

IBM Data Science Final Capstone

Battle of Neighborhoods

1. Introduction and Business Problem

Introduction:

People seeking better career opportunities migrate from one city to another city. It is often difficult for them to find a neighborhood which is nearby to their new work location and suitable for their family in the new city. Most of the cases people love their current neighborhood because of its amenities, venues and tourist spots, so they look for neighborhood which is more similar to their current neighborhood.

Business Problem:

Today migration from one place to another place for better career opportunities is a most common phenomenon. Such a drastic change in the neighborhood from their currently habituated neighborhood is usually difficult for an individual and his family to adjust. On a positive note every city has its own unique and diverse neighborhoods which usually people like to explore. We can find aspects in which both the cities are similar based on the information like venues, tourist spots, restaurants etc. The goal is to analyze neighborhoods in cities and find similar neighborhoods to help people migrating to new cities to live comfortably as they are doing in their current city. In my case study, A person from Toronto wants to migrate to New York seeking a better career opportunity. By, Analyzing the Current Neighborhood of that person in Toronto we should be able to suggest a Similar Neighborhood for that person in New York.

2. Data Acquisition and Data Cleaning

Data Acquisition:

1. The data on Neighborhoods and Boroughs in New York City is obtained from an open source spatial data repository. Link - https://geo.nyu.edu/catalog/nyu_2451_34572.
2. The data on Neighborhoods and Borough in Toronto the information is taken from the Wikipedia list. Link - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
3. The Geo Spatial Coordinates CSV of the Toronto Neighborhoods and Borough is obtained from different dataset of Latitude and Longitude of different Postal Codes.
4. The information such as venues, restaurants, parks, hotels etc. is collected using the Foursquare API for the neighborhoods in Toronto and New York. Foursquare API provides location-based experiences with information like venues, check-ins. Foursquare API provides the venue information in the JSON format.

Data Cleaning:

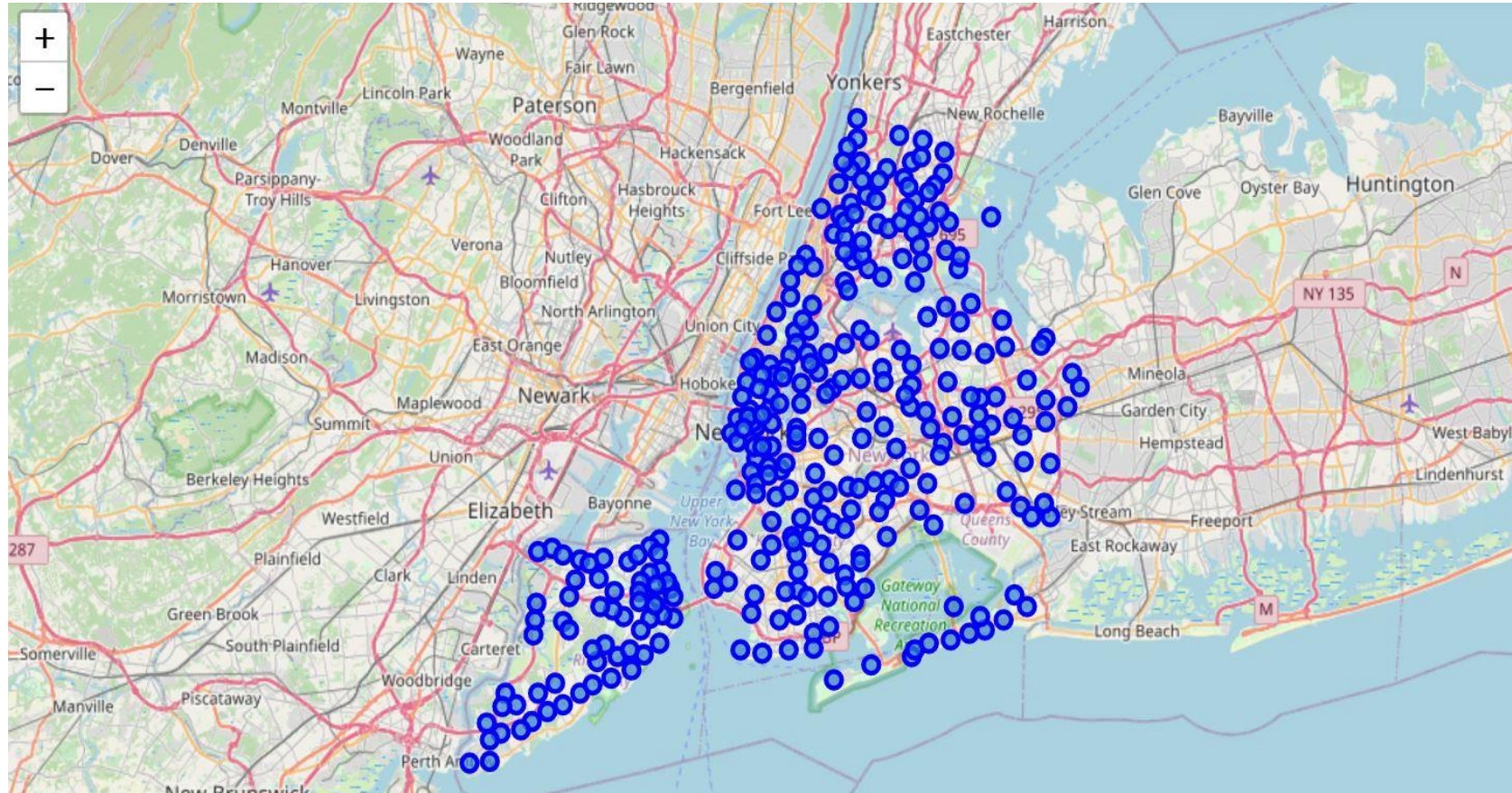
1. The data on Neighborhoods and Boroughs in New York City is obtained in the JSON Format. So pandas.io.json is used to transform this JSON data into Pandas Dataframe
2. The data of Neighborhoods of Toronto is obtained by Web Scraping the List of Postal Code of Canada : M using Beautiful Soap python package. Beautiful Soap is a popular python package for web scraping and parsing HTML files. The Resultant Pandas DataFrame of Neighborhood of Toronto is created from the Wikipedia lists web page.
3. The Geo Spatial Coordinates CSV is loaded into Pandas Dataframe to obtain the latitude and longitude information of Postal Code of Borough and Neighborhoods of Toronto

4. The Pandas Dataframe on Borough and Neighborhoods of Toronto is merged with the Geo Spatial Coordinates CSV of Toronto to obtain a single Dataframe consisting Neighborhoods and Borough of Toronto and their respective Latitude and Longitude information
5. The FourSquare API venue information is obtained through API call using requests python package through FourSquare API credentials. The obtained result is JSON using pandas.io.json the JSON is transformed into a pandas Dataframe.
6. The Final Dataframe after combining the FourSquare API venue data with the Dataframe of Neighborhoods and Borough of Toronto and New York Cities Separately and renaming the Columns of the resulting Dataframe { name – Venue , lat – Venue Latitude , lng – Venue Longitude and categories – Venue Category }. The Dataframe obtained for Newyork after keeping only Manhattan Borough and The Dataframe obtained for Toronto after keeping only Downtown Toronto Borough

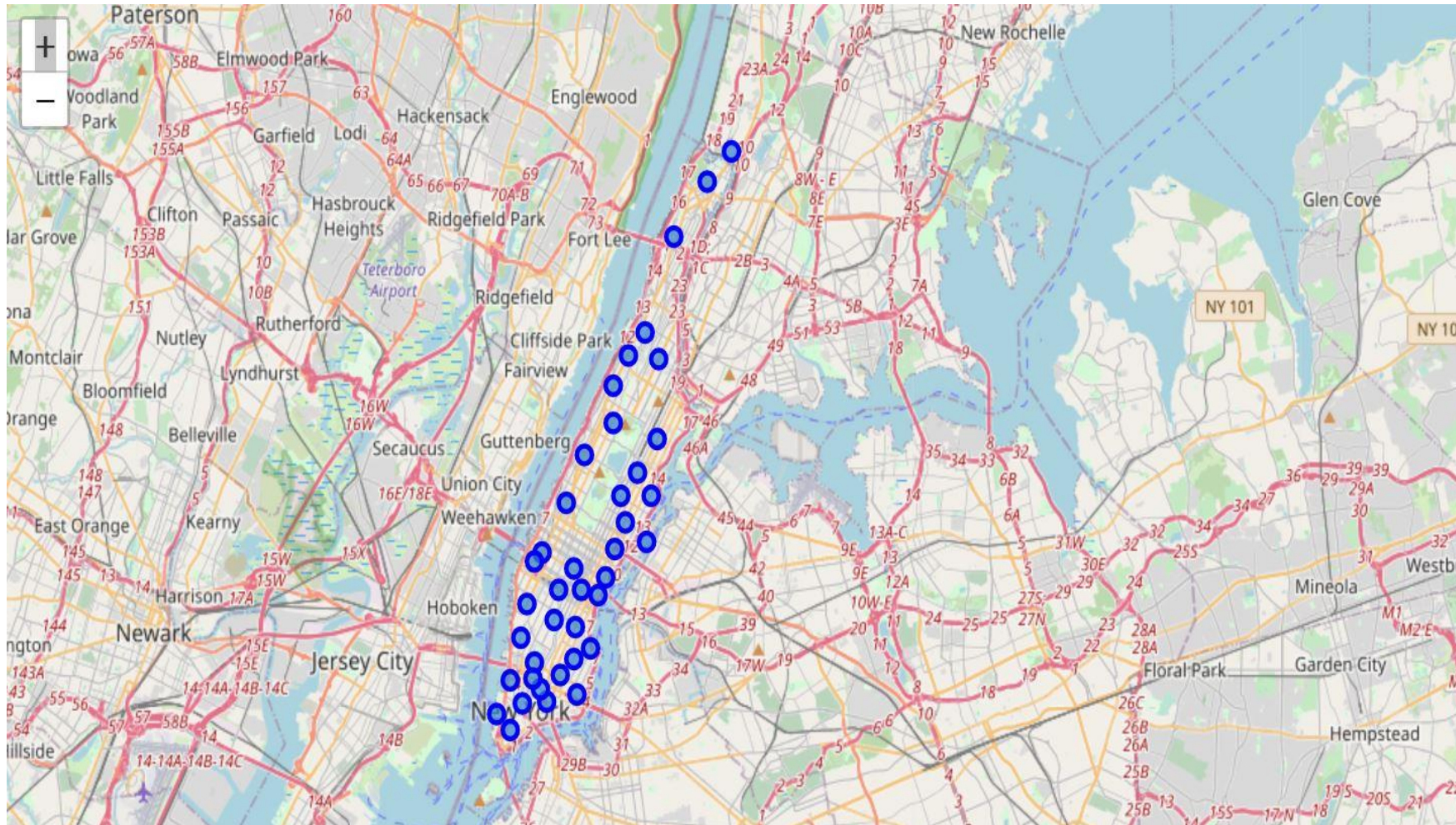
3. Methodology

Data Analysis:

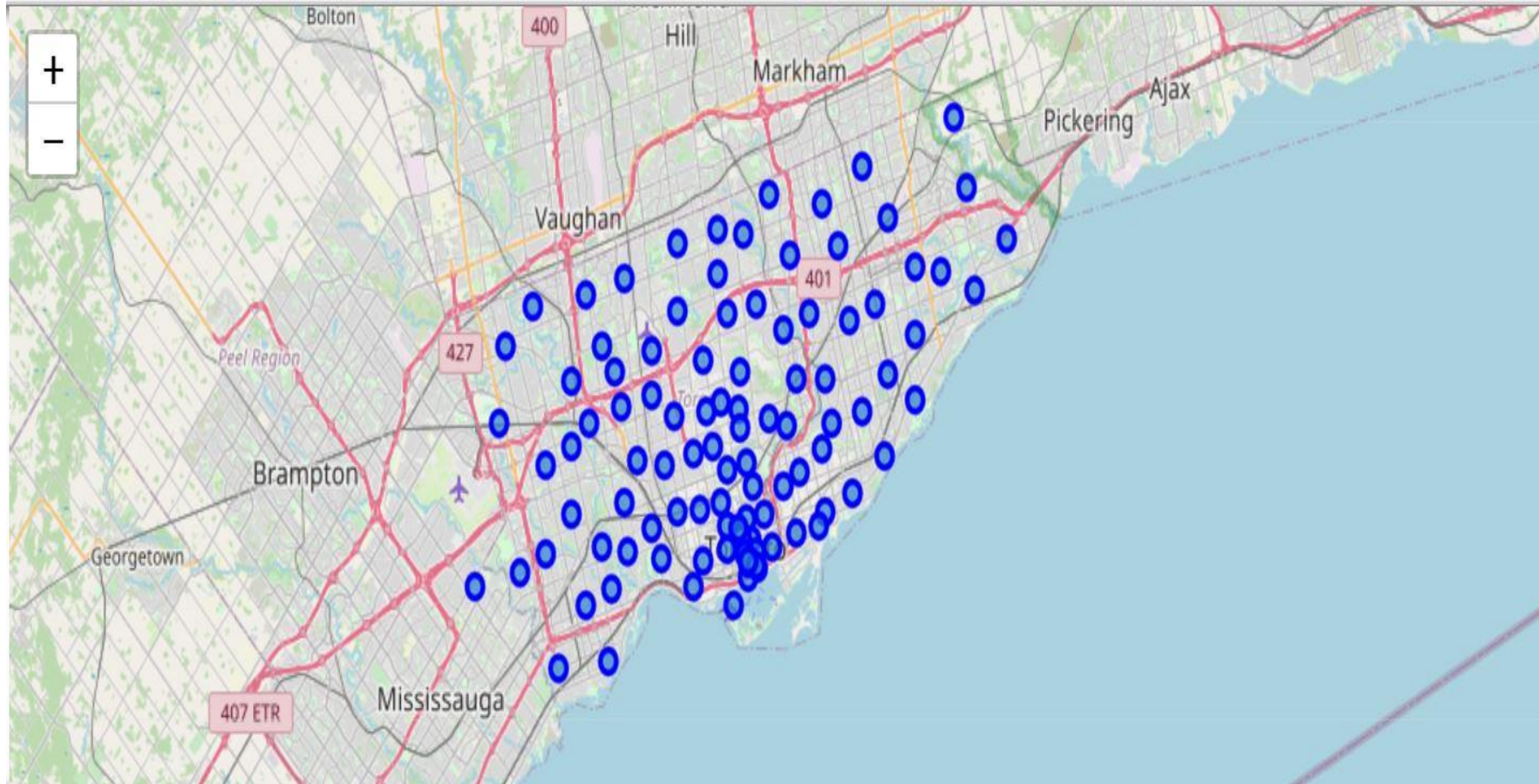
Separate Data Analysis is performed for the Venue Data Extracted for the Neighborhoods of Toronto and Neighborhoods of New York. The areas Highlighted in Blue in the Map are the Neighborhoods of Newyork that is plotted using Folium from the Newyork Borough Neighborhoods Dataframe



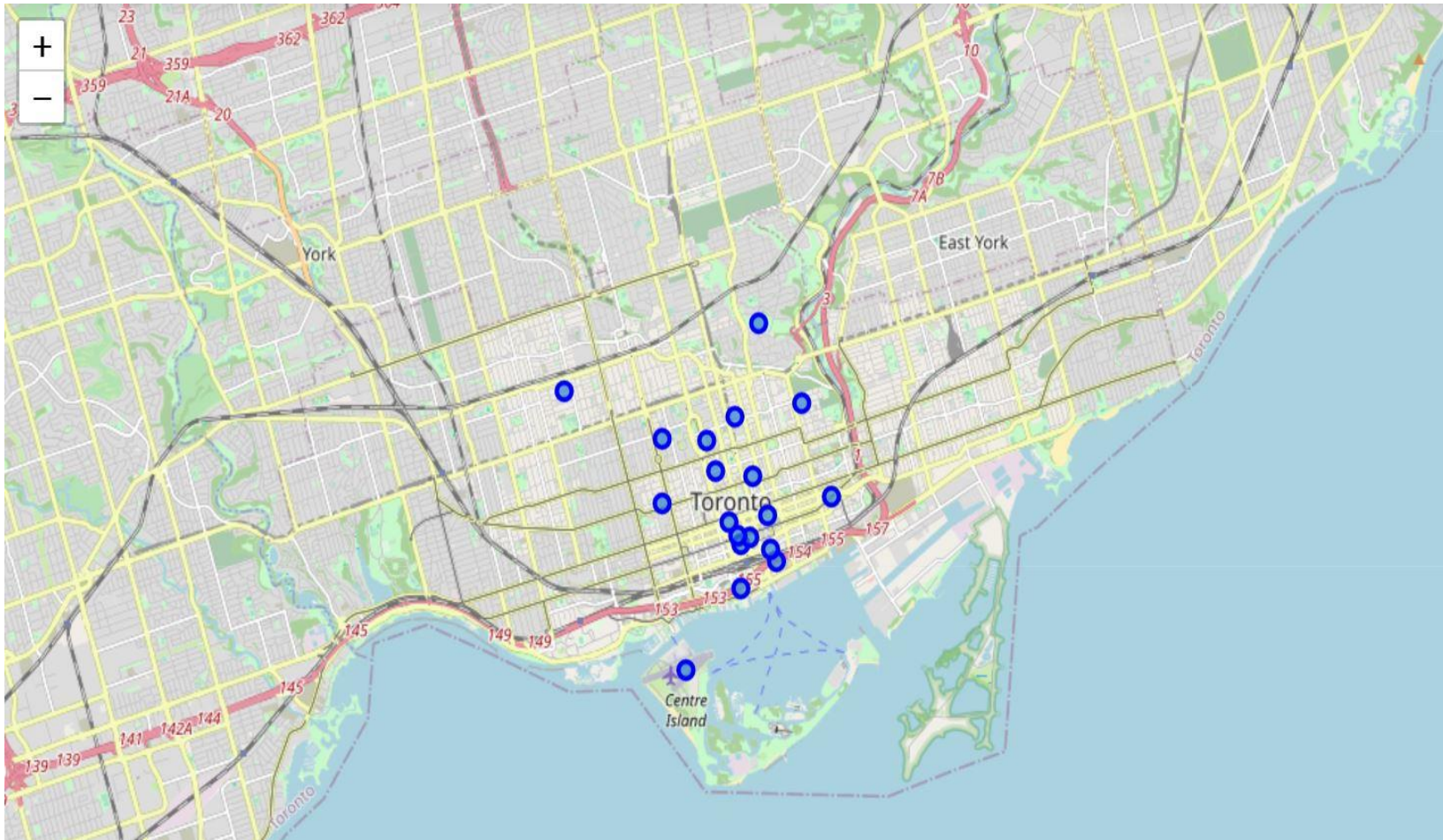
A Dataframe of Borough Manhattan is created with it's neighborhoods containing latitude and longitude. The Manhattan Neighborhoods are separately viewed using Folium for better understanding the Manhattan Neighborhoods



The areas highlighted in Blue in the Map are the Neighborhoods of Toronto that is plotted using Folium from the Toronto Borough Neighborhoods Data



A Dataframe of Borough Downtown Toronto is created with it's neighborhoods containing latitude and longitude. The Downtown Toronto Neighborhoods are separately viewed using Folium for better understanding the Downtown Toronto Neighborhoods



The response of the Foursquare API venues information for the Neighborhoods of Downtown Toronto result after converting into Dataframe, The Venues are Categorized into Venue Categories using data obtained from the Foursquare API.The number of Venues returned in each neighborhood is counted for Downtown Toronto and Manhattan

Manhattan

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Battery Park City	55	55	55	55	55	55
Carnegie Hill	87	87	87	87	87	87
Central Harlem	45	45	45	45	45	45
Chelsea	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Civic Center	88	88	88	88	88	88
Clinton	100	100	100	100	100	100
East Harlem	43	43	43	43	43	43
East Village	100	100	100	100	100	100
Financial District	100	100	100	100	100	100
Flatiron	91	91	91	91	91	91
Gramercy	66	66	66	66	66	66
Greenwich Village	100	100	100	100	100	100

Downtown Toronto

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Berczy Park	55	55	55	55	55	55
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	16	16	16	16	16	16
Central Bay Street	60	60	60	60	60	60
Christie	16	16	16	16	16	16
Church and Wellesley	77	77	77	77	77	77
Commerce Court, Victoria Hotel	100	100	100	100	100	100
First Canadian Place, Underground city	100	100	100	100	100	100
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
Kensington Market, Chinatown, Grange Park	55	55	55	55	55	55
Queen's Park, Ontario Provincial Government	39	39	39	39	39	39
Regent Park, Harbourfront	47	47	47	47	47	47

Modelling:

The Final Dataframe of Manhattan containing neighborhoods with latitude and longitude information can be used to find the venues surrounding each neighborhoods within 500 meters radius by requesting the data from FourSquare API. The JSON File returned from the Foursquare API is converted into pandas Dataframe.

Preprocessing:

This Dataframe contains the venues information along with venue categories to corresponding neighborhoods. One hot encoding should be performed. The Venue Category that is present for a neighborhood is assigned 1 and else is assigned 0. Resultant Dataframe of Manhattan Neighborhoods venue data. Resultant Dataframe of Downtown Toronto Neighborhoods venue data.

The Venue data after one hot encoding is grouped by Neighborhoods and the mean of the frequency of occurrence of venues in each venue category for the Neighborhoods is calculated. The Frequency of occurrence of each venue category for the Manhattan. The Frequency of occurrence of each venue category for the Downtown Toronto

Finally, a Dataframe of 10 common venues is created for the neighborhoods in Manhattan and 10 common venues for the neighborhoods in Downtown Toronto

The Similarity between the Manhattan and Downtown Toronto Neighborhoods will be found using clustering the Neighborhoods that are similar using K-means clustering which is an unsupervised machine learning algorithm. K-means clustering uses a predefined cluster size for clustering. The Cluster Size of 5 shall be used to cluster the Neighborhoods of Manhattan and Downtown Toronto into 5 clusters. Based on the clusters of neighborhoods obtained from the similar venues the neighborhood from Manhattan can be shortlisted based on the amenities, venues, tourist spots etc. which is similar to the neighborhood current city Downtown Toronto.

4. Results

The resultant 5 clusters obtained from K-means clustering. Examining each cluster of Manhattan Neighborhood and Downtown Toronto Neighborhood.

Cluster 1 Manhattan

	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
5	Manhattanville	Coffee Shop	Seafood Restaurant	Italian Restaurant	Mexican Restaurant	Park	Deli / Bodega	Sushi Restaurant	Lounge	Boutique	Supermarket
8	Upper East Side	Italian Restaurant	Bakery	Exhibit	Gym / Fitness Center	Spa	Hotel	Juice Bar	Pizza Place	Yoga Studio	Wine Shop
9	Yorkville	Coffee Shop	Italian Restaurant	Gym	Bar	Deli / Bodega	Sushi Restaurant	Mexican Restaurant	Japanese Restaurant	Diner	Wine Shop
10	Lenox Hill	Italian Restaurant	Coffee Shop	Pizza Place	Café	Sushi Restaurant	Cocktail Bar	Burger Joint	Gym / Fitness Center	Gym	Mexican Restaurant
12	Upper West Side	Italian Restaurant	Wine Bar	Coffee Shop	Bakery	Bar	Dessert Shop	American Restaurant	Seafood Restaurant	Ice Cream Shop	Bagel Shop
13	Lincoln Square	Plaza	Gym / Fitness Center	Café	Italian Restaurant	Concert Hall	Performing Arts Venue	Theater	American Restaurant	Wine Shop	Indie Movie Theater
18	Greenwich Village	Italian Restaurant	Café	Gym	Coffee Shop	Sushi Restaurant	Wine Bar	Bakery	Comedy Club	Ice Cream Shop	Jazz Club
19	East Village	Pizza Place	Coffee Shop	Bar	Japanese Restaurant	Juice Bar	Mexican Restaurant	Dessert Shop	Cocktail Bar	Gym / Fitness Center	Grocery Store
21	Tribeca	Italian Restaurant	Park	Wine Bar	Café	Spa	American Restaurant	Art Gallery	Steakhouse	Skate Park	Scenic Lookout

Cluster 1 Downtown Toronto

	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
14	Downtown Toronto	0	Airport Service	Airport Lounge	Airport Terminal	Harbor / Marina	Boat or Ferry	Boutique	Coffee Shop	Sculpture Garden	Bar	Airport Gate

Cluster 2 Manhattan

	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
37	Stuyvesant Town	Park	Gas Station	Baseball Field	Gym / Fitness Center	German Restaurant	Harbor / Marina	Cocktail Bar	Coffee Shop	Heliport	Farmers Market

Cluster 2 Downtown Toronto

	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
1	Downtown Toronto	1	Coffee Shop	Restaurant	Bakery	Italian Restaurant	Café	Pub	Pizza Place	Grocery Store	Playground	Indian Restaurant
2	Downtown Toronto	1	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Restaurant	Yoga Studio	Pub	Men's Store	Mediterranean Restaurant	Hotel	Smoke Shop
4	Downtown Toronto	1	Clothing Store	Coffee Shop	Café	Restaurant	Bubble Tea Shop	Middle Eastern Restaurant	Japanese Restaurant	Cosmetics Shop	Hotel	Ramen Restaurant
5	Downtown Toronto	1	Coffee Shop	Café	American Restaurant	Gastropub	Cocktail Bar	Lingerie Store	Cosmetics Shop	Clothing Store	Seafood Restaurant	Italian Restaurant
6	Downtown Toronto	1	Coffee Shop	Cocktail Bar	Bakery	Beer Bar	Café	Seafood Restaurant	Cheese Shop	Restaurant	Shopping Mall	Cosmetics Shop

Cluster 3 Manhattan

	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
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Coffee Shop	Spa	American Restaurant	Salon / Barbershop	Optical Shop	Bar	Dim Sum Restaurant
2	Washington Heights	Café	Bakery	Mobile Phone Shop	Spanish Restaurant	Pizza Place	Grocery Store	Chinese Restaurant	Latin American Restaurant	Tapas Restaurant	New American Restaurant
3	Inwood	Mexican Restaurant	Bakery	Café	Lounge	Restaurant	Pizza Place	Park	Chinese Restaurant	American Restaurant	Frozen Yogurt Shop
4	Hamilton Heights	Pizza Place	Deli / Bodega	Café	Coffee Shop	Mexican Restaurant	Yoga Studio	Cocktail Bar	Indian Restaurant	Sushi Restaurant	Sandwich Place
6	Central Harlem	Chinese Restaurant	Art Gallery	Bar	Fried Chicken Joint	French Restaurant	Seafood Restaurant	Gym / Fitness Center	African Restaurant	American Restaurant	Cycle Studio
7	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Pizza Place	Steakhouse	Deli / Bodega	Latin American Restaurant	Taco Place	Cocktail Bar	Beer Bar
20	Lower East Side	Chinese Restaurant	Park	Art Gallery	Pharmacy	Café	Cocktail Bar	Coffee Shop	Japanese Restaurant	Juice Bar	Latin American Restaurant
22	Little Italy	Chinese Restaurant	Italian Restaurant	Spa	Bubble Tea Shop	Hotel	Mediterranean Restaurant	Café	Pizza Place	Thai Restaurant	Bakery
25	Manhattan Valley	Yoga Studio	Pizza Place	Bar	Thai Restaurant	Chinese Restaurant	Coffee Shop	Health Food Store	Peruvian Restaurant	Park	Noodle House

Cluster 3 Downtown Toronto

	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	Downtown Toronto	2	Park	Trail	Playground	Women's Store	Cosmetics Shop	Distribution Center	Discount Store	Diner	Dessert Shop	Department Store

Cluster 4 Manhattan

	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	Marble Hill	Sandwich Place	Gym	Coffee Shop	Yoga Studio	Deli / Bodega	Supplement Shop	Steakhouse	Seafood Restaurant	Pizza Place	Department Store

Cluster 4 Downtown Toronto

	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
17	Downtown Toronto	3	Grocery Store	Café	Park	Italian Restaurant	Restaurant	Candy Store	Baby Store	Diner	Nightclub	Coffee Shop

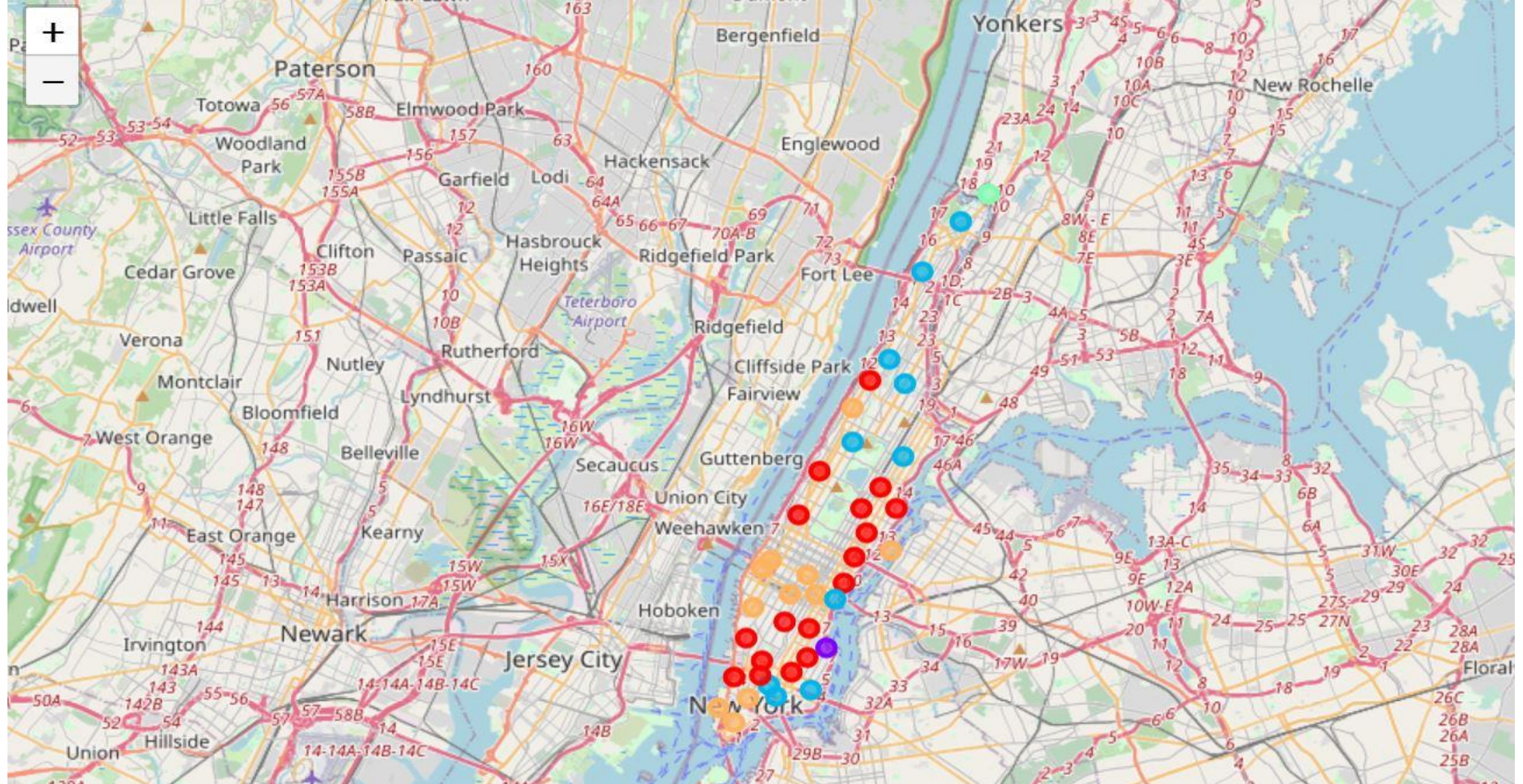
Cluster 5 Downtown Toronto

	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
3	Downtown Toronto	4	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Theater	Café	Beer Store	Farmers Market	Hotel
7	Downtown Toronto	4	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Salad Place	Ice Cream Shop	Bubble Tea Shop	Burger Joint	Bar	Thai Restaurant
18	Downtown Toronto	4	Coffee Shop	Sushi Restaurant	Diner	Yoga Studio	Mexican Restaurant	Burger Joint	Burrito Place	Café	Park	College Auditorium

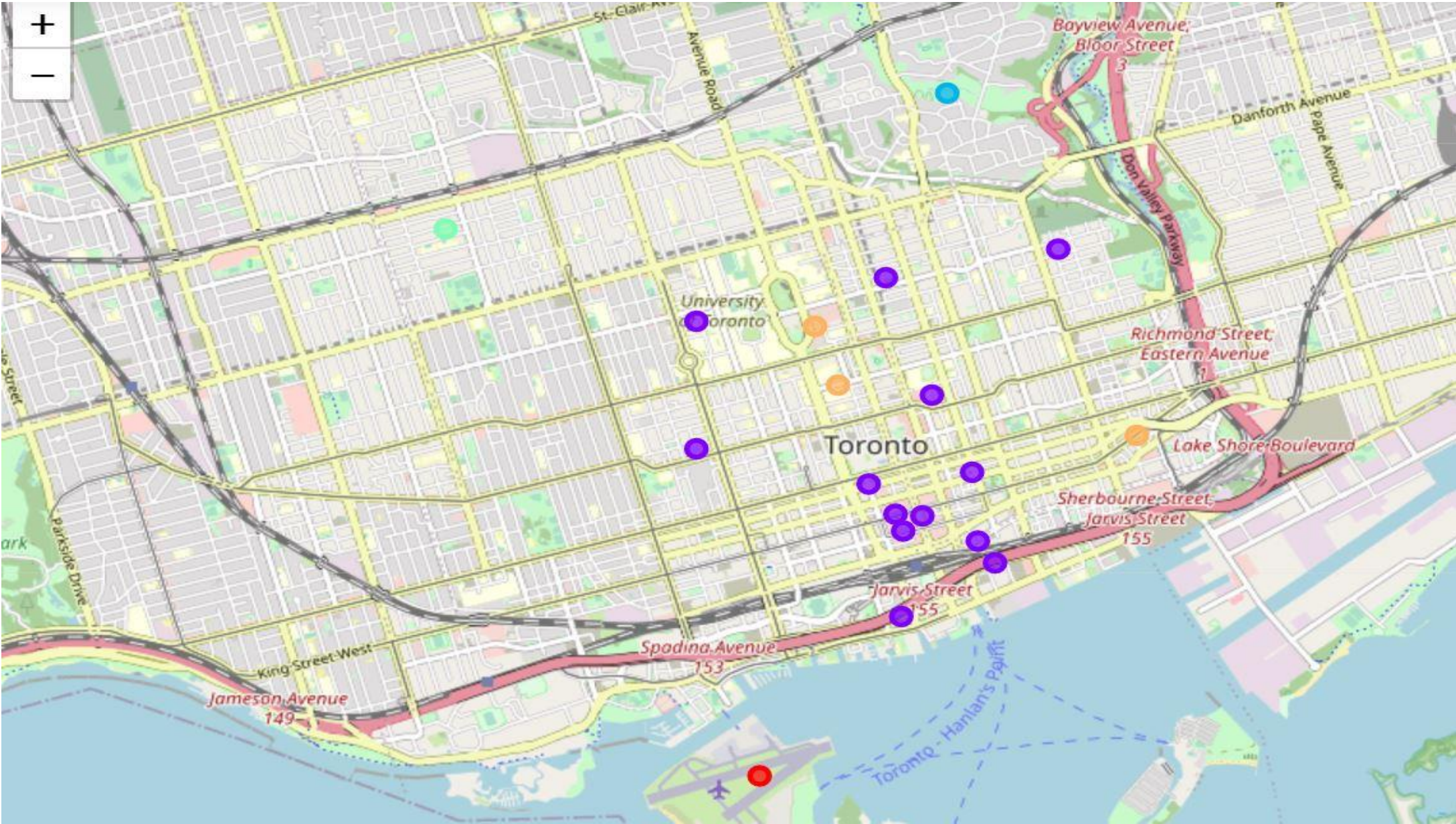
Cluster 5 Manhattan

	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
11	Roosevelt Island	Park	Deli / Bodega	Coffee Shop	Sandwich Place	Greek Restaurant	Liquor Store	Outdoors & Recreation	Dog Run	Scenic Lookout	Gym / Fitness Center
14	Clinton	Theater	Gym / Fitness Center	Coffee Shop	Wine Shop	Hotel	Gym	Italian Restaurant	Sandwich Place	Spa	Pizza Place
15	Midtown	Hotel	Clothing Store	Coffee Shop	Theater	Pizza Place	Cuban Restaurant	Spa	Tailor Shop	Pilates Studio	American Restaurant
16	Murray Hill	Sandwich Place	Japanese Restaurant	Coffee Shop	Burger Joint	Hotel	Mediterranean Restaurant	Chinese Restaurant	Pizza Place	Gym / Fitness Center	Bar
17	Chelsea	Art Gallery	Coffee Shop	Italian Restaurant	Ice Cream Shop	American Restaurant	Seafood Restaurant	Boutique	Market	Juice Bar	Bakery
26	Morningside Heights	Park	American Restaurant	Coffee Shop	Bookstore	Deli / Bodega	Burger Joint	Tennis Court	Pizza Place	Supermarket	Café
28	Battery Park City	Park	Hotel	Gym	Memorial Site	Wine Shop	Coffee Shop	Shopping Mall	Food Court	Burger Joint	Gourmet Shop
29	Financial District	Coffee Shop	Hotel	Salad Place	Gym / Fitness Center	Cocktail Bar	Japanese Restaurant	Falafel Restaurant	Park	Pizza Place	Café
32	Civic Center	Coffee Shop	Hotel	Spa	Yoga Studio	Gym / Fitness Center	French Restaurant	Park	Sushi Restaurant	Cocktail Bar	Café

Visualizing Clustered Neighborhood using folium for Manhattan



Visualizing Clustered Neighborhood using folium for Downtown Toronto



The Cluster 1 in Manhattan most common venues are Restaurant, Pizza place, Coffee Shop whereas the Cluster 1 in Downtown Toronto most common venue is Airport.

The Cluster 2 in Manhattan most common venue is Park and cluster 2 in Downtown Toronto most common venue is Coffee Shop.

The Cluster 3 in Manhattan most common venue is Restaurant, Pizza place and cluster 3 in Downtown Toronto most common venue is park.

The Cluster 4 in Manhattan most common venue is Sandwich place and cluster 3 in Downtown Toronto is grocery store.

The Cluster 5 in Manhattan most common venue is Park, Coffee shop, Cluster 5 in Downtown Toronto is Coffee Shop.

Each Cluster is Color coded and is plotted using Folium for Manhattan and Downtown Toronto.

5. Discussion

Based on the most common venues the Restaurants, Parks, Coffee Shop are similar in both Manhattan and Downtown Toronto. But based on the Clusters there is an another observation that Manhattan is more diverse in terms of availability of restaurants when compared to Downtown Toronto, There are Chinese restaurants, Italian Restaurants, Mexican Restaurants in Manhattan. This depicts that there are people of different cultures living in Manhattan. So, the people migrating from Downtown Toronto to Manhattan, New York will find it more similar as far the venues, Parks for children and Coffee shops are concerned and provides a great opportunity to explore the people from diverse cultures living in Manhattan. If the family is more interested in amenities like theatre, Park, Coffee Shop then Cluster 5 of Manhattan is more suitable but if they are interested to explore the diverse culture in Manhattan then Cluster 3 and Cluster 1 is more Suitable.

6. Conclusion

This project helps a person migrating to new city and is unaware of the neighborhoods. But due to the love with his current neighborhood he wants to find a neighborhood similar to his current neighborhood. Based on the venue data we could find a similar neighborhood to the current neighborhood and in addition to that suggest more neighborhood options to the person who really wants to explore a diverse culture existing in the new place. We can further expand this idea such that based on the user input we can immediately give recommendations possible suitable neighborhood based on his current neighborhood information and also based on his favorite type of restaurant and food, favorite weekend venue etc. to recommend more suitable neighborhoods during their migration to a new city.