SIMILAR NEIGHBOURHOODS BETWEEN MANHATTAN, NEW-YORK AND TORONTO

INTRODUCTION:

New York and Toronto, two major cities of the world. Lots of businesses and companies have their offices in and around these cities, which urges lots of people from around the world from all walks of life to come and settle here. Additionally, many people must move among these two cities too, i.e., New York to Toronto or vice versa for job requirements or business needs.

Suppose if a resident of New York must permanently move to Toronto and he/she is searching for a place to live there. That person would like to move to similar area as their current area in New York based on nearby venues including coffee shops or schools or parks or types of restaurants, etc.

So using this project, we can cluster similar areas of New York and Toronto together and the person is able to easily identify the similar areas in Toronto to his/her current area in New York so that the person can make the decision.

DATA:

For this purpose, first we would require the list of all the areas of Toronto and New York. I took the base data from the following URLs:

New York - https://cocl.us/new_york_dataset

Toronto - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This data gives us the list of all the Neighbourhoods and Boroughs.

Out of New York data, I have filtered out Manhattan data which looks like below:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Figure 1 Manhattan Data

For Toronto data, all those boroughs are selected with Toronto in them. Then I have added the Latitude and Longitude data also to the Toronto Data from the file: Geospatial_Coordinates.csv from the link https://cocl.us/Geospatial_data.

Once the data for Toronto is complete, it will look like below:

	Postal Code	Borough	Neighbourhood	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
...	...	...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

Figure 2Toronto Data

We will drop the Postal Code column from Toronto Data so that it is in the same shape as of Manhattan data.

Once this data is prepared, it will be clubbed for both the cities. so that the clustering can be done together. This is done because we want to cluster the combined data so that combined clusters can be formed for New York and Toronto neighbourhoods.

So, once the data for both the cities are clubbed together, the respective latitude and longitude of all the neighbourhoods is put in the Foursquare API to get the nearby venues.

METHODOLOGY:

Now, since the base data is prepared, and we have got the venues data also for all the Neighbourhoods, we can start exploring the data and try to see any important patterns from it.

Firstly, one hot encoding is done on the data based on the venues and then the data is modified to have the top 10 most common venues for all the neighbourhoods. This data looks like below:

	Neighbourhood	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 Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Gym	Memorial Site	Plaza	Wine Shop	Boat or Ferry	Gourmet Shop	Food Court
1	Berczy Park	Coffee Shop	Cocktail Bar	Café	Cheese Shop	Pub	Beer Bar	Restaurant	Bakery	Seafood Restaurant	Breakfast Spot
2	Brockton, Parkdale Village, Exhibition Place	Café	Coffee Shop	Breakfast Spot	Bakery	Restaurant	Grocery Store	Gym	Climbing Gym	Italian Restaurant	Performing Arts Venue
3	Business reply mail Processing Centre, South C...	Yoga Studio	Gym / Fitness Center	Comic Shop	Recording Studio	Restaurant	Park	Skate Park	Burrito Place	Farmers Market	Fast Food Restaurant
4	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Plane	Boutique	Sculpture Garden	Coffee Shop	Harbor / Marina	Rental Car Location	Boat or Ferry

Figure 3 Top 10 most common venues for all neighbourhoods

The basic methodology followed behind this process is, if we club the neighbourhood data of 2 different cities and get clusters out of them, it will cluster the similar neighbourhoods of both the cities together on the basis of their venues. Hence, one can find his/her area in the clusters and can see the neighbourhoods of the other city falling under the same cluster. This will help the person to filtering the area for searching a place to stay.

As discussed above, both the cities neighbourhood data is clubbed together already, and top 10 most common venues have been filtered and a data frame has been formed. Now we can simply apply K-means clustering on this data to form different clusters based on the venues data. The optimum value of K can be identified using elbow method. Here, we have taken the value of K as 5 which gives the optimum result.

Once the clusters are formed, we can add a column specifying the cluster each neighbourhood belongs to in the final data. This will help us identify each cluster. Now we can filter out all four clusters separately.

	Borough	Neighbourhood	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 Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	3	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Miscellaneous Shop	Steakhouse	Shopping Mall	Supplement Shop	Seafood Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	3	Chinese Restaurant	Bakery	Cocktail Bar	Bubble Tea Shop	Vietnamese Restaurant	Salon / Barbershop	Coffee Shop	Ice Cream Shop	Optical Shop
2	Manhattan	Washington Heights	40.851903	-73.936900	3	Café	Bakery	Grocery Store	Chinese Restaurant	Mobile Phone Shop	Gym	Mexican Restaurant	Supermarket	Sandwich Place
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Lounge	Restaurant	Café	Bakery	Deli / Bodega	Frozen Yogurt Shop	Chinese Restaurant	Caribbean Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	3	Pizza Place	Café	Coffee Shop	Mexican Restaurant	Deli / Bodega	Yoga Studio	Caribbean Restaurant	Sushi Restaurant	Bakery

Figure 4 CLuster Labels column depicting specific cluster

Mapping the clusters will give following result:

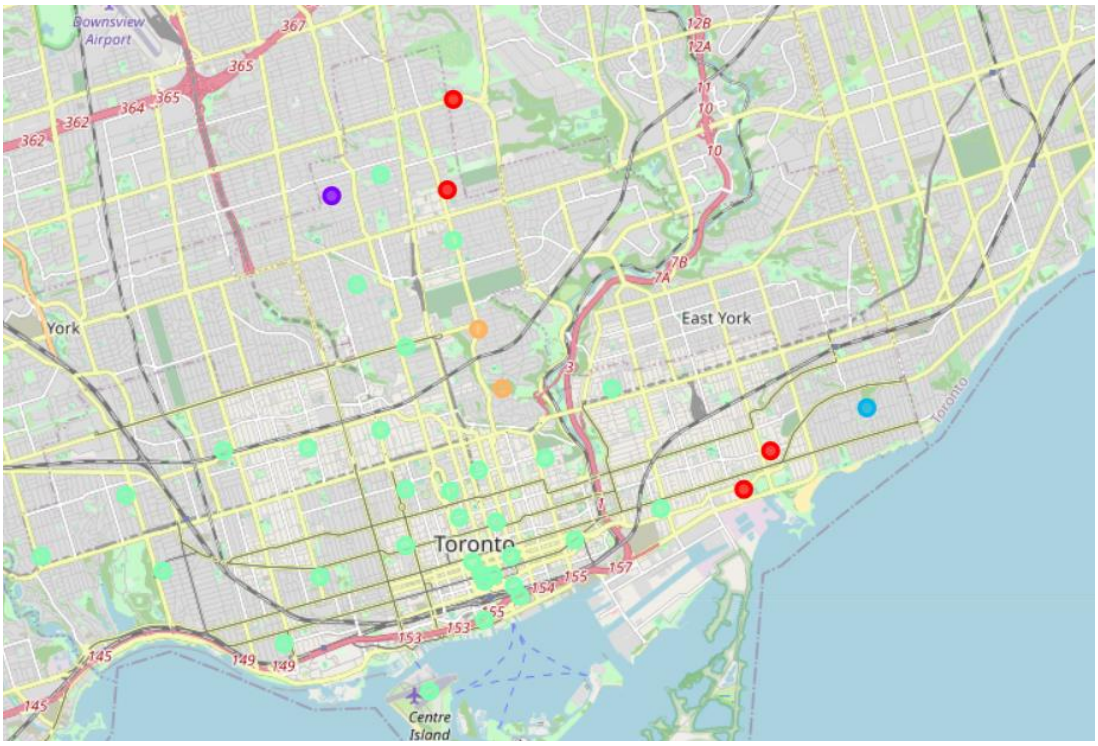

Figure 5 Toronto clusters

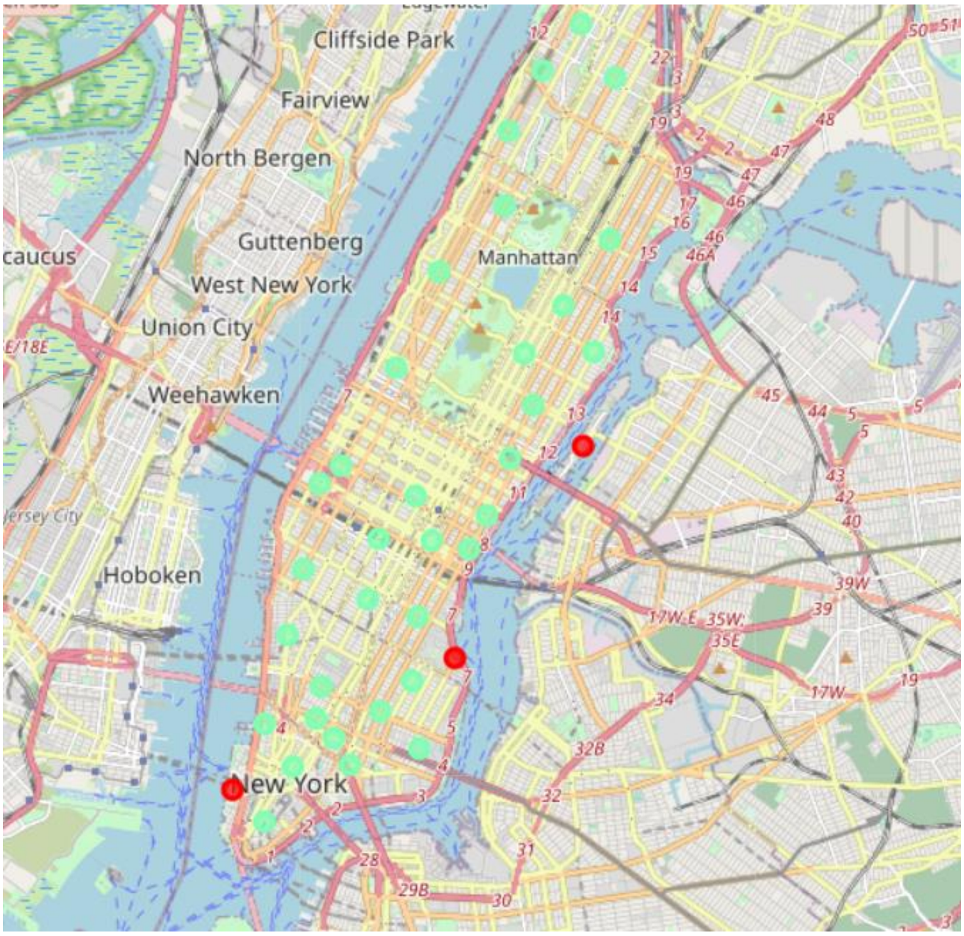

Figure 6 Manhattan, New York Clusters.

RESULTS:

As we can see from this process, five different clusters were formed for the neighbourhoods of both the areas. Out of which, 2 clusters have most of the common areas of both the cities which are like each other which are marked with read and light green. The resulting clusters example are below:

	Neighbourhood	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 Most Common Venue	10th Most Common Venue
11	Roosevelt Island	Park	Playground	Liquor Store	Japanese Restaurant	Residential Building (Apartment / Condo)	Restaurant	Baseball Field	Sandwich Place	Scenic Lookout	School
28	Battery Park City	Park	Coffee Shop	Hotel	Gym	Memorial Site	Plaza	Wine Shop	Boat or Ferry	Gourmet Shop	Food Court
37	Stuyvesant Town	Park	Playground	Heliport	Baseball Field	Gas Station	Gym / Fitness Center	Cocktail Bar	Harbor / Marina	Pet Service	Fountain
55	India Bazaar, The Beaches West	Park	Fast Food Restaurant	Fish & Chips Shop	Italian Restaurant	Pet Store	Pub	Restaurant	Sandwich Place	Movie Theater	Burrito Place
58	Lawrence Park	Bus Line	Park	Swim School	Yoga Studio	Drugstore	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Egyptian Restaurant	Electronics Store
60	Davisville North	Hotel	Breakfast Spot	Pizza Place	Food & Drink Shop	Sandwich Place	Gym / Fitness Center	Park	Department Store	Electronics Store	Donut Shop
78	Business reply mail Processing Centre, South C...	Yoga Studio	Gym / Fitness Center	Comic Shop	Recording Studio	Restaurant	Park	Skate Park	Burrito Place	Farmers Market	Fast Food Restaurant

Figure 7 Cluster 1

	Neighbourhood	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 Most Common Venue	10th Most Common Venue
59	Roselawn	Home Service	Garden	Yoga Studio	English Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Egyptian Restaurant

Figure 8 Cluster 2

	Neighbourhood	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 Most Common Venue	10th Most Common Venue
44	The Beaches	Pub	Trail	Health Food Store	Neighborhood	Yoga Studio	Empanada Restaurant	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant

Figure 9 Cluster 3

	Neighbourhood	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 Most Common Venue	10th Most Common Venue
0	Marble Hill	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Miscellaneous Shop	Steakhouse	Shopping Mall	Supplement Shop	Seafood Restaurant	Donut Shop
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Bubble Tea Shop	Vietnamese Restaurant	Salon / Barbershop	Coffee Shop	Ice Cream Shop	Optical Shop	American Restaurant
2	Washington Heights	Café	Bakery	Grocery Store	Chinese Restaurant	Mobile Phone Shop	Gym	Mexican Restaurant	Supermarket	Sandwich Place	Bank
3	Inwood	Mexican Restaurant	Lounge	Restaurant	Café	Bakery	Deli / Bodega	Frozen Yogurt Shop	Chinese Restaurant	Caribbean Restaurant	American Restaurant
4	Hamilton Heights	Pizza Place	Café	Coffee Shop	Mexican Restaurant	Deli / Bodega	Yoga Studio	Caribbean Restaurant	Sushi Restaurant	Bakery	School
5	Manhattanville	Coffee Shop	Seafood Restaurant	Bus Stop	Italian Restaurant	Mexican Restaurant	Chinese Restaurant	Park	Bus Station	BBQ Joint	Fried Chicken Joint
6	Central Harlem	African Restaurant	American Restaurant	Cosmetics Shop	Chinese Restaurant	French Restaurant	Bar	Seafood Restaurant	Market	Juice Bar	BBQ Joint
7	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Deli / Bodega	Latin American Restaurant	Park	Spa	Sandwich Place	Liquor Store	Taco Place
8	Upper East Side	Italian Restaurant	Bakery	Gym / Fitness Center	Coffee Shop	Juice Bar	Spa	French Restaurant	Exhibit	Yoga Studio	Wine Shop
9	Yorkville	Coffee Shop	Italian Restaurant	Gym	Bar	Deli / Bodega	Sushi Restaurant	Wine Shop	Pizza Place	Japanese Restaurant	Mexican Restaurant
10	Lenox Hill	Coffee Shop	Italian Restaurant	Pizza Place	Sushi Restaurant	Cocktail Bar	Café	Gym / Fitness Center	Gym	Burger Joint	Thai Restaurant

Figure 10 Cluster 4

	Neighbourhood	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 Most Common Venue	10th Most Common Venue
69	Moore Park, Summerhill East	Park	Trail	Restaurant	Yoga Studio	Empanada Restaurant	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Duty-free Shop
73	Rosedale	Park	Playground	Trail	Yoga Studio	Empanada Restaurant	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Duty-free Shop

Figure 11 Cluster 5

We can see from the data that Cluster 1 and Cluster 4 are the clusters with similar areas from both the cities. Also, we can notice from the most common venues in these clusters that Cluster 1 has more of parks, sports fields, activity centres like swimming pool, yoga studios, gyms, and a few cafes.

While in Cluster 4, the most common venues are various restaurants, bakeries, pubs, etc. So, we can see a clear difference between these two neighbourhoods and a person can decide the area in the other city based on the Clusters in which his/her current area of current city is fitting.

It also can happen that there is no similar area to the current area but then too looking at the other clusters with most common areas can help him/her making an informed decision.

DISCUSSION:

One important thing which can be observed is that there are few areas which are not matching with the ones with the other city at all. While this is completely possible as these are two different countries all together, but probably more data can be taken to bring out some similarities if possible.

Other parameters can also be looked upon for any areas like conveyance, etc., which can help further strengthen the decision of choosing an area.

CONCLUSION:

After doing this comparison, we can conclude that this clustering can help someone to discover areas in other cities to stay. Using the Foursquare API and neighbourhood data, one can do this operation on any two cities to find the common neighbourhoods among them. One can even apply this any number of cities together to figure out similar areas among multiple cities.