Capstone Project – Wallmart sales

Table of Contents

1. Problem Statement
2. Project Objective
3. Data Description
4. Data Pre-processing Steps & inferences
5. Model Evaluation and Techniques
6. Inferences from the Same
7. Future Possibilities limitations

Problem Statement

The dataset we are given  consists historical sales data for 45 Walmart stores located in different regions.

Retail giants like Walmart consider this data as their biggest asset as this helps them predict future sales and customers and helps them lay out plans to generate profits and compete with other organizations. Walmart is an American multinational retail corporation that has almost 11,000 stores in over 27 countries, employing over 2.2 million associates (Wikipedia, n.d.).

Catering to their customers with the promise of ‘everyday low prices’, the range of products sold by Walmart draws its yearly revenue to almost 500 billion dollars thus making it extremely crucial for the company to utilize extensive techniques to forecast future sales and consequent profits.

Following are the task to be done in the project

1. 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.

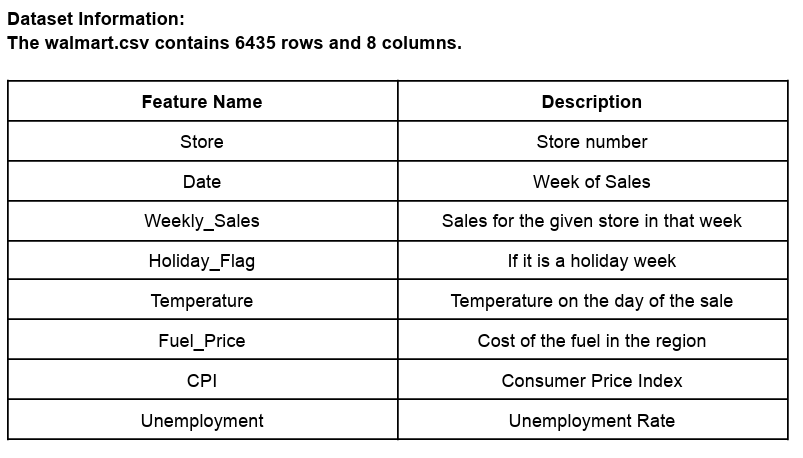
2. Use predictive modeling techniques to forecast the sales for each store for the next 12 weeks

Project Objective

The main focus of this project is to predict Walmart’s sales based on the available historic data and identify whether factors like temperature, unemployment, fuel prices, etc affect the weekly sales of particular stores under study. This study also aims to understand whether sales are relatively higher during holidays like Christmas and Thanksgiving than normal days so that stores can work on creating promotional offers that increase sales and generate higher revenue.

There is additional information in the data set about the factors that might influence the sales of a particular week. Factors like Consumer Price Index (CPI), temperature, fuel price, promotional markdowns for the week, and unemployment rate have been recorded for each week to try and understand if there is a correlation between the sales of each week and their determinant factors.

Correlation testing has been performed to understand if there is a correlation between the individual factors and weekly sales and whether such factors have any impact on sales made by Walmart.

Data Description

Data Preprocessing Steps

1. Exploratory Data Analysis

It is crucial to have an in-depth understanding of the dataset that is used in the analysis to understand the models that would give the most accurate prediction. But in this project the EDA is very specific to the Problem statement.

* Progamming Language—Python is used
* Libraries – Numpy, Pandas, Seaborn, Plotly, Matplot, Statsmodel.

‘

* info’ and ‘isnull’ function is used to get datatypes and null values of all columns. The data set has no null values . Since date column was found to be in ‘object’ datatype , it has been changed to ‘datetime’.
* ‘groupby’ is used to group the weeklysales according to stores
* Scatter plot, subplot, lineplot , heatmap are being used to understand graphical relation between features and weekly sales
* ‘Statsmodel’ library is being used to get various statistical insights of weeklysales of each store Such as ( p-values, stationarity, seasonality, autocorrelation)

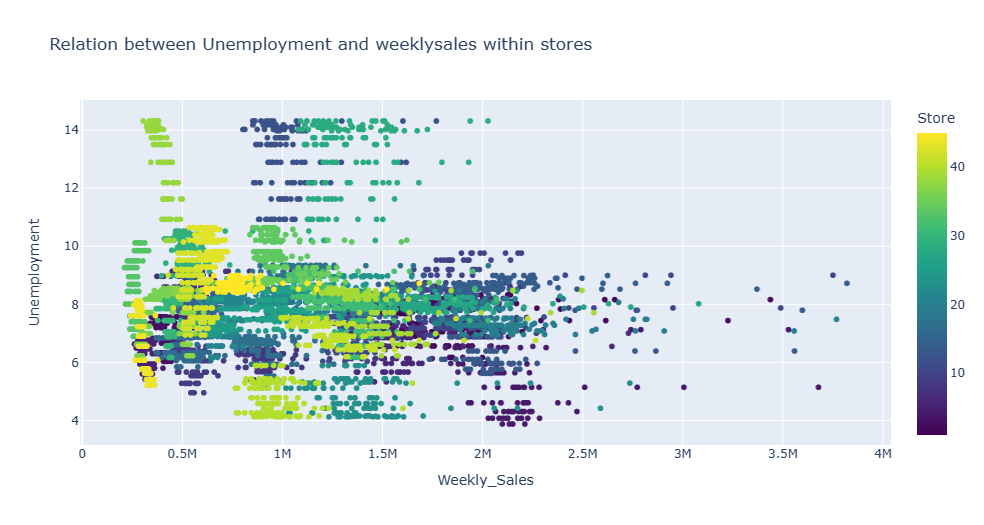
Top performing stores according to historical data

Top – store no 4

Bottom – store no 33

Impact of unemployment on sales

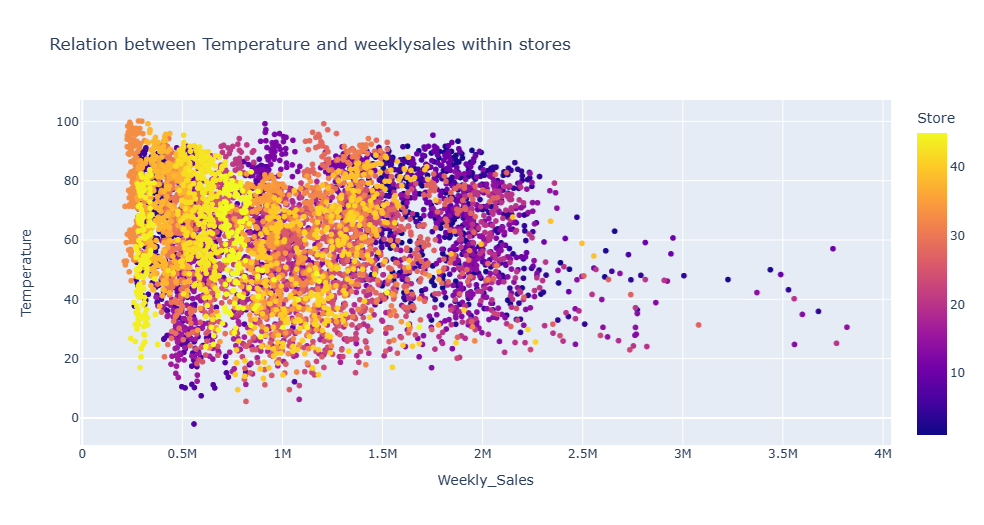
* There seems to be a visible decrease in sales when the unemployment index is higher than 11
* 2 stores affected by unemployment store [33,38]
* It has a negative correlation i.e (-0.11)



Impact of temperature on sales

Temperatures between 40 to 70 degrees Fahrenheit are considered as favourable for humans to live in considering they are not as hot or cold.

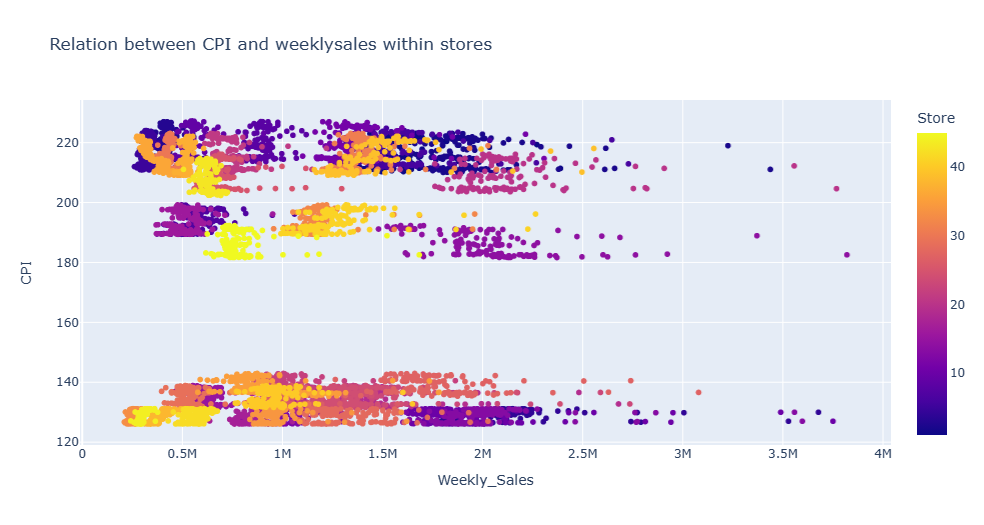
As seen below, the highest sales occur for most store types between the range of 40 to 80 degrees Fahrenheit, thus proving the idea that pleasant weather encourages higher sales. Sales are relatively lower for very low and very high temperatures but seem to be adequately high for favourable climate condition.



Impact of CPI on sales

A higher CPI generally means that the price of goods has increased and that an individual needs to spend more money to maintain the same standard of living.

In our scatter plot below, we can identify three different clusters around different ranges of CPI; while there seems to be no visible relationship between the change in CPI and weekly sales for Walmart stores (sales still occur at high CPI rates), the only negligible observation that can be made is the high amount of sales for store (25-35) when CPI is at a low rate of 140



Model Evaluation and Technique

Considering the seasonality, stationarity and autocorrelation we used ARIMA model for the time series analysis.

Arima model is best suited for forecasting the sales

For the ML model the test to train ratio is 20:80

Forecast graph shows the actual and predicted trend

In this project 5 random stores has been selected and ML algorithm is applied to each store to forecast the sales for the next 12 weeks

Limitations

A huge constraint of this study is the lack of sales history data available for analysis. The data for the analysis only comes from a limited number of Walmart stores between the years 2010 and 2013. Because of this limited past history data, models cannot be trained as efficiently to give accurate results and predictions. Because of this lack of availability, it is harder to train and tune models as an over-constrained model might reduce the accuracy of the model. An appropriate amount of training data is required to efficiently train the model and draw useful insights.

Future Possibilities

With growing technology and increasing consumer demand , Walmart can shift its focus on the e-commerce aspects of the business. Taking inspiration from Amazon’s business model, Walmart can grow its online retail business massively and gather huge profits. With already established stores and warehouses, it is easier for the organization to create a nationwide reach, limiting the presence of physical stores and helping their customers save on fuel costs by delivering goods at their doorstep. It also makes it a lot easier to identify consumer buying patterns.

Another aspect that would be worth exploring with this study is identifying trends with sales for each of the stores and predicting future trends based on the available sales data.