

Applied Data Science Capstone

Toronto and Manhattan neighbourhood comparison

Eduardo Palomero-López

June, 2020

Contents

- Introduction
- Data description
- Methodology
- Results
- Discussion
- Conclusion

Introduction

During this Capstone course, we had the opportunity of analyse neighbourhoods of Manhattan (NYC) and Toronto.

It was used k-means with k=5 to get clusters of neighbourhoods in these two cities.

For NYC it was observed that there were a variety of neighbourhoods, with 2 clusters with higher number of nodes, but, mostly, they are quite balanced.

But for Toronto, most neighbourhoods fall into one cluster while all other barely have 1 or 2 neighbourhoods. There is one cluster with 2 neighbourhoods and 3 clusters with only one neighbourhood.

Does this mean that Toronto neighbourhoods are quite similar between them?

If we were able to run this analysis with both cities, which Manhattan area will be more similar to Toronto?

Under this assumption, would we still have most Toronto neighbourhoods in same cluster? Or they will be divided and falling into other clusters?

This exercise has been thought as theoretical work for algorithm evaluation

Data description

Data used is the same as it was used during this Capstone.

For NYC:

Neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood.

It will be needed to filter Manhattan neighbourhoods from this data.

Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572.

For Toronto:

Data will be obtained from table in Wikipedia. The table is a list of the postal codes and neighbourhoods in Toronto, including coordinates, neighbourhood and Borough.

It will be needed to filter Toronto neighbourhoods from this data and discard all other.

We will use BeautifulSoup to parse the data and put into a pandas dataframe to be used in the analysis.

Methodology

First we would need to get data from different data sources. Then, it will be needed to filter interesting data and discard all other and pre-process the data to get pandas dataframe that will be used for analysis.

These data sources come from different formats.

New York data is in a JSON file, so it is needed to parse the data from JSON and put into pandas dataframe. There is also the need of removing all neighbourhoods that are not in Manhattan.

Toronto data will be get from Wikipedia. Using BeautifulSoup library we will be able to get the data from the web, parse it and put into a pandas dataframe. There is also the need of filtering the data as it includes further than Toronto.

Using a Foursquare account connection, we get venue data for each neighbourhood. Ten top venue categories will be considered.

For the analysis, we will use K-means algorithm for neighbourhood clustering. Parameter k will be set to 5 as per recommendation in previous exercises within this Capstone course.

Both Toronto and Manhattan will be analysed separately.

Then, a third analysis will be performed in order to compare both cities.

Results

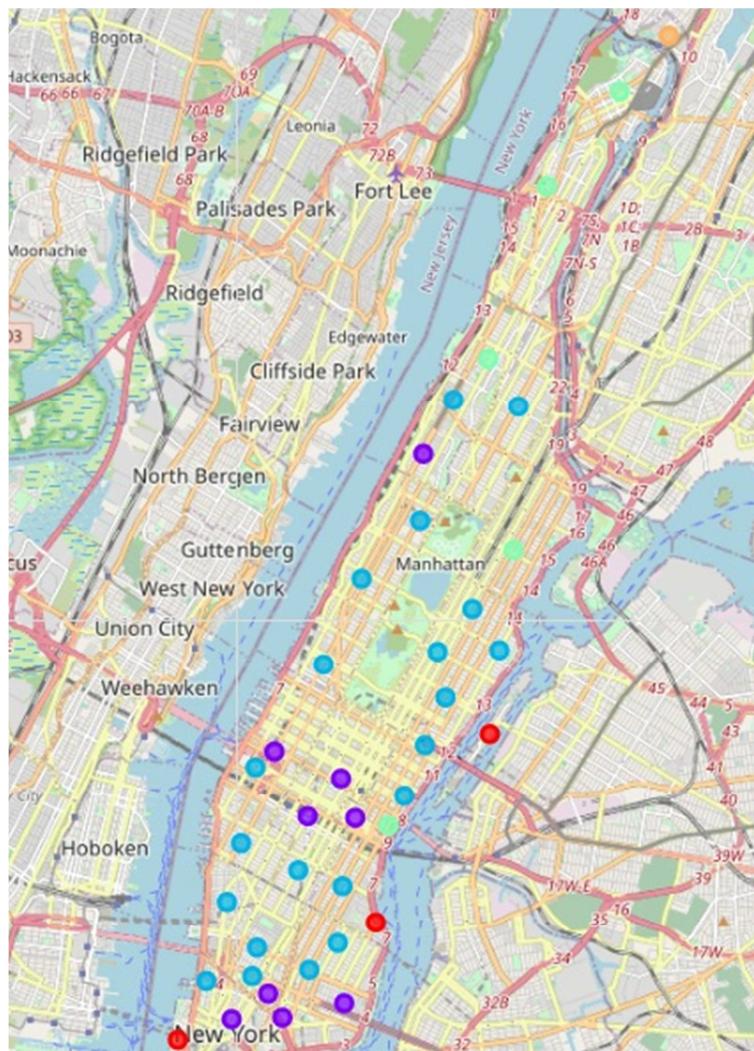
Manhattan Analysis

There are 40 neighbourhoods in Manhattan area.

After requesting Foursquare data, we have 3132 venues for all these 40 neighbourhoods. They belong to 331 unique categories.

Then, k-means algorithm is then executed for 5 clusters ($k=5$).

Then each neighbourhood is assigned to a cluster as in the following map:



It is observed the following distribution:

- Cluster 0 (Red):
 - 3 neighbourhood in the cluster, spotted in the East and South, all 3 river side

	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
11	Roosevelt Island	Park	Deli / Bodega	Dog Run	Coffee Shop	Greek Restaurant	Outdoors & Recreation	Monument / Landmark	Sandwich Place
28	Battery Park City	Park	Hotel	Memorial Site	Gym	Coffee Shop	Playground	Food Court	Mexican Restaurant
37	Stuyvesant Town	Boat or Ferry	Park	Playground	Heliport	Fountain	Farmers Market	Skating Rink	Gas Station

- Cluster 1 (Purple):

- Second most common, with 9 neighbourhoods, mostly at centre and South-East

	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
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Bubble Tea Shop	Coffee Shop	Optical Shop	Bar
14	Clinton	Theater	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Hotel	Gym	Cocktail Bar
15	Midtown	Hotel	Coffee Shop	Theater	Bakery	Clothing Store	Sushi Restaurant	Japanese Restaurant
16	Murray Hill	Hotel	Sandwich Place	Pizza Place	Japanese Restaurant	Coffee Shop	Gym / Fitness Center	Mediterranean Restaurant
20	Lower East Side	Chinese Restaurant	Bakery	Café	Ramen Restaurant	Art Gallery	Coffee Shop	Cocktail Bar
22	Little Italy	Spa	Bubble Tea Shop	Mediterranean Restaurant	Italian Restaurant	Ice Cream Shop	Pizza Place	Bakery
26	Morningside Heights	Park	Coffee Shop	American Restaurant	Bookstore	Burger Joint	Sandwich Place	Deli / Bodega
32	Civic Center	Coffee Shop	Hotel	Cocktail Bar	Spa	Park	American Restaurant	Gym / Fitness Center
33	Midtown South	Korean Restaurant	Hotel	Dessert Shop	Burger Joint	Coffee Shop	Japanese Restaurant	Café

- Cluster 2 (Blue):

- Most of the neighbourhoods of Manhattan are grouped in this cluster

	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
5	Manhattanville	Coffee Shop	Seafood Restaurant	Park	Bakery	Mexican Restaurant	Chinese Restaurant	Deli / Bodega
6	Central Harlem	African Restaurant	Seafood Restaurant	Gym / Fitness Center	American Restaurant	Bar	French Restaurant	Chinese Restaurant
8	Upper East Side	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Bakery	Yoga Studio	French Restaurant	Spa
9	Yorkville	Coffee Shop	Gym	Italian Restaurant	Bar	Sushi Restaurant	Pizza Place	Deli / Bodega
10	Lenox Hill	Coffee Shop	Pizza Place	Italian Restaurant	Sushi Restaurant	Cafe	Cocktail Bar	Gym / Fitness Center
12	Upper West Side	Bar	Italian Restaurant	Dessert Shop	Wine Bar	Indian Restaurant	Middle Eastern Restaurant	Bagel Shop
13	Lincoln Square	Plaza	Italian Restaurant	Cafe	Gym / Fitness Center	Theater	Performing Arts Venue	Concert Hall
17	Chelsea	Coffee Shop	Art Gallery	Ice Cream Shop	Cafe	American Restaurant	Bakery	Pizza Place
18	Greenwich Village	Italian Restaurant	Cafe	Sushi Restaurant	Burger Joint	Spa	Comedy Club	Dessert Shop
19	East Village	Mexican Restaurant	Bar	Coffee Shop	Cocktail Bar	Speakeasy	Ice Cream Shop	Juice Bar
21	Tribeca	Park	American Restaurant	Italian Restaurant	Wine Bar	Greek Restaurant	Spa	Cafe
23	Soho	Italian Restaurant	Coffee Shop	Sandwich Place	Mediterranean Restaurant	Clothing Store	Spa	Ice Cream Shop
24	West Village	Italian Restaurant	Wine Bar	American Restaurant	Pizza Place	Park	Jazz Club	Cocktail Bar
25	Manhattan Valley	Coffee Shop	Yoga Studio	Pizza Place	Mexican Restaurant	Bar	Park	Spa
27	Gramercy	Bar	Coffee Shop	Pizza Place	Bagel Shop	Mexican Restaurant	Italian Restaurant	Cocktail Bar
29	Financial District	Coffee Shop	Pizza Place	Hotel	American Restaurant	Italian Restaurant	Cafe	Juice Bar
30	Carnegie Hill	Coffee Shop	Cafe	Yoga Studio	Bookstore	Gym / Fitness Center	Gym	Italian Restaurant
31	Noho	Pizza Place	Coffee Shop	Grocery Store	Japanese Restaurant	Italian Restaurant	Sushi Restaurant	Mexican Restaurant
34	Sutton Place	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Park	Furniture / Home Store	Hotel	Gym
35	Turtle Bay	Italian Restaurant	Coffee Shop	Park	Cafe	Sushi Restaurant	Seafood Restaurant	Deli / Bodega
38	Flatiron	Gym / Fitness Center	Mediterranean Restaurant	Cafe	Italian Restaurant	Coffee Shop	Yoga Studio	Spa
39	Hudson Yards	Hotel	Italian Restaurant	American Restaurant	Gym / Fitness Center	Cafe	Gym	Nightclub

- Cluster 3 (Green):
 - 5 neighbourhoods in the cluster, 4 of them at the North

	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
2	Washington Heights	Café	Bakery	Deli / Bodega	Mobile Phone Shop	Chinese Restaurant	Grocery Store	Mexican Restaurant	Bank
3	Inwood	Mexican Restaurant	Lounge	Bakery	Café	Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
4	Hamilton Heights	Pizza Place	Mexican Restaurant	Coffee Shop	Deli / Bodega	Café	Sushi Restaurant	Bakery	Cocktail Bar
7	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Deli / Bodega	Sandwich Place	Latin American Restaurant	Pizza Place	French Restaurant
36	Tudor City	Café	Park	Mexican Restaurant	Pizza Place	Sushi Restaurant	Deli / Bodega	Greek Restaurant	Thai Restaurant

- Cluster 4 (Orange):
 - Only one neighbourhood in the cluster, Marble Hill, placed at the North of Manhattan

	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
0	Marble Hill	Sandwich Place	Gym	Coffee Shop	Yoga Studio	Pharmacy	Supplement Shop	Steakhouse	Seafood Restaurant

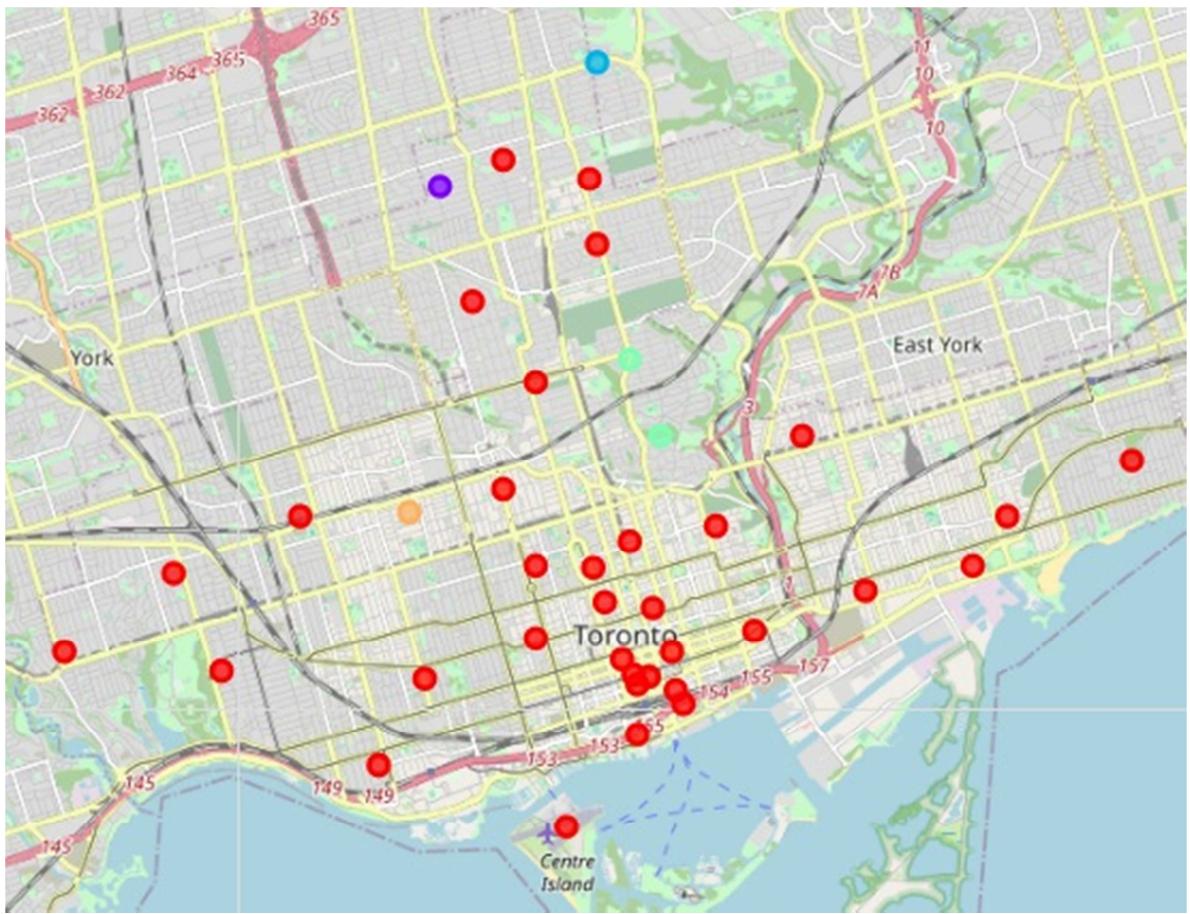
Toronto Analysis

There are 39 neighbourhoods in Toronto area.

After requesting Foursquare data, we have 1618 venues for these 39 neighbourhoods. They belong to 237 unique categories.

Then, k-means algorithm is then executed for 5 clusters (k=5).

Then each neighbourhood is assigned to a cluster as in the following map:



It is observed the following distribution:

- Cluster 0 (Red):
 - Majority of the neighbourhoods of Toronto are grouped in this cluster

	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
0	East Toronto	0	Trail	Health Food Store	Pub	Yoga Studio	Diner	Discount Store	Distribution Center
1	East Toronto	0	Greek Restaurant	Italian Restaurant	Coffee Shop	Bookstore	Restaurant	Ice Cream Shop	Furniture / Home Store
2	East Toronto	0	Sandwich Place	Fast Food Restaurant	Gym	Pet Store	Brewery	Burrito Place	Restaurant
3	East Toronto	0	Café	Coffee Shop	Brewery	Gastropub	Bakery	American Restaurant	Convenience Store
5	Central Toronto	0	Gym	Hotel	Park	Food & Drink Shop	Sandwich Place	Department Store	Breakfast Spot
6	Central Toronto	0	Clothing Store	Coffee Shop	Yoga Studio	Chinese Restaurant	Spa	Sporting Goods Shop	Diner
7	Central Toronto	0	Dessert Shop	Sandwich Place	Sushi Restaurant	Cafe	Pizza Place	Coffee Shop	Gym
9	Central Toronto	0	Coffee Shop	Light Rail Station	Pub	Pizza Place	Bank	Bagel Shop	Restaurant
11	Downtown Toronto	0	Café	Coffee Shop	Pizza Place	Restaurant	Bakery	Italian Restaurant	Pub
12	Downtown Toronto	0	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Restaurant	Gay Bar	Café	Pub
13	Downtown Toronto	0	Coffee Shop	Park	Bakery	Pub	Breakfast Spot	Café	Restaurant
14	Downtown Toronto	0	Clothing Store	Coffee Shop	Bubble Tea Shop	Japanese Restaurant	Italian Restaurant	Café	Middle Eastern Restaurant
15	Downtown Toronto	0	Coffee Shop	Café	Cocktail Bar	Gastropub	American Restaurant	Restaurant	Hotel
16	Downtown Toronto	0	Coffee Shop	Cocktail Bar	Cheese Shop	Beer Bar	Bakery	Seafood Restaurant	Restaurant
17	Downtown Toronto	0	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Japanese Restaurant	Department Store	Salad Place
18	Downtown Toronto	0	Coffee Shop	Café	Restaurant	Gym	Deli / Bodega	Clothing Store	Hotel
19	Downtown Toronto	0	Coffee Shop	Aquarium	Café	Hotel	Restaurant	Brewery	Sporting Goods Shop
20	Downtown Toronto	0	Coffee Shop	Hotel	Café	Restaurant	Italian Restaurant	Salad Place	Seafood Restaurant
21	Downtown Toronto	0	Coffee Shop	Restaurant	Café	Hotel	American Restaurant	Gym	Deli / Bodega
23	Central Toronto	0	Jewelry Store	Trail	Sushi Restaurant	Mexican Restaurant	Yoga Studio	Diner	Event Space
24	Central Toronto	0	Sandwich Place	Café	Coffee Shop	BBQ Joint	Donut Shop	Burger Joint	Indian Restaurant
25	Downtown Toronto	0	Café	Bar	Italian Restaurant	Japanese Restaurant	Bookstore	Bakery	Restaurant
26	Downtown Toronto	0	Café	Coffee Shop	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Mexican Restaurant	Gaming Cafe	Bar

- Cluster 1 (Purple):
 - Only 1 neighbourhood in this cluster

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	
22	Central Toronto	1	Ice Cream Shop	Home Service	Garden	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant

- Cluster 2 (Red):
 - Only 1 neighbourhood in this cluster

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	
4	Central Toronto	2	Park	Swim School	Bus Line	Farmers Market	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant

- Cluster 3 (Green):
 - Only 2 neighbourhoods in this cluster

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	
8	Central Toronto	3	Gym	Park	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
10	Downtown Toronto	3	Park	Playground	Trail	Department Store	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant

- Cluster 4 (Orange):
 - Majority of the neighbourhoods of Toronto are grouped in this cluster

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	
30	Downtown Toronto	4	Grocery Store	Café	Park	Coffee Shop	Nightclub	Candy Store	Baby Store	Italian Restaurant

Analysis of Toronto and Manhattan together

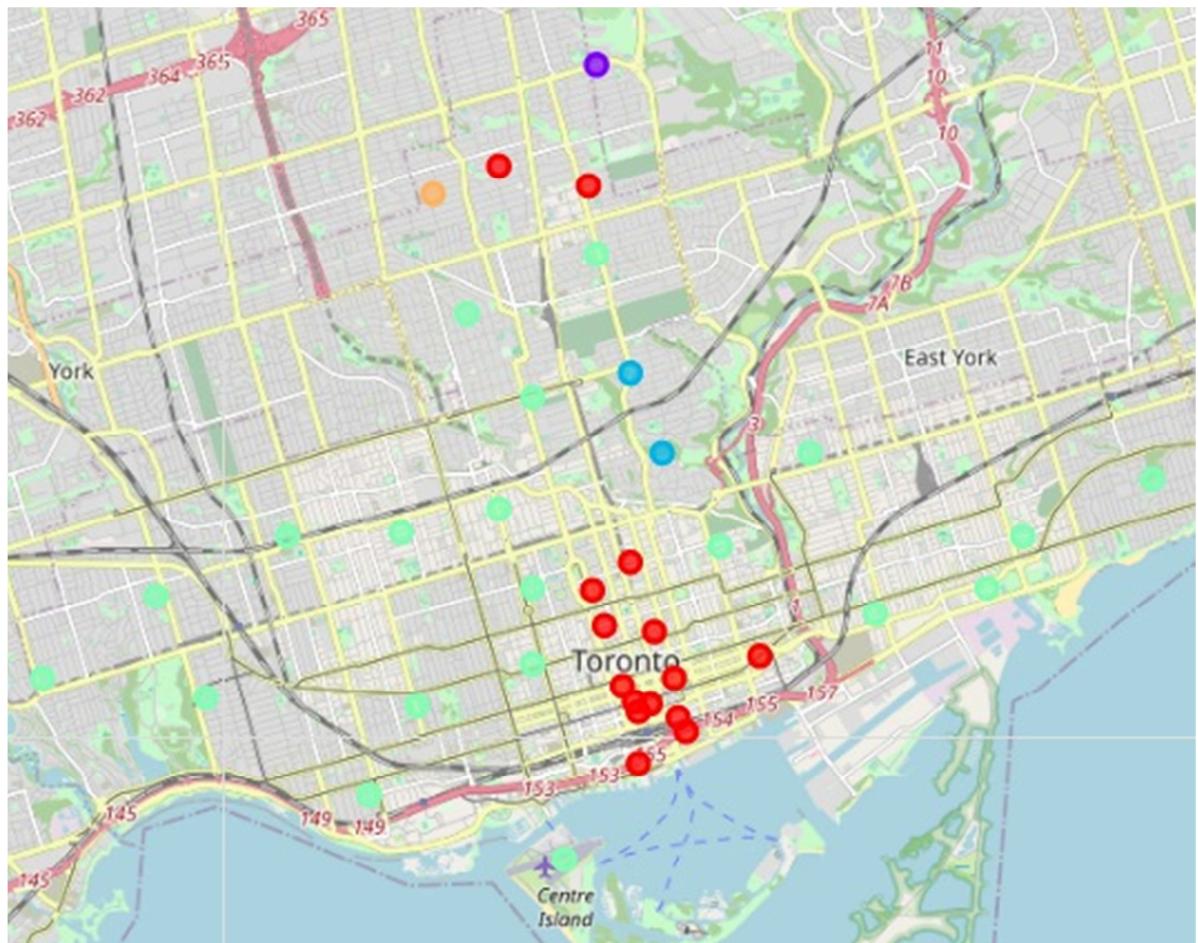
There are 79 neighbourhoods in Toronto plus Manhattan areas.

After requesting Foursquare data, we have 4750 venues for these 79 neighbourhoods. They belong to 237 unique categories.

Then, k-means algorithm is then executed for 5 clusters (k=5).

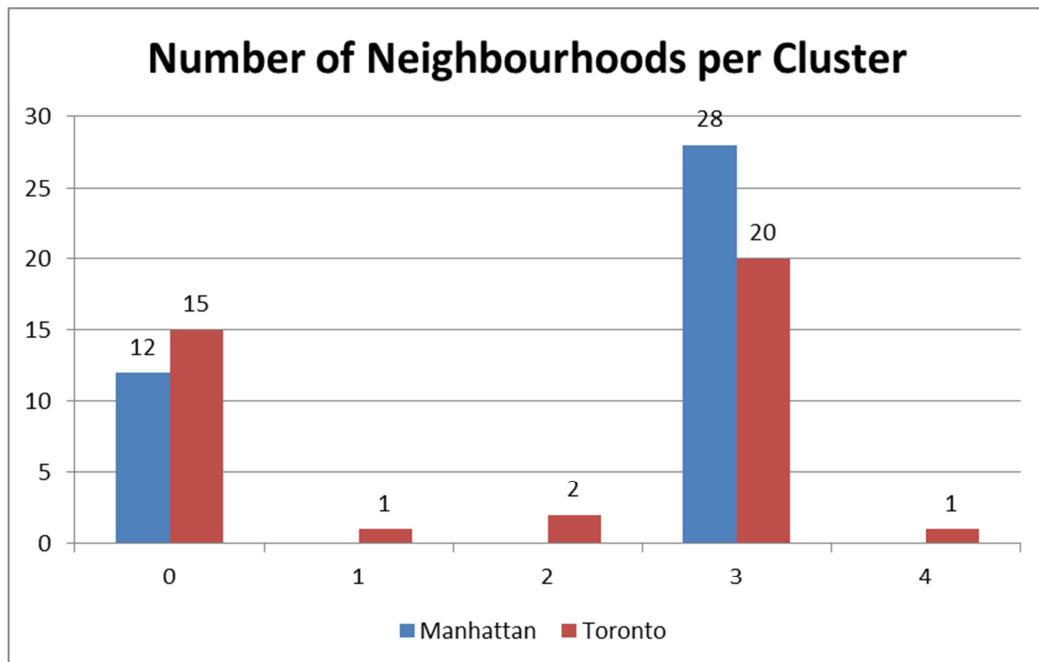
Then each neighbourhood is assigned to a cluster as in the following maps, one centred in Manhattan and the other one centred in Toronto:





Note that cluster distribution has changed quite in both areas.

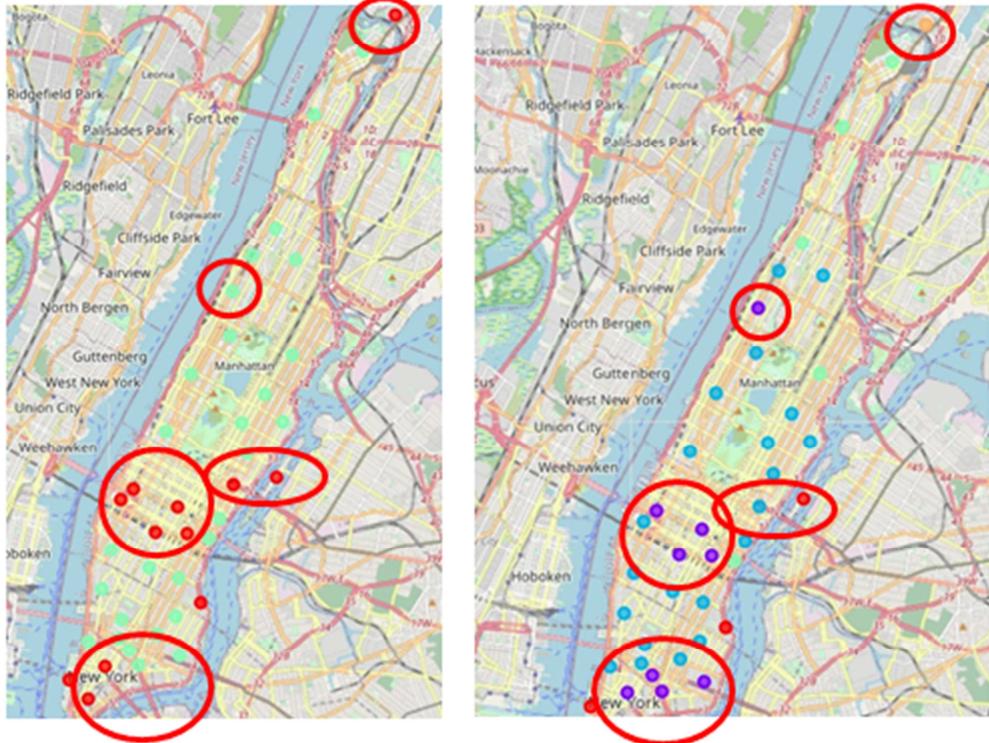
New distribution is shown in next graph:



For Manhattan, next picture illustrates the main differences:

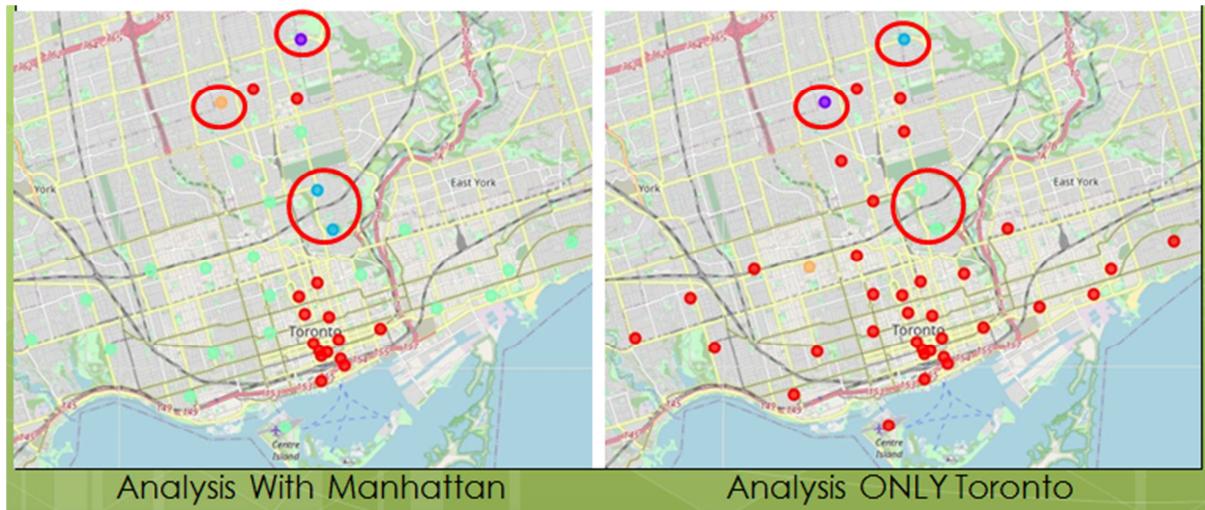
- For combined analysis, only 2 cluster are present in Manhattan area
- All other 3 clusters have been merged with these other 2
- Cluster 0 includes 12 neighbourhoods
- Cluster 3 has 28 neighbourhoods

Analysis With Toronto Analysis ONLY Manhattan



For Toronto, main highlights can be observed in next picture

- All 5 clusters are present in Toronto
- Similarly to Manhattan majority of neighbourhoods are in cluster 0 and 3
- Cluster 0 has 15 neighbourhoods
- Cluster 3 has 20 neighbourhoods
- Other 3 clusters (1, 2 and 4) have same neighbourhoods as they had in the analysis for Toronto only
- Another Cluster has been integrated into Cluster 3
- The big cluster (red one in Toronto only analysis) has been split into two, matching the two clusters in Manhattan



The following neighbourhoods are the ones that are so peculiar in Toronto that they cannot be clustered with any Manhattan neighbourhood.

- Lawrence Park (cluster 1)
- Moore Park, Summerhill East & Rosedale (cluster 2)
- Roselawn (cluster 4)

Discussion

Manhattan had several defined areas different between them.

When compared with Toronto, Manhattan has not so many differences and all neighbourhoods fall into 2 clusters.

Toronto was very similar, with 34 neighbourhoods in the same cluster.

When compared with Manhattan, it is observed 2 different areas (clusters 0 and 3).

Clusters 1, 2 and 3 from ONLY Toronto analysis are kept into new clusters 1, 2 and 4

Conclusion

We can conclude that Toronto and Manhattan are quite similar in terms of venues, which is what was analysed here.

The following neighbourhoods are the ones that are so peculiar in Toronto that they cannot be clustered with any Manhattan neighbourhood.

- Lawrence Park (cluster 1)
- Moore Park, Summerhill East & Rosedale (cluster 2)
- Roselawn (cluster 4)

For further analysis are recommended the following:

- Check why these neighbourhoods in Toronto not clustered with any in Manhattan are so peculiar
- Perform same analysis changing k parameter, to check what would be the behaviour of these neighbourhoods