

# **Capstone Project - The Battle of Neighborhoods**

Baichuan Chu

August 25, 2020

## **1. Introduction**

This project is mainly analyzing the neighborhood in both Toronto and New York City. This analysis will give some people ideas on either traveling or having international businesses. Today, more and more people are engaging in globalization. For instance, there are people who travel around the world in order to study. There are also people who are interested in investing in foreign countries. This analysis will analyze the two major cities in North America so that people can make wise decisions.

## **2. Questions:**

- Within Toronto and New York City, which boroughs of those two cities contain the most neighborhood?
- What are the differences between the two boroughs?(Mainly discuss the differences in their venues.)

- Which one of the cities is good for people to travel and study?
- Which one of the cities is good for people to open up a Chinese restaurant?

### 3. **Data**

I used several data sets in this project. My project starts with analyzing Toronto's neighborhoods. The first data set is a Wikipedia site that give me a lot of information of Toronto. I also need the coordinates of each neighborhood in Toronto. This is my second data set. With the two of the Toronto's data sets, I could finish the Toronto's analysis. Second part of the project is analyzing the New York City. I used the link provided within practice labs to find out information of New York's neighborhoods.

Toronto data set: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Toronto's coordinate: [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)

New York data set: [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

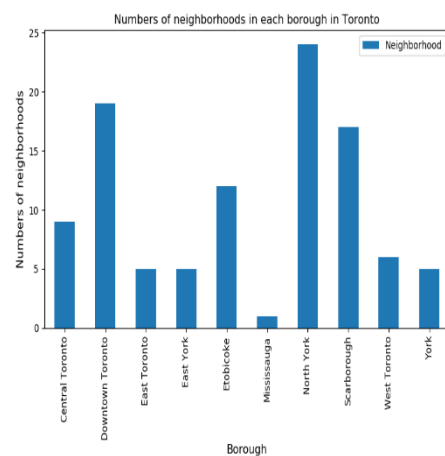
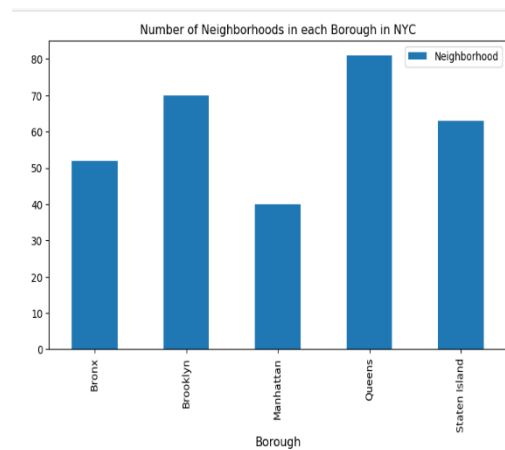
### 4. **Methodology:**

- Clean and sort the data from the sources I provided above.
- Using Python libraries to create data frames and merge longitudes and latitudes into the tables.

- Use matplotlib to create chart and pick the area for analysis.

Charts attached below in order to determine which specific area of Toronto and New York will be analyzed. In my case, I choose Queen and North York for this project.

- Then using FourSqaure API to find out the venues around the targeted area. With the venues' information, I could compare the two area briefly.
- Last step was to make clusters and segregate each neighborhood into different clusters in order to make final decisions.



## 5. Python Libraries used:

- Pandas
- Numpy
- Requests

- BeautifulSoup
- Geocoder
- Folium
- Matplotlib

## 6. Result:

- Within the analysis, I segregate both Toronto and New York into 5 clusters using Kmean clustering method.

### i. Toronto: The first cluster in Toronto contains 3

neighborhoods, and the most common venues are parks.

The second cluster in Toronto has 21 venues. Gyms, Coffee

shops, and parks are most seen venues. The third, fourth,

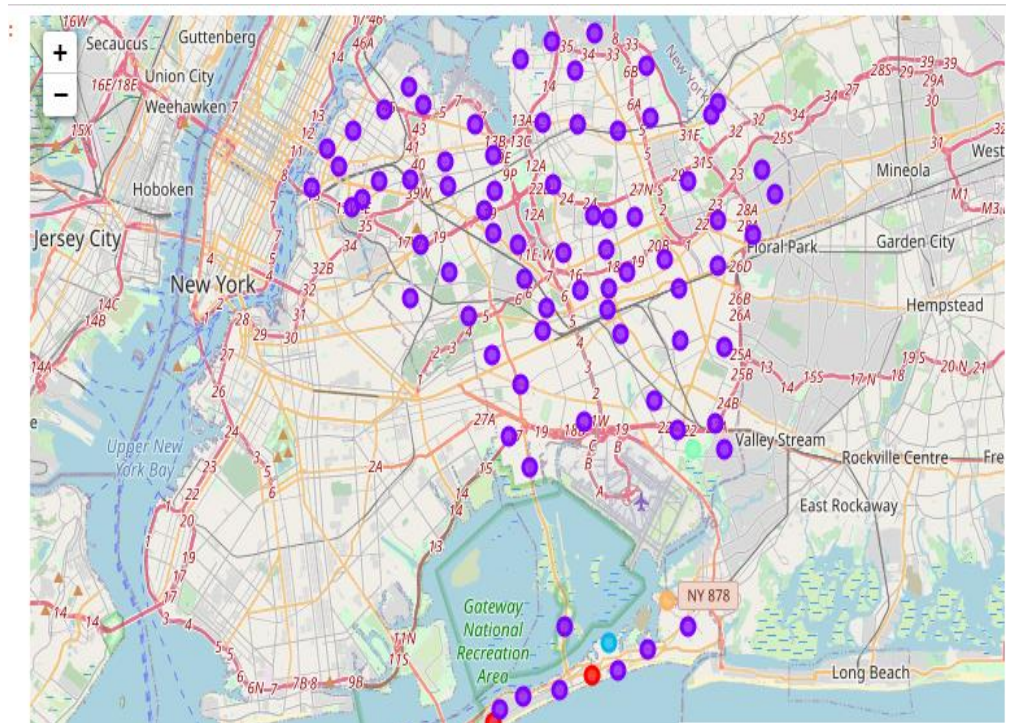
and fifth clusters only contain 1 neighborhood, and they are

gym, golf course, and Furniture.



**iii.** New York: Unlike North York in Toronto, Queen is much more interesting to explore. The first cluster contains 3 neighborhoods, and they are all beaches. Then second cluster is huge and has 80 clusters. They are mostly restaurant. Third, fourth, and fifth clusters contain only 1 neighborhood, and they are park, deli, and playground.

**iv.**



This is the New York Map with neighborhoods in it created by using folium. We can see that New York is much more compact and has so many venues around.

## 7. Discussion

- In Toronto, there are 10 boroughs and 103 neighborhoods. North York is the largest borough among the ten boroughs. North York has 236 venues and 24 neighborhoods. One interesting fact is that North York does not have much restaurant business within it. Within my cluster section for North York, we can see that there are very few restaurants as the first most common venues. The most often seen one is park and coffee shops.
- For New York, it is a different situation. New York only has 5 boroughs but 306 neighborhoods. Queen is about the biggest borough among all others. It contains more than 80 neighborhoods. Within Queen, there are 2088 venues in 81 neighborhoods. In my clustering section, we can see that most of the venue in Queen are restaurants businesses, coffee shops, and gyms. This is different from what we see in North York Toronto.

## **8. Conclusion**

This project is made to answer the questions I posted at the beginning.

1. The biggest boroughs in Toronto and New York are North York and Queen.
2. There are many differences:

- Queen is much bigger than North York in terms of neighborhoods.
- Queen contains way more venues than North York does.
- The venues in different area are very different. North York mostly contains parks and coffee shops. Queen is mostly filled with foods and gyms. This could lead some business opportunities.

3. For a student and a traveler's perspective, New York is a better place to stay. Within only Queen area, there are so many options and things to do. People can either get some food or have a place to rest. It's more interesting to walk through all venues.

4. Unlike Queen, North York does not have that many venues within it. This also give me a sense of opportunity. People who are aiming to have businesses overseas could consider North York, Canada. North York is the biggest borough in Toronto. Besides that, there isn't many venues and business going there. Coming back to my question, I believe it's a good idea to open a Chinese restaurant in North York. From the cluster analysis, we barely see any Chinese restaurants listed in the 1<sup>st</sup> common venues in all North York

neighborhoods. In other words, there isn't many competitors from North York just terms of Chinese restaurant. We know Toronto is a city with a lot of Chinese people; thus, opening a Chinese restaurant in North York could be a beneficial idea for a lot of foreign business man.