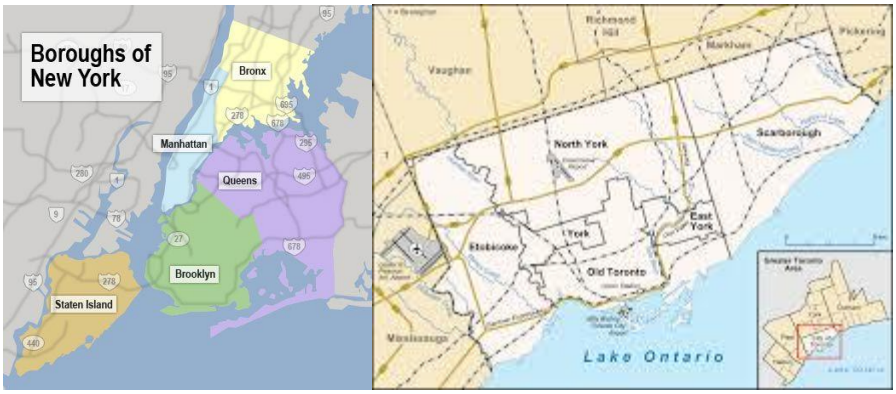
CapStone Project:
Battle of the Neighborhoods

By Lspdbe

https://github.com/lspdbe/Coursera_Capstone/

Introduction:

This project will analyze neighborhoods between Toronto and New York City. A company is looking to move its headquarters to either of both. The company wants insight into the neighborhoods and local businesses in the cities so that its employees may have the best possible living standards. This project will explore the similarities and differences between certain neighborhoods in the two cities, and determine which neighborhoods best fit the culture of the company.



Data:

The data used in this project will be acquired from Wikipedia for Toronto, and from the NYU website for New York. Data includes postal codes, neighbourhood names, and coordinates. Foursquare API feature will be used to collect venue information for each neighbourhood. With details about each venue, insights can be generated regarding their quality. Also, various python packages will be used to create maps and machine learning models to further provide insights into this project.

Toronto:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M → Toronto Neighborhoods

http://cocl.us/Geospatial_data → Toronto coordinates

https://geo.nyu.edu/catalog/nyu_2451_34572 → NY Neighborhoods

Python Geolibrary → NY Coordinates

Methodology:

Project flow:

- HTTP requests to Foursquare API server using the zip codes of Toronto and New York City neighborhoods, this allows to pull location information
- Using the Foursquare API search feature, locations could be explored and data could be collected on numerous venues.
- Using Folium: the neighbourhood clusters can be visualized for New York and Toronto in an interactive leaflet map
- Two neighborhoods are chosen (Scarborough for Toronto, and Long Island City for New York). These neighborhoods are extensively compared to derive insights using Pandas, NumPy and Scikit-learn.
- K-means clustering analysis is applied to form clusters of different venue categories in and around the neighborhoods. The formed clusters are analyized individually, in group and compared to each other to come up with conclusions for the project.

Python packages used in the project:

- Pandas: library for data analysis
- NumPy: library to handle data in a vectorized manner
- JSON: library to handle JSON files
- Geopy: used to retrieve location data
- Requests: library to handle http requests
- Matplotlib: python plotting module
- Sklearn: pyhton machine learning library
- Folium: map rendering library

Results:

Scarborough Borough in Toronto (Canada):

Using K-means clustering, the neighborhoods in Scarborough were grouped into 3 clusters.

Cluster 0 has 1 neighborhood, and the most common venues are a park, an Asian restaurant and a playground.

Cluster 1 has 16 neighborhoods, and the most common venues are friend chicken joints, general entertainment, and fast food restaurants.

Cluster 2 also has 1 neighborhood, and the most common venues are discount stores, department stores and playgrounds.

Queens Borough in New York (United States)

Using K-means clustering, the neighborhoods in Queens were grouped into 5 clusters.

Cluster 0 has 6 neighbourhoods, and the most common venues are delis, pizza places and bus stations.

Cluster 1 has 1 neighbourhood, and the most common venues are delis, basketball courts and history museums.

Cluster 2 has 3 neighbourhoods, and the most common venues are American restaurants, bus stations and coffee shops.

Cluster 3 has 70 neighbourhoods, and the most common venues are pizza places, delis and bodegas.

Cluster 4 has 1 neighbourhood, and the most common venues are surf spots, metros stations and coffee shops.

Discussion:

Toronto has 11 boroughs and 103 neighborhoods, and the geographical coordinates are 43.653963, -79.387207.

Scarborough is located at 43.773077, -79.25777. In Scarborough, 89 venues were located in 17 neighborhoods. OF these, there are 81 distinct venues which fall into 53 categories.

New York has 5 boroughs and 306 neighborhoods, and the geographical coordinates are 40.7127281, -74.0060152.

Queens is located at 40.6524927, -73.791421415. In Queens we can find Long Island City at 40.75021734610528, -73.93920223915505. Here we can find 73 venues.

In Queens, there are 2145 venues in 81 neighborhoods.

For both boroughs, clusters with only one neighbourhood were identified. These results pose further research to find out why they stand out against the other clusters.

Conclusion:

Following the analysis above, Queens seems to be the better choice compared to Scarborough to relocate the offices of the company. It offers a greater amount of gyms, grocery stores, restaurants which benefit bot individual and families of the employees of the company.