Walmart Store Sales

a. 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. You are a data scientist, who has to come up with useful insights using the data and make prediction models to forecast the sales for X number of months/years.

b. Project Objective

1. Using the above data, come up with useful insights that can be used by each of the stores to improve in various areas.
2. Forecast the sales for each store for the next 12 weeks.

c. Data Description

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

d. Data Pre-processing Steps and Inspiration

Gathering of the data, cleaning data & analyzing the data and weekly sales of each store and plotting charts and scatter plots between parameters.

1. Converts date column of type object to datetime type
2. Plotting date vs weekly sales for all store
3. Mean of sales grouped by weeks

e. Choosing the Algorithm for the Project

Algorithm Chosen: ARIMA, SARIMAX

f. Motivation and Reasons for Choosing the Algorithm

Autoregressive Integrated Moving Average (ARIMA): Auto Regressive Integrated Moving Average, ARIMA, models are among the most widely used approaches for time

series forecasting. It is actually a class of models that ‘explains’ a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

Seasonal Autoregressive Integrated Moving Average (SARIMA): Seasonal autoregressive integrated moving average (SARIMA) models extend basic ARIMA models and allow for the incorporation of seasonal patterns.

g. Assumptions

A common assumption in many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time.

h. Model Evaluation and Techniques

Model Evaluation: RMSE, RMSE/MEAN

i. Inferences from the Same

RMSE: 13031.85, RMSE/MEAN = 0.03836

j. Future Possibilities of the Project

With the data set, we have found the predicted sales of walmart stores, we can now extend to even departments and find which department is contributing to the high sales in the particular store