

Comparing New York City and Toronto for an ideal hotel location

Applied Data Science Capstone Project

Business Problem

- ▶ To identify ideal location for a new hotel in New York City and Toronto.
- ▶ Understand similarities and dissimilarities between points of interests between both the locations using k-means clustering.
- ▶ Evaluate and suggest most preferable location to attract hotel guests.

Exploration of Dataset, Segmenting & clustering of neighborhoods in New York and Toronto

- ▶ Dataset for New York city was obtained from:
https://cocl.us/new_york_dataset
- ▶ Dataset for Toronto was obtained from:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- ▶ Foursquare location data is leveraged to show location coordinates i.e. latitudes and longitudes of interests in each neighborhood .
- ▶ Only the location of interests for the borough of Manhattan in New York city and downtown areas of Toronto city are examined.
- ▶ Top 10 venues from each neighborhood is clustered using K-means algorithm.
- ▶ Folium library is used to visualize the neighborhoods in New York City and Toronto

Example of New York city dataset & location data captured from FourSqaure

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
5	Manhattan	Manhattanville	40.816934	-73.957385
6	Manhattan	Central Harlem	40.815976	-73.943211
7	Manhattan	East Harlem	40.792249	-73.944182
8	Manhattan	Upper East Side	40.775639	-73.960508
9	Manhattan	Yorkville	40.775930	-73.947118
10	Manhattan	Lenox Hill	40.768113	-73.958860
11	Manhattan	Roosevelt Island	40.762160	-73.949168
12	Manhattan	Upper West Side	40.787658	-73.977059
13	Manhattan	Lincoln Square	40.773529	-73.985338
14	Manhattan	Clinton	40.759101	-73.996119

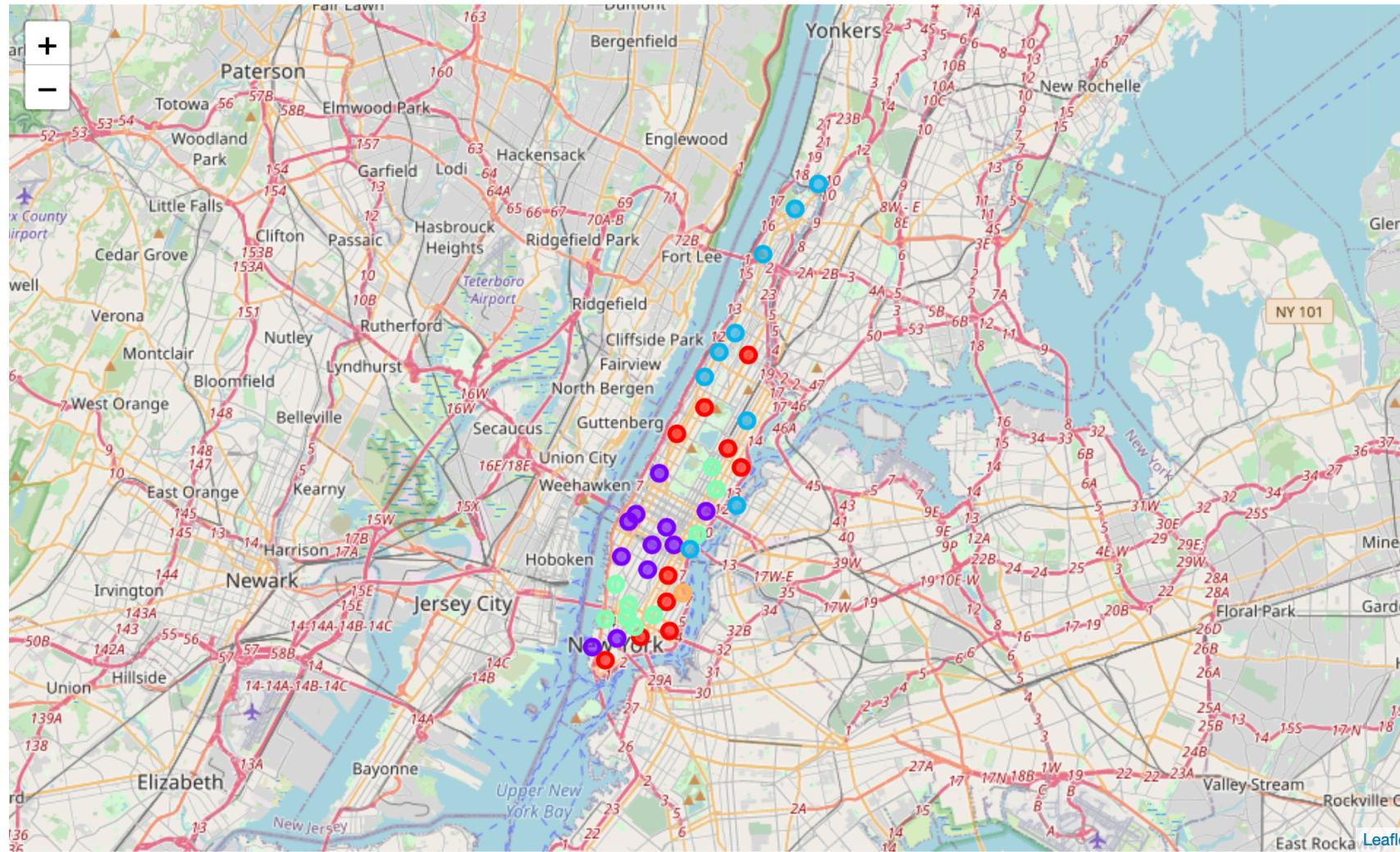
- ▶ All major districts are considered in the borough of Manhattan of New York city
- ▶ Neighborhood of ‘Marble Hill’ is identified for locating other neighboring points of interests as shown below:

	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

Example of identifying Top 10 neighborhoods in Manhattan, New York City

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	Gym	Boat or Ferry	Plaza	Gourmet Shop	Food Court	Shopping Mall
1 Carnegie Hill	Coffee Shop	Café	Italian Restaurant	Pizza Place	Yoga Studio	Gym	Gym / Fitness Center	Bookstore	Wine Shop	Vietnamese Restaurant
2 Central Harlem	Chinese Restaurant	American Restaurant	Bar	French Restaurant	Seafood Restaurant	Gym / Fitness Center	African Restaurant	Cycle Studio	Fried Chicken Joint	Library
3 Chelsea	Coffee Shop	Art Gallery	American Restaurant	Ice Cream Shop	French Restaurant	Café	Bakery	Market	Italian Restaurant	Theater
4 Chinatown	Chinese Restaurant	Bakery	Dessert Shop	Vietnamese Restaurant	Cocktail Bar	Bar	Spa	Bubble Tea Shop	Ice Cream Shop	American Restaurant

Clustering of locations of interest in Manhattan, New York City



Examining clusters in Manhattan, New York City

Cluster 1

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merg
```

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 Cluster 2

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1]] + list(range(5, manhattan_mer
```

Cluster 3

12 13 Lin

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_mer
```

Cluster 4

25 0 Mε Cluster 4

```
16 : manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_m
27 2 Was : 1st Most 2nd Most 3rd Most 4th Most 5th Most 6th Most 7th Most 8th Most 9th Most 10th Most]
```

28

Upper East Side Italian Coffee Shop Bakery Gym / Fitness Juice Bar Yoga Studio French Spa Women's Gym Sustainable

18 Greenwich Village Italian Restaurant Sushi Café Clothing Store Ice Cream Shop Chinese Restaurant French Restaurant Bar Caribbean Restaurant Coffee Shop

34	S	7	East	Village	Restaurant	Restaurant		Store	Shop	Restaurant	Restaurant		Restaurant
			21	Tribeca	Park	Italian Restaurant	American Restaurant	Greek Restaurant	Wine Bar	Spa	Café	Coffee Shop	Burger Joint
38		11	Re									Playground	

11

Restaurant Shop Restaurant Restaurant Shop Shop Restaurant

26	MOR	23	Soho	Italian Restaurant	Clothing Store	Sandwich Place	Coffee Shop	Mediterranean Restaurant	Bakery	Ice Cream Shop	Hotel	Café	French Restaurant
36	TUE	24	West Village	Italian	New American	Cocktail Bar	Park	Wine Bar	American	Bakery	Jazz Club	Coffee	Pizza Place

24

21 Italian Rice Pilaf Italian Saffron Rice Grocery French Sandwich Quiche B Mexican American Waffles

31	Nono	Pizza Place	Restaurant	Coffee Shop	Store	Restaurant	Place	Cocktail Bar	Restaurant	Restaurant	Wine Bar
35	Turtle Bay	Italian Restaurant	Coffee Shop	Park	Café	Sushi Restaurant	Deli / Bodega	Seafood Restaurant	French Restaurant	Turkish Restaurant	Garde

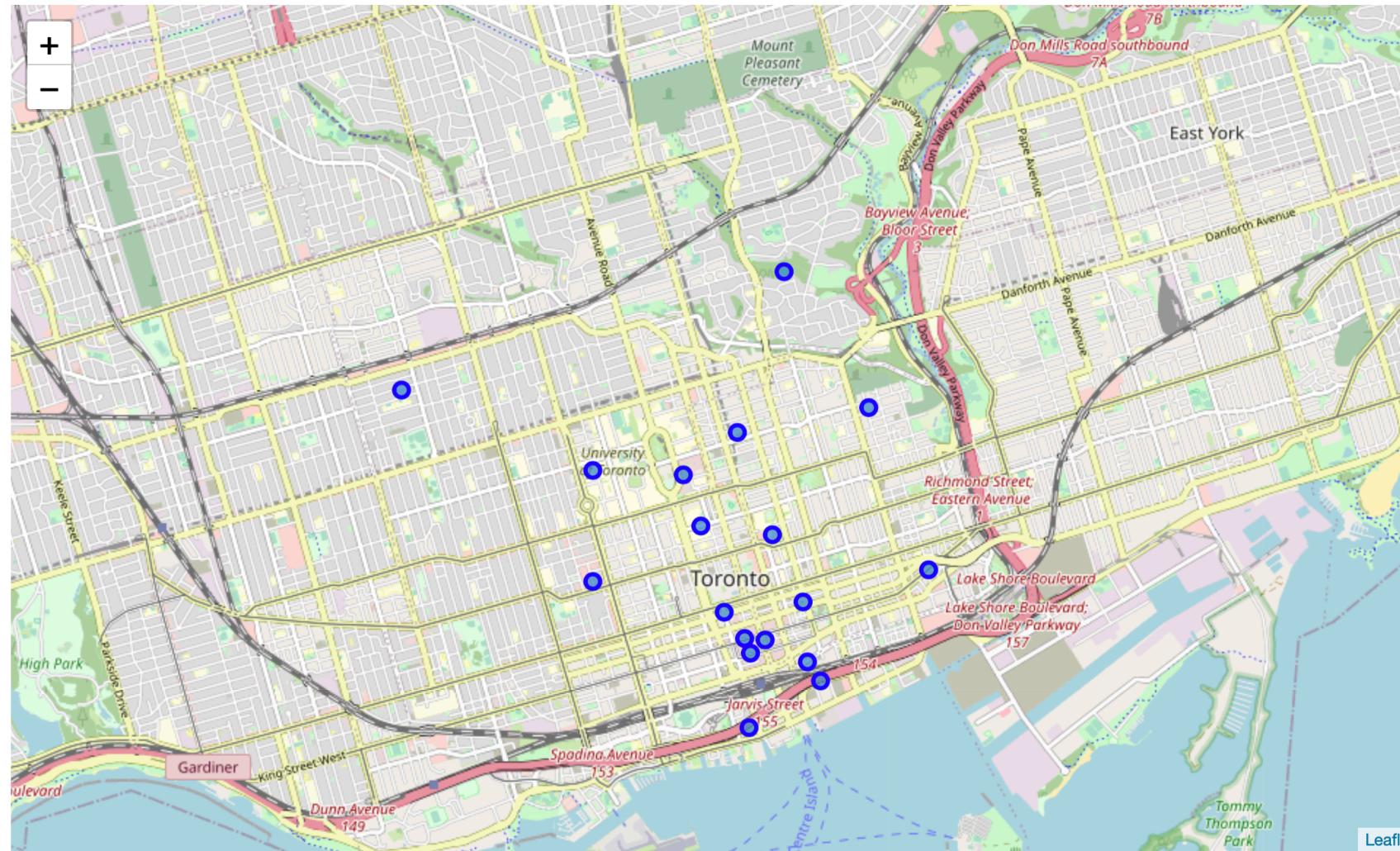
Example of Toronto dataset & location data captured from FourSqaure

- ▶ All major neighborhoods are considered for Toronto

```
Toronto_data3.head(15) #show DataFrame with the 15 columns
```

	Postal Code	Borough	Neighborhood	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
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724
12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
13	M3C	North York	Don Mills	43.725900	-79.340923
14	M4C	East York	Woodbine Heights	43.695344	-79.318389

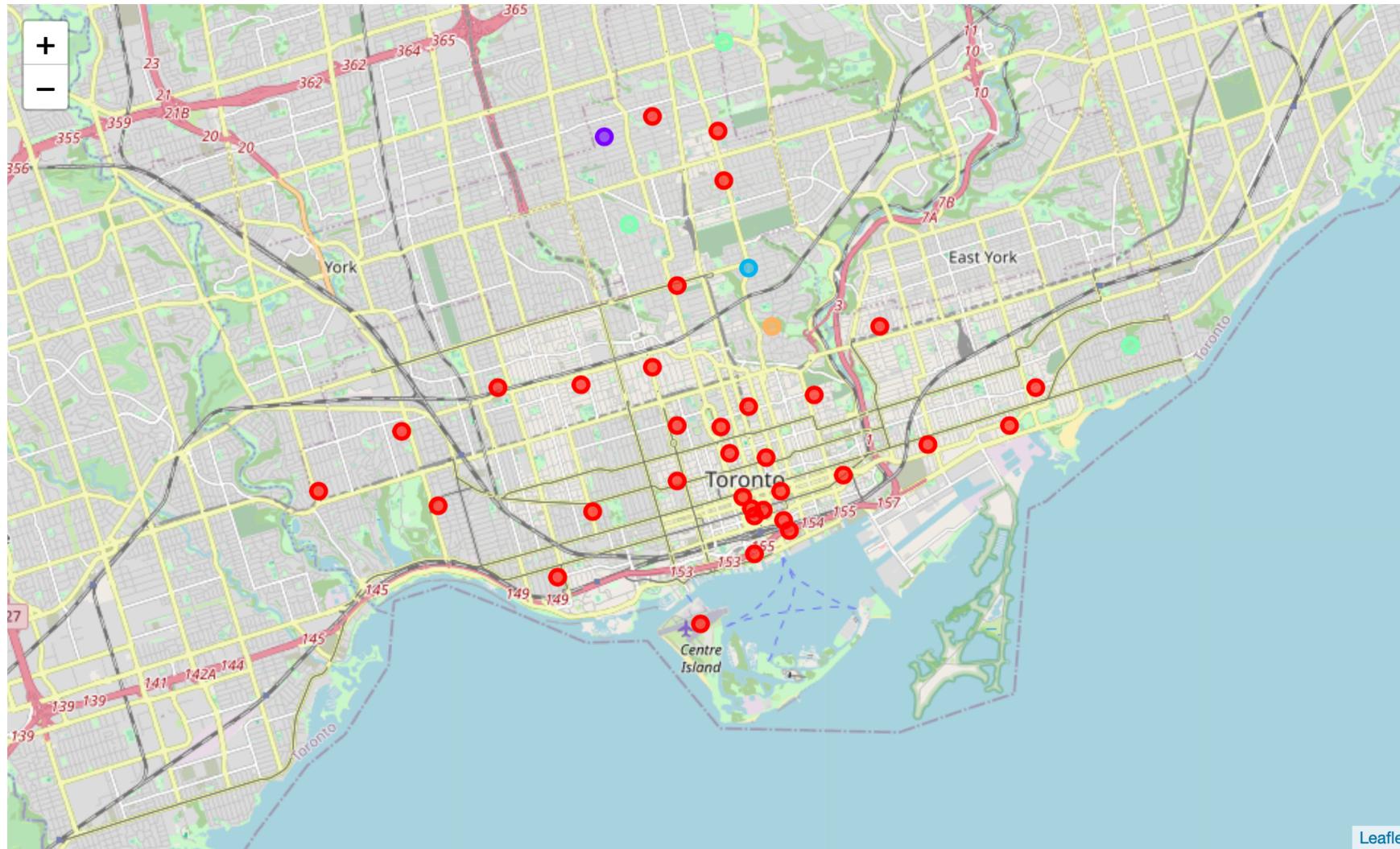
Identifying all locations of interest in Toronto



Example of identifying locations of interest in downtown Toronto and location data from FourSquare

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Dominion Pub and Kitchen	43.656919	-79.358967	Pub

Clustering neighborhoods in downtown Toronto using k-means



Examining clusters in downtown Toronto using k-means

Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]
```

	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	0	Coffee Shop	Park	Pub	Bakery	Theater	Café	Restaurant	Breakfast Spot	Event Space	Hotel						
1	Downtown Toronto	0	Coffee Shop	Sushi Restaurant	Bank	Beer Bar	Smoothie Shop	Sandwich Place	Burrito Place	Restaurant	Café	Park						
2	Downtown Toronto	0	Clothing Store	Coffee Shop	Bubble Tea Shop	Cosmetics Shop	Japanese Restaurant	Italian Restaurant	Middle Eastern Restaurant	Café	Bookstore	Bakery						
3	Downtown Toronto	0	Coffee Shop	Café	Cocktail Bar	Gastropub	American	Moroccan	Cosmetics	Creperie	Department	Clothing						
5	Downtown Toronto	0	Coffee Shop	Cocktail Bar	Restaurant	toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]												
6	Downtown Toronto	0	Coffee Shop	Sandwich Place	Italian Restaurant	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	
7	Downtown Toronto	0	Grocery Store	Café	Pub	19	Central Toronto	1	Garden	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center
8	Downtown Toronto	0	Coffee Shop	Café	Restaurant	toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]												
9	West Toronto	0	Pharmacy	Bakery	Grocery Store	toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]												
10	Downtown Toronto	0	Coffee Shop	Aquarium	Cafe	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	
11	West	0	Bar	Restaurant	Cafe	29	Central Toronto	2	Gym	Trail	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run

Cluster 4

Conclusion

- ▶ Locations of interests (interesting venues) are found to be very high for Cluster 1 of Toronto compared to Clusters 1 through 4 for New York city.
- ▶ It is preferable to house the location of new hotel in Cluster 1 of Toronto.
- ▶ However, since several clusters have good amount of locations of interest for New York City, a new hotel may be housed in either of these clusters.