

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

[https://github.com/kirbyburns/DSC680 Project 1 Retail Analytics](https://github.com/kirbyburns/DSC680)

Which Domain?

This project will focus on retail sales trends and the correlating features that affect the changes. I have no experience in the retail industry, much less in retail analytics, but I'm interested in how merchants determine when, aside from the typical Black Friday/Cyber Monday sales, to put which items on sale.

I am using the following references to begin this project:

Nicasio, Francesca. August 20, 2019. "Retail Analytics: How to Use Data to Win More Sales and Customers." <https://www.vendhq.com/blog/how-retailers-can-use-data-to-boost-productivity-customer-service-sales/>

Bobriakov, Igor. July 22, 2018. "Top 10 Data Science Use Cases in Retail." <https://medium.com/activewizards-machine-learning-company/top-10-data-science-use-cases-in-retail-6483accc6042>

Alloy Client Solutions. March 21, 2019. "Retail analytics trends in 2019 and beyond." <https://medium.com/alloytech/retail-analytics-trends-in-2019-and-beyond-c671783c67b7>

Jones, Joshua. March 2, 2018. "Fast Forward: How retailers will use data and analytics to succeed in 2018 and beyond." <https://www.strategywise.com/fast-forward-how-retailers-will-use-data-and-analytics-to-succeed-in-2018-and-beyond/>

Hufford, Jillian. June 3, 2019. "How to Perform Sales Trend Analysis For Your Retail Business." <https://www.nchannel.com/blog/how-to-perform-sales-trend-analysis/>

Prevedere.com. N.D. "10 Mistakes to Avoid to Better Leverage Retail Analytics and Forecasting." <https://www.prevedere.com/leveraging-retail-analytics-and-forecasting/>

Sisene.com. N.D. "What is Retail Analytics?" <https://www.sisense.com/glossary/retail-analytics/>

Chatterjee, Sharmistha. May 27, 2019. "Traditional vs Deep Learning Algorithms in Retail Industry." <https://towardsdatascience.com/traditional-vs-deep-learning-algorithms-in-retail-industry-i-b7b7f86793d4>

Maitra, Sarit. October 18, 2019. "Clustering & Machine Learning Combination in Sales Prediction." <https://towardsdatascience.com/clustering-machine-learning-combination-in-sales-prediction-330a7a205102>

Bosier, Fabian. November 12, 2019. "Pandas' groupby explained in detail." <https://towardsdatascience.com/pandas-groupby-aggregate-transform-filter-c95ba3444bbb>

Which Data?

The dataset I will be using was uploaded to Kaggle.com by Manjeet Singh two years ago and is found at: <https://www.kaggle.com/manjeetsingh/retaildataset#Features%20data%20set.csv>

This dataset contains weekly sales totals from 45 different stores from February 2010 to October 2012. These data include indicators as to whether or not a particular department had noteworthy markdowns or sales leading up to and during the week indicated.

Research Questions? Benefits? Why analyze these data?

My primary research question is, “What drives sales?” Drilling down, I will try to find specific reasons behind sales highs and lows between different stores and between different departments within those stores.

The benefit to this study is in gaining insight as to when a store’s customer base is likely to make a purchase and what things, if any, a store can do to influence those purchases. Obviously, having the ability to influence sales has one MAJOR benefit: increased profits.

Initially, I will analyze the data visually, to spot any obvious trends and to see where each department flourished. Then I will attempt to identify correlating features that affect sales. I will most likely try heatmaps and confusion matrices to see how the features correlate to one another.

My interest in this domain is in speaking with a friend who works for an analytics firm primarily focused on retail analytics.

What Method?

At first, I’ll be using R and/or Python for visualizations. I will most likely try to create visualizations in Power BI, as well, as many companies utilize Power BI and are looking for analysts who are fluent in it. I will most likely use R for finding correlations, as I have always found it easier to use than Python.

Potential Issues?

For me, one of the biggest challenges I see would be if I could not find any real significant correlations, as my hypothesis would be that there are likely several correlating features. This will go off track if I focus on one thing more than the others and lose sight of time management.

Concluding Remarks

I’m interested in getting a job in analytics, and since retail analytics seems to be a booming focus in the realm of data science, I wanted to tackle this domain to get some experience understanding what stores can do to influence customers to buy from them.