

Battle of the Neighbourhoods

Toronto vs New York City

for The best location for the best organization.

Introduction

This project will analyse neighbourhoods between Toronto, Canada and New York City, New York. An Organization 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 Fortune 500 company'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:

1. 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).

2. 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.

3. Folium- Python visualization library would be used to visualize the neighbourhoods cluster

Distribution of Seattle city over an interactive leaflet map.

4. Extensive comparative analysis of two randomly picked neighbourhood's 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 neighbourhoods. These clusters from

Each of those two chosen neighbourhoods would be analysed individually collectively and

Comparatively to derive the conclusions.

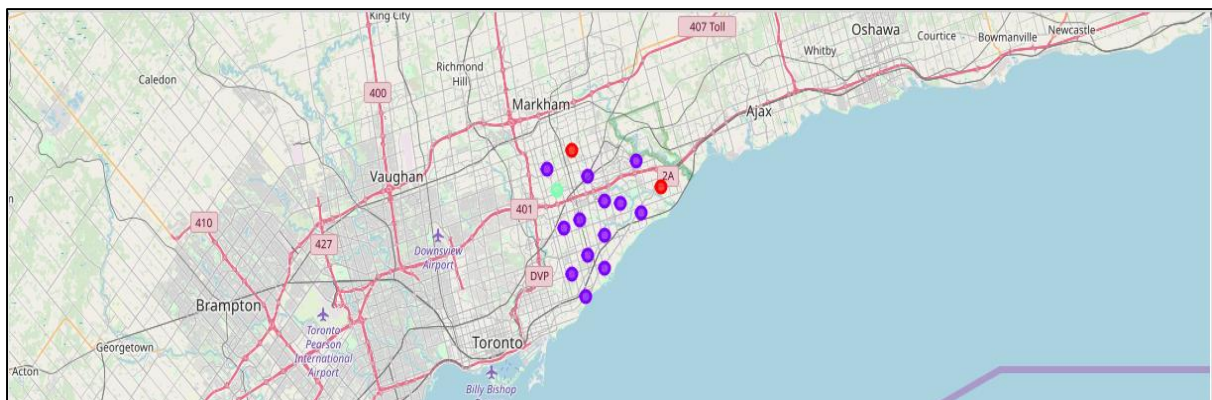
The following are the Python packages I used:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a 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 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:

Scarborough has 81 distinct venues in 54 categories, the most common venues in Scarborough are coffee shops, international restaurants, and Pharmacy

Queens has 1805 distinct venues in 269 categories, the most common venues in Queens are bodegas, bakeries, coffee shops, and parks

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 Organization. Queens offer way more in choices for restaurants, gyms, grocery stores, and extracurricular activities for individuals and families of the company's employees.