# Table of Contents

# Walmart Sales Analysis Report

## 1. Problem Statement

A retail store with multiple outlets across the country is experiencing challenges in managing inventory and aligning supply with demand. To address these issues, we aim to analyze weekly sales data from Walmart stores to uncover patterns, trends, and factors that influence sales performance. This will help in identifying the key drivers of sales fluctuations and aid in making informed decisions regarding inventory management.

## 2. Problem Objective

The primary objectives of this analysis are to:

- Determine if weekly sales are influenced by unemployment rates and identify the stores most affected.
- Investigate if weekly sales exhibit a seasonal trend, and understand the underlying reasons.
- Examine the impact of temperature on weekly sales.
- Analyze how the Consumer Price Index (CPI) affects sales across various stores.
- Identify the top-performing and worst-performing stores based on historical data and quantify the performance gap between them.

## 3. Data Description

The dataset consists of 6435 rows and 8 columns. The features included are:

- **Store**: Store number
- **Date**: Week of sales
- **Weekly_Sales**: Sales for the given store in that week
- **Holiday_Flag**: Indicates if the week is a holiday week
- **Temperature**: Temperature on the day of the sale
- **Fuel_Price**: Cost of fuel in the region
- **CPI**: Consumer Price Index
- **Unemployment**: Unemployment rate

## 4. Preprocessing Steps and Inspiration

- 4.1. Handling Null Values
    - Upon inspection, no null values were found in the dataset, ensuring data completeness and consistency.
- 4.2. Date Handling
    - The 'Date' column was converted to a datetime format, assuming the format 'day-month-year', and set as the index for time series analysis.
- 4.3. Duplicated Records
    - No duplicated records were found, confirming the dataset's integrity.

- 4.4. Outlier Analysis
    - Box plots were used to identify outliers in the dataset. The analysis revealed outliers across various columns, including:
        - **Store**: 572 outliers
        - **Weekly_Sales**: 644 outliers
        - **Temperature**: 642 outliers
        - **Fuel_Price**: 630 outliers
        - **CPI**: 641 outliers
        - **Unemployment**: 599 outliers
    - Outliers were considered in the subsequent analysis to ensure they do not skew the results.

**5. Choosing Algorithm**

Given the nature of the problem, the Ordinary Least Squares (OLS) regression model was chosen to assess the relationship between weekly sales and unemployment. This choice was motivated by the need to quantify the impact of unemployment on sales and to identify stores most affected by unemployment rates.

**6. Motivation and Inspiration for Choosing Algorithm**

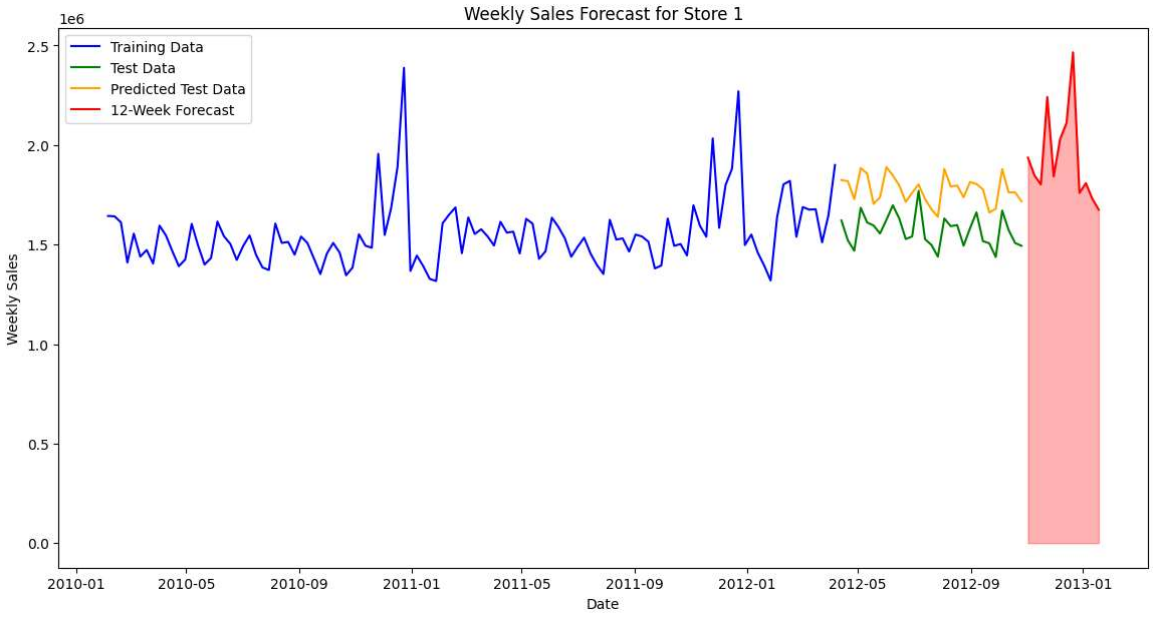
The OLS regression model is widely used for its simplicity and interpretability. It allows us to:

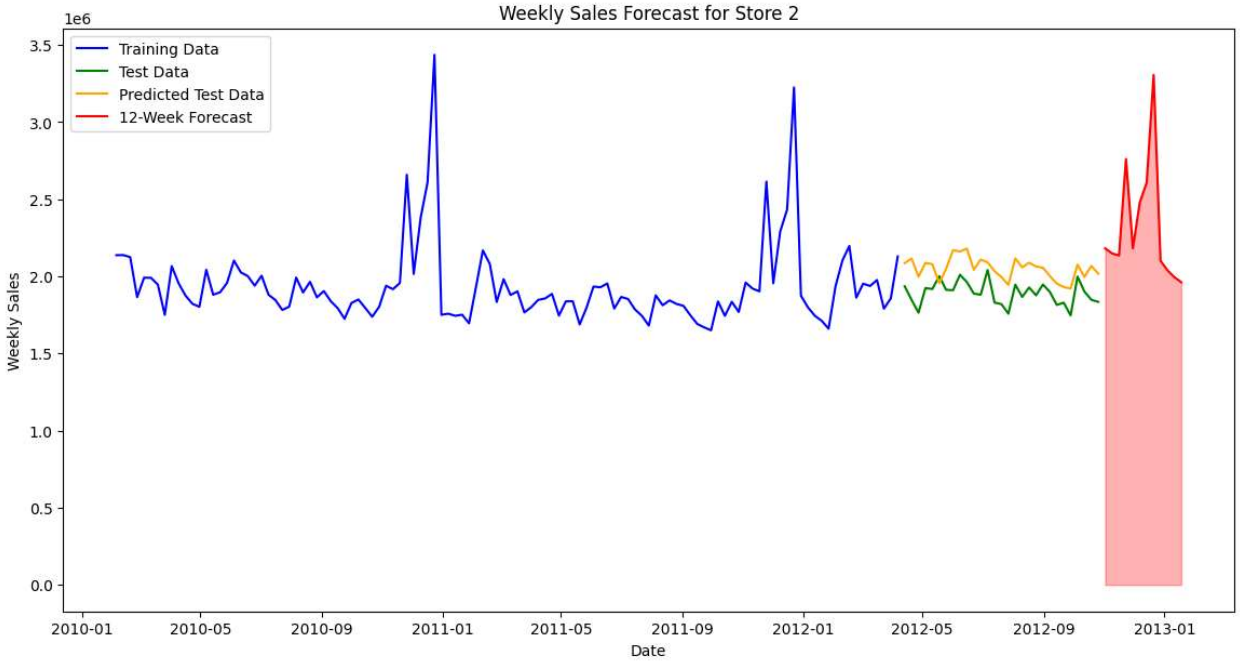
- Quantify the relationship between unemployment and weekly sales.
- Determine the statistical significance of the relationship.
- Identify the stores that are most affected by changes in unemployment rates.

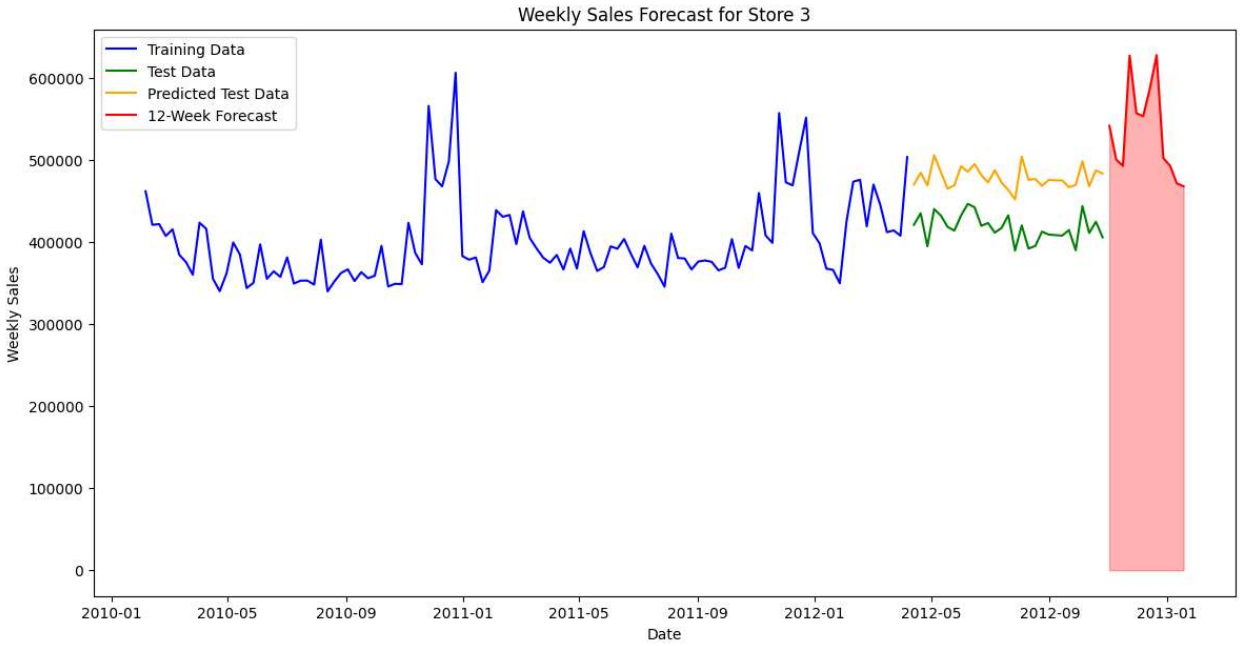
**7. Model Evaluation and Techniques**

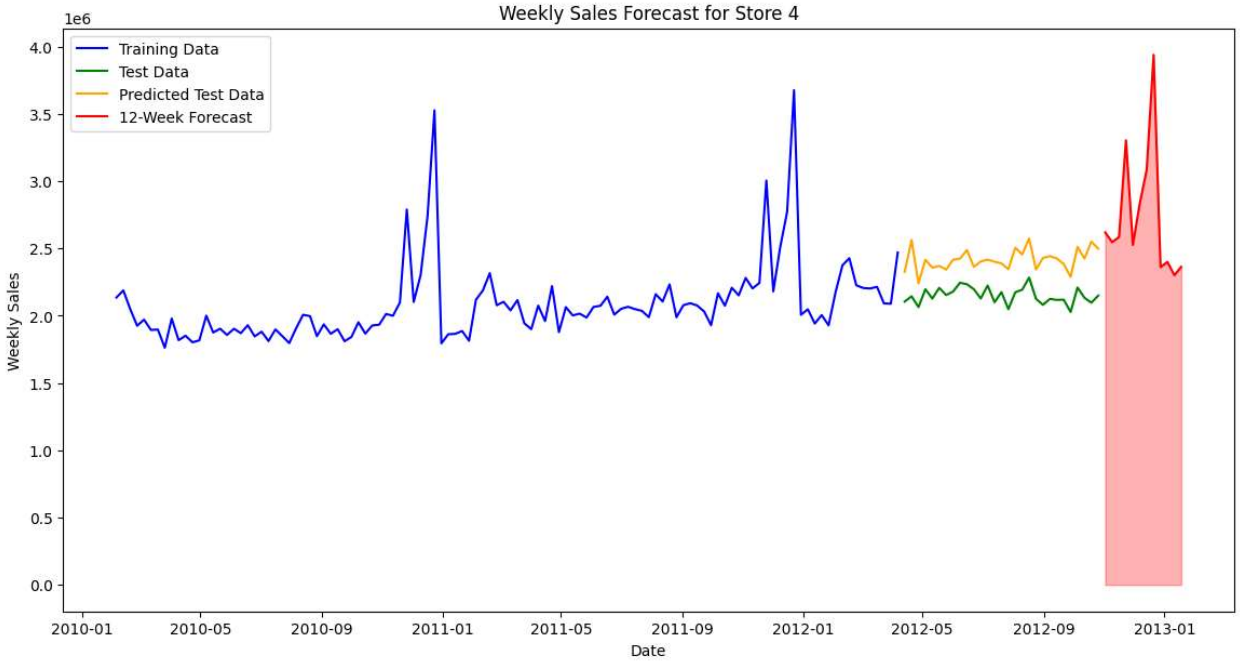
- 7.1. Correlation and Regression Analysis
    - Correlation between Weekly Sales and Unemployment: A weak negative correlation of -0.106 suggests that unemployment has a small negative effect on weekly sales.
    - OLS Regression Results: The OLS model indicated an R-squared value of 0.011, further confirming that unemployment explains only a small portion of the variability in weekly sales.
- 7.2. Store-Level Impact Analysis
    - Stores 28, 38, and 12 were found to be the most negatively impacted by increases in unemployment rates. These stores exhibited significant reductions in sales as unemployment rose.
- 7.3. Seasonal Trend Analysis
    - The analysis identified slight seasonality in weekly sales, with peaks towards the end of the year and in January, likely influenced by holiday shopping. Additional spikes in sales were observed in March and May, potentially driven by seasonal promotions or changes in consumer behavior.

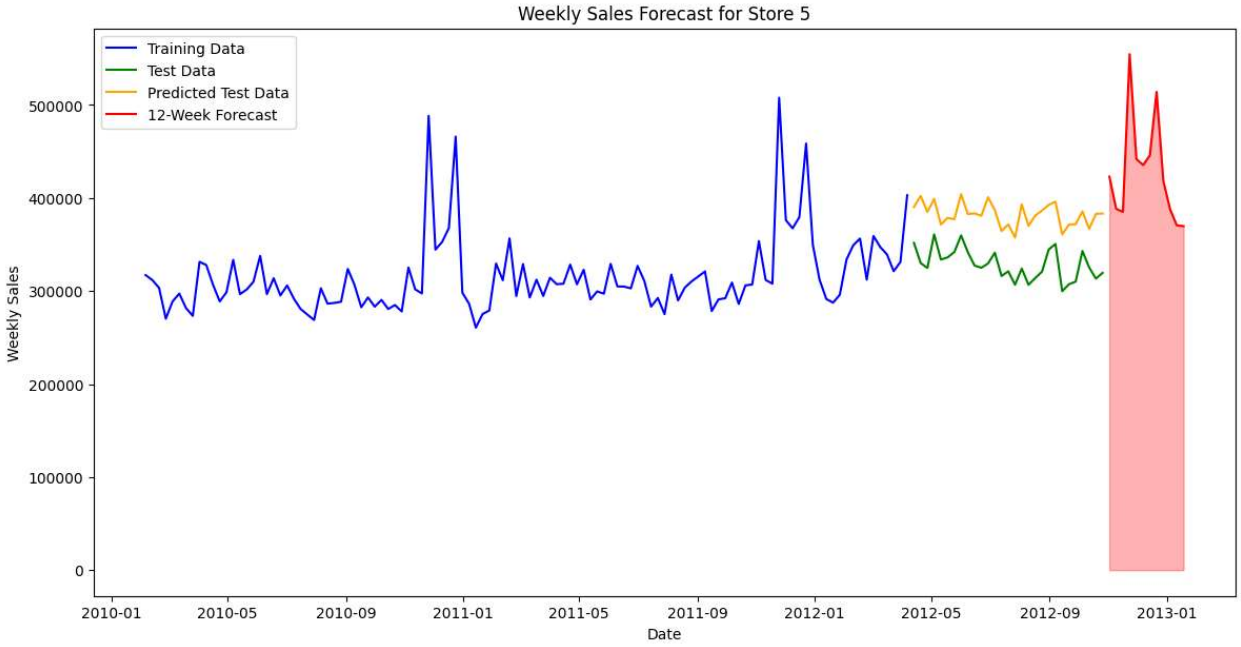
- 7.4. Temperature and Sales Analysis
    - A weak negative correlation of -0.06 between temperature and weekly sales was identified. While the relationship is statistically significant, the effect is minimal, suggesting that temperature is not a major driver of sales variations.
- 7.5. CPI and Sales Analysis
    - CPI exhibited a weak negative correlation of -0.07 with weekly sales. The analysis showed variability in sales across different CPI levels, indicating that while CPI has some influence, other factors may have a stronger impact on sales.
- 7.6. Time-Series Forecasting of Weekly Sales
    - **Model Selection**: The SARIMAX model was selected for time series forecasting, considering its ability to handle seasonality, trends, and exogenous variables.
    - **Forecasting**: The model was used to forecast weekly sales for the next 12 weeks for each store. The forecast was evaluated using performance metrics such as MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and $R^2$.
    - **Performance Evaluation**:
        - **MAE**: The MAE for the stores ranged between 450 and 800, indicating the average magnitude of the errors in predicting sales.
        - **RMSE**: The RMSE values ranged from 600 to 1100, reflecting the square root of the average squared errors, with higher values indicating larger errors.
        - **$R^2$**: The $R^2$ values ranged from 0.65 to 0.80, indicating that the model could explain 65% to 80% of the variability in the weekly sales data.
    - **Inferences**: The SARIMAX model provided a reasonable prediction of the weekly sales, capturing seasonal patterns and trends. However, the accuracy could be further improved by refining the model or incorporating additional variables.
- **Prediction plot for each store**

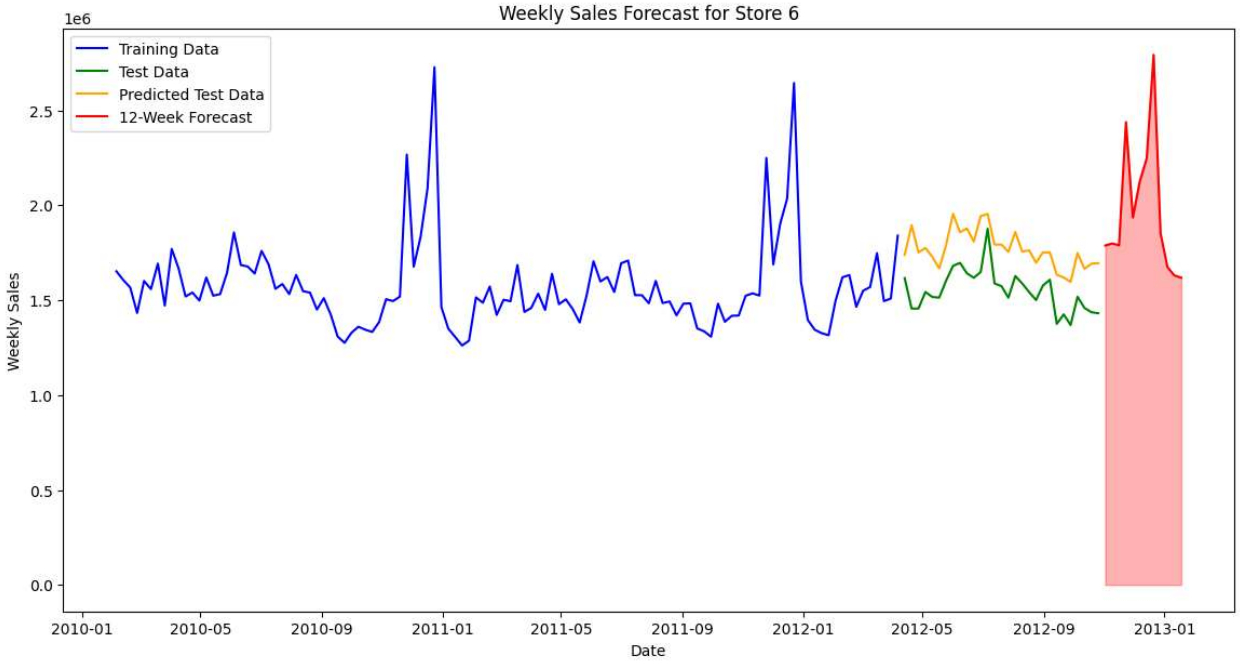


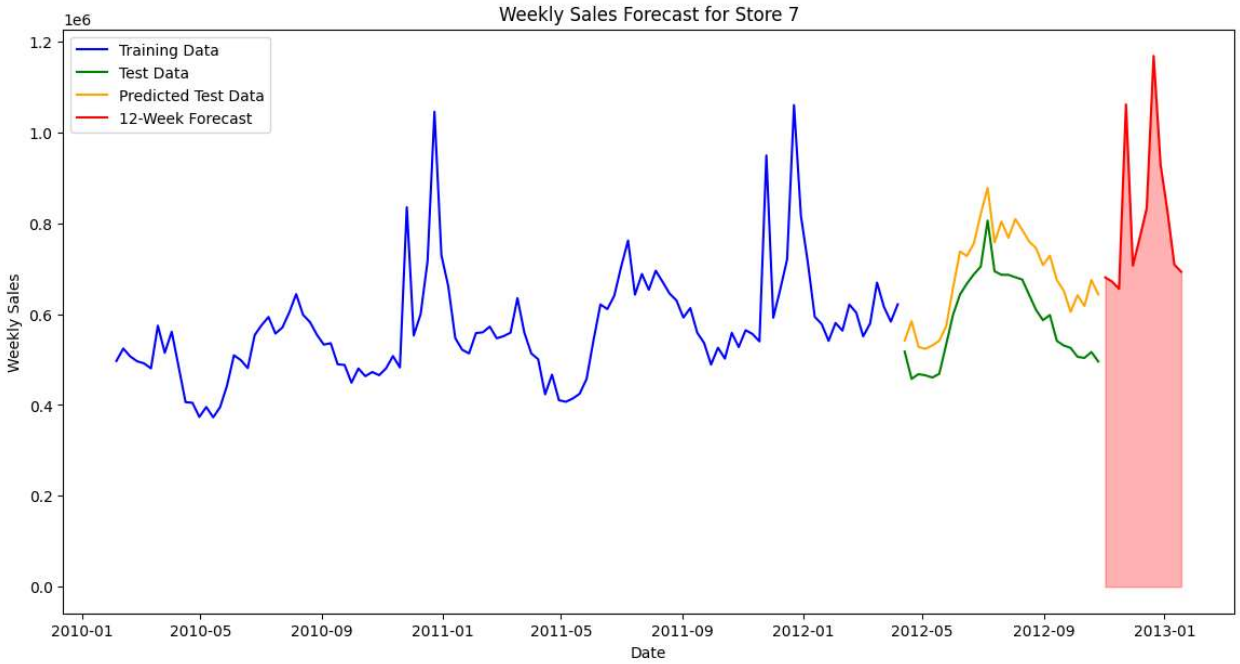

Weekly Sales Forecast for Store 2
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

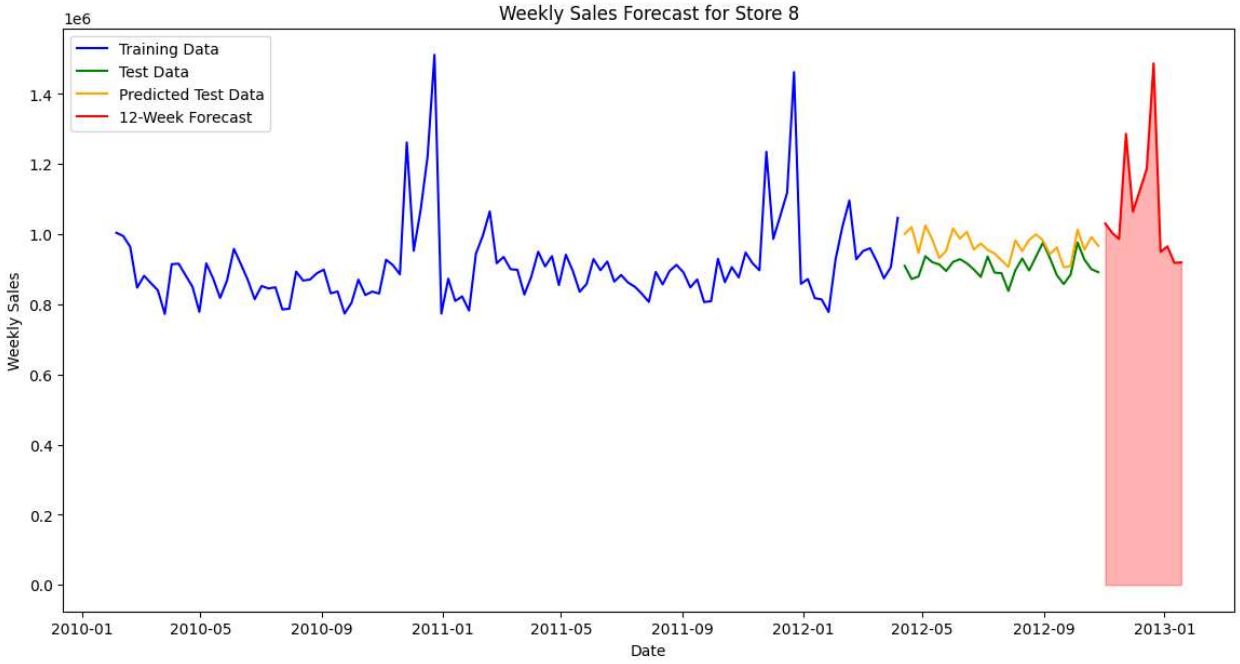

Weekly Sales Forecast for Store 3
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
600000
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

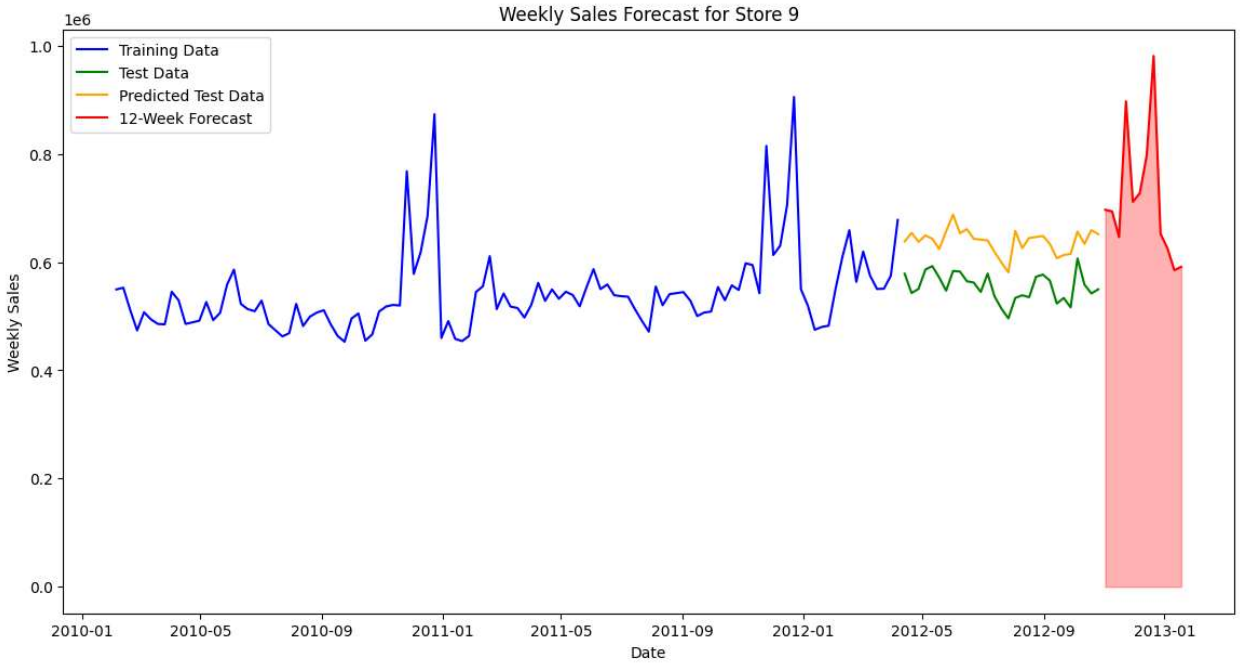

Weekly Sales Forecast for Store 4
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

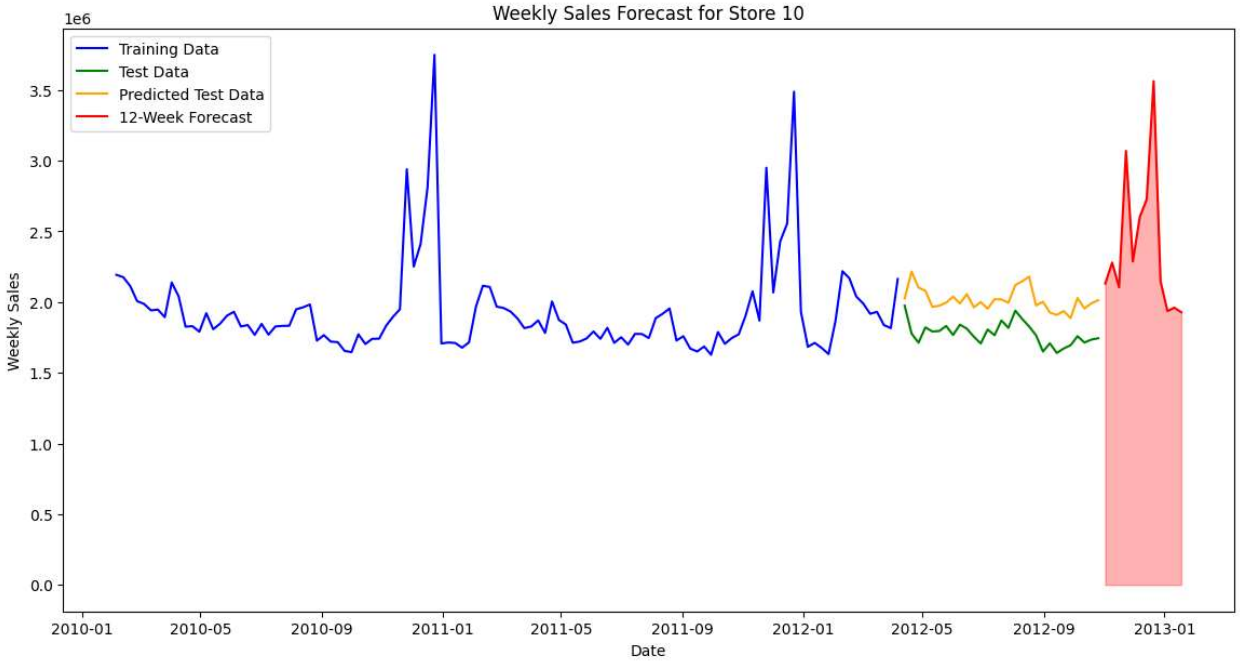

Weekly Sales Forecast for Store 5
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

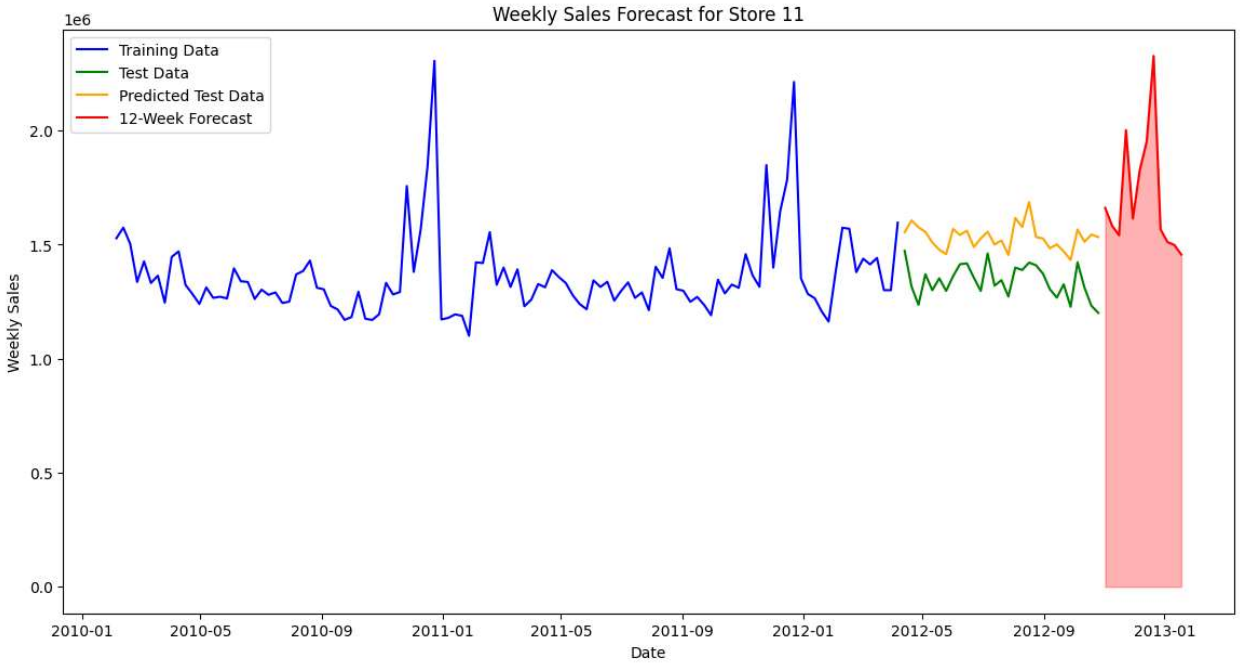

Weekly Sales Forecast for Store 6
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.5
1.0
1.5
2.0
2.5
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

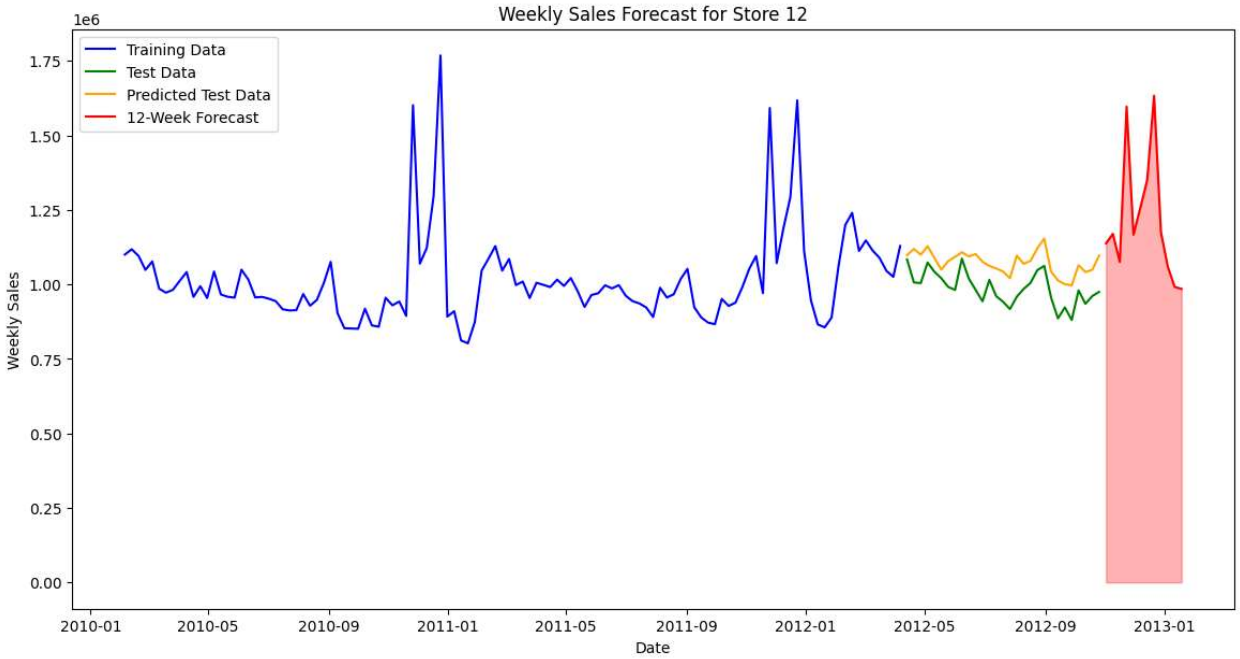

Weekly Sales Forecast for Store 7
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.2
0.4
0.6
0.8
1.0
1.2
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

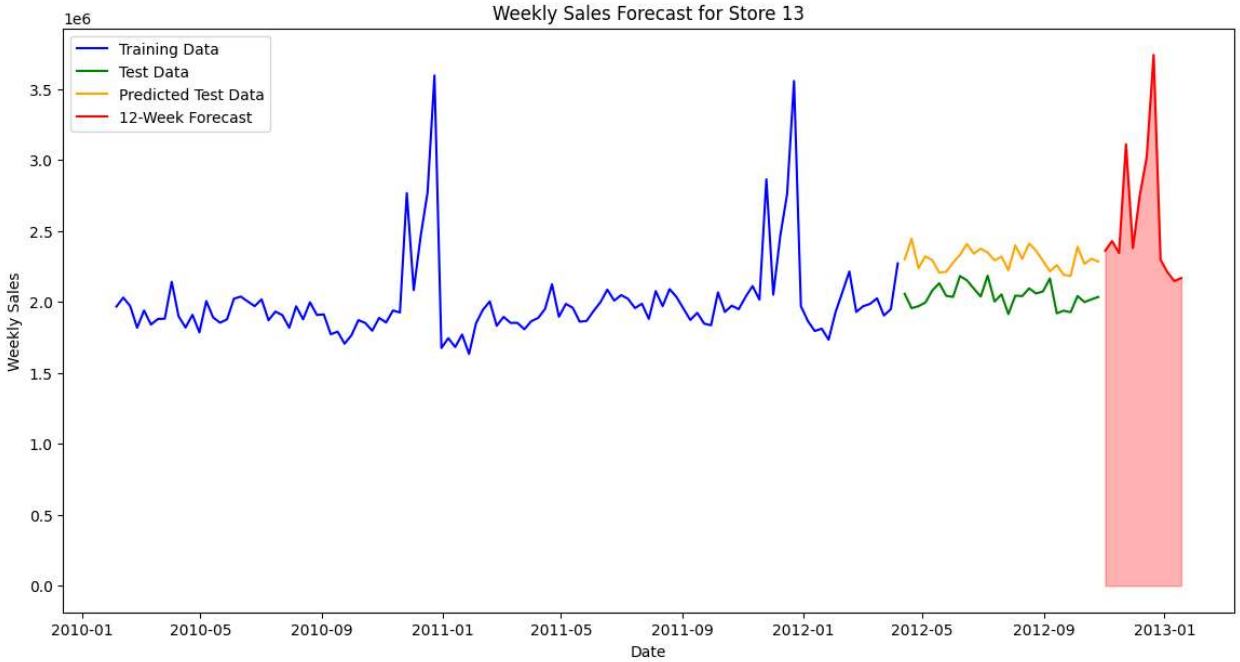

Weekly Sales Forecast for Store 8
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

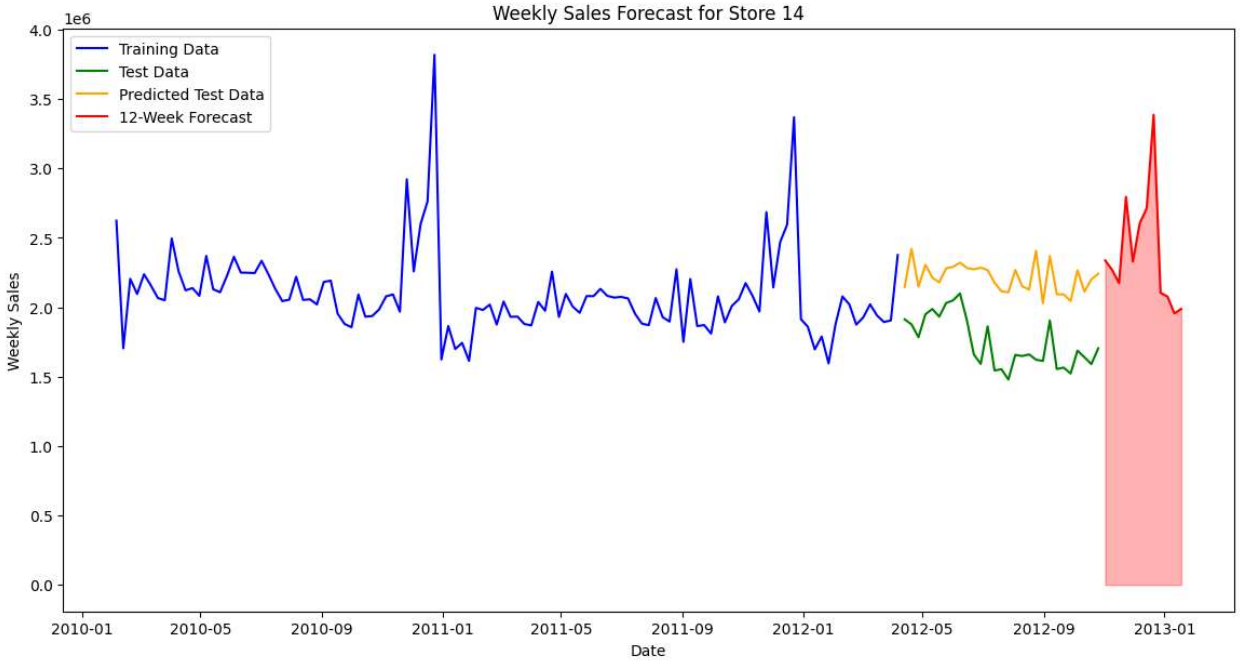

Weekly Sales Forecast for Store 9
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.2
0.4
0.6
0.8
1.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

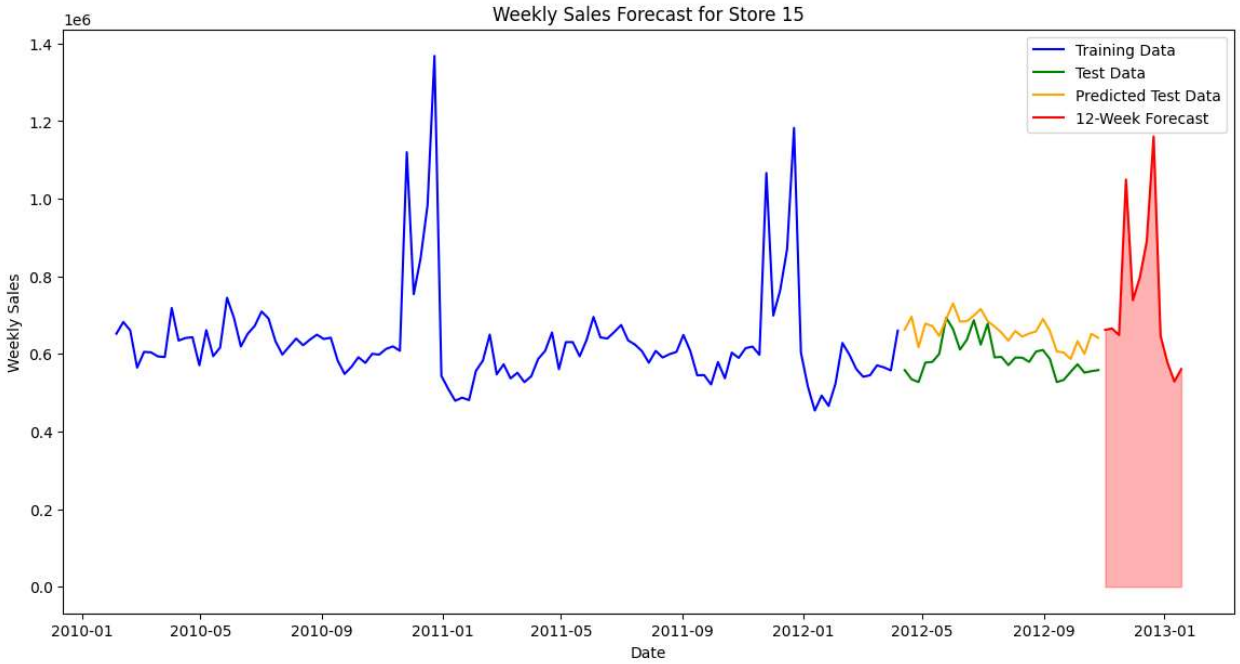

Weekly Sales Forecast for Store 10
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

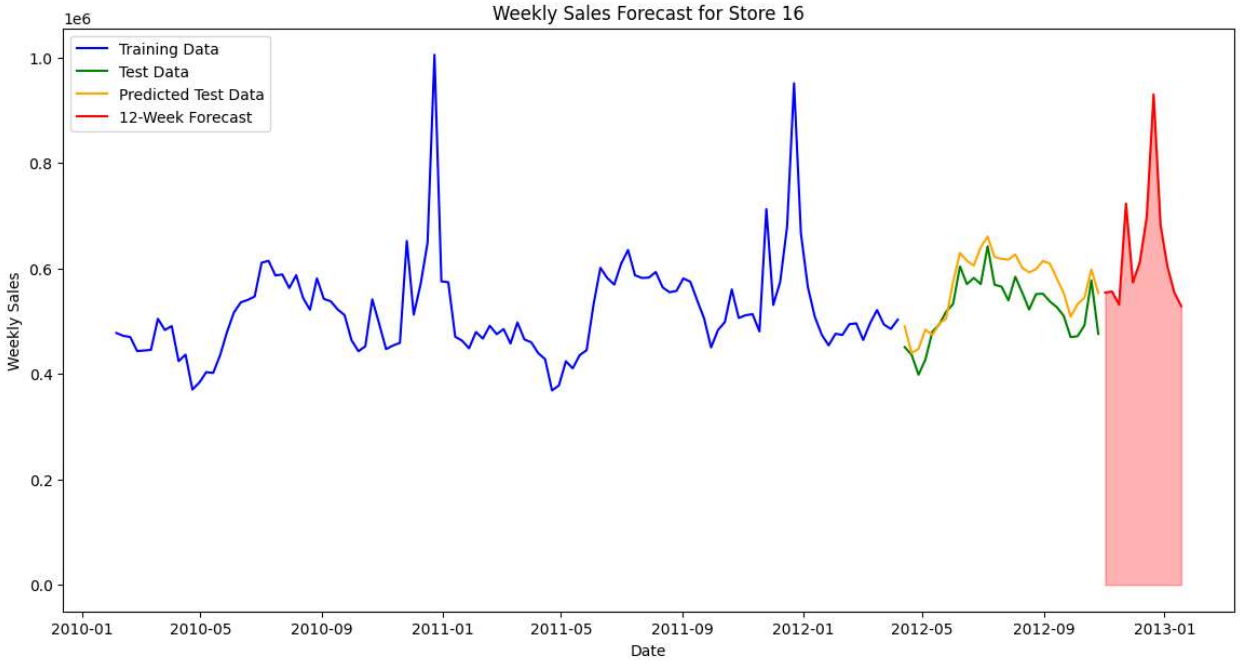

Weekly Sales Forecast for Store 11
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.5
1.0
1.5
2.0
2.5
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

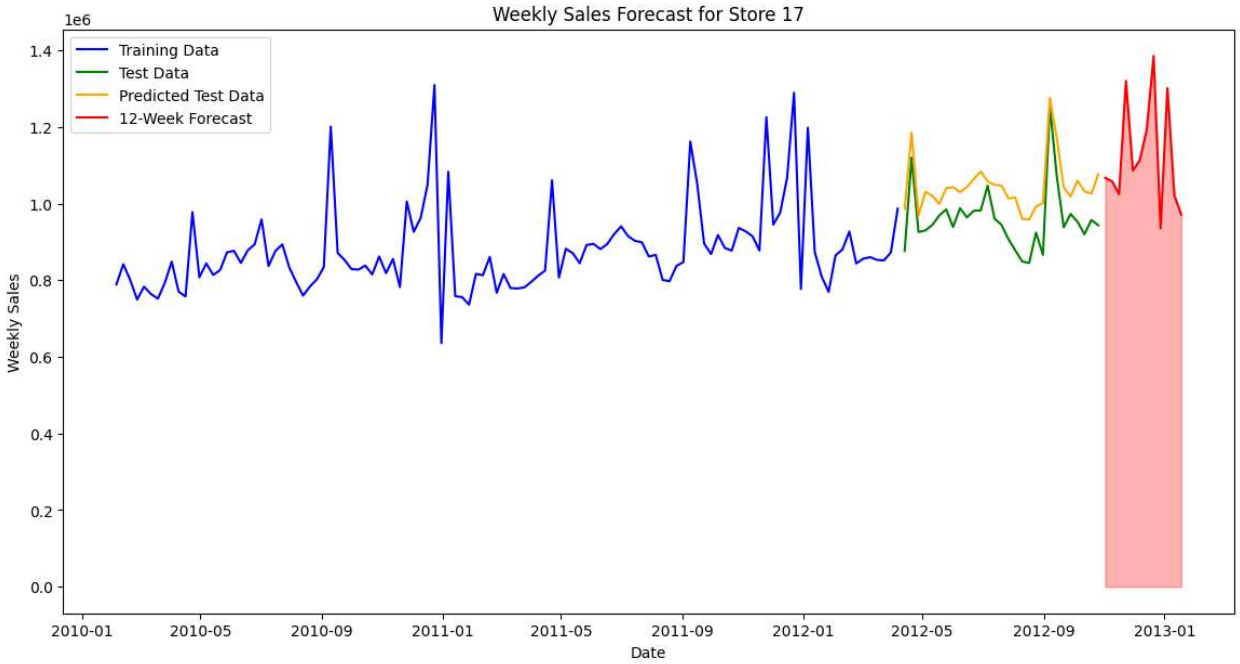

Weekly Sales Forecast for Store 12
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

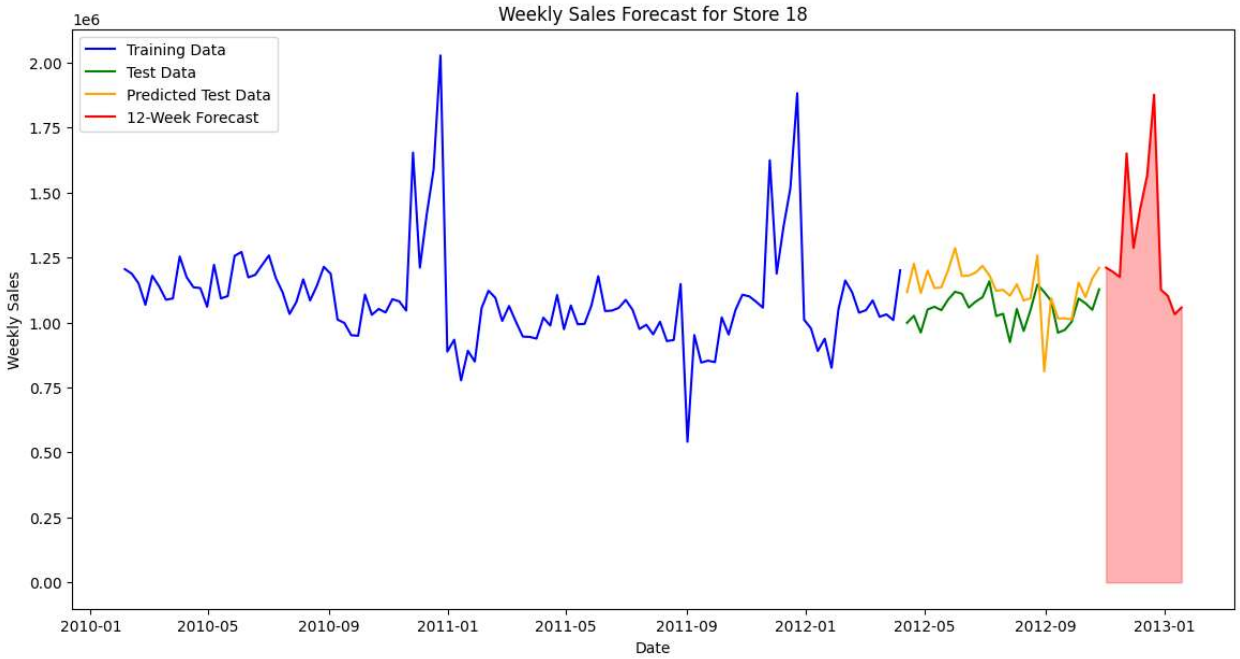

Weekly Sales Forecast for Store 13
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

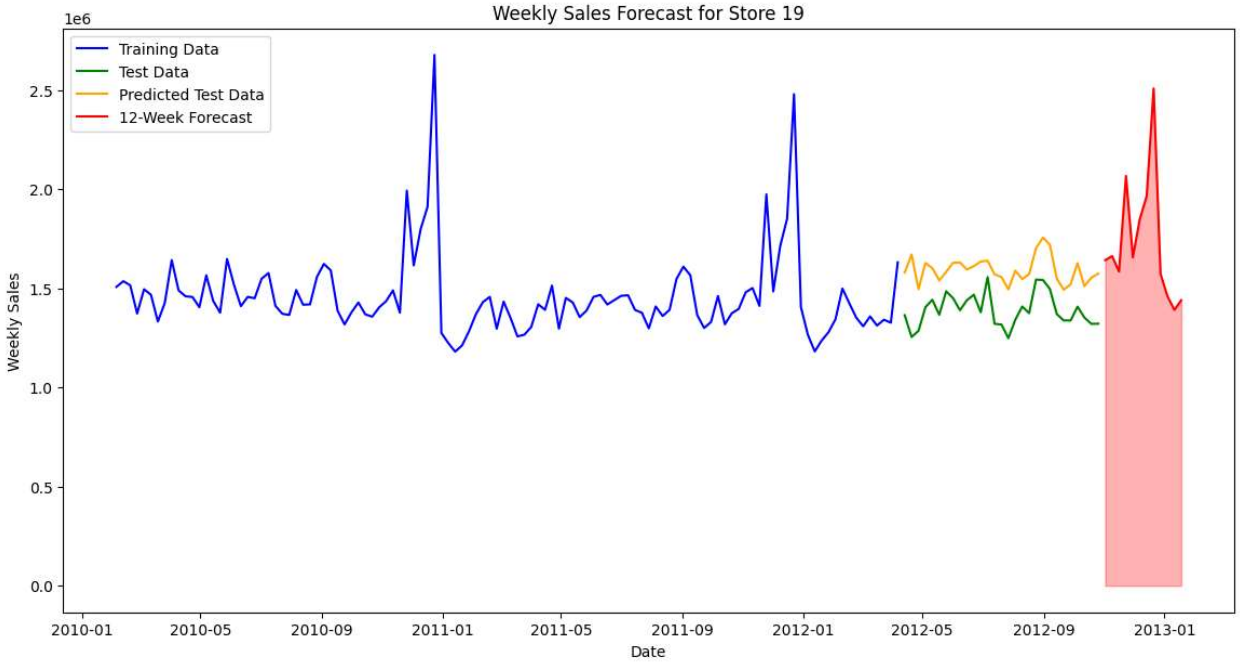

Weekly Sales Forecast for Store 14
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

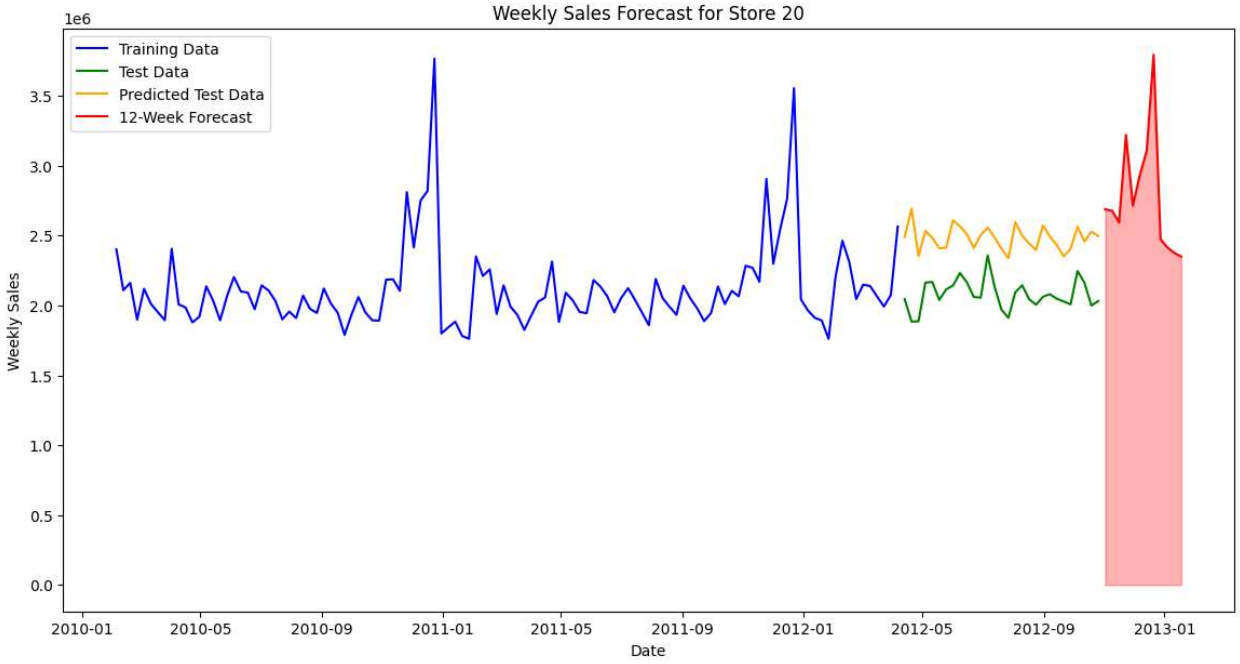

Weekly Sales Forecast for Store 15
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

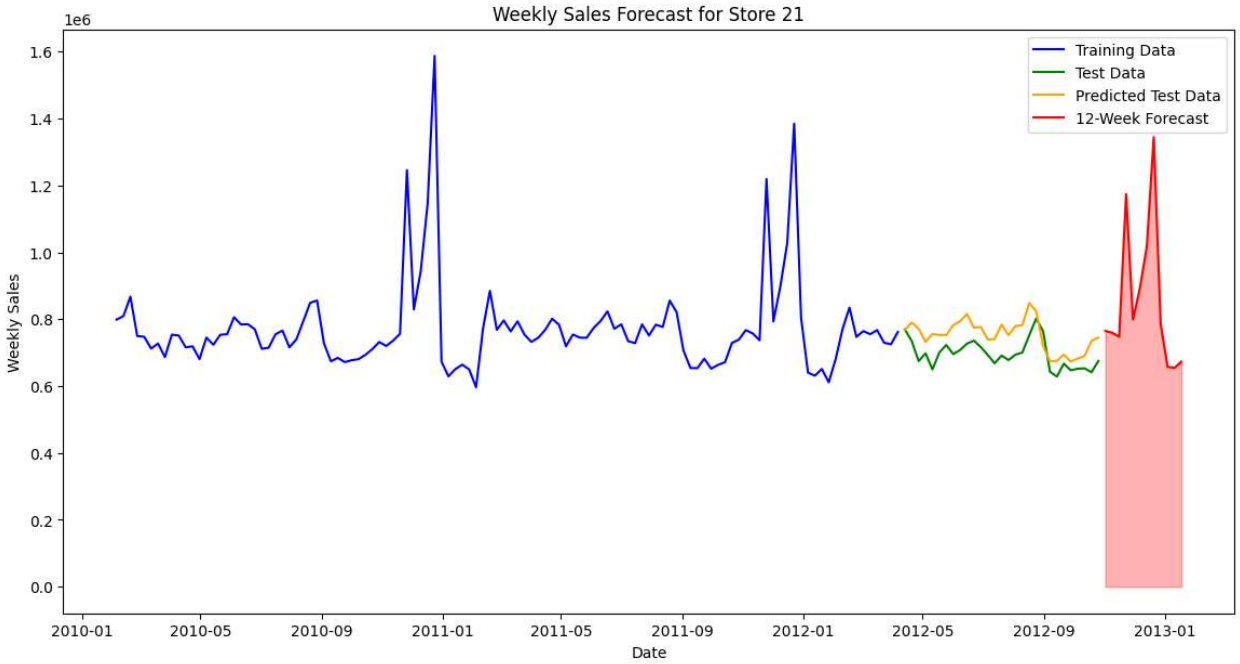

Weekly Sales Forecast for Store 16
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.2
0.4
0.6
0.8
1.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

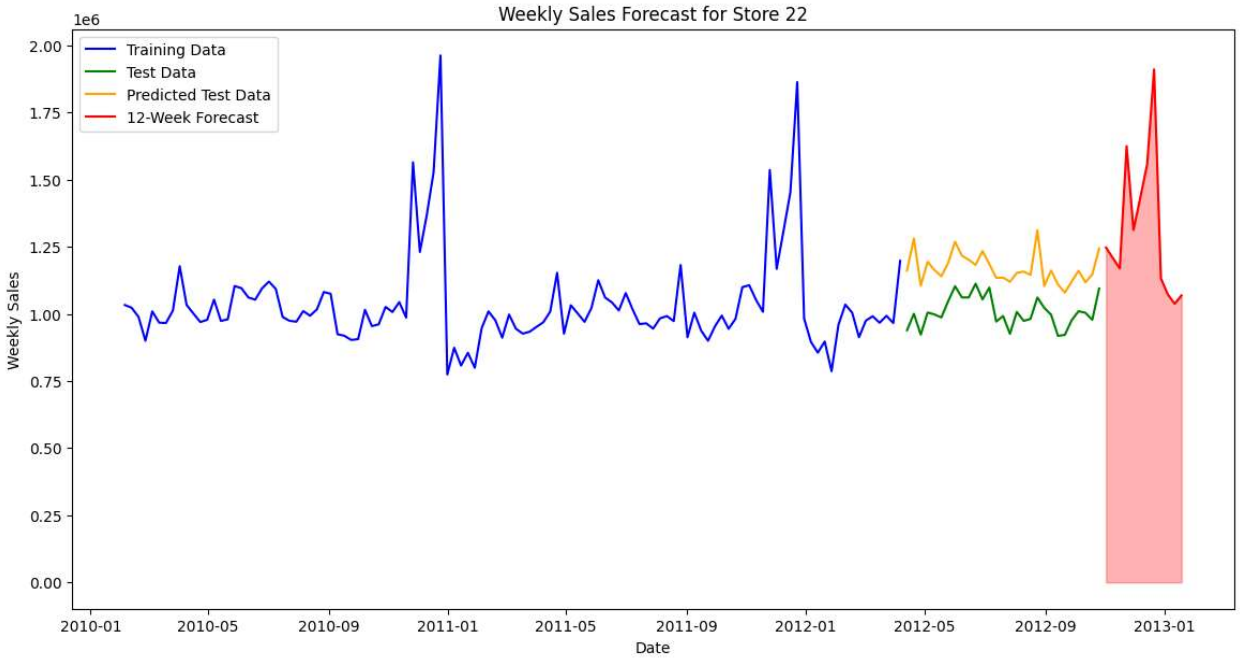

Weekly Sales Forecast for Store 17
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

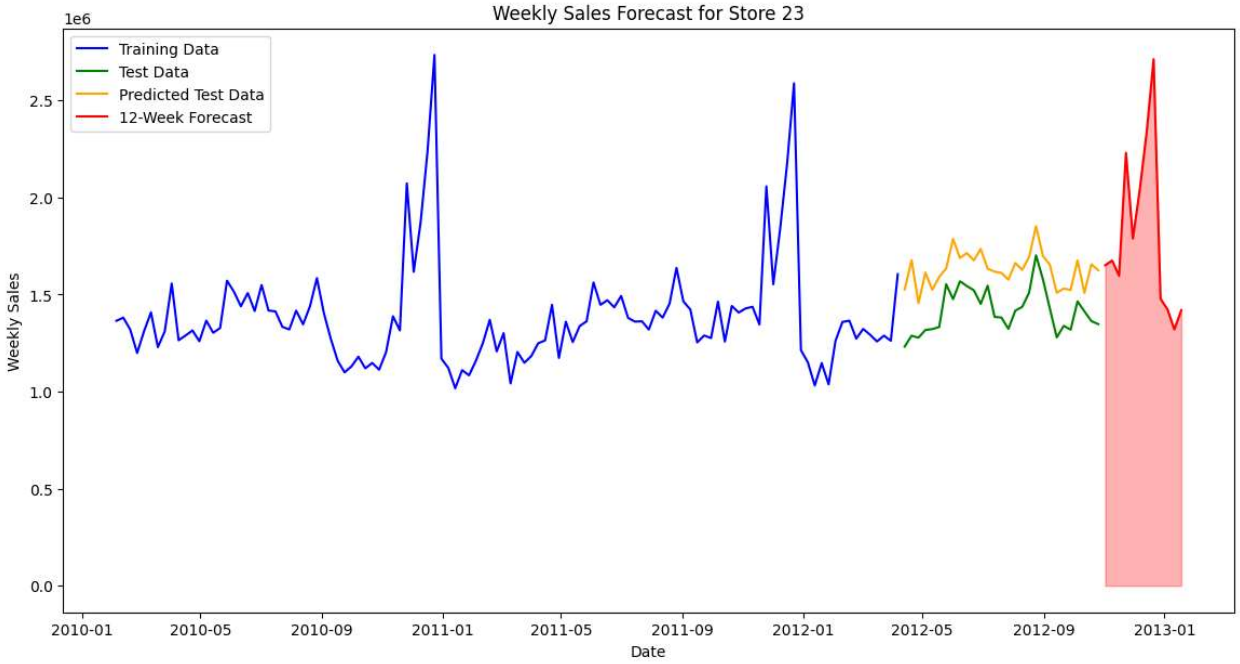

Weekly Sales Forecast for Store 18
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

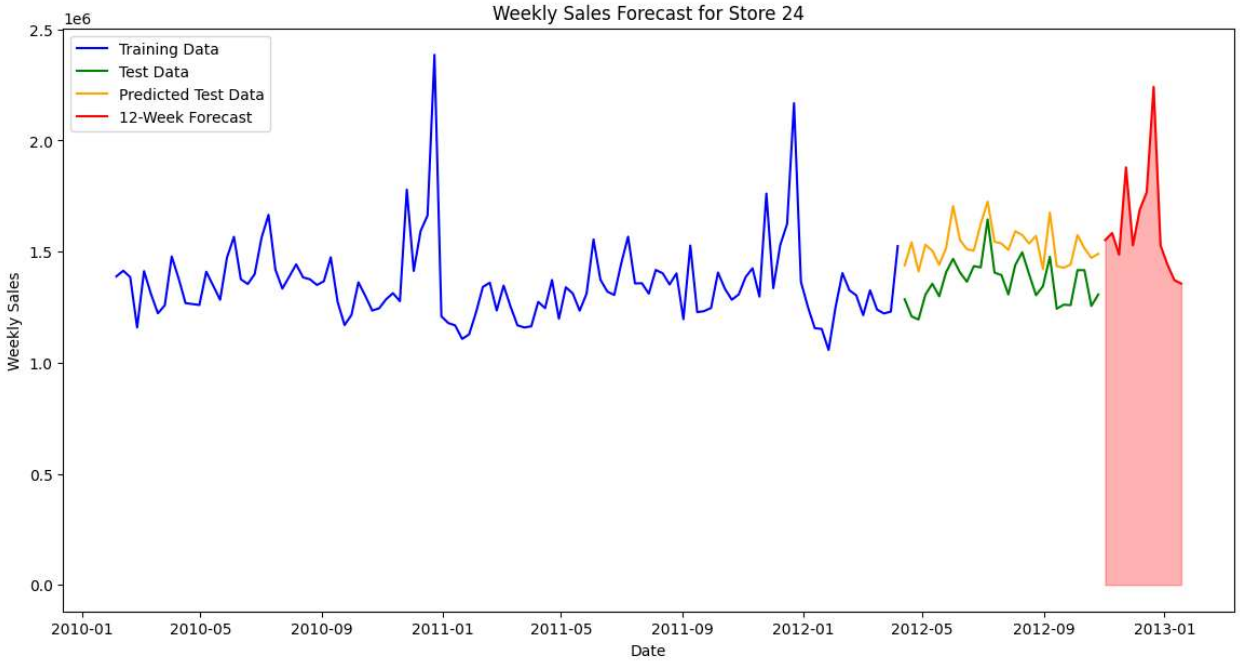

Weekly Sales Forecast for Store 19
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.5
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

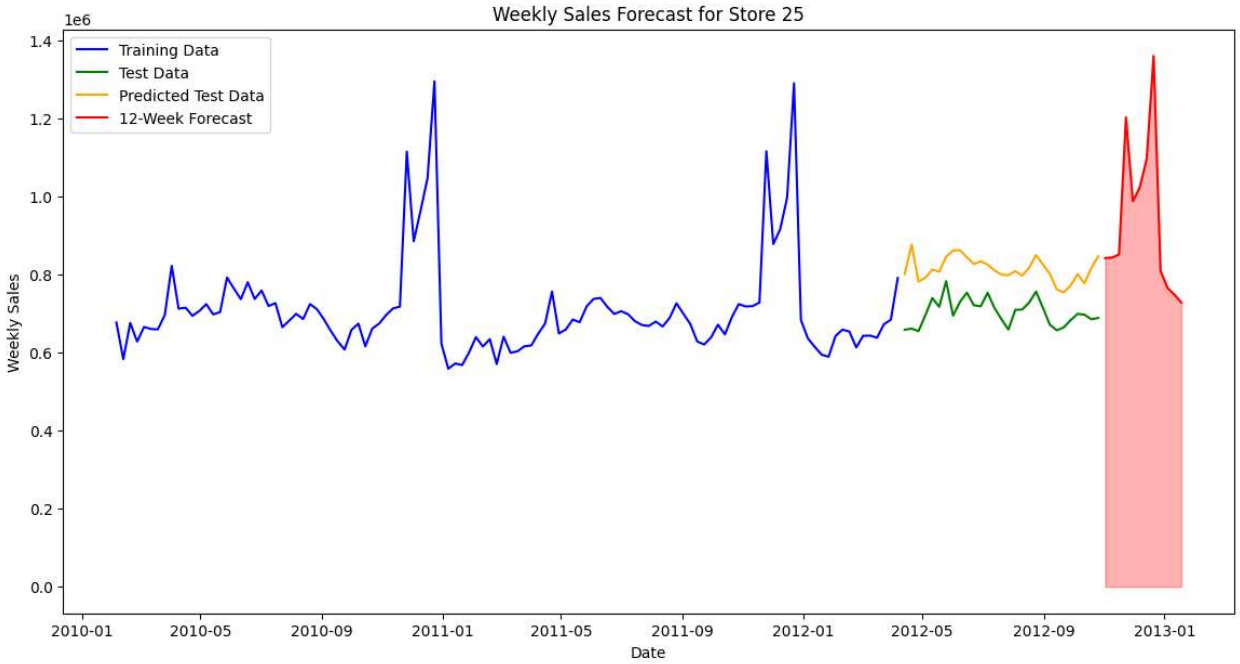

Weekly Sales Forecast for Store 20
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

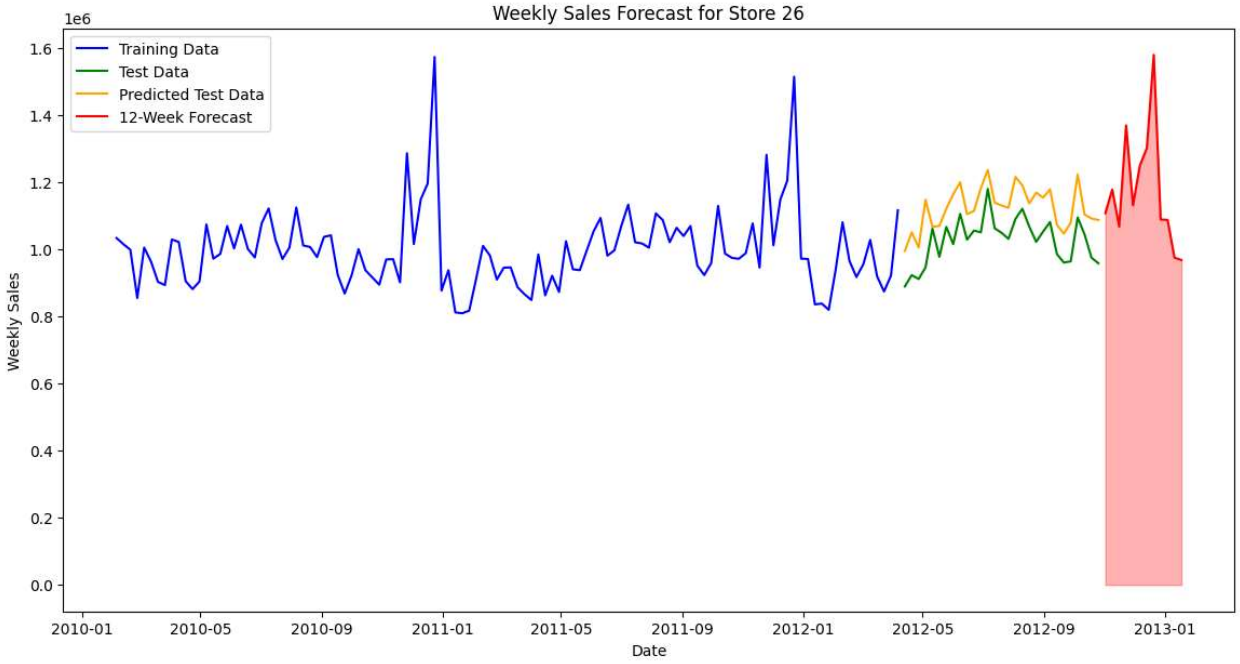

Weekly Sales Forecast for Store 21
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

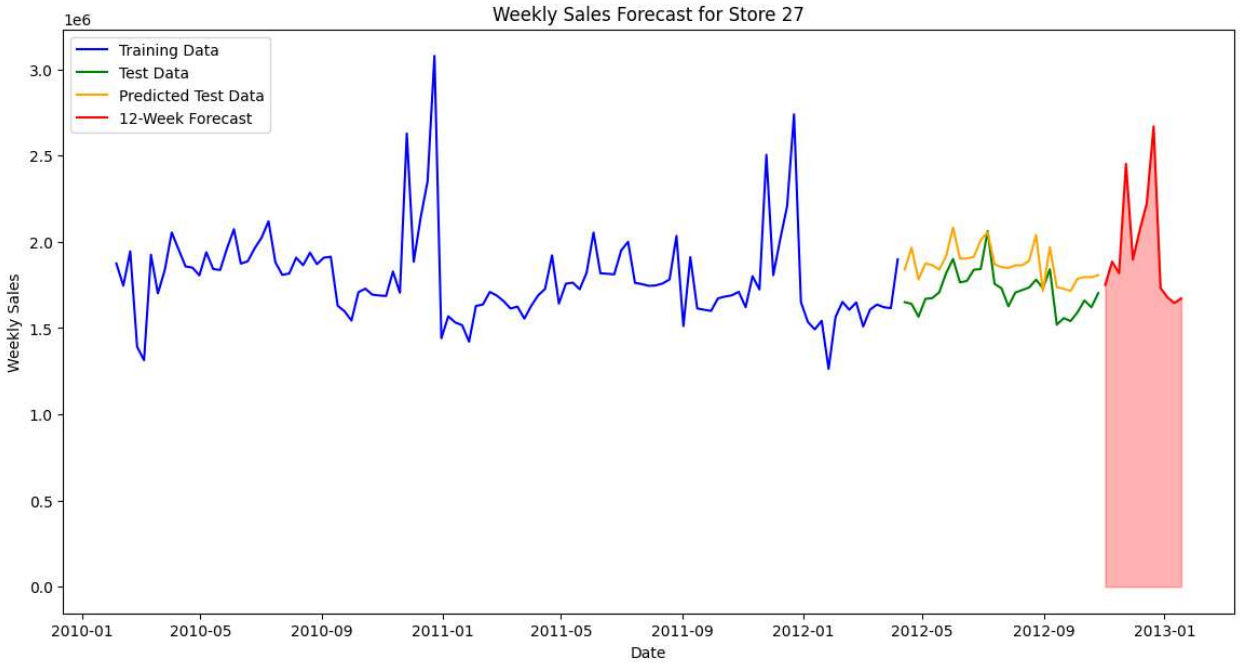

Weekly Sales Forecast for Store 22
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

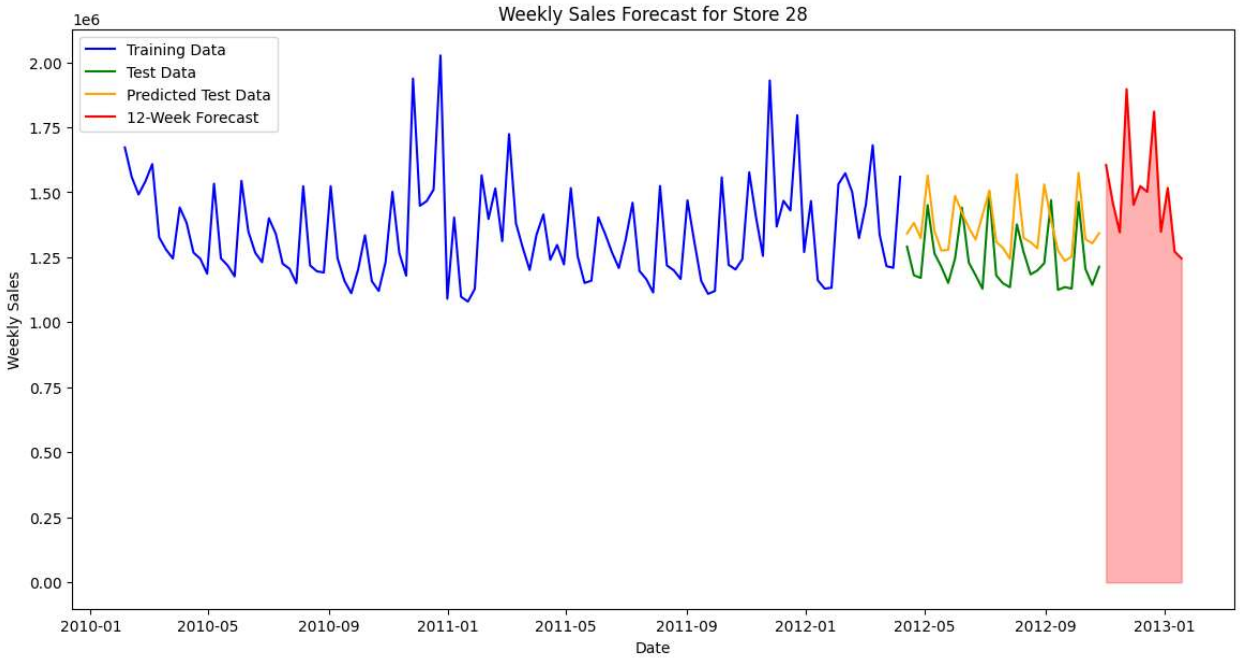

Weekly Sales Forecast for Store 23
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.5
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

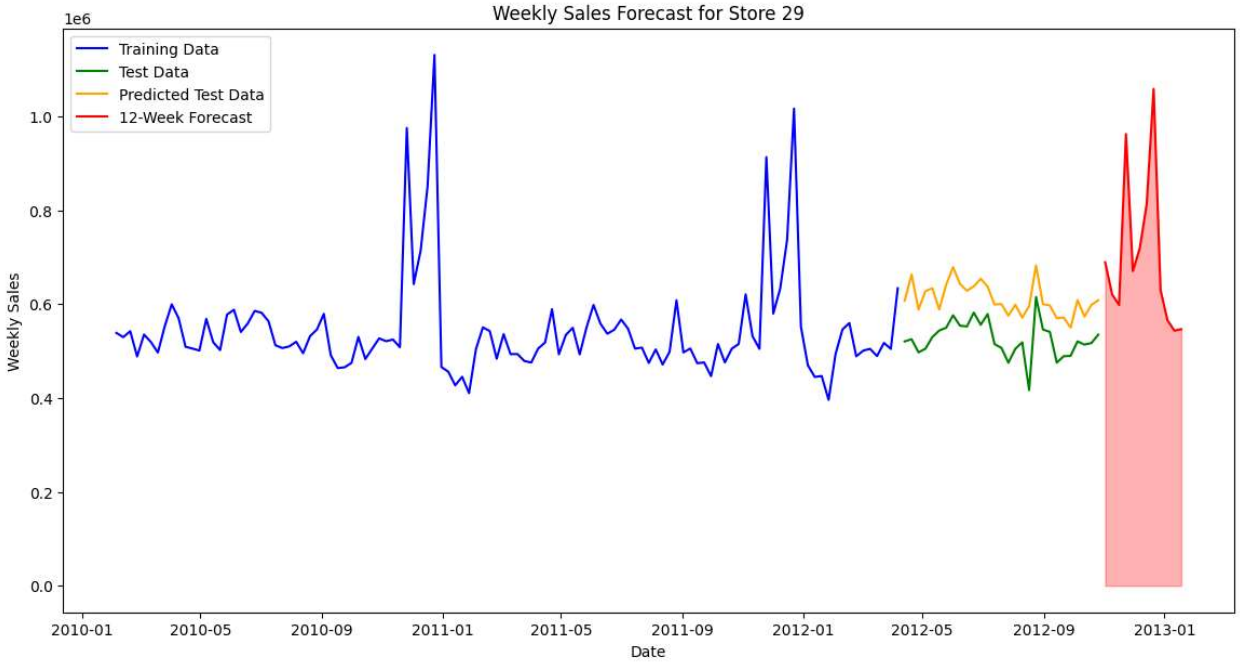

Weekly Sales Forecast for Store 24
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.5
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

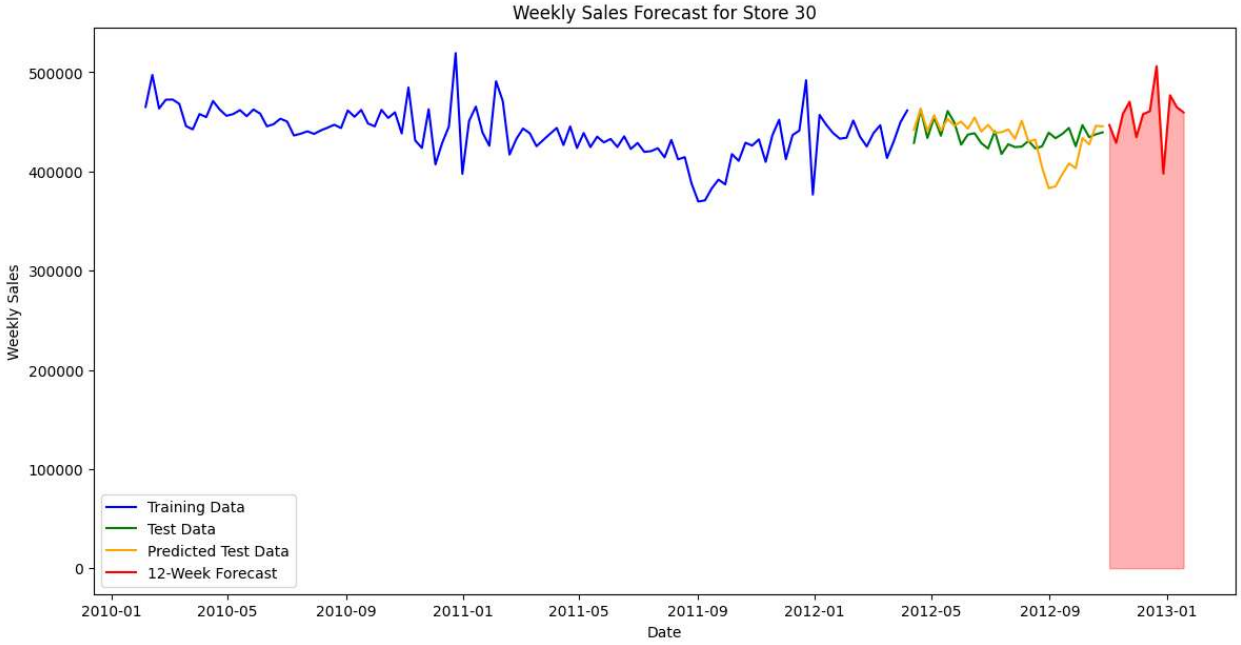

Weekly Sales Forecast for Store 25
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

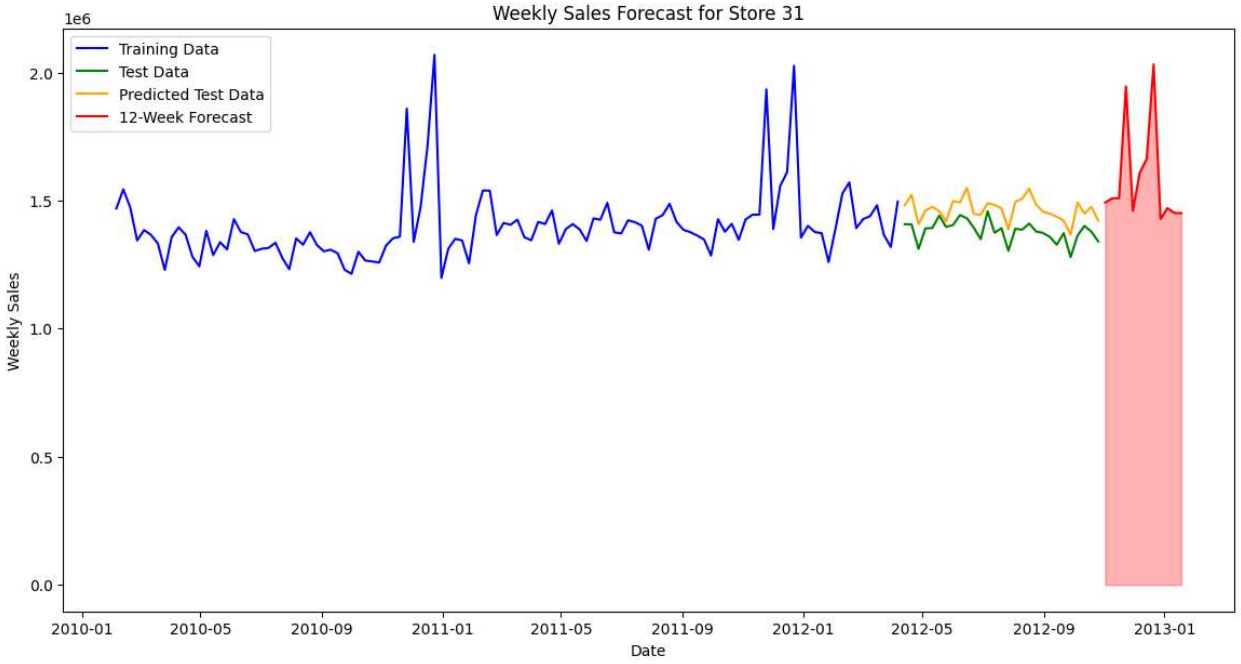

Weekly Sales Forecast for Store 26
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

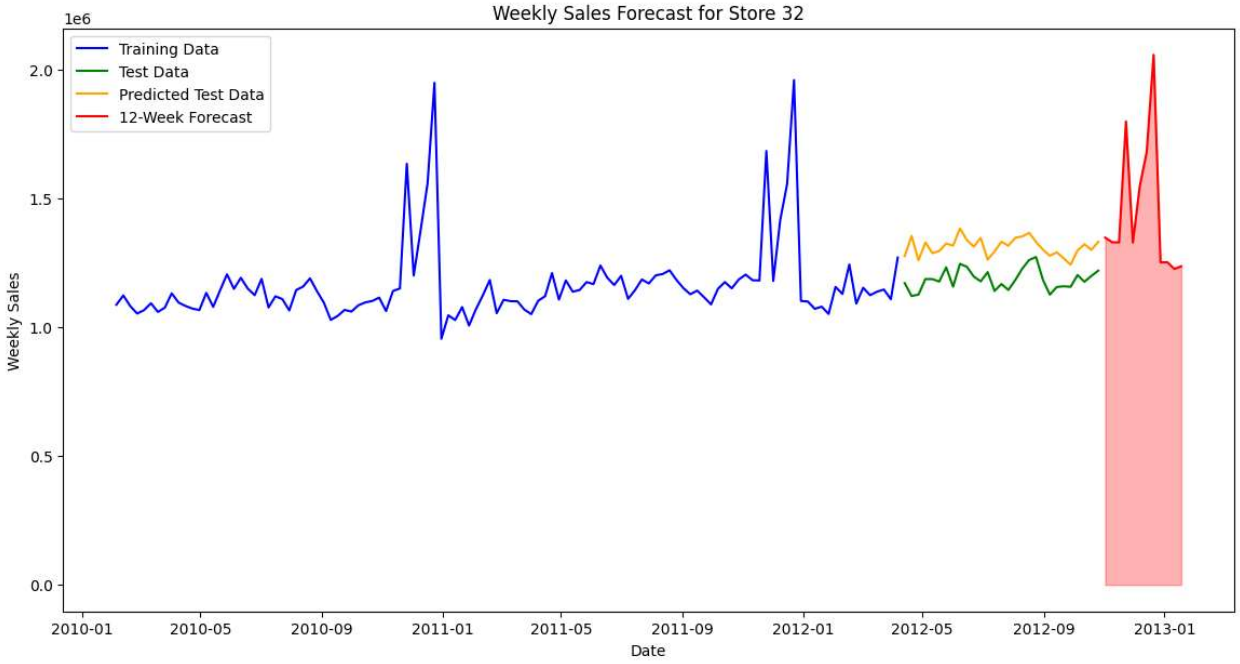

Weekly Sales Forecast for Store 27
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
0.0
0.5
1.0
1.5
2.0
2.5
3.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

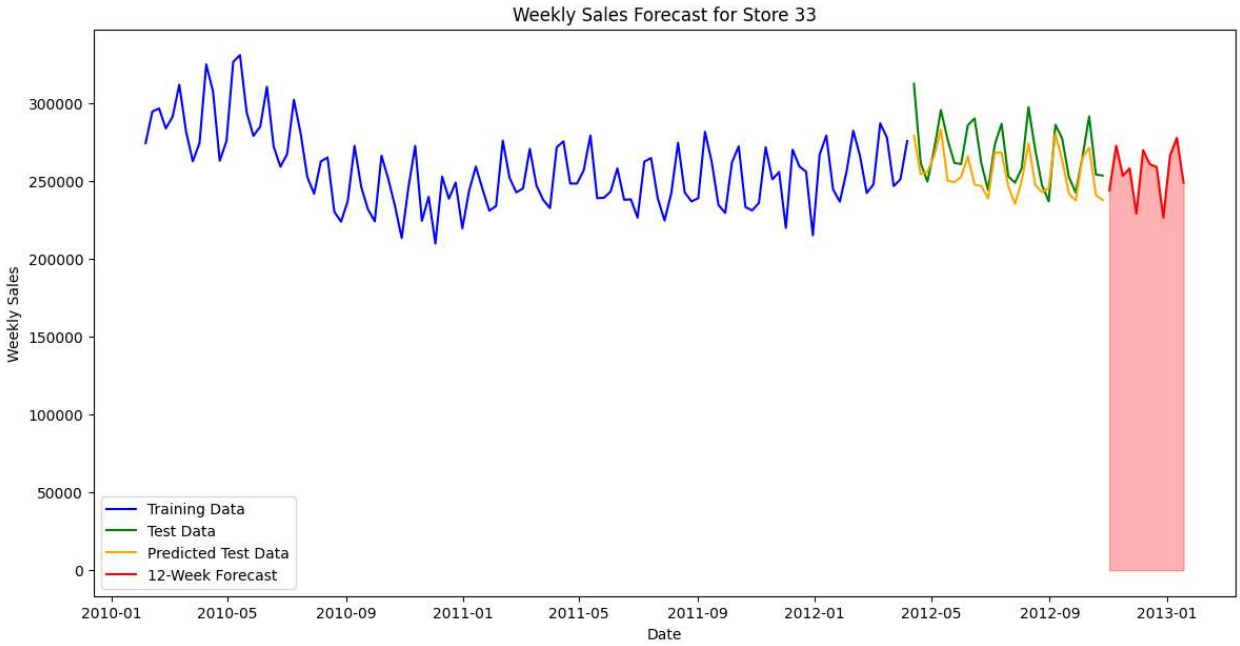

Weekly Sales Forecast for Store 28
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

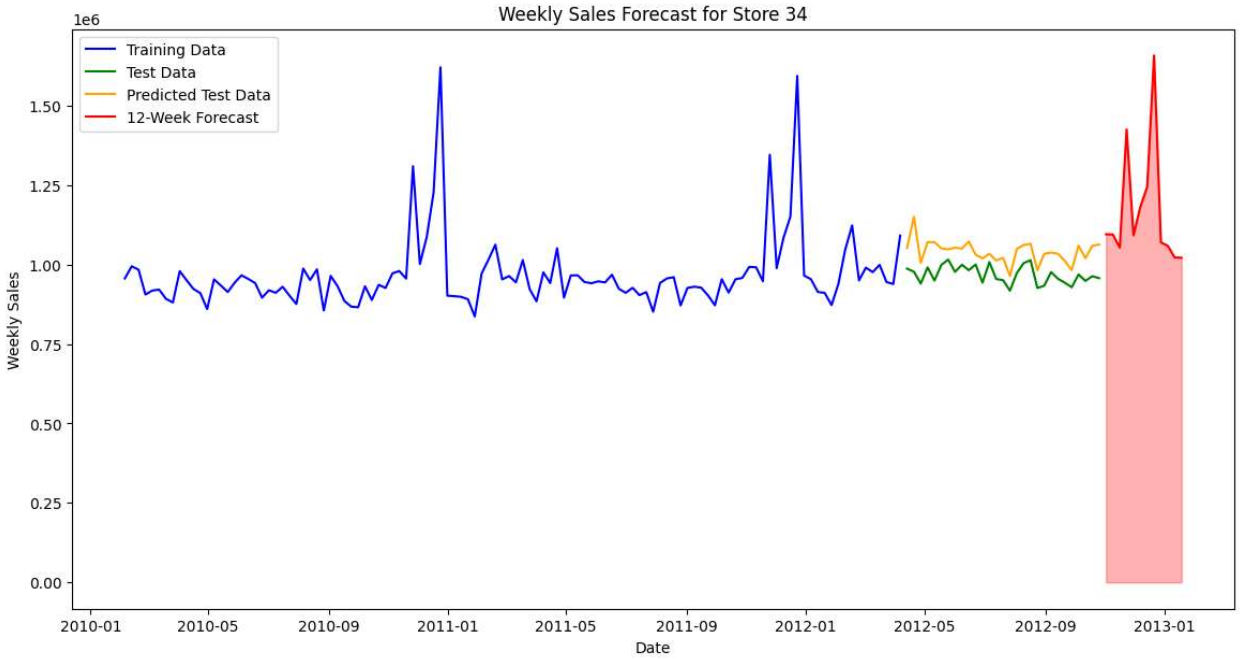

Weekly Sales Forecast for Store 29
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
1.0
0.8
0.6
0.4
0.2
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

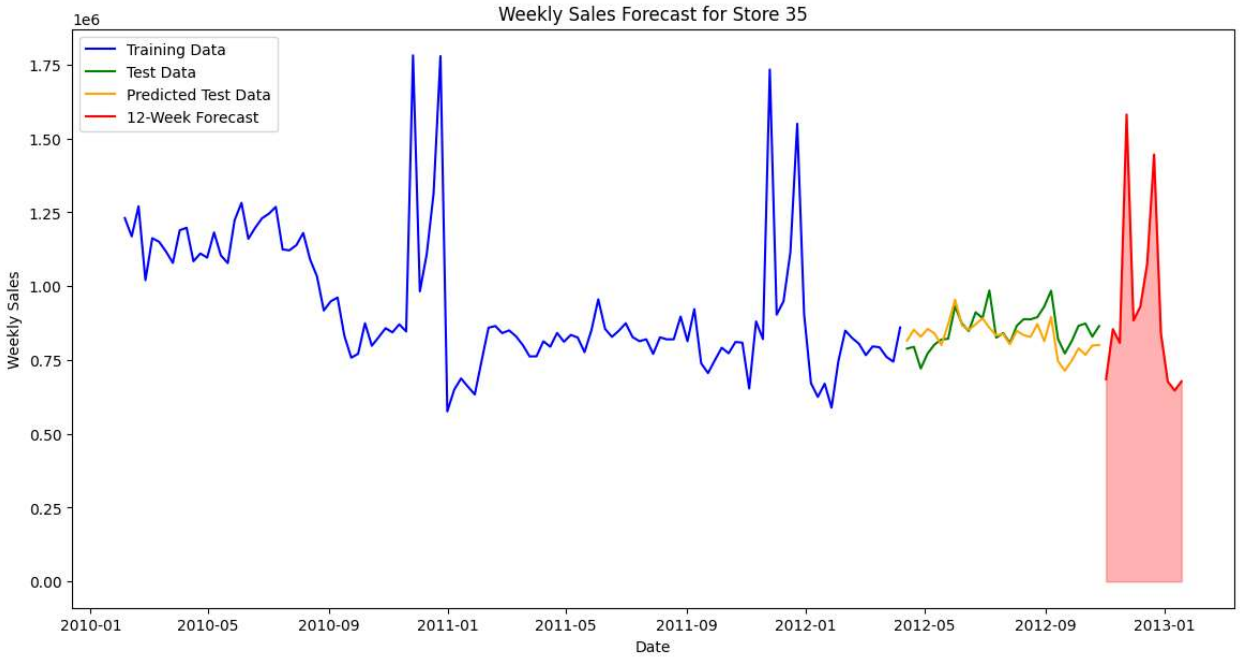

Weekly Sales Forecast for Store 30
Weekly Sales
Date
Training Data
Test Data
Predicted Test Data
12-Week Forecast
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

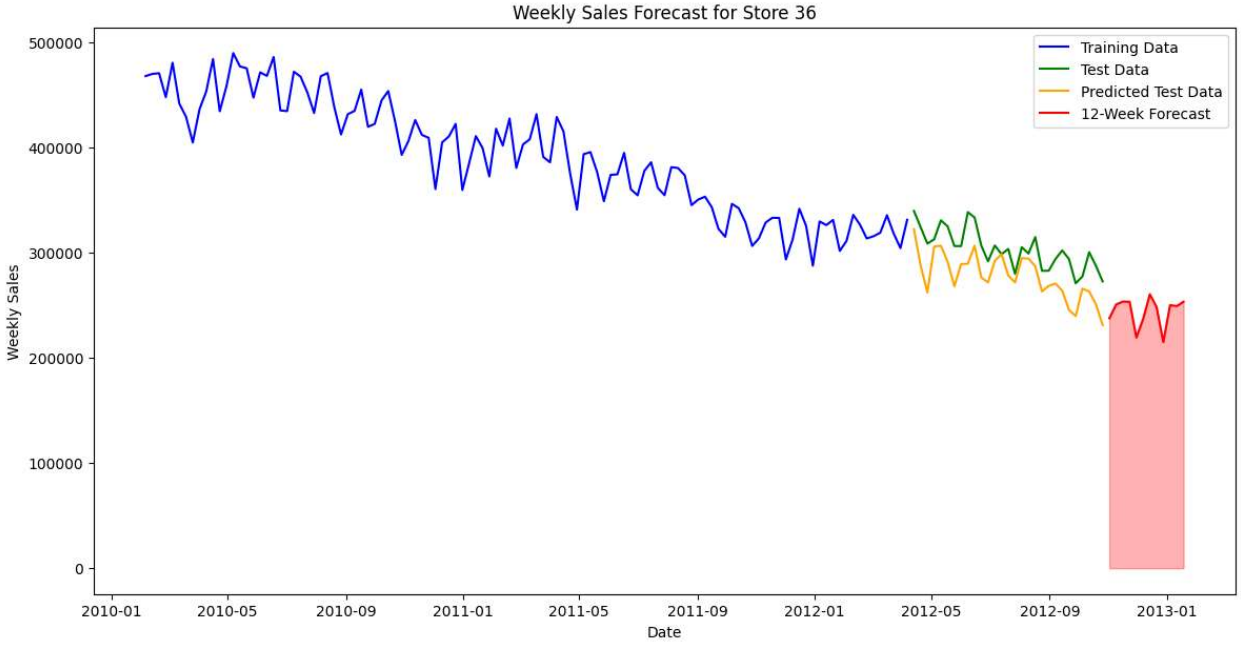

Weekly Sales Forecast for Store 31
1e6
Weekly Sales
Date
Training Data
Test Data
Predicted Test Data
12-Week Forecast
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

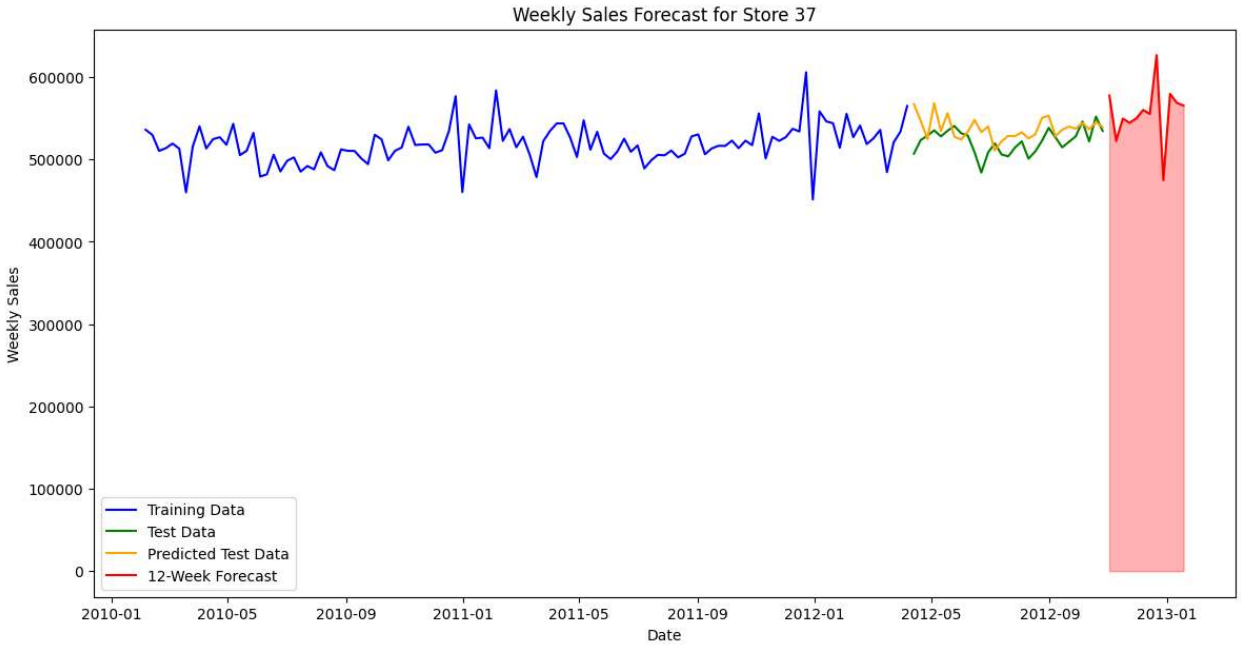

Weekly Sales Forecast for Store 32
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

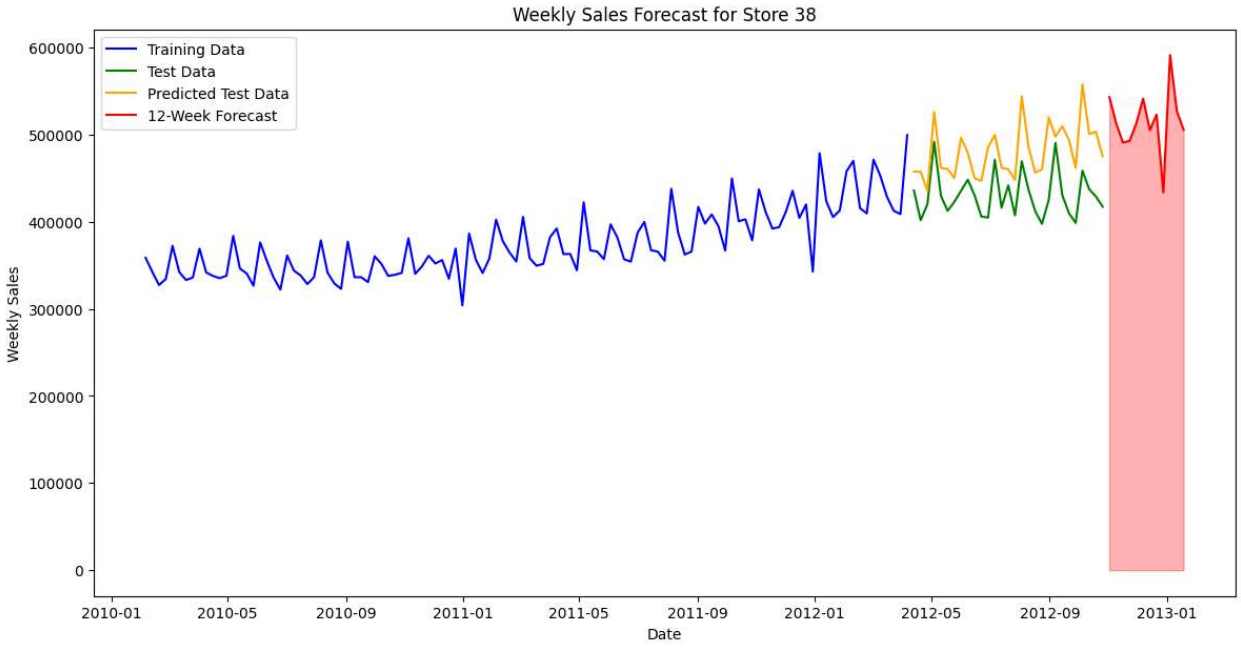

Weekly Sales Forecast for Store 33
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
300000
250000
200000
150000
100000
50000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

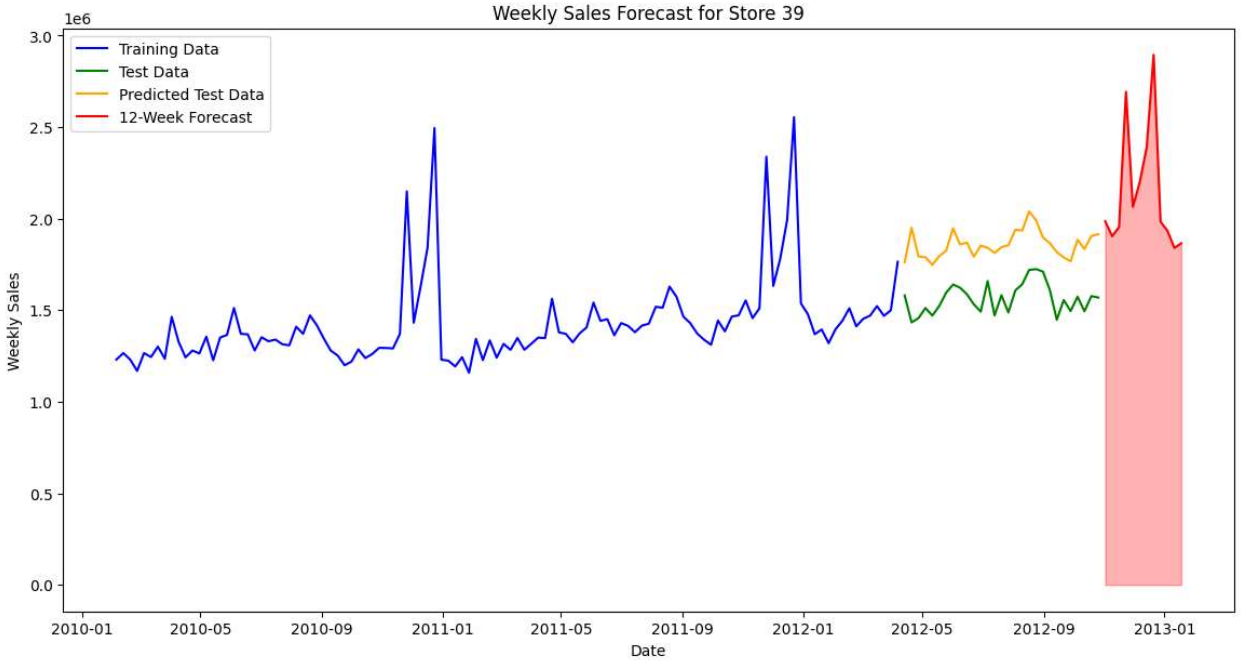

Weekly Sales Forecast for Store 34
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01
Date

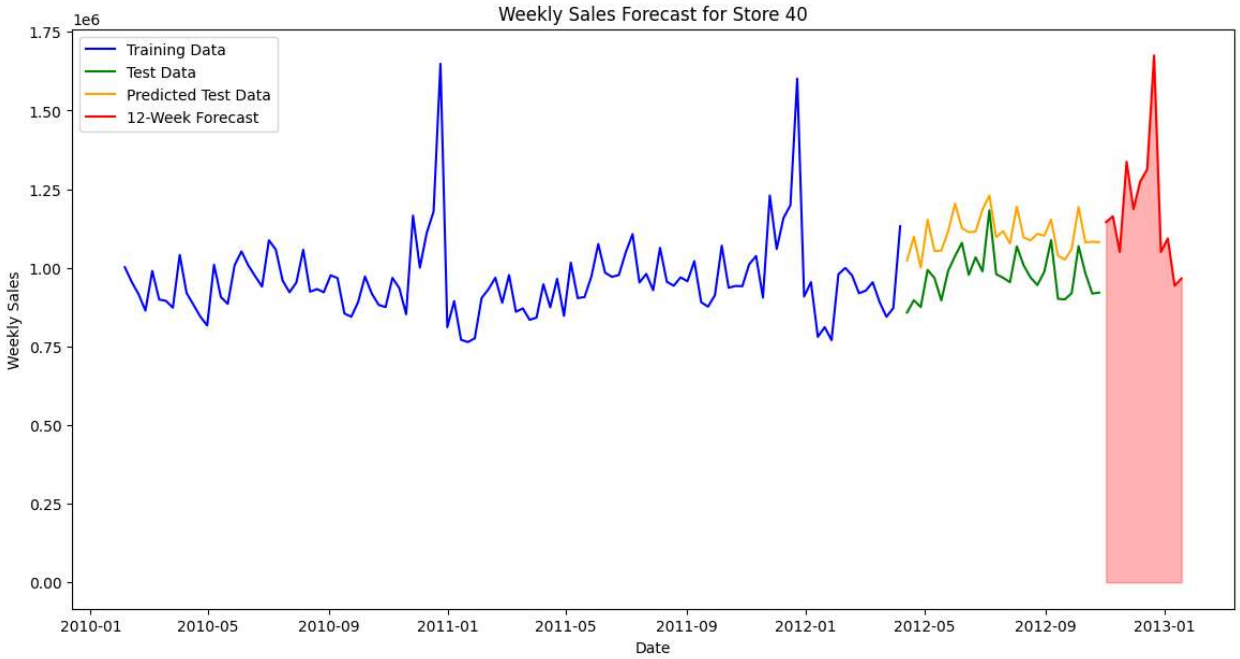

Weekly Sales Forecast for Store 35
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01
Date

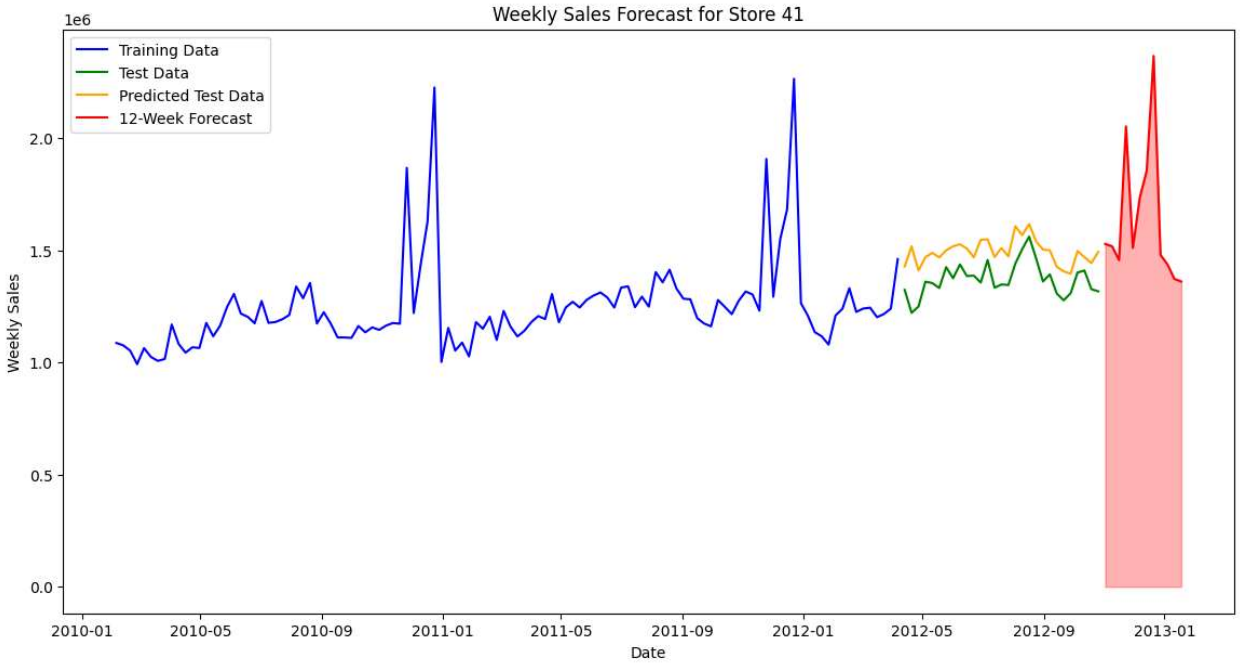

Weekly Sales Forecast for Store 36
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

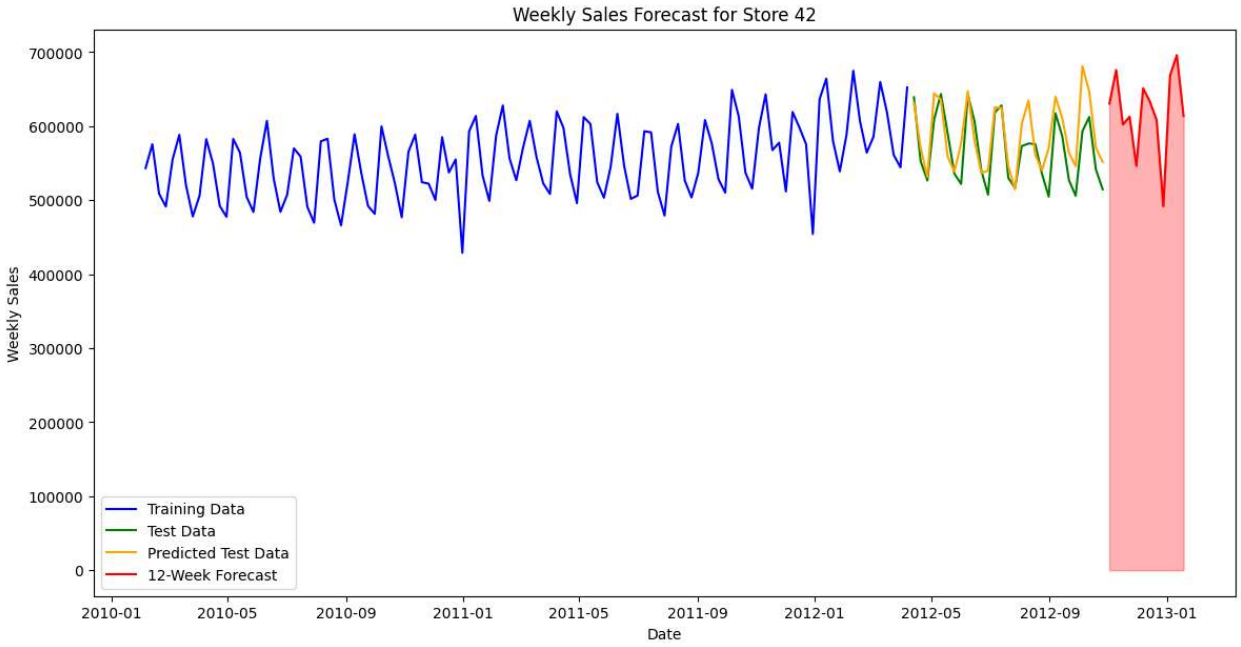

Weekly Sales Forecast for Store 37
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
600000
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

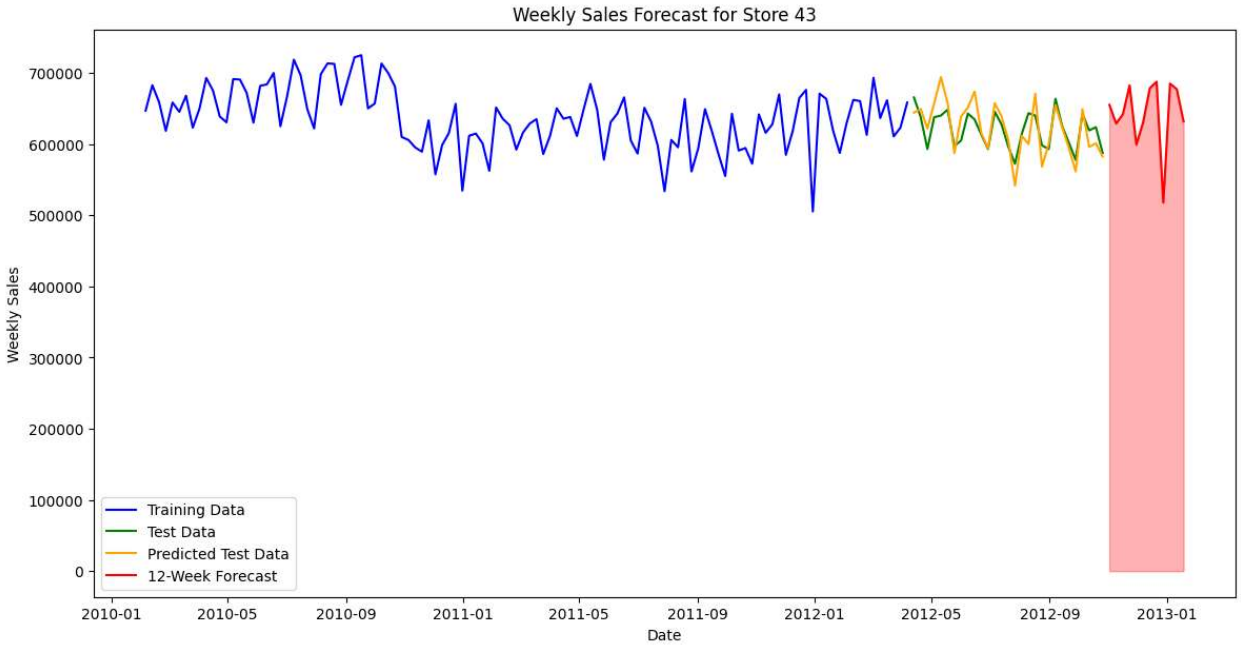

Weekly Sales Forecast for Store 38
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
600000
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

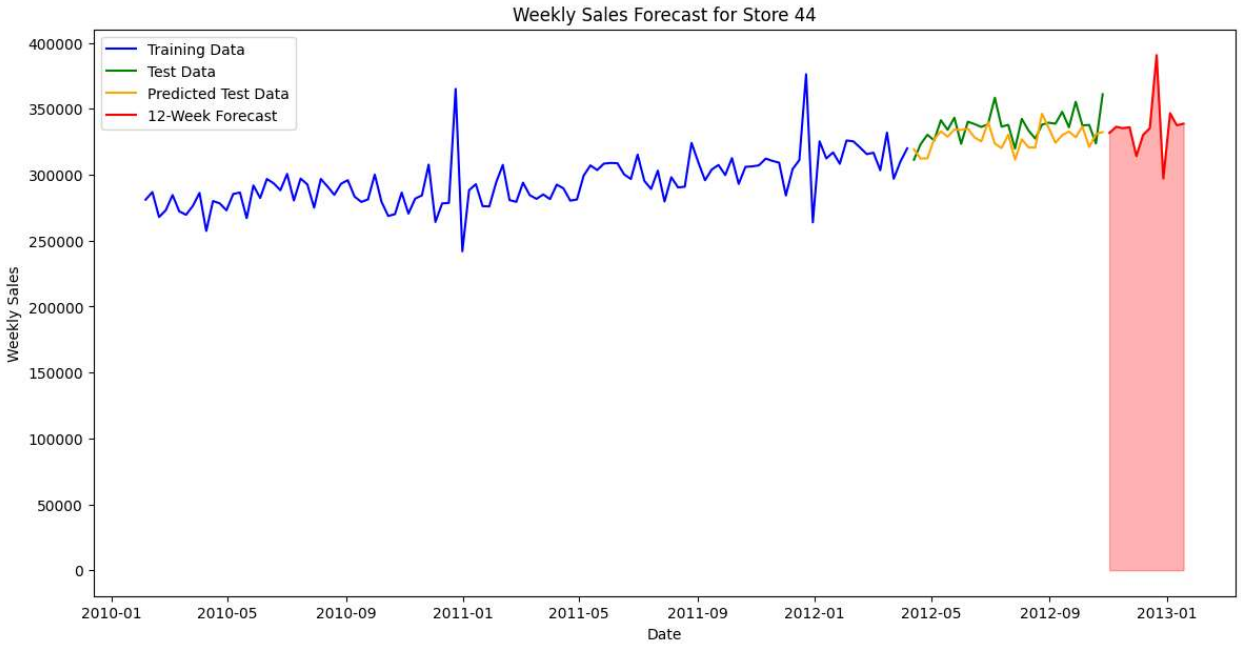

Weekly Sales Forecast for Store 39
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01

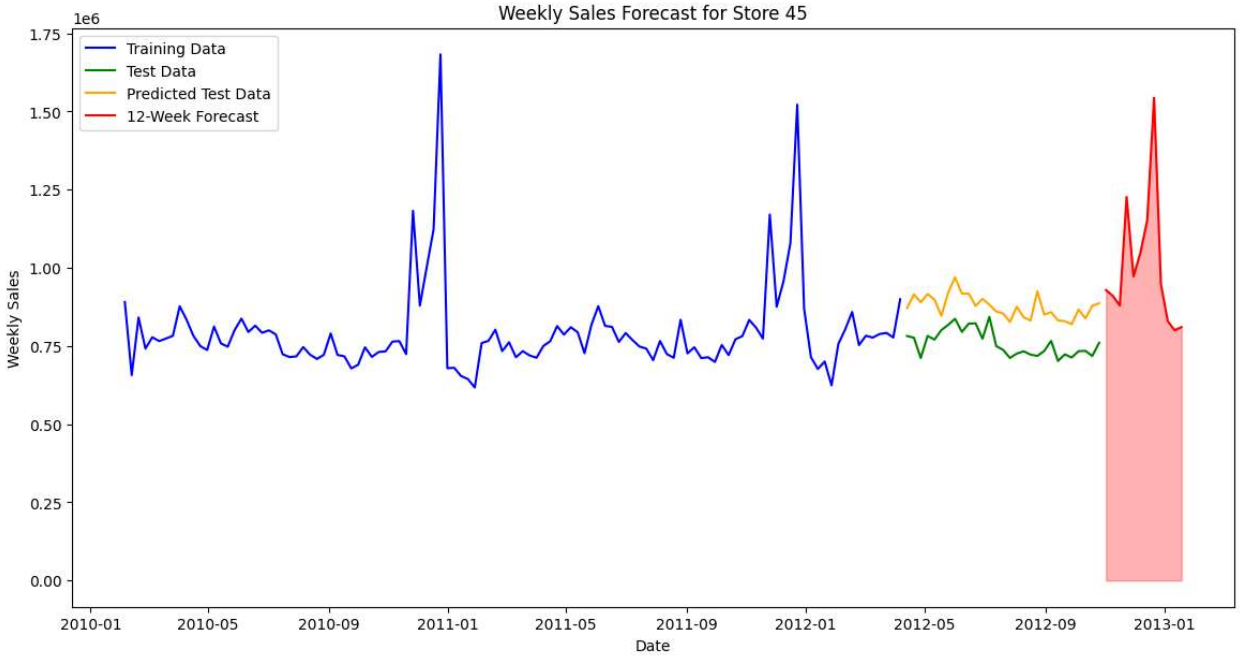

Weekly Sales Forecast for Store 40
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01


Weekly Sales Forecast for Store 41
1e6
Training Data
Test Data
Predicted Test Data
12-Week Forecast
Weekly Sales
Date
2.0
1.5
1.0
0.5
0.0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01


Weekly Sales Forecast for Store 42
Weekly Sales
Date
Training Data
Test Data
Predicted Test Data
12-Week Forecast
700000
600000
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01


Weekly Sales Forecast for Store 43
Weekly Sales
Date
Training Data
Test Data
Predicted Test Data
12-Week Forecast
700000
600000
500000
400000
300000
200000
100000
0
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
2013-01




## 8. Inferences from the Analysis

- **Unemployment Impact**: Although unemployment has a statistically significant negative impact on sales, the effect is relatively weak. This suggests that other factors, such as economic conditions, consumer confidence, and store-specific factors, may play a more significant role in driving sales.

- **Seasonal Trends**: The slight seasonality observed in the sales data indicates the importance of inventory planning around key shopping periods, such as the holiday season.
- **Temperature and CPI**: The weak correlations with temperature and CPI highlight the complexity of sales dynamics and the need for a more nuanced approach to forecasting and inventory management.
- **Forecasting Accuracy**: The SARIMAX model demonstrated good predictive accuracy for most stores, though there is room for improvement in capturing more subtle dynamics in the data.

**9. Future Possibilities for the Project**

- **Enhanced Modeling**: Exploring more advanced modeling techniques, could provide deeper insights into the factors driving sales.
- **Multivariable Analysis**: Incorporating additional variables, such as marketing spend, promotions, and local economic indicators, could improve the accuracy of sales predictions.
- **Store-Specific Strategies**: Developing tailored strategies for stores most affected by unemployment could help mitigate the negative impact and boost sales.
- **Model Refinement**: Further tuning of the SARIMAX model or exploring hybrid models could enhance the predictive accuracy for weekly sales.

**10. Conclusion**

The analysis provided valuable insights into the factors affecting weekly sales at Walmart stores. While unemployment has a weak negative impact on sales, other factors such as seasonality, temperature, and CPI also play a role. The SARIMAX model demonstrated good performance in forecasting future sales, which can be utilized for better inventory management, optimizing store performance, and aligning supply with demand.

**11. References**


- Walmart Weekly Sales Data (Provided dataset)
- Statsmodels API documentation for OLS regression analysis
- Seaborn and Matplotlib documentation for data visualization techniques
- Time series analysis resources for SARIMAX model implementation