Capstone 2: Customer Purchasing Behaviors

Problem: Customer purchasing optimization is a common problem for a lot of companies, both large and small. Luckily most large companies should have a wealth of data available to them in order to predict the possible purchasing behavior for many of their clients so they can make informed decisions about stocking their shelves, providing customer recommendations on items, and predicting when a customer is losing interest in their service and planning to buy from a competitor. My proposal is to analyse and demonstrate some of these metrics on public data found on the internet as well as provide an algorithm that can predict sales in the future.

Client: Walmart is looking to hire a Data Scientist to analyze their data and design tools for their grocery store data in order to help make their products as visible and convenient as possible to their customers. Part of this will require finding items that are commonly bought together and be able to stock those items in the same vicinity as each other. They are also in having some kind of forecasting method to predict their sales in the future so they can help push items when people are most likely to buy.

Data: Inspired by the paper, Customer Shopping Pattern Prediction:A Recurrent Neural Network Approach by Hojjat Salehinejad and Shahryar Rahnamayan, they used a Recurrent Neural Network to predict shopping patterns using the Ta Feng Grocery Dataset. After doing some internet searching I managed to find a copy or sample of the Ta Feng Grocery dataset hosted on Kaggle.

Approach: I intend to do basic exploratory analysis, followed by clustering customers into demographic groups using Kmeans clustering, and analyzing some of those groups, provide a few metrics such as propensity, affinity, customer lifetime value (CLV), and possibly customer churn or attrition. Finally I plan on finishing the project with trying to apply a Recurrent Neural Network to predict sales using Keras.

Deliverables: For deliverables I intend to write a blog post, provide code on GitHub, and maybe have a video if time permits.