**Migration to the US**

**Muhammad Tallal**

**2 Feb, 2020**

**Business Problem:**

The business problem that we are going to discuss is migration to a different country. Let’s say a person wants to migrate to another country and he wants to know which city would be preferable for him if he would compare it to the current city he is living in. That way he can make in informed decision. In this project we consider that the person is currently living in Toronto and wants to shift to the US. He is considering options New York and Chicago. So we are going to take neighbourhoods of these two cities and compare them with Toronto and see which city is more similar to Toronto based on venues. These venues of the neighbourhoods will be retrieved through Foursquare api. This project can be extended or modified for any such use cases where we need to find similarities or differences among cities or states. Therefore this project can be used for any person who is looking for either migrating to a different city or country or maybe shifting or extending his business.

**Data:**

The neighbourhood data of given cities will be retrieved using the following websites:

[Toronto Neighbourhoods](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

[New York Neighbourhoods](https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm)

[Chicago Neighbourhoods](https://www.seechicagorealestate.com/chicago-zip-codes-by-neighborhood.php)

This neighbourhood data will be used to find venues in different neighbourhoods of different cities using Foursquare api. Foursquare api will return the longitude and latitude data of different category of venues which will be then used to compare the cities.

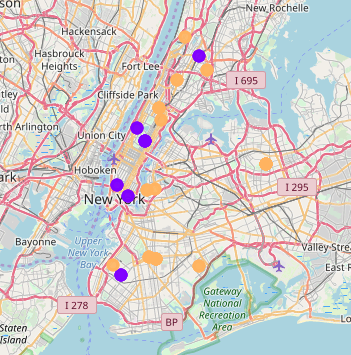
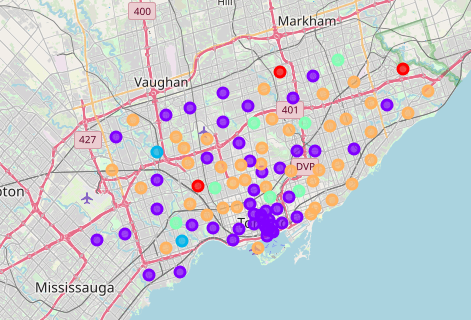
**Methodology:**

First step is to data wrangling of all the cities. So we will get the neighbourhood data and get the venues in those neighbourhoods using Foursquare api. Then it will be determined which type of venues are most common in different neighbourhoods. This data will be prepared for all the cities i.e., Toronto, New York and Chicago. Then we will merge data of Toronto and New York, and Toronto and Chicago to perform k-Means clustering. This will give us clusters of neighbourhoods and we determine which city is more similar to Toronto based on the fact that how many neighbourhoods from different cities are placed in the same clusters. This is also visualized using Folium so that the analysis becomes more easy. The main algorithm used for this project is k-means clustering.

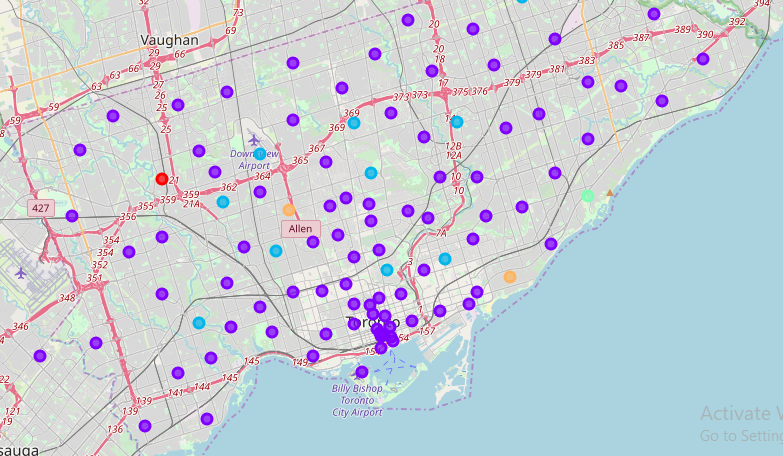
**Results:**

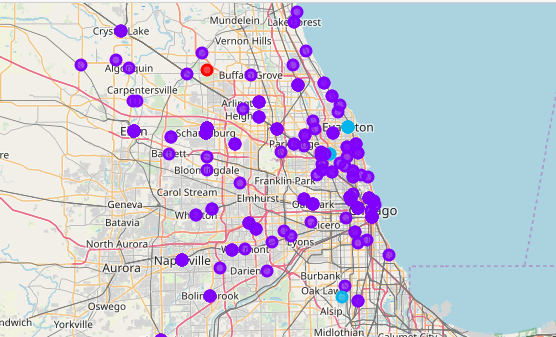
After analysis different neighbourhoods of different cities were clustered together and then these clusters were displayed using different colours on map.

Following maps show Toronto and New York comparison:



Following maps show Toronto and Chicago comparison:





**Discussion and Conclusion:**

As evident from the maps, Chicago is more closer to Toronto as compared to New York. So if a person is shifting or extending his business to the USA then Chicago is a better option for him rather than New York,

This project can be extended or modified for any different cities or countries to analyze which cities or countries resemble each other for any such use cases.