Data Section

Scarborough is a region for new immigrants in Canada to live. Scarborough is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship.

Data Source : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

For this project we make use of the following:

Foursquare API:

Four-square API would be used in this project as a gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

Libraries Which are Used to Developed the Project:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data. Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module