The Battle of Neighborhood

1. Introduction:

    Immigration is defined as the movement of people to live from their homeland or region to another country they are not native to. There are specific economic factors contributing to immigration, including the desire to get higher wage rates, to improve living standards, to have better job opportunities and to gain education. There are also significant non-economic factors and these include Leaving home for persecution, ethnic cleansing, genocide, war, natural disasters and political control (such as dictatorship). Immigration has become increasingly common worldwide throughout history, with improved transportation and technology. In both the home country and the host country, immigration numbers affect.

    a. Benefits of Immigration:

        Immigration has many advantages to it. Immigrants primarily choose to leave their homeland to improve their quality of life. Economic factors for immigration include higher pay rates, improved work conditions, a higher standard of life and incentives for employment. Immigrants also frequently leave their homelands to escape poverty religious persecution, oppression, ethnic cleansing, genocide, wars, or a political structure (e.g. repressive dictatorship). Whatever the reasoning behind immigration is, it provides the immigrant with a new beginning in life and more opportunities for growth than previously available. Success in a new country is not guaranteed and often requires hard work and sacrifices, but many immigrants are willing to take risks for themselves to be able to have a better future.

2. Problem Description:

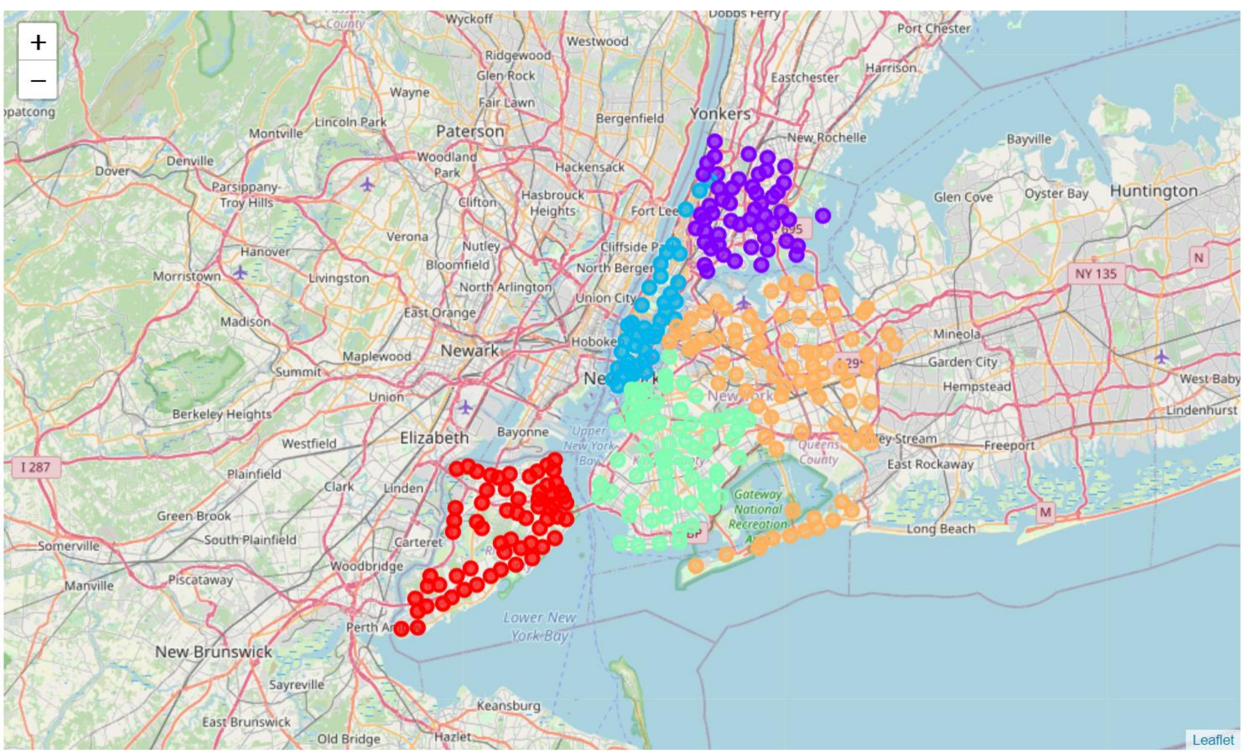
    a. The neighborhoods of New York City and the city of Toronto are compared in this project. Both cities are very diverse and are the financial capitals of their respective countries. The similarity or dissimilarity is compared.
    b. Both the cities are segmented and clustered their neighborhoods based on Borough in each city.
    c. A detailed overview of most common venues in each cluster are discussed to prove the similarity and dissimilarity

3. Data
    a. Data set for New York city is available in following link: https://geo.nyu.edu/catalog/nyu_2451_34572
    b. Data set for Toronto city is available in following link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
4. Data Acquisition:
    a. The data for New York has all the information Neighbourhood name, location det ails (Latitude and longitude)
    b. Data for New York city has 5 boroughs and 306 neighborhoods.
    c. Table has Borough details of New York city

Borough
Bronx
Manhattan
Brooklyn
Queens
Staten Island

    d.

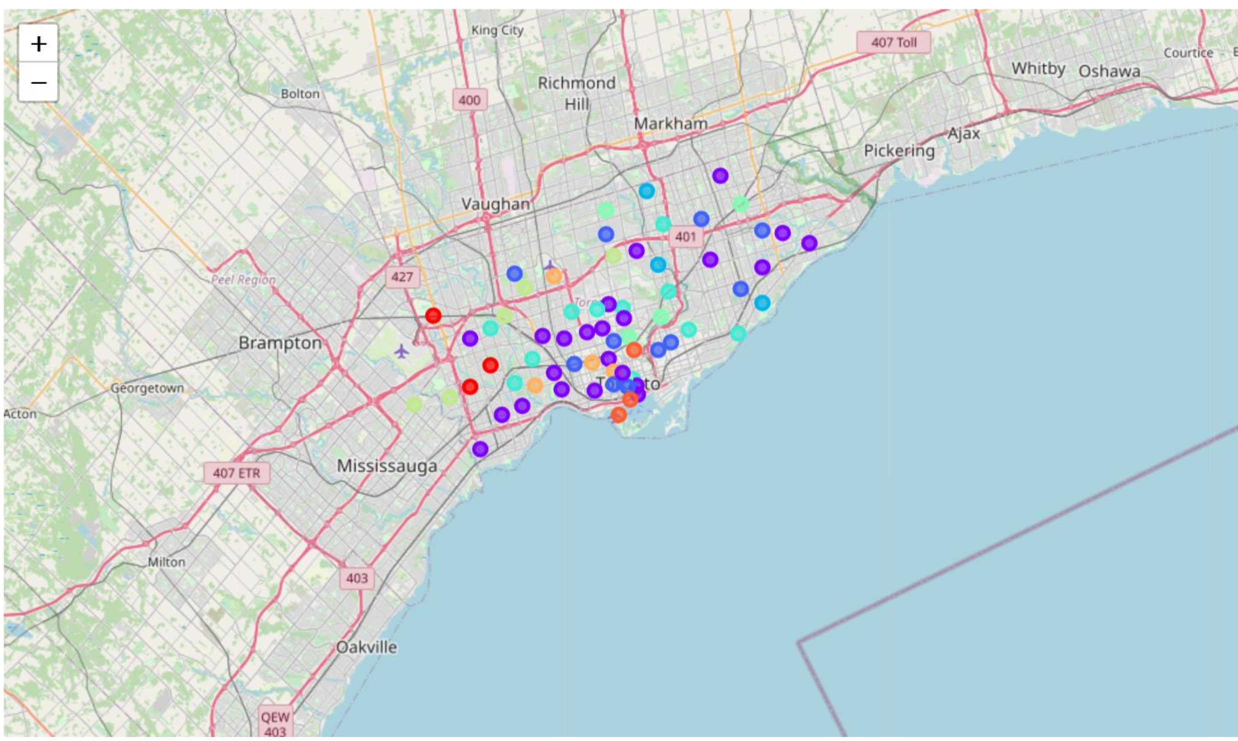
    e. New York city Map



f. The data of Toronto city has postal code, Neighbourhood name. I used Nominatim library from geopy to get the location details of Toronto City. Toronto city has 9 boroughs.

g. Map of Toronto City:

Borough
North York
Downtown Toronto'
Etobicoke
Scarborough
East York
York
East Toronto
West Toronto
Central Toronto



h.

i. Four-square API was is used to search for near by venues in each neighbourhood of Toronto city and New York city in the radius of 500 mt. The venue name and Venue category are extracted. The total number of venues in each category are counted and also the most frequent venue category for each neighbourhood is classified.

5. Data Analysis:
    a. After retrieving all the data in the form of JSON file from Foursquare API. The data of near by venues in the radius of 500 mt of both Toronto city and New York city are merged.
    b. According to the total number of venues in each category I used K-means clustering algorithm to cluster the data.
    c. Total number of Clusters specified are 6. (Some iterations are made to reduce the outliers). After the iteration total number of clusters is fixed to 6

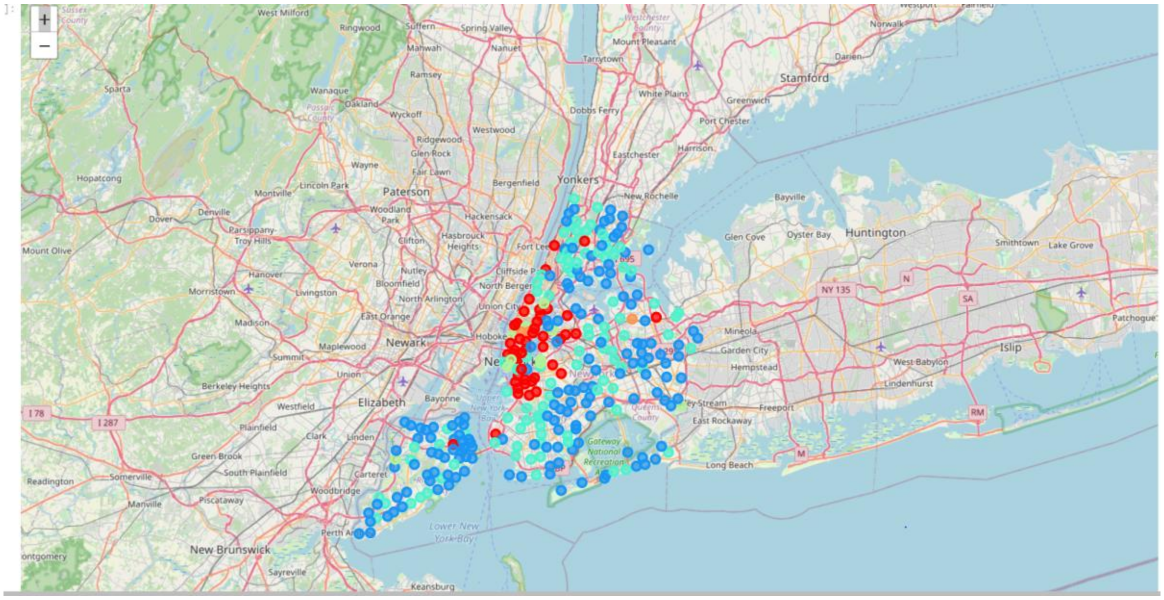
```
58]: from sklearn.cluster import KMeans
kclusters = 6
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(all_venues_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
allvenues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
```

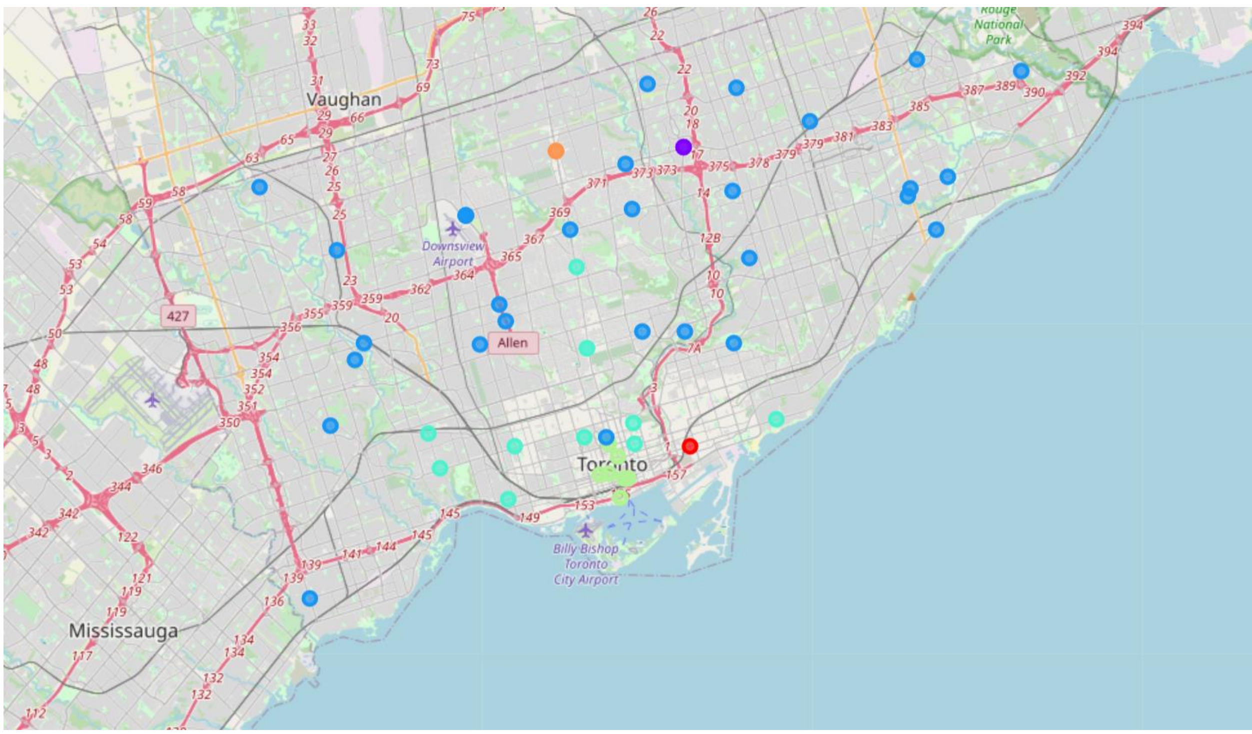
6. Results:
    a. Total Neighbourhoods in Each cluster are shown in the image:

```
Total number of neighborhoods in cluster 0 is 46
Total number of neighborhoods in cluster 1 is 2
Total number of neighborhoods in cluster 2 is 184
Total number of neighborhoods in cluster 3 is 105
Total number of neighborhoods in cluster 4 is 17
Total number of neighborhoods in cluster 5 is 5
```

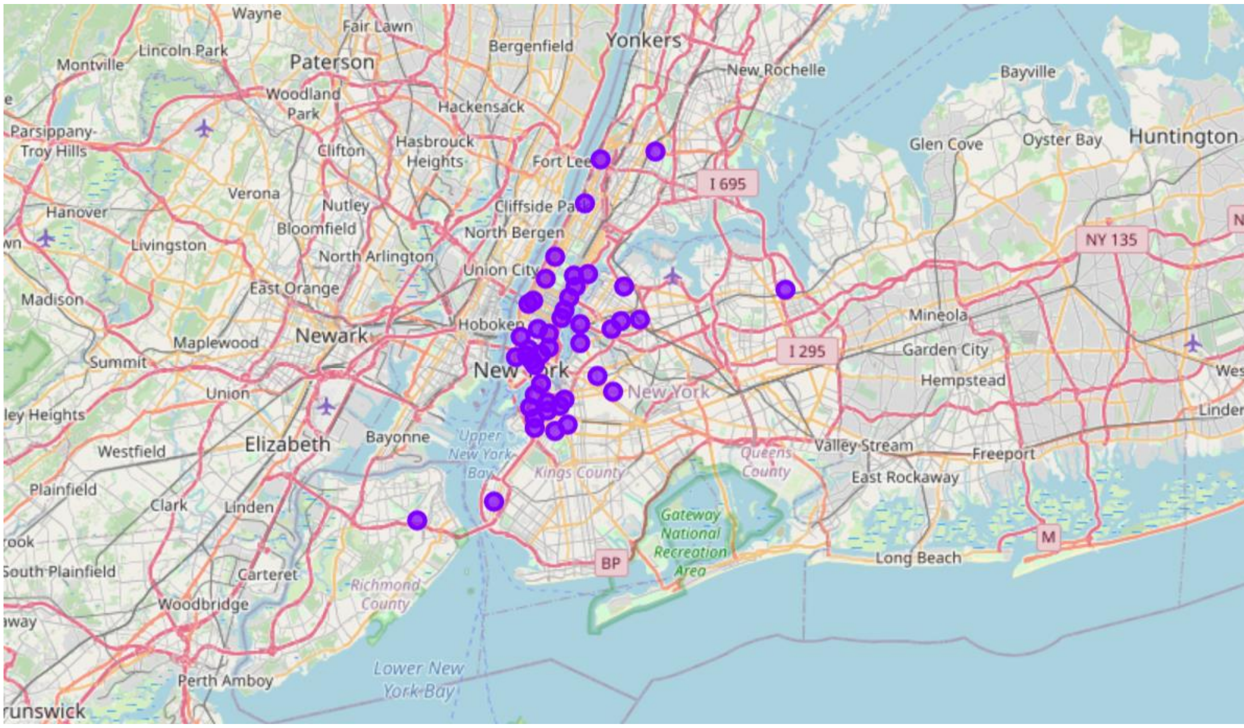
    b. Cluster map of New york city:



c. Cluster Map of Toronto City:



d. Cluster 0: This has more bars, pizza places. This has all clusters in New york city as shown in figure below:



e. Cluster1:

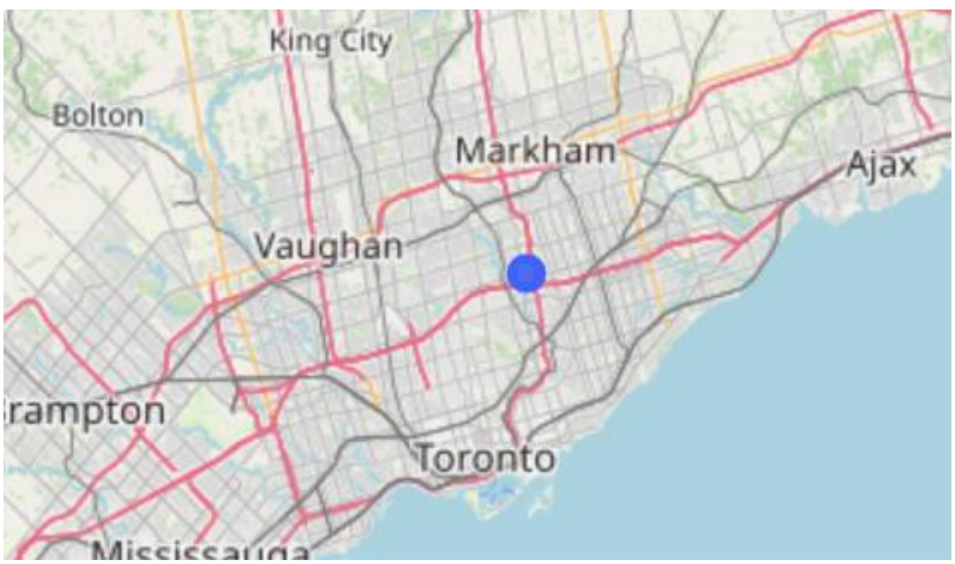
Has only 2 neighbourhoods and may be a possible outlier :

Clustr :1

```
[193]: allvenues_sorted.loc[allvenues_sorted['Cluster Labels'] == 1, allvenues_sorted.columns[[1] + list(range(5, allvenues_sorted.shape[1]))]]
```

[193]:

	City	Longitude	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
311	Toronto	-79.3459	Clothing Store	Coffee Shop	Japanese Restaurant	Fast Food Restaurant	Restaurant	Bank	Bakery	Sandwich Place	Women's Store	Convenience Store
313	Toronto	-79.3459	Clothing Store	Coffee Shop	Japanese Restaurant	Fast Food Restaurant	Restaurant	Bank	Bakery	Sandwich Place	Women's Store	Convenience Store



f.

g. Cluster :2

Is distributed across New York and Toronto with pharmacy stores, grocery stores, Bus station.

Cluster:2

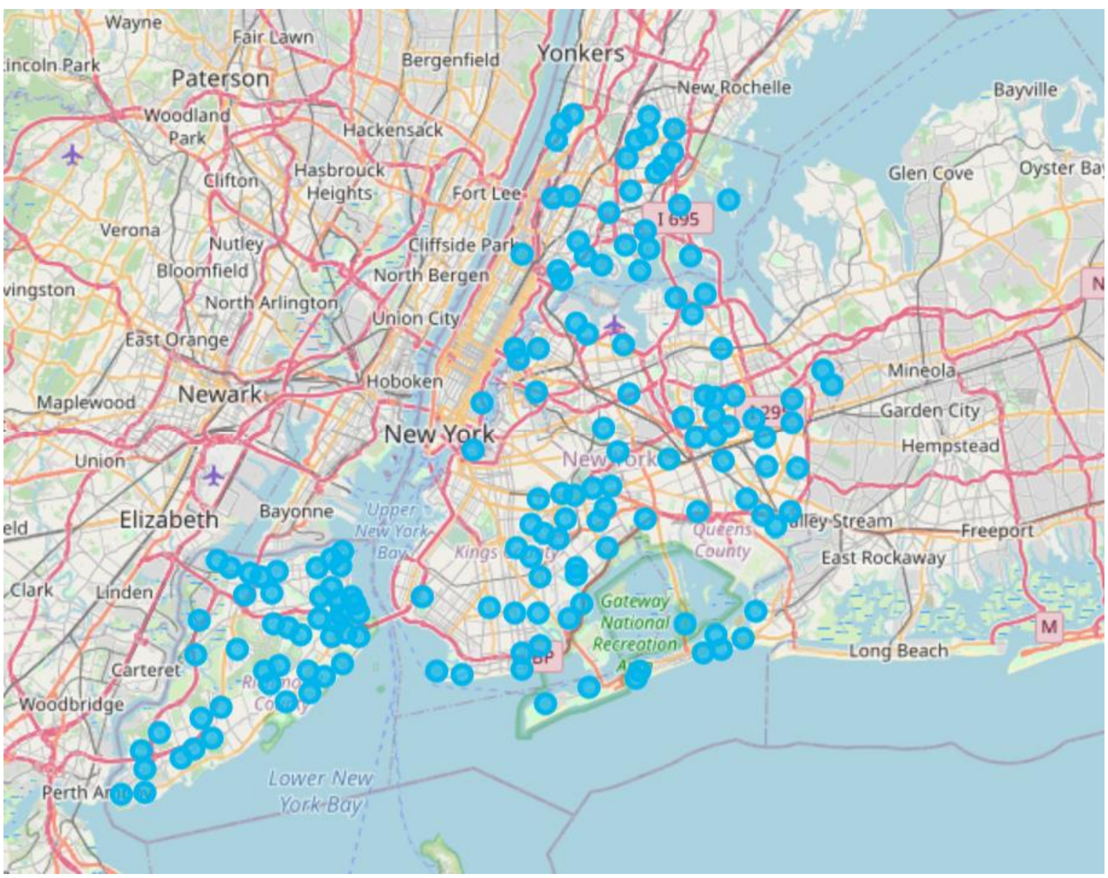
```
[95]: allvenues_sorted.loc[allvenues_sorted['Cluster Labels'] == 2, allvenues_sorted.columns[[1] + list(range(5, allvenues_sorted.shape[1]))]]
```

[95]:

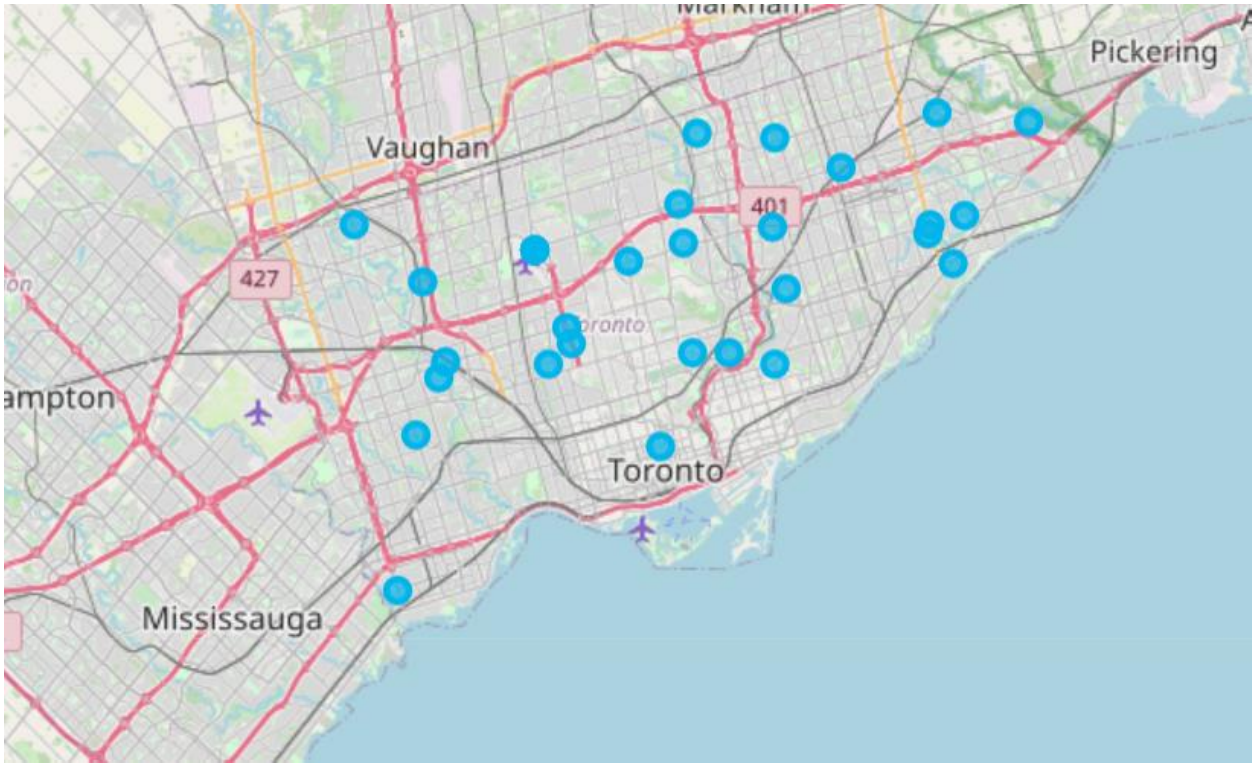
	City	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	New York	-73.8472	Pharmacy	Food	Dessert Shop	Donut Shop	Laundromat	Ice Cream Shop	Pizza Place	Sandwich Place	Gas Station	Factory
1	New York	-73.8299	Bus Station	Basketball Court	Pizza Place	Fast Food Restaurant	Grocery Store	Chinese Restaurant	Park	Baseball Field	Pharmacy	Bagel Shop
2	New York	-73.8278	Caribbean Restaurant	Deli / Bodega	Bus Station	Diner	Food & Drink Shop	Automotive Shop	Bakery	Donut Shop	Bowling Alley	Pizza Place
3	New York	-73.9056	Bus Station	River	Plaza	Yoga Studio	Fish & Chips Shop	Event Service	Event Space	Exhibit	Factory	Falafel Restaurant
4	New York	-73.9126	Bus Station	Park	Medical Supply Store	Gym	Baseball Field	Food Truck	Bank	Plaza	Farm	Fast Food Restaurant
...	...	...	...	...	...	...	...	...	...	...	...	...
349	Toronto	-79.2785	Chinese Restaurant	Bakery	Cantonese Restaurant	Hong Kong Restaurant	Train Station	Vietnamese Restaurant	Korean Restaurant	Asian Restaurant	Shopping Mall	Food Court
353	Toronto	-79.318	Grocery Store	Chinese Restaurant	Fast Food Restaurant	Pizza Place	Electronics Store	Coffee Shop	Breakfast Spot	Bank	Sandwich Place	Pharmacy
355	Toronto	-79.5452	Pizza Place	Skating Rink	Pub	Pharmacy	Coffee Shop	Gym	Sandwich Place	Yoga Studio	Ethiopian Restaurant	Event Service
356	Toronto	-79.1658	Park	Fast Food Restaurant	Yoga Studio	Cultural Center	Event Service	Event Space	Exhibit	Factory	Falafel Restaurant	Farm
358	Toronto	-79.3871	Burger Joint	Sushi Restaurant	Coffee Shop	Yoga Studio	Hobby Shop	Burrito Place	Café	Dance Studio	Bubble Tea Shop	Grocery Store

184 rows × 12 columns

New York:



Toronto:



h. Cluster:3

Has a lot of Pizza places, sand witch hubs located in both New York and Toronto.

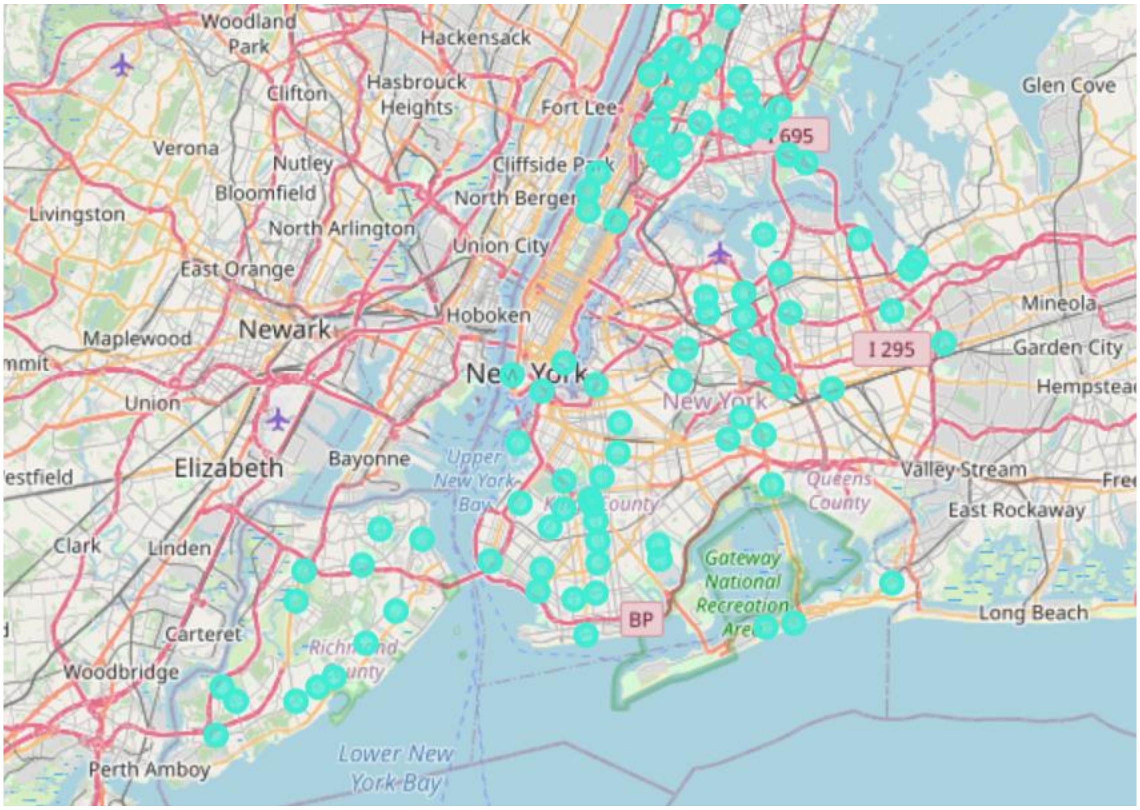
Cluster :3

```
198]: allvenues_sorted.loc[allvenues_sorted['Cluster Labels'] == 3, allvenues_sorted.columns[[1] + list(range(5, allvenues_sorted.shape[1]))]]
```

198]:

	City	Longitude	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	New York	-73.9028	Pizza Place	Bar	Latin American Restaurant	Sandwich Place	Bakery	Supermarket	Mexican Restaurant	Spanish Restaurant	Fast Food Restaurant	Donut Shop
6	New York	-73.9107	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Video Game Store	Tennis Stadium	Seafood Restaurant	Miscellaneous Shop	Bank	Pharmacy
7	New York	-73.8673	Pub	Deli / Bodega	Pizza Place	Playground	Italian Restaurant	Food & Drink Shop	Grocery Store	Park	Bar	Bakery
8	New York	-73.8794	Pizza Place	Bank	Park	Pharmacy	Food Truck	Bus Station	Pet Store	Caribbean Restaurant	Sandwich Place	Mobile Phone Shop
11	New York	-73.8548	Bus Station	Italian Restaurant	Frozen Yogurt Shop	Pizza Place	Mexican Restaurant	Bank	Sandwich Place	Coffee Shop	Gym / Fitness Center	Metro Station
...	...	...	...	...	...	...	...	...	...	...	...	...
348	Toronto	-79.4397	Tibetan Restaurant	Café	Italian Restaurant	Pharmacy	Indian Restaurant	Restaurant	Diner	Pizza Place	Bakery	Tattoo Parlor
350	Toronto	-79.3973	Italian Restaurant	Sushi Restaurant	Coffee Shop	Trail	Park	Convenience Store	Pub	Bank	Bar	Café
351	Toronto	-79.3987	Café	Coffee Shop	Restaurant	Italian Restaurant	Japanese Restaurant	Gym	Bookstore	Bakery	Yoga Studio	Music School
352	Toronto	-79.4759	Café	Coffee Shop	Bakery	Bank	Pizza Place	Bookstore	Gastropub	Liquor Store	Falafel Restaurant	Boutique
357	Toronto	-79.3721	Restaurant	Diner	Gastropub	Grocery Store	Hotel	Pub	Pet Store	Rock Club	Bakery	Bank

New York map:


Woodland Park
Hackensack
Clifton
Hasbrouck Heights
Fort Lee
Glen Cove
Verona
Nutley
Bloomfield
North Bergen
Livingston
North Arlington
Union City
East Orange
Hoboken
Mineola
Newark
Maplewood
I 295
Garden City
Union
Elizabeth
Bayonne
Upper New York Bay
Valley Stream
Queens County
East Rockaway
Clark
Linden
Gateway National Recreation
BP
Carteret
Long Beach
Woodbridge
Perth Amboy
Lower New York Bay

Toronto Map:


Vaughan
401
427
Toronto
Toronto

i. Cluster:4

Located in both New York and Toronto. This cluster has all kinds of venues like art gallery, pub, restaurant’s, Café etc.

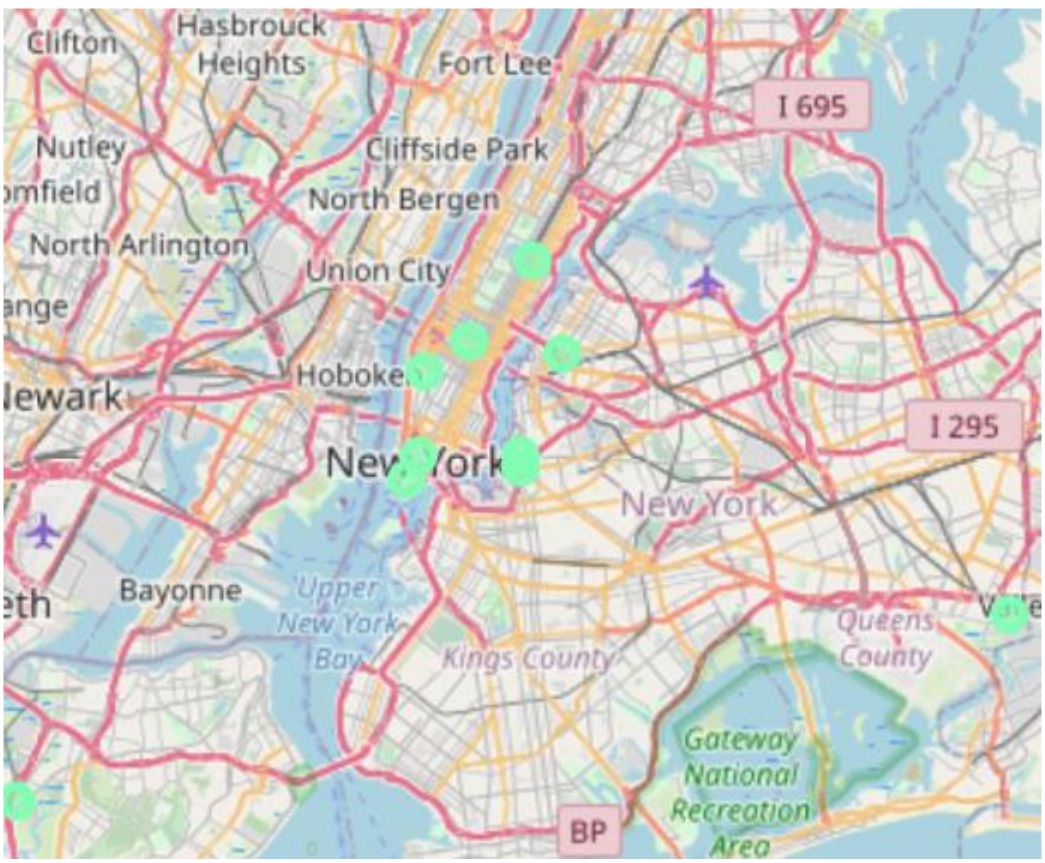
Cluster:4 ¶

201]:
```
allvenues_sorted.loc[allvenues_sorted['Cluster Labels'] == 4, allvenues_sorted.columns[[1] + list(range(5, allvenues_sorted.shape[1]))]]
```

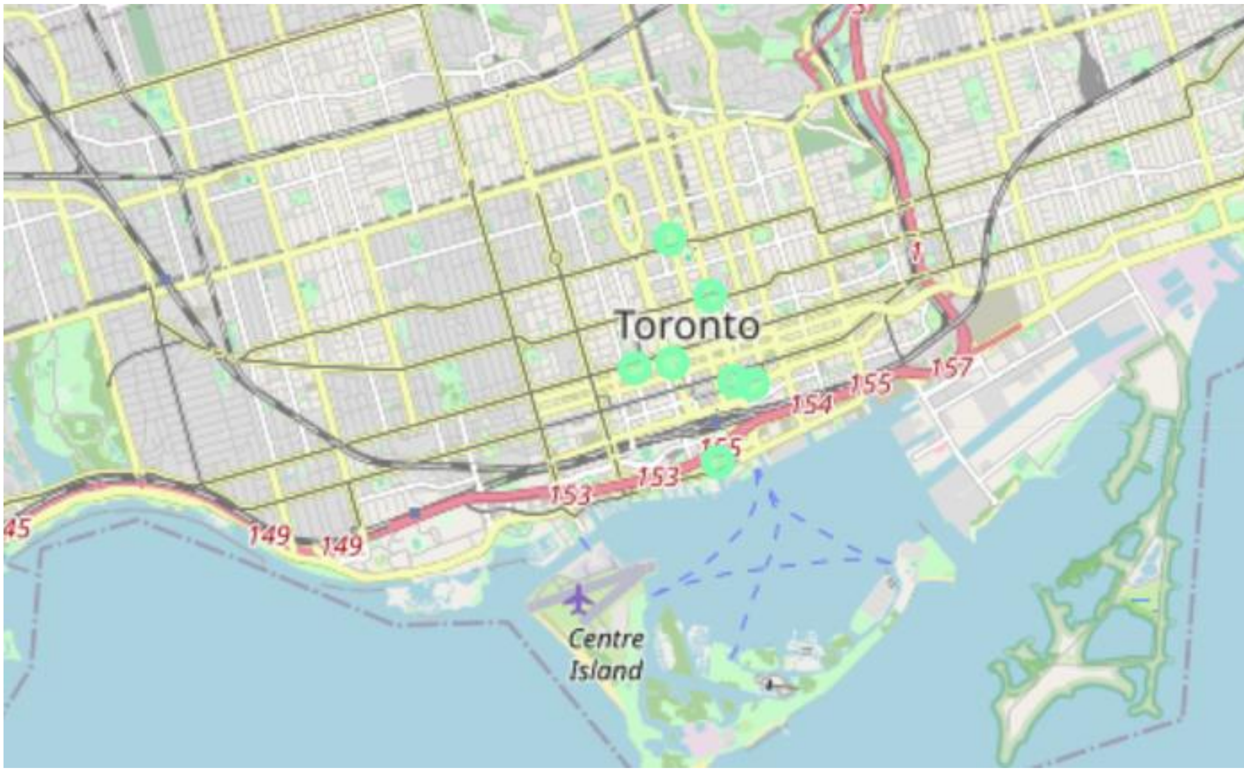
201]:

	City	Longitude	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
96	New York	-73.9588	Coffee Shop	Pizza Place	Wine Bar	Yoga Studio	American Restaurant	Bakery	Jewelry Store	Vegetarian / Vegan Restaurant	Bar	Juice Bar
97	New York	-73.958	Coffee Shop	Bar	American Restaurant	Pizza Place	Yoga Studio	Breakfast Spot	Wine Bar	Japanese Restaurant	Mexican Restaurant	Cycle Studio
114	New York	-73.9817	Hotel	Coffee Shop	Theater	Bakery	Clothing Store	Sushi Restaurant	Japanese Restaurant	Pizza Place	Cuban Restaurant	Burger Joint
116	New York	-74.0031	Coffee Shop	Art Gallery	Ice Cream Shop	Bakery	Café	American Restaurant	Bar	Italian Restaurant	Cocktail Bar	Seafood Restaurant
128	New York	-74.0107	Coffee Shop	Pizza Place	Hotel	Café	Italian Restaurant	American Restaurant	Event Space	Mexican Restaurant	Cocktail Bar	Sandwich Place
139	New York	-73.9392	Coffee Shop	Hotel	Pizza Place	Bar	Café	Mexican Restaurant	Supermarket	Deli / Bodega	Gym / Fitness Center	Office
169	New York	-73.7353	Coffee Shop	Clothing Store	Sandwich Place	Middle Eastern Restaurant	Café	Bar	Chinese Restaurant	Hotel	Ramen Restaurant	Tanning Salon
244	New York	-74.1896	Coffee Shop	Art Gallery	Ice Cream Shop	Bakery	Café	American Restaurant	Bar	Italian Restaurant	Cocktail Bar	Seafood Restaurant
247	New York	-73.9533	Coffee Shop	Café	Yoga Studio	Bookstore	Gym / Fitness Center	Italian Restaurant	Gym	Wine Shop	Pizza Place	Japanese Restaurant
249	New York	-74.0054	Coffee Shop	Hotel	Cocktail Bar	Park	American Restaurant	Spa	Gym / Fitness Center	French Restaurant	Café	Yoga Studio
307	Toronto	-79.3799	Coffee Shop	Café	Restaurant	Hotel	Italian Restaurant	History Museum	Park	Music Venue	Chinese Restaurant	Sandwich Place
318	Toronto	-79.3754	Coffee Shop	Café	Restaurant	Italian Restaurant	Japanese Restaurant	Gastropub	Hotel	Cocktail Bar	Gym	Bakery

New York Map:



Toronto Map:



j. Cluster:5

Having only 5 neighbours hood this cluster is a possible outlier:

Cluster 5

```
allvenues_sorted.loc[allvenues_sorted['Cluster Labels'] == 5, allvenues_sorted.columns[[1] + list(range(5, allvenues_sorted.shape[1]))]]
```

	City	Longitude	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
115	New York	-73.9783	Korean Restaurant	Hotel	Sandwich Place	Coffee Shop	Pizza Place	Bar	Japanese Restaurant	Gym / Fitness Center	Mediterranean Restaurant	Fried Chicken Joint
180	New York	-73.8128	Korean Restaurant	Hotel	Sandwich Place	Coffee Shop	Pizza Place	Bar	Japanese Restaurant	Gym / Fitness Center	Mediterranean Restaurant	Fried Chicken Joint
250	New York	-73.9887	Korean Restaurant	Hotel	Dessert Shop	American Restaurant	Café	Burger Joint	Coffee Shop	Japanese Restaurant	Hotel Bar	Gym / Fitness Center
338	Toronto	-79.4138	Bubble Tea Shop	Korean Restaurant	Pizza Place	Ramen Restaurant	Dessert Shop	Vietnamese Restaurant	Japanese Restaurant	Sushi Restaurant	Coffee Shop	Greek Restaurant
347	Toronto	-79.4138	Bubble Tea Shop	Korean Restaurant	Pizza Place	Ramen Restaurant	Dessert Shop	Vietnamese Restaurant	Japanese Restaurant	Sushi Restaurant	Coffee Shop	Greek Restaurant

7. Discussion:
   a. Cluster 1, cluster 5 are possible outliers.
   b. Cluster 2 is good of residential purpose as I have all amenities in neighbourhood like bus stations, pharmacy stores, grocery stores.
   c. Cluster 4 can be compared as downtown of Toronto and Manhattan area for New York with all types of venues

d. Cluster 0 and cluster3 have similar types of venues like restaurant's with international cuisine and may be common places of hang out in both cities.

8. Conclusion
   a. Both the cities being finical capital of respective countries share a common culture. Although New York is big compared to Toronto. The cities similar in terms of distribution.