

Walmart Sales Data Analysis Report

1. Introduction

This report presents a comprehensive data analysis of Walmart sales data. The objective is to understand sales patterns, identify key factors influencing sales, and provide actionable insights.

2. Data Exploration and Understanding

The dataset contains 6,435 entries and 8 columns, including `Store`, `Date`, `Weekly_Sales`, `Holiday_Flag`, `Temperature`, `Fuel_Price`, `CPI`, and `Unemployment`. There are no missing values in the dataset.

- `Store` : Store number (1 to 45)
- `Date` : Week of sales
- `Weekly_Sales` : Sales for the given week
- `Holiday_Flag` : 1 if the week is a holiday week, 0 otherwise
- `Temperature` : Average temperature in the region
- `Fuel_Price` : Cost of fuel in the region
- `CPI` : Consumer Price Index
- `Unemployment` : Unemployment rate

3. Data Cleaning and Transformation

During the data cleaning and transformation phase, the following steps were performed:

- The `Date` column was converted to datetime objects.

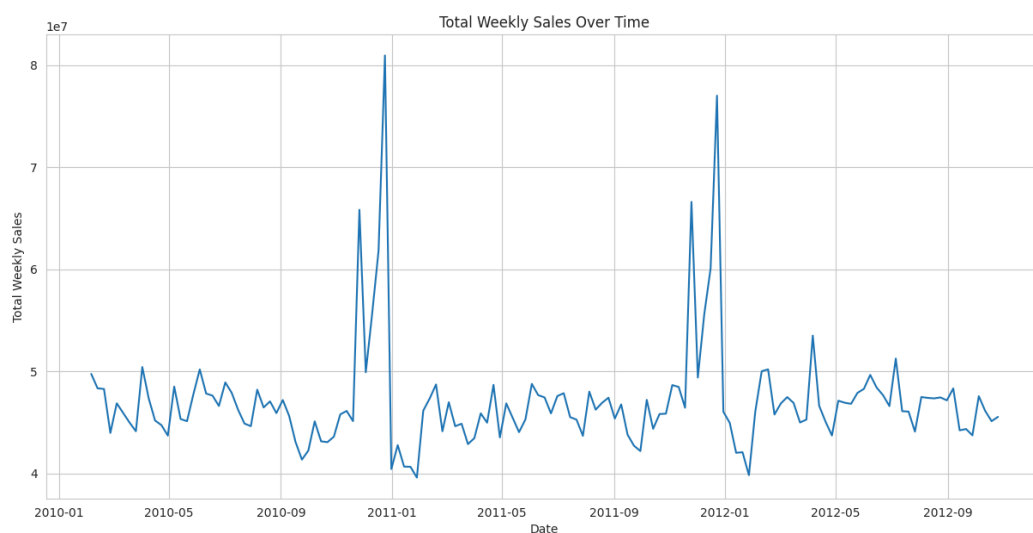
- New features were extracted from the `Date` column: `Year` , `Month` , `Week` , and `Quarter` .
- Outliers in `weekly_Sales` were identified using the IQR method. Approximately 34 outliers were found.
- Categorical variables `Holiday` (from `Holiday_Flag`) and `Store_Size` (based on `weekly_Sales`) were created.
- `Temp_Category` was created from `Temperature` .

4. Exploratory Data Analysis and Visualizations

Several visualizations were generated to understand the trends and relationships within the data.

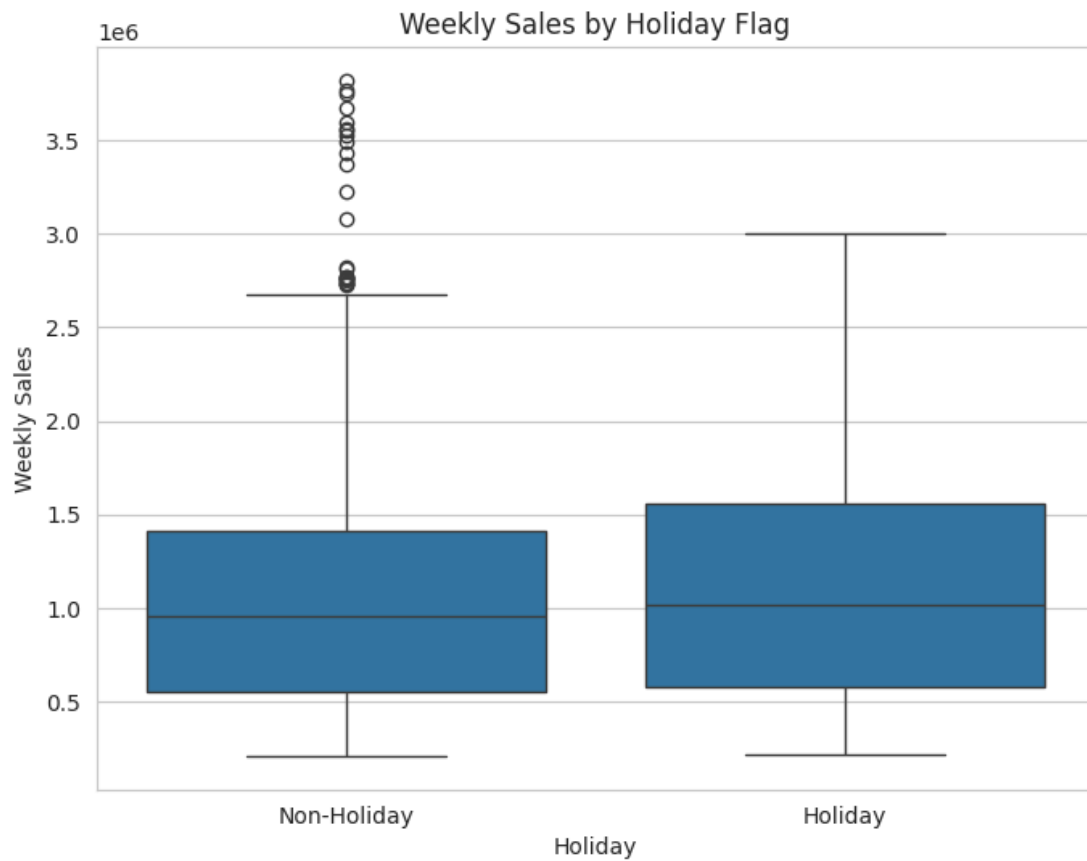
4.1 Total Weekly Sales Over Time

This plot shows the overall trend of weekly sales across all stores over the given period.



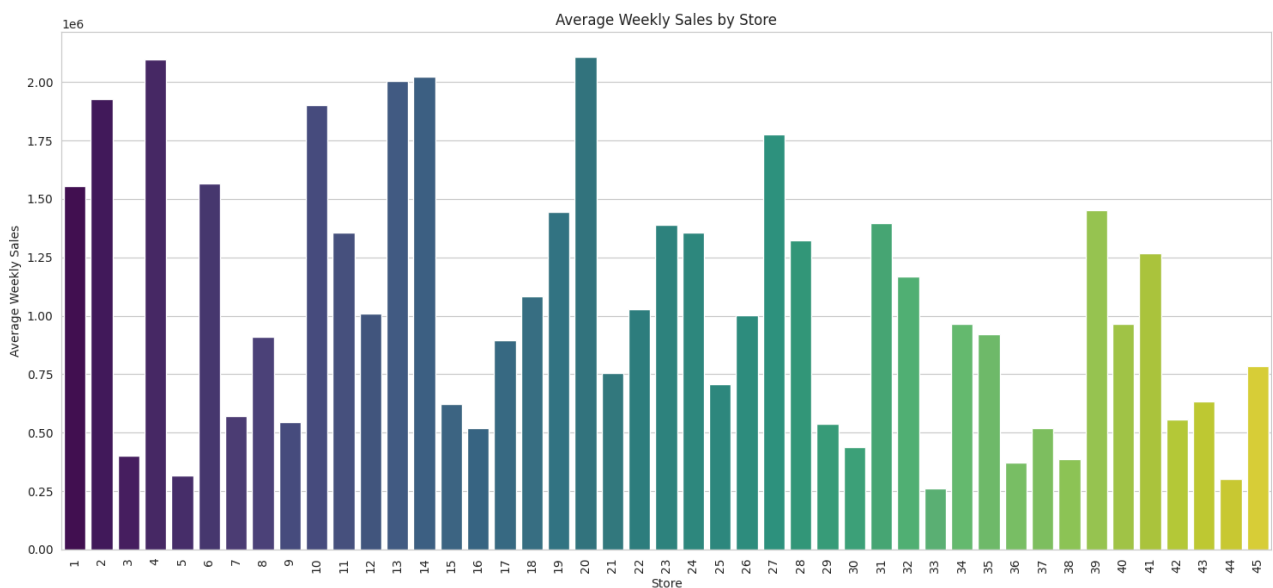
4.2 Weekly Sales by Holiday Flag

This box plot compares weekly sales during holiday weeks versus non-holiday weeks.



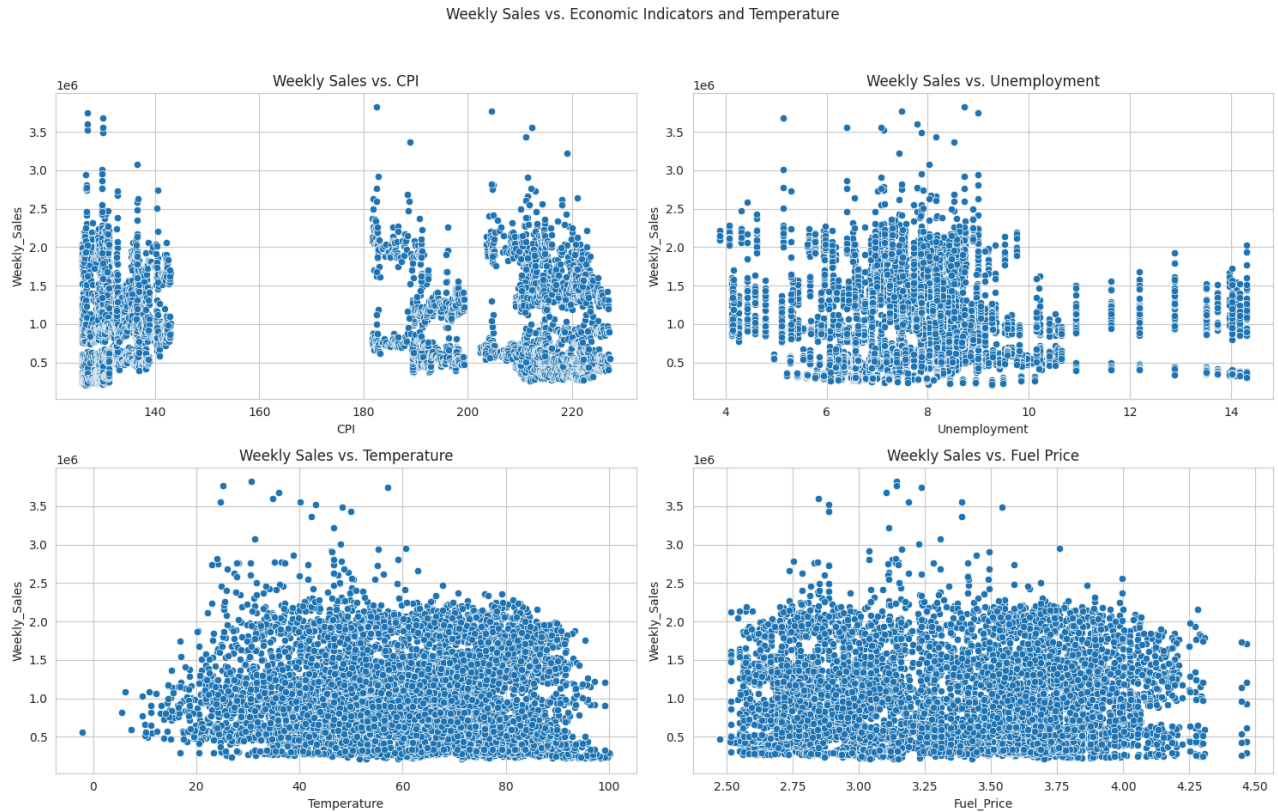
4.3 Average Weekly Sales by Store

This bar plot displays the average weekly sales for each store, sorted in descending order.



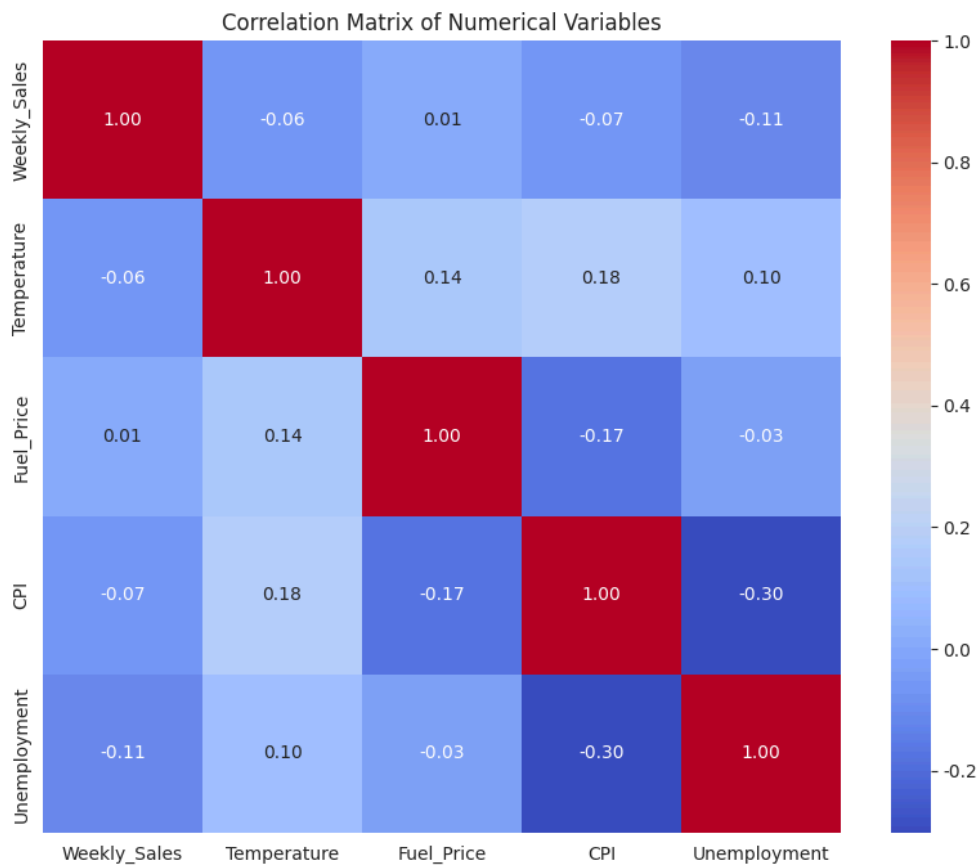
4.4 Weekly Sales vs. Economic Indicators and Temperature

These scatter plots illustrate the relationship between weekly sales and key economic indicators (CPI, Unemployment, Fuel Price) and Temperature.



4.5 Correlation Matrix of Numerical Variables

This heatmap shows the correlation coefficients between `weekly_sales` and other numerical variables.



5. Statistical Analysis and Insights

5.1 T-test for Weekly Sales (Holiday vs. Non-Holiday)

A t-test was conducted to compare weekly sales during holiday and non-holiday weeks.

- **Mean Weekly Sales (Holiday):** \$1,122,887.89
- **Mean Weekly Sales (Non-Holiday):** \$1,041,256.38
- **T-statistic:** 2.68
- **P-value:** 0.008

Conclusion: There is a statistically significant difference in weekly sales between holiday and non-holiday weeks, with holiday weeks generally showing higher sales.

5.2 OLS Regression Results (Weekly Sales vs. Economic Indicators)

An Ordinary Least Squares (OLS) regression model was built to understand the impact of CPI, Unemployment, Temperature, and Fuel Price on Weekly Sales.

Regression Summary:

| OLS Regression Results | | | | | |
|---|------------------|---------------------|-----------|-------|-----------|
| ===== | | | | | |
| Dep. Variable: | Weekly_Sales | R-squared: | 0.024 | | |
| Model: | OLS | Adj. R-squared: | 0.024 | | |
| Method: | Least Squares | F-statistic: | 40.09 | | |
| Date: | Wed, 23 Jul 2025 | Prob (F-statistic): | 3.21e-33 | | |
| Time: | 00:49:26 | Log-Likelihood: | -94273. | | |
| No. Observations: | 6435 | AIC: | 1.886e+05 | | |
| Df Residuals: | 6430 | BIC: | 1.886e+05 | | |
| Df Model: | 4 | | | | |
| Covariance Type: | nonrobust | | | | |
| ===== | | | | | |
| | coef | std err | t | P> t | [0.025 |
| 0.975] | | | | | |
| ----- | | | | | |
| - | | | | | |
| const | 1.744e+06 | 7.96e+04 | 21.918 | 0.000 | 1.59e+06 |
| 1.9e+06 | | | | | |
| CPI | -1585.8180 | 195.164 | -8.126 | 0.000 | -1968.404 |
| -1203.232 | | | | | |
| Unemployment | -4.121e+04 | 3972.667 | -10.375 | 0.000 | -4.9e+04 |
| -3.34e+04 | | | | | |
| Temperature | -885.6699 | 396.195 | -2.235 | 0.025 | -1662.345 |
| -108.995 | | | | | |
| Fuel_Price | -1.225e+04 | 1.58e+04 | -0.778 | 0.437 | -4.31e+04 |
| 1.86e+04 | | | | | |
| ===== | | | | | |
| Omnibus: | 370.623 | Durbin-Watson: | 0.113 | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 437.779 | | |
| Skew: | 0.639 | Prob(JB): | 8.66e-96 | | |
| Kurtosis: | 3.010 | Cond. No. | 2.16e+03 | | |
| ===== | | | | | |
| Notes: | | | | | |
| [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. | | | | | |
| [2] The condition number is large, 2.16e+03. This might indicate that there are strong multicollinearity or other numerical problems. | | | | | |

Insights from Regression Analysis:

- CPI:** A negative coefficient (-1585.82) suggests that as CPI increases, Weekly_Sales tend to decrease, holding other factors constant. This might indicate that higher prices lead to reduced purchasing power and thus lower sales.

- **Unemployment:** A significant negative coefficient (-41,210) indicates that as Unemployment increases, Weekly_Sales tend to decrease, which is intuitive as higher unemployment generally means less disposable income and reduced consumer spending.
- **Temperature:** The coefficient for Temperature (-885.67) is negative and statistically significant, suggesting a minor inverse linear relationship with Weekly_Sales. Further analysis might be needed to check for non-linear relationships or optimal temperature ranges.
- **Fuel_Price:** The coefficient for Fuel_Price (-12,250) is negative but not statistically significant (p-value > 0.05), suggesting a weak or no linear relationship with Weekly_Sales in this model.
- **R-squared:** The R-squared value of 0.024 indicates that only 2.4% of the variance in Weekly_Sales can be explained by these independent variables. This suggests that other factors not included in this model have a much larger impact on weekly sales.

5.3 Top 5 Stores by Average Weekly Sales

- **Store 20:** \$2,107,676.87
- **Store 4:** \$2,094,712.96
- **Store 14:** \$2,020,978.40
- **Store 13:** \$2,003,620.31
- **Store 2:** \$1,925,751.34

These stores consistently show the highest average weekly sales.

5.4 Average Weekly Sales by Quarter

- **Quarter 1:** \$1,006,136.18
- **Quarter 2:** \$1,040,806.41
- **Quarter 3:** \$1,023,251.31
- **Quarter 4:** \$1,128,773.58

Sales are highest in Quarter 4, likely due to holiday shopping.

5.5 Average Weekly Sales by Store Size

- **Large:** \$3,021,672.28
- **Medium:** \$1,793,443.41
- **Small:** \$780,983.88

Larger stores, as categorized by their weekly sales volume, tend to have significantly higher average weekly sales.

6. Conclusion and Recommendations

- **Holiday Impact:** Holiday weeks significantly boost sales. Walmart should strategize promotions and inventory management around these periods.
- **Economic Indicators:** Unemployment and CPI have a statistically significant negative impact on sales. Monitoring these indicators can help forecast sales trends.
- **Store Performance:** Identify and analyze the strategies of top-performing stores (e.g., Store 20, 4, 14, 13, 2) to replicate their success across other locations.
- **Seasonal Trends:** Q4 shows the highest sales, reinforcing the importance of holiday season planning.
- **Store Size:** Larger stores contribute more to overall sales. Further investigation into the characteristics of these stores could provide insights for optimizing smaller or medium-sized stores.

Further Analysis: While the current model provides some insights, the low R-squared value suggests that other factors, such as local competition, marketing efforts, product availability, or specific events, might have a more substantial impact on weekly sales. Future analysis could incorporate these variables for a more robust model.