CAPSTONE PROJECT – **WALMART**

Table of Contents

1. Problem Statement
2. Project Objective
3. Data Description
4. Data Pre-processing Steps and Inspiration
5. Choosing the Algorithm for the Project
6. Motivation and Reasons For Choosing the Algorithm
7. Assumptions
8. Model Evaluation and Techniques
9. Inferences from the Same
10. Future Possibilities of the Project
11. Conclusion
12. References

Problem Statement

A retail store that has multiple outlets across the country are facing issues in managing the inventory - to match the demand with respect to supply. You are a data scientist, who has to come up with useful insights using the data and make prediction models to forecast the sales for 3 months/years.

Project Objective

1. To search for patterns and seasonality in the data.
2. To look for factors affecting Weekly sales.
3. Forecasting future sales.

Data Description

* The data consists of 7 columns. Each carrying their own characteristics.
* All the columns have some or the other effect on Weekly Sales.
* Following are the features in the dataset:

1. Store
2. Date
3. Weekly Sales
4. Holiday
5. Temperature
6. Fuel Price
7. CPI
8. Unemployment

* Data Visualization and Insights:

1. Stores which are affected by unemployment –

Store no. 4,9,13,16,17,23,26,34,37,39,40,41,42,44

1. Date v/s Weekly Sales:

Shows a seasonal pattern due to the effect of Christmas Eve and New year. More number of people go out for shopping in the month of November and December on occasion of Christmas and New Year.

1. Temperature v/s Weekly Sales:

Temperature between 40 and 80 Sales is not affected but above 80 and below 40 has some effect. As Temperature is increasing the sales are decreasing and if temperature decreases below

1. Consumer Price Index v/s Weekly Sales:
2. Store 1,6,8,9,16,19,20,26,29,31,32,35,37,38,40,43:

These Stores are constantly moving with ups and downs. But are neither giving an increasing nor decreasing pattern.

1. Store 2,3,4,5,7,13,17,34,39,41,42,44:

With increase in CPI, Weekly Sales are also increasing. But some stores and showing regular increase and decrease in Weekly Sales in upward direction.

1. Store 10,11,12,14,15,21,27,36,45:

Showing decreasing pattern with increase in CPI.

1. Store 18:

Small amount of fall between CPI value 131 and 133 but after that the graph is moving constantly.

1. Store 22, 23:

Steep jumps are seen and steep falls are also present.

1. Store 25:

It showed significant increase between CPI value 211 and 212 but went down steeply after it.

1. Store 28:

Going through tough times after CPI value went above 130.7.

1. Store 30:

Till 216 it was in decreasing pattern but then showed increase in Sales.

1. Top 10 performing stores

|  |  |
| --- | --- |
| Store no. | Weekly Sales |
| 1. Store 20 | 3.013978e+08 |
| 1. Store 4 | 2.995440e+08 |
| 1. Store 14 | 2.889999e+08 |
| 1. Store 13 | 2.865177e+08 |
| 1. Store 2 | 2.753824e+08 |
| 1. Store 10 | 2.716177e+08 |
| 1. Store 27 | 2.538559e+08 |
| 1. Store 6 | 2.237561e+08 |
| 1. Store 1 | 2.224028e+08 |
| 1. Store 39 | 2.074455e+08 |

1. Worst performing Stores:

Store 33 with weekly sales of 37160221.96 is the Worst performing store.

1. Holiday v/s Weekly Sales:

On holiday’s the sales are significantly higher.

1. Month v/s Weekly Sales:

From last 3 years data we can say that the company has generated more revenue in the month of April.

1. Date v/s Temperature

From the graph Temperature hasn't drastically changed. It is fluctuating as per season. Temperature shouldn’t affect much in weekly sales.

1. Date v/s CPI

From the graph CPI is gradually increasing. It can affect future sales. It is better if we can negotiate with our suppliers due to rise in CPI. As rising CPI means rise in inflation.

1. Date v/s Unemployment

From the graph we can say that Unemployment rate is falling. This can be advantageous for us as people may come to the store and buy more items due to change in their lifestyle.

Data Processing Steps:

1. Firstly we check data type of all the columns.
2. We found that all columns have data types as per their characteristic except for ‘Date’ column. The data type for ‘Date’ column is ‘object’.
3. Changing the data type of ‘Date’ column to ‘datetime’.
4. Adding columns ‘Year’, ‘Month’ and ‘Week’ for further cases.
5. To pass df.describe command to know mean, standard deviation, etc. of all the columns.
6. Checking for null values. But luckily there are no null values present in this dataset.
7. Check for outliers, but in this dataset we cannot remove the outliers as they are giving us some valuable information.

Time Series Analysis

Time series is set of data points over a given period of time. Where time is an independent variable and the goal is to Forecast for the Future. It is just not about Forecasting but also to analyse Trend, seasonality and variability.

Time Series can be used in the following sectors:

* Rainfall measurements
* Sales Forecasting
* Temperature Forecasting
* Stock trading

**Prophet**

I chose Prophet method for forecasting and the reasons are as shown below:

1. Prophet model possesses less error.
2. Prophet can handle non-linear and non-stationary trends.
3. It is able operate with complex and multiple seasonality, and holiday effects.
4. Prophet models are robust to outliers and missing values.
5. It can be customized as per user’s preferences.
6. It has informative tables and visualization tools such as uncertainty intervals, cross-validation, and performance metrics.

Assumptions

1. The model assumes that the time series data has seasonality, which means the patterns occur periodically over time.
2. It assumes that the data is stationary, which means statistical properties of the data do not change over time.
3. It further assumes that trend, seasonality and other components can be added to get the overall forecast.
4. It assumes that the trend component of the time series data follows piecewise linear function, which means that the trend can change at different points in time.

Model Evaluation and Technique

1. Only One value per week will be considered. In case if a week has two values then second will be kept and first will be deleted.
2. Only keep Date column and Weekly Sales column. And rename Date column as ‘ds’ and Weekly Sales column as ‘y’.
3. Import Prophet library.
4. Call an object and fit the model.
5. Use ‘make\_future\_dataframe’ function and enter “period = 12, freq = ‘W’”. Here ‘period’ means how many forecasts do we need and ‘freq’ means whether it is daily, weekly, monthly or yearly data.
6. Then assign predict function to the data.
7. Extract ‘yhat’ column from forecasted data.
8. For validation we’ll check whether the model has predicted well or not and to do that we’ll need ‘Mean Absolute Percentage Error(MAPE)’.
9. We’ll import MAPE library from sklearn.
10. After importing we’ll pass both the data frames having the same date into the MAPE function.

Evaluation Report:

* Here in my case, the MAPE value is ranging between 0.02-0.08. Reminder, I’m getting a range because I have evaluated each store.

Inferences from Project

* As you can see in my python file the error shown by the model for each store is very low.
* The model has been trained really well.
* Here I didn’t use ARIMA or SARIMA model for evaluation as their computation time is greater and there were outliers in the data.

Future Possibilities

* This model can be used for forecasting.
* It can help us strategize our sales.
* We can manage our inventory based on forecasting.
* Visualisation and forecasting can help us plan our marketing effectively.
* To be aware of how external factors like unemployment, temperature, inflation, etc. can affect our sales. And can make us aware about future issues.
* It is necessary to come up with a strategy for stores whose sales are gradually decreasing.
* Stores which are going through tough times should adopt market basket analysis based on type of population and their past data.

Conclusion

This project has helped us give a broader perspective about the market and the market condition. There are stores which require more attention due to their underperformance. We can conduct a survey in the locality where these stores are situated. And from that we can analyse the true problem.

At the same time it becomes necessary for us to covey these problems to government officials if the market conditions are severe.

References

* <https://machinelearningmastery.com/time-series-forecasting-with-prophet-in-python/>
* <https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/>
* <https://towardsdatascience.com/getting-started-predicting-time-series-data-with-facebook-prophet-c74ad3040525>