Applied Data Science Capstone project

The Battle of Neighborhoods

Introduction & Business problem

Background

The City of New York, is the most populous city in the United States. New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centers. Its iconic sites include skyscrapers such as the Empire State Building and sprawling Central Park.

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's northwestern shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower.

New York & Toronto being the major cities of USA & CANADA, we will have lot of people migrating between these two places.

Problem description

As part of this project, we would like to compare neighborhoods in New York & Toronto cities and help people in choosing similar neighborhoods when they migrate from New York to Toronto or Vice Versa. We will utilize the Foursquare API to explore the neighborhoods and segment them. We will use k-means to cluster the neighborhoods into 5 clusters.

We would be restricting the comparison to Manhattan Borough which is at the core of the New York City and the Boroughs which have Toronto in its name in Toronto to restrict number of api calls that we make using Foursquare api.

We will be performing the below steps as part of this project.

- Download and Explore Datasets for New York & Toronto cities
- Explore Neighborhoods in New York & Toronto cities
- Cluster Neighborhoods
- Examine Clusters

Note: It's straightforward to remove the condition and compare all the Boroughs in New York and Toronto.

Success Criteria

The success criteria of this project is to help end users identify similar neighborhoods in New York and Toronto cities and make informed choices when they migrate from New York to Toronto or Vice versa.

Data acquisition & cleansing

As part of this project, we would be comparing neighborhoods in New York & Toronto cities. We would need the Boroughs, latitude & longitude coordinates of each neighborhood in these two cities.

We will use Four Square api to explore each neighborhood and will be getting top 100 venues within 500m radius

New York city data

New York has a total of 5 boroughs and 306 neighborhoods. We need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset is available in the below link

https://cocl.us/new_york_dataset

We will be using the below command to get the dataset

!wget -q -O 'newyork_data.json' https://cocl.us/new_york_dataset

We will transform the data that we get from the above link into a pandas data frame and then use Foursquare api to get venue details for each neighborhood

Toronto city data

We will scrape the below Wikipedia page using BeautifulSoup api in order to obtain Borough, Neighborhood and postal codes in Toronto and then transform the data into a pandas data frame

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,

We will use the csv file present in the below link to get latitude & longitude of each neighborhood of Toronto using Postal Code as the Key

http://cocl.us/Geospatial_data

Methodology

Exploratory data analysis & Inferential statistics

Download New York data

New York data is present in the link https://cocl.us/new_york_dataset. We use the below command to download this data.

!wget -q -O 'newyork_data.json' https://cocl.us/new_york_dataset

We parse the JSON file that we get using the above command and create a pandas data frame with column names 'Borough', 'Neighborhood', 'Latitude', 'Longitude' and store the New York data in the data frame.

New York data is stored in the data frame as below.

	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

New York data contains 5 boroughs and 306 neighborhoods

Download Toronto data

We have Postal Code, Borough & Neighborhood information of Toronto city in the link http://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. We scrape this page using BeautifulSoup library and create a data frame with column names 'PostalCode', 'Borough', 'Neighborhood' and store the Toronto data in this data frame

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Scarborough,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Scarborough

We need to get the latitude and longitude for each of the postal code in the above data frame. Latitude and Longitude details for each postal code of Toronto is present in the link http://cocl.us/Geospatial_data , we use the below command to get this data

lat_log = pd.read_csv("http://cocl.us/Geospatial_data")

Parse the csv file and create a data frame like below.

	PostalCode	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

Merge the Toronto data frame and the data frame with Latitude and Longitude to get a data frame like below.

	PostalCode	Borough	Neighborhood	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	Scarborough,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Scarborough	43.773136	-79.239476

Toronto data contains 11 boroughs and 103 neighborhoods.

Merge New York & Toronto data

As we will be comparing neighborhoods of New York with Toronto, we will go ahead and merge these 2 data frames.

As we can make only limited number of api calls to Foursquare api in day, we will restrict our comparison of neighborhoods to Manhattan Borough in New York and all the Borough which contains Toronto in its name in Toronto city. We will filter the New York & Toronto data with the above conditions and create a new data frame merged with this data.

The new data frame has 5 Boroughs and 78 neighborhoods. We will do further analysis on this data.

Note: We can easily extend this analysis to all the Boroughs of New York & Toronto just by removing the conditions before merging the data frames.

Four Square api to fetch Venues for each neighborhood

Define Four Square credentials and version. We use the below url to fetch venues for each neighborhood

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

Create a function to fetch venues for the neighborhoods in Manhattan in New York and neighborhoods name starting with Toronto in CANADA. Limit the number of venues returned by Four square api to 100 within a radius of 500m

Once we get the venues for each neighborhood, we get the below data frame

	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

Group the above data frame based on neighborhood count and create a data frame like below

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Battery Park City	100	100	100	100	100	100
Berczy Park	56	56	56	56	56	56
CN Tower,Downtown Toronto,Downtown Toronto,Downtown Toronto,King and Spadina,Railway Lands,South Niagara	16	16	16	16	16	16
Cabbagetown,St. James Town	46	46	46	46	46	46
Carnegie Hill	100	100	100	100	100	100
Central Harlem	44	44	44	44	44	44
Central Toronto	67	67	67	67	67	67
Chelsea	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Chinatown,Grange Park,Kensington Market	100	100	100	100	100	100
Church and Wellesley	82	82	82	82	82	82
Civic Center	100	100	100	100	100	100
Clinton	100	100	100	100	100	100
Commerce Court,Downtown Toronto	100	100	100	100	100	100
Deer Park,Central Toronto,Rathnelly,South Hill,Central Toronto	15	15	15	15	15	15
Design Exchange,Toronto Dominion Centre	100	100	100	100	100	100
Dovercourt Village,West Toronto	14	14	14	14	14	14
Downtown Toronto	194	194	194	194	194	194
Downtown Toronto,Downtown Toronto	100	100	100	100	100	100

Mode: Command Ln 1, Col 1 Capstone_Project.i

We have 385 unique venue categories in the above data frame.

Analyze each neighborhood

We will analyze each neighborhood in the data frame, as Venue category is a categorical variable, we use one hot encoding to transform Venue category data as below.

	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Add neighborhood column back to the data frame and group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. We have a new data frame as below.

	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Arepa Restaurant
0	Battery Park City	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.010000	0.000000	0.00	0.00	0.000000
1	Berczy Park	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.00	0.000000
2	CN Tower,Downtown Toronto,Downtown Toronto,Dow...	0.000000	0.000000	0.00	0.000000	0.000000	0.0625	0.0625	0.0625	0.125	0.1875	0.125	0.000000	0.000000	0.00	0.00	0.000000
3	Cabbagetown,St. James Town	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.021739	0.000000	0.00	0.00	0.000000
4	Carnegie Hill	0.030000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.010000	0.000000	0.00	0.00	0.000000
5	Central Harlem	0.000000	0.000000	0.00	0.000000	0.068182	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.045455	0.000000	0.00	0.00	0.000000
6	Central Toronto	0.014925	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.00	0.000000
7	Chelsea	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.030000	0.010000	0.00	0.00	0.000000
8	Chinatown	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.040000	0.000000	0.00	0.00	0.000000
9	Chinatown,Grange Park,Kensington Market	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	0.00	0.00	0.000000
10	Church and Wellesley	0.012195	0.000000	0.00	0.012195	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.012195	0.000000	0.00	0.00	0.000000
11	Civic Center	0.030000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.030000	0.010000	0.00	0.00	0.000000
12	Clinton	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.040000	0.000000	0.00	0.00	0.000000
13	Commerce Court,Downtown Toronto	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.040000	0.000000	0.00	0.00	0.000000
14	Deer Park,Central Toronto,Rathnelly,South Hill...	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.066667	0.000000	0.00	0.00	0.000000
15	Design Exchange,Toronto	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.030000	0.000000	0.00	0.00	0.000000

We can then create a data frame with top 10 most common venues for each neighborhood as below.

	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	Coffee Shop	Hotel	Memorial Site	Wine Shop	Gym	Italian Restaurant	Clothing Store	Women's Store	Pizza Place
1	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Bakery	Beer Bar	Steakhouse	Cheese Shop	Farmers Market	Café	Museum
2	CN Tower,Downtown Toronto,Downtown Toronto,Dow...	Airport Service	Airport Lounge	Airport Terminal	Boutique	Harbor / Marina	Coffee Shop	Bar	Boat or Ferry	Sculpture Garden	Airport Gate
3	Cabbagetown,St. James Town	Restaurant	Coffee Shop	Flower Shop	Pub	Italian Restaurant	Café	Bakery	Pizza Place	Bank	Gastropub
4	Carnegie Hill	Coffee Shop	Pizza Place	Cosmetics Shop	Café	Yoga Studio	Wine Shop	Gym	Grocery Store	Japanese Restaurant	Bakery

Modeling

We will use K-Means to cluster the neighborhood into 5 clusters

We will add cluster labels to the data frame and will have a new data frame as below.

	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	0	Sandwich Place	Discount Store	Coffee Shop	Yoga Studio	Deli / Bodega	Department Store	Diner	Spa	Bank	Clothing Store
1	Manhattan	Chinatown	40.715618	-73.994279	0	Chinese Restaurant	Cocktail Bar	Salon / Barbershop	Vietnamese Restaurant	American Restaurant	Spa	Dumpling Restaurant	Bubble Tea Shop	Ice Cream Shop	Bakery
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Café	Bakery	Mobile Phone Shop	Grocery Store	Spanish Restaurant	Tapas Restaurant	Coffee Shop	Park	New American Restaurant	Deli / Bodega
3	Manhattan	Inwood	40.867684	-73.921210	0	Mexican Restaurant	Café	Lounge	Pizza Place	Pharmacy	Chinese Restaurant	Frozen Yogurt Shop	Spanish Restaurant	Bakery	American Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Pizza Place	Café	Mexican Restaurant	Coffee Shop	Deli / Bodega	Bakery	School	Sandwich Place	Caribbean Restaurant	Chinese Restaurant

Now will go ahead and examine the 5 clusters that we have created using K-Means

Cluster 1

We have 4 neighborhoods and all of them are from Manhattan itself.

	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
15	Midtown	Hotel	Clothing Store	Coffee Shop	Cocktail Bar	Theater	Bookstore	Japanese Restaurant	Spa	Steakhouse	Bakery
22	Little Italy	Bakery	Café	Italian Restaurant	Sandwich Place	Bubble Tea Shop	Clothing Store	Salon / Barbershop	Mediterranean Restaurant	Women's Store	Optical Shop
23	Soho	Clothing Store	Boutique	Women's Store	Art Gallery	Shoe Store	Sporting Goods Shop	Men's Store	Mediterranean Restaurant	Italian Restaurant	Furniture / Home Store
33	Midtown South	Korean Restaurant	Hotel	Cosmetics Shop	Dessert Shop	Japanese Restaurant	Hotel Bar	Coffee Shop	American Restaurant	Yoga Studio	Gym / Fitness Center

Cluster 2

We have only 1 neighborhood from 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
8	Moore Park,Central Toronto	Summer Camp	Playground	Trail	Women's Store	Ethiopian Restaurant	Drugstore	Dry Cleaner	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant

Cluster 3

We have 2 neighborhoods and one of them is from Manhattan and the other one from Toronto. This cluster is useful for our study. We can say Rosedale in Manhattan is like Forest Hill North, Central Toronto in 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
10	Rosedale	Park	Playground	Trail	Building	Women's Store	Ethiopian Restaurant	Drugstore	Dry Cleaner	Dumpling Restaurant	Duty-free Shop
23	Forest Hill North,Central Toronto	Park	Trail	Sushi Restaurant	Jewelry Store	Women's Store	English Restaurant	Drugstore	Dry Cleaner	Dumpling Restaurant	Duty-free Shop

Cluster 4

We have only 1 neighborhood from Manhattan and not much useful for our study

	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
4	Lawrence Park	Park	Swim School	Photography Studio	Bus Line	Women's Store	Ethiopian Restaurant	Dry Cleaner	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant

Cluster 5

We have 31 neighborhoods in this cluster and has a mix of neighborhoods from New York and Toronto. If anyone from one of the below neighborhoods is planning to move from Toronto to New York or Vice Versa, he can look at the below study and see which neighborhoods are similar and make informed choices.

	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
20	Lower East Side	Coffee Shop	Café	Pizza Place	Sandwich Place	Bakery	Japanese Restaurant	Cocktail Bar	Chinese Restaurant	Ramen Restaurant	Art Gallery
26	Morningside Heights	Park	Coffee Shop	Bookstore	American Restaurant	Food Truck	Deli / Bodega	Sandwich Place	Tennis Court	Burger Joint	Café
37	Stuyvesant Town	Bar	Playground	Park	Heliport	Baseball Field	German Restaurant	Basketball Court	Gas Station	Harbor / Marina	Cocktail Bar
1	East Toronto,Riverdale	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Furniture / Home Store	Yoga Studio	Pizza Place	Sports Bar	Brewery	Spa
3	East Toronto	Café	Coffee Shop	Brewery	Bakery	Gym / Fitness Center	American Restaurant	Italian Restaurant	Gastropub	Park	Smoke Shop
5	Central Toronto	Pizza Place	Dessert Shop	Coffee Shop	Sandwich Place	Park	Clothing Store	Gym / Fitness Center	Gym	Sushi Restaurant	Diner
6	Central Toronto	Pizza Place	Dessert Shop	Coffee Shop	Sandwich Place	Park	Clothing Store	Gym / Fitness Center	Gym	Sushi Restaurant	Diner
7	Central Toronto	Pizza Place	Dessert Shop	Coffee Shop	Sandwich Place	Park	Clothing Store	Gym / Fitness Center	Gym	Sushi Restaurant	Diner
9	Deer Park,Central Toronto,Rathnelly,South Hill...	Pub	Coffee Shop	Fried Chicken Joint	Light Rail Station	Liquor Store	Sushi Restaurant	Supermarket	Bagel Shop	American Restaurant	Restaurant
11	Cabbagetown,St. James Town	Restaurant	Coffee Shop	Flower Shop	Pub	Italian Restaurant	Café	Bakery	Pizza Place	Bank	Gastropub
12	Church and Wellesley	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Café	Mediterranean Restaurant	Fast Food Restaurant	Gym	Gastropub
13	Harbourfront,Regent Park	Coffee Shop	Bakery	Café	Park	Breakfast Spot	Mexican Restaurant	Pub	Farmers Market	Brewery	Bank

14	Downtown Toronto,Downtown Toronto	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Middle Eastern Restaurant	Restaurant	Ramen Restaurant	Italian Restaurant	Fast Food Restaurant	Lingerie Store
15	St. James Town	Coffee Shop	Italian Restaurant	Café	Hotel	Restaurant	Gastropub	Cosmetics Shop	Breakfast Spot	Cocktail Bar	Clothing Store
16	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Bakery	Beer Bar	Steakhouse	Cheese Shop	Farmers Market	Café	Museum
17	Downtown Toronto	Coffee Shop	Café	Italian Restaurant	Restaurant	Park	Seafood Restaurant	Bakery	Grocery Store	Sandwich Place	Hotel
18	Downtown Toronto,Downtown Toronto,Downtown Tor...	Coffee Shop	Café	Steakhouse	Bar	Thai Restaurant	American Restaurant	Asian Restaurant	Restaurant	Gym	Breakfast Spot
19	Downtown Toronto,Toronto Islands,Union Station	Coffee Shop	Aquarium	Hotel	Café	Italian Restaurant	Pizza Place	Brewery	Sporting Goods Shop	Bakery	Scenic Lookout
20	Design Exchange,Toronto Dominion Centre	Coffee Shop	Café	Hotel	Restaurant	Italian Restaurant	Gastropub	American Restaurant	Gym	Bakery	Bar
21	Commerce Court,Downtown Toronto	Coffee Shop	Café	Hotel	Restaurant	American Restaurant	Seafood Restaurant	Steakhouse	Deli / Bodega	Gastropub	Gym
22	Central Toronto	Pizza Place	Dessert Shop	Coffee Shop	Sandwich Place	Park	Clothing Store	Gym / Fitness Center	Gym	Sushi Restaurant	Diner
24	The Annex,Central Toronto,Yorkville	Coffee Shop	Café	Sandwich Place	Pizza Place	Pub	Park	Vegetarian / Vegan Restaurant	American Restaurant	Burger Joint	Pharmacy
25	Downtown Toronto,University of Toronto	Café	Bookstore	Bakery	Restaurant	Japanese Restaurant	Bar	Sandwich Place	Sushi Restaurant	Beer Bar	Nightclub
28	Downtown Toronto	Coffee Shop	Café	Italian Restaurant	Restaurant	Park	Seafood Restaurant	Bakery	Grocery Store	Sandwich Place	Hotel
29	First Canadian Place,Underground city	Coffee Shop	Café	Restaurant	Steakhouse	Hotel	Bar	American Restaurant	Gastropub	Asian Restaurant	Deli / Bodega
30	Downtown Toronto	Coffee Shop	Café	Italian Restaurant	Restaurant	Park	Seafood Restaurant	Bakery	Grocery Store	Sandwich Place	Hotel
32	Little Portugal,Trinity	Bar	Coffee Shop	Asian Restaurant	Café	New American Restaurant	Bakery	Pizza Place	French Restaurant	Vietnamese Restaurant	Men's Store
33	West Toronto,Exhibition Place,Parkdale Village	Café	Coffee Shop	Breakfast Spot	Music Venue	Bar	Furniture / Home Store	Stadium	Falafel Restaurant	Burrito Place	Yoga Studio
35	Parkdale,Roncesvalles	Breakfast Spot	Gift Shop	Restaurant	Movie Theater	Eastern European Restaurant	Bar	Bank	Dog Run	Dessert Shop	Coffee Shop
36	Runnymede,Swansea	Pizza Place	Coffee Shop	Café	Sushi Restaurant	Italian Restaurant	Fish Market	Gastropub	Bar	Supplement Shop	French Restaurant
37	East Toronto	Café	Coffee Shop	Brewery	Bakery	Gym / Fitness Center	American Restaurant	Italian Restaurant	Gastropub	Park	Smoke Shop

Results & Discussion

We have compared neighborhoods in Manhattan Borough in New York with neighborhoods which contains Toronto in its name in Toronto and have created 5 clusters as below.

Cluster 1	4 neighborhoods from Manhattan
Cluster 2	1 neighborhood from Toronto
Cluster 3	1 neighborhood from Toronto & 1 neighborhood from Toronto
Cluster 4	1 neighborhood from Manhattan
Cluster 5	31 neighborhoods from Manhattan & Toronto

We can see that we have close to 31 neighborhoods in Cluster 5 (17 from Toronto & 14 from Manhattan) so people migrating from Toronto to New York or Vice versa in the above neighborhoods can look for similar neighborhoods in Manhattan or Toronto depending on from where they are migrating.

We have limited this study only to a few neighborhoods in New York and Toronto because of the restrictions on the number of api calls that we can make to Four square api but we can easily extend this

project to compare all the neighborhoods in New York and Toronto and create clusters. We can use this clusters data to find similar neighborhoods in New York and Toronto.

We can also extend this project to compare different cities across the World and see how similar or dissimilar they are and help people while they are migrating. We only did the comparison on the most common venues but we can extend these criteria to include lot of other factors like Cost of Living, Safety etc. and make comparisons between different neighborhoods in different cities.