

Capstone Project:

Machine Learning

Using Python

Submitted by

Kannu Priya

Project 1

a) Problem Statement:

The retail store with multiple outlets across the country is facing inventory management issues in matching the demand with supply. The goal is to provide useful insights and build prediction models to forecast sales for the next 12 weeks.

b) Project Objective:

The objective is to analyze the provided dataset and derive insights to help each store improve in various areas. Additionally, the project aims to forecast the sales for each store for the next 12 weeks.

c) Data Description:

The dataset, 'walmart.csv', contains 6435 rows and 8 columns. The features include:

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

There are 45 stores in total and each store has 143 records.

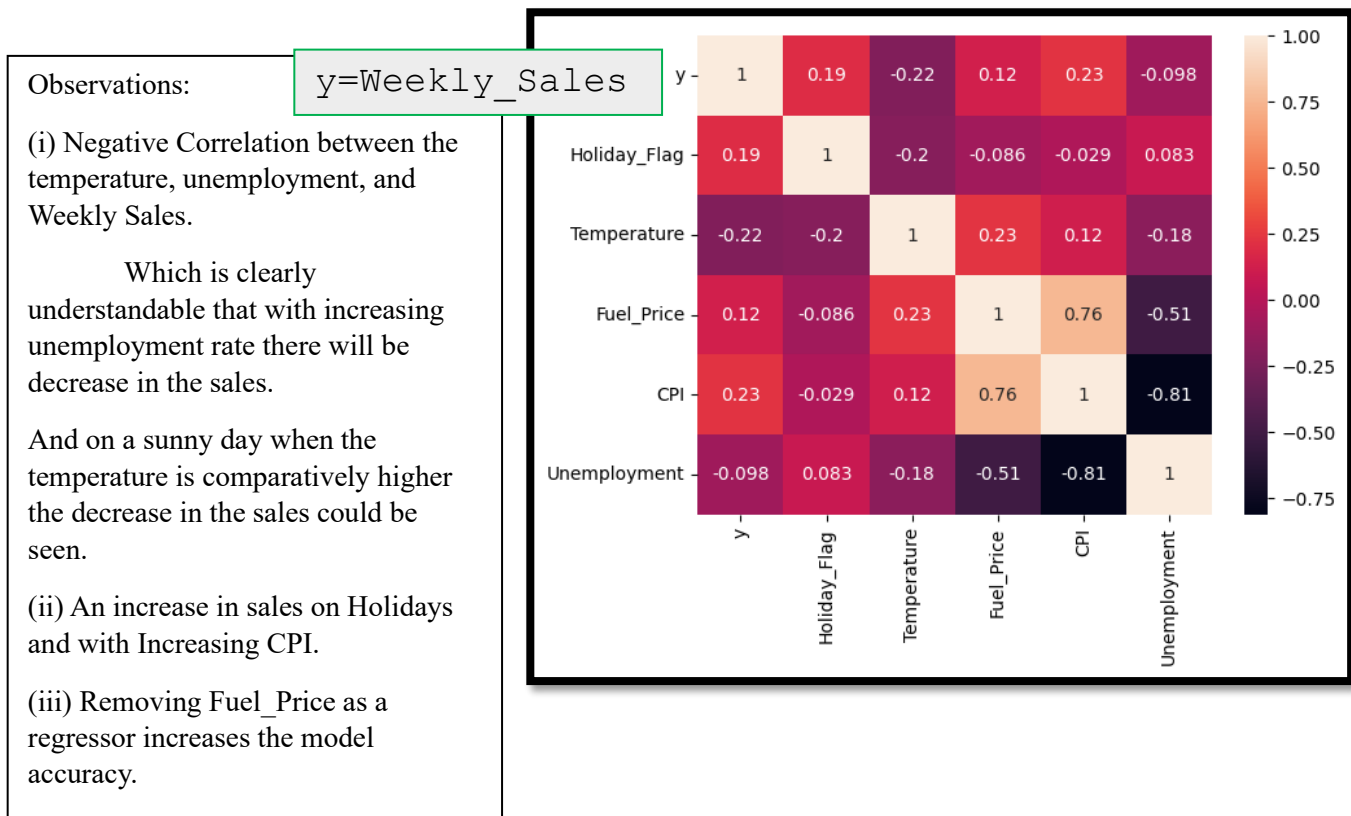
No Null values and No Duplicates were found.

d) Data Pre-processing Steps and Inspiration:

Data pre-processing steps may include handling duplicates, handling null values, data type conversion, and feature engineering. Inspiration can be drawn from the relationships between variables, identifying patterns, and exploring the impact of different factors on sales.

Steps:

- (i) After loading the dataset, converted the data type of 'Date' column to 'datetime' from 'Object'.
- (ii) No Null values and No Duplicates were found.
- (iii) Checking for correlation between the features.



e) Choosing the Algorithm for the Project:

The chosen algorithm for this project is the Prophet model, a time series forecasting algorithm developed by Facebook's Core Data Science team.

f) Motivation and Reasons for Choosing the Algorithm:

Prophet is suitable for time series forecasting tasks as it can handle trends, seasonality, and other complex patterns. It provides an intuitive interface and automated functionality for modeling time series data, making it an effective choice for sales forecasting in this project.

Adding the Regressors to the model help in making the model more wholesome and takes the effect of other features on the target as well.

```
prophet_model = Prophet()
prophet_model.add_regressor("Holiday_Flag")
prophet_model.add_regressor("Temperature")
prophet_model.add_regressor("CPI")
prophet_model.add_regressor("Unemployment")
prophet_model.fit(train_data)
```

g) Assumptions:

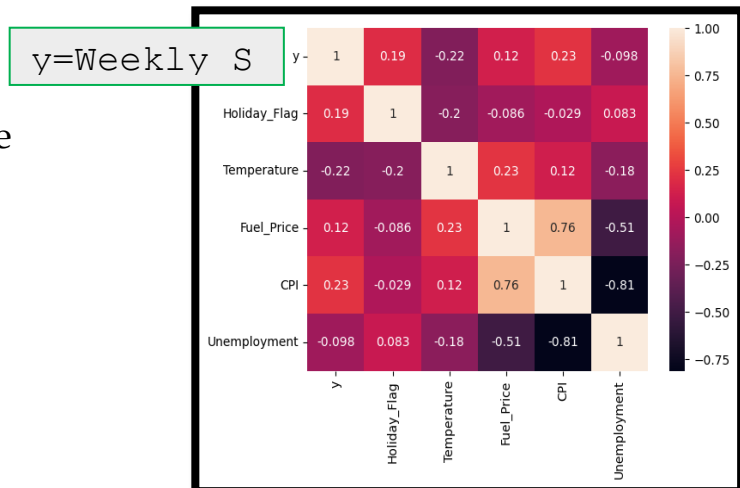
Assumptions:

i) Assuming the historical sales patterns will continue in the future.

I used this assumption to generate some random data to test the model.

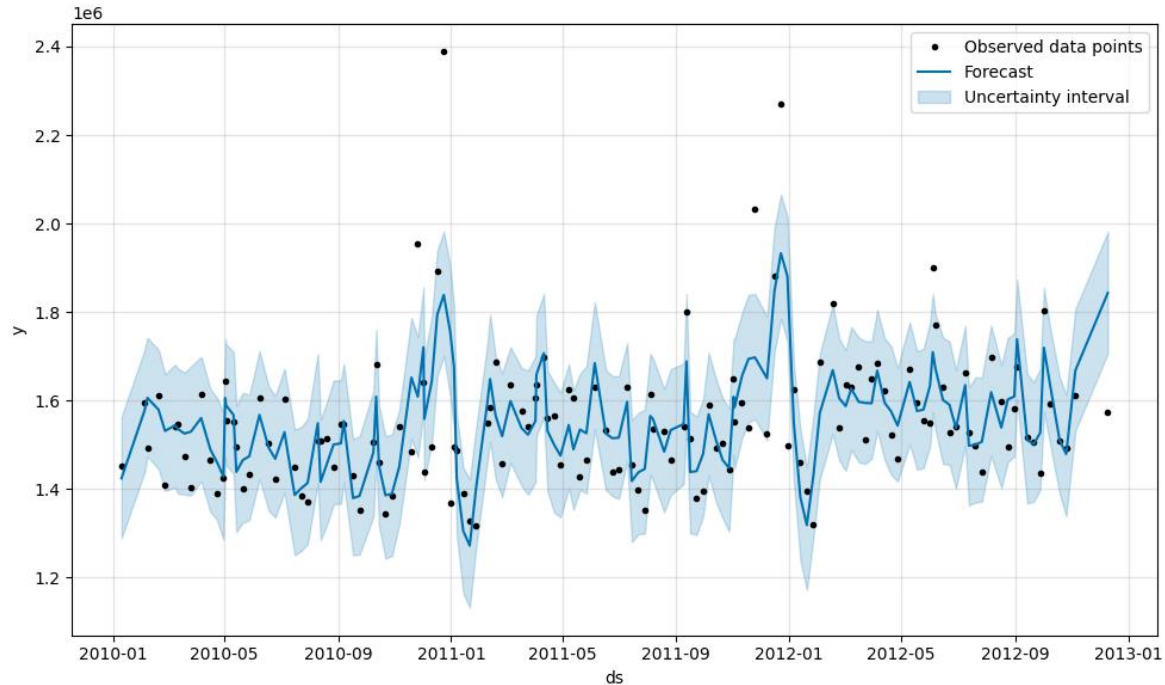
ii) Assuming that the provided data is representative of the overall sales behavior.

iii) External factors such as holidays and temperature have an impact on sales, which can be seen in the correlation matrix.



h) Model Evaluation and Techniques:

Cross-validation techniques and visual inspection of the forecasted sales were used for evaluation.



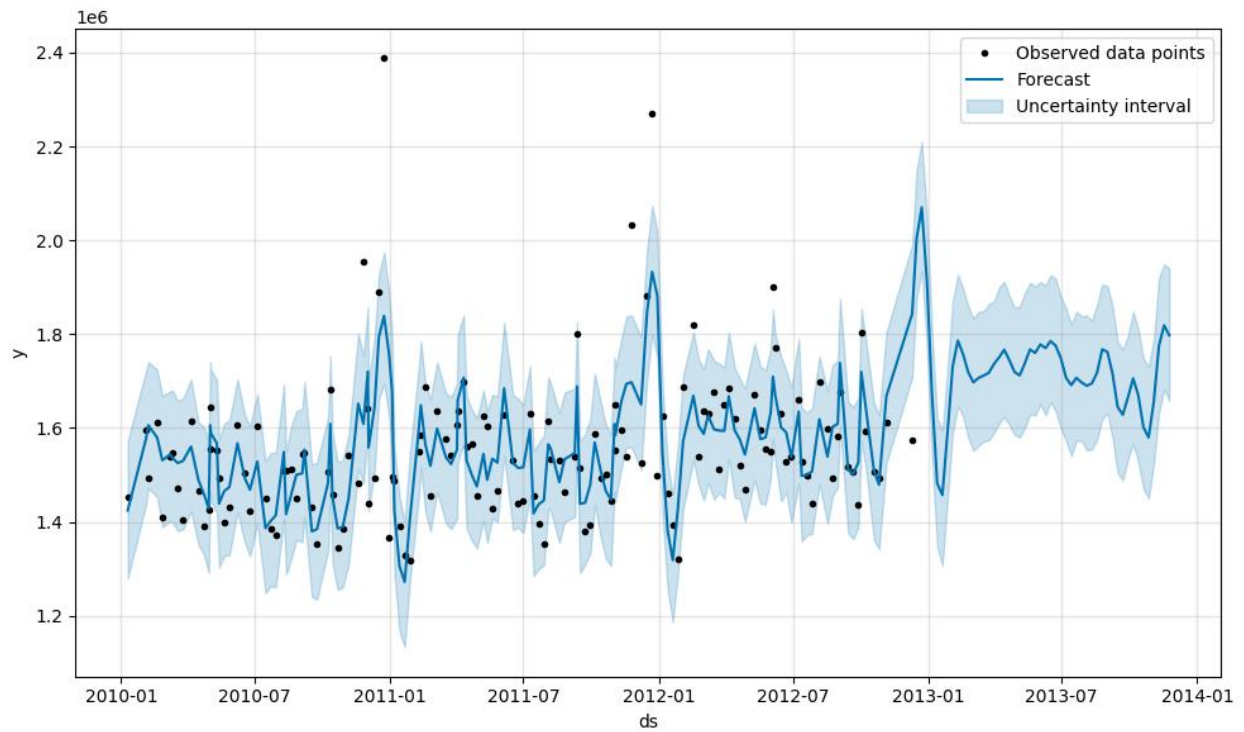
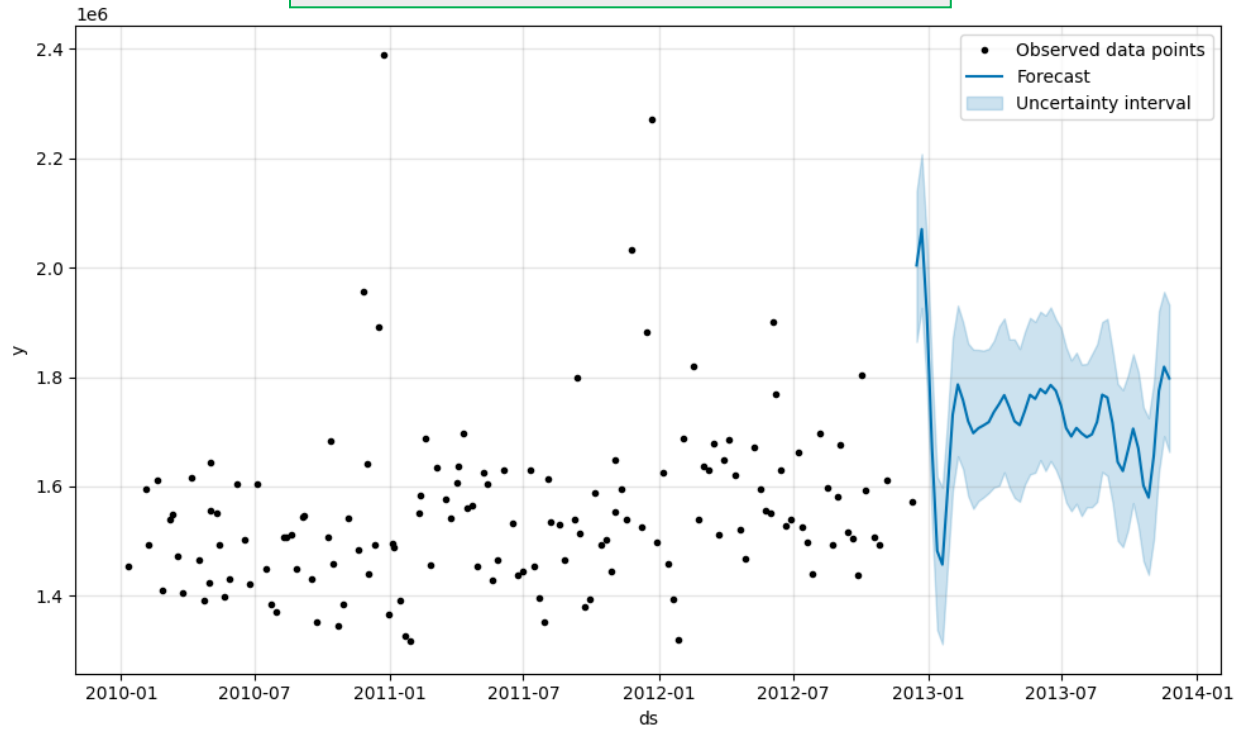
i) Inferences from the Model:

Inferences from the Prophet model can include insights on sales trends, seasonality patterns, the impact of holidays, temperature, and other factors on sales. These insights can help the retail store identify areas for improvement and make informed decisions regarding inventory management.

(i) Yearly seasonality pattern is identified. Sales tend to increase from end of September and increases till January. Then the sales reduce till February ending or March starting.

(ii) Impact of features like 'Unemployment,' 'High Temperature' can be seen on the sales.

Generated 50 Random Values of
all the features to test the
model.



j) Future Possibilities of the Project:

Future possibilities of the project may include expanding the forecasting horizon beyond 12 weeks, incorporating additional external factors that influence sales, implementing advanced forecasting techniques, and integrating the forecasted sales with supply chain management systems for efficient inventory management. Having more data for testing would be more beneficial.