Capstone Project - Battle of the Neighborhoods

Introduction

This project will analyze neighborhoods between Toronto, Canada and New York City, United States. My family and I want to move abroad and we are considering these two options. We want insight into the neighborhoods and local businesses in the cities so that we 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 our needs.

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.

The following websites will be used to acquire the datasets: 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 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.
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 common venues are Fast Food Restaurant, Vietnamese Restaurant and Hobby Shop. Cluster 1 has 2 neighborhoods, and the most common venues are Indian and Vietnamese Restaurants. Cluster 2 has 1 neighborhood, and the most common venues are Skating Rink and General Entertainment.

Queens Borough in New York City

I used k-means to group the Queens borough into 5 clusters. Cluster_0 has 80 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 a Deli/Bodega. Cluster_2 has 1 neighborhood and the most common venue are Brewery and Food & Drink Shop. Cluster_3 has 1 neighborhood and the most common venues are Gym/ Fitness Center and Gym. Cluster_4 has 1 neighborhood and the most common venues are Surf Spot and Metro Station.

Discussion

Toronto has 11 boroughs and 103 neighborhoods. The geographical coordinates of Toronto, Canada are 43.7170226, -79.4197830350134. In Scarborough, Forsquare Found 90 venues in 17 neighborhoods.

New York City has 5 boroughs and 306 neighborhoods. The geographical coordinates of New York City are 40.7308619, -73.9871558. Foursquare Found 2117 venues in 81 neighborhoods.

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 greater number of neighborhoods 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 my family. Queens offer way more in choices for restaurants, gyms, grocery stores, and extracurricular activities for my wife, my children and I.