The Battle of Neighbourhoods

Toronto VS New York City

Julian Fan

1. Introduction

New York (also called New York City, short form NYC) is the biggest city in the United States, located in the state of New York. Over 8 million people live in it, and over 22 million people live in the bigger New York metropolitan area. It is in the south end of the state of New York, which is in the northeastern United States. It is the financial capital of the US since it is home to the nation's stock market, Wall Street, and the One World Trade Center. It is also the home of the United Nations Headquarters. The central and oldest part of the city is Manhattan.

Toronto is the capital city of the province of Ontario in Canada. It is also the largest city in both Ontario and Canada. Found on the north-west side of Lake Ontario, the City of Toronto has a population of over 3 million people and even more people live in the regions around it. All together, the Greater Toronto Area is home to over 6 million people making it the biggest metropolitan area in Canada. Toronto is the fourth-largest metropolitan area in North America behind Los Angeles, New York, and Chicago

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. New York is clearly the more famous, with its well-known history and touristy flair. But how does it compare to Toronto when it comes to cost of living? A general guess is that it's the pricer of the two, but it's not until you take a look at the numbers that you realize just how much pricier it really is. The comparison between two cities are shown in the figure below.

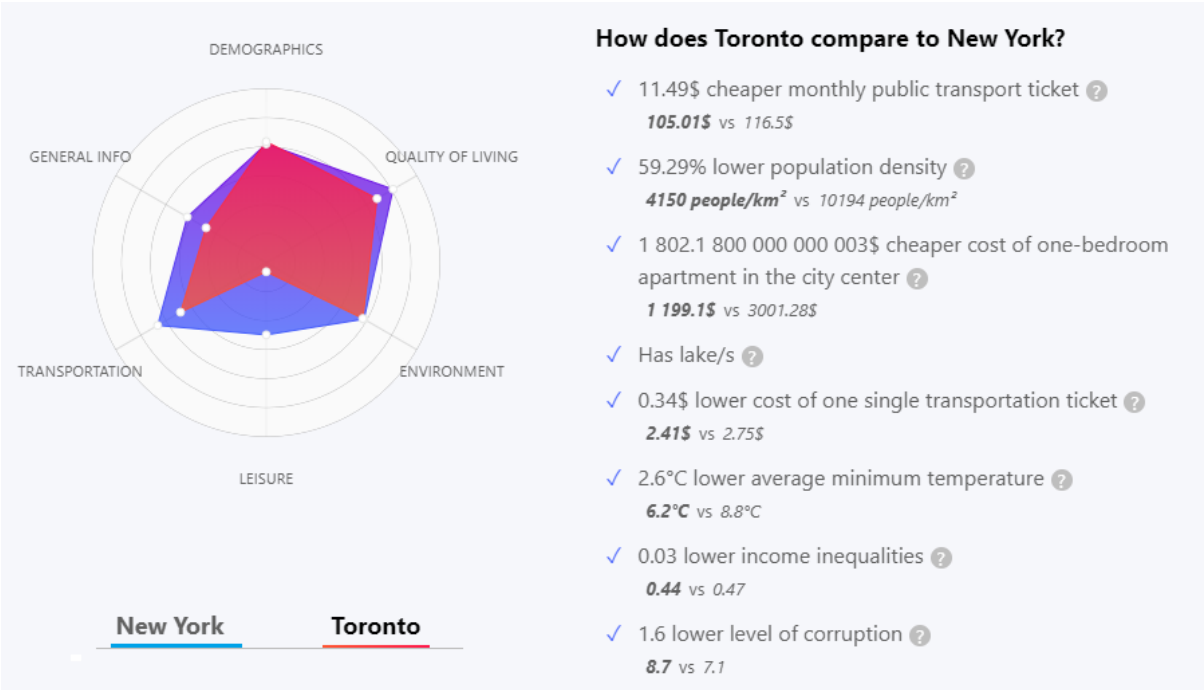

Figure 1. How does Toronto compare to New York

What we try to do in this report is to explore the neighborhoods of both cities, and show the similarity and difference of their neighborhoods areas by the clustering method on the combined data from both cities. It could provide good guidance for selecting right neighborhood which fits well for their living style. It is particularly useful for people relocating from New York to Toronto or from Toronto to New York.

2. Data

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

The data downloaded are the neighborhoods located in New York and Toronto. Moreover, their specific coordinates are merged. Only Manhattan neighborhoods and boroughs that contain the string "Toronto" are taken into account. 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.

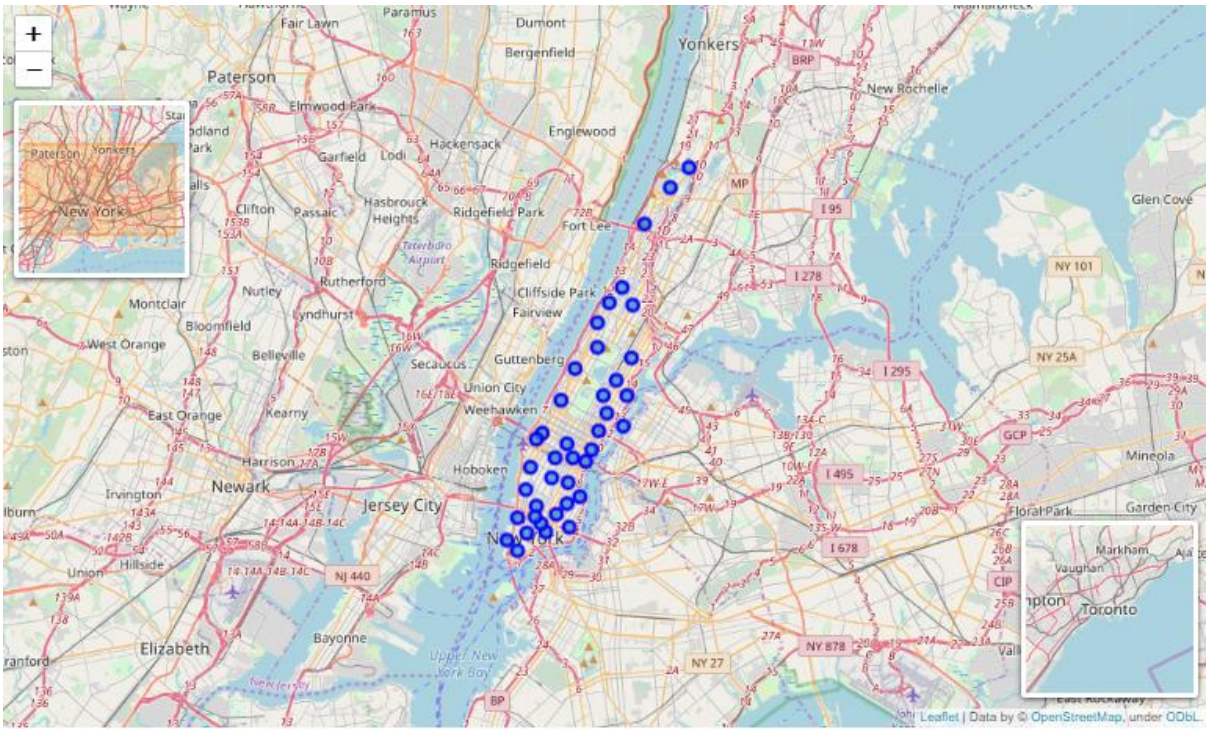

Figure 2. Neighborhoods in Manhattan

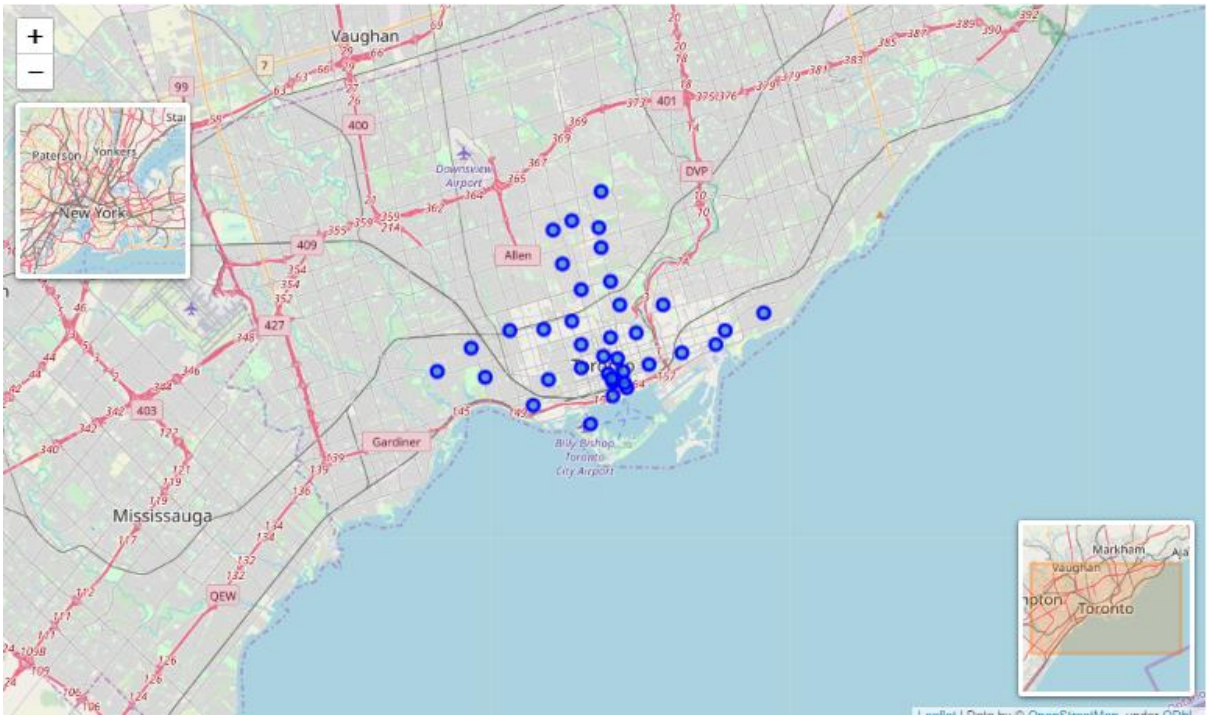

Figure 3. Neighborhoods In Central Toronto

The neighborhood data of two cities acquired from the Foursquare API will be combined and then used for clustering analysis. The similarities will be determined based on the frequency of the categories found in the neighborhoods. These similarities found are a strong indicator for a user and can help him to decide whether to move in a particular neighborhood.

	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

Figure 4. Information around the mixed neighborhoods from Foursquare API

3. Methodology

3.1. 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

Figure 5. Normalized Information around the mixed neighborhoods

3.2. Unsupervised Learning

We will do a clustering analysis on the combined neighborhood data of both cities. The purpose of doing the clustering to found similarities among the neighborhoods of two cities. In this case K-Means is used due to its simplicity and efficiency.

K-Means is a clustering algorithm. This algorithm search clusters within the data and the main objective function is to minimize the data dispersion for each cluster. Thus, each group found represents a set of data with a pattern inside the multidimensional features.

Contrary to supervised learning where we have the ground truth to evaluate the model's performance, clustering analysis doesn't have a solid evaluation metric that we can use to evaluate the outcome of different clustering algorithms. Moreover, since k-Means requires k as an input and doesn't learn it from data, there is no right answer in terms of the number of clusters that we should have in any problem. Sometimes domain knowledge and intuition may help but usually that is not the case. In the cluster-predict methodology, we can evaluate how well the models are performing based on different K clusters since clusters are used in the downstream modeling.

Elbow method gives us an idea on what a good k number of clusters would be based on the sum of squared distance (MSE) between data points and their assigned clusters' centroids. We pick k at the spot where SSE starts to flatten out and forming an elbow, which indicates K=5. Then, further clustering analysis of each cluster is done.

Figure 6. K=5 based on Elbow method

4. Clustering Results

Clustering analysis result is given in figure 7, in which every cluster information is clearly listed. The borough and most common venues are given in the figure. The clustering information are also added on the maps clearly in figure 8 and 9. For visualization purposes, the geographical data is again plotted but with different colors. Each color represents the cluster for which that neighborhood belongs.

In this image it is evident that cluster algorithm is not segmenting the neighborhoods for location areas. This means that it is not true that geolocation of neighborhoods is correlated with the categories of the venues around each neighborhood. Yet, it is possible to see which neighborhoods within Manhattan, New York are more similar to the neighborhoods within Toronto. Those neighborhoods that are similiar among them belong to the same cluster. Hence, they have the same color in the image above. In fact, all three minors clusters are from Toronto.

To identify the proportion of the neighborhoods assigned to each cluster, a waffle chart is implemented in figure 10. . There are two major clusters and three minor clusters. To study the characteristics of each cluster, a bar char is plotted for each cluster, as shown in Figure 11.

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.

.

clst

1

	Borough	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
0	Manhattan	Marble Hill	Sandwich Place	Coffee Shop	Donut Shop	Supplement Shop	Steakhouse	Big Box Store	Miscellaneous Shop	Shopping Mall	R
1	Manhattan	Chinatown	Chinese Restaurant	Vietnamese Restaurant	Cocktail Bar	American Restaurant	Spa	Optical Shop	Salon / Barbershop	Bakery	
2	Manhattan	Washington Heights	Bakery	Café	Mobile Phone Shop	Grocery Store	Spanish Restaurant	Donut Shop	Pizza Place	Coffee Shop	
...	...	...	...	...	...	...	...	...	...	...	...
72	West Toronto	Little Portugal, Trinity	Bar	Coffee Shop	Men's Store	Restaurant	Asian Restaurant	Vietnamese Restaurant	New American Restaurant	Bakery	R
74	West Toronto	High Park, The Junction South	Mexican Restaurant	Bar	Thai Restaurant	Café	Italian Restaurant	Cajun / Creole Restaurant	Furniture / Home Store	Discount Store	
77	East Toronto	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	S

55 rows × 12 columns

[54]:

2

	Borough	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
29	Manhattan	Financial District	Coffee Shop	Hotel	American Restaurant	Gym	Food Truck	Pizza Place	Event Space	Wine Shop	Café
40	East Toronto	The Beaches	Pub	Trail	Health Food Store	Women's Store	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
41	East Toronto	The Danforth West, Riverdale	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Brewery	Bakery	Liquor Store	Lounge
...	...	...	...	...	...	...	...	...	...	...	...
73	West Toronto	Brockton, Exhibition Place, Parkdale Village	Café	Coffee Shop	Breakfast Spot	Italian Restaurant	Caribbean Restaurant	Bakery	Furniture / Home Store	Climbing Gym	Restaurant
75	West Toronto	Parkdale, Roncesvalles	Gift Shop	Coffee Shop	Cuban Restaurant	Restaurant	Eastern European Restaurant	Italian Restaurant	Bookstore	Movie Theater	Ba
76	West Toronto	Runnymede, Swansea	Coffee Shop	Café	Sushi Restaurant	Italian Restaurant	Pizza Place	Gastropub	Restaurant	Tea Room	Diner

19 rows × 12 columns

3

	Borough	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
50	Downtown Toronto	Rosedale	Park	Playground	Trail	Women's Store	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
63	Central Toronto	Forest Hill North, Forest Hill West	Park	Trail	Sushi Restaurant	Jewelry Store	Women's Store	Electronics Store	Doner Restaurant	Donut Shop	Drugstore

4

	Borough	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
62	Central Toronto	Roselawn	Ice Cream Shop	Garden	Women's Store	Empanada Restaurant	Doner Restaurant	Donut Shop	Drugstore	Dry Cleaner	Dumpling Restaurant

5

	Borough	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
48	Central Toronto	Moore Park, Summerhill East	Playground	Restaurant	Women's Store	English Restaurant	Doner Restaurant	Donut Shop	Drugstore	Dry Cleaner	Dumpling Restaurant

Figure 7. Clustering result

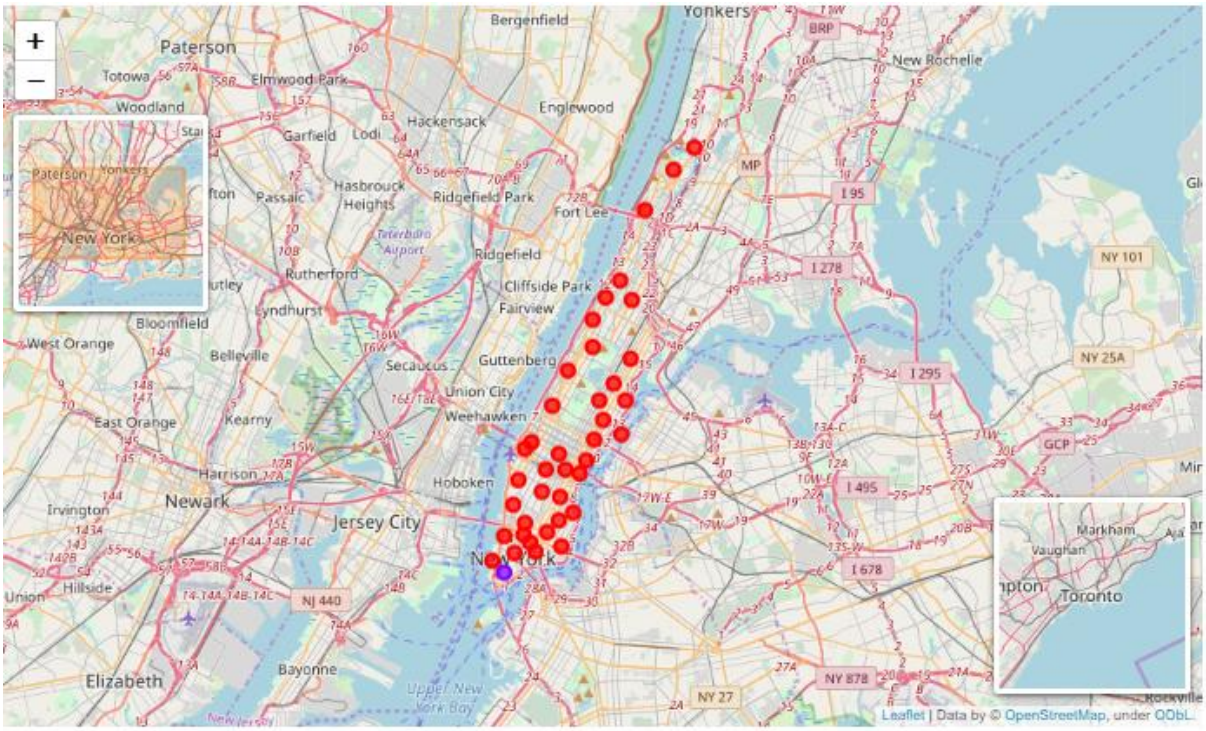

Figure 8. Cluster display on Manhattan

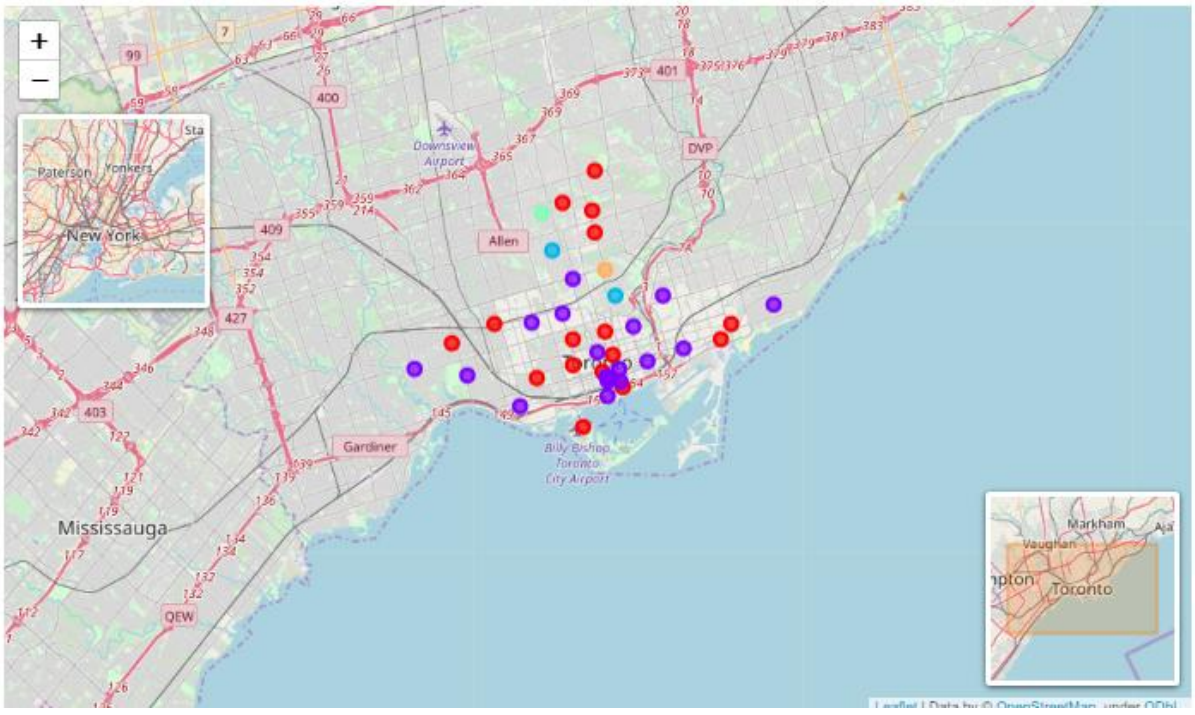

Figure 9. Cluster display on Toronto

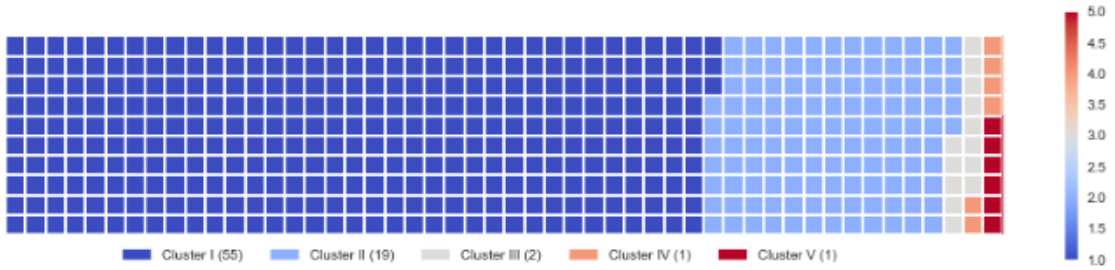

Figure 10. Cluster Waffle chart

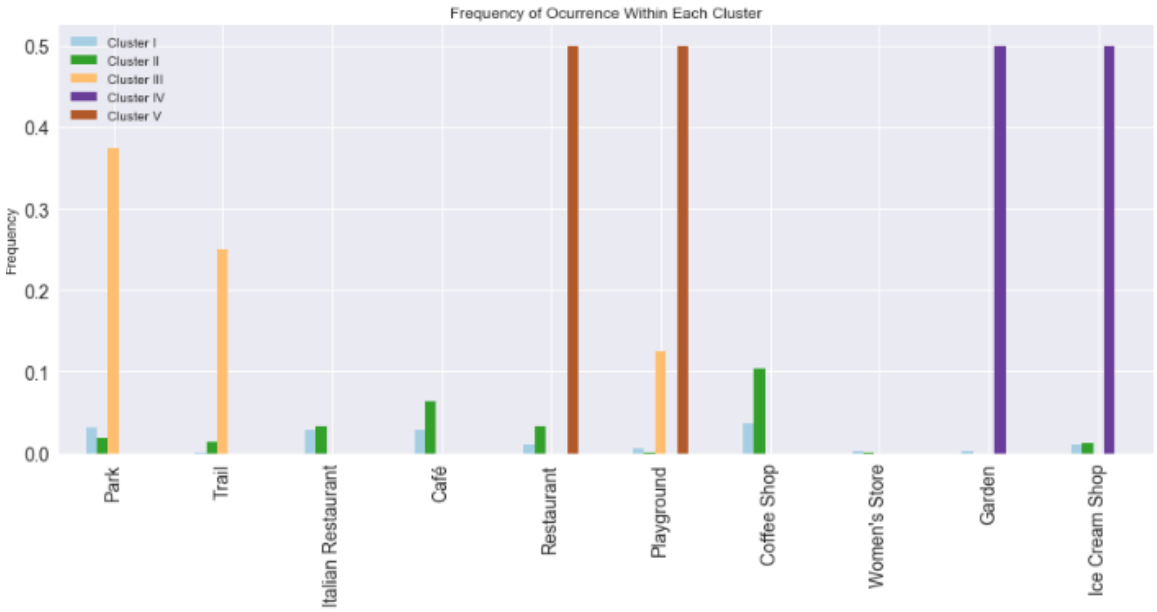

Figure 11. Cluster bar chart

5. Conclusion

In this work a segmentation between the neighborhoods in Manhattan, New York and the neighborhoods near to the center of Toronto. The data is downloaded and the venues around the neighborhoods is acquired using the Foursquare API. One Hot Encoding is used for converting the categories of the venues into a feature matrix. Then, all venues are grouped by neighborhoods and at the same time the mean is calculated. Hence, the resulting features used are the frequency of occurrence from each category in a neighborhood.

The K-Means clustering algorithm is used for finding similarities between all the neighborhoods listed in the feature matrix. 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.

.