When Will Customer Buy: An Evaluation of sales and what takes out money from customers pocket?

Rahul Gupta - DSC 680 - Winter 2020

https://github.com/rahulgupta271/DSC680_Project_1_Market_analytics_holiday_trend

Project 1 Check In

Any Surprises from the Domain?

As a way to get more acquainted with developments in retail sales and retail data analysis, I chose this project. Apparently, there are a million and one views on retail, and sadly, before realizing it really has nothing to do with the details, I'm taking a look at, I'll be reading about some idea I've never heard of and eventually get sidetracked.

The data I'm using is fairly straightforward for the most part and covers a lot of characteristics besides just sales. The consumer price index, unemployment, gasoline prices and even normal temperatures are included. If either of these correlates to shifts in retail sales, it will be important to see.

Data What I Thought it Would Be?

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

Any Adjustments?

I made no changes to my original query. I'd also like to see how sales can be influenced by markdowns and holidays, but I'm still interested to see how external factors can drive or impede sales as well. Without any degree of speculation, it may be hard to determine, since I have no knowledge about what store each store number represents. For instance, if the temperature has fallen and store 13 has declining sales while store 35 has a rise, it may be because store 13 is Ron Jon swimwear and store 35 is Burlington Coat Factory.... Only speculating that external factors could influence sales based more on those stores' products than the external factor itself.

I will try to conduct individual research on single stores, just to see whether external variables have an aggregate effect or whether improvements in particular stores could be isolated.

Method Working?

I've only been using Python (Jupyter Lab) to explore the data so far. I intend to continue that, but I want to try to build a nice Power BI dashboard of this information if time permits, as well. I will pursue it for future projects if I don't have time to do it for this one.

Challenges?

So far, trying to sort out what is and is not important is the biggest challenge I've encountered. I don't know the names of these stores or what they sell, as I described above, so I think that any external variables can influence individual stores more than aggregate retail sales. I feel like I'm going to do more research in this case than initially expected.

Concluding Remarks

I'm also curious in seeing if I can find any connections between external variables and retail sales. Doing research from Every store will be more work, so maybe I'm going to try the top 5 or 10 stores in terms of total revenue and see where that takes me.