**Capstone Project**

**The Battle of Neighborhoods**

**Author: Himani Verma**

# Introduction

I live in in Toronto with family. My husband has been offered his dream job a thousand miles from home. While my job provides me the flexibility to work from a remote location, his new job doesn't have this facility. So, we have decided to relocate somewhere near to his work location.

Along with the excitement that I have moving to a new city, I am a littile anxious and nervous as well.  
I was born and raised in Toronto and I have never been away from my parents for a long time. All my friends are here, and my kid is also studying at a school here. All the things that we need are easily available nearby.

Both NYU and Toronto cities are very diverse and are the financial capitals of their respective countries. We would like to compare the neighborhoods of the two cities and determine how similar or dissimilar they are before we move there.

**Audience:**

1. Anyone who is moving to a new city they have never been to. It would help them with settling down in the new environment and exploring new life changing options.
2. When considering business development strategies, it is often helpful to examine downtown or business district relative to those in peer cities or cities that embody similar qualities. This comparison can help you discover business development opportunities that will strengthen economic vitality in your business district.
3. This Project will help to determine what kinds of retail or service business activities are supported in like places.
4. It will also provide real-life examples of districts that have defined themselves with a clear position in their market in terms of goods and services offered and of primary consumer segments served.

# Objective

We will use Foursquare Data and Machine Learning segmentation and Clustering to study area classification in details.

The aim of this project is to segment areas of New York City and Toronto based on the most common places captured from Foursquare. Using segmentation and clustering, we hope we can determine:

1. The similarity or dissimilarity of both cities
2. Classification of area located inside the city whether it is residential, tourism places, or others

# Data

We are going to use data from below Sources.

**Toronto** Dataset:

* web url: "<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>"
* **Geocoder\_File**: Geospatial\_Coordinates.csv

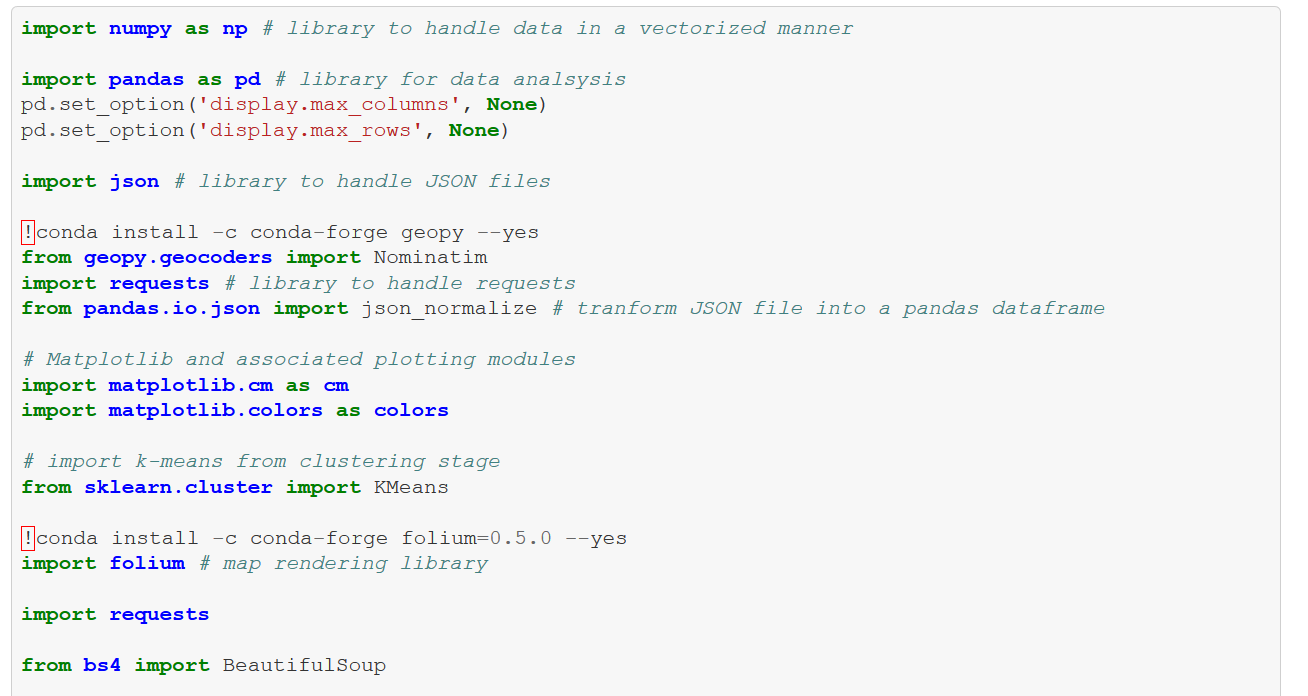
**New York** Dataset:

* **Filename**: nyu\_2451\_34572-geojson.json

The above files have been saved in local path for using in the notebook

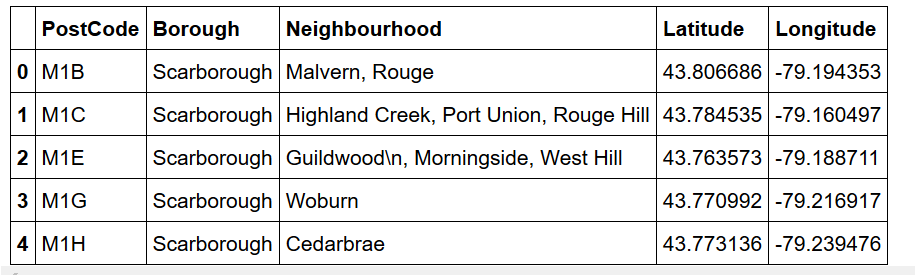
**Required Libraries:**

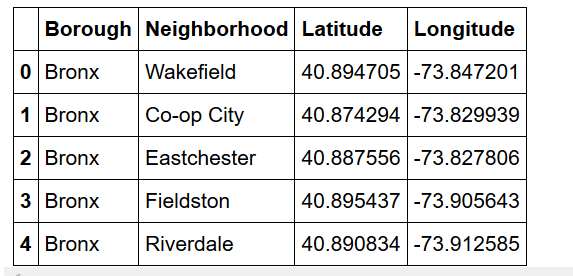
Python has a variety of libraries that we can use in this Project.



Let’s have a look at some sample data:

**Toronto** dataframe:



**New York** Dataframe:

# Methodology

In this project, we will use the basic methodology.

We have plotted maps of New York City and Toronto using latitude and longitude values.

After this we will use the Foursquare API to explore neighborhoods in both cities.

We will use explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters

**K-means** clustering algorithm will be used to complete this task.

**Folium** is being used to explore neighborhoods of these cities.

**Segment and form Clusters:**

Toronto Segment**: East Toronto**

New York Segment**: Manhattan**

A grouped dataframe is obtained one for **East Toronto** and one for **Manhattan** using

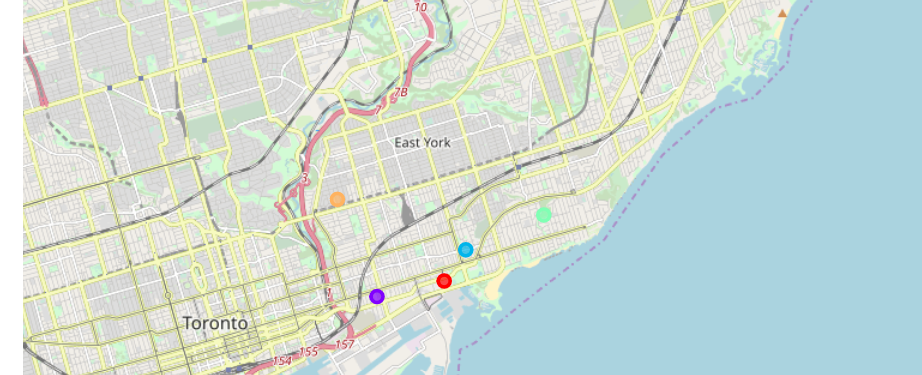
Foursquare most common venue API.

And K-means clustering Algorithm is used to find clusters. Then is generated on these clusters using Folium.

**East Toronto Neighborhood Venues Dataframe:**



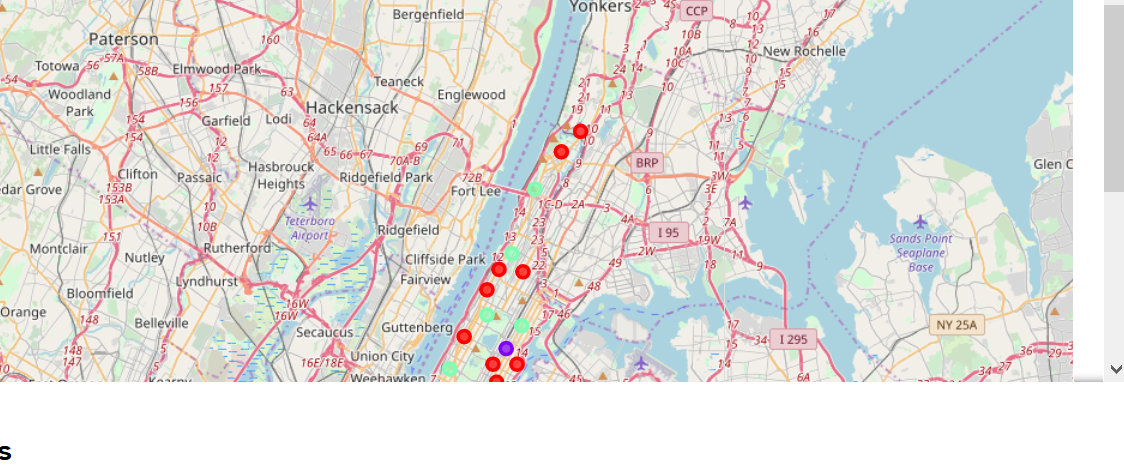
**East Toronto Cluster Map:**



**Manhattan Neighborhood Venues Dataframe:**

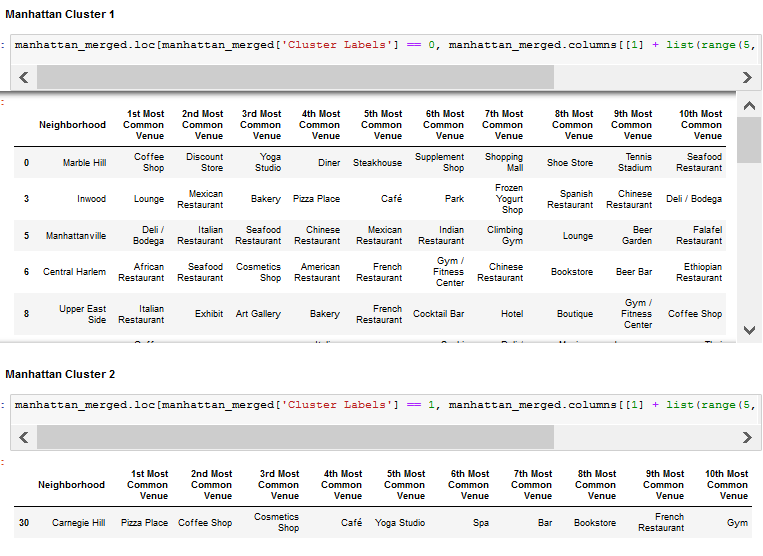
****

**Manhattan Cluster map:**



# Results





# Discussion

The K-means algorithm is a technique to cluster n objects, based on attributes, into k partitions, k < n. Clustering is the classification of objects into different groups or, more precisely, the partitioning of a dataset into clusters (subsets) so that the data in each cluster (ideally) share some common trait—often proximity according to some defined distance measure

Based on cluster for each cities above, we believe that classification for each cluster can be done better with calculation of venues categories (most common) in each cities. Referring to each cluster, we can't determine clearly what represent in each cluster by using Foursquare - Most Common Venue data.

However, for the sake of this project we assumed each cluster as follow:

* Cluster 1: East Toronto: Residential
* Cluster 2: East Toronto: Mix
* Cluster 3: East Toronto: Health and Fitness
* Cluster 1: Manhattan: Sports, Residential, Shopping
* Cluster 2: Manhattan: Health and Fitness
* Cluster 3: Manhattan: Residential

What is lacking at this point is a systematic, quantitative way to identify and distinguish different district and to describe the correlation most common venues as recorded in Foursquare. The reality is however more complex: similar cities might have or might not have similar common venues. A further step in this classification would be to find a method to extract these common venues and integrate the spatial correlations between different of areas or district.

We believe that the classification we propose is an encouraging step towards a quantitative and systematic comparison of the different cities. Further studies are indeed required to relate the data acquired, then observe it to more meaningful and objective results.

# Conclusion

We can conclude that foursquare API's are useful in capturing data about various cities in the world. Using it, we refer to our main objectives, which is to determine;

* The similarity or dissimilarity of both cities
* Classification of area located inside the city whether it is residential, tourism places, or others

In conclusion, both cities East Toronto and Manhattan are the two most popular cities. However, to declare both cities are similar or dissimilar base on common venues visited is quite difficult.

Both cities is similar in some venues also dissimilar in certain venues. And for classification based on common venues, again we must have more systematic or quantitative way to identify and declare this. Comparison can be made, but no such method or quantitative data to determine this. We hope in the future, a method to determine it can be establish and explore for references.

For this particular statement we can conclude that it might take little bit of initial adjustment, but we will live a comfortable life in New York City. Looking forward to relocating.