CAPSTONE PROJECT - WALMART PROJECT

Table of Contents

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.

Problem Objective

The Objective of this project is to…

1. You are provided with the weekly sales data for their various outlets. Use statistical analysis, EDA, outlier analysis, and handle the missing values to come up with various insights that can give them a clear perspective on the following:

a. If the weekly sales are affected by the unemployment rate, if yes - which stores are suffering the most?

b. If the weekly sales show a seasonal trend, when and what could be the reason?

c. Does temperature affect the weekly sales in any manner?

d. How is the Consumer Price index affecting the weekly sales of various stores?

e. Top performing stores according to the historical data.

f. The worst performing store, and how significant is the difference between the highest and lowest performing stores.

2. Use predictive modelling techniques to forecast the sales for each store for the next 12 weeks.

Data Description

The data set available is **walmart.csv**

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 includes the following steps:

1. Load the dataset “Walmart Dataset.csv”
2. Check the shape of the given dataset
3. Check the info of the dataset using **walmart.info ()**, if there is any mismatch in datatype change it. In the given dataset the ‘Date’ datatype is given as ‘Object’
4. Using the **pd.to_datatime.** Convert the **dtype** from **Object** to **datatime64[ns]**
5. Now again check the info of the dataset. Whether all the Columns are perfectly matched with dataset or not
6. Check for null values using walmart.isnull.sum()
7. Check for duplicate values using walmart.duplicated.sum()

Choose the Algorithm for the Project

I have chosen ARIMA and SARIMA algorithm for this project

Because ARIMA, SARIMA is most widely used algorithm for TIME SERIES Forecast Method.

They both differ in handling Seasonality.

Assumptions

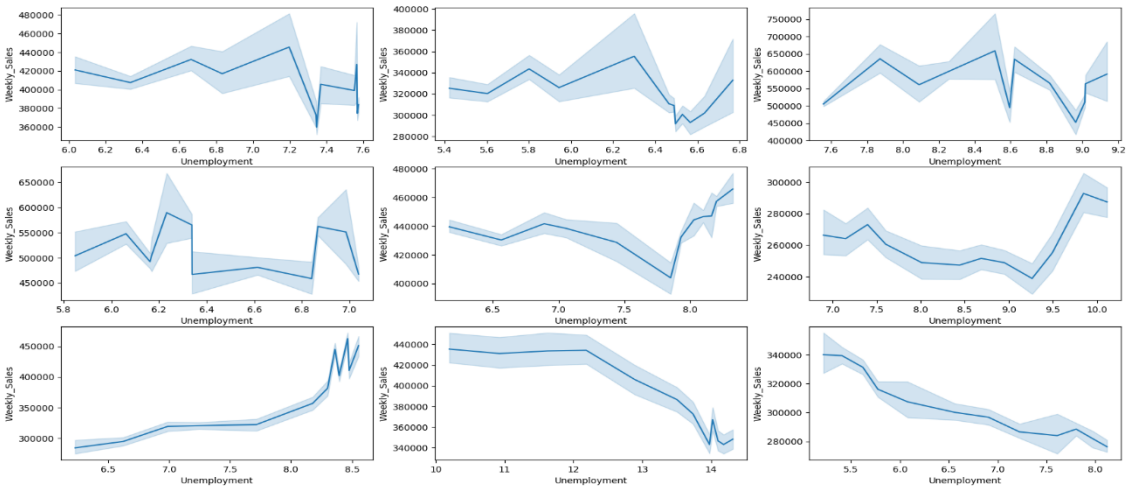
- I assumed that there are no outliers in the data
- I assumed that there is multi correlation between the variables
- I assumed that every year ending there is an increase in sales

Model Evaluation and Technique

- I used Time Series Model by using Arima & Sarima for forecasting

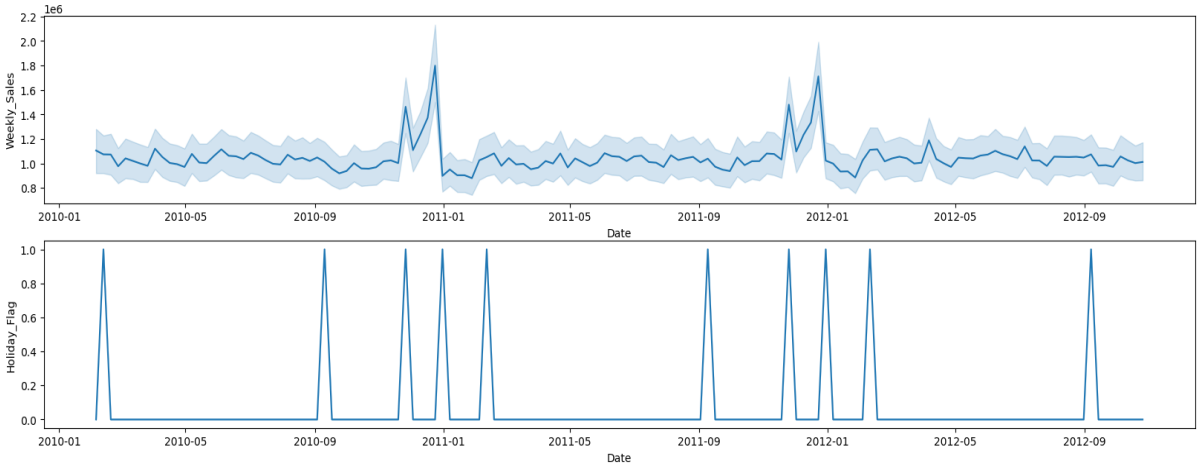
Inferences from the Project

a. If the weekly sales are affected by the unemployment rate, if yes - which stores are suffering the most?



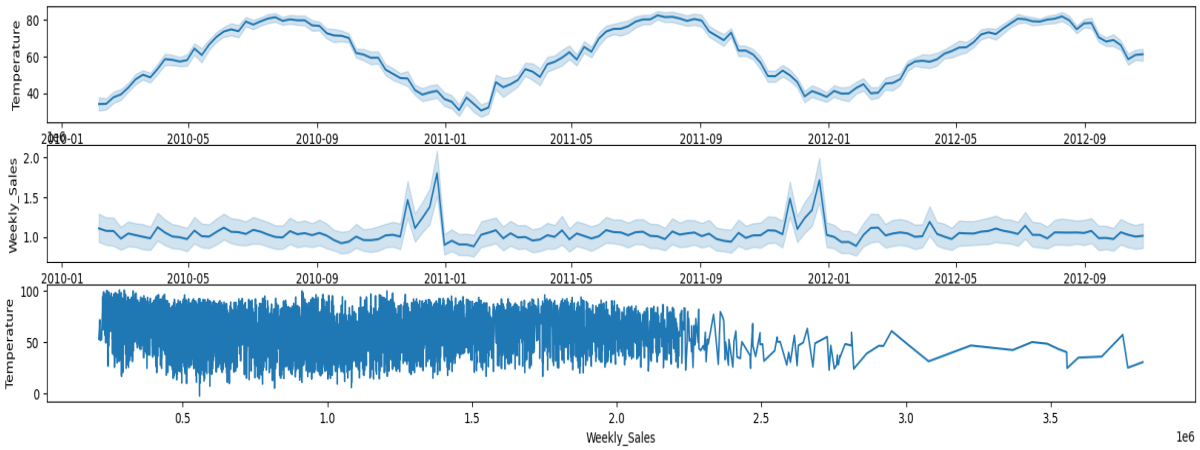
Here we can see that store 38 and store 44 has high impact when unemployment increases sales decrease

b. If the weekly sales show a seasonal trend, when and what could be the reason?



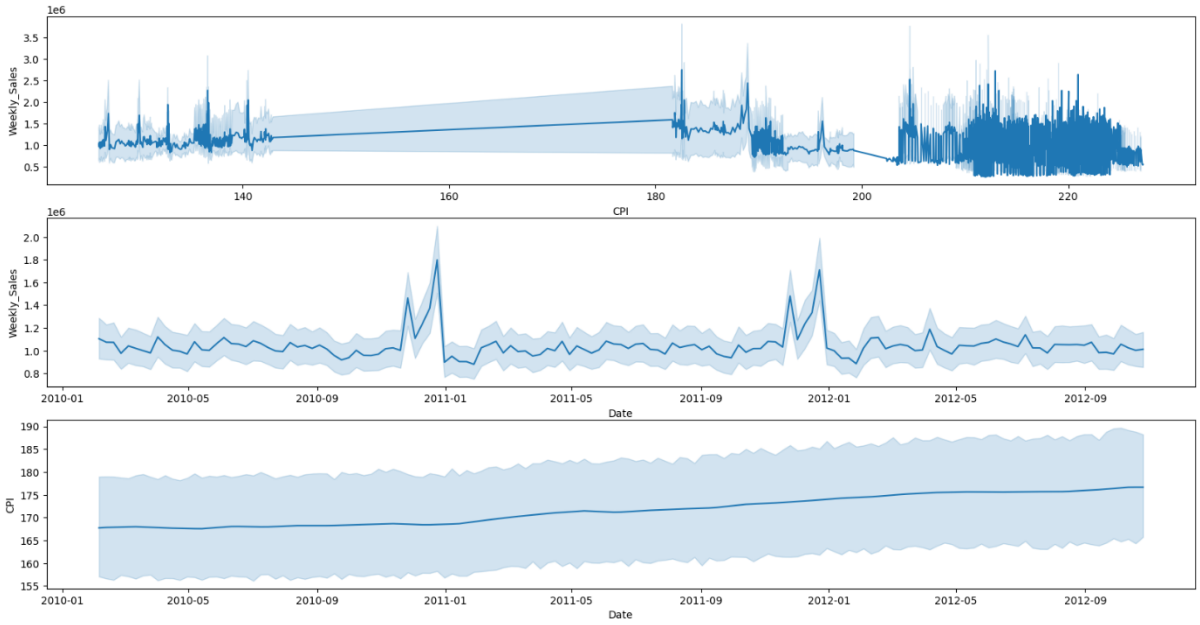
From above two graphs we found that at the end of each year there is increase in sales. It may due to festivals at the end of the year

c. Does temperature affect the weekly sales in any manner?



When the Temperature is getting low, The Weekly Sales got increased

d. How is the Consumer Price index affecting the weekly sales of various stores?



The CPI is increasing in the observed period of time. But there is no upward or downward by weekly_sales

e. Top performing stores according to the historical data

```
Top 10 best performed stores:
[20, 4, 14, 13, 2, 10, 27, 6, 1, 39]
```

f. The worst performing store, and how significant is the difference between the highest and lowest performing stores

highest_sale_of_a_store: 2107676.87
lowest_sale_of_a_store: 259861.69

the difference between the highest and lowest performing stores: 1847815.18

2. Use predictive modelling techniques to forecast the sales for each store for the next 12 weeks

```
2012-05-27    1.046908e+06
2012-06-03    1.050355e+06
2012-06-10    1.059231e+06
2012-06-17    1.048544e+06
2012-06-24    1.050580e+06
2012-07-01    1.038604e+06
2012-07-08    1.065624e+06
2012-07-15    1.050816e+06
2012-07-22    1.064544e+06
2012-07-29    1.063476e+06
2012-08-05    1.087439e+06
2012-08-12    1.046084e+06
Freq: W-SUN, Name: predicted_mean, dtype: float64
```

Future Possibilities

I used Time series forecasting Method based on given data. I got the output consisting of predicted mean values for each week from May 27, 2012, to August 12, 2012 for about 12 weeks.

Conclusion

This project consists of Weekly Sales & Temperature of Walmart. How the weekly sales will vary along with Temperature. I used Time Series Forecasting model for calculating the Predicted Mean for next 12 Weeks using the given Data.

References

- **ChatGPT**
- **Referred some Time Series Forecasting Projects**
- **https://github.com/topics/time-series-analysis?o=asc&s=stars**