5-2022

SKU Time Series Forecasting Methods for FMCGs

Mohammad Al Orbani
mma4034@rit.edu

SKU Time Series Forecasting Methods for FMCGs

by

Mohammad Al Orbani

A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

May 2022

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Capstone Approval

Student Name**:** Mohammad Al Orbani

Graduate Capstone Title**:** SKU Time Series Forecasting Methods for FMCGs

Graduate Capstone Committee:

Name: Dr. Sanjay Modak **Date:**

Chair of committee

Name: Dr. Ioannis Karamitsos **Date:**

Member of committee

Acknowledgments

I had like to thank Dr. Ioannis Karamitsos as well as Dr. Sanjay Modak for their thorough support throughout the Capstone project. As well as every RIT Professor that has taught in the Data Analytics course as they helped shape the person I am today where I can finally give back to the world with my knowledge and help organizations drive digital transformation for a better tomorrow. I had like to thank my family for their thorough support and encouragement to pursue a master's in what I stand for and advocate. Finally, I had like to thank my peers and work colleagues who helped me through and through and for always making time to address any doubts I had because I was an entry-level in the market.

Abstract

This research aims at using forecasting algorithm that predicts the demand that is to be needed on a monthly basis while factoring in occasional inconsistent patterns, seasonality, and non-stationary and cyclical patterns of the data. The prediction is to predict around 3000 SKUs in 19 end markets and since the data is necessary for marketing enhancement and strategies, the Forecasting accuracy must be high. Since market strategies will be based on those predictions and revenue will be lost in the case of an error. Hence, we need to keep in mind that the model is not overfitted and that it wouldn't give a reasonable accuracy when tested on another SKU. In this study, I will use encrypted data from the organization as such the name SKUs are in numbers instead of names where the trends are there while the region and SKUS will remain undisclosed as well as the numbers wouldn't be the same. The algorithms used were FBProphet and SARIMA for the given SKUs. They were able to forecast at a MAPE accuracy of 77% and 87% respectively.

Keywords: FMCG, SKU, ML, SARIMA, ARIMA, YOY

Table of Contents

List of Figures

List of Tables

Chapter 1

1.1 Introduction

Forecasting has been a lifelong challenge in the FMCG industry as inaccuracies cost! Logistical costs are associated with importing and transporting the product and if it ends up being stocked for too long then it would be wasted as it has expired especially when so much cost has gone into the product to be on the shelves. As of now, human assumptions are creating the forecast numbers in contrast to using algorithms to forecast sales and then basing assumptions on our forecast numbers. FMCG companies have been struggling with stock-out for years. The concept of stock-outs negatively impacts sales and leads to dissatisfied customers. SKU is the bottom level of the product level and we have a strategic dilemma about how to maintain the stock-out. Recent studies indicated that customers prefer to move to another shop and never return instead of waiting to buy the preferable product (Huang et al, 2013). To resolve this problem retailers, need to balance the lost revenue from stock-outs against the expense of safety stocks. One solution to this problem is optimizing the demand forecasting of products using Forecasting methods.

In addition, the task of creating such forecasting (manual) is time-consuming as we have over 3000 SKUs from 19 different end markets and each SKU follows its trend. As each SKU tends to have an offset from year to year however seasonality is likely to stay the same and there are cases where the is non-stationary when we are launching a new SKU which can be hard to forecast. The task requires a lot of communication, time, and effort which falls down to costing the business a lot of capital. In the industry of FMCGs each little, insignificant insight holds a significant impact and as such we have to automate a sophisticated ML algorithm to rely on facts and statistics rather than intuition in this stage. As insights and intuition should come at a later stage where it would add rather than subtract value. FMCGs have large consumption rates and hence for sustainability and cost reasons, it is much more efficient if we are to predict the exact demands needed. The best warehouse to have been none!

Demand planning is an important part of FMCGs for supply chain and management strategy purposes. As such, we are expected to achieve high accuracy for such predictions based on events and seasonality.

We usually have to call and ask the reasoning behind the given forecasts whether it makes sense or not and that is where the human errors come in. These processes are lengthy, inaccurate, and exhausting the human resource in the company which is why the automation of such a process is necessary. Another source of inaccuracy is the communication issue where demand is not properly communicated within the organization and the additional human resource spent to account for any mismatches or changes in assumptions from the end market. The project aims at exploring the various factors that could influence the data prediction to be as accurate as possible while having a one-stop solution that fits all the SKUs as this could cause losses for the organization if not addressed.

Human expertise is needed now more than ever because all these dull, tedious tasks can now be automated, and human resources will be relocated to a different function where they will be able to produce higher value for the organization. The human resource will be allocated to a more efficient role as such they are able to not only produce more profit and revenue but with lower costs as well because of the previous functions' automation which forsaken an additional human resource that was previously needed.

An unsupervised forecasting algorithm will be used due to us lacking the future attributes which would have been used in a supervised forecasting algorithm. Looking at the literature review the best accuracy is achieved by neural networks however we are unable to do so as the process needs to be supervised by an individual which is disastrous in our case. In addition to the infeasibility of the resources needed to do so for 3000 SKUs which will not be cost-effective. Hence our choice would be SARIMA since it accounts for seasonality, non-stationary data, and cyclical patterns while the fact that it wouldn't overfit the data can account for the inconsistency of the data. SARIMA overcomes ARIMA's non-seasonal forecasting on month-on-month intervals in our case and is univariate which is suitable for our need. As well as FBProphet which has gained a reputation in the market to perform well against time series forecasting.

1.2 Project goals

Automation of Demand forecasting as algorithms are unbiased, unlike human processes and predictions. We are to look for trends and analyze respectively. Factor in contributing attributes of the dataset that are not yet considered.

Importance:

- In the industry of FMCGs each little, insignificant insight holds a significant impact and as such we have to automate a sophisticated ML algorithm to rely on facts and math rather than intuition in this stage
- As Insights and intuition should come at a later stage where it would add rather than subtract value

Relevance:

- FMCGs have large consumption rates and hence for sustainability and cost reasons, it is much more efficient if we are to predict the exact demands needed
- The best warehouse to have been none!

Significance:

- Demand planning is an important part of FMCGs for supply chain and management strategy purposes
- As such, we are expected to achieve high accuracy for such predictions based on events and seasonality

1.3 Aims and Objectives

- Explanatory Data Analysis (EDA) – How to make the most out of our dataset to turn up insights and help management with decision making.
- As such we will find new information about our data to make the most out of it.
- We will use the EDA to discover the most impactful SKU in the dataset to make the best use of our resources and achieve maximum ROI.
- The main measure of accuracy used will be MAPE as it accounts for the volume we are dealing with as such we are aiming for a forecasting accuracy above 85% since this is the market standard for a good forecast.
- We aim to forecast monthly sales numbers for the next year.
- Hence, we will be using Facebook Prophet which has been known to achieve decent accuracy in the data science industry
- As well as SARIMA which is an algorithm well known in the industry as well for quite some time
- As such we will test both out and come up with a forecast from both and compare the accuracies
- The project aimed to make the employees' lives easier and improve morale
- The project will enable the organization to do more work with less time and effort
- The project also aims to increase productivity
- It also aimed to demonstrate using the proof of value for organizations
- It is a facilitator for organizations in digital transformation

1.4 Research Methodology

To be able to achieve favorable results, the CRISP-DM method is used by following the below steps

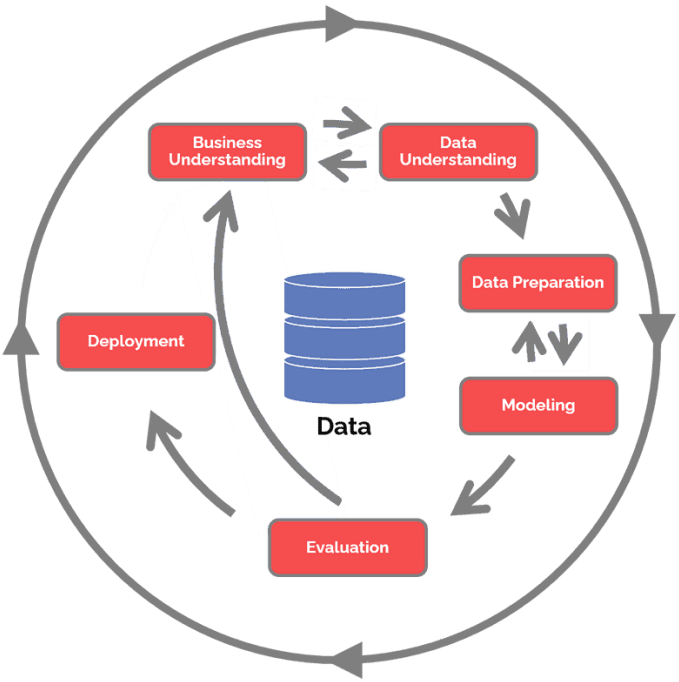


Figure 1: CRISP-DM Methodology

Stage 1: Business Understanding

Business Objective: To be able to forecast multiple SKUs using the same forecasting model while achieving a high degree of accuracy. We need to avoid overfitting while achieving a decent accuracy that the organization can use with trust. The forecasting algorithm will be unsupervised and will account for seasonality, non-stationary data, and cyclical patterns. As well as the inconsistency of the data and the various factors that could be leading to these inconsistencies.

Stage 2: Data Understanding

This phase will include an initial collection of data, encryption of the data, description of the data, exploration of the data, and making sense of the correlations made. The collection of data will include the data trends demonstrated by the historical data and develop a comprehensive understanding of the type of data whether it was such as categorical or sales numbers. The most relevant data for the given algorithms to be used are the SKU index numbers, the number of units sold, and the month and year the number of SKUs were sold in.

Index	SKUs are numbered instead of having the Brand Name on it with 80 SKUs in total
Tipping	Is the technology used in the filters' paper which is named by our R&D division they are to be kept in mind when manufacturing a new SKU
Blend	The blend used in the Tobacco sold
Pack type	How are the sticks packaged we have different packaging for different SKUs
GDB	Is the brand a Global Drive brand this is a Boolean column
Sticks	Number of sticks per pack, Numeric
Filter Design	Discarded since all of the data indicates non-tube SKUs in the dataset due to it being outdated, Boolean
Capsule	Specifies whether the given SKU has a menthol capsule in the filter or not, Boolean
SKU Family Group	Under which size group does the SKU fall under, Character
Length	These are usually indicated by some metric but in the Tobacco industry the case is there are certain standards to be followed and hence a specific name is given to each length, Char
Circumference	These are usually indicated by some metric but in the Tobacco industry the case is there are certain

	standards to be followed and hence a specific name is given to each Circumference, Char
Tar Delivery	Tar delivery in milligrams per stick, Integer
Delivery	Nicotine delivery as per filter used in the SKU, Char
Price class	There are 4 different price classes Premium, Aspirational Premium, Value for Money, and Low, character
Date	Date sorted in mm/dd/yyyy which are aggregated to represent a single month for each row, Date
Value	Value representing the number of sticks sold, Integer

Table 1: Data Description

Stage 3: Data Preparation

The data available after the collection phase will not be ready to be fed into a forecasting algorithm, power query will be used for cleaning and encryption the data. The following steps will be followed to get the needful data preparation:

1. Clean the data – Remove null values, detect outliers, treat them if necessary, and keep the valuable SKUs rather than using prediction for all SKUs needlessly to optimize our resources. The data had to be cleaned since it was pivoted as is the market standard on excel for data reporting. Unpivoting and creating a date column instead of having them in column headings.

2. Transform the data – Changing the data type, unpivoting the table as the columns are the months to give us a dates column, deleting null SKUs as well as SKUs with who is been discontinued or with insufficient data, and encrypting the data

3. Integrate data – joining different tables from different end markets and brands

Stage 4: Modelling

The ML algorithms will be used to predict sales volume for each of the top-performing SKUs for the next year. Since multiple SKUs are being forecasted, we do not run the risk of overfitting and we can aim for a higher degree of accuracy. R will be used for the modeling of the SARIMA model as it will be used since

the data is seasonal and non-stationary in nature. In addition to SARIMA's ability to account for a yearly offset which is accounted for. The yearly offset is a common trend in FMCGs due to the nature of brands and marketing strategies set out for the year. Power BI will be used for the visualization to give the user easily readable results and it is the resource recommended by the organization.

Stage 5: Evaluation

In the final step after training the model with the train set of data, the model needs to be

evaluated with the test set to calculate the accuracy of the model. For a prediction model, the test data sets will help to compare the actual results with the predicted results.

1.5 Limitations of the Study

Due to the nature of time series Forecasting, only 1 SKU can be forecasted at a given time while having over 3000+ SKUs spread out over 19 countries Hence forecasting for each single SKU wouldn't be feasible. As such for our business case we are going to pick the SKU that represents our top contender to sales as such has the highest impact on total revenue and cost. The collection of the data is aggregated by the business for monthly sales instead of daily sales for reporting purposes and as such we are constricted in analysis, weekly and daily seasonalities would not be accounted for. Sales data is represented in units sold rather than the amount sold for dollars which can be misleading if we are aiming for maximum value extraction and strategic purposes.

Chapter 2 – Literature Review

2.1 Introduction

Relevant works on Forecasting applied to demand forecasting for fast-moving consumer items were discovered during the literature review for this study. The advantages of a successful model and the business sectors covered are emphasized over traditional statistical approaches. The importance of predictions is a recent study that identifies the first-rate forecasts as a critical achievement component that has a significant impact on publications and firm governance of manufacturing, supply chains, logistics, stores, and so on. According to the literature, businesses employ statistical techniques and tools to estimate product demand, with past data from previous revenue serving as their primary source of data. However, according to Kandananond (2012), using system learning to anticipate client goods is of higher quality than using traditional techniques, notable processes like the autoregressive incorporated shifting average (ARIMA) version of client goods. System engineering is used to conduct a study on a specific application. Analyze the process of learning an algorithm. Simulation of a person learning about a procedure.

Other studies have attempted to anticipate product sales by including promotional data directly. Forecasting performance was addressed in the paper as part of their examination of promotional profitability (Rinne,1988). Their model was devoid of dynamic and competitive impacts, as well as any evidence of forecasting accuracy. looked at a small number of product categories and created static models without considering competitive impacts (Preston, 1990). There was no model comparison exercise. To anticipate daily milk sales, Kuo (2001) suggested a fuzzy neural network model. The neural network model is combined with an algorithm that learns imprecise rules regarding promotional effects from marketing specialists.

Wu and Zheng (2015) provide a forecasting model based on system learning that is more accurate than out-of-date statistical models for products with extremely risky demands and

extremely short life cycles. Fast-fashion merchants Especially when statistics other than income records are added.

In the F&B industry, Fujimaki (2016) developed a model for a beverage retailing company that combined forecasting and price optimization strategies, resulting in enhanced forecasts, increased decision-making reliability, and a 16% increase in sales. According to Tsoumakas (2018), Forecasting strategies are more powerful and adaptable than traditional statistical strategies for predicting time collection because they have more processing power and the ability to address more variables. The authors propose several Forecasting methodologies for achieving better revenue estimates.

Islek and Oguducu (2015) took a different approach to develop a demand forecasting technique based entirely off of the Bayesian Networks, a method that verifies conditional possibilities in accordance with the statistical techniques used, yielding in improved accuracy for revenue forecasts for a range of retail merchandise.

A variety of factors, including the product's own features and the economic environment, influence a product's income, and as a result, there was no pattern in the historical data used for estimates. They offered an income prediction version based on Forecasting techniques with multiple variables, including statistics from early orders, historical income, promotions, and product traits, which resulted in greater precision in income forecasting and more frequent replenishment at the retail stores, and therefore improving supply chain effectiveness to drive higher profits (Guo, 2013). Lu (2014) agrees that new variables should be included in forecast models, adding that various factors, including inventory market indexes, influence the accuracy of income forecasts, and that, as a result, variable selection is critical. employed a Support Vector Machine (SVM) to increase the capability of income management for electronic products that are easily replaceable and prone to large fluctuations in demand (Lu, 2014). Qu et al. (2017) used several statistics sources in their Forecasting algorithms, including each internal statistic as well as outside variables such as holidays, with the main results being a higher income forecast and an adjustment of the inventory levels of semi-luxurious goods' stores, which have seasonal traits and large versions in buy stimulus. Lu and Shao (2012) used Extreme Learning Machine

methods to obtain more precise revenue estimates for short-lifecycle merchandise, allowing pc merchandise and accessories retail chains to help them reformulate their strategy to make for a successful business model. Lee, Kim, Park, and Kang (2014) used a joint method of Forecasting techniques in a study on electronics, whose era had no revenue history, to advance a forecast income model for modern items, which resulted in more accuracy than other ways available. Lu and Chang (2014) have developed a cross-income prediction version for data-era commodities that incorporates SVM and is more accurate and reliable than earlier proposed solutions. Chen and Lu (2017) demonstrate the advantages of Forecasting for contemporary businesses with short product life cycles by boosting a computer retailer's revenue forecasting accuracy and stock management. The apparel sector, according to studies, is looking for improved demand forecasting. Choi (2014) developed a practical ELM set of rules for fashion retail income management, allowing for improved precision and efficiency in the estimation process through Forecasting algorithms on old and catalog income data, Tehrani and Ahrens (2016) reduced the income projection residual rate for fashion with a high output manufacturing scale, minimizing losses and increasing profitability. The research affirmed sales growth for a web-based fashion clothes and accessory store, where forecasting algorithms analyzed lost revenue history and forecasted future demand for brand new products launch (Ferreira et al, 2016). Another research used Extreme Learning Machine (ELM) techniques to apply to historical sales data with the volumes and characteristics of the products sold to meet an online fashion product store's commercial enterprise goal of having snappy and timely revenue forecasts for a large volume of SKUs (Yu et al, 2011). Liu and colleagues (2013) demonstrate the evolution of fashion retail income estimates, showcasing ELM methodologies and benefits like efficient stock management.

Tsoumakas (2018) also highlights the advantages of Forecasting technologies for predicting food business earnings. Accurate short-term forecasting allows for more efficient inventory level reduction and hence reduced warehousing costs, the elimination of expired merchandise in stores, and avoiding profit loss due to out-of-stock status. The key advantages noticed were a reduction in human bias because of the increased use of forecasting algorithms and data science in the forecasting, higher accuracies were achieved as a result, and the ability to discover independent variables affecting the demand. The article identified two potential drawbacks for agencies: the likely lack of precise old information and the wide number of learning algorithms

accessible, which could make selecting the best one challenging. The main limitation is the number of unprecedented attributes that happen within the consumers which are impossible to quantify and predict demand for.

2.2 Takeaways from the Literature Review

I have come to discover how multiple industries tackled the challenge of forecasting sales and stocks needed. The F&B, Fashion, and Electronic industries have had several useful methods for so and the accuracy of forecasting is as critical to these businesses as it is to FMCGs in the sense that they expire or go out of date quickly and logistics are time-consuming in nature hence time is of the essence!

These papers have demonstrated the value of Forecasting for businesses and how is it being done. Algorithms such as SVM, ARIMA, Logistic regression, deep learning, and Bayesian networks were used where they fall under time series as well as the classification for forecasting. FBProphet seems to not exist in the world of literature as of yet and I will be comparing it with the market standard for seasonal data forecasting (SARIMA). Forecasting has integrated properly into businesses and enabled them to make business decisions such as the pricing of an SKU and how that would affect their demand. Digital transformation has driven the business to improve efficiency and achieve an overall higher ROI.

Chapter 3- Project Description

3.1 Project Description

The Forecasting algorithms are accounting for numerous aspects to consider such as Promotions, seasonality, and competition's release of SKUs are examples of data that may have an impact on sales. However, they are not directly accounted for in a column, but they are represented in the value of sales numbers. In the time-series dataset, a demand forecasting model will be developed to estimate sales in each store where data is divided by 4 years to train the algorithm and 2 years to test the algorithm and the method used for accuracy is MAPE. As forecasting is computing and labor-intensive, the top-performing SKU will be forecasted for maximum value. The dataset will be preprocessed, which will include eliminating null values and errors. Additionally, data preparation was carried out in order to clean the data and prepare it for analysis.

Chapter 4- Data Analysis

4.1 Explanatory Data Analysis

In this chapter we will come to explain how we have studied our data to capitalize on our data and provide actionable insights.

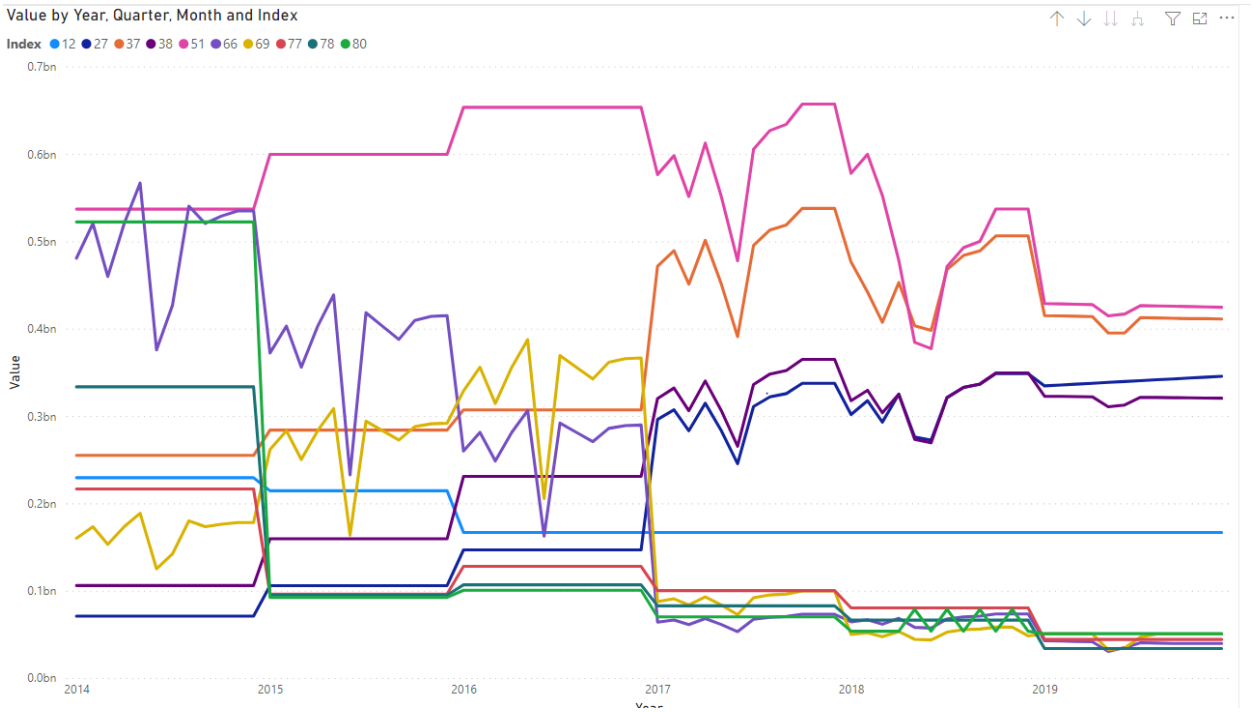


Figure 2: Sales YOY

As we can see seasonalities differ from one SKU to another as they operate in different End Markets or countries. Seasonalities tend to depend on taxes/tariffs imposed by the government as these tend to change from one end market to another. As well as cultural ones such as Ramadan and Eid where these would cause variables and are not dependent on a specific month such as April as the dates shift year on year. Here we have shown a visual of the most impactful SKUs in our End Market Year on Year, SKU 51 seems to be the best performing SKU and as such seems to be our top contender to forecast.

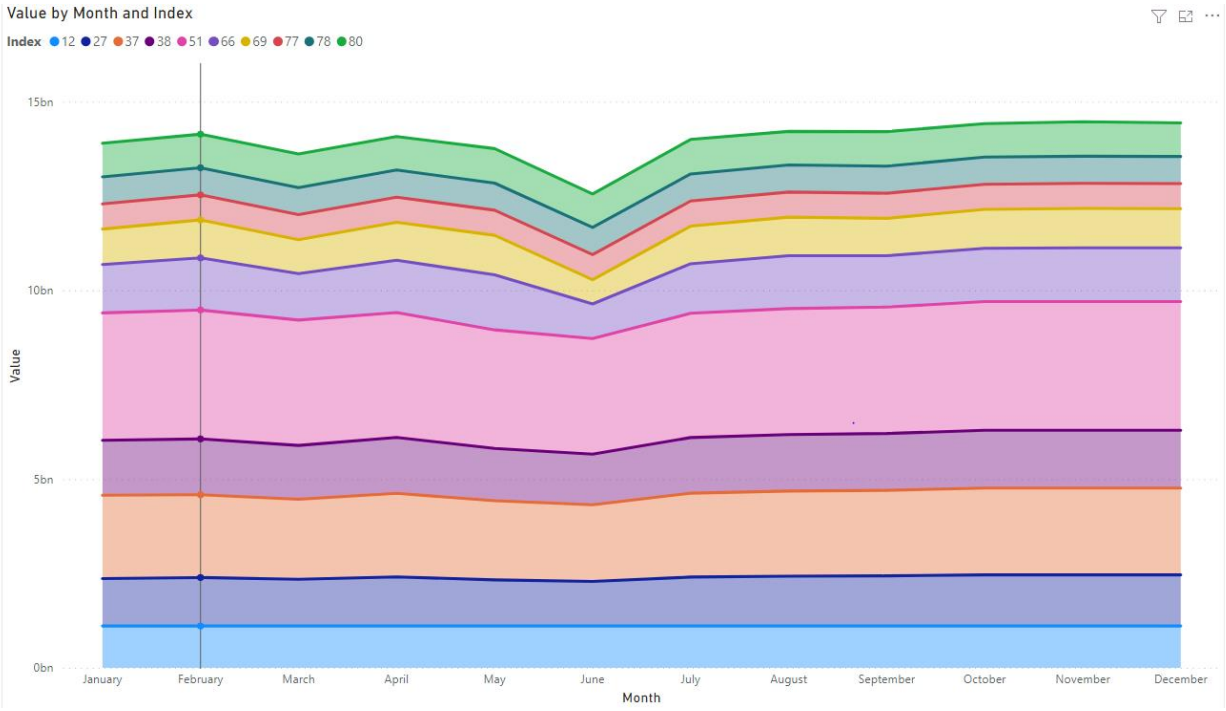


Figure 3: Monthly sales seasonality (aggregated)

Here again, we have shown the top 10 performers in the end market however they are shown in terms of months so we can have an idea of the seasonality and see if exists or not, to begin with. We can draw the conclusion from the visual that there's an obvious drop in sales in June for each single SKU and December seems to be the month that the SKUs sell the best.

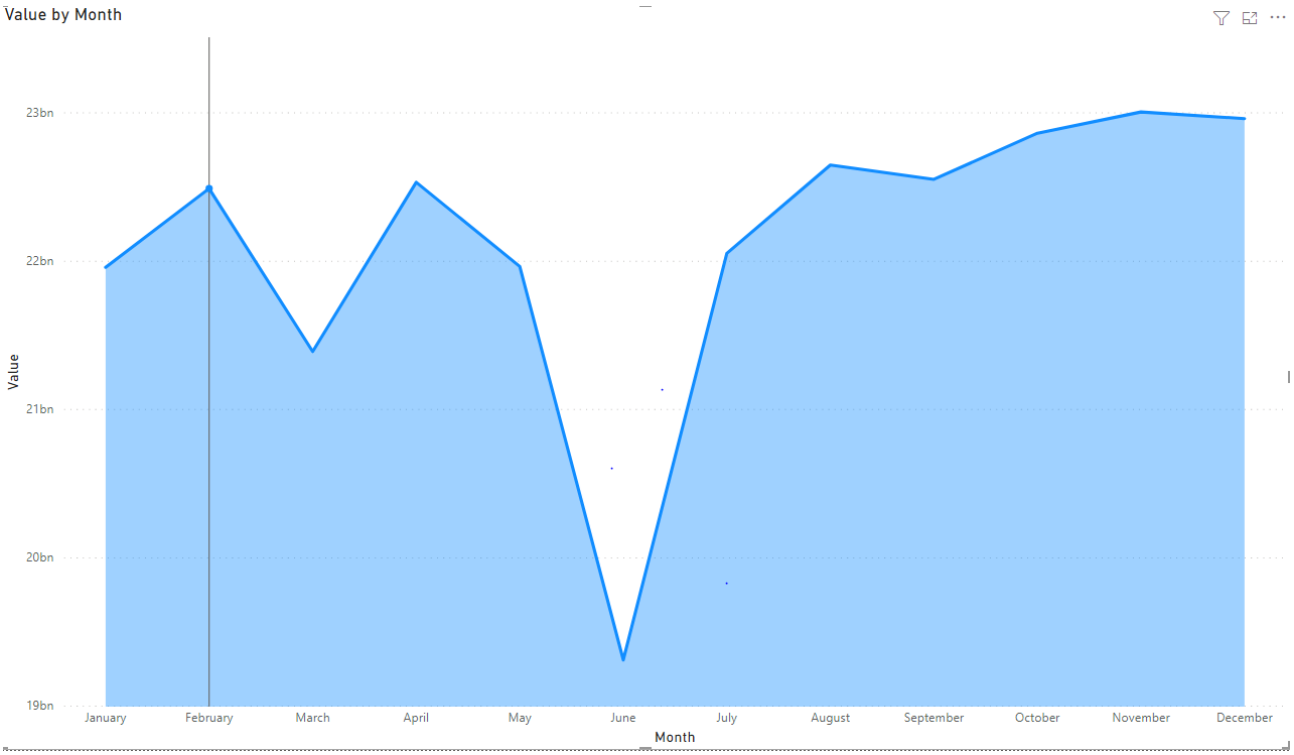


Figure 4: Total Sales of all SKUs

Now using the visual with the total sales for all SKUs we can see another seasonality much more clearly in march where we witness a drop in sales before returning to normal and later on dropping significantly in June. Now looking at this visual alongside the previous one we can focus the management sights on the top 10 SKUs since they make up more than 60% of the sales. As such more human resources, analysis, and Marketing investments will be focused on these SKUs.

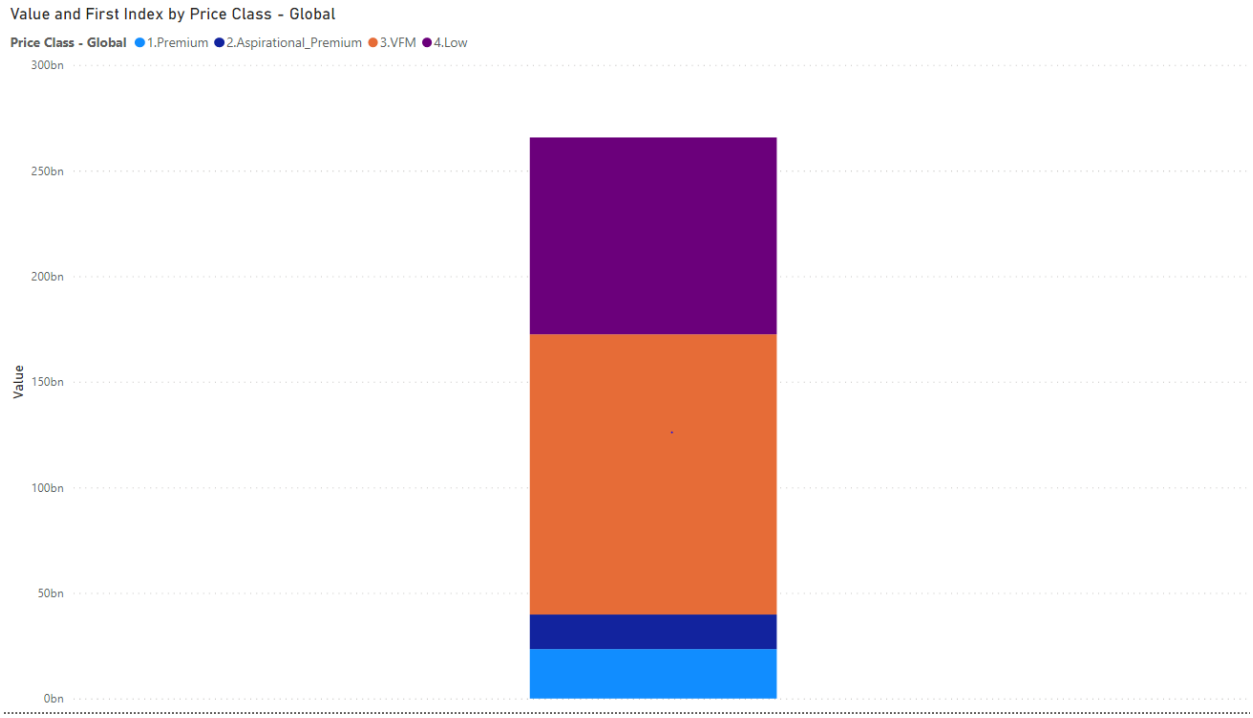


Figure 5: Sales volumes per price class

Low-priced brands seem to be the best performing in sales with more than 50% of total sales volumes. Low was previously filled in as 4. Low as well as Low by the database administrators which created confusion in the visual since Power BI classified them as different categories in the visualization phase which had to be amended. This visualization also shows the total sales from 2014 to 2020 for all SKUs and how are our brands perceived in the eyes of the consumer and how can we capitalize on this data.

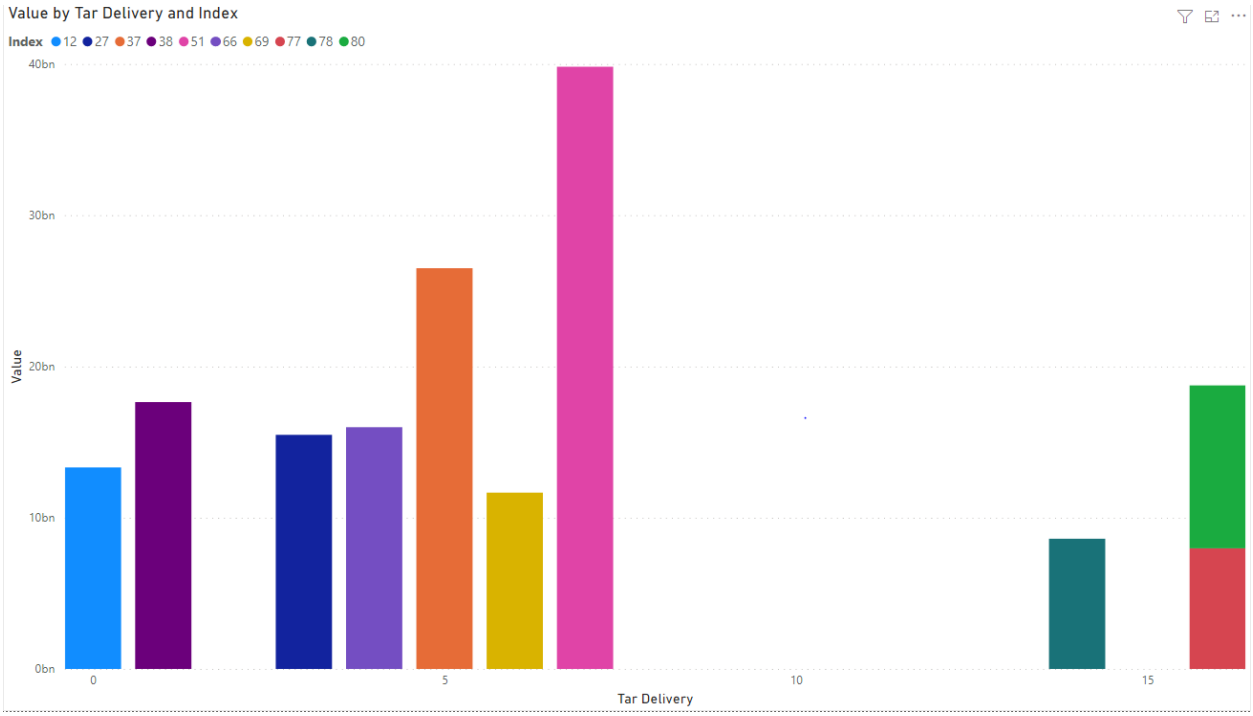


Figure 6: Sales by tar delivery (aggregated)

This is a visual showing the top 10 performing tar deliveries in terms of sales volumes. As it can be clearly shown that the majority of the sales lie in the extremes of either high levels of tar delivery or extremely low ones in comparison. As is the case that tar deliveries from 8 mg to 14 mg yield minimal sales in comparison. However, it can be seen that clearly the top-performing Tar delivery is 7 mg by miles at almost double in sales compared to the third-best performing SKUs tar delivery and hence management knows where to shift their focus and strategy towards to yield the most value out of SKUs in terms of tar delivery and possibly capitalize on top-performing tar delivery with investing in them to yield greater results with the least possible investment.

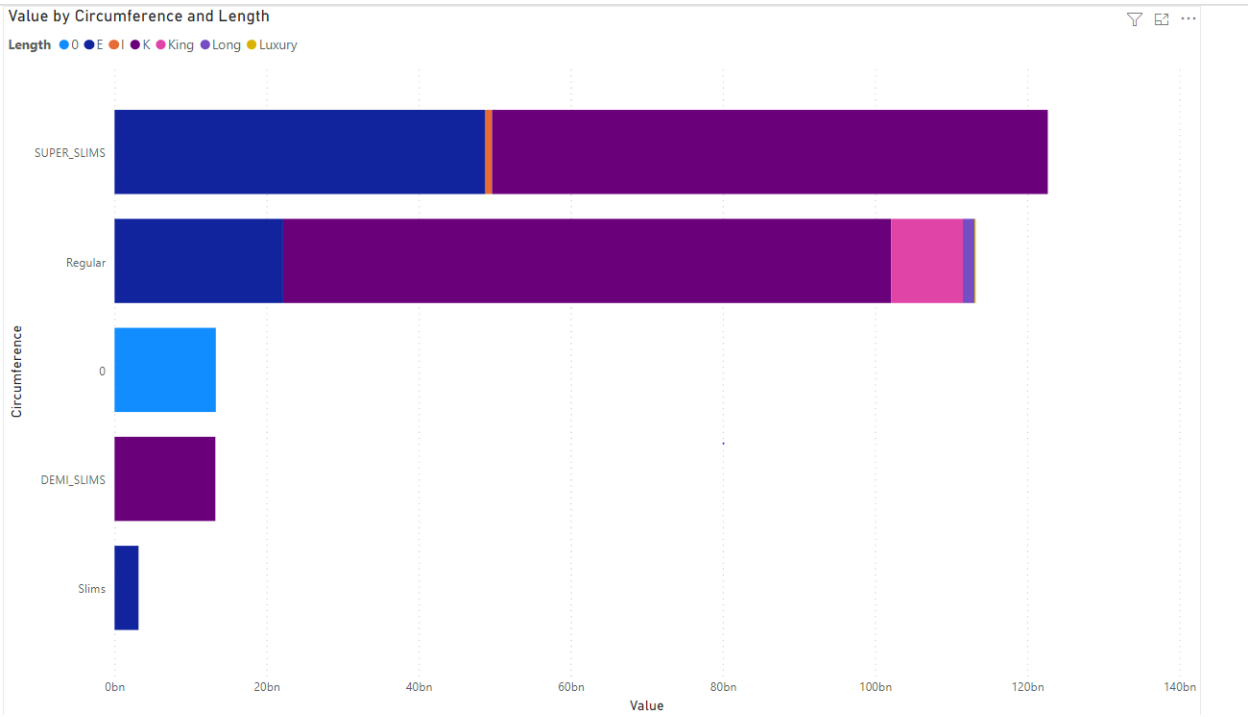


Figure 7: Analyzing sales by Length and Circumference

Regular Long SKUs seem to be the best performing and the second-best performing would be Super Slims with Long. Hence from the visual, we can summarize that Long sticks sell better than any other attribute in the visual and is the key focus of the business.

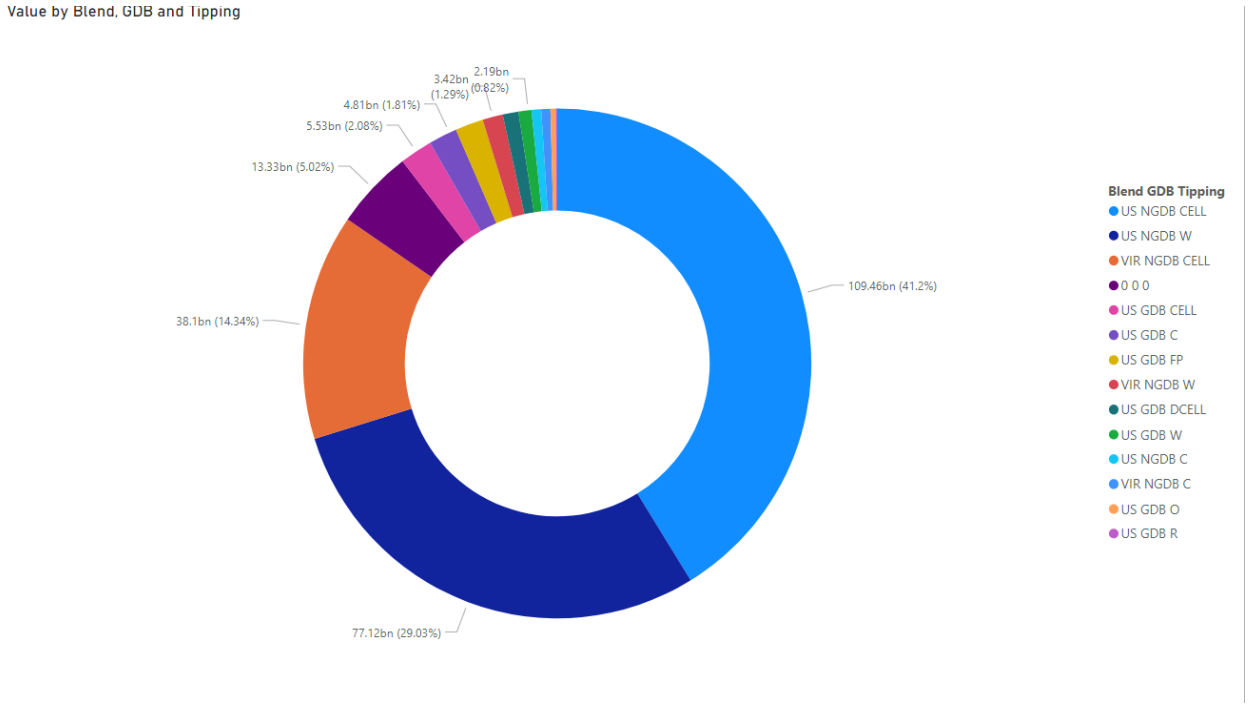


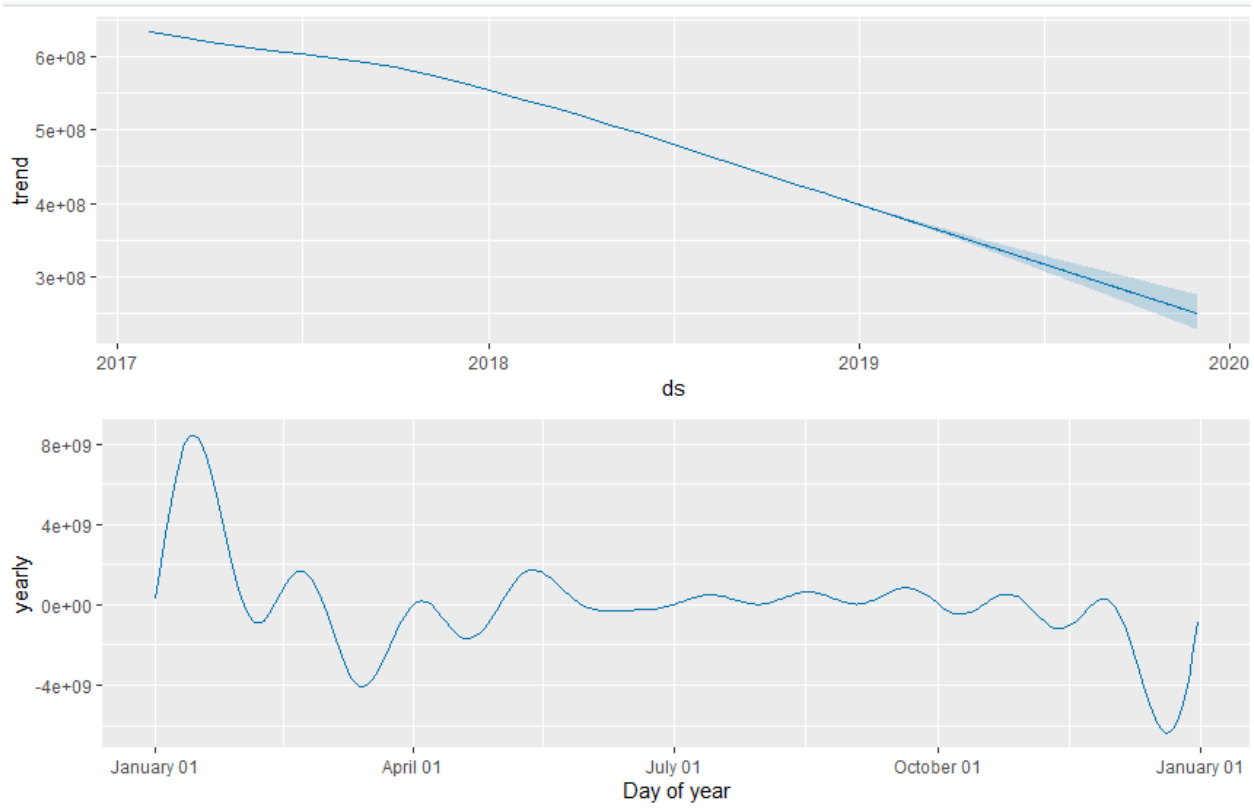
Figure 8: Analyzing Sales by stick's tipping

Surprisingly the top 3 best performing SKUs are not a Global Drive Brand and hence the conclusion could be made that not every NGDB is not a necessary competent SKU. As such Management needs to prioritize this SKU and not just GDB. It also happens to be a US blend tobacco and tipping of CELL to be the best performing combination followed by the US tobacco again however with tipping of W. As such the best selling tobacco is very well the US blend and this is beneficial in the release of a new SKU and how to make for a successful product in a sense our data is acting a recipe for future SKUs moreover success!

4.2 FBProphet

Since SKU 51 is the one that yielded the best values and it would make sense to analyze the top-performing SKU that would yield the most value and allocate forecasting resources where they matter the most. As such the visualization tells us that SKU 51 has absolutely no outliers making it a reasonable time series data to forecast. The visualization also tells us the sales for SKU 51 lie at 0.55 billion sticks monthly on average.

Figure 9: SKU Boxplot




```
> MAPE(foreacc1$trend,dataacc$value)
[1] 0.2374723
```

Figure 10: FBP Forecast & seasonality

FBP which stands for Facebook prophet is known for handling seasonalities well as well as non-stationary data. Since the FB Prophet is a library on R we were able to treat it as a black box that reallocated resources towards the analysis rather than the forecasting alone. With that being said FB Prophet was able to account for weekly and daily seasonalities however the data was aggregated to show monthly seasonality. We were able to obtain a decent MAPE forecast accuracy of almost 77% which is good nonetheless however in our application it is always better to have higher accuracy. As we can see since we had the monthly seasonality the fbprophet tells us the months' seasonalities as well as the forecast output as well as the Year on Year trend.

4.2 SARIMA

SARIMA is a time series machine learning algorithm. Effectively used for predicting time series data.

We create an SARIMA model based on the data provided by AR, MA algorithm.

AR(Auto-Regressive):

The AR algorithm determines the linear regression of (Present fitted values) vs. (Past fitted values) for prediction purposes. The AR is represented by the parameter 'p'.

If the order is p =3, it means we are using first three lags (past values) to make prediction.

MA (Moving Average):

The linear regression of the (Present value of residuals) vs. MA is discovered by MA (Past value of residuals) for prediction purposes. The MA is represented by parameter 'q'.

If the order is q = 2, it means that we are using two past residuals values to make prediction.

I (Integrated):

The meaning of integrated is just differencing.

If the data is not stationary and contains trend, we difference our current values with previous values to remove trend and make data stationary.

If the order of I (represented as 'd parameter') is 2, it means the differencing is made two times.

Accuracy metrics:

Here we are using MAPE (Mean Absolute Percentage Error) which is a statistical measure to define the accuracy of a machine learning algorithm on a particular dataset. MAPE can be considered as a loss function to define the error termed by the model evaluation.

AIC :

The AIC is designed to find the model that explains the most variation in the data, while penalizing for models that use an excessive number of parameters. Once you've fit several regression models, you can compare the AIC value of each model. The lower the AIC, the better the model fit.

Step used for the implementation of the SARIMA Model are as follows:

1. Importing the required Libraries
2. Loading and Reading the Dataset
3. Initial EDA Analysis
4. Plot the series and search for possible outliers.
5. Stabilize the variance by transforming the data (Box-Cox) (Not needed in our case)
6. Analyse the stationarity of the transformed series.
7. Identify the seasonal model by analyzing and exploring the seasonal coefficients residuals of the ACF and PACF
8. Visualization of Best fitted Model on original data
9. Now, splitting the data into training and testing sets
10. Fitting Tuned model using auto Arima function on the training set
11. Visualization of Forecasting on test data using tuned model
12. Analysis of Residuals of Tuned model using q-q line, histogram, etc.

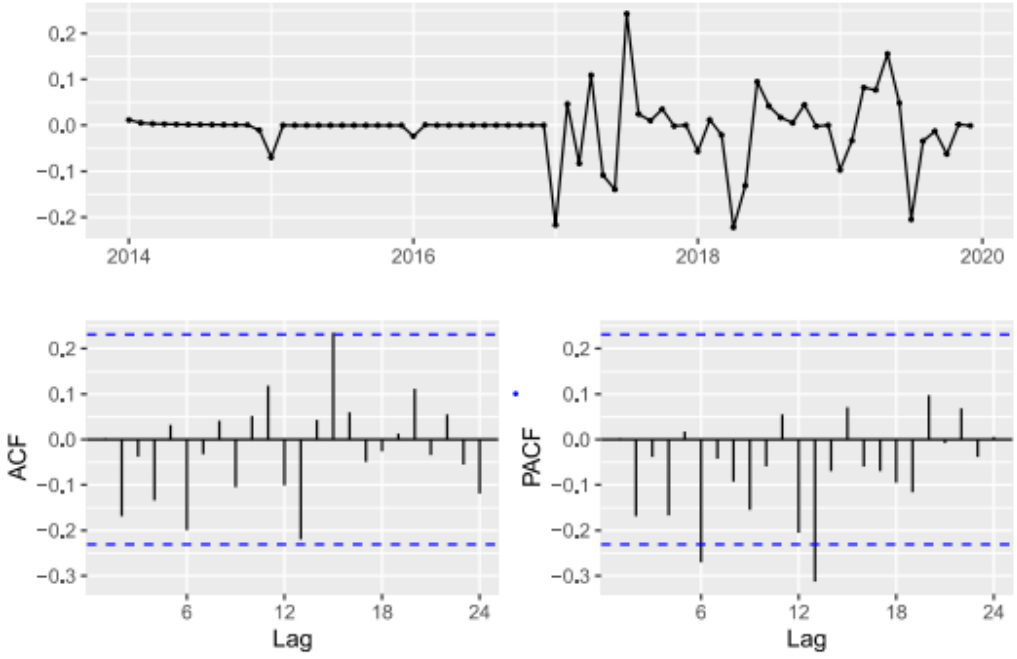


Figure 11: ACF PACF of the tuned SARIMA Model

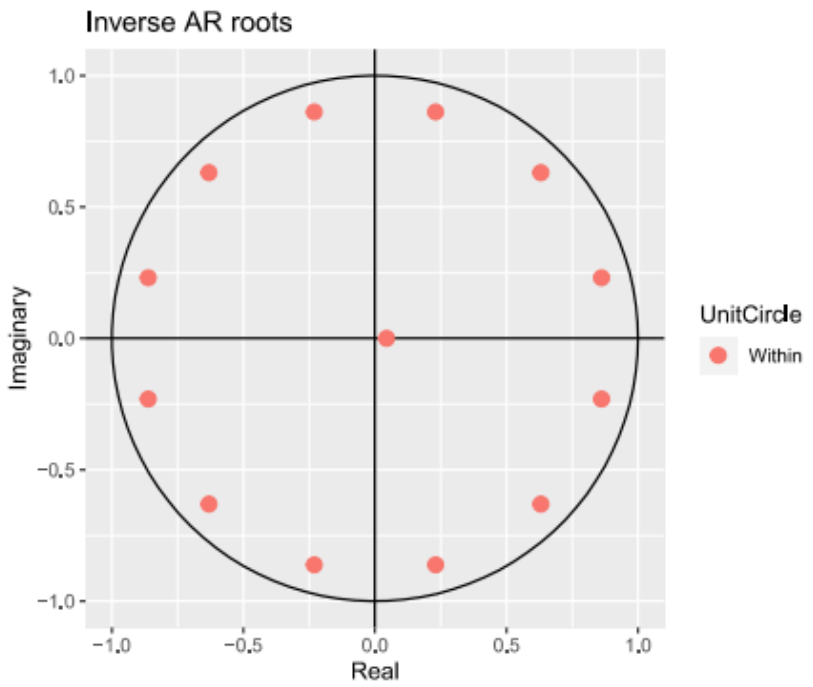


Figure 12: Inverse AR roots

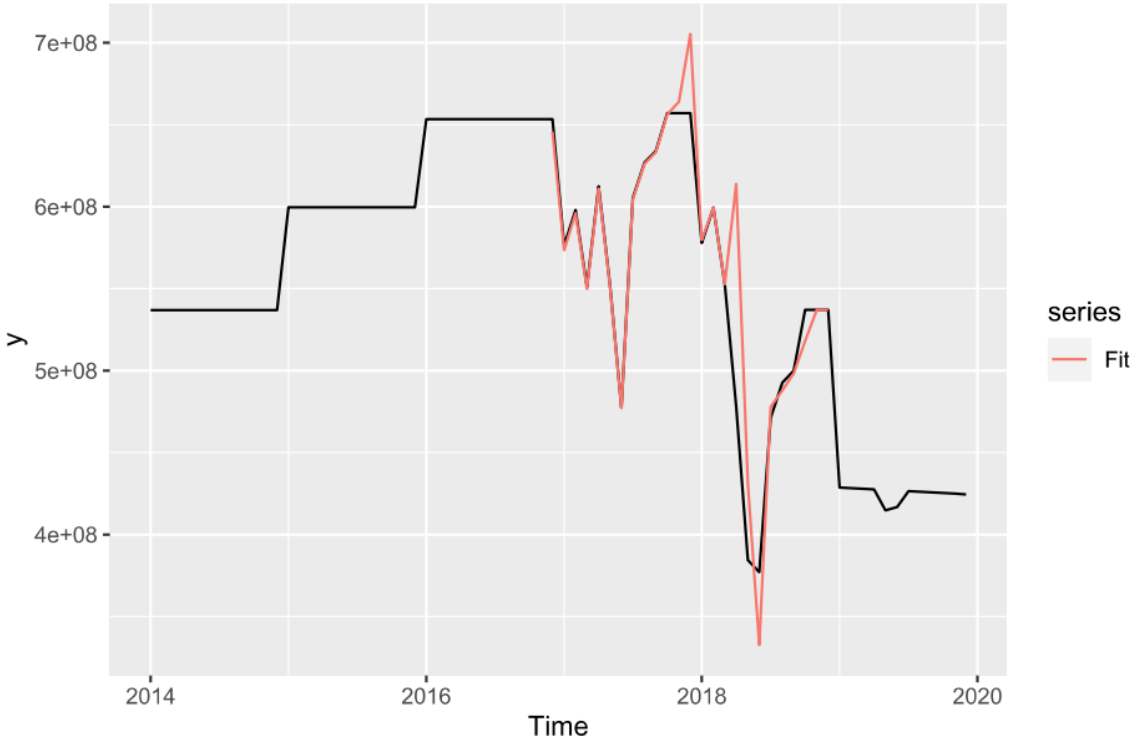
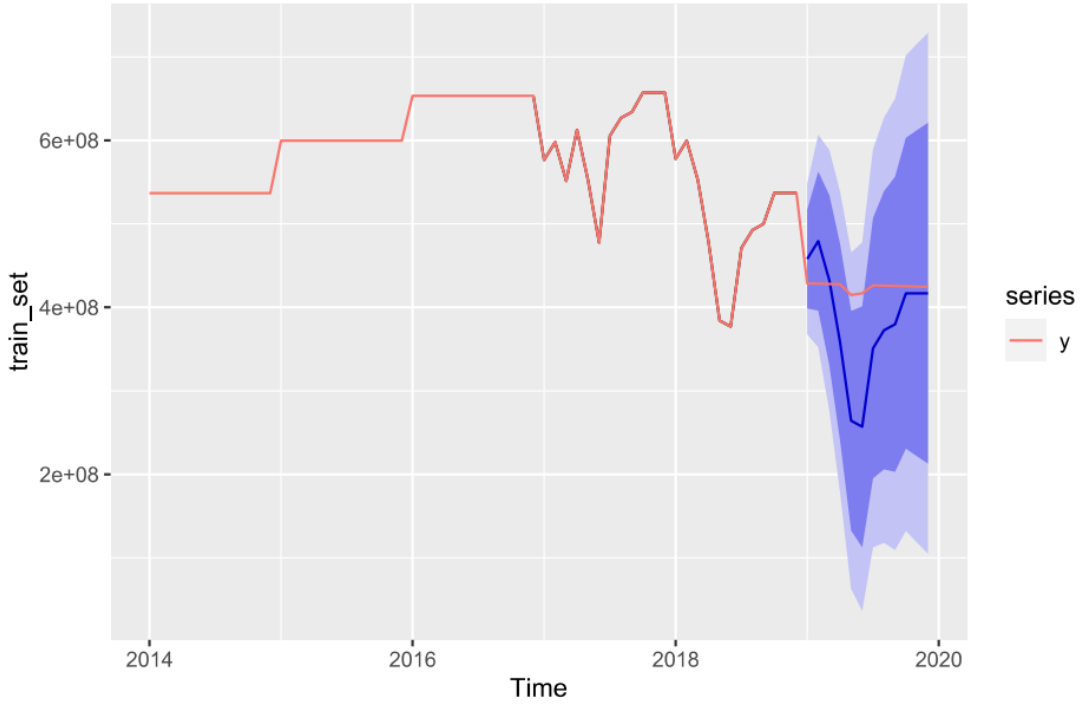


Figure 13: SARIMA Model fit

Figure 14 presents the fit of the SARIMA Model used, which is a perfect balance between bias and variance, where we have avoided overfitting and the model is complex enough for the seasonality of data. Therefore, this model fits the nature of our data nature. Our data lack stationarity because the mean is actually dependent on time. Due to data quality issues the model was trained from 2017 to 2019, where the year 2019 was forecasted. Our data definitely follows seasonality as it was previously observed in the EDA. As well as the overall year on year trend observed as the SKU's overall demand is decreasing due to the population's awareness of the health risks associated with tobacco as such management is looking for a sustainable alternative of revenue.




```
## [1] 0.1310738
```

Figure 14: SARIMA Model forecast

I was able to obtain a decent accuracy of forecast despite the data irregularity due to logistical difficulties and the region that the SKU lay in. The accuracy obtained is an excellent one at 87% MAPE accuracy which is one of the best metrics in time series forecasting since it accounts for a percentage change in absolute value between the forecasted and actual values rather than just values which is not meaningful in

contrast with MAE, especially in our use case where we have a sheer amount of volumes and percentage accuracy is much more impactful. As such we have eased the human element however it is not eliminated as the forecasted values can lie anywhere between + or – 13% which would need the human resource/element to decide on.

Chapter 5

5.1 Conclusion

The study combined different statistical, numerical, and predictive techniques to drive value for organizations whether it is for shareholder's meetings by showing them the growth of our brand in trend instead of trying to communicate the information with technicalities or whether it is for supply chain forecasting as an inaccurate forecast can be costly in terms of product and logistical costs. As such the research aimed for a higher degree of accuracy.

We can conclude that the SARIMA had the most accurate forecast when compared to FBP in the case of SKU 51 with 87% accuracy in contrast with 77% accuracy with MAPE as a metric for both. However, each SKU has it is own target audience and it is own data time-series trends and specifications as such we are not disregarding an algorithm in favor of the other but rather making studies SKU by SKU basis to decide the best algorithm to forecast our monthly sales for the next 12 months.

5.2 Recommendations

Using the same methodology for different SKUs and testing out both algorithms as mentioned previously because of how differently our data react from one SKU to another as the SKU could very well be in a different region and as such we will be testing on SKU by SKU cases using R to run both codes together simultaneously and picking the lower MAPE error given by each algorithm and as such we would have automated the process of forecasting and reallocating human resources in analysis where they drive the most value.

References

1. Chen I.-F., Lu C.-J. (2016). Sales forecasting by combining clustering and machine-learning techniques for computer retailing. The Natural Computing Applications Forum.

2. Choi, Tm; Hui, Cl., Liu, N., Ng, Sf., Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. Decision Support Systems.

3. Elcio Tarallo, Getúlio K. Akabane, Camilo I. Shimabukuro, Jose Mello, Douglas Amancio, Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research.

4. Ferreira K.J., Lee B.H.A., Simchi-Levi D. (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Institute for Operations Research and the Management Sciences, v. 18, n. 1, 69-88.*

5. Fujimaki R., Muraoka Y., Ito S., Yabe A. (2016). From prediction to decision making - Predictive optimization technology. NEC Technical Journal.

6. Huang, T., Fildes, R. and Soopramanien, D. The value of competitive information forecasting FMCG retail product sales and the variable selection problem (LUMS Working Paper 2013:1). Lancaster University: The Department of Management Science.

7. İşlek, İ., & Öğüdücü, Ş. G. (2015, June). A retail demand forecasting model based on data mining techniques. In *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)* (pp. 55-60). IEEE.

8. Kandananond, K. (2012). A comparison of various forecasting methods for autocorrelated time series. *International Journal of Engineering Business Management*, *4*, 4.

9. R.J. Kuo. (2001). A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. *European Journal of Operational Research, 129.*

10. Lee H., Kim S.G., Park H.-W., Kang P. (2014). Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forecasting & Social Change, Vol.86, 49-64.*

11. Liu, N; Ren, Sy; Choi, Tm; Hui, Cl; Ng, Sf. (2013). Sales Forecasting for Fashion Retailing Service Industry: A Review. *Mathematical Problems in Engineering.*

12. Lu, C.-J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing, n. 128, 491-499.*

13. Lu, CJ., Chang, CC. (2014). A Hybrid Sales Forecasting Scheme by Combining Independent Component Analysis with K-Means Clustering and Support Vector Regression. *Scientific World Journal.*

14. Lu C.J., Shao Y.E. (2012). Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine. Mathematical Problems in Engineering.

15. J. Preston, A. Mercer. (1990).The evaluation and analysis of retail sales promotions.
16. European Journal of Operational Research, 47, pp. 330-338

17. Qu T., Zhang J.H., Chan F.T.S., Srivastava R.S., Tiwari M.K., Park W.-Y. (2017). Demand prediction and price optimization for semi-luxury supermarket segment. Computers & Industrial Engineering.

18. H. Rinne, M. Geurts, (1988). Forecasting model to evaluate the profitability of price promotions. *European Journal of Operational Research, 33.*

19. Tehrani, Af., Ahrens, D. (2016). Enhanced predictive models for purchasing in the fashion field by using kernel machine regression equipped with ordinal logistic regression. *Journal of Retailing and Consumer Services.*

20. Tsoumakas, G. (2018). A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, *52*(1), 441-447.

21. Visin, F., Kastner, K., Cho, K., Matteucci, M., Courville, A., & Bengio, Y. (2015). Renet: A recurrent neural network based alternative to convolutional networks. *arXiv preprint arXiv:1505.00393.*

22. Yu Y., Choi T.-M., Hui C.-L. (2011). An intelligent fast sales forecasting model for fashion products. Expert Systems with Applications.

23. Zheng, M., Shu, Y., & Wu, K. (2015). On optimal emergency orders with updated demand forecast and limited supply. *International Journal of Production Research*, *53*(12), 3692-3719.

24. Z.X. Guo, W.K. Wong, Min Li. (2013). A multivariate intelligent decision-making model for retail sales forecasting,Decision Support Systems,Volume 55, Issue 1,2013,Pages 247-255, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2013.01.026.