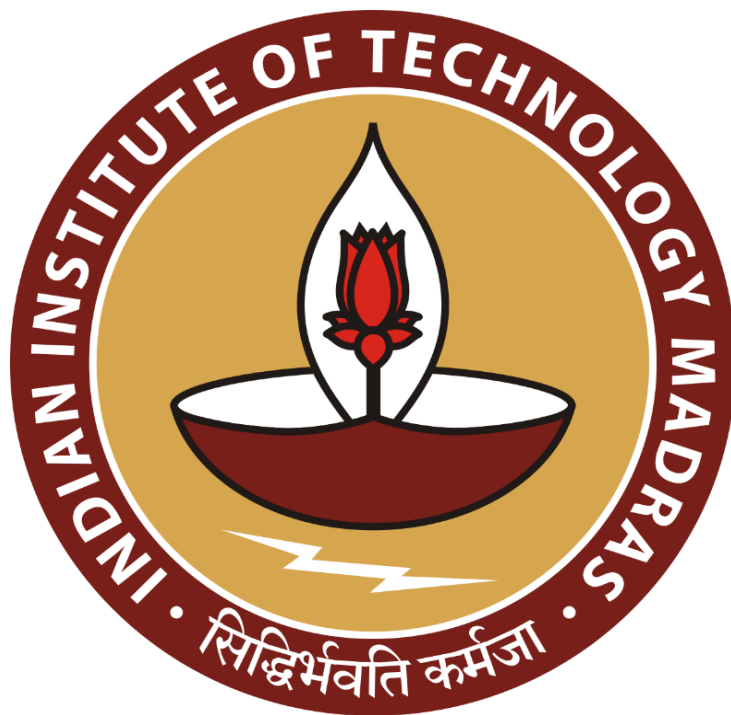# Data-driven Sales and Demand Prediction for Electrical Appliance Supplier

Final Report

Submitted by

Trivedhan Sivaprakash

21F1001724

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

Declaration Statement

I would first and foremost like to extend my appreciation to Ohm Muruga Services for being very forthcoming in providing the necessary data that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the information of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfilment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.

Signature of the Candidate

X
_____
Trivedhan Sivaprakash

Name: Trivedhan Sivaprakash

Date: 29th October, 2023

Executive Summary

This final report presents the outcomes of the project aimed at developing a sales and demand forecasting model for Ohm Muruga Electrical Appliance Supply Company. Facing revenue declines during the COVID-19 pandemic, this project was initiated to check for possible patterns and perform a Sales Analysis and identify trends that can be implemented and improved.

The Primary Sales and Purchase Order Data between the months January 2023 to June 2023 was collected and cleaned for a performing the analysis. The predictive Forecasting model for Sales and Purchase Order Data, which was constructed using Excel, helped in prediction of the revenue trend. The forecasting model employed linear regression model and improved upon the result by using Exponential Smoothing Techniques.

The Sales and Purchase Datasets were analyzed comprehensively for any notable trends or important points. After this, the Sales and Demand Forecasting was conducted and was used to predict the data for the next three months. Based on this forecast certain improvement points and recommendations were made to improve Profitability and ensure a strong supply chain. Pareto analysis and other Graphing tools helped in identifying key features which were then used for better understanding of insights.

In conclusion, this project has empowered Ohm Muruga Electrical Appliance Supply Company to adapt to market fluctuations, improve profitability, and make informed decisions.

Methodology and Explanation used for the Analysis

Analysis and Examination of the Datasets

1. Dataset Collection from the Primary Source

The primary Sales and Purchase Order Data was obtained from Ohm Muruga Services with the express permission of the Chairman. The sales data comprises of 516 records of data spanning between 1st January, 2023 to 30th June 2023. The purchase order dataset comprised of dataset 112 records of data spanning between the aforementioned timeline. This data needed to be cleaned before analysis.

2. Pre-processing of the Datasets

Initially, dataset was obtained in the format of pdfs with individual pdfs for each month worth of data. With the aid of online tools, the pdfs were parsed into individual excel worksheets. These worksheets were then concatenated into two single pdf comprising the sales and purchase order data for all the six months for ease of work. Irrelevant text information present around the dataset was removed until only the table can be extracted and converted into excel format.

Once this was done, the dataset had the following columns listed in the table below.

Columns in Both Datasets
Date (dd-mm-yyyy)
Particulars
Voucher Type
Voucher No
Sales/Purchase Amount

Table 1: Columns in both Datasets

Based on the type of data present in the columns and their dependencies with the other columns, the irrelevant columns were discarded and the relevant columns were kept for understanding. The columns Voucher Type and Voucher No. were removed from both the datasets as they do not contribute to the data and were found to not be correlated with the other columns.

After this, additional columns were added in the dataset to further segregate the data based on the day of the week and the month for a more comprehensive analysis.

3. Exploratory Data Analysis (EDA)

The dataset derived after pre-processing was then analyzed for any consistent patterns, trends and relations based on which insights could be made. This also helped in deriving any potential hypothesizes for analysis and further investigation. This also helped in obtaining a comprehensive understanding of the provided data.

Sales Forecasting of the Dataset

Sales forecasting is the process of estimating future sales performance based on historical data, market analysis, and relevant factors. This acts as a strategic tool for businesses to anticipate demand, allocate resources efficiently, and make informed decisions. Utilizing statistical methods, machine learning algorithms, or simpler quantitative techniques, sales forecasting aims to project future sales figures. Accurate sales forecasting empowers businesses to optimize inventory levels, allocate marketing budgets effectively, and enhance overall operational efficiency, contributing to sustained growth and competitiveness.

For Forecasting, the following conditions must be met before performing the prediction.

1. The Time difference between the values of each sale or purchase must be the same i.e., the time period between consecutive records must be identical.
2. The Sales data must be listed in a chronological order without the presence of any blank values. In case of any blank values, the values need to be imputed or estimated for ensuring a smooth prediction.

Before processing the data, the previously mentioned contention points were cleared to process the data for Time Series Sales Forecasting to ensure a more accurate and effective prediction of forecast.

For the first point, the data provided had a **Average Time Period of 2 days** with a Deviation of 2 days. To resolve this, the data was **aggregated** into timespans of 2 days and 5 days to check the variance of prediction with respect to the time difference.

To counter the second point, the values where the aggregate comes out to be zero was imputed with the average estimate to ensure that the prediction remains smooth and to avoid a larger variation.

Once this is done, the sales forecasting can be processed in Excel. This uses Linear Regression to compute Sales and Purchase Order Forecast. The formula used for this is as follows;

SALES FORECAST = FORECAST(predict_for_dates, known_dates, known_sales)

In this, the historical values of sales with respect to evenly spaced dates is provided along with the dates for which the prediction is needed. The result would be the predicted value for the unknow date. This formula incorporates linear regression algorithm to identify the trendline.

Additionally, Excel also offers Exponential smoothing Forecasting for a more accurate prediction. This method assigns extra weights to more recent observations while assigning

exponentially decreasing weights for distant observations. The obtained prediction will be the linear sum of weights. This method is found to be unreliable for long term predictions.

The formula used for this is as follows;

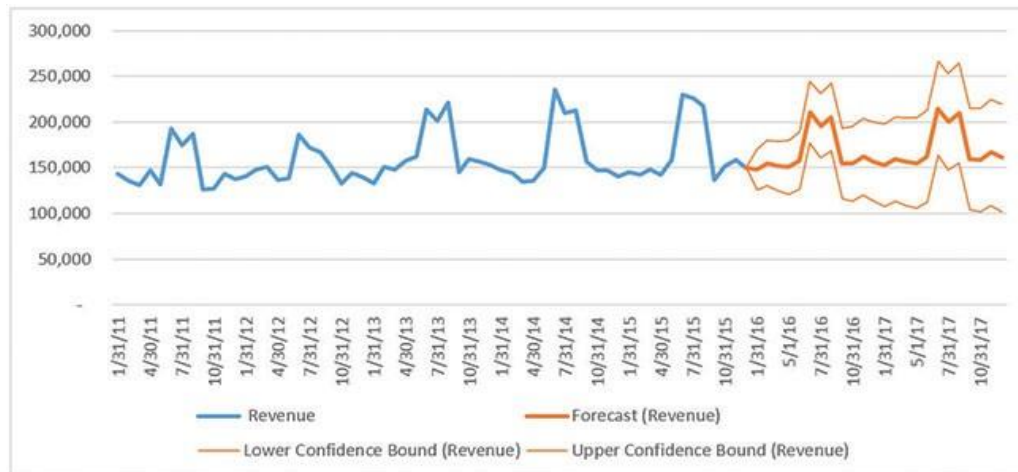SALES FORECAST = FORECAST.ETS(target_date, values, timeline)

Figure 1: An Example of a Forecast

Here, similar to the previous formula, historical data along with the future data to be predicted is required for the exponential prediction. More to this, the forecast can also be used to predict the confidence interval for each of the predictions. This interval can be added and subtracted to the prediction to obtain the upper and lower limits to the prediction. Therefore, the overall prediction can have a range of potential values that can be present.

SALES FORECAST = FORECAST.ETS.CONFINT(target_date, values, timeline)

Overall, the forecast would resemble the image present above and this can be used for deriving observations and drawing out the necessary conclusions.

Results and Findings

Exploratory Data Analysis and Forecasting of Sale Order Dataset

Exploratory Data Analysis (EDA) serves as a foundational phase in data analysis, aiming to unearth patterns, relationships, and anomalies within a dataset. Its purpose is to comprehend the structure and characteristics of the data before formal modeling, enabling informed decisions about subsequent analyses.

1. Data Summary

The sales dataset comprises of list of records generated by SS Services that outlines the customer's request to purchase the respective goods from their organization. This dataset has 515 records spanning from 1st Jan to 30th June, 2023.

	Date	Particulars	Vch Type	Vch No.	Debit\nAmount	Credit\nAmount
0	2023-01-02	Sri Lakshmi Traders Electricals & Hardwares(Arni)	Remi & Polar	SSA/0675/22-23	29331.0	0.0
1	2023-01-02	Mahadev Electricals,Hardwares	Remi & Polar	SSA/0676/22-23	11835.0	0.0
2	2023-01-02	New Darling T.V. Center(Arni)	Remi & Polar	SSA/0677/22-23	1933.0	0.0
3	2023-01-02	Babu Elctricals(Nemili)	Remi & Polar	SSA/0678/22-23	12144.0	0.0
4	2023-01-04	Sri Lakshmi Traders Electricals & Hardwares(Arni)	Remi & Polar	SSA/0679/22-23	32125.0	0.0

Figure 2: Sales Order Dataset

This dataset also comprises of Date, Particulars, Sales Amount. The Date was (as specified before) split into days and months for a more comprehensive analysis of the data.

Post this, the **Debit Amount/Sales Amount** column with the numerical variable (the sales column) was then analyzed for any trends present.

Parameters	Value
Mean	10331.74742
Std.Deviation	7502.083589
Minimum	124
Maximum	43096
Count	515
Sum	53,20,849.92

Table 2: Descriptive Statistics of the Debit Amount Column

On a glance, the dataset has a considerable variability as indicated by the large Standard Deviation present. This results in a wide range of data present from the central reference point (mean).

2. Distribution of Data

In order to understand the plot more, it is imperative that the key variables are explored comprehensively. In this dataset, the variables of importance are **Debit Amount/Sales Amount** and the **Particulars** columns. Through these columns, the data can be inferred and understood. Hence, certain visualization tools were employed to understand the structure of the data present in these columns, paving the way for a more profound understanding of its underlying characteristics.
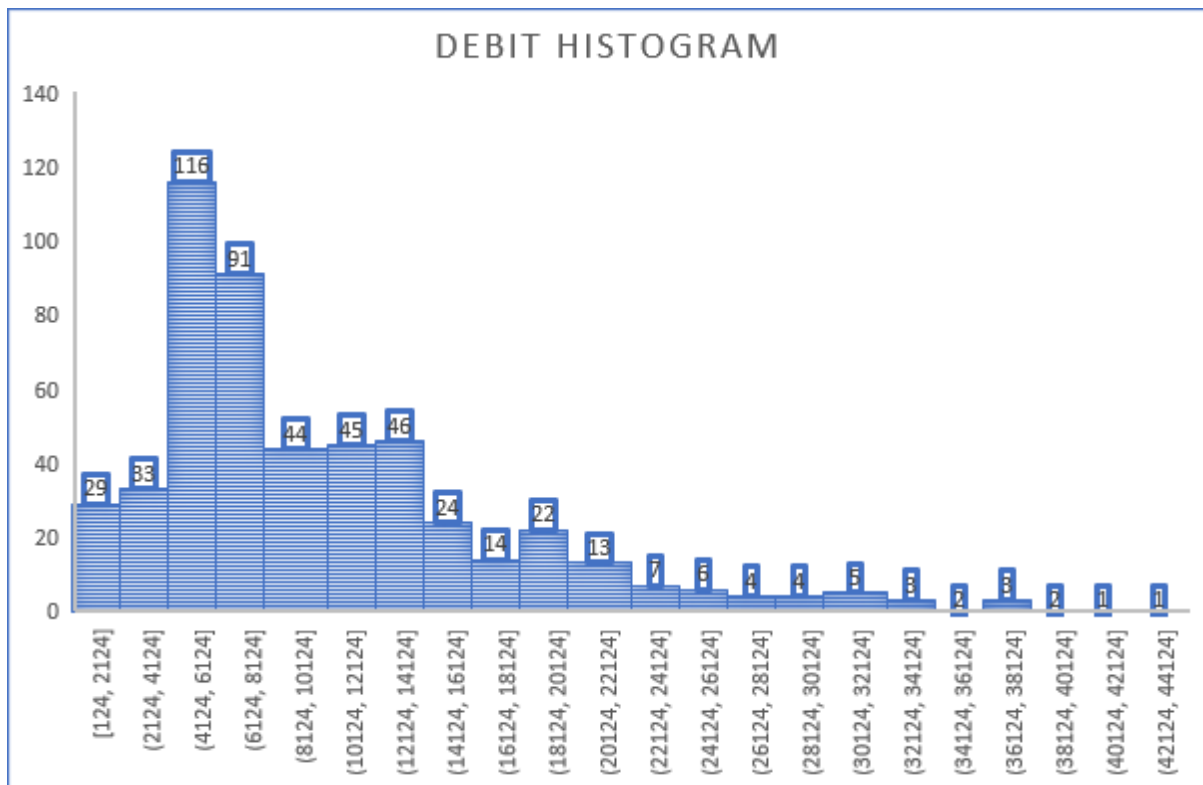

Chart 1: Debit Amount Histogram of Sales Data

The histogram above was drafted with a bin size of 2000. The chart above can be seen to be having a longer tail and reaching a peak at the bin (4124 – 6124). The **skewness was calculated to be 1.60** and **the kurtosis was calculated to be 2.91** for the following variable. This positive value in skewness and kurtosis indicates the longer tail on the right side of the distribution with the tails being heavy when compared to a normal distribution. These values correlate with the histogram present above.

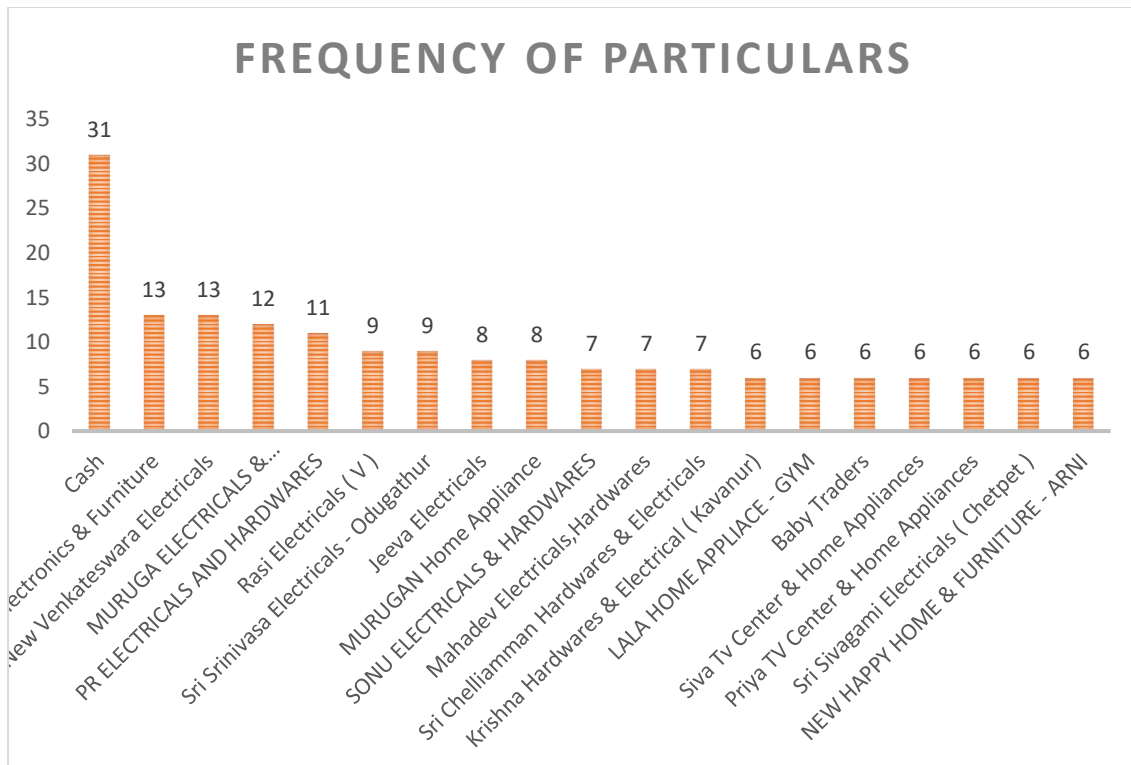The next plot indicates the distribution of the purchasers who had placed the orders for the duration of the dataset.

FREQUENCY OF PARTICULARS

Chart 2: Distribution of the Purchasers in Sales Dataset

Through this plot, it can be inferred that the number of **in hand purchases (cash)** supersedes that of any bulk purchases made by other distributors. The orders placed by various other distributors to Ohm Muruga Services is rearranged in descending order to form a pareto distribution for ease of understanding.

It can be noted that the top purchase made is roughly equal to the next three purchases. After this, the rest of the purchases made by various other distributors are roughly equalizes. This indicates that the Service Center is predominantly accessible to customers. This behavior is acceptable for any good shop. This also leads to a division of sales between on hand customer and distributors. This can be used for further analysis and derivations.

3. Further Analysis of the Dataset

Post analyzing the distribution of the categorical and the numerical variables, the interrelation between these variables were studied for the dataset to glean out valuable insights pertaining to the data. The following charts were obtained from this analysis.
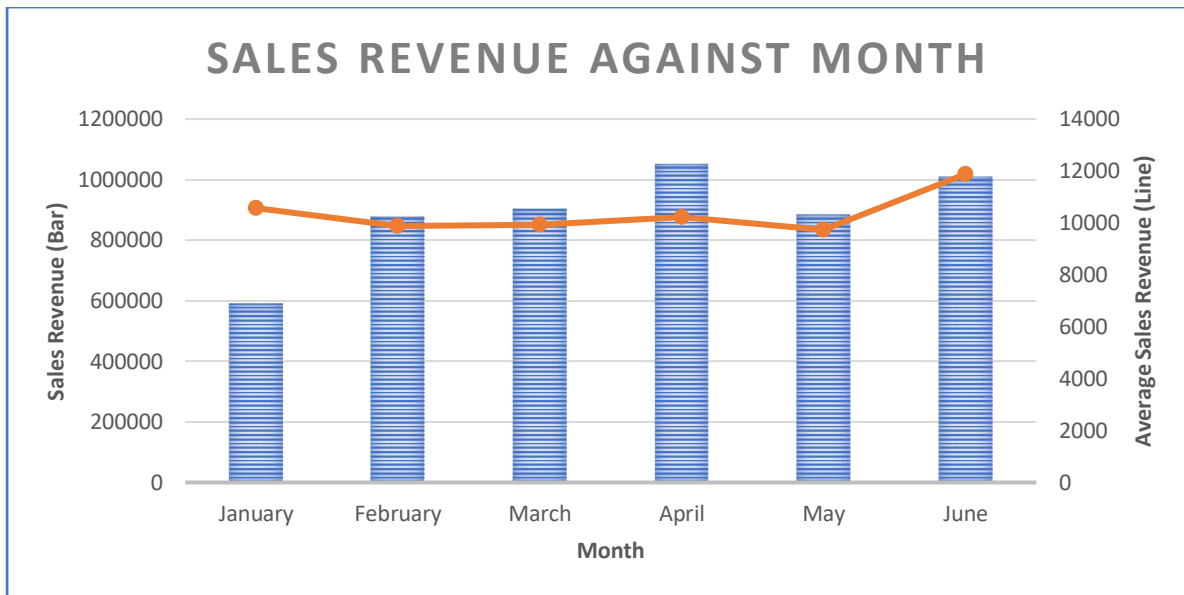
ss

SALES REVENUE AGAINST MONTH

Chart 3: Sales Revenue vs Month

The Sales Distribution was plotted for each month in this combo graph. The line indicates the average sales distribution per month whereas the graph indicates the total Sales Revenue. The **Monthly Sales Revenue shows a steady incline** however, the **Average Sales Revenue per Month remains somewhat stable**. This can indicate that the number of purchases had increased in consecutive months to accommodate the change in the sales revenue per month.

To verify this, the chart of purchase made by various vendors were summed and placed in a vertical pareto chart for a more comprehensive understanding of the data.
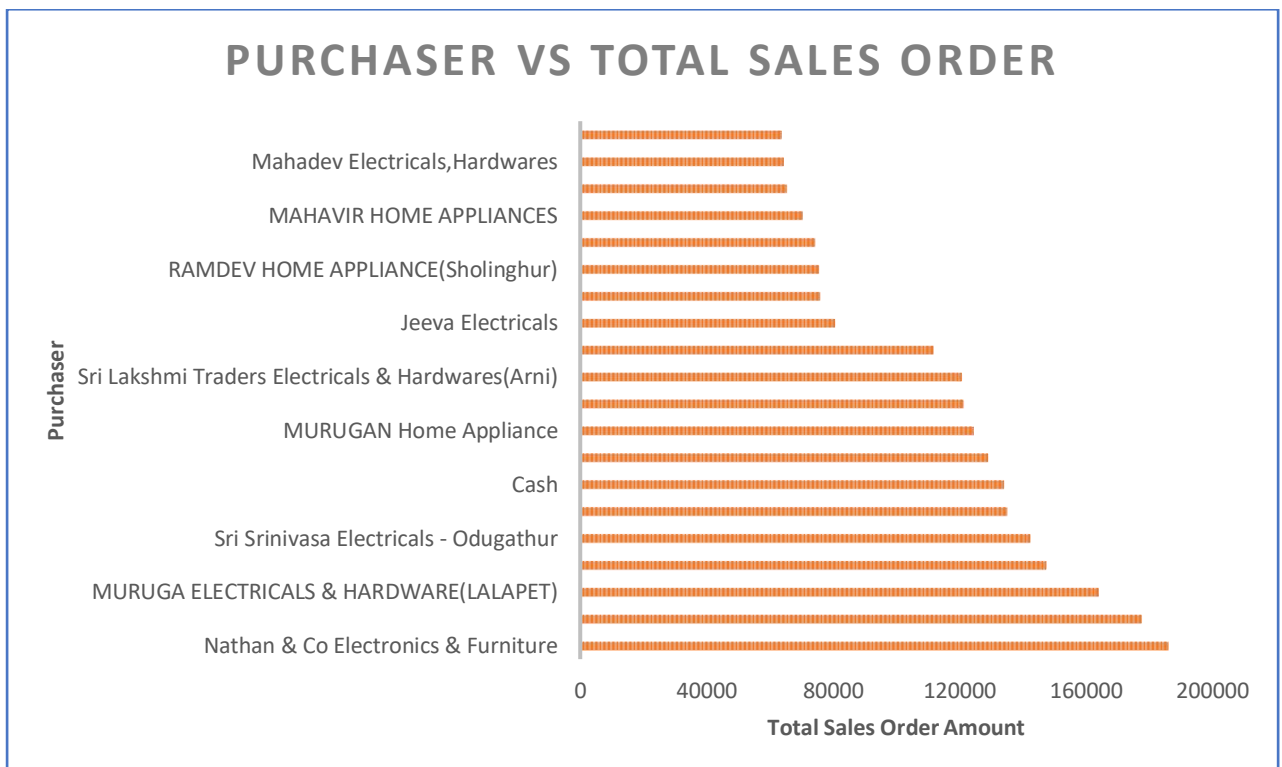
PURCHASER VS TOTAL SALES ORDER

Chart 4: Pareto Chart of Sales Revenue vs the Vendor for the provided Time period

Comparing this with Chart 2 can help in identifying the disparity between the number of purchases and the total sales revenue for each vendor. It can be noted that while the number of **in hand purchases** is greater than any of the other vendor, its corresponding total sales order amount is relatively lesser than some of the vendors. This indicates that the quantity of the product purchased during each transaction is less for in hand purchases in comparison to vendors. This goes well in hand with live observation as well.

Upon further analysis of the sales made on respective days of the week, the following chart was obtained on plotting the number of Sales made on working days for each month.
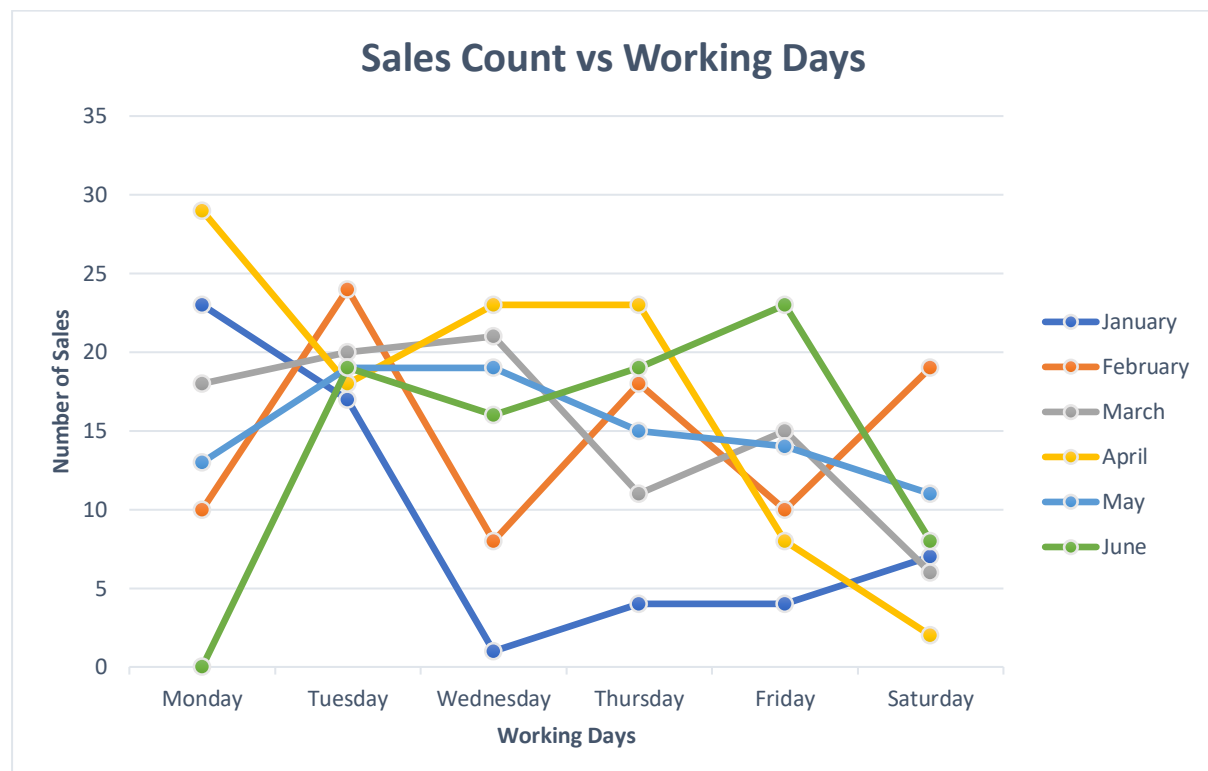

Chart 5: Number of Sales on Working Days for Each Month

This chart helps in understanding the purchases made on each day and the variations of those in each month. It can be noted that Tuesday ends up being the busiest of the days whereas Saturday is the least busy day of the working days. The average was calculated for this and provided in the table listed below.

Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Average	15.5	19.5	14.66666667	15	12.3333	8.833333333

Table 3: Average of Sales vs Day of the Week

This helps in understanding the days of the week where the center is has the most number of purchases and the least.

4. Sales Forecasting of Sale Order Dataset

Sales forecasting for the Sales Order Dataset was done for two instances of time period. This is to identify the time period which offers the most accurate predictions and ease of calculation. From those, the one with the smoothest prediction was identified to be the one with 5 days as time period.
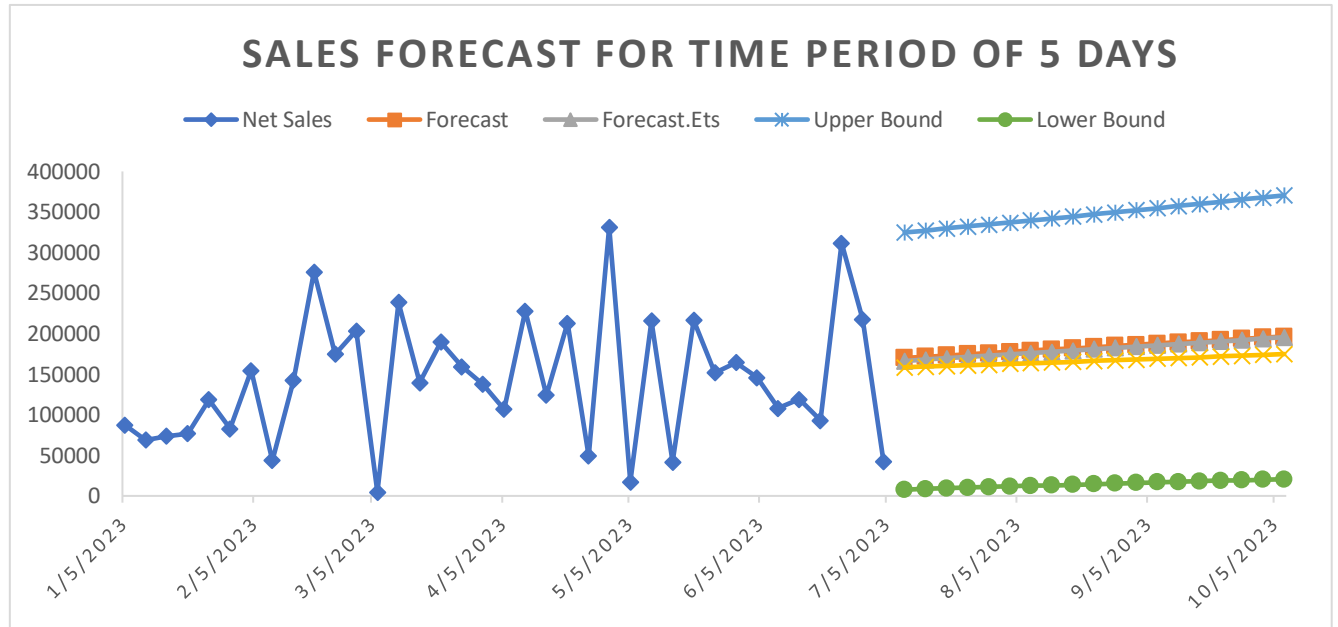

Chart 6: Sales Forecast for Time Period of 5 Days

The Forecast here employs both Linear Regressor as well as Exponential Smoothening techniques. It predicts an upward growth for the organization. However, it has a huge confidence interval which resulting in a larger variance. The prediction takes the values till the month of June and extends further till the month of October for prediction.

Exploratory Data Analysis and Forecasting of Purchase Order Dataset

Exploratory Data Analysis (EDA) serves as a foundational phase in data analysis, aiming to unearth patterns, relationships, and anomalies within a dataset. Its purpose is to comprehend the structure and characteristics of the data before formal modeling, enabling informed decisions about subsequent analyses.

1. Data Summary

The purchase dataset comprises of list of records generated by SS Services that outlines it's request to purchase the respective goods from the vendors. This dataset has 112 records spanning from 1st Jan to 30th June, 2023.

	Date	Particulars	Vch Type	Vch No.	Debit\nAmount	Credit\nAmount
0	2023-01-02	MAHAVIR HOME APPLIANCES	Purchase	126	0.0	28910
1	2023-01-05	MAHAVIR HOME APPLIANCES	Purchase	127	0.0	743
2	2023-01-06	MAHAVIR HOME APPLIANCES	Purchase	128	0.0	41483
3	2023-01-10	SARASWATHI MARKETING	Purchase	129	0.0	8100
4	2023-01-11	MAHAVIR HOME APPLIANCES	Purchase	130	0.0	56663.2

Figure 3: Purchase Order Dataset

This dataset also comprises of Date, Particulars, Purchase Amount. The Date was (as specified before) split into days and months for a more comprehensive analysis of the data.

Post this, the **Credit Amount/Amount** column with the numerical variable (the sales column) was then analyzed for any trends present.

Parameters	Value
Mean	40145.51964
Std.Deviation	35422.59946
Minimum	708
Maximum	208447
Count	112
Sum	4496298.2

Table 4: Descriptive Statistics of the Credit Amount Column

On a glance, the dataset has a considerable variability as indicated by the large Standard Deviation present similar to the Sales Order Data.

2. Distribution of Data

In order to understand the plot more, it is imperative that the key variables are explored comprehensively. In this dataset, the variables of importance are **Credit Amount/Purchase Amount** and the **Particulars** columns. Through these columns, the data can be inferred and understood. Hence, certain visualization tools were employed to understand the structure of the data present in these columns, paving the way for a more profound understanding of its underlying characteristics.
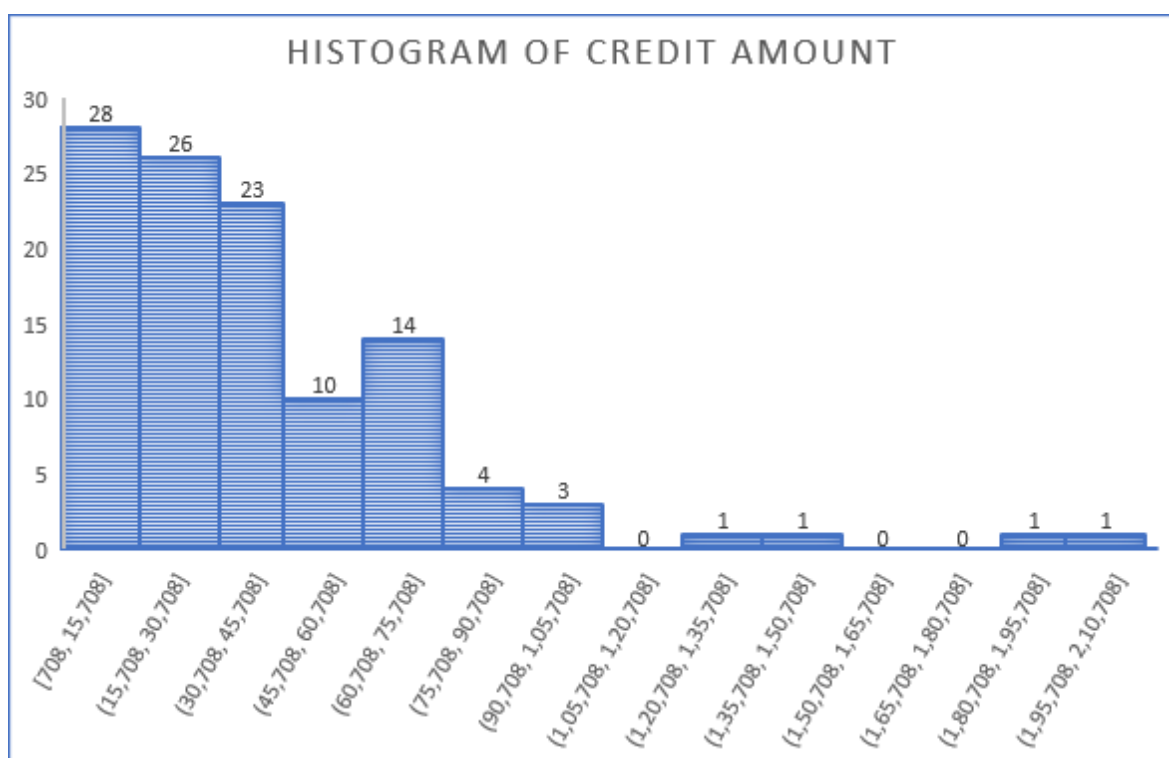

Chart 7: Histogram of the Credit Amount

The histogram above was drafted with a bin size of 15000. The chart above can be seen to be having a longer tail and reaching a peak at the bin (4124 – 6124). The **skewness was calculated to be 2.09** and **the kurtosis was calculated to be 6.69** for the following variable. This positive value in skewness indicates the longer tail on the right side of the distribution where as a positive kurtosis indicates a distribution with far heavier tails and more pronounced peaks in contrast to the normal distribution.

However, in contrast to the sales dataset, for purchase orders, Ohm Muruga Service Center relies solely on just 5 authorized dealers. The distribution amongst these dealers is indicated in the pie chart listed below.
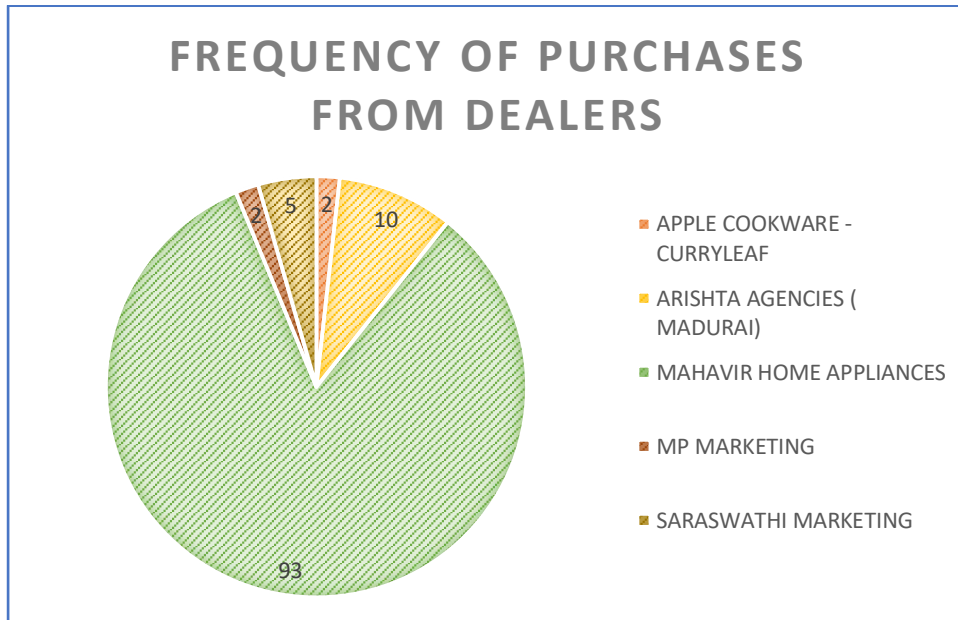

Chart 8: Frequency of Purchases from Dealers

Through this plot, it can be inferred that out of the 5 distributors, the primary distributor **Mahavir Home** Appliance comprises of more than 80 % of its purchase and supersedes purchases from other distributors.

3.   Further Analysis of the Dataset

Post analyzing the distribution of the categorical and the numerical variables, the interrelation between these variables were studied for the dataset to glean out valuable insights pertaining to the data. The following charts were obtained from this analysis.
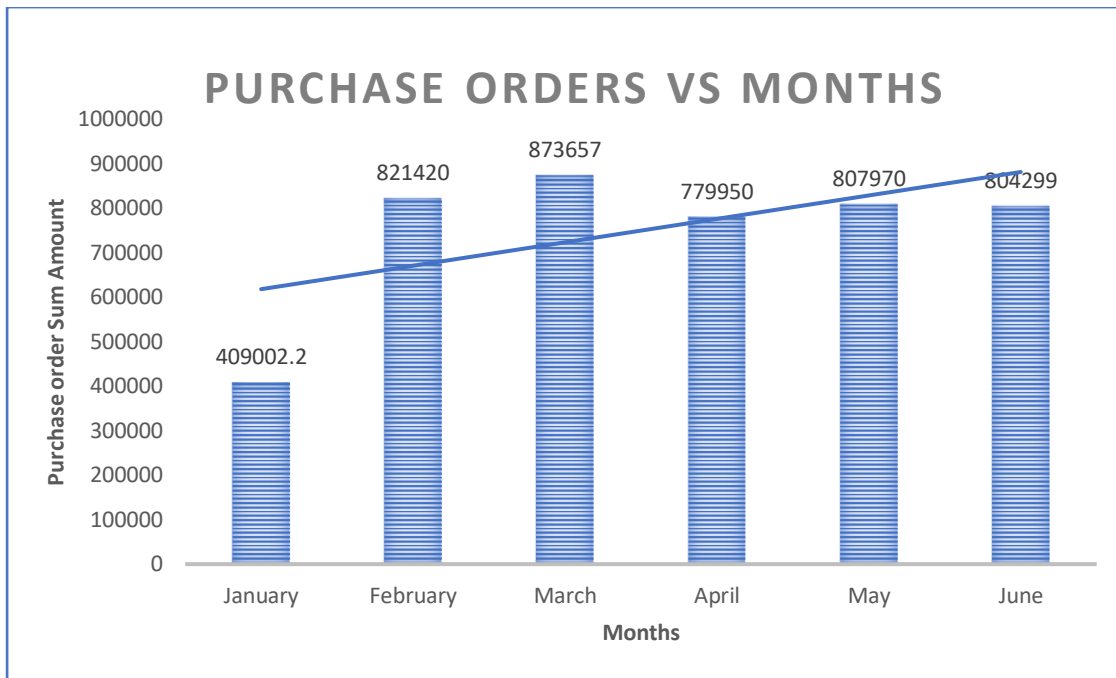
Chart 9: Purchase Orders placed each month

The Purchase Distribution was plotted for each month in this bar graph. Upon extending the trend line, it can be observed that the sum of purchase keeps increasing with consecutive months. To check this with the average of purchases made in each month, it is noted to be stable.
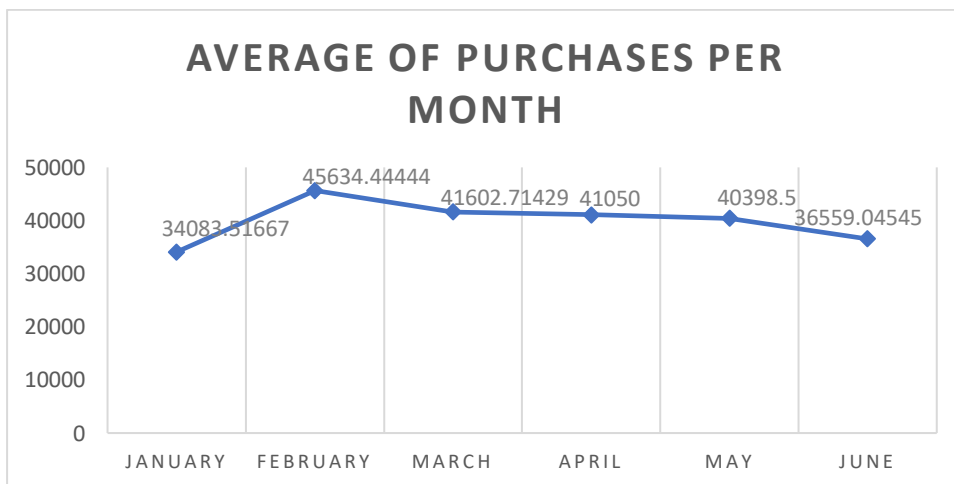

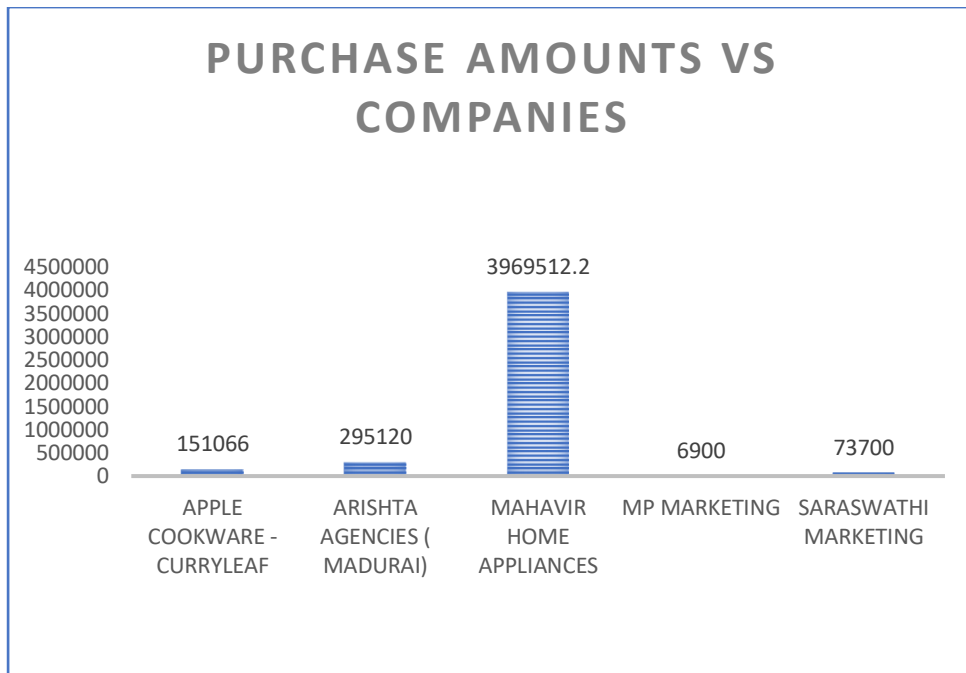*Chart 10: Average of Purchases made each month*

Chart 11: Pareto Chart of Purchase Revenue vs the Vendor for the provided Time period

Comparing this with Chart 8 can help in identifying the similarity between the number of purchases and the total purchase amount for each vendor. As expected, the primary vendor for purchase orders has the most amount by a large margin. This indicates that the quantity of the product purchased during each transaction is less for in hand purchases in comparison to vendors. This goes well in hand with live observation as well.

Upon further analysis of the number of purchases made on respective months for each of the suppliers, the following chart was obtained.
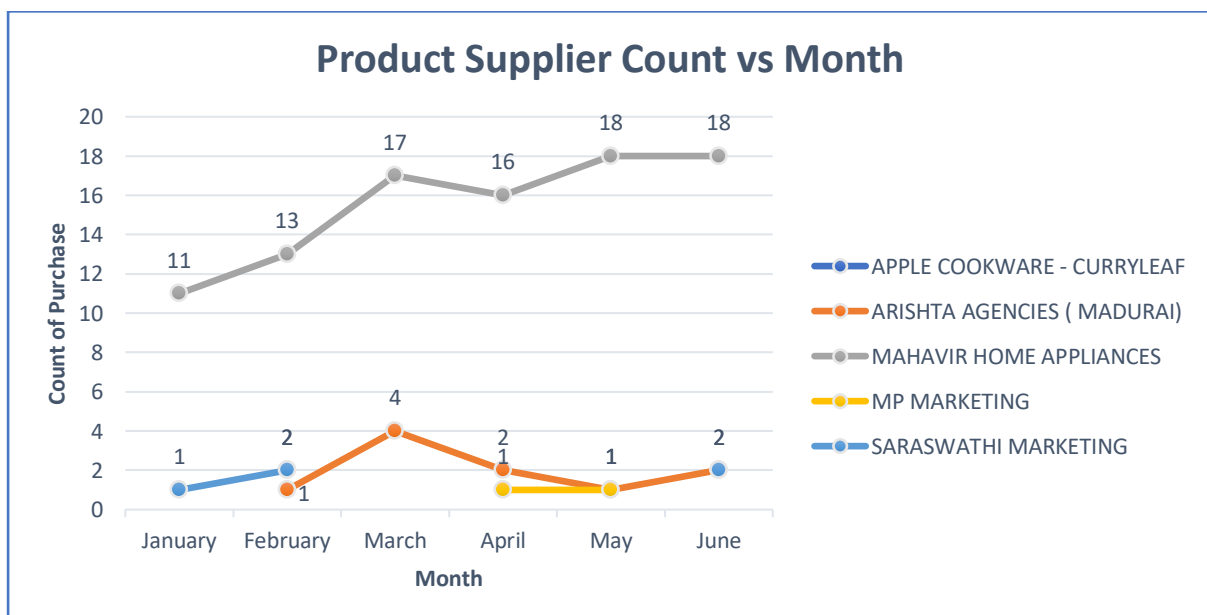

Chart 12: Number of Purchases on for Each Month for each Vendors

This chart helps in the individual purchases made from each of the vendors and their variance per month. It can be noted that the reliance successive count of purchases made by successive months. Its dependence on Mahavir Home Appliance increases with each month as well. However, from other vendors, the purchases are not as prominent. It peaked at March with a total of 4 purchase order placed. Other than that, Mahavir Home Appliance remains the predominant supplier for Ohm Muruga Services.

4. Sales Forecasting of Purchase Order Dataset

Sales forecasting for the Sales Order Dataset was done for two instances of time period. This is to identify the time period which offers the most accurate predictions and ease of calculation. From those, the one with the smoothest prediction was identified to be the one with 5 days as time period.
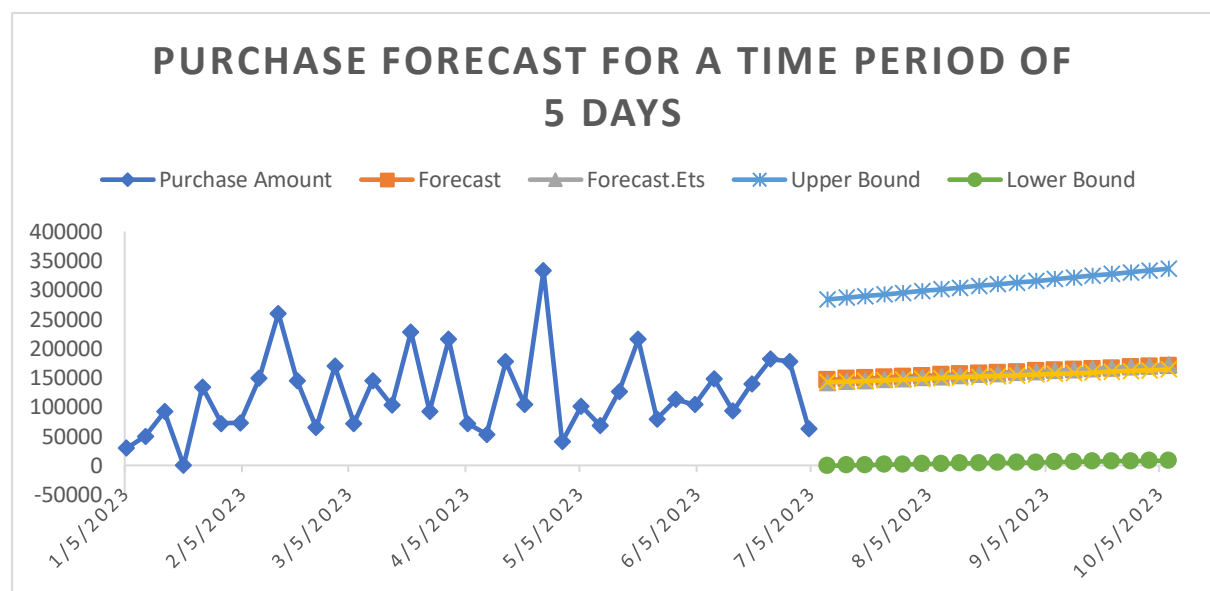

Chart 13: Purchase Forecast for Time Period of 5 Days

The Forecast here employs both Linear Regressor as well as Exponential Smoothening techniques. It predicts a somewhat moderate growth for the organization in terms of the number of purchases made. However, it has a huge confidence interval which resulting in a larger variance. The prediction takes the values till the month of June and extends further till the month of October.

Interpretation and Recommendations

Interpretations of Findings

Upon performing the following analysis, the following critical points emerged that helped in shredding the dynamics of the sales performance of the Organization. The following interpretations were observed on analyzing the Sales Dataset.

a) Most of the purchases made from Ohm Service Center lie in the range of INR 4000 – INR 8000.
b) Most frequent as well as the greatest number of purchases are made by customers on demand rather than on bulk by some external distributors.
c) Even though in-demand purchases tend to be the most in frequency, external distributors purchase items with more expense.
d) Tuesdays tend to be their busiest days whereas Saturdays tend to have reduced number of sales.
e) The Sales Forecast predicts an increasing demand thereby resulting in a rise of further purchase by vendors.

The following interpretations listed below were observed upon analyzing the Purchase Order Dataset.

a) Most of the purchases made by Ohm Service Center lie in the range of INR 400 – INR 40,000.
b) There were only 5 vendors from which offers the products. Out of which, there is an over 83 % dependence in one of the Vendors in contrast to others.
c) There is a variation in the purchase amount with respect to the month, but the average purchase amount spent per month remains the same.
d) Purchase Forecast indicates a very gradual/no incline of purchasing amount. This lies in contrast to the increase in sales forecast predicted above.

Key Recommendations

With a deeper understanding of the Sales and Purchase Order Data, the following recommendations can be implemented to improve the performance of the Sales.

1. There exists a disparity between the Sales and the Purchase Order. This can result in increase of demand without a corresponding increase in the supply. To compensate this imbalance, **key suppliers need to be notified of this to increase the purchase.** Suppliers can be offered incentives and other bulk discounts to increase the goods.

2. The dependency on a single vendor for the goods must be reduced. The **supplier base needs to be diversified** to create a really stable supply chain. A comprehensive analysis

of the suppliers needs to be performed and other reliable suppliers needs to be added with suitable incentives to improve the stability of the supply chain.

3. Since, In-demand purchases are greater than bulk purchases, **critical items need to be prioritized** to meet up with the immediate needs. The most purchased item in the in-demand purchases needs to be in high stock in contrast to other items.

4. The service center sees a higher sale on Tuesday than the other days by a margin of 25%. This means there can be drives to increase the customer sales on other days as well. **Staff levels needs to be optimized** to accommodate the higher purchases on Tuesdays.

<u>Conclusion</u>

The interpretation and recommendation section serves to conclude this Sales Forecast and Analysis report, transforming data into actionable insights that can be implemented into the daily working to make visible changes to the Profits.