

Capstone Project - Battle of the Neighbourhoods

Introduction

This project will analyse neighbourhoods between Toronto, Canada and New York City, New York. A Corporation is looking to move its headquarters to either Toronto or New York City. The company wants insight into the neighbourhoods 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 neighbourhoods in the two cities, and determine which neighbourhoods best fit the culture of the Corporation's employees.

Data

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

I used the following datasets from these websites:
Toronto Neighbourhoods

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

Toronto Latitude and Longitude

http://cocl.us/Geospatial_data
New York City neighbourhoods

https://geo.nyu.edu/catalog/nyu_2451_34572 New York City Latitude and Longitude = Python Geolibrar

Methodology Work Flow

- HTTP requests would be made to this Foursquare API server using zip codes of the Seattle city neighbourhoods to pull the location information (Latitude and Longitude).
- Foursquare API search feature would be enabled to collect the nearby places of the neighbourhoods. Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.
- Folium- Python visualisation library would be used to visualise the neighbourhoods cluster distribution of Seattle city over an interactive leaflet map.
- Extensive comparative analysis of two randomly picked neighbourhoods would be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.
- Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighbourhoods. These clusters from each of those two chosen neighbourhoods would be analysed 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 vectorised 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

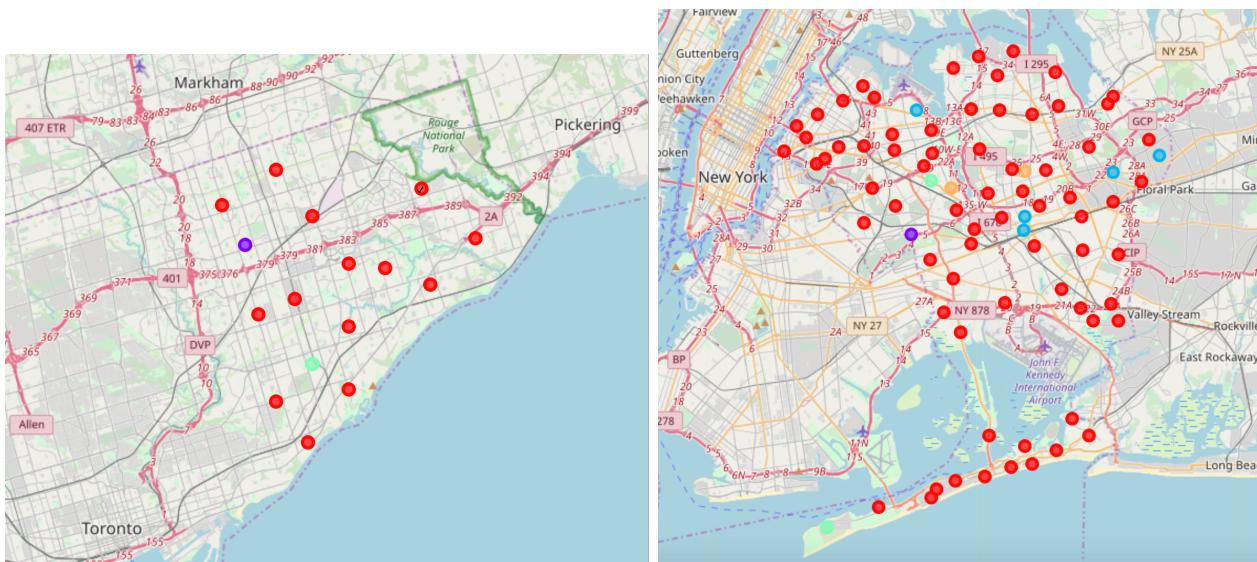
Results

Scarborough Borough in Toronto, Canada

I use k-means to group the neighbourhoods in Scarborough into 3 clusters. Cluster_0 has 15 neighbourhoods and the most common venues are skating rinks, international cuisine restaurants and breakfast spots. Cluster 1 has 1 neighbourhood 1 neighbourhood , and the most common venues are pizza place and noodle house. Cluster 2 has 1 neighbourhood, and the most common 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 neighbourhoods 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 neighbourhood and the most common venue is a dance studio. Cluster_2 has 5 neighbourhoods and the most common venue are donut shops and international cuisine restaurants. Cluster_3 has 2 neighbourhoods and the most common venues are the beach and a bakery. Cluster_4 has 2 neighbourhoods and the most common venues are gyms and donut shops.



Discussion

Toronto has 11 boroughs and 103 neighbourhoods. The geographical coordinate of Toronto, Canada are 43.7170226, -79.4197830350134. In Scarborough borough, found 85 venues in 17 neighbourhoods

In Scarborough borough, the neighbourhoods with the most venues are L'Amoreaux West and Steeles West. There are 79 distinct venues in 50 categories.

New York City has 5 boroughs and 306 neighbourhoods. The geographical coordinate of New York City are 40.7308619, -73.9871558. Foursquare found 2108 venues in 81 neighbourhoods in Queens borough.

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

Conclusion

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