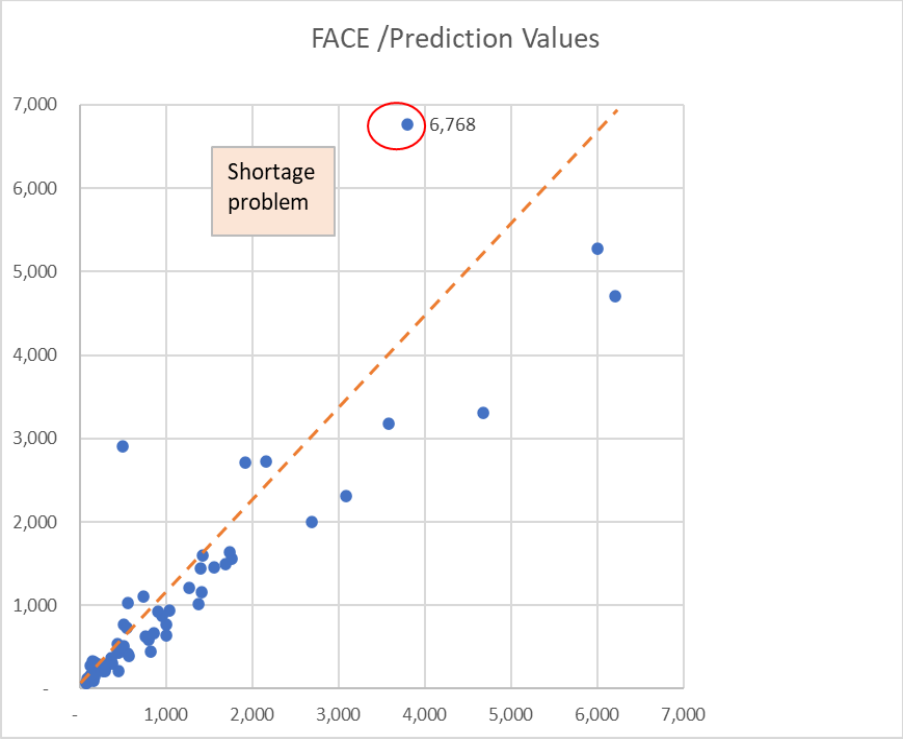

Figure 5.5. Face group detailed predictions

In the figure of 5.6 we want to check on if is there any effects of discounts on outliers in Face group. Because if applied too high discount on an item than unusual, the item can react and boost its sales. One of them which circled with blue line, have discount 30% and percentage error of prediction is 500%. The reason of the high error is this item is delisted, not produced any more. There is no significant discount and error relation on it.

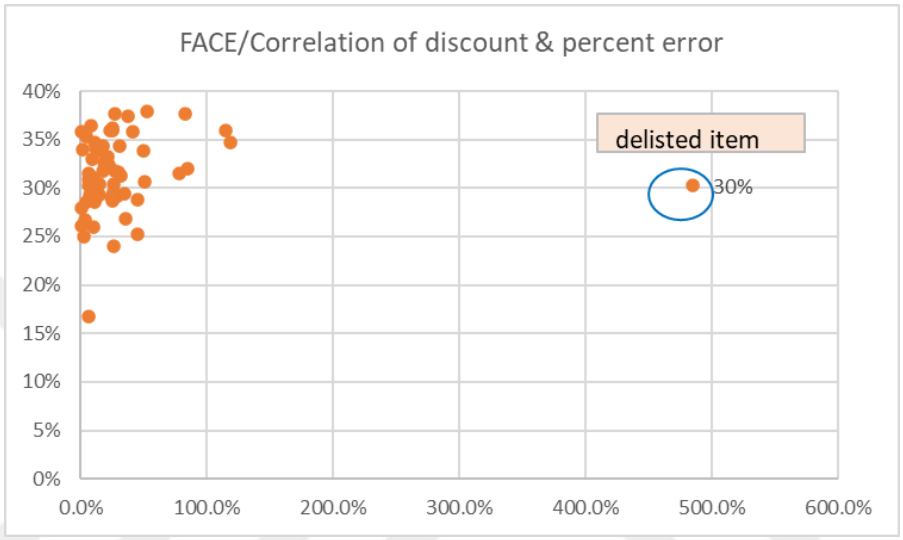

Figure 5.6. Distribution of errors & discount in face detail

6.CONCLUSION

The current demand planning process results need lots of quick fixes, many meetings and time consuming process for the marketing and demand planning departments and to obtain a good forecast. In the current situation, there is forecasted on the total demand but there are high deviation for predicting volume of product and promotional demand. Besides, there is not much accuracy in the current process caused by low scoring from the center of company where the all products produced.

The developed baseline method is using to determine the demand for now, and demand planning department implement this method in the current process using basic calculations in Microsoft Excel. The method uses historical products' sales input. Based on the promotion pressure, it is determined with the same period which close the same discount level. We clean and organized the new and delisted items' sales and outliers to obtain a baseline. The discount rates unfortunately are not so significant to predict the sales unit because promotions and discount strategy are not various values mostly they are close to each other.

We used family as categoric variable, price and revenue variables and month data as input for the prediction model and tested which variables are most important for predicting the volume. We finalized and determined that product name, month, family, price and previous month sales units are the most important variables. We didn't apply hyperparameter-tuning for all algorithms because of our instances are not so many and there was a risk to face with overfitting problem.

We used 10-fold cross-validation to measure the performance of our forecast models. The performance of the current forecasting method and the Random Forest model 's accuracy difference is significant. We reached out almost 20% improvement, from 44.9% to 25.7% for MAPE.

Our new model is more efficient and have low MAPE for Perfume and Hair families. Error value of the Perfume family reached 22.3% and reached 20.5% for Hair family. These two families will reach the best results by using our new model the reason is

that, these two groups items are change and relaunch so much times and thanks to that we observe their sales in time in stable and fair conditions. For body and make-up categories MAPE values are 30.1% and 28.1% which are also higher than the rest of. As we explained before for these two families there are too many items that have transitions and relaunches mcuh more. So we couldn't catch the transition demand so easy and effective.

Our new model needs historical data as input to make a prediction. In case of new launch product forecasting due to we have no sales data or not much sales data available, we calculated a simulated prediction through a product which affected close and the same sales unit level. But in our new model to obtain a good forecast for new products is diffucult and complicated. This means that the forecasting model is not able to make a prediction for new launch items. The supply chain planning department have a high deviate percentage error for this type of products' forecasting. I recommend doing more research in the field of new product forecasting is crucial.

REFERENCES

[1] M. I. Gómez, V. R. Rao, and E. W. Mclaughlin, ‘‘Empirical analysis of budget and allocation of trade promotion in the U.S. supermarket industry,’’ J.Marketing Res. vol. 44, no.3, pp. 410–424, 2007.

[2] K. L. Ailawadi, B. A. Harlam, J. César, and D. Trounce, ‘‘Promotion profitability for a retailer: The role of promotion, brand, category, and store characteristics,’’ J. Marketing Res., vol. 43, no. 4, pp. 518–535, 2006.

[3] M. Natter, T. Reutterer, A. Mild, and A. Taudes, ‘‘Practice prize report - An assortment wide decision-support system for dynamic pricing and promotion planning in DIY retailing,’’ Marketing Sci., vol. 26, no.4, pp. 576–583, 2007.

4] R. C. Blattberg and S. A. Neslin, Sales Promotion: Concepts, Methods, and Strategies. Englewood Cliffs, NJ, USA: Prentice-Hall, 1990.

[5] R. C. Blattberg, R. Briesch, and E. J. Fox, ‘‘How promotions work’’ Marketing Sci., vol. 14, no. 3, pp. G122–G132, 1995.

[6] K. H. van Donselaar, J. Peters, A. de Jong, and R. A. C. M. Broekmeulen, ‘‘Analysis and forecasting of demand during promotions for perishable items,’’ Int. J. Prod. Econ., vol. 172, pp. 65–75, Feb. 2016.

[7] H. Louis, A.X. Sun, and O. Urcan, ‘‘Do analysts sacrifice forecast accuracy for informativeness?’’ Manage. Sci., vol. 59, no. 7, pp. 1688–1708, 2013.

[8] Bohdan Pavlyshenko. Using Stacking Approaches for Machine Learning Models. In Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2018.

[9] C. A. Palacios, S. M. Romero, J. L. Rojo-Alvarez, “Forecasting Promotional Sales Within the Neighbourhood” Universidad Politécnica de Madrid, June 2019.

[10] Armstrong, J.S. Combining forecasts: The end of the beginning or the beginning of the end Int. J. Forecast. 1989, 5, 585–588.

[11] James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An Introduction to Statistical Learning; Springer: Cham, Switzerland, 2013; Volume 112.

[12] Chung-Jui Tu, Li-Yeh Chuang, Jun-Yang Chang, Cheng-Hong Yang, et al. Feature selection using pso-svm. International Journal of Computer Science, 2007.

[13] Tao Zhang, Tianqing Zhu, Ping Xiong, Huan Huo, Zahir Tari, and Wanlei Zhou. Correlated differential privacy: Feature selection in machine learning. IEEE Transactions on Industrial Informatics, 2019.

[14] Pearson documentation https://en.wikipedia.org/wiki/Pearson_ correlation_coefficient Accessed: 2022-04-25.

[15] Sai Nikhil Boyapati, Ramesh Mummidi. Predicting sales using Machine Learning Techniques. Faculty of Computing, Blekinge Institute of Technology, Karlskrona, Sweden, May 2020.

[16] Monkeylearn blog. What is data cleaning? https://monkeylearn.com/blog/data-cleaning-steps/ . Accessed: 2022-05-22. (Para. 3-5)

[17] Kedar Potdar, Taher S Pardawala, and Chinmay D Pai. A comparative study of categorical variable encoding techniques for neural network classifiers. International journal of computer applications, 175(4):7–9, 2017.

[18] Tom Weustink, Forecasting promotional demand in the FMCG industry. Industrial Engineering and Management (MSc), September 2020.

[19] Hastie, T., Tibshirani, R., James, G., & Witten, D. (2006). An Introduction to Statistical Learning, Springer Texts. https://doi.org/10.1016/j.peva.2007.06.00

[20] Breiman L. (2001). Randomforest 2001. Machine Learning. https://doi.org/10.1017/CBO9781107415324.004

[21] Vishal Morde. XGBoost Algorithm : Long May She Reign!Medium platform. https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d

[22] Marketing-Insider (2020), Product life cycle strategies (PLC) and characterist