Capstone Project - Battle of the Neighborhoods

Introduction

This project analyzes neighborhoods between Toronto, Canada vs. New York City, New York. The goal of this project is to find a city to open a restaurant, by analyzing the neighborhoods in the cities. This project will explore the similarities and differences between certain neighborhoods in the two cities, and determine which neighborhoods is a better option to open a restaurant.

Data

The data used in this project will be obtained from the cities' Wikipedia website pages. The datasets consists of information such as the postal codes, neighborhood names, borough, latitude, and longitude information for each neighborhood. Foursquare API search feature will also be used to collect neighborhood venue information. In addition, details about local venues and locality will be provided to analyze the information of a neighborhood. Besides Foursquare, various python packages will be used to create maps and machine learning models to further provide analysis into our neighborhood battle project.

The following datasets are used in this report:

- NYC Neighborhoods - https://geo.nyu.edu/catalog/nyu_2451_34572
- New York City Latitude and Longitude = Python Geolibrar
- Toronto Neighborhoods - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- Toronto Latitude and Longitude - http://cocl.us/Geospatial_data

Methodology – please see below the steps

1. Import all the necessary Python packages.
2. Read the data from the Wikipedia page and clean the data.
3. Use Foursquare API server to pull the location information (Latitude and Longitude).
4. Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods.
5. Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of each city over an interactive leaflet map.
6. 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.
7. 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:

1) Pandas - Library for Data Analysis
2) NumPy – Library to handle data in a vectorized manner
3) JSON – Library to handle JSON files

4) Geopy – To retrieve Location Data
5) Requests – Library to handle http requests
6) Matplotlib – Python Plotting Module
7) Sklearn – Python machine learning Library
8) Folium – Map rendering Library
9) BeautifulSoup
10) Lxml

Results

The data of Toronto, Canada is used via k-means to group the neighborhoods into 3 clusters. Cluster_0 has 15 neighborhoods and the most common venues are coffee shop and breakfast spots. Cluster 1 has 1 neighborhood, and the most common venues are pizza place and noodle house. Cluster 2 has 1 neighborhood, and the most common venues are Italian restaurants and discount stores.

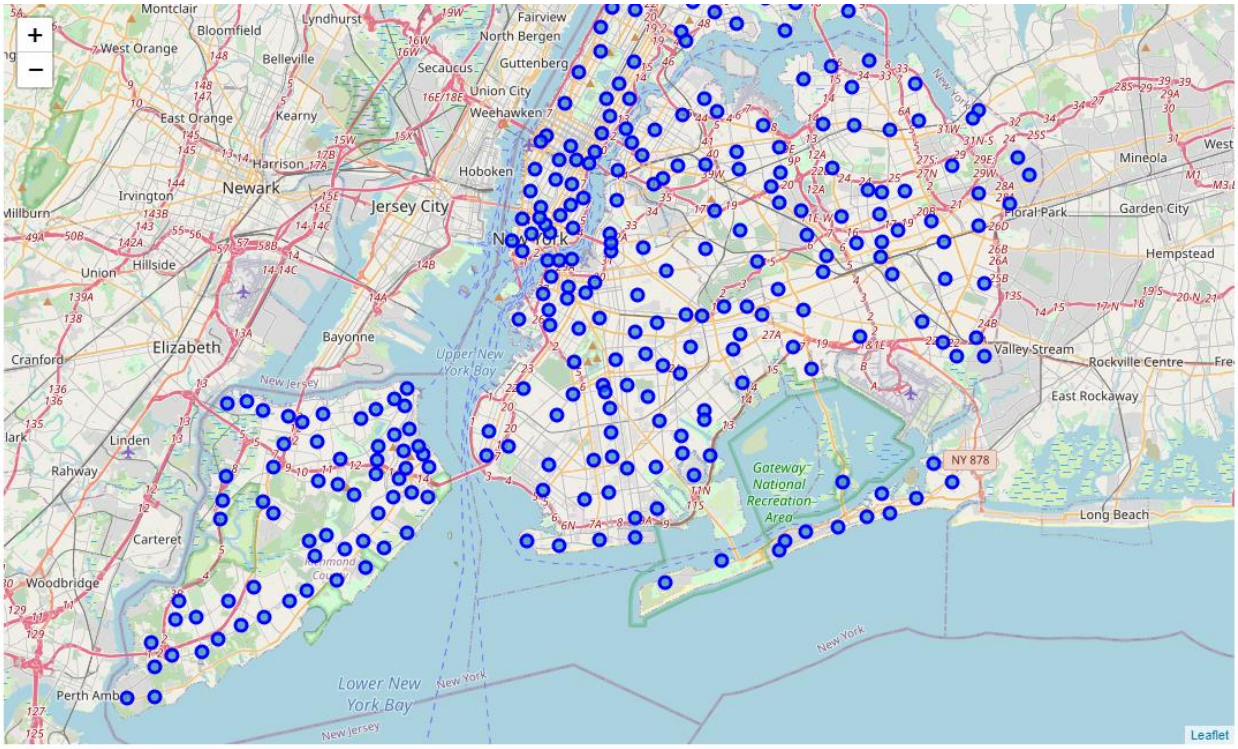
The data of Queens, New York City is used via k-means to group the neighborhoods into 5 clusters. Cluster_0 has 14 neighborhoods and consist of many Pizza Place, Hotel, Hotpot Restaurant, Deli, Etc. Cluster_1 has 1 neighborhood and the most common venue is a Deli. Cluster_2 has 4 neighborhoods and the most common venue are donut shops, restaurant, clothing store and Indian restaurant. Cluster_3 has 1 neighborhood and the most common venue is a gym. Cluster_4 has 61 neighborhoods and the most common venues are bars and grocery stores.

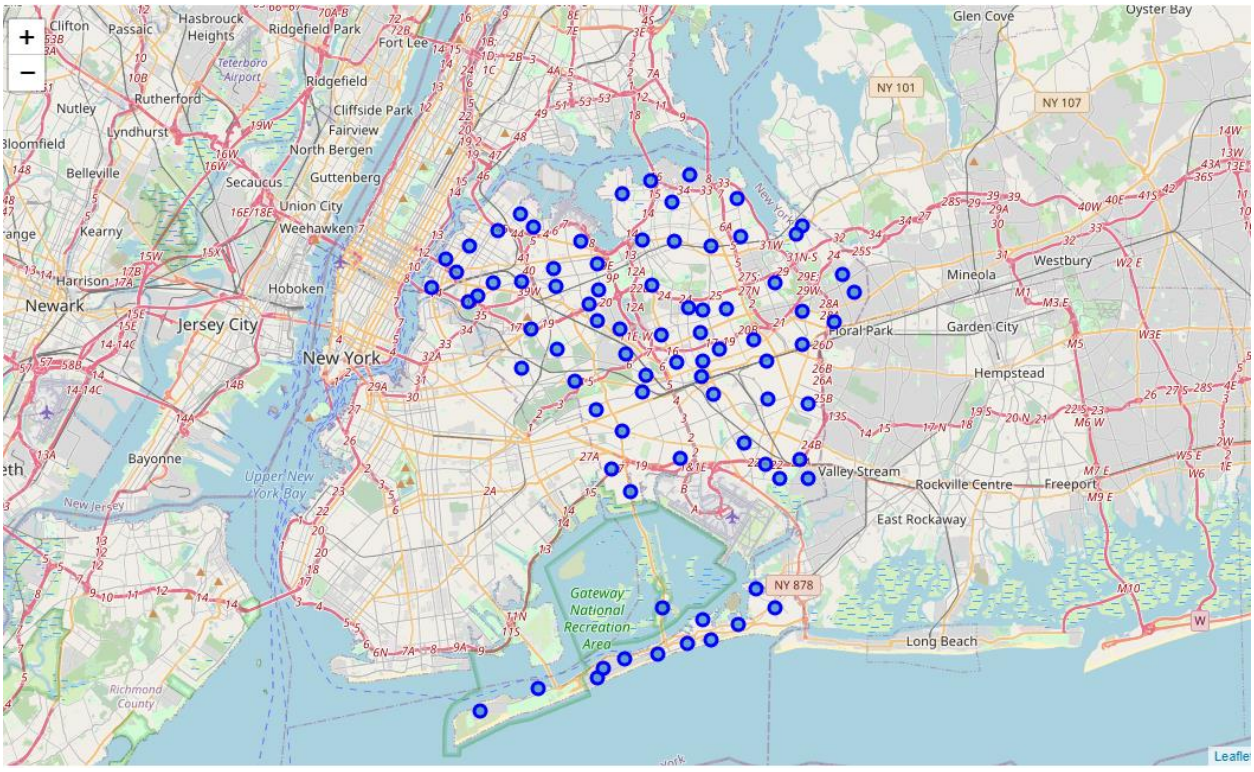
Discussion

Toronto has 11 boroughs and 103 neighborhoods. The geographical coordinate of Toronto, Canada are 43.7170226, -79.4197830350134. In Scarborough borough, there are 85 venues in 17 neighborhoods and 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 5 boroughs and 306 neighborhoods. The geographical coordinate of New York City are 40.7308619, -73.9871558. There are 2108 venues in 81 neighborhoods in Queens borough.

Both Scarborough borough and Queens borough consist of neighborhood clusters that contain majority of the neighborhoods, and the remaining cluster had 1-5 neighborhoods. Queens borough had a significant more number of neighborhoods and significant more diverse venues than Scarborough.





Conclusion

In conclusion, based on the quantity of venues and diversity of venues, I would suggest Queens as a better option to open a restaurant. Queens offer more neighborhoods to open the restaurant, and

Queens has a more diverse environment that will provide a large population of tourists and visitors to the restaurant. In addition, comparatively, there will be less competition to open the restaurant in Queens since there are many Chinese restaurants in Toronto.