

Segmentation & Clustering of New York and Toronto

Capstone Project - The Battle of Neighborhoods

Nikhila Anumukonda



- INTRODUCTION
- DATA ACQUISITION & PREPARATION
 - CLUSTERING
- CLUSTER ANALYSIS
 - CONCLUSION

CONTENTS

INTRODUCTION

- PROBLEM STATEMENT
- TARGET AUDIENCE

PROBLEM STATEMENT & TARGET AUDIENCE

- A young business entrepreneur from India is looking to migrate to Toronto or New York and setup a business in Toronto or New York. He needs to decide between the two options New York or Toronto. In this scenario, he wants to understand which categories of business are popular in Toronto and New York, which categories of business are rare and niche. He also wants to understand the neighboring business venues.
- **The audience have the following profile.**
- He is a Indian citizen with dreams to migrate to New York or Toronto because they are the business centers in their respective countries.
- He does not know much about the New York and Toronto and their neighborhoods, hence he wants to understand feasibility of setting up business in the neighborhoods.
- He is exploring what kind of business has less competition and which neighborhoods.
- He has a family with skills to run the business.
- He likes to live in a neighborhood with ample open spaces.

- NEIGHBORHOOD DATA
 - VENUE DATA

DATA ACQUISITION AND PREPARATION

Data Acquisition

Neighborhood Data (Toronto)

Source for Canada Data : Wikipedia page titled “List of postal codes of Canada: M”.

Source for Geospatial Data : https://cocl.us/geospatial_data

The following steps were performed on the Toronto Data for Data preparation

- Scrape the data from the Wiki page
- combine the neighborhoods which have same postcode
- replace the name of the neighborhoods which are “Not Assigned” with the name of the Borough
- merge the neighborhoods which have same Borough
- filter the data having Borough “Toronto”.
- Read Geospatial data and merge with Toronto data to get Longitudes and Latitudes.

	Postcode	Borough	Neighbourhood	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	The Beaches West, India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923

Data Acquisition

Neighborhood Data (New York)

Source for New York Data : https://cocl.us/new_york_dataset.

The dataset is a JSON file that specifies the name of each neighborhood, its coordinates—latitude and longitude, its borough combine the neighborhoods which have same postcode

The following steps were performed on the Toronto Data for Data preparation

- Neighborhoods that share the same name but are located in different boroughs will include the borough name in the name of these neighborhoods

	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
5	Bronx	Kingsbridge	40.881687	-73.902818

Data Acquisition

Venue Data (Toronto and New York)

Source for New York Data : Foursquare API

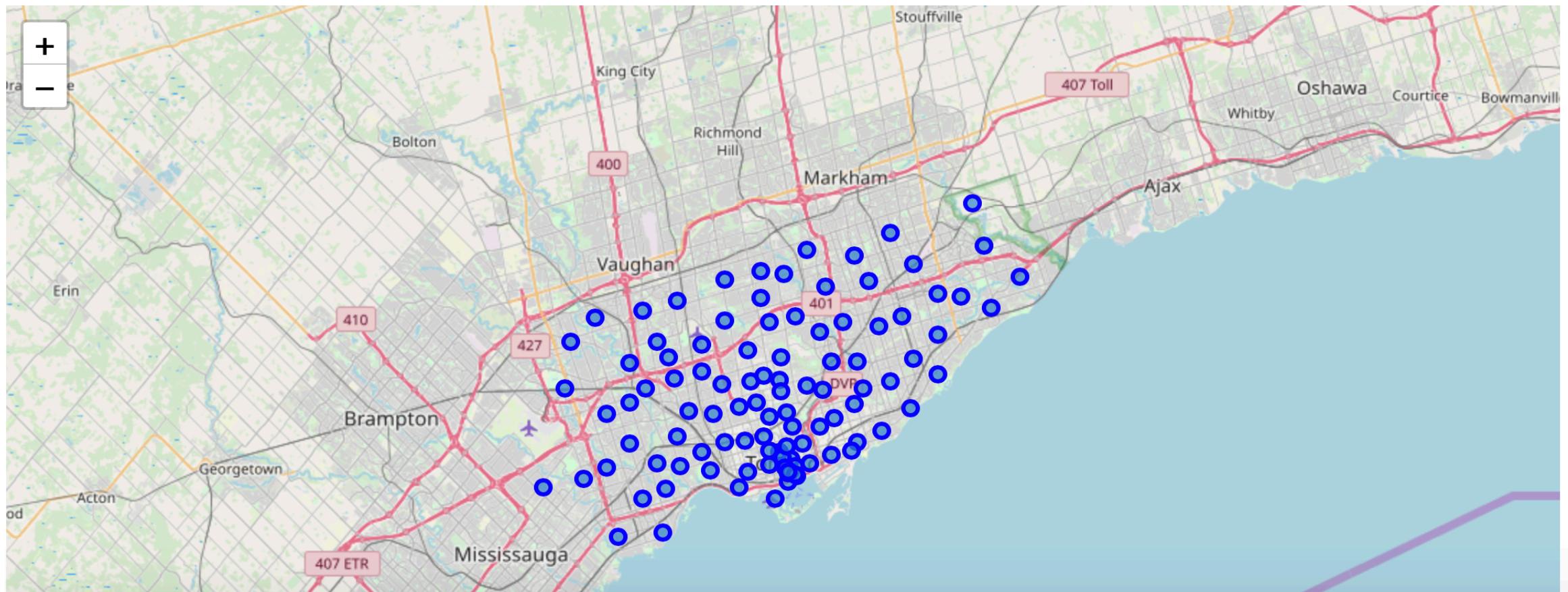
To retrieve data from Foursquare using their API, a URL should be prepared and used to request data related a specific location. Search criteria, Client Id, Client Secret, Version, Latitude, Longitude, Radius, Limit are passed.

For example : https://api.foursquare.com/v2/venues/search?client_id=1234&client_secret=1234&v=20180605&ll=40.89470517661,-73.84720052054902&radius=500&limit=100

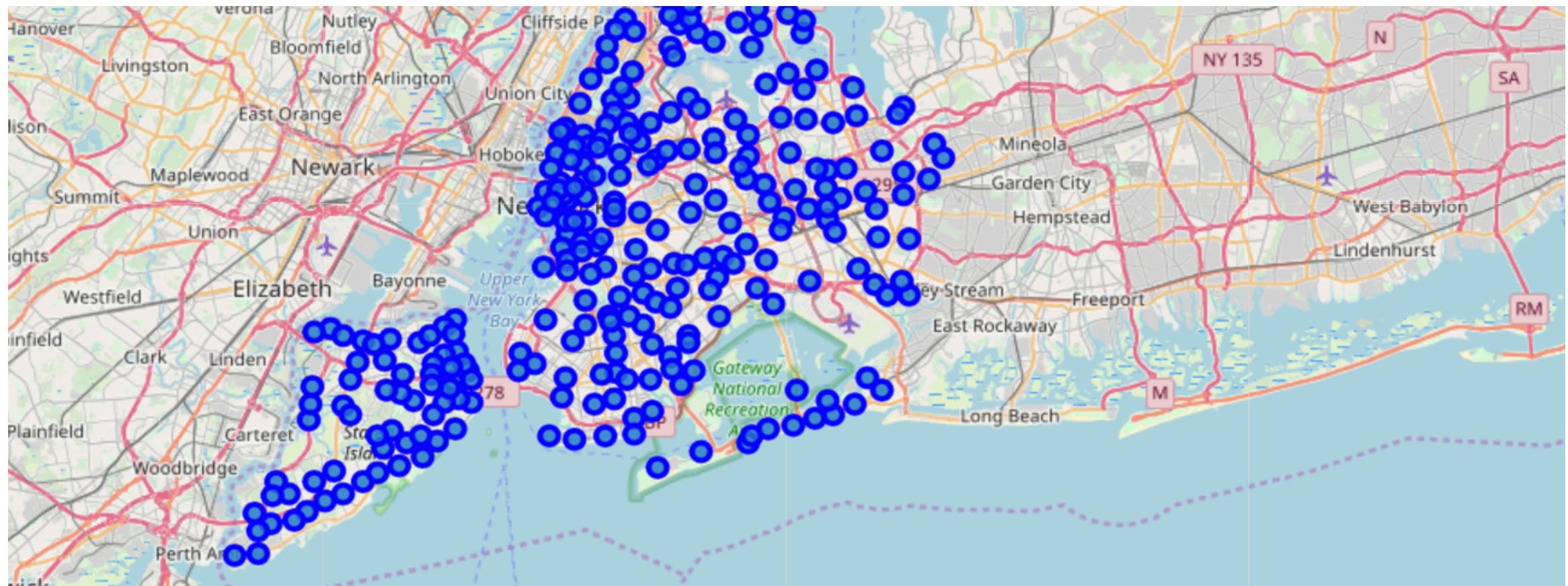
- search indicates the API endpoint used,
- client_id and client_secret are credentials used to access the API service and are obtained when registering a Foursquare developer account,
- v indicates the API version to use,
- ll indicates the latitude and longitude of the desired location,
- radius is the maximum distance in meters between the specified location and the retrieved venues, and limit is used to limit the number of returned results if necessary.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653191	-79.357947	Gym / Fitness Center
3	Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
4	Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

MAP OF TORONTO NEIGHBORHOODS



MAP OF NEW YORK NEIGHBORHOODS





**Most Common Venue
Categories**

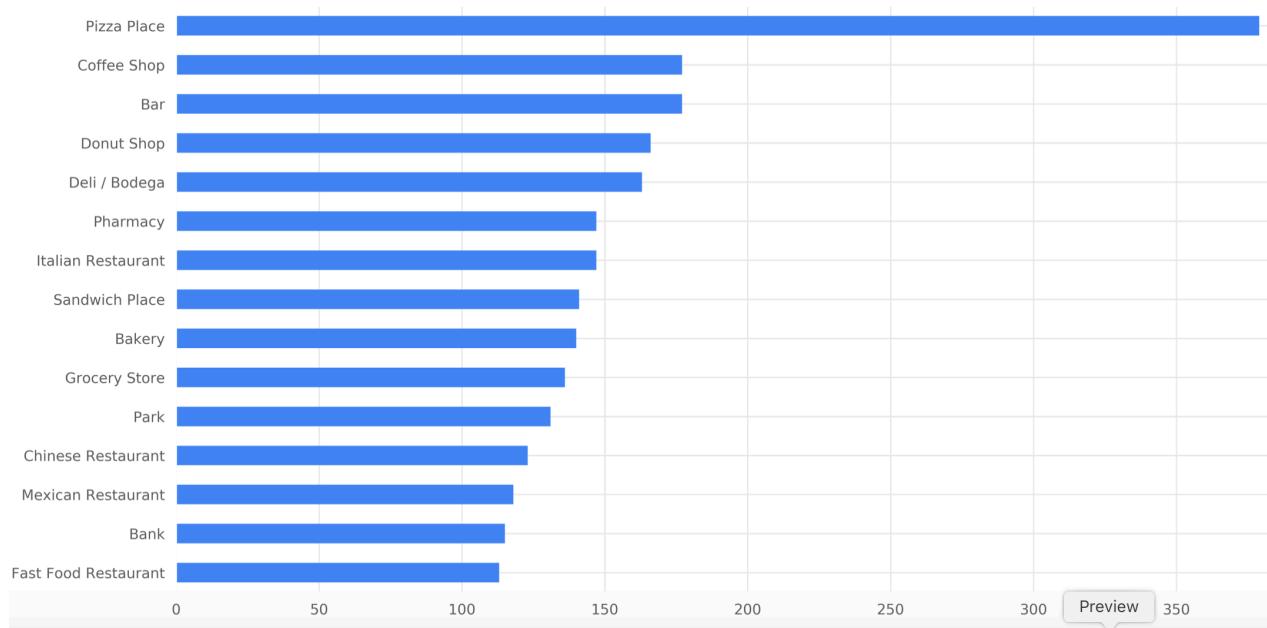
**Most widespread Venue
Categories**

Most Rare Categories

METHODOLOGY

Most Common Venue Categories - NY

New York City

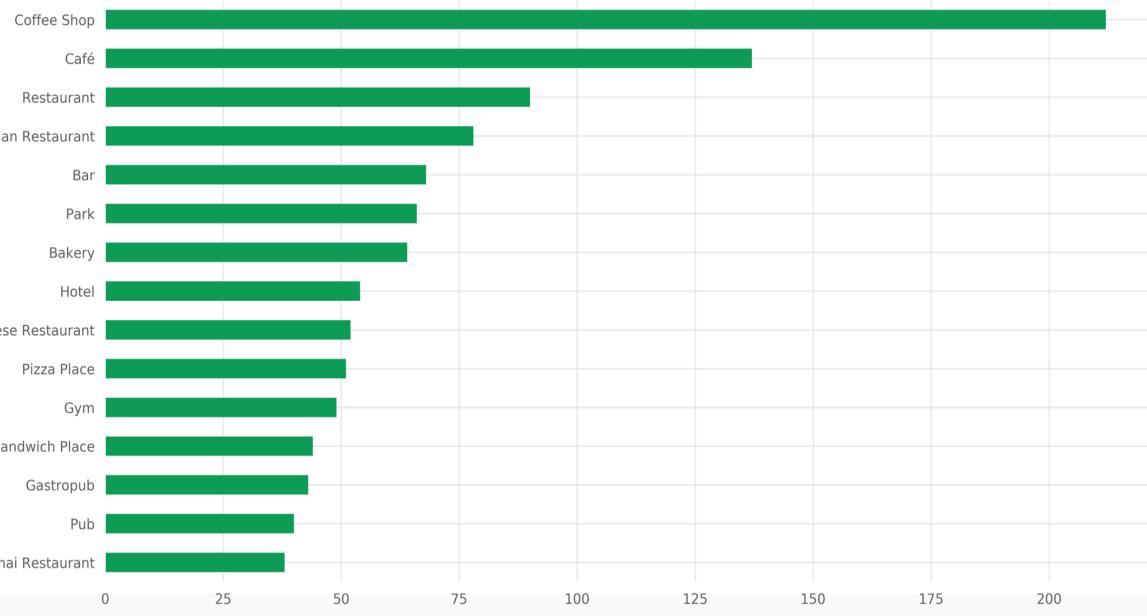


We can see that the most common category is

- “Pizza Place” with 400 venues in NYC;
- In the second rank, the category “Coffee shop” appears with 175 venues.
- In the third rank comes the “Bar” category.

Most Common Venue Categories - Toronto

Toronto

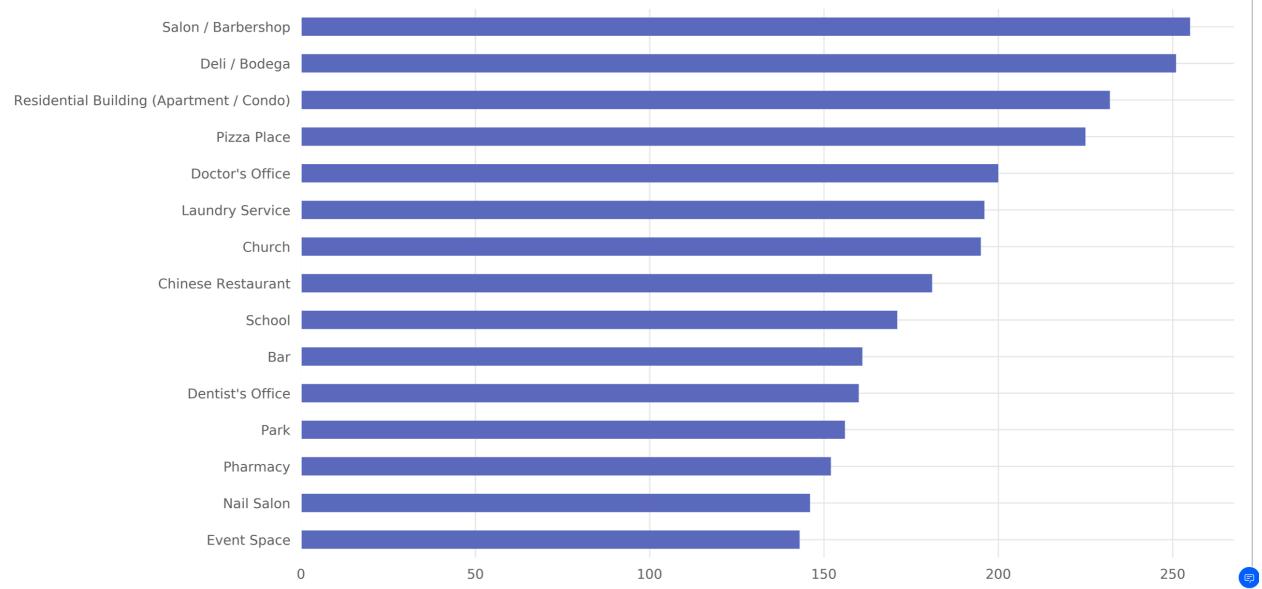


For Toronto, the Top 3 most common venue category is Coffee Shop, Park and Café.

- Coffee Shop with 250 venues,
- restaurant with 87 venues and
- Café with 137 venues.

Most widespread Venue Categories - NY

New York City



- The most widespread category is “Salon / Barbershop” and exist in 250 neighborhoods out of the 306 neighborhoods.
- Next comes the “Deli / Bodega” category with venues in 250 neighborhoods also.
- In the third place comes the “Residential Building (Apartment / Condo)” category with venues in 230 neighborhoods.

Most widespread Venue Categories - Toronto

Toronto



- In the first place comes the “Coffee Shop” category with venues in 80 neighborhoods.
- Then comes “Residential Building (Apartment / Condo)” category with venues in 80 neighborhoods. And the third most-widespread category is “Park” with venues in 75 neighborhoods.

RARE CATEGORIES IN NEW YORK

```
|: nyc_bot_vc = nyneighborhoods_venues['venue_category'].value_counts(normalize=False)
nyc_bot_vc = nyc_bot_vc.tail(15).to_frame('Count')
nyc_bot_vc.index.names = ['Venue Category']
nyc_bot_vc
```

	Count
Venue Category	
Memorial Site	1
Empanada Restaurant	1
Motel	1
Varenyky restaurant	1
Winery	1
Flea Market	1
Waterfront	1
Tailor Shop	1
Laundry Service	1
Shopping Plaza	1
Music Store	1
Austrian Restaurant	1
Shabu-Shabu Restaurant	1
Lake	1
Tennis Stadium	1

Certain Venue categories are Rare. It might be feasible to select one of these categories to set up business as there would be less competition. For example:

- Motel
- Laundry Service
- Music store
- Austrian restaurant

RARE CATEGORIES IN TORONTO

Rare Categories in Toronto

```
|: tor_bot_vc = tor_venues['Venue Category'].value_counts(normalize=False)
tor_bot_vc = tor_bot_vc.tail(15).to_frame('Count')
tor_bot_vc.index.names = ['Venue Category']
tor_bot_vc
```

Venue Category	Count
Coworking Space	1
Churrascaria	1
Soup Place	1
Tanning Salon	1
IT Services	1
Cantonese Restaurant	1
Pool Hall	1
Bed & Breakfast	1
Indonesian Restaurant	1
Belgian Restaurant	1
College Rec Center	1
Farm	1
Theme Restaurant	1
River	1
Dive Bar	1

Certain Venue categories are Rare. It might be feasible to select one of these categories to set up business as there would be less competition.

For example:

- Salon
- IT Services
- Theme Restaurant

CLUSTERING OF
NEIGHBORHOODS

RESULTS

K-means clustering algorithm

- Preparing the data

Prepare the data for K-means clustering algorithm :

The following steps are performed on the data since K Means takes numerical values only.

- **apply one-hot encoding on “Venue Category”**

	Neighborhood_	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	.
0	Harbourfront	0	0	0	0	0	0	0	0	0	..
1	Harbourfront	0	0	0	0	0	0	0	0	0	..
2	Harbourfront	0	0	0	0	0	0	0	0	0	..
3	Harbourfront	0	0	0	0	0	0	0	0	0	..
4	Harbourfront	0	0	0	0	0	0	0	0	0	..

- **aggregation on “Neighborhood”**

	Neighborhood_	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	.
0	Adelaide, King, Richmond	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.020202	..
1	Berczy Park	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	..
2	Brockton, Exhibition Place, Parkdale Village	0.011905	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.011905	..

This process is applied for both Toronto, New York datasets separately

Grouping rows by neighborhood and taking the mean of the frequency of occurrence of each category for each neighborhood

7 MOST COMMON VENUES IN TORONTO NEIGHBORHOOD

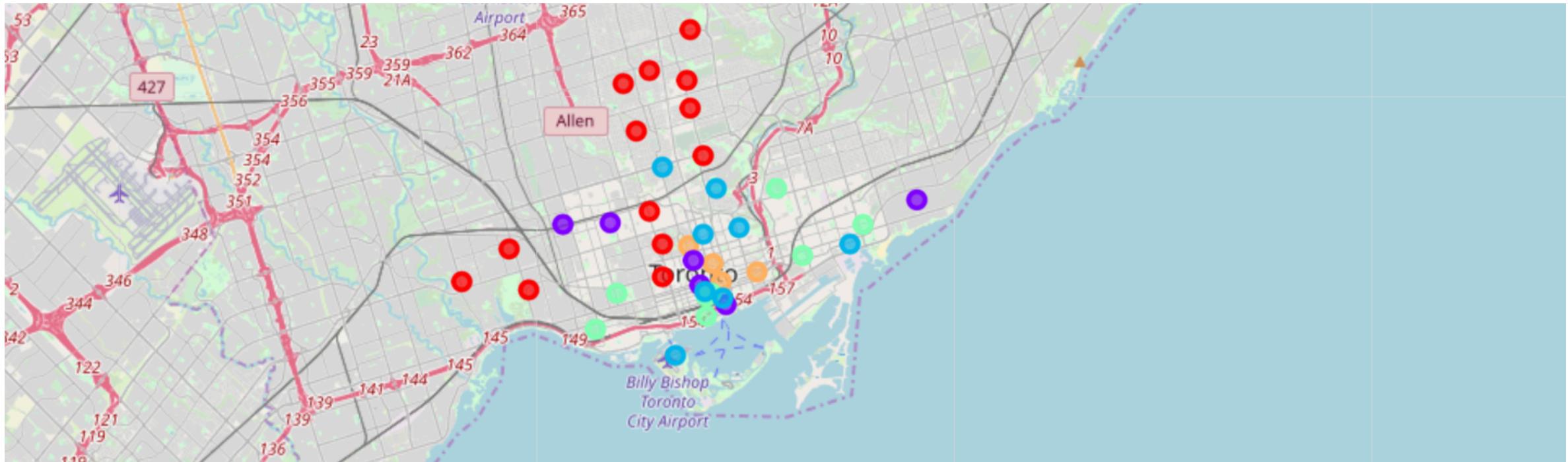
	Neighborhood_	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
0	Adelaide, King, Richmond	Coffee Shop	Café	Restaurant	Sushi Restaurant	Hotel	Theater	Pizza Place
1	Berczy Park	Coffee Shop	Restaurant	Café	Hotel	Beer Bar	Japanese Restaurant	Breakfast Spot
2	Brockton, Exhibition Place, Parkdale Village	Coffee Shop	Café	Gift Shop	Bar	Bakery	Restaurant	Japanese Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Fast Food Restaurant	Bar	Burrito Place	Bakery	Light Rail Station	Brewery	Pub
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Rental Car Location	Airport Service	Sculpture Garden	Coffee Shop	Harbor / Marina	Airport Terminal	Boat or Ferry
5	Cabbagetown, St. James Town	Coffee Shop	Café	Grocery Store	Park	Restaurant	Pizza Place	Diner
6	Central Bay Street	Coffee Shop	Clothing Store	Italian Restaurant	Japanese Restaurant	Art Gallery	Sushi Restaurant	Park
7	Chinatown, Grange Park, Kensington Market	Bar	Café	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Bakery	Mexican Restaurant	Dessert Shop
8	Christie	Café	Grocery Store	Park	Italian Restaurant	Indian Restaurant	Diner	Coffee Shop
9	Church and Wellesley	Coffee Shop	Japanese Restaurant	Burger Joint	Restaurant	Café	Gastropub	Park
10	Commerce Court, Victoria Hotel	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	American Restaurant	Concert Hall
11	Davisville	Italian Restaurant	Coffee Shop	Pizza Place	Gym	Sandwich Place	Restaurant	Dessert Shop
12	Davisville North	Pizza Place	Coffee Shop	Gym	Café	Park	Sandwich Place	Breakfast Spot

7 MOST COMMON VENUES IN NEW YORK NEIGHBORHOOD

Neighborhood	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category	
0	Allerton	Pizza Place	Donut Shop	Sandwich Place	Food	Bar	Discount Store	Fast Food Restaurant
1	Annadale	Pizza Place	Restaurant	Park	Sushi Restaurant	Pharmacy	Train Station	American Restaurant
2	Arden Heights	Mexican Restaurant	Spa	Bagel Shop	Liquor Store	Sandwich Place	Elementary School	Food
3	Arlington	Deli / Bodega	Juice Bar	Intersection	Supermarket	Coffee Shop	Home Service	Boat or Ferry
4	Arrochar	Italian Restaurant	Deli / Bodega	Baseball Field	Pizza Place	Beach	Martial Arts Dojo	Toll Plaza
5	Arverne	Surf Spot	Beach	Sandwich Place	Donut Shop	Supermarket	Playground	Furniture / Home Store
6	Astoria	Bar	Greek Restaurant	Coffee Shop	Middle Eastern Restaurant	Bakery	Grocery Store	Seafood Restaurant
7	Astoria Heights	Rental Car Location	Italian Restaurant	Bakery	Supermarket	Chinese Restaurant	Home Service	Pizza Place
8	Auburndale	Korean Restaurant	Cosmetics Shop	Pizza Place	Greek Restaurant	Mobile Phone Shop	Bank	Gym / Fitness Center
9	Bath Beach	Bank	Pizza Place	Bakery	Cantonese Restaurant	Coffee Shop	Bubble Tea Shop	Chinese Restaurant
10	Battery Park City	Coffee Shop	Park	Hotel	Plaza	Women's Store	Grocery Store	Memorial Site
11	Bay Ridge	Pizza Place	Spa	Italian Restaurant	Greek Restaurant	Bar	Bagel Shop	Cosmetics Shop
12	Bay Terrace, Queens	Clothing Store	American Restaurant	Mobile Phone Shop	Lingerie Store	Cosmetics Shop	Kids Store	Donut Shop



CLUSTERING OF THE TORONTO NEIGHBORHOODS



TORONTO CLUSTERS

■ Cluster 1

Postcode	Longitude	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
23	M5P -79.411307	0	Jewelry Store	Mexican Restaurant	Sushi Restaurant	Trail	Electronics Store	Dog Run	Doner Restaurant

■ Cluster 2 – part of cluster dataset of 35 clusters

Postcode	Longitude	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
0	M4E -79.293031	1	Pub	Gastropub	Bar	Sandwich Place	Breakfast Spot	Coffee Shop	Japanese Restaurant
1	M4K -79.352188	1	Greek Restaurant	Coffee Shop	Pub	Café	Ice Cream Shop	Italian Restaurant	Fast Food Restaurant
2	M4L -79.315572	1	Indian Restaurant	Grocery Store	Gym	Coffee Shop	Fast Food Restaurant	Sandwich Place	Café
3	M4M -79.340923	1	Café	Coffee Shop	Bar	Diner	Bakery	Sandwich Place	Italian Restaurant
5	M4P -79.390197	1	Pizza Place	Coffee Shop	Gym	Café	Park	Sandwich Place	Breakfast Spot
6	M4R -79.405678	1	Clothing Store	Coffee Shop	Café	Sporting Goods Shop	Dessert Shop	Diner	Restaurant
7	M4S -79.388790	1	Italian Restaurant	Coffee Shop	Pizza Place	Gym	Sandwich Place	Restaurant	Dessert Shop
8	M4T -79.383160	1	Park	Grocery Store	Thai Restaurant	Sandwich Place	Candy Store	Gym / Fitness Center	Gym
9	M4V -79.400049	1	Coffee Shop	Sushi Restaurant	Italian Restaurant	Restaurant	Sandwich Place	Gym	Pizza Place

TORONTO CLUSTERS

■ Cluster 3

Postcode	Longitude	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
22	M5N -79.416936	2	Playground	Garden	Home Service	Pet Store	Electronics Store	Dive Bar	Dog Run

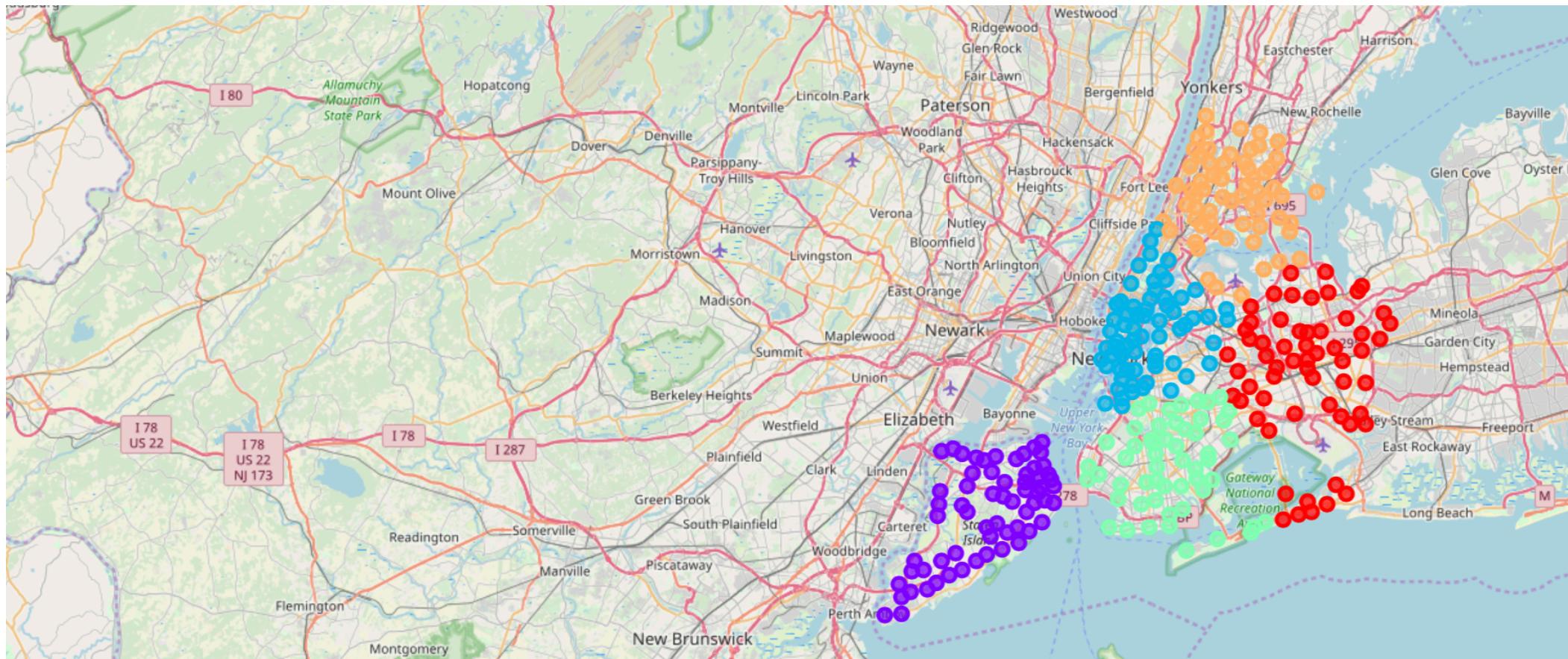
■ Cluster 4

Postcode	Longitude	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
10	M4W -79.377529	3	Park	Trail	Playground	Candy Store	Yoga Studio	Dive Bar	Dog Run

■ Cluster 5

Postcode	Longitude	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
4	M4N -79.38879	4	Locksmith	Electronics Store	Swim School	Business Service	Coffee Shop	Park	Doner Restaurant

CLUSTERING OF THE NEW YORK NEIGHBORHOODS





NEW YORK BOROUGHS

- BRONX
- MANHATTAN
- BROOKLYN
- QUEENS
- STATEN ISLAND

CLUSTERS FOR NEW YORK

■ Cluster 1

Borough	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
71 Brooklyn	Fried Chicken Joint	Latin American Restaurant	Fast Food Restaurant	Donut Shop	Grocery Store	Ice Cream Shop	Pizza Place
90 Brooklyn	Pizza Place	Supermarket	Fried Chicken Joint	Grocery Store	Donut Shop	Chinese Restaurant	Liquor Store
133 Queens	Italian Restaurant	Sandwich Place	Park	Sushi Restaurant	Ice Cream Shop	Pharmacy	Fast Food Restaurant

■ Cluster 2

Borough	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
197 Staten Island	Clothing Store	Italian Restaurant	Bakery	Bar	Sporting Goods Shop	Art Gallery	Ice Cream Shop
198 Staten Island	Mexican Restaurant	Chinese Restaurant	Park	Playground	Boat or Ferry	Convenience Store	Art Gallery
199 Staten Island	Bank	Harbor / Marina	Discount Store	Mexican Restaurant	Sandwich Place	Pizza Place	Supermarket
200 Staten Island	Beach	Grocery Store	Italian Restaurant	Train Station	Deli / Bodega	Supermarket	Athletics & Sports
201 Staten Island	Bar	Coffee Shop	Pizza Place	Ice Cream Shop	Italian Restaurant	Breakfast Spot	Salon / Barbershop

CLUSTERS FOR NEW YORK

■ Cluster 3

Borough	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
49	Brooklyn	Bar	Coffee Shop	Cocktail Bar	Pizza Place	Café	Record Shop
61	Brooklyn	Bar	Pizza Place	Coffee Shop	Wine Bar	American Restaurant	Breakfast Spot
62	Brooklyn	Bar	Mexican Restaurant	Coffee Shop	Pizza Place	Café	Deli / Bodega
63	Brooklyn	Pizza Place	Deli / Bodega	Coffee Shop	Chinese Restaurant	Bar	Café
64	Brooklyn	Park	Yoga Studio	Italian Restaurant	Bakery	Wine Shop	Pizza Place
65	Brooklyn	Italian Restaurant	Cocktail Bar	Bar	Playground	Yoga Studio	Cosmetics Shop

■ Cluster 4

Borough	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
46	Brooklyn	Pizza Place	Spa	Italian Restaurant	Greek Restaurant	Bar	Bagel Shop
47	Brooklyn	Donut Shop	Chinese Restaurant	Japanese Restaurant	Italian Restaurant	Bakery	Sushi Restaurant
48	Brooklyn	Pizza Place	Bakery	Chinese Restaurant	Mexican Restaurant	Latin American Restaurant	Bank

CLUSTERS FOR NEW YORK

■ Cluster 5

Borough	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category	
0	Bronx	Supermarket	Pharmacy	Gas Station	Fried Chicken Joint	Caribbean Restaurant	Pizza Place	Donut Shop
1	Bronx	Mattress Store	Restaurant	Pizza Place	Fast Food Restaurant	Paper / Office Supplies Store	Grocery Store	Men's Store
2	Bronx	Caribbean Restaurant	Fast Food Restaurant	Deli / Bodega	Diner	Bowling Alley	Historic Site	Seafood Restaurant
3	Bronx	Plaza	Park	Art Gallery	Coffee Shop	Home Service	Pizza Place	Athletics & Sports

Combining NYC and Toronto Data Neighborhoods

Most common venue categories in each neighborhood after Clustering K=5

The below Table shows the number of neighborhoods in each cluster. The clustering algorithm was run on 406 neighborhoods in NYC and Toronto.

Neighborhood_	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
Wingate_NY	2	Pizza Place	Donut Shop	Deli / Bodega	Grocery Store	Fried Chicken Joint	Pharmacy	Sandwich Place
Woodhaven_NY	2	Deli / Bodega	Latin American Restaurant	Pizza Place	Pharmacy	Supermarket	Fried Chicken Joint	Sandwich Place
Woodlawn_NY	2	Pub	Bar	Deli / Bodega	Pizza Place	Italian Restaurant	Discount Store	Grocery Store
Woodrow_NY	2	Grocery Store	Donut Shop	Diner	Bank	Bakery	Park	Liquor Store
Woodside_NY	0	Thai Restaurant	Bar	Grocery Store	Latin American Restaurant	Bakery	Chinese Restaurant	Pub
Yorkville_NY	0	Gym	Italian Restaurant	Coffee Shop	Pizza Place	Thai Restaurant	Ice Cream Shop	Deli / Bodega
Adelaide, King, Richmond_TOR	0	Coffee Shop	Café	Restaurant	Sushi Restaurant	Thai Restaurant	Bar	Steakhouse
Berczy Park_TOR	0	Coffee Shop	Hotel	Café	Restaurant	Japanese Restaurant	Beer Bar	Cocktail Bar
Brockton, Exhibition Place, Parkdale Village_TOR	0	Coffee Shop	Café	Gift Shop	Bakery	Bar	Restaurant	Nightclub
Business Reply Mail Processing Centre 969 Eastern_TOR	0	Fast Food Restaurant	Bar	Park	Burrito Place	Light Rail Station	Bakery	Brewery

Cluster 1

Some records that belong to the first cluster

Neighborhood_	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
Lighthouse Hill_NY	0	Trail	Clothing Store	Bagel Shop	Spa	Café	History Museum	Art Museum
Lincoln Square_NY	0	Plaza	Italian Restaurant	Theater	French Restaurant	Concert Hall	Gym / Fitness Center	Park
Little Italy_NY	0	Italian Restaurant	Cosmetics Shop	Café	Ice Cream Shop	Cocktail Bar	Sandwich Place	Hotel
Little Neck_NY	0	Italian Restaurant	Chinese Restaurant	Bank	Coffee Shop	Bakery	Middle Eastern Restaurant	Mobile Phone Shop
Long Island City_NY	0	Hotel	Coffee Shop	Pizza Place	Mexican Restaurant	Gym / Fitness Center	Bubble Tea Shop	Indian Restaurant
Lower East Side_NY	0	Art Gallery	Ice Cream Shop	Mexican Restaurant	Coffee Shop	Japanese Restaurant	Asian Restaurant	Pizza Place
Manhattan Beach_NY	0	Harbor / Marina	Dessert Shop	Turkish Restaurant	Restaurant	Seafood Restaurant	Lake	Sandwich Place
Manhattan Valley_NY	0	Park	Indian Restaurant	Coffee Shop	Pizza Place	Chinese Restaurant	Bar	Wine Shop
Manhattanville_NY	0	Chinese Restaurant	Park	Mexican Restaurant	Italian Restaurant	Coffee Shop	Pizza Place	Café
Midtown_NY	0	Coffee Shop	Hotel	Plaza	Pizza Place	Japanese Restaurant	Sandwich Place	Bakery

Cluster2

Some records that belong to the second cluster

Cluster Labels 1st Most Common Category 2nd Most Common Category 3rd Most Common Category 4th Most Common Category 5th Most Common Category 6th Most Common Category 7th Most Common Category

Neighborhood_

Todt Hill_NY	1	Mountain	Zoo Exhibit	Festival	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service
--------------	---	----------	-------------	----------	--------------------	-----------------------	----------------------	---------------

CLUSTER 3

- Some records that belong to the third cluster

Neighborhood	Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
Willowbrook_NY	2	Bagel Shop	Chinese Restaurant	Pizza Place	Zoo Exhibit	Farmers Market	English Restaurant	Entertainment Service
Wingate_NY	2	Pizza Place	Donut Shop	Deli / Bodega	Grocery Store	Fried Chicken Joint	Pharmacy	Sandwich Place
Woodhaven_NY	2	Deli / Bodega	Latin American Restaurant	Pizza Place	Pharmacy	Supermarket	Fried Chicken Joint	Sandwich Place
Woodlawn_NY	2	Pub	Bar	Deli / Bodega	Pizza Place	Italian Restaurant	Discount Store	Grocery Store
Woodrow_NY	2	Grocery Store	Donut Shop	Diner	Bank	Bakery	Park	Liquor Store

Cluster 4

Some records that belong to the fourth cluster

Cluster Labels 1st Most Common Category 2nd Most Common Category 3rd Most Common Category 4th Most Common Category 5th Most Common Category 6th Most Common Category 7th Most Common Category

Neighborhood_

Willowbrook_NY	2	Bagel Shop	Chinese Restaurant	Pizza Place	Zoo Exhibit	Farmers Market	English Restaurant	Entertainment Service
Wingate_NY	2	Pizza Place	Donut Shop	Deli / Bodega	Grocery Store	Fried Chicken Joint	Pharmacy	Sandwich Place
Woodhaven_NY	2	Deli / Bodega	Latin American Restaurant	Pizza Place	Pharmacy	Supermarket	Fried Chicken Joint	Sandwich Place
Woodlawn_NY	2	Pub	Bar	Deli / Bodega	Pizza Place	Italian Restaurant	Discount Store	Grocery Store
Woodrow_NY	2	Grocery Store	Donut Shop	Diner	Bank	Bakery	Park	Liquor Store

Cluster 5

Some records that belong to the fifth cluster

Cluster Labels	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
<u>Neighborhood_</u>							
Bayswater_NY	4	Playground	Construction & Landscaping	Park	Athletics & Sports	Fast Food Restaurant	English Restaurant
Clason Point_NY	4	Park	Boat or Ferry	Moving Target	Zoo Exhibit	Festival	Entertainment Service
Howland Hook_NY	4	Harbor / Marina	Resort	Park	Zoo Exhibit	Farmers Market	Empanada Restaurant
Malba_NY	4	Park	Playground	Cosmetics Shop	Rest Area	Food Truck	Harbor / Marina
Forest Hill North, Forest Hill West_TOR	4	Park	Sushi Restaurant	Jewelry Store	Gym / Fitness Center	Dry Cleaner	Trail
Rosedale_TOR	4	Park	Trail	Candy Store	Playground	Flea Market	Falafel Restaurant
							Empanada Restaurant

RARE VENUE CATEGORIES IN NY AND TORONTO

```
nyc_tor_bot_vc = pd.concat([NYneighborhoods_venues, tor_venues])['Venue Category'].value_counts(normalize=False)
nyc_tor_bot_vc = nyc_tor_bot_vc.tail(15).to_frame('Count')
nyc_tor_bot_vc.index.names = ['Venue Category']
nyc_tor_bot_vc
```

:

Venue Category	Count
Souvenir Shop	1
Syrian Restaurant	1
Volleyball Court	1
Plane	1
Auditorium	1
Mountain	1
Vape Store	1
Cooking School	1
Gun Range	1
Baby Store	1

Cluster Analysis

Cluster 1:

Category	% of venues
Coffee Shop	4.93001
Italian Restaurant	3.0693
Bar	2.83454
Pizza Place	2.79976
Café	2.74759
Bakery	2.47805
Park	2.09547

Cluster 2:

Category	% of venues
Mountain	33.3333
Dog Run	33.3333
Park	33.3333

Cluster 5:

Category	% of venues
Park	28.2051
Playground	10.2564
Trail	7.69231
Harbor / Marina	5.12821
Bus Stop	5.12821
Dry Cleaner	2.5641
Candy Store	2.5641

Cluster 3:

Category	% of venues
Pizza Place	6.68512
Donut Shop	3.72852
Pharmacy	3.49548
Deli / Bodega	3.42266
Sandwich Place	2.88378
Bank	2.79639
Chinese Restaurant	2.66531

Cluster 4:

Category	% of venues
Beach	28.5714
Deli / Bodega	5.71429
Baseball Field	5.71429
Park	4.28571
Spa	4.28571
Pier	2.85714
Pub	2.85714

Cluster Analysis

- While Park constitute 2% of venues in the neighborhoods of the first cluster, they constitute 33% of the venues in the second cluster, 4% of the venues in the fourth cluster, and 28% of the venues in the fifth cluster.
- Pizza places appear in the most common categories of the first, and third clusters only.
- Italian restaurants appear in the most common categories of the First cluster only.
- Chinese restaurant appear in the most common categories of the third cluster only
- Beach is a very popular category in cluster four.

Cluster 1:

Category	% of venues
Coffee Shop	4.93001
Italian Restaurant	3.0693
Bar	2.83454
Pizza Place	2.79976
Café	2.74759
Bakery	2.47805
Park	2.09547

Cluster 2:

Category	% of venue
Mountain	33.3333
Dog Run	33.3333
Park	33.3333

Cluster 3:

Category	% of venues
Pizza Place	6.68512
Donut Shop	3.72852
Pharmacy	3.49548
Deli / Bodega	3.42266
Sandwich Place	2.88378
Bank	2.79639
Chinese Restaurant	2.66531

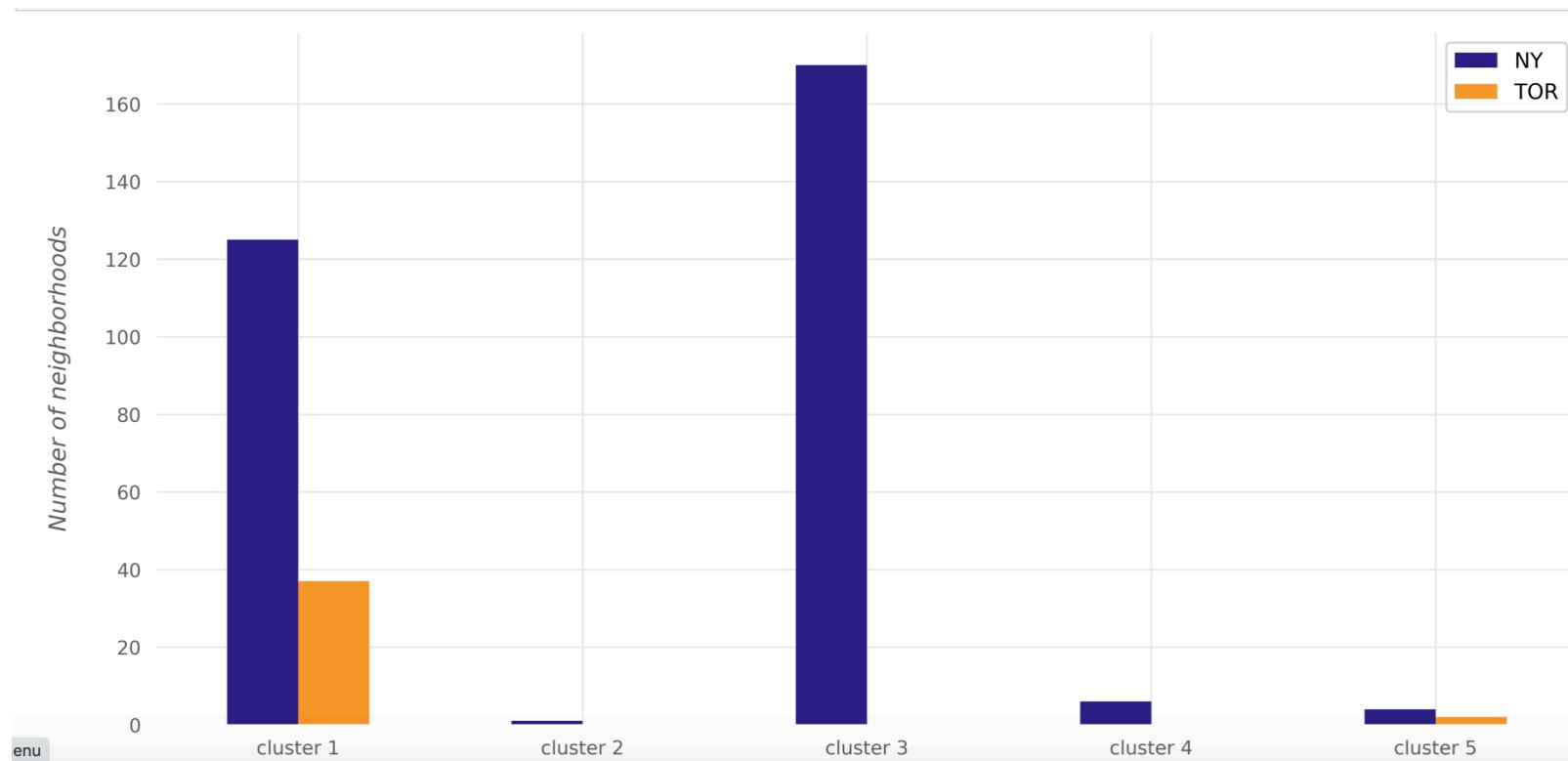
Cluster 4:

Category	% of venues
Beach	28.5714
Deli / Bodega	5.71429
Baseball Field	5.71429
Park	4.28571
Spa	4.28571
Pier	2.85714
Pub	2.85714

Cluster 5:

Category	% of venues
Park	28.2051
Playground	10.2564
Trail	7.69231
Harbor / Marina	5.12821
Bus Stop	5.12821
Dry Cleaner	2.5641
Candy Store	2.5641

CLUSTER ANALYSIS



The number of NYC and Toronto neighborhoods in each cluster

CONCLUSION

- The results showed that there are venue categories that are common in some of the clusters than the others
- The most common venue categories differ from one cluster to the other.
- There are rare venue categories in each city
- Some venues are more popular than other venues in the same cluster.
- K-Means clustering algorithm can be effectively applied to find similarity or dissimilarity of neighborhoods.