1. Introduction and Business Problem

Introduction:

People seeking better career opportunities migrate from one city to another city. It is often difficult for them to find a neighborhood which is nearby to their new work location and suitable for their family in the new city. Most of the cases people love their current neighborhood because of its amenities, venues and tourist spots, so they look for neighborhood which is more similar to their current neighborhood.

Business Problem:

Today migration from one place to another place for better career opportunities is a most common phenomenon. Such a drastic change in the neighborhood from their currently habituated neighborhood is usually difficult for an individual and his family to adjust. On a positive note every city has its own unique and diverse neighborhoods which usually people like to explore. We can find aspects in which both the cities are similar based on the information like venues, tourist spots, restaurants etc. The goal is to analyze neighborhoods in cities and find similar neighborhoods to help people migrating to new cities to live comfortably as they are doing in their current city. In my case study, A person from Toronto wants to migrate to New York seeking a better career opportunity. By, Analyzing the Current Neighborhood of that person in Toronto we should be able to suggest a Similar Neighborhood for that person in New York.

2. Data Acquisition and Data Cleaning

Data Acquisition:

1. The data on Neighborhoods and Boroughs in New York City is obtained from an open source spatial data repository. Link - https://geo.nyu.edu/catalog/nyu_2451_34572.

- Borough – Name of the Borough
- Neighborhood – Name of the Neighborhood in Borough
- Latitude – Latitude of the Neighborhood
- Longitude – Longitude of the Neighborhood

2. The data on Neighborhoods and Borough in Toronto the information is taken from the Wikipedia list. Link - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Borough – Name of the Borough
- Neighborhood – Name of the Neighborhood in Borough

- Postal Code – Postal Code of the Neighborhood in Borough

3. The Geo Spatial Coordinates CSV of the Toronto Neighborhoods and Borough is obtained from different dataset of Latitude and Longitude of different Postal Codes.

- Postal Code – Postal Codes of the Neighborhoods
- Latitude – Latitude of the Neighborhoods
- Longitude – Longitude of the Neighborhoods

4. The information such as venues, restaurants, parks, hotels etc. is collected using the Foursquare API for the neighborhoods in Toronto and New York. Foursquare API provides location-based experiences with information like venues, check-ins. Foursquare API provides the venue information in the JSON format.

- name – Name of the Venue in the Neighborhood
- categories – Type of the Venue in the Neighborhood
- lat – Latitude of the Venue in the Neighborhood
- lng – Longitude of the Venue in the Neighborhood

Data Cleaning:

1. The data on Neighborhoods and Boroughs in New York City is obtained in the JSON Format (See Fig 2.1). so pandas.io.json is used to transform this JSON data into Pandas Dataframe (See Fig 2.2).

```
{'type': 'FeatureCollection',
 'totalFeatures': 306,
 'features': [{'type': 'Feature',
   'id': 'nyu_2451_34572.1',
   'geometry': {'type': 'Point',
    'coordinates': [-73.84720052054902, 40.89470517661]},
   'geometry_name': 'geom',
   'properties': {'name': 'Wakefield',
    'stacked': 1,
    'annoline1': 'Wakefield',
    'annoline2': None,
    'annoline3': None,
    'annoangle': 0.0,
    'borough': 'Bronx',
    'bbox': [-73.84720052054902,
     40.89470517661,
     -73.84720052054902,
     40.89470517661]}},
```

Figure 2.1 New York City JSON format data from Spatial Data Repository

	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

Figure 2.2 Panda DataFrame of Neighborhoods and Borough of New York

2. The data of Neighborhoods of Toronto is obtained by Web Scraping the List of Postal Code of Canada : M using Beautiful Soap python package. Beautiful Soap is a popular python package for web scraping and parsing HTML files. The Resultant Pandas DataFrame of Neighborhood of Toronto is created from the Wikipedia lists web page. (See Fig 2.3).

	Postal code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Figure 2.3 Pandas DataFrame of Neighborhoods and Borough of Toronto

3. The Geo Spatial Coordinates CSV is loaded into Pandas Dataframe to obtain the latitude and longitude information of Postal Code of Borough and Neighborhoods of Toronto (See Fig 2.4).

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Figure 2.4 Geo Spatial Coordinates CSV of Toronto converted to Pandas Dataframe

The Pandas Dataframe on Borough and Neighborhoods of Toronto is merged with the Geo Spatial Coordinates CSV of Toronto to obtain a single Dataframe consisting Neighborhoods and Borough of Toronto and their respective Latitude and Longitude information (See Fig 2.5).

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304
11	M1R	Scarborough	Wexford, Maryvale	43.750072	-79.295849

Figure 2.5 Dataframe of Neighborhoods of Toronto with corresponding Latitude and Longitude

4. The FourSquare API venue information is obtained through API call using requests python package through FourSquare API credentials. The obtained result is JSON using pandas.io.json the JSON is transformed into a pandas Dataframe (See Fig 2.6 and Fig 2.7).

	name	categories	lat	lng
0	Rosedale Park	Playground	43.682328	-79.378934
1	Whitney Park	Park	43.682036	-79.373788
2	Alex Murray Parkette	Park	43.678300	-79.382773
3	Milkman's Lane	Trail	43.676352	-79.373842

Figure 2.7 Venues in Neighborhood and their Categories DataFame Toronto

	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	Dunkin'	Donut Shop	40.877136	-73.906666

Figure 2.7 Venues in Neighborhood and their Categories DataFame New York

5. The Final Dataframe after combining the FourSquare API venue data with the Dataframe of Neighborhoods and Borough of Toronto and New York Cities Seperately and renaming the Columns of the resulting Dataframe { name – Venue , lat – Venue Latitude , lng – Venue Longitude and categories – Venue Category }. The Dataframe obtained for Newyork after keeping only Manhattan Borough (See Fig. 2.8) and The Dataframe obtained for Toronto after keeping only Downtown Toronto Borough (See Fig. 2.9)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Figure 2.8 Manhattan Borough Venues Dataframe

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rosedale	43.679563	-79.377529	Rosedale Park	43.682328	-79.378934	Playground
1	Rosedale	43.679563	-79.377529	Whitney Park	43.682036	-79.373788	Park
2	Rosedale	43.679563	-79.377529	Alex Murray Parkette	43.678300	-79.382773	Park
3	Rosedale	43.679563	-79.377529	Milkman's Lane	43.676352	-79.373842	Trail
4	St. James Town, Cabbagetown	43.667967	-79.367675	Cranberries	43.667843	-79.369407	Diner

Figure 2.9 DownTown Toronto Borough Venues Dataframe

3. Methodology

Data Analysis:

Separate Data Analysis is performed for the Venue Data Extracted for the Neighborhoods of Toronto and Neighborhoods of New York. The areas Highlighted in

Blue in the Map are the Neighborhoods of Newyork that is plotted using Folium from the Newyork Borough Neighborhoods Dataframe (See Fig 3.1).
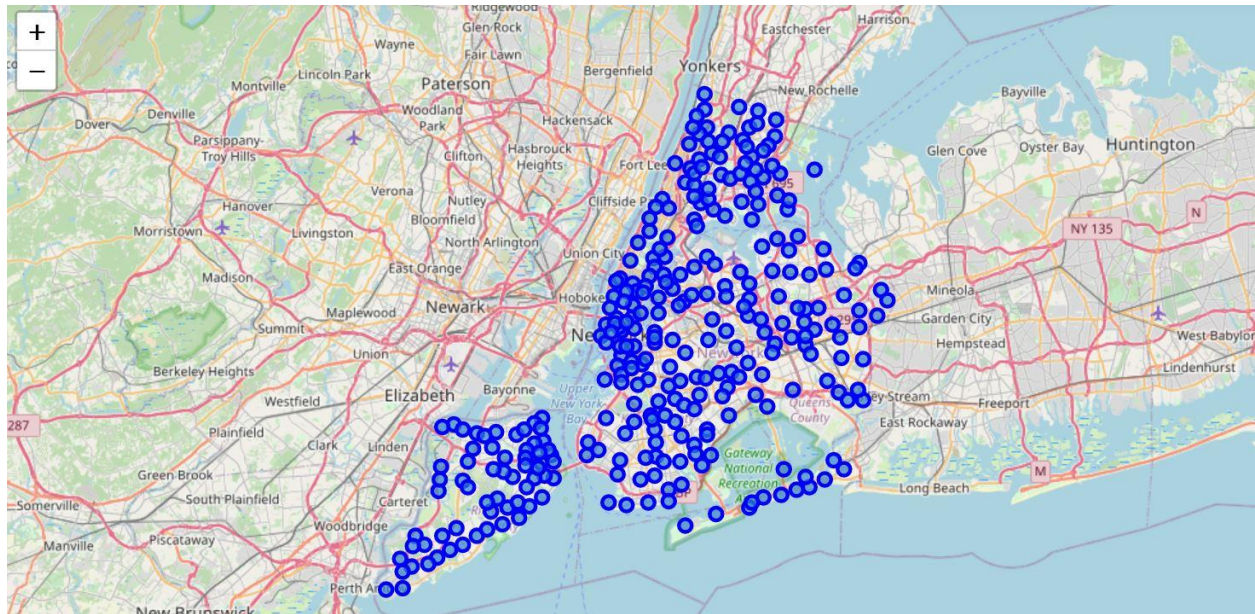

Figure 3.1 New York Neighborhoods

A Dataframe of Borough Manhattan is created with it's neighborhoods containing latitude and longitude. The Manhattan Neighborhoods are separately viewed using Folium for better understanding the Manhattan Neighborhoods (See Fig 3.2).
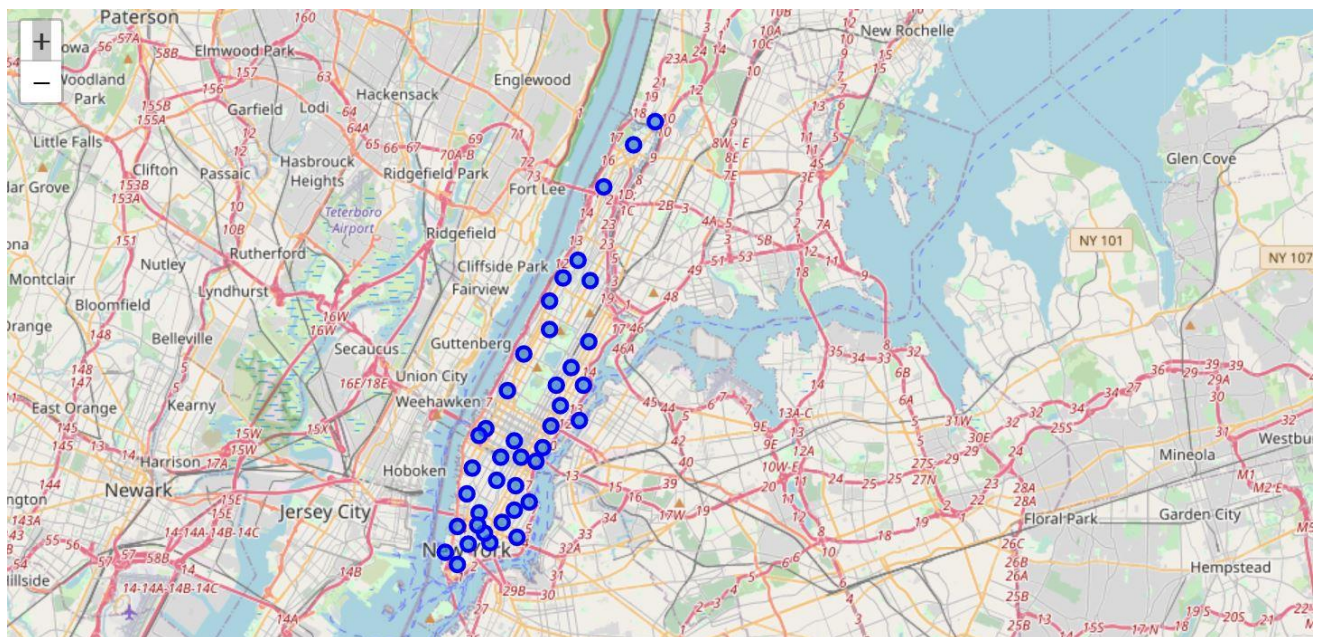

Figure 3.2 Manhattan Neighborhoods

The areas highlighted in Blue in the Map are the Neighborhoods of Toronto that is plotted using Folium from the Toronto Borough Neighborhoods Data (See Fig 3.3).
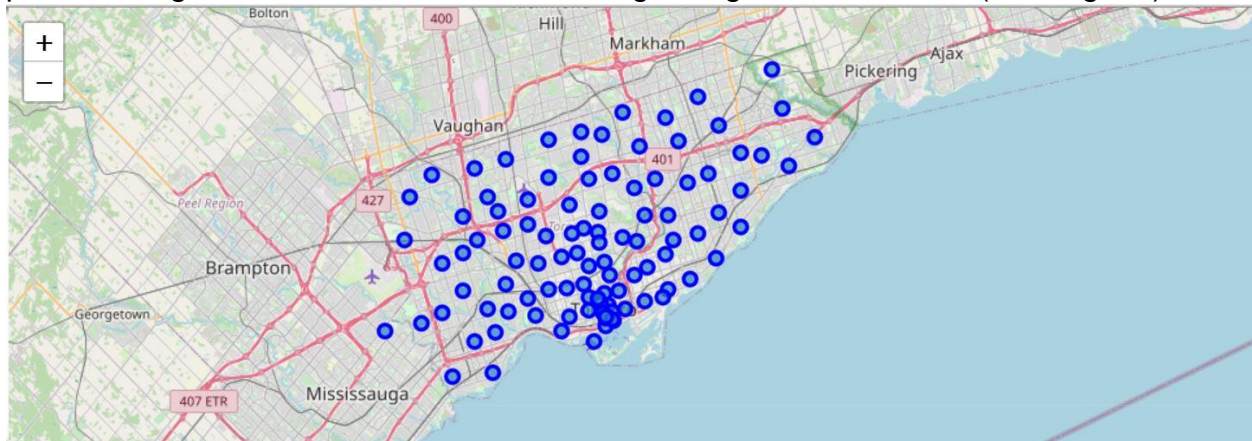

Figure 3.3 Toronto Neighborhoods

A Dataframe of Borough Downtown Toronto is created with it's neighborhoods containing latitude and longitude. The Downtown Toronto Neighborhoods are separately viewed using Folium for better understanding the Downtown Toronto Neighborhoods (See Fig 3.4).

Figure 3.4 Downtown Toronto Neighborhoods

The response of the Foursquare API venues information for the Neighborhoods of Downtown Toronto result after converting into Dataframe, The Venues are Categorized into Venue Categories using data obtained from the Foursquare API.The number of Venues returned in each neighborhood is counted for Downtown Toronto (See Fig. 3.5) and Manhattan (See Fig. 3.6)

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Battery Park City	55	55	55	55	55	55
Carnegie Hill	87	87	87	87	87	87
Central Harlem	45	45	45	45	45	45
Chelsea	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Civic Center	88	88	88	88	88	88
Clinton	100	100	100	100	100	100
East Harlem	43	43	43	43	43	43
East Village	100	100	100	100	100	100
Financial District	100	100	100	100	100	100
Flatiron	91	91	91	91	91	91
Gramercy	66	66	66	66	66	66
Greenwich Village	100	100	100	100	100	100

Figure 3.5 Manhattan Count of Venues in each Neighborhood

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	55	55	55	55	55	55
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	16	16	16	16	16	16
Central Bay Street	60	60	60	60	60	60
Christie	16	16	16	16	16	16
Church and Wellesley	77	77	77	77	77	77
Commerce Court, Victoria Hotel	100	100	100	100	100	100
First Canadian Place, Underground city	100	100	100	100	100	100
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
Kensington Market, Chinatown, Grange Park	55	55	55	55	55	55
Queen's Park, Ontario Provincial Government	39	39	39	39	39	39
Regent Park, Harbourfront	47	47	47	47	47	47

Figure 3.6 Downtown Toronto Count of Venues in each Neighborhood

Modelling:

The Final Dataframe of Manhattan containing neighborhoods with latitude and longitude information can be used to find the venues surrounding each neighborhoods within 500 meters radius by requesting the data from FourSquare API. The JSON File returned from the Foursquare API is converted into pandas Dataframe.

Preprocessing:

This Dataframe contains the venues information along with venue categories to corresponding neighborhoods. One hot encoding should be performed. The Venue Category that is present for a neighborhood is assigned 1 and else is assigned 0. Resultant Dataframe of Manhattan Neighborhoods venue data (See Fig 3.7). Resultant Dataframe of Downtown Toronto Neighborhoods venue data (See Fig 3.8).

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	0

Figure 3.7 One Hot Encoding of Manhattan Venue Data

Neighborhood	New American Restaurant	Nightclub	Noodle House	Office	Opera House	Optical Shop	Organic Grocery	Other Great Outdoors	Park	Performing Arts Venue	Pet Store
Rosedale	0	0	0	0	0	0	0	0	0	0	0
Rosedale	0	0	0	0	0	0	0	0	1	0	0
Rosedale	0	0	0	0	0	0	0	0	1	0	0
Rosedale	0	0	0	0	0	0	0	0	0	0	0
St. James Town, Cabbagetown	0	0	0	0	0	0	0	0	0	0	0

Figure 3.8 One Hot Encoding of Downtown Toronto Venue Data

The Venue data after one hot encoding is grouped by Neighborhoods and the mean of the frequency of occurrence of venues in each venue category for the Neighborhoods is calculated. The Frequency of occurrence of each venue category for the Manhattan (See Fig. 3.9). The Frequency of occurrence of each venue category for the Downtown Toronto (See Fig. 3.10)

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant
0	Battery Park City	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Carnegie Hill	0.000000	0.000000	0.00	0.000000	0.011494	0.000000	0.000000	0.000000	0.011494
2	Central Harlem	0.000000	0.000000	0.00	0.044444	0.044444	0.000000	0.000000	0.000000	0.000000
3	Chelsea	0.000000	0.000000	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000
4	Chinatown	0.000000	0.000000	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000
5	Civic Center	0.000000	0.000000	0.00	0.000000	0.022727	0.011364	0.000000	0.000000	0.000000
6	Clinton	0.000000	0.000000	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000
7	East Harlem	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	East Village	0.000000	0.000000	0.00	0.000000	0.020000	0.000000	0.000000	0.010000	0.010000

Figure 3.9 Frequency of occurrence of each venue category Manhattan

Neighborhood	Yoga Studio	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum
Berczy Park	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.018182	0.000000
CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.0625	0.0625	0.0625	0.125	0.1875	0.125	0.000000	0.000000	0.00	0.000000	0.000000
Central Bay Street	0.016667	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.000000	0.016667
Christie	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.000000	0.000000
Church and Wellesley	0.025974	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.012987	0.000000	0.00	0.000000	0.000000
Commerce Court, Victoria Hotel	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.040000	0.000000	0.00	0.010000	0.000000

Figure 3.10 Frequency of occurrence of each venue category Downtown Toronto

Finally, a Dataframe of 10 common venues is created for the neighborhoods in Manhattan (See Fig. 3.11) and 10 common venues for the neighborhoods in Downtown Toronto (See Fig. 3.12)

	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	Battery Park City	Park	Hotel	Gym	Memorial Site	Wine Shop	Coffee Shop	Shopping Mall	Food Court	Burger Joint	Gourmet Shop
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Pizza Place	Bar	Bookstore	Grocery Store	Gym	Gym / Fitness Center	Japanese Restaurant
2	Central Harlem	Chinese Restaurant	Art Gallery	Bar	Fried Chicken Joint	French Restaurant	Seafood Restaurant	Gym / Fitness Center	African Restaurant	American Restaurant	Cycle Studio
3	Chelsea	Art Gallery	Coffee Shop	Italian Restaurant	Ice Cream Shop	American Restaurant	Seafood Restaurant	Boutique	Market	Juice Bar	Bakery
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Coffee Shop	Spa	American Restaurant	Salon / Barbershop	Optical Shop	Bar	Dim Sum Restaurant

Figure 3.11 Top 10 Common Venues Downtown Toronto

	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	Berczy Park	Coffee Shop	Cocktail Bar	Bakery	Beer Bar	Café	Seafood Restaurant	Cheese Shop	Restaurant	Shopping Mall	Cosmetics Shop
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Harbor / Marina	Boat or Ferry	Boutique	Coffee Shop	Sculpture Garden	Bar	Airport Gate
2	Central Bay Street	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Salad Place	Ice Cream Shop	Bubble Tea Shop	Burger Joint	Bar	Thai Restaurant
3	Christie	Grocery Store	Café	Park	Italian Restaurant	Restaurant	Candy Store	Baby Store	Diner	Nightclub	Coffee Shop
4	Church and Wellesley	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Restaurant	Yoga Studio	Pub	Men's Store	Mediterranean Restaurant	Hotel	Smoke Shop

Figure 3.12 Top 10 Common Venues Downtown Toronto

The Similarity between the Manhattan and Downtown Toronto Neighborhoods will be found using clustering the Neighborhoods that are similar using K-means clustering which is an unsupervised machine learning algorithm. K-means clustering uses a predefined cluster size for clustering. The Cluster Size of 5 shall be used to cluster the Neighborhoods of Manhattan and Downtown Toronto into 5 clusters. Based on the clusters of neighborhoods obtained from the similar venues the neighborhood from Manhattan can be shortlisted based on the amenities, venues, tourist spots etc. which is similar to the neighborhood current city Downtown Toronto.

4. Results

The resultant 5 clusters obtained from K-means clustering. Examining each cluster of Manhattan Neighborhood and Downtown Toronto Neighborhood.

Cluster 1 Manhattan (See Fig. 4.1) Downtown Toronto (See fig. 4.2)

	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
5	Manhattanville	Coffee Shop	Seafood Restaurant	Italian Restaurant	Mexican Restaurant	Park	Deli / Bodega	Sushi Restaurant	Lounge	Boutique	Supermarket
8	Upper East Side	Italian Restaurant	Bakery	Exhibit	Gym / Fitness Center	Spa	Hotel	Juice Bar	Pizza Place	Yoga Studio	Wine Shop
9	Yorkville	Coffee Shop	Italian Restaurant	Gym	Bar	Deli / Bodega	Sushi Restaurant	Mexican Restaurant	Japanese Restaurant	Diner	Wine Shop
10	Lenox Hill	Italian Restaurant	Coffee Shop	Pizza Place	Café	Sushi Restaurant	Cocktail Bar	Burger Joint	Gym / Fitness Center	Gym	Mexican Restaurant
12	Upper West Side	Italian Restaurant	Wine Bar	Coffee Shop	Bakery	Bar	Dessert Shop	American Restaurant	Seafood Restaurant	Ice Cream Shop	Bagel Shop
13	Lincoln Square	Plaza	Gym / Fitness Center	Café	Italian Restaurant	Concert Hall	Performing Arts Venue	Theater	American Restaurant	Wine Shop	Indie Movie Theater
18	Greenwich Village	Italian Restaurant	Café	Gym	Coffee Shop	Sushi Restaurant	Wine Bar	Bakery	Comedy Club	Ice Cream Shop	Jazz Club
19	East Village	Pizza Place	Coffee Shop	Bar	Japanese Restaurant	Juice Bar	Mexican Restaurant	Dessert Shop	Cocktail Bar	Gym / Fitness Center	Grocery Store
21	Tribeca	Italian Restaurant	Park	Wine Bar	Café	Spa	American Restaurant	Art Gallery	Steakhouse	Skate Park	Scenic Lookout

Figure 4.1 Cluster 1 Manhattan

	Borough	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
14	Downtown Toronto	0	Airport Service	Airport Lounge	Airport Terminal	Harbor / Marina	Boat or Ferry	Boutique	Coffee Shop	Sculpture Garden	Bar	Airport Gate

Figure 4.2 Cluster 1 Downtown Toronto

Cluster 2 Manhattan (See Fig 4.3) Downtown Toronto (See Fig 4.4)

	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
37	Stuyvesant Town	Park	Gas Station	Baseball Field	Gym / Fitness Center	German Restaurant	Harbor / Marina	Cocktail Bar	Coffee Shop	Heliport	Farmers Market

Figure 4.3 Cluster 2 Manhattan

	Borough	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
1	Downtown Toronto	1	Coffee Shop	Restaurant	Bakery	Italian Restaurant	Café	Pub	Pizza Place	Grocery Store	Playground	Indian Restaurant
2	Downtown Toronto	1	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Restaurant	Yoga Studio	Pub	Men's Store	Mediterranean Restaurant	Hotel	Smoke Shop
4	Downtown Toronto	1	Clothing Store	Coffee Shop	Café	Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Japanese Restaurant	Cosmetics Shop	Hotel	Ramen Restaurant
5	Downtown Toronto	1	Coffee Shop	Café	American Restaurant	Gastropub	Cocktail Bar	Lingerie Store	Cosmetics Shop	Clothing Store	Seafood Restaurant	Italian Restaurant
6	Downtown Toronto	1	Coffee Shop	Cocktail Bar	Bakery	Beer Bar	Café	Seafood Restaurant	Cheese Shop	Restaurant	Shopping Mall	Cosmetics Shop

Figure 4.4 Cluster 2 Downtown Toronto

Cluster 3 Manhattan (See Fig 4.5) Downtown Toronto (See Fig 4.6)

	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
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Coffee Shop	Spa	American Restaurant	Salon / Barbershop	Optical Shop	Bar	Dim Sum Restaurant
2	Washington Heights	Café	Bakery	Mobile Phone Shop	Spanish Restaurant	Pizza Place	Grocery Store	Chinese Restaurant	Latin American Restaurant	Tapas Restaurant	New American Restaurant
3	Inwood	Mexican Restaurant	Bakery	Café	Lounge	Restaurant	Pizza Place	Park	Chinese Restaurant	American Restaurant	Frozen Yogurt Shop
4	Hamilton Heights	Pizza Place	Deli / Bodega	Café	Coffee Shop	Mexican Restaurant	Yoga Studio	Cocktail Bar	Indian Restaurant	Sushi Restaurant	Sandwich Place
6	Central Harlem	Chinese Restaurant	Art Gallery	Bar	Fried Chicken Joint	French Restaurant	Seafood Restaurant	Gym / Fitness Center	African Restaurant	American Restaurant	Cycle Studio
7	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Pizza Place	Steakhouse	Deli / Bodega	Latin American Restaurant	Taco Place	Cocktail Bar	Beer Bar
20	Lower East Side	Chinese Restaurant	Park	Art Gallery	Pharmacy	Café	Cocktail Bar	Coffee Shop	Japanese Restaurant	Juice Bar	Latin American Restaurant
22	Little Italy	Chinese Restaurant	Italian Restaurant	Spa	Bubble Tea Shop	Hotel	Mediterranean Restaurant	Café	Pizza Place	Thai Restaurant	Bakery
25	Manhattan Valley	Yoga Studio	Pizza Place	Bar	Thai Restaurant	Chinese Restaurant	Coffee Shop	Health Food Store	Peruvian Restaurant	Park	Noodle House

Figure 4.5 Cluster 3 Manhattan

	Borough	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	Downtown Toronto	2	Park	Trail	Playground	Women's Store	Cosmetics Shop	Distribution Center	Discount Store	Diner	Dessert Shop	Department Store

Figure 4.6 Cluster 3 Downtown Toronto

Cluster 4 Manhattan (See Fig. 4.7) Downtown Toronto (See Fig. 4.8)

	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	Marble Hill	Sandwich Place	Gym	Coffee Shop	Yoga Studio	Deli / Bodega	Supplement Shop	Steakhouse	Seafood Restaurant	Pizza Place	Department Store

Figure 4.7 Cluster 4 Manhattan

	Borough	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
17	Downtown Toronto	3	Grocery Store	Café	Park	Italian Restaurant	Restaurant	Candy Store	Baby Store	Diner	Nightclub	Coffee Shop

Figure 4.8 Cluster 4 Downtown Toronto

Cluster 5 Manhattan (See Fig. 4.9) Downtown Toronto (See Fig. 4.10)

	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
11	Roosevelt Island	Park	Deli / Bodega	Coffee Shop	Sandwich Place	Greek Restaurant	Liquor Store	Outdoors & Recreation	Dog Run	Scenic Lookout	Gym / Fitness Center
14	Clinton	Theater	Gym / Fitness Center	Coffee Shop	Wine Shop	Hotel	Gym	Italian Restaurant	Sandwich Place	Spa	Pizza Place
15	Midtown	Hotel	Clothing Store	Coffee Shop	Theater	Pizza Place	Cuban Restaurant	Spa	Tailor Shop	Pilates Studio	American Restaurant
16	Murray Hill	Sandwich Place	Japanese Restaurant	Coffee Shop	Burger Joint	Hotel	Mediterranean Restaurant	Chinese Restaurant	Pizza Place	Gym / Fitness Center	Bar
17	Chelsea	Art Gallery	Coffee Shop	Italian Restaurant	Ice Cream Shop	American Restaurant	Seafood Restaurant	Boutique	Market	Juice Bar	Bakery
26	Morningside Heights	Park	American Restaurant	Coffee Shop	Bookstore	Deli / Bodega	Burger Joint	Tennis Court	Pizza Place	Supermarket	Café
28	Battery Park City	Park	Hotel	Gym	Memorial Site	Wine Shop	Coffee Shop	Shopping Mall	Food Court	Burger Joint	Gourmet Shop
29	Financial District	Coffee Shop	Hotel	Salad Place	Gym / Fitness Center	Cocktail Bar	Japanese Restaurant	Falafel Restaurant	Park	Pizza Place	Café
32	Civic Center	Coffee Shop	Hotel	Spa	Yoga Studio	Gym / Fitness Center	French Restaurant	Park	Sushi Restaurant	Cocktail Bar	Café

Figure 4.9 Cluster 5 Manhattan

	Borough	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
3	Downtown Toronto	4	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Theater	Café	Beer Store	Farmers Market	Hotel
7	Downtown Toronto	4	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Salad Place	Ice Cream Shop	Bubble Tea Shop	Burger Joint	Bar	Thai Restaurant
18	Downtown Toronto	4	Coffee Shop	Sushi Restaurant	Diner	Yoga Studio	Mexican Restaurant	Burger Joint	Burrito Place	Café	Park	College Auditorium

Figure 4.10 Cluster 5 Downtown Toronto

Visualizing Clustered Neighborhood using folium for Manhattan (See Fig. 4.11) and for Downtown Toronto (See Fig 4.12)
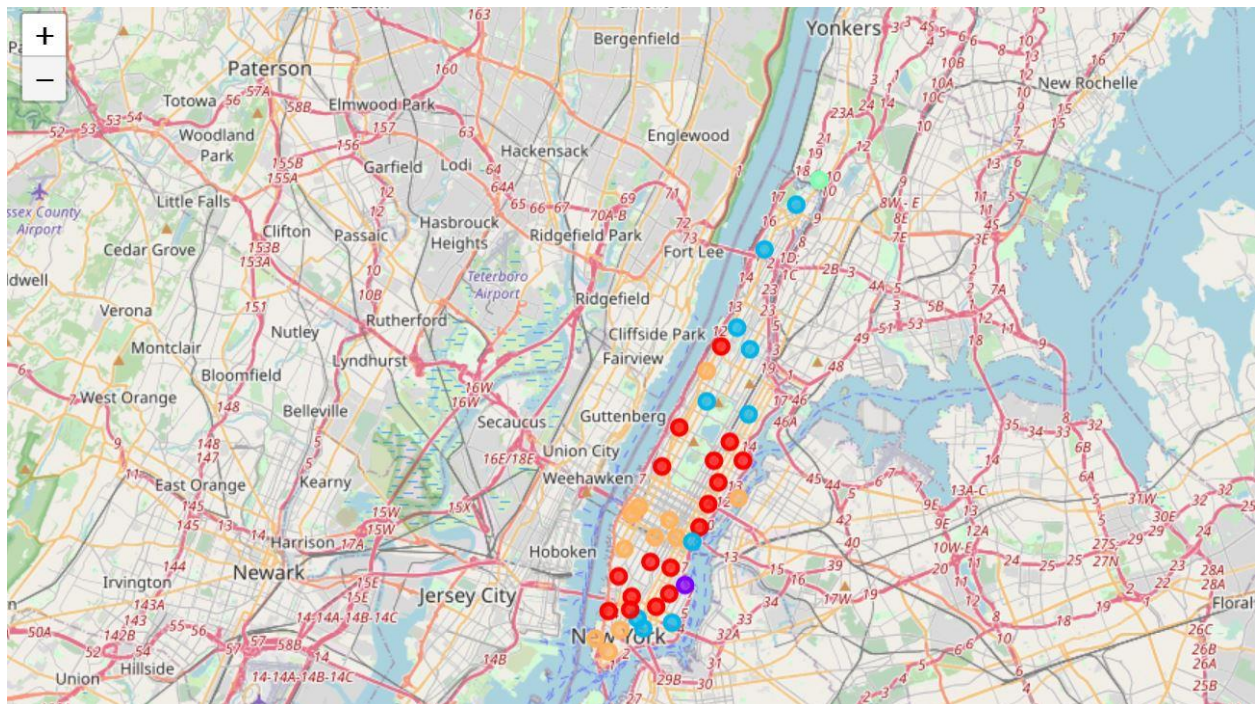

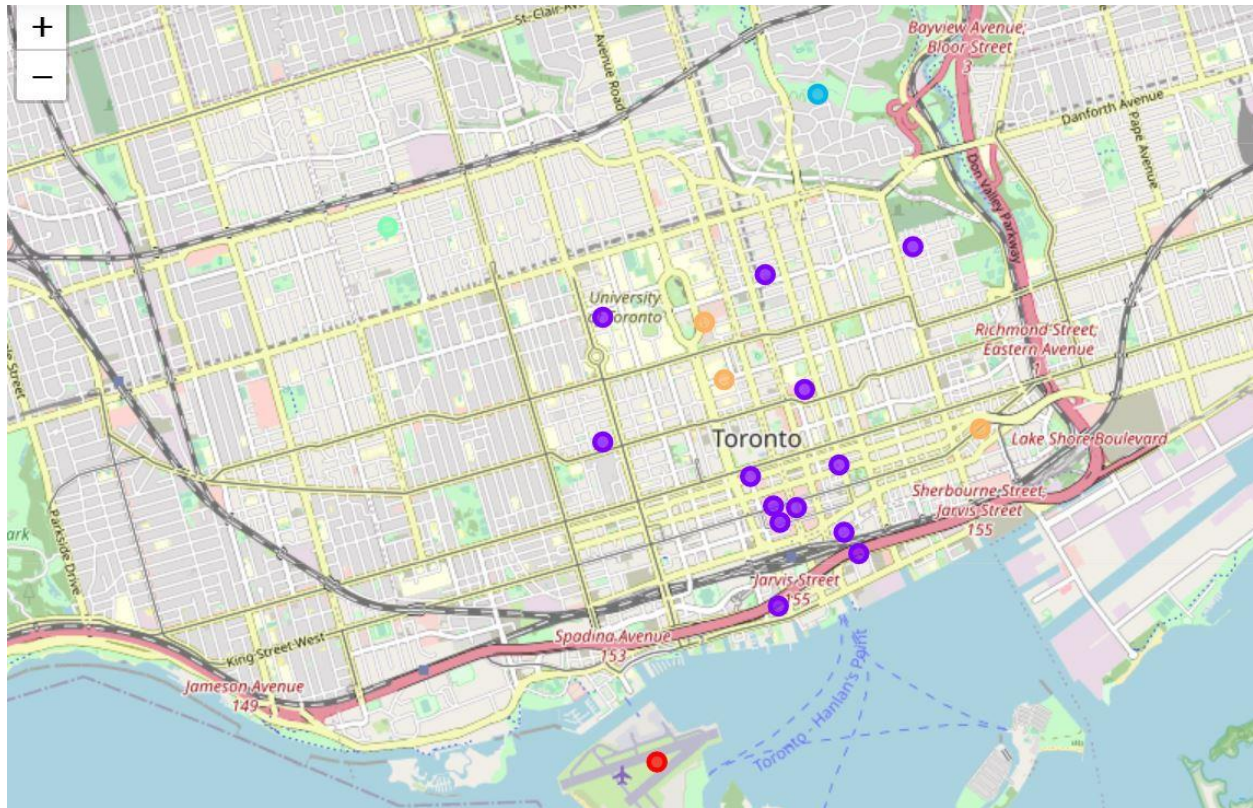*Figure 4.11 Manhattan Clustered Neighborhoods*

Figure 4.12 Downtown Toronto Clustered Neighborhoods

The Cluster 1 in Manhattan (See Fig. 4.1) most common venues are Restaurant, Pizza place, Coffee Shop whereas the Cluster 1 in Downtown Toronto (See Fig. 4.2) most common venue is Airport.

The Cluster 2 in Manhattan (See Fig. 4.3) most common venue is Park and cluster 2 in Downtown Toronto (See Fig. 4.4) most common venue is Coffee Shop.

The Cluster 3 in Manhattan (See Fig. 4.5) most common venue is Restaurant, Pizza place and cluster 3 (See Fig. 4.6) in Downtown Toronto most common venue is park.

The Cluster 4 in Manhattan (See Fig. 4.7) most common venue is Sandwich place and cluster 3 in Downtown Toronto (See Fig. 4.8) is grocery store.

The Cluster 5 in Manhattan (See Fig. 4.9) most common venue is Park, Coffee shop, Cluster 5 in Downtown Toronto (See Fig. 4.10) is Coffee Shop.

Each Cluster is Color coded and is plotted using Folium for Manhattan (See Fig. 4.11) and Downtown Toronto (See Fig. 4.12).

5. Discussion

Based on the most common venues the Restaurants, Parks, Coffee Shop are similar in both Manhattan and Downtown Toronto. But based on the Clusters there is an another observation that Manhattan is more diverse in terms of availability of restaurants when compared to Downtown Toronto, There are Chinese restaurants, Italian Restaurants, Mexican Restaurants in Manhattan. This depicts that there are people of different cultures living in Manhattan. So, the people migrating from Downtown Toronto to Manhattan, New York will find it more similar as far the venues, Parks for children and Coffee shops are concerned and provides a great opportunity to explore the people from diverse cultures living in Manhattan. If the family is more interested in amenities like theatre, Park, Coffee Shop then Cluster 5 of Manhattan is more suitable but if they are interested to explore the diverse culture in Manhattan then Cluster 3 and Cluster 1 is more Suitable.

6. Conclusion

This project helps a person migrating to new city and is unaware of the neighborhoods. But due to the love with his current neighborhood he wants to find a neighborhood similar to his current neighborhood. Based on the venue data we could find a similar neighborhood to the current neighborhood and in addition to that suggest more neighborhood options to the person who really wants to explore a diverse culture existing in the new place. We can further expand this idea such that based on the user input we can immediately give recommendations possible suitable neighborhood based on his current neighborhood information and also based on his favorite type of restaurant and food, favorite weekend venue etc. to recommend more suitable neighborhoods during their migration to a new city.