Retail Analysis with Walmart Data

Problem Statement :

One of the leading retail stores in the US, Walmart, would like to predict the sales and demand accurately. There are certain events and holidays which impact sales on each day. There are sales data available for 45 stores of Walmart. The business is facing a challenge due to unforeseen demands and runs out of stock some times, due to the inappropriate machine learning algorithm. An ideal ML algorithm will predict demand accurately and ingest factors like economic conditions including CPI, Unemployment Index, etc.

Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labour Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. Part of the challenge presented by this competition is modelling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data. Historical sales data for 45 Walmart stores located in different regions are available.

Objectives :

To perform these tasks, you can use any of the different Python libraries such as NumPy, SciPy, Pandas, scikit-learn, matplotlib.

Prerequisites :

- **Basics of Python**
- **Application of Python libraries in data science**
- **Perform analysis on a dataset**
- **Knowledge of Data Frame**
- **Train and perform prediction on a dataset**

Discerption:

This is the historical data that covers sales from 2010-02-05 to 2012-11-01, in the file Walmart_Store_sales. Within this file you will find the following fields:

- **Store - the store number**
- **Date - the week of sales**
- **Weekly_Sales - sales for the given store**
- **Holiday Flag - whether the week is a special holiday week 1 – Holiday week 0 – Non-holiday week**
- **Temperature - Temperature on the day of sale**
- **Fuel Price - Cost of fuel in the region**
- **CPI – Prevailing consumer price index**
- **Unemployment - Prevailing unemployment rate**

ANALYSIS:

Understand the dataset:,

Load dataset and convert date into datetimeformat and make 3 new columns fo date, month and year

In [3]:
```
data = pd.read_csv("Walmart_Store_sales.csv")
data.head()
```

Out[3]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	05-02-2010	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	12-02-2010	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	19-02-2010	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	0	46.50	2.625	211.350143	8.106

In [4]:
```
data['Date'] = pd.to_datetime(data['Date'])
data.head()
```

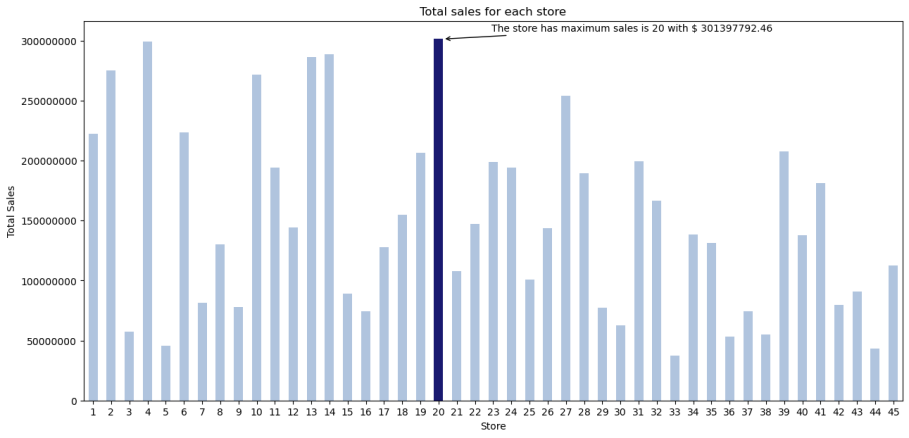
Out[4]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	2010-05-02	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	2010-12-02	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	2010-02-19	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	2010-02-26	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	2010-05-03	1554806.68	0	46.50	2.625	211.350143	8.106

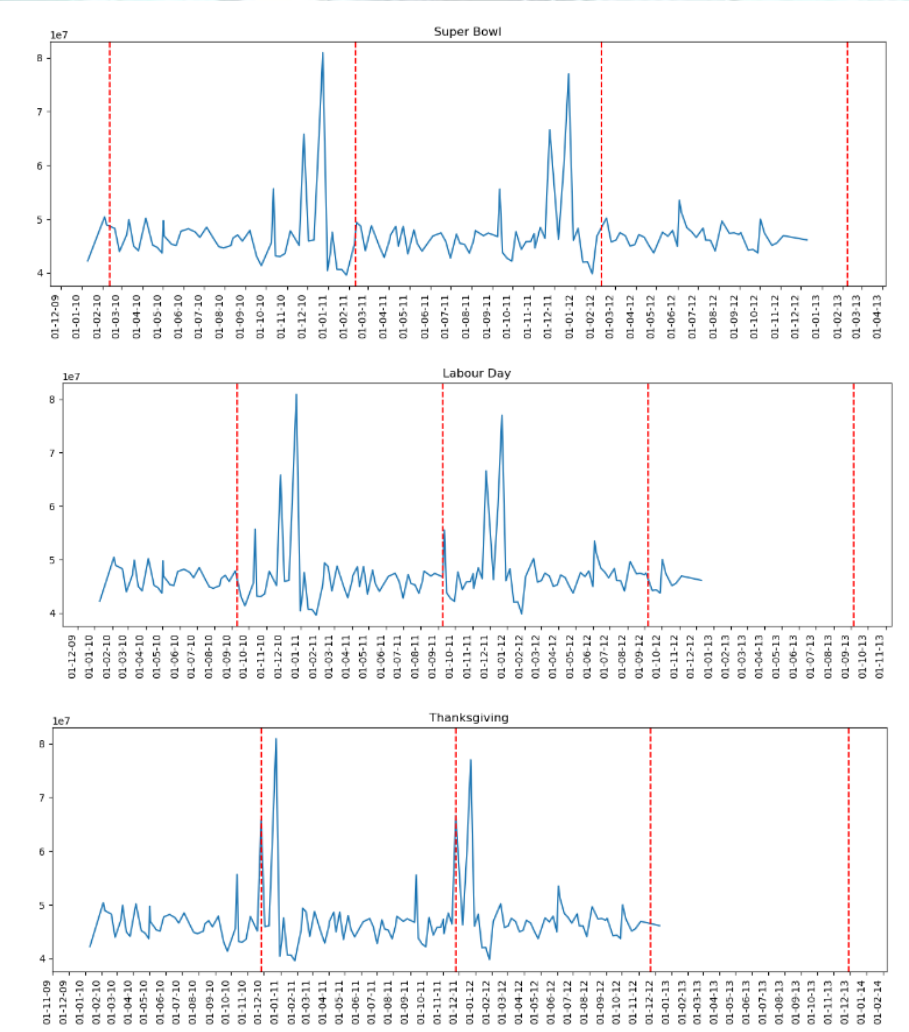
Basic data exploratory analysis:

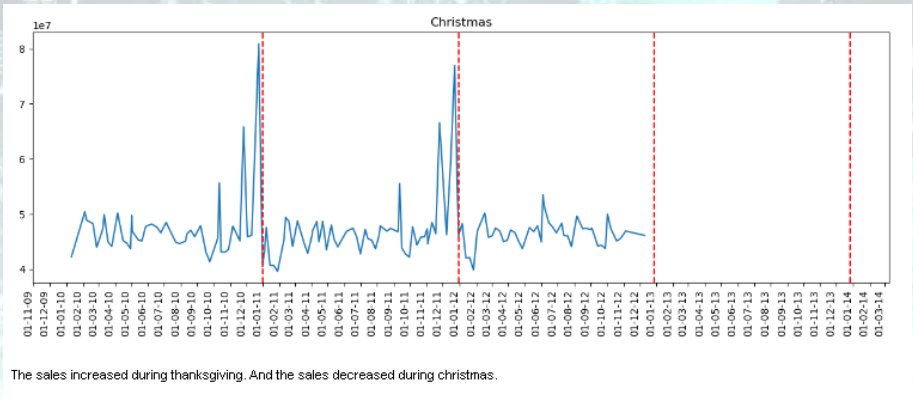
Commented [VT1]:

TO FIND WHICH STORE AS MAXIMUM SALES



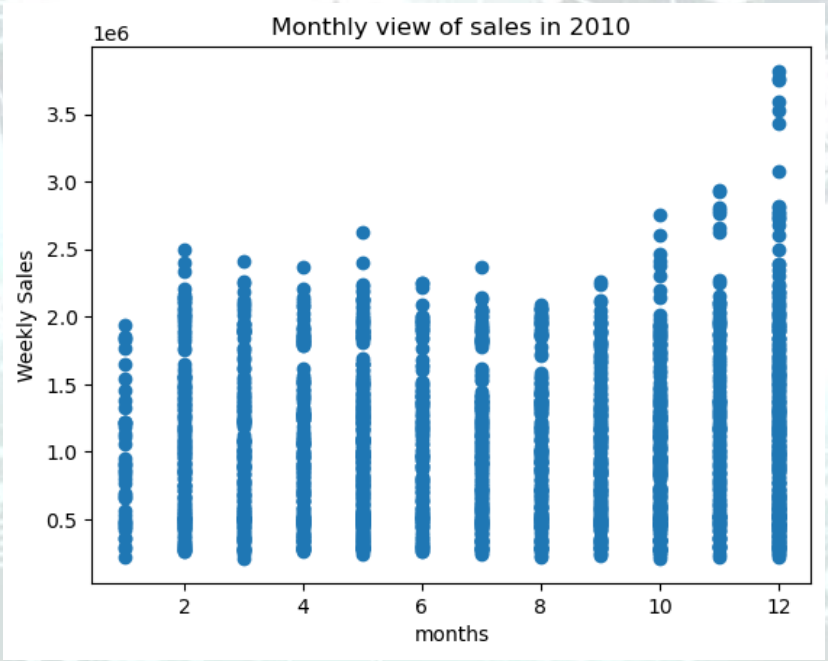
Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together

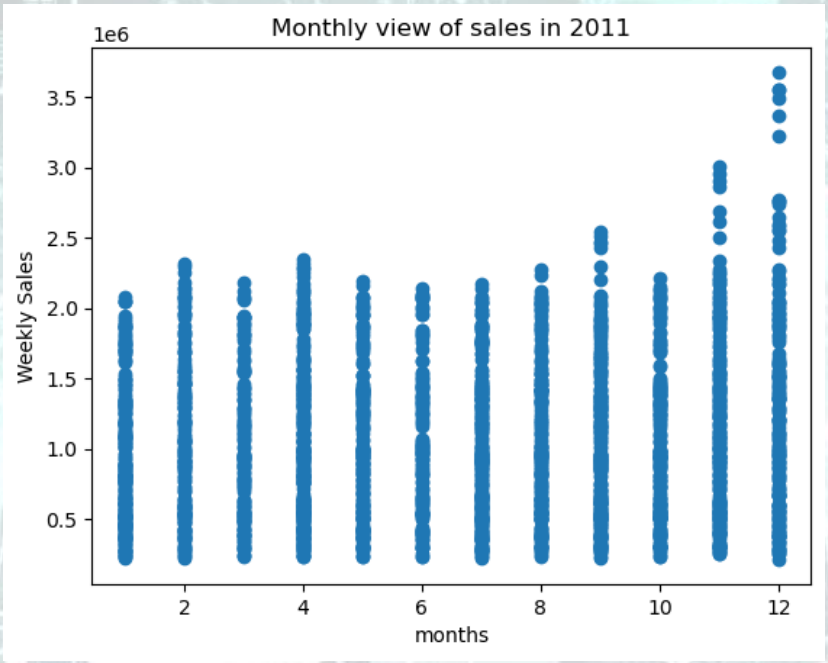


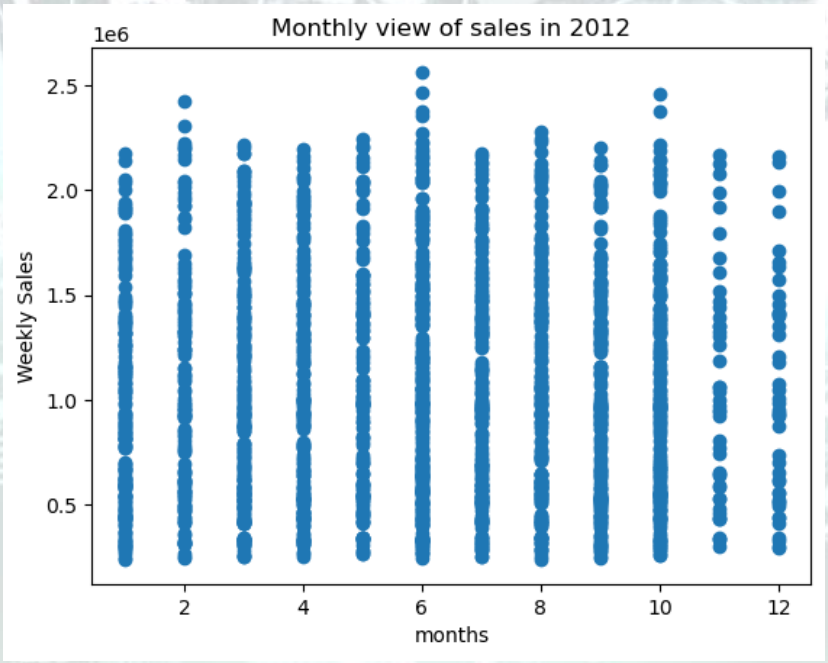


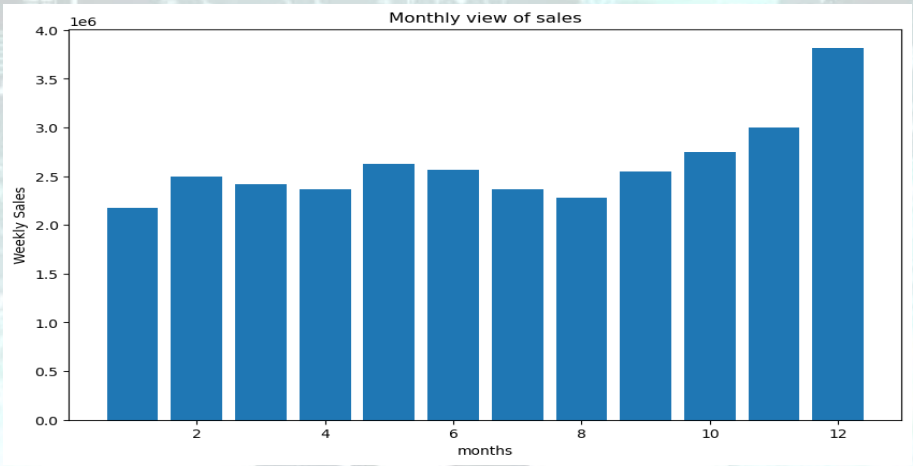
The sales increased during thanksgiving. And the sales decreased during christmas.

Monthly and semester view of sales in units and give insights

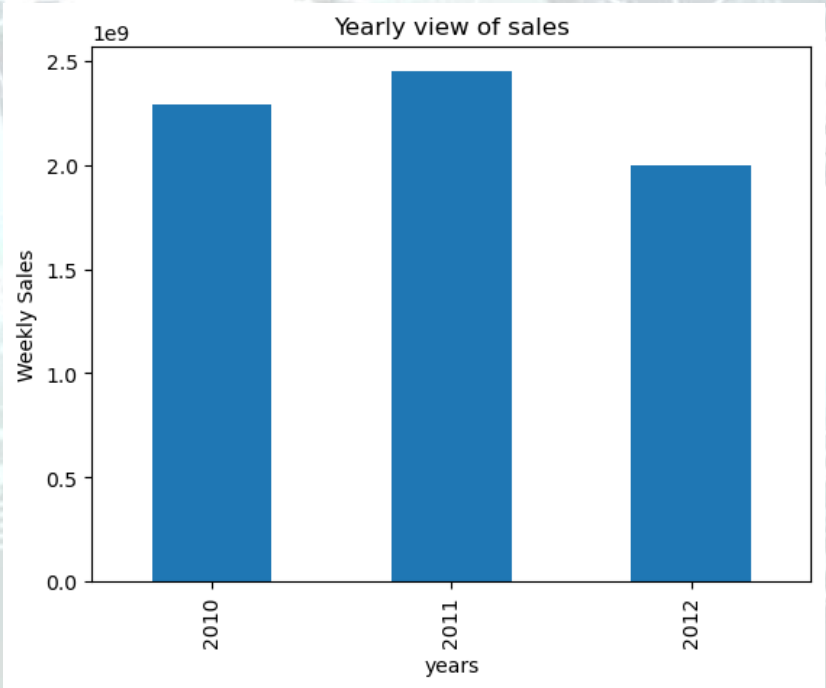



Monthly view of sales in 2011
1e6
Weekly Sales
months
0.5
1.0
1.5
2.0
2.5
3.0
3.5
2
4
6
8
10
12


Monthly view of sales in 2012
1e6
Weekly Sales
months
0.5
1.0
1.5
2.0
2.5
2
4
6
8
10
12



Yearly view of the sales



Changing dates into days by creating new variable

```
In [21]: data['Day'] = pd.to_datetime(data['Date']).dt.day_name()
         data.head()
```

Out[21]:

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	Day	Month	Year
0	1	2010-05-02	1643690.90	0	42.31	2.572	211.096358	8.106	Sunday	5	2010
1	1	2010-12-02	1641957.44	1	38.51	2.548	211.242170	8.106	Thursday	12	2010
2	1	2010-02-19	1611968.17	0	39.93	2.514	211.289143	8.106	Friday	2	2010
3	1	2010-02-26	1409727.59	0	46.63	2.561	211.319643	8.106	Friday	2	2010
4	1	2010-05-03	1554806.68	0	46.50	2.625	211.350143	8.106	Monday	5	2010

The dataset contains the information about the Changing dates into days by creating new variable.

<u>OBSERVATION OF THE DATA SET :</u>

- **Store number 20 has the maximum sale of sum total $ 301397792.46**
- **Store number 14 has maximum standard deviation of STD 317569.9994755081**
- **Store number 4 has the good quarterly growth in Q3 2012 of 25656119.35**
- **Some holiday have negative impact on the sales**
    1. **Super bowl day total sales : 1456882278.34**
    2. **Labour day total sales : 140727684.68**
    3. **Thanksgiving day total sales : 132414608.5**
    4. **Christmas day total sales : 86474980.039999**
- **Sales increased during the Thanksgiving day and decreased during Christmas**
- **In monthly view sales of 2010,2011 and 2012 the December month has the highest**
- **In the yearly view sales the 2011 has the highest sale , next comes 2010 and 2012 has least sales**


DATA SCIENCE
DATA ANA
DATA
PROCES
DATA PRESE
DATA APPLICATION

THANK YOU