Retail Sales Report

Data Analysis on sales from
02/2010 - 08/2012

Presented By : Dave Pekerow

Retail Store Sales Report

Task: Clean a dataset containing sales data for a big box retail chain and analyze sales, unemployment rate, CPI, fuel price, and other fields.

Data download: https://www.kaggle.com/datasets/mikhail1681/walmart-sales

Data Cleaning Practice

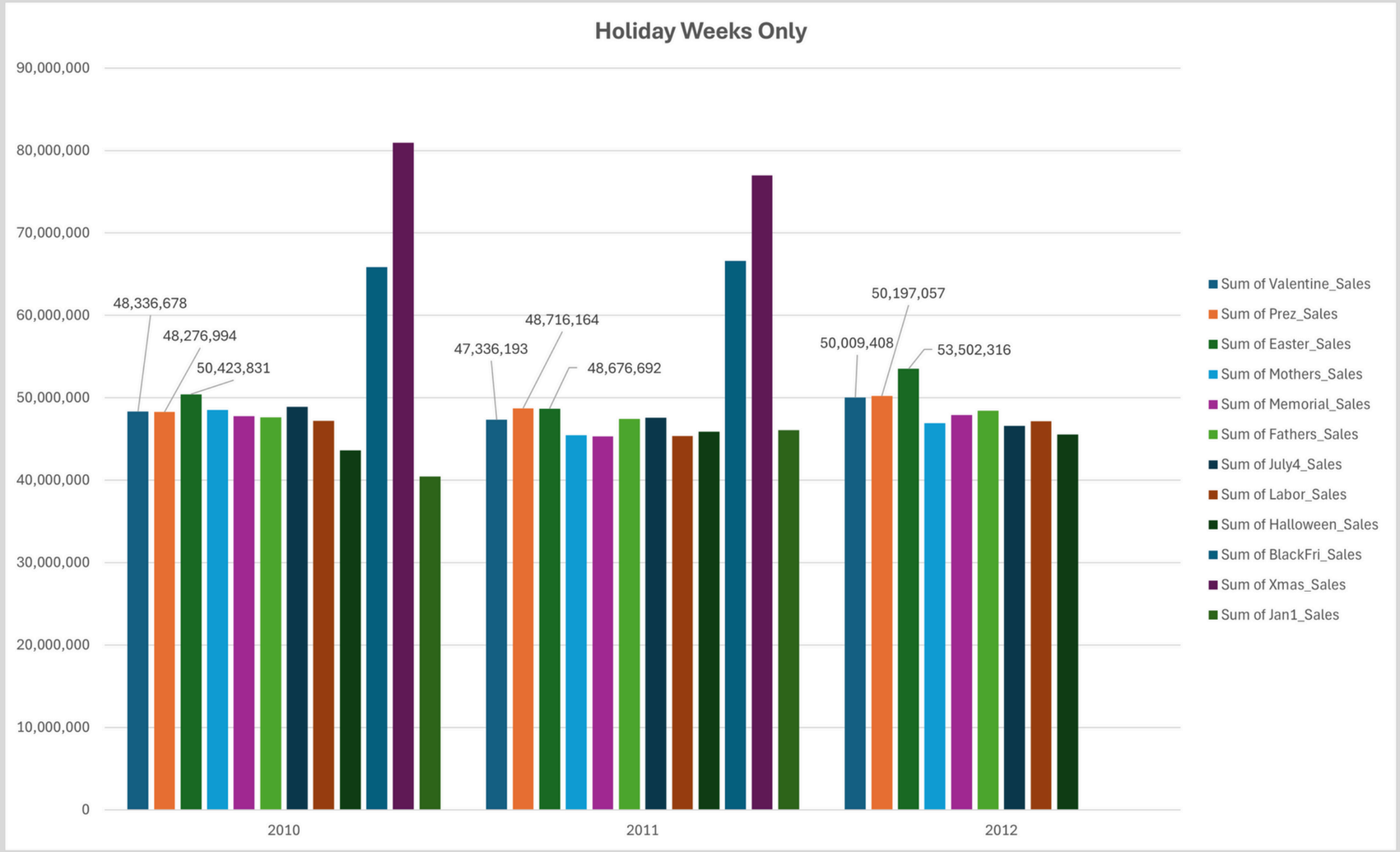
Using Python, R, Excel, or another data cleaning tool of your choice, clean the data to meet the following criteria:

- Data is sorted first by store number (ascending) and second by date (ascending)
- Date is in the format MM-DD-YYYY
- Weekly Sales is rounded to the nearest 2 decimal places
- Temperature is rounded to the nearest whole number
- Fuel Price is rounded to the nearest 2 decimal places
- CPI is rounded to the nearest 3 decimal places
- Unemployment is rounded to the nearest 3 decimal places
- Ensure that there is no missing data

Business Questions:

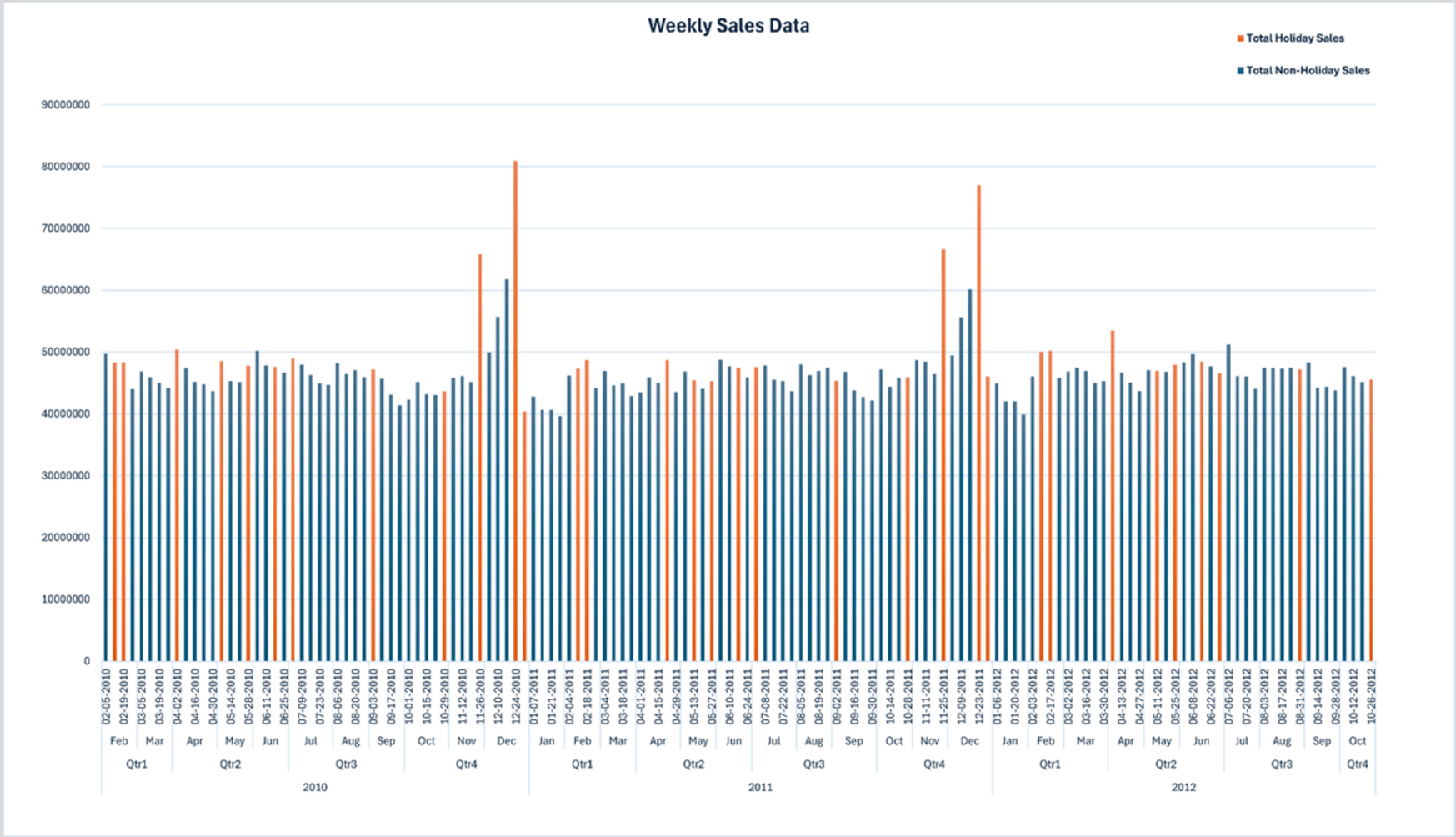
1. Which holidays affect weekly sales the most?
2. Which stores in the dataset have the lowest and highest unemployment rate? What factors do you think are impacting the unemployment rate?
3. Is there any correlation between CPI and Weekly Sales? How does the correlation differ when the Holiday Flag is 0 versus when the Holiday Flag is 1?
4. Why do you think Fuel Price is included in this dataset? What conclusions can be made about Fuel Price compared to any of the other fields?

Which holidays affect weekly sales the most?



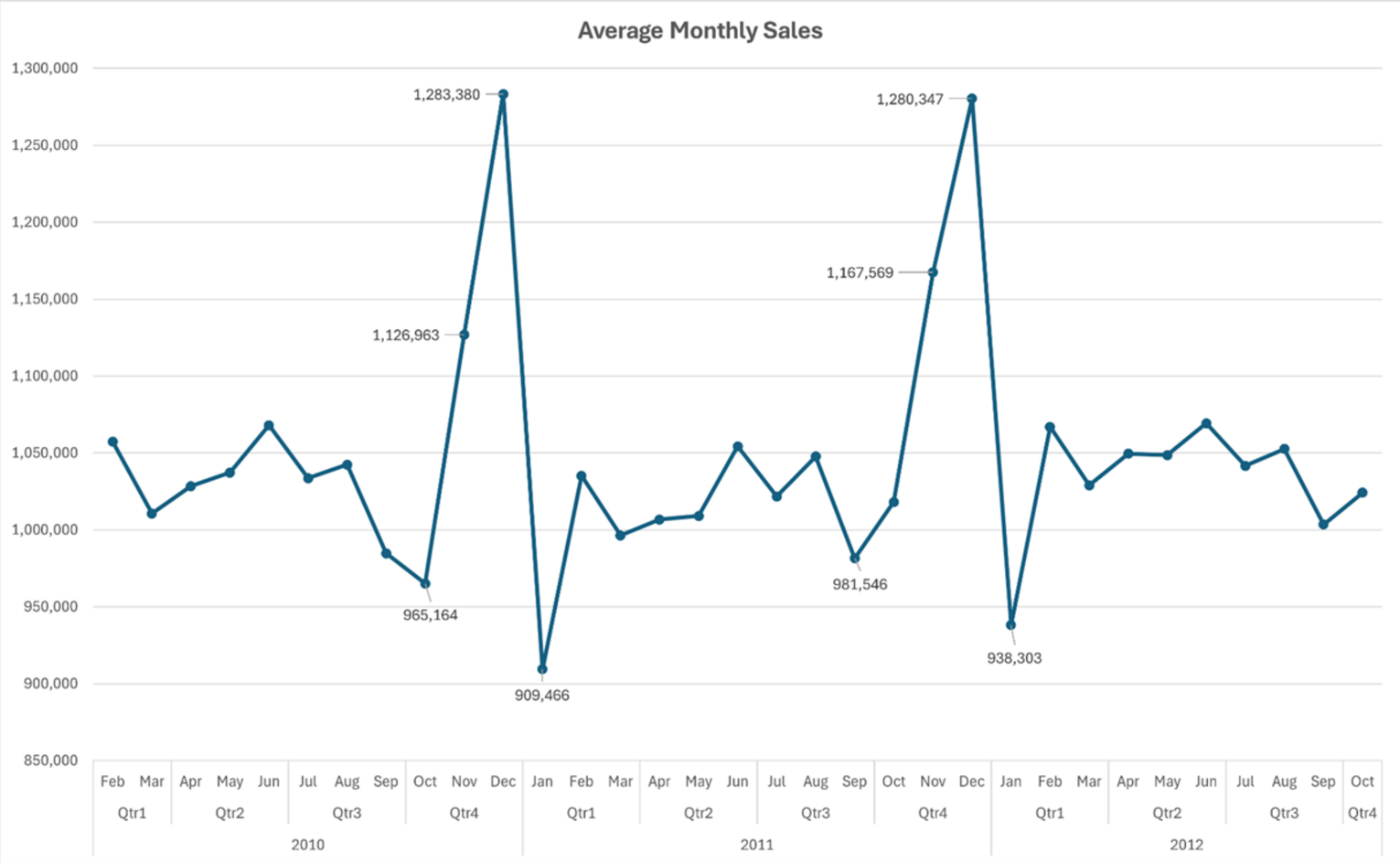
Black Friday and Christmas week are dominant. Valentine’s and Presidents’ Days, together with Easter, are the next strongest holiday sales events.

Which holidays affect weekly sales the most?



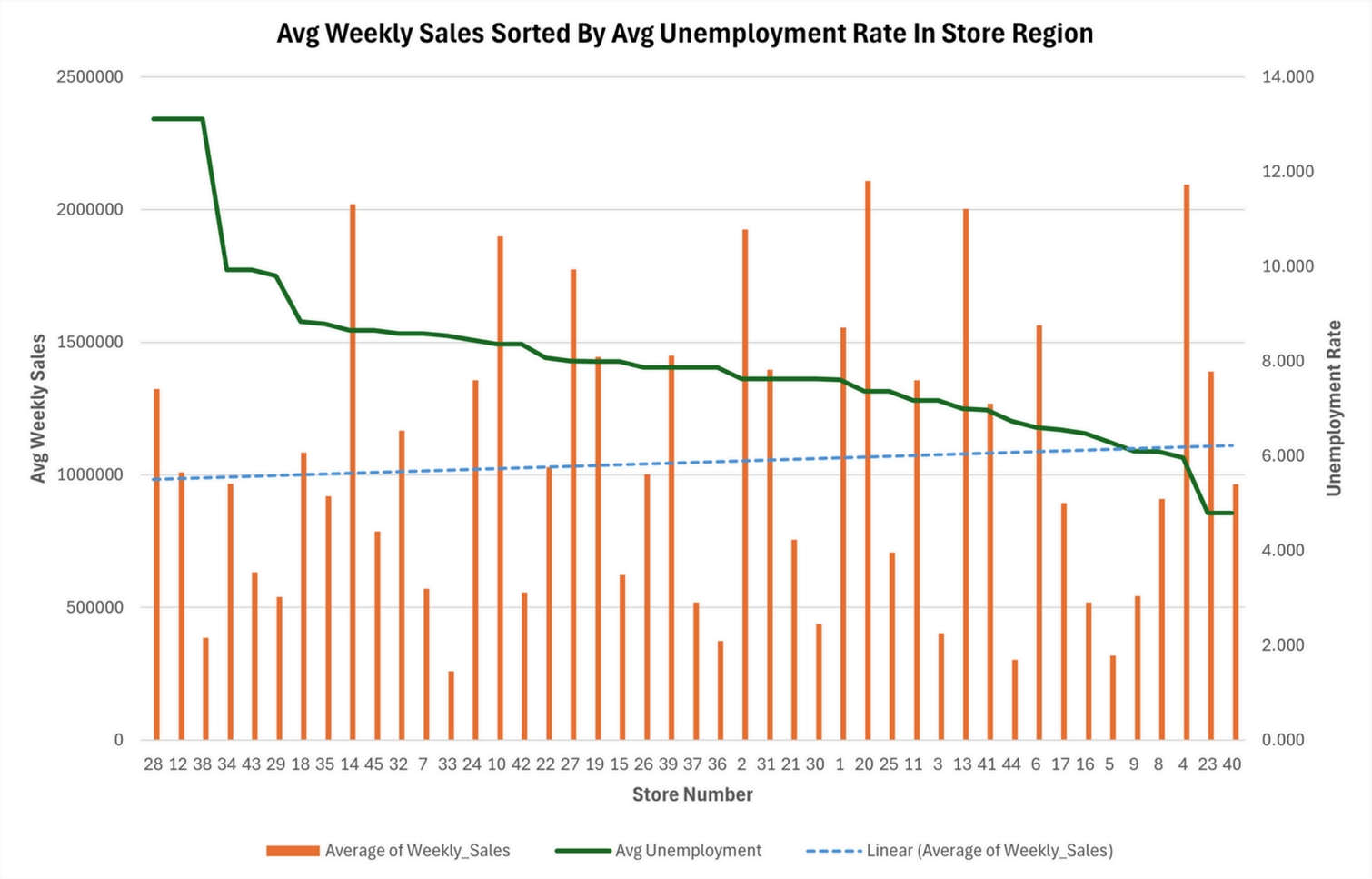
Weekly sales throughout the period show that Valentine’s Day, Presidents’ Day, and Easter sales are typical of any week. Overall, store sales remain steady, averaging $40-50 million, with the post-Thanksgiving season standing out as the only significant holiday sales event.

Which holidays affect weekly sales the most?



Though less granular than the weekly view, the monthly sales chart shows clear spikes in November and December each year, followed by a sharp drop in January. Sales stabilize in February and remain consistent throughout the rest of the year, with other holidays not generating similar numbers.

Which stores in the dataset have the lowest and highest unemployment rate?

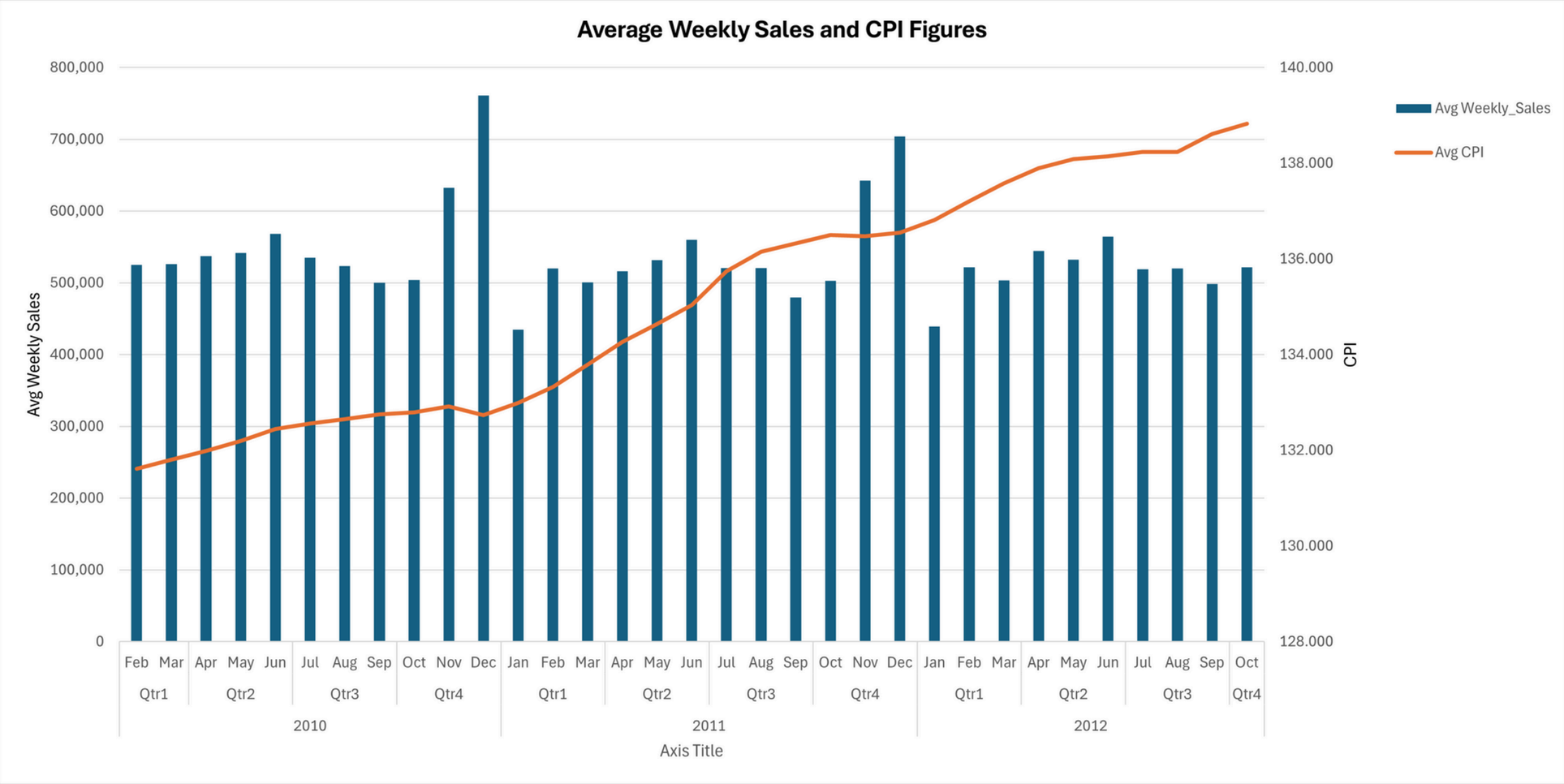


- Stores 28, 12, and 38 all had an average unemployment rate of 13.116% over the entire period, more than 3 points higher than the next-highest unemployment rate, 9.935%, shared by two stores.
- The average store weekly sales figure for all stores is $1,048,563.
- Store 28 had above average weekly sales, with $1,323,522.
- Store 12 had close to average weekly sales, with $1,009,002.
- Store 38 had well below average weekly sales, with $385,732.

This pattern was present throughout the dataset. Stores with high regional unemployment sometimes had good sales, while others with low unemployment had poor sales—but the opposite was true just as often. When we rank the stores by unemployment rate, the trendline for sales shows a slight upward trend. However, because sales vary so much from store to store, this small increase is probably just due to chance. Other factors like weather, population demographics, commuting times, and store quality could all influence sales independently.

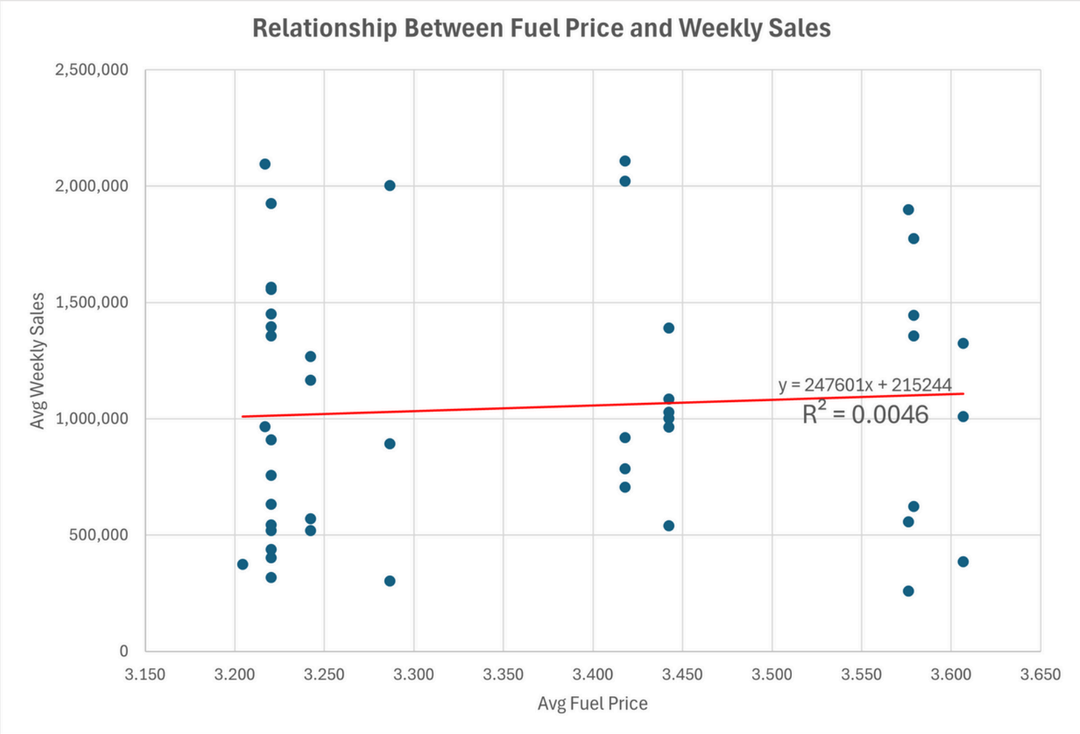
This question seeks to establish a relationship between unemployment data and retail sales trends. There is no obvious correlation between the two, and a detailed examination of data from individual stores bears this out.

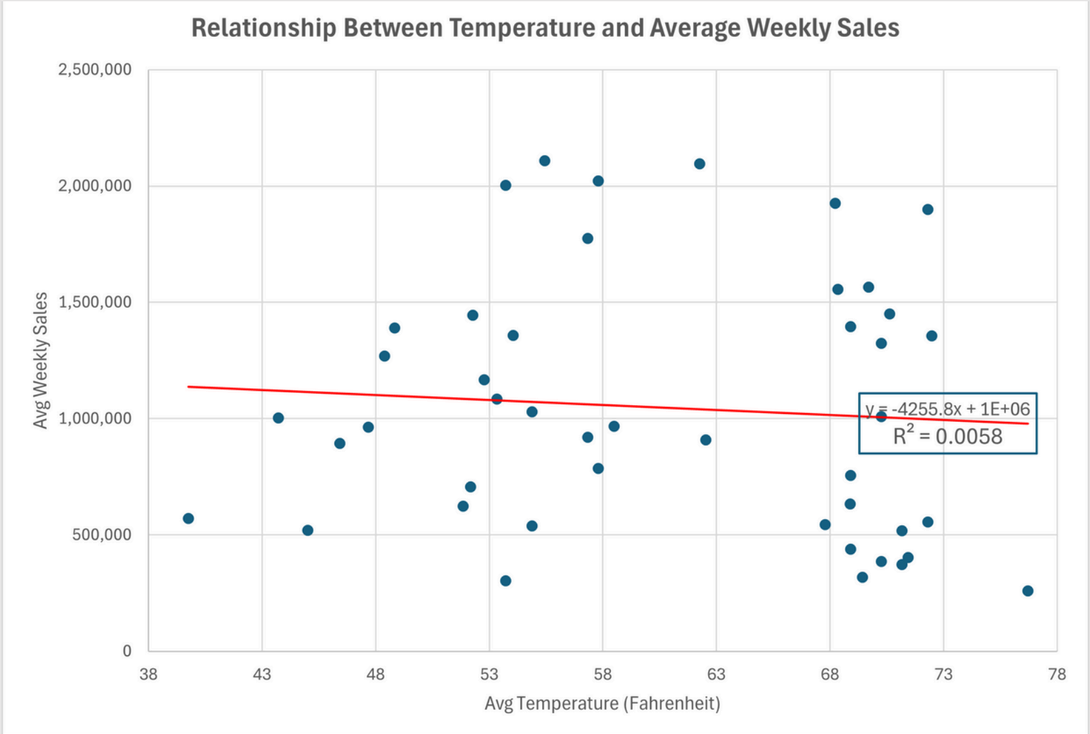
Is there any correlation between CPI and Weekly Sales, and do holidays make a difference?



Weekly sales fluctuate independently of the rising CPI. There's no clear connection between the two, and the variance between stores doesn't follow this trend. We've already established that most holiday weeks aren't much different from non-holiday weeks.

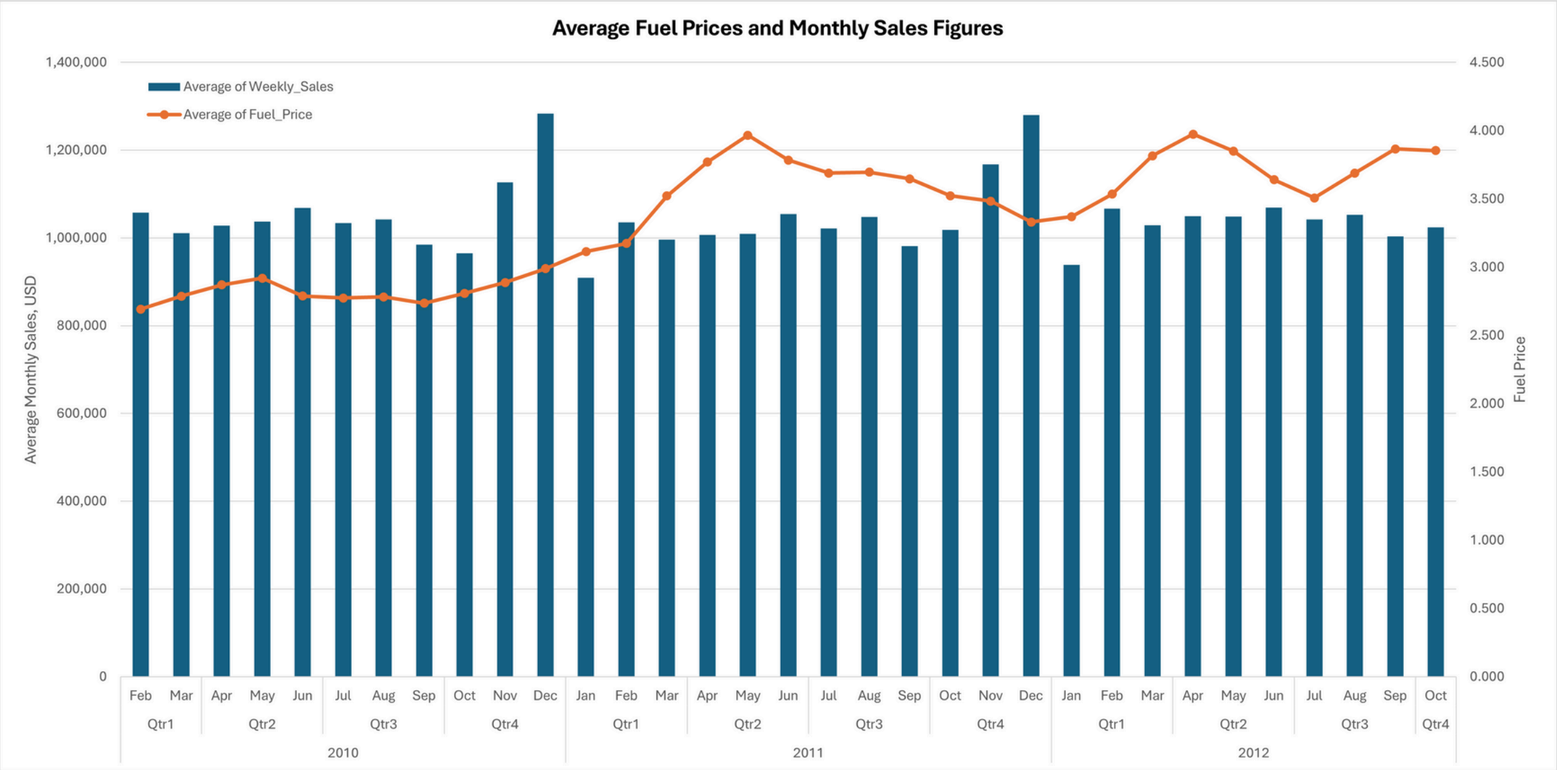
Does Fuel Price Affect Weekly Sales Data?





Regression analysis between fuel price and weekly sales shows no significant correlation ($R^2 = 0.0046$). While fuel price might be relevant in regions where vehicle dependence is high, the data shows wide variations in sales regardless of fuel costs. Similarly, average temperature was found to have minimal impact on sales ($R^2 = 0.0058$). These findings suggest that factors other than fuel price and temperature—like local demand or store-specific characteristics—are more likely driving sales performance.

Does Fuel Price Affect Weekly Sales Data?



This chart compares companywide average monthly sales with average fuel prices over time. Despite a noticeable rise in fuel prices through 2011, monthly sales remain relatively stable, with only seasonal peaks in November and December. The lack of a clear relationship between rising fuel prices and sales further supports the earlier analysis that external factors like fuel costs have minimal impact on overall sales performance.

Group	Store	Avg Fuel Price	Avg Temp	Average Weekly Sales
1	14	3.418	57.804	$ 2,020,978
1	45	3.418	57.804	$ 785,981
2	21	3.220	68.902	$ 756,069
2	30	3.220	68.902	$ 438,580
2	31	3.220	68.902	$ 1,395,901
3	12	3.607	70.252	$ 1,009,002
3	28	3.607	70.252	$ 1,323,522
3	38	3.607	70.252	$ 385,732
4	10	3.576	72.301	$ 1,899,425
4	42	3.576	72.301	$ 556,404

This Excel table categorizes stores into groups based on shared fuel prices and temperatures, highlighting geographic proximity. Each group displays significant variations in weekly sales, underscoring the complex relationship between environmental factors and retail performance. Despite similar local conditions, sales figures vary dramatically within groups, indicating that other local or store-specific factors might play a more substantial role in influencing sales than fuel prices or climate alone.

Conclusions

- Holiday sales analysis: Black Friday and Christmas week drive the highest sales, while other holidays underperform.
- Large chain stores: The data suggests that general merchandise stores like Wal-Mart may not attract as many holiday shoppers as expected, despite offering high-ticket items.
- Store performance: The dataset does not explain why some stores perform better than others, even within the same region.
- Macroeconomic factors: Macroeconomic data doesn't provide insights into store-level sales variations, suggesting other factors, like inventory and staffing, are more influential.
- Focus on internal factors: To improve store performance, internal factors such as inventory fulfillment and human resources should be the focus, rather than broader economic trends.