Capstone Project - Project Walmart

By Angelo Alexander

Email: [angelolife2002@gmail.com](mailto:angelolife2002@gmail.com)

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

# 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.

# Project Objective

The objective of this project is to predict the expected weekly sales for the next 12 weeks to manage inventory according the previous store sales data.

# Data Description

The dataset available is Intellipaat’s website.

Dataset Information:

The walmart.csv contains 6435 rows and 8 columns.

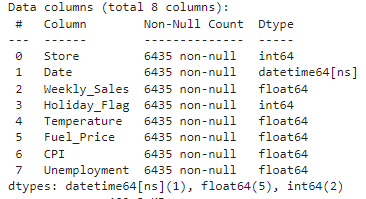
|  |  |
| --- | --- |
| Feature Name | Description |
| Store | Store number |
| Date | Week of Sales |
| Weekly Sales | Sales for the given store in that week |
| Holiday Flag | If it is a holiday week |
| Temperature | Temperature on the day of the sale |
| Fuel Price | Cost of the fuel in the region |
| CPI | Consumer Price Index |
| Unemployment | Unemployment Rate |

# Data Preprocessing

Steps And Inspiration

The preprocessing of the data included the following

1. The data has no null values, from the 6435 Rows, and 8 columns.
2. The date type is in object type; hence we make it to Data Time type using pd.to\_Datetime.
3. The date format is in American format with YYYY/MM/DD
4. The dataset has redundant dates, because each store’s weekly sales have its own data which makes the dataset big.
5. Since we have to make prediction for each and every store, we make them into a separate data frame before we make prediction (we do this in the last column which is in a for loop).
6. The dataset has outliers which are necessary for prediction so we do not remove them.

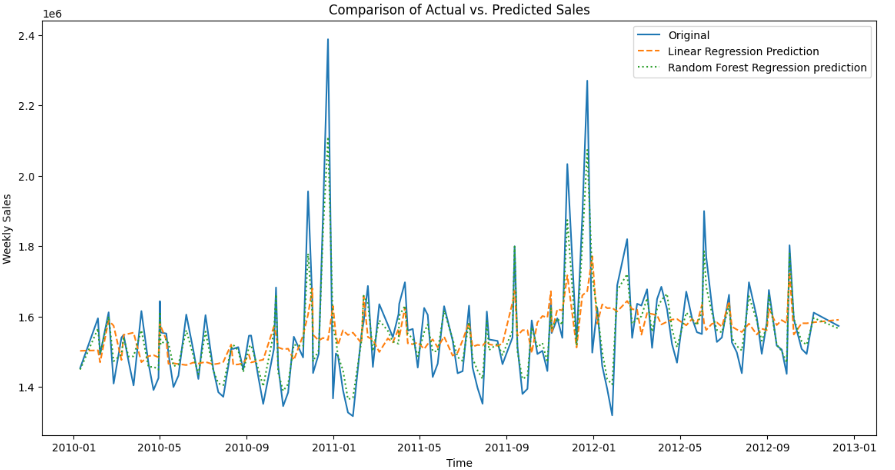


# Choosing the Algorithm for the Project

Description for the FBProphet algorithm for the project.

I have chosen the FBProphet algorithm for this project for the following reasons:

1. The dataset we have has multiple columns, so to forecast for the future we need the other columns for it to predict the data.
2. I have tried to do it with Linear Regression
3. I have also tried to do it with Random Forest Regression
4. And have compared the difference between the two.
5. I have introduced lag 1 and lag 2 to the dataset.
6. Which have significantly improved the predictive capabilities of the model.
7. But for predicting the future without the other columns to support we use the FBProphet library.
8. FBProphet is a powerful time series forecasting algorithm that can capture complex patterns in the data such as seasonality, trends, and the effect of holidays.
9. Which uses the Bayesian approach that allows for uncertainty estimation in the predictions



# Assumptions

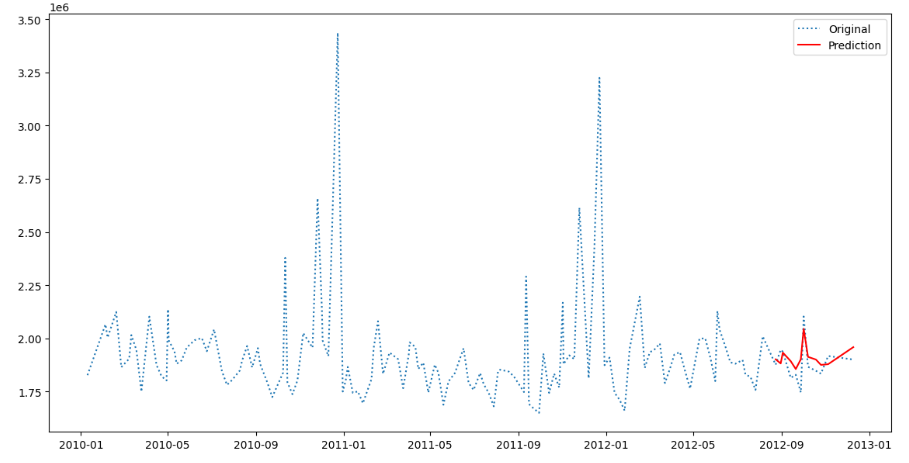
The following assumptions were made in order to create the model for Walmart project.

1. I have done Educated Assumption from the correlation plot.
2. From which we see that the other columns don’t significantly correlate to the weekly sales.
3. Thus, for making prediction we ignore them at the end.

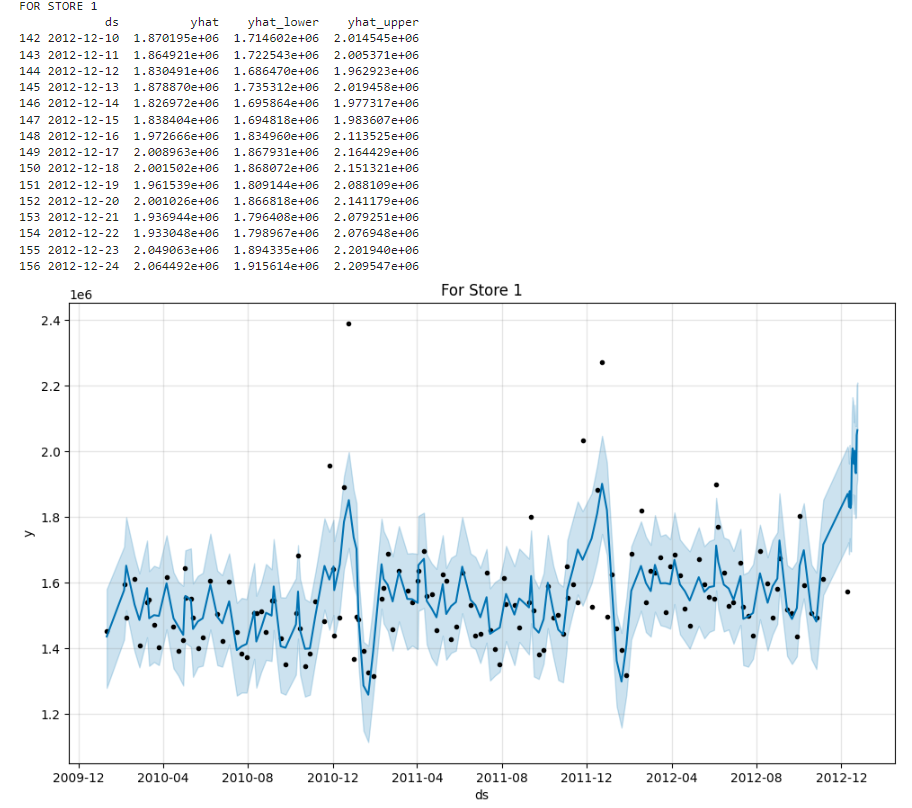
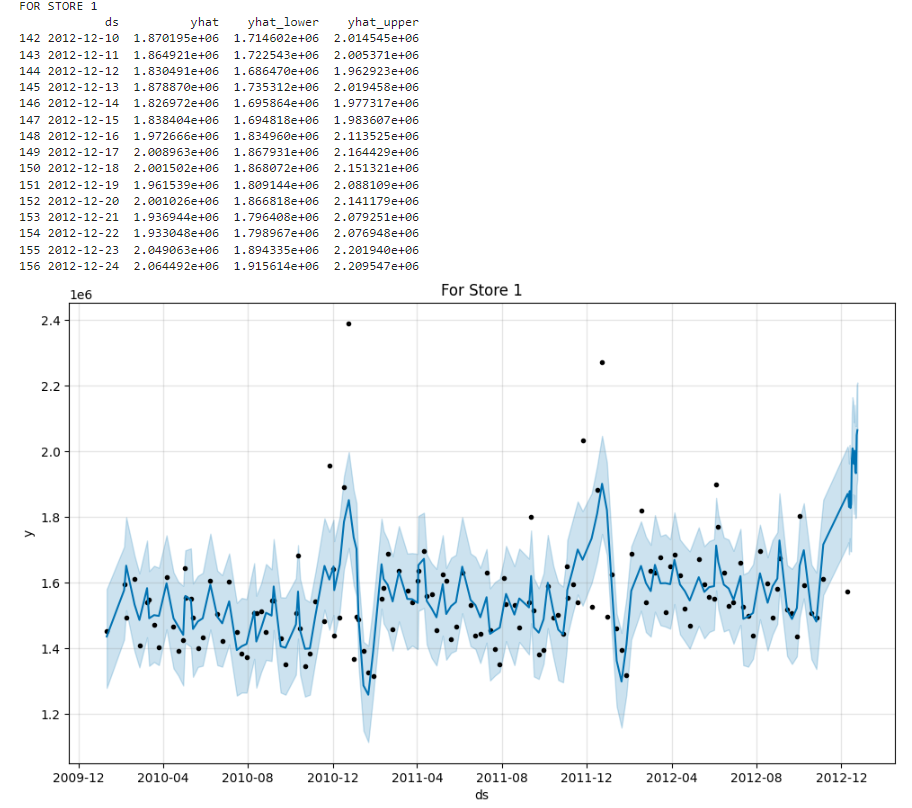
Model Evaluation and Technique

The following techniques and steps were involved in the evaluation of the model

1. The model’s prediction from Linear Regression and Random Forest Regression has been place on top of each other to see how good they predict the sales.



## The evaluation report suggests the following:



1. The sales increases till the New year and drop after 2 weeks which then improves till average sales.
2. The Sales drop during March for some stores.
3. The Sales is at peak at Christmas and New year holidays.
4. The sales drop after New year Holidays.

# Inferences from the Project

1. The Inventory has needs to be more during the holidays.
2. For some Stores the Inventory can be reduced during March and April.
3. For Most Stores the inventory can be reduced after the New year.

# Future Possibilities

1. The future possibilities are to add more columns for making the prediction model.
2. The prediction has a big error range which can be further reduced.

# Conclusion

1. I have learnt about Time Series models and the working of it through this project.
2. Practical doing the project has helped me in solving errors that arise while actually doing.

# References

1. <https://www.geeksforgeeks.org/>
2. <https://facebook.github.io/prophet/docs/quick_start.html#python-api>
3. <https://pandas.pydata.org/docs/reference/frame.html>
4. <https://www.kaggle.com/learn/time-series>