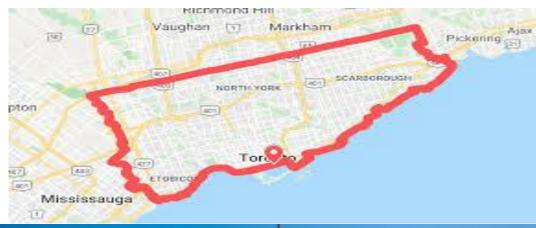# Capstone Project - Battle of the Neighborhoods

<mark>Introduction</mark> :

This project will analyze neighborhoods in Toronto. One of company is looking move its headquarters to Toronto. The company wants insight into the neighborhoods and local businesses in the cities so that its employees may have the optimum living standards and quality of life. This project will explore neighborhoods in Toronto, and determine which neighborhoods best fit the culture of the Fortune 500 company's employees.
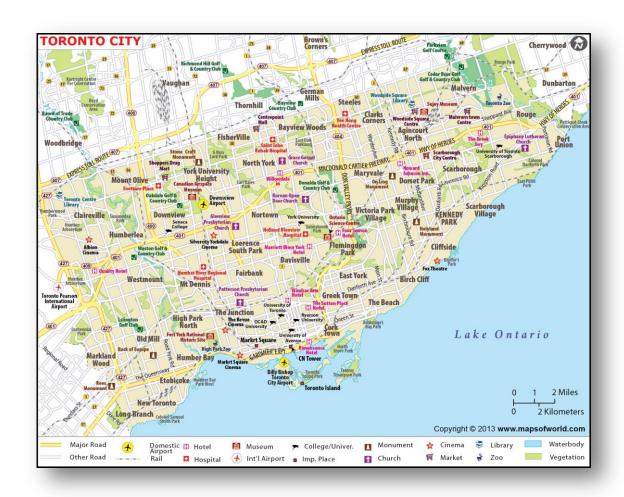
Data:

Data The data used for this project will be acquired from the respective cities Wikipedia website pages. The datasets consists of the postal codes, neighborhood names, latitude, and longitude information for each neighborhood. Foursquare API search feature will be used to collect neighborhood venue information. Details about local venues and locality will be provide insight into the qualities of a neighborhood. In addition to Foursquare, various python packages will be used to create maps and machine learning models to further provide insights into our neighborhood battle project.

I used the following datasets from these websites:
Toronto Neighborhoods
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. Toronto
Latitude and Longitude - http://cocl.us/Geospatial_data

<mark>Methodology</mark> :

1. HTTP requests would be made to this Foursquare API server using zip codes of the Seattle city neighborhoods to pull the location information (Latitude and Longitude).
2. Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods. 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.
3. Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Seattle city over an interactive leaflet map.
4. Extensive comparative analysis of two randomly picked neighborhoods world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

The following are the Python packages I used:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

Results:

I use k-means to group the neighborhoods in Scarborough into 3 clusters. Cluster_0 has 3 neighborhoods and the most common venues are skating rinks, international cuisine restaurants and breakfast spots. Cluster 1 has 1 neighborhood 1 neighborhood , and the most common venues are pizza place and noodle house. Cluster 2 has 38 neighborhood, and the most common venues are Chinese restaurants and discount stores.

Discussion:

Toronto has 11 boroughs and 103 neighborhoods. The geographical coordinate of Toronto, Canada are 43.7170226, -79.4197830350134. In Scarborough borough, found 85 venues in 17 neighborhoods In Scarborough borough, the neighborhoods with the most venues are L'Amoreaux West and Steeles West. There are 79 distinct venues in 50 categories.

Conclusion:

In conclusion, based on the quantity of venues and variety of venues, I would choose a good place to the company.