The Battle of Neighbourhoods

Toronto VS New York City

Julian Fan

Introduction

- For years, Toronto and New York City have been pitted against each other in a no-holds-barred comparison to determine which is the better city, and what are their similarities and differences.

- To explore the neighborhoods of both cities, and show the similarity and difference of their neighborhoods by the clustering method on the combined data from both cities.

- It could provide good guidance for selecting right neighborhood which fits well for the living style. It is particularly useful for people relocating from New York to Toronto or from Toronto to New York.

Data Source & Preparation

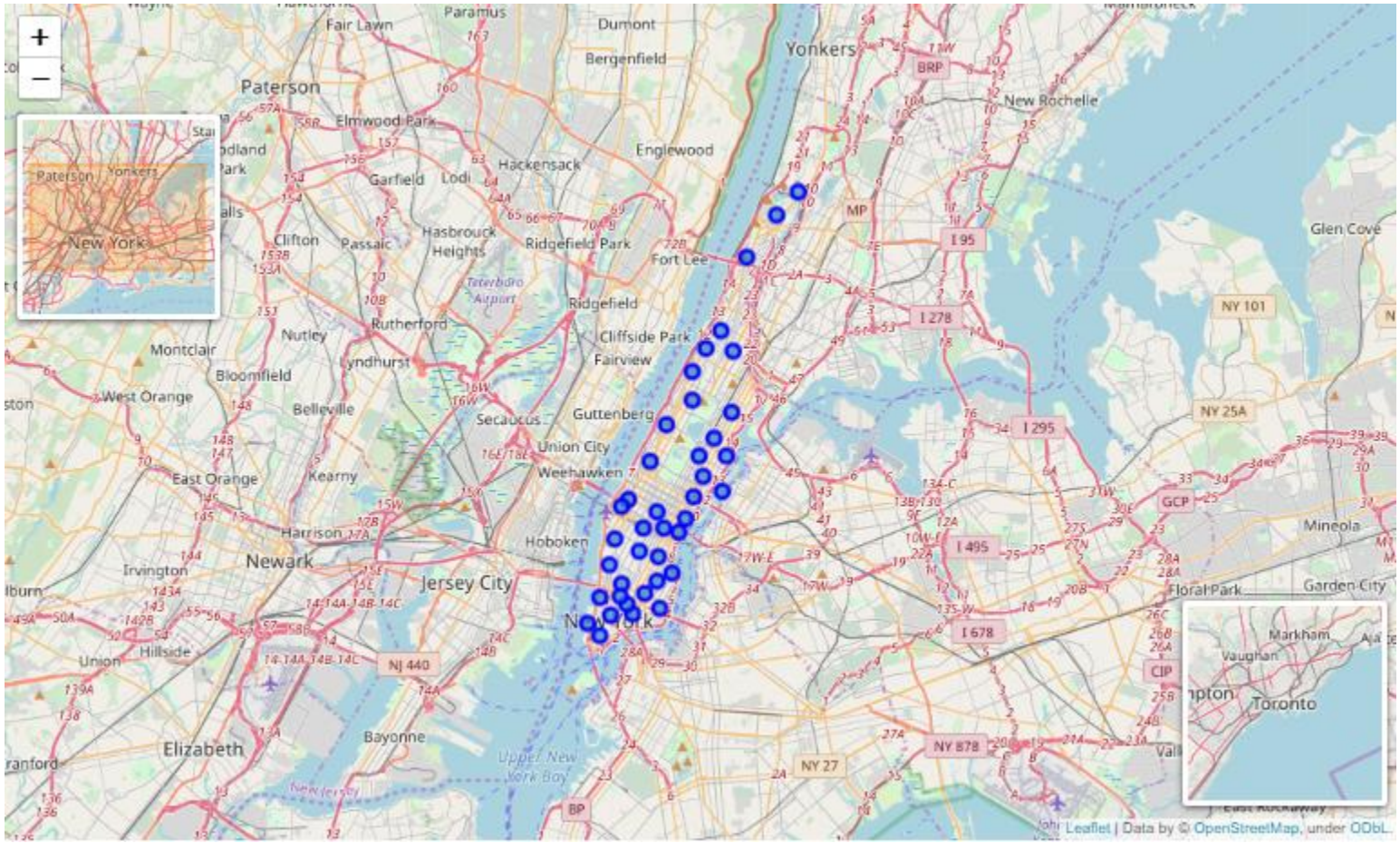
- For this project the Foursquare API will be used. A list of neighborhoods in New York and Toronto is downloaded and their respective location in longitude and latitude coordinates is obtained. The sources are the following:
- New York neighborhoods:
- https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json
- Toronto neighborhoods:
- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Data Source & Preparation

- A Foursquare API GET request is sent in order to acquire the surrounding venues that are within a radius of 500m. The data is formatted using one hot encoding with the categories of each venue. Then, the venues are grouped by neighborhoods computing the mean of each feature.

	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	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Asian Restaurant	Bakery	American Restaurant	Sushi Restaurant	Steakhouse	Hotel	Restaurant
1	Battery Park City	Park	Coffee Shop	Hotel	Gym	Memorial Site	Wine Shop	Plaza	Playground	Pizza Place	Italian Restaurant
2	Berczy Park	Coffee Shop	Italian Restaurant	Cheese Shop	Seafood Restaurant	Bakery	Beer Bar	Steakhouse	Café	Cocktail Bar	Farmers Market
3	Brockton, Exhibition Place, Parkdale Village	Café	Coffee Shop	Breakfast Spot	Italian Restaurant	Caribbean Restaurant	Bakery	Furniture / Home Store	Climbing Gym	Restaurant	Grocery Store
4	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Auto Workshop	Gym / Fitness Center	Fast Food Restaurant	Comic Shop	Brewery	Burrito Place	Smoke Shop	Skate Park	Garden

Neighborhoods in Manhattan



Neighborhoods In Central Toronto

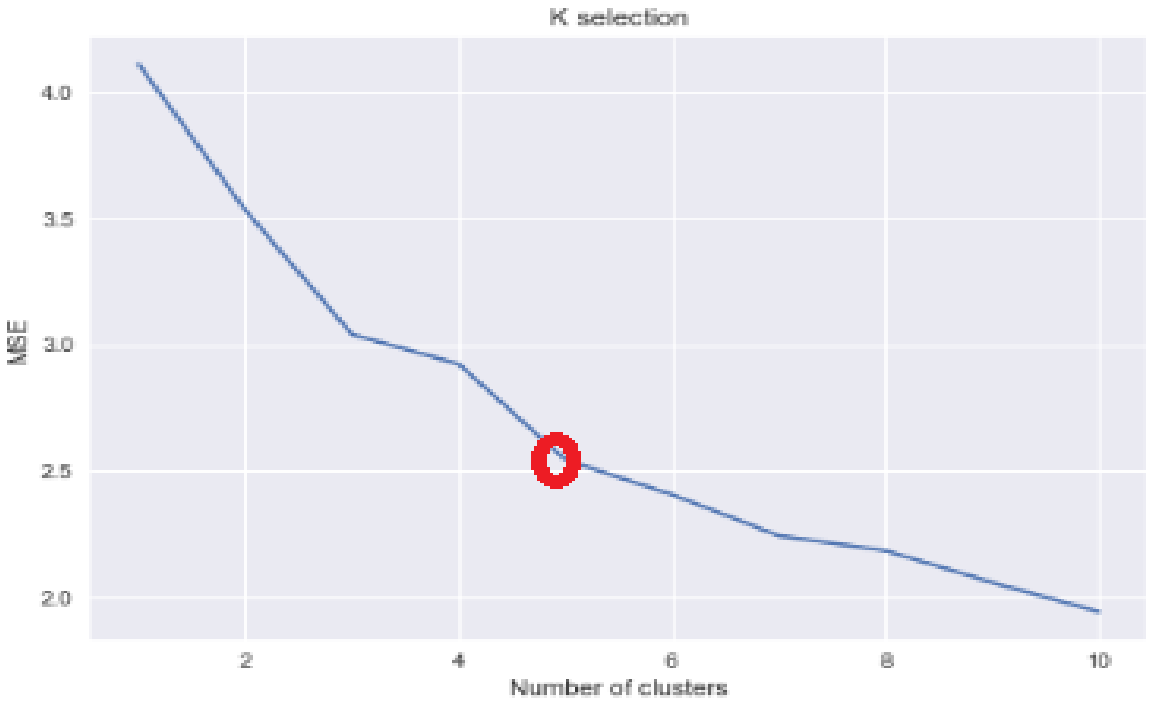
Methodology: Feature Extraction

For feature extraction, One Hot Encoding is used in terms of categories. Therefore, each feature is a category that belongs to a venue. Each feature becomes binary, this means that 1 means this category is found in the venue and 0 means the opposite. Then, all the venues are grouped by the neighborhoods, computing at the same time the mean. This will give us a venue for each row and each column will contain the frequency of occurrence of that particular category.

	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Animal Shelter	Antique Shop	Aquarium
0	Adelaide, King, Richmond	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.030000	0.0	0.0	0.0
1	Battery Park City	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.010000	0.0	0.0	0.0
2	Berczy Park	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
75	Washington Heights	0.0	0.011628	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.011628	0.0	0.0	0.0
76	West Village	0.0	0.010000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.040000	0.0	0.0	0.0
77	Yorkville	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0

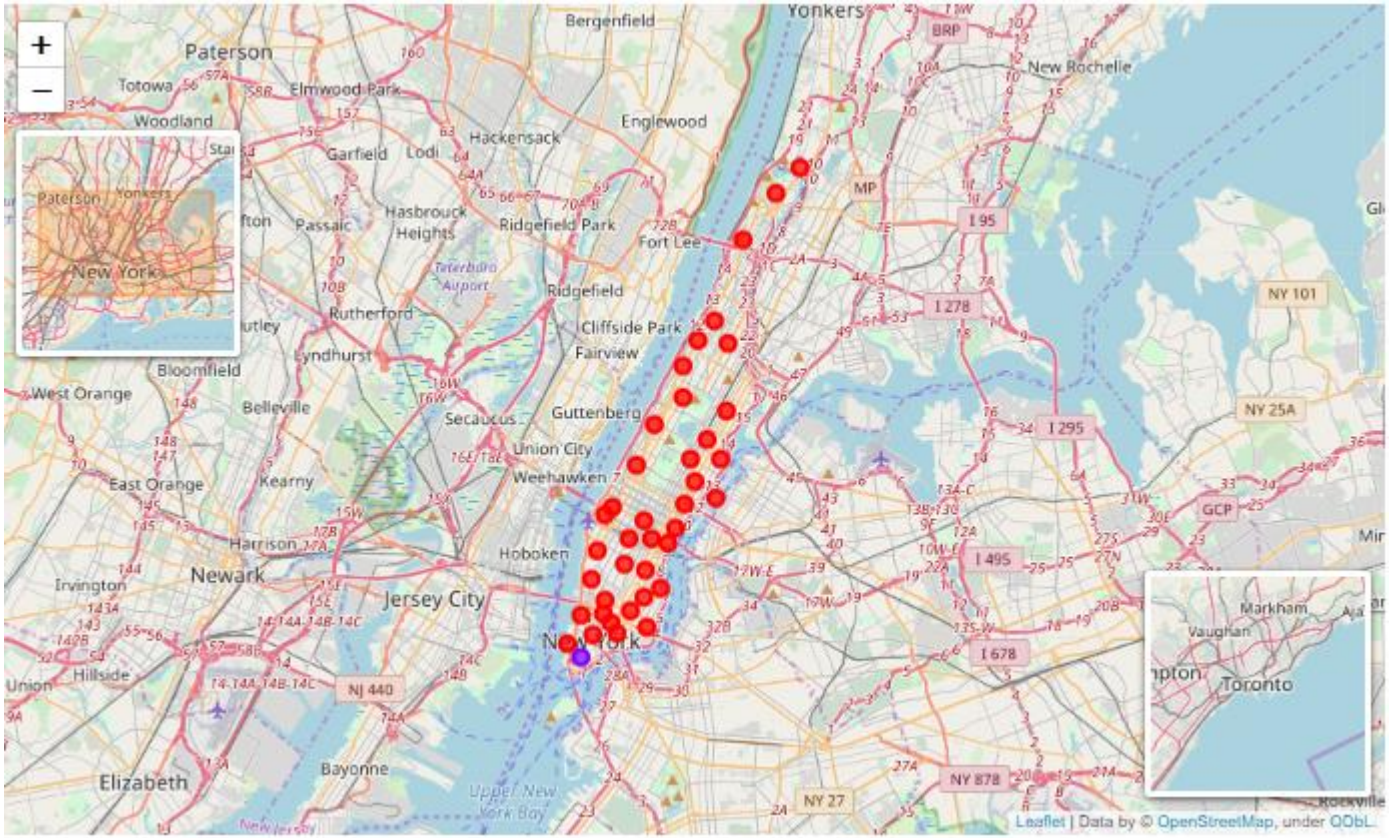
78 rows × 375 columns

Unsupervised Learning: K-Means



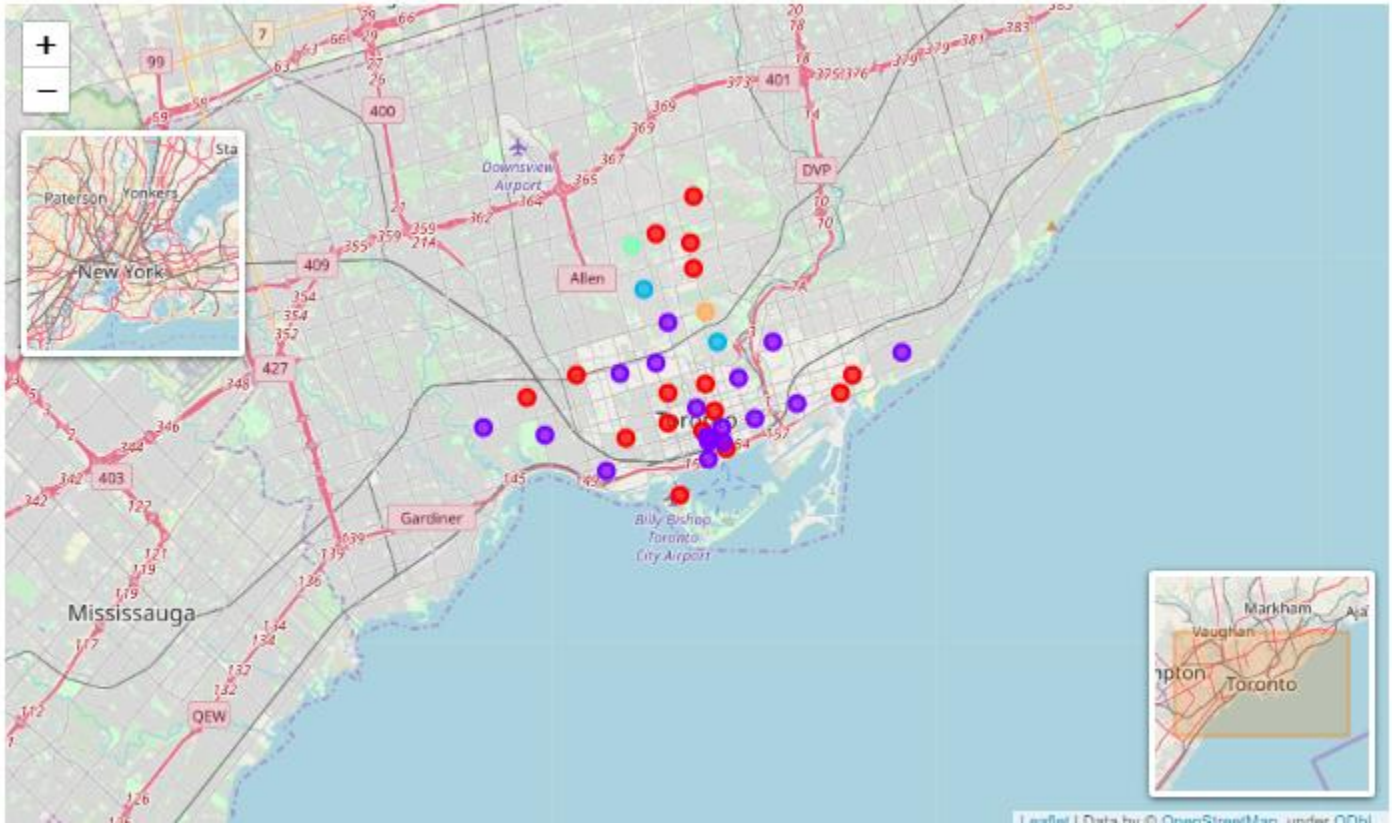
K=5 based on Elbow method

Clustering Result



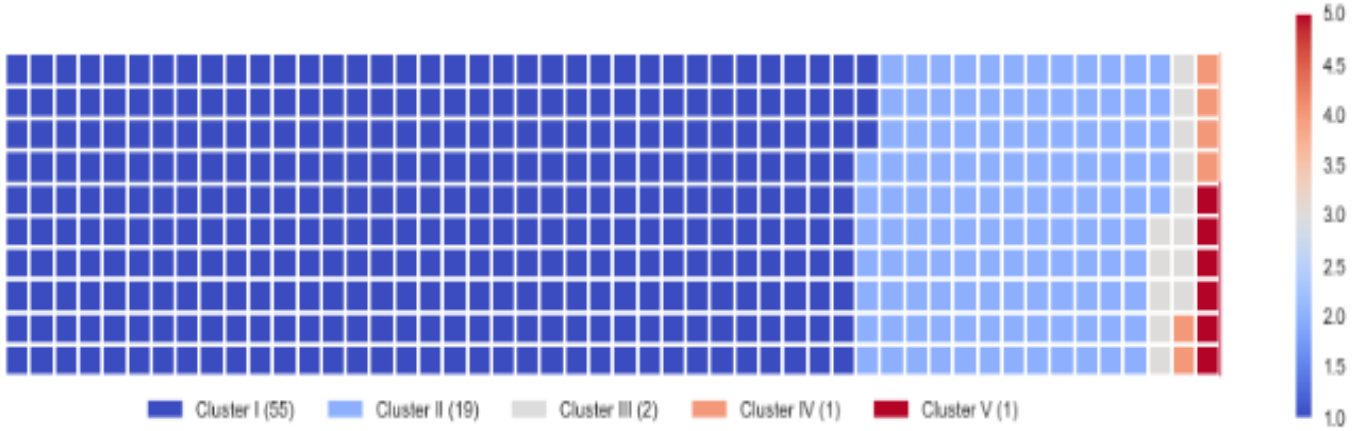
Cluster display on Manhattan

Clustering Result



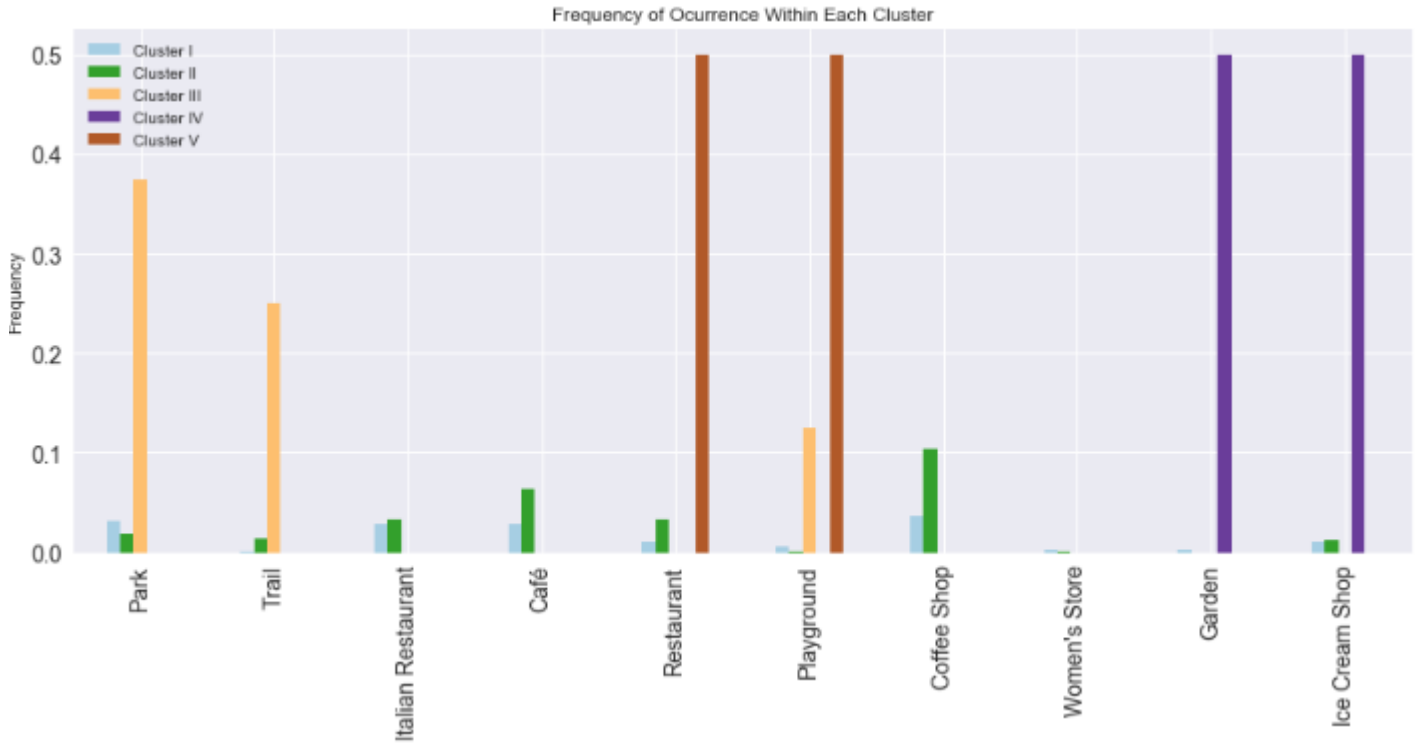
Cluster display on Toronto

Clustering Result



Cluster Waffle chart

Clustering Result



Cluster bar chart

Conclusion

- The K-Means clustering algorithm is used for finding similarities between all the neighborhoods. The elbow method is used for selecting the appropriate number of clusters.

- 5 clusters are obtained using K-means. Most of the neighborhoods (55) in Manhattan and Toronto fall into Cluster 1, where there balanced distribution on restaurants, Café, playground and park. Cluster 2 consists of 19 neighborhoods from Manhattan and Toronto, where there are a lot of Cafe and coffee shops. There are 3 special Clusters in Toronto. Cluster 3 is characterized by its outdoor attractions: park, trail and playground. Cluster 4 is a special garden-dominated area with plenty of Ice Cream Shops. Cluster 5 is dominated by restaurants and playgrounds.

- Hopefully the analysis results can provide a good guidance for people who want to move from Manhattan to Toronto and vice versa to get an idea what is the best suitable place for them