IBM Coursera Capstone Project

Report Prepared By : Kinjal Goswami

- ## Introduction / Business Problem

There are many people who are working in various cities (say New York and Toronto) across the world. Let's say a person got a job offer from a great company with great career prospects in other city or other borough of his/her current city. If person accept the job offer then he/she has to shift to new location. I think a person would love to shift a location which is exactly or almost similar to his/her last location because he/she loves the great amenities and other types of venues that exist in his/her current neighbourhood like school, gym, swimming pool, Amusement park, restaurants, coffee-shops, spencer etc. So I'll find out what are borough-neighbourhoods are very similar to give current location. A person can shift within the city or from one city to another city. If person is shifting within the city then he/she must be changing the borough and my task would be finding neighbourhoods in that borough which are very similar to his/her last location. If a person is shifting from one city to another city then my task would be finding the similar neighbourhoods.

Target Audience :-

The Job Seekers Who wants to find new Locations for job.

- ## Data :

To Solve this Problem I will Need below Data :

1. List of Neighbourhoods in Toronto, Canada.
2. Latitude and Longitude of this Neighbourhoods.
3. Perfect locations for Jobs.

Extracting The Data:

Scrapping the Data of Neighbourhoods from Wikipedia.

Getting Latitude and Longitude data of these neighbourhoods using geocoder.

Using Foursquare API getting venue data.

- # Methodology

First, I need to get this data from Wikipedia. Scrapping Data by using Pandas Data Frame or Table. Using This I just getting Cities and Their Postal Codes. To Find All Coordinators I used Foursquare API. I am getting Data for New York City and Toronto. After that I Apply EDA (Exploratory Data Analysis) to find Common Data between Them.

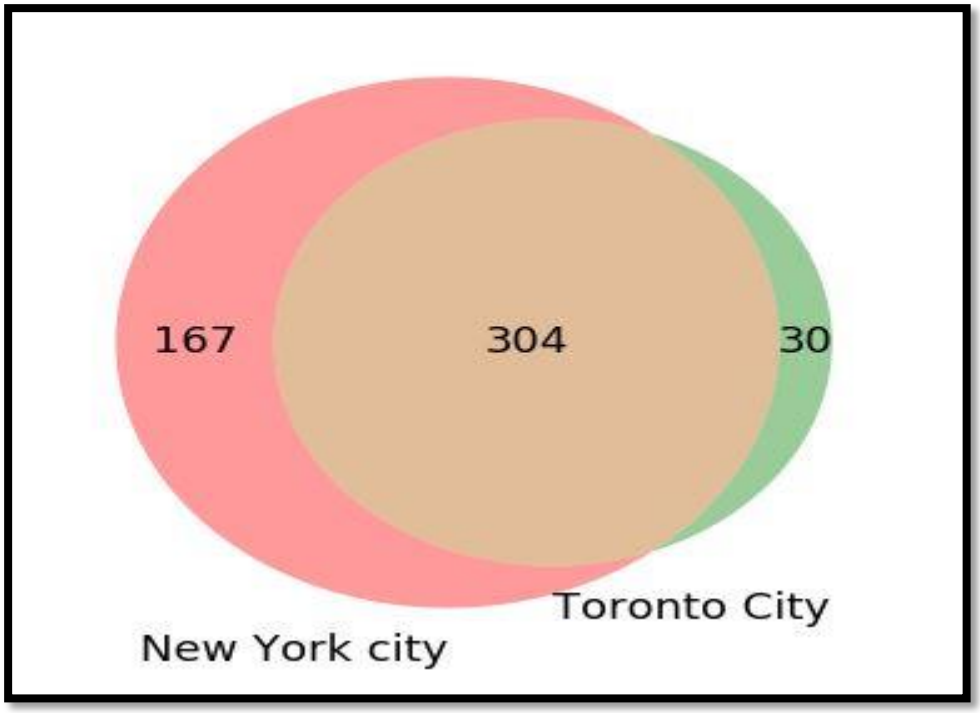
To find the most similar city, cosine similarity will be used. This kind of recommendation comes under item-item based recommender system. Finding Common Venues Categories.

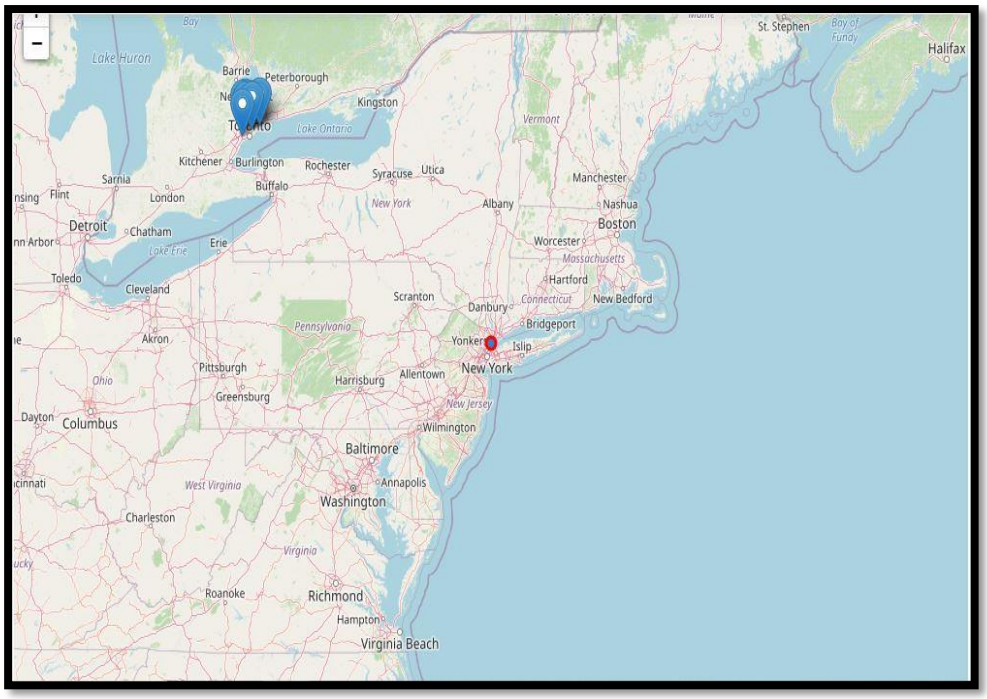
Let us one person is living in Bronx, Riverdale, New York city. Now he/she wants to move to Toronto city for some reason and he/she wants to find the most similar neighbourhood to his/her current location in Toronto city.

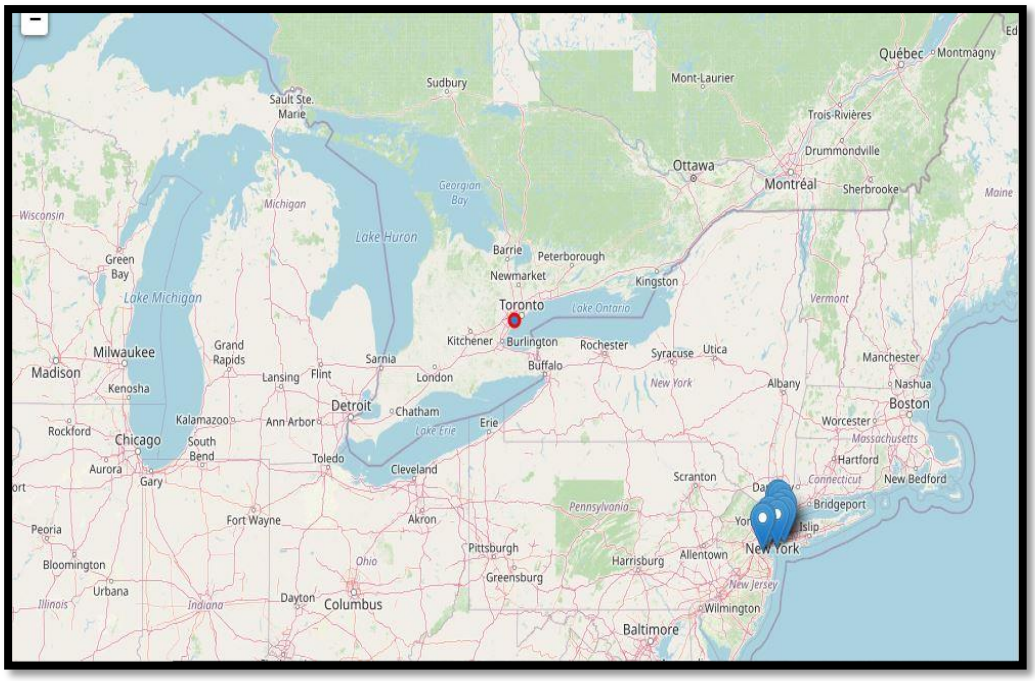
- We will be finding only top 7 most similar boroughs.

1. Find following details of his/her location from above data:
2. index value
3. latitude
4. longitude
5. venues near this location

- **Result**


167
304
30
Toronto City
New York city


Lake Huron
Barrie
Peterborough
Kingston
Lake Ontario
Kitchener
Burlington
Rochester
Syracuse
Utica
Buffalo
Sarnia
Flint
London
Detroit
Chatham
Erie
Lake Erie
Toledo
Cleveland
Akron
Pittsburgh
Ohio
Greensburg
Dayton
Columbus
Harrisburg
Allentown
Scranton
Pennsylvania
New York
Albany
Vermont
Manchester
Nashua
Boston
Worcester
Massachusetts
Hartford
Connecticut
New Bedford
Danbury
Bridgeport
Yonkers
Islip
New York
New Jersey
Wilmington
Baltimore
Annapolis
Washington
West Virginia
Charleston
Virginia
Roanoke
Richmond
Hampton
Virginia Beach
St. Stephen
Halifax



- **Conclusion :**

In this Project, we have gone through the process of identifying the process of business problem, specifying the data required, extracting and preparing data, performing the machine Learning.

- **References :**

Wikipedia, Foursquare developer documentation.