**CAPSTONE PROJECT**

**WALMART PROJECT**

**SUBMITTED**

**BY**

**DHINESH S R**

**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

12. References

**1 Problem Statement**

A retail store that has multiple outlets across the country is facing issues in managing the inventory - to match the demand with respect to supply.

**2 Project Objective**

The objective of our project is to develop a statistical Model based on the dataset available. We are using the historical sales data for 45 Walmart stores located in different regions to predicting the Weekly sales for each store. Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays.

The objective of this data to following inferences:

1. 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:
   1. If the weekly sales are affected by the unemployment rate, if yes - which stores are suffering the most?
   2. If the weekly sales show a seasonal trend, when and what could be the reason?
   3. Does temperature affect the weekly sales in any manner?
   4. How is the Consumer Price index affecting the weekly sales of various stores?
   5. Top performing stores according to the historical data.
   6. 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

**3 Data Description**

A Walmart dataset could refer to a collection of data related to the operations and activities of Wal-Mart.

**Store Information:** Walmart store column has store number, 45 stores each with 143 observations.

**Date:** Date column has week of sales, data from 2010 until 2012.

**Weekly-Sales:** A weekly sale in a Walmart dataset involves summarizing and analyzing the sales data for each week. Here’s how you can describe weekly sales in a Walmart dataset: Aggregated Sales, Average Sales, Minimum & Maximum Sales, Sales Growth, Seasonal pattern, Trend Analysis, Day-of-Week Sales etc.

**Holiday-Flag: “**Holiday Flag” in a Walmart dataset typically involves understanding how this flag is used to mark holidays or special events and its impact on sales or other relevant data. Here’s how you can describe holiday Flag in a Walmart dataset: Flag description (1 & 0), Holiday Dates, Sales impact, Customer behavior, promotional events etc.

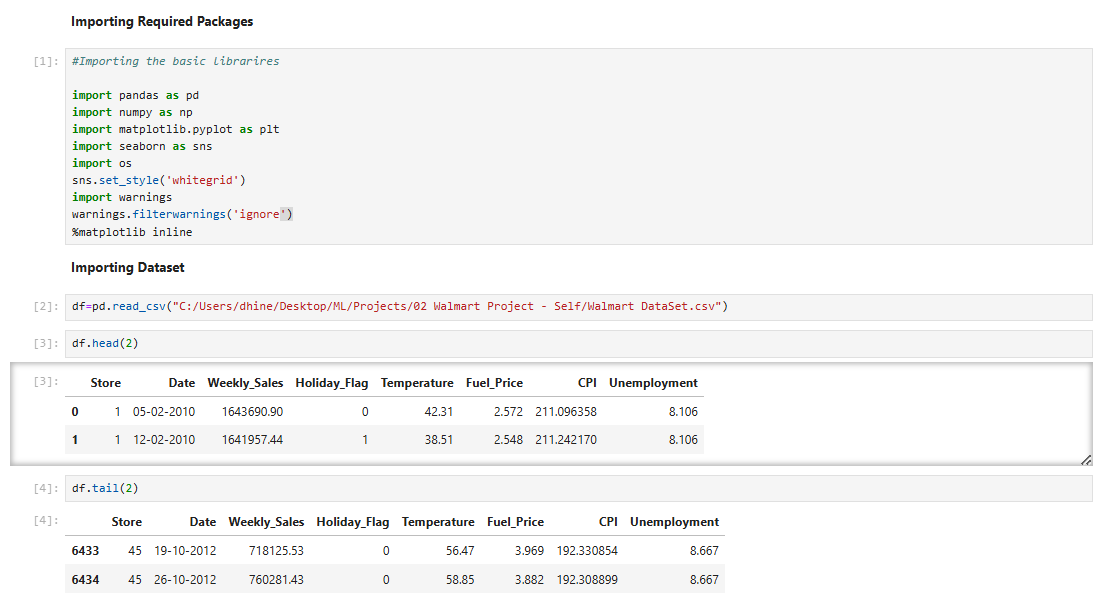
**Temperature:** Temperature data in a Walmart dataset involves providing insights and analysis about how temperature variations might impact aspects of Wal-Mart’s operations, such as sales, customer traffic, and inventory management. Here’s how you can describe the temperature data in Walmart dataset: Temperature trends (e.g., summer, Winter), Weather related events, Sales impact, Strategies etc.

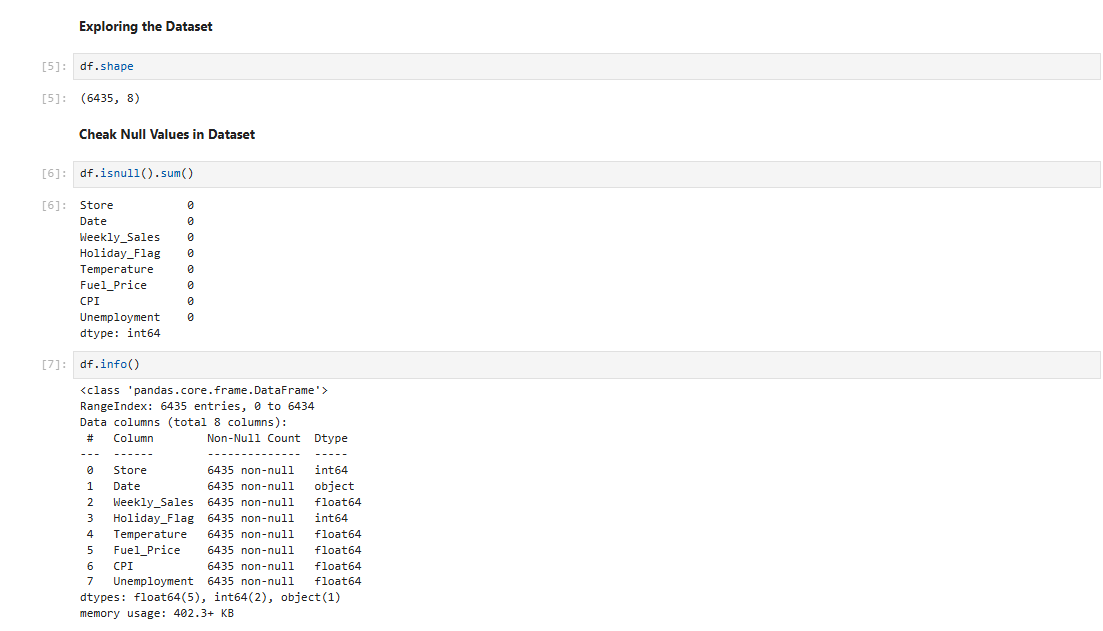
**Fuel Price:** Fuel price data in a Walmart dataset involves providing insights and analysis about how fluctuations in fuel price might impact aspects of Wal-Mart’s operations, particularly in terms of transportation costs and potentially consumer behavior. Here’s how you can describe the fuel price data in Walmart dataset: Fuel price trends, Transportation costs, Product demand etc.

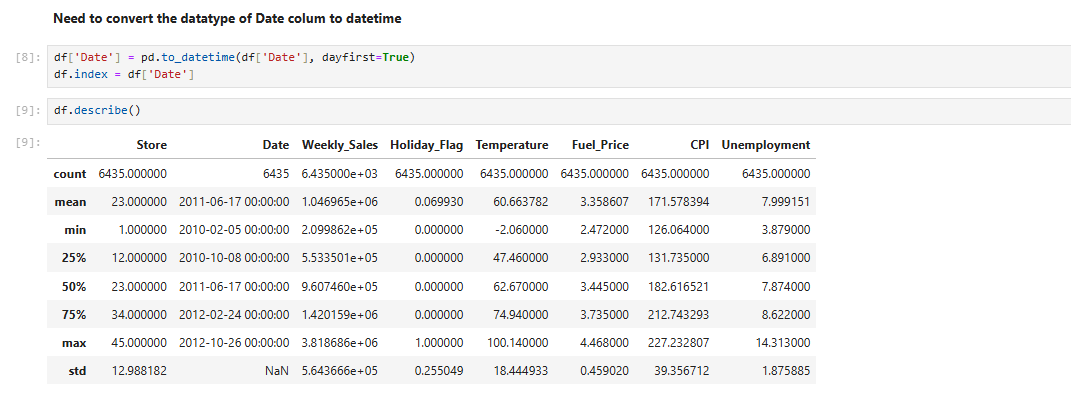
**CPI:** Consumer price index (CPI) data in a Walmart dataset involves providing insights and analysis about how CPI fluctuations might impact various aspects of Wal-Mart’s operations, including pricing strategies, consumer behavior and inflation effects. Here’s how you can describe the CPI data in Walmart dataset: CPI index, pricing strategies, inflations effects, product demand etc.

**Unemployment:** Unemployment data in a Walmart dataset involves providing insights and analysis about how fluctuations in unemployment might impact various aspects of Wal-Mart’s operations, particularly in terms of consumer behavior and economic conditions. Here’s how you can describe the unemployment data in Walmart dataset: Unemployment trends, consumer behavior, pricing & promotions, product demand etc.

**4. Data Pre-processing Steps and Inspiration**



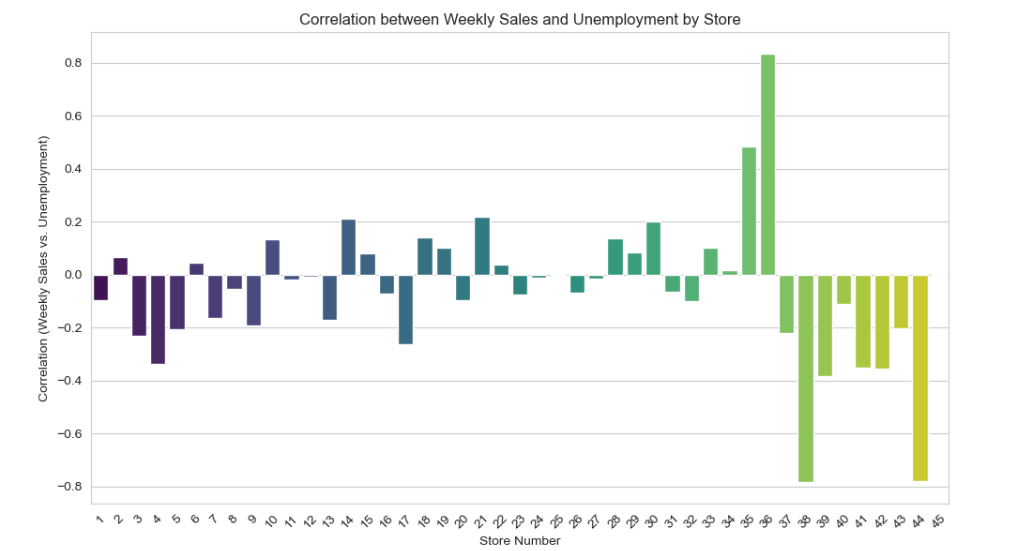


****

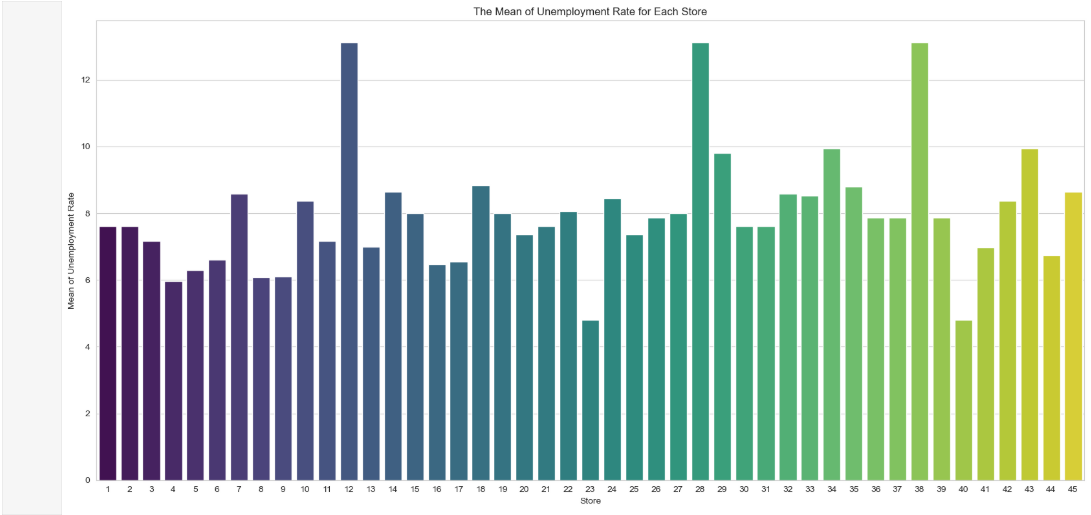
**EDA Results**

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

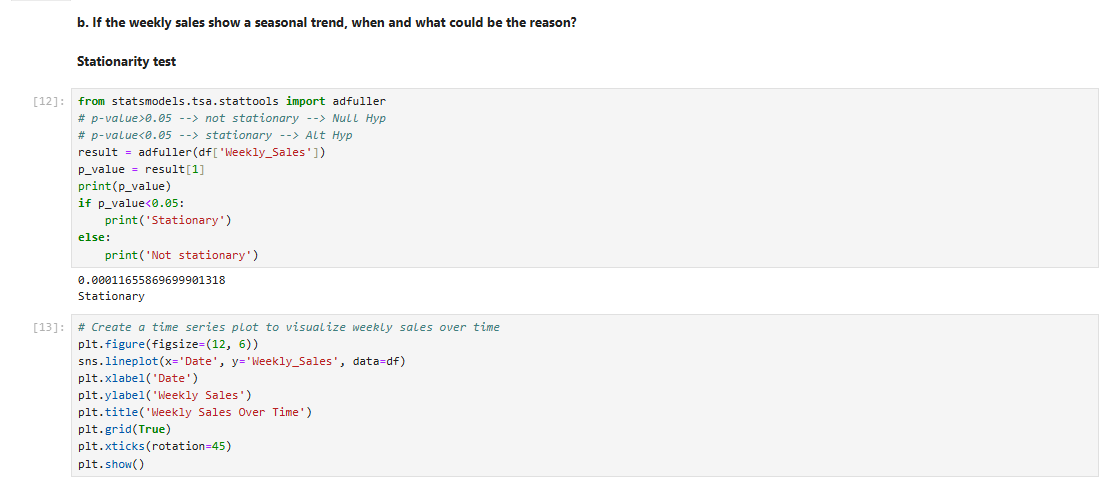


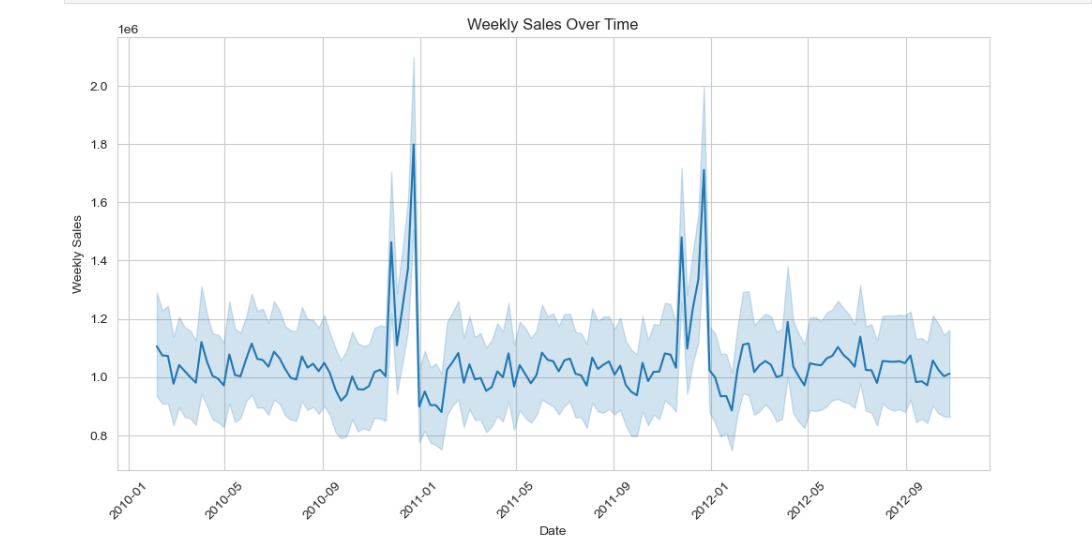




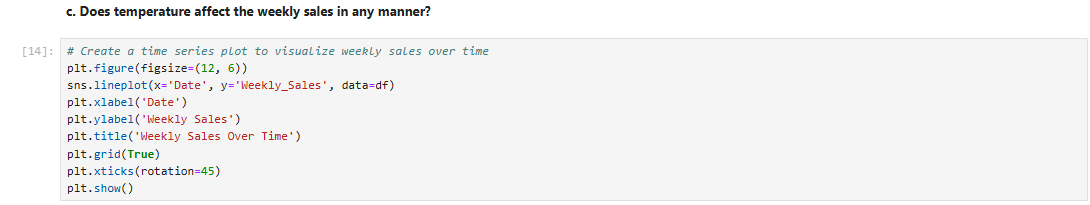


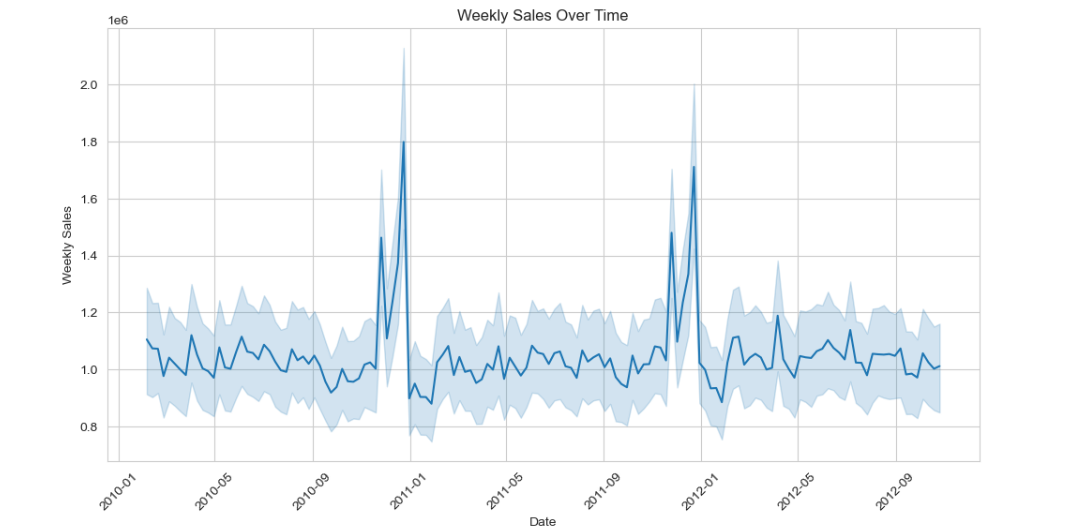
If the weekly sales show a seasonal trend, when and what could be the reason?

****



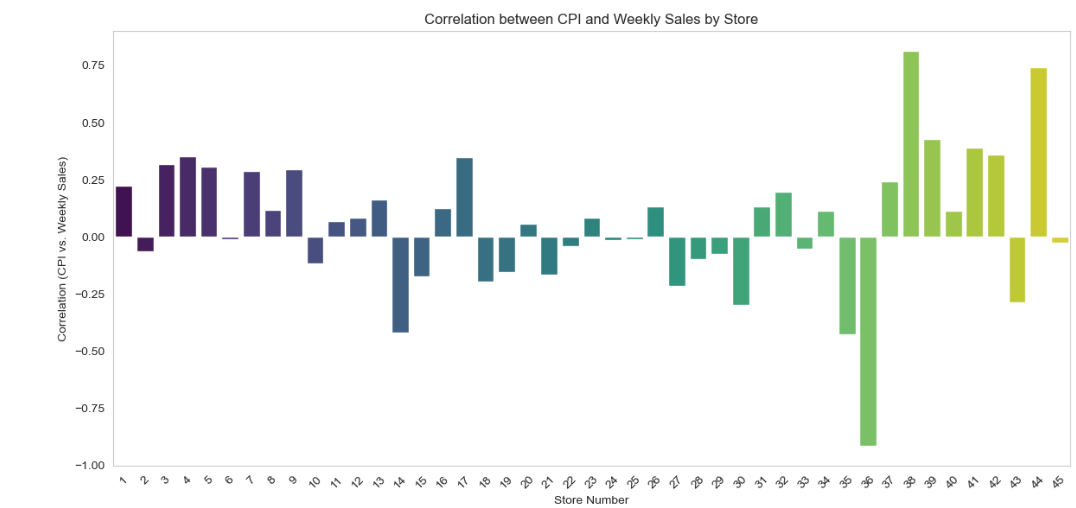
Does temperature affect the weekly sales in any manner?



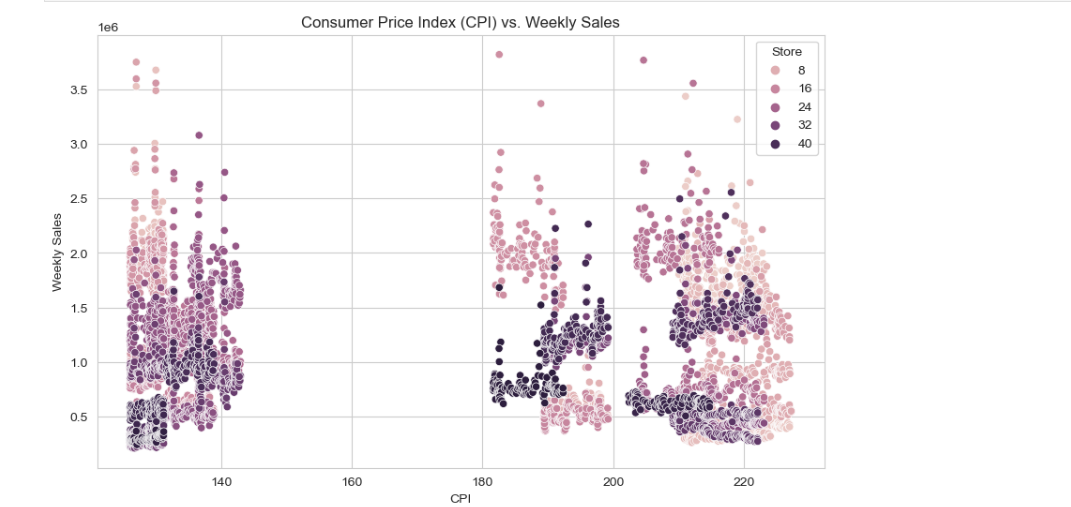


How is the Consumer Price index affecting the weekly sales of various stores?



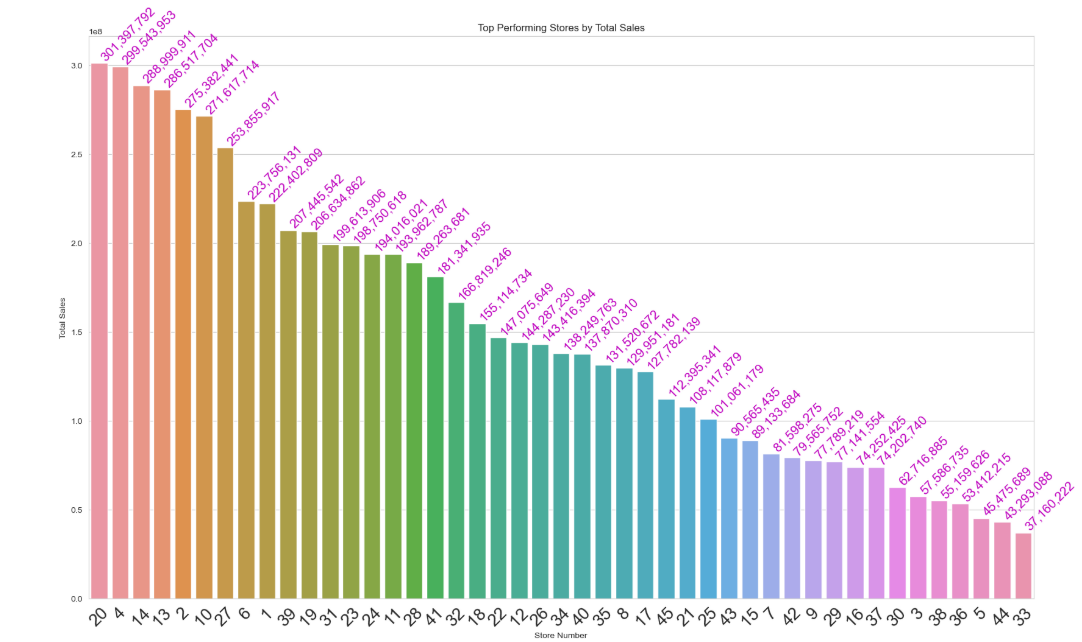




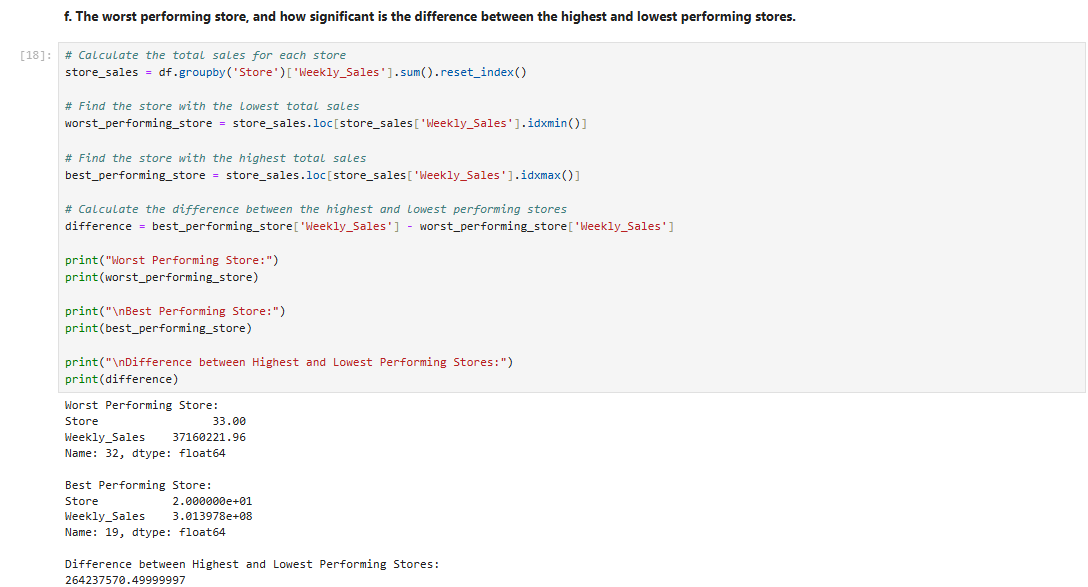


Top performing stores according to the historical data.

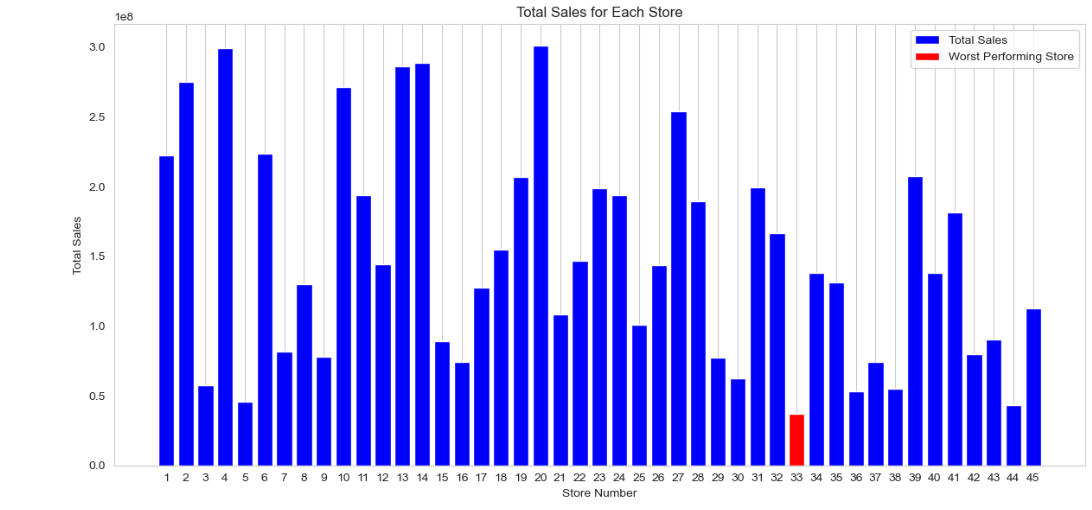




The worst performing store, and how significant is the difference between the highest and lowest performing stores.

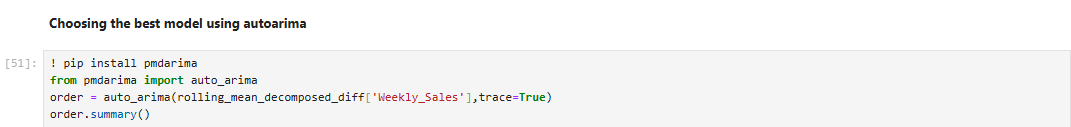






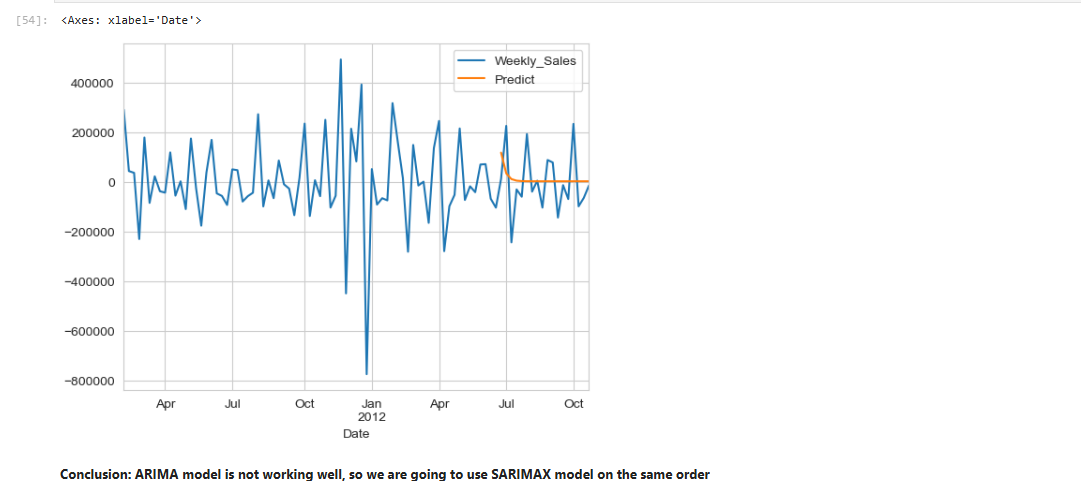
**5. Choosing the Algorithm for the Project**

Choosing the best model using autoarima

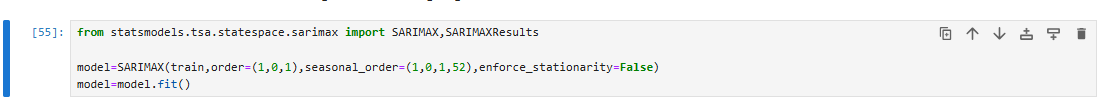
****

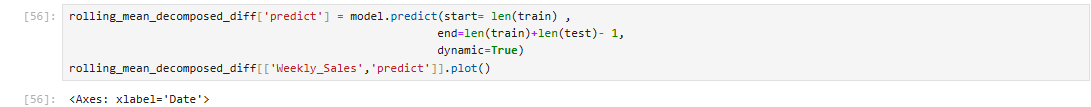
**6.Motivation and Reasons For Choosing the Algorithm**

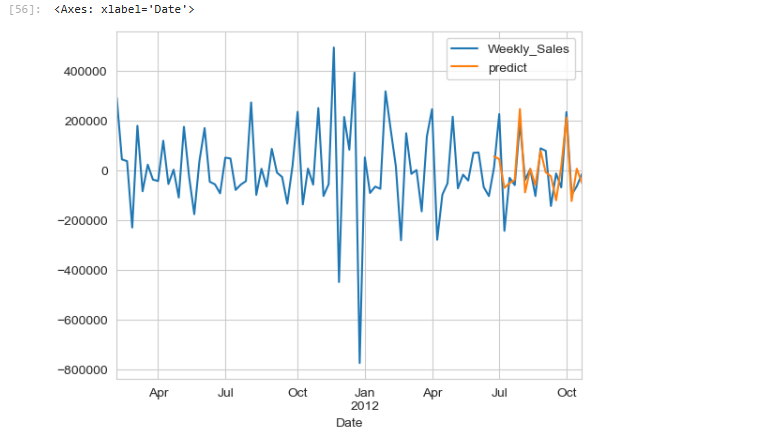




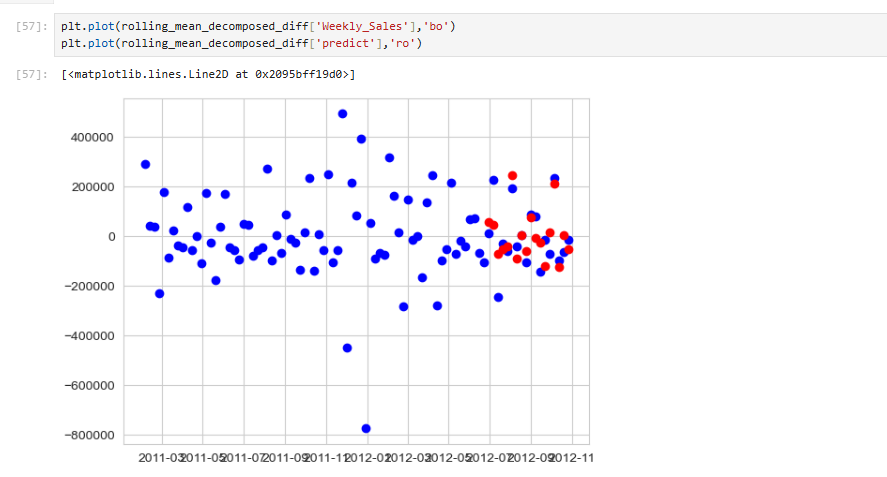
ARIMA model is not working well, so we are going to use SARIMAX model on the same order

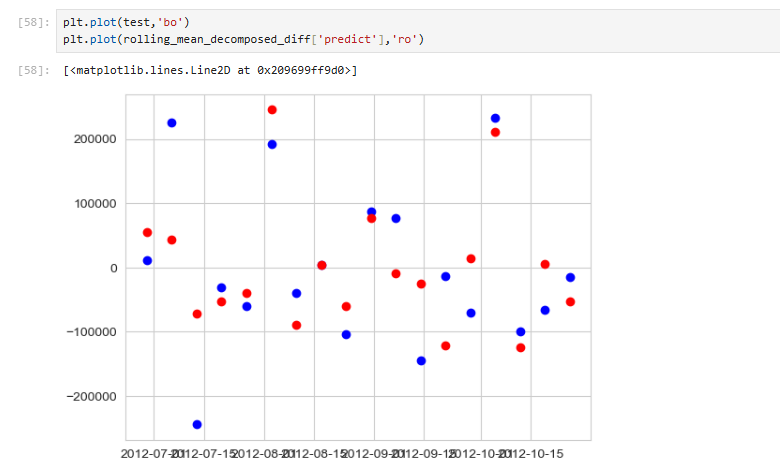
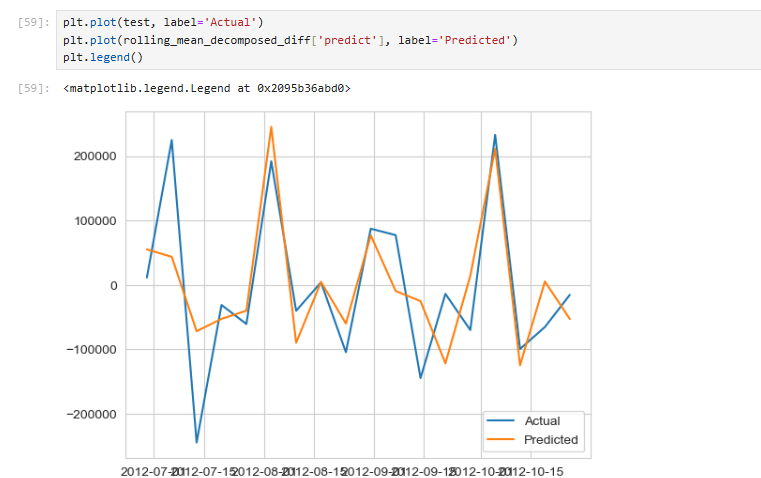






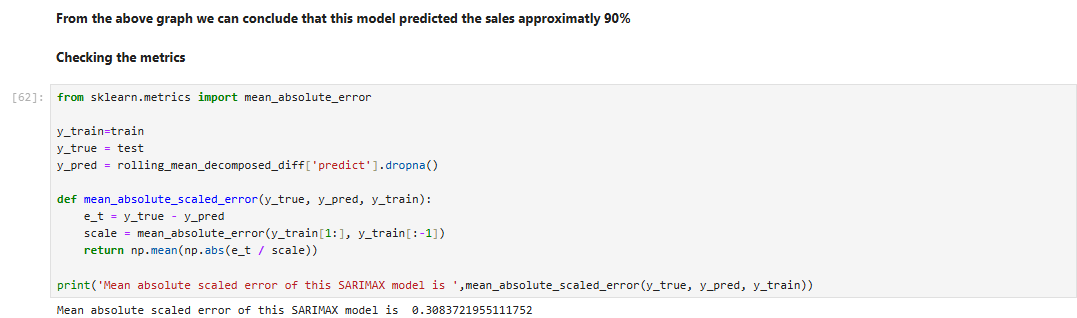
**7. Assumptions**

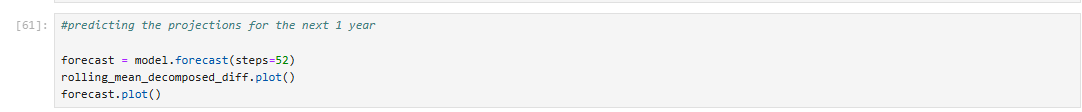


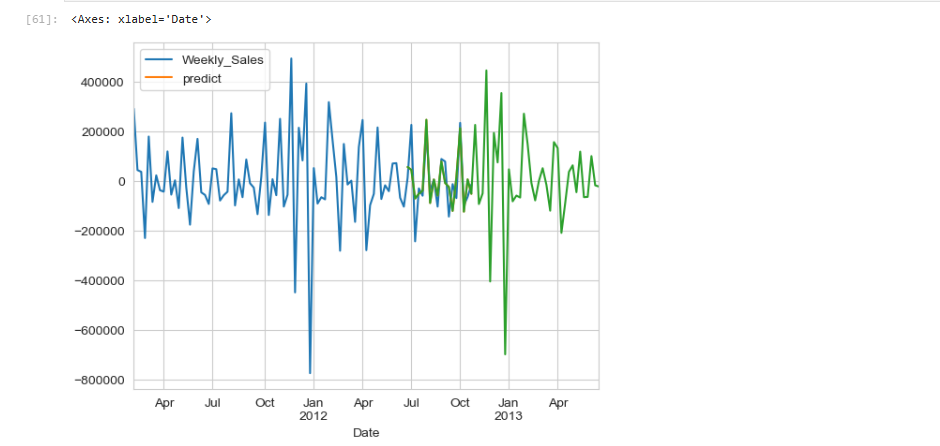
 

We can check the model performance from the above graph, it predicted nearly possible values.

**8. Model Evaluation and Techniques**

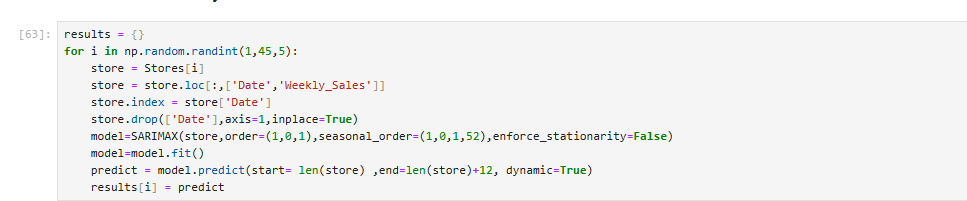


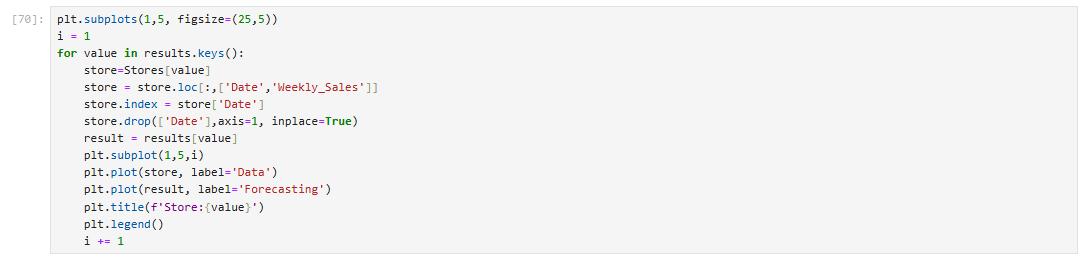


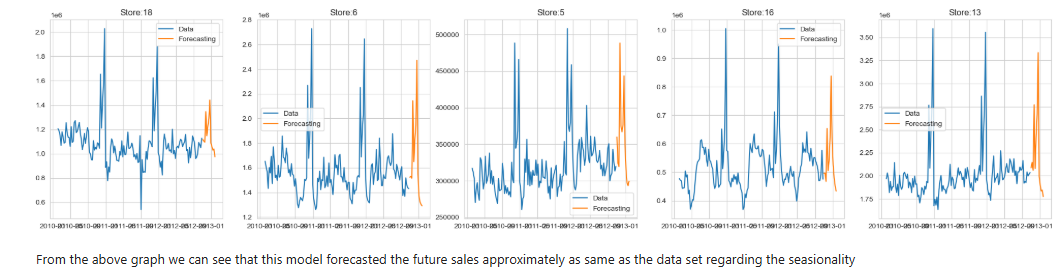


From the above graph we can conclude that this model predicted the sales approximately 90%

**9. Inferences from the Same**







**10. PRIDICTION OF SALES FOR EACH STORE NEXT 12 WEEKS**

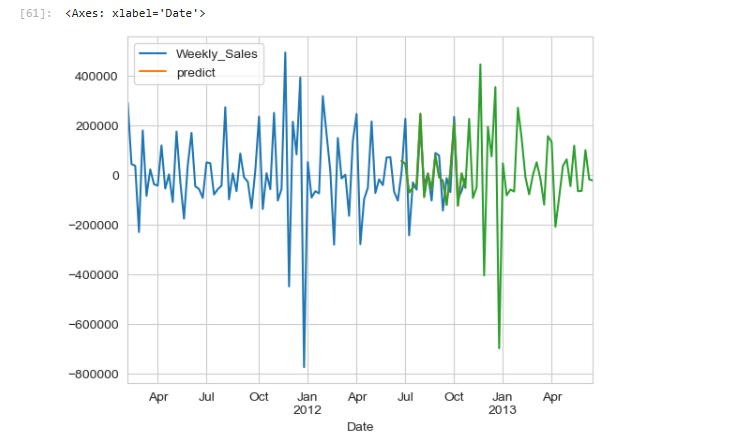
1.Time Series Analysis: Examine your historical sales data to understand any trends, seasonality, and stationarity. Ensure your data is appropriately preprocessed for time series modeling.

2.Train a SARIMA Model: You can use libraries like stats models in Python to fit a SARIMA model to your sales data. The SARIMA model has several hyper parameters that need to be tuned, including the order of differencing (d), autoregressive (p), integrated (d), and moving average (q) components, as well as seasonal components (P, D, Q, s). You'll need to select appropriate values for these hyper parameters.

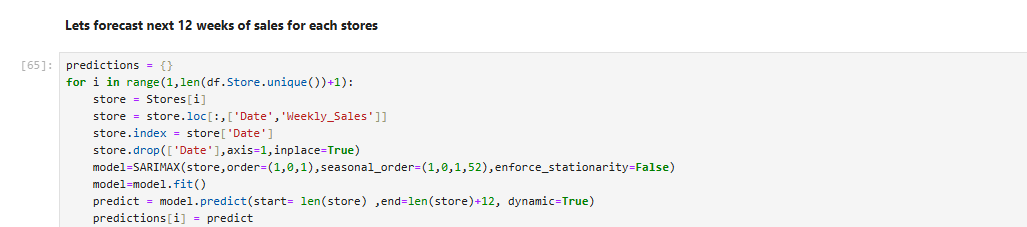
3.Model Validation: Split your historical data into a training set and a validation set. Use the training set to fit the SARIMA model and the validation set to evaluate its performance. You can use metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) to assess the model's accuracy.

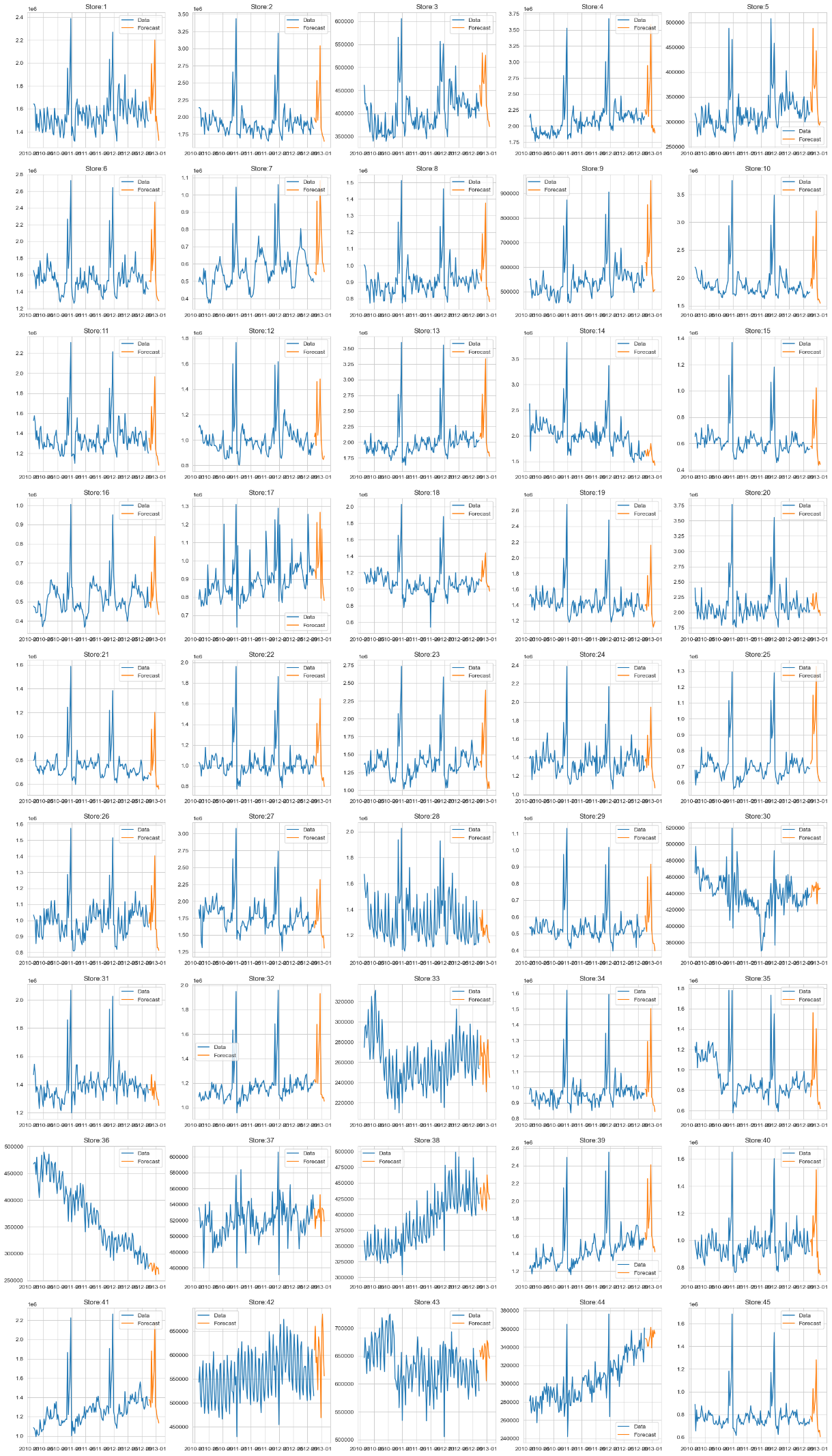
4.Model Forecasting: Once you have a well-tuned SARIMA model, you can use it to forecast sales for the next 12 weeks. You'll need to provide the model with the necessary input data, which typically includes the historical sales data.

****

****

**PRIDICTION OF SALES FOR EACH STORE NEXT 12 WEEKS**

****



**11. Conclusion**

From the above graph we can see that the Stores 14,20,28,30,31,37,45 are having poor forecasting.

**12. References**

* Our Session Live Class (Handson AirPassengers Times Series)
* Chat GPT
* Kaggle