

The Influence of Weather and Economic Trends on Walmart Sales

Aarush Kapoor
School of Computing and Data Science

Abstract—This study investigates the multifaceted influences of weather and economic trends on Walmart sales, analyzing how environmental and economic factors impact consumer behavior and retail performance. By integrating sales data from Kaggle and weather data from Visual Crossing, the research explores the relationships between temperature, precipitation, extreme weather conditions, holidays, and unemployment rates with weekly sales. Using data visualization techniques and multiple linear regression, the study reveals nuanced insights: sales fluctuate with temperature extremes, decline significantly during severe weather events, spike during holidays like Thanksgiving and Christmas, and correlate negatively with high unemployment rates. The predictive model explains approximately 49% of sales variance, highlighting the complex interplay of external factors in retail dynamics. These findings provide critical strategic insights for retailers, demonstrating the importance of comprehensive, data-driven approaches to understanding and anticipating consumer behavior.

Keywords—Retail Sales, Weather Impact, Economic Trends, Consumer Behavior, Walmart Analytics

I. INTRODUCTION

This project examines the "Influence of Weather and Economic Trends on Walmart Sales," focusing on how factors like temperature, precipitation, extreme weather, holidays, and unemployment impact consumer behavior. Understanding these relationships is vital for large retailers like Walmart to optimize operations, manage inventory, and enhance marketing strategies.

Temperature and precipitation can affect foot traffic and online sales, while extreme weather may cause sudden shifts in demand. Holidays drive significant sales spikes influenced by weather and economic conditions, and unemployment rates can shift spending toward budget-friendly products.

Recent economic research, such as Roth Tran's (2022) work [1], highlights the significant and persistent effects of weather on sales, with some extreme weather events potentially reducing revenue by up to 40%, underscoring the importance of understanding these complex interactions. This report uses retail data from Kaggle [2] and weather data from Visual Crossing [3] to explore how these factors interact and impact Walmart's sales, providing actionable insights for strategic planning.

II. DATASETS

A. Source of dataset

The first dataset, sourced from Kaggle [2], contains historical sales data for 45 Walmart stores in the U.S., covering the period from February 5, 2010, to November 1, 2012. This dataset was created as part of a machine learning competition aimed at developing predictive algorithms for sales and demand forecasting, integrating economic

conditions and special events. It includes fields such as Store, Date, Weekly Sales, Temperature, and Unemployment, among others. The dataset is credible and widely used in data science projects for its depth and relevancy to retail analysis. For this study, data cleaning and feature engineering were applied to focus on the most impactful fields related to consumer behavior and sales trends.

The second dataset was obtained from Visual Crossing [3], a trusted provider of comprehensive weather data. This dataset includes detailed weather attributes, such as dates, temperatures, precipitation data, windspeed, and visibility, which are crucial for analyzing the effect of weather on consumer behavior. Visual Crossing's [3] data is derived from a network of weather stations, ensuring reliability and accuracy for time-series analyses. The selected weather fields were preprocessed to align with the sales data, allowing for correlations between weather conditions and sales performance. This integration facilitates a deeper understanding of how external factors influence retail sales, aiding in more strategic decision-making.

B. Character of the datasets

The Walmart sales dataset includes fields that are key to understanding sales trends, such as store number, date of sales, weekly sales, holiday indicator, and economic metrics like temperature, fuel price, CPI, and unemployment rate. The primary fields used for the analysis were the store number, weekly sales, temperature, date, and unemployment rate. The dataset spans from February 5, 2010, to November 1, 2012, covering 45 Walmart stores in different regions. Data cleaning was performed to remove rows with missing or corrupted information to maintain data quality, as well as to focus on only the specific columns that were needed to answer the topic questions. The temperature data was retained in Fahrenheit as provided, without conversion. To combine this dataset with weather data, matching by store and date was used.

Field Name	Description	Unit/Format
Store	Identifier for the Walmart store	Integer
Date	The date of the weekly sales data	Date (DD-MM-YYYY)
Weekly_Sales	Total sales for the store in a week	Float (USD)
Holiday_Flag	Indicator of whether the week was a holiday	1 (Holiday), 0 (Non-holiday)
Tempurature	Temperature on the day of sale	Float (Fahrenheit)
Fuel_Price	Cost of fuel in the region	Float (USD per gallon)

CPI	Consumer Price Index (CPI)	Float (Index)
Unemployment	Unemployment rate in the region	Float (%)

The Visual Crossing [3] weather dataset contains a variety of weather metrics, including datetime (timestamp), temperature (tempmax, tempmin, temp), perceived temperature (feelslikemax, feelslikemin, feelslike), humidity, precipitation information (precip, precipprob, precipcover, preciptype), wind data (windspeed, windgust, winddir), visibility, pressure, and other environmental factors like solar radiation and UV index. For the analysis, the key fields were datetime, tempmin, temp, precip, windspeed, and visibility, as these factors were most relevant to understanding weather impacts on consumer behavior. The original data was in Fahrenheit for temperature, and no conversion to Celsius was applied. Data cleaning ensured that rows with missing values were excluded, and feature engineering focused on developing new variables, such as the temperature range and categorizing weather types, to enhance the analysis with the Walmart dataset.

Field Name	Description	Unit/Format
Datetime	Timestamp for the recorded weather data	Date (MM-DD-YYYY)
Tempmax	Maximum temperature recorded	Float (Fahrenheit)
Tempmin	Minimum temperature recorded	Float (Fahrenheit)
Temp	Average temperature recorded	Float (Fahrenheit)
Precip	Total precipitation recorded	Float (inches)
Windspeed	Average wind speed	Float (mph)
Visibility	Visibility distance in miles	Float (miles)

III. METHODOLOGY

A. Data Visualization and Initial Analysis

To address the research question, initial data exploration and visualization were performed using Python in Google Colab. The libraries used include pandas, matplotlib, numpy, and scipy. The main aim of this phase was to create visual representations of the data, which provided insights into potential relationships between weather patterns and Walmart sales.

Tools and Techniques:

- Google Colab: An interactive cloud-based coding platform that supports Python programming and provides a convenient environment for data analysis.
- pandas: Used for data manipulation and cleaning, allowing for efficient data handling and preprocessing.
- matplotlib: Utilized for generating scatter plots, box plots, and bar graphs to visually assess the distribution and patterns within the data.
- numpy: Employed for numerical operations, such as calculating means and standard deviations for data analysis.
- scipy: Applied for statistical calculations and to conduct hypothesis testing when needed.

Advantages:

- Interactive Environment: Google Colab allows real-time collaboration and ease of access.
- Comprehensive Python Libraries: Libraries like pandas, matplotlib, and scipy are well-suited for data manipulation, visualization, and statistical analysis.
- Customizability: The flexibility of Python enables the creation of detailed and specific data visualizations.

Disadvantages:

- Dependency on Internet: Google Colab requires a stable internet connection for uninterrupted access.
- Performance: For very large datasets, processing in Colab may be slower compared to local setups with higher computing power.

Adjustments for Improved Results:

- Feature Scaling: Standardization of variables was performed using StandardScaler to ensure uniformity and improve model performance. In this instance, sales on the y-axis of graphs are based on a normalized scale such as 0 to 1 rather than from 0 to the highest recorded sale.

These tools were chosen due to their robustness and integration in Python for data analysis and visualization. They allow for a clear depiction of relationships between variables, which was essential for the exploratory phase of this study.

B. Regression Analysis for Sales Prediction

To quantify the relationship between extreme weather conditions and Walmart sales, a regression model was used. The model selected was a multiple linear regression model, which allows for prediction based on multiple variables. Multiple linear regression is a statistical method used to model the relationship between a dependent variable (e.g., Walmart sales) and one or more independent variables (e.g., temperature, precipitation, and wind speed). The main assumption of this model is that there is a linear relationship

between the independent variables and the dependent variable.

Assumptions:

- Linearity: The relationship between the dependent variable and the predictors is assumed to be linear.
- Independence: Observations are assumed to be independent of each other.
- Homoscedasticity: The variance of the error terms should be constant across all levels of the independent variables.
- Normality: The residuals (errors) of the model should be normally distributed.

Advantages:

- Simplicity and Interpretability: Multiple linear regression is straightforward to implement and easy to interpret, providing a clear understanding of how each independent variable impacts the dependent variable.
- Scalability: It can handle multiple predictors simultaneously, which is beneficial for analyzing the effect of multiple weather variables on sales.

Disadvantages:

- Limited to Linear Relationships: The model may not capture complex relationships between variables as effectively as non-linear models.
- Sensitive to Outliers: Outliers can significantly impact the model's performance and predictions.

Python Implementation:

The scikit-learn library was used for applying the regression model. The implementation process involved:

- Importing the LinearRegression class from sklearn.linear_model.
- Splitting the dataset into training and testing sets using train_test_split from sklearn.model_selection.
- Training the model using the .fit() method and evaluating its performance using metrics such as R^2 and Mean Absolute Error (MAE).

Adjustments for Improved Results:

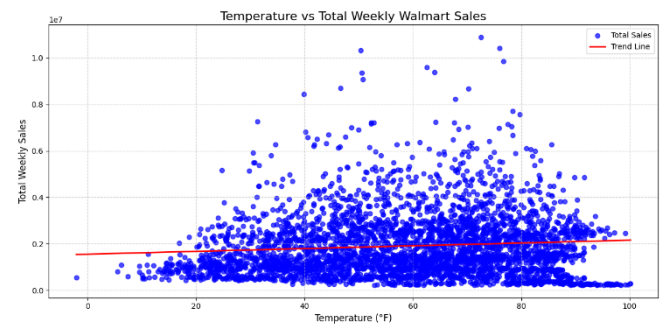
- Feature Scaling: Standardization of variables was performed using StandardScaler to ensure uniformity and improve model performance.

By employing this regression model, we were able to make informed predictions about how extreme weather conditions impact Walmart sales, contributing valuable insights into consumer behavior patterns under varying weather conditions.

IV. RESULTS

1. How does temperature affect Walmart Sales?

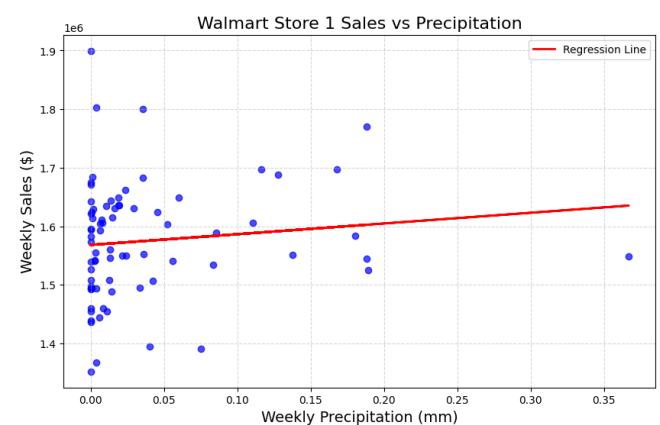
To answer this question, a scatterplot was made comparing temperature to total weekly Walmart sales:



From these results we can see that there is a very slight positive correlation, that being the higher the temperature the higher the sales. However, we can observe that sales start to drop when temperatures are either below 20 degrees or above 90 degrees Fahrenheit. A possible explanation for this is because the warmer the weather is, the more likely people are to be outside. People are more likely to stay inside in extreme temperatures, however, such as freezing cold or extremely hot conditions.

2. How does precipitation affect Walmart Sales?

To answer this question, another scatter plot was used to compare precipitation levels to Walmart sales. However, only temperature is shown in the Kaggle [2] dataset, so precipitation data was collected from Visual Crossing's [3] database, and matched by dates to the Walmart sales data. For this graph, only one Walmart store was looked at, that being the Walmart Supercenter in Colorado Springs.



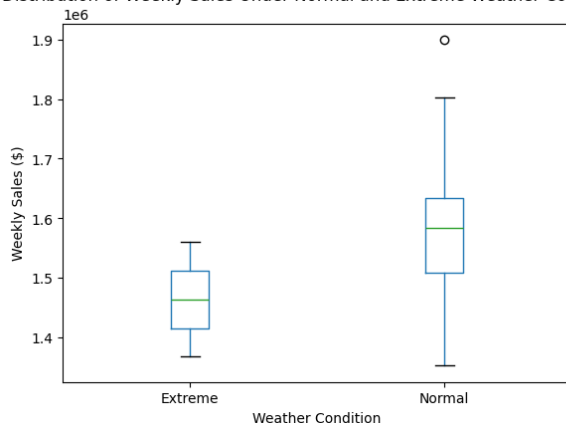
From these results we can see that there isn't much correlation between precipitation and Walmart sales. There are a lot more datapoints for low precipitation due to days without rain being more common than with, however the sales for any given week don't seem to be affected when it does rain. A possible explanation for this could be that rain is not a factor that determines whether people go shopping or not. People might go shopping for groceries and products when they need those groceries and products, independent of whether it is raining or not. Another contributing factor could be that in this day and age most people drive to stores or take transportation methods

that don't involve being outside such as walking or biking, which minimizes the affect rain will have on people going out to shop.

3. How do extreme weather conditions affect Walmart Sales?

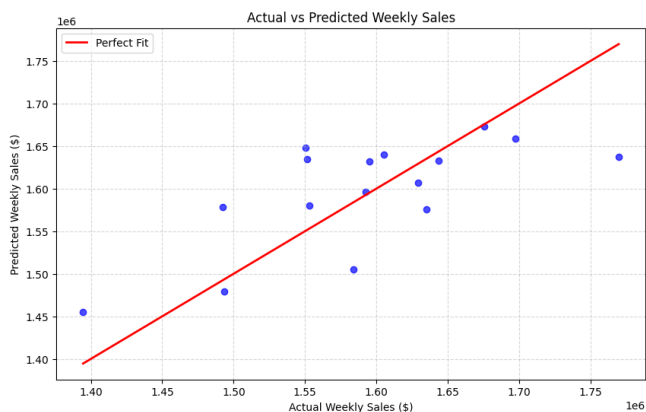
A box plot was used to answer this question, with one box plot used to represent sales during extreme weather conditions, and another to represent overall sales regardless of weather conditions. To determine what extreme weather conditions were, data was filtered to only include datapoints where either the precipitation was higher than 70%, the wind speed was greater than 50 mph, the visibility was less than 5 miles, or the minimum temperature was less than 10 degrees Fahrenheit.

Distribution of Weekly Sales Under Normal and Extreme Weather Conditions



As is shown by this visualization, Walmart's sales during extreme weather conditions are much lower than their sales during normal weather conditions. From this correlation, we can infer that extreme weather conditions do impact Walmart sales. Things like heavy precipitation and low visibility impact safety on roads and can be a contributing factor to less people driving unless necessary. Freezing temperatures and high wind speeds can account for days where there might be storms, which is another safety concern for those not staying inside. For these reasons, a possible explanation for this correlation is that safety risks are much higher during extreme weather conditions, causing more people to stay inside and less people to go shopping.

The following regression model was based on this data to predict Walmart's sales during regular and extreme weather conditions:



Mean Squared Error (MSE): 3785396113.69

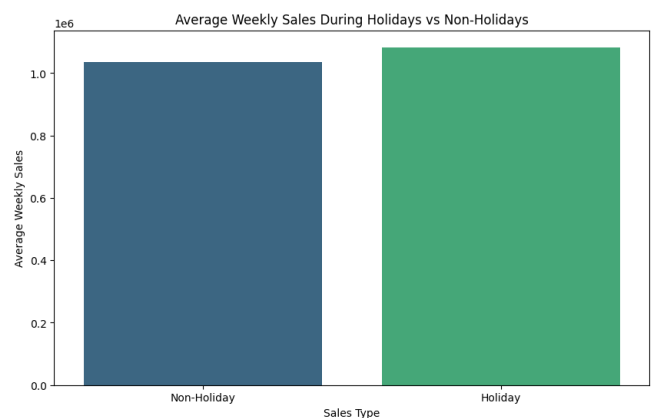
R-squared (R^2): 0.49

The regression model used is a Linear Regression model designed to predict Walmart's weekly sales based on weather-related features: precipitation, windspeed, visibility, and minimum temperature. After cleaning the dataset by removing rows with missing values, the data was split into training (80%) and testing (20%) subsets. The model was trained on the training data, learning the relationships between the features and the target variable, and evaluated on the test set using Mean Squared Error (MSE) and R-squared (R^2). These metrics measure the average error of predictions and how well the model explains variance in sales, respectively. A scatter plot of actual versus predicted sales showed a positive correlation, indicating the model captures some patterns in the data.

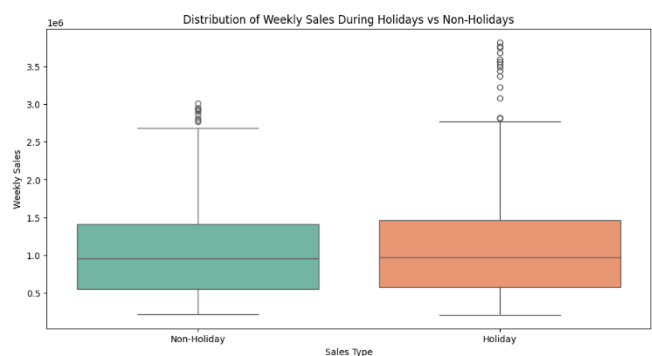
The visualization highlights the model's performance, with the red line representing a perfect prediction. While many data points align closely with this line, deviations suggest the model is not fully accurate, as evidenced by the R^2 value. This outcome indicates that the selected weather features explain a portion of the variance in weekly sales, but other unaccounted factors, such as promotions or holidays, likely influence sales patterns. Overall, the model provides a useful starting point for understanding the relationship between weather and Walmart's weekly sales.

4. What is the effect of holidays on Walmart Sales?

To answer this question, two graphs were made, those being a bar chart and a box plot. Dates flagged as holidays in the Walmart sales data were Superbowls, Labour Day, Thanksgiving, and Christmas.



From our bar chart, we can see that Holiday sales are larger than Non-Holiday sales, although not by that much.



Our box plot on the other hand give us more insight into our data. This also shows higher sales trends for Holidays as opposed to Non-Holidays, but it also shows us a significant amount of positive outliers during Holidays. This could mean that although Holiday sales are generally higher, there is a bigger chance of there being an unusually high week of sales during a Holiday as well.

To give us a better understanding, a third visualization was used to show total Walmart sales over time.



This visualization gives us a very clear understanding of our outliers. Every year there are two significant spikes in sales, those being around Thanksgiving and Christmas. The spike in sales during Thanksgiving is most likely due to Black Friday sales, while the spike in sales during Christmas/New Years time is most likely due to people buying gifts and Holiday items and decorations. The sales from these two events are likely to be what is influencing our bar chart and box plots the most in showing the difference between Holiday and Non-Holiday sales.

5. How does unemployment affect Walmart Sales?

This last question observes another economic trend that could affect Walmart's sales, in which data was graphed on a scatterplot to analyze. This scatterplot compares unemployment rates to total weekly Walmart sales:



The first thing to be noted about this visualization is the frequency of vertical lines made up of multiple data points. This is due to many stretches of days having the same unemployment rate. Aside from this, there seems to be a very slight negative correlation in the data, where the higher the unemployment, the lower the sales. The sales seem to drop around an unemployment rate of 10%, where positive outliers become nonexistent, and high weeks for sales are much fewer compared to average and lower weeks. This could be due to less money being made in our economy in general stemming from unemployment, leading to less money being spent in retail stores like Walmart.

V. DISCUSSION

A. Methodological Limitations

The current study revealed several important insights into factors affecting Walmart sales, but also exposed several methodological limitations that warrant careful consideration. The linear regression model used to predict sales based on weather conditions achieved only a moderate explanatory power, with an R-squared value of 0.49. This indicates that while weather parameters provide some predictive value, a significant portion of sales variance remains unexplained.

B. Key Limitations and Insights

Weather Impact Analysis

- The temperature and sales relationship showed a nuanced pattern, with sales declining at extreme temperatures (below 20°F or above 90°F). However, the analysis was limited to a single dataset, which may not be representative of all Walmart stores or geographic regions.
- Precipitation analysis was constrained by data availability, with the study focusing on a single Walmart Supercenter in Colorado Springs. This localized approach limits the generalizability of the findings.

Extreme Weather Conditions

- The study demonstrated a clear impact of extreme weather on sales, but the definition of "extreme" was relatively narrow. Future research could benefit from a more comprehensive and flexible definition of extreme weather conditions.

Holiday and Economic Factors

- While the holiday sales analysis revealed interesting patterns, particularly around Thanksgiving and Christmas, the study could explore more deeply the underlying mechanisms driving these sales spikes.
- The unemployment analysis suggested a slight negative correlation with sales, but the relationship could be more thoroughly investigated.

C. Suggestions for Future Research

Expanded Geographic and Temporal Scope

- Incorporate data from multiple Walmart stores across different regions
- Collect data over a more extended period to capture long-term trends
- Include a more diverse range of geographic locations to account for regional variations

Enhanced Predictive Modeling

- Integrate additional features beyond weather and economic indicators
- Explore non-linear relationships between variables
- Incorporate advanced techniques like ensemble methods or neural networks

Comprehensive Economic Analysis

- Detailed analysis of how unemployment impacts consumer behavior
- Integration of additional economic indicators (e.g., inflation, consumer confidence)
- Exploration of seasonal economic variations

Advanced Weather Impact Studies

- Use more granular weather data
- Create more sophisticated categorizations of weather conditions
- Investigate interaction effects between different weather parameters

VI. CONCLUSION

This research into Walmart sales reveals the complex interplay of external factors that shape retail performance. Temperature is shown to have impacts on shopping behavior, with sales declining during temperatures below 20 degrees or above 90 degrees Fahrenheit.

Notably, extreme weather events—characterized by heavy precipitation, high wind speeds, low visibility, or freezing temperatures—substantially reduce sales, likely due to increased safety concerns that keep customers at home, while precipitation surprisingly showed minimal impact on sales—suggesting that people shop regardless of light rain due to modern transportation methods.

Holidays emerged as critical sales periods, with Thanksgiving and Christmas producing dramatic sales spikes, likely due to the Black Friday phenomenon and Christmas gift-buying season, which particularly stand out as key drivers of exceptional sales weeks. These insights highlight the importance of strategic planning around seasonal peaks.

Economic indicators, particularly unemployment rates, showed a meaningful correlation with sales. As unemployment goes past 10%, a noticeable contraction in consumer spending becomes apparent, underscoring the direct link between economic health and retail performance.

The most significant finding is the multifaceted nature of sales dynamics. No single factor determines sales performance; instead, a complex interaction of weather, economic conditions, and seasonal trends shapes consumer behavior. Our predictive model, explaining approximately 49% of sales variance, provides a robust framework for understanding these intricate relationships.

For retailers like Walmart, these insights are invaluable. They offer a strategic roadmap for adapting to changing external conditions, optimizing inventory management, and developing more targeted marketing approaches. The research challenges simplistic views of sales performance, presenting it as a dynamic system influenced by a wide array of external factors.

In essence, this study provides a nuanced understanding of the forces driving retail sales, demonstrating the critical importance of comprehensive, data-driven approaches to business strategy. While rain might not deter shoppers, severe weather conditions can dramatically reshape retail landscapes.

ACKNOWLEDGMENT

The author would like to express their sincere gratitude to several individuals and organizations that made this research possible:

First and foremost, I thank Wentworth Institute of Technology and the School of Computing and Data Science for providing the necessary resources and academic support for this research project.

Special appreciation is also given to Professor Weijie Pang for their invaluable guidance, insightful feedback, and continuous support throughout the research process.

I am grateful to Visual Crossing [3] for providing the precipitation data that was crucial to our analysis. Their comprehensive database was instrumental in supplementing the Walmart sales dataset.

My thanks also go to the Kaggle [2] community and the original dataset contributors, whose data made this research possible. The availability of open-source data continues to be vital for academic research and data science investigations.

This research was conducted as part of COMP3125 Data Science Fundamentals at Wentworth Institute of Technology. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author's and do not necessarily reflect the views of the supporting organizations.

REFERENCES

- [1] B. R. Tran, The Impact of Weather on Retail Sales, www.frbsf.org, Aug. 22, 2022. <https://www.frbsf.org/research-and-insights/publications/economic-letter/2022/08/impact-of-weather-on-retail-sales/>
- [2] M. Yasser H, Walmart Dataset, [www.kaggle.com](https://www.kaggle.com/datasets/yasserh/walmart-dataset/data), 2021. <https://www.kaggle.com/datasets/yasserh/walmart-dataset/data>
- [3] Visual Crossing, Weather Data Services: Visual Crossing, www.visualcrossing.com, <https://www.visualcrossing.com/weather/weather-data-services>