# **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 the Algorithm for the Project

6. Motivation and Reasons for Choosing the Algorithm

7. Assumptions

8. Model Evaluation and Techniques

9. Inferences from the Same

10. Future Possibilities of the Project

11. Conclusion

# **Problem Statement**

Walmart is one of the largest and most well-known retail corporations in the world that operates a chain of hypermarkets, so as a huge corporation it has been facing some problems.

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.

There are two problems which we have to analyse that are:

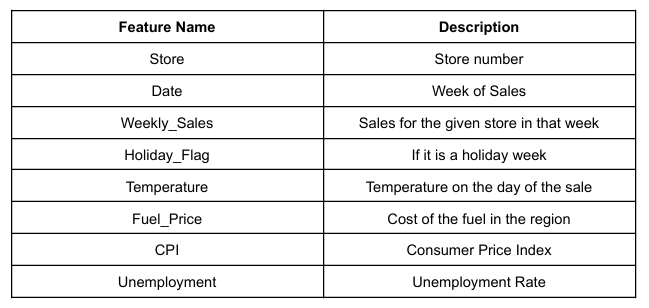
1. You are provided with the weekly sales data for their various outlets. Use statistical analysis, EDA, outlier analysis, and handle the missing values to come up with various insights that can give them a clear perspective on the following: a. If the weekly sales are affected by the unemployment rate, 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 modelling techniques to forecast the sales for each store for the next 12 weeks.

# **Project Objective**

The objective of this project is to analyse the insight trends and statistical analysis on the basis of weekly sales and build a model to forecast the sales for each store for the next 12 weeks.

# **Data Description**

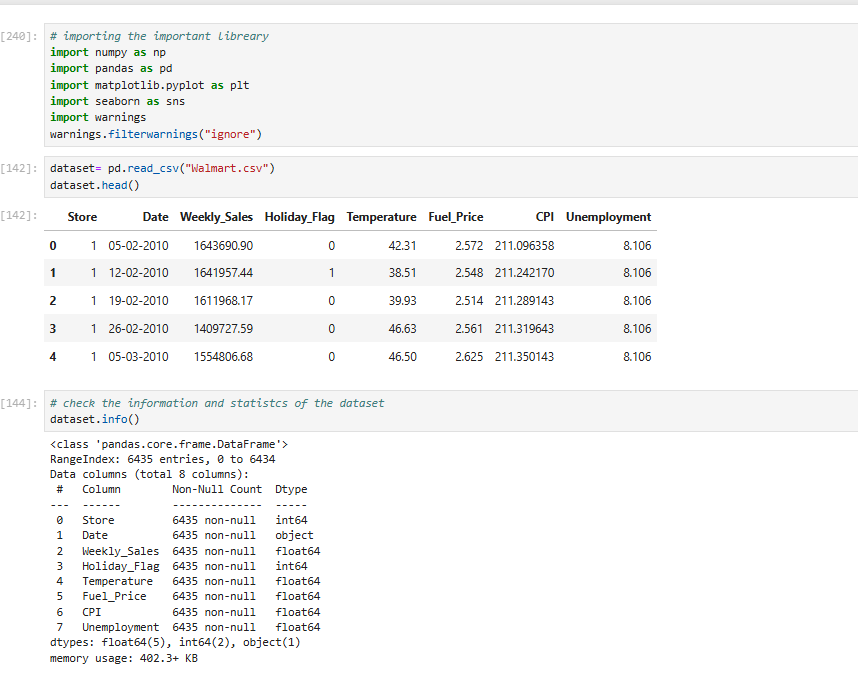
The Datasets for this project are define below:

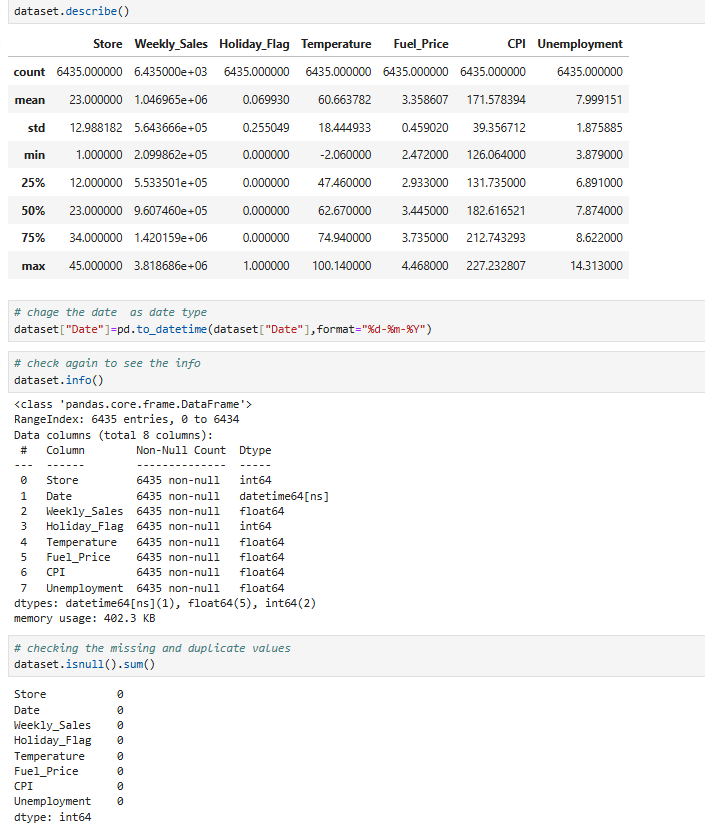


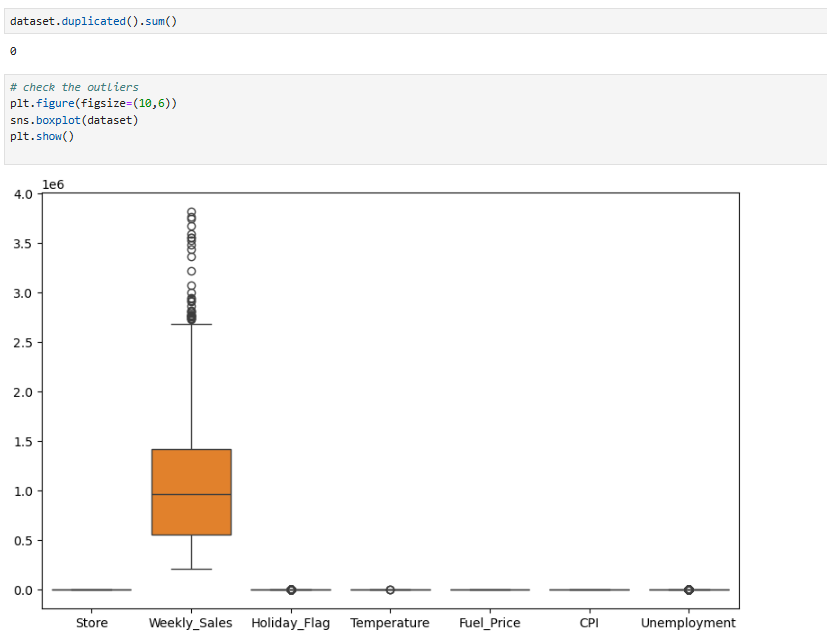
# **Data Pre-processing Steps and Inspiration**

We need to do some cleaning process and preprocessing steps before the model building. the steps are below:

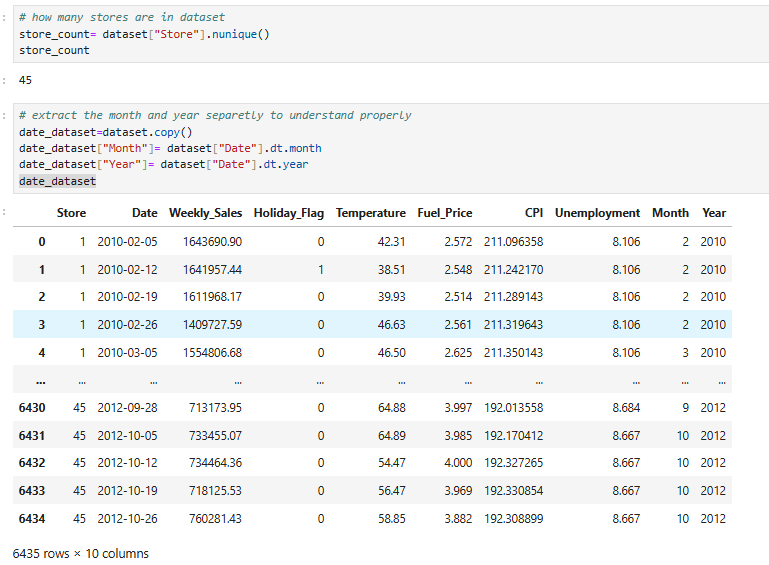
# From importing the libraries to importing the dataset. Checking the null values and analyse the dataset.







# Analysis the dataset in different aspect

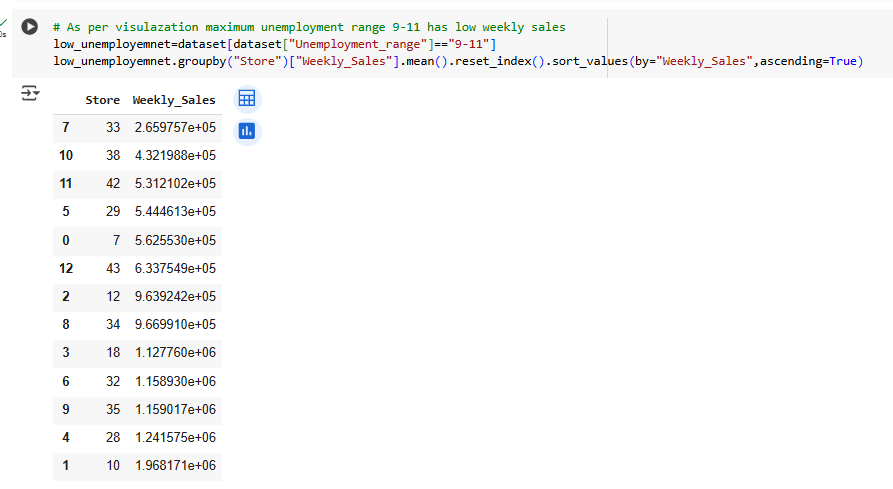


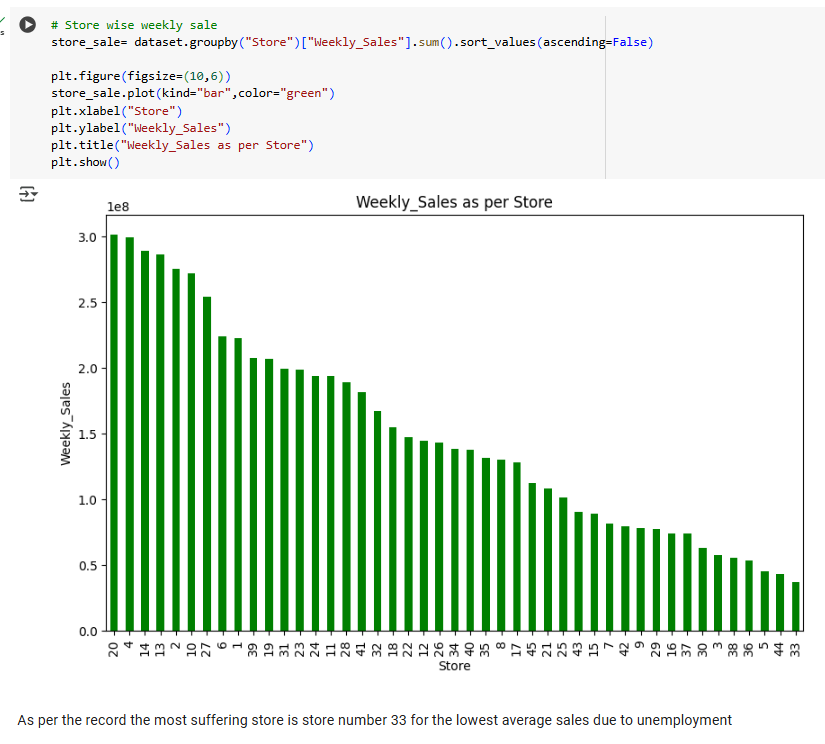
# Various types of data analysis for the insight that can give a clear perspective on the following:

1. If the weekly sales are affected by the unemployment rate, if yes - which stores are suffering the most:



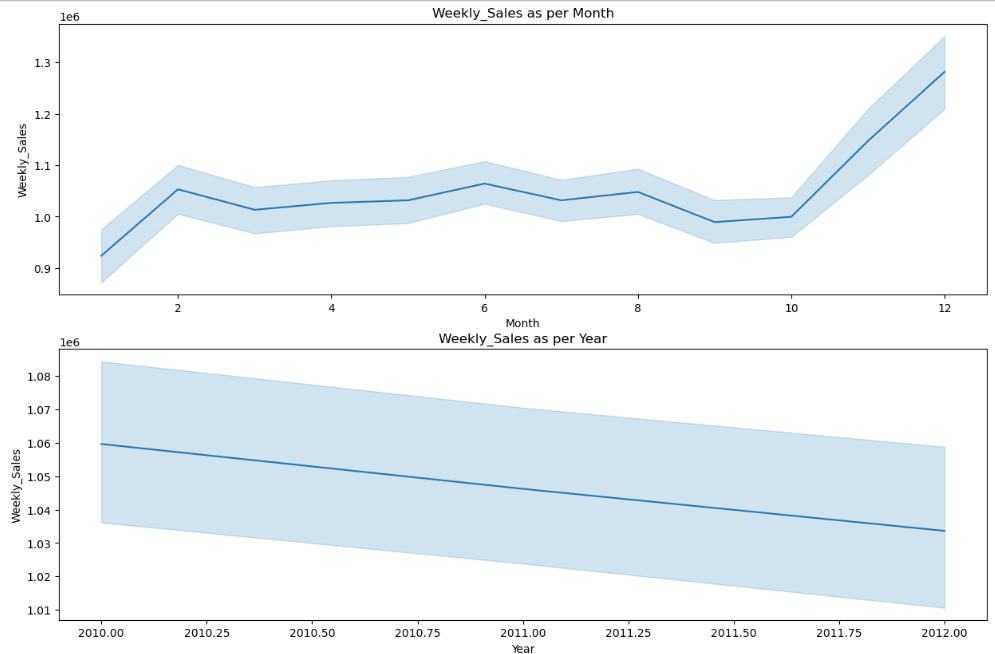


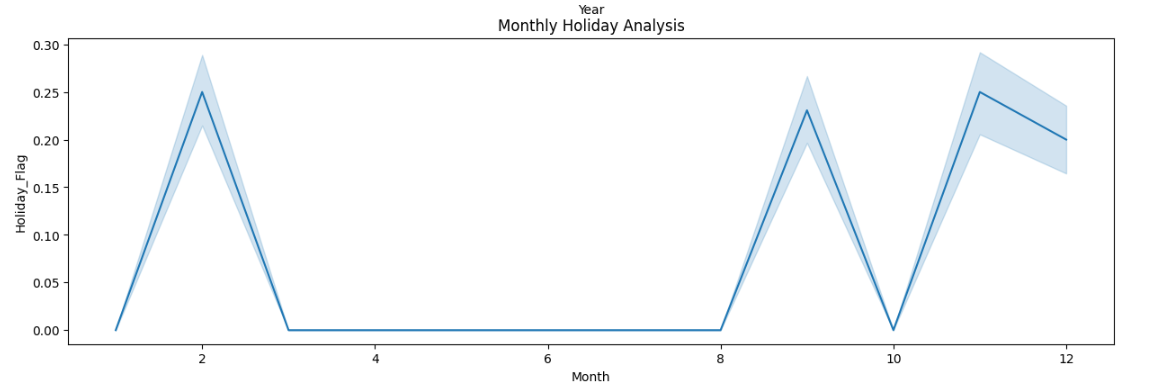




1. If the weekly sales show a seasonal trend, when and what could be the reason:

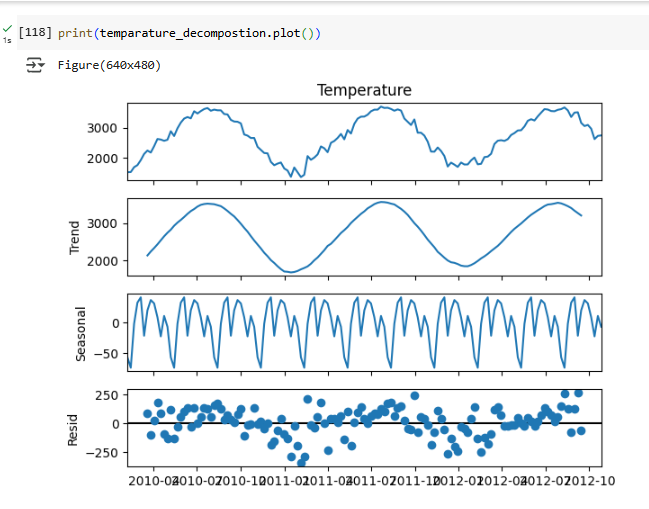


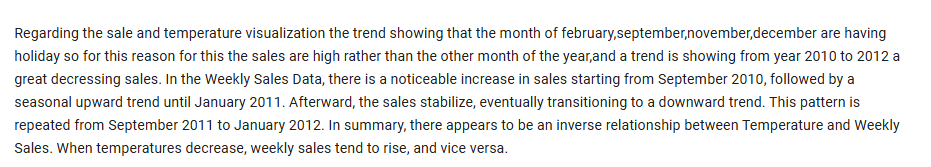






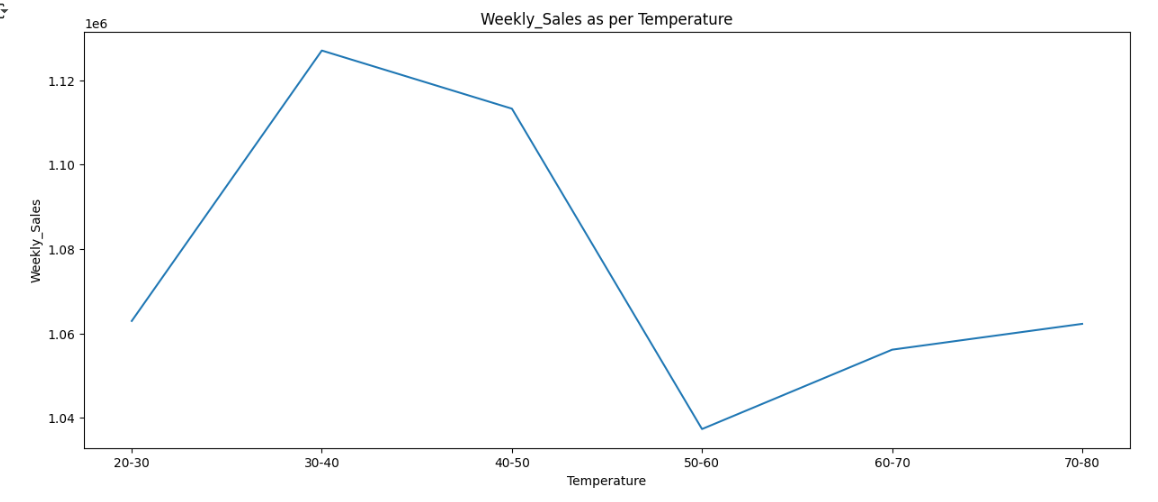


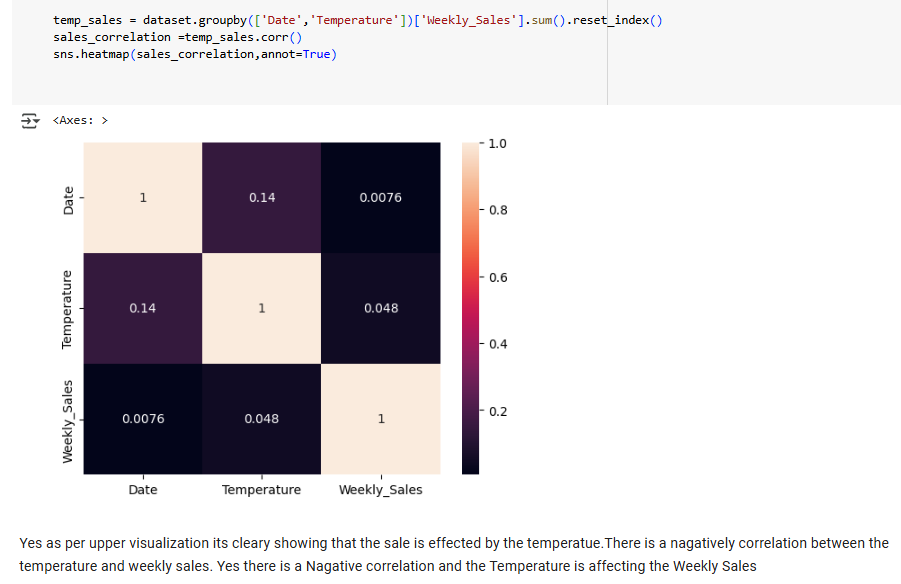




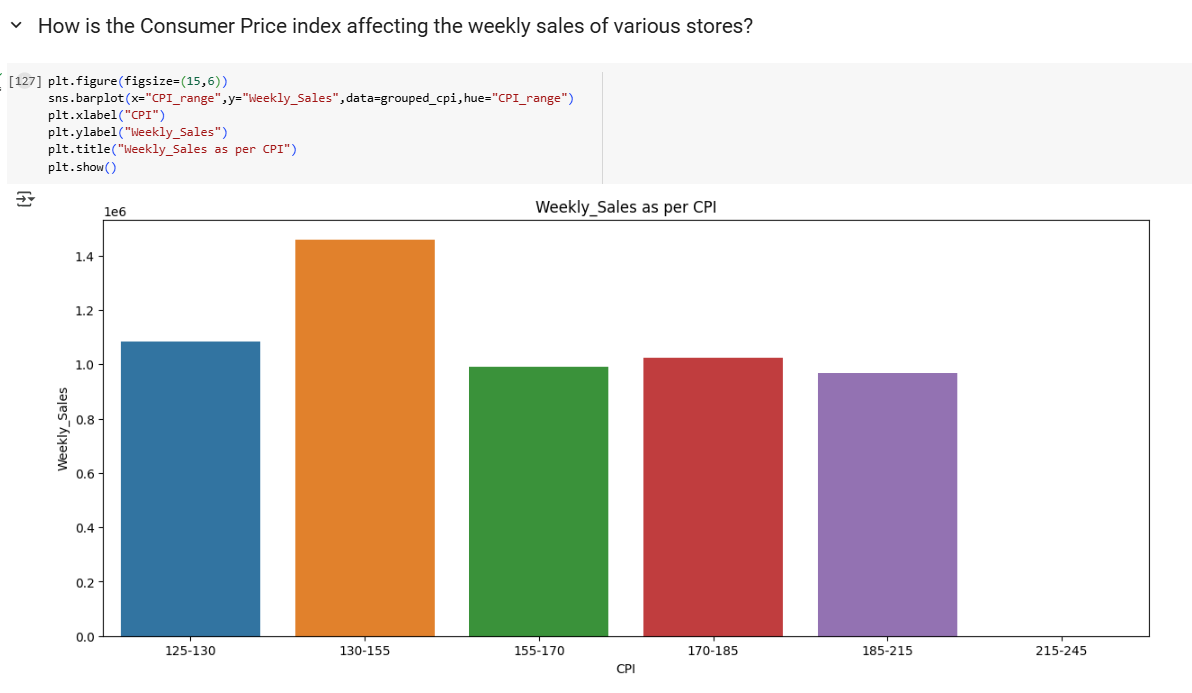
1. Does temperature affect the weekly sales in any manner?

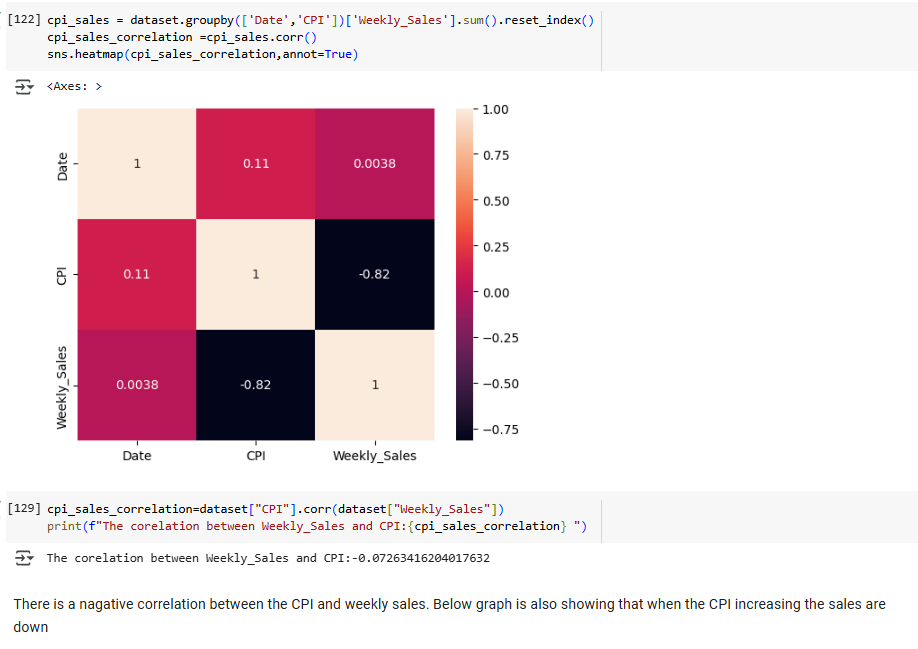






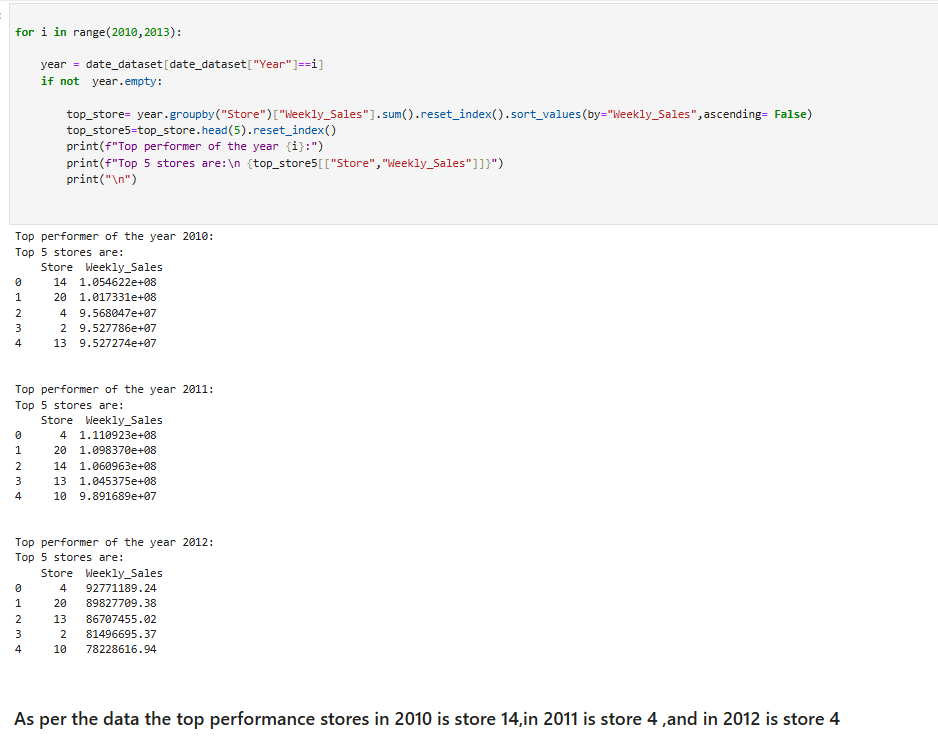
1. How is the Consumer Price index affecting the weekly sales of various stores:





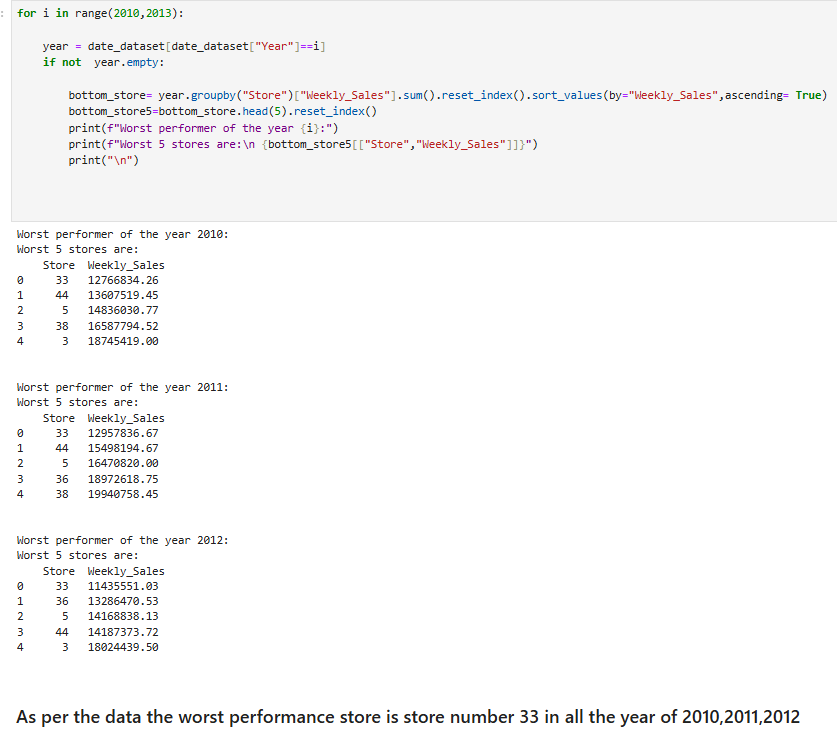


1. Top performing stores according to the historical data.



1. The worst performing store, and how significant is the difference between the

highest and lowest performing stores.

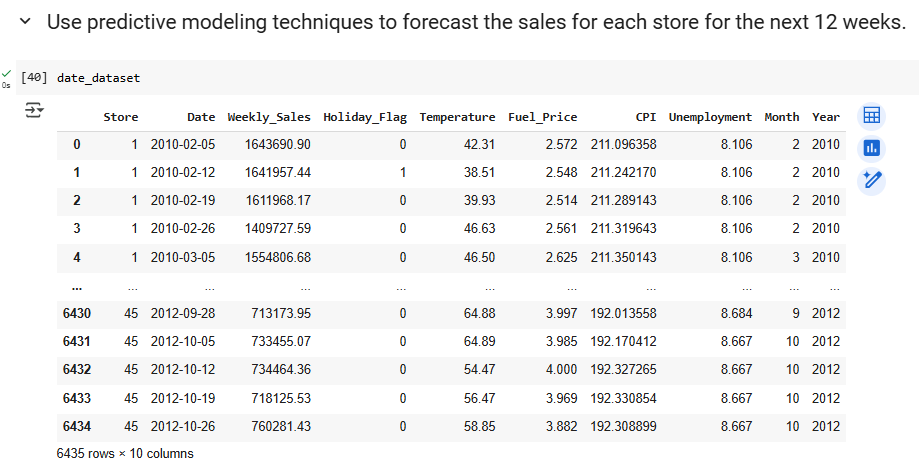


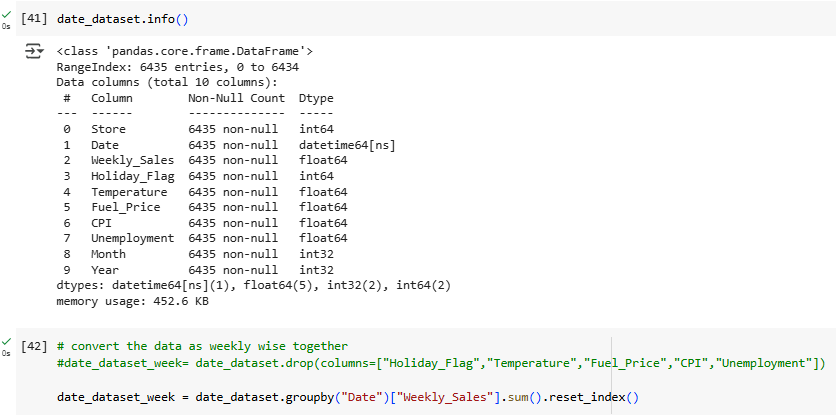


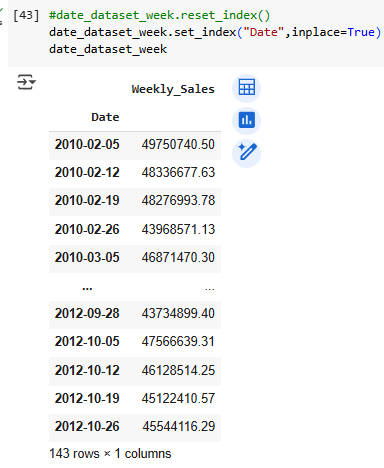
# There are few visualization with the weekly sales and other attributes:

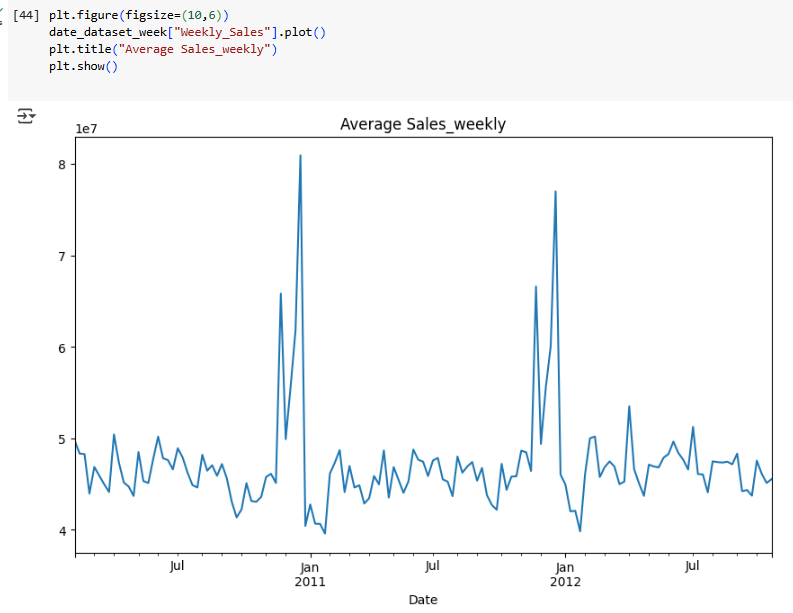


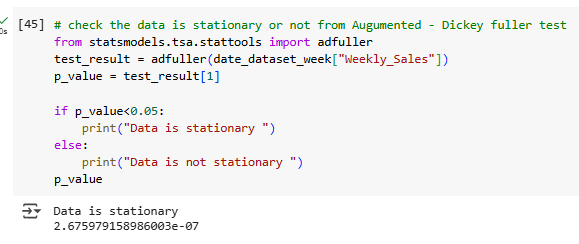
# Use predictive model to forecast the sales for each store for the next 12 week





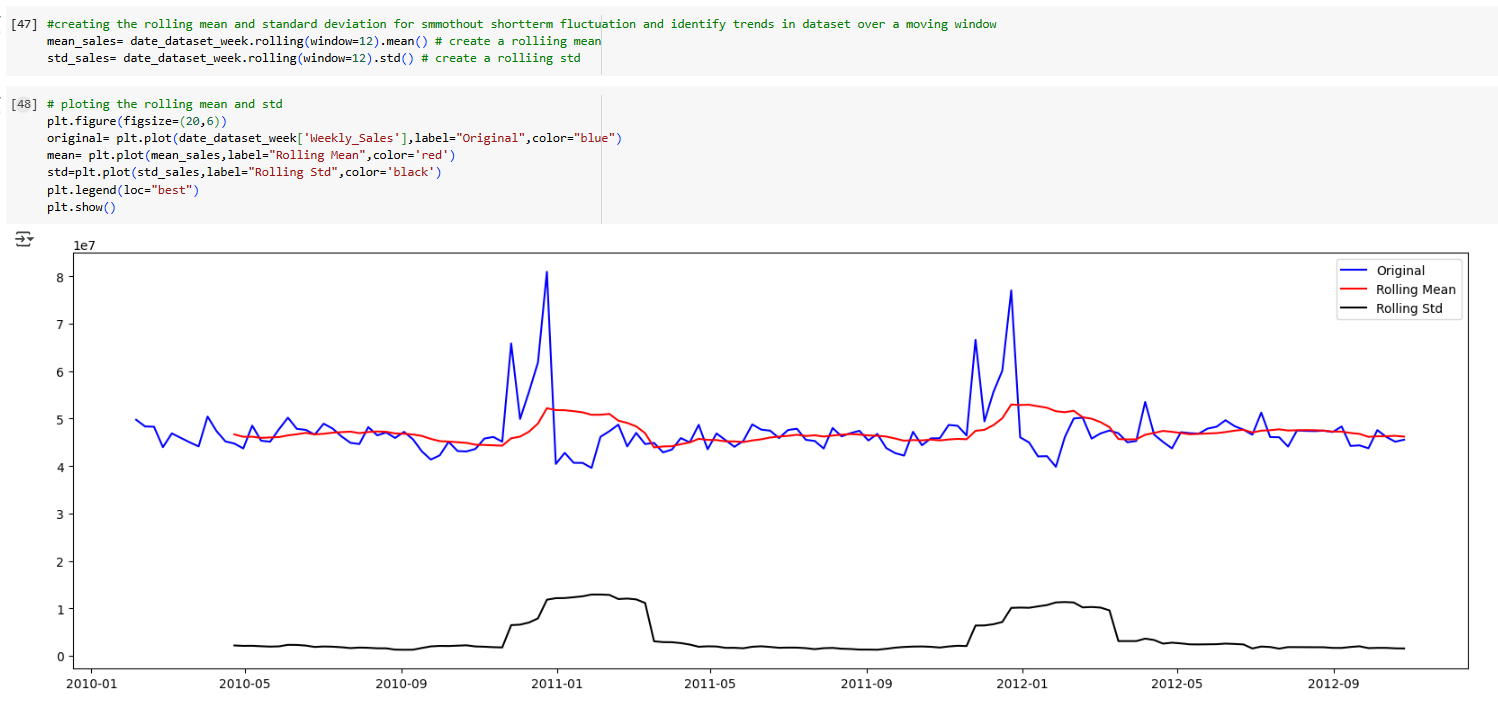


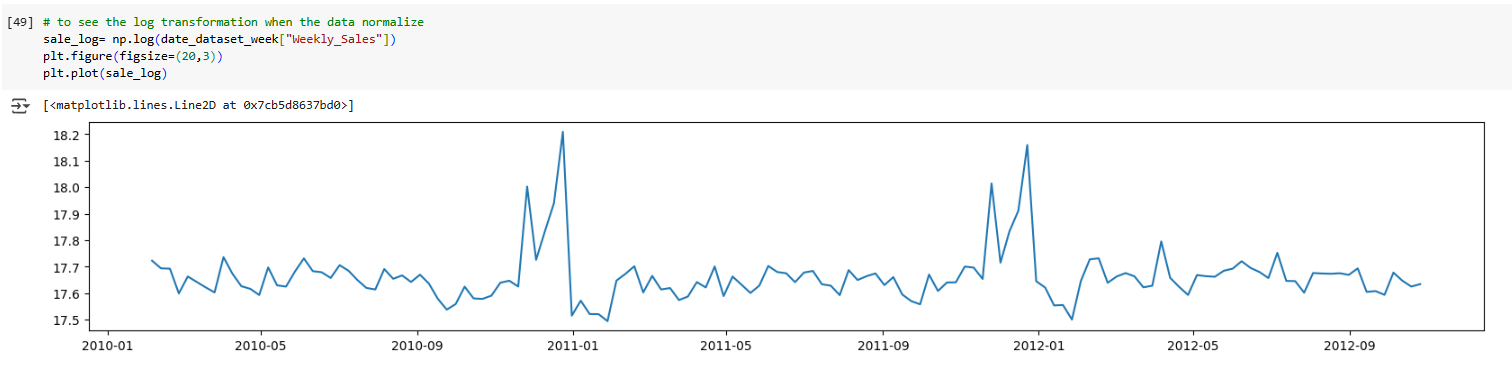




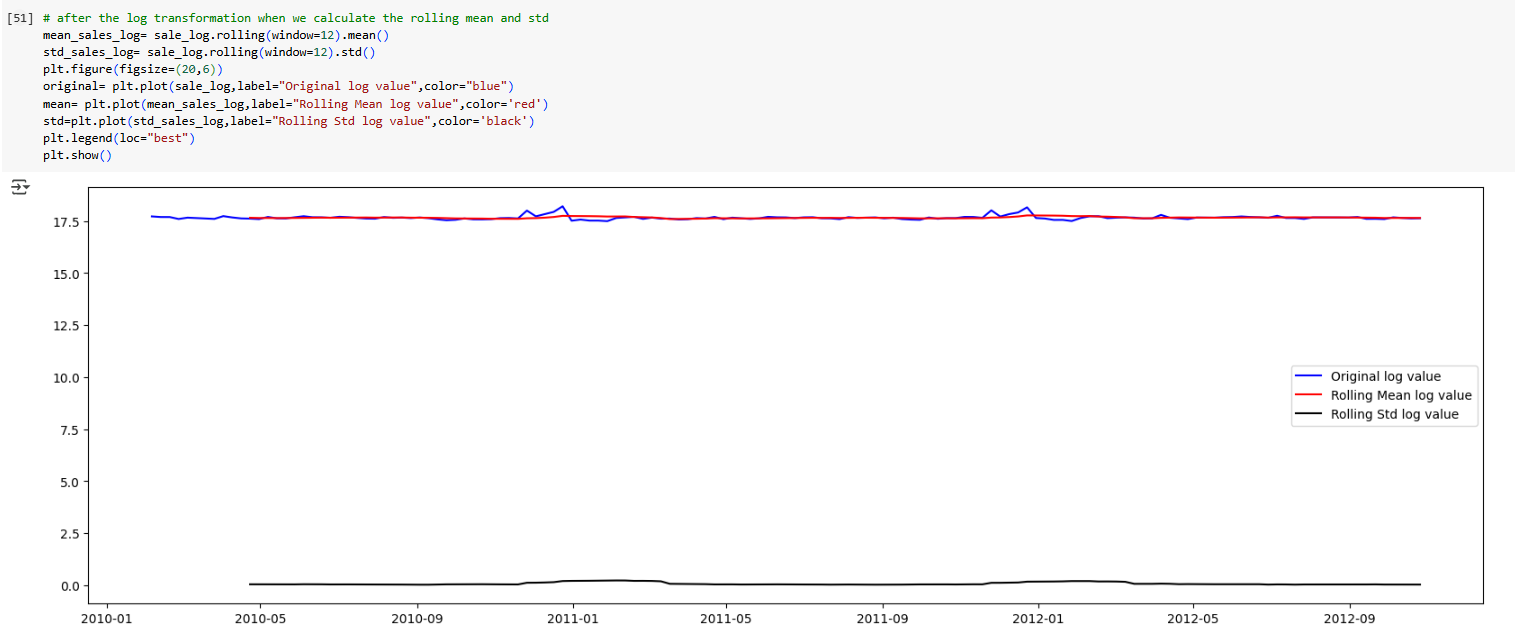
# 

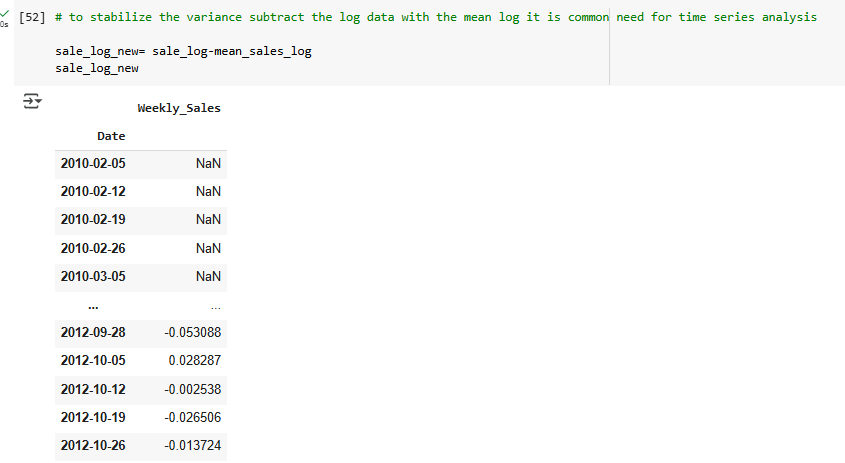
# Update rolling mean and standard deviation to predict the time series better

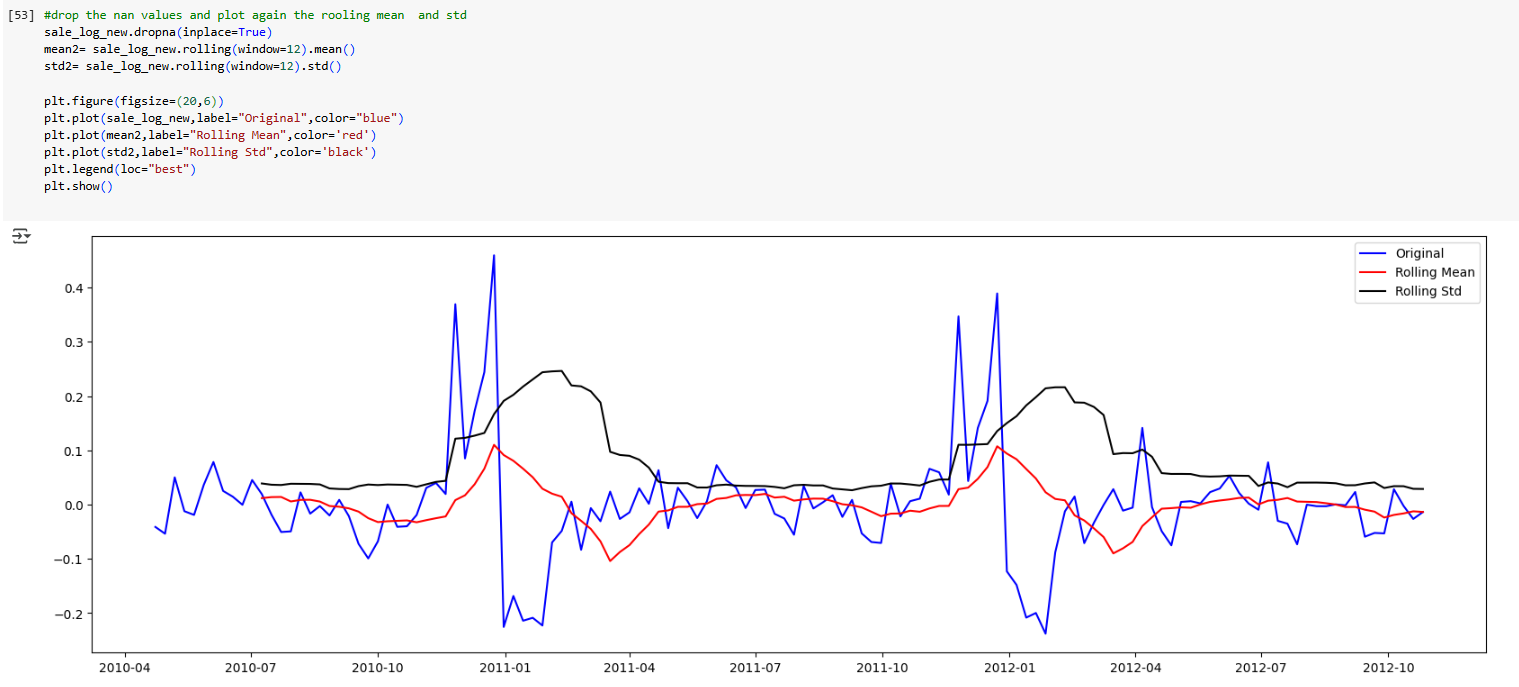


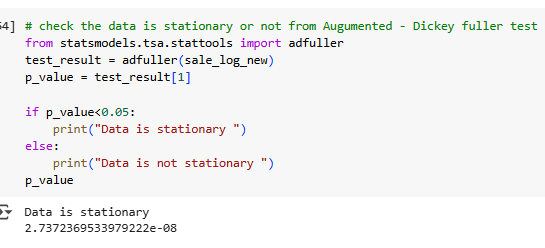


# 

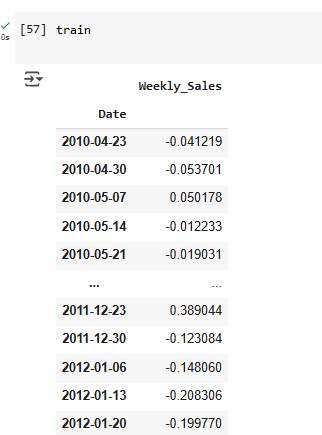


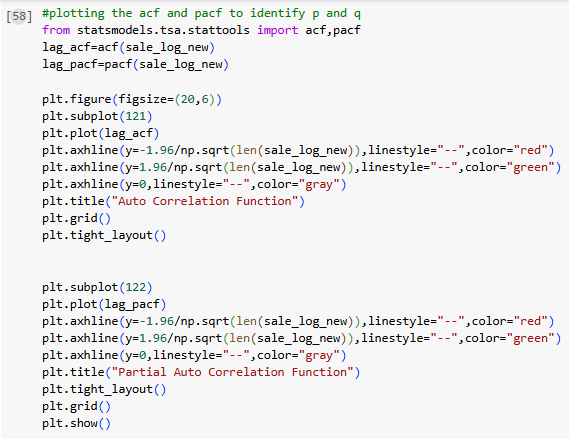


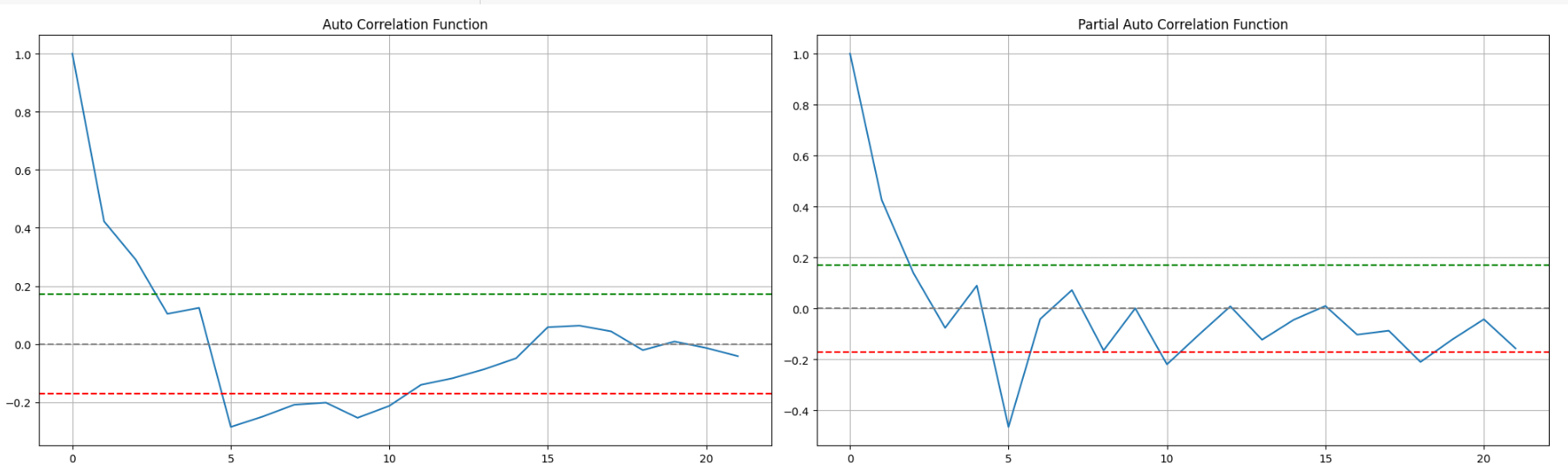


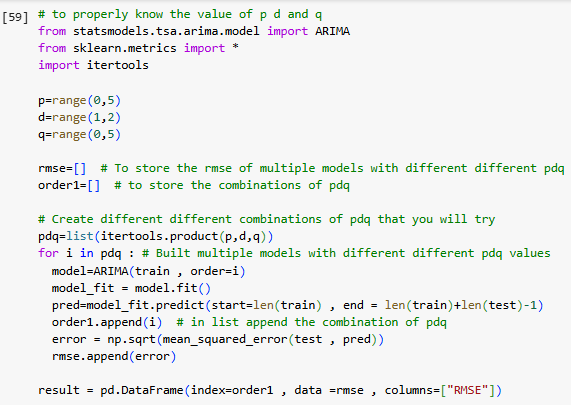


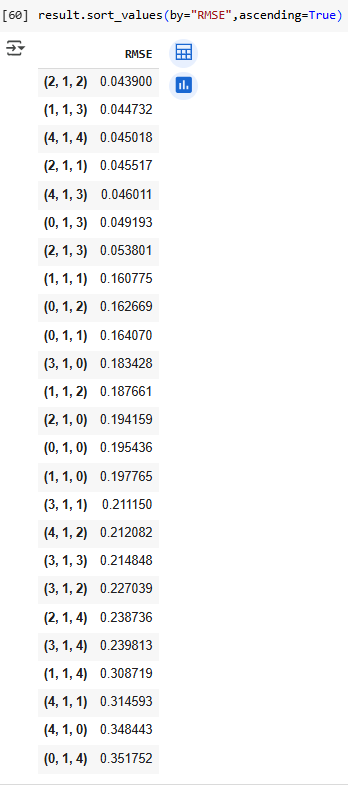
# 





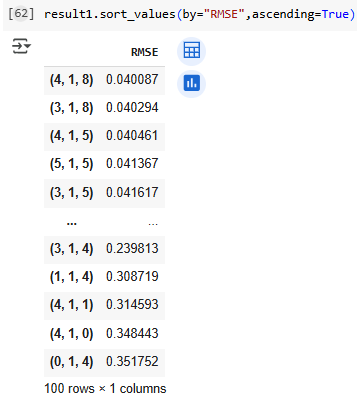






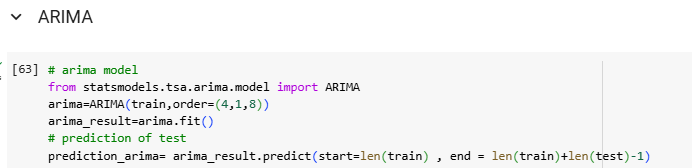
# To get more optimize RMSE we can range the pdq more

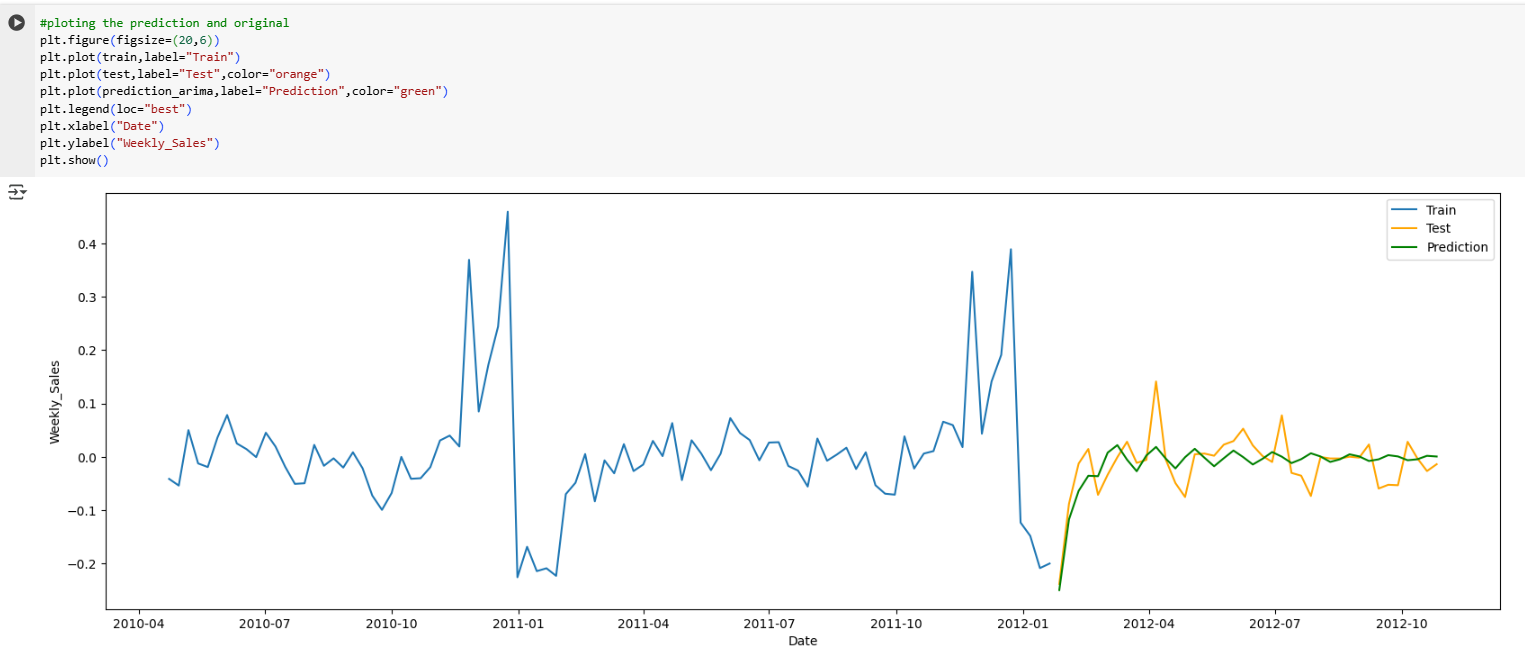
# 

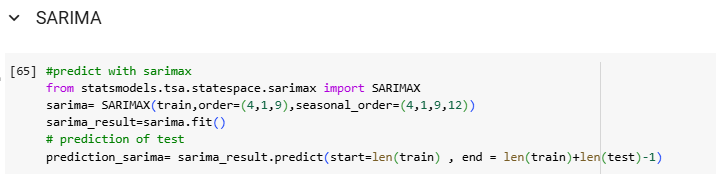


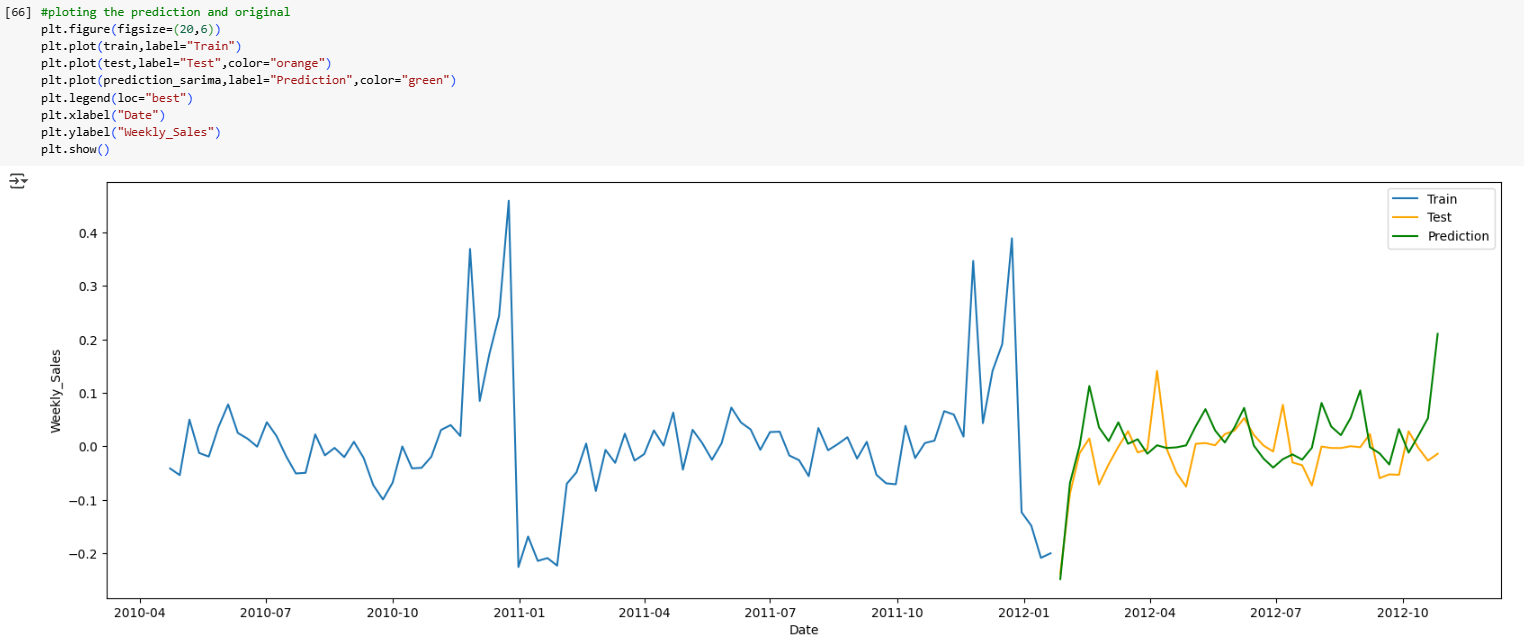
**Choosing the Algorithm for the Project**

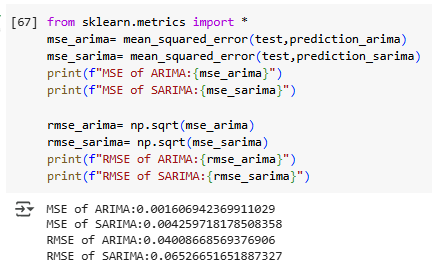
There are two algorithm ARIMA and SARIMA lets see which one is better



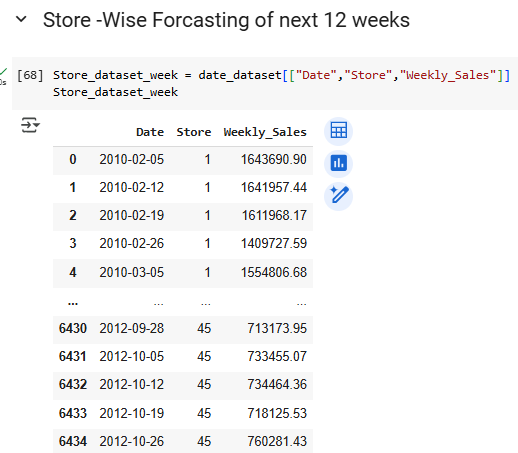


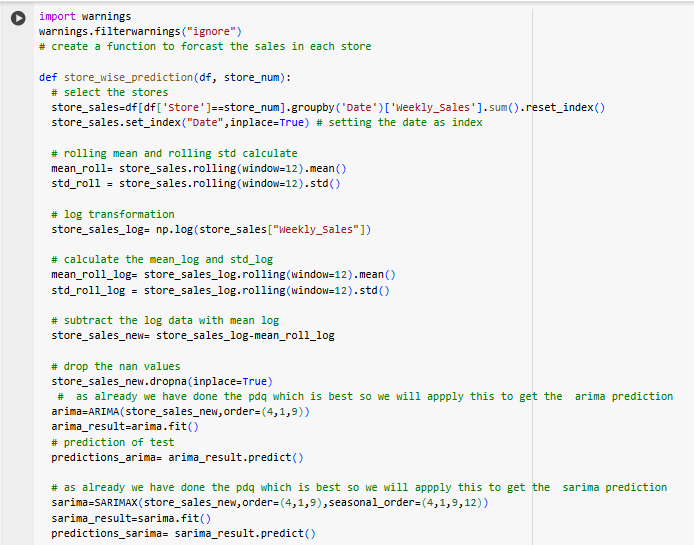






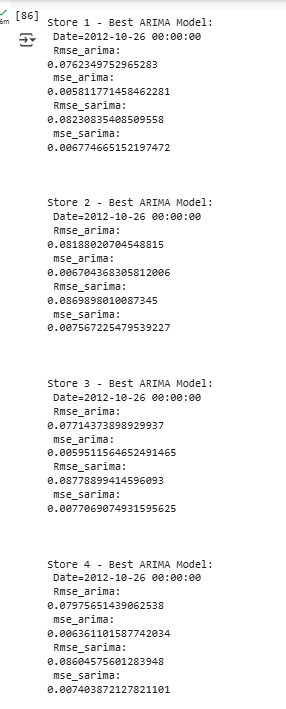
# Store wise prediction

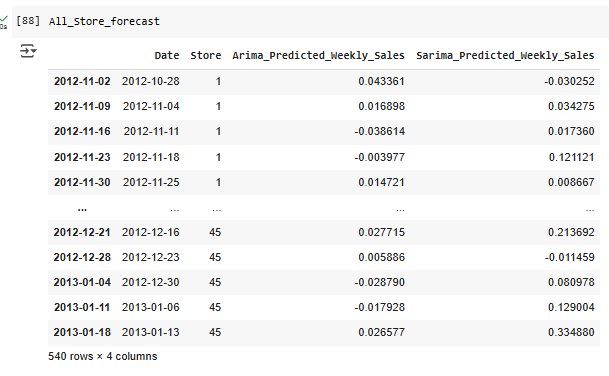




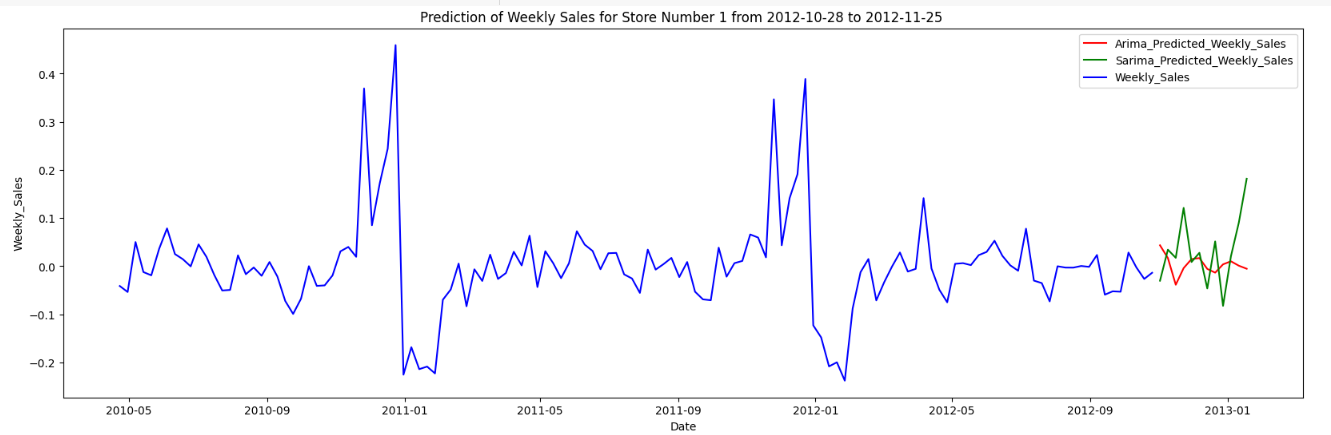


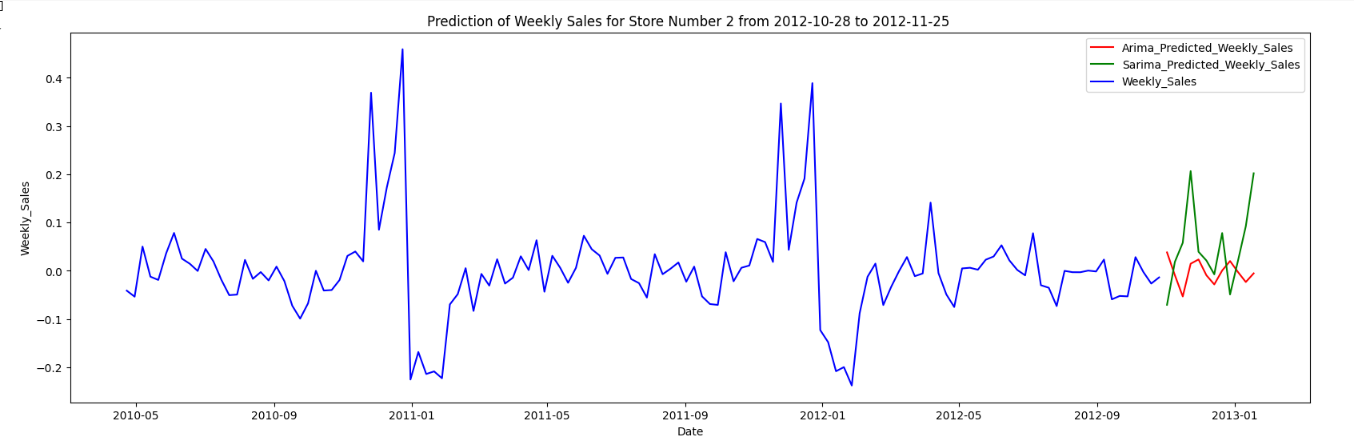
# Its showing only 5 stores but all of the store will be printout



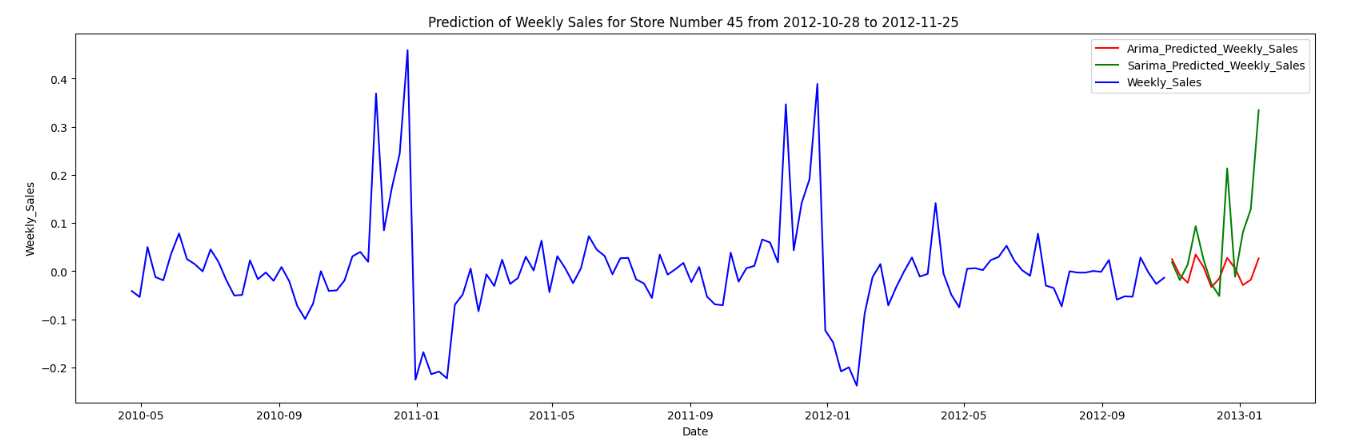








# There is all the Visualization so only showing here some:



# **Motivation and Reasons for Choosing the Algorithm**

We evaluated time series models ARIMA, SARIMA for some reasons:

* The model uses past values in the series to predict future values. The number of past observations considered is represented by the parameter *p* (AR order).
* ARIMA can model non-stationary data by differencing, which transforms the series into a stationary one (i.e., the statistical properties like mean and variance remain constant over time). The order of differencing is represented by *d*.
* The model uses past forecast errors to predict future values. The number of past forecast errors used is represented by the parameter *q* (MA order).
* SARIMA is an extension of ARIMA designed to handle seasonal components in the time series data.
* The length of the seasonal cycle (e.g., m=12 for monthly data with yearly seasonality).
* SARIMA can handle datasets that exhibit seasonality (e.g., sales data that peaks every holiday season). It adds seasonal components to the ARIMA model to capture these repeating cycles.
* Both ARIMA and SARIMA can be applied to a wide variety of time series problems, making them suitable for different domains

# **Assumptions**

There are few assumptions for these recommendation projects:

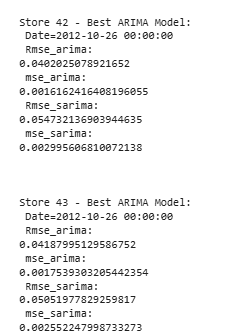
* The model assumes that there is enough data for predictions.
* The model assume that user’s past behaviour is good indicator for future prediction.
* The model assume that the data is accurate and reliable.
* The model assume that the user’s preferences are stable.

# **Model Evaluation and Techniques**

There is an important feature that is evaluation technique is need to complete the project. How much the actual value and model prediction value score is accurate that is more important. There are many evaluation technique is there like

* RMSE
* MSE

I have seen RMSE(Root Mean Square Error) here to see the error value.



# **Inferences from the Same**

After evaluating from the model there are some inferences from the model:

* Inferences from accuracy shows how the model predicts ratings, which help us to understand user’s recommendation.
* By this model we understand the both the arima and sarima prediction of future 12 weeks.

# **Future Possibilities of the Project**

The future of recommendation system can explore several possibilities like

* In future recommendation adapt in real time user ‘s changing behaviour, mood, context.
* Developing this type of Model we can predict the future numbers of any datas

# **Conclusion**

The project successfully demonstrated how predictive modelling can help Walmart sales forecasting, and the sales trends are related to the various factors like temperature, cpi, unemployment etc.