Comparison of Neighborhoods of New York and Toronto

INTRODUCTION

New York and Toronto are both most populous cities in the Unites States and Canada respectively. New York - the most densely populated city in the US, a Global power city which has been described as the Cultural, Financial and media capital of the World. On the other hand, Toronto is also a global city which lies as the centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Both being the populous cities, tourism world is always interested in these places. Lets not get deeper into the tourism world but will analyze the attractive venues in and around the neighborhood of both cities to get to know well about the places.

With this analysis we will get to know which one is better over the other or which city has more venues in nearby distance and details on those venues in which one may be interested to have a visit. We will get to know if both cities are similar with respect to these venues or totally different!!!.

DATA

We need to follow first 3 basic steps here

1. What Data is required?
2. From where to Collect the Data?
3. Is data collected is sufficient and is in correct format? If not, what next?

What Data is required?

To do the comparison of NY and Toronto Neighborhoods, we first need the neighborhood details of NY and Toronto. All the neighborhoods with their Boroughs and postalcodes to get the proper location of the each neighbourhood. Then the venues in those neighbourhoods. Type of venues, number of venues and details on those venues.

From where to collect the data?

NY and Toronto neighbourhood details are readily available in various websites. We can get the details directly from the trusted websites. Regarding the venues in each neighbourhood, **FourSquare API** comes handy. This has all the venue details in each neighbourhood.

Verify the data

Next step is to verify if the data is sufficient enough to start the analysis. Yes, for our analysis these data is fair enough. Now check the data format. Is it in proper format to start the processing. Partially Yes. Because NY data is in proper format which can be processed. But Toronto data is available in Wikipedia page from where we have to segregate and pick the required neighborhood data alone. So, pre-processing of data is required. Once the pre-processing is done, we will have both dataframes ready to start the processing and analysis.

METHODOLOGY

As a first step, we will extract the NY and Toronto Neighborhood data.

NY:

Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. This data set is readily available from https://geo.nyu.edu/catalog/nyu_2451_34572. We will download the data and transform it to pandas dataframe. Below is the sample of dataframe created which has Borough Bronx with the latitude and longitude values.

	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

Fig3.1. NY Dataframe

Let's create a map of NewYork using **Folium** Maps to visualize the neighbourhoods

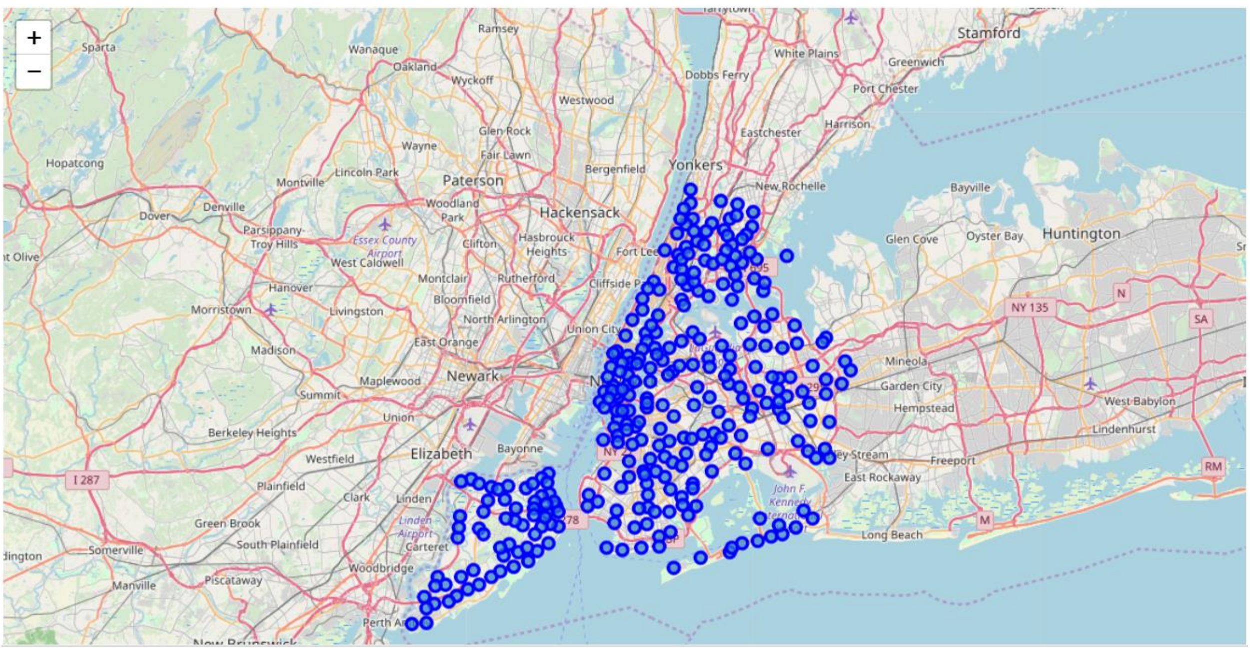


Fig3.2. NewYork Map with Neighborhoods superimposed on top

We can see from the map that there are plenty of neighborhoods. Let's simplify the map and cluster only the neighborhoods of Manhattan only.

Following is the refined dataframe that has Manhattan data.

	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

Fig3.3. Manhattan Dataframe

Let's now create the map of Manhattan with its neighborhoods

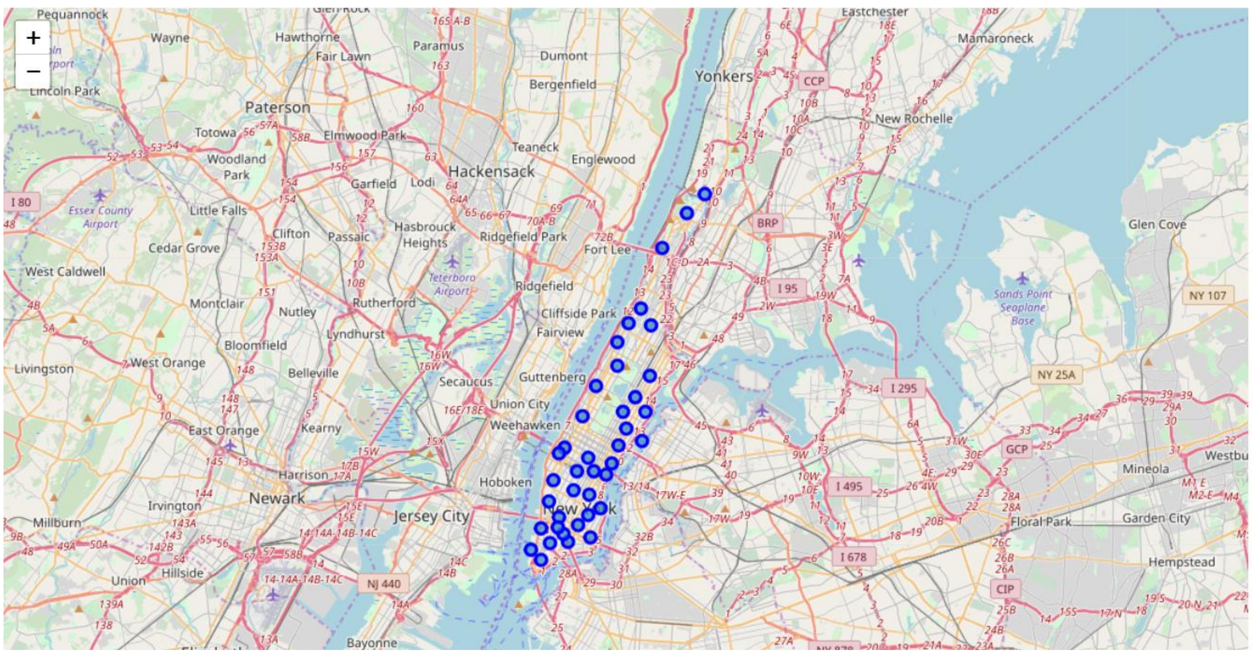


Fig3.4. Manhattan Map with Neighborhoods superimposed on top

We will follow the same steps to get the Neighborhood visualization of Toronto.

Toronto:

Like, NY we don't have Toronto Neighborhood data readily available. But we have got a Wikipedia page(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)where there is a list of Toronto places presented in tabular format. So first step here is to segregate the table data from Wikipedia page and transfer into readable pandas dataframe. After data cleaning and putting into dataframe, Toronto data looks like below

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Fig3.5. Toronto Dataframe(Sample)

Next step would be fetching the latitude and longitude values of all the neighborhoods and adding into the existing dataframe.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Fig3.6. Toronto Dataframe(With Latitude and Longitude values)

Let’s now create a visualization of Toronto Neighborhood using Folium Maps.

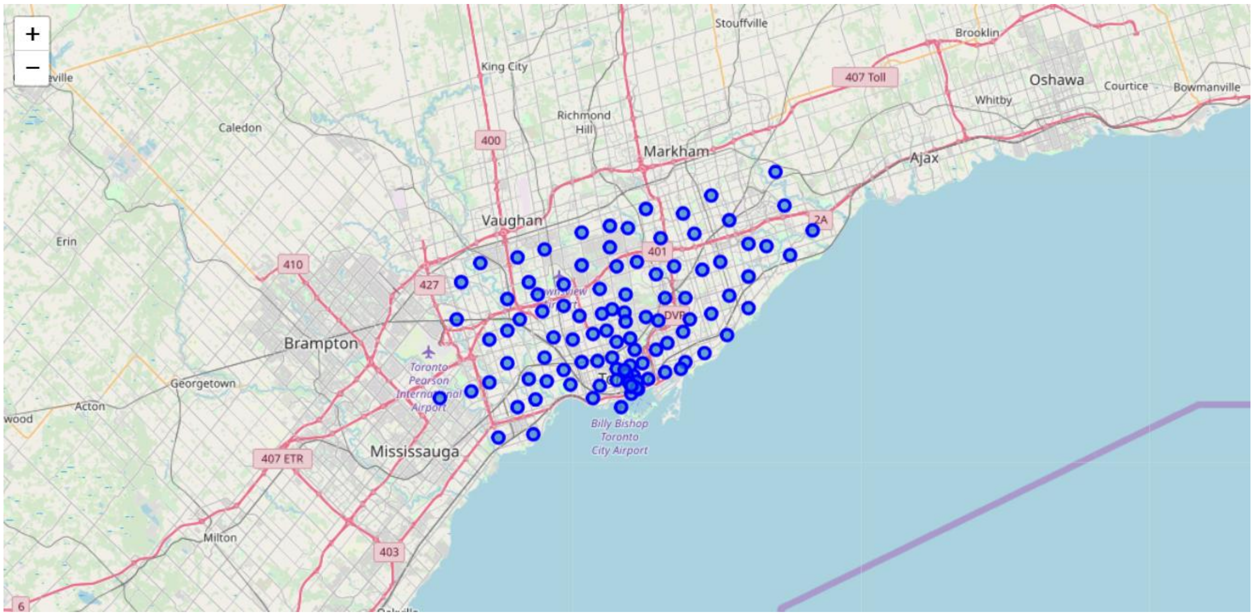


Fig3.7.Toronto Map with Neighborhoods superimposed on top

ANALYSIS

Let us now begin the analysis part.

We will use k-means to cluster the neighborhoods(here we will have 5 clusters)

Below dataframe shows the 10 most common venues in the neighborhoods

	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 Most Common Venue	10th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	4	Discount Store	Coffee Shop	Yoga Studio	Deli / Bodega	Supplement Shop	Steakhouse	Shopping Mall	Shoe Store	Seafood Restaurant	Sandwich Place
1	Manhattan	Chinatown	40.715618	-73.994279	0	Chinese Restaurant	Bubble Tea Shop	American Restaurant	Cocktail Bar	Vietnamese Restaurant	Dim Sum Restaurant	Hotpot Restaurant	Salon / Barbershop	Noodle House	Bakery
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Mobile Phone Shop	Pizza Place	Tapas Restaurant	Sandwich Place	Chinese Restaurant	Shoe Store	Mexican Restaurant	Grocery Store
3	Manhattan	Inwood	40.867684	-73.921210	4	Lounge	Mexican Restaurant	Café	Pizza Place	Park	Bakery	Frozen Yogurt Shop	Restaurant	Deli / Bodega	Chinese Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	4	Mexican Restaurant	Deli / Bodega	Coffee Shop	Café	Pizza Place	Liquor Store	Cocktail Bar	Sandwich Place	School	Chinese Restaurant

Fig4.1.NY Dataframe with 10 most common venues

Let’s now visualize the 5 clusters on the map

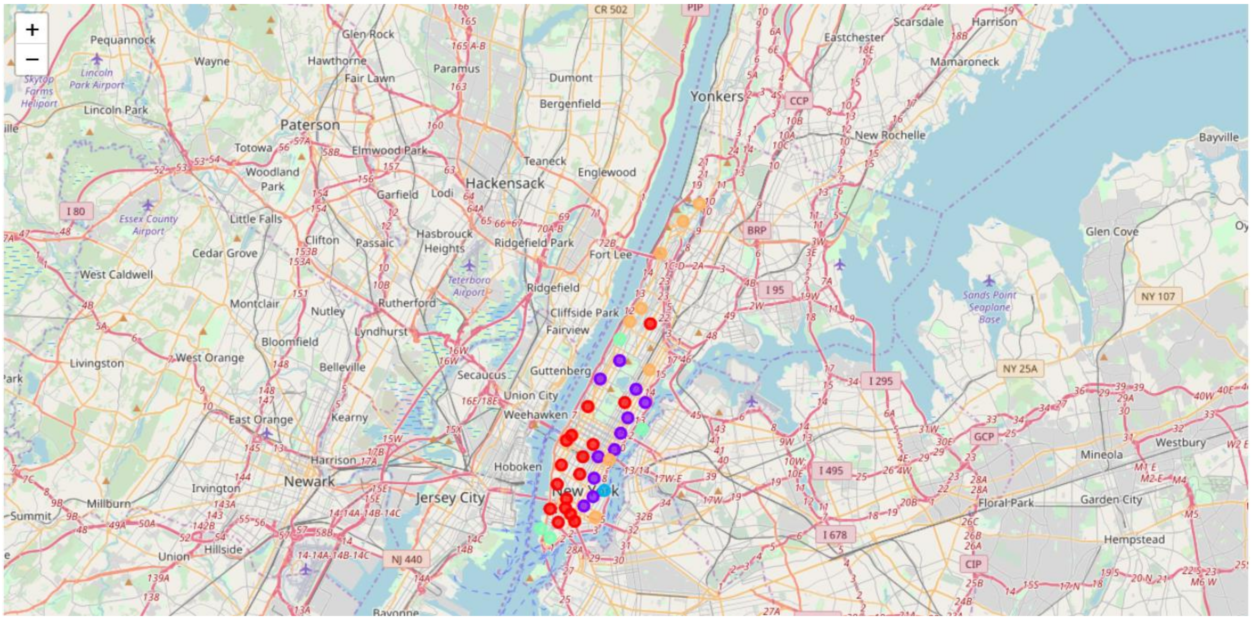


Fig4.2. NY Manhattan map with 5 clusters of neighborhoods grouped together

Different color shows the grouped clusters.

So we have got an overview of the neighborhoods. Lets analyze the venues in those neighborhoods now.

We will be using **FourSquare API** for this

NY:

We will get the venues of Marble Hill(First Neighborhood of NY Manhattan) first. After all processing we get the below dataframe with venue details

	name	categories	lat	lng
0	Arturo's	Pizza Place	40.874412	-73.910271
1	Bikram Yoga	Yoga Studio	40.876844	-73.906204
2	Tibbett Diner	Diner	40.880404	-73.908937
3	Starbucks	Coffee Shop	40.877531	-73.905582
4	Land & Sea Restaurant	Seafood Restaurant	40.877885	-73.905873

Fig4.3. Marble Hill Venues

We will now get the venues(atleast 100 – we can change the desired number of venues to be returned by the FourSquare API) for all the neighborhoods of Manhattan.

```
----Battery Park City----
                venue  freq
0         Coffee Shop  0.08
1                Park  0.07
2               Hotel  0.05
3  Italian Restaurant  0.03
4           Wine Shop  0.03


----Carnegie Hill----
            venue  freq
0     Pizza Place  0.06
1  Cosmetics Shop  0.05
2     Coffee Shop  0.05
3            Café  0.04
4     Yoga Studio  0.03


----Central Harlem----
                  venue  freq
0    African Restaurant  0.07
1        Cosmetics Shop  0.07
2     French Restaurant  0.04
3  Gym / Fitness Center  0.04
4   American Restaurant  0.04
```

Fig4.4. Sample venue data of 3 Neighborhoods of Manhattan

For our analysis, we will be taking first 5 Neighborhood data. There are around 100 venues returned for each neighborhood. We will take only 3 – Restaurant(which includes all types of restaurants in that neighborhood), Coffee shops and Studios for analysis.

	Restaurant	Coffee	Studio
Neighbourhood			
Battery Park City	0.10	0.08	0.00
Carnegie Hill	0.26	0.05	0.04
Central Harlem	0.29	0.00	0.02
Chelsea	0.26	0.07	0.01
Chinatown	0.41	0.02	0.02

Fig.4.5 Venues for analysis

We will draw a bar chart with the sample venues.

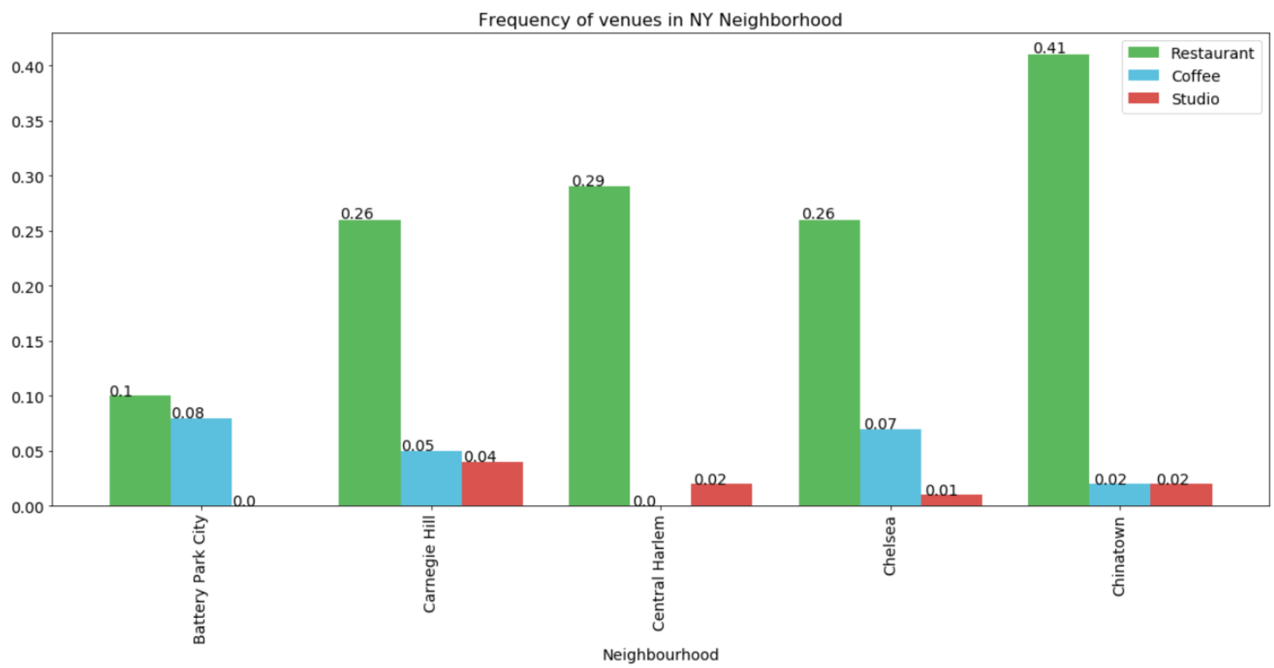


Fig.4.6. Bar chart representing 3 common venues of 5 NY Manhattan Neighborhoods

Toronto:

We will use same k-means to cluster the Toronto neighborhoods(here we will have 5 clusters)

Below dataframe shows the 10 most common venues in the neighborhoods

	Neighborhood	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	Adelaide, King, Richmond	Coffee Shop	Café	American Restaurant	Thai Restaurant	Steakhouse	Gym	Clothing Store	Asian Restaurant	Bakery	Bar
1	Berczy Park	Coffee Shop	Cocktail Bar	Restaurant	Seafood Restaurant	Bakery	Italian Restaurant	Steakhouse	Farmers Market	Café	Pub
2	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Café	Coffee Shop	Pet Store	Climbing Gym	Performing Arts Venue	Stadium	Burrito Place	Bar	Caribbean Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Garden	Park	Pizza Place	Recording Studio	Restaurant	Butcher	Burrito Place	Brewery
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Lounge	Airport Terminal	Airport Service	Boutique	Sculpture Garden	Plane	Boat or Ferry	Airport Gate	Airport Food Court	Airport

Fig4.7.Toronto Dataframe with 10 most common venues

Let’s now visualize the 5 clusters on the map

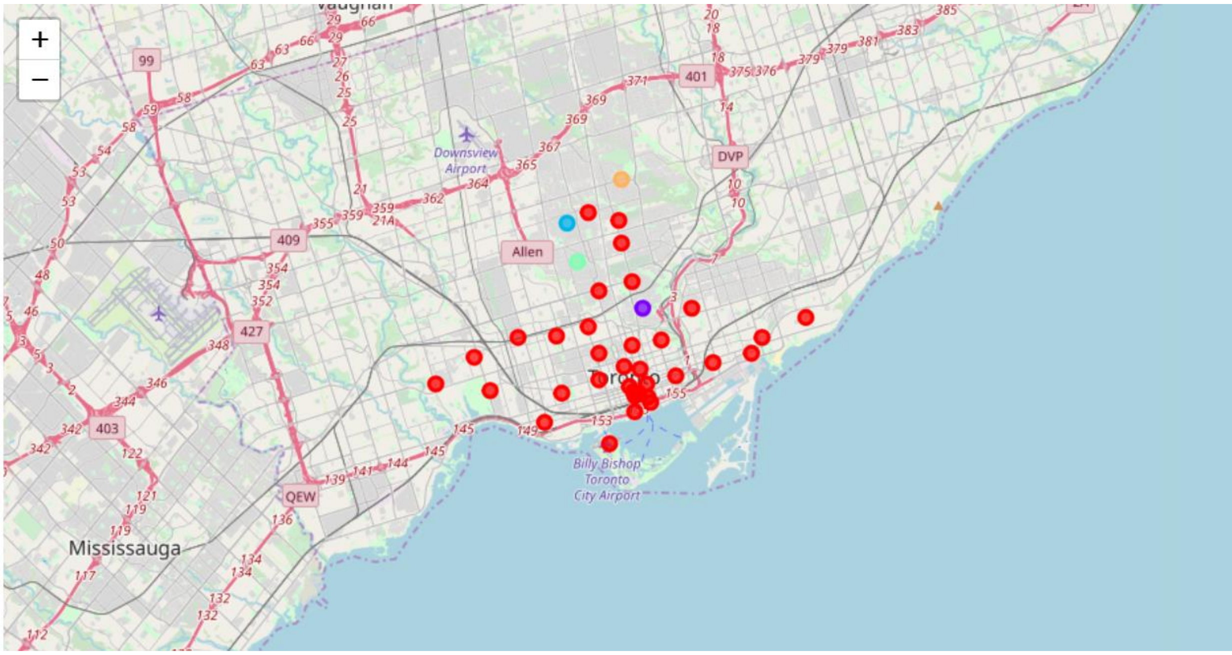


Fig4.8. Toronto map with 5 clusters of neighborhoods grouped together

Different color shows the grouped clusters.

So we have got an overview of the neighborhoods. Lets analyze the venues in those neighborhoods now.

We will be using **FourSquare API** for this

We will get the venues of Toronto Neighborhoods. After all processing we get the below dataframe with venue details

	name	categories	lat	lng
0	The Big Carrot Natural Food Market	Health Food Store	43.678879	-79.297734
1	Grover Pub and Grub	Pub	43.679181	-79.297215
2	Starbucks	Coffee Shop	43.678798	-79.298045
3	Glen Stewart Park	Park	43.675278	-79.294647
4	Upper Beaches	Neighborhood	43.680563	-79.292869

Fig4.9. Toronto Venues

We will now get the venues(atleast 100 – we can change the desired number of venues to be returned by the FourSquare API) for all the neighborhoods of Toronto.

```
----Adelaide, King, Richmond----
                venue  freq
0         Coffee Shop  0.06
1                Café  0.05
2          Steakhouse  0.04
3     Thai Restaurant  0.04
4  American Restaurant  0.04


----Berczy Park----
               venue  freq
0        Coffee Shop  0.07
1       Cocktail Bar  0.05
2         Restaurant  0.05
3  Italian Restaurant  0.04
4     Farmers Market  0.04


----Brockton, Exhibition Place, Parkdale Village----
                  venue  freq
0        Breakfast Spot  0.11
1                  Café  0.11
2           Coffee Shop  0.11
3          Climbing Gym  0.05
4  Performing Arts Venue  0.05
```

Fig4.10. Sample venue data of 3 Neighborhoods of Toronto

For our analysis, we will be taking first 5 Neighborhood data. There are around 100 venues returned for each neighborhood. We will take only 3 – Restaurant(which includes all types of restaurants in that neighborhood), Coffee shops and Studios for analysis.

Neighbourhood	Restaurant	Coffee	Studio
Adelaide, King, Richmond	0.27	0.06	0.00
Berczy Park	0.23	0.07	0.00
Brockton, Exhibition Place, Parkdale Village	0.10	0.11	0.00
Business Reply Mail Processing Centre 969 Eastern	0.12	0.00	0.12
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	0.00	0.00	0.00

Fig.4.11 Venues for analysis

We will draw a bar chart with the sample venues.

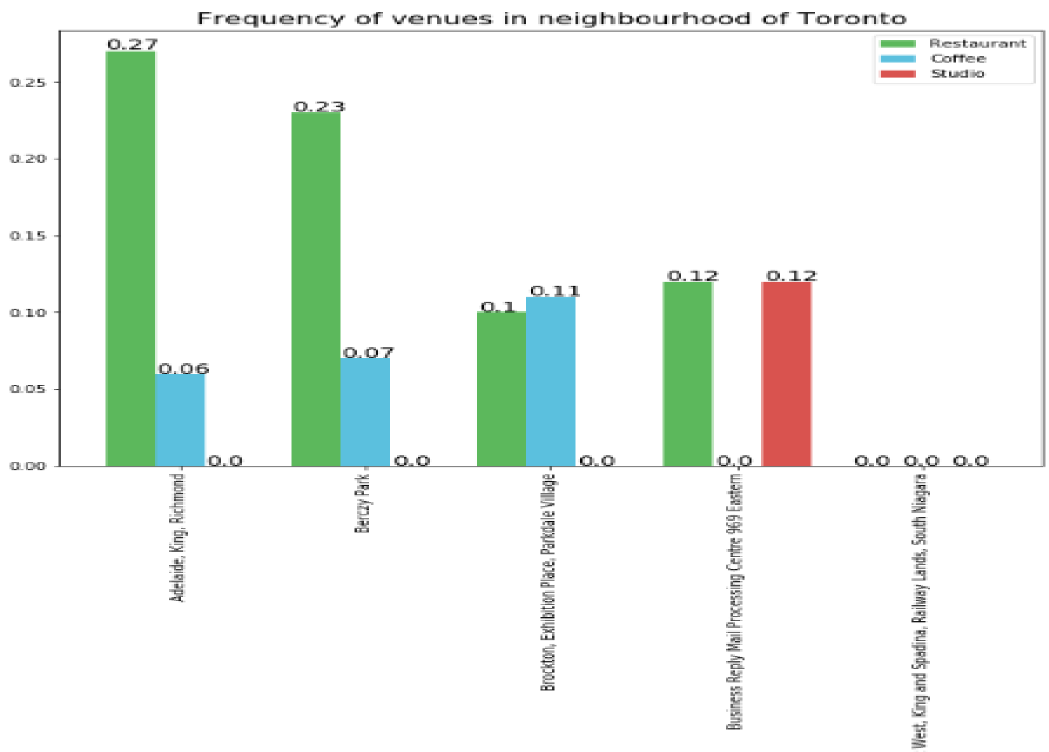
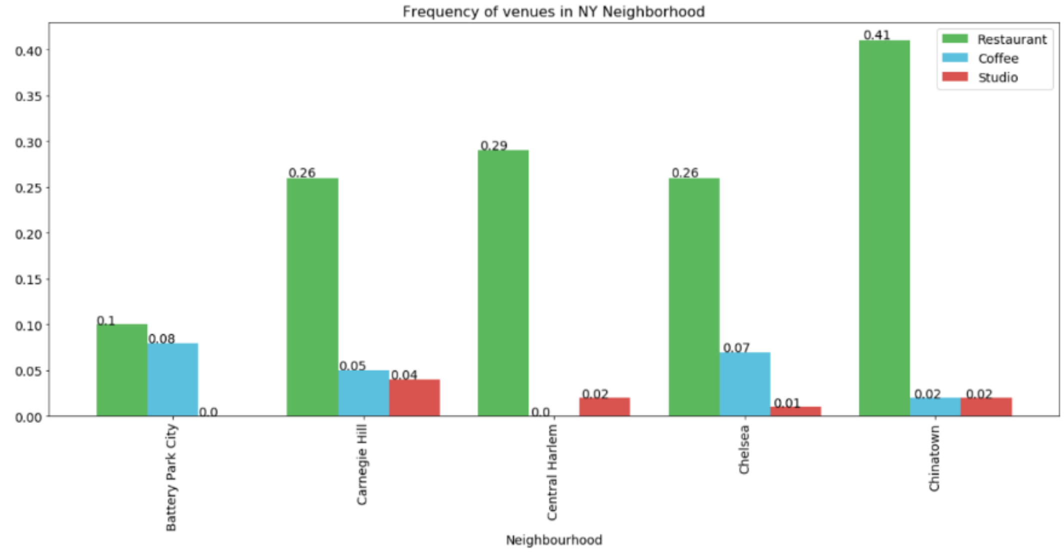
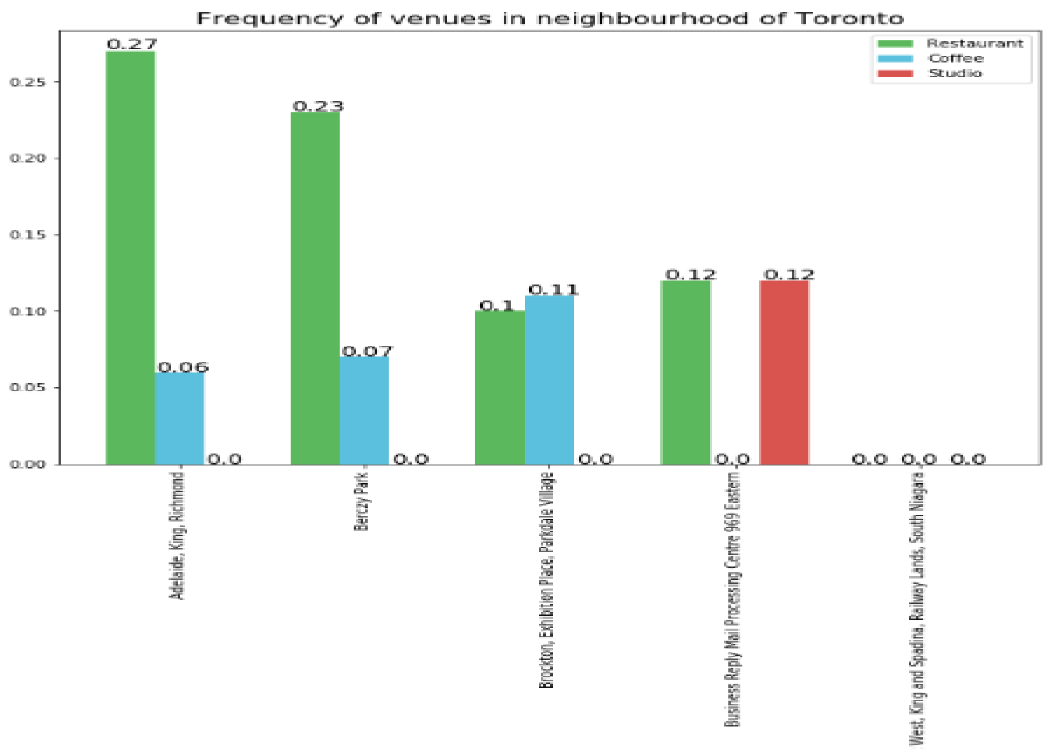


Fig.4.12. Bar chart representing 3 common venues of 5 Toronto Neighborhoods

RESULT

Lets compare the Bar chart that we generated for the 3 common venues of both Neighborhoods for comparison.





We had taken 3 most common venues – Restaurants, Coffee shops and Studios for our comparison study.

From the above bar graphs, we see that Restaurants(which includes all types of restaurants) are equally distributed in both the city neighbourhoods. So taking Restaurant as comparison parameter, we can say both cities are similar

Taking Coffee shops are comparison parameter, we see that Toronto has more coffee shops in its neighborhoods than Manhattan NY. We can say its somewhat dissimilar with this criteria.

Taking Studios(Yoga,Cycle,Dance,Music,et.,) as comparison parameter, we see NY neighborhoods has evenly distributed studios though less in number but Toronto has more number of studios in one particular neighborhood only. So we can say both cities are dissimilar regards to studios.

CONCLUSION:

Comparing New York and Toronto neighborhoods and the most common venues in those neighborhoods, we can conclude that both cities are similar with respect to few venues like restaurants and dissimilar with respect to other venues like Studios. We can take more venues as comparison parameter and repeat the process to get a better understanding of the similarity of the both cities.