

# Retail Sales Forecasting



By  
Shreya Mukerjee

# Index

---

Introduction

---

Data Collection and Usage

---

Exploratory Data Analysis

---

Prediction Model - FbProphet

---

Result and Recommendation

---

References

# Introduction

---

In the retail industry, forecasting is essential for retailers to stay competitive and optimize their pricing strategies, inventory management, and other key business decisions.

---

The project develops a forecasting model for retail sales that addresses the unique challenges of the industry, such as limited historical data, seasonal fluctuations, and the impact of markdowns on sales

---

Our model can be customized to meet the specific needs of individual retailers, taking into account factors such as store size, seasonality, weekly sales, pricing, promotions, and inventory.

# Data

The dataset was collected from Kaggle for 45 stores located in different regions - each store contains multiple departments. Here is the link:

<https://www.kaggle.com/datasets/manjeetsingh/retaildataset?select=sales+data-set.csv>

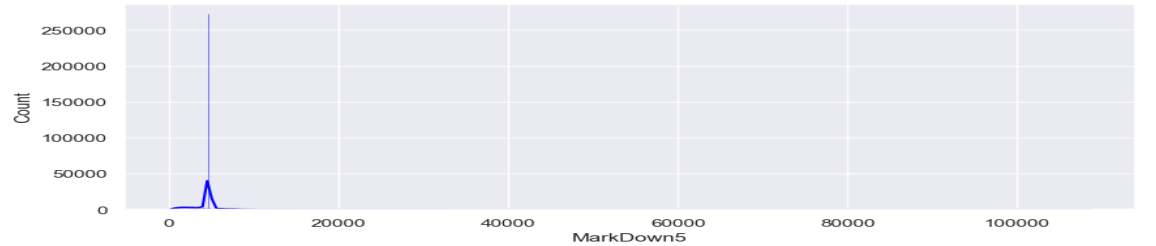
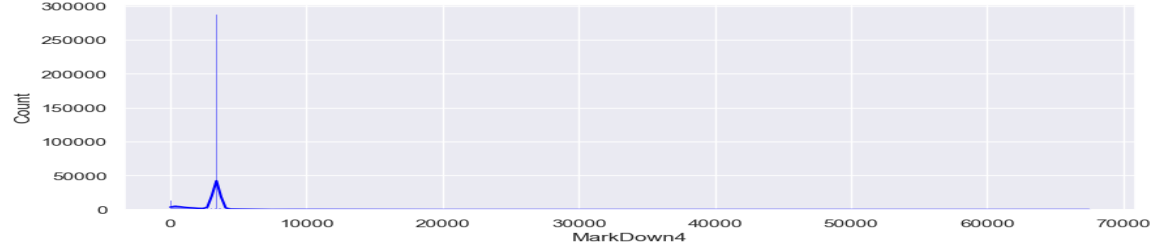
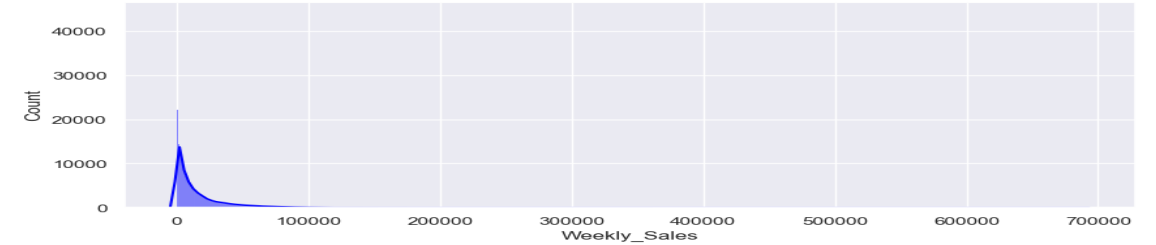
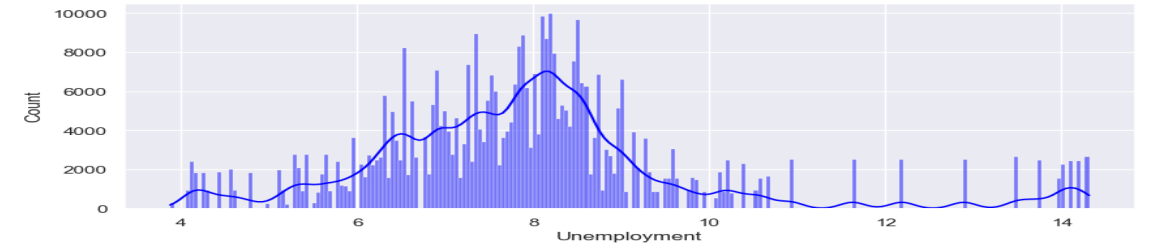
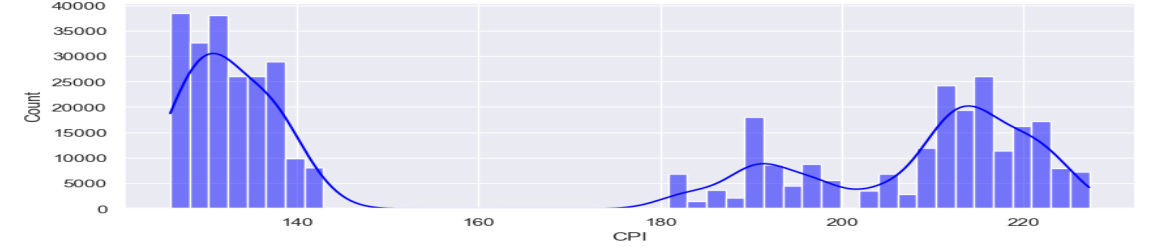
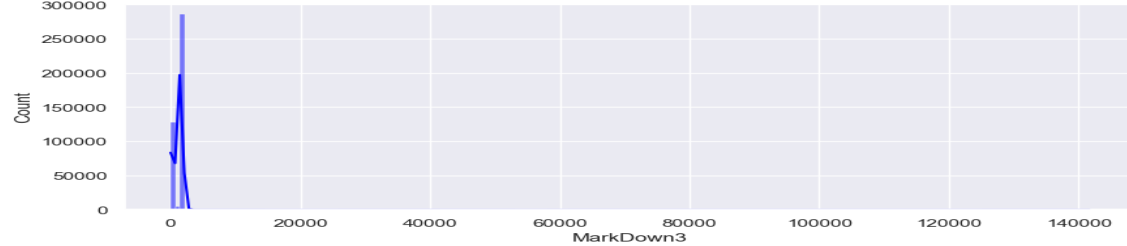
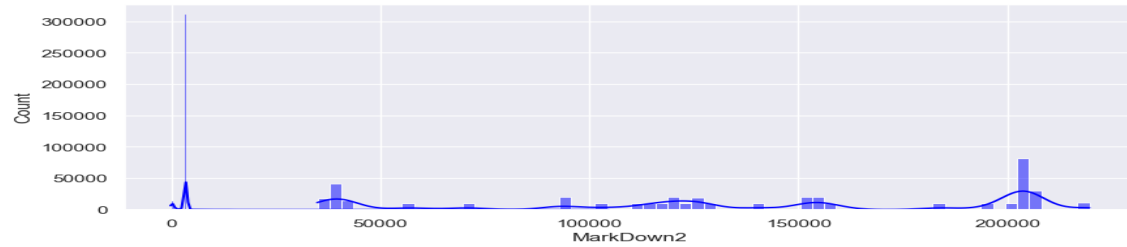
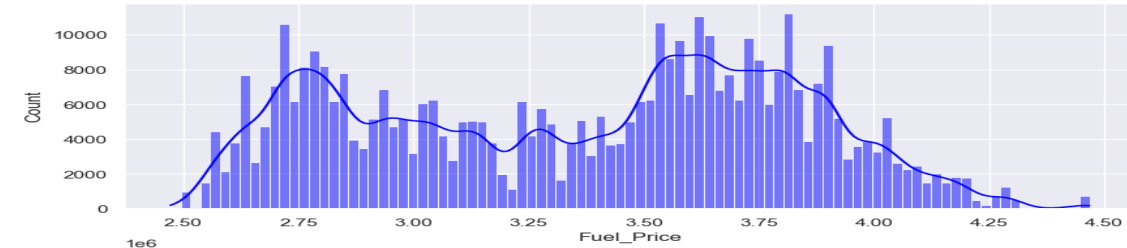
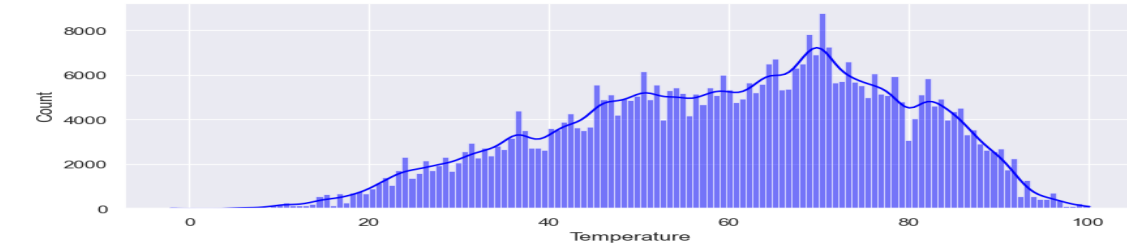
The dataset is divided into three parts where :

- Stores- Anonymized information about the 45 stores, indicating the type and size of store
- Features- This dataset has columns representing Store number, date, temperature (of that day) Markdown (amount of discount or price reduction offered on certain products)
- Sales – Store, Date, Department, Weekly\_sales, Is\_holiday

On merging the datasets there were around 421570 rows ranging from 2010-02-05 to 2012-11-01.

Weekly\_sales is our dependent variable whereas the rest of our variables are our independent variables

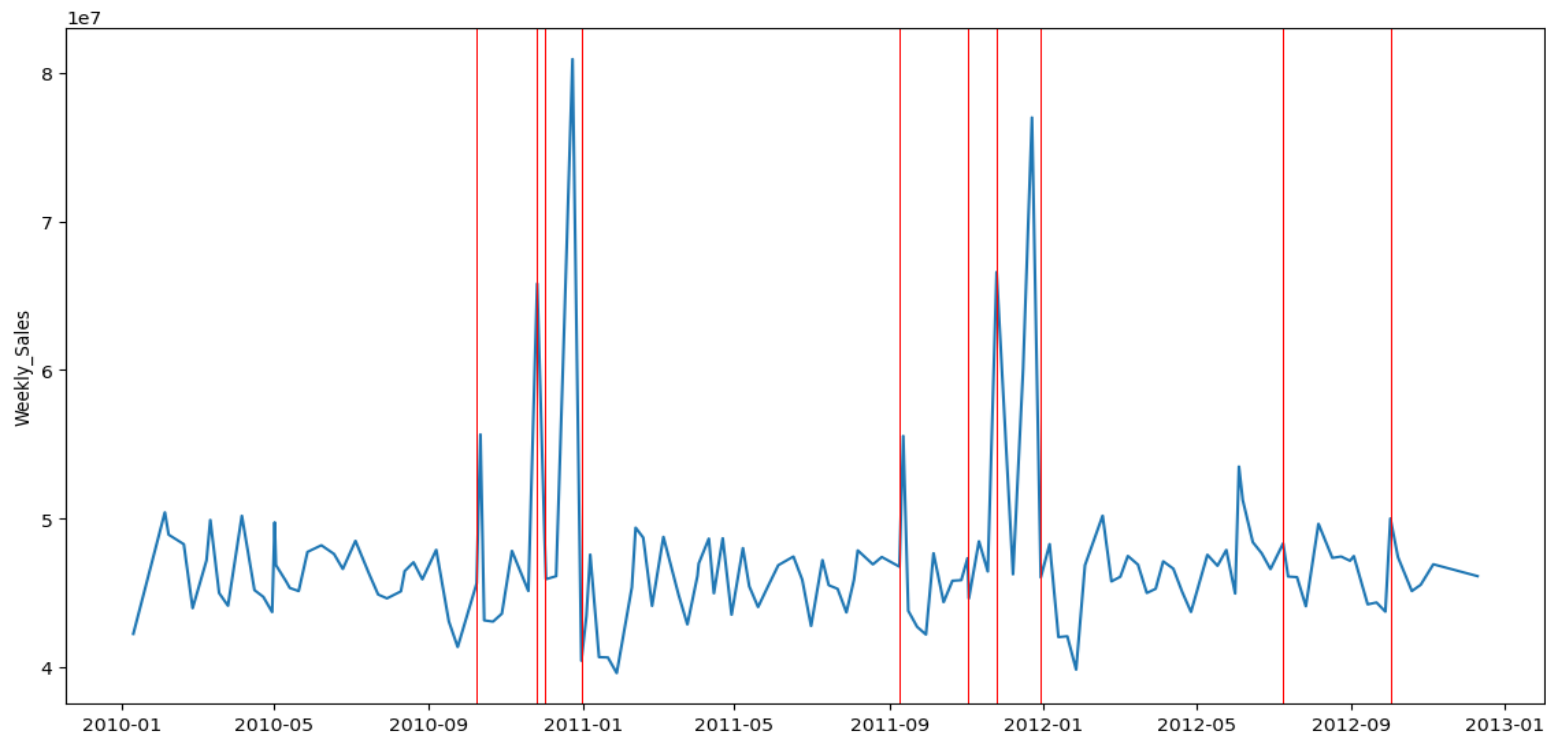
# EDA- Univariate Analysis



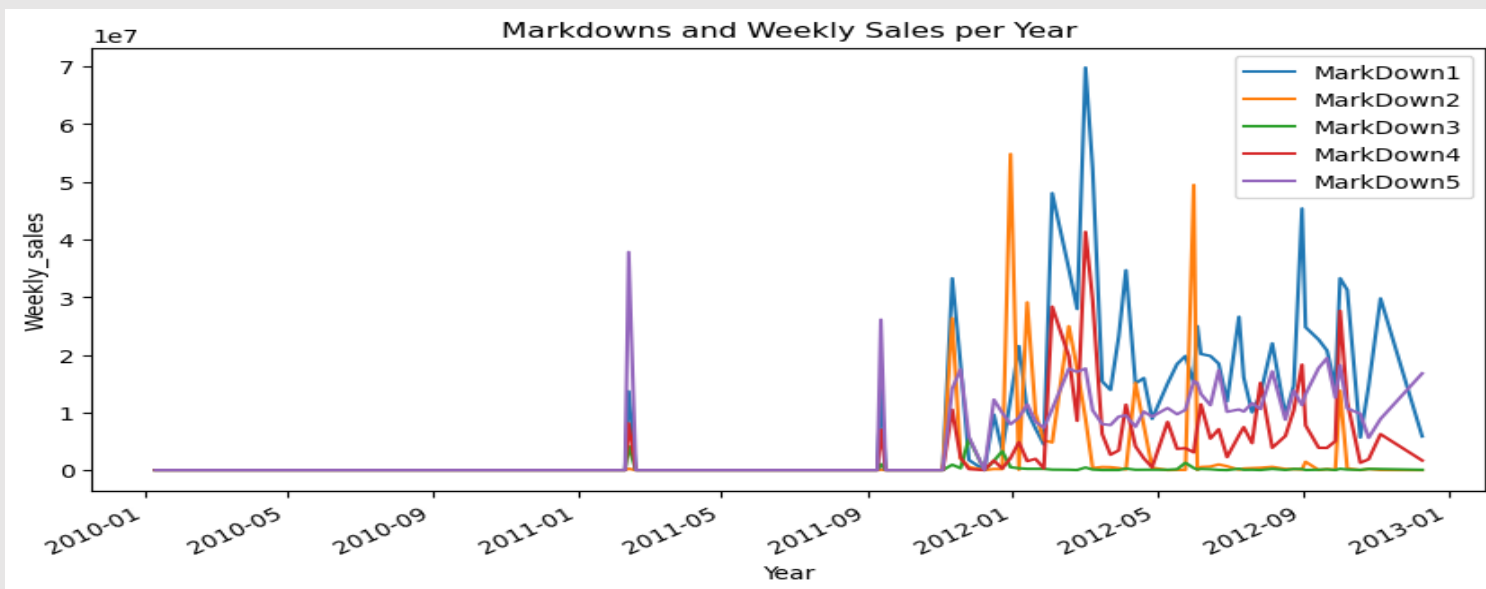
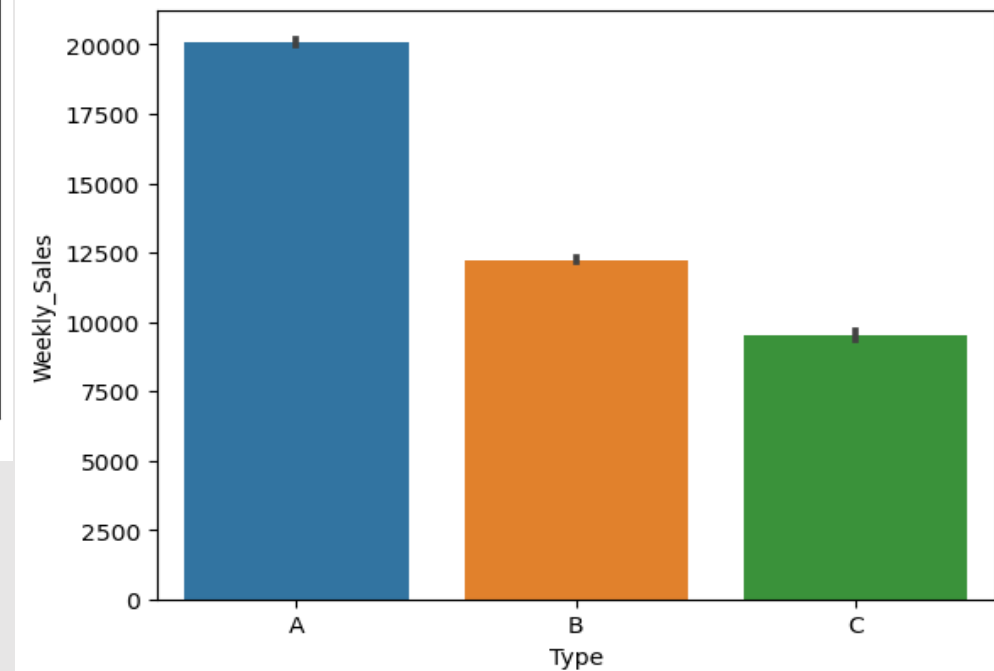
# Insights from univariate analysis

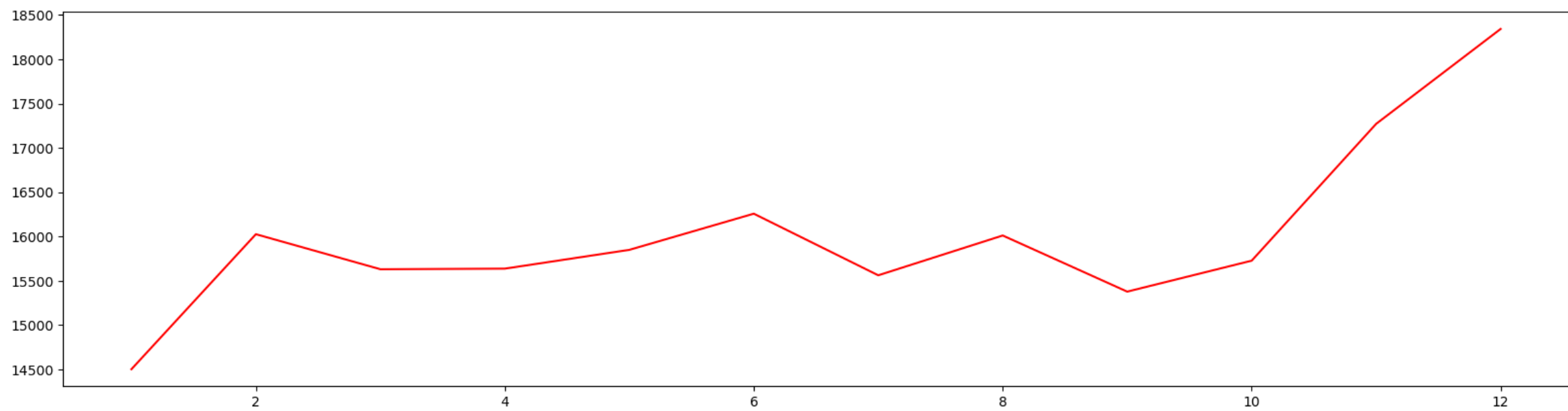
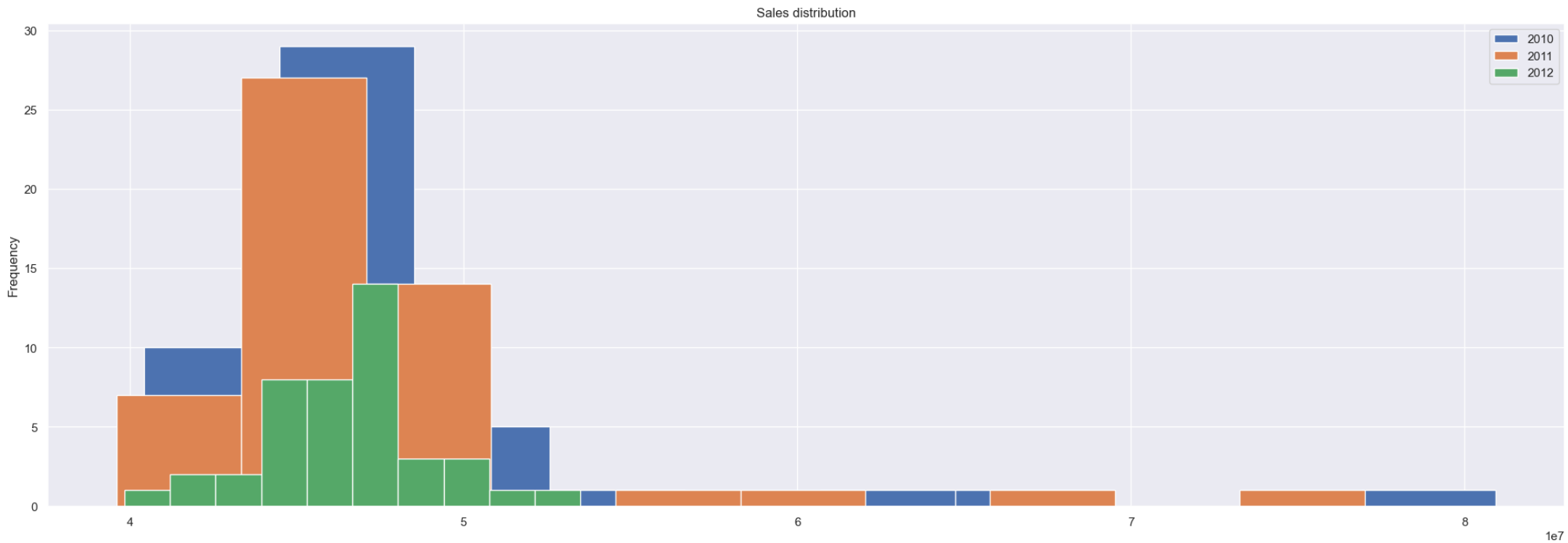
The Consumer Price Index (CPI) measures the monthly change in prices paid by U.S. consumers. We can see 3 peaks in its distribution which hints that there might be seasonality in the data.

We can again see peaks and troughs in fuel prices which depicts that prices go high at a certain 'season' than others.



## EDA – Bivariate Analysis







# Insights from bivariate analysis

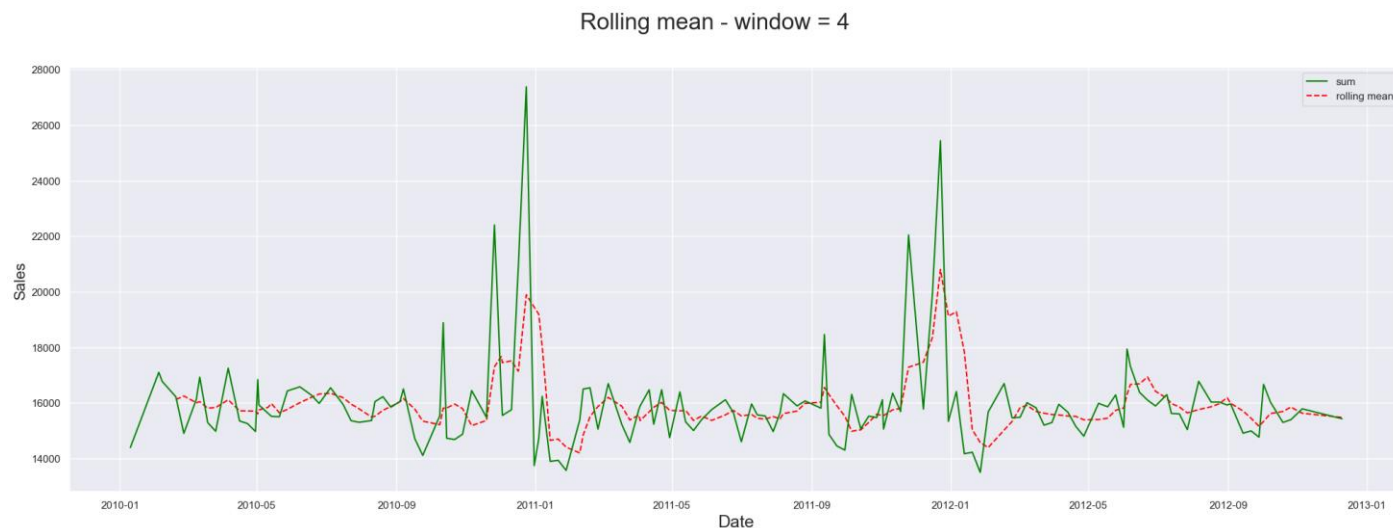
In the Year v/s weekly\_sales diagram, Holidays are marked by red vertical lines. We can clearly see that store sales spike during Thanksgiving, Labor Day and Christmas. This is a clear indicator of seasonal component being present in our data.

Markdowns are nothing but discounts. These discounts are given during holiday seasons. We can clearly see a pattern of offering more discounts during festive seasons to boost up sales.

The line graph displays that monthly sales are highest during festive time

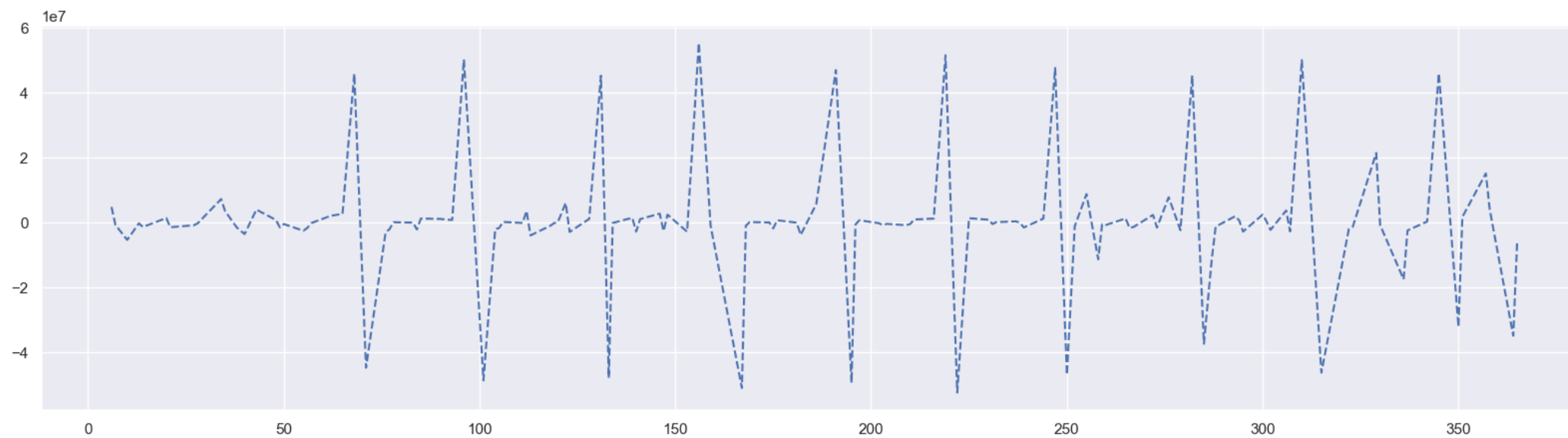
In the Sales distribution, 2012 doesn't show much of a pattern because it doesn't have enough data for holiday sales

# Time series Analysis

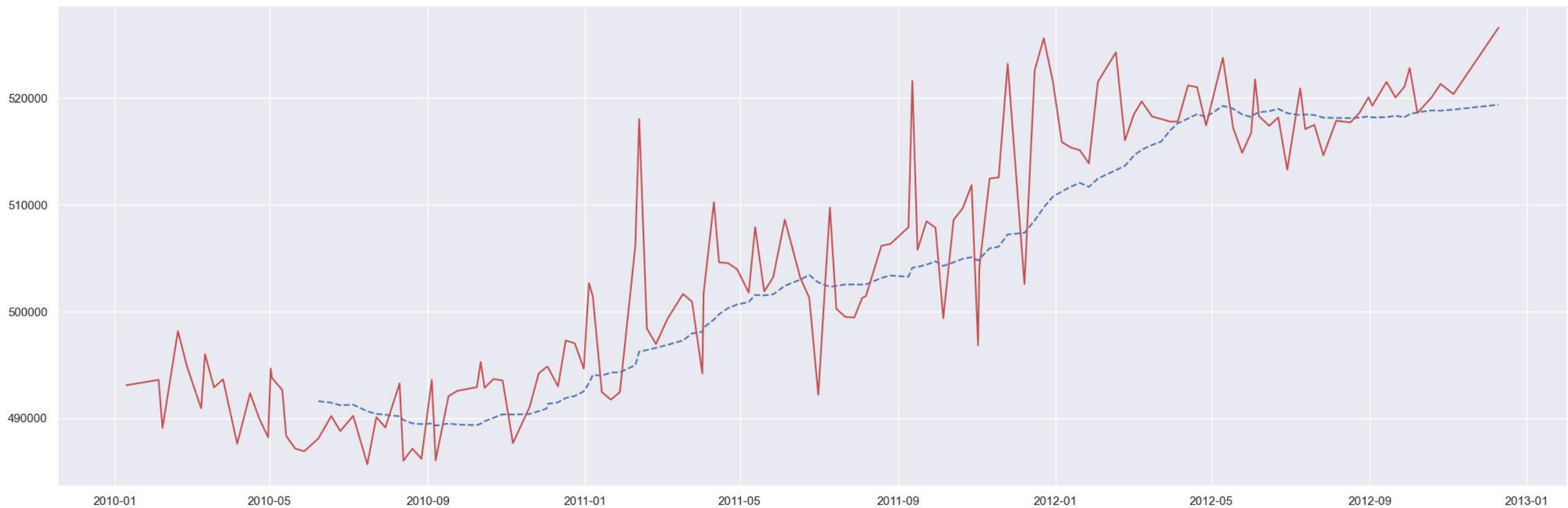


## Insights

Rolling mean(or Moving Average) taken at  $k = 4$  shows a much smoother graph for weekly sales.



Difference of Weekly Sales sum by day of year compared with a previous day of year



The above diagram shows the analysis of CPI over the tenure provided. We can observe that not only payment capability has increased over the years but also that there are visible spikes during festive seasons. The blue dotted line indicates moving average for  $k = 2$ .

# Model

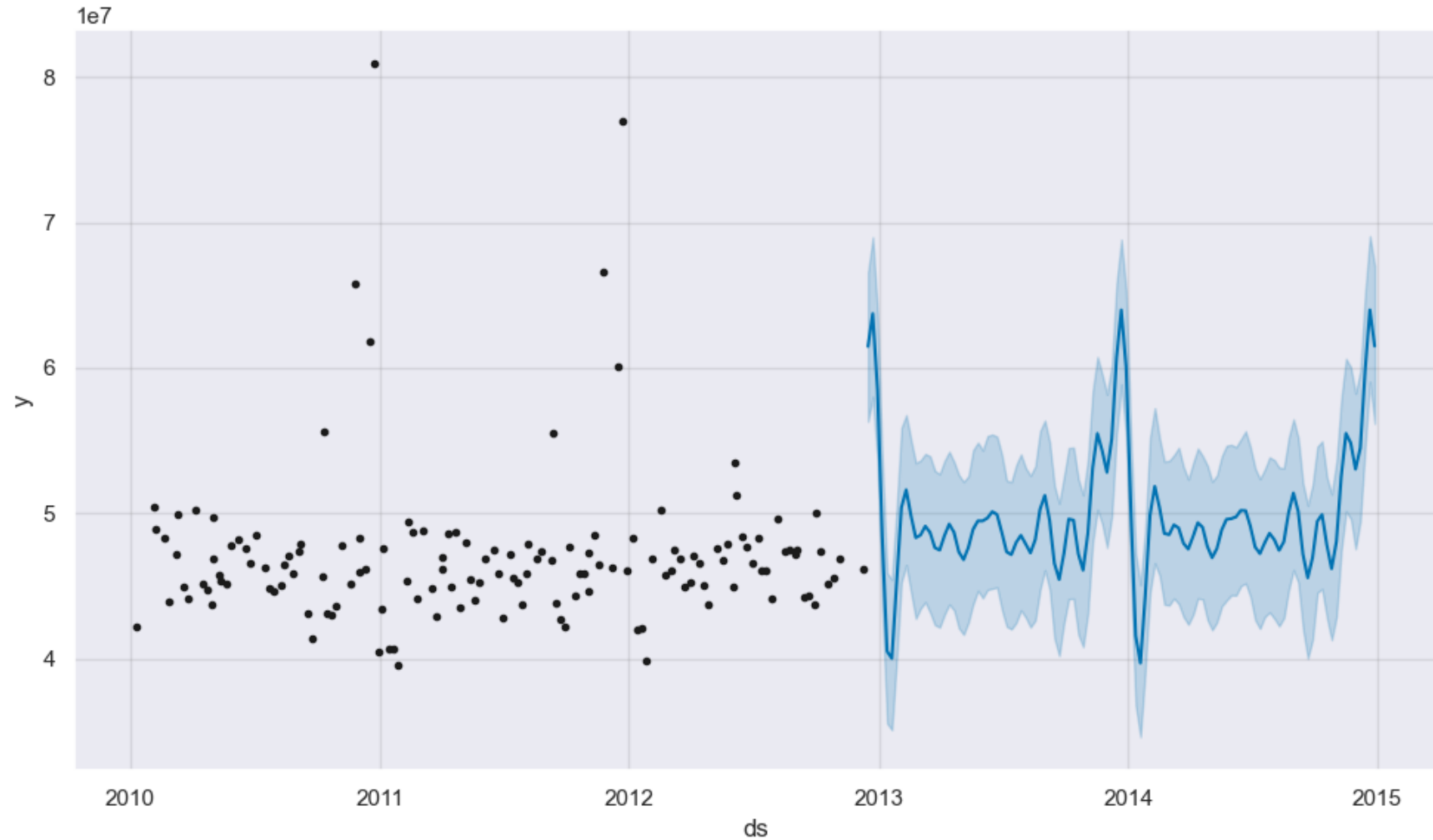
---

Chosen Fbprophet model to forecast weekly sales.

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

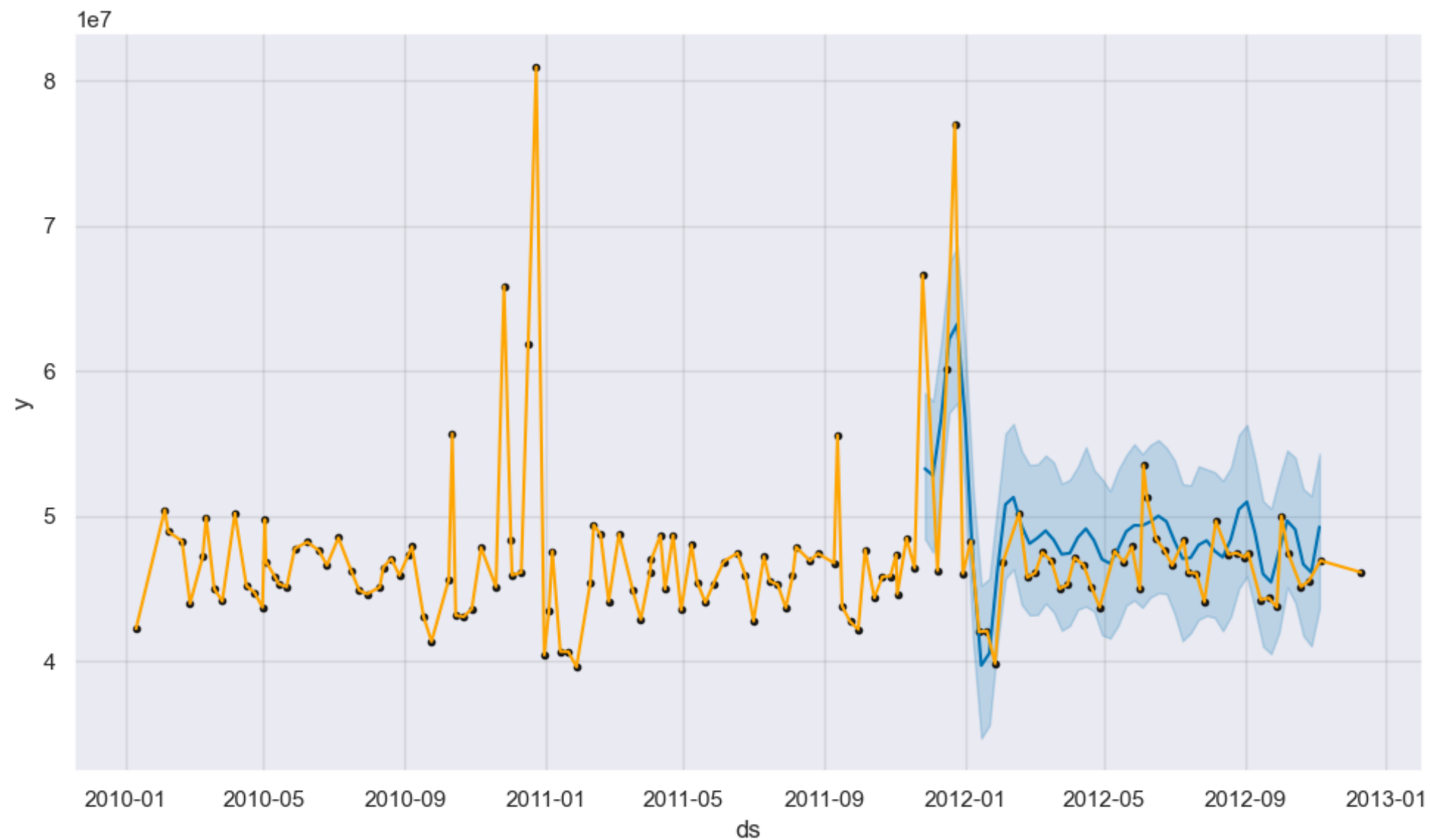
Its fully automatic and accurate and fast.

# Forecast



Here, weekly sales have been forecasted for next three years(2013, 2014, 2015) alongwith the lower and upper boundaries using the FbProphet model.

# Forecast



Here, forecast has been done for next 3 months of 2012 using FbProphet model

# Results

The following patterns were noticed during our analysis :

- Sales is highest for months when there are festivals like Thanksgiving, Black Friday, Christmas
- Larger the size of the store, more is the sales
- When temperature is b/w 20 - 60, the companies have recorded highest sales
- Weekly Sales increased noticeably when the stores started offering promotions

Recommendations for improving sales –

- To improve sales in Type B and Type C, the stores can send out better promotional offers that should align with holiday season
- Type B and Type C stores can increase the size of the store and their interiors to attract more customers

# References

- <https://facebook.github.io/prophet/>
- Instructions to download Prophet can be found here  
<https://stackoverflow.com/questions/50808322/prophet-fbprophet-package-in-python>
- Data can be downloaded from here  
<https://www.kaggle.com/datasets/manjeetsingh/retaildataset?select=sales+data-set.csv>