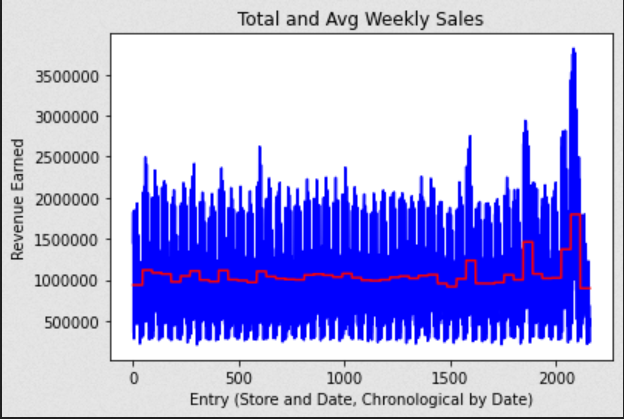
Trends in Retail Sales

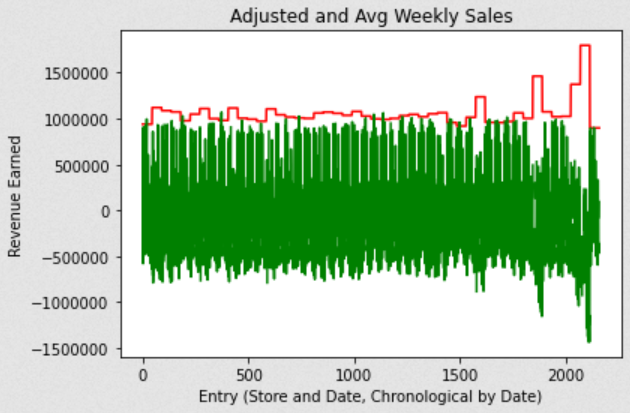


Image Credits to Charles Thompson (Pixabay)

By Andrew, Emerie, and Samantha

For this data science project, our team received data from Kaggle about the weekly retail sales of stores and features of those stores, although there is no further context for which company is being represented and exactly where the data came from. Entries in store features spanned from 2010 to 2013, but weekly sales records only covered the year of 2010, so our sales analysis was restricted to data from 2010. We decided to investigate how three variables affected store sales: fuel prices, unemployment, and whether there was a holiday that week. We would use these findings to see if certain combinations of these could be linked to the best and worst sales weeks.

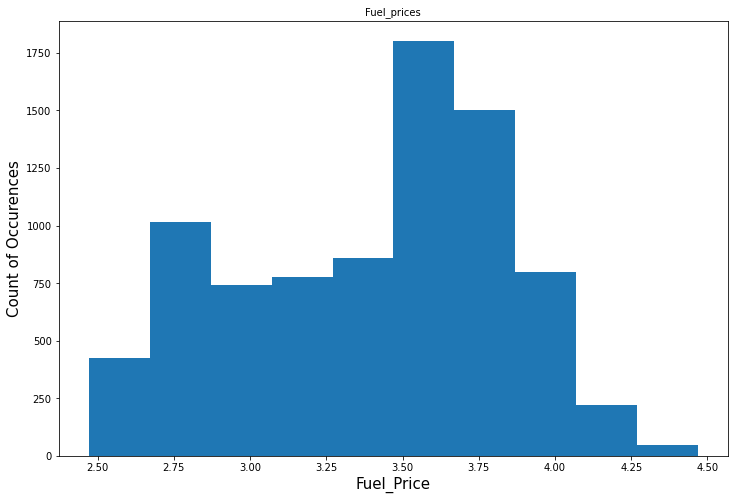




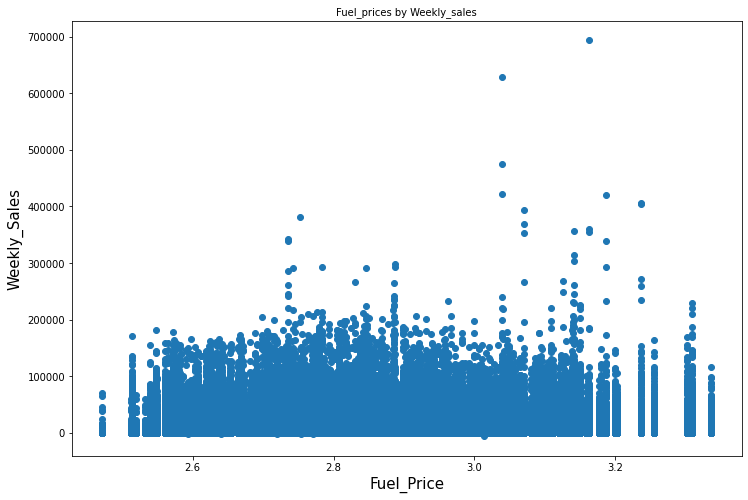
The first step of our analysis was to calculate the average sales per week and subtract this from total weekly sales to get adjusted weekly sales values. These indicate how well a store is doing in a given week when compared to the average. Sales appeared fairly consistent with a few spikes in revenue towards the end of the year. We also noticed that stores could lose revenue in a given week, resulting in negative values.



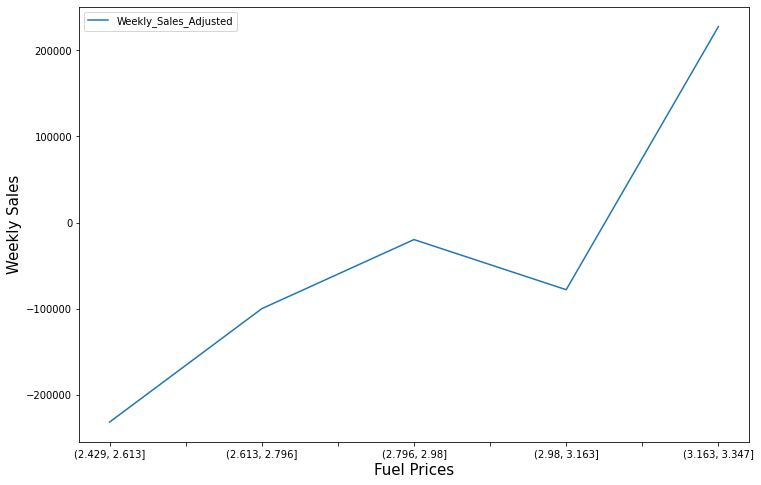
Holidays, represented by an ‘IsHoliday’ boolean, had some unusual data. Weekly records for October 9, November 26, and December 31 were listed as holidays, likely representing Columbus Day, Thanksgiving, and New Year’s Eve. The week containing December 2 was listed as a holiday, though it’s uncertain which one. However, there were separate sales entries for December 3, which was not listed as a holiday. Holidays were associated with a few weeks’ worth of boosted revenue. However, the average sales were very similar between holiday and non-holiday weeks, suggesting that this variable has a slight but not statistically significant impact on weekly sales as a whole.



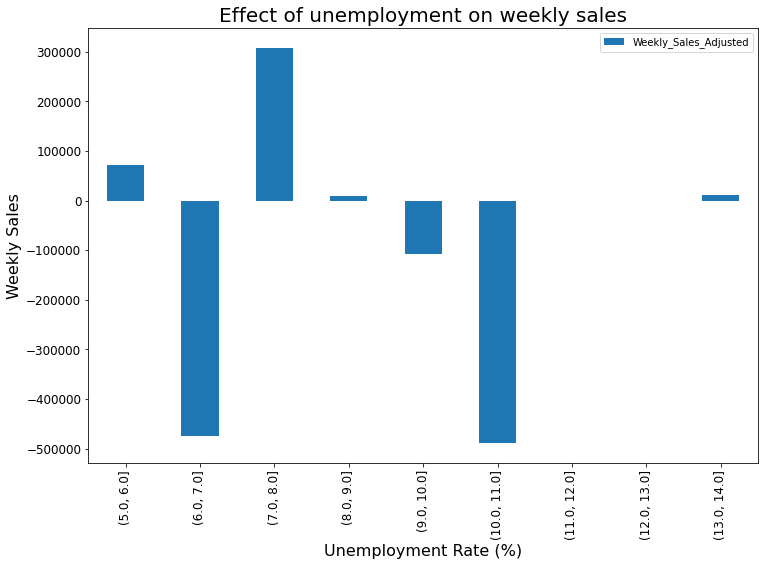
Initially with fuel prices, we just wanted to see the distribution of the fuel prices and see where the majority of the fuel prices were occurring. From the histogram above, we can see that the majority of fuel prices happen around $2.75 as well as between $3.5 and $4. This can be correlated to weekly sales in the next graph in our analysis.



In this scatter plot, we are comparing the fuel price to the weekly sales and trying to see the correlation between these two variables. Comparing it to the previous histogram, we can see that there are some values that aren’t being represented in the scatter plot, that’s due to the lack of aligned weekly sales to fuel prices. But from this scatter plot, we can see that there is no correlation between the two variables since the clusters of points are generally within the same area.

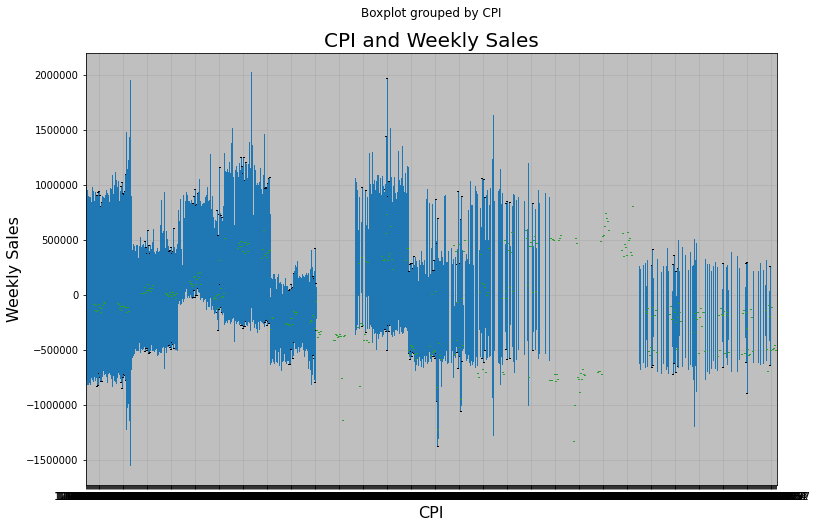


Lastly, we compare the fuel prices to the adjusted weekly sales which shows the difference of weekly sales based on the average weekly sale. And here we can see the trend that increases as the fuel prices increase. Before ~$2.8 we can see that there is a negative trend where there is a loss of weekly sales compared to the average and after, we can see a positive trend where there is a gain of weekly sales.



To explore the potential effects of unemployment rate on weekly sales, we binned the rates so that they may display a clearer depiction of how far above or below weekly sales are relative to the average using the adjusted weekly sales. There appears to be significant variation at each unemployment rate as revealed by the barchart. At the lowest included unemployment rate of 5-6%, weekly sales were about $9,000 above average. When unemployment is increased by 1, weekly sales significantly drop to around $50,000 below average then peaks at 7-8% unemployment. After the peak, sales dropped until 11%. Interestingly, sales increase to above average at a very high unemployment rate of 14% along with adjusted weekly sales and unemployment having a very weak and positive relationship (r^2 = 0.000597). This outlier creates a new aspect of the data to be explored.

There is a general observation that as unemployment rises, consumer spending decreases and debt increases. So why would sales appear to increase at such a high unemployment rate? We hypothesize that inflation could possibly explain this trend.



Our dataset included the consumer price index (CPI) for each week. CPI is used as a measure of inflation by measuring the average change in prices over time (average prices paid). It is expected that as unemployment rises, inflation or CPI should decrease because consumers are spending less and businesses are reducing prices in order to promote sales. Therefore, we should see a negative relationship between unemployment and CPI. However, the relationship between unemployment and CPI was found to be weak and positive (r^2 = 0.124352). Weekly sales appear to remain fairly constant as CPI increases then slightly decreases at high CPI. Overall, there did not appear to be a strong influence from unemployment on weekly sales and weak relationship between unemployment and inflation.

With each variable having a minimal effect on sales, it was difficult to approach finding which combinations of our variables led to best and worst sales weeks. We decided to use an ANOVA test to find potential relationships between our variables and sales. Because ANOVA requires categorical independent variables, we had to ‘bin’ the continuous fuel price and unemployment rate values into categorical groups, for example binning the fuel prices by using a 10 cent increment. The ANOVA did not find statistically significant interactions between our three variables.

Holidays, fuel prices, and unemployment rates did not appear to have statistically significant effects on weekly retail sales. Had we received more information about the stores reporting the data, such as their parent company or the kinds of items sold, we might have been able to offer additional hypotheses about the trends found in the weekly revenue.