# Capstone Project – The Battle of Neighborhoods

# Report – Anthony Giorgio

## 1. Introduction

Toronto is the provincial capital of Ontario. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario, while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

New York City, often called simply New York and abbreviated as NYC, is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), New York City is also the most densely populated major city in the United States. Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

### The problem

Now let me explain the context of this Capstone project through a scenario. Suppose a friend who live on the west side of the city of Toronto in Canada, receive a job offer from a great company in New York, Manhattan borough. He has to move to New York City. He love his neighborhood in Toronto beacause of its variety of venues for food, parks, schools and entertainment places. Consequently he want to move in a similar zone in Manhattan borough. The aim of this project is to study and analyze the neighborhoods of Toronto city and New York city and group them into similar clusters. Finally I will use those information to find the most similar neighborhood of the two borough of the two cities.

## 2. The Data

For this project we need the following data :

1. New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
   - Data source : https://cocl.us/new_york_dataset
2. Toronto data that contains list Boroughs, Neighborhoods along with their latitude and longitude
   - Data source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M , http://cocl.us/Geospatial_data

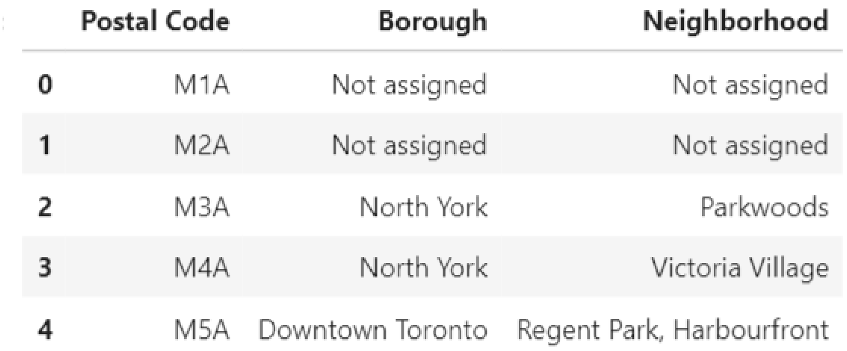
## 3. Methodology

1. Get the data of both cities into pandas dataframe and process them (I will consider only the
2. Use the Foursquare API to find all venues for each neighborhood and analyze them
3. Use cluster algorithm to find similar neighborhood
4. Compare clusters between the two cities

# 4. Analysis

## 4.1 Toronto data

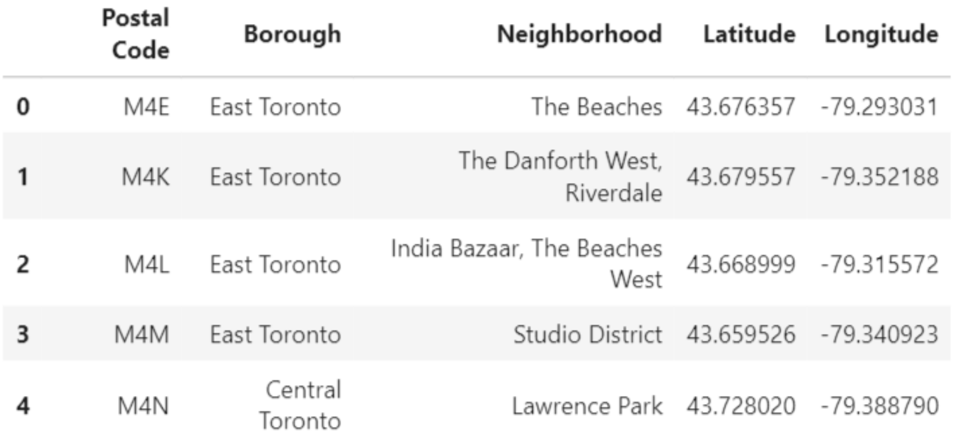
To start with our analysis, let’s firtsly scrape the Wikipedia page and gathering data into the below Pandas dataframe:



| | Postal Code | 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 | Regent Park, Harbourfront |

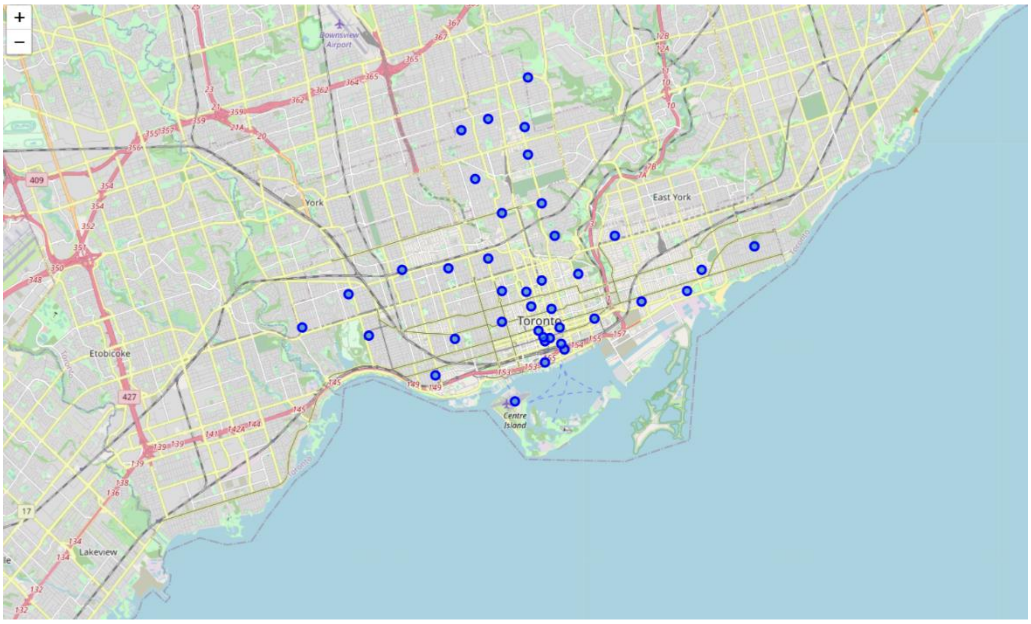
The dataframe consist of three columns: PostalCode, Borough, and Neighborhood. Let's drop cells with a borough that is "Not assigned". We notice that more than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma. Moreover if a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

We also fetched the coordinate data for all the neighborhoods in Toronto using the csv file and put it into a dataframe. Next, we combine both the dataframes i.e. adding the coordinate data to the original dataframe. Finally, we filter only boroughs that contain the word Toronto. The resulti is the following:



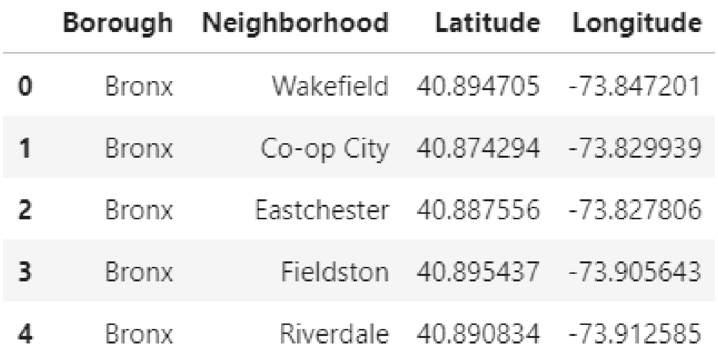
| | Postal Code | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 |
| 1 | M4K | East Toronto | The Danforth West, Riverdale | 43.679557 | -79.352188 |
| 2 | M4L | East Toronto | India Bazaar, The Beaches West | 43.668999 | -79.315572 |
| 3 | M4M | East Toronto | Studio District | 43.659526 | -79.340923 |
| 4 | M4N | Central Toronto | Lawrence Park | 43.728020 | -79.388790 |

Now, we use geopy library to get the latitude and longitude values of Toronto. We then use the python folium library to visualize geographic details of Toronto and its boroughs.



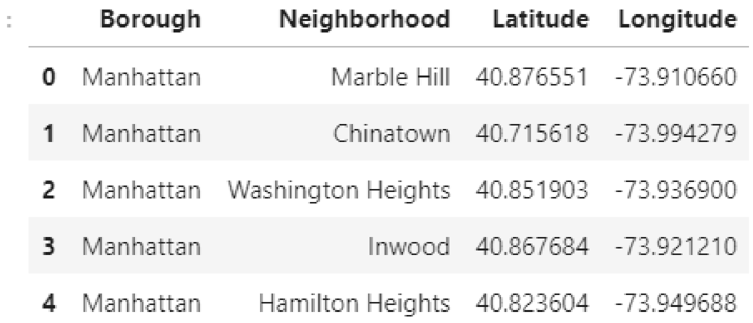
## 4.2 New York data

Similarly to Toronto data, after loaded the data needed, we create a Pandas dataframe as follow:



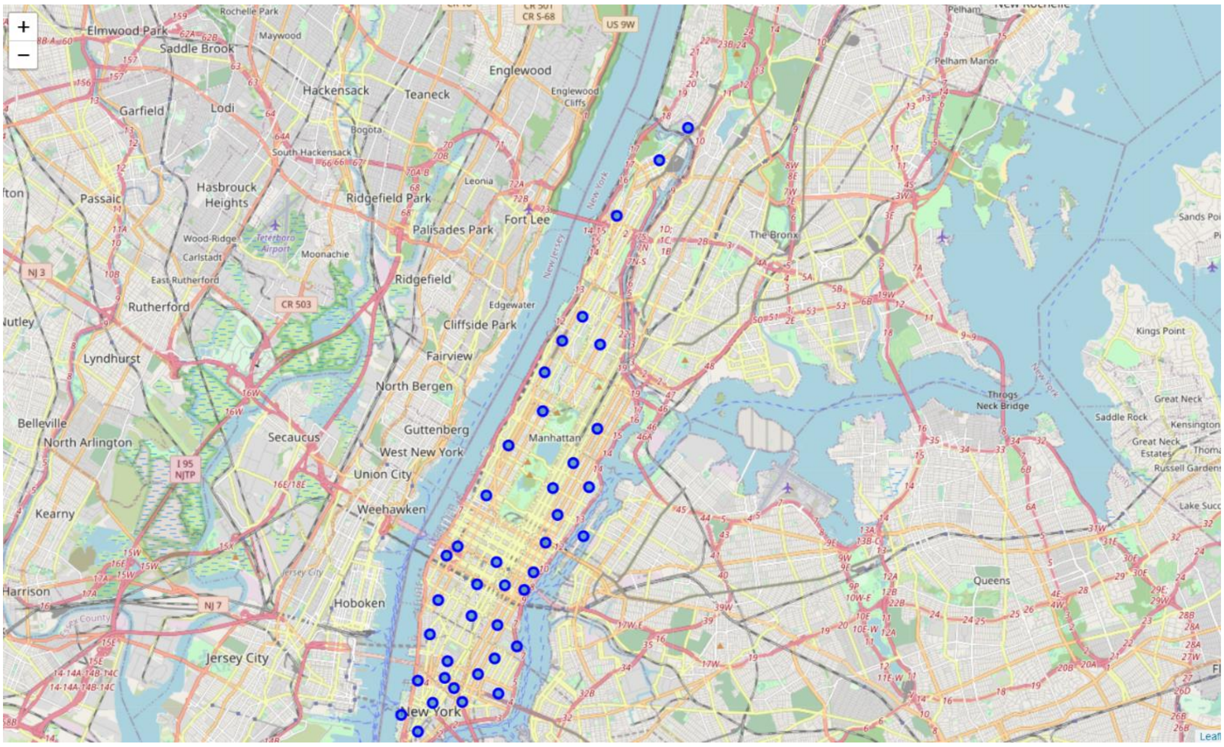
| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| **0** | Bronx | Wakefield | 40.894705 | -73.847201 |
| **1** | Bronx | Co-op City | 40.874294 | -73.829939 |
| **2** | Bronx | Eastchester | 40.887556 | -73.827806 |
| **3** | Bronx | Fieldston | 40.895437 | -73.905643 |
| **4**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

Now, for our purpose, let's slice the original dataframe and create a new dataframe of the Manhattan data.



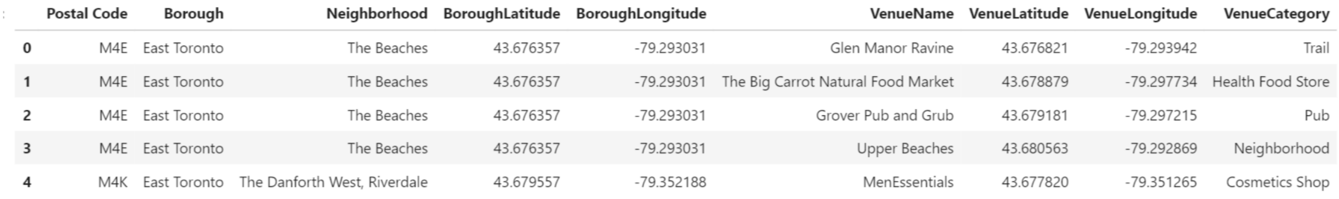
| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Manhattan | Marble Hill | 40.876551 | -73.910660 |
| 1 | Manhattan | Chinatown | 40.715618 | -73.994279 |
| 2 | Manhattan | Washington Heights | 40.851903 | -73.936900 |
| 3 | Manhattan | Inwood | 40.867684 | -73.921210 |
| 4 | Manhattan | Hamilton Heights | 40.823604 | -73.949688 |

Let's visualize Manhattan the neighborhoods:



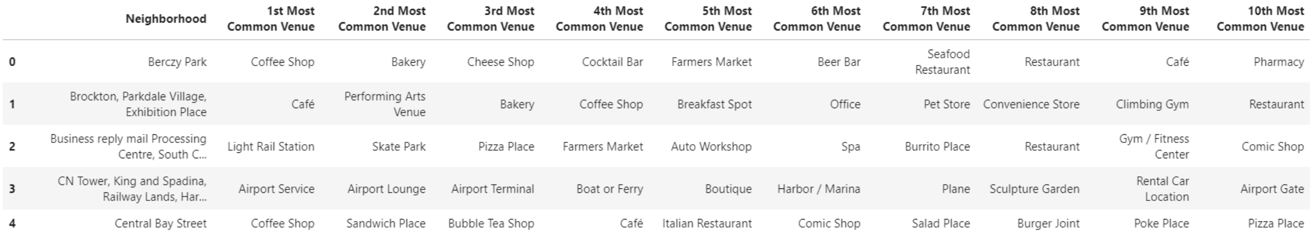
## 4.3. Toronto areas analysis

We utilize Foursquare API to explore the neighborhoods and segment them. let's get the top 50 venues that are in the selected Toronto borough s within a radius of 500 meters.



| | Postal Code | Borough | Neighborhood | BoroughLatitude | BoroughLongitude | VenueName | VenueLatitude | VenueLongitude | VenueCategory |
|---|---|---|---|---|---|---|---|---|---|
| 0 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | Glen Manor Ravine | 43.676821 | -79.293942 | Trail |
| 1 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | The Big Carrot Natural Food Market | 43.678879 | -79.297734 | Health Food Store |
| 2 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | Grover Pub and Grub | 43.679181 | -79.297215 | Pub |
| 3 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | Upper Beaches | 43.680563 | -79.292869 | Neighborhood |
| 4 | M4K | East Toronto | The Danforth West, Riverdale | 43.679557 | -79.352188 | MenEssentials | 43.677820 | -79.351265 | Cosmetics Shop |

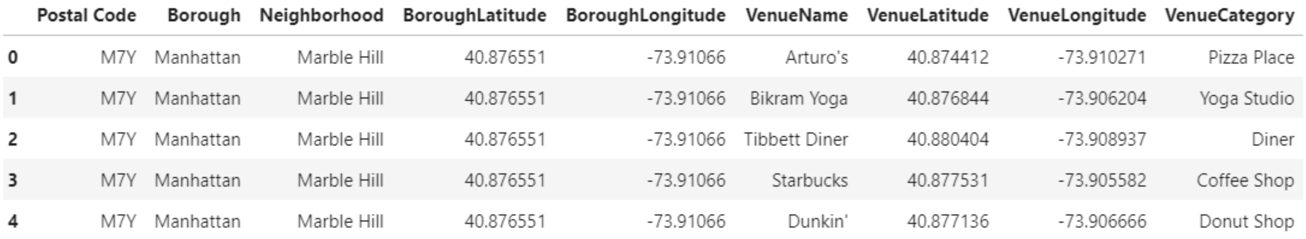
We use One Hot Encoding, use the neighborhood to group data, and find out the top ten venues present in each neighborhood.



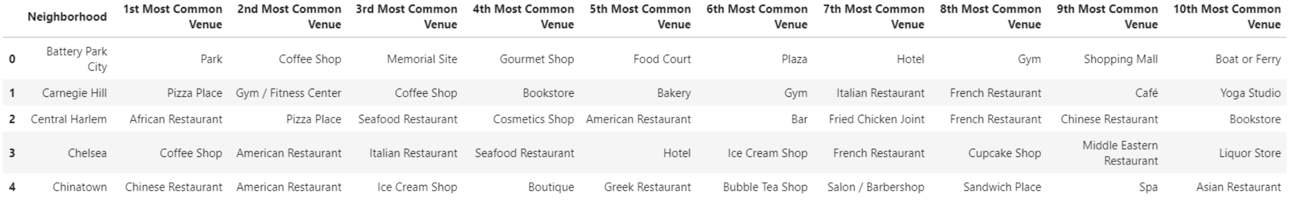
| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Berczy Park | Coffee Shop | Bakery | Cheese Shop | Cocktail Bar | Farmers Market | Beer Bar | Seafood Restaurant | Restaurant | Café | Pharmacy |
| 1 | Brockton, Parkdale Village, Exhibition Place | Café | Performing Arts Venue | Bakery | Coffee Shop | Breakfast Spot | Office | Pet Store | Convenience Store | Climbing Gym | Restaurant |
| 2 | Business reply mail Processing Centre, South C... | Light Rail Station | Skate Park | Pizza Place | Farmers Market | Auto Workshop | Spa | Burrito Place | Restaurant | Gym / Fitness Center | Comic Shop |
| 3 | CN Tower, King and Spadina, Railway Lands, Har... | Airport Service | Airport Lounge | Airport Terminal | Boat or Ferry | Boutique | Harbor / Marina | Plane | Sculpture Garden | Rental Car Location | Airport Gate |
| 4 | Central Bay Street | Coffee Shop | Sandwich Place | Bubble Tea Shop | Café | Italian Restaurant | Comic Shop | Salad Place | Burger Joint | Poke Place | Pizza Place |

## 4.4. Manhattan areas analysis

We do the same for Manhattan's neighborhoods, getting the two dataframe:



| | Postal Code | Borough | Neighborhood | BoroughLatitude | BoroughLongitude | VenueName | VenueLatitude | VenueLongitude | VenueCategory |
|---|---|---|---|---|---|---|---|---|---|
| 0 | M7Y | Manhattan | Marble Hill | 40.876551 | -73.91066 | Arturo's | 40.874412 | -73.910271 | Pizza Place |
| 1 | M7Y | Manhattan | Marble Hill | 40.876551 | -73.91066 | Bikram Yoga | 40.876844 | -73.906204 | Yoga Studio |
| 2 | M7Y | Manhattan | Marble Hill | 40.876551 | -73.91066 | Tibbett Diner | 40.880404 | -73.908937 | Diner |
| 3 | M7Y | Manhattan | Marble Hill | 40.876551 | -73.91066 | Starbucks | 40.877531 | -73.905582 | Coffee Shop |
| 4 | M7Y | Manhattan | Marble Hill | 40.876551 | -73.91066 | Dunkin' | 40.877136 | -73.906666 | Donut Shop |

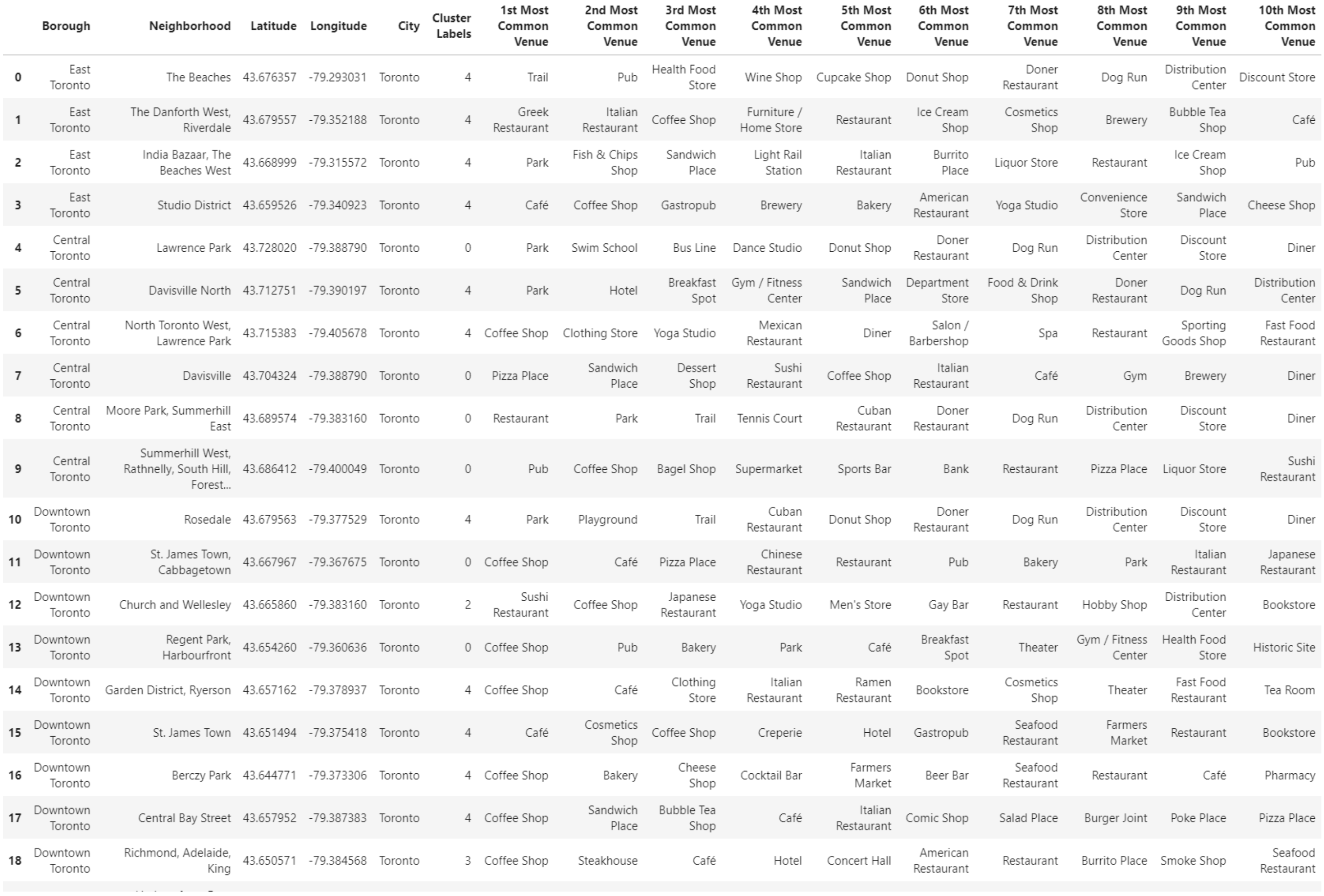


| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Battery Park City | Park | Coffee Shop | Memorial Site | Gourmet Shop | Food Court | Plaza | Hotel | Gym | Shopping Mall | Boat or Ferry |
| 1 | Carnegie Hill | Pizza Place | Gym / Fitness Center | Coffee Shop | Bookstore | Bakery | Gym | Italian Restaurant | French Restaurant | Café | Yoga Studio |
| 2 | Central Harlem | African Restaurant | Pizza Place | Seafood Restaurant | Cosmetics Shop | American Restaurant | Bar | Fried Chicken Joint | French Restaurant | Chinese Restaurant | Bookstore |
| 3 | Chelsea | Coffee Shop | American Restaurant | Italian Restaurant | Seafood Restaurant | Hotel | Ice Cream Shop | French Restaurant | Cupcake Shop | Middle Eastern Restaurant | Liquor Store |
| 4 | Chinatown | Chinese Restaurant | American Restaurant | Ice Cream Shop | Boutique | Greek Restaurant | Bubble Tea Shop | Salon / Barbershop | Sandwich Place | Spa | Asian Restaurant |

## 4.5. Cluster Neighborhood

Now, firstly let’s aggregate the two dataframe in a singol one. We have some common venue categories in the neighborhoods. We use the unsupervised learning K-means algorithm to cluster the neighborhoods. K-Means algorithm is one of the most common method for clustering in unsupervised learning. Let's run k-means to cluster the neighborhood into 5 clusters.

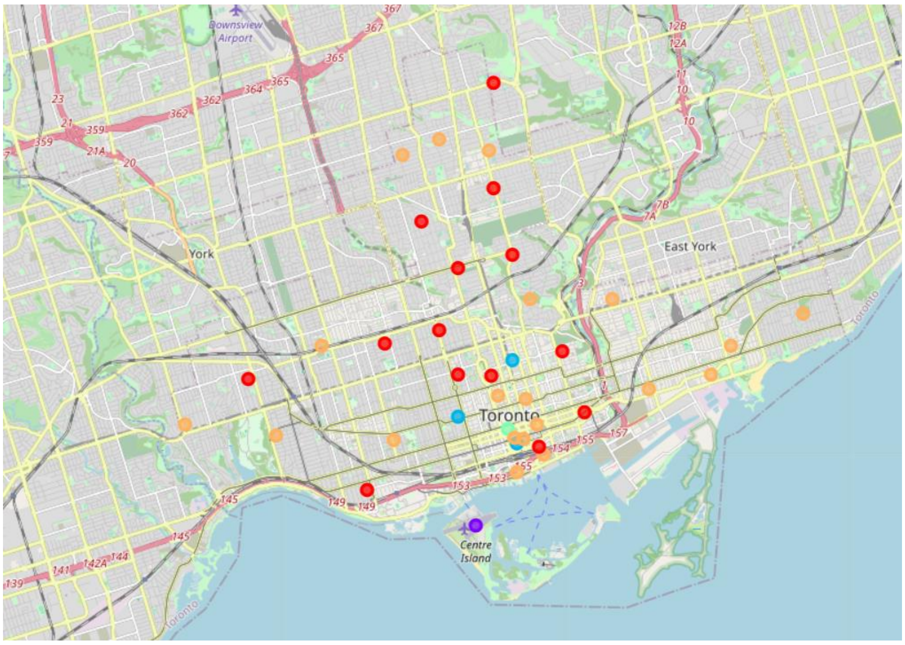
We can visualize the first 18 rows of the final result in the following table:

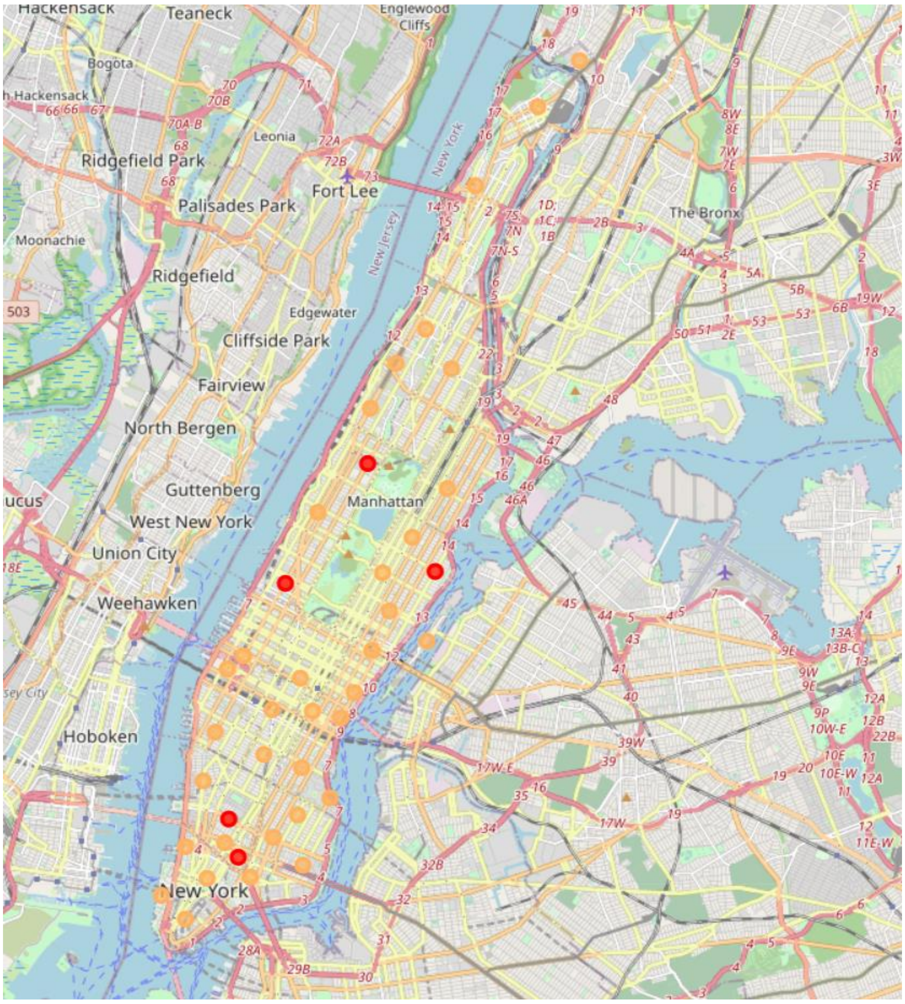


| | Borough | Neighborhood | Latitude | Longitude | City | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | East Toronto | The Beaches | 43.676357 | -79.293031 | Toronto | 4 | Trail | Pub | Health Food Store | Wine Shop | Cupcake Shop | Donut Shop | Doner Restaurant | Dog Run | Distribution Center | Discount Store |
| **1** | East Toronto | The Danforth West, Riverdale | 43.679557 | -79.352188 | Toronto | 4 | Greek Restaurant | Italian Restaurant | Coffee Shop | Furniture / Home Store | Restaurant | Ice Cream Shop | Cosmetics Shop | Brewery | Bubble Tea Shop | Café |
| **2** | East Toronto | India Bazaar, The Beaches West | 43.668999 | -79.315572 | Toronto | 4 | Park | Fish & Chips Shop | Sandwich Place | Light Rail Station | Italian Restaurant | Burrito Place | Liquor Store | Restaurant | Ice Cream Shop | Pub |
| **3** | East Toronto | Studio District | 43.659526 | -79.340923 | Toronto | 4 | Café | Coffee Shop | Gastropub | Brewery | Bakery | American Restaurant | Yoga Studio | Convenience Store | Sandwich Place | Cheese Shop |
| **4** | Central Toronto | Lawrence Park | 43.728020 | -79.388790 | Toronto | 0 | Park | Swim School | Bus Line | Dance Studio | Donut Shop | Doner Restaurant | Dog Run | Distribution Center | Discount Store | Diner |
| **5** | Central Toronto | Davisville North | 43.712751 | -79.390197 | Toronto | 4 | Park | Hotel | Breakfast Spot | Gym / Fitness Center | Sandwich Place | Department Store | Food & Drink Shop | Doner Restaurant | Dog Run | Distribution Center |
| **6** | Central Toronto | North Toronto West, Lawrence Park | 43.715383 | -79.405678 | Toronto | 4 | Coffee Shop | Clothing Store | Yoga Studio | Mexican Restaurant | Diner | Salon / Barbershop | Spa | Restaurant | Sporting Goods Shop | Fast Food Restaurant |
| **7** | Central Toronto | Davisville | 43.704324 | -79.388790 | Toronto | 0 | Pizza Place | Sandwich Place | Dessert Shop | Sushi Restaurant | Coffee Shop | Italian Restaurant | Café | Gym | Brewery | Diner |
| **8** | Central Toronto | Moore Park, Summerhill East | 43.689574 | -79.383160 | Toronto | 0 | Restaurant | Park | Trail | Tennis Court | Cuban Restaurant | Doner Restaurant | Dog Run | Distribution Center | Discount Store | Diner |
| **9** | Central Toronto | Summerhill West, Rathnelly, South Hill, Forest... | 43.686412 | -79.400049 | Toronto | 0 | Pub | Coffee Shop | Bagel Shop | Supermarket | Sports Bar | Bank | Restaurant | Pizza Place | Liquor Store | Sushi Restaurant |
| **10** | Downtown Toronto | Rosedale | 43.679563 | -79.377529 | Toronto | 4 | Park | Playground | Trail | Cuban Restaurant | Donut Shop | Doner Restaurant | Dog Run | Distribution Center | Discount Store | Diner |
| **11** | Downtown Toronto | St. James Town, Cabbagetown | 43.667967 | -79.367675 | Toronto | 0 | Coffee Shop | Café | Pizza Place | Chinese Restaurant | Restaurant | Pub | Bakery | Park | Italian Restaurant | Japanese Restaurant |
| **12** | Downtown Toronto | Church and Wellesley | 43.665860 | -79.383160 | Toronto | 2 | Sushi Restaurant | Coffee Shop | Japanese Restaurant | Yoga Studio | Men's Store | Gay Bar | Restaurant | Hobby Shop | Distribution Center | Bookstore |
| **13** | Downtown Toronto | Regent Park, Harbourfront | 43.654260 | -79.360636 | Toronto | 0 | Coffee Shop | Pub | Bakery | Park | Café | Breakfast Spot | Theater | Gym / Fitness Center | Health Food Store | Historic Site |
| **14** | Downtown Toronto | Garden District, Ryerson | 43.657162 | -79.378937 | Toronto | 4 | Coffee Shop | Café | Clothing Store | Italian Restaurant | Ramen Restaurant | Bookstore | Cosmetics Shop | Theater | Fast Food Restaurant | Tea Room |
| **15** | Downtown Toronto | St. James Town | 43.651494 | -79.375418 | Toronto | 4 | Café | Cosmetics Shop | Coffee Shop | Creperie | Hotel | Gastropub | Seafood Restaurant | Farmers Market | Restaurant | Bookstore |
| **16** | Downtown Toronto | Berczy Park | 43.644771 | -79.373306 | Toronto | 4 | Coffee Shop | Bakery | Cheese Shop | Cocktail Bar | Farmers Market | Beer Bar | Seafood Restaurant | Restaurant | Café | Pharmacy |
| **17** | Downtown Toronto | Central Bay Street | 43.657952 | -79.387383 | Toronto | 4 | Coffee Shop | Sandwich Place | Bubble Tea Shop | Café | Italian Restaurant | Comic Shop | Salad Place | Burger Joint | Poke Place | Pizza Place |
| **18** | Downtown Toronto | Richmond, Adelaide, King | 43.650571 | -79.384568 | Toronto | 3 | Coffee Shop | Steakhouse | Café | Hotel | Concert Hall | American Restaurant | Restaurant | Burrito Place | Smoke Shop | Seafood Restaurant |

## 5. Results: Visualizing the resulting clusters

Now using some python libraries we can visualize the clusters on a map of both cities and make a comparison





## 6. Discussion

The aim of this analysis was to find out similar neighborhoods for a person relocating in New York city. The maps above show us that if your friend want to move from a neighborhood in Toronto to a neighborhood in Manhattan, he has to choose the neighborhood with the same color displayed if he want to find the same kind of venues of his living zone. Consequently he's lucky if he live in an orange or red zone of the map above in Toronto.

## 7. Conclusion

The scope of the analysis has been achieved. However the model created can easily be replicated again and again with data from other cities by using the Foursquare API. This show us the potentiality of Data Science in real life problems