**Capstone Project-**

**Walmart Project**

**Table of Contents:**

1. Problem Statement
2. Project Objective
3. Data Description
4. Data pre-processing Steps and Inspiration
5. Choosing an algorithm for the project
6. Motivation and reasons for choosing the algorithm
7. Model evaluation and techniques
8. Inferences from the same
9. Future possibilities of the project
10. Conclusion
11. References
12. **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.

1. **Project Objective-**

To forecast items demand in advance to fulfil its supply.

1. **Data Description-**

This is the historical data that covers sales from 2010-02-05 to 2012-11-01, in the file Walmart\_Store\_sales.

**Dataset Information:**

The walmart.csv contains 6435 rows and 8 columns.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

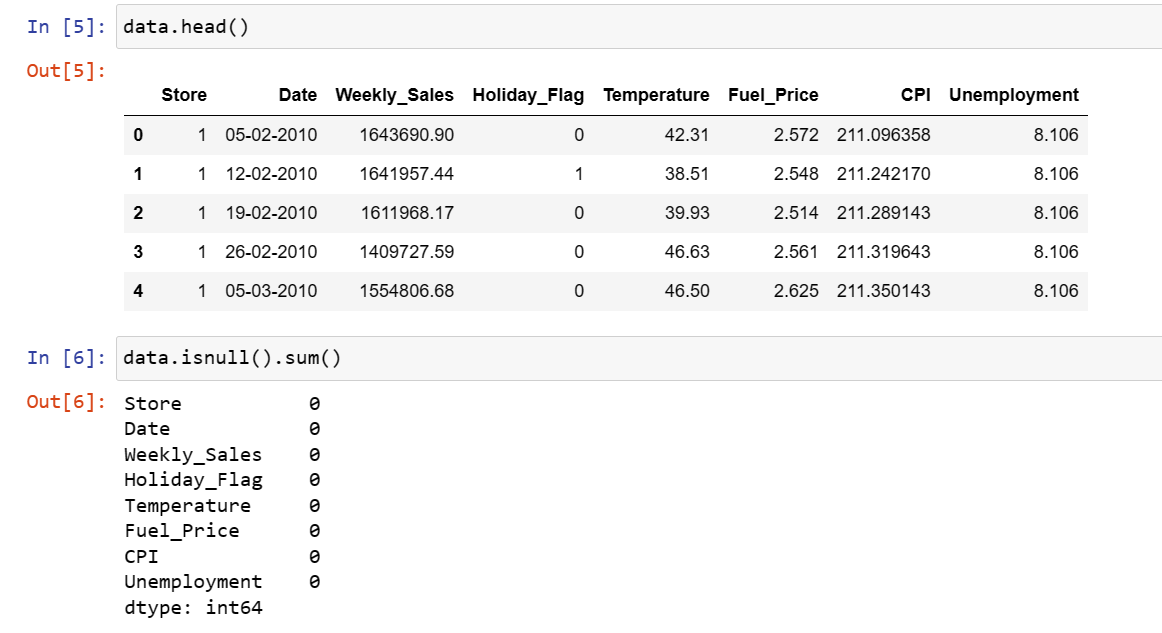
Fuel\_Price - Cost of fuel in the region

CPI – Prevailing consumer price index

Unemployment - Prevailing unemployment rate

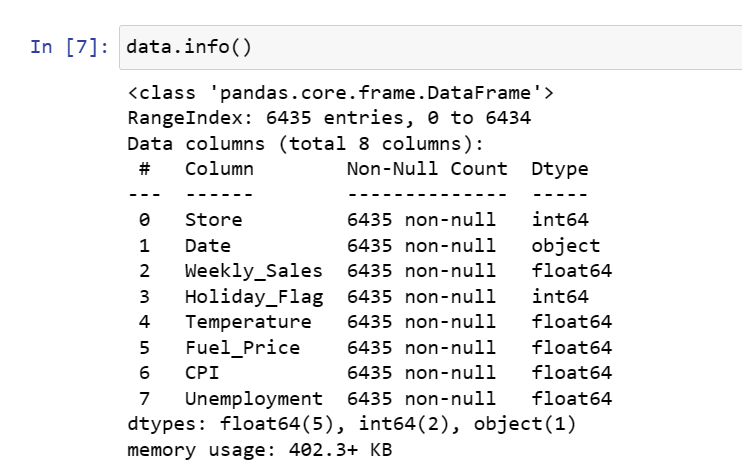
1. **Data pre-processing Steps and Inspiration-**

**I firstly checked my dataset for the missing/null values**

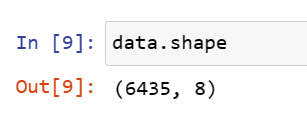
****

By running the above code , I discovered theres no missing values in the dataset.

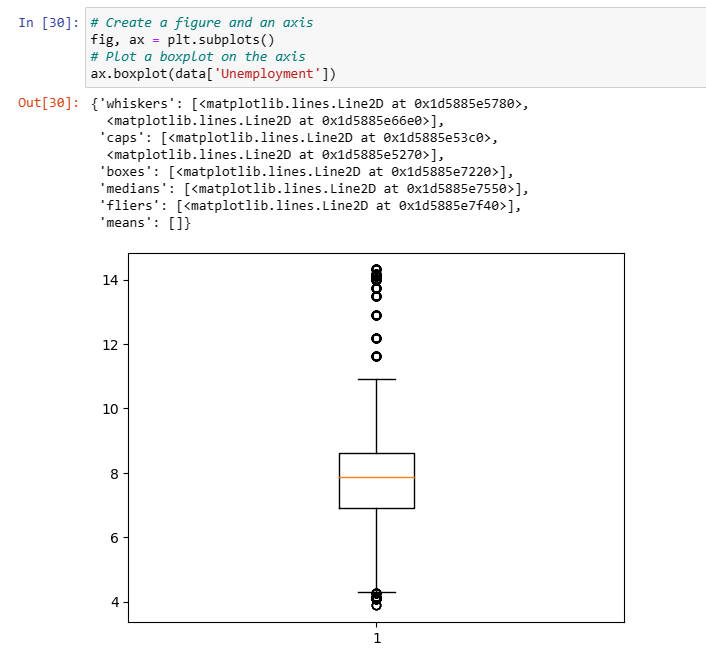
I checked for the datatypes of the dataset.



I found the number of rows ans columns in the dataset

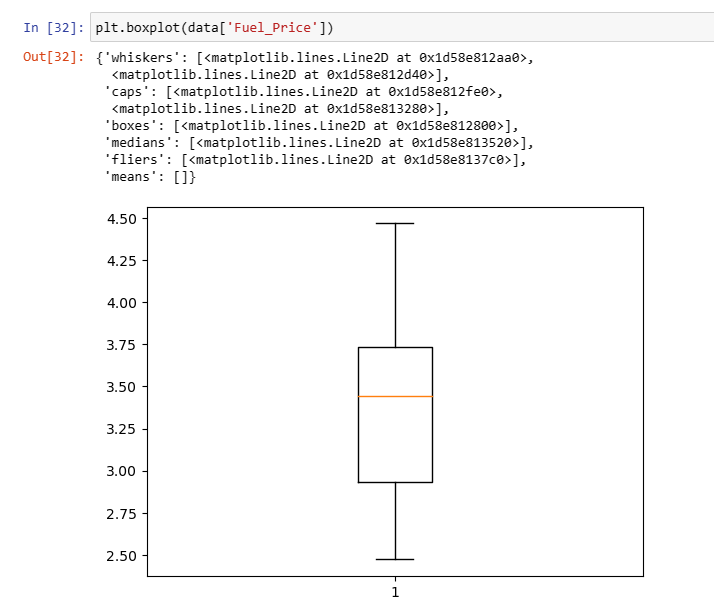


**Checking outliers in Unemployement field**

.

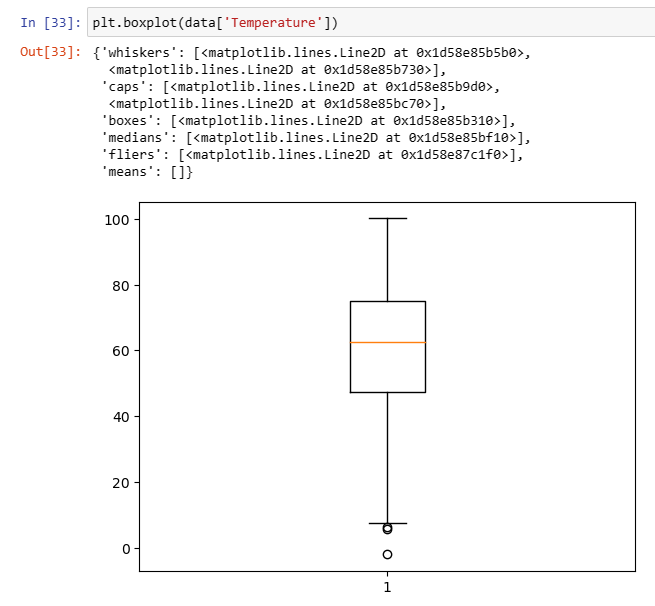
**There are outiers in unemployement .outliers are<5.5 and>11**

Checking outliers for Fuel\_price field



**There are no outliers in fuel price**

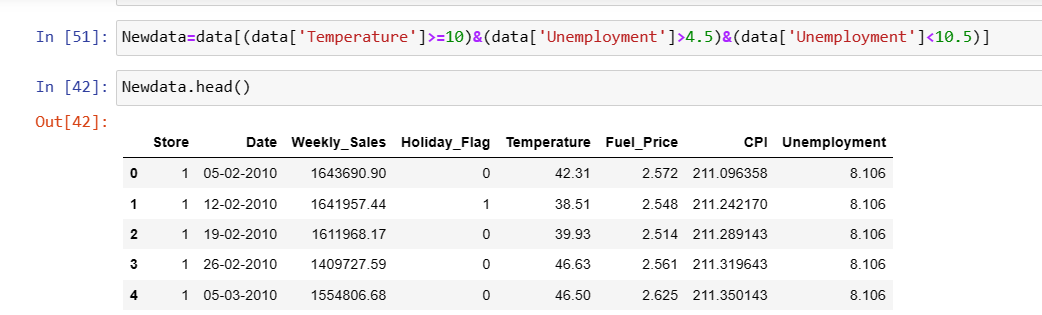
**Checking outliers in Temperature field:**

****

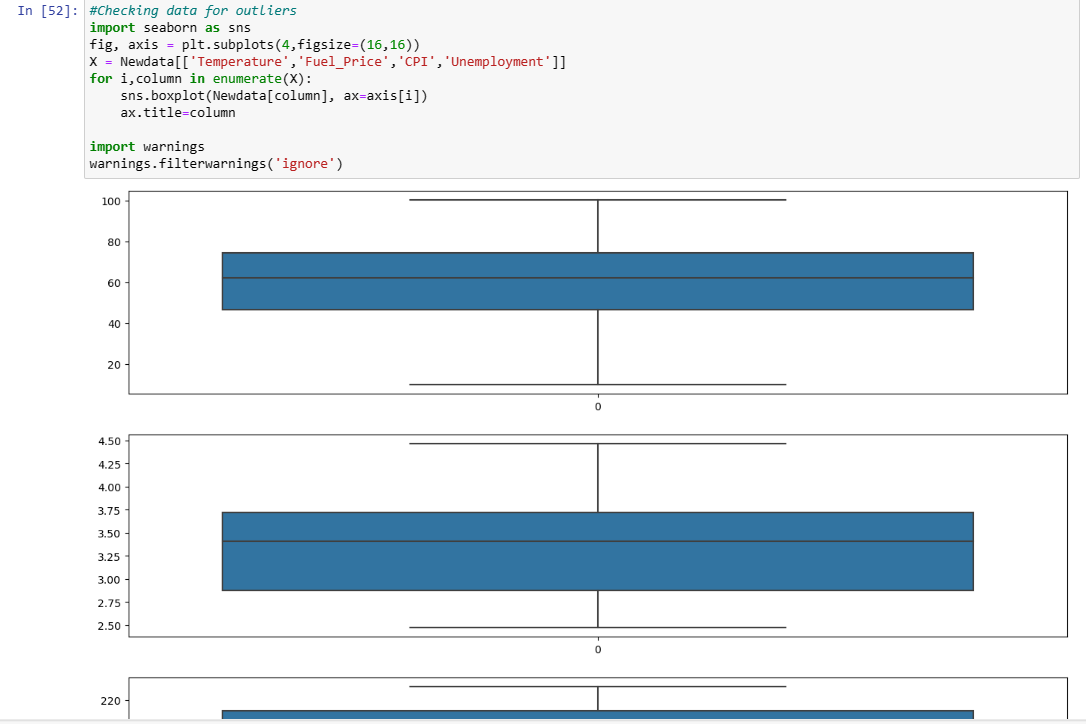
**All outliers in temperature are <10**

**Handling outliers:**

**Removing the records with outlier values of Temperature and unemployement:**

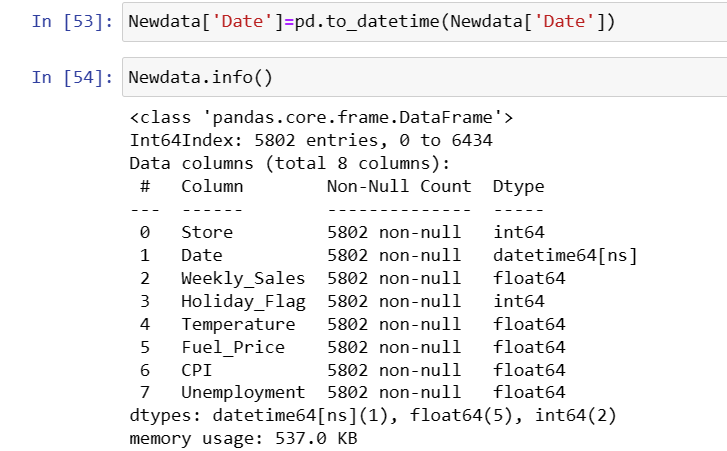
****

**Checking outliers again:**

****

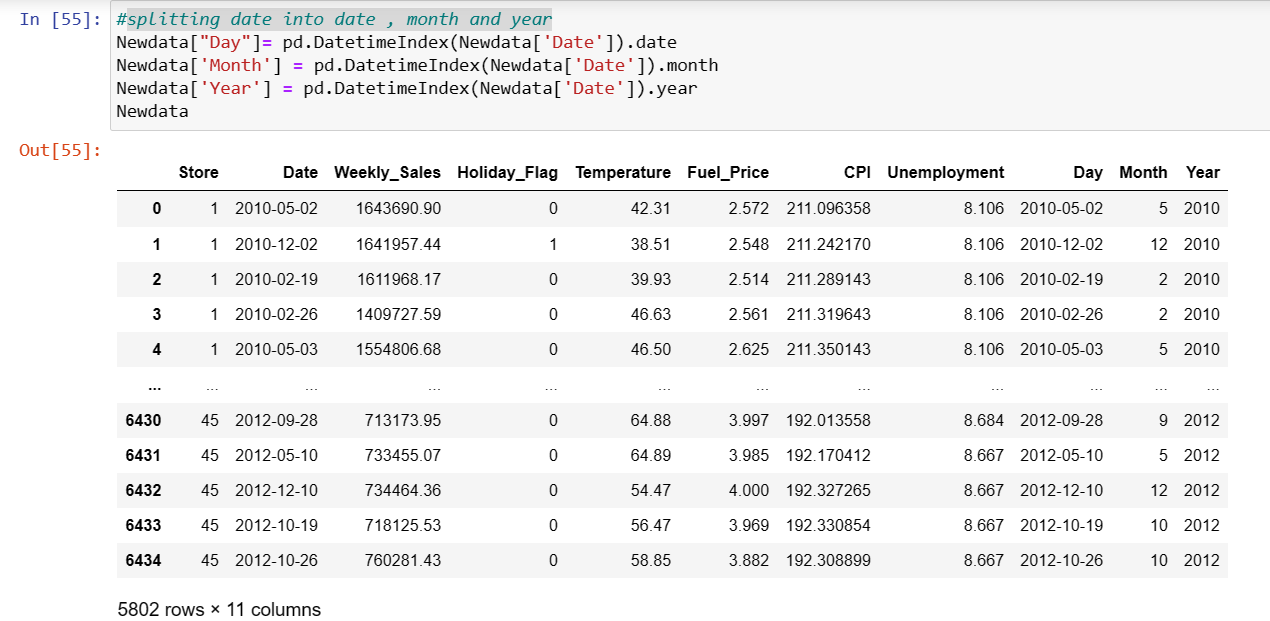
**Outliers have been removed**

**Converting data type of date:**

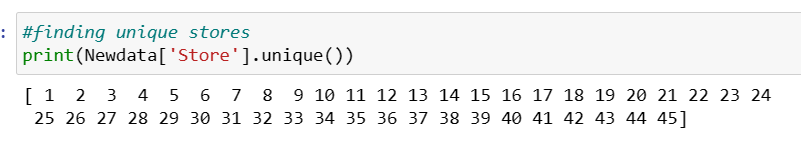
****

**Feature Engineering:**

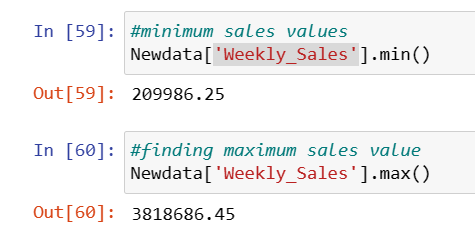
**splitting date into date , month and year:**

****

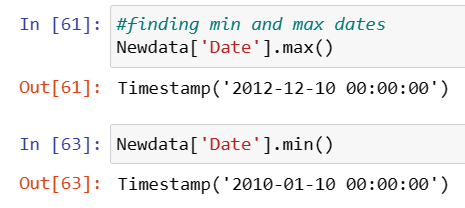
**finding unique stores**

****

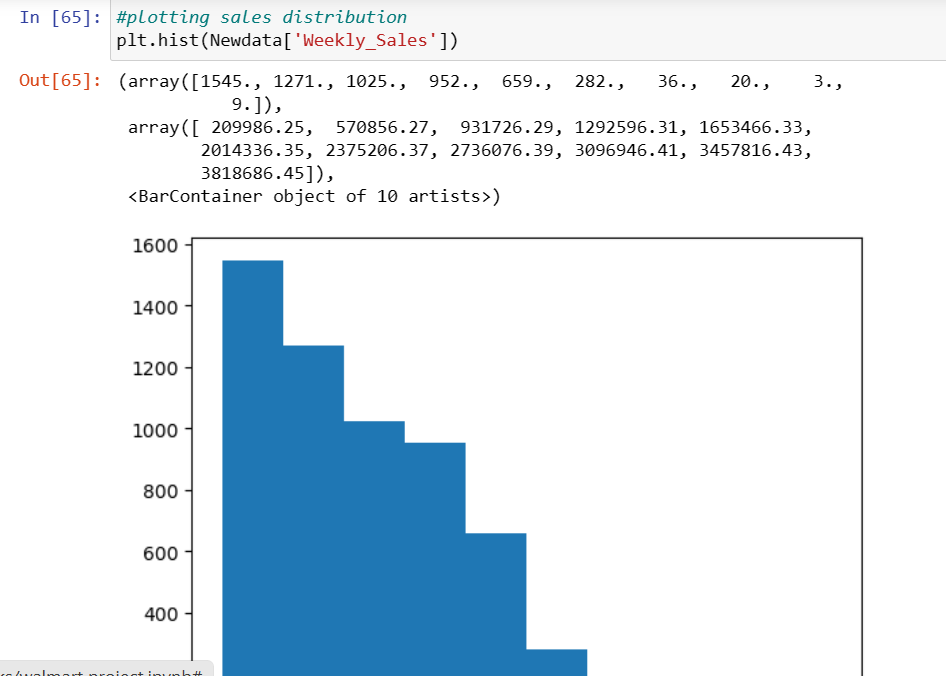
Finding Minimum and maximum sales values



Finding minimum and maximum dates:

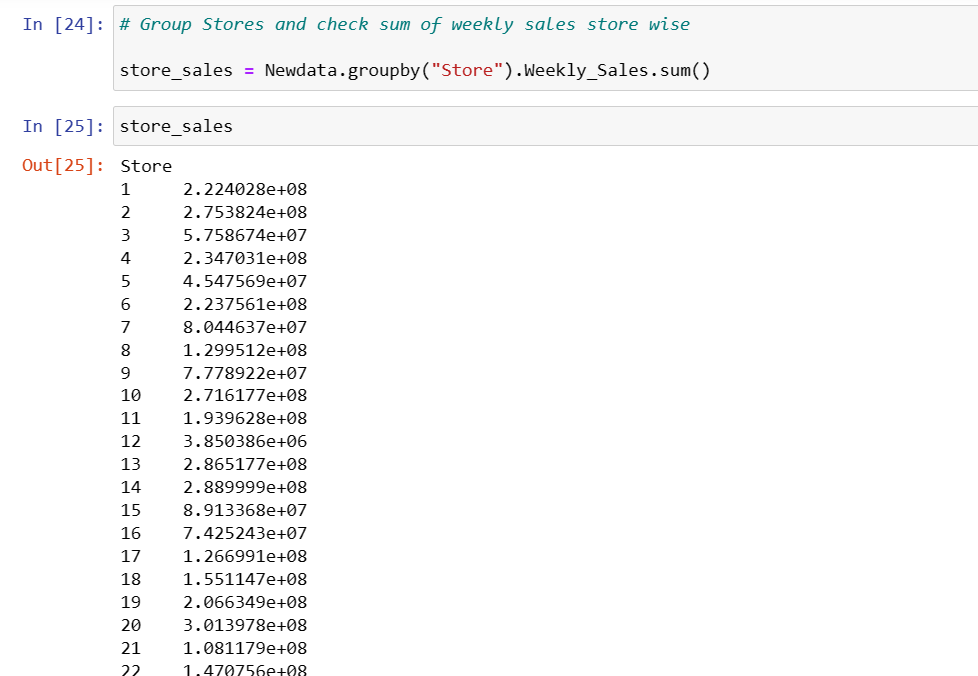


Checking Distribution of the target column

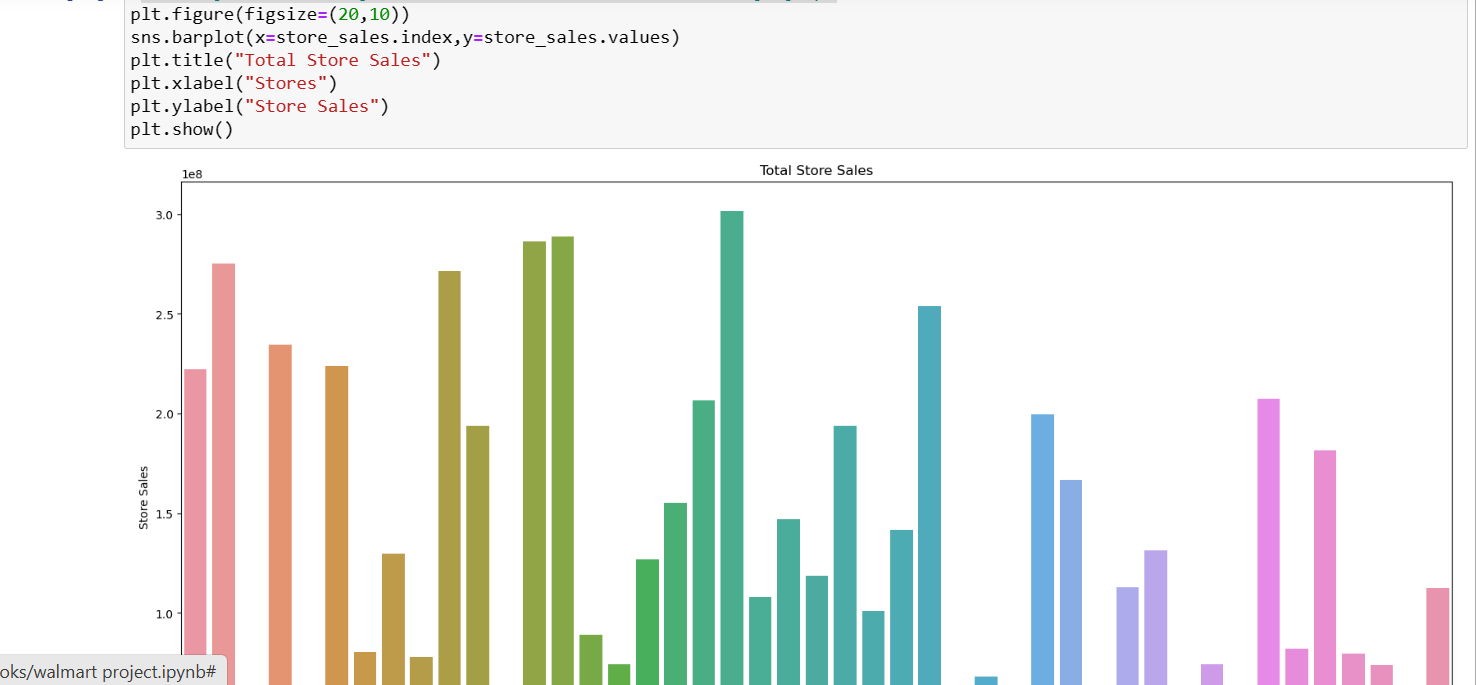


Data is not normally distributed.Its rightly skewed.

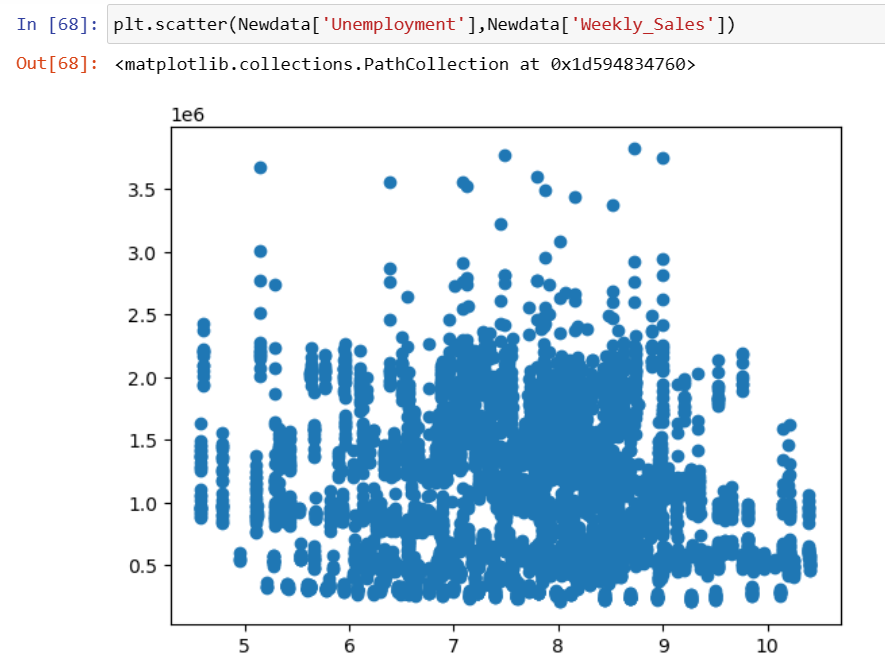
**Checking total sales store wise:**

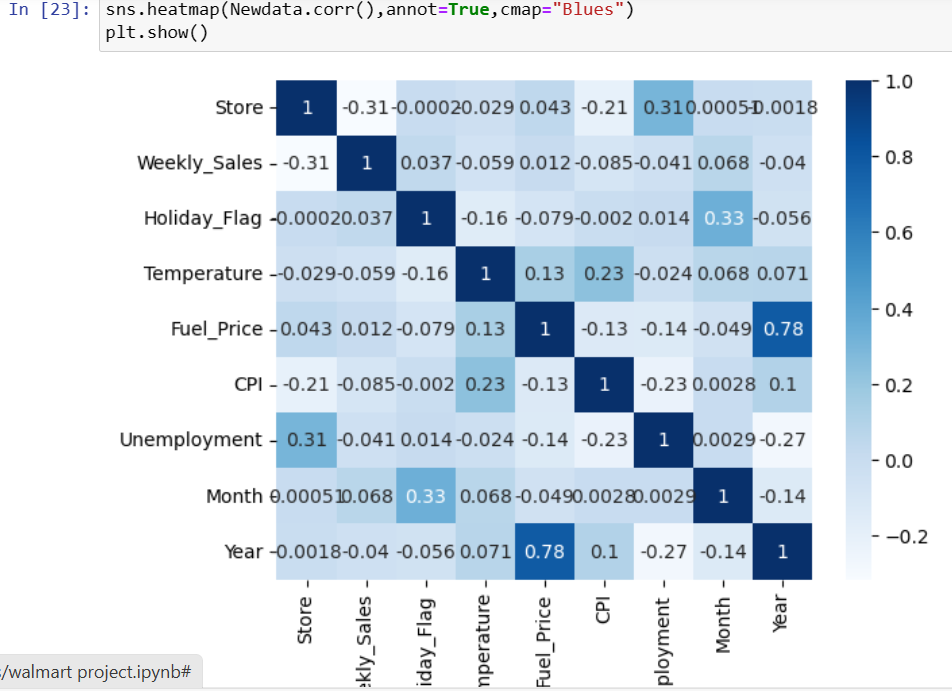


**checking store wise highest and lowest sales store wise through graph**

****

1. **Find out relationship between Unemployment and weekly sales:**





Heatmap shows a negative relationship between the sales and unemployment.

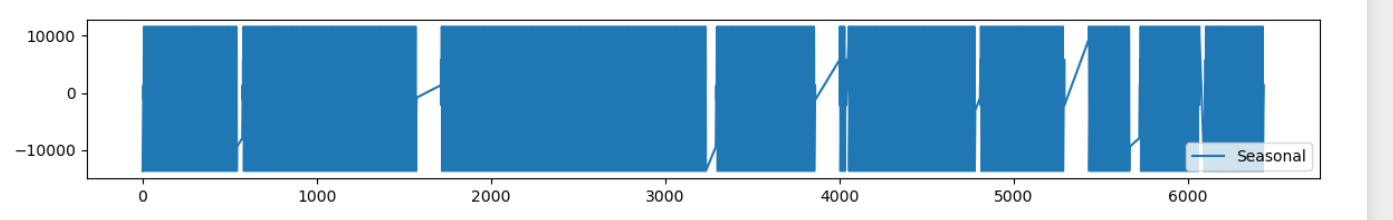
Checking the correlation between the two store wise:



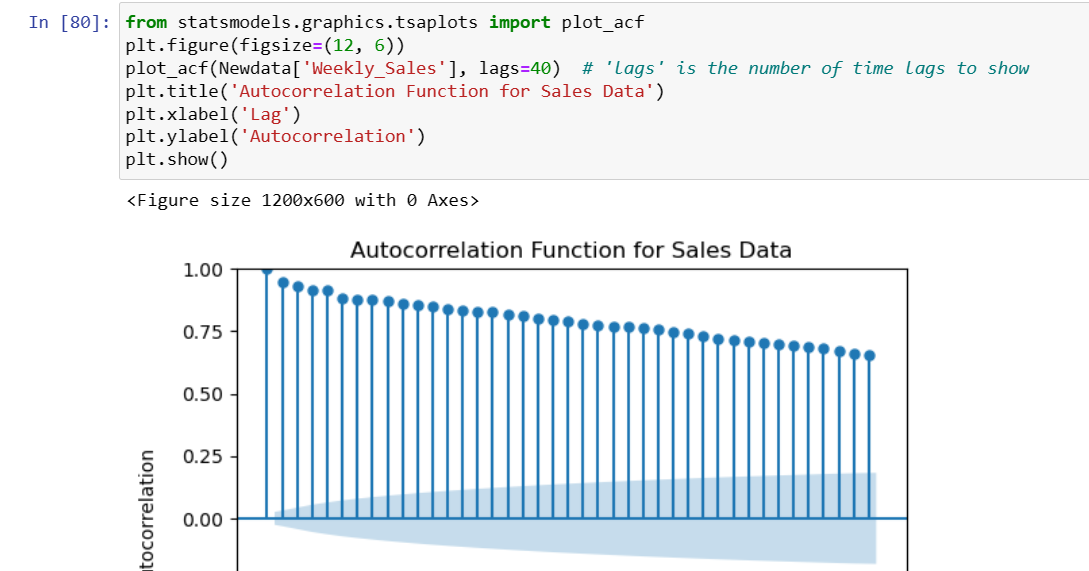
**Which store is affected the most by unemployment is-store 44**



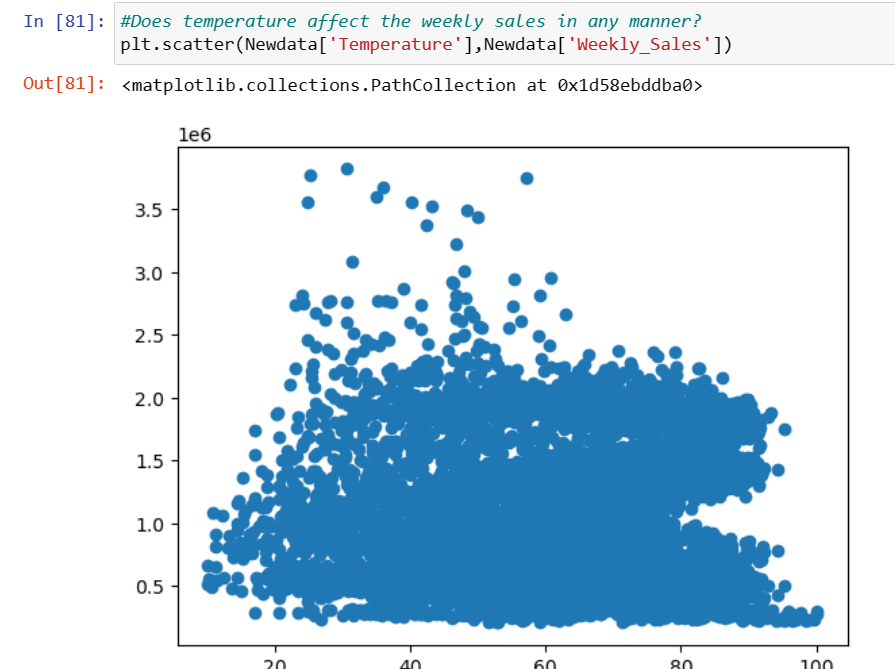
b. Checking Seasonality in Weekly Sales data:



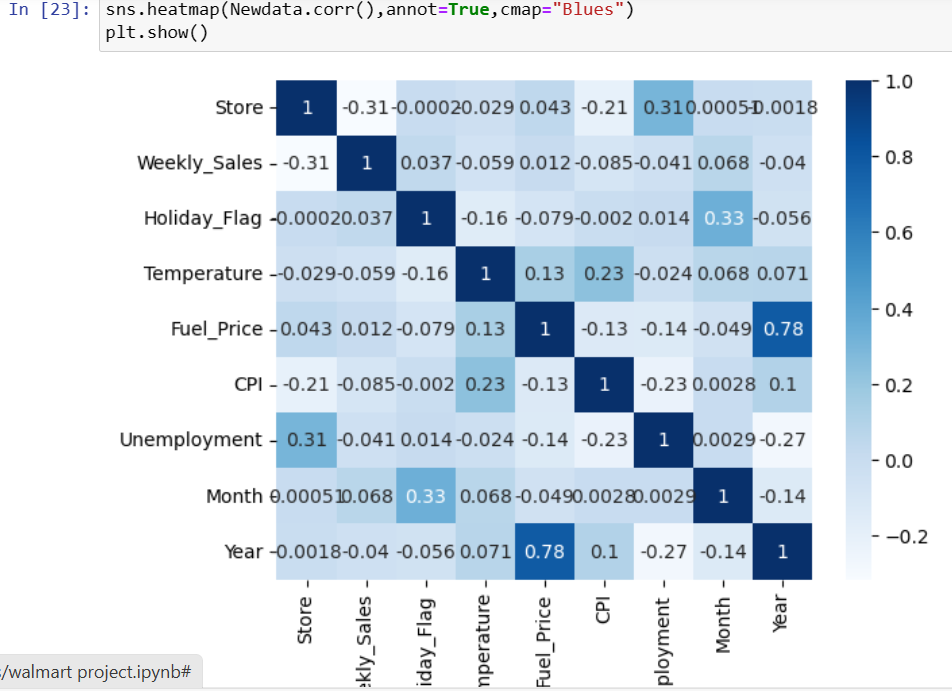
Checking seasonality through ACF Graph



**c. Does Temperature affect Weekly sales?**



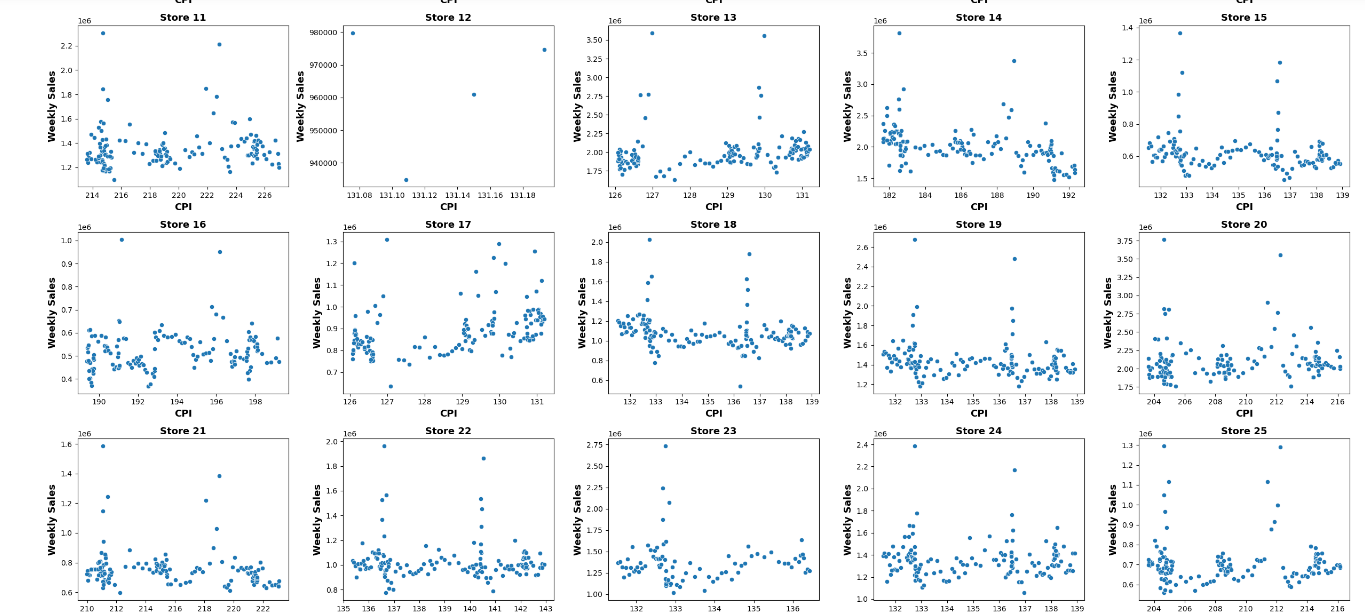
Lets check the effect of temperature on sales through heatmap.

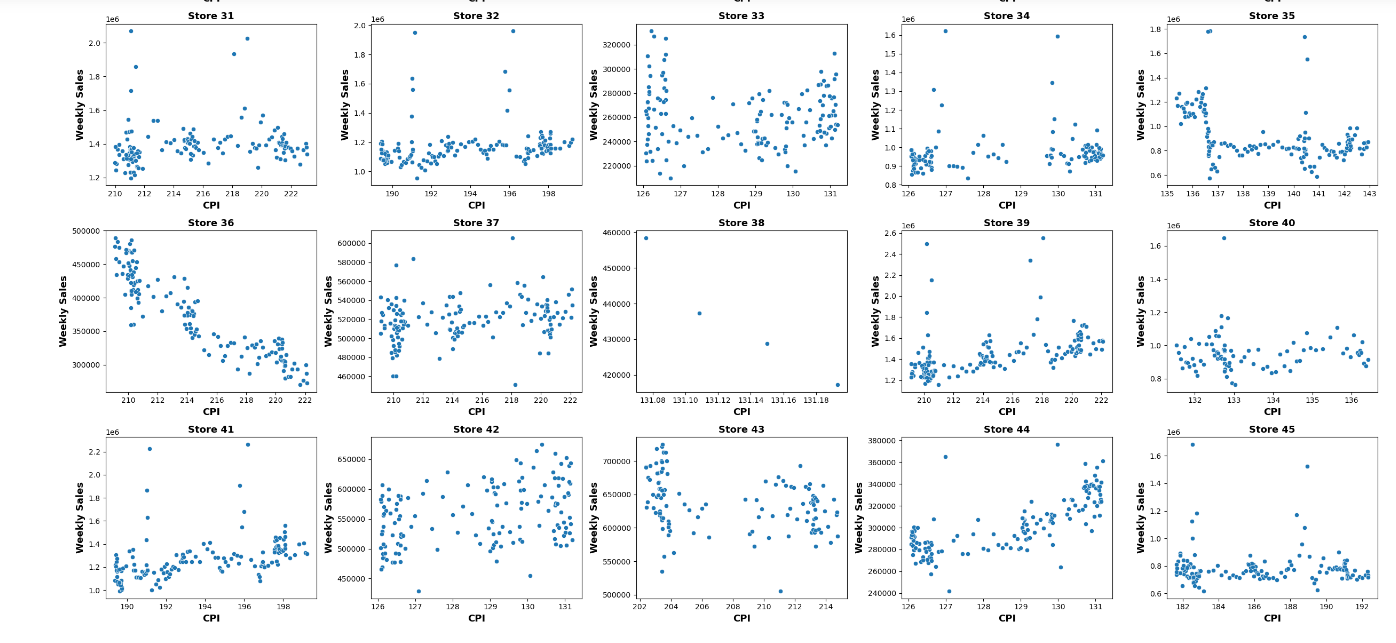


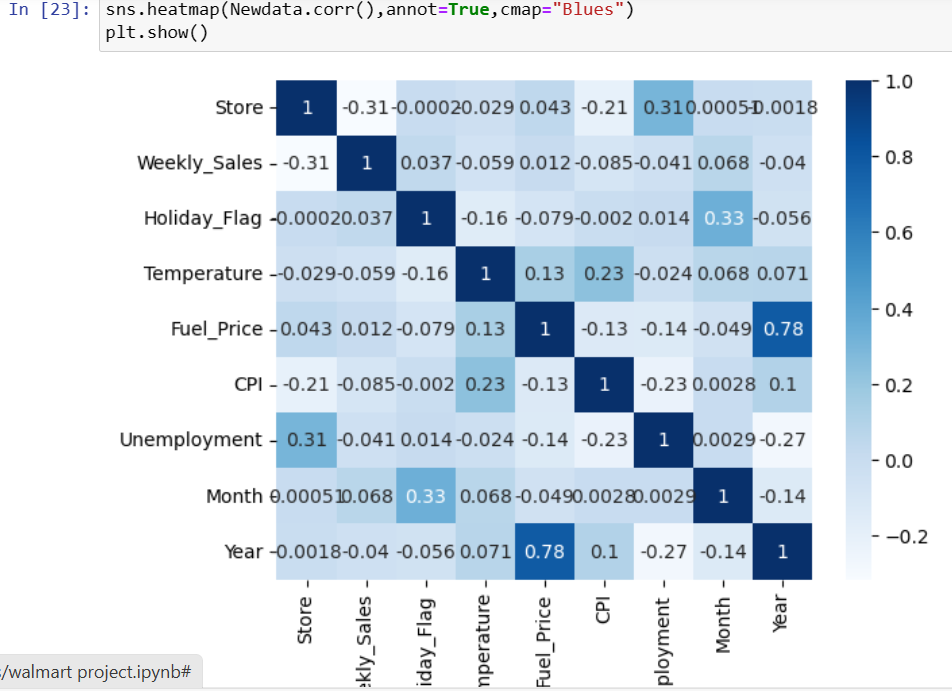
As per heatmap, there is a negative relationship between sales and temperature.

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



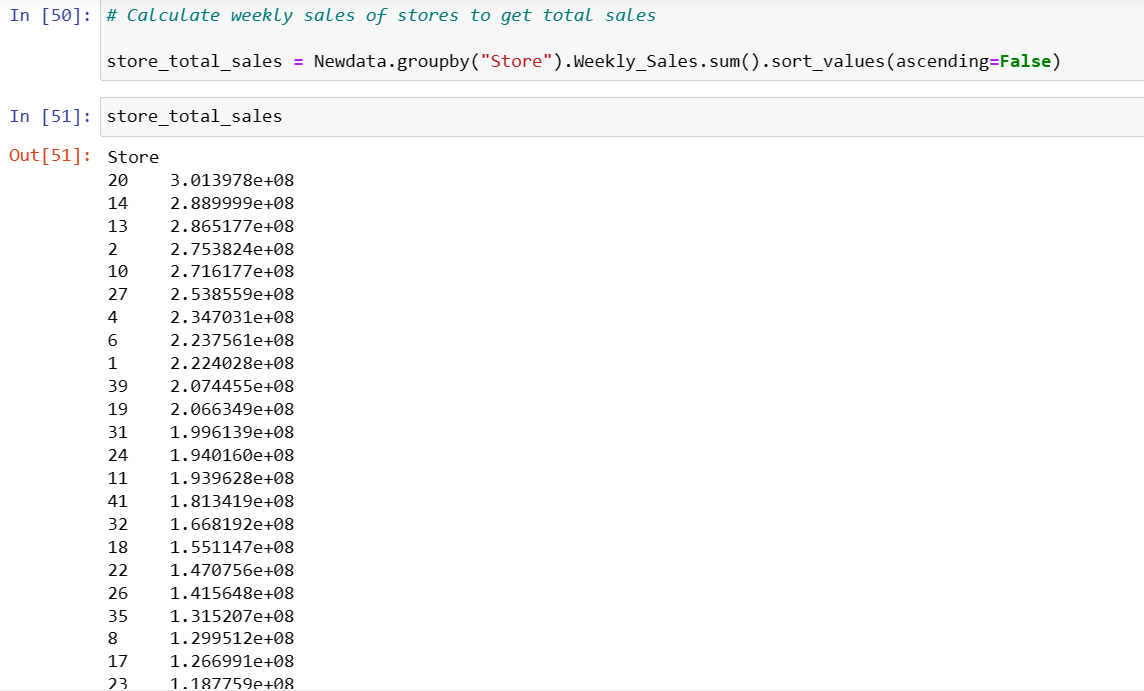


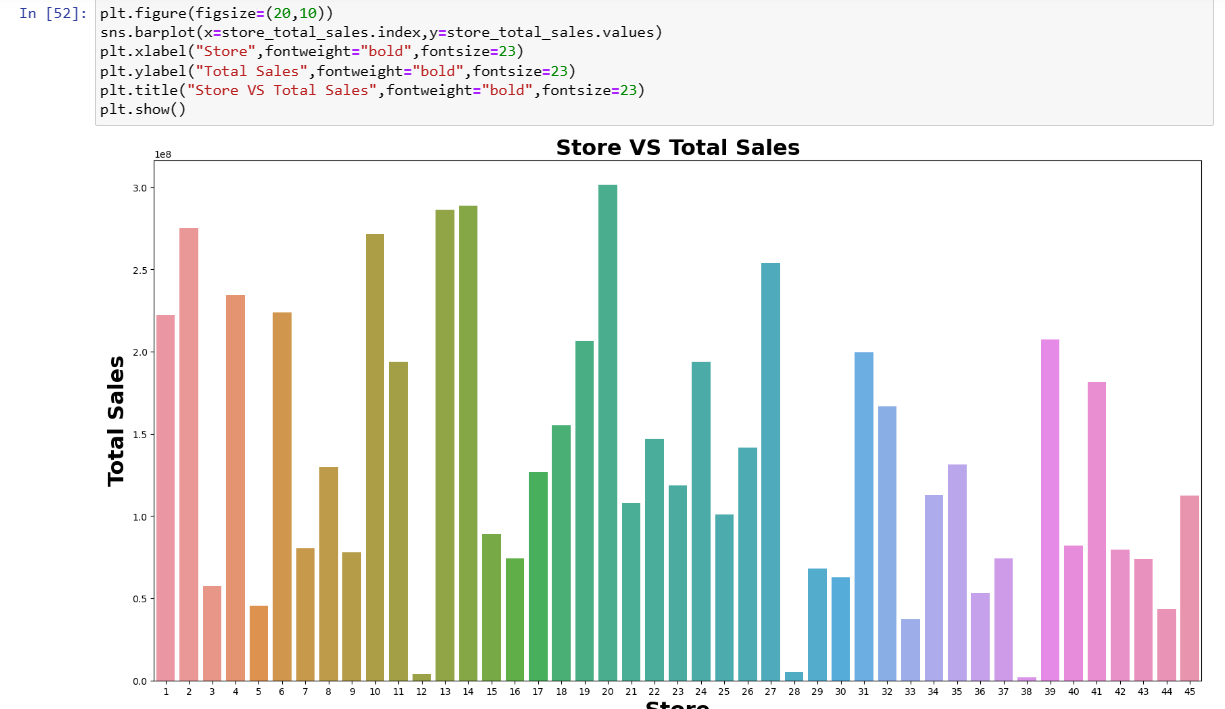


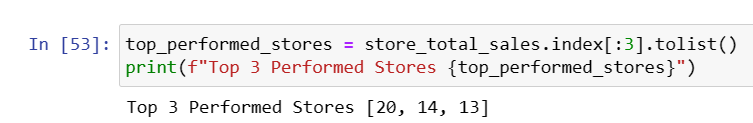


As per heatmap, there’s a negative relationship between CPI and Sales.

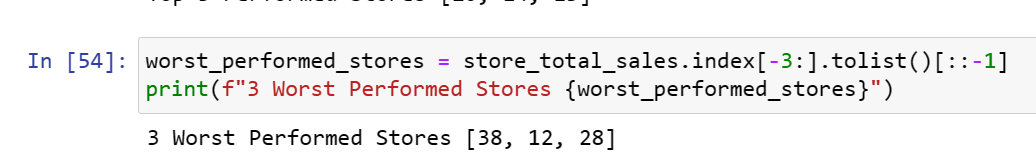
**e. Top performing stores according to the historical data-**

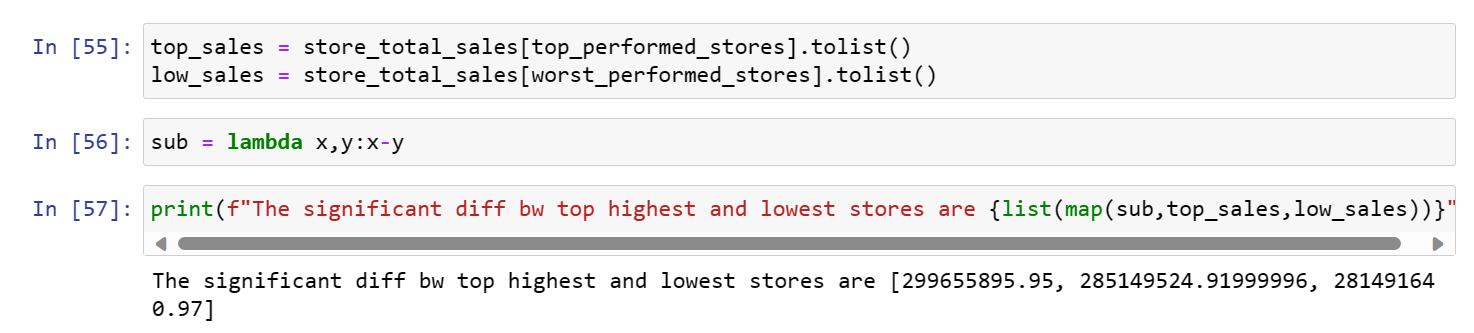






f. Worst performing stores, and difference between highest and lowest performing stores

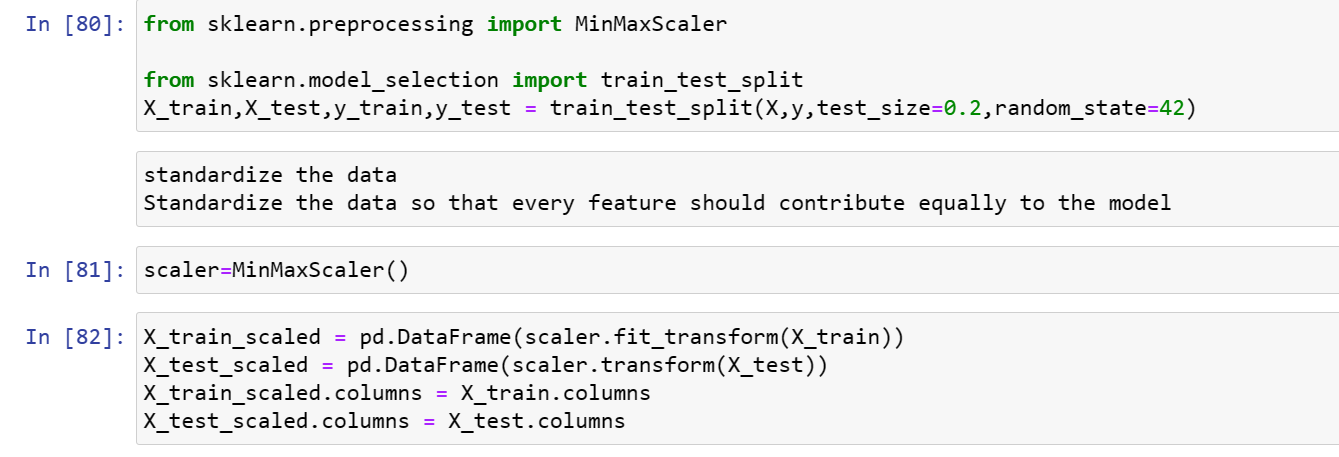




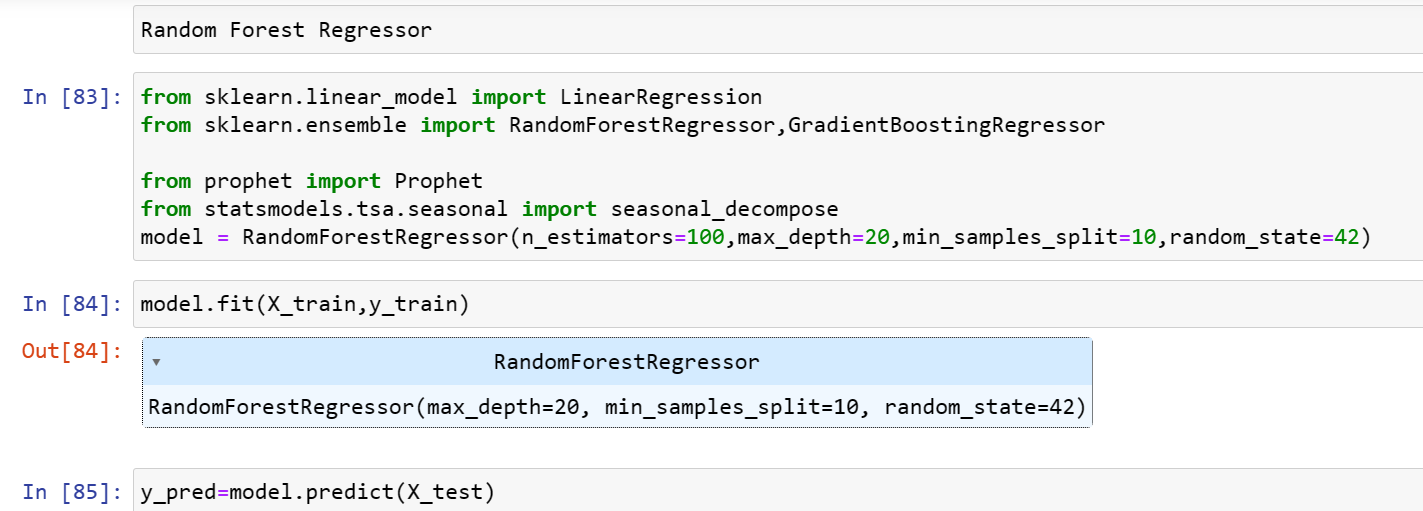
Choosing the Algorithm for the project:

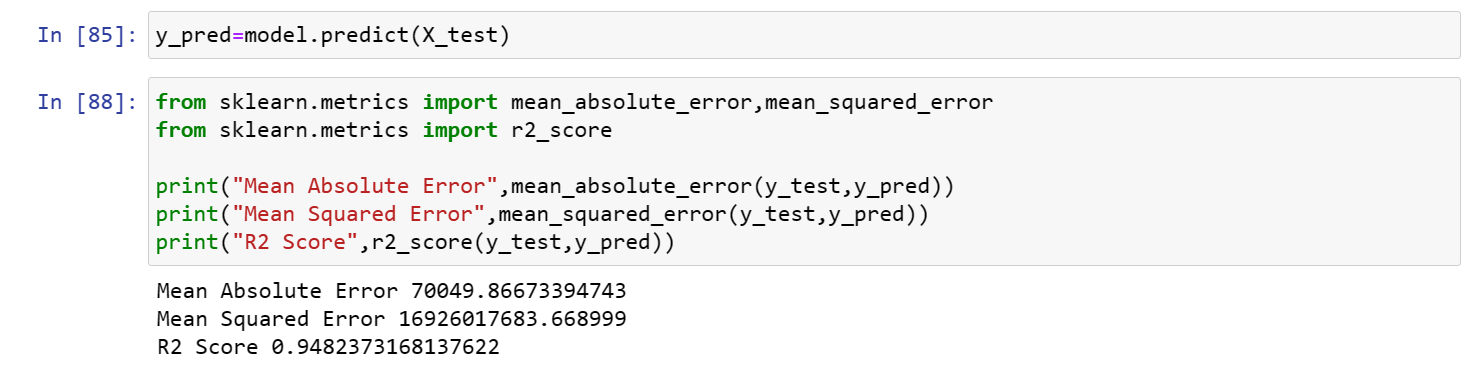
## Model Building

Scaling data



Random Forest:

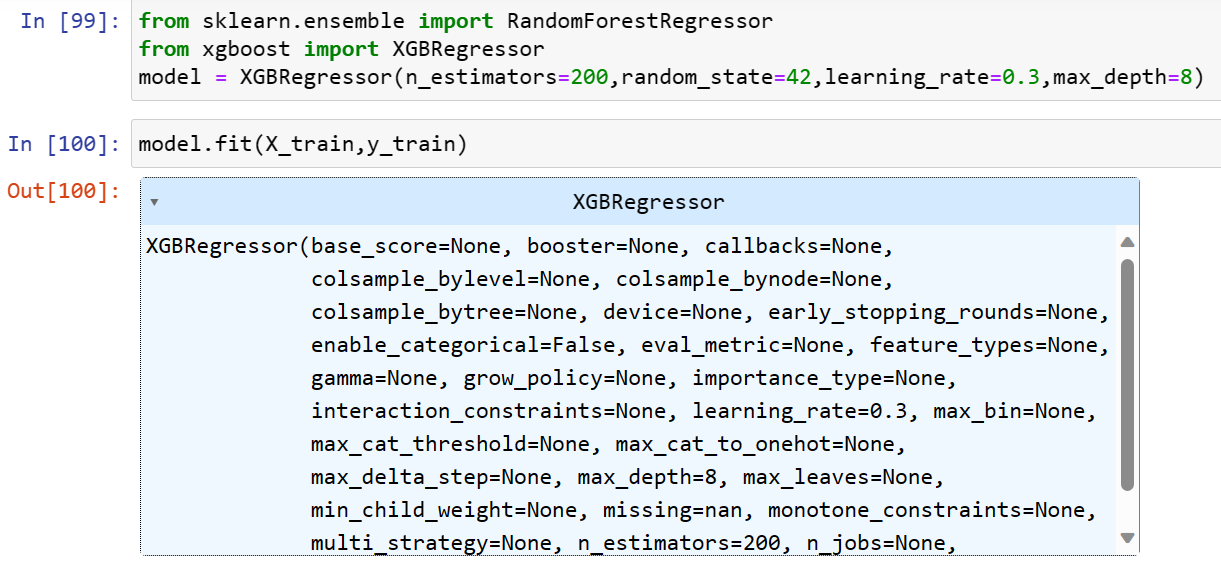


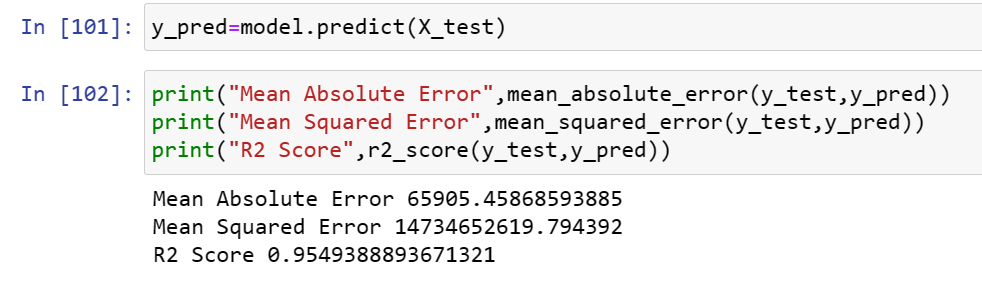


Conclusion:

Random Forest is not suitable for our problem as R2 score is very low and Errors are very high

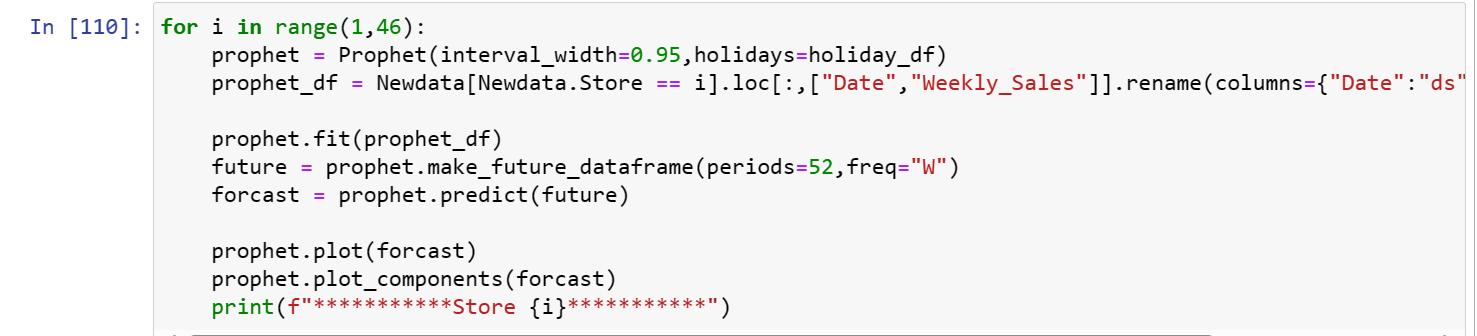
XGBoost:



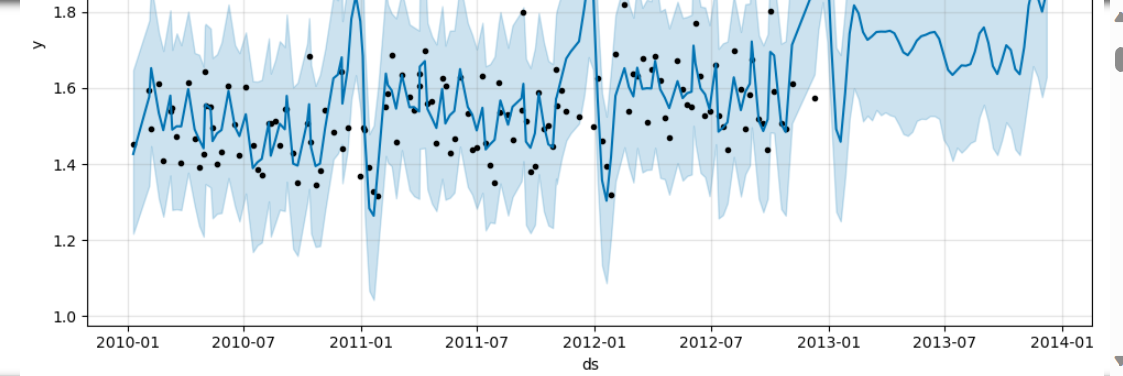


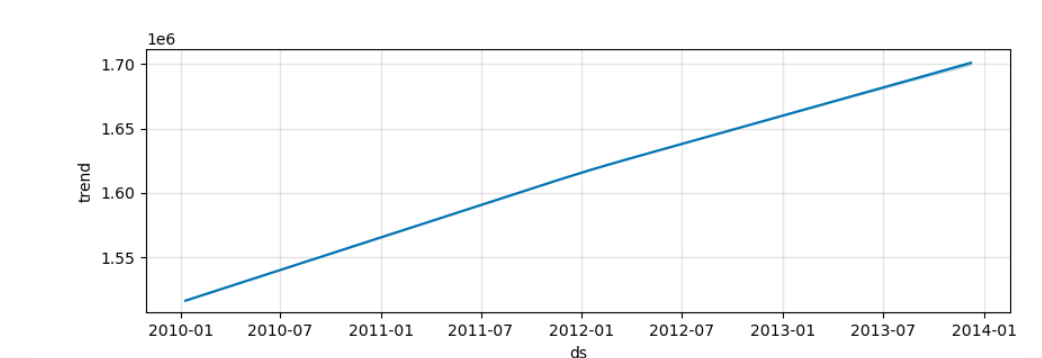
Result- XGboost is showing the highest R2 score amonst the three and lowest error.

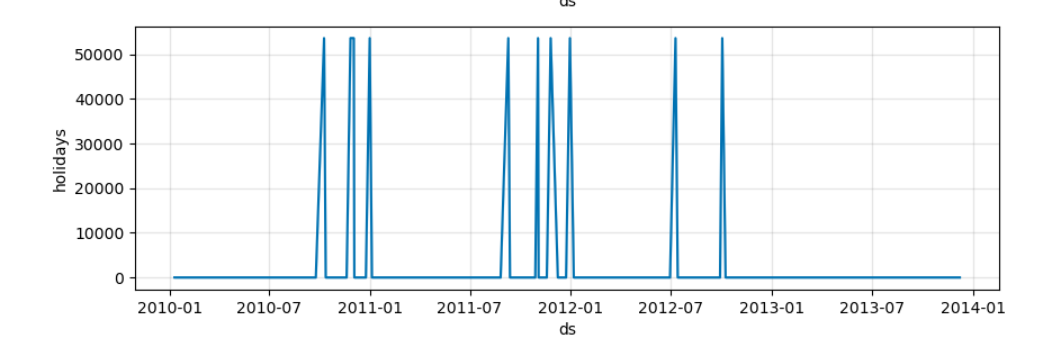
2. Use predictive modeling techniques to forecast the sales for each store for the next 12 weeks-

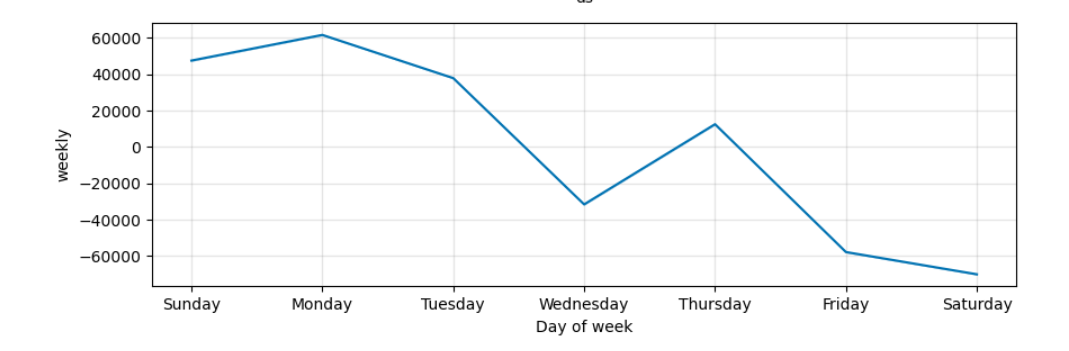


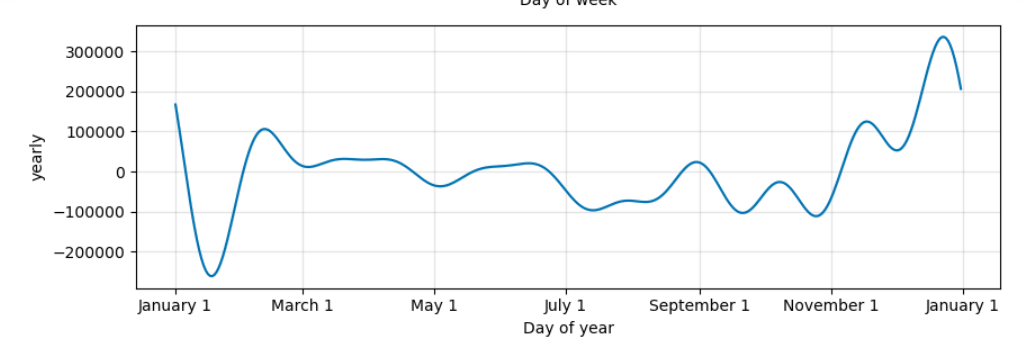
Store 45-











## Future possiblities of the project

* **Data:** If we get information like location of the store and other we may be able to forcast precisely about the weekly sales store sales.
* **Model Improvement:** I have used bagging, boosting and other models to make good perdictions.XGB (ensemble model) was the one who got good score.We can improve by using other models or neural network models which can overcome overfitting and are less sensitive to outliers.
* **Forcasting:** Models like ARIMA,SARIMA,Gated Recurrent Units,Long-short Memory can improve forcasting.

Conclusion

* **The insights which i got about walmart store wise weekly sales are:**
* Weekly sales are increasing on holiday weeks.
* Unemployment has negative relationship with weekly sales,indicates that Greater the Unemployment rate,lesser the sales.
* Consumer price index have also negative relationship with weekly sales,it can have negative impact on weekly store sales
* Temperature have also negative relationship with weekly sales.indicates temperature can have negative impact weekly store sales
* Fuel Price have positive relationship with weekly store sales
* The stores like 4, 20, 14 have performed great all the time and stores like 33, 44, 5 have performed worst all the time.This can help walmart to make better decisions about the stores.While forcasting most of the stores have upward trend.
* By using Prophet model,we are able to capture most of predictions.
* By checking how Unemployment is changing over the time,we found Unemployment has downward trend.This helps us to understand that future sales may be increased or stores affected by Unemployment will imporve the weekly sales.
* By checking yearly sales,2012 have less sales than preivous year.

References:

<https://youtu.be/KvLG1uTC-KU>

<https://www.bing.com/videos/riverview/relatedvideo?&q=xgboost&&mid=7F0393A02A0A12A5F0777F0393A02A0A12A5F077&&FORM=VRDGAR>

<https://analyticsvidya.com/>