When Will They Buy: An examination of retail sales and what drove customers to spend their hard-earned money?

Robert K Burns
DSC 680 - Winter 2019

Project 1 Check In

https://github.com/kirbyeburns/DSC680_Project_1_Retail_Analytics

Any Surprises from the Domain?

I chose this project as a way to get more familiar with trends in retail sales and retail data analysis. There are apparently a million and one opinions on retail, and unfortunately, I'll be reading about some concept I've never heard of and ultimately get sidetracked before realizing it really has nothing to do with the data I'm taking a look at.

For the most part, the data I'm using is pretty straightforward and covers a lot of features besides just sales. It includes the Consumer Price Index, Unemployment, Fuel Prices and even average temperatures. It will be interesting to see if any of these correlate to changes in retail sales.

Data What I Thought it Would Be?

The data is pretty clean with very few "NA"s. I have not worked with time interval data much, but considering how each line of the sales data is by week, this will give me a good opportunity to get my hands dirty with it.

Any Adjustments?

I haven't made any adjustments to my initial query. I'd still like to see how markdowns and holidays can affect sales, but I'm also curious to see how external factors can also drive or hinder sales. It may be difficult to ascertain without some degree of speculation, as I have no information on what store each store number represents. For example, if the temperature has dropped and store 13 has declining sales while store 35 has an increase, it could be because store 13 is Ron Jon swimwear and store 35 is Burlington Coat Factory….just speculating that outside factors may affect sales based more on the products of those stores than the outside factor, themselves.

I may try to run individual analysis on single stores, just to see if outside factors have an overall affect or if they could be isolated to changes at specific stores.

Method Working?

So far, I've just used Python (Jupyter Lab) to explore the data. I plan to continue that, but if time permits, I want to try and create a nice Power BI dashboard of this information, as well. If I don't have time to do it for this project, I'll pursue it for future projects.

Challenges?

So far, the biggest challenge I've encountered is trying to sort through what is and is not important. As I mentioned above, I don't know the names of these stores or what they sell, so I think that any outside factors may affect individual stores more so than overall retail sales. In this case, I feel like I'll be doing more analysis than originally thought.

Concluding Remarks

I'm still interested to see if I can find any correlations between outside factors and retail sales. It would be more work to perform analysis on EVERY store, so I'm going to pursue perhaps the top 5 or 10 stores in terms of overall sales and see where that gets me.