Walmart Sales Prediction

Table of contents

01

Problem Statement

02

Project Objective

03

Data Description

04

Data Pre-processing Steps and Inspiration

05

Choosing the Algorithm

06

Reasons for Choosing the Algorithm

07

Assumptions

08

Model Evaluation and Techniques

09

Inferences from the Same

10

Future Possibilities

11

Conclusion

12

References

Problem Statement

A retail store with multiple outlets across the country is experiencing significant challenges in managing its inventory. These issues are primarily related to accurately matching the demand for products with the available supply at each outlet. The misalignment between demand and supply has resulted in various operational inefficiencies, impacting both customer satisfaction and the store's profitability.
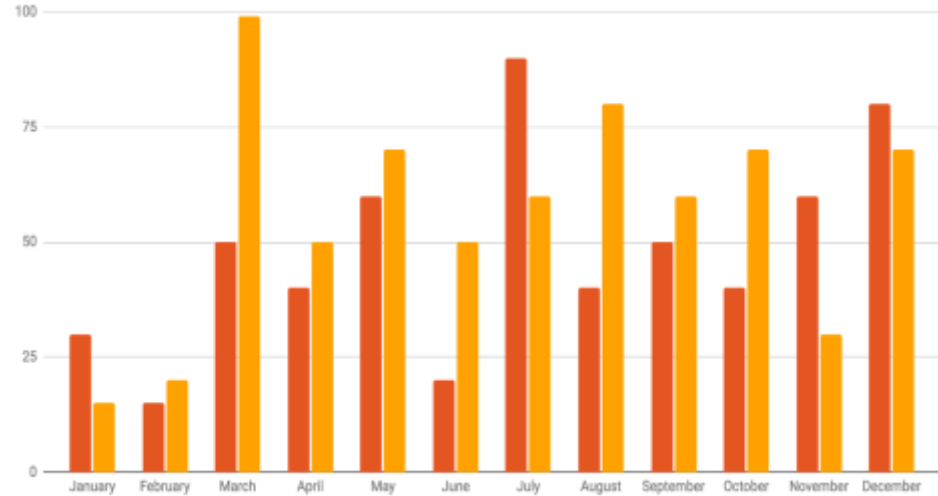
Problem Object

The objective of this project is to conduct a comprehensive analysis of the retail chain's sales data to understand the underlying factors affecting weekly sales and to develop predictive models for forecasting future sales. Specifically, the project aims to address the following key questions:

Unemployment Rate Impact: Investigate if the weekly sales are influenced by the unemployment rate and identify which stores are most affected by changes in unemployment.

Seasonal Trends: Determine whether weekly sales exhibit a seasonal trend, identify the time periods of these trends, and understand the potential reasons behind these seasonal patterns.

Temperature Influence: Assess the effect of temperature on weekly sales and analyze if there is any significant relationship between temperature variations and sales performance.

Consumer Price Index (CPI) Effect: Examine how the Consumer Price Index affects weekly sales across different stores, and determine the extent of its impact.

Store Performance Analysis: Identify the top-performing stores based on historical sales data and evaluate the factors contributing to their success.

Worst Performing Store: Determine the worst performing store and analyze the significance of the difference between the highest and lowest performing stores.

Sales Forecasting: Utilize predictive modeling techniques to forecast the sales for each store for the next 12 weeks, providing actionable insights for inventory management and strategic planning.

By addressing these objectives, the project seeks to uncover the key drivers of sales performance, identify areas for improvement, and provide reliable sales forecasts to enhance the retail chain's inventory management and operational efficiency.

Data Description

DataSet Details

Feature	Description
Store	Store number
Date	Week of Sales
Weekly_Sales	Sales for the given store in that week
Holiday_Flag	If it is a holiday week
Temperature	Temperature on the day of the sale
Fuel_Price	Cost of the fuel in the region
CPI	Consumer Price Index
Unemployment	Unemployment Rate

Data Pre-processing Steps and Inspiration

1.Loading the Data: Import the dataset and examine its structure to understand the available features and their types.

2.Handling Missing Values: Identify and handle any missing values in the dataset. This can involve techniques such as imputation or removing rows/columns with missing data.

3.Data Type Conversion: Ensure that all columns have appropriate data types. Convert the 'Date' column to date time format.

4.Feature Engineering: Create new features if necessary, such as extracting month and year from the 'Date' column to help identify seasonal trends.

5.Outlier Detection and Treatment: Identify and treat outliers in the data, as they can significantly impact the performance of predictive models.

6.Aggregation and Grouping: Aggregate data as required for analysis, such as summing weekly sales per store, and grouping data by different features for deeper insights.

Data Pre-processing Steps and Inspiration

Unemployment Rate Impact: Investigate if the weekly sales are influenced by the unemployment rate and identify which stores are most affected by changes in unemployment.
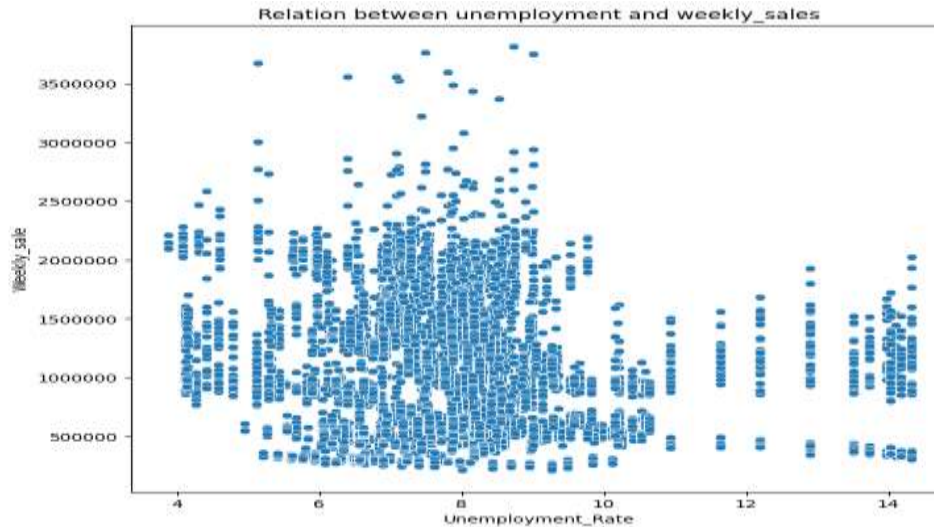
In the scatter plot, it is evident that as the unemployment rate increases, the weekly sales decrease. This observation suggests a negative correlation between unemployment rate and weekly sales

Data Pre-processing Steps and Inspiration

Unemployment Rate Impact: identify which stores are most affected by changes in unemployment.

Weekly_sales as per store
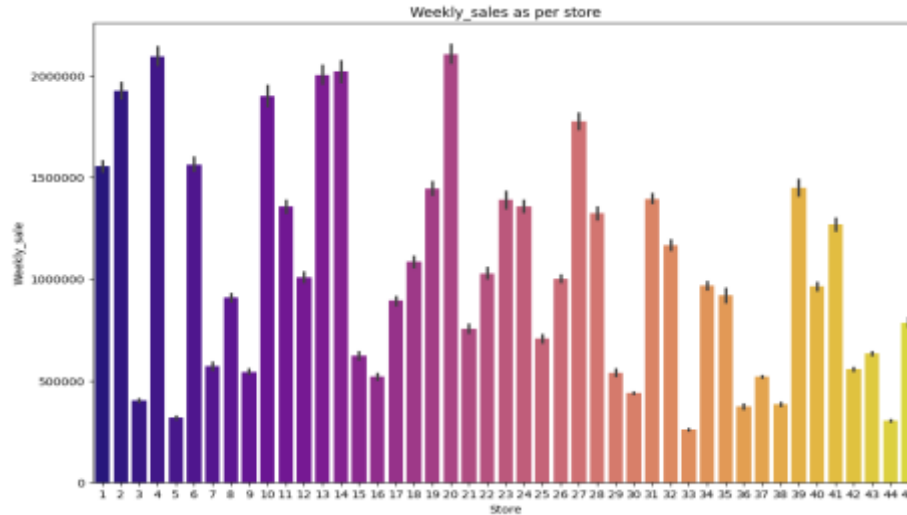
In the bar plot, it's evident that Store Number 33 exhibits lower weekly sales compared to other stores. This insight can be gleaned from the relative height of the bar representing Store 33, which is noticeably shorter in comparison to bars representing other stores.
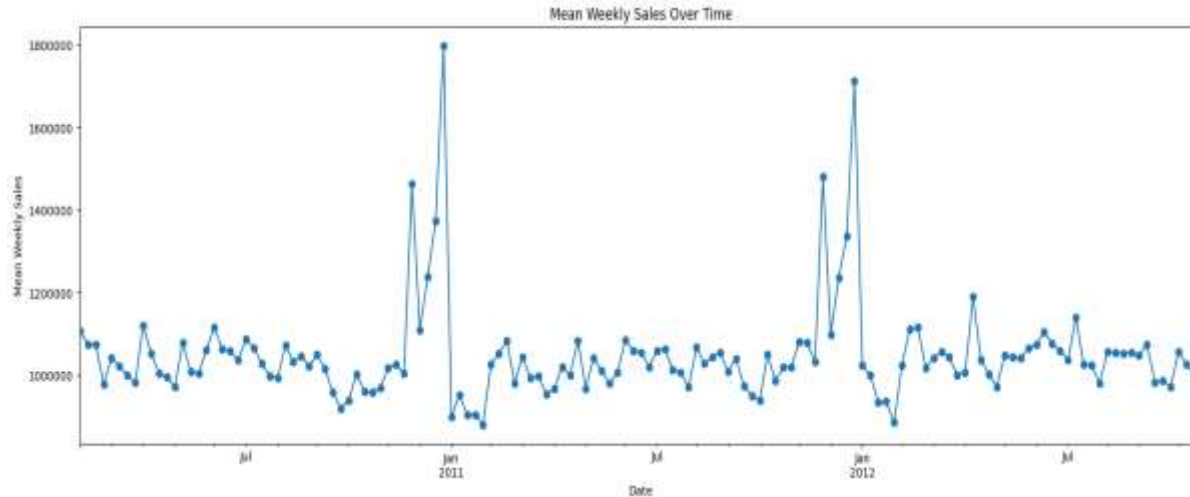
Data Pre-processing Steps and Inspiration

Seasonal Trends: Determine whether weekly sales exhibit a seasonal trend, identify the time periods of these trends, and understand the potential reasons behind these seasonal patterns.
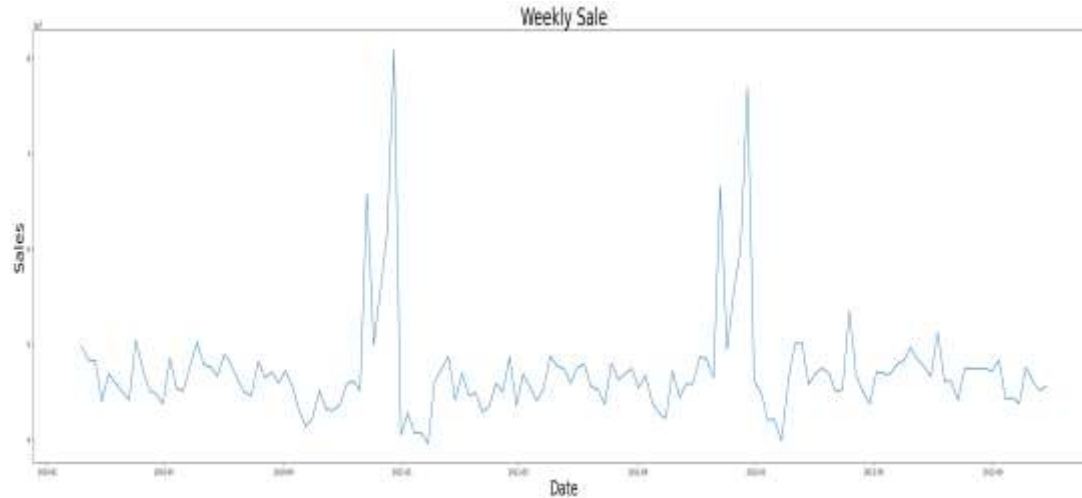
Weekly Sale

Over a two-year period, the company's weekly sales displayed a seasonal tendency that appears in the picture. The tendency expresses itself throughout the year as recurrent cycles that occur regularly at prearranged intervals. Notably, sales exhibit winter and summer peaks as well as spring and fall decreases. Constant cyclical fluctuation, indicating predictable differences in sales performance at different times of the year, is a defining characteristic of seasonal patterns.

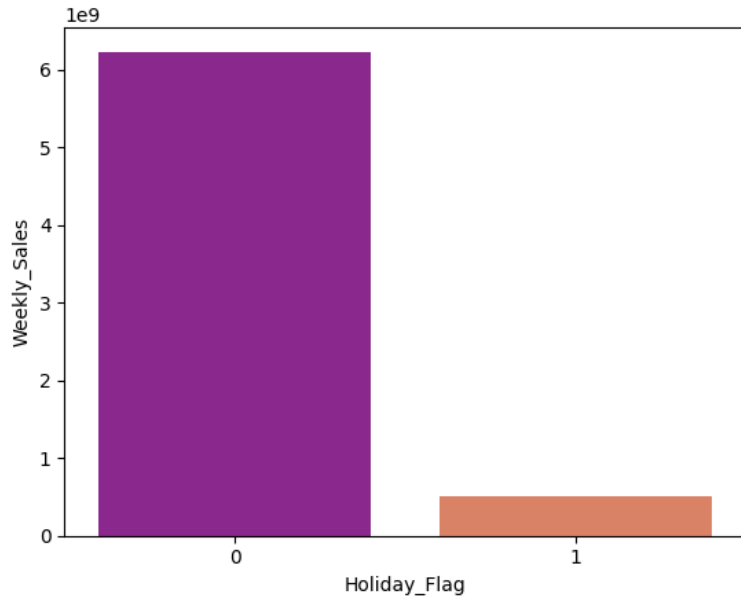# Data Pre-processing Steps and Inspiration

Seasonal Trends: Determine whether weekly sales exhibit a seasonal trend, identify the time periods of these trends, and understand the potential reasons behind these seasonal patterns.

Mean Weekly Sales Over Time

Since we can readily get wintertime sales, which are at their peak from November to January, I have calculated an average weekly sales throughout time.

Data Pre-processing Steps and Inspiration

Impact of Holidays on Sales: Analyzing how holiday weeks influence sales and understanding store performance during these periods.

The Sales are High In Normal Days, it is Quite common because of that holidays are always low compared to working days.

Data Pre-processing Steps and Inspiration

Temperature Influence: Assess the effect of temperature on weekly sales and analyze if there is any significant relationship between temperature variations and sales performance.

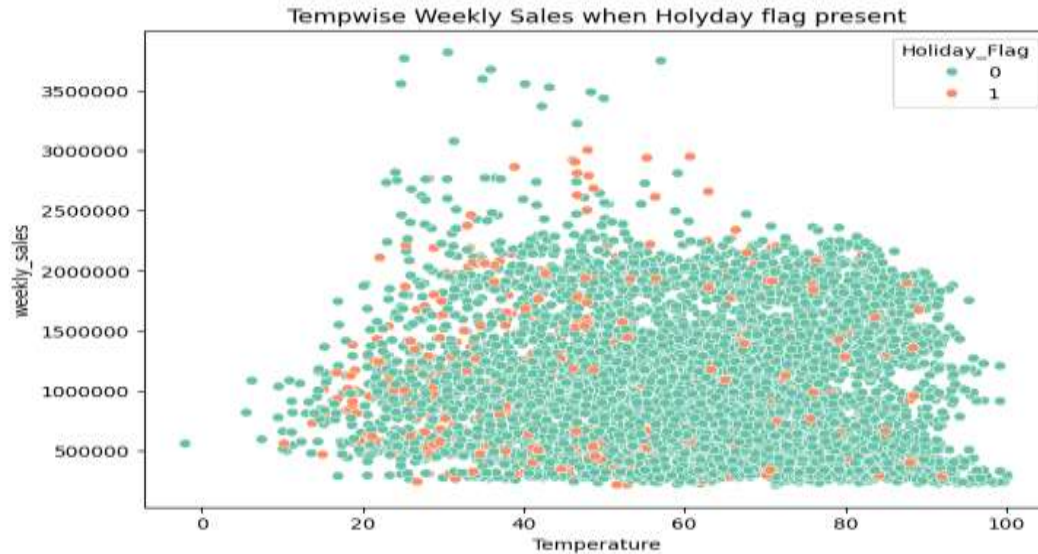Temperature vs. Weekly Sales
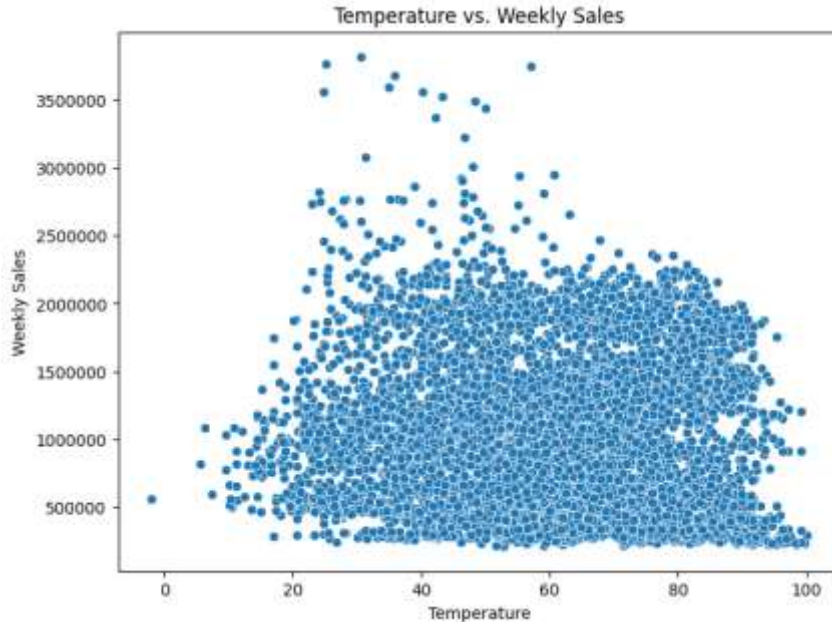
Stores perform well in moderate temp that is 26-50
Sales went peaks in Moderate to High temp that is 51-75
Sales went down in Very High and Low temp that is 0-25 & 76-100

Positive Correlation between the temp and weekly sales ,meaning that sales tend to be higher when the temp is higher. For example ,people may be more likely to go shopping when the weather is nice.

# Data Pre-processing Steps and Inspiration

Weekly Sales by Temperature When Holiday Flag is Active: This analysis explores how weekly sales vary with temperature during holiday weeks

Overall the graph shows that there is positive correlation between temperature and weekly sales ,but there is lot of scatter in the data.
The holyday flag does not seem to have strong effect on weekly sales.

Data Pre-processing Steps and Inspiration

Consumer Price Index (CPI) Effect: Examine how the Consumer Price Index affects weekly sales across different stores, and determine the extent of its impact
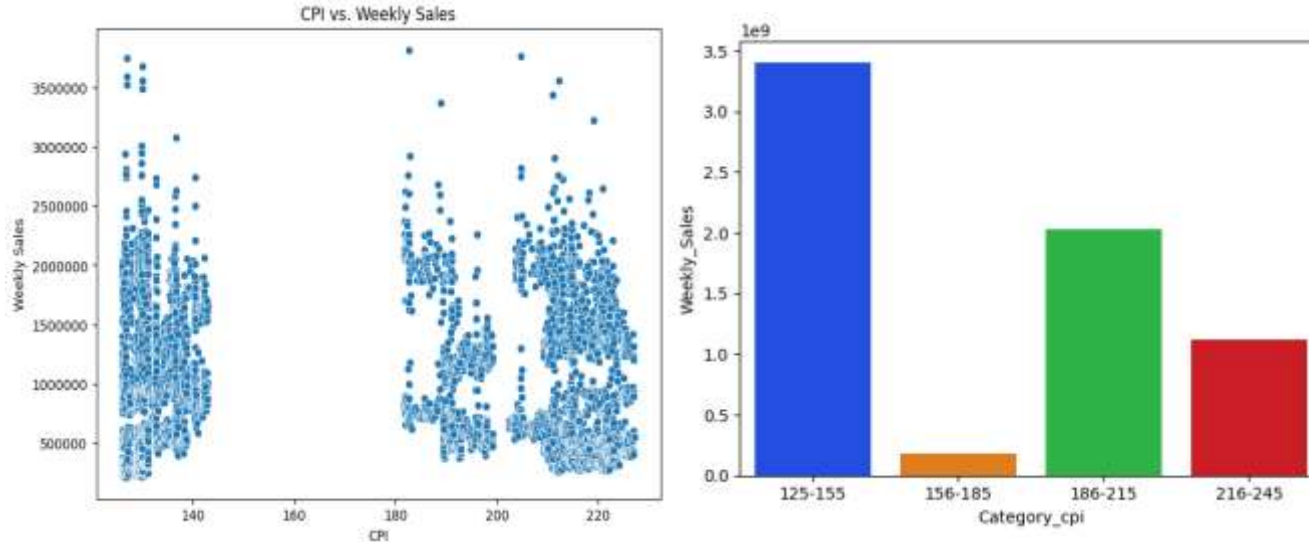
when inflation is lower the performance of stores are high

in initial time of inflation the sale were drop significantly

The scatter plot shows that there is positive relation between the CPI and weekly sales. This means that as the CPI increses,weekly sales also tend to increase.
There is lot of variation in the data,however.This means that there are some weeks where sales are high even wheb CPI is low, and vice versa.
this could be due to number of factor, such as seasonal fluctuation.

Data Pre-processing Steps and Inspiration

Store Performance Analysis: Identify the top-performing stores based on historical sales data and evaluate the factors contributing to their success.

Worst Performing Store: Determine the worst performing store and analyze the significance of the difference between the highest and lowest performing stores.
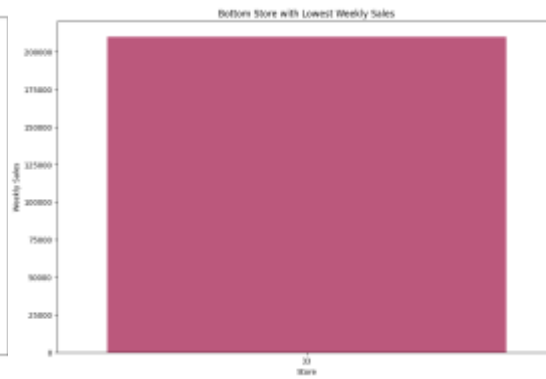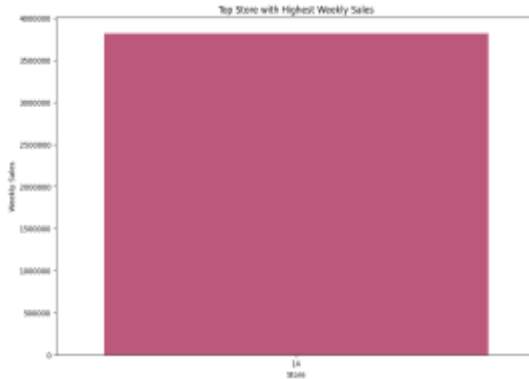

Mean weekly sales of the worst performing store (Store 33): 259861.692027972 Mean weekly sales of the best performing store (Store 14): 2020978.400979021

The mean difference between the best store and the worst store is: 1761116.708951049

Upon visualizing the bar plot, it is evident that Store number 14 exhibits the highest weekly sales among all stores. And 33 is the lowest

Data Pre-processing Steps and Inspiration

Correlation: statistical summary that measures the strength and direction of the relationship between variables
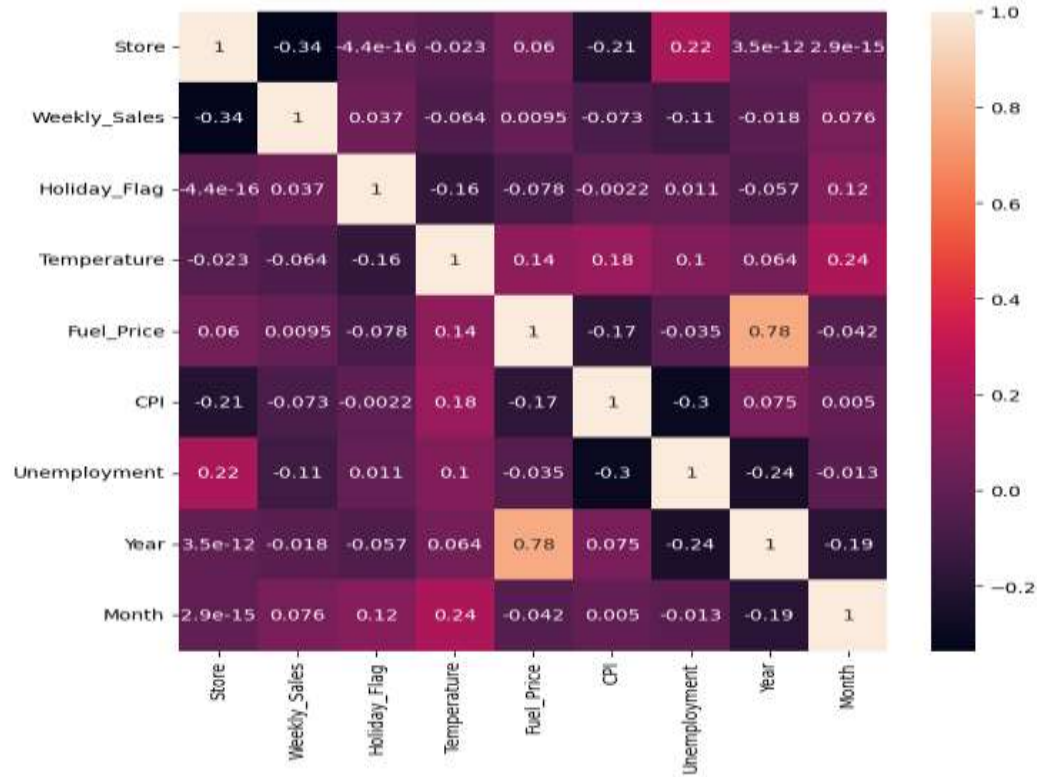
Choosing the Algorithm Reasons for Choosing the Algorithm

I have used ARIMA (Autoregressive Integrated Moving Average) model for my project can be justified by several factors.

1. ARIMA is tailored for time series forecasting tasks.

2. It captures both seasonality and underlying trends effectively.

3. Handles data stationarity through differencing.

4. Accommodates various types of time series data flexibly.

5. Parameters have clear interpretations for better understanding.

6. Demonstrates good performance in diverse applications.

7. Widely available in popular libraries for easy implementation.

Assumptions

1. **Stationarity**: Time series data is assumed to be stationary or can be made stationary through differencing.

2. **Linear Relationships**: Relationships between past and present observations are assumed to be linear.

3. **Homoscedasticity**: Constant variance of residuals across time series data.

4. **Independence**: Observations are assumed to be independent and identically distributed.

5. **No Multicollinearity**: Absence of multicollinearity among predictor variables.
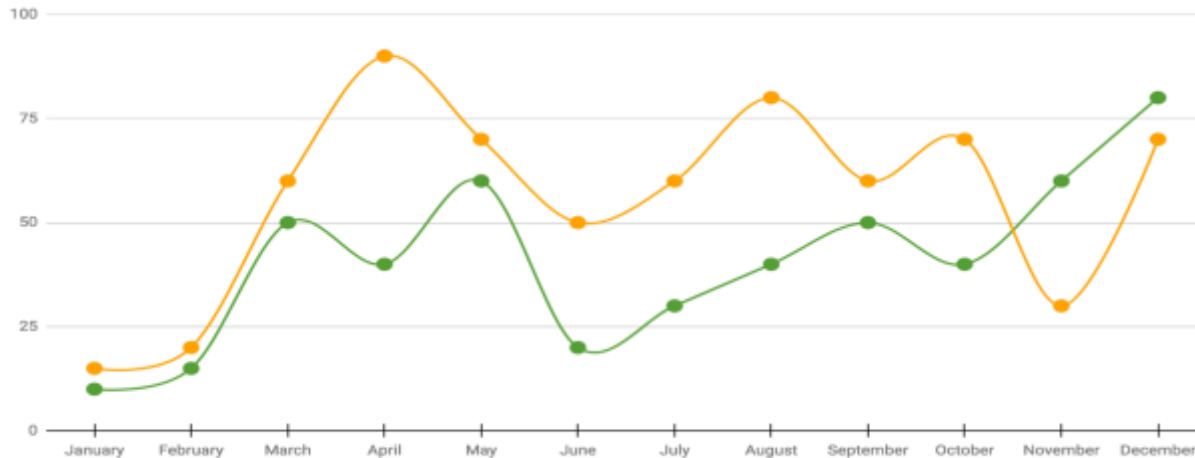
Model Evaluation and Techniques

1. perform ADF test for stationarity

2. Check stationarity

3. Stationaries the series (if necessary)

4. Plotting the components of the data

5. Creating the Rolling mean

6. Creating the Rolling standard deviation

7. Applying log transformation to the data

8. Checking the Data stationarity again after the log transform

9. Calculate autocorrelation and partial autocorrelation values

10. Building the ARIMA MODEL

11. Predicted actual and prediction store wise

ARIMA Model - Actual vs Predicted
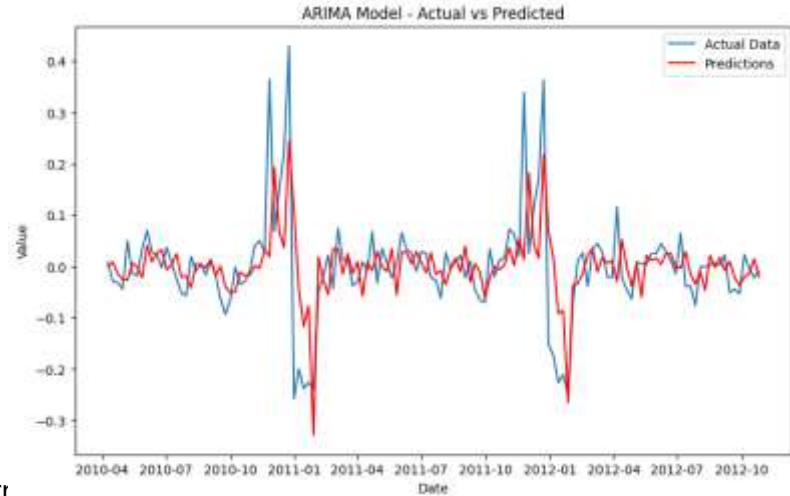
Future Possibilities

With the help of the Arima model and the techniques I have followed to predict the predicted sales for the model, we can easily forecast the sales for each store for the next 12 weeks.

Utilizing the ARIMA (Autoregressive Integrated Moving Average) model and associated techniques for sales forecasting presents several promising future possibilities

1. **Improved Inventory Management**: Accurate sales forecasts generated by the ARIMA model can aid in optimizing inventory levels, ensuring that each store maintains optimal stock levels to meet customer demand while minimizing overstocking and stock outs.
2. **Enhanced Resource Allocation**: With reliable sales predictions, retailers can allocate resources more efficiently, such as staffing levels, promotional activities, and marketing campaigns, to maximize sales potential and operational efficiency.
3. Strategic Planning: The ability to forecast sales for each store accurately enables retailers to make informed strategic decisions, such as expansion plans, store openings/closures, and product assortment adjustments, to align with anticipated demand trends.
4. **Seasonal and Trend Analysis**: ARIMA models can capture seasonal patterns and underlying trends in sales data, providing valuable insights into consumer behavior and preferences over time. This information can inform product development, marketing strategies, and pricing decisions.
5. **Integration with Advanced Technologies**: As technology evolves, ARIMA models can be integrated with advanced analytics techniques, such as machine learning and artificial intelligence, to enhance forecasting accuracy and adaptability to changing market conditions.

Conclusion

To sum up, this study has given important new information on what influences weekly sales in a chain of retail stores with several locations. Several important discoveries have been made possible by thorough data analysis and predictive modelling:

The investigation conducted indicated that external factors, including temperature, the unemployment rate, and the consumer price index (CPI), had a noteworthy impact on weekly sales. While favorable weather conditions might result in increased sales, stores with higher unemployment rates may see fewer sales.

Seasonal Trends: Based on the sales data, seasonal trends were found, with some seasons of the year typically exhibiting larger or lower sales volumes. Comprehending these seasonal fluctuations is essential for efficient inventory control and resource distribution.

Store Performance: By highlighting the best-performing stores using past sales data, the study made it possible to look more closely at the tactics and elements that made them successful. The fact that the worst-performing shop has been identified further emphasises how critical it is to close performance gaps and put focused improvement initiatives in place.

Predictive Sales Forecasting: The project effectively projected sales for each store over the course of the following 12 weeks by utilizing predictive modelling techniques. The retail chain will be able to better fulfil customer demand and streamline operations with the help of these projections, which offer practical insights for strategic planning, resource allocation, and inventory management.

Overall, this project demonstrates the power of data-driven decision-making in improving inventory management and operational efficiency in retail. By understanding the underlying factors driving sales performance and leveraging advanced analytics, the retail chain can make informed decisions to drive growth, enhance customer satisfaction, and stay competitive in the dynamic retail landscape.

References

Google

Stack Overflow

Statsmodel

I have used for dataset understanding

For small Coding doubt I have used Stack overflow

I have used for Statistical tools and Arima model