



Integrated Retail Analytics for Store Optimization

ABHISHEK YADAV

MASTER'S IN COMPUTER SCIENCE (AI)

Project objective

1. To utilize machine learning and data analysis techniques to optimize store performance, forecast demand, and enhance customer experience through segmentation and personalized marketing strategies.
2. To foster long-term customer loyalty and satisfaction by delivering personalized experiences that exceed expectations. By continuously refining our strategies based on customer feedback and market insights, we can position ourselves as a trusted partner in meeting the evolving needs of our customers while driving sustainable business success in an increasingly competitive retail landscape.



Project Components

1. About Dataset
2. Data Preprocessing and Feature Engineering
3. Exploratory Data Analysis
4. Anomaly Detection in Sales Data
5. Time-Based Anomaly Detection
6. Customer Segmentation Analysis & Segmentation Quality Evaluation
7. Personalization Strategies
8. Market Basket Analysis
9. Demand Forecasting
10. Impact of External Factors
11. Real-World Application and Strategy Formulation
12. Conclusion

About Dataset

There are 3 Tables –Features , Sales and Stores

Content

You are provided with historical sales data for 45 stores located in different regions - each store contains a number of departments. The company also runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks.

Features data-set.csv :

Contains additional data related to the store, department, and regional activity for the given dates.

- Store - the store number
- Date - the week
- Temperature - average temperature in the region
- Fuel_Price - cost of fuel in the region
- Markdown1-5 - anonymized data related to promotional markdowns. Markdown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA
- CPI - the consumer price index
- Unemployment - the unemployment rate
- IsHoliday - whether the week is a special holiday week

sales data-set.csv:

Anonymized information about the 45 stores, indicating the type and size of store

stores data-set.csv:

Historical sales data, which covers to 2010-02-05 to 2012-11-01. Within this tab you will find the following fields:

- Store - the store number
- Dept - the department number
- Date - the week
- Weekly_Sales - sales for the given department in the given store
- IsHoliday - whether the week is a special holiday week

Data Preprocessing and Feature Engineering

Changing datatype: Convert data type of dates column to datetime to extract days, months, weeks & year details.

Merging Datasets: Joined all the tables i.e. features, stores and sales table in a single table.

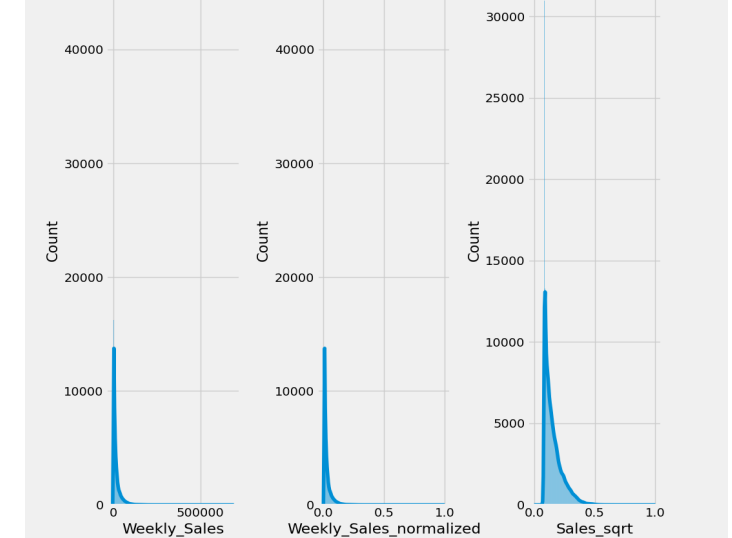
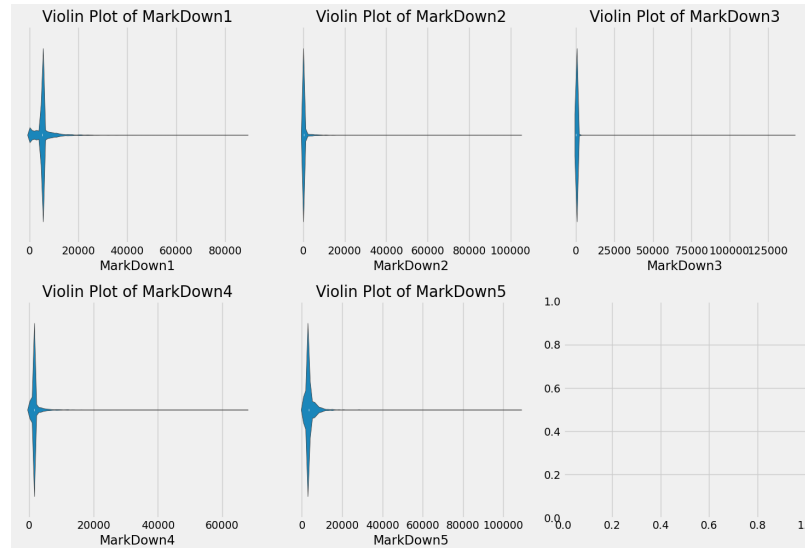
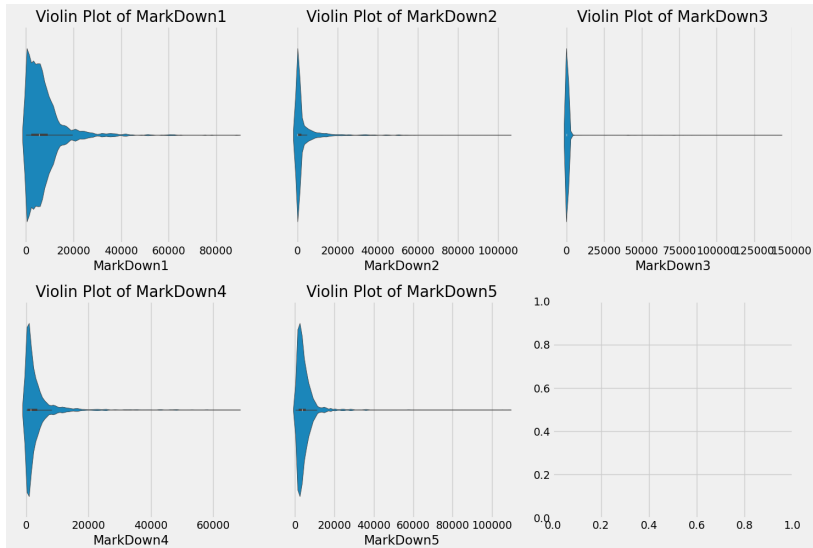
Handle Missing Values: Use techniques like median imputation due to anomalies to address missing values in your markdown columns .

This step ensures that you have a complete dataset for analysis.

Normalize Sales Data: Apply one of the normalization techniques discussed earlier to bring all sales data on a consistent scale. Applied Min-Max Scaling on the sales data.

Square root Transformation to handle skewness in numerical features.

	0
Store	0
Dept	0
Date	0
Weekly_Sales	0
IsHoliday	0
Type	0
Size	0
Temperature	0
Fuel_Price	0
MarkDown1	270889
MarkDown2	310322
MarkDown3	284479
MarkDown4	286603
MarkDown5	270138
CPI	0
Unemployment	0

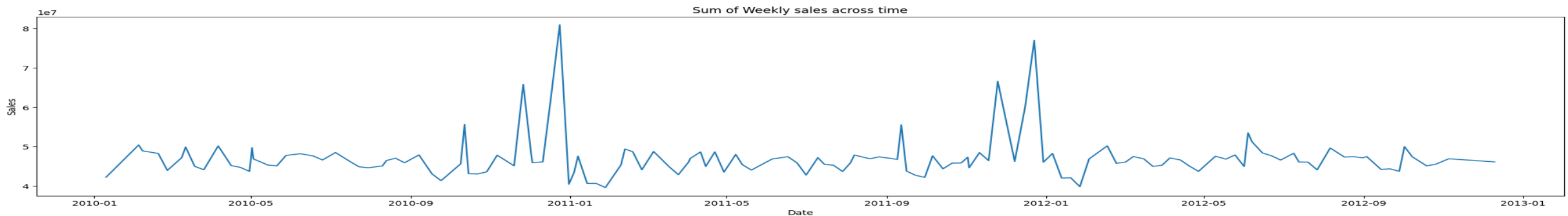
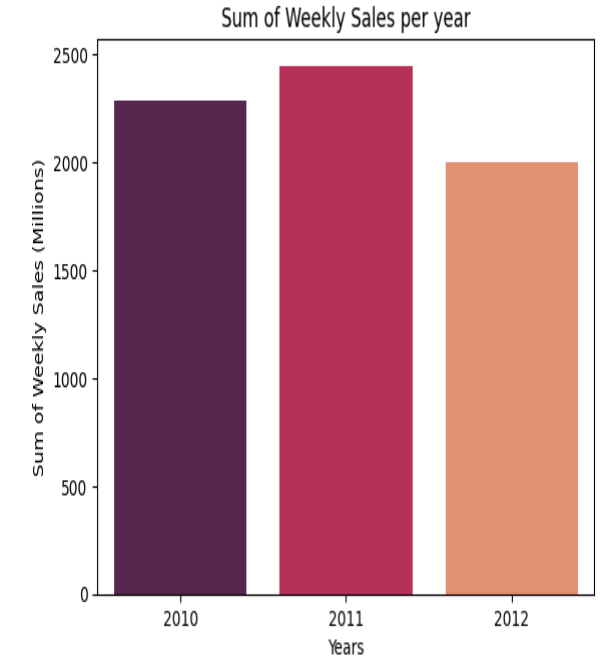
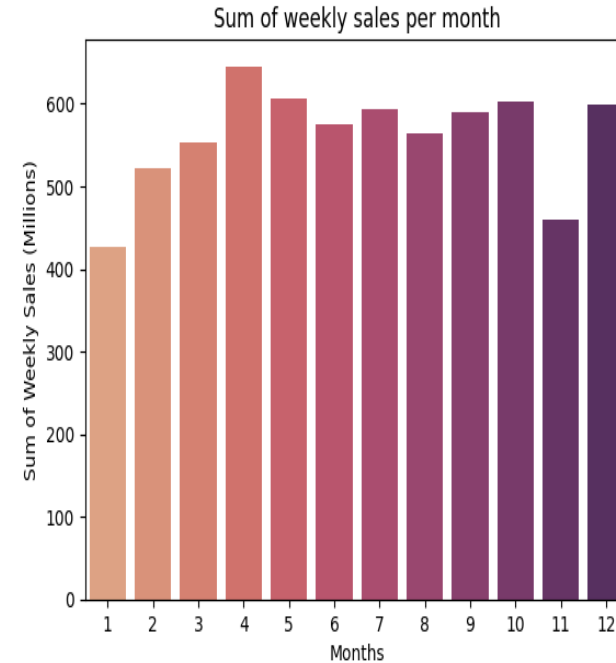


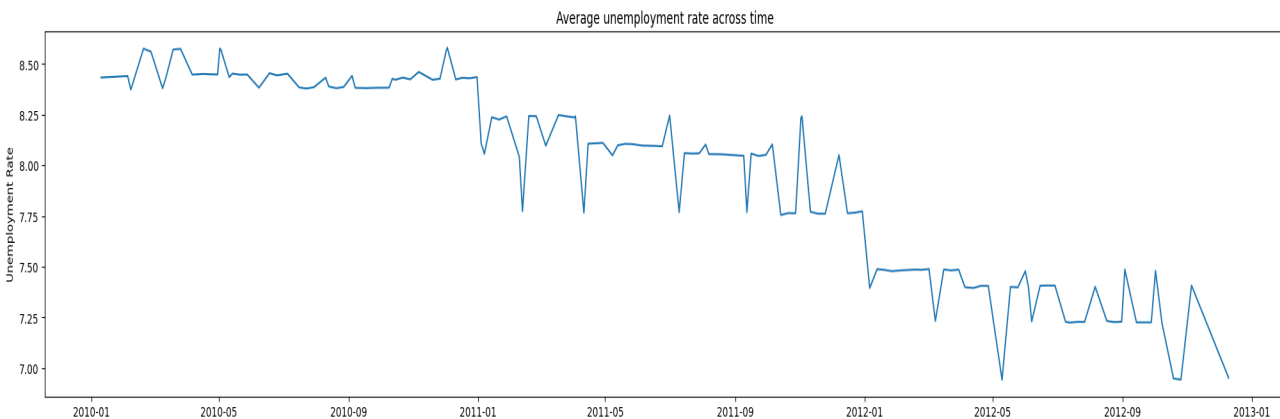
Exploratory Data Analysis

Weekly Sales Over Time

Findings:-

- Sales tend to rise significantly during special offers and holidays (e.g., Black Friday, Christmas, New Year), followed by a decline.
- Subsequently, sales regain momentum in certain months.
- November indicates reduced buying, possibly due to anticipation of upcoming sales.
- December 2013 data appears incomplete; despite the spike, overall sales did not reach substantial heights due to limited available data.

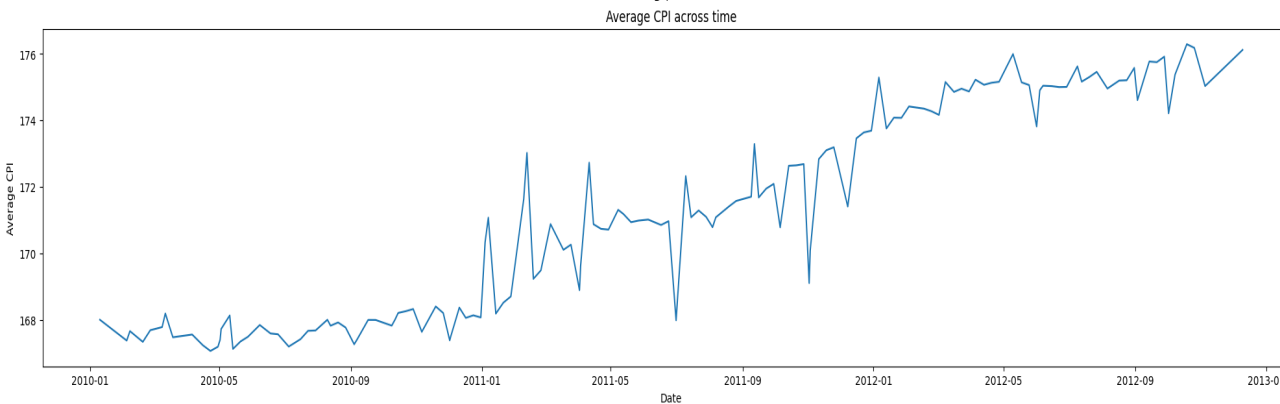




Trend in average unemployment rate across time

Findings:-

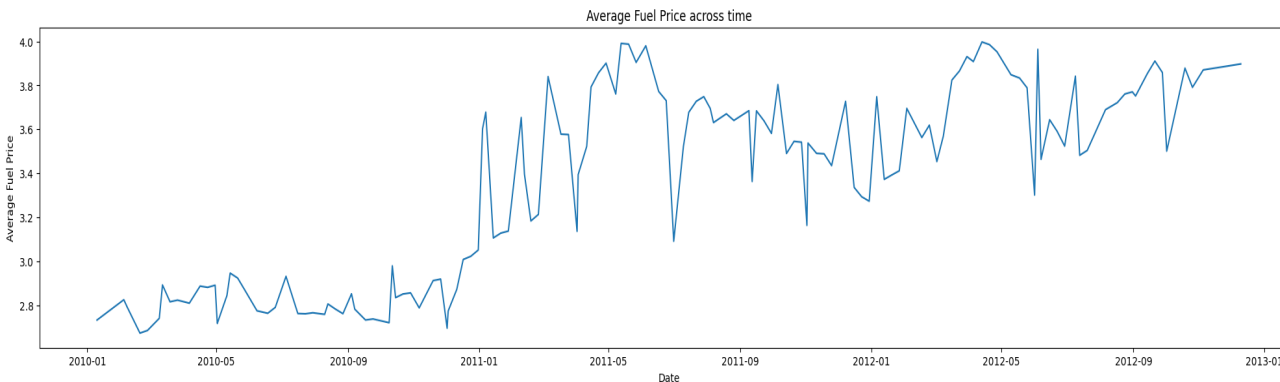
- Over the period spanning from January 2010 to December 2012, we have noted a gradual decrease in the unemployment rate.



Trend in average consumer price index rate across time

Findings:-

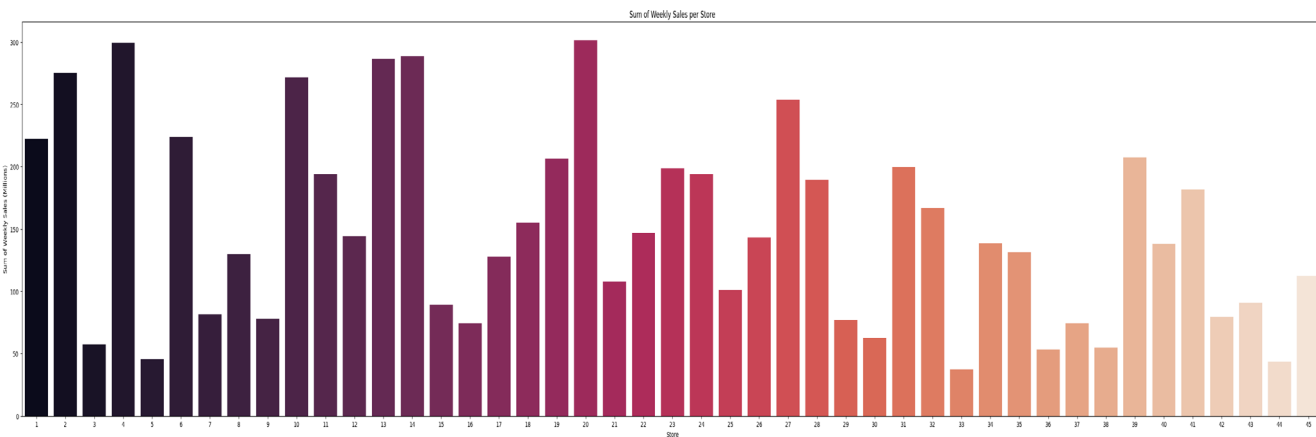
- Over the period spanning from January 2010 to December 2012, we have noted a gradual increase in the consumer price index due to inflation.



Trend in average fuel price rate across time

Findings:-

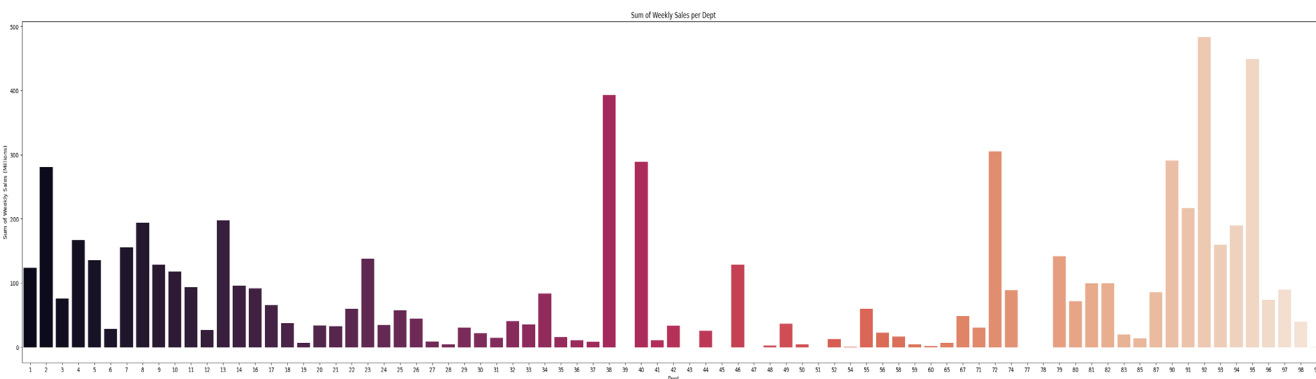
- From January 2010 to December 2010, we observed a relatively stable fuel price rate with no significant changes. However, starting from January 2011 through December 2012, there was a gradual increase in fuel prices attributed to inflationary factors.



Which stores are the highest earners

Findings:-

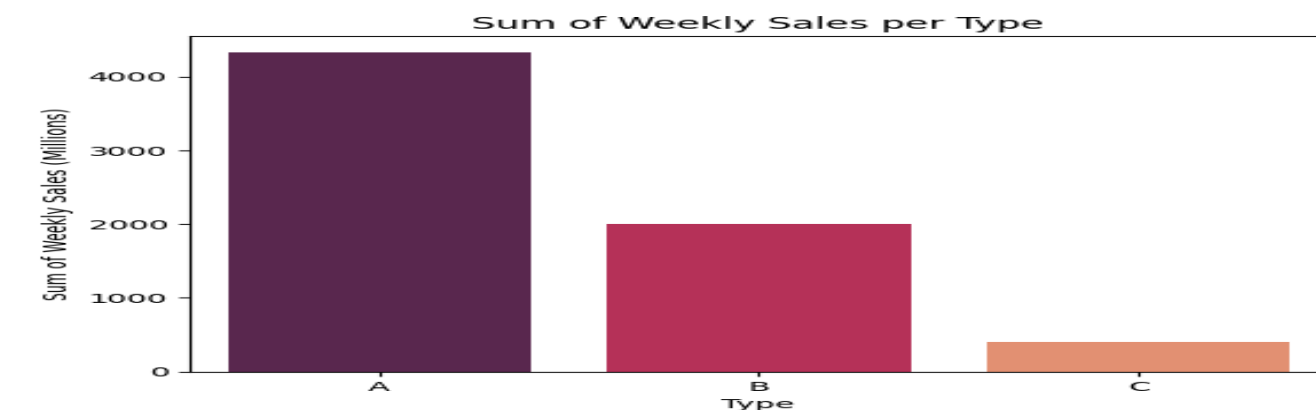
Store number 20 and 4 were identified as the top earners based on their weekly sales, while store number 44 and 5 emerged as the least profitable.



Which departments are the highest earners

Findings:-

Certain departments (1-15, 38, 40, 72, and 90-95) demonstrate notably higher weekly sales, with an average increase of 25%.

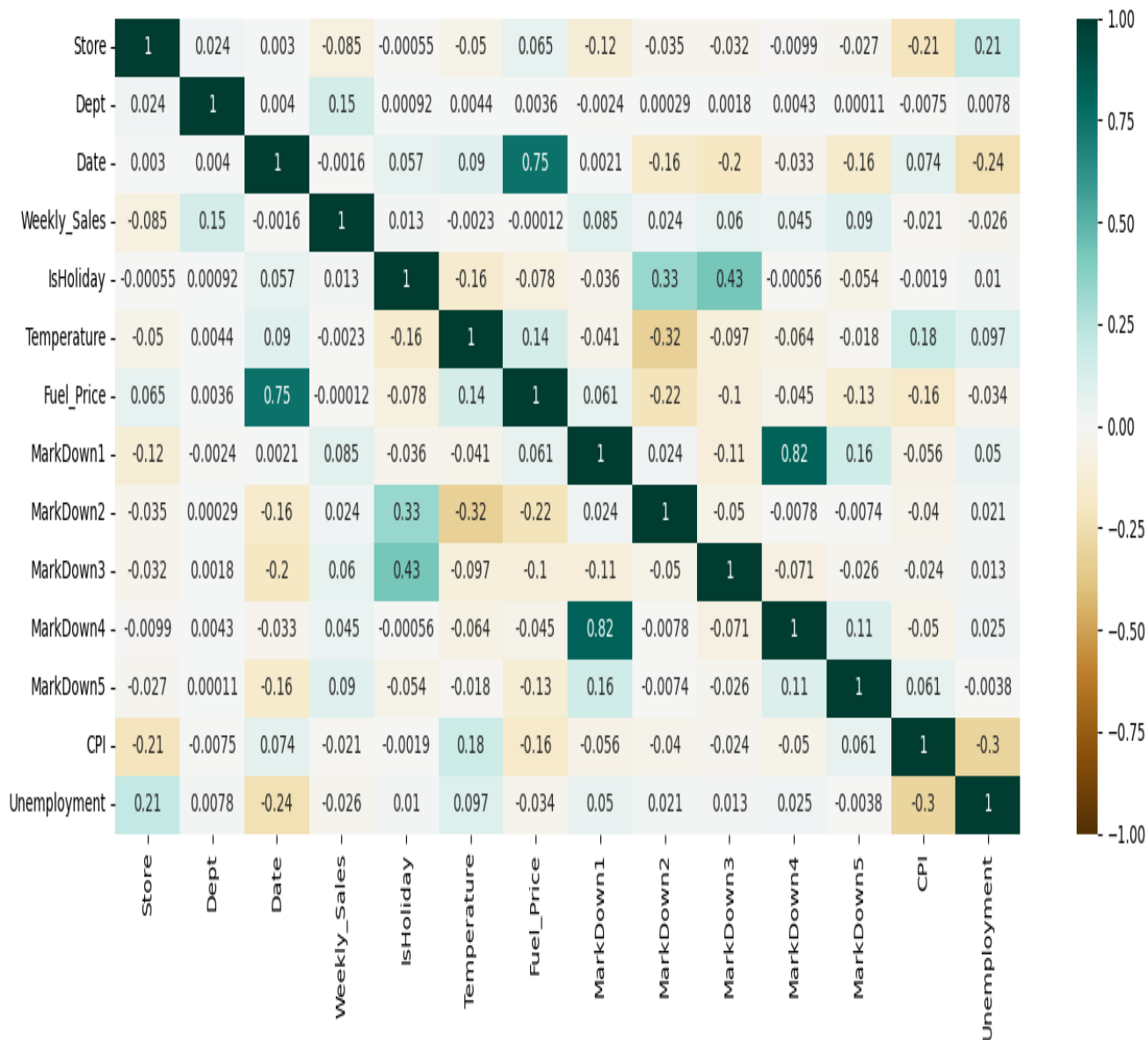


Which type of store is the highest earner

Findings:-

Store type A, succeeded by store type B, were identified as the store types with the highest earnings, while store type C exhibited the lowest weekly sales.

Correlation heatmap

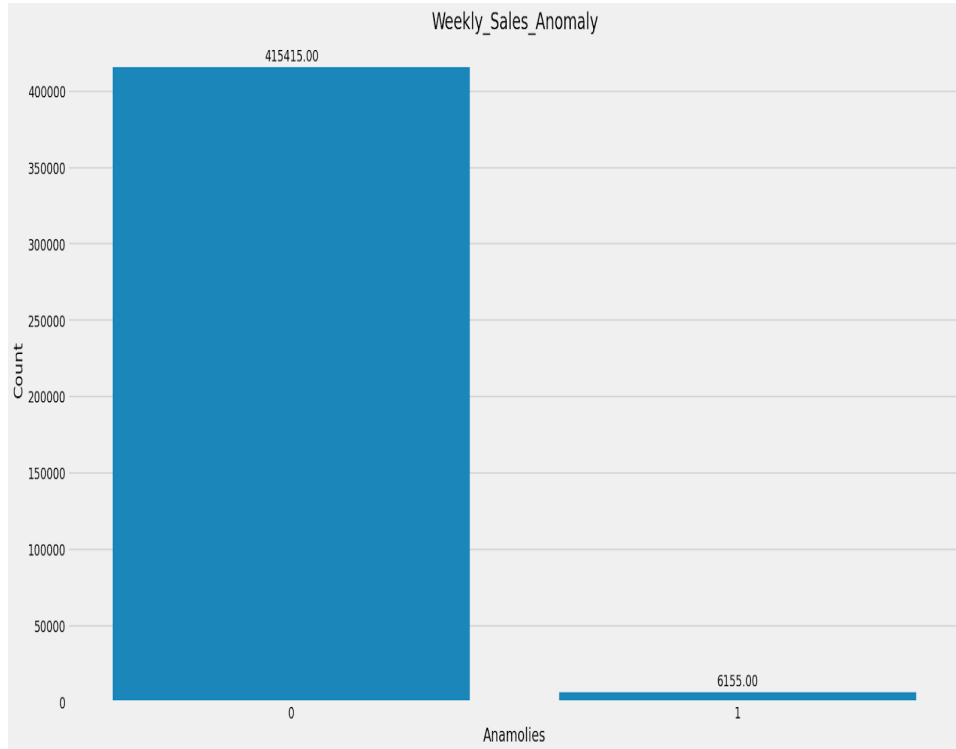


Correlation of factors and weekly sales

Findings:-

- Weekly sales do not exhibit a strong correlation with other parameters.
- Consumer Price Index (CPI) and Unemployment display a negative correlation.
- Unemployment and Fuel Price are also negatively correlated.

Anomaly Detection in Sales Data



Handled Anomalous Data Points Using Interquartile Range (IQR) Technique

Objective-

- Identify unusual sales patterns across stores and departments.
- Investigate potential causes (e.g., holidays, markdowns, economic indicators).
- Implement anomaly handling strategies to clean the data for further analysis.

What is Anomaly

- Anomalies are deviations from expected behavior or patterns in data, and their presence can indicate issues or security threats.

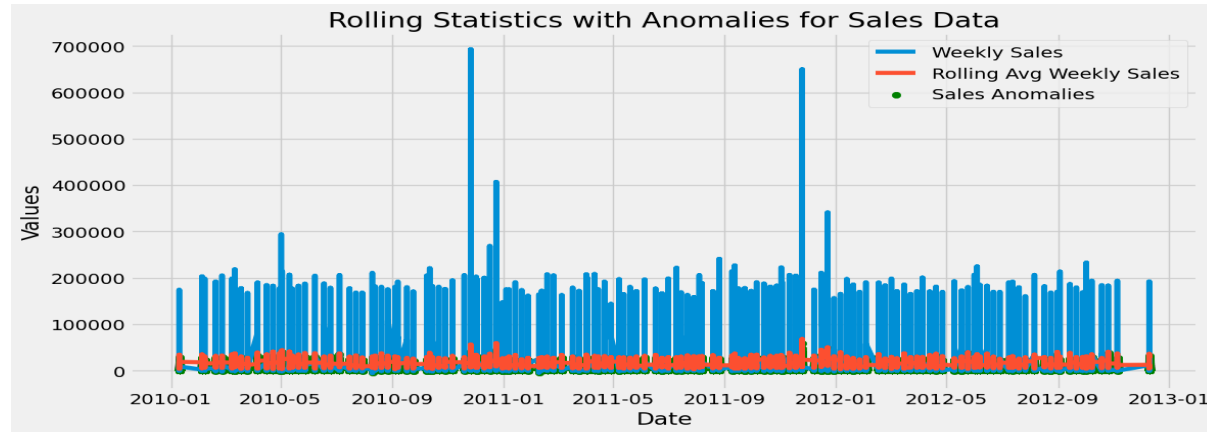
Method used for Detecting Anomalies in Retail Data

- **Descriptive statistics** are among the foundational methods for anomaly detection in retail. Techniques like mean, median, and standard deviation can be used to calculate summary statistics, which help identify point anomalies.

Mitigation Strategies

- Data Cleaning Data Normalization Data Imputation

Time-Based Anomaly Detection

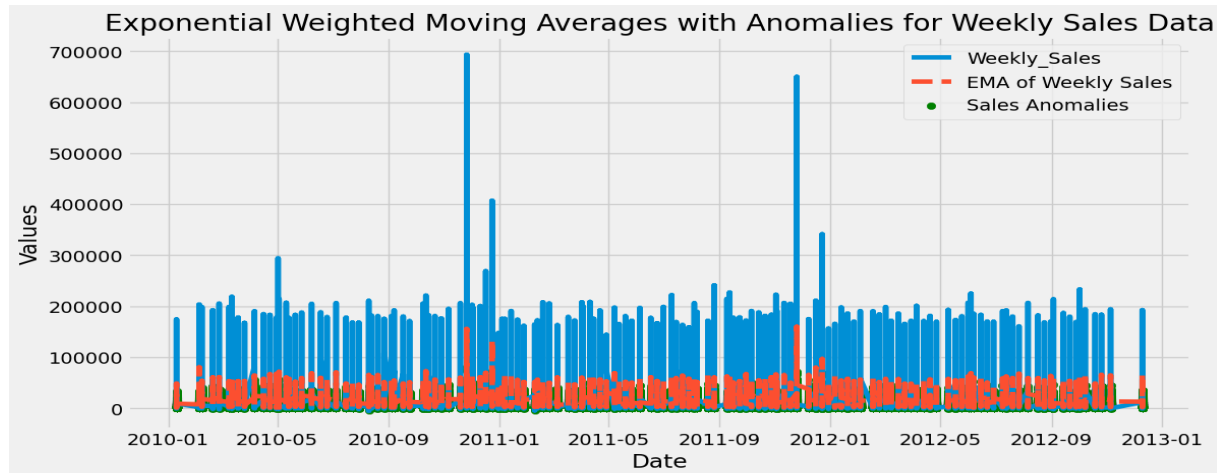


Objective-

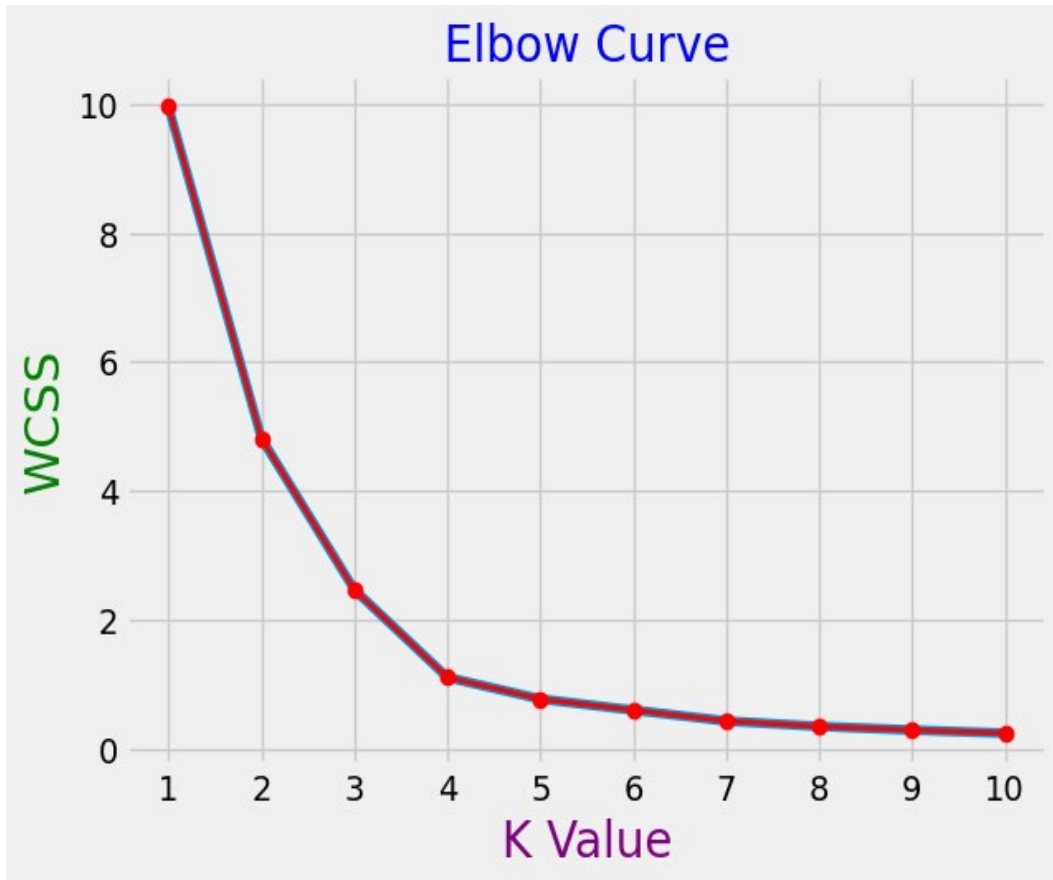
- Analyze sales trends over time.
- Detect seasonal variations and holiday effects on sales.
- Use time-series analysis for understanding store and department performance over time.
- Detect anomalies in time series Sales data using specialized methods.

Approach:

- **Create Rolling Statistics:** Calculate rolling averages, moving sums, and standard deviations to track data variations over time.
- **Apply Exponential Smoothing:** Use exponential moving averages (EMA) and anomaly detection techniques to highlight unusual patterns in time series data.
- **Highlight Anomalies:** Set thresholds and visual cues to identify significant deviations from smoothed values, aiding in anomaly detection and trend analysis.



Customer Segmentation Analysis



To identify the best number of clusters, I utilized the Elbow method.

Objective-

- Segment stores or departments based on sales patterns, markdowns, and regional features.
- Analyze segment-specific trends and characteristics

What is Segmentation

- Customer segmentation is a critical marketing strategy that involves dividing a customer base into distinct groups based on specific criteria or characteristics.

Algorithm used for Segmentation in Retail Data

Applied the K-Means Clustering algorithm to segment stores and departments. K-Means Clustering is an unsupervised machine learning technique widely used for grouping data into clusters based on similarity.

Segmentation Quality Evaluation

Objective-

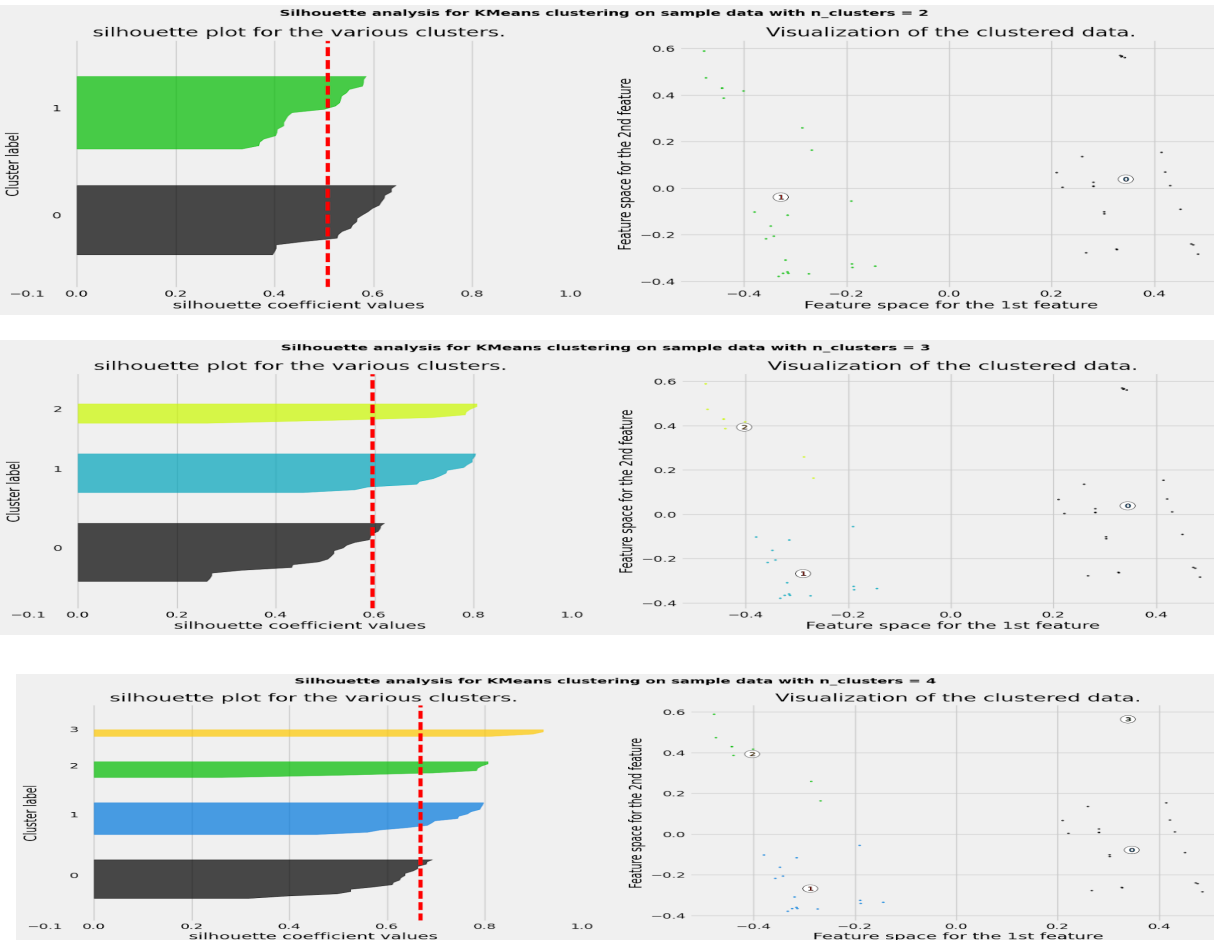
- Evaluate the effectiveness of the customer segmentation.
- Use metrics to assess the quality of segments in terms of homogeneity and separation.

Significance of Segmentation Quality Evaluation

- The significance of segment quality evaluation lies in its ability to assess the effectiveness of clustering techniques in partitioning data into meaningful groups. It allows us to measure how well the clusters represent the underlying structure of the data and how distinct they are from each other.
- This information is crucial for making informed decisions and drawing actionable insights from the clustered data, such as personalized marketing strategies

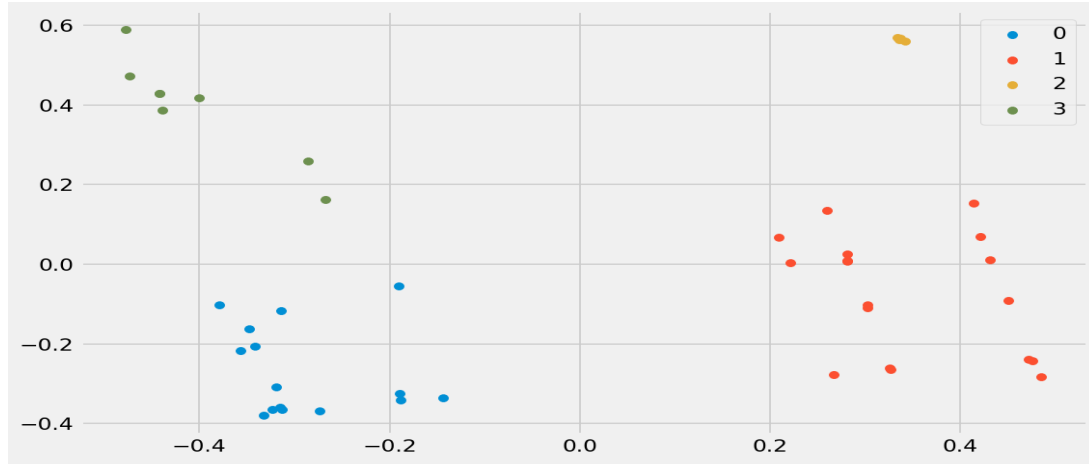
Metric used for Segmentation Quality Evaluation in Retail Data

- Utilized the silhouette score metric to evaluate the quality of segments and select the most suitable cluster configuration.
- A silhouette score closer to 1 indicates better segmentation quality, and after analysis, we determined the optimal number of clusters to be 4 for both store and department segments.

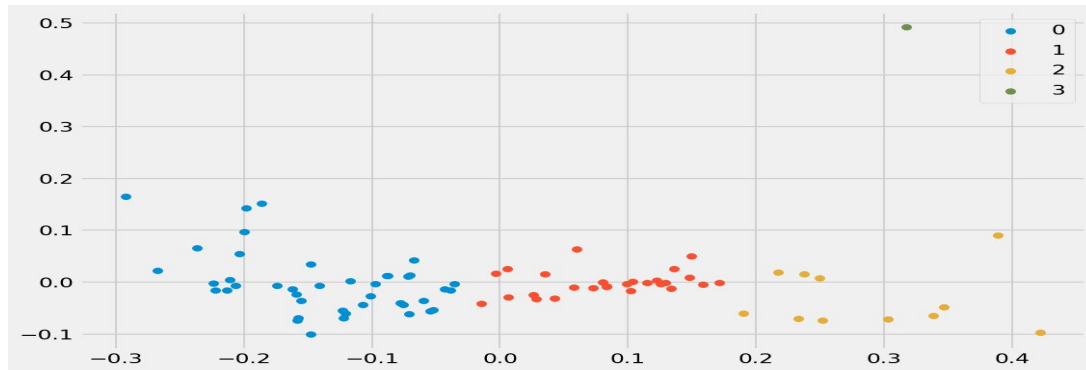


The optimal number of cluster is 4 according to silhouette score

Personalization Strategies



Store Segments Plot



Department Segments Plot

Objective-

- Develop personalized marketing strategies based on the store and department segments.
- Propose inventory management strategies tailored to store and department needs.

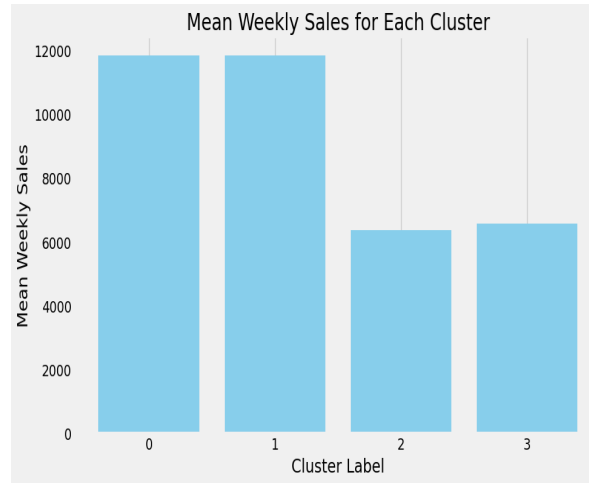
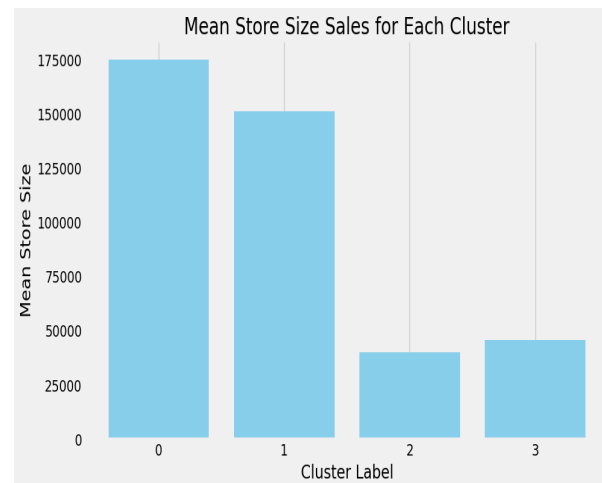
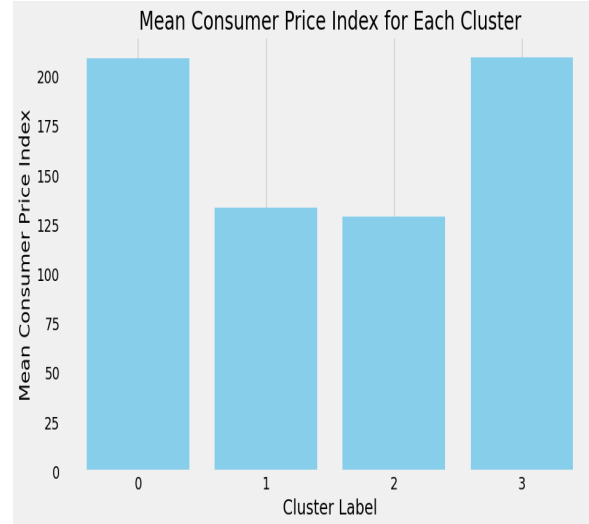
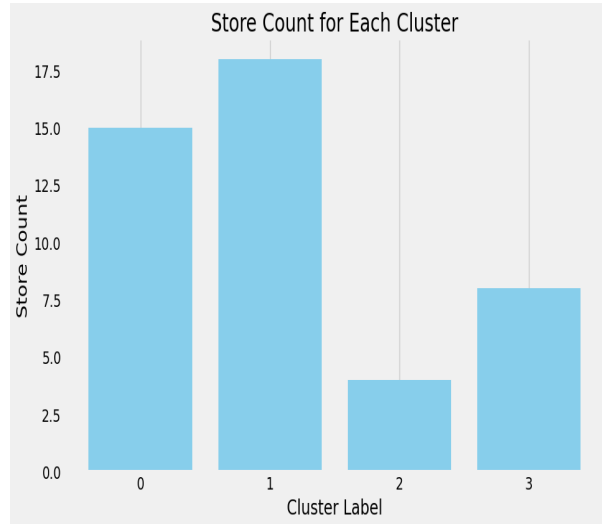
Significance of Personalization Strategies

- The significance of personalization strategies lies in their ability to tailor products, services, and marketing efforts to meet the specific needs and preferences of individual customers. By personalizing interactions and experiences, businesses can enhance customer satisfaction, loyalty, and engagement.

Significance of Inventory management Strategies

- Inventory management strategies are significant for several reasons:
- Cost Optimization: Effective inventory management helps minimize storage costs by preventing overstocking or stockouts, thus optimizing inventory levels and reducing holding costs.
- Improved Cash Flow: By maintaining optimal inventory levels, businesses can free up capital that would otherwise be tied up in excess inventory, leading to improved cash flow and liquidity.

Personalization Strategies for Store Segments



Personalization Strategies for Store Segments

Cluster 0: Premium Space Retailers & Sizeable Luxury Stores

- Marketing campaigns should highlight the spaciousness and luxury ambiance of the stores.
- Emphasize premium product lines and exclusive offerings to attract affluent customers.

Cluster 1: Value-Oriented Stores

- Focus marketing efforts on value-conscious consumers by highlighting competitive pricing and promotions.
- Leverage social media platforms and targeted digital advertising to reach budget-conscious shoppers.

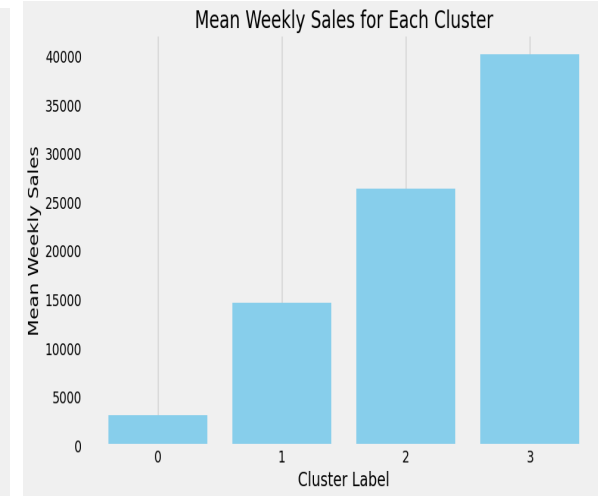
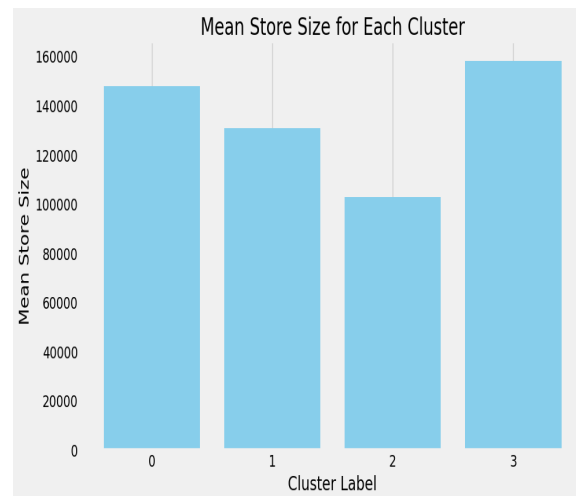
Cluster 2: Budget-Friendly Stores

- Promote budget-friendly product lines and emphasize affordability in marketing messaging.
- Utilize cost-effective marketing channels such as email campaigns and local advertising.

Cluster 3: Compact Elegance Stores

- Position the stores as boutique destinations offering curated selections of premium products.
- Highlight the elegance and sophistication of the store ambiance in marketing materials.

Personalization Strategies for Department Segments



Personalization Strategies for Department Segments

Cluster 0: Sizeable Luxury Departments

- Highlight the spaciousness and luxury ambiance of the departments in marketing materials.
- Promote premium product lines and exclusive offerings to attract affluent customers.

Cluster 1: Premium Selection Departments

- Emphasize the quality and exclusivity of products in marketing campaigns to appeal to discerning customers.
- Utilize digital marketing channels and social media platforms to reach a wider audience of luxury shoppers.

Cluster 2: Elite Departments

- Position the departments as boutique destinations offering curated selections of premium products.
- Showcase the convenience and efficiency of compact physical spaces in marketing materials.

Cluster 3: Grand Outlets Department

- Emphasize the variety and breadth of products available in the departments to appeal to a wider customer base.
- Utilize value-driven marketing messages to highlight affordability and attract price-conscious shoppers.

Market Basket Analysis

	Antecedent	Consequent	Support	Confidence	Lift
0	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...	30	0.012898	1.000000	9.508576
1	[1, 2, 4, 6, 8, 10, 11, 13, 19, 20, 24, 25, 27...	37	0.012627	0.979021	77.531469
2	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...	35	0.012357	0.958042	9.109615
3	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...	36	0.012357	0.958042	9.109615
4	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14...	29	0.011365	0.881119	8.378186

The output of the analysis consists of association rules that reveal relationships between departments. Each rule specifies a set of items (antecedent) that are associated with another set of items (consequent). For example, a rule might indicate that when department 29 is present, items in stores 1, 2, 3, ..., 45 tend to be purchased together.

Objective-

- In market basket analysis we have tried to find out the association within different departments with the different stores in the given sales data.
- Develop cross-selling strategies based on these inferences.

Significance of Market Basket Analysis

- Market basket analysis is a strategic data mining technique used by retailers to enhance sales by gaining a deeper understanding of customer purchasing patterns. This method involves examining substantial datasets, such as historical purchase records, to unveil inherent product groupings and identify items that customers tend to buy together.
- By recognizing these patterns of co-occurrence, retailers can make informed decisions to optimize inventory management, devise effective marketing strategies, employ cross-selling tactics.

Algorithms Used in Market Basket Analysis in Retail Data

- The Apriori Algorithm widely uses and is well-known for Association Rule mining, making it a popular choice in market basket analysis.

Market Basket Analysis

Implementation:

- The Apriori algorithm is applied to the sales data with specific parameters such as minimum support, confidence, lift, and length. This algorithm efficiently discovers frequent itemset and generates association rules based on these itemset.

Association Results:

- The association results are printed out, showing each association rule along with its support, confidence, and lift. For example, a rule might indicate that when department 29 is present, there is a high confidence (e.g., 95%) that stores 1, 2, 3, ..., 45 will also be present in the same transaction, with a lift value indicating the strength of the association.

Overall, our approach encompasses standard practices in market basket analysis and leverages the Apriori algorithm to uncover meaningful associations within the provided sales data, which can lead to actionable insights for optimizing product placement, cross-selling strategies, and marketing efforts across different departments and stores.

Market Basket Analysis

Develop cross-selling strategies based on these inferences

Optimize Store Layout:

- Utilize the associations between departments and stores to optimize the layout of each store. Place related departments in close proximity to each other within the store to encourage cross-store purchases. For example, if departments 29 and 30 are frequently purchased together across various stores, consider placing them adjacent to each other to facilitate cross-department sales.

Create Bundled Offers:

- Identify departments that exhibit strong associations across multiple stores, such as departments 35 and 36. Create bundled offers or package deals that include products from both departments to incentivize customers to make cross-department purchases.

Targeted Promotions:

- Develop targeted promotions and marketing campaigns based on the associations between departments and stores. For instance, if department 37 frequently co-occurs with departments 1, 2, 4, and 40 across different stores, create promotions that span these departments to capitalize on customer purchase patterns.

Dynamic Pricing Strategies:

- Adjust pricing strategies based on the associations between departments and stores. Offer discounts or special pricing for products that are frequently purchased together across different stores to stimulate sales and increase customer satisfaction.

Demand Forecasting

Objective-

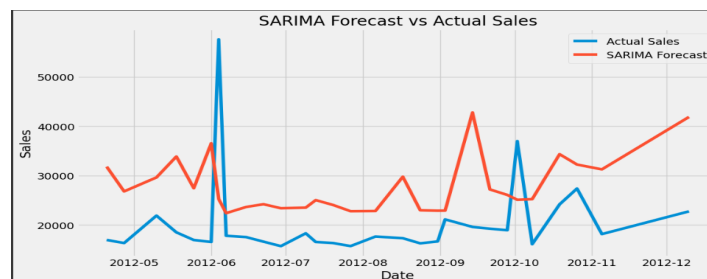
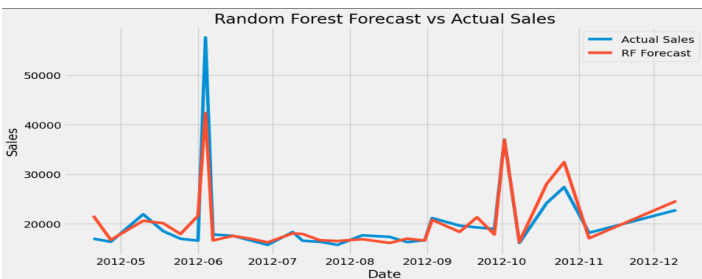
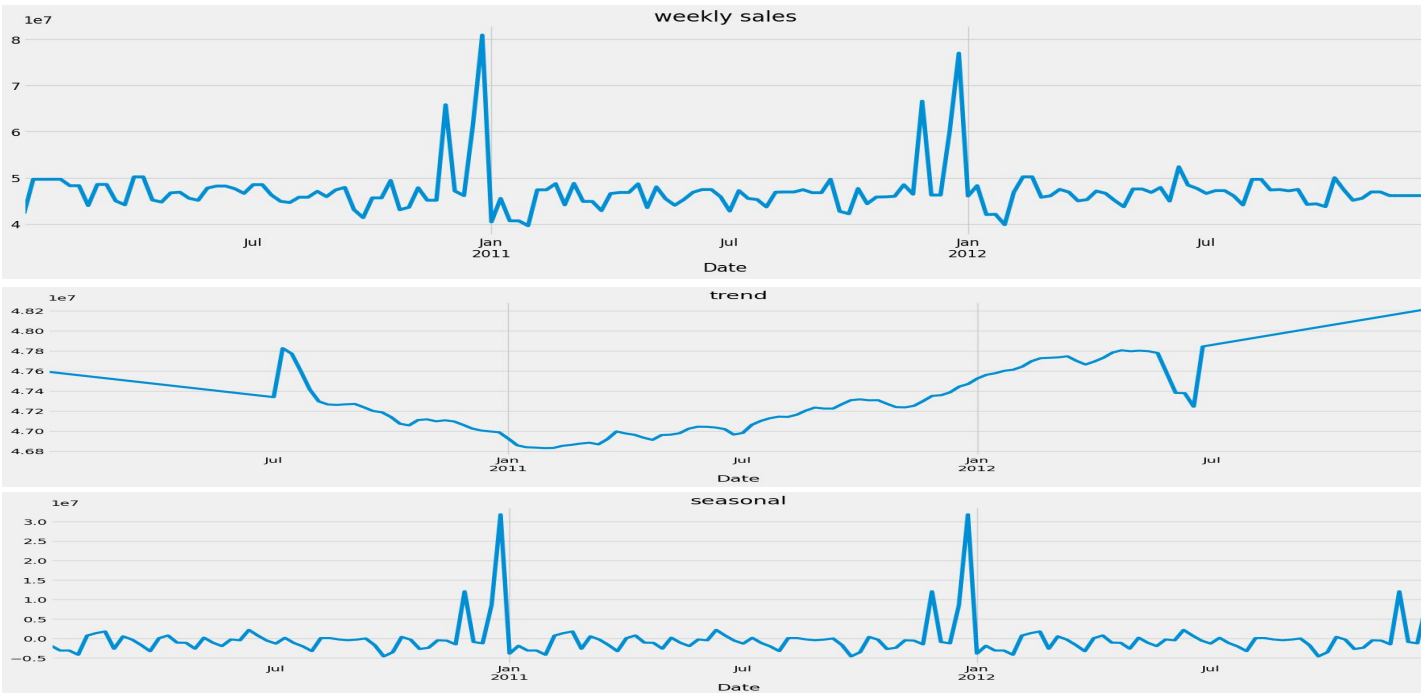
- Build models to forecast weekly sales for each store and department.
- Incorporate factors like CPI, unemployment rate, fuel prices, and store/dept attributes.
- Explore short-term and long-term forecasting models.

Significance of Demand Forecasting

- Demand forecasting is a process used by businesses to predict future customer demand for their products or services.
- It involves analyzing historical sales data, market trends, and other relevant factors to estimate the quantity of goods or services that customers are likely to purchase over a specific period of time.

Models Used for Demand Forecasting in Retail Data

- **Short-term forecasting model(SARIMA)**- Predicting future events or values over a relatively brief period, often within weeks or months, for immediate operational decisions.
- **Long-term forecasting model(Random Forest)**- Predicting future trends and values over an extended period, typically spanning months to years, for strategic planning and decision-making.



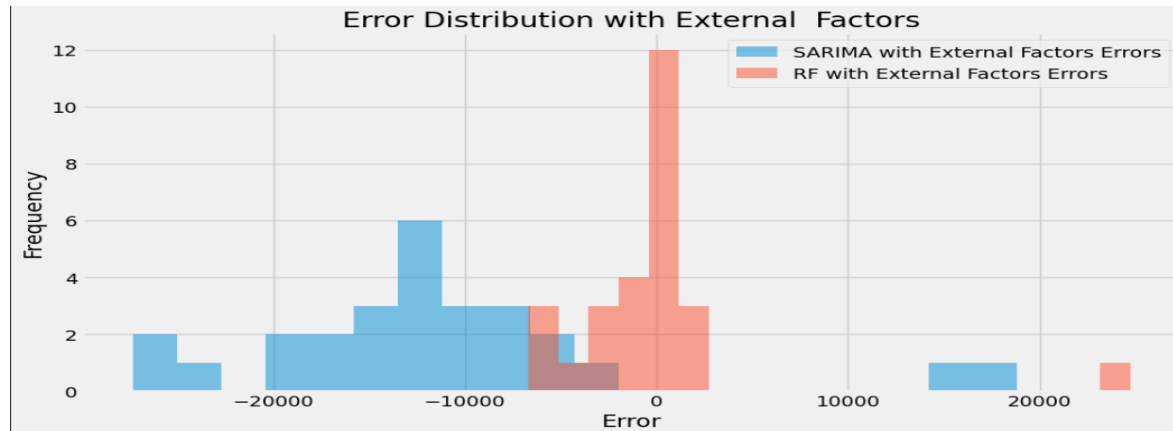
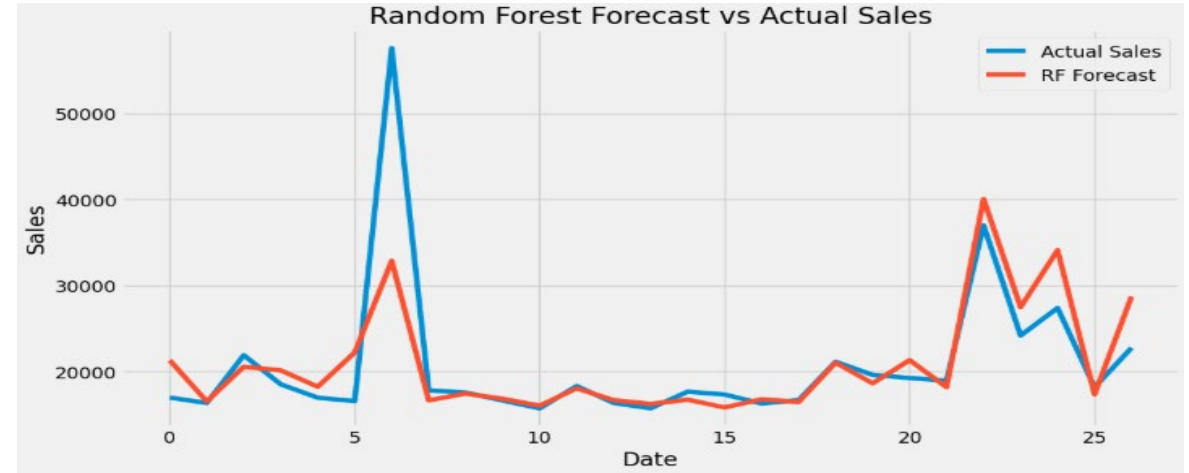
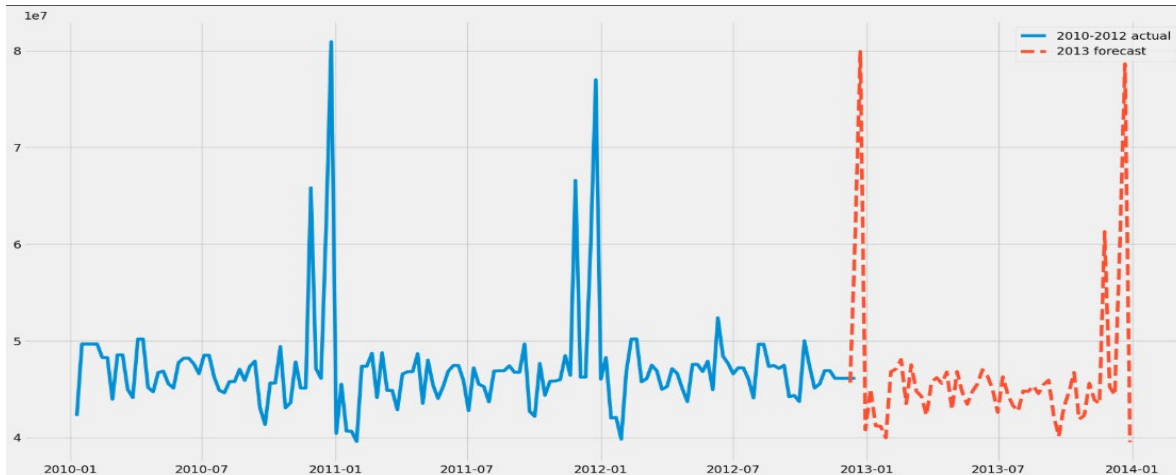
```
print("RF MAE:", rf_mae)
print("RF RMSE:", rf_rmse)

RF MAE: 1848.5777172413796
RF RMSE: 3442.0996185843583
```

```
print("SARIMA MAE:", sarima_mae)
print("SARIMA RMSE:", sarima_rmse)

SARIMA MAE: 10497.416422910876
SARIMA RMSE: 12275.935543289825
```

Demand Forecasting



The error distribution plot demonstrates that the distribution in long-term forecasting random forest model adheres to a Gaussian curve.

Forecasting for 2013

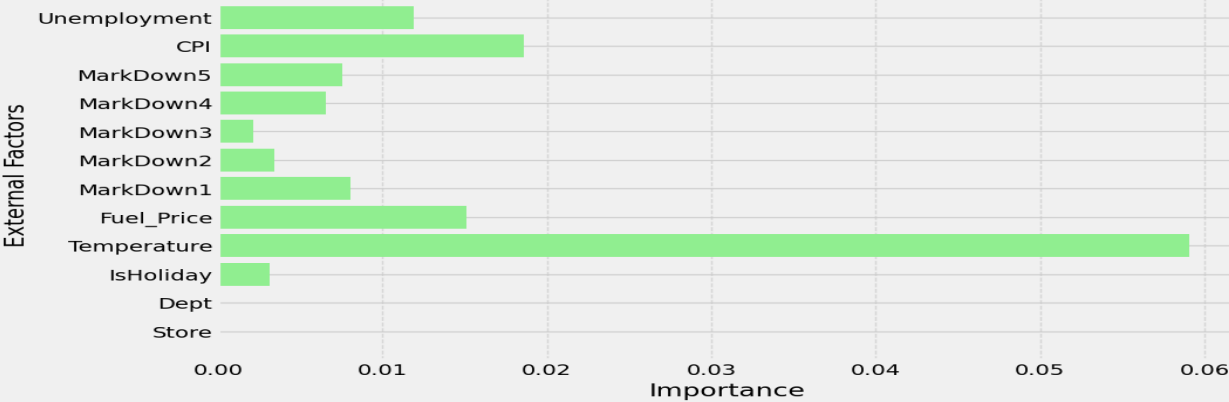
Forecasting using the Holt-Winters Model:

To perform forecasting using the Holt-Winters Model, also known as Triple Exponential Smoothing.

- We can leverage the Holt-Winters Model to perform forecasting and predict future weekly sales for each store and department, taking into account the level, trend, and seasonality components of the data.

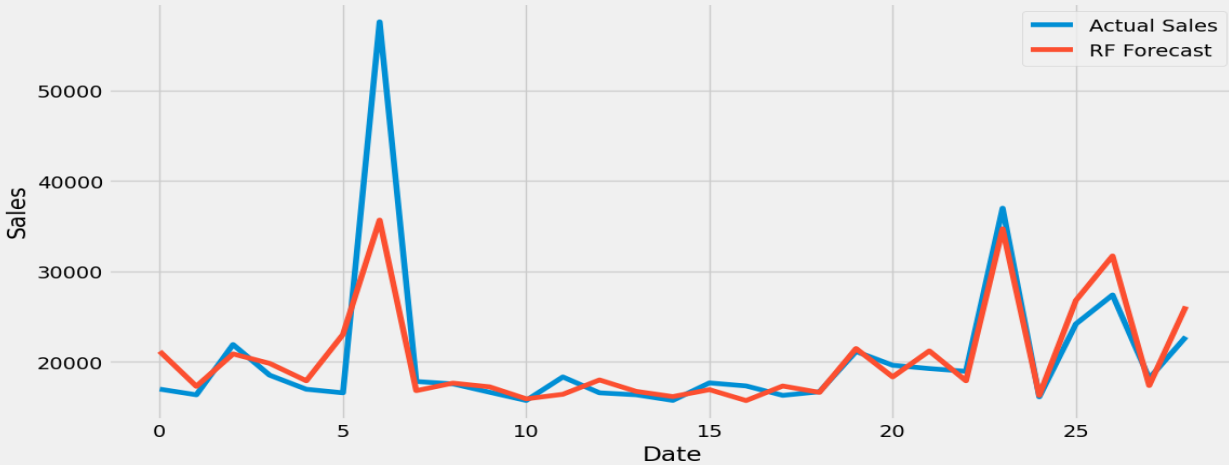
Impact of External Factors

Impact of External Factors on Demand



As it can be observed, the top 3 external factors which impact the demand forecasting in long term forecasting were temperature, consumer price index and fuel price.

Random Forest Forecast vs Actual Sales



Objective-

- Examine how external factors (economic indicators, regional climate) influence sales.
- Incorporate these insights into the demand forecasting models.

Impact of external factors in Demand Forecasting

Temperature

- It plays a significant role in shaping consumer demand, especially for seasonal products and services.
- During colder months, there is typically an increased demand for products such as winter clothing, heating appliances, and comfort foods. Conversely, warmer temperatures can lead to higher demand for items like outdoor recreational equipment, summer clothing, and cold beverages.

Consumer Price Index (CPI) and Fuel Price Index

- The CPI is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.
- Changes in the CPI reflect inflationary or deflationary pressures in the economy, which directly impact consumers' purchasing power.
- Fluctuations in fuel prices directly affect operational costs for businesses involved in transportation and logistics, leading to changes in product pricing and availability.

Real-World Application and Strategy Formulation

Real World Application and Strategy Formulation in Inventory Management

Seasonal Forecasting:

- Utilize temperature data to forecast seasonal demand fluctuations. Allocate inventory space and resources accordingly to meet the anticipated demand for seasonal products.

Dynamic Pricing:

- Incorporate CPI data to adjust pricing strategies in response to changes in consumer purchasing power. Offer promotions or discounts during periods of low inflation to stimulate sales.

Supply Chain Optimization:

- Monitor fuel price fluctuations to anticipate changes in transportation costs. Optimize supply chain routes and distribution networks to minimize the impact of fuel price changes on logistics expenses.

Real-World Application and Strategy Formulation

Real World Application and Strategy Formulation in Marketing Strategies

Targeted Campaigns:

- Tailor marketing campaigns based on regional climate conditions to promote relevant products. For example, launch winter clothing promotions in colder regions and summer gear promotions in warmer areas.

Price Sensitivity Analysis:

- Analyze consumer response to price changes using CPI data. Adjust pricing strategies to maximize revenue while remaining competitive in the market.

Promotional Timing:

- Coordinate marketing efforts with fluctuations in fuel prices. Launch promotions or special offers during periods of low fuel prices to offset potential decreases in consumer discretionary spending.

Real-World Application and Strategy Formulation

Real World Application and Strategy Formulation in Store Optimization

Product Placement:

- Arrange store layouts to reflect seasonal demand patterns identified through temperature data. Highlight relevant products prominently to attract customer attention.

Inventory Turnover:

- Use CPI data to optimize inventory levels and minimize carrying costs. Adjust stock levels based on changes in consumer spending patterns to ensure optimal turnover rates.

Customer Experience:

- Leverage insights from fuel price data to offer convenience-oriented services, such as online ordering with flexible delivery options, to mitigate the impact of increased transportation costs on consumers.

Real-World Application and Strategy Formulation

Challenges in implementing these strategies may include

Data Integration:

- Ensuring seamless integration of diverse data sources (temperature, CPI, fuel prices) into existing systems for accurate forecasting and decision-making. Resource Allocation: Allocating resources effectively to implement dynamic pricing strategies and targeted marketing campaigns based on real-time data analysis.

Operational Flexibility:

- Adapting supply chain and inventory management processes to respond quickly to changes in external factors, such as sudden fluctuations in fuel prices or unexpected shifts in consumer behavior.

Competitive Pressures:

- Navigating competitive pressures and market dynamics while adjusting pricing strategies to maintain profitability and market share. Addressing these challenges requires a combination of technological capabilities, strategic planning, and organizational agility to leverage the insights gained from external factors effectively.

Conclusion

- 1) This project aims to optimize store performance, forecast demand accurately, and enhance customer experience through segmentation and personalized marketing strategies.
- 2) By incorporating external factors such as regional climate, consumer price index, and fuel prices into our analysis, we have gained valuable insights into the dynamics shaping consumer behavior and purchasing patterns.
- 3) A comprehensive approach to inventory management, marketing, and store optimization, we seek to capitalize on these insights to drive sustainable growth and competitive advantage.
- 4) Tailoring our strategies to meet the specific needs and preferences of different store and department segments, we can deliver more relevant and engaging experiences that resonate with our target audience.
- 5) By Leveraging data-driven techniques such as dynamic pricing and real-time inventory management, we can adapt swiftly to changing market conditions and consumer trends, ensuring optimal product availability and pricing competitiveness.