**Capstone Project - Battle of the Neighborhoods**

**Introduction / Business Problem**

A Chinese technology company is embarking on global expansion and is looking for a startup setup in either Downtown of Toronto or Brooklyn of 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 Chinese technology startup’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, 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.

I used the following datasets 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
* New York City Latitude and Longitude = Python Geolibrar

**Methodology**

**Approach:**

1. Under the business requirements and determine the analytical approach to assess whether the neighborhood is suitable, with some key considerations as below:

* Dining – Chinese restaurants for team gathering, pizza place for quick lunch and coffee shop for client meetings
* Daily needs – pharmacy and grocery store
* Location – nearby CBD or financial center to provide back-office support to clients

1. Understand the data requirements and collect the data as described in the “Data” section above.
2. The data was analyzed, cleaned, and processed. As a result, some features have been extracted and used.
3. HTTP requests would be made to this Foursquare API server using zip codes of the Toronto and New York city neighborhoods to pull the location information (Latitude and Longitude).
4. 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.
5. Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Toronto and New York city neighborhoods over an interactive leaflet map.
6. 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**

**Downtown Borough in Toronto, Canada**

I use k-means to group the 18 neighborhoods in Downtown (Toronto) into 3 clusters:

|  |  |
| --- | --- |
| * Cluster #1 has 16 neighborhoods and the most common venues are coffee shop, café and park |  |
| * Cluster #2 has 1 neighborhood, and the most common venues are airport lounge, airport service and airport terminal |  |
| * Cluster #3 has 1 neighborhood, and the most common venues are Japanese restaurant, sushi restaurant and coffee shop |  |

**Brooklyn Borough in New York City**

I used k-means to group the 70 neighborhoods in Brooklyn (New York) into 5 clusters:

|  |  |
| --- | --- |
| * Cluster #1 has 60 neighborhoods and the most common venues are pizza places, deli and Chinese restaurants |  |
| * Cluster #2 has 1 neighborhood and the most common venues are pizza place, Chinese restaurant and bagel shop |  |
| * Cluster #3 has 6 neighborhoods and the most common venues are pizza place, harbor/ marina and donut shop |  |
| * Cluster #4 has 2 neighborhoods and the most common venues are pizza place, Caribbean restaurant and pharmacy |  |
| * Cluster #5 has 1 neighborhoods and the most common venues are bakery, Chinese restaurant and pizza place |  |

**Discussion**

Toronto has 11 boroughs and 103 neighborhoods. The geographical coordinate of Toronto, Canada are 43.653963, -79.387207.

|  |  |
| --- | --- |
|  | In Downtown borough:   * Foursquare found 1278 venues in 18 neighborhoods * The neighborhoods with the most venues are Habourfront East, Adelaide, Chinatown, Commerce Court, Design Exchange, First Canadian Place and Ryerson * There are 744 distinct venues in 205 categories |

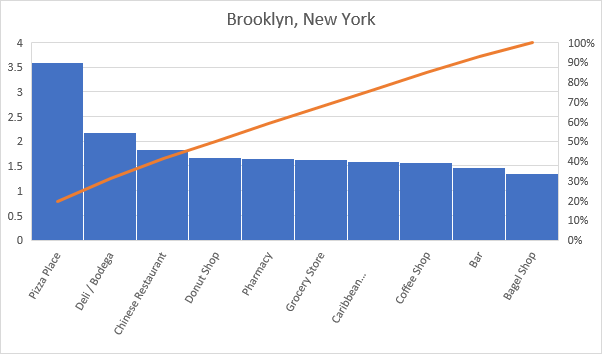
New York City has 5 boroughs and 306 neighborhoods. The geographical coordinate of New York City are 40.7308619, -73.9871558.

|  |  |
| --- | --- |
|  | In Brooklyn borough:   * Foursquare found 2809 venues in 70 neighborhoods; many of the neighborhoods are homogenous and are very similar to each other * The neighborhoods with the most venues are Downtown, Brooklyn Heights, Carroll Gardens, Cobble Hill, Greenpoint, North Side and South Side * There are 2312 distinct venues in 287 categories |

Both Downtown of Toronto and Brooklyn of New York consist of neighborhood cluster that contain majority of the neighborhoods, and the remaining clusters had only around 10% neighborhoods. Brooklyn of New York had a significant more number of neighborhoods and venues than Downtown of Toronto.

**Conclusion**

In conclusion, based on the quantity of venues and variety of venues, I would choose Brooklyn of New York over Downtown of Toronto as a choice to setup the startup of the Chinese technology company. Brooklyn of New York offer way more in choices for Chinese & other restaurants, pharmacy, grocery stores, various venues and extracurricular activities for individuals and families of the company’s employees.



Brooklyn of New York can better fulfill the business requirements as below:

* Dining – Chinese restaurants for team gathering, pizza place for quick lunch and coffee shop for client meetings (as per the above)
* Daily needs – pharmacy and grocery store (as per the above)
* Location – nearby CBD or financial center to provide back-office support to clients (given its geographical coordinate, such as Downtown neighborhood, is very close to Manhattan)