Comparing the Venues of Downtown Toronto and Manhattan

1. Introduction

The goal of this project is to study, analyze, cluster, and compare the neighborhoods of Downtown Toronto in Ontario, Canada and Manhattan in New York City, USA for various business located in each city. We will analyze the businesses are common in both cities, what kinds of businesses are more common in one of the two cities than the other city, and what kinds of businesses are not common in both cities.

This project will enable us to get a better understanding of similarities and differences between the two cities what types of businesses are more likely to thrive in both cities, what are the neighborhoods that are suitable for each type of business, and what types of businesses are not very desirable in each city. This allows business people to take better and more effective decisions regarding where to open their businesses.

Downtown Toronto is the main central business district of Toronto, Ontario, Canada. Located entirely within the district of Old Toronto, it is approximately 17 square kilometers in area, bounded by Bloor Street to the northeast and Dupont Street to the northwest, Lake Ontario to the south, the Don Valley to the east, and Bathurst Street to the west. It is also the location of the municipal government of Toronto and the Government of Ontario.

Figure 1. An aerial view of Downtown Toronto.

The area is made up of Canada's largest concentration of skyscrapers and businesses that form Toronto's skyline. Downtown Toronto has the third most skyscrapers in North America exceeding 200 metres (656 ft) in height, behind New York City and Chicago.

Manhattan is the most densely populated of the five boroughs of New York City. Manhattan serves as the city's economic and administrative center, cultural identifier and historical birthplace. Manhattan Island is divided into three informally bounded components, each aligned with the borough's long axis: Lower, Midtown, and Upper Manhattan.

Figure 2. An aerial view of Manhattan.

Anchored by Wall Street, Manhattan is home to the world's two largest stock exchanges by total market capitalization: the New York Stock Exchange and NASDAQ. Many multinational media conglomerates are based in Manhattan, and the borough has been the setting for numerous books, films, and television shows.

2. Data Acquisition and Preparation

Based on criteria listed above the following data was utilized:

- Neighborhood data for Downtown Toronto. (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Neighborhood data for Manhattan. (https://cocl.us/new_york_dataset)
- Geocoding API for longitude and latitudes. (https://developers.google.com/maps/documentation/geocoding/start)
- Foursquare API for the number of venues within the certain radius of each borough. (https://foursquare.com/developers)

2.1 Neighborhood Data (Downtown Toronto)

The neighborhood data for Downtown Toronto is derived from the source listed above. The read_html() function was used to scrap the data from the web page and format it into a table as seen below.

	PostalCode	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

Figure 3. Postal codes with neighborhood and borough names.

2.2 Geocoding data (Downtown Toronto)

Using geolocater for Toronto we obtained the latitudes and longitude of each postal code. This data was inserted into a dataframe and merged with the data scraped earlier using the dataframe.join() function on Postal Code.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Figure 4. Merged data frame with latitude and longitude.

Once this was done, we focused on Downtown Toronto and obtained a map of the neighborhoods using the latitude and longitude data.

Figure 5. Map of Neighborhoods in Downtown Toronto.

2.3 Neighborhood Data (Manhattan)

The neighborhood data for Manhattan is derived from the source listed above. The JSON file that specifies the name of each neighborhood, its coordinates—latitude and longitude and its boroughs is read into a data frame.


```
{'type': 'Feature',
 'id': 'nyu_2451_34572.1',
 'geometry': {'type': 'Point',
  'coordinates': [-73.84720052054902, 40.89470517661]},
 'geometry_name': 'geom',
 'properties': {'name': 'Wakefield',
  'stacked': 1,
  'annoline1': 'Wakefield',
  'annoline2': None,
  'annoline3': None,
  'annoangle': 0.0,
  'borough': 'Bronx',
  'bbox': [-73.84720052054902,
   40.89470517661,
   -73.84720052054902,
   40.89470517661]}}
```


Figure 6. Extracted data from the JSON file.

This data is used to obtain the data frame below.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 7. Dataframe obtained from JSON file.

2.4 Geocoding data (Manhattan)

Using geolocater for Manhattan we obtained a map of the neighborhoods using the latitude and longitude data.

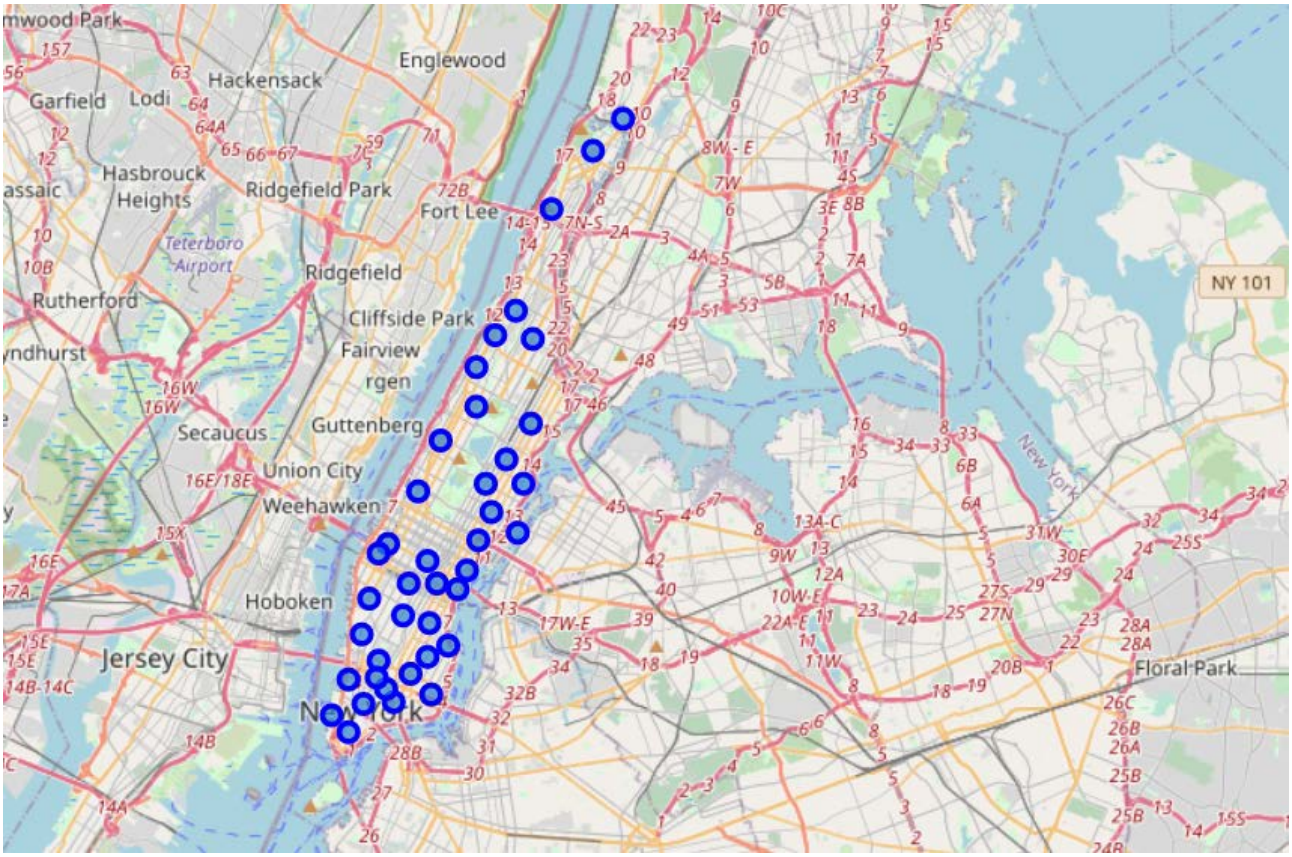

Figure 8. Map of Neighborhoods in Manhattan.

2.5 Venues Data (Downtown Toronto & Manhattan)

The foursquare API was utilized to obtain the venues data for each city and their categories.

	name	categories	lat	lng
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Cooper Koo Family YMCA	Distribution Center	43.653249	-79.358008
3	Body Blitz Spa East	Spa	43.654735	-79.359874
4	Impact Kitchen	Restaurant	43.656369	-79.356980

Figure 9. Sample venues data for Downtown Toronto.

	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	Astral Fitness & Wellness Center	Gym	40.876705	-73.906372

Figure 10. Sample venues data for Manhattan.

3. Exploratory Data Analysis

3.1 Most Common Venue Categories

We can analyze the most common businesses in each city and the frequency of each category. We can see below that the most common venue in Downtown Toronto is 'Coffee Shop' with 'Café' being the second most common.

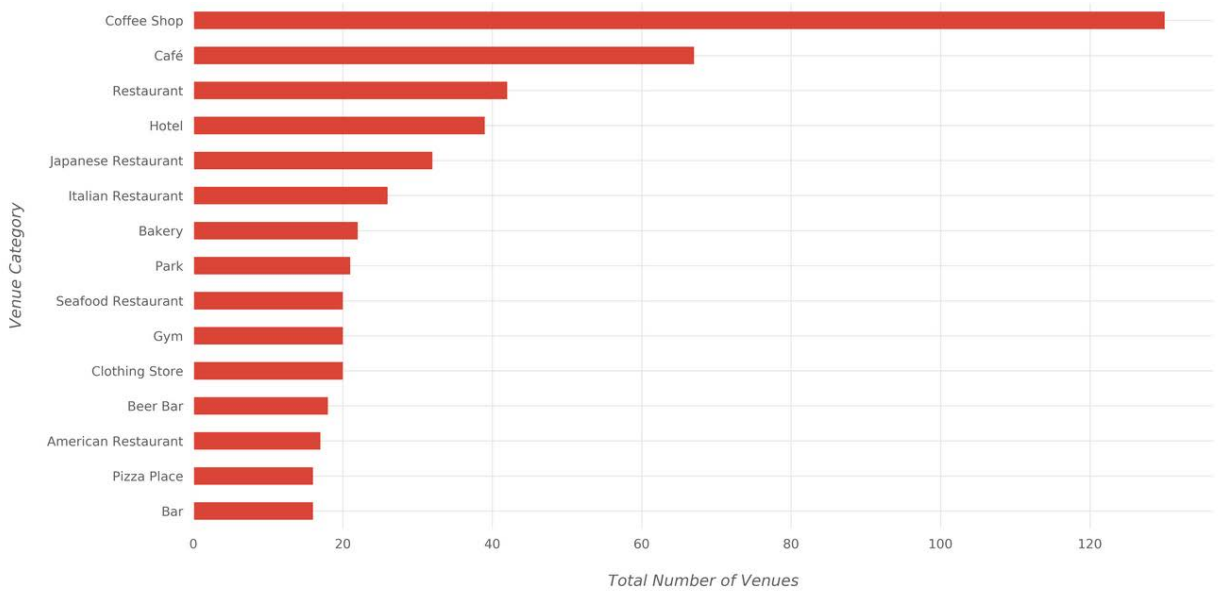

Figure 10. Most common venue categories in Downtown Toronto.

We can see below that the most common venue in Manhattan is also 'Coffee Shop' with 'Italian Restaurant' close in second.

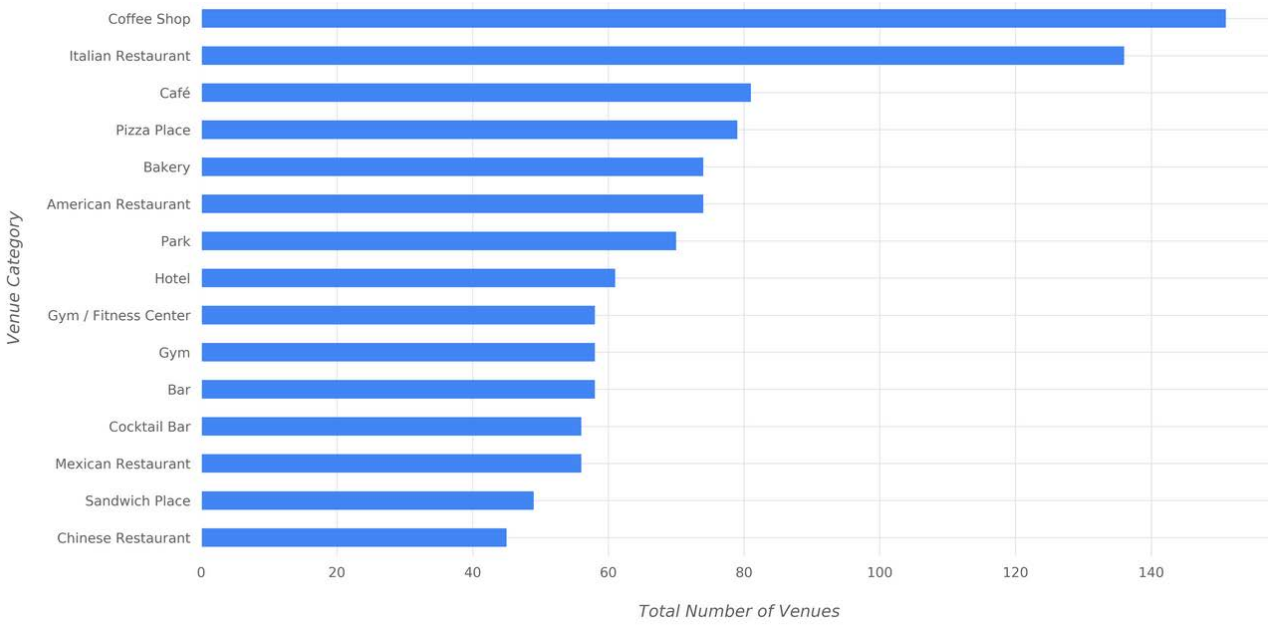

Figure 11. Most common venue categories in Manhattan.

3.2 Most Number of Venue Categories in each Neighborhood

We can also analyze the venue categories that exist in most neighborhoods. This different than what we obtained above, because it tells how saturated any one business is in a particular neighborhood compared to any other business. This will help people make decisions before opening any particular business in any particular neighborhood with high competition. This will help business owners to think of a more creative business to start amidst the saturation.

We can see below that the most common business in every neighborhood in Downtown Toronto is a coffee shop.

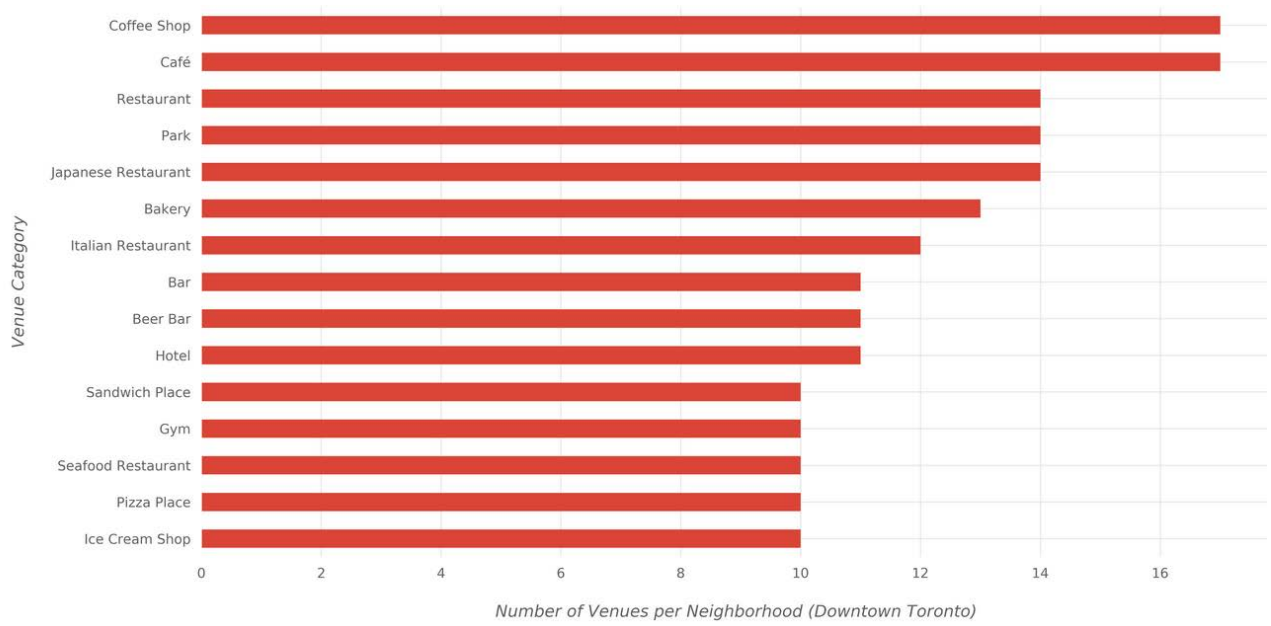

Figure 12. Common businesses in each neighborhood in Toronto.

We can see below that the most common business in every neighborhood in Manhattan is also a coffee shop.

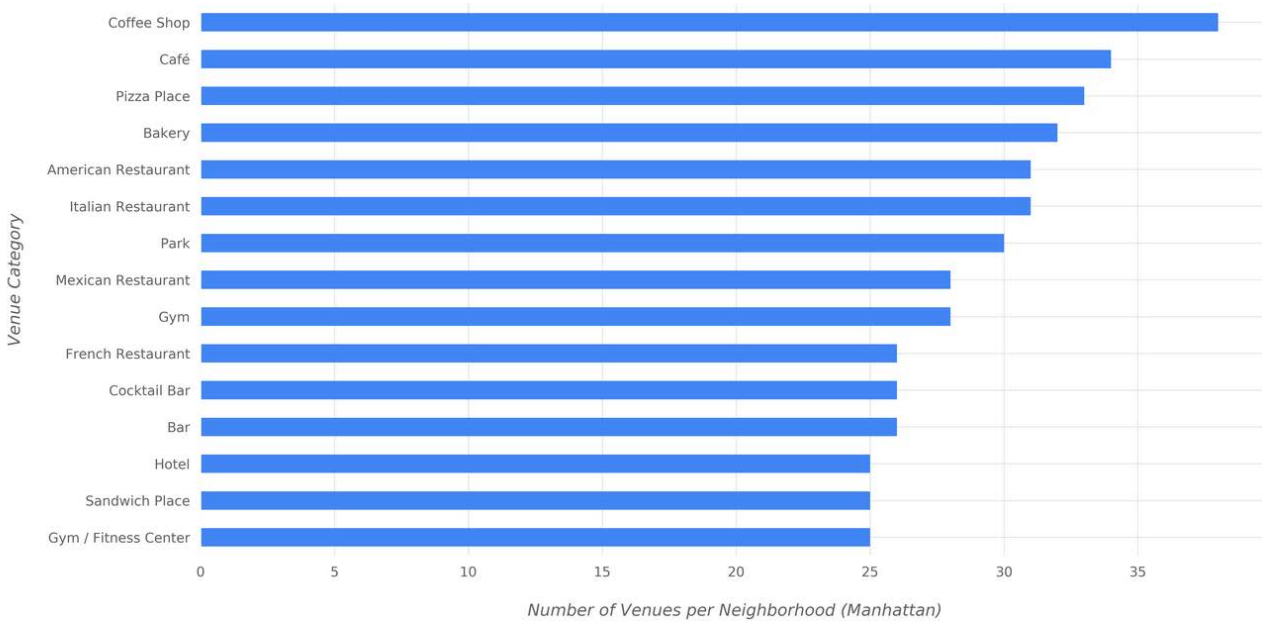

Figure 13. Common businesses in each neighborhood in Manhattan.

3.3 Unique Venues in each city

Here we explore the differences in each city with businesses that are low in frequency and may stand out.

We can see below the number of business unique to certain neighborhoods in Downtown Toronto.

Unique Venue Category	Number of Venues
Skating Rink	1
Optical Shop	1
Taiwanese Restaurant	1
College Arts Building	1
Health & Beauty Service	1
Candy Store	1
College Auditorium	1
Chocolate Shop	1
Baby Store	1
Doner Restaurant	1
Snack Place	1
Taco Place	1
German Restaurant	1
Flower Shop	1
Butcher	1

Figure 14. Unique businesses in Downtown Toronto.

We can see below the number of business unique to certain neighborhoods in Manhattan.

Unique Venue Category	Number of Venues
Perfume Shop	1
Drugstore	1
Used Bookstore	1
Gaming Cafe	1
Kitchen Supply Store	1
Rest Area	1
Adult Boutique	1
Czech Restaurant	1
Duty-free Shop	1
Eye Doctor	1
Photography Studio	1
Whisky Bar	1
Auditorium	1
Brazilian Restaurant	1
College Cafeteria	1

Figure 15. Unique businesses in Manhattan.

3.3 Unique Venues between each city

Here we explore the similarities between each city and we can find the analysis of the most unique venues between each city below.

List of Common Venues	Number of Venues
Video Store	1
Whisky Bar	1
Persian Restaurant	1
Photography Studio	1
Eye Doctor	1
Physical Therapist	1
Camera Store	1
Volleyball Court	1
Athletics & Sports	1
Doner Restaurant	1
Stables	1
Veterinarian	1
Empanada Restaurant	1
Check Cashing Service	1
Coworking Space	1

Figure 16. Unique venues between each city.

4. Cluster Analysis

To identify groups (clusters) with similar characteristics, the unsupervised learning method to our data, namely K-Means algorithm, was applied to our data. We used clustering to find similar venues in Downtown Toronto and Manhattan. Clustering algorithm works with numerical features. Hence, one-hot encoding was applied on the "Venue Category" feature and the result of the encoding will be used for clustering. One-hot encoding will be applied on the data for each city.

Based on clustering results, two maps are created one showing the venues in Downtown Toronto and other for Manhattan.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0	Coffee Shop	Park	Bakery	Pub	Café
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	4	Coffee Shop	College Cafeteria	Diner	Smoothie Shop	Beer Bar
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Clothing Store	Coffee Shop	Cosmetics Shop	Bubble Tea Shop	Café
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Café	Coffee Shop	Restaurant	Clothing Store	Cocktail Bar
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	0	Coffee Shop	Cocktail Bar	Café	Cheese Shop	Seafood Restaurant

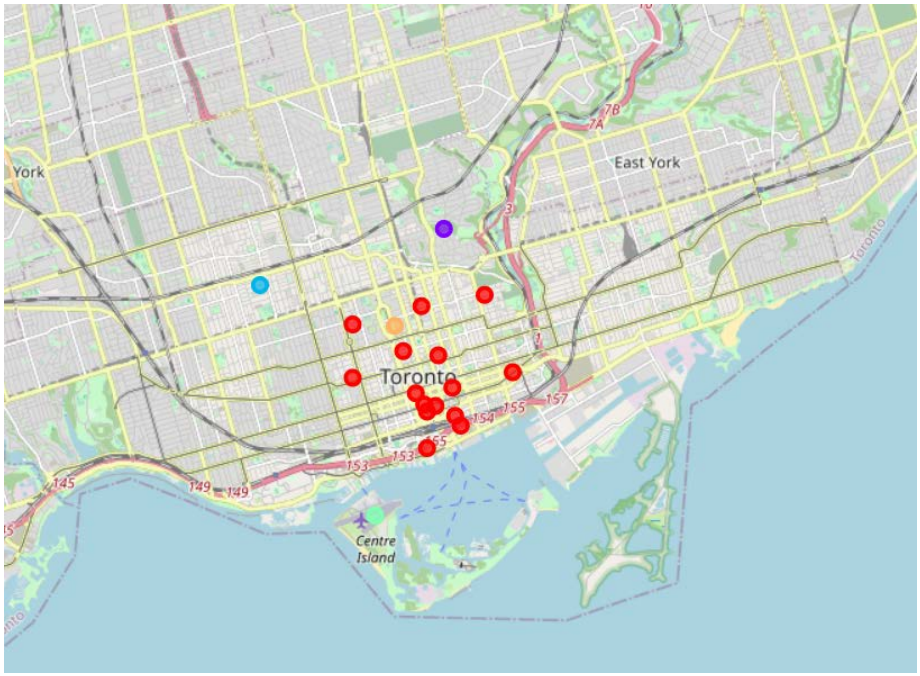

Figure 17. Clustering applied to Downtown Toronto.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	3	Gym	Coffee Shop	Seafood Restaurant	Tennis Stadium	Big Box Store
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	Bakery	Cocktail Bar	Salon / Barbershop	Vietnamese Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Chinese Restaurant	Mobile Phone Shop	Coffee Shop
3	Manhattan	Inwood	40.867684	-73.921210	4	Mexican Restaurant	Lounge	Café	Restaurant	Park
4	Manhattan	Hamilton Heights	40.823604	-73.949688	4	Pizza Place	Café	Coffee Shop	Deli / Bodega	Mexican Restaurant

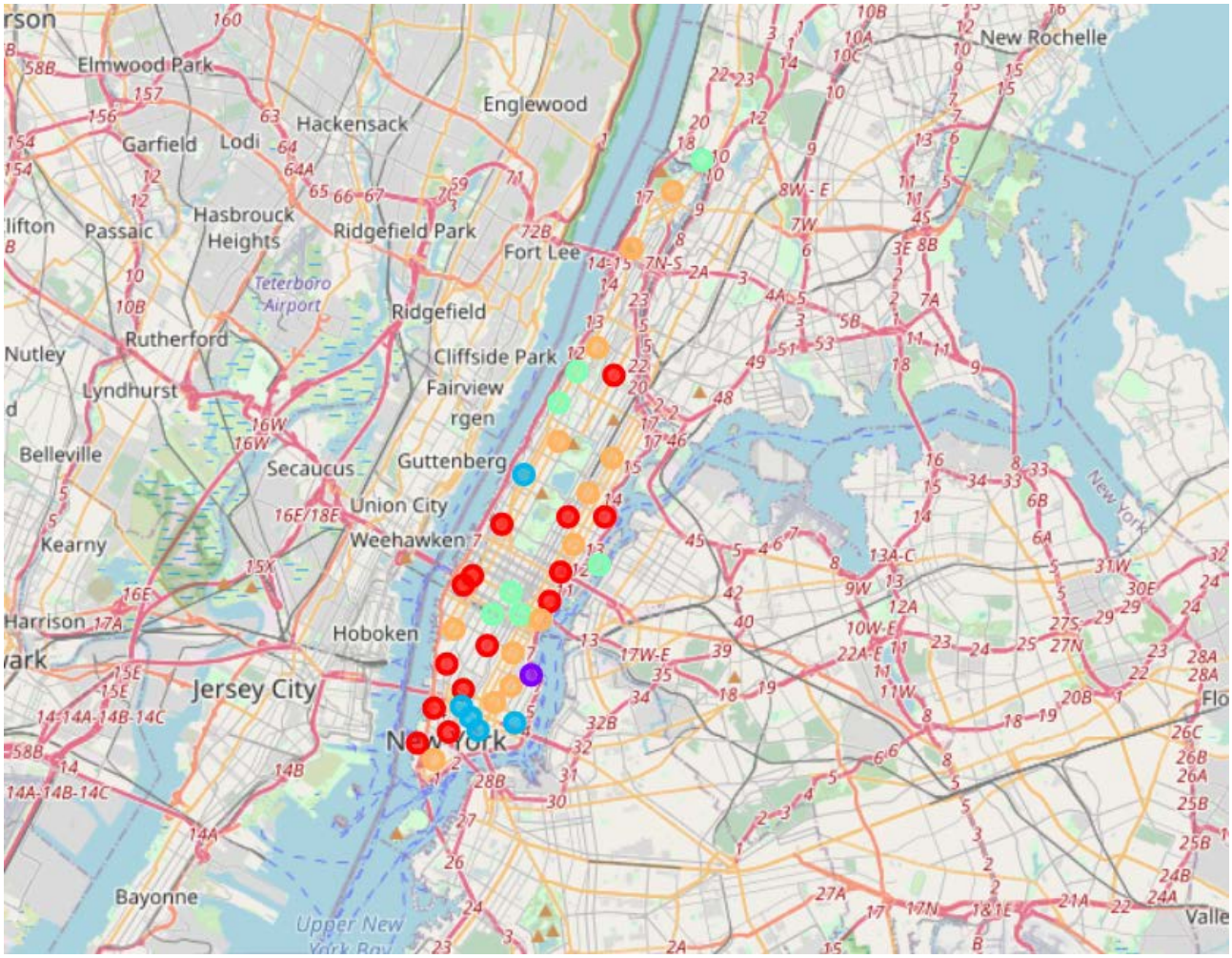

Figure 18. Clustering applied to Manhattan.

5. Results and discussion

The clustering algorithm grouped neighborhoods of NYC and Toronto in 5 clusters based on the similarity between their venues.

After defining the venue categories, one can perform deeper analysis to find the best locations to open a business.

What could be done better?

Foursquare doesn't represent the full picture, since many venues are not on the list. Boroughs have too complex geometry, thus defining the closest venues within the certain radius brings additional error to our analysis.

6. Conclusion

To conclude, the basic data analysis was performed to compare venues for the cities of Downtown Toronto and Manhattan. During the analysis, several important statistical features of the boroughs were explored and visualized. New York City and Toronto were clustered into multiple groups based on the categories (types) of the venues in these neighborhoods. The results showed that there are venue categories that are more common in some cluster than the others; the most common venue categories differ from one cluster to the other. If a deeper analysis—taking more aspects into account—is performed, it might result in discovering different style in each cluster based on the most common categories in the cluster.