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.

	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor , Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park , Ontario Provincial Government	43.662301	-79.389494

:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

2.3 Data Preparation and Understanding

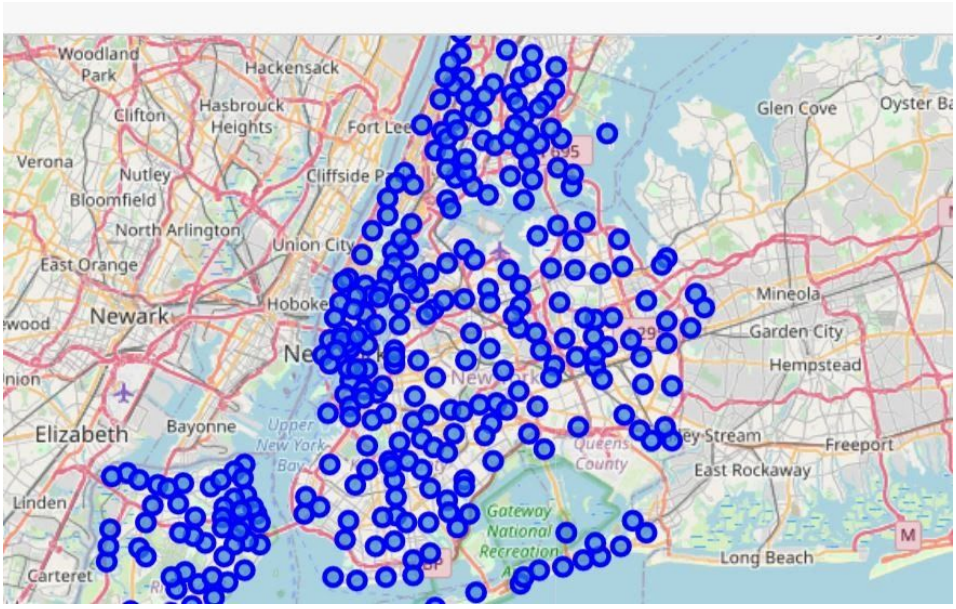
The data of New York and Toronto can be used into a FourSquare account and a lot of useful information about neighborhoods, venues there tips, etc can be received from their servers. These venues can be clustered, visualized and compared to give great results.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Parkwoods	43.753259	-79.329656	TTC stop - 44 Valley Woods	43.755402	-79.333741	Bus Stop
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

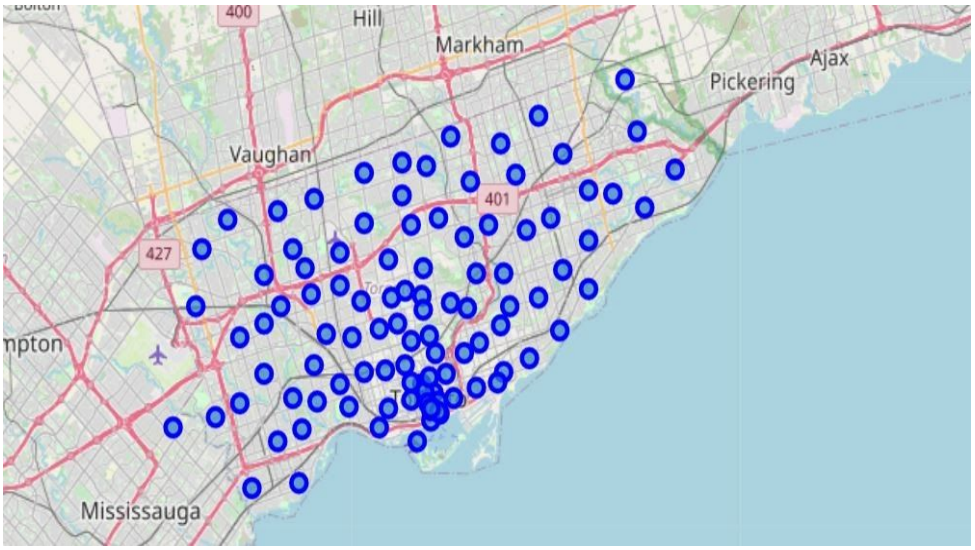
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

3. Methodology

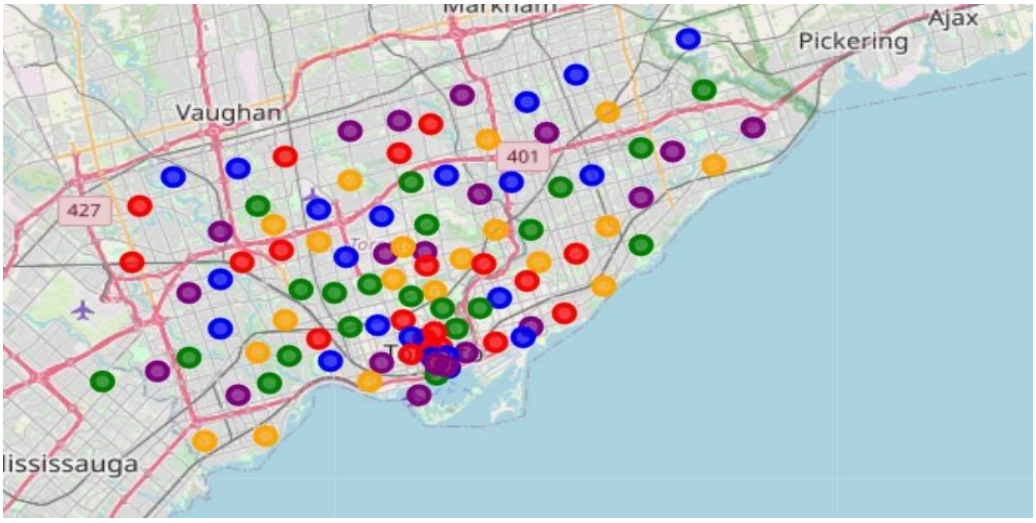
So here we can already see the neighborhoods are more intense in New York City than in Toronto. Not only that the density is high and nearly constant throughout the city of New York, while in Toronto it is concentrated around a single point and slowly disperses away as we go far.
This here can show us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



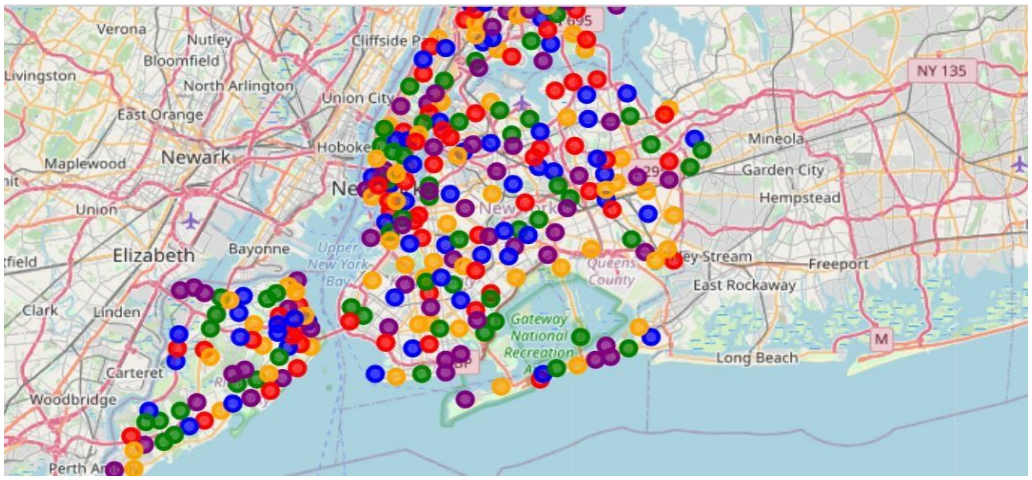
While in Toronto the Neighborhoods reduce significantly and grow wider as we go apart from the epicenter.



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

	Postal code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th
0	M3A	North York	Parkwoods	43.753259	-79.329656	0.0	Park	Food & Drink Shop	Bus Stop	Women's Store	Dim Sum Restaurant	Dance Studio	Deli / Bodega	Depart
1	M4A	North York	Victoria Village	43.725882	-79.315572	1.0	Coffee Shop	Hockey Arena	Grocery Store	Portuguese Restaurant	Dim Sum Restaurant	Dance Studio	Deli / Bodega	Depart
2	M5A	Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636	1.0	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Restaurant	Café	Th
3	M6A	North York	Lawrence Manor , Lawrence Heights	43.718518	-79.464763	1.0	Clothing Store	Furniture / Home Store	Accessories Store	Coffee Shop	Vietnamese Restaurant	Boutique	Miscellaneous Shop	
4	M7A	Downtown Toronto	Queen's Park , Ontario Provincial Government	43.662301	-79.389494	1.0	Coffee Shop	Sushi Restaurant	Diner	Yoga Studio	Beer Bar	Japanese Restaurant	Café	H

Toronto

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th M
0	Bronx	Wakefield	40.894705	-73.847201	0.0	Pharmacy	Donut Shop	Ice Cream Shop	Dessert Shop	Gas Station	Sandwich Place	Laundromat	Women's Store	Ev
1	Bronx	Co-op City	40.874294	-73.829939	3.0	Baseball Field	Bus Station	Restaurant	Park	Basketball Court	Liquor Store	Bagel Shop	Gift Shop	Sal
2	Bronx	Eastchester	40.887556	-73.827806	3.0	Caribbean Restaurant	Deli / Bodega	Diner	Bus Station	Intersection	Bowling Alley	Seafood Restaurant	Fast Food Restaurant	Do
3	Bronx	Fieldston	40.895437	-73.905643	3.0	Bus Station	River	Plaza	Women's Store	Field	Ethiopian Restaurant	Event Service	Event Space	Exh
4	Bronx	Riverdale	40.890834	-73.912585	3.0	Bus Station	Park	Bank	Plaza	Gym	Medical Supply Store	Baseball Field	Home Service	Playgrou

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.