**IBM Coursera Capstone Project**

**Report**

**28 April 2020**

**Divyansh Agarwal**

**Battle Of Neighborhoods(2)**

**1.Introduction**

**----------------------------------------------------------**

**1.1 Introduction**

I have been pursuing this course with IBM and coursera, and I would like to thank IBM for providing such high quality free software platforms like IBM Watson Studio and Skills Network Labs, as I was able to learn a lot of things. In this Capstone project I will try my best to put all the knowledge and try to get the best results. I will be comparing two cities Toronto and New York and find similarities and differences between the two based on venues and neighborhood.

**1.2 Business Problem/Interpretation**

New York City and Toronto are the well known cities around the world for their various features.I will be comparing the two cities based on their neighborhoods and venue, and try to find a relation between the two cities. That is, try to find similarity or differences between the two. We will use geospatial data to analyse and visualize the neighborhoods of the cities and compare them with the various methods and techniques(Those taught in the course).

**1.3 Interest**

This information and conclusion will be useful to people who wants to invest or start business in the North-American Continent and will also help in various works. Also for various research purposes and these financial capitals can be analysed on various levels as well.

**2. Data**

**----------------------------------------------------------**

**2.1 Data Collection**

The data we require is geojson data or geospatial data of two cities which can be further be used to retrieve data from the FourSquare website to gather information like Venues, Neighborhood and Tips etc. The geospatial data for New York was given by our instructor and can be easily accessed while the data for Toronto will be taken from Wikipedia as a list of postal codes.

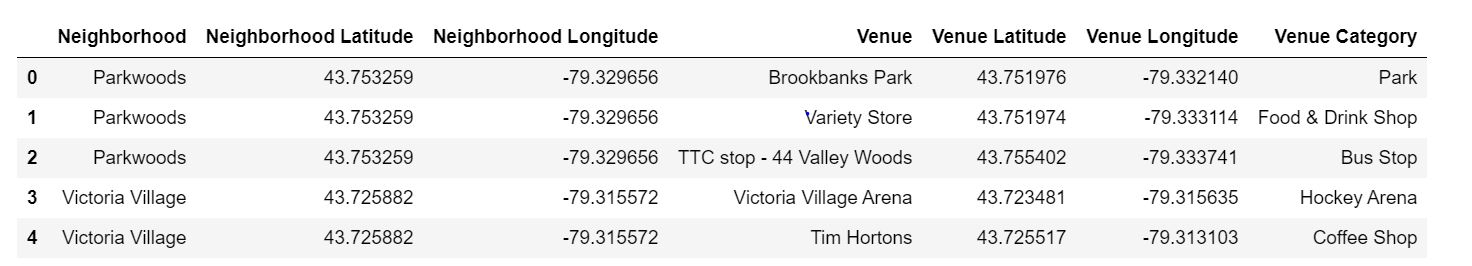
**2.2 Data PreProcessing**

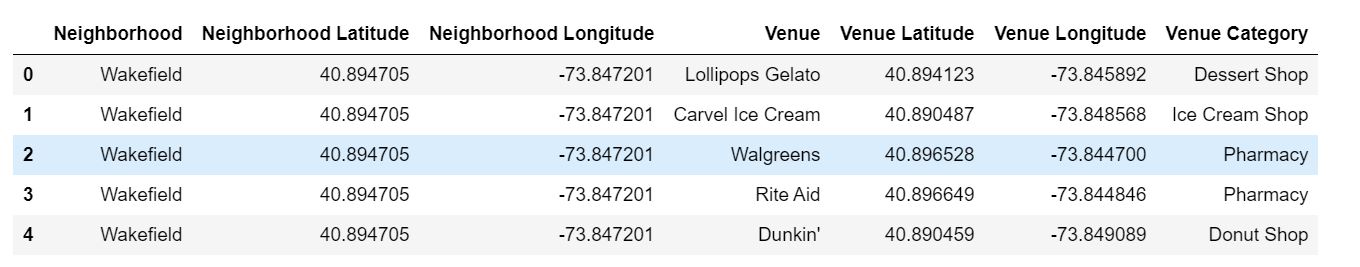
The data for New York City is already ready for use while the data for Toronto needs to be processed. The data is in the form of an HTML List/Table. To process this data we can use a very helpful library in Python named BeautifulSoup which can convert this HTML data into pandas DataFrame. This dataframe can be further modified accordingly.



**2.3 Data Preparation and Understanding**

The data of New York and Toronto can be used into a FourSquare account and a lot of useful information about neighborhoods, venues there tips, etc can be received from their servers. These venues can be clustered, visualized and compared to give great results.



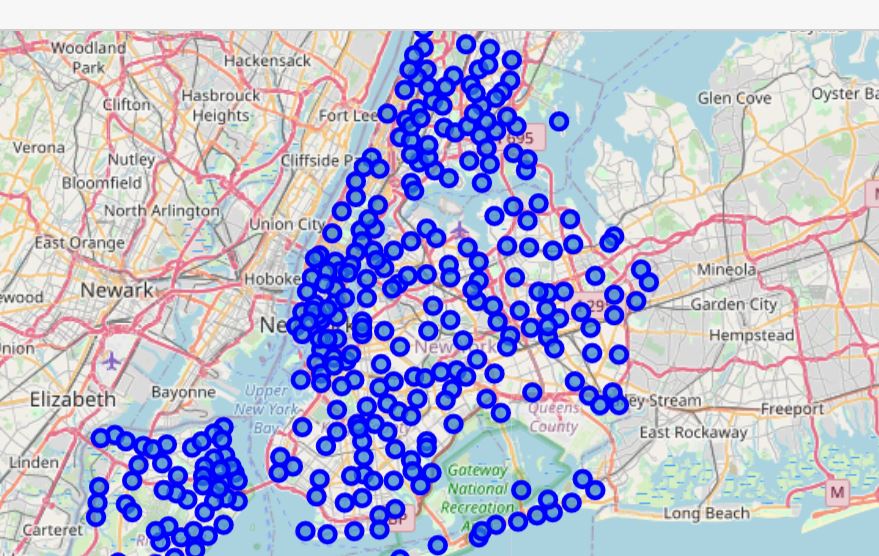


**3. Methodology**

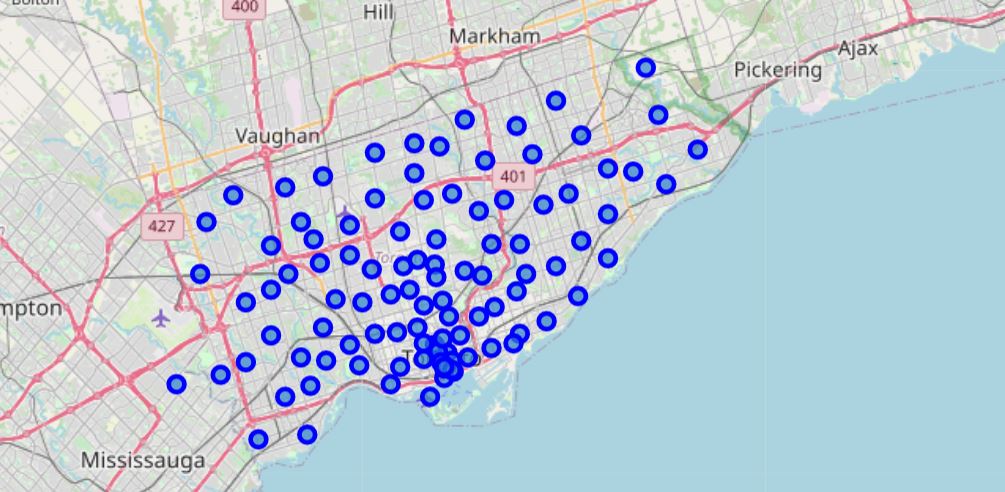
**----------------------------------------------------------**

So here we can already see the neighborhoods are more intense in New York City than in Toronto. Not only that the density is high and nearly constant throughout the city of New York, while in Toronto it is concentrated around a single point and slowly disperses away as we go far.

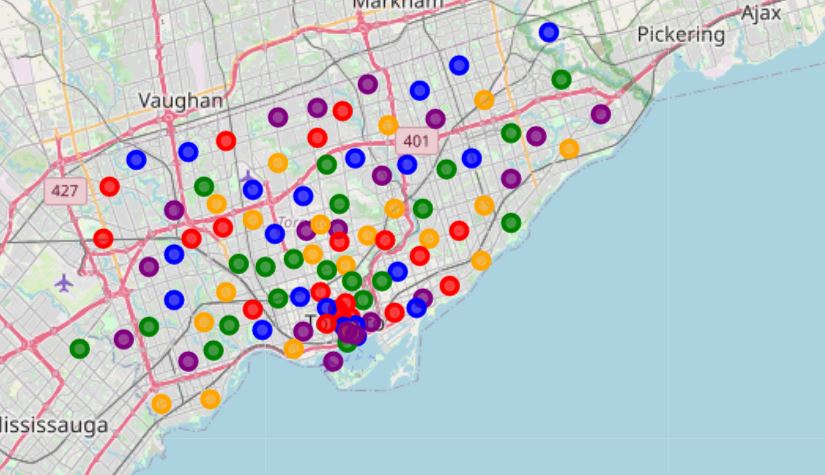
This here can show us that in New York, there are a lot of densely populated communities and are evenly spread throughout the city.

****

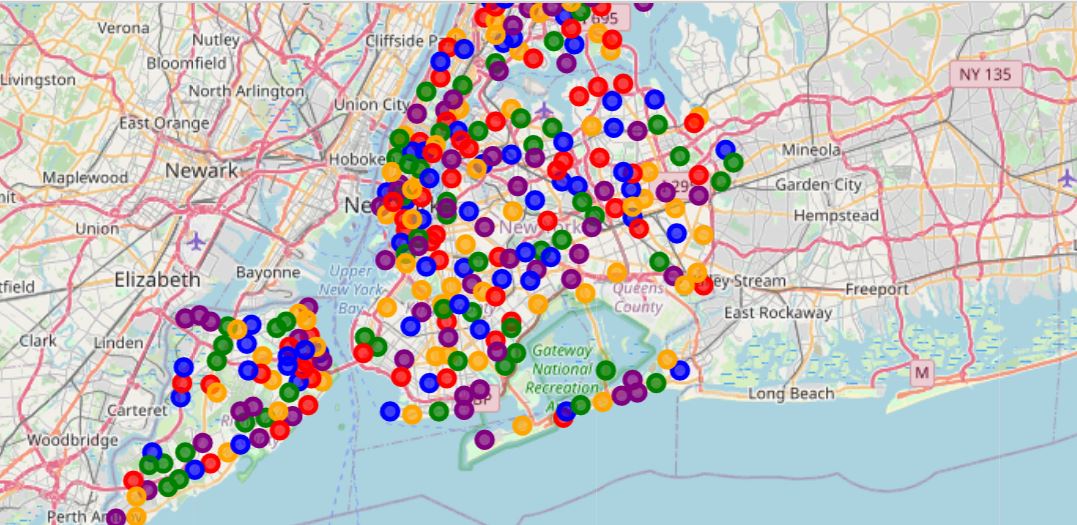
While in Toronto the Neighborhoods reduce significantly and grow wider as we go apart from the epicenter.

****

After this I clustered the points in both the cities into a total of 5 clusters each. The result was calculated through K-means clustering on the basis of 10 most common venues using the FourSquare website and hence an expected result was achieved.



In New York the clustering was even throughout the 5 of them while in Toronto the clustering was random.



**4. Result**

**----------------------------------------------------------**

So the result of this project is that New York has a more dense population and chances of growing business with high competition as well as high variety of customers while Toronto has less densely populated and can start a business as there is no competition but also a low variety of people available.

But for sure we can state these cities are not the same, they haven no such similarities and are totally different in Nature.

**5. Discussions**

**----------------------------------------------------------**



Toronto



New York

In both the diagrams you can evidently see that the most frequently visited locations are very different and these venues can give rise to very logical conclusions that is how the people of these cities' lifestyles are. That is how these people think and live.

**6. Conclusions**

**----------------------------------------------------------**

In conclusion I would like to say that both the cities have no such similarity on the basis of geographical presentation of venues and neighborhood.

Both the cities have various features making them special on their own with different kinds of work areas and possibilities.