1. **Introduction**

The main problem in my project is finding an ideal neighborhood in Scarborough, Toronto to open up a new restaurant. We must take into consideration many different aspects such as price of rent/owning in all neighborhoods, how populated each neighborhood is and how many people visit it. The target audience is, therefore, anyone who is looking to open up a new restaurant in Scarborough and wants to assess all the different aspects that would affect her/him.

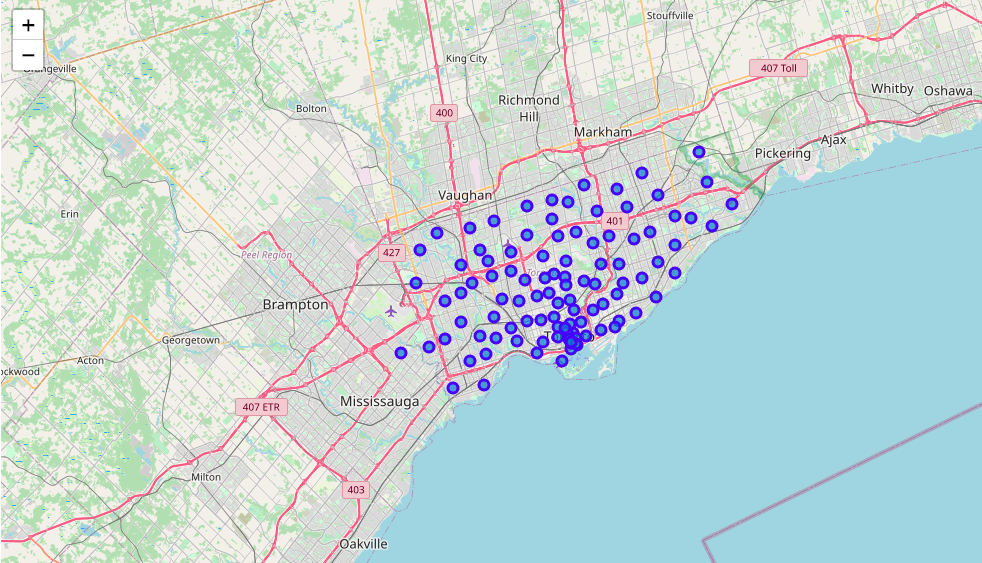
1. **Data Acquisition and Cleaning**

I needed two main sources of data for this project. In the beginning, I needed to get a hold of all the different postal codes in Scarborough, each postal code representing a neighborhood. In order to do so, I got a list of all the postal codes available in the boroughs around the Toronto area and cleaned it out to have the Scarborough codes left on their own. Then, I retrieved the detailed data for each possible venue (availability, popularity and numbers of restaurants and food services available in each area) from Foursquare. I created my own account with my own Client ID and Client Secret and started searching.

1. **Data Analysis**

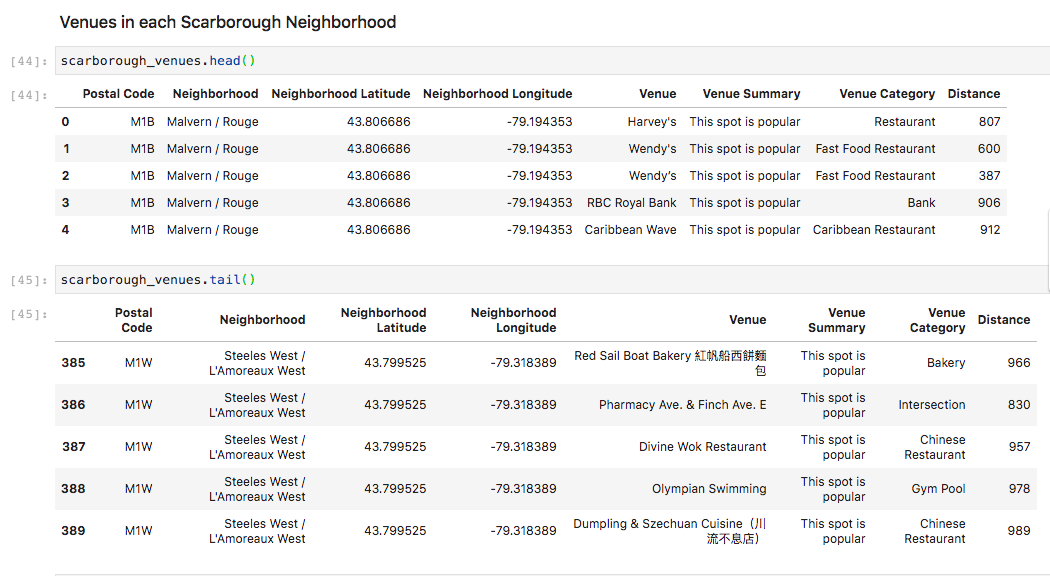
After creating the data frame that includes all postal codes for all areas in Scarborough, I created two maps: One map showing the greater area of Toronto and the other zooming in on the Scarborough area specifically.

Toronto Map:

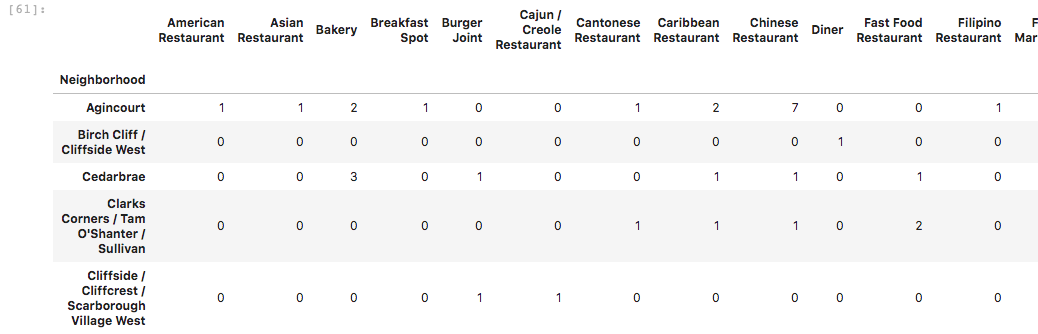


Scarborough Map:

After that, I searched Foursquare to be able to look into each neighborhood that I have retrieved to be a part of Scarborough. That will give me many results, the first being all the names of the venues. This will allow me to create a dataframe in Python that includes a list of venues, their names, locations, categories, whether the spot is popular or not and of course, our initial data: the postal code.



After that, I create a list of features that includes all the different types of food servers in the area, such as restaurants, grocery stores, bakeries and so on. I do this in order to create another data frame that includes a count of each of these food servers for each neighborhood.



1. **Result**

Based on the analysis above and after conducting a K-means examination, I was able to conclude that Neighborhood G1 is the best one as it is the most popular neighborhood with the larger number of visitors and restaurant joints. I based my analysis on the 5 centers that I chose for each cluster and included a count for total joints and total sum.

