Subrata Mandal Suman
05-06-2024

Analyzing the Influence of Economic Indicators and Seasonal Factors on Walmart's Weekly Sales: A Linear Regression Approach

Introduction

As one of the world's largest retail companies, Walmart holds a significant global market position (1). In the competitive world of retail operations, understanding the factors that influence sales is crucial for planning and performance. Weekly sales at large retail chains like Walmart are influenced by various factors ranging from economic indicators to weather conditions (2,3). This study focuses on identifying and analyzing these factors to provide a better picture of the impacts on Walmart's weekly sales.

This research addresses the effect of weekly sales at Walmart stores using various economic indicators such as the Consumer Price Index (CPI), unemployment rates, fuel prices, and other variables like temperature, month, and holiday effects. This is a significant area of study because understanding these factors that influence weekly sales can lead to more effective business decisions. We often think that holidays increase sales because people are off work and shop more. We can use this idea to see if weekly sales are significantly higher on holidays. Also, the weather affects sales, if it's too cold or too hot, people prefer to stay at home. Additionally, economic factors can influence consumer behavior and purchasing pattern. Higher fuel prices mean people have less income; a higher CPI indicates increased prices in general; and higher unemployment can reduce overall spending. Sales trends can be seasonal, for example, December might see higher sales due to the holiday season. By breaking down the 'Date' into month and year, we will analyze patterns. Each of these questions uses the linear regression model to explore and answer questions about factors influencing the weekly sales of Walmart.

Previous research has emphasized various influences on retail sales, including macroeconomic conditions such as consumer price index (CPI) and unemployment rates, as well as other factors like local weather conditions and holiday effects. One study has demonstrated the effectiveness of machine learning models in forecasting retail sales (4). Furthermore, research by Vyas and AS emphasize the significance of incorporating seasonality into sales predictions, particularly for Walmart (5). Another study by Chenghao Yu, has explored the use of different regression models to predict Walmart's sales, focusing on model accuracy and efficiency (6). Others have integrated machine learning models like XGBoost to enhance predictive performance further (7). However, many of these studies have not fully analyzed the interaction of CPI and unemployment with

different months, temperature, fuel price, and holidays, which can provide deeper insights into consumer buying patterns. Additionally, there remains a gap in the literature regarding the interaction effects between economic indicators and other factors and how they collectively impact weekly sales. This study seeks to bridge this gap by employing a linear regression model.

The structure of this paper is organized in a proper way. The next section discusses description of data and data collection procedure, statistical formulation, and methods, including the experimental design. After that, we will present our main results, which highlight key findings and their implications for sales strategies. Then, the discussion and limitations section interpret these results in the context of existing literature and our limitations. This paper also suggests future research directions.

Materials and Methods

Description of Data:

The Walmart dataset from Kaggle (8) used for the analysis comprises weekly sales data from 45 Walmart stores across the United States and different dates, including economic indicators and seasonal factors influencing sales. This dataset contains a total sample size of 6,435 with each representing weekly sales data for a specific store on a given date and includes observations of weekly sales alongside variables such as the Consumer Price Index (CPI), unemployment rates, fuel prices, temperature, date, and holidays flag. This comprehensive data set provides the best framework for analyzing various factors affecting weekly sales.

Experimental Design and Data Collection Procedures:

This dataset was compiled to observe the influence of multiple economic and seasonal factors on the weekly sales. Data collection spanned from 2010 to 2012. Each store's data entries include the date of transaction, weekly sales figures in USD, local temperature in degrees Fahrenheit in the region, fuel prices cost per gallon, CPI, and the unemployment rate. Additionally, the data distinguishes between holiday and non-holiday in the specific week. Weekly sales were automatically captured at each Walmart store, which records all transactions made during the week. Temperature, fuel price, CPI, and unemployment were gathered from external sources such as government, state reports, and weather stations. Since the data was collected from Walmart stores, it represents a diverse set of geographic areas across the United States. The data also includes economic indicators, which are specific to the U.S. economy. Moreover, Walmart stores are characterized by offering a wide range of products, from groceries to electronics, which attracts a broad segment of consumers. Therefore, this data represents to American consumers shopping in large retail environments.

Statistical Formulation and Methods:

The primary scientific questions addressed in this analysis are: How do various factors such as economic indicators (CPI, unemployment rate, and fuel price), and holidays, months, and temperature influence weekly sales at Walmart? Is there an interaction effect between CPI and unemployment rates on Walmart's weekly sales? The data set contains "Date" variable, and we break down the "Date" into month and year to identify is there any relationship of month on weekly sales.

Before answering these questions, exploratory data analysis is performed to visualize the data distribution and underlying relationships between variables. These visualizations are pivotal in understanding the data structure and guiding the statistical analysis. The frequency distribution along with the percentage is constructed for categorical variable, and for continuous variables, mean and standard deviation gives the idea about the variation of the data. We construct a scatter plot of weekly sales vs. CPI by month and a scatter plot of unemployment vs. weekly sales by month to see the relationship between weekly sales of economic variables stratified by month.

After exploratory data analysis, we employ multiple linear regression model, a statistical model useful for understanding the relationship between one dependent variable and two or more independent variables, to answer these questions. The model included not only the main effects of each variable but also considered the interaction effect between CPI and unemployment rates to capture any effects. The specific model formula used is:

$$\text{Weekly Sales} = \beta 0 + \beta 1 \times \text{CPI} + \beta 2 \times \text{Unemployment} + \beta 3 \times \text{Holiday\_Flag} + \beta 4 \times \text{Temperature} + \beta 5 \times \text{Month} + \beta 6 \times \text{Fuel\_Price} + \beta 7 \times (\text{CPI} \times \text{Unemployment}) + \epsilon$$

Where ϵ is an error term, assumed to be normally distributed.

In our model, the main covariate of interest is Consumer Price Index (CPI). When examining the relationship between the Consumer Price Index (CPI) and Weekly Sales, several other covariates can potentially confound this relationship. Both CPI and unemployment are economic indicators that can influence consumer spending. Higher unemployment reduces income, which can affect sales independently of CPI, thereby confounding the relationship between CPI and Weekly Sales. The cost of fuel can impact consumer spending patterns, as higher fuel prices decrease the amount of income consumers are willing to spend on shopping. Since CPI is also influenced fuel price, can confound the relationship between CPI and Weekly Sales. Seasonal (Month) fluctuations can impact on both CPI and Weekly Sales. Holidays can significantly impact both CPI and Weekly Sales due to changes in consumer spending habits during these periods. Temperature can influence both Weekly Sales and CPI independently. Extreme temperatures prohibit shopping

activities, leading to fluctuations in sales and price levels. Additionally, temperature variations impact supply chain efficiency and costs, that is reflected in the CPI. We also include interaction term between CPI and unemployment. For this reason, we include these variables in our model.

Model diagnostics and validation are conducted to ensure the appropriateness of the regression model, included in the Appendix section C. All analysis is performed using the R Programming language.

Results

Summary of Descriptive Statistics:

Table 1: Descriptive Statistics of Categorical Variables

Variables	Frequency (n)	Percentage (%)
Holiday Flag		
0	5985	93.01
1	450	6.99
Month		
Jan	360	5.60
Feb	540	8.40
March	585	9.09
Apr	630	9.79
May	540	8.40
Jun	585	9.09
Jul	630	9.79
Aug	585	9.09
Sep	585	9.09
Oct	585	9.09
Nov	360	5.60
Dec	450	6.99

Table 1 represents most of the data points, 93.01%, correspond to non-holiday weeks, with only 6.99% representing holiday weeks. Monthly data distribution shows relatively even spread across the year, with March, June, August, September, and October each accounting for 9.09% of the data. January and November share the lowest representation at 5.60%, while December and May align with the percentage at 6.99% and 8.40%, respectively. This variation in monthly data distribution may reflect seasonal sales cycles or data collection operation at Walmart.

Table 2: Descriptive Statistics of Continuous Variables

Variables	Mean	Standard Deviation
Weekly Sales	1046965	564366.6
Temperature	60.66	18.44
Fuel Price	3.36	0.46
CPI	171.58	39.36
Unemployment	7.99	1.88

Table 2 represents the mean weekly sales are at $1,046,965, but depicts considerable variation, as indicated by a high standard deviation of $564,366.6. The temperature averages 60.66 degrees Fahrenheit with a moderate variation (SD = 18.44), suggesting fluctuations in weather conditions. Fuel prices and CPI are relatively stable, averaging 3.36 dollars and 171.58 respectively, with low to moderate variations. Unemployment stands at 7.99% on average, with a standard deviation of 1.88, indicating relatively steady employment conditions over the period.

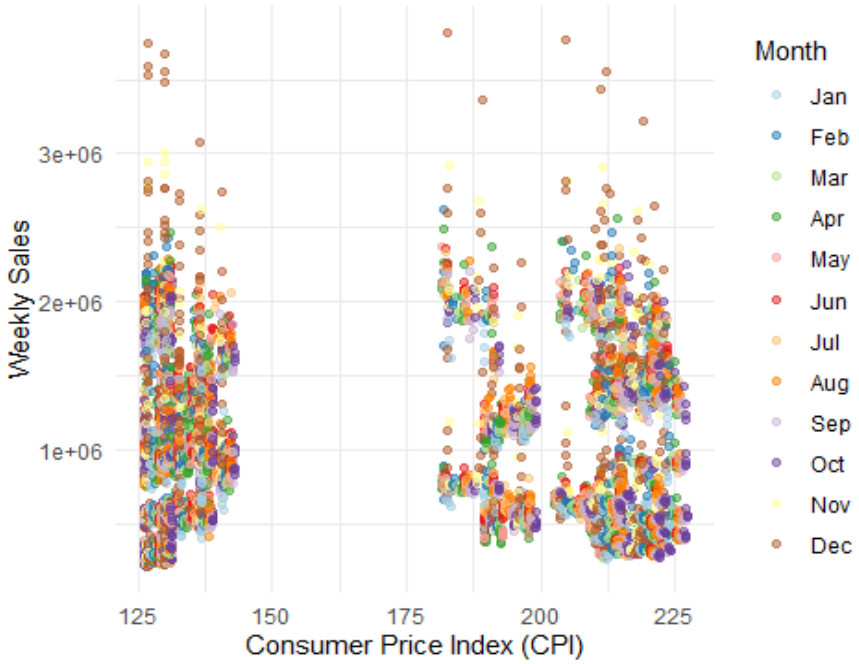


Figure 1: Scatter plot of Weekly Sales vs CPI by Month

Figure 1 shows a dispersed relationship between CPI and Weekly Sales, with no clear trend indicating a direct correlation. Data points are spread across various CPI values, suggesting that additional factors may influence sales. The color differentiation by month does not expose any

consistent seasonal pattern, indicating variability in sales respond to economic conditions across the year.

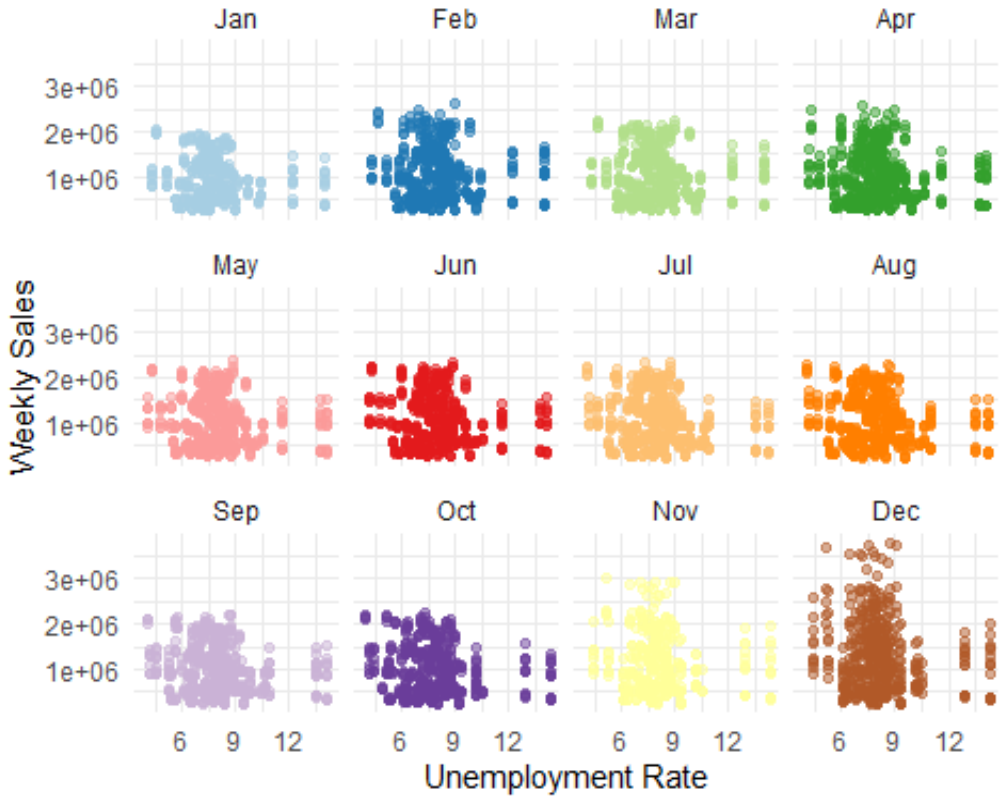


Figure 2: Scatter Plot of Unemployment vs. Weekly Sales by Month

Figure 2 displays varying relationships between unemployment rates and weekly sales across different months. The data points are spread widely for most months, indicating no strong correlation between unemployment and sales, highlighting those sales are influenced by other factors beyond unemployment alone within each month. However, some months, particularly May through August, show denser clusters at lower unemployment rates, suggesting better sales performance during lower unemployment periods in these months.

Table 3: Linear Regression Results

Coefficients	Estimate	Std. Error	95% CI	t value	P value
Intercept	2523984.4	155435.2	(2219279.53, 2828689.22)	16.238	0.000 ***
CPI	-9003.2	999.1	(-10961.72, -7044.74	-9.012	0.000 ***
Unemployment	-182941.1	18971.4	(-220131.38, -145750.85)	-9.643	0.000 ***
Holiday	33622.6	29664.1	(-24528.88, 91774.10)	1.133	0.257
Temperature	1028.1	697.8	(-339.79, 2395.97)	1.473	0.140
Month02	126975.3	38232.4	(52026.97, 201923.56)	3.321	0.000 ***
Month03	79282.1	37851.6	(5080.28, 153483.98)	2.095	0.036*
Month04	81656.6	39006.6	(5190.73, 158122.52)	2.093	0.036*
Month05	77122.6	42254.4	(-5710.07, 159955.34)	1.825	0.068
Month06	102134.3	45115.7	(13692.54, 190576.12)	2.264	0.024*
Month07	65955.9	46656.7	(-25506.73, 157418.53)	1.414	0.157
Month08	80403.8	46831.0	(-11400.69, 172208.20)	1.717	0.086
Month09	21305.8	44704.0	(-66328.95, 108940.57)	0.477	0.634
Month10	50517.9	40439.9	(-28757.91, 129793.64)	1.249	0.212
Month11	206869.6	42618.6	(123322.84, 290416.29)	4.854	0.000 ***
Month12	353859.5	39444.4	(276535.38, 431183.68)	8.971	0.000 ***
Fuel Price	17920.4	16156.7	(-13752.00, 49592.85)	1.109	0.267
CPI:Unemployment	968.7	129.4	(715.12, 1222.35)	7.488	0.000 ***

*** indicates highly significant

Linear Regression results shows the coefficient for CPI is negative (-9003.2) with a highly significant p-value ($p < 0.001$), suggesting that an increase in the Consumer Price Index is associated with a decrease in weekly sales. This reflects the impact of increasing prices on consumer purchasing behavior. Similarly, unemployment has a significant negative effect on sales (-182941.1, $p < 0.001$), indicating that higher unemployment rates may lead to reduced consumer spending at Walmart. On the other hand, the holiday shows a positive coefficient (33622.6), although it is not statistically significant ($p = 0.257$), suggesting that holidays can potentially increase sales, but the effect is weak.

Several months (February, March, April, June, November, and December) show statistically significant positive effects on sales compared to January, with the effects ranging from roughly 79282.1 to 353859.5. This variability reflects seasonal shopping behaviors.

The interaction term between CPI and Unemployment is positive (968.7) and significant ($p < 0.001$). This suggests that the negative impact of unemployment on sales is moderated or lessened when CPI is higher, which can indicate economic variation where price increase and

unemployment rates interact to affect consumer behavior. On the other hand, temperature and Fuel Price have positive coefficients, suggesting they positively influence weekly sales, but neither is statistically significant in this model, indicating weak effects.

Discussion and Limitations

The analysis conducted in this study using linear regression revealed significant insights into the factors influencing Walmart's weekly sales. Our findings indicate that both the Consumer Price Index (CPI) and unemployment rates have substantial negative effects on sales, suggesting that higher prices and unemployment prohibit consumer spending at Walmart. Interestingly, the interaction between CPI and unemployment suggests a dynamic effect where the impact of unemployment on sales is moderated under conditions of higher CPI, potentially indicating an interplay between these economic indicators during periods.

Holidays, although expected to boost sales, did not show a statistically significant effect, which suggests that any increase in sales during holidays is not distinctly different from non-holiday periods. Seasonal variations, represented by monthly changes, showed significant impacts, with certain months like February, November, and December experiencing higher sales, aligning with holiday shopping behaviors.

Despite these insights, the study has several limitations. The dataset was from 2010 to 2012, which doesn't fully capture the current economic conditions or consumer pattern. Additionally, the model accounted for linear relationships and interactions but did not explore non-linear dynamics that may provide deeper insights. The focus on only Walmart stores is also limiting the generalization of the findings to other retail corporations. Moreover, the influence of online shopping trends and their impact on physical store sales was also not considered, which can provide insights into changing sales patterns.

The linear regression model employed in this study provides a foundational understanding of the factors influencing weekly sales. However, further exploration into nonlinear relationships and higher-order interactions and polynomials can yield deeper insights. Future studies can benefit from incorporating machine learning techniques and time series analysis, which can handle nonlinear relationships and interactions more effectively than traditional regression models, which are from some literature reviews (4–7). Investigating the role of online sales, particularly in response to the increasing trend of e-commerce, would also be efficient (9,10). By addressing and exploring these limitations, future research can continue to explore our understanding of retail sales variation and contribute to more effective business strategies in retail corporations.

Author Statement

This study uses publicly available data, which contains no ethical or privacy concerns, available at Kaggle (8).

Appendix

We also performed a simple model to check the relationship of various factors and weekly sales but without interaction term. In this section, we also include histogram, scatter plots, and box plots for all variables in the Walmart's weekly sales dataset. In this section, we also diagnostics of the main model, which was included in our main analysis. We focused on the main model to identify the factors influencing weekly sales.

Appendix A:

We also used another simple model, including less variable and without interaction term.

This simple model is: Weekly Sales = $\beta 0$ + $\beta 1$×CPI + $\beta 2$×Month + $\beta 3$×Holiday_Flag + $\beta 4$× Unemployment + ϵ

Where ϵ is an error term, assumed to be normally distributed.

Table 1: Simple Model Regression Results

Coefficients	Estimate	Std. Error	95% CI	t value	P value
Intercept	1555771.6	57868.9	(1442329.38, 1669213.89)	26.884	0.000 ***
CPI	-1662.0	183.6	(-2021.96, -1302.05)	-9.051	0.000 ***
Month02	130565.8	38337.8	(55410.96, 205720.57)	3.406	0.000 ***
Month03	97674.0	37022.2	(25098.16, 170249.91)	2.638	0.008**
Month04	109437.6	36513.1	(37859.71, 181015.56)	2.997	0.003**
Month05	113240.7	37600.9	(39530.39, 186950.93)	3.012	0.003**
Month06	144436.9	37016.4	(71872.52, 217001.33)	3.902	0.000 ***
Month07	110841.2	36508.8	(39271.67, 182410.70)	3.036	0.002**
Month08	123965.5	37015.2	(51403.27, 196527.74)	3.349	0.000 ***
Month09	60890.6	37647.6	(-12911.25, 134692.41)	1.617	0.106
Month10	72733.2	37016.8	(167.97, 145298.35)	1.965	0.049*
Month11	225052.0	41863.1	(142986.32, 307117.59)	5.376	0.000 ***
Month12	361513.5	39534.3	(284013.08, 439013.88)	9.144	0.000 ***
Holiday	31473.0	29767.0	(-26880.29, 89826.27)	1.057	0.290
Unemployment	-43900.5	3858.4	(-51464.28, -36336.79)	-11.378	0.000 ***

*** indicates highly significant

The simple model regression results show significant relationships between the weekly sales and most predictors. The coefficient for CPI is negative, indicating a decrease in the weekly sales with

an increase in CPI. Unemployment also has a significant negative effect on the weekly sales. The most monthly coefficients are significant, implying strong monthly variations.

Appendix B:

Additional Figures:

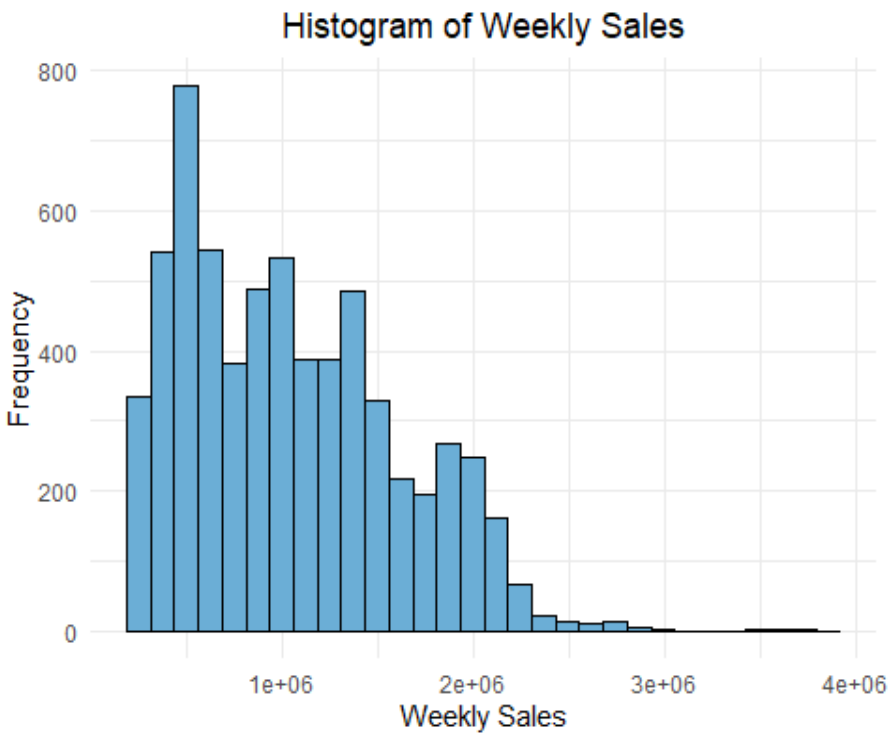


Figure 1: Histogram of Weekly Sales

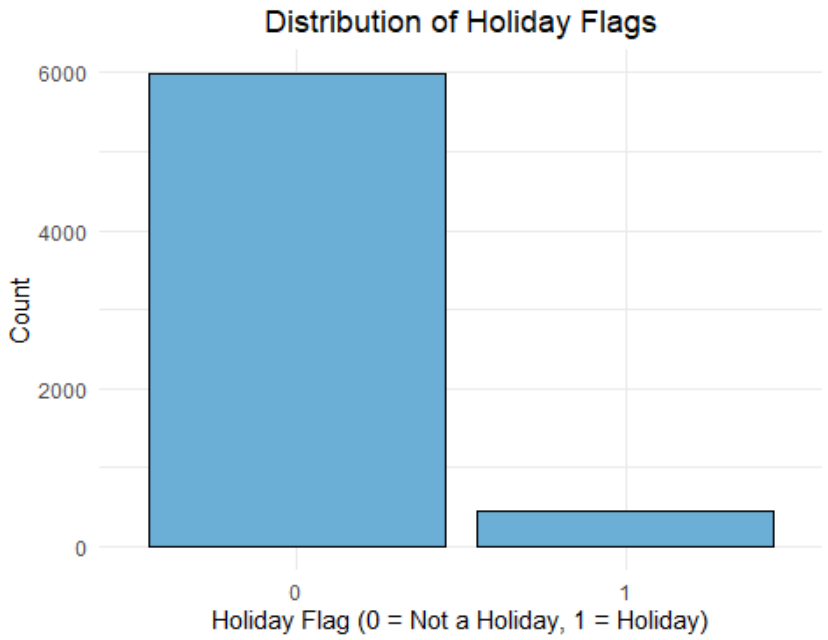


Figure 2: Distribution of Holiday Flags

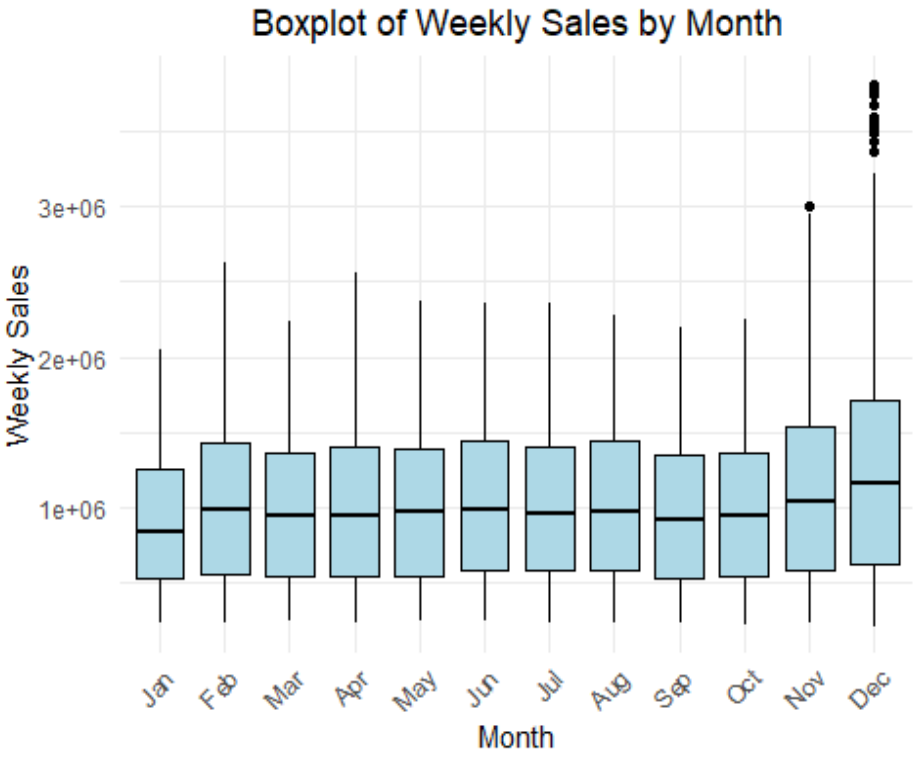


Figure 3: Boxplot of Weekly Sales by Month

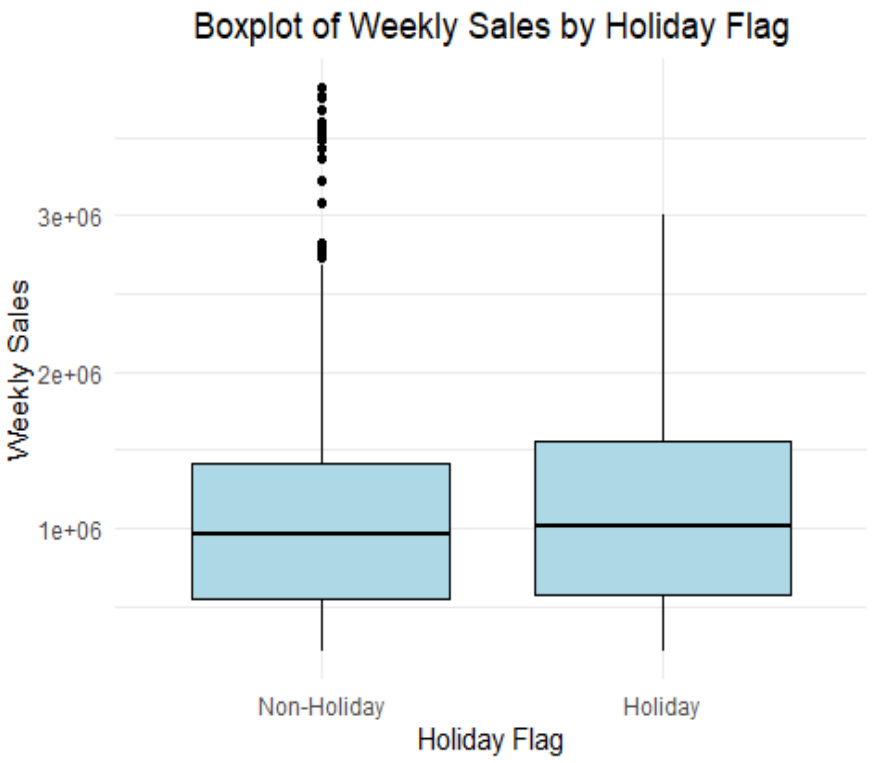


Figure 4: Boxplot of Weekly Sales by Holiday Flag

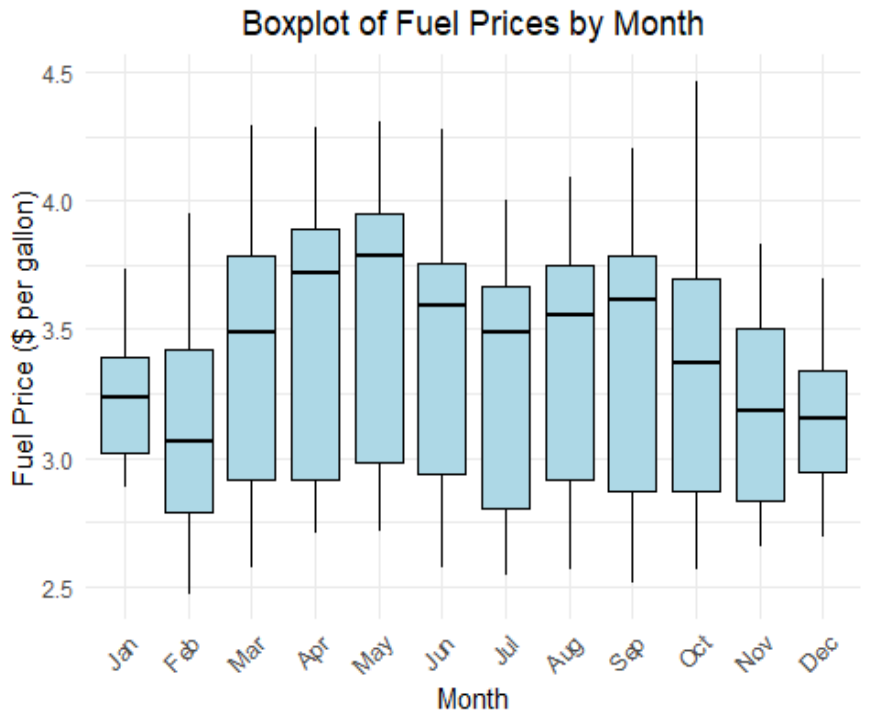


Figure 5: Boxplot of Fuel Prices by Month

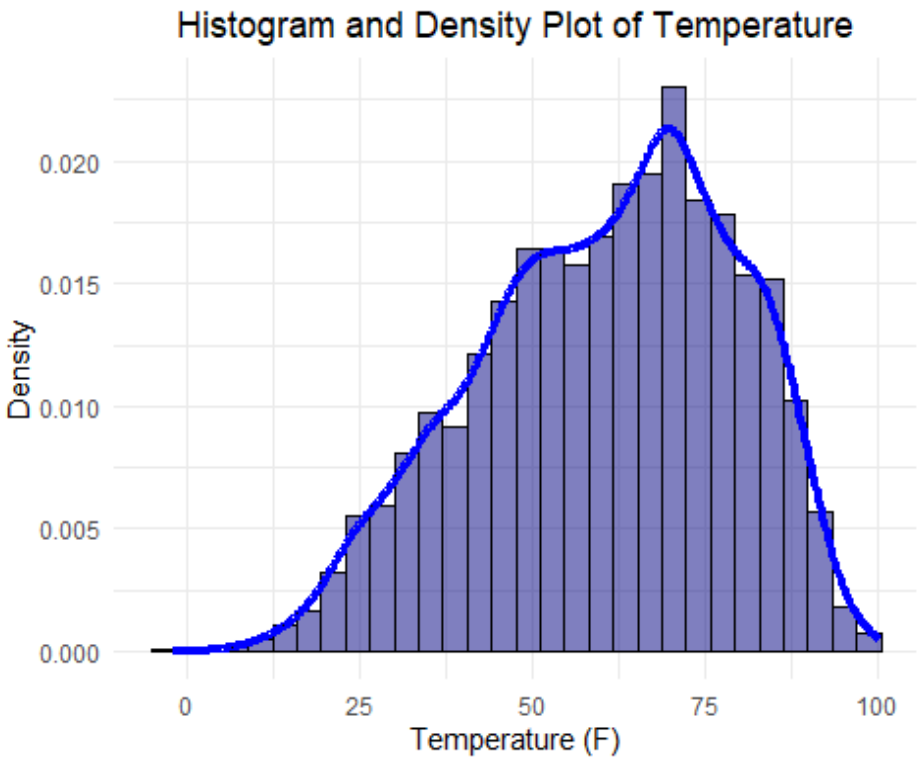


Figure 6: Histogram and Density Plot of temperature

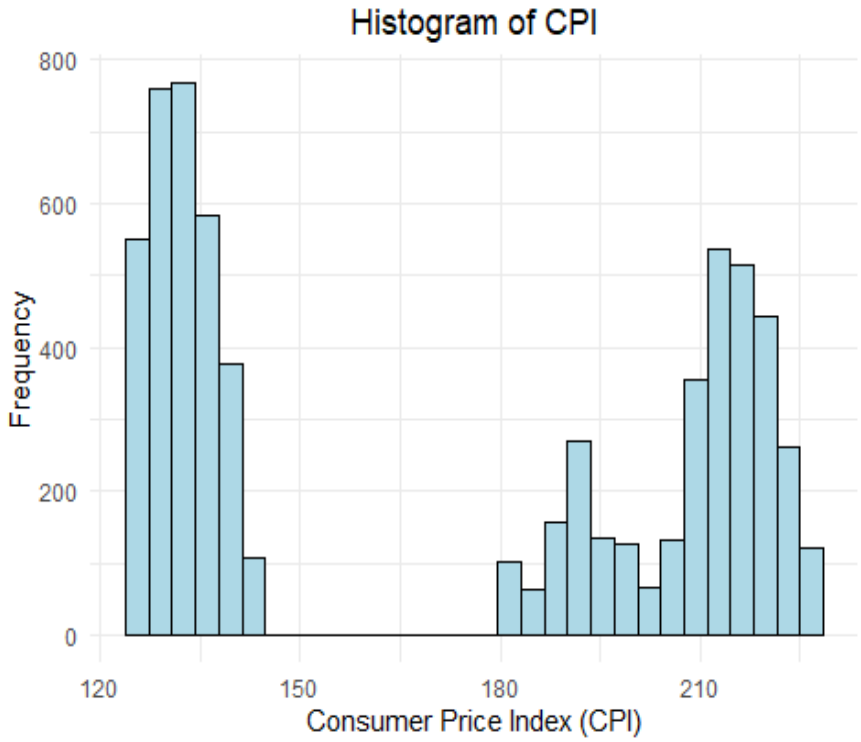


Figure 7: Histogram of CPI

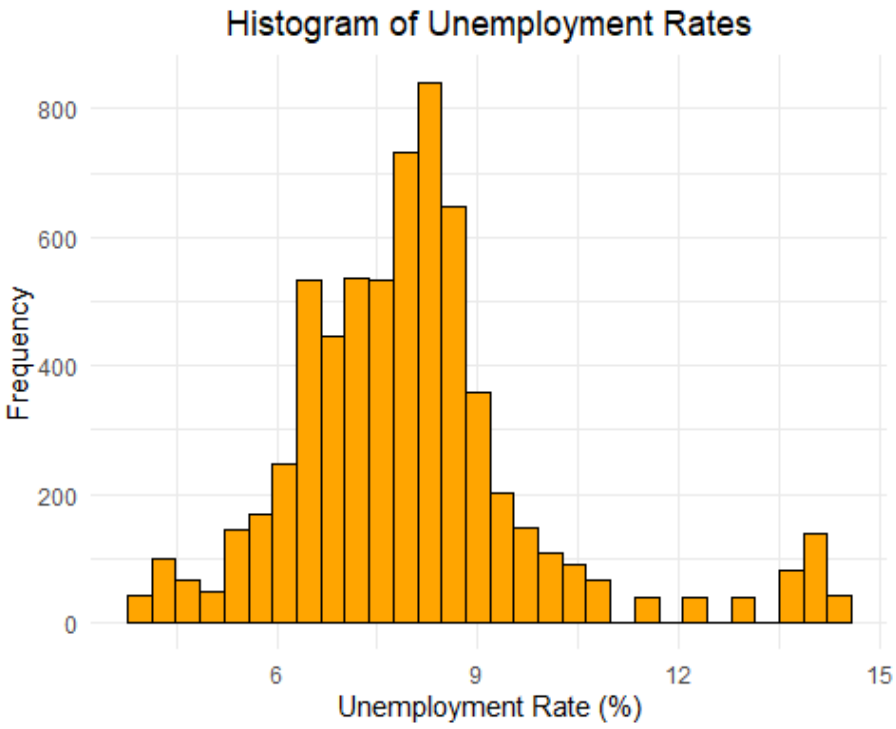


Figure 8: Histogram of Unemployment Rates

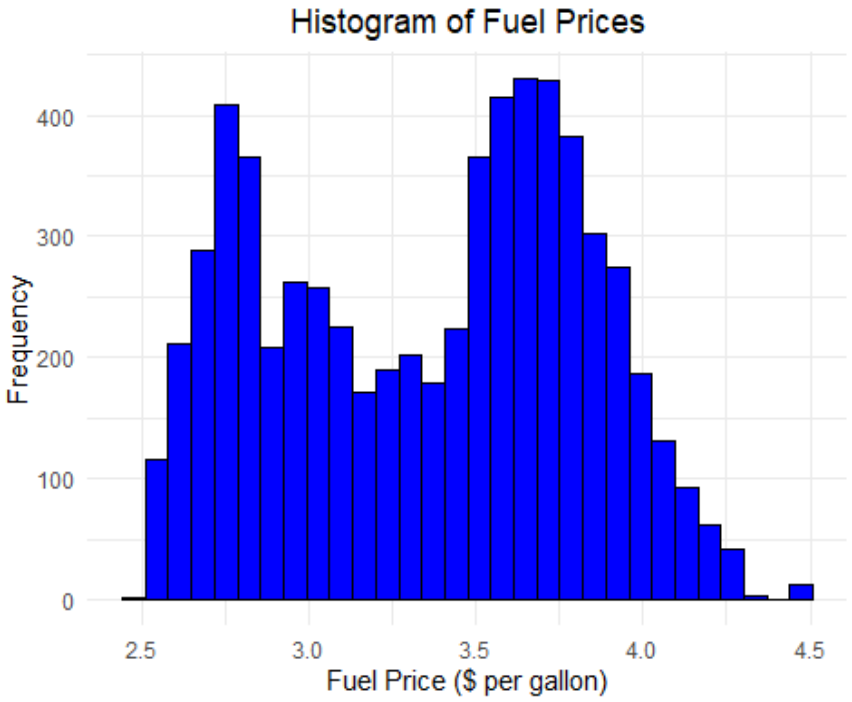


Figure 9: Histogram of Fuel Prices

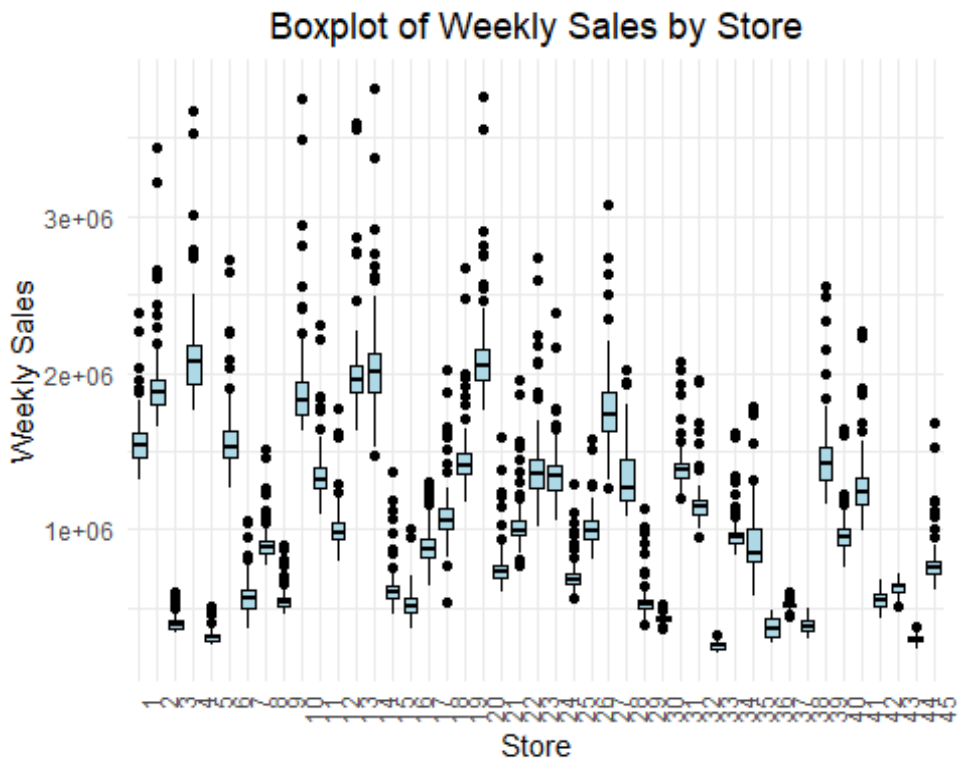


Figure 10: Boxplot of Weekly Sales by Store

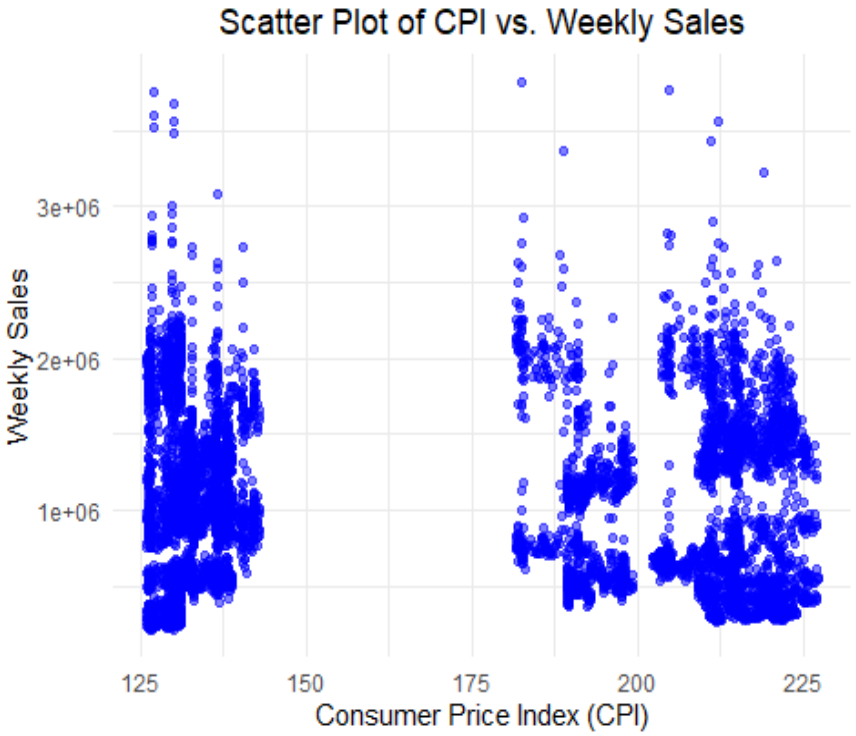


Figure 11: Scatter Plot of CPI vs. Weekly Sales

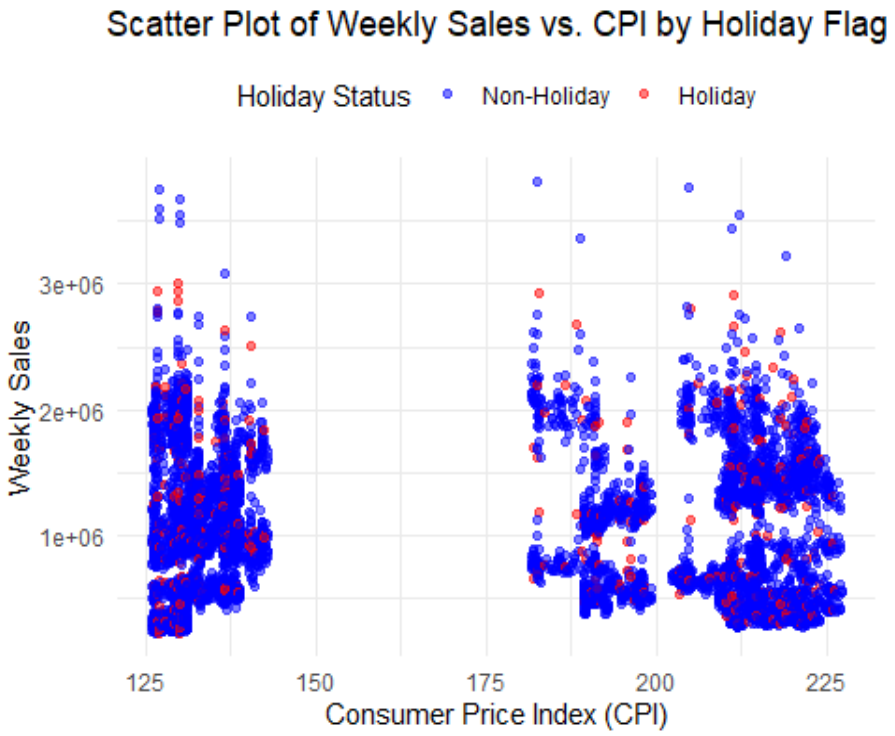
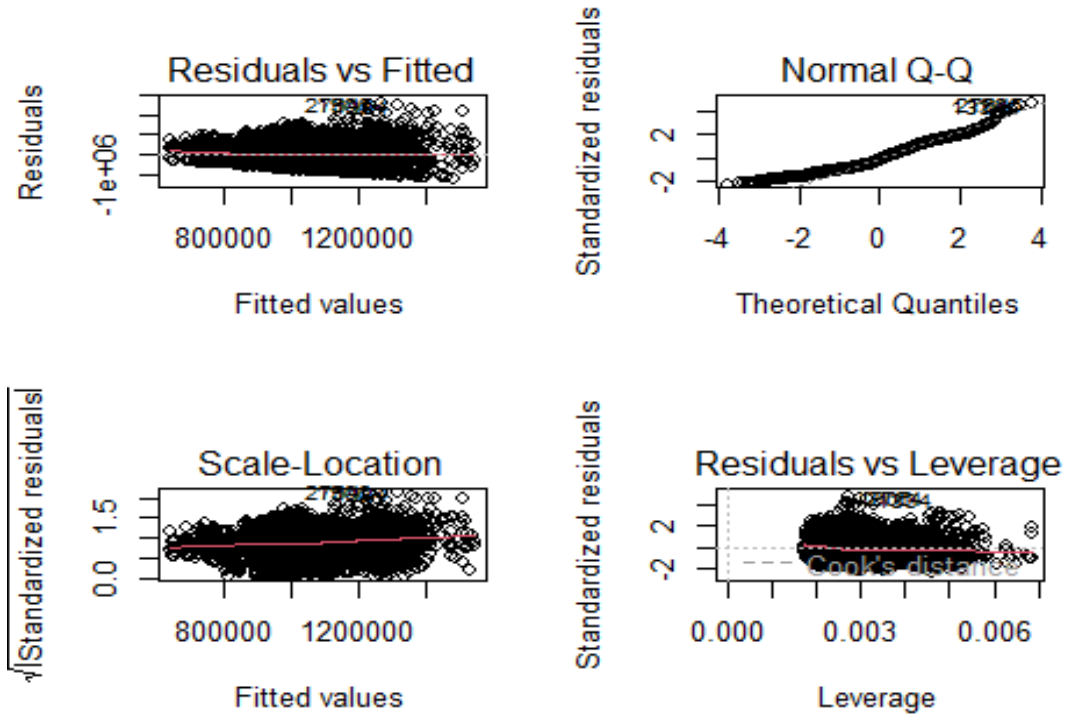


Figure 12: Scatter Plot of Weekly Sales vs. CPI by Holiday Flag

Appendix C:

Diagnostics of the Main model used in our main analysis:



Residuals vs Fitted plot shows the residuals seem to be evenly distributed across different values of fitted line, which is a good sign. However, there is a slight fan shape, suggesting potential issues with homoscedasticity. Normal Q-Q plot shows most points on the line, but there are some deviations at both tails (especially the right tail). This suggests some deviation from normality, particularly with extreme values. In the Scale-Location plot, the line is horizontal with some minor deviations, suggesting mild issues with homoscedasticity. Residuals vs Leverage plot shows a few points with high leverage, but none seems to significantly exceed the Cook's distance, which is positive. However, the existence of these high leverage points can influence the model's performance.

Addressing the mild heteroscedasticity and deviations from normality seen in the residuals, we can use transformations or robust regression techniques or non-parametric methods. Applying a transformation to the weekly sales can help make stable variance across the range of predictions, also using a robust regression technique can reduce influential points influence on the regression estimates.

References

1. Walmart: The world's largest retailer | Produce News [Internet]. [cited 2024 May 5]. Available from: https://theproducenews.com/headlines/walmart-worlds-largest-retailer
2. Wang CH. Considering economic indicators and dynamic channel interactions to conduct sales forecasting for retail sectors. Comput Ind Eng. 2022 Mar 1;165:107965.
3. Forecasting Success: The Impact of Weather on Retail Sales – Weather Source [Internet]. [cited 2024 May 5]. Available from: https://weathersource.com/blog/forecasting-success-the-impact-of-weather-on-retail-sales/
4. Yi S. Walmart Sales Prediction Based on Machine Learning. Highlights in Science, Engineering and Technology. 2023;47:87–94.
5. Vyas R, As R. Seasonal Sales Prediction and Visualization for Walmart Retail Chain Using Time Series and Regression Analysis: A Comparative Study. In: 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN). IEEE; 2022. p. 1–6.
6. Yu C. Walmart Sales Forecasting using Different Models. Highlights in Science, Engineering and Technology. 2024;92:302–7.
7. Latha SB, Dastagiraiah C, Kiran A, Asif S, Elangovan D, Reddy PCS. An Adaptive Machine Learning model for Walmart sales prediction. In: 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT). IEEE; 2023. p. 988–92.
8. Walmart Sales [Internet]. [cited 2024 May 5]. Available from: https://www.kaggle.com/datasets/mikhail1681/walmart-sales

9. Yuan M, Pavlidis Y, Jain M, Caster K. Walmart online grocery personalization: Behavioral insights and basket recommendations. In: International Conference on Conceptual Modeling. Springer; 2016. p. 49–64.

10. Bustillo M, Fowler GA. Wal-Mart sees stores as online edge. The Wall Street Journal. 2009;(B1).