
BATTLE OF THE NEIGHBORHOODS

New York City vs. Toronto



BUSINESS PROBLEM TO SOLVE

- Explore information of New York City and the city of Toronto; segmented and clustered their neighborhoods.
- Choose a city of my own choice to open a restaurant.
- In this project, I narrow down the goal to open a Chinese restaurant of a chosen city.



OBJECTIVE

- Collect neighborhoods information using Foursquare
- Using k-means clustering algorithm to form clusters of the neighborhoods
- Identify the similarities and differences between neighborhoods of New York City vs. City of Toronto
- Make a decision for the better location to open a Chinese restaurant



DATA

- 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

- Import all the necessary Python packages.
- Read the data from the Wikipedia page and clean the data.
- Use Foursquare API server to pull the location information (Latitude and Longitude).
- Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods.
- Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of each city over an interactive leaflet map.
- Extensive comparative analysis of two randomly picked neighborhoods would be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.
- 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.



PYTHON PACKAGE USED

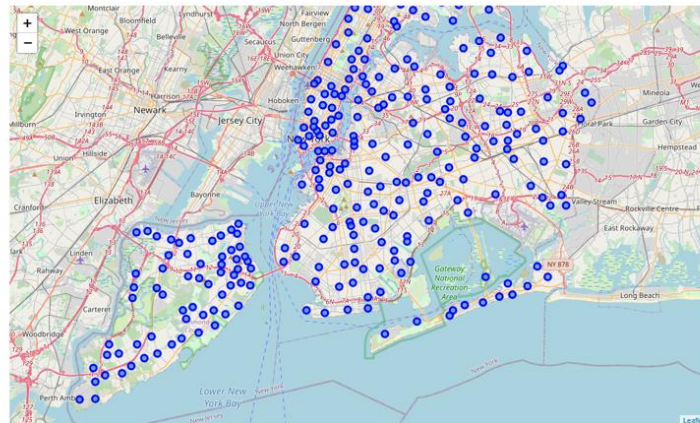
- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library
- BeautifulSoup
- Lxml



RESULTS

- 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.

MAP OF NEW YORK CITY



COMPARISON BETWEEN NEIGHBORHOODS

- Toronto has 103 unique neighborhoods.
- The most common venues in Toronto are Breakfast spot, pizza place, coffee shop, sandwich place and café.
- New York City has 306 unique neighborhoods.
- The most common venues in New York City are more diverse. I have clustered the neighborhood of Queens into 5 clusters:
 - Cluster 0: pizza place, deli, beach, hotel, etc.
 - Cluster 1: Deli
 - Cluster 2: Clothing store, restaurant, donut shop
 - Cluster 3: Gym
 - Cluster 4: Bakery, restaurant, park, bar, etc.



CONCLUSION

- Based on the quantity and diversity of venues, New York City, especially the Queens neighborhood, is a better choice than Toronto to open a new Chinese restaurant. New York City has less restaurants to compete against, and there are more diverse venues such as gym, park, beach to attract more visitors/tourists, who might pay a visit to the new Chinese restaurant.

