# Capstone Project - Battle of the Neighborhoods

**Introduction**

This project will analyze neighborhoods between Toronto, Canada and New York City, New York. A Fortune 500 company is looking to move its headquarters to either Toronto or New York City. 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 the similarities and dissimilarities between certain neighborhoods in the two cities, and determine which neighborhoods best fit the culture of the Fortune 500 company’s employees.



# Data

The data used for this project will be acquired from the respective cities Wikipedia website pages. The datasets consist 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.

Datasets from the following websites were used for the analysis:

Toronto Neighborhoods - https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M. Toronto Latitude and Longitude - <http://cocl.us/Geospatial_data>

New York City neighborhoods - https://geo.nyu.edu/catalog/nyu\_2451\_34572

New York City Latitude and Longitude = Python Geolibrar

## Work Flow:

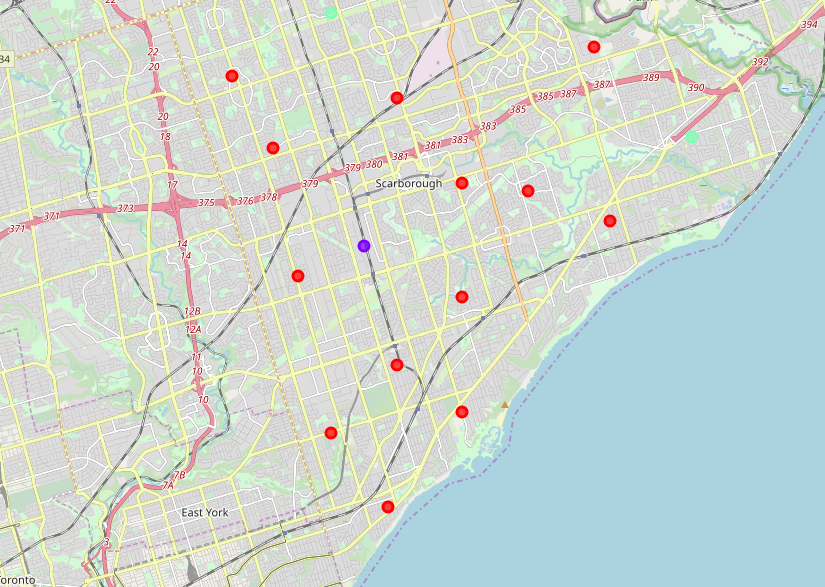
1. HTTP requests will be made to this Foursquare API server using zip codes of the Seattle city neighborhoods to pull the location information (Latitude and Longitude).
2. Foursquare API search feature will be enabled to collect the nearby places of the neighborhoods. Due to http request limitations 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 will 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 will be carried out to derive the desirable insights from the outcomes using python’s scientific libraries Pandas, NumPy and Scikit-learn.
5. Unsupervised machine learning algorithm K-mean clustering will be applied to form the clusters of diﬀerent categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods will be analyzed individually collectively and comparatively to derive the conclusions.

## The following Python packages were 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

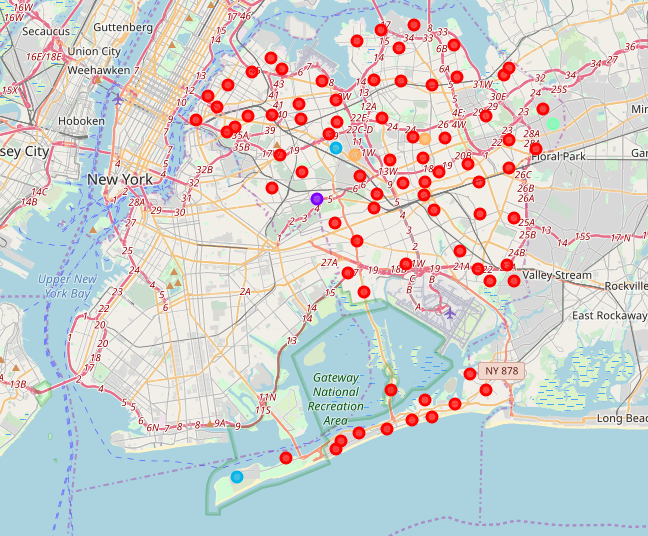
## Scarborough Borough in Toronto, Canada

K-means was used to group the neighborhoods in Scarborough into 3 clusters. Cluster\_0 has 13 neighborhoods and the most common venues are international cuisine restaurants and breakfast spots. Cluster 1 has 1 neighborhood, and the most common venues are international cuisine restaurants. Cluster 2 has 2 neighborhoods, and the most common venues are eating places and discount stores.



## Queens Borough in New York City

K-means was used to group the Queens borough into 5 clusters. Cluster\_0 has 75 neighborhoods and consist of many international cuisine restaurants and grocery stores. The most common venues are pizza places, deli, and Chinese restaurants. Cluster\_1 has 1 neighborhood and the most common venue is Food & Drink shop. Cluster\_2 has 2 neighborhoods and the most common venue are bakeries and landmarks. Cluster\_3 has 1 neighborhood and the most common venues are international cuisine restaurants. Cluster\_4 has 2 neighborhoods and the most common venues are gyms and delis.



Toronto has 11 boroughs and 103 neighborhoods. The geographical coordinates of Toronto, Canada are 43.653963, -79.387207. In Scarborough borough, found 87 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 51 categories.

New York City has 5 boroughs and 306 neighborhoods. The geographical coordinate of New York City are 40.7127281, -74.0060152. Foursquare found 1744 distinct venues in 81 neighborhoods in Queens borough.

Many of the neighborhoods are homogenous and are very similar to each other. Both Scarborough and Queens borough consist of neighborhood cluster that contain majority of the neighborhoods, and the remaining cluster had 1-5 neighborhoods. Queens borough had a significant greater number of neighborhoods and venues than Scarborough.

# Conclusion

In conclusion, based on the quantity of venues and variety of venues, Queens would be a better choice over Scarborough to relocate the headquarters of the Fortune 500 company. Queens oﬀer way more in choices for restaurants, gyms, grocery stores, and extracurricular activities for individuals and families of the company’s employees.