A Consumer Business Comparison of New York and Toronto Based on the Top Venues of Their Key Boroughs

Jie Wu

1 Abstract

The consumer business in New York and Toronto are studied and compared by analyzing the top venues of their key boroughs—Manhattan and Toronto borough. The venue data are collected from Foursquare based on which the neighborhoods are clustered and the distinguishing categories of the venues are identified. The analysis reflects a more developed and well-rounded consumer business in New York suggesting a better opportunity of starting up or expanding a consumer business in Toronto. The insight obtained from this study should be useful to consumer business managers and/or investors exploring business opportunities in one of these two cities.

2 Introduction

New York and Toronto, the financial hubs of United States and Canada, are often compared by people living in North America. There are numerous blogs and articles in the internet favoring one of them from the perspective of employment and salary, cost of living, environment, education etc. In this report the consumer business of the key boroughs of these two cities, Manhattan and Toronto borough are studied and compared, aiming to reveal the characteristics of the consumer business in the these two cities and suggest the prospect of starting up a consumer business in them. To achieve this goal the report answers the following two questions:

- Is there any consumer business that one borough has but the other doesn't? If so managers and/or investors may consider starting up or expanding the business to the other borough.
- For the consumer business that both boroughs have how is it developed and geographically distributed? The level of development and the pattern of distribution may hint at the prospects of expanding that business in the borough.

3 Methodology

The consumer business of the key boroughs of New York and Toronto are depicted using the venue location data retrieved from Foursquare. The key boroughs chosen are Manhattan and Toronto borough respectively, the latter including east Toronto, west Toronto, central Toronto and downtown Toronto. The choice of borough is partly based on the business districts, financial status and population of the borough compared to the other boroughs of the cities and partly for the sake of comparison—each borough contains roughly the same number of neighborhoods, 40 for Manhattan borough and 38 for Toronto borough. To avoid verbose exposition in the following the word borough will be dropped and the two boroughs are simply referred to as Manhattan and Toronto (the latter should not be confused with Toronto as a city).

First a list of top 100 venues within 500 meters of the neighborhood's geographic coordinates (i.e. latitude and longitude) are collected from Foursquare for each neighborhood. The rankings are based on the popularity of the venues among Foursquare users. Giving that Foursquare is among the top three location data providers hosting over 60 million users the bias should be minimal. Moreover the number of venues considered is 100 which should be sufficient to portraying the consumer business in the neighborhood. The radius of the search is chosen according to the typical size of a neighborhood. Therefore the choice of 500 meters should be appropriate to not only include the vast majority of the consumer business within a particular neighborhood but also prevent the analysis from being affected by the consumer business of nearby neighborhoods.

After the top 100 venues are collected for each neighborhood within a distance of 500 meters the categories of the venues are retrieved and consolidated to reveal the richness of the consumer business of New York and Toronto. Then the top 10 categories of venues are identified for each neighborhood based on the mean of the frequency of their occurrence. The purpose of the identification is to characterize the consumer business of the neighborhoods and group neighborhoods with similar characteristics. The neighborhood clustering is determined using the k-means clustering algorithm where the coordinates of the neighborhoods are the mean of the occurrence frequency of the various venue categories. Finally the neighborhoods are and plotted in the maps of the two cities and labeled with their cluster labels. The distributions of the various neighborhood clusters of New York and Toronto are compared to reflect the distribution of the types of consumer business and the business opportunities are assessed.

4 Result

The comparison is carried out in the following four aspects.

4.1 The variety of consumer business

Manhattan hosts 330 categories of venues according to Foursquare's venue categorization, which is way more than the 238 categories of venues hosted by Toronto. This large difference is partly attributed to the greater variety of restaurants, stores and entertainments. For example Manhattan has Szechuan restaurant, Lebanese restaurant, Himalayan restaurant, Australian restaurant and Russian restaurant none of which can be found in Toronto. In addition Manhattan has duty-free shop, used bookstore and herbs & spices store while Toronto does not. The large difference is also attributed to some geographical difference—Manhattan has beach thus offers ferry services and allows marina recreational activities. On the one hand the geographically restriction makes Manhattan the obvious choice of investment for marina related services and entertainment. On the other hand the greater variety of restaurants and services in Manhattan suggests Toronto an unexplored area thus an opportunity for the business that can only be found in Manhattan.

4.2 The distinguishing venue categories of neighborhood clusters

Figure 1 to Figure 5 give the top 10 venue categories for the 5 neighborhood clusters of Manhattan. The distinguishing venue categories are

Cluster 1: park, hotel

Cluster 2: theater

Cluster 3: various restaurants

Cluster 4: boat/ferry, harbor/marina

Cluster 5: bar, various shop/store

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

	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	Manhattan	0	Sandwich Place	Park	Gym	Farmers Market	Dry Cleaner	School	Scenic Lookout	Greek Restaurant	Liquor Store	Train
15	Manhattan	0	Hotel	Food Truck	Steakhouse	Theater	Coffee Shop	Spa	Cocktail Bar	American Restaurant	Clothing Store	Bookstore
26	Manhattan	0	Coffee Shop	Food Truck	Park	American Restaurant	Bookstore	Burger Joint	Deli / Bodega	Tennis Court	Sandwich Place	New American Restaurant
28	Manhattan	0	Coffee Shop	Park	Hotel	Wine Shop	Italian Restaurant	Fountain	Memorial Site	Burger Joint	Food Truck	Food Court
29	Manhattan	0	Coffee Shop	Hotel	Steakhouse	Wine Shop	Gym	Bar	Food Truck	Pizza Place	Italian Restaurant	Juice Bar

Figure 1: Neighborhood cluster #1 of Manhattan (red in the distribution plot Figure 11)

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

	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
13	Manhattan	1	Theater	Gym / Fitness Center	Plaza	Italian Restaurant	Concert Hall	French Restaurant	Café	Opera House	Performing Arts Venue	Park
14	Manhattan	1	Theater	Coffee Shop	Italian Restaurant	American Restaurant	Gym / Fitness Center	Spa	Hotel	Wine Shop	Gym	Dog Run
39	Manhattan	1	Coffee Shop	Italian Restaurant	Café	Gym / Fitness Center	Restaurant	Theater	Hotel	American Restaurant	Gym	Art Gallery

Figure 2: Neighborhood cluster #2 of Manhattan (purple in the distribution plot Figure 11)

The distinctions between neighborhoods of Manhattan are quite clear-cut, implying that different neighborhoods tend to have a focus on the consumer services provided. This is good for consumers as they would have many choices thus can easily find a venue they like without traveling to other neighborhoods. From business management point of view the choice of neighborhood to start up a new business or expand an existing one is also clear, since for example consumers who would

```
# cluster 3
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.s
```

	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
5	Manhattan	2	Deli / Bodega	Seafood Restaurant	Italian Restaurant	Mexican Restaurant	BBQ Joint	Beer Garden	Food & Drink Shop	Park	Burger Joint	Bar
8	Manhattan	2	Italian Restaurant	Exhibit	Art Gallery	Coffee Shop	Juice Bar	Bakery	Gym / Fitness Center	French Restaurant	Boutique	Hotel
9	Manhattan	2	Coffee Shop	Bar	Gym	Italian Restaurant	Pizza Place	Mexican Restaurant	Deli / Bodega	Sushi Restaurant	Japanese Restaurant	Dessert Shop
10	Manhattan	2	Italian Restaurant	Coffee Shop	Sushi Restaurant	Pizza Place	Gym / Fitness Center	Sporting Goods Shop	Burger Joint	Gym	Art Gallery	Cosmetics Shop
12	Manhattan	2	Italian Restaurant	Bar	Coffee Shop	Wine Bar	Bakery	Burger Joint	Indian Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Gym / Fitness Center
16	Manhattan	2	Coffee	Hotel	Italian	Japanese	Sandwich	Spa	Gym	French	Salon /	Bar

Figure 3: Neighborhood cluster #3 of Manhattan (blue in the distribution plot Figure 11)

```
# cluster 4
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.s
```

	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
37	Manhattan	3	Bar	Park	Playground	Boat or Ferry	Basketball Court	Heliport	Baseball Field	Gas Station	Beer Garden	Harbor / Marina

Figure 4: Neighborhood cluster #4 of Manhattan (green in the distribution plot Figure 11)

like to have dinner would most likely go to Cluster 3 neighborhoods and those who would like to have a drink would probably go to Cluster 5 neighborhoods.

Figure 6 to Figure 10 give the top 10 venue categories for the 5 neighborhood clusters of Toronto. The distinguishing venue categories are

Cluster 1: park, playground, Falafel and Ethiopian restaurant

Cluster 2: park, brewery, Yoga studio

Cluster 3: coffee shop, various restaurants

Cluster 4: trail, jewelry store, bus line

Cluster 5: music venue, home service, garden, Falafel and Ethiopian restaurant

Compared to those of Manhattan the distinctions between neighborhoods of Toronto are not so clear-cut. The restaurants especially Falafel and Ethiopian restaurant are hosted by neighborhoods from almost all clusters, and somewhat surprisingly shops and stores do not play an important role in distinguishing neighborhoods which is a bit counterintuitive. This implies that consumers may not go to specific neighborhoods for shopping, dining or drinking. It hence offers

```
# cluster 5
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.s
```

	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	Manhattan	4	Discount Store	Coffee Shop	Yoga Studio	Pizza Place	Supplement Shop	Steakhouse	Shopping Mall	Shoe Store	Seafood Restaurant	Sandwich Place
1	Manhattan	4	Chinese Restaurant	Bubble Tea Shop	Cocktail Bar	American Restaurant	Dim Sum Restaurant	Vietnamese Restaurant	Noodle House	Salon / Barbershop	Bakery	Hotpot Restaurant
2	Manhattan	4	Café	Bakery	Mobile Phone Shop	Sandwich Place	Supermarket	Gym	Chinese Restaurant	Shoe Store	Mexican Restaurant	Tapas Restaurant
3	Manhattan	4	Café	Mexican Restaurant	Pizza Place	Lounge	Wine Bar	Restaurant	Bakery	Frozen Yogurt Shop	Deli / Bodega	Chinese Restaurant
4	Manhattan	4	Mexican Restaurant	Deli / Bodega	Coffee Shop	Café	Pizza Place	Liquor Store	Cocktail Bar	Sandwich Place	School	Chinese Restaurant

Figure 5: Neighborhood cluster #5 of Manhattan (orange in the distribution plot Figure 11)

```
# 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
8	Central Toronto	0	Playground	Park	Restaurant	Yoga Studio	Diner	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store
10	Downtown Toronto	0	Park	Playground	Trail	Yoga Studio	Diner	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store

Figure 6: Neighborhood cluster #1 of Toronto (red in the distribution plot Figure 12)

little guidance to the location choice of new business venture for a particular type of consumer business.

4.3 The distribution of neighborhood clusters

As can be seen from Figure 11 and Figure 12 below the neighborhoods belonging to different clusters are mixed together in Manhattan except for the cluster featured by theater, which aggregate in the west border facing Newark. The cluster rich in restaurant spread over the island. This is convenient for residents and visitors of Manhattan. But it may not be an encouraging news for food vendors and restaurant managers who are considering starting up a new business and expanding an existing business. Because there is little void to fill as suggested by the wide spread of the cluster. Similar distribution is observed for the cluster featured by bars, shops and stores.

In contrast Figure 6 to Figure 9 which depict the top 10 venue categories for the 4 neighborhood clusters of Toronto show a much greater extent of aggregation. Moreover the minority clusters i.e. cluster 1 featured by park and playground, cluster 4 distinguished by trail and jewelry store and cluster 5 branded by music venue and home service appear to be in the outlier region. This is not convenient for residents and visitors of Toronto since they might have to travel for a long distance

```
# cluster 2
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, 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
2	East Toronto	1	Sandwich Place	Park	Brewery	Steakhouse	Italian Restaurant	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant	Liquor Store	Pet Store
4	Central Toronto	1	Bus Line	Park	Lake	Dim Sum Restaurant	Swim School	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space
37	East Toronto	1	Light Rail Station	Yoga Studio	Garden	Pizza Place	Recording Studio	Restaurant	Burrito Place	Brewery	Skate Park	Farmers Market

Figure 7: Neighborhood cluster #2 of Toronto (purple in the distribution plot Figure 12)

```
# cluster 3
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, 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	10t Mos Commo Venu
0	East Toronto	2	Neighborhood	Coffee Shop	Pub	Dance Studio	Discount Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Ethiopia Restaurar
1	East Toronto	2	Greek Restaurant	Coffee Shop	Ice Cream Shop	Bookstore	Italian Restaurant	Grocery Store	Brewery	Bubble Tea Shop	Restaurant	Caribbea Restaurar
3	East Toronto	2	Café	Coffee Shop	Yoga Studio	American Restaurant	Bakery	Italian Restaurant	Convenience Store	Coworking Space	Juice Bar	Ne America Restaurar
5	Central Toronto	2	Hotel	Gym	Park	Breakfast Spot	Sandwich Place	Restaurant	Food & Drink Shop	Burger Joint	Yoga Studio	Dumplin Restaurar
6	Central Toronto	2	Coffee Shop	Clothing Store	Sporting Goods Shop	Yoga Studio	Bagel Shop	Cosmetics Shop	Gym / Fitness Center	Chinese Restaurant	Dessert Shop	Dine
7	Central Toronto	2	Dessert Shop	Sandwich Place	Pizza Place	Coffee Shop	Restaurant	Café	Pharmacy	Seafood Restaurant	Sushi Restaurant	Italia Restaurar

Figure 8: Neighborhood cluster #3 of Toronto (blue in the distribution plot Figure 12)

to visit the venue they like. Nonetheless this is a good news to investors and business managers because there seems to be a lot of void to fill, for example the downtown Toronto neighborhoods seems to lack jewelry stores and cocktail bars.

5 Discussion

The study does not take into the population density into consideration, nor does it account for the fact that Manhattan has much less residential area than Toronto. The demographics of people working and/or living in the two cities are certainly factors that can never be ignored when assessing consumer business. Further study may include the population and residents' median income of the neighborhoods as attributes in addition to the venue categories for clustering.

For the attributes that have already been considered i.e. the venue categories there is also room for improvement. Since the focus is consumer business some venue categories offering public services such as park and public transit may be excluded. This could potentially result in easier and

```
# cluster 4
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, 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
23	Central Toronto	3	Trail	Jewelry Store	Sushi Restaurant	Bus Line	Yoga Studio	Dog Run	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space

Figure 9: Neighborhood cluster #4 of Toronto (green in the distribution plot Figure 12)

```
# cluster 5
toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, 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
22	Central Toronto	4	Music Venue	Home Service	Garden	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store

Figure 10: Neighborhood cluster #5 of Toronto (orange in the distribution plot Figure 12)

more meaningful identification of the distinguishing categories for different neighborhoods, although business enjoying better environment and more convenient transportation facilities would attract more patrons thus are more popular hence should be taken into account when choosing the location for new business venture.

Moreover for both the existing and potential attributes of neighborhoods it is highly desirable to have a higher level of categorification that Foursquare fails to do. For example mini golf, boxing gym, volleyball court and Tennis stadium can be categorized as sports facilities. Too high a specialization for the venue may render the data less useful in comparing neighborhoods and cities.

Regarding the clustering algorithm the method used in the study is K-means clustering. Given the nature of the attributes (all continuous and non-negative float numbers) this seems to be an appropriate choice. Although there apparently seems no ground to consider hierarchical clustering DBSCAN is an alternative to try. However experiment shows that DBSCAN results in only one cluster containing the majority of the neighborhoods and many outliers. One factor that restricts the effectiveness of DBSCAN might be the number of neighborhoods (40 for Manhattan and 38 for Toronto). It is worth trying DBSCAN clustering on the hundreds of neighborhoods of the entire city of New York and Toronto in further study, albeit the interpretation of the characteristics of consumer business in the two cities might become more difficult due to the substantial increase of data and variety.

One disturbing observation must be mentioned, which is that the code is run at different time of the day yielding different clustering result. It seems like venues are ranked by Foursquare based on the patronage over a very short period of time or even instantly. If this is the case then the data retrieved from Foursquare, although useful to consumers looking for venues to visit, have little value from the business assessment perspective. Ideally the rankings should be based on the patronage accumulated over the trailing 12 months (TTM).

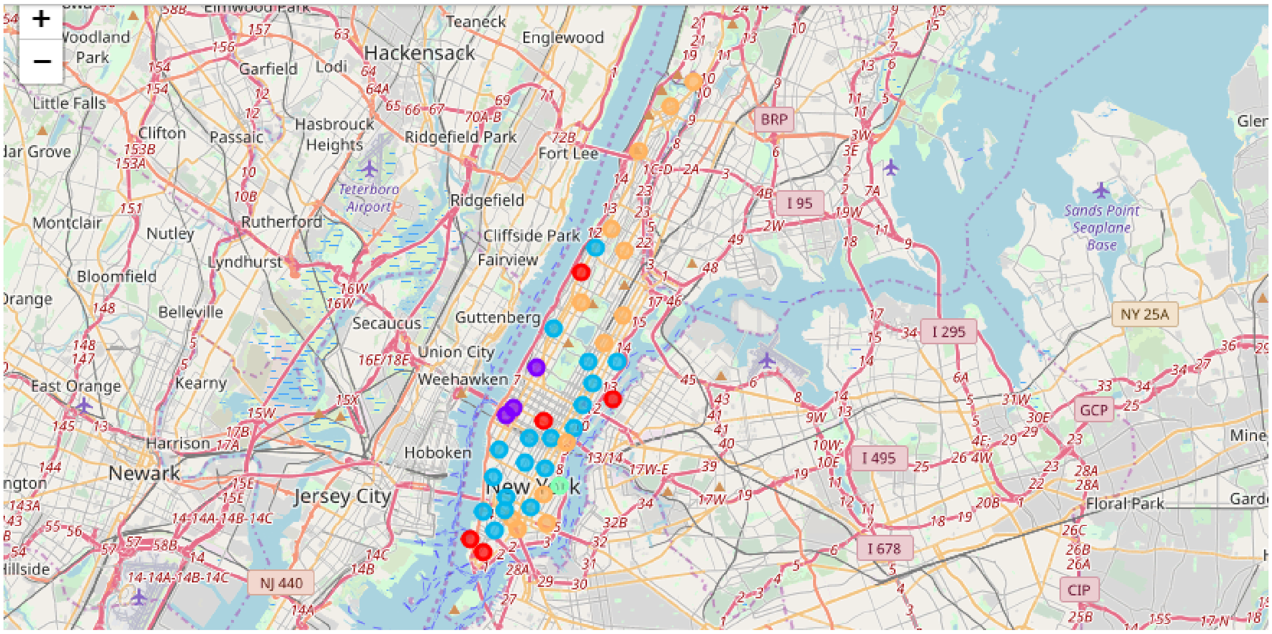

Figure 11: Neighborhood cluster distribution of Manhattan (cluster color code is purple, blue, green, orange and red for cluster 1 to 5)

6 Conclusion

In this study the consumer business in New York and Toronto are studied and compared by analyzing the top venues of their key boroughs—Manhattan and Toronto borough. The venue data for the neighborhoods collected from Foursquare are analyzed and compared from three perspectives—the variety of consumer business, the distinguishing venue categories of neighborhood clusters and the distribution of neighborhood clusters. The analysis reveals more developed and well-rounded consumer business in New York suggesting a better opportunity of starting up or expanding a consumer business in Toronto.

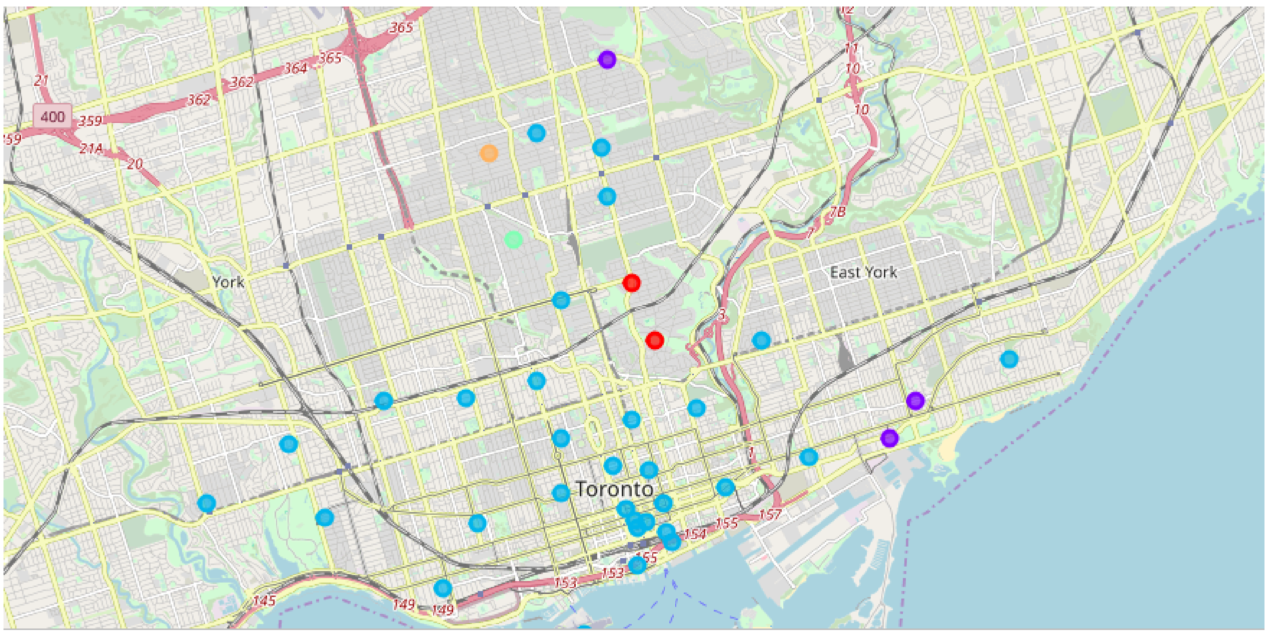

Figure 12: Neighborhood cluster distribution of Toronto (cluster color code is purple, blue, green, orange and red for cluster 1 to 5)