
# Santa Clara University

## Analysis on Walmart Sales

Anqi Wei, Prakatheeswari Ravi, Sindhu Raghavendra Moktali, Uma Rajiv Kanetkar

Santa Clara University

Marketing Analytics, MKTG 2505

Dr. Sujata Ramnarayan

March 12, 2023

**Table of Contents**

## 1.Abstract

This project uses advanced statistics to analyze Walmart's sales data for strategic insights. It identifies trends and patterns, investigating factors like holidays and temperature to suggest ways for Walmart to improve sales. Using generalized linear models and clustering analysis, it develops accurate predictions and identifies key factors for sales performance.

## 2.Introduction To Company

Walmart is a global retailer with over 11,000 stores in 27 countries, employing over 2.3 million people and serving over 265 million customers per week. Its business model focuses on providing low prices and a wide variety of products, including groceries, apparel, electronics, and home goods. Walmart has also expanded into e-commerce, with rapidly growing online sales.

## 3.Dataset Description

This dataset includes weekly sales data for various Walmart stores over a period of time. The data includes the store number, the date of the sale, the weekly sales amount, whether the sale occurred during a holiday period or not, the temperature, fuel price, CPI (Consumer Price Index), unemployment rate, and the holiday name. The dataset can be used to analyze sales trends and patterns across different stores, as well as to identify factors that impact sales performance.

| | Store | Date | Weekly_Sales | Holiday_Flag | Temperature | Fuel_Price | CPI | Unemployment |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 05-02-2010 | 1643691 | 0 | 42.31 | 2.572 | 211.0964 | 8.106 |
| 2 | 1 | 12-02-2010 | 1641957 | 1 | 38.51 | 2.548 | 211.2422 | 8.106 |
| 3 | 1 | 19-02-2010 | 1611968 | 0 | 39.93 | 2.514 | 211.2891 | 8.106 |
| 4 | 1 | 26-02-2010 | 1409728 | 0 | 46.63 | 2.561 | 211.3196 | 8.106 |
| 5 | 1 | 05-03-2010 | 1554807 | 0 | 46.50 | 2.625 | 211.3501 | 8.106 |

Table 1: Dataset

## 4.Situation Analysis (EDA) and Problem Identification

From the perspective of Walmart,we want to understand the sales patterns and trends to identify areas for potential improvement or optimization.

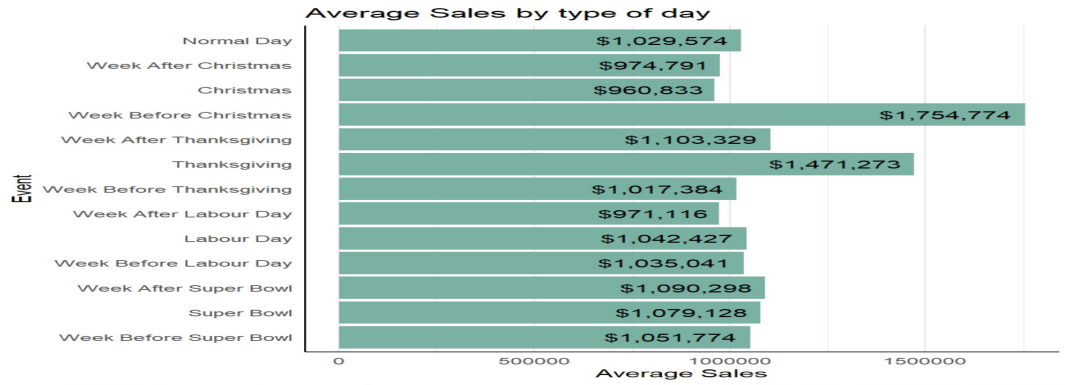



**Figure 1:** Yearly Total Sales

**4.1 Insight:** The Average Weekly Sales is highest during the week before Christmas with average total sales of $1,754,774 followed by Thanksgiving with the average total sales of $1,471,273.

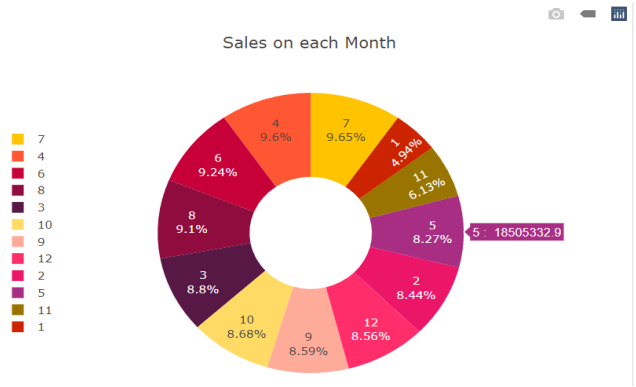



**Figure 2:** Total Monthly Sales

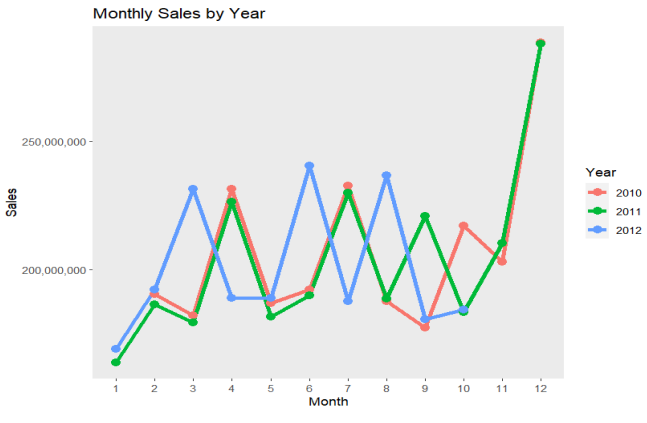



**Figure 3:** Monthly Sales Across Years

**4.2 Insight:** Monthly sales vary across months with the highest sales in the month of July followed by April. The monthly sales are different for each year as well. In the year 2012, the sales trend seems to be drastically different than other two years.

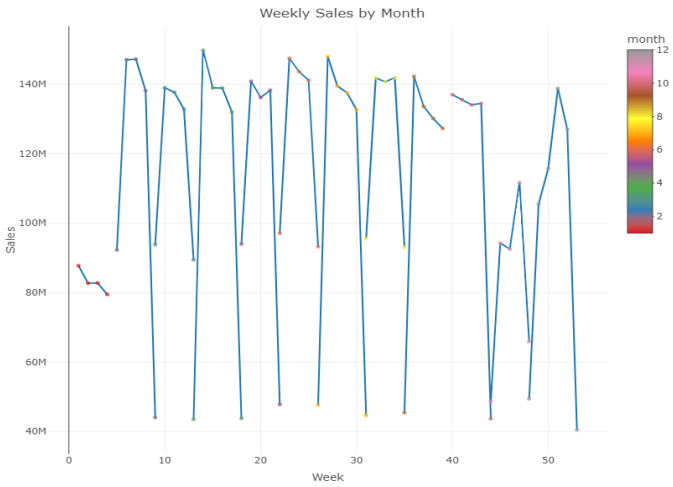



**Figure 4: Weekly sales By Month**

**4.3 Insight:** The first week and last week of every month has less sales than the mid weeks.

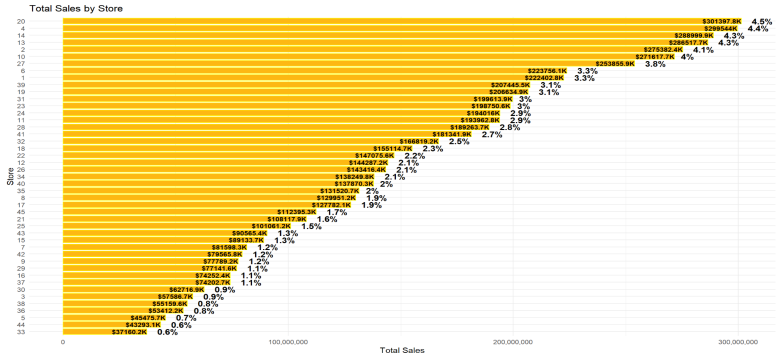
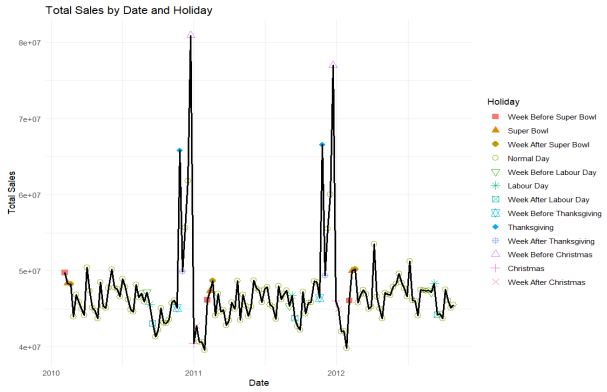



**Figure 5:** Weekly sales Across Different Stores

**4.4 Insight:** The Weekly sales are different across different stores and which may be due to various reasons like location, popularity, store promotions, customer experience etc.



**4.5 Insight:** There were significant increases in sales during the weeks leading up to holidays. This phenomenon is likely due to the fact that shoppers tend to begin their purchases a few weeks ahead of the holiday, in anticipation of the event.

## 5. Approaches and Techniques

We utilized Multi-Linear Predictive Regression Analysis in our study to reduce the possibility of omitted variable bias and to utilize all the variables present in the dataset. By incorporating multiple significant variables such as Temperature, Holidays, CPI, Unemployment, and Fuel Prices, we aimed to analyze their causal impact on the average weekly sales of Walmart. We also included interaction terms and used clustering to identify patterns or similarities. This could potentially help in identifying factors that are driving sales in certain groups of stores or in certain years. This approach allowed us to gain insights, potentially enabling more informed decision-making processes.

## 6. Models and Insights

### 6.1 Method 1: Generalized Linear Model

The aim of the study is to develop a model that accurately predicts weekly sales based on various predictor variables. Several linear regression models were developed with varying combinations of predictor variables. The first three models included single predictor variables, namely CPI, Holiday_Flag, and Temperature. The fourth model incorporated multiple predictor variables, namely Temperature, Fuel_Price, CPI, and Unemployment. The models' R-squared values were relatively low, indicating a poor fit to the data.

The subsequent step involved adding the Store variable as a predictor variable to the regression models which created two new models, with the latter incorporating interaction terms between

some of the variables which greatly improved the models' predictive power. This is evidenced by the impressive R-squared value of 0.95 and with a corresponding MAPE score of 7.58%.

Based on the summary of the model,  during Thanksgiving and Holiday Week before Christmas, sales skyrocketed with the highest coefficient of 624707 and 1234736 respectively . Whereas other festivals didn't have much effect on the weekly sales. Unemployment and CPI interaction term has a negative correlation on the weekly sales. Temperature has a positive correlation with the weekly sales. Few stores were negatively correlated with the weekly sales variable.

### 6.2 Method 2: K-Means Clustering

| | cluster | total_weekly_sales | avg_weekly_sales | avg_unemployment | avg_CPI | avg_fuel | num_stores |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 16093248 | 1788138.7 | 7.398963 | 147.2947 | 3.454542 | 9 |
| 2 | 5 | 13343122 | 833945.1 | 7.876862 | 156.8441 | 3.362208 | 16 |
| 3 | 2 | 9248697 | 1541449.4 | 7.418508 | 216.4939 | 3.219699 | 6 |
| 4 | 4 | 4893832 | 543759.1 | 7.398301 | 215.9454 | 3.217977 | 9 |
| 5 | 1 | 3534521 | 706904.2 | 11.249113 | 128.6797 | 3.594221 | 5 |

On performing K-means clustering on stores, it was found that the unemployment rate has a significant impact on the revenue generation through sales. Lower unemployment leads to higher sales on average and visa-versa. CPI and fuel price don't play a significant role in affecting sales.

## 7. Conclusion

Based on the analysis, we can conclude that there are several factors that impact Walmart's weekly sales. Holidays, especially Thanksgiving and the week before Christmas, have a significant positive impact on sales. However, other festivals did not have a significant effect. Fuel prices and unemployment have a weak negative correlation with Sales. There are some outliers in the non-holiday weeks, indicating some weeks with exceptionally high or low sales. The analysis also identified some stores that are negatively correlated with weekly sales.

## 8.Recommendations

**Focus on holiday sales:** The analysis showed that sales significantly increased during weeks leading up to Thanksgiving and Christmas. This suggests that Walmart should focus on these holiday periods and create targeted marketing campaigns and promotions to encourage shoppers to start their holiday shopping early.

**Pay attention to store locations:** Walmart should pay attention to tailor their marketing and product offerings to each store's location to better meet the needs of the local customers.

**Improve low-performing stores:** Walmart could explore ways to improve the performance of stores with lower sales, such as optimizing the product mix or store layout. Additionally, some strategies from high-performing stores can be employed.

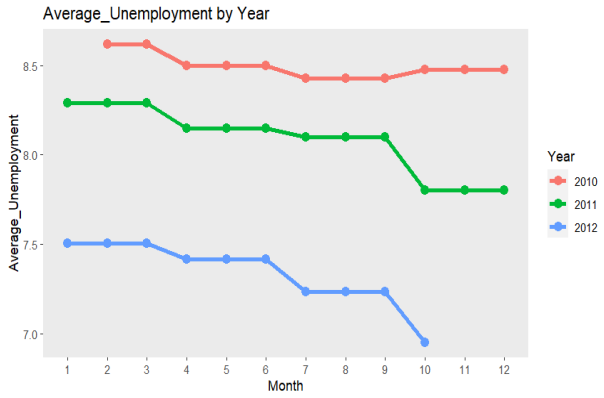
**Monitor unemployment and CPI:** The interaction term between unemployment and CPI had a negative correlation with weekly sales. Walmart should monitor these economic factors and adjust their pricing and promotions accordingly.

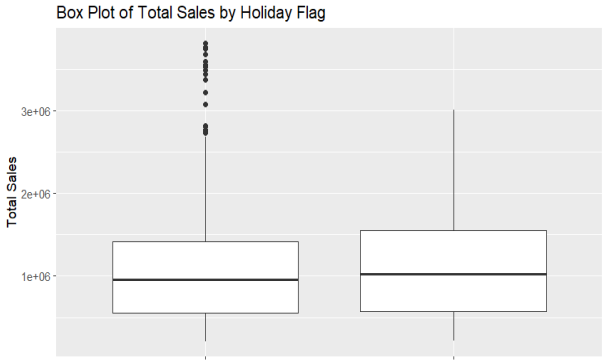
**Improve inventory management:** The analysis showed that the first and last week of every month had less sales than the mid weeks. This suggests that Walmart should improve their inventory management to ensure that they have enough stock throughout the month and avoid stock outs during peak sales periods.

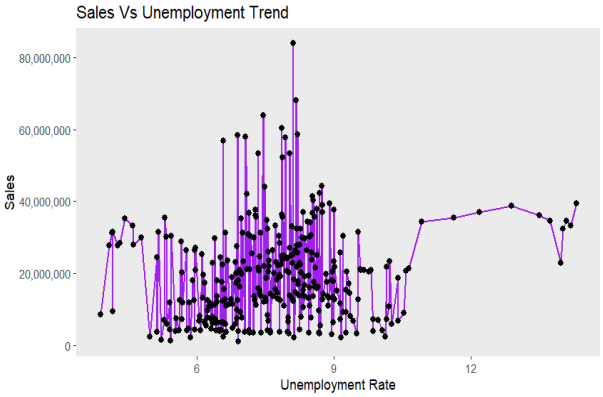
**9.References**:

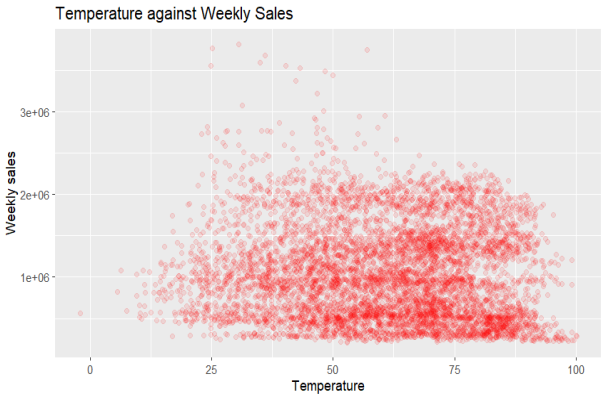

1. Bari, A., Chaouchi, M., & Jung, T. (n.d.). How to utilize linear regressions in predictive analytics. https://www.dummies.com/programming/big-data/data-science/how-to-utilize-linear-regressions-in-predictive-analytics/
2. Baum, D. (2011). How higher gas prices affect consumer behavior. https : / / www.sciencedaily.com/releases/2011/05/110512132426.htm
3. Ellis, L. (2019). Simple eda in r with inspectdf. https://www.r-bloggers.com/2019/05/part-2-simple-eda-in-r-with-inspectdf/
4. Wikipedia, t. f. e. (n.d.). Walmart. https : / / en . wikipedia . org / w / index . php ? title = Walmart&oldid=1001006854

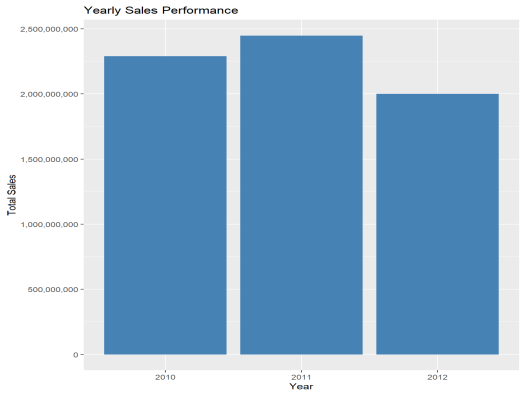
# Appendix1

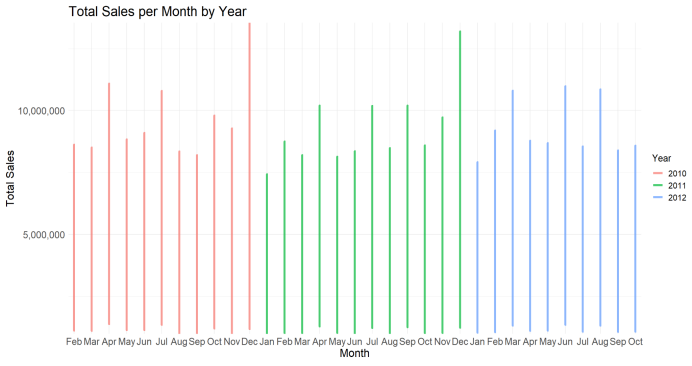

Average_Unemployment by Year
Average_Unemployment
Month
Year
2010
2011
2012

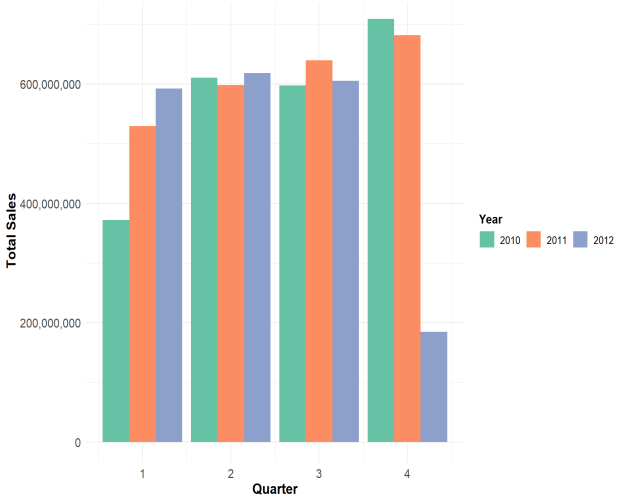

Box Plot of Total Sales by Holiday Flag
Total Sales

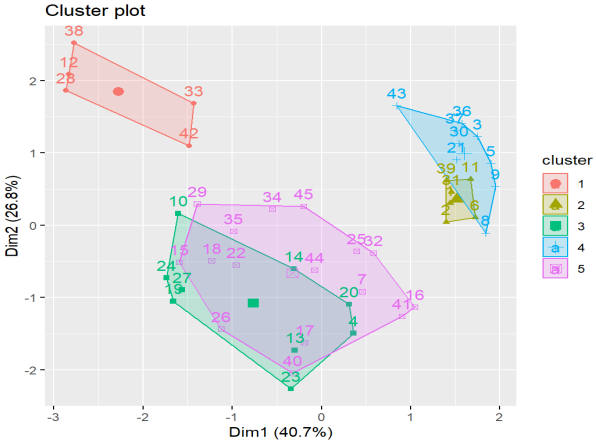

Sales Vs Unemployment Trend
Sales
Unemployment Rate

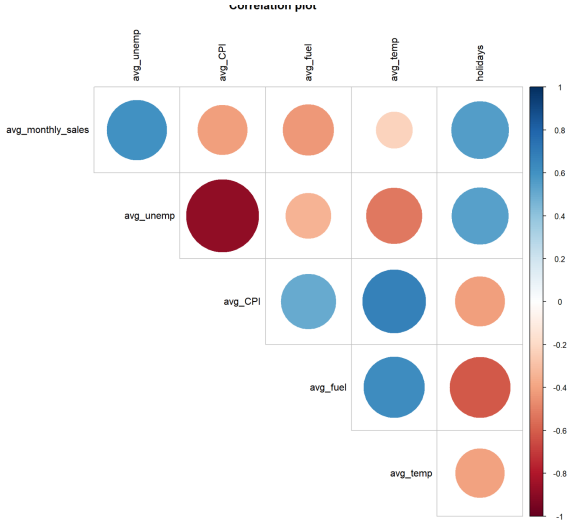

Temperature against Weekly Sales
Weekly sales
Temperature


Yearly Sales Performance
Total Sales
Year


Total Sales per Month by Year
Total Sales
Month
Year
2010
2011
2012


Total Sales
Quarter
Year
2010
2011
2012


Cluster plot
Dim2 (26.8%)
Dim1 (40.7%)
cluster
1
2
3
4
5


Correlation plot
avg_unemp
avg_CPI
avg_fuel
avg_temp
holidays
avg_monthly_sales
avg_unemp
avg_CPI
avg_fuel
avg_temp
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1