**WALMART SALES ANLYSIS**



**ABOUT**

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 sometimes, 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.

**In this project we focused retail analysis with Walmart data and answer the following questions:**

1. Which stores have maximum and sales?
2. Which store has maximum standard deviation i.e., the sales vary a lot? Also, find out the coefficient of mean to standard deviation.
3. Which store/s has good quarterly growth rate in Q3’2012?
4. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together.
5. Provide a monthly and semester view of sales in units and give insights.
6. Build prediction to forecast demand.

**DATASET**

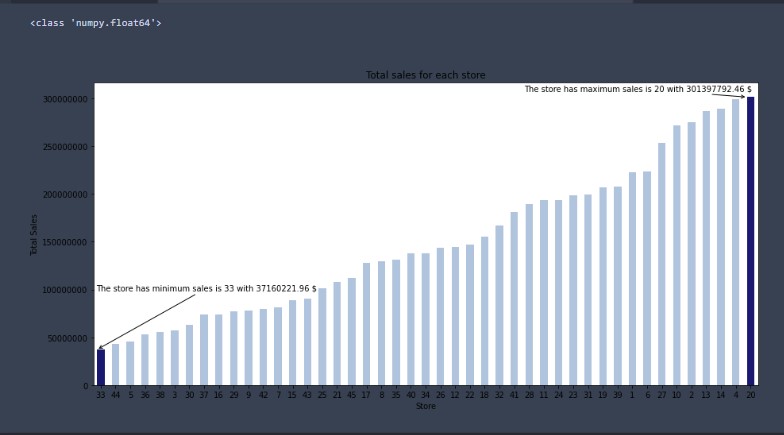
This is the historical data that covers sales from 2010-02-05 to 2012-11-01, in the file Walmart *Stor*e sales. Within this file you will find the following fields:

1. Store - the store number
2. Date - the week of sales
3. Weekly Sales - sales for the given store
4. Holiday Flag - whether the week is a special holiday week 1 – Holiday week 0 – non-holiday week
5. Temperature - Temperature on the day of sale
6. Fuel Price - Cost of fuel in the region
7. CPI – Prevailing consumer price index
8. Unemployment - Prevailing unemployment rate

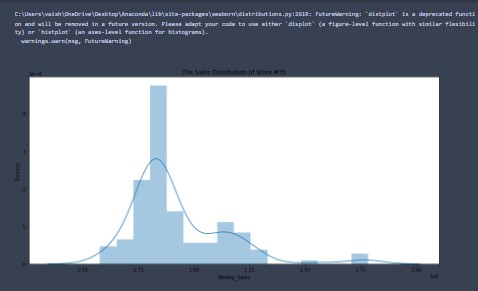
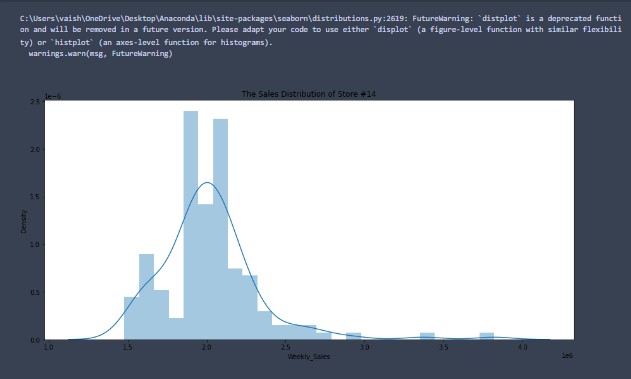
**IMPLEMENTATION**

**Steps:** In this project, we used [RandomForestRegressor](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html) and [Linear Regression](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html) to predict of sales. The data have been split into training and testing with a ratio of 80:20.

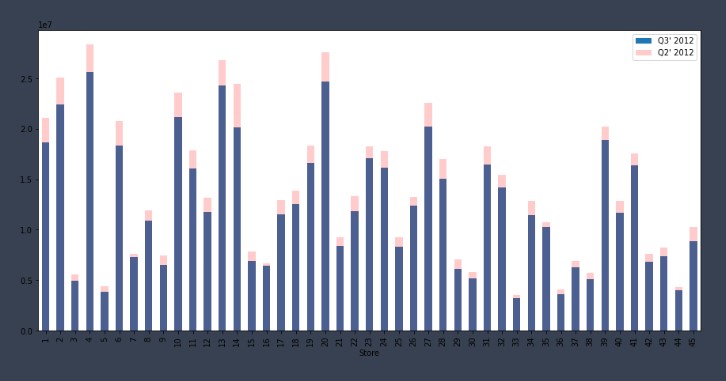
1. Data Preparation
2. Which store has minimum and maximum sales?



1. Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation?



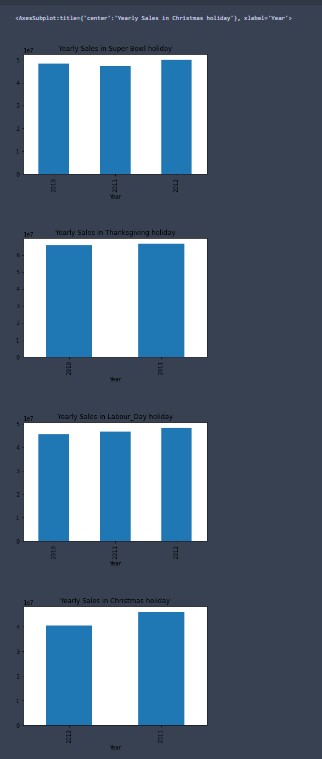
1. Which store/s has good quarterly growth rate in Q3’2012



1. 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

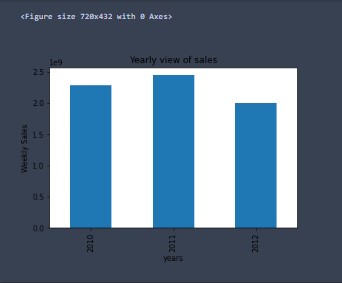
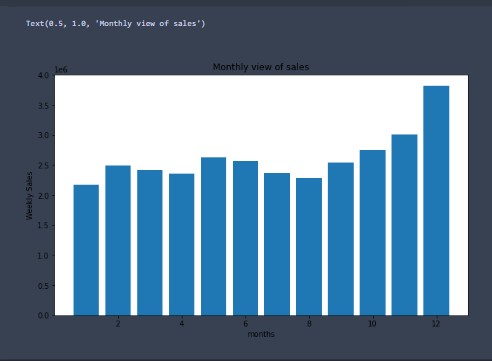
**Holiday Events**

* Super Bowl: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13
* Labour Day: 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13
* Thanksgiving: 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13
* Christmas: 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

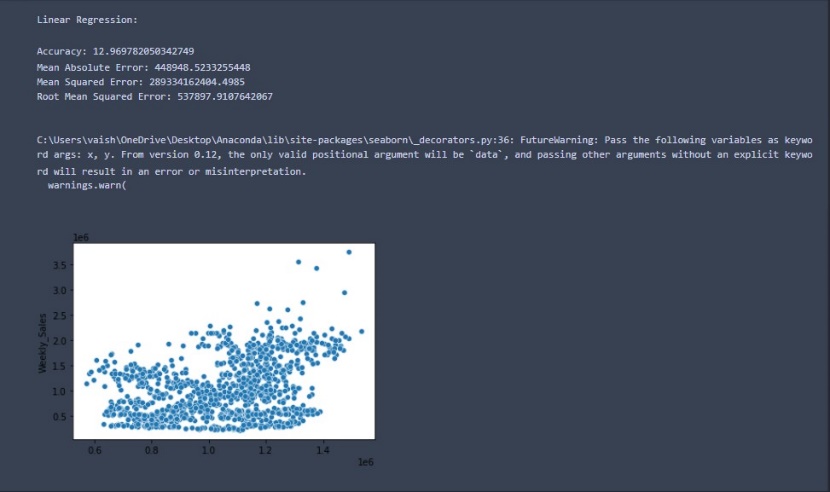


* **The sales increased during Christmas and Thanksgiving**

1. Provide a monthly and semester view of sales in units and give insights



1. Building Models
2. Linear Regression: Utilize variables like date and restructure dates as 1 for 5 Feb 2010 (starting from the earliest date in order). Hypothesize if CPI, unemployment, and fuel price have any impact on sales.



1. Random Forest Regression: Change dates into days by creating new variable.



**CONCLUSION**

The details of the results show in the code.

