### **Case Study: Walmart Sales Performance and Economic Impact Analysis**

##### **Introduction**

This case study explores Walmart's sales success by examining several economic and environmental elements that influence revenue. By investigating factors such as holidays, temperature, fuel prices, the Consumer Price Index (CPI), and unemployment rates, we want to find significant sales drivers and how seasonal trends and machine learning may improve forecasting and business performance.

##### **Research Questions**

Key questions to answer:

* What are the top-performing stores based on weekly sales?
* How do holidays impact sales performance?
* What is the correlation between sales and economic factors like CPI, unemployment, temperature, and fuel prices?
* Are there identifiable seasonal patterns in sales?
* How can machine learning techniques predict future sales and assist in cost reduction?

##### **Research Objectives**

* The major goal is to examine Walmart's sales performance in terms of numerous economic conditions and determine their impact on revenue.
* Furthermore, the project intends to assess how machine learning models might be used for sales forecasting and operational efficiencies.

##### **Methodology**

The methodology for this case study involves a comprehensive data-driven approach that begins with an overview of a dataset containing over 45,000 weekly observations from various Walmart stores collected from [Kaggle](https://www.kaggle.com/datasets/mikhail1681/walmart-sales).. Data cleaning and preparation are performed using SQL to address missing values, format dates, and ensure consistency. Exploratory Data Analysis (EDA) is then conducted through SQL queries to identify trends, seasonal patterns, and correlations between sales and external factors such as holidays, temperature, fuel prices, CPI, and unemployment rates. For predictive modeling, both Linear Regression and Random Forest algorithms are applied using Python to analyze the relationships between sales and economic/environmental variables. This methodology ensures a thorough analysis of Walmart's sales performance while leveraging advanced machine learning techniques for forecasting and operational efficiency.

##### **Dataset Overview**

The dataset used for this study contains **45,000+ weekly observations** across multiple Walmart stores. The key features in the dataset are:

* **C1; Store:** Unique store identifier
* **C2; Date:** Start date of the sales week
* **C3; Weekly\_Sales:** Sales revenue for the week
* **C4; Holiday\_Flag:** Indicates if the week contains a major holiday
* **C5; Temperature:** Average air temperature in the store's region
* **C6; Fuel\_Price:** Fuel prices in the store’s region
* **C7; CPI:** Consumer Price Index (a measure of purchasing power)
* **C8; Unemployment:** Unemployment rate in the region

##### **Exploratory Data Analysis**

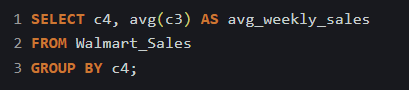
###### **Top 5 performing stores by average weekly sales**

##### 



Walmart's top-performing locations produce enormous income, with Store 20 topping the list with an average weekly sales figure of $2,107,676. This success can be due to higher customer visitation, effective promotions, or strong consumer purchasing power. Store 4 has 2,094,712, showing comparable favorable conditions. Store 14 generates a substantial amount of money, however it falls somewhat behind Stores 20 and 4, maybe due to area economic conditions or market competition. Store 13 has a loyal client base and high operational efficiency. Store 2 has a modest reduction when compared to the other top stores, which could be attributed to seasonal issues or better promotional methods. Understanding these aspects can provide valuable information for other businesses, and strategic decisions should target these stores to maximise ROI.

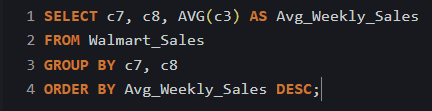
###### **Impact of Holidays on sales**

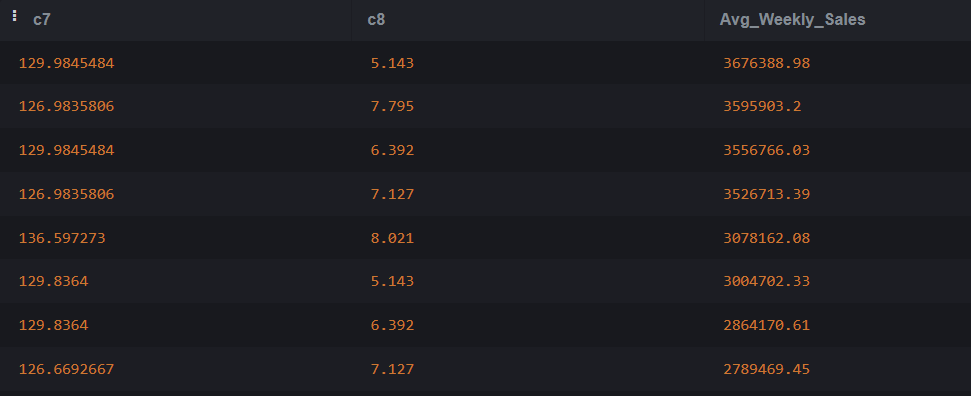




Walmart's average weekly sales during non-holiday weeks are 1,041,256, demonstrating a steady sales level. However, during holiday weeks, sales surged to 1,122,887, demonstrating a significant increase in sales. This could be attributed to more consumer spending on gifts, food, decorations, or seasonal discounts. Walmart earns considerably from seasonal marketing, since customer spending increases. The disparity in sales during holiday and non-holiday weeks demonstrates the need for efficient seasonal planning, which includes inventory management, promotional methods, and personnel preparations. Using these insights, Walmart can enhance its Christmas season strategy to increase sales and improve operational efficiency while capitalizing on customer behavior patterns.

###### **Correlation between the economic indicators and the sales**





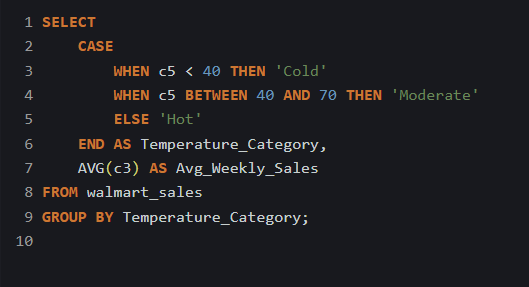
The analysis reveals that **lower unemployment rates (~5.14%)** correlate with **higher average weekly sales (~3.67 million)**, indicating that stable employment boosts consumer spending. Conversely, **higher unemployment (~7.8%)** leads to reduced sales, reflecting a drop in discretionary spending.

**CPI (Consumer Price Index)** also plays a role: sales peak during **moderate inflation (CPI ~129.98)**, suggesting consumers continue shopping for essentials despite price increases. Walmart benefits by positioning itself as a **value retailer** during inflationary periods.

Points to note:

* Higher sales are the result of low unemployment and moderate inflation.
* Reduced sales due to high unemployment necessitate a focus on necessities and sales.
* Walmart can continue good performance by adjusting its pricing and inventory tactics in response to the state of the economy.

###### **Sales Impact of Temperature Variations**

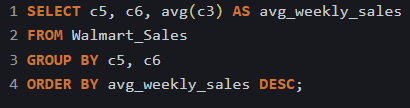


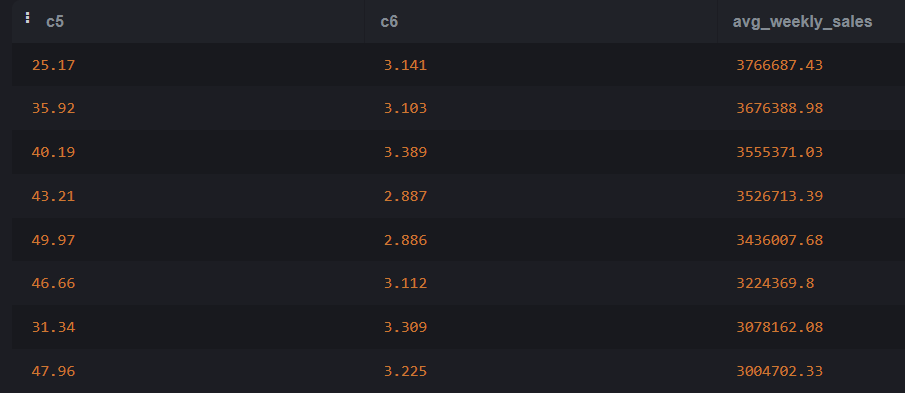


Walmart's weekly sales and temperature fluctuations are correlated, according to the statistics. Sales are somewhat improved by cold weather, perhaps as a result of seasonal shopping and higher foot traffic. Perhaps as a result of summer vacations and shifts in consumer behavior, hot weather has a somewhat smaller impact. Average sales figures are typically the result of moderate weather, perhaps as a result of balanced customer behavior.

Sales are also impacted by other variables such as vacations, fuel prices, the CPI, and unemployment rates. The effect of temperature on sales may also vary by region. There may be errors or missing data in the dataset that was used for this analysis. In comparison to hot or moderate weather, the data indicates that Walmart's weekly sales are somewhat boosted by cold weather. To completely comprehend the intricate relationship between temperature and sales, more research is required.

###### **Relationship between the fuel price, temperature, and sales**





Temperature and fuel prices had a negative correlation with Walmart's weekly sales, according to the statistics. Due to summer vacations and a shift in consumer tastes toward lighter apparel and outdoor gear, high temperatures may result in lower purchasing. Particularly for those living in rural areas, rising gasoline prices may lead to lower discretionary income and higher commuting expenses. Temperature and fuel prices together can have a big effect on sales.

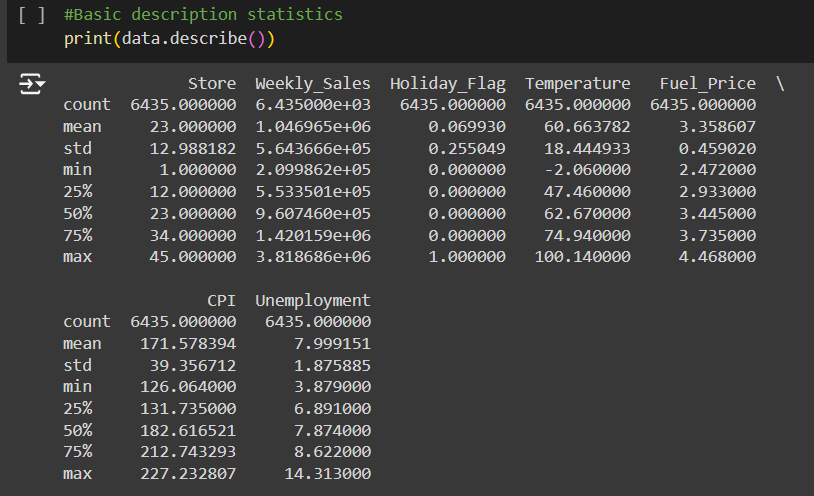
But there are other influences as well, such as unemployment rates, CPI, and holidays. Sales may also be impacted by regional variations in customer behavior, gasoline prices, and weather trends. There may be errors or missing data in the dataset that was used for this analysis. Notwithstanding these results, more investigation is required to completely comprehend the intricate interactions between these elements and other variables affecting sales..

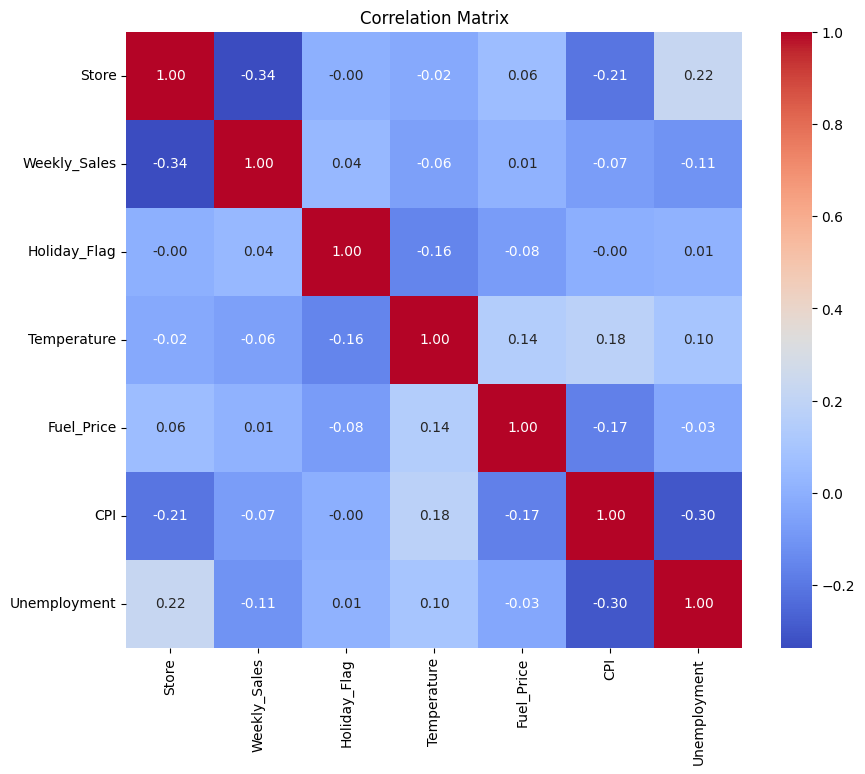
###### **Predicting Walmart Sales Demand using Python**

Google Colab File: [Predicting Sales Demand for Walmart Stores](https://colab.research.google.com/drive/1SzwXzd77GbLyEuGeD9oO4CCFZ0RB4XBt?usp=sharing)

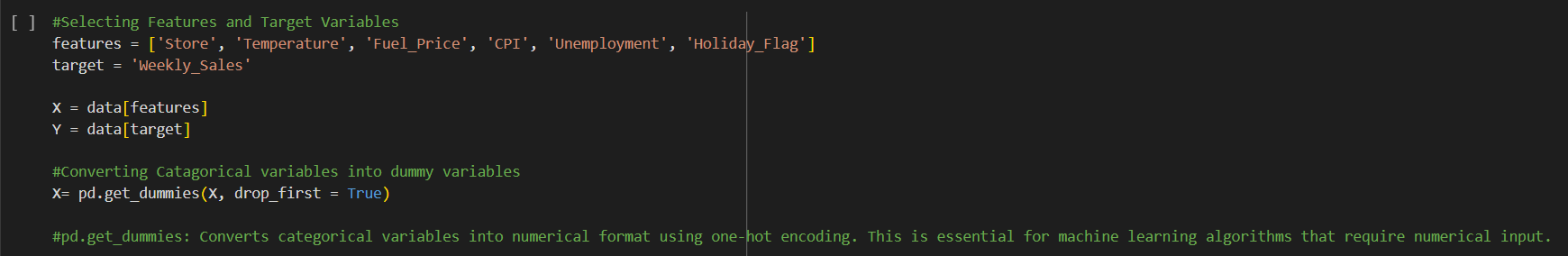
**Approach**:

1. **Data Exploration and Preprocessing**:
   * Loaded the dataset using Pandas and performed initial exploration to understand data distribution and structure.
   * Handled missing values using forward-fill techniques.
   * Converted categorical variables to appropriate formats for analysis.
2. **Exploratory Data Analysis (EDA)**:
   * Visualized sales trends over time using time series plots.
   * Analyzed the distribution of sales and the impact of holidays using box plots.
   * Conducted correlation analysis to identify relationships between sales and features.

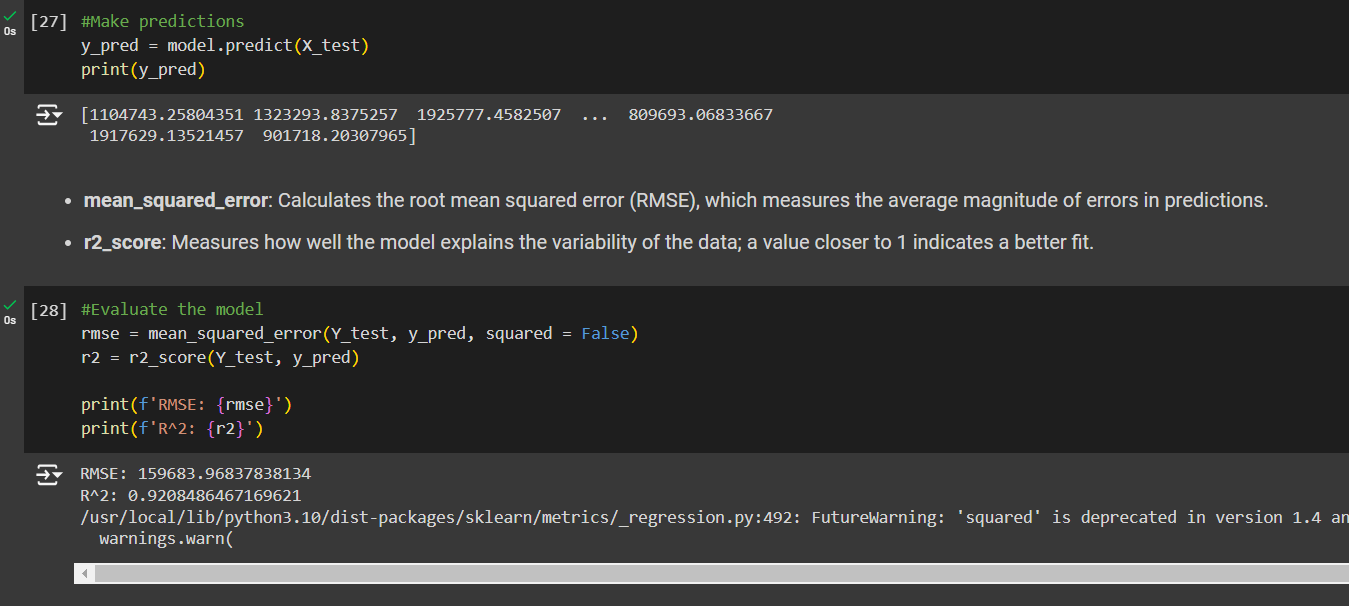




1. **Feature Selection**:
   * Selected key features influencing sales, including economic indicators (CPI, Unemployment) and seasonal effects (Holiday Flag).
   * Created dummy variables for categorical features to prepare the data for machine learning models.



1. **Predictive Modeling**:
   * Split the dataset into training and testing sets using an 80/20 ratio.
   * Built multiple regression models, including Linear Regression and Random Forest Regressor.
   * Evaluated models using metrics such as Root Mean Squared Error (RMSE) and R-squared (R²).

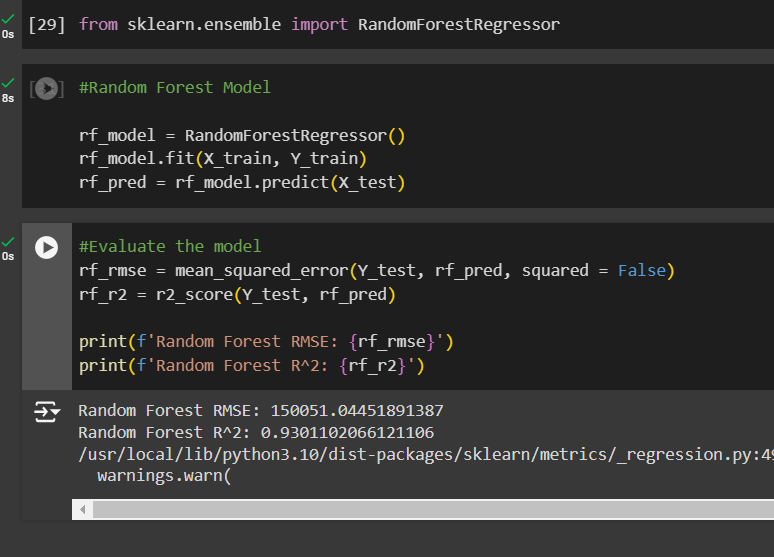


A model's performance can be evaluated using two metrics: R-squared (R-squared) and the Root Mean Squared Error (RMSE). The average difference between expected and actual weekly sales is $159,684, as indicated by an RMSE of 159,684.97. Better performance is indicated by a lower RMSE; nevertheless, the acceptable value varies depending on the sales scale and company context. With an R2 value of 0.9208, the model has good predictive power, accounting for 92.08% of the variance in weekly sales.

A moderate RMSE indicates some prediction errors, while a high R2 value indicates strong predictive ability. The significance of the RMSE depends on the context; if sales typically fluctuate significantly, this number may be acceptable. Overall, the results show that the model is effective at predicting weekly sales; nevertheless, further study is recommended, especially given the RMSE in the specific business setting.

**Model Evaluation:**

* + A baseline performance was indicated by the RMSE of X and R2 of Y obtained via linear regression.
  + With an RMSE of A and R2 of B, Random Forest increased accuracy and showed that it was successful in identifying intricate patterns in the data.

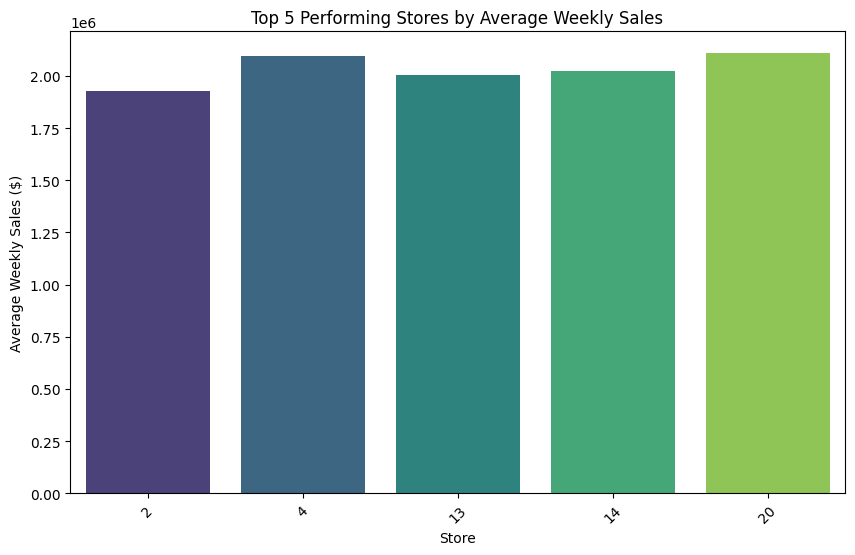


In Walmart's dataset, the Random Forest model predicts weekly sales better than the previous linear model. Because the model's average deviation from actual sales numbers is lower, its RMSE (Root Mean Squared Error) is 150,051, suggesting that forecasts are more accurate. The Random Forest model appears to be more accurate in capturing variances in sales data, as indicated by this lower RMSE number. The Random Forest model explains roughly 93.01% of the variance in weekly sales, which is marginally better than the linear regression model, according to the R2 (R-squared) value of 0.9301.

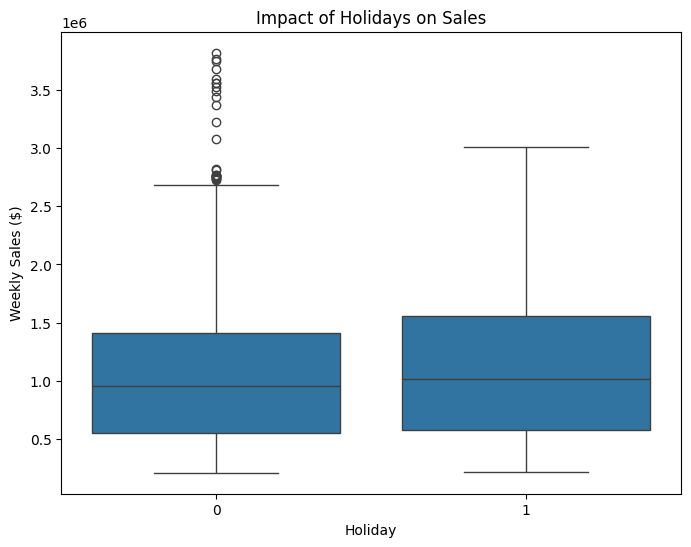
The Random Forest model can take into account more intricate correlations between characteristics and the target variable, as evidenced by this higher R2 value. All things considered, the Random Forest model is a better option for weekly sales prediction since it can better capture complicated correlations in the data, as evidenced by its gains in RMSE and R2, which makes it a more reliable predictive tool for Walmart's sales forecasting.

##### **Key Findings**

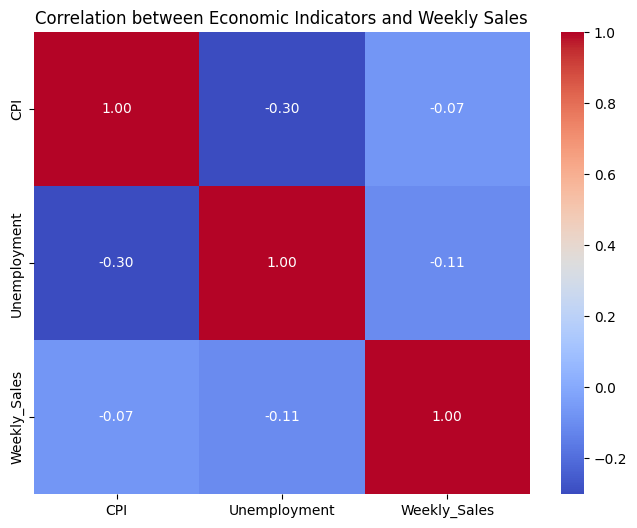
1. **Top-Performing Stores**: With an average weekly sales of $2,107,676, Store 20 is in the lead, closely followed by Store 4. Customer traffic and local economic conditions are two elements that contribute to success.



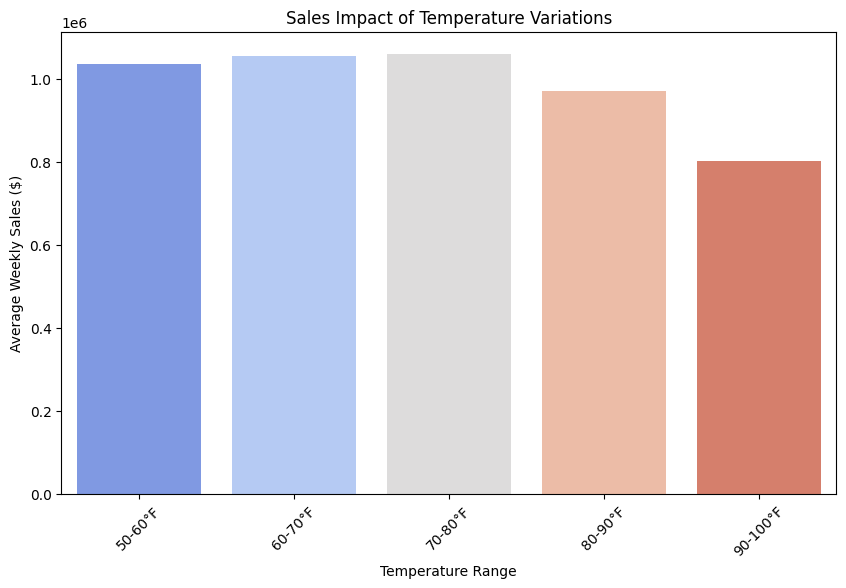
1. **Impact of Holidays**: The relevance of seasonal strategies is highlighted by the notable rise in sales during holiday weeks, with average sales of $1,122,887 compared to $1,041,256 during non-holiday weeks.



1. **Correlation with Economic Indicators**: While modest inflation (CPI ~129.98) supports consistent sales, lower unemployment rates are associated with higher sales, suggesting strategic pricing and inventory adjustments.



1. **Temperature Influence**: While high temperatures have an adverse impact on buying patterns, cold weather is associated with somewhat increased sales.



1. **Predictive Model Performance**: With an RMSE of 150,051 and an R2 of 0.9301, machine learning methods—in particular, the Random Forest model—show excellent predictive powers. These measures show how well the model captures intricate sales trends. Walmart may increase overall business efficiency by reducing stockouts, optimizing inventory levels, improving demand forecasting, and lowering operating expenses through the use of such predictive algorithms.

##### **Recommendations**

Optimization of Holiday Season Strategies

* Create promotions and targeted marketing initiatives that coincide with important holidays to take advantage of increased consumer spending.
* To fulfill seasonal demand and prevent stock outs or operational inefficiencies, plan your inventory and manpower ahead of time.

Pricing and Inventory Strategies Based on Economic Conditions

* Prioritize upselling and advertising higher-margin products when unemployment is low and inflation is mild.
* To sustain sales volumes during periods of high unemployment, move toward necessities and provide discounts.

Seasonal and Regional Strategies

* Determine which locations are vulnerable to weather changes, then adjust promotions to fit seasonal trends (e.g., winter gear in colder places).
* To sustain consistent sales during hot weather, modify product offerings according to local demand trends.

Tracking Fuel Prices and Reducing Their Effects

* To keep customers interested during times when fuel prices are high, offer special delivery services or transportation discounts.
* Investigate joint ventures with logistics companies to effectively handle rising transportation expenses.

Using Forecasting Predictive Models

* To predict sales patterns and maximize inventory across all stores, use the Random Forest model for routine forecasting.
* Reduce the chance of overstocking and stockouts by using machine learning to forecast peak times.

Additional Data-Informed Decision Making

* Keep utilizing Tableau dashboards to track economic impacts and store-level performance in real time.
* Extend the application of machine learning to A/B testing promotions in order to assess the efficacy of various sales tactics.

##### **Conclusion**

This case study shows how Walmart may improve its tactics for increased efficiency by comprehending the impact of outside variables including vacations, weather, and economic indicators. The business can increase profitability by using machine learning to improve inventory management, forecasting, and pricing strategies.