The Impact of Holidays on Weekly Sales Patterns

Team members: Chloe Pham, Gracie Nguyen, Nam Nguyen, Thao Nguyen.

Date: 05/12/2025

I. Introduction

This report investigates:

How do holidays impact weekly sales patterns across Walmart stores?

We aim to determine if holiday weeks show statistically significant sales differences compared to non-holiday weeks and explore holiday effects across stores and departments. Holidays like Black Friday and Christmas drive consumer spending, impacting retail revenue and operations (e.g., inventory, staffing) while reflecting economic trends in consumer confidence. Understanding these patterns helps retailers optimize strategies and informs policymakers about spending behaviors.
Our objectives are to: (1) quantify holiday effects on sales, (2) test statistical significance of holiday vs. non-holiday sales, and (3) offer insights for retail optimization.

II. Data Description

This study uses the Walmart Sales Forecasting dataset from Kaggle, consisting of two key files: *train.csv* and *features.csv*. The *train.csv* file contains 421,570 weekly sales records from 45 Walmart stores, spanning February 2010 to October 2012. Each record includes the store number, department, date, weekly sales amount, and a binary holiday indicator (IsHoliday). The *features.csv* file provides 8,190 store-date records with contextual variables: temperature, fuel prices, consumer price index (CPI), unemployment rates, promotional markdowns (MarkDown1-5), and IsHoliday.

We merged *train.csv* and *features.csv* on Store and Date using an inner join to a unified dataset of 6,435 records. The resulting dataset includes nine variables: Store, Date, Temperature, Fuel_Price, CPI, Unemployment, IsHoliday, IsMarkdown, and Weekly_Sales. The IsMarkdown variable, derived from MarkDown1-5, indicates whether a store had promotional markdowns (True/False) on a given date, simplifying the analysis of promotional effects.

1. Data Preprocessing and Quality

We performed data quality checks on *train.csv,* confirming no missing values or duplicate records across its 421,570 entries. To analyze store-level trends, we aggregated train.csv by Store

and Date, summing `Weekly_Sales` throughout all Dept of one Store. We also dropped `Dept` and `IsHoliday` columns, as they were redundant after aggregation.

The *features.csv* dataset has missing values in MarkDown1-5 (approximately 50–60% non-null), CPI, and Unemployment (7,605/8,190 non-null). We addressed MarkDown1-5 by creating the binary `IsMarkdown` variable, where True indicates at least one non-null markdown. Missing CPI and Unemployment values were retained for initial analysis but may require imputation for further statistical tests.

The merged dataset's structure is summarized below, showing no missing values in key columns after preprocessing.

```
RangeIndex: 6435 entries, 0 to 6434
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Store         6435 non-null   int64
 1   Date          6435 non-null   datetime64[ns]
 2   Temperature   6435 non-null   float64
 3   Fuel_Price    6435 non-null   float64
 4   CPI           6435 non-null   float64
 5   Unemployment  6435 non-null   float64
 6   IsHoliday     6435 non-null   bool
 7   IsMarkdown    6435 non-null   bool
 8   Weekly_Sales  6435 non-null   float64
dtypes: bool(2), datetime64[ns](1), float64(5), int64(1)
memory usage: 364.6 KB
```

	Store	Date	Temperature	Fuel_Price	CPI	Unemployment	IsHoliday	IsMarkdown	Weekly_Sales
0	1	2010-02-05	42.31	2.572	211.096358	8.106	False	False	1643690.90
1	1	2010-02-12	38.51	2.548	211.242170	8.106	True	False	1641957.44
2	1	2010-02-19	39.93	2.514	211.289143	8.106	False	False	1611968.17
3	1	2010-02-26	46.63	2.561	211.319643	8.106	False	False	1409727.59
4	1	2010-03-05	46.50	2.625	211.350143	8.106	False	False	1554806.68

Figure 1: Summary of the merged dataset, showing 6,435 records and 9 variables with no missing values in critical columns.

2. Preliminary Findings

The merged dataset provides a comprehensive foundation for analyzing sales trends across 45 individual retail stores. As a first step in our exploratory data analysis, we aimed to identify a store that best represents the overall sales behavior in the dataset. To achieve this, we created a bar graph plotting the average weekly sales of each store and selected the median store based on this distribution. Store 34 appeared as the most representative, with average weekly sales around $1 million. This selection allows us to conduct a focused yet generalizable investigation into broader sales patterns. To research deeper into the behavior of Store 34, we generated a time series visualization depicting its weekly sales over the entire observation period.

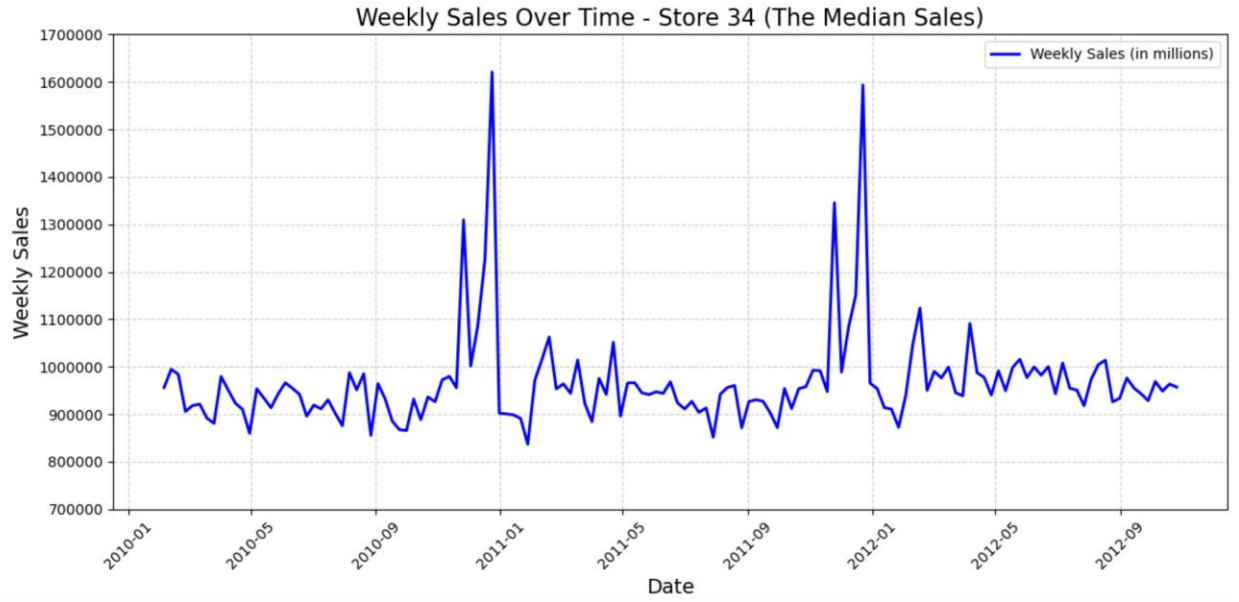

Figure 2: Weekly sales for Store 34, showing high volatility with peaks likely tied to holiday periods like November–December.

The store exhibits significant sales volatility, with pronounced peaks likely corresponding to holiday seasons (e.g., Black Friday and Christmas), which suggests that consumer spending at this store is highly seasonal, driven by well-timed promotional events and broader holiday-related demand. These visual insights indicate that holidays play a pivotal role in driving spikes in sales, while non-holiday periods display more stable but comparatively lower sales volumes. The structure of the dataset, combined with these preliminary visualizations, provides a strong starting point for more statistical analyses.

III. Time Series Model Application

To address the research question, *"How do holidays impact weekly sales patterns across Walmart stores?"*, we applied an Autoregressive (AR) model with four lags (AR(4)). The choice of four lags corresponds to a one-month lag structure, assuming weekly data and aiming to capture short-term temporal dependencies in sales patterns. By including four previous weeks of sales data, the model allows us to account for autocorrelation and assess how current sales are influenced not only by recent trends but also by the presence of holiday weeks.

This modeling approach helps isolate the impact of holidays while controlling for underlying seasonality and persistence in sales behavior. The AR(4) model results are summarized below:

AutoReg Model Results

Dep. Variable:	Weekly_Sales	No. Observations:	143
Model:	AutoReg-X(4)	Log Likelihood	-1749.267
Method:	Conditional MLE	S.D. of innovations	70665.771
Date:	Sun, 11 May 2025	AIC	3534.534
Time:	22:45:05	BIC	3587.355
Sample:	03-05-2010	HQIC	3555.999
	- 10-26-2012		

	coef	std err	z	P>|z|	[0.025	0.975]
const	4.508e+05	1.86e+05	2.427	0.015	8.68e+04	8.15e+05
Weekly_Sales.L1	-0.3085	0.087	-3.529	0.000	-0.480	-0.137
Weekly_Sales.L2	0.0816	0.069	1.185	0.236	-0.053	0.216
Weekly_Sales.L3	0.0915	0.072	1.265	0.206	-0.050	0.233
Weekly_Sales.L4	0.4326	0.078	5.526	0.000	0.279	0.586
IsMarkDown	2.571e+04	1.39e+04	1.845	0.065	-1606.789	5.3e+04
Feb	3.006e+05	5.08e+04	5.915	0.000	2.01e+05	4e+05
Mar	1.803e+05	4.02e+04	4.487	0.000	1.02e+05	2.59e+05
Apr	2.203e+05	4.33e+04	5.083	0.000	1.35e+05	3.05e+05
May	2.122e+05	4.35e+04	4.879	0.000	1.27e+05	2.97e+05
Jun	2.128e+05	4.26e+04	4.991	0.000	1.29e+05	2.96e+05
Jul	1.741e+05	4.34e+04	4.009	0.000	8.9e+04	2.59e+05
Aug	2.211e+05	4.55e+04	4.857	0.000	1.32e+05	3.1e+05
Sep	1.78e+05	4.41e+04	4.037	0.000	9.16e+04	2.64e+05
Oct	2.054e+05	4.66e+04	4.406	0.000	1.14e+05	2.97e+05
Nov	3.31e+05	4.7e+04	7.040	0.000	2.39e+05	4.23e+05
Dec	4.394e+05	4.58e+04	9.586	0.000	3.5e+05	5.29e+05

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-1.1200	-0.0000j	1.1200	-0.5000
AR.2	1.2563	-0.0000j	1.2563	-0.0000
AR.3	-0.1738	-1.2700j	1.2818	-0.2717
AR.4	-0.1738	+1.2700j	1.2818	0.2717

Figure 3: Result of AR(4) Model

Using the AR(4) model results, we focus on the coefficients of the monthly dummy variables and the `IsMarkDown` dummy, as these capture seasonal and promotional effects likely tied to holidays. The model predicts `Weekly_Sales` based on 143 weekly observations from March 5, 2010, to October 26, 2012, across Walmart stores. Below, the impact of holidays on weekly sales patterns is interpreted, provide specific insights for each month, and address the role of markdowns.

1. General Interpretation

The monthly dummy variables (February to December, with January as the reference) capture seasonal sales patterns, many of which are driven by holidays. The coefficients represent the average change in weekly sales relative to January, holding other factors constant. Significant coefficients indicate months where holidays or seasonal events are likely to boost sales. The `IsMarkDown` dummy reflects the impact of promotional markdowns, which may be holiday related.

2. Detailed Monthly Holiday and Mark Down Impacts

Compared to January, several months demonstrate significantly higher sales, with February, November, and December standing out most prominently. All coefficients are statistically significant ($p < 0.001$), indicating strong confidence in these seasonal effects. February shows a sharp increase in sales, with a coefficient of 300,600, reflecting the combined impact of the Super Bowl and Valentine's Day. These events spur consumer spending on food, gifts, flowers, and dining-related products. November, with a coefficient of 331,000, marks the beginning of the holiday shopping season, driven by Thanksgiving and the highly anticipated Black Friday sales. Consumers typically make large purchases of electronics, home goods, and holiday preparations during this time. December tops the chart with a coefficient of 439,400, representing the highest seasonal sales peak. This increase is largely attributable to Christmas and Hanukkah, which generate substantial spending on gifts, decorations, food, and travel. These three months clearly benefit from major national holidays and cultural traditions that significantly boost retail performance.

Other months also display meaningful, though more moderate, increases relative to January, each shaped by their own seasonal and retail events. September, with a coefficient of 178,000, shows a modest sales uplift linked to Labor Day, which stimulates demand for outdoor goods, appliances, and clearance items as summer winds down. March (180,300) and April (220,300) reflect the influence of Easter, which may fall in either month, and St. Patrick's Day in March. These holidays drive sales in candy, clothing, and home decor. May (212,200) and June (212,800) benefit from family-focused holidays such as Mother's Day, Memorial Day, and Father's Day, along with the start of summer, encouraging purchases of gifts, grills, furniture, and summer gear. July (174,100) experiences a smaller bump, as Independence Day is a single-day event, one associated with food, party supplies, and patriotic merchandise. August (221,100) is notable for its strong performance despite lacking a major federal holiday, due to back-to-school shopping, a key retail driver of apparel, electronics, and school supplies. October (205,400), driven by Halloween, rounds out the seasonal calendar with increased sales of costumes, decorations, and candy. Overall, while every month shows a significant increase over January, the magnitude varies in line with the intensity and duration of associated holidays and consumer shopping behavior.

The impact of markdowns on weekly sales is measurable but less pronounced than the seasonal effects tied to major holidays. The variable `IsMarkDown` has a coefficient of 25,710 with a p-value of 0.065, indicating that while markdowns are associated with an average increase of approximately $25,710 in weekly sales, the result is only marginally significant at conventional statistical thresholds. This suggests that while markdowns do contribute positively to sales, their effectiveness is more variable and context dependent. Markdowns are frequently employed during key retail periods such as post-Thanksgiving and post-Christmas, where they serve both to clear excess inventory and to draw in price-sensitive customers. However, compared to the larger sales lifts seen in months like November and December - driven by underlying holiday demand - the incremental boost from markdowns appears modest.

The analysis reveals clear patterns in how various holidays influence retail sales. The peak holiday seasons occur in November and December, which show the largest sales increases with coefficients of 331,000 and 439,400, respectively. These gains are driven primarily by Thanksgiving, Black Friday, and Christmas, underscoring the critical importance of these months to Walmart's annual revenue. These periods demand proactive planning around inventory, staffing, and supply chain readiness to capture heightened consumer demand. Secondary peaks are observed in February, April, and August, with coefficients ranging from approximately 200,000 to 300,000. These increases are linked to Valentine's Day in February, Easter (which may fall in April), and back-to-school shopping in August - events that are significant, though not as intense as year-end holidays. Moderate effects are seen in months like July and September, with coefficients near 174,000 to 178,000. These reflect single-day holidays such as Independence Day and Labor Day, which drive short bursts of consumer spending but lack the sustained impact of multi-week seasonal events. In contrast, the role of markdowns is more limited. While they offer a modest average sales boost of around $25,710 per week, the marginal statistical significance ($p = 0.065$) suggests markdowns function more as a supportive tactic, especially when paired with larger holiday promotions, rather than a primary driver of holiday-related sales surges.

2. Model testing with in-sample value

To confirm our model finding, we generated in-sample predictions

.

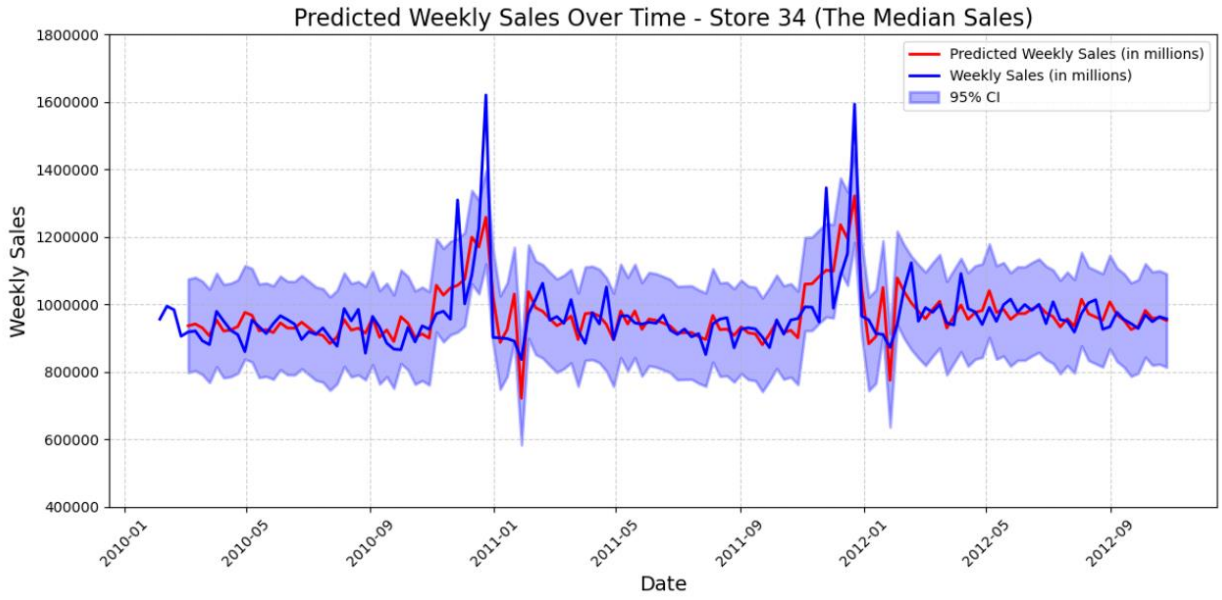

Figure 4: Predicted Weekly Sales with in-sample values

The graph presents a time series analysis of weekly sales at Store 34 from early 2010 through late 2012. Actual sales are represented by the blue line, predicted sales by the red line, and a 95% confidence interval (CI) is shaded around the predicted values.

a. Model Fit and Performance

The model shows a generally good fit, with predicted sales closely following the actual sales trend throughout most of the period. In the majority of weeks, actual values fall within the 95% confidence interval, indicating that the model successfully captures the underlying trend and normal variability in the data. This alignment suggests that the model is well-calibrated for forecasting average weekly sales levels under typical conditions.

b. Performance During Sales Spikes

While the model handles routine fluctuations reasonably well, it struggles with large, sudden spikes in sales. Notably, periods such as late 2010, mid-2011, and late 2011 exhibit sharp increases in actual sales that are underestimated by the model and fall outside the predicted range. These deviations likely correspond to key seasonal events - such as major holidays or promotional periods - that are not fully captured by the model. This pattern indicates that the model lacks certain explanatory variables that influence extreme short-term surges.

c. Seasonal Trends and Volatility

The model does pick up on recurring seasonal patterns, evidenced by periodic peaks in the predicted line. However, it consistently underestimates the magnitude of those seasonal peaks, suggesting that while it detects cyclical trends, it fails to fully quantify their impact. The

underlying sales data is also quite volatile, with frequent short-term shifts. The model accommodates this volatility within its confidence bounds most of the time but fails to predict the most extreme cases accurately.

Overall, the model provides a reliable forecast of general trends and moderate fluctuations in weekly sales at Store 34. However, it underperforms in predicting sharp, event-driven spikes.

IV. Strategic Recommendations

To maximize sales performance across Walmart stores, we propose a seasonally informed and performance-sensitive strategy rooted in insights from the AR(4) model and sales trend analysis. The data highlights significant fluctuations in weekly sales associated with specific calendar events, particularly national holidays and seasonal campaigns. These fluctuations offer strategic entry points for Walmart to drive traffic, increase conversion, and better allocate operational resources.

1. Maximize Holiday Season Performance (November–December)

The period from mid-November through the end of December consistently shows the highest sales volumes across most stores, reflecting the powerful influence of Black Friday, Cyber Monday, and Christmas. To capitalize on this surge, Walmart should proactively increase inventory levels and staffing in its top-performing locations well ahead of the season. In addition, launching early promotional campaigns and exclusive holiday bundles can help extend the shopping window, attracting early-bird consumers and reducing last-minute logistics pressure. Given the rise in e-commerce, investing in online order processing, curbside pickup, and delivery logistics during this time is also critical for maintaining customer satisfaction and capturing a larger market share.

2. Capitalize on Smaller Seasonal Peaks (February, March, May)

Beyond the traditional holiday season, other months show modest but consistent upticks in sales tied to culturally significant events like Super Bowl, Valentine's Day (February), the start of spring (March), and Mother's Day (May). Walmart can boost revenue during these months by executing thematic promotional campaigns tailored to these occasions. This includes curated marketing efforts focused on gift items, seasonal apparel, beauty products, home décor, and gardening supplies. Strategic placement in stores, email marketing, and localized advertising can help Walmart attract occasion-driven shoppers and diversify revenue streams beyond Q4.

3. Strengthen January Sales Recovery

January typically experiences a decline in sales due to post-holiday fatigue and budget-conscious consumers. However, this presents a unique opportunity for Walmart to reset and refocus its promotional strategy. Targeted campaigns centered on New Year resolutions - such as discounts on health, fitness, and wellness products - can align with consumer goals. Additionally, holding

inventory clearance events can move excess holiday stock while encouraging traffic during a slow retail month. Introducing loyalty rewards or bundled savings may also incentivize repeat visits during this otherwise soft period.

4. Tailor Strategy Based on Store Performance

Not all stores exhibit the same responsiveness to seasonal campaigns. Data shows that top-performing stores experience greater sales volatility and are more sensitive to promotions and holiday-related activity. These locations should follow an aggressive, dynamic promotional calendar with localized marketing and flexible staffing. In contrast, stores with lower or more stable sales patterns may benefit more from cost-efficient operations and a focus on steady-selling categories such as groceries, personal care, and household essentials. Segmenting stores by performance and tailoring strategies accordingly can lead to more efficient resource deployment and improved overall profitability.

In summary, by aligning operational planning and promotional strategies with seasonal sales behavior and store-level performance, Walmart can enhance customer satisfaction, increase profitability, and sustain competitive advantage throughout the year.

IV. Conclusion

This study confirms that holidays exert a substantial and statistically significant influence on weekly sales patterns across Walmart stores. Through rigorous data preprocessing, exploratory analysis, and an AR(4) time series model, we identified clear seasonal peaks in sales associated with key holidays, particularly in November and December. These months, driven by Thanksgiving, Black Friday, and Christmas, produced the most dramatic increases in weekly sales - underscoring their critical importance in retail planning and revenue generation.

Our findings also highlight the broader seasonal impact of holidays like Valentine’s Day, Easter, and back-to-school season, which create secondary but still meaningful boosts in consumer spending. While promotional markdowns do contribute positively to weekly sales, their effect is notably smaller and less consistent compared to the strong, recurring influence of holidays.

These insights have practical implications for retailers: strategic inventory management, targeted promotions, and optimized staffing around holiday periods can help capitalize on predictable surges in demand. Moreover, policymakers and economists can use these patterns as indicators of consumer confidence and spending behavior during holiday periods.

In summary, holidays are powerful drivers of retail sales, and understanding their impact can support more effective decision-making in both business and economic contexts.