Retail Analytics

M SALMAN ALI KHAN

DATED: OCTOBER 7TH,2023

Content

- Introduction
- Executive Summary
- Metadata
- Data Wrangling
- Data Transformation
- Exploratory Data Analysis
- Data Visualization
- Data Modeling
- Analysis Conclusion
- Considerations
- Recommendations

Introduction

- Retail, the process of selling goods and services directly to consumers, is a cornerstone of the global economy. Retail analysis involves the systematic examination of data and factors related to retail operations, including sales, customer behavior, inventory management, pricing strategies, and market trends.
- This analysis provides retailers with critical insights into their performance, helping them make informed decisions to enhance profitability, efficiency, and customer satisfaction.
- It enables retailers to adapt to changing market dynamics and consumer preferences, optimize inventory levels to minimize costs, set competitive pricing strategies, and identify growth opportunities.
- Therefore, retailers can tailor their marketing efforts, create personalized shopping experiences, and respond effectively to holiday demands.

Executive Summary

Project Objective:

The objective of this project is to conduct a comprehensive analysis of retail data, including weekly sales, markdowns, and various store attributes while considering key influencing factors like CPI, temperature, fuel price, unemployment, and holidays.

Key Insights:

- Weekly Sales and Total Markdown exhibit left-skewed distributions, indicating occasional high-value outliers, possibly due to special events or promotions while Fuel Price, CPI, and store size follow normal distributions, suggesting stability in these factors.
- Store 20, 12, and 4 perform exceptionally well in terms of sales. Type A stores lead in sales, while Type C stores have comparatively lower sales.
- Seasonal patterns show peak weekly sales in December, high markdowns in February, and elevated fuel prices in April. Fuel Price negatively influences Weekly Sales but exhibits a U-shaped relationship with Gross Markdown.
- Holidays correlate with higher sales and markdowns. Warmer temperatures above 25 degrees positively impact Weekly Sales.

Executive Summary

Immediate Future Actions:

- Develop a targeted markdown strategy to align with high-sales periods based on Total Markdown insights. Optimize pricing by focusing on moderate fuel price periods to reduce markdowns and improve profitability.
- Leverage holidays for increased sales and tailor marketing and inventory management accordingly. Utilize warm weather periods (>25°C) for stimulating consumer spending through weather-based marketing.
- Monitor CPI changes to adapt markdowns, considering higher CPI as a sign of consumer purchasing power. Concentrate sales efforts on Departments A, B, and C to maximize overall sales performance.
- Plan for peak weekly sales in the second week of December and adjust strategies accordingly. Prepare for higher markdowns in the first week of February and higher fuel prices in April, making pricing and inventory adjustments.
- Maintain consistent sales and markdown patterns throughout the week, including holidays. Continuously gather and analyze data, especially related to Fuel Price, CPI, and weather conditions, to adapt strategies to evolving market dynamics.

About the Dataset

The open-source dataset is taken from Kaggle:

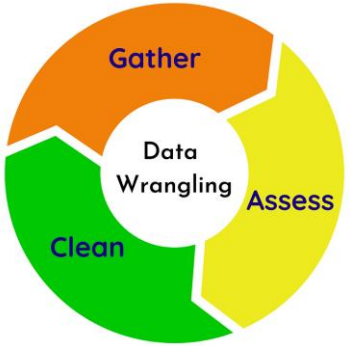
https://www.kaggle.com/datasets/manjeetsingh/retaildataset

- This dataset comprises historical sales records for 45 stores situated in diverse regions, each of these stores featuring multiple departments.
- Additionally, the company conducts several promotional markdown events throughout the year, particularly preceding major holidays such as the Super Bowl, Labor Day, Thanksgiving, and Christmas.
- It is noteworthy that week's encompassing these significant holidays carry five times the evaluation weight compared to regular non-holiday weeks.

Data Wrangling

The csv data file was extracted using readr package in R. The Data Wrangling operations were performed using dplyr, lubridate and tidyr in R. Some of the performed operations were:

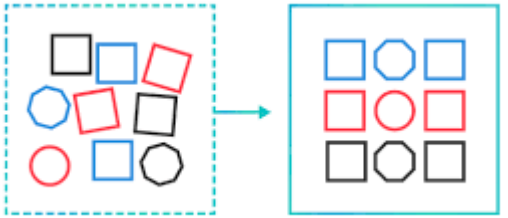
- Missing Values replacement and removal
- Removal of Duplicates
- Filtering and Sorting Columns
- Correcting Data Types



Data Transformation

Data Transformation was done by using dplyr and tidyr in R. Some of the performed operations were:

- Joining Multiple Files
- Dealing with Outliers
- Addition of calculated columns for further comparisons
- Creation of specific pivot tables for analysis and visualizations



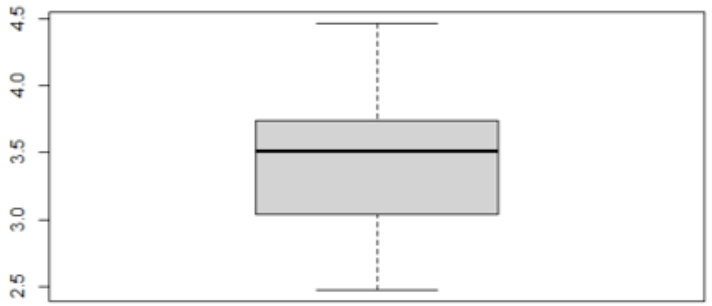
Exploratory Data Analysis

Exploratory Data Analysis

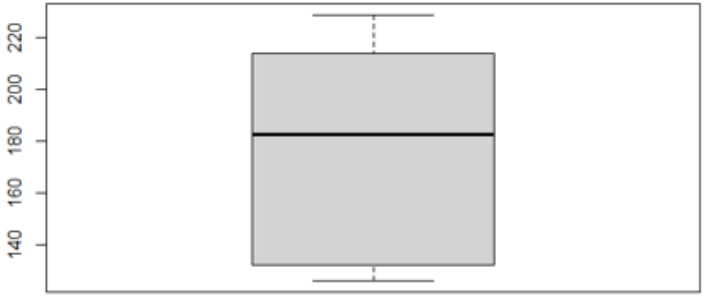
Exploratory Data Analysis was done by using dplyr, ggplot2 and tidyr in R. Some of the performed operations were:

- Summary Statistics
- Arranging, Sorting and Grouping by different Statistics to better explore data
- Box plot Creation

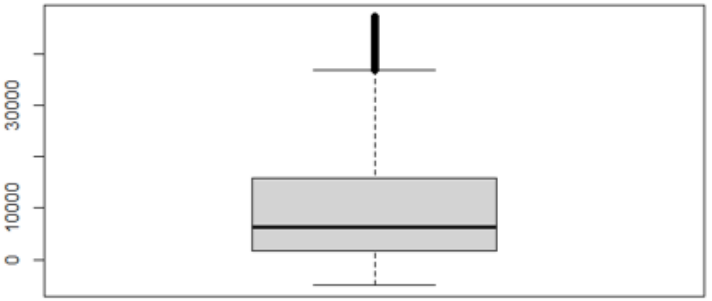
Box Plots



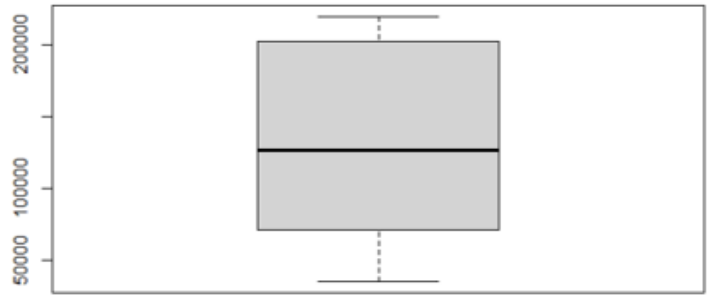
Fuel Price ($)
Slightly Right skewed



CPI ($)
Normally Distributed

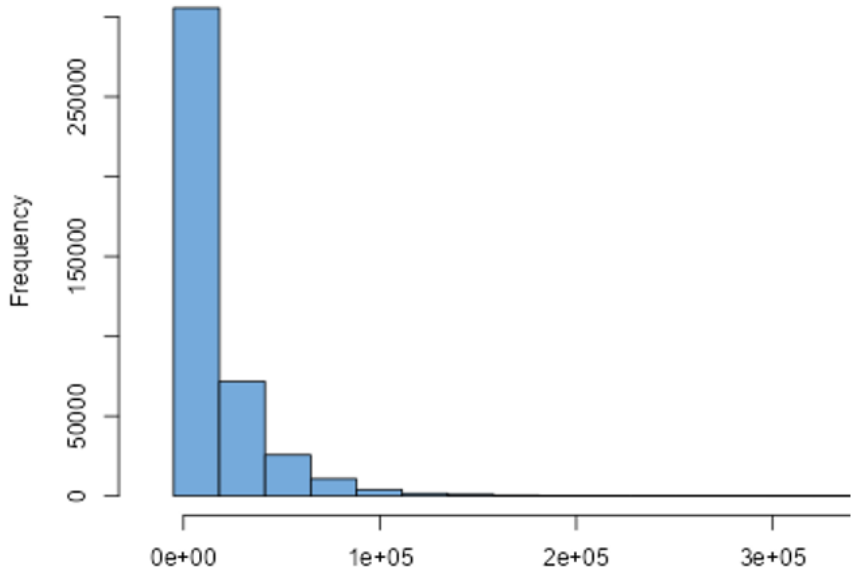


Weekly Sales ($)
Left skewed with many outliers



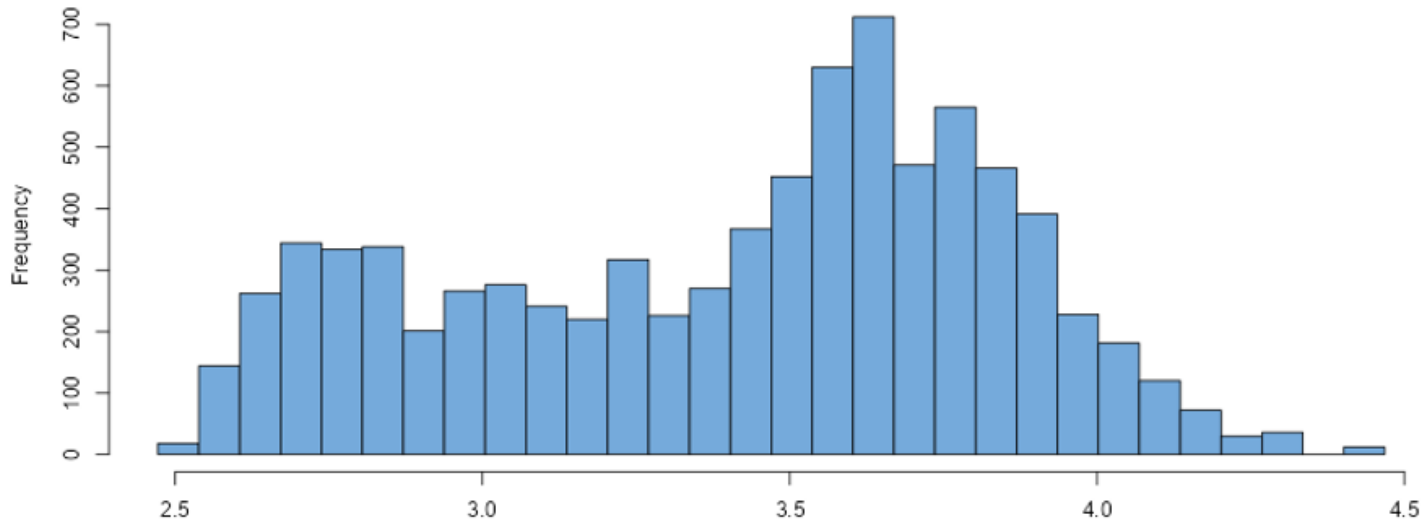
Store Size ($)
Normally Distributed

Distribution Histograms



Weekly Sales ($)

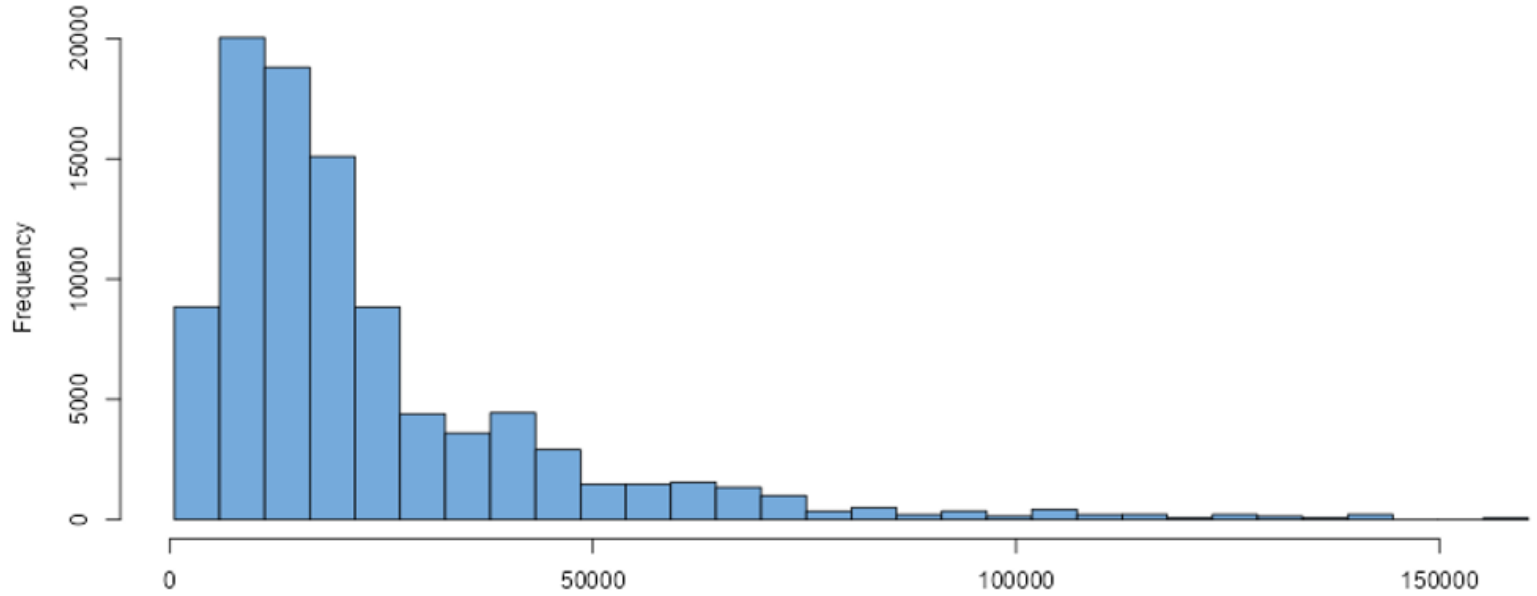
Left skewed with many outliers



Fuel Price ($)

Mostly Normally Distributed

Distribution Histograms



Exploratory Data Analysis Results

- Weekly Sales and Total Markdown are left-skewed distributions. Left-skewed distributions imply that there are a few high-value outliers or extreme values in the data.
- For Weekly Sales, this suggests that there are certain weeks where sales significantly exceed the typical values, possibly due to special events, promotions, or holidays. Total Markdown being left-skewed implies that there are periods when substantial markdowns are applied, likely during clearance sales or promotional campaigns.
- Fuel Price, CPI and store size is normally distributed. Normally distributed variables indicate a more balanced and symmetrical distribution of data points around the mean.
- Fuel Price following a normal distribution suggests that fuel prices tend to vary around an average price without extreme fluctuations. A normally distributed CPI suggests that consumer price levels are relatively stable, with most values clustering around the mean. Store Size being normally distributed indicates that store sizes are diverse but generally cluster around the average store size.

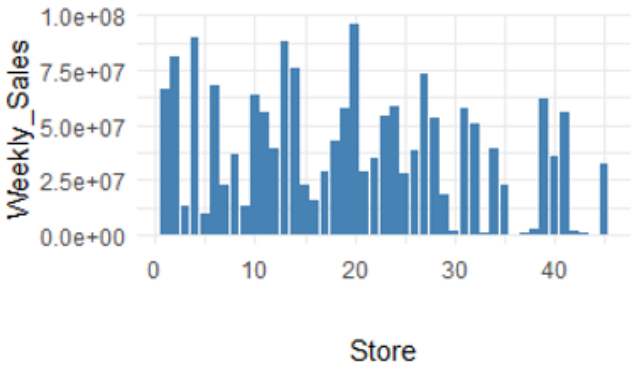
Exploratory Data Analysis Results

- When building predictive models, it's important to account for the skewness in Weekly Sales and Total Markdown, possibly by using transformation techniques like logarithmic transformations to make the distributions more symmetric.
- For normally distributed variables, traditional linear models may work well, while for left-skewed variables, regression models may need adjustments or alternative modeling approaches to handle the skewed data effectively.
- The left-skewed Total Markdown distribution could indicate the need for a strategic approach to markdowns to better align with customer behavior and preferences during high-sales periods.
- Understanding the left-skewed Weekly Sales distribution can help in planning for inventory management, staffing, and marketing efforts during peak sales weeks.

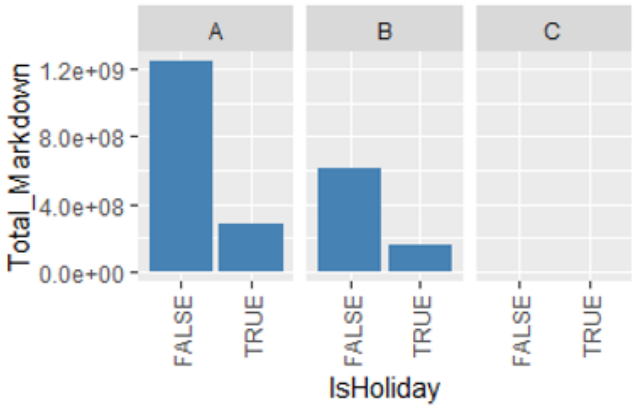
Data Visualization

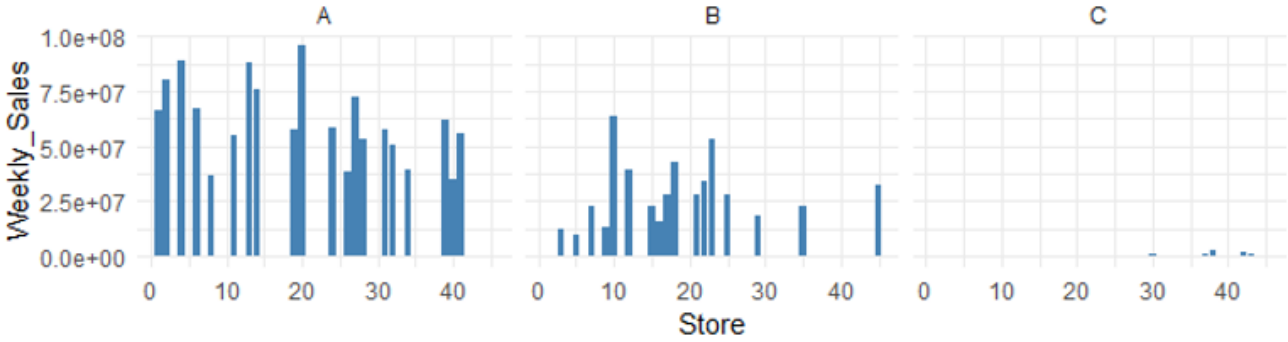
Weekly Sales against Stores and Markdown against Holidays

Store 20, 12 and 4 have made the most



The total markdown is higher for working days





Type A Stores 20, 13 and 4 have the highest

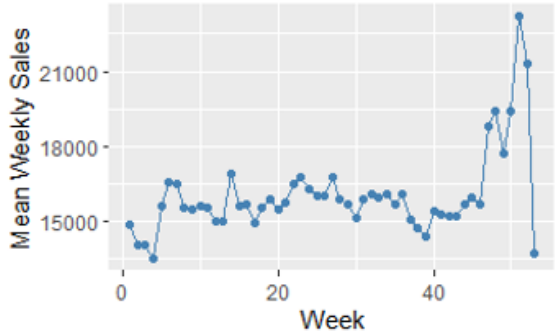
Type B Stores 10, 23 and 18 have the highest sales

Type C Stores 38 and 42 have the highest sales

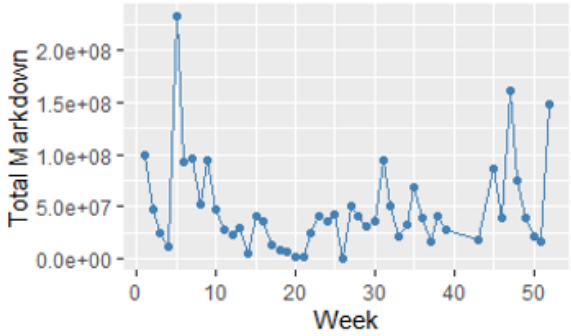
Type A stores have the highest sales followed by Type B. Type C Stores have very low sales compared to the other two.

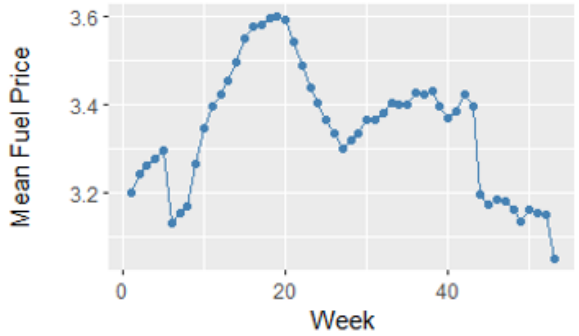
Mean Weekly Sales, Gross Markdown, Fuel Price and CPI over time.

The Mean Weekly Sales are highest in the second week of December

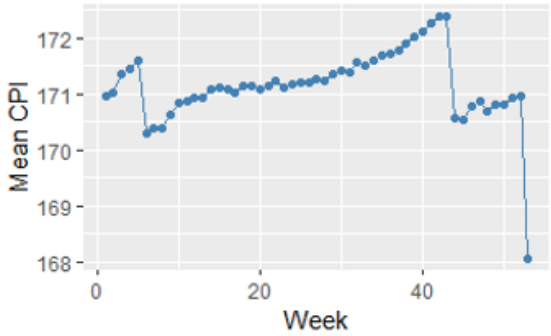


The Gross Markdown is highest in the first week of February





The Average Fuel Price is the highest in April



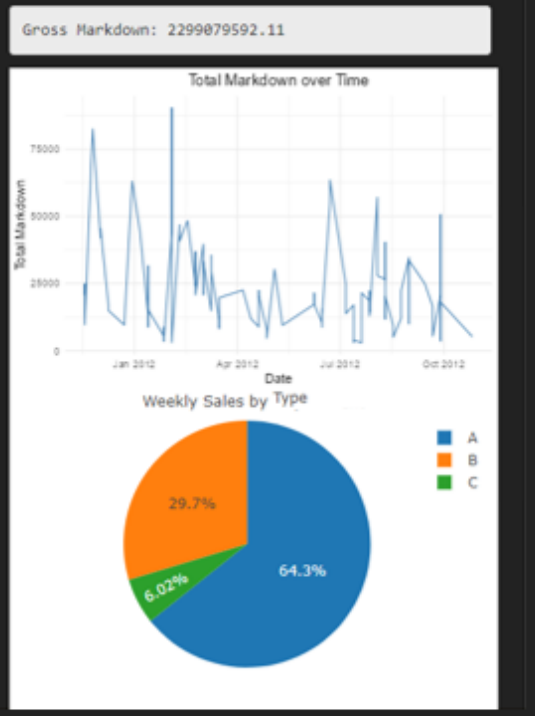
The Mean CPI Price is the highest in the later half of October

Time series Dashboard

The Weekly Sales and Total Markdown remain consistent throughout the year with some high peaks at times

The Highest Weekly Sale is observed in January 2011 whereas unusually high Total Markdown is observed in October 2012 and February 2012.

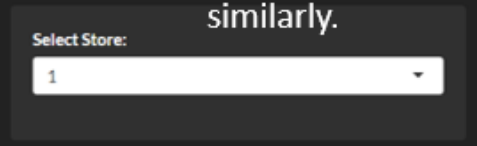
The highest total fuel price is months March, August and December but is exceptionally low in May and October.



Most Sales are made by Department A, B and C in this order.

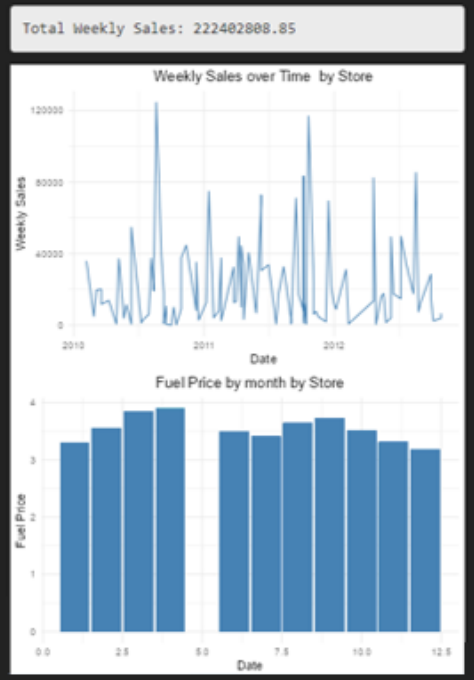
Store Dashboard

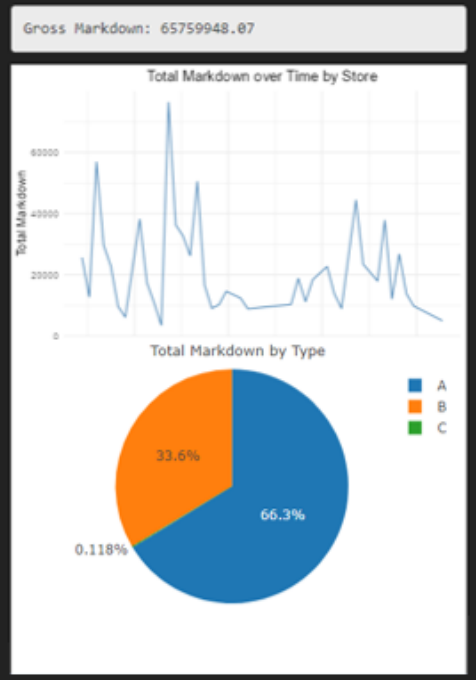
This visualization studies store 1, other stores can be studies similarly.



The Weekly Sales and Total Markdown generally remain consistent with some peaks.

The Fuel Price for a Store is consistent throughout the year.

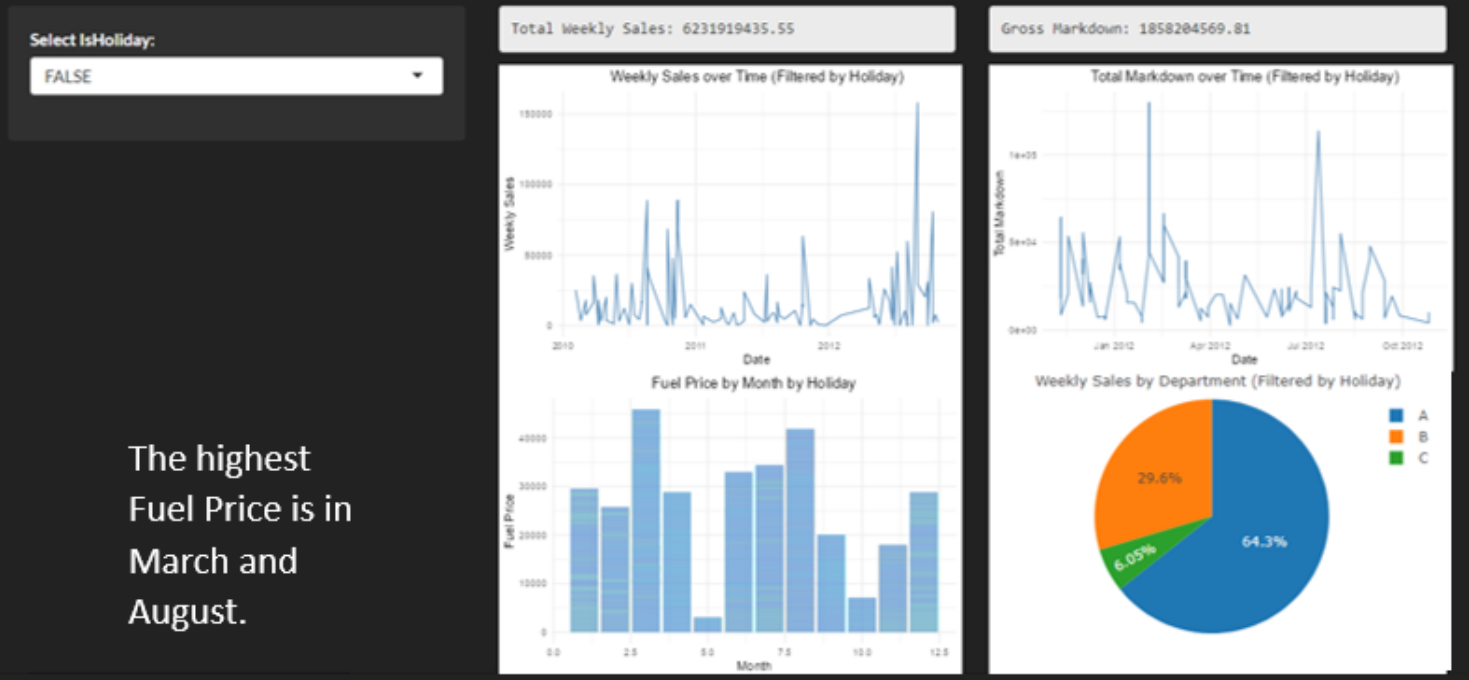




Type A Stores have the highest Gross Markdown followed by B. Type C has negligible Gross Markdown in Comparison.

Working Days Dashboard

The Weekly Sales and Total Markdown follow similar pattern on Working days as the for all

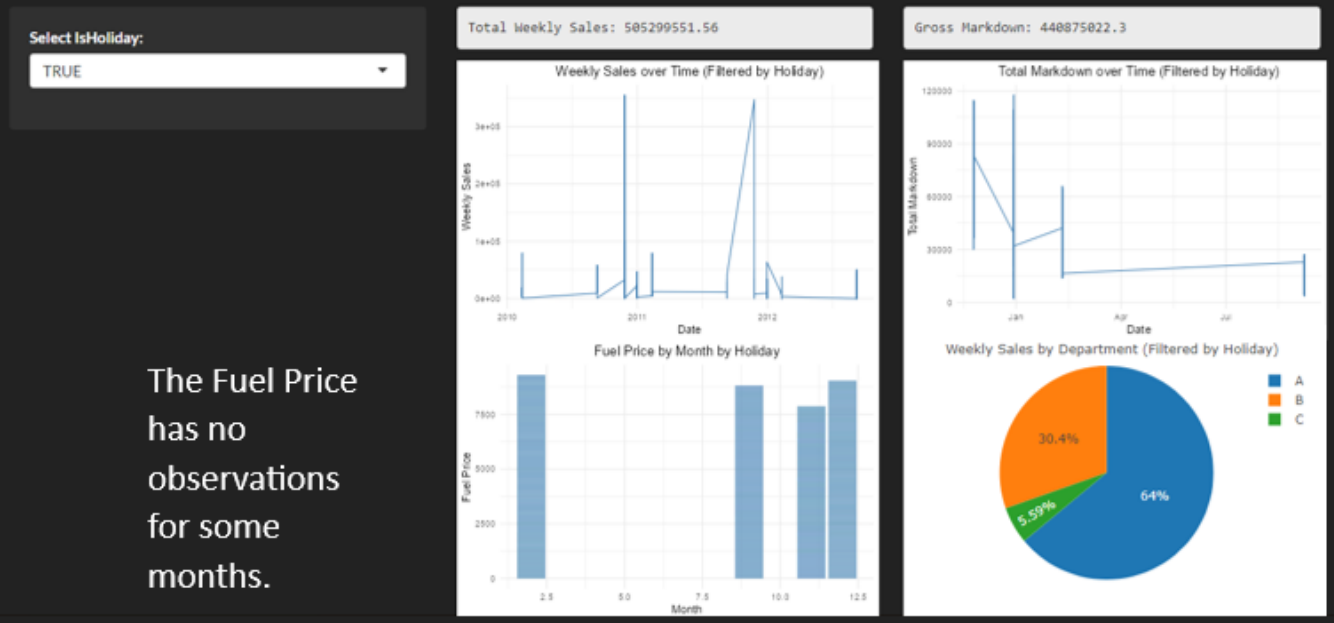


The highest Fuel Price is in March and August.

Most Sales are made by Department A, B and C in this order.

Holidays Dashboard

The Weekly Sales and Total Markdown follow similar pattern on Holidays as the for all



The Fuel Price has no observations for some months.

The distribution of weekly sales is similar to working

The distribution of weekly sales does not relate to holidays much but the Total Markdown is lower for holidays than working days.

Key Data Insights

1. Store 20, 12, and 4 have demonstrated the highest sales performance, showcasing their effectiveness in generating revenue.
2. Type A stores, represented by 20, 13, and 4, exhibit the most substantial sales figures. For Type B stores, 10, 23, and 18 lead in sales, while Type C stores, specifically 38 and 42, have notably lower sales in comparison.
3. The mean weekly sales reach their zenith during the second week of December. Gross Markdowns peak in the initial week of February, while average fuel prices show their highest values in April. The mean CPI price attains its highest point in the latter half of October.
4. Weekly sales and total markdowns display consistent patterns throughout the year, punctuated by occasional spikes in sales.

Key Data Insights

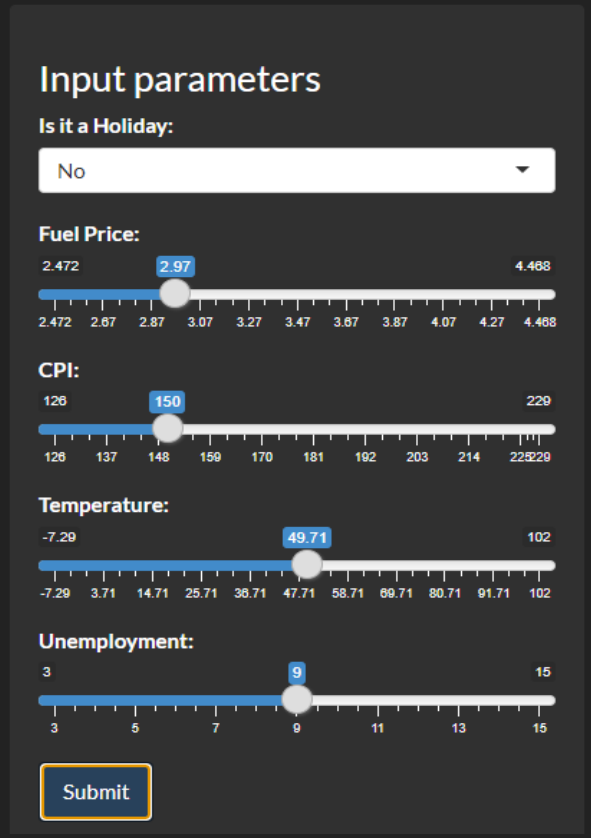
5. The highest weekly sales occur in January 2011, with notably elevated total markdowns observed in October 2012 and February 2012. Fuel prices are highest in March, August, and December but unusually low in May and October. The majority of sales are attributed to Departments A, B, and C, in that order.
6. Weekly sales and total markdowns exhibit similar patterns on working days as they do throughout the week. It's important to note that there are some months with no observations for fuel prices. Additionally, the distribution of weekly sales appears relatively consistent across holidays, while total markdowns are lower during holiday periods compared to working days.

Data Modeling

- A Shiny web application's server logic is used for predicting two different variables: Weekly Sales and Total Markdown, using XGBoost regression models.
- These models aim to predict specific outcomes and each model is trained with a set of input features and corresponding labels from the dataset.
- The XGBoost model for Weekly Sales prediction uses a more comprehensive set of input features, including "IsHoliday," "Fuel_Price," "CPI," "Temperature," and "Unemployment." This model provides users with forecasts of weekly sales based on various economic and environmental factors.
- the XGBoost model for Total Markdown prediction is trained similarly to the Weekly Sales model but focuses on predicting Total Markdown. The purpose of the Model is to assist businesses in optimizing their markdown strategies. The output is a prediction of Total Markdown based on input data.

Data Model for predicting Weekly Sales

Predict Weekly Sales



Model Result

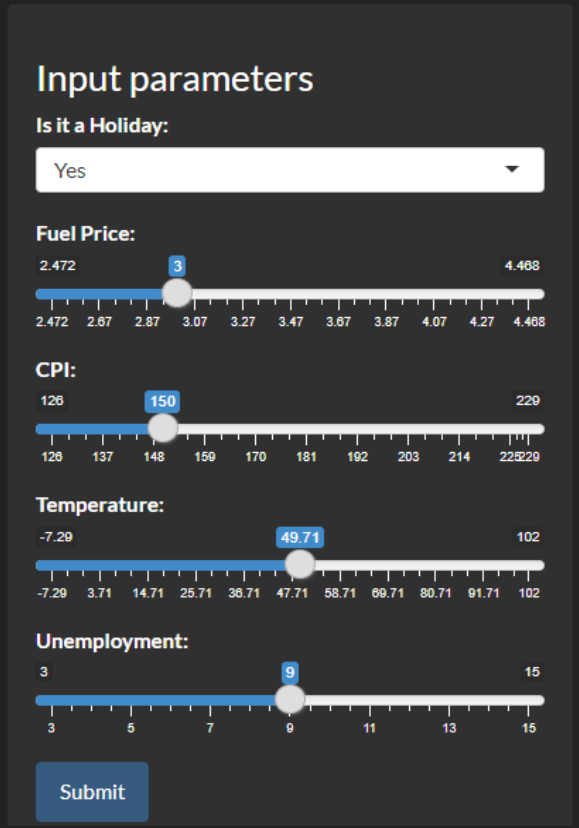
```
[1] "Calculation complete for Weekly Sales."
```

Prediction

16823.16

Data Model for predicting Total Markdown

Predict Total Markdown



Model Result

```
[1] "Server is ready for Total Markdown calculation."
```

Data Modeling Results

1. Increasing Fuel Price is associated with a decrease in Weekly Sales, indicating a negative correlation. This suggests that higher fuel prices may lead to reduced consumer spending on goods, impacting sales negatively.
2. Gross Markdown exhibits a U-shaped relationship with Fuel Price. Moderate fuel prices are associated with the lowest gross markdown, while both excessively high and low fuel prices result in increased markdowns. This suggests that optimizing pricing strategies during periods of moderate fuel prices can lead to lower markdowns and improved profitability.
3. Both sales and markdown are greater on holidays, indicating a positive relationship between them. This highlights the importance of catering to holiday shopping trends and consumer behavior.

Data Modeling Results

4. The CPI (Consumer Price Index) does not exhibit a predictable effect on Weekly Sales. This implies that changes in consumer price levels do not have a significant and consistent impact on the observed sales patterns. However, Increasing CPI is associated with a decrease in gross markdown. This implies that as consumer price levels rise, there is a potential reduction in the need for significant markdowns, possibly indicating higher consumer purchasing power.
5. Higher temperatures, specifically those exceeding 25 degrees, are linked to better sales. This suggests a positive correlation between temperature and Weekly Sales, indicating that warmer weather may encourage higher consumer spending.
6. Temperature and Unemployment do not consistently exhibit predictable effects on Gross Markdown, indicating that their impact may be less significant or subject to other contextual factors.

Analysis Conclusions

1. Weekly Sales and Total Markdown exhibit left-skewed distributions, indicating occasional high-value outliers, possibly due to special events or promotions.
2. Fuel Price, CPI, and store size follow normal distributions, suggesting stability in these factors. Skewness in Weekly Sales and Total Markdown warrants data transformation in predictive models.
3. Store 20, 12, and 4 perform exceptionally well in terms of sales. Type A stores lead in sales, while Type C stores have comparatively lower sales.
4. Seasonal patterns show peak weekly sales in December, high markdowns in February, and elevated fuel prices in April.
5. Fuel Price negatively influences Weekly Sales but exhibits a U-shaped relationship with Gross Markdown. CPI's inversely impacts Gross Markdown.
6. Holidays correlate with higher sales and markdowns. Warmer temperatures above 25 degrees positively impact Weekly Sales

Considerations to keep in mind

1. The dataset is not inclusive of all the details regarding weekly sales and markdown. As a result, the accuracy of this information is confined to the scope of the provided dataset.
2. Apart from the discussed metrics there are many other metrics that effect sales and markdown.

Recommendations

1. Develop a strategic approach for markdowns that aligns with customer behavior during high-sales periods, leveraging insights from the Total Markdown distribution.
2. Optimize pricing strategies by considering the U-shaped relationship between Gross Markdown and Fuel Price, focusing on moderate fuel price periods to reduce markdowns and enhance profitability.
3. Capitalize on the positive relationship between holidays and both sales and markdowns. Tailor marketing, inventory management, and staffing efforts to align with holiday shopping trends.
4. Consider weather conditions, particularly temperatures exceeding 25 degrees, as potential opportunities to stimulate higher consumer spending. Implement weather-based marketing strategies during such periods.
5. Focus marketing and sales efforts on Departments A, B, and C, which contribute the most to overall sales. Tailor strategies to optimize sales within these departments.

Recommendations

6. Keep a watchful eye on changes in the Consumer Price Index (CPI) as it may influence gross markdowns. Higher CPI levels may indicate greater consumer purchasing power and potentially reduce the need for significant markdowns.
7. Plan for peak weekly sales during the second week of December and adjust staffing, inventory, and marketing strategies accordingly.
8. Be prepared for elevated gross markdowns in the initial week of February and higher fuel prices in April. Adjust pricing and inventory management during these periods.
9. Maintain consistency in sales and markdown patterns between working days and the entire week. Ensure that the distribution of weekly sales remains relatively consistent across holidays.
10. Continue to gather data and monitor trends, especially regarding Fuel Price, CPI, and weather conditions, to adapt strategies as market dynamics evolve.

THE END