Introduction

This project will analyze neighborhoods between Toronto, Canada and New York City, New York. A software company is going to open its headquarters in Toronto or New York City. The company wants its employees may have excellent living standards and quality of life. This job will explore the dissimilarities between two neighborhoods in the two cities, and determine which community is the best fit for the companies 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 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 provide insights into our neighborhood battle project further.

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

New York City neighborhoods - https://geo.nyu.edu/catalog/nyu_2451_34572

New York City Latitude and Longitude = Python Geolibrar

Methodology

Work Flow:

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 spots 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 is 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 would 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 would be analyzed individually collectively and comparatively to derive the conclusions.

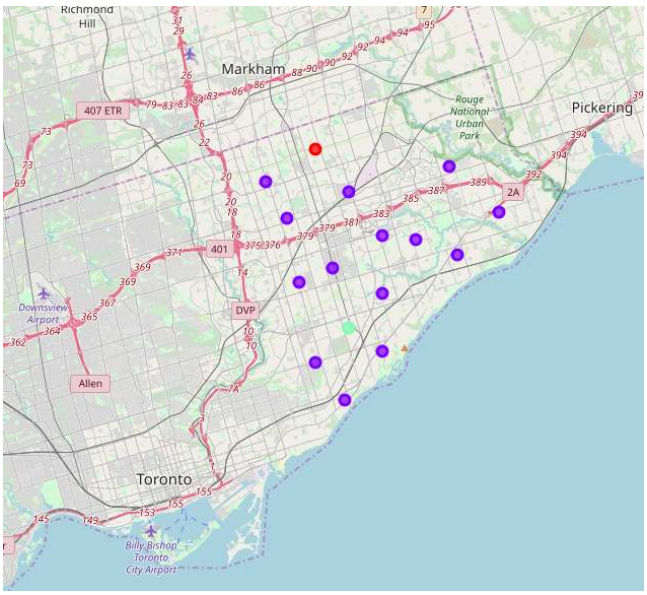
The following are the Python packages I used:

- 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

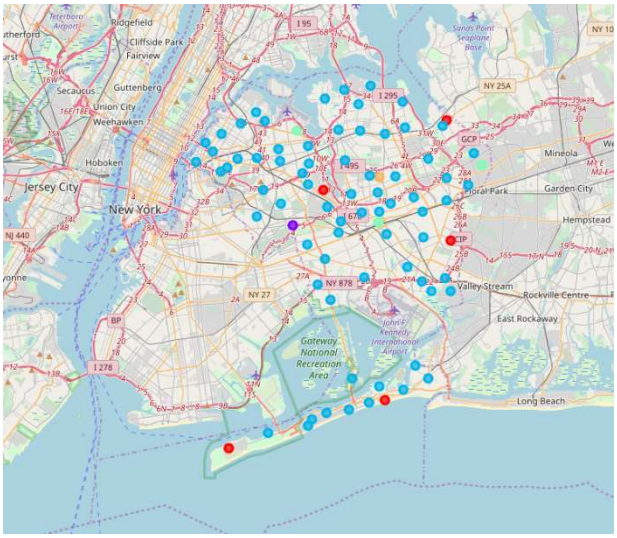
Scarborough Borough in Toronto, Canada

I use k-means to group the neighborhoods in Scarborough into 3 clusters. Cluster_0 has 15 neighborhoods and the most popular venues are skating rinks, international cuisine restaurants and breakfast spots. Cluster 1 has one neighborhood, and the most popular venues are pizza place and noodle house. Cluster 2 has one neighborhood, and the most popular venues are Chinese restaurants and discount stores.



Queens Borough in New York City

I used k-means to group the Queens borough into 5 clusters. Cluster_0 has 81 neighborhoods and consist of many international cuisine restaurants and grocery stores. The most popular venues are pizza places, deli, and Chinese restaurants. Cluster_1 has one neighborhood and the most popular venue is a dance studio. Cluster_2 has five neighborhoods and the most popular venue are donut shops and international cuisine restaurants. Cluster_3 has two neighborhoods and the most popular venues are the beach and a bakery. Cluster_4 has two neighborhoods and the most popular venues are gyms and donut shops.



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, the neighborhoods with the most venues are L'Amoreaux West and Steeles West. There are 79 distinct venues in 50 categories. New York City has five boroughs and 306 neighborhoods. The geographical coordinate of New York City is 40.7308619, -73.9871558. Foursquare found 2108 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 consists of neighborhood cluster that contains the majority of the neighborhoods, and the remaining cluster had 1-5 neighborhoods. Queens borough had a significant number of neighborhoods and venues than Scarborough.

Conclusion

In conclusion, based on the number of venues and a variety of venues, I would choose Queens over Scarborough as a choice to relocate the headquarters of the Software company. Queens offers way more in options for restaurants, gyms, grocery stores, and extracurricular activities for individuals and families of the company's employees.