Demand Forecasting with Seasonality

Final Submission for the BDM capstone Project

Submitted by

Name: Vasuki Durairaj

Roll number: 22f2001500

IITM Online BS Degree Program

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Table of Contents

Demand Forecasting with Seasonality

1 Executive Summary and Title

Accurate demand forecasting is vital to optimize inventory level and reduce logistic costs, storage costs. This project deals with how the seasonality affects the demand by using several forecasting methods. SS Medicart is a private Pharmacy located in Madurai, Tamilnadu. The company operates in the Business-to-Customer (B2C) sector specializing in procuring medicines from pharmaceutical distributors and selling them to individual Customers. The Pharmacy services primarily to local people near their area by selling schedule H drugs (Prescription drugs), Over the Counter (OTC) drugs and Fast Moving Consumer Goods (FMCG). Even though the area in which the Pharmacy is located has a ward population of around 50,000. The major business issues are reduced profit margin, losing Customers due to lack of Customer satisfaction and repetitive stock out situation.

The issues will be addressed by analysing the data via the following steps. Actual sales data for 4 financial years from April 2020 to March 2024 were used to construct the Dataset.

Firstly, the Dataset is analysed for the seasonal pattern or trend over the years with respect to season. The correlation between sale quantity and season is tested using Chi square test method.

Secondly, Forecasting methods like *Linear forecast, Moving average, Seasonality with Trend Forecast, Time Series Decomposition, Exponential Smoothing, Holt's Winter Exponential Smoothing, Holt's Exponential Smoothing, Multiple Linear Regression* are carried out to forecast the medicine demand.

Finally, the measuring errors like MAD, MSE, RMSE, MAPE are calculated for both training and test data. The best method is chosen for Demand Forecasting and to optimize inventory management. Therefore, predicting demand for the top five category products will detect potential stockouts, plan drug supply and respond to needful patients.

The Segmentation of the Customers helps to improve Happy Customers and define focused Customer Care. The Customer Segmentation is carried out according to their spending in the period. They are classified into four groups. The total number of visits greater than 40 and more than ₹100000 spent in the period is the major factors of classification. Different strategies are suggested to improve Customer satisfaction such as identifying the Value Customers, Demand Forecasting for Group1 and Group 2 Customers and increasing the discounts for Group 3 and Group 4 Customers.

The expected outcome helps the Pharmacy to improve in marketing strategies and Customer count, identifying the demand pattern and to categorize the investment. Limitation of the project includes individual demand forecasting seems to be a wide topic and hence it is done as a sample simulation block for a Group1 Customer. We only consider seasonality factor in the sale but other factors like promo, stock availability are not provided in the Dataset and hence the relationship of these factors with sales is not calculated. Other factors improving profit margin such as using generic medicine instead of branded medicine are excluded.

2 Detailed Explanation of Analysis Process

Although Pharmaceutical companies have fairly enough data, they do not effectively utilize these data to assist in business decisions. Demand forecasting is an important utilization of the recorded data. Hence data mining is done to find the hidden relationship between elements and to find the best model for forecasting.

2.1 Problem Statements

The following are identified based on the concerns informed by the Pharmacy owner.

1. To improve the profit margin of the business.
2. Enhance better Customer satisfaction and identify marketing strategies.
3. To reduce the repetitive stock out situation due to which the business is facing loss of Customers.

2.2 Definitions

Pharmacy	-	SS Medicart, Madurai
Customer	-	Who buys Medicines from the Pharmacy
Dealer or Distributor	-	Agency or Company Supply medicines to the Pharmacy

2.3 Dataset

The total Sales Report of the last four financial year periods which is from April 2020 to Mar 2024 is given by the Pharmacy in an MS Excel format. Also, the Purchase report provides details of the Dealers or Distributor supplies.

Refer the Links for Sales Report Dataset and its Calculations.

https://docs.google.com/spreadsheets/d/1OwyAg7iu9AG_37DcrIWVSC5qifddY8lN/edit?usp=sharing&ouid=104125864226783131288&rtpof=true&sd=true

https://docs.google.com/spreadsheets/d/1zMHcGyqs7wVSb47f8LBwTTXFLbIGAy0y/edit?usp=sharing&ouid=104125864226783131288&rtpof=true&sd=true

https://docs.google.com/spreadsheets/d/1M4ZpkhcLZaJKi5uPA8YMV33vNvqcCPnF/edit?usp=sharing&ouid=104125864226783131288&rtpof=true&sd=true

Refer to the Links for Purchase Report Dataset and its Calculations.

https://docs.google.com/spreadsheets/d/1R0vJT6lNK3L_23tEbZ4RglelOIvCnTr7/edit?usp=sharing&ouid=104125864226783131288&rtpof=true&sd=true

Refer the Link for Meta Data for the Datasets.

https://docs.google.com/document/d/117ZrcbB3eET449DtW8mbdm4YqQK32e8C/edit?usp=sharing&ouid=104125864226783131288&rtpof=true&sd=true

2.4 Data Preprocessing

Data cleaning is the major process where the data is organised in a suitable form in order to get better results. In this project, the Dataset is cleaned and processed using MS Excel.

2.4.1 Sales Report Preprocessing

The features that contain Names are cleaned by clustering. The same Names with different spells are corrected through open refine(tool) and some rows manually. The Phone numbers of the Customers are masked for privacy and security reasons. Using Delimiter, the value for the Total Quantity in a Pack feature is taken from the Medicine Name which includes the numbers that represent the total number of medicines in the pack. Null Value and unknown value rows in each feature are removed for accurate prediction. New features like Season, Medicine Category, Medicine Use, Year, Season Number, Sale Amount, Total Quantity in a Pack etc are added to the Sales Report table. Refer sections 2.4.3 and 2.4.4 for medicine categories and seasons.

2.4.2 Purchase Report Preprocessing

For the Purchase report the Dealer names are clustered and renamed manually by creating a Dealer Name table as there are only few Dealers and the same is corrected in the Purchase data by VLOOKUP function. Since in the Purchase data report, date is not available, a separate column is maintained for year.

2.4.3 Medicine Categories

There are a total of 1276 medicines names in the report. In order to predict the sale of Medicine, the categorisation of medicine is carried out and split into 18 major categories. A separate table is created for Medicine name, its pharmacological usage and its category. For example, the medicine PAN 40MG TABLETS 15'S is used for Peptic Ulcer and is categorised under Digestive. The medicine categories are Anti- allergy, Anti-infective, Cardio, Dental, Dermal, Diabetic, Digestive, Endocrine, Excretory, Genito Urinary System, Neuro, Ocular and Others (cotton, syringe, etc.)

Out of 18 categories of medicines the top five categories are chosen based on the sale quantity which is considered for further analysis. Refer Figure 1 for top five categories.

2.4.4 Defining Seasons

The data is categorized into four seasons to determine the seasonal behaviour.

(1) Monsoon – June, July, August, September
(2) Post Monsoon – October, November, December
(3) Summer – March, April, May
(4) Winter – January, February

2.5 Data Analysing Process

2.5.1 Sales Report Analysis

Qty		Rate		DayNum		Season num		Sale Amount	
Mean	18.83871613	Mean	108.6026474	Mean	3.9943339	Mean	2.214945172	Mean	322.4108008
Standard Error	0.106193068	Standard Error	0.556135743	Standard Error	0.011526255	Standard Error	0.006335333	Standard Error	6.872019718
Median	15	Median	85	Median	4	Median	2	Median	87
Mode	30	Mode	105	Mode	5	Mode	1	Mode	105
Standard Deviation	18.39409857	Standard Deviation	96.33035243	Standard Deviatior	1.99650582	Standard Deviation	1.097366699	Standard Deviation	1190.328242
Sample Variance	338.3428621	Sample Variance	9279.536798	Sample Variance	3.98603549	Sample Variance	1.204213672	Sample Variance	1416881.324
Minimum	1	Minimum	4.65	Minimum	1	Minimum	1	Minimum	0.045
Maximum	300	Maximum	3500	Maximum	7	Maximum	4	Maximum	61740
Sum	565218	Sum	3258405.23	Sum	119842	Sum	66455	Sum	9673291.256
Count	30003	Count	30003	Count	30003	Count	30003	Count	30003

Table 1: Descriptive Statistics for sales report.

From Table 1 it is evident that the pharmacy is selling more than 5,60,000 quantities of medicine in the given analysis period. Therefore, the Pharmacy purchases more than 1,40,000 medicines on average per year. The Pharmacy could increase its revenue by demand forecasting the quantity of medicine. If the medicines are ordered in large quantities with the same Dealers in a year, the Pharmacy could avail larger Dealer discounts from Dealers. Moreover, the mode for quantity is 30. Hence most of the Customers buy 30 medicines which represents the Customer mentality to buy for a month or for half month.

From Table 1, it is estimated that the average price of the medicine is in the range of ₹108+/-96 per pack. This is considered an affordable range. The Pharmacy also sale some medicine on a maximum price of ₹3500. There is high variance in the price range of medicines. Procuring medicine from the Dealers located in the same area which is nearby to the Pharmacy area would reduce the travel cost.

Table 1 shows the mode for the Season Num is 1 so, the sale is higher in Monsoon season. It also shows that sales on Fridays are higher. The Pharmacy earns on an average of ₹ 322 for a sale.

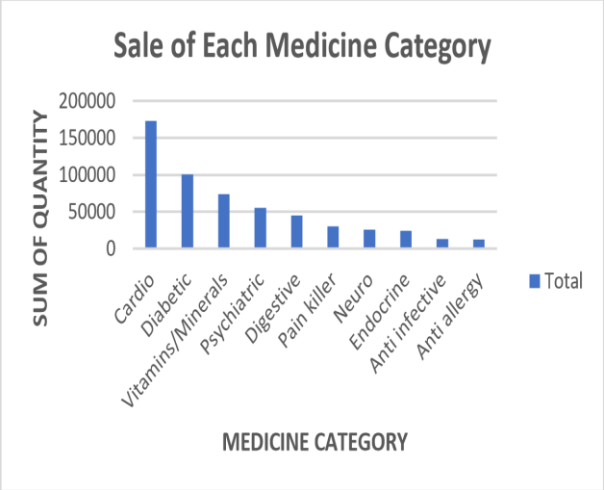


Figure 1: Sale of each medicine category

Medicine Category	Sum of Qty
Cardio	173348
Diabetic	100551
Vitamins/Minerals	73821
Psychiatric	55092
Digestive	44865
Pain killer	30209
Neuro	25775
Endocrine	24482
Anti infective	13098
Anti-allergy	12407

Table 2: Medicine Category and Sale quantity

From the above Figure 1 and Table 2, more quantities of medicines are purchased by Customers in the medicine categories of Cardio, Diabetic, Vitamins/Minerals, Psychiatric and

Digestive. The analysis forecast the total quantity of these top five categories of medicine in accordance with the seasons.

2.5.2 Purchase Report Analysis

The Purchase Report is available only for three financial years from April 2021 to March 2024. As shown in Table 3, the medicine purchased from a Dealer at single order was a maximum of 2196 medicines. This reduced the Dealer discount since the order quantity was very less.

The descriptive Statistics of the Purchase Report data (Table 4 below), shows that the variance for Quantity is high. Hence, it is understood that the medicines were not ordered in a planned manner.

The Dealers had provided free medicines for larger orders. As per the Table 4, the maximum of 870 free medicines were given by the Dealer for single order. This shows that when order quantity increases with the Dealer, free medicine or Dealer-discount might increase. Nearly 2% of the total medicine is free of cost from the distributor in the three years. This could be improved by ordering based on seasonal demand.

PALEPU PHARMA PVT LTD	SP.S.S MEDICAL HALL	VASAVI MEDICARE SOLUTIONS PVT.	KASIVISWANATHAN MEDICAL	IRAGU ENTERPRISES
714	1993	2196	851	850
710	1917	1123	710	780
704	1350	825	595	593

Table 3: Top three quantity medicine ordered with the distributor in a single purchase.

Quantity		*Free*	
Mean	27.95375786	Mean	0.547799945
Standard Error	0.326410944	Standard Error	0.085610281
Median	20	Median	0
Mode	30	Mode	0
Standard Deviation	44.15002134	Standard Deviation	11.57956195
Sample Variance	1949.224384	Sample Variance	134.086255
Range	1469	Range	870
Minimum	1	Minimum	0
Maximum	1470	Maximum	870
Sum	511414	Sum	10022
Count	18295	Count	18295

Table 4: Descriptive statistics for purchase data.

2.5.3 Analysis of Sales Pattern

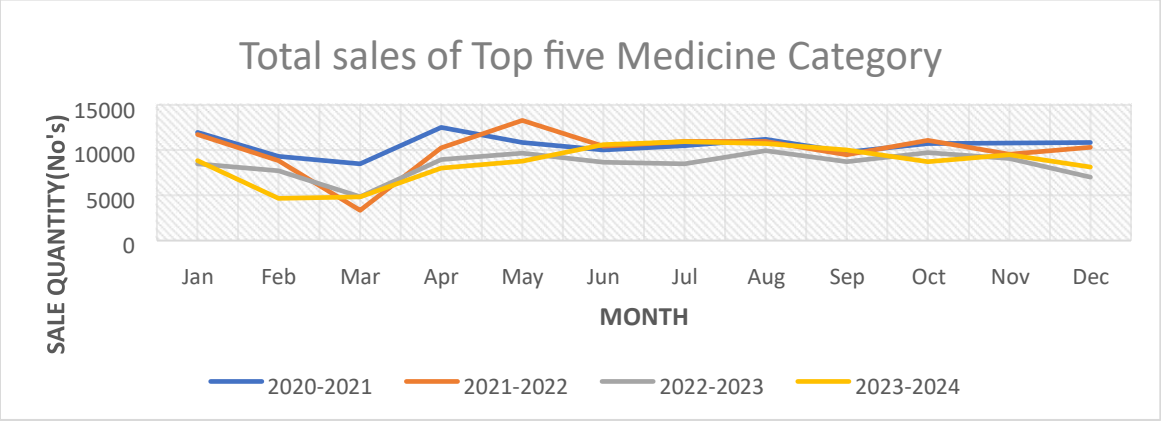


Figure 2: Total sales of the top five medicine categories.

The graph above shows the sale of the total quantity of the top five medicine categories for every year which shows reducing sale till March and then the sale grows up till May and follows a high quantity sale pattern (for all four years) till September and then decreases till the end of the year.

Figure 3: Seasonal quantity of sale

The sales increase from winter to monsoon and attend its peak in monsoon and start to decrease till summer. In Figure 3 the seasonality pattern is clearly visible. There is a continuing pattern observed in response to the season. Hence, to prove the association between season and top five medicine category sales, correlation between season and total sale quantity of each category is calculated using *Chi square test*.

2.5.4 Correlation between Sales & Seasons

Correlation

Observed	Cardio	Diabetic	Digestive	Psychiatric	Vitamins/Minerals	Grand Total
Monsoon	63109	35672	16332	19624	26753	161490
Post Monsso	46706	23738	11848	14351	18876	115519
Summer	36855	23500	9650	12386	16693	99084
Winter	26678	17641	7035	8731	11499	71584
Grand Total	173348	100551	44865	55092	73821	447677

Season/Medicine Category	Cardio	Diabetic	Digest	Psychi	Vitami	Grand Total
Monsoon	63109	35672	16332	19624	26753	161490
Post Monsson	46706	23738	11848	14351	18876	115519
Summer	36855	23500	9650	12386	16693	99084
Winter	26678	17641	7035	8731	11499	71584
Grand Total	173348	100551	44865	55092	73821	447677

Expected	Cardio	Diabetic	Digestive	Psychiatric	Vitamins/Minerals
Monsoon	62531.62106	36272	16184.1	19873.273	26629.36289
Post Monsso	44730.88323	25946	11577	14215.992	19048.84124
Summer	38366.97716	22255	9929.94	12193.469	16338.74415
Winter	27718.51856	16078	7173.96	8809.2659	11804.05172

Ho = Season and Medicine category are not associated.

P Value 2.13E-132

P value <0.05
Hence reject Ho
Result:
There is association between season and category medicine

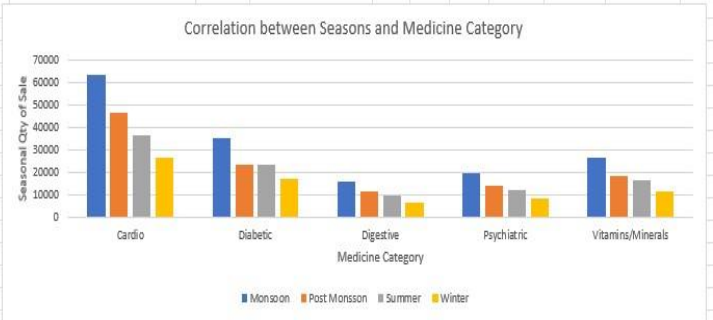


Figure 4: Correlation

P value thus obtained is 2.13E-132 which is less than 0.05 hence there is association between season and medicine sale. The tabulated Chi square value of 21.02 is less than the calculated value of 654.74 null hypothesis is rejected. Hence association between season and category medicine is proved. The graph in Figure 4 clearly shows in every category of medicine the monsoon sale is higher followed by post monsoon sale and then the summer and finally the winter sale.

2.5.5 Classification of Customers

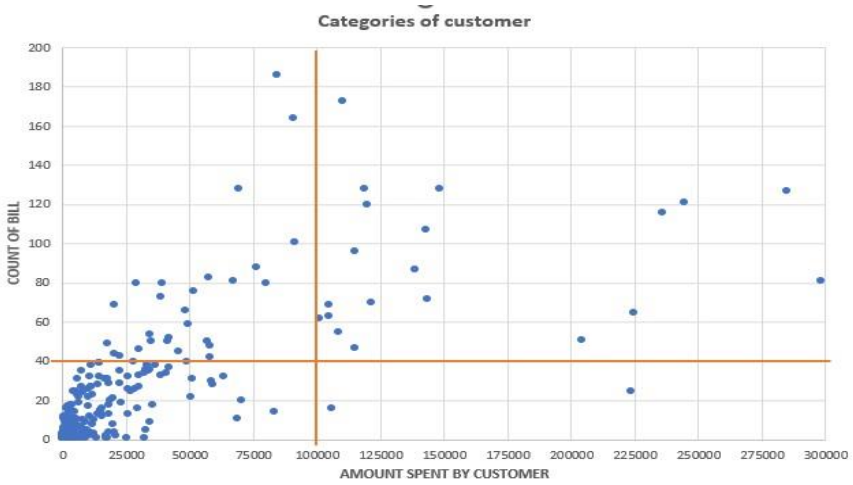


Figure 5: Categories of Customer

In the Figure 5 above it is visible that there are more floating customers. So, we are binning the customers into four groups in reference to the value from the graph. Number of visits greater than 40 times in the 4 years and more than ₹1,00,000 spent in the period are the two factors taken for classification of Customers.

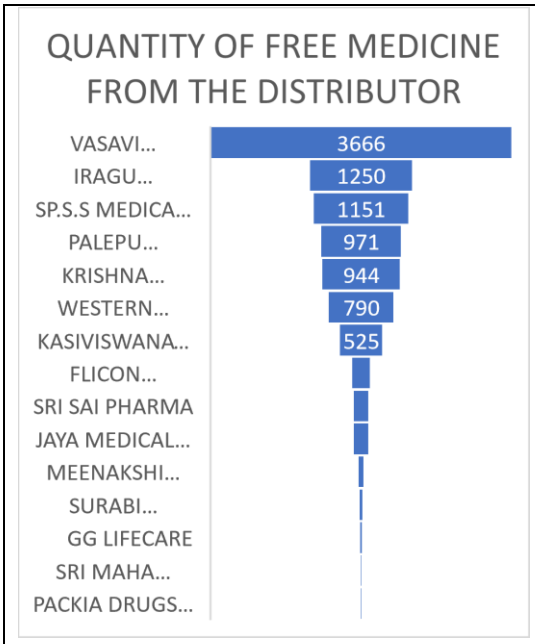
Group 1: Amount spent > ₹1,00,000 and Number of visits > 40 in 4 years

Group 2: Amount spent < ₹1,00,000 and Number of visits > 40 in 4 years

Group 3: Amount spent < ₹1,00,000 and Number of visits < 40 in 4 years

Group 4: Amount spent > ₹1,00,000 and Number of visits < 40 in 4 years

2.5.6 Dealer Transactions Analysis



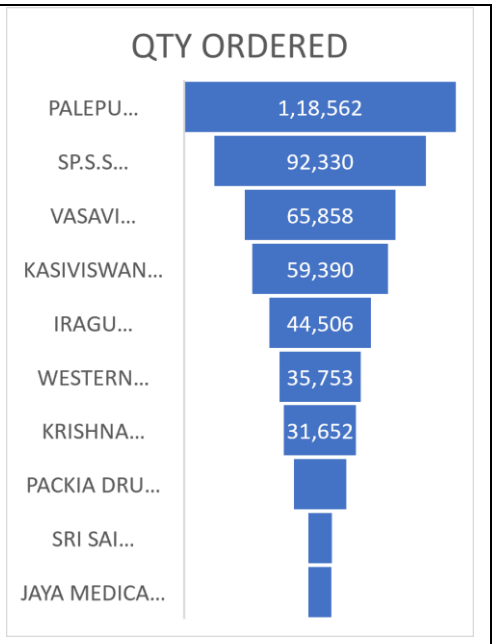


Figure 6: left: Quantity of free medicines from each distributor

Figure 7: right: Quantity of medicines ordered with distributors

With reference to Figure 6 and 7 the Pharmacy is purchasing more than 21% of total medicine from PALEPU PHARMA PVT LTD and getting 9.6% of free medicine of total free medicines. However, the Pharmacy is purchasing 11% of total medicines are from VASAVI MEDICARE and getting 36.5% of free medicines. Though, the order from VASAVI MEDICARE is less, they are offering a better percentage of free medicines.

2.5.7 Analysis for Problem Statements 1 & 3

Demand Forecasting is important from the perspective of the pharmaceutical industry. Pharmacy is subject to a high degree of demand fluctuation due to seasonality. While non-availability of medicines is not usually life-threatening but inconvenience for the Customer which results in loss of Customer loyalty. In this analysing process, *Linear forecast, Moving average, Seasonality with trend forecast, Time series decomposition, Exponential smoothing, Holt's winter exponential smoothing, Holt's exponential smoothing, Multiple linear regression,* The best accurate time series quantitative forecasting model is chosen by comparing the quantitative indicators.

2.5.7.1 Demand Forecasting Models

1) *3month Moving Average*

In the moving average method, the forecast demand (L_{t+1}) is the average of the most recent 3 month sale.

$L_{t+1} = (D_t + D_{t-1} + D_{t-N+1}) / N$ where N is 3 for a 3 month average. L_{t+1} is the demand calculated. D_t is the observed demand in time period t.

Moving Average performs best with static data, but it does not capture the trend and seasonality.

2) *Linear Forecast*

Linear Forecast gives better results when the sale follows an increasing or decreasing trend at a constant rate. The forecast demand is calculated with finding slope and intercept for the history of data. L_t = intercept + (slope * t) where L_t is the demand calculated. 't' is the time period.

3) Forecasting with seasonality and trend

This method is followed in time series analysis for predicting future demands with history of data following seasonal patterns as well as trends. L_t = seasonality index * Linear forecast, where L_t is the demand calculated. Seasonality index (SI) is calculated as

$$SI = \frac{Seasonal\ Average}{Average\ of\ all\ seasons} \times 100$$

4) Time series decomposition

It undercovers fluctuation within data underlying trends and recurring seasonal patterns. The data is broken down into trend, seasonal and residual components. By isolating these components, it is easy to understand the trend and recurring seasonality and forecast the demand.

$Y_t = T_t + S_t + C_t$, where Y_t is the demand forecasted and T_t is the Trend component at period t S_t and is the seasonal component at time period t and C_t is the cyclical component at time period t.

5) Exponential smoothing

Exponential smoothing works by exponentially decreasing weights for the past data. It gives high weightage to recent observations. It helps to capture short term variation and trends in data.

$$\hat{Y}_{t+1} = \alpha \cdot Y_t + (1 - \alpha) \cdot \hat{Y}_t$$

$Y\hat{}_{t+1}$ is the forecast for next time period. Y_t is the actual value observed at t. $Y\hat{}_t$ is the forecast for current time period t. α is the smoothing parameter that ranges between 0 and 1.

6) Holt's Winter Exponential Smoothing

Holt's winter method is used to analyse the seasonal effect in the data. This method has three components: trend, level, and seasonality. If the seasonal effect in the data shows increasing or decreasing trend Multiplicative Holt's winter exponential smoothing is used and additive method is used when the seasonal pattern shows increase or decrease with irregular tendency.

$$L_t = \alpha \times \left(\frac{Y_t}{S_{t-m}}\right) + (1-\alpha) \times (L_{t-1} + T_{t-1})$$
$$T_t = \beta \times (L_t - L_{t-1}) + (1-\beta) \times T_{t-1}$$
$$S_t = \gamma \times \left(\frac{Y_t}{L_t}\right) + (1-\gamma) \times S_{t-m}$$
$$\hat{Y}_{t+h} = (L_t + h \times T_t) \times S_{t-m+(h}$$

Where L_t is the level component and T_t is the trend component at time t, m is the seasonal length. S_t is the seasonal component at t, $\hat{Y}_t$ is the forecast for current time period t, $\hat{Y}_{t+h}$ is the forecast value in h period.

Optimum α, β, γ are found using Solver in Excel.

7) Holt's Exponential Smoothing

Holt's exponential is an extension of exponential smoothing where two components level and trend are calculated but no seasonality. Initial value for level and trend is obtained by calculating linear regression between demand and time period which is of the form.

$D_t = at + b$ where 'b' is the initial level component and 'a' is the initial trend component.

$$L_t = \alpha \times Y_t + (1-\alpha) \times (L_{t-1} + T_{t-1})$$
$$T_t = \beta \times (L_t - L_{t-1}) + (1-\beta) \times T_{t-1}$$
$$\hat{Y}_{t+h} = L_t + h \times T_t$$

α, β are the smoothing constant for level and trend respectively and it ranges from 0 to 1.

Where L_t is the level component and T_t is the trend component at time t, $\hat{Y}_t$ is the forecast for current time period t and $\hat{Y}_{t+h}$ is the forecast value in h period.

8) Multiple Linear Regression

Multiple linear regression for categorical variables can be done by using dummy variables. For 4 seasons 3 dummy variables are used. For Monsoon sale, season 1 variable takes the value 1 and season 2 and season 3 variable is zero. For Post Monsoon, season 2 variable takes value 1 and season1 and season 3 variable takes the value 0. For Summer, season 3 variable takes the value 1 and season 1 and season 2 variable takes the value 0. For Winter, all the variables season 1, season 2 and season 3 takes the value 0.

Regression is done by the data analysis pack in MS Excel. And the intercept, the coefficient of time and the coefficient of three seasons are fetched.

$\hat{Y}_t = \beta_0 + t_0 * T + \beta_1$*(season1 variable) + β_2*(season2 variable) + β_3*(season3 variable)

Where, β_0 is the intercept, β_1, β_2, β_3 are the coefficients of three seasons and t_0 is the coefficient for time period T. $\hat{Y}_t$ is the forecast for current time period t.

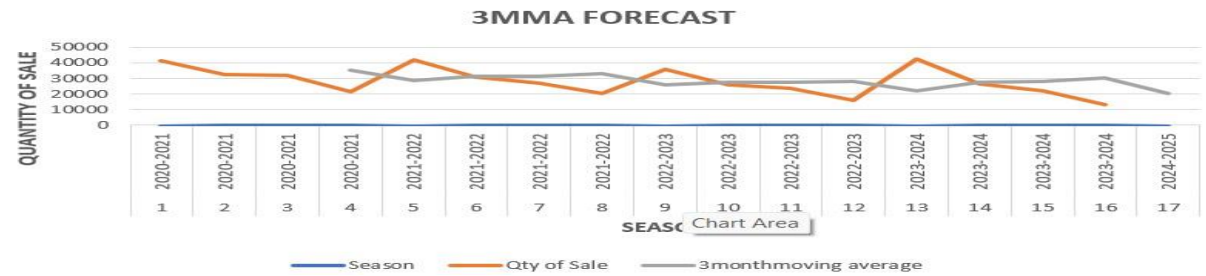


Figure 8: Represent the 3 month moving average.

The orange curve is the total sale of top 5 medicine categories with respect to season and the grey curve is the moving average curve. In this method the demand is forecasted as per the previous three month sale. Hence the forecast is not so accurate in detecting seasonality and trend.

Figure 9: Represent Linear forecast method.

The orange curve is the total sale of top 5 medicine categories with respect to season and the grey curve is the linear forecast curve. This method captures the trend in the sale. Since the trend is decreasing the sale predicted is decreasing in nature.

Figure 10: Represent Forecasting with seasonality and trend.

The orange curve is the total sale of top 5 medicine categories with respect to season and the yellow curve is the forecast curve. Since the trend is decreasing, the pattern follows a decrease form with seasonality. The forecast is capturing the seasonality better than the former models.

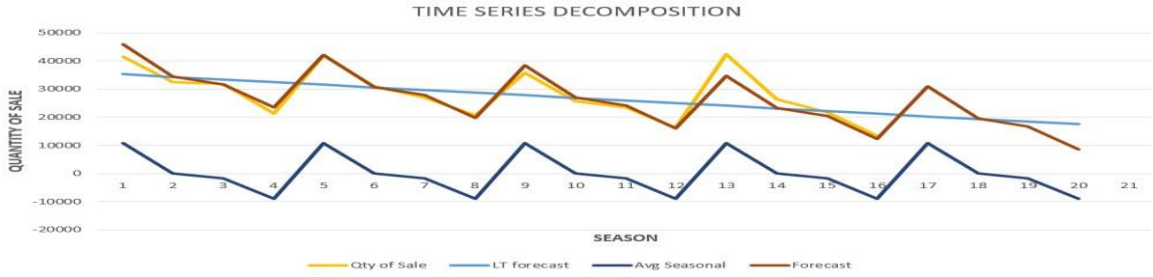


Figure 11: Represent Time Series Decomposition.

The yellow curve is the total sale of top 5 medicine categories with respect to season and the red curve is the forecast curve. Since the forecast well captures the trend and seasonality in the data and MAPE is less than the last method, this TSD model is predicting sales better.

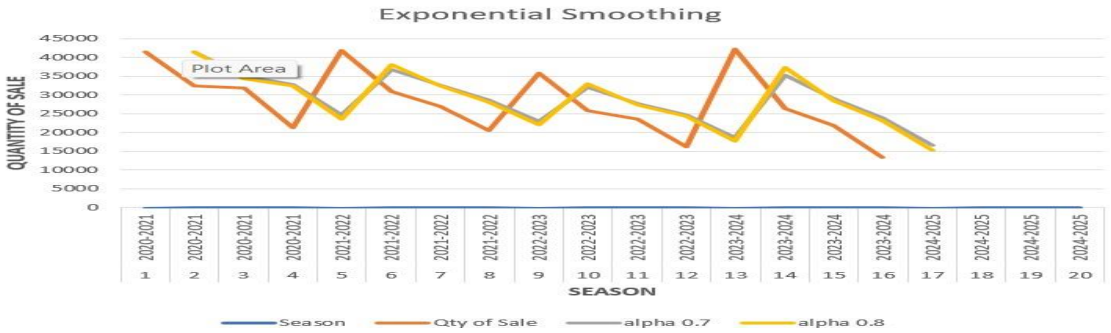


Figure 12: Represent exponential smoothing.

The orange curve is the total sale of top 5 medicine categories with respect to season and the yellow and grey curve is the forecast curve with alpha value 0.7 and 0.8. Even though it is detecting seasonality to some extent the error is higher in predicting the demand.

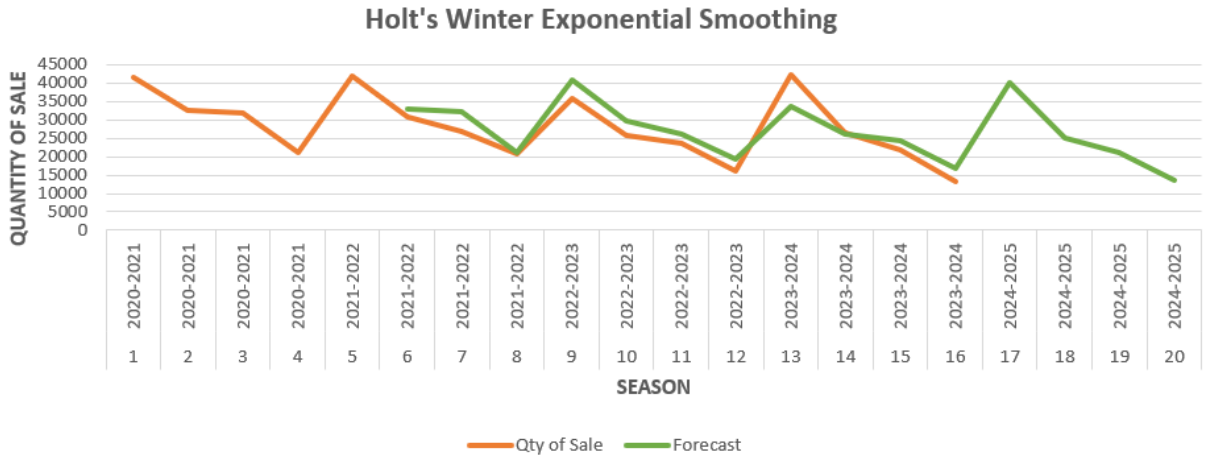


Figure 13: Represent Holt's winter exponential smoothing.

The orange curve indicates the total sale of top 5 medicine categories with respect to season and the green curve is the forecast curve with $\alpha=0.2$, $\beta=0.04$, $\gamma=0.9$

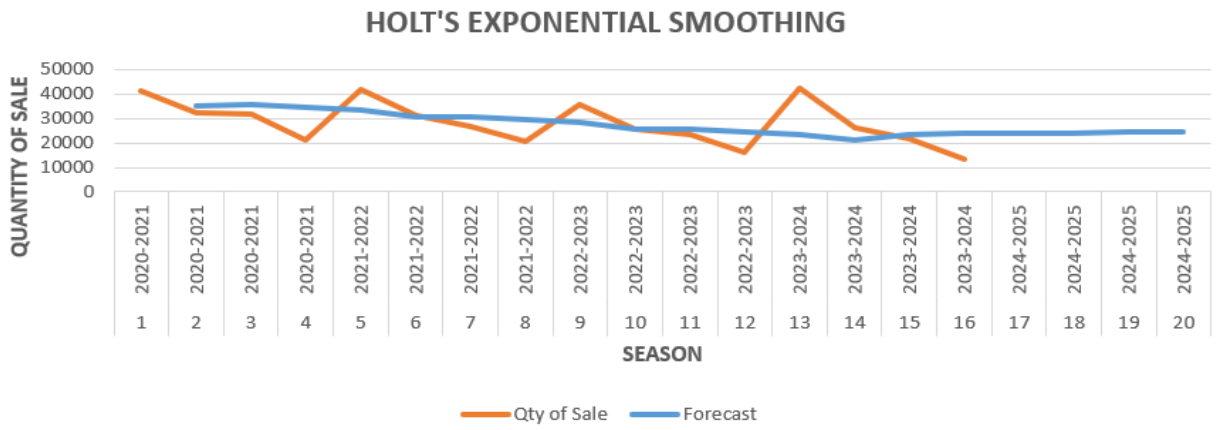


Figure14: Represent Holt's exponential smoothing.

The orange curve indicates the total sale of top 5 medicine categories with respect to season and the blue curve is the forecast curve with α=0.08, β=1

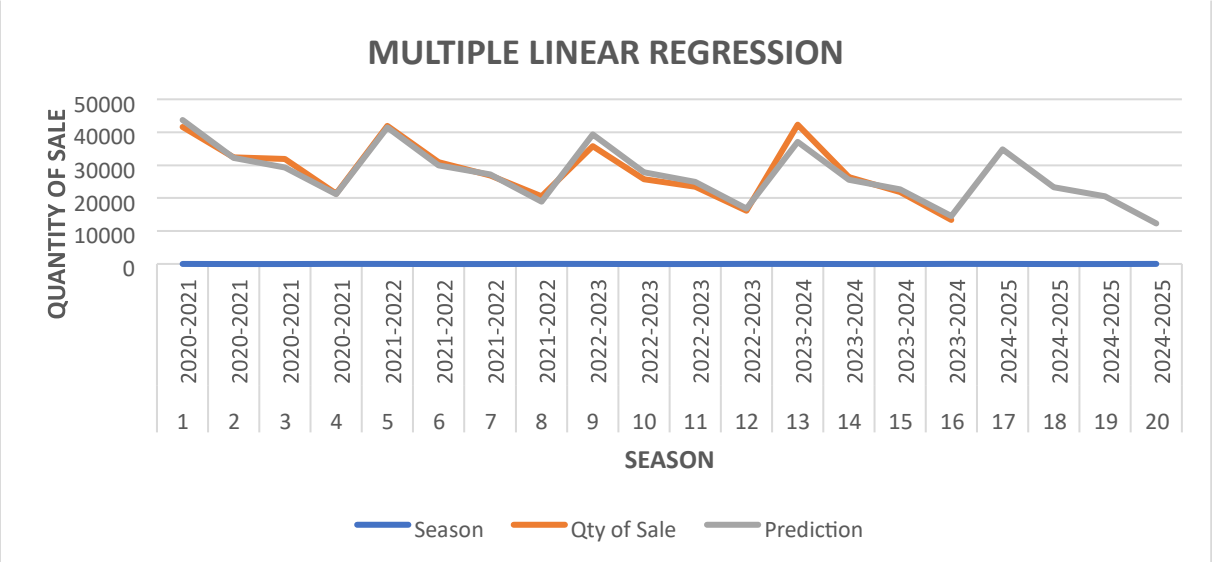


Figure 15: Represent Multiple linear regression.

The orange curve indicates the total sale of top 5 medicine categories with respect to season and the grey curve is the forecast curve.

The four year data is split into training and testing data. From 2020 to 2023 is treated as training data and the last year 2023-2024 is treated as testing data. The forecasting models above are applied in training and testing data and the forecasting techniques are evaluated based on their accuracy in forecasting actual demand. The optimum model to be chosen based on accuracy.

2.5.7.2 Measuring Forecast Accuracy

Thus, different measures of forecast accuracy like MAD (Mean Absolute Deviation), MSE (Mean Squared Error), RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) are calculated in order to choose the best model.

Forecasting is not 100% accurate of the actual value. Hence measuring the accuracy of each forecasting method is done by measuring MAD, MSE, RMSE, MAPE values. The lower the value means the better the forecast.

1) Mean Absolute Deviation (MAD) measures the average of absolute difference between the forecast and actual demand. Smaller the value of MAD better the forecast method accuracy.

$$MAD = \frac{\sum_{i=1}^{n} |Forecast_i - Actual_i|}{n}$$

Where n is the number of time period, Forecast$_i$ is the forecast demand for the specific time period i and Actual$_i$ is the actual demand for the specific time period i.

2) Mean Squared Error (MSE) is the average square of the absolute difference between the forecast and actual demand. The error with higher values will get highlighted in this measure.

$$MSE = \frac{\sum_{i=1}^{n}(Forecast_i - Actual_i)^2}{n}$$

3) Root Mean Squared Error (RMSE) is the square root of Mean squared error (MSE). RMSE is the measure of the spread of the residuals. Lower the value of RMSE better the forecast method.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Forecast_i - Actual_i)^2}{n}}$$

4) Maximum Absolute Percentage Error (MAPE) compares error in percentages. It gives indication of relative magnitude of forecast error in comparison to the actual forecast error.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Actual_i - Forecast_i}{Actual_i}\right|$$

2.5.8 Analysis for Problem Statement 2

A table with high value Customers (Group 1 and Group 2) and their phone numbers are given to the Pharmacy. So that Pharmacy could market their product according to their demand. For Group 3 and 4 the way to increase the sale is by increasing the discount percentage.

Group 1 and Group 2 are Loyal Customers who visit the Pharmacy more than 40 times in the period. The total number of Group 1 and Group 2 Customers are 50. Hence in order to improve satisfaction of the loyal customer, a new customer friendly approach is suggested. This includes calculating their demand for the top medicine they purchase and approaching them for home delivery of medicine when they have the demand. So, most of the Customers in Group 1 and 2 will accept the sale order.

For Demand forecasting of Group 1 and Group 2 Customer different forecasting methods such as *3 Months Moving Average, STL, Linear regression, Exponential Smoothing, Forecasting with Trend and Seasonality, Holt's Winter Exponential, Multiple Linear Regression* are carried out and the best model for each Customer is selected by the simulator.

In this process the Demand Forecast model is Customer specific since Sales Pattern changes for every customer. Hence, the optimum model to be identified for each Customer specifically. For example, the Customer named Mr. Arumugam's (Group1) demand is forecasted by all 7 methods mentioned above. The STL forecast values are taken as the demand as it gives less measuring error. Refer Section 4 for measuring errors comparison.

STL (Seasonal and Trend Decomposition Using Loess):
In this method time series data is decomposed into three components Trend, Seasonal and remainder.
Seasonal component = Actual Sale – Month Average
Trend = Linear trend for the actual sale.
Remainder= Actual sale – (Seasonal component + Trend)
Forecast = Seasonal component + Trend.

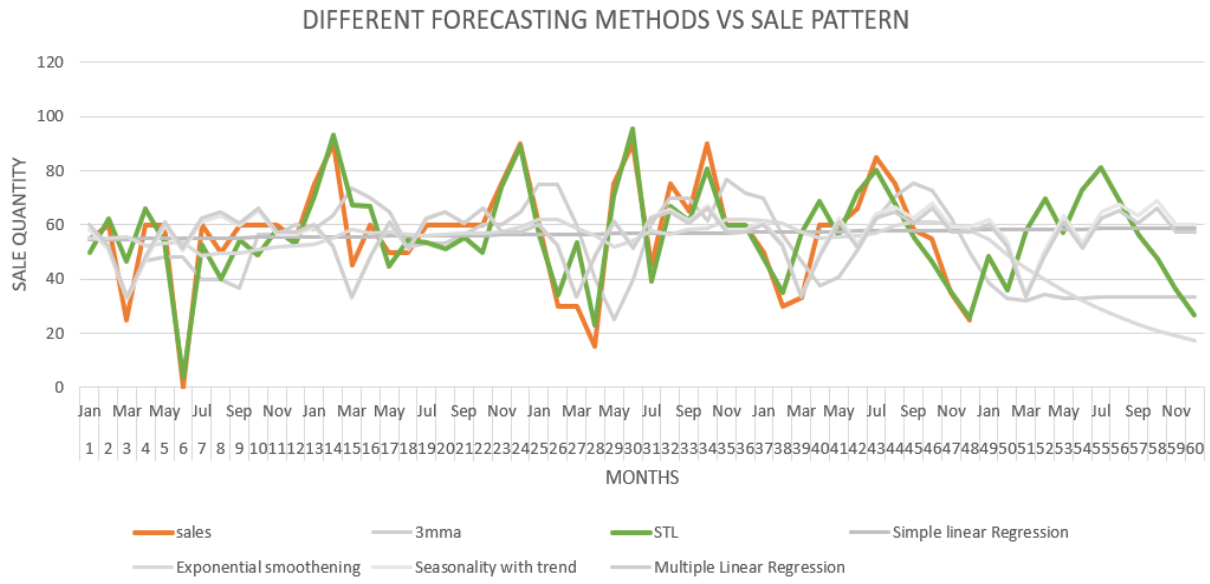


Figure 16: STL (Seasonal and Trend Decomposition using Loess) forecast method.

Figure 16 represents the results of Forecasting models for the Customer Mr. Arumugam's top category medicine demand. As per the graph, the STL Model performs better than the other models. The STL method is shown in green and the other models shown in grey. Customer's Purchase Pattern is shown in red.

Of all the methods, the best MAD value of 6.6 for training and 6.2 for testing is obtained and RMSE value of 8.6 for training and 8.6 is obtained. Since, MAPE calculation is sensitive to zero it is not calculated because individual demand may have zero values for a month or two. Hence the demand for the Loyalty Customer could be found through STL method and Customer Care could be improved for those Customers.

Group 3 and Group 4 are floating customers hence by increasing the discount percentage from 10% to 12% those floating customers could become regular customers.

3 Results and Findings

3.1 Result For Problem Statement 1 & 3

The Table 4 below depicts the measurement errors, obtained by using the eight forecasting techniques for the total sale of top five Medicine Category vs Season.

Out of these eight forecasting techniques, the Multiple Linear Regression model gives the best minimum score for measuring errors. Hence for forecasting the seasonal demand Multiple linear regression model is carried out to forecast the demand.

Technique Error		MAD	MSE	RMSE	MAPE
Moving Average	Training	8097	90521341	9514	34%
	Testing	11177	1.86E+08	13634	52%
Linear Forecast	Training	5237	41668962	6455	21%
	Testing	7430	1.01E+08	10063	29%
Forecasting with seasonality and trend	Training	2173	10187748	3192	**7%**
	Testing	3026	18104371	4255	**9%**
Time Series Decomposition	Training	1344	3343721	1829	**4%**
	Testing	3292	17423105	4174	**11%**
Exponential smoothing	Training	8534	91495752	9565	31%
	Testing	13023	2.16E+08	14695	51%
Holt's Exponential	Training	5418.892	45280551	3631.09	13%
	Testing	8895.826	1.21E+08	4885.65	15%
Holt's Winter Exponential	Training	3239.28	13184827	3631.09	13%
	Testing	3762.09	23869575	4885.65	15%
Multiple Linear Regression	Training	1294.192	2739294	1655.081	**5%**
	Testing	2024.5	7593948	2755.712	**7%**

Table 5: Measurement errors of forecasting models for seasonal demand

The MAPE value for the Time Series Decomposition, Forecasting with Seasonality and Trend and Multiple Linear Regression are less than 10 %. However, the MAPE for Multiple Linear Regression is the smallest among them. Hence, the Multiple Linear Regression method is recommended to forecast the seasonal demand.

3.2 Result For Problem Statement 2

To address the Problem statement, the Customer Category Table is prepared with Customer Names and Phone Numbers and this table categorises Customers into Group 1 and Group 2. The Pharmacy can prepare marketing strategies according to the Groups. Table 6 below shows the first 12 rows of the Customer Category table.

S.No	Group1 Customer Name	Phone Number	Group2 Customer Name	Phone Number
1	ARUMUGAM	XXXXXXXXXX	ABDUL NAZAR	XXXXXXXXXX
2	BALU SAMY	XXXXXXXXXX	BASEER	XXXXXXXXXX
3	BANUMATHI	XXXXXXXXXX	CHITHRA	XXXXXXXXXX
4	JOESPH	XXXXXXXXXX	CHOCKALAINGAM	XXXXXXXXXX
5	KAVITHA	XXXXXXXXXX	GANDHIMATHI	XXXXXXXXXX
6	LAKSHMANAN	XXXXXXXXXX	GNANAMBAL	XXXXXXXXXX
7	LAKSHMI	XXXXXXXXXX	JAILANI	XXXXXXXXXX
8	NATARAJAN	XXXXXXXXXX	JAYA KUMAR	XXXXXXXXXX
9	PALANIVEL	XXXXXXXXXX	JEYA KUMAR	XXXXXXXXXX
10	RADHA KRISHNAN	XXXXXXXXXX	JOHNSON	XXXXXXXXXX
11	RAJASEKARAN	XXXXXXXXXX	KADHAR MAIDHIN	XXXXXXXXXX
12	RAMANATHAN	XXXXXXXXXX	MARIAPPAN	XXXXXXXXXX

Table 6: Customer Category table

Demand forecasting for specific Customer is carried out to improve the satisfaction of the loyal Customers. For example, the Demand Forecast for Customer Mr. Arumuagm (S.No 1) is calculated for the period April 2024 to March 2025 using Seasonal and Trend Decomposition technique as the measuring error of the technique is minimum. Refer Table 7.

Technique Error		MAD	MSE	RMSE
Moving Average	Training	17.88	517.51	22.75
	Testing	19.28	408.63	20.21
STL (Seasonal and Trend Decomposition using Loess)	Training	**6.61**	**73.73**	**8.59**
	Testing	**6.22**	**74.74**	**8.65**
Simple Linear Regression	Training	13.88	378.95	19.47
	Testing	10.73	330.79	18.19
Exponential Smoothing	Training	14.78	404.54	20.11
	Testing	16.41	383.14	19.57
Forecasting with Seasonality and Trend	Training	11.94	301.92	17.38
	Testing	13.89	284.73	16.87
Holt's Winter Exponential	Training	43.33	2930.54	54.13
	Testing	35.42	1680.52	40.99
Multiple Linear Regression	Training	12.20	305.50	17.48
	Testing	56.41	270.72	16.45

Table 7: Measurement errors of forecasting model for individual customer demand

A MS Excel Simulator is given to the Pharmacy, to calculate Demand Forecast for the specific Customer for the upcoming months for the most purchased medicine category. This will facilitate the Pharmacy to meet the Customer needs in a timely manner. The Table 8 represents the Simulator Output for the Customer Mr. Arumugam's medicine demand for the year.

CUSTOMER NAME:		ARUMUGAM
PHONE NUMBER:		XXXXXXXXXX
YEAR	MONTH	DEMAND
2024-2025	Jan	48.01
2024-2025	Feb	35.58
2024-2025	Mar	57.90
2024-2025	Apr	69.47
2024-2025	May	57.04
2024-2025	Jun	72.86
2024-2025	Jul	80.93
2024-2025	Aug	68.50
2024-2025	Sep	55.82
2024-2025	Oct	47.38
2024-2025	Nov	36.20
2024-2025	Dec	26.27

Table 8: Simulation results of the individual customer demand

4 Interpretation of Results and Recommendation

The result shows that the Pharmacy had a total income of Rs. 97,93,253 in the four years period while receiving the maximum of 4% of discounts from the Dealers. If the Pharmacy places orders in high quantities, the Dealers offer more discounts and free medicine.

Currently the Pharmacy places order monthly or weekly or as required. But this project study recommends the Pharmacy to place order for the seasons (which is three months) instead of weekly or monthly. The Pharmacy can also estimate the quantity of medicines for every season using the recommended techniques in this study. This enables the Pharmacy to receive additional discounts from the Dealers. Even a small increment of 4% to 7% of Dealers discount would result in 75% of additional income for each year. The current income for a year is Rs. 24,48,313. When 7% discount is given the income would become Rs. 42,84,548 for a year.

The Pharmacy can avoid Stock out and Over stocking Situation if it forecasts the Demand and places orders accordingly. The Stock Out Situation would result in Customer Dissatisfaction and loss in Customer relationship. In case of Overstocked medicines expired, the Pharmacy loses 35% of the medicines cost on return to the Dealers.

In addition to the above benefit, if the medicines are ordered for seasons instead of month or week, the logistic costs also will be reduced.

As per the Figure 6, VASAVI MEDICARE and IRAGU ENTERPRISES are identified as the Dealers who provide more percentage of free medicines. If Pharmacy places orders with these Dealers, would benefit in terms of free medicines, and thus profit.

This study also discusses opportunities to improve Customer satisfaction. For Group 1 & 2 Customers, the Forecast Demand approach helps to optimize the stock availability and provide timely supply of medicines. The benefits and profit from the Demand Forecast approach enable the Pharmacy to provide discounts for Group 3 & 4 Customers to attract them.

The following are additionally recommended to the Pharmacy may be helpful to them to develop business further and to improve Customer relationship.

- Advertisement of discounts and offers
- Online Sale of medicines and Free home delivery (ePharmacy)
- Ordering through WhatsApp

--End of Report--