Capstone Project - Battle of the Neighborhoods

Introduction/Business Problem

Toronto and New York being the financial capital of Canada and the US respectively are one of the densely populated cities in the world. Both these cities see a lot of diversity resulting from the movement of a lot of immigrants from several parts of the world for work and settlement. These are one of the most immigrant-friendly cities, still different in so many aspects, which we are going to observe in this work.

•This project will analyze neighborhoods between Toronto and New York City, New York. A 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 company's employees.

• Hence the problem is how to determine the right neighbourhood for company to set up its headquarter.

Data Requirements

•The data used for this project will be acquired from the respective wikipedia website pages. The dataset consist of postal codes,neighborhood names,latittude 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 further provide insights into our neighborhood battle project.

I used data 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

• Foursquare database: https://Foursquare.com

Methodology

❖ Scraping the wikipedia page using BeautifulSoup library and pandas library and cleaning it and transforming it into a pandas dataframe.

- 1.Web scraping of the neighborhood data from postal codes of Canada Wiki-link. Clean the data by removing the missing values and store the data in a python Dataframe consisting of three columns namely: PostalCode, Borough, and Neighborhood.
- 2.Web scraping of the neighborhood data from postal codes of New York Wiki-link. Clean the data by removing the missing values and store the data in a python Dataframe consisting of three columns namely: PostalCode, Borough, and Neighborhood
- 3.Take the help of long-lat data from the geospatial data wikilink and append the geographical coordinates in the above dataframes to get new respective dataframes for further analysis.

❖ The data sets of the two neighbourhoods, the Scarborough and the Queen's were visualized to gain knowledge of the distribution of variables and using folium library to see how the neighbourhoods in this borough are spatially distributed.

❖ Getting location data using the Foursquare API. It will be used to retrieve information of the common venues in Toronto and New York neighborhoods. The API will return a JSON file which will be further converted into a Python Dataframe.

- ❖ Exploratory Data Analysis
  - • Plot and find the relationship between Neighborhoods and various places for both cities.
  - • List down the most common venues for both cities.

- ❖ 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 5 clusters. Cluster_0 has 2 neighborhoods and the most common venues are American restaurant and burger point. Cluster 1 has 15 neighborhood and the most common venues are bar, spa ,bakery, park. Cluster 2 has 1 neighborhood, and the most common venues are college stadium. Cluster 3 has 1 neighborhood, and the most common

venues are middle eastern restaurant. Cluster 4 has 1 neighborhood, and the most common venues are Indian restaurant.

Queens Borough in New York City :

I used k-means to group the Queens borough into 5 clusters. Cluster_0 has 81 neighborhoods and the most common venues are pizza places, deli, and Chinese restaurants. Cluster_1 has 1 neighborhood and the most common venue is a Bodega. Cluster_2 has 3 neighborhoods and the most common venue are donut shops .Cluster_3 has 1 neighborhoods and the most common venues are the gym and fitness center. Cluster_4 has 1 neighborhoods and the most common venues are surf spot.

Discussion

- Toronto has 15 boroughs and 103 neighborhoods. The geographical coordinate of Toronto are 43.7170226, -79.4197830350134. Foursquare found 97 venues in 17 neighborhoods in Scarborough borough.
- New York City has 5 boroughs and 306 neighborhoods. The geographical coordinate of New York City are 40.7308619, -74.9871558. Foursquare found 71 venues 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 more 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 the headquarters of the company. Queens offer way more in choices for restaurants, gyms, grocery stores, and extracurricular activities for individuals and families of the company’s employees.