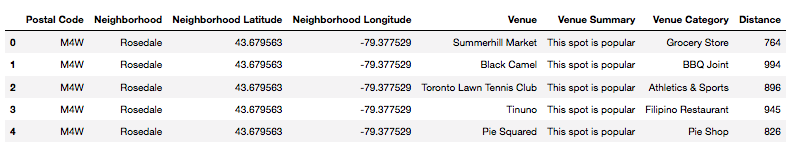
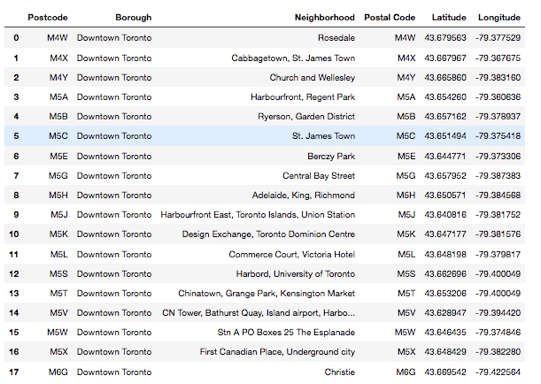
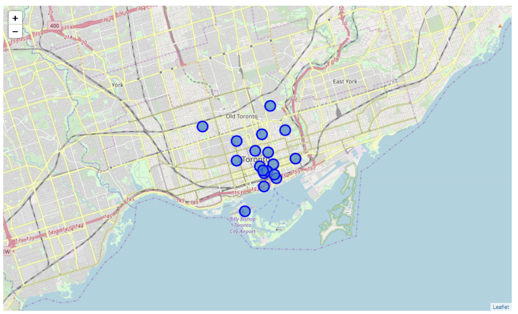
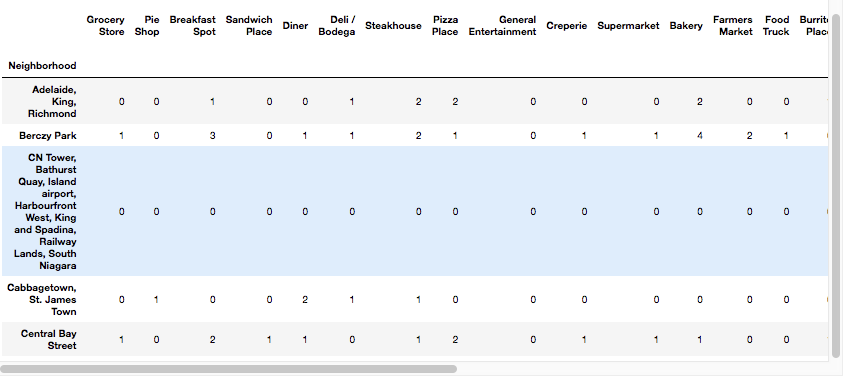
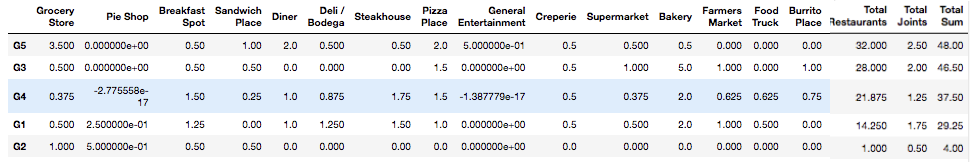
**A RECOMMENDER SYSTEM FOR GROCERIES BUSINESS IN DOWNTOWN TORONTO**

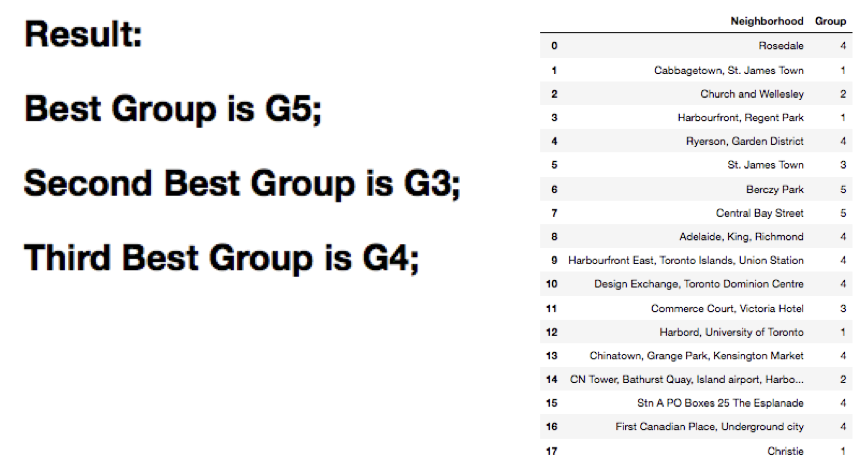
1. **Introduction** 
   1. **Business Problem:** There is a groceries contractor in one of the boroughs of Toronto (Downtown Toronto). This contractor provides places such as: Different types of Restaurants, Bakery, Breakfast Spot, and Café with fresh and high-quality groceries. The contractor wants to build a grocery markets it buys from villagers and farmers inside the borough, so that they will support more customers and also bring better "Quality of Service" to the old customers.
2. **Data acquisition and cleaning** 
   1. We will need geo-locational information about that specific borough and the neighborhoods in that borough. We assume it is “Downtown Toronto" in Toronto. This is easily provided for us by the contractor, because the contractor has already made up his mind about the borough.
   2. We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. A typical request from Foursquare will provide us with the following information:



1. **Methodology**
   1. Identifying Postal Codes (and then Neighborhoods) in “Downtown Toronto” 
   2. Drawing map 
   3. Connecting to Foursquare and Retrieving Locational Data for Each Venue in Every Neighborhood
      1. After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 1000 meter. It means that we have asked Foursquare to find venues that are at most 1000 meter far from the center of the neighborhood.
   4. Processing the Retrieved Data and Creating a DataFrame for All the Venues inside the Downtown Toronto
      1. When the data is completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue. After this stage, the column "Venue's Category" wil be One-hot encoded and different venues will have different feature-columns. After On-hot encoding we will integrate all restaurant columns to one column "Total Restaurants" and all food joint columns to "Total Joints" column. 
   5. Complete data for running K-means clustering



1. **Results**
   1. Now, we focus on the centers of clusters and compare them for their "Total Restaurants" and their "Total Joints". The group which its center has the highest "Total Sum" will be our best recommendation to the contractor. {Note: Total Sum = Total Restaurants + Total Joints.} This algorithm although is pretty straightforward yet is strongly powerful.



1. **Conclusion**: In this study, I analyzed for recommendation system for grocery business contractors in Toronto area. I identified Group 5 cluster such as Berczy Park and Central Bay Street neighborhoods is the best area to open new grocery markets. I built K-means clustering models to predict the best area to open grocery markets.