
BATTLE OF THE NEIGHBOURHOODS

Varun Moturi

ABSTRACT

Greater Toronto Area (GTA) welcomes more than 100,000 immigrants every year. When migrating to a vast area like GTA, immigrants often find themselves in a dilemma when it comes to deciding a location to buy house. They would need to account for all the amenities in their neighbourhood including grocery stores, restaurants, movie theatres, etc. In order to help immigrants understand the neighbourhoods in GTA, I created a model that clusters similar neighbourhoods and visually represent them to the user.

1. INTRODUCTION

1.1 *Problem*

With a population of 2.7 million, Toronto is known to be the most populous city in Canada and fourth most populated city in North America. The city of Toronto is part of the Greater Toronto Area (GTA) which includes the following regions:

- Halton
- Peel
- City of Toronto
- York
- Durham

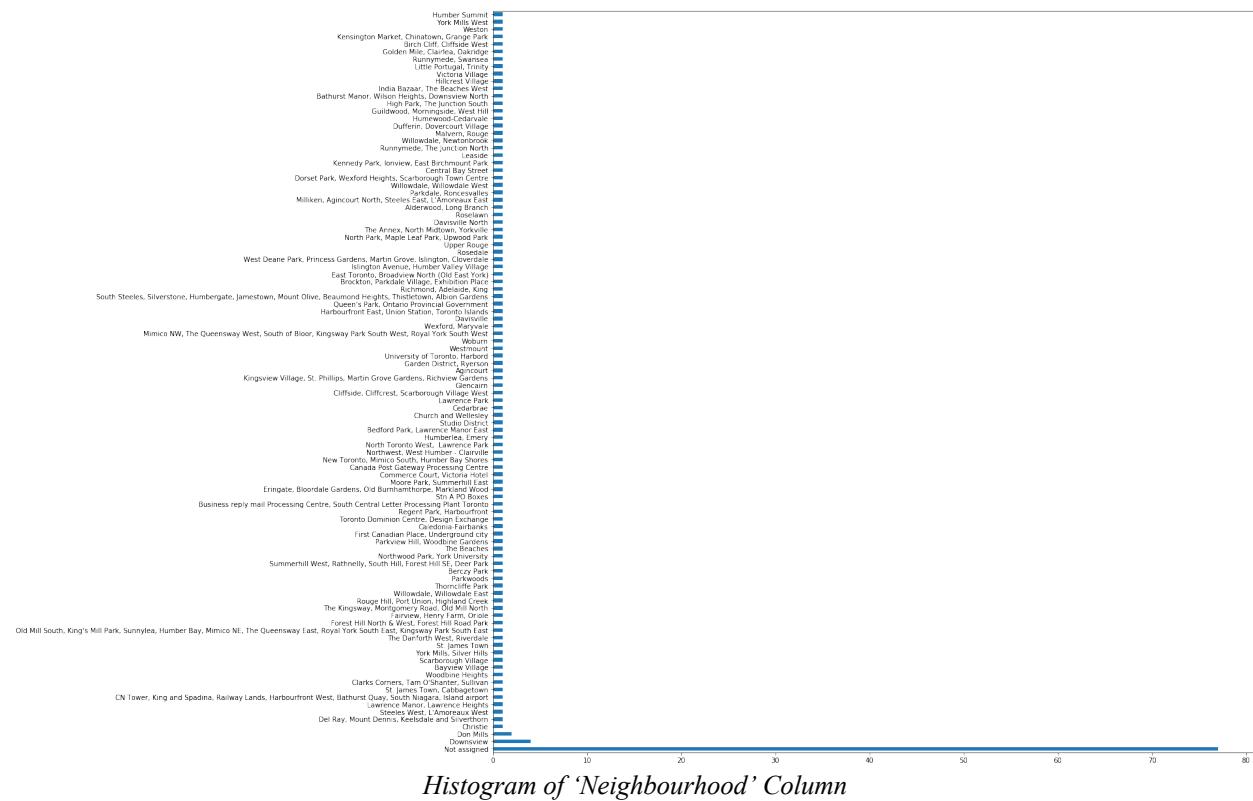
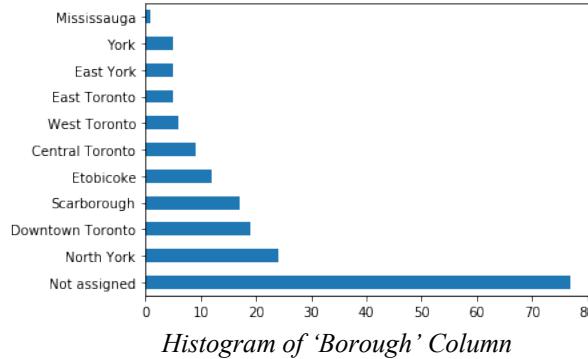
These regions are further divided into subdivisions called Boroughs. Being the most populated area in Canada, GTA welcomes more than 100,000 immigrants every year. Immigrants usually face a challenge when it comes to choosing a neighbourhood which would best suit their need. Immigrant families tend to rent a room at a random location to get a better understanding of the city before they choose to settle in a neighbourhood. This can be solved by providing the immigrants with enough knowledge about all the neighbourhoods so they can choose the one that is best suited for them.

2. DATA

The data is scraped from a publicly available dataset on Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). This dataset consists of the following attributes such as 'Postal Codes', 'Borough' and 'Neighbourhood'. The dataset contains all the subdivisions of GTA along with its neighbourhoods. Google Maps Geocoding API will be leveraged to obtain the latitude and longitude of each neighbourhood. As described in the problem description, venues surrounding each neighbourhood will be retrieved using the Foursquare API. Neighbourhoods will then be clustered based on the venues surrounding them. Each Neighbourhood will be displayed with its most popular venues. A map of GTA will be displayed with similar neighbourhoods giving users more alternatives.

3. DATA CLEANING

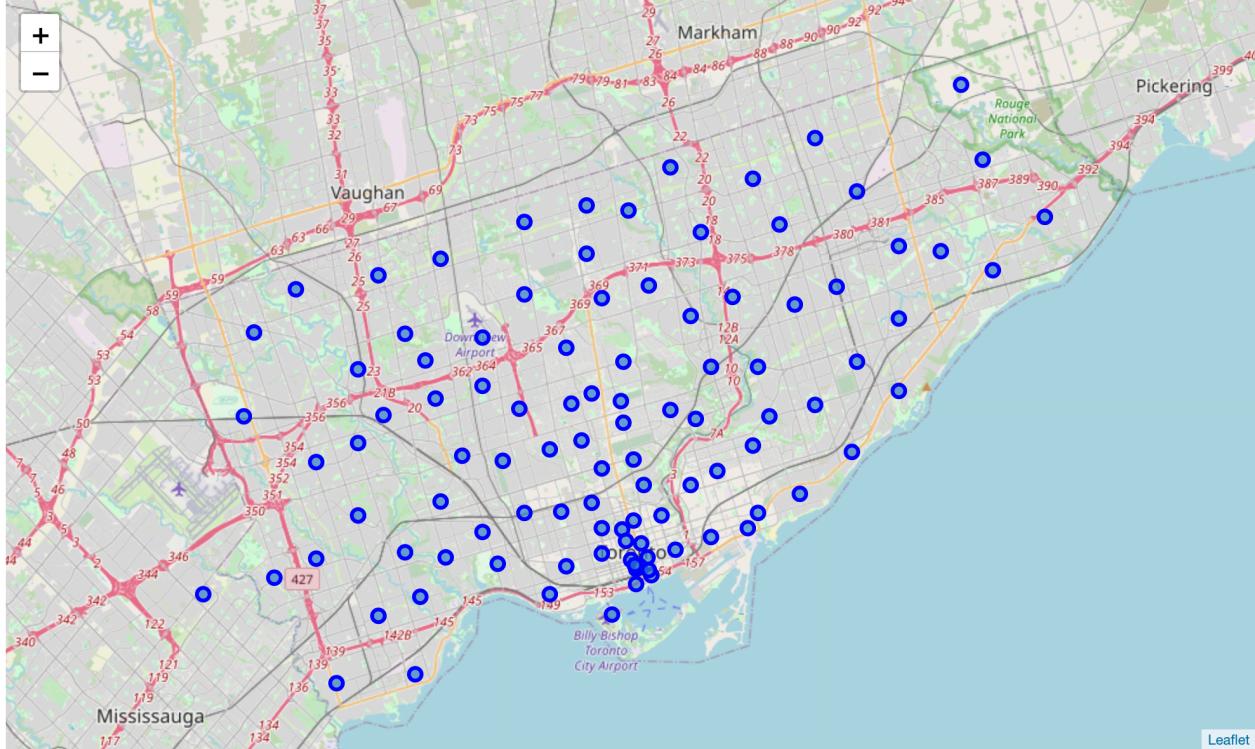
The dataset retrieved a total of 180 postal codes from the online source. The dataset consisted of many missing values represented as ‘Not Assigned’.



As seen above, there were approximately 77 Boroughs and Neighborhoods that were not assigned. To make the data more reliable, I dropped all the rows where ‘Borough’ was ‘Not Assigned’ and assigned ‘Borough’ as ‘Neighborhood’ if ‘Neighborhood’ was ‘Not Assigned’. This reduced the dataset from 180 rows to 103 rows.

4. DATA CLUSTERING

After retrieving the latitudes and longitudes of each neighborhood, the ‘folium’ library was used to plot them on a map.



As shown above, the neighborhoods in the dataset are well dispersed throughout the Greater Toronto Area (GTA). The Foursquare API was used to get the points of interest (venues) which are located within 500 m of each neighbourhood. A total of 271 venue categories and 2124 venues were discovered through the Foursquare API.

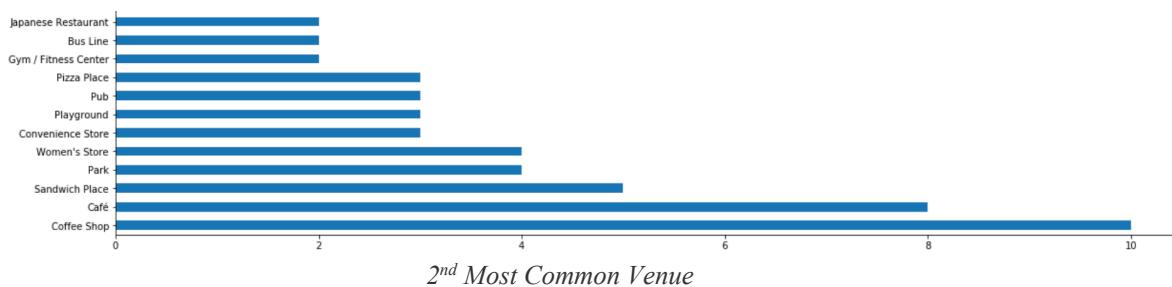
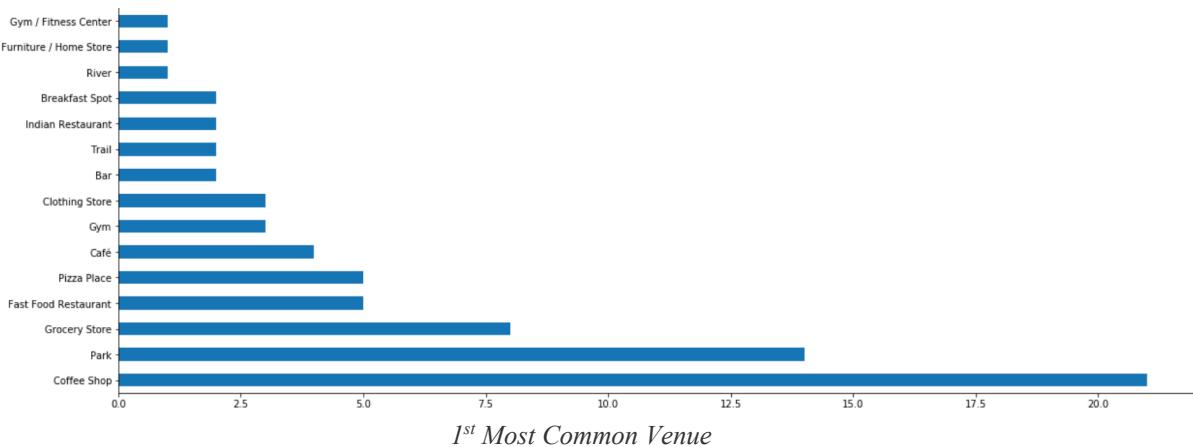
```
print(toronto_venues.shape)
toronto_venues.head()
(2124, 7)
```

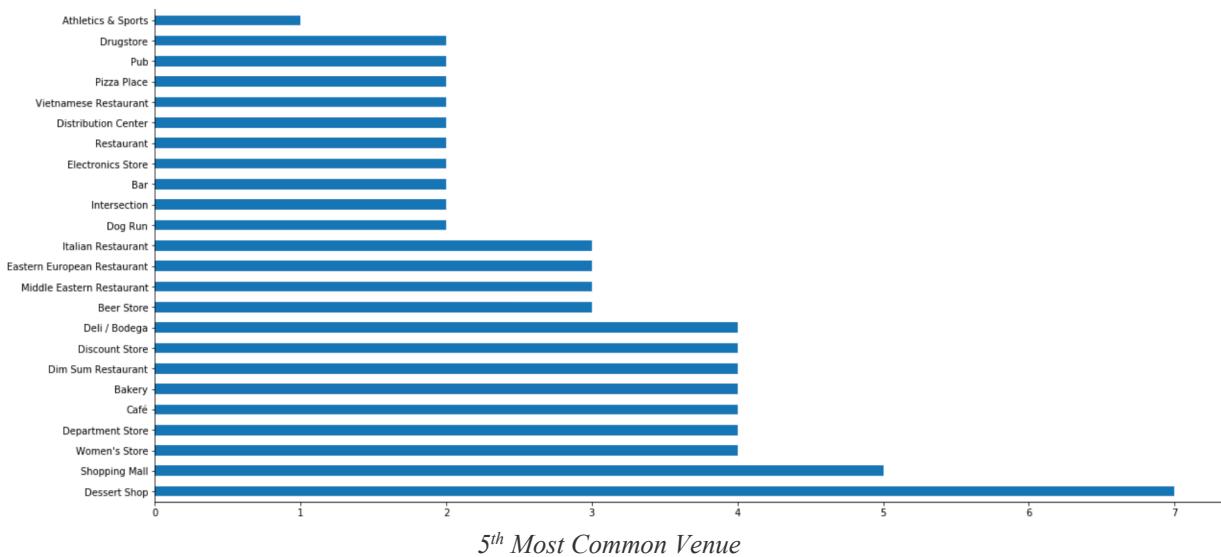
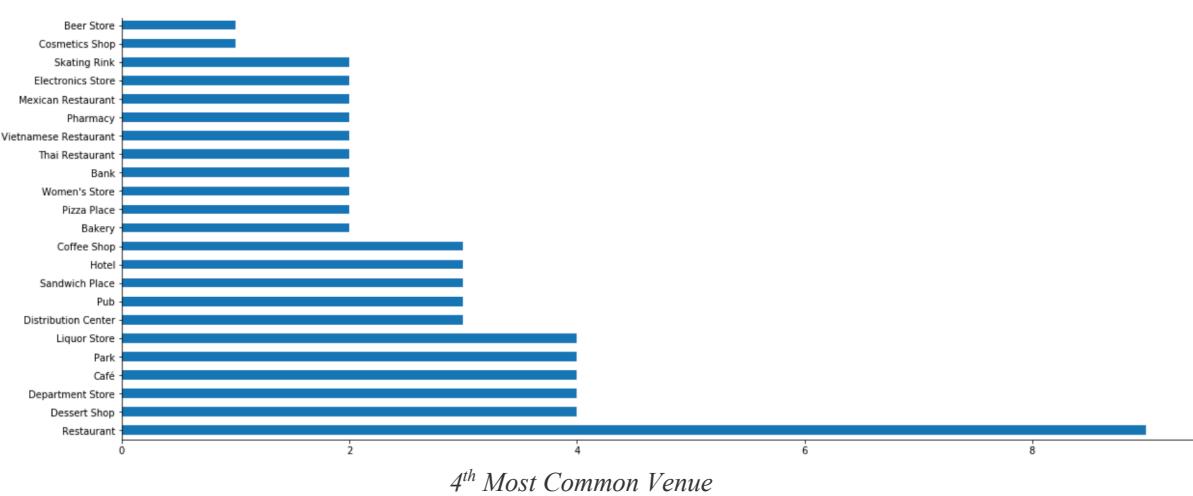
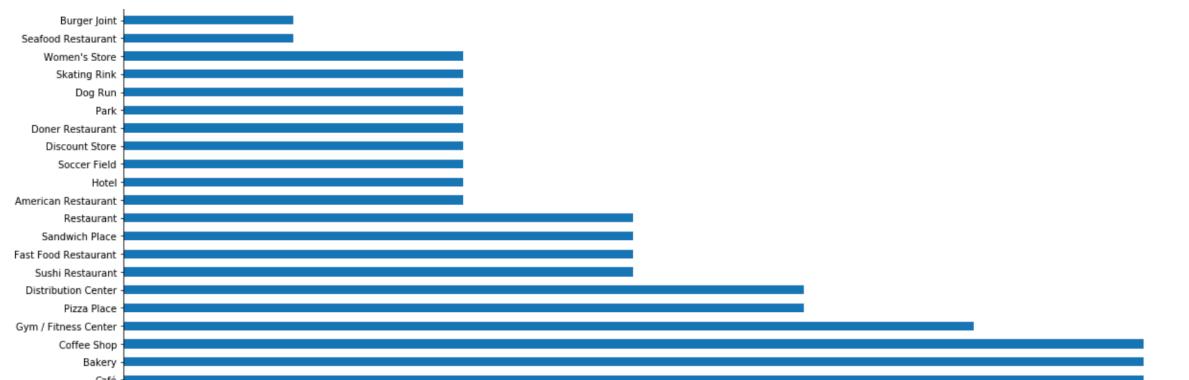
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
5	Victoria Village	43.725882	-79.315572	The Frig	43.727051	-79.317418	French Restaurant
6	Regent Park, Harbourfront	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
7	Regent Park, Harbourfront	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
8	Regent Park, Harbourfront	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
9	Regent Park, Harbourfront	43.654260	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
10	Regent Park, Harbourfront	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

As the new dataset consisted of categorical variables, ‘one-hot encoding’ was performed to convert the categorical data to numerical data. The data was processed and grouped based the most occurring venue in each neighbourhood. In my case, I listed the top 10 most seen venues in each neighbourhood.

	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	Agincourt	Breakfast Spot	Lounge	Latin American Restaurant	Skating Rink	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run
1	Alderwood, Long Branch	Pizza Place	Gym	Coffee Shop	Skating Rink	Pub	Sandwich Place	Pool	Women's Store	Dim Sum Restaurant	Deli / Bodega
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Diner	Pharmacy	Deli / Bodega	Fried Chicken Joint	Middle Eastern Restaurant	Shopping Mall	Supermarket	Pizza Place
3	Bayview Village	Café	Japanese Restaurant	Bank	Chinese Restaurant	Women's Store	Department Store	Dim Sum Restaurant	Diner	Discount Store	Distribution Center
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Sandwich Place	Coffee Shop	Thai Restaurant	Indian Restaurant	American Restaurant	Sushi Restaurant	Hobby Shop	Pizza Place	Pharmacy

Histograms were created for all the columns in this dataset to visualize the top venues in each neighborhood



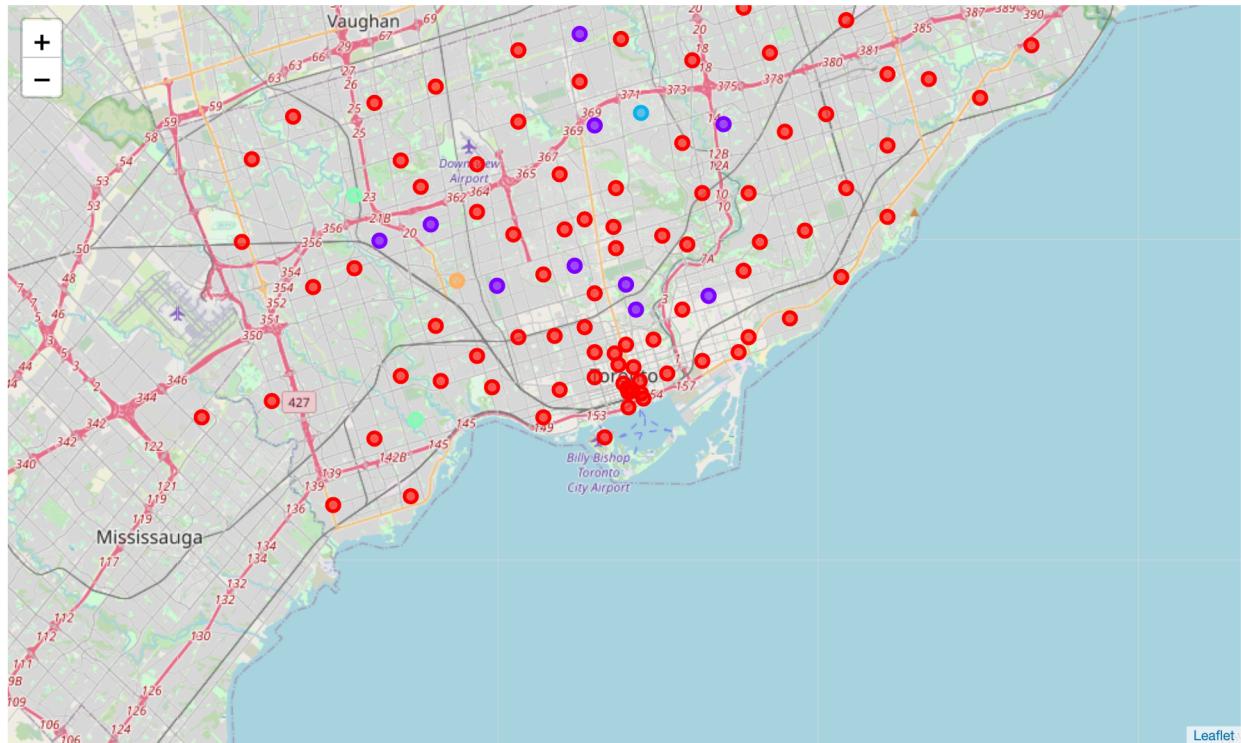


As seen above, 'Coffee shop' ranked 1st and 2nd among the most common venues followed by 'Café', 'Restaurant' and 'Desert Shop'. Looking at the data, the user can understand that most of the neighbourhoods have coffee shops near them.

Once the top venues for each neighbourhood were found, the ‘k-nearest neighbours’ algorithm was performed to cluster the neighbourhoods based the venues. This resulted in a list of cluster labels for each neighbourhood.

	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	6th Most Common Venue	7th Most Common Venue	8th Common Venue
0	M3A	North York	Parkwoods	43.753259	-79.329656	1.0	Park	Food & Drink Shop	Distribution Center	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Dis
1	M4A	North York	Victoria Village	43.725882	-79.315572	0.0	Hockey Arena	French Restaurant	Coffee Shop	Portuguese Restaurant	Women's Store	Department Store	Dessert Shop	Din Rest
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0.0	Coffee Shop	Pub	Bakery	Park	Café	Theater	Breakfast Spot	Event Sp
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	0.0	Clothing Store	Women's Store	Accessories Store	Event Space	Vietnamese Restaurant	Coffee Shop	Boutique	Miscella
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.0	Coffee Shop	Sushi Restaurant	Yoga Studio	Burrito Place	Bar	Italian Restaurant	Japanese Restaurant	Be

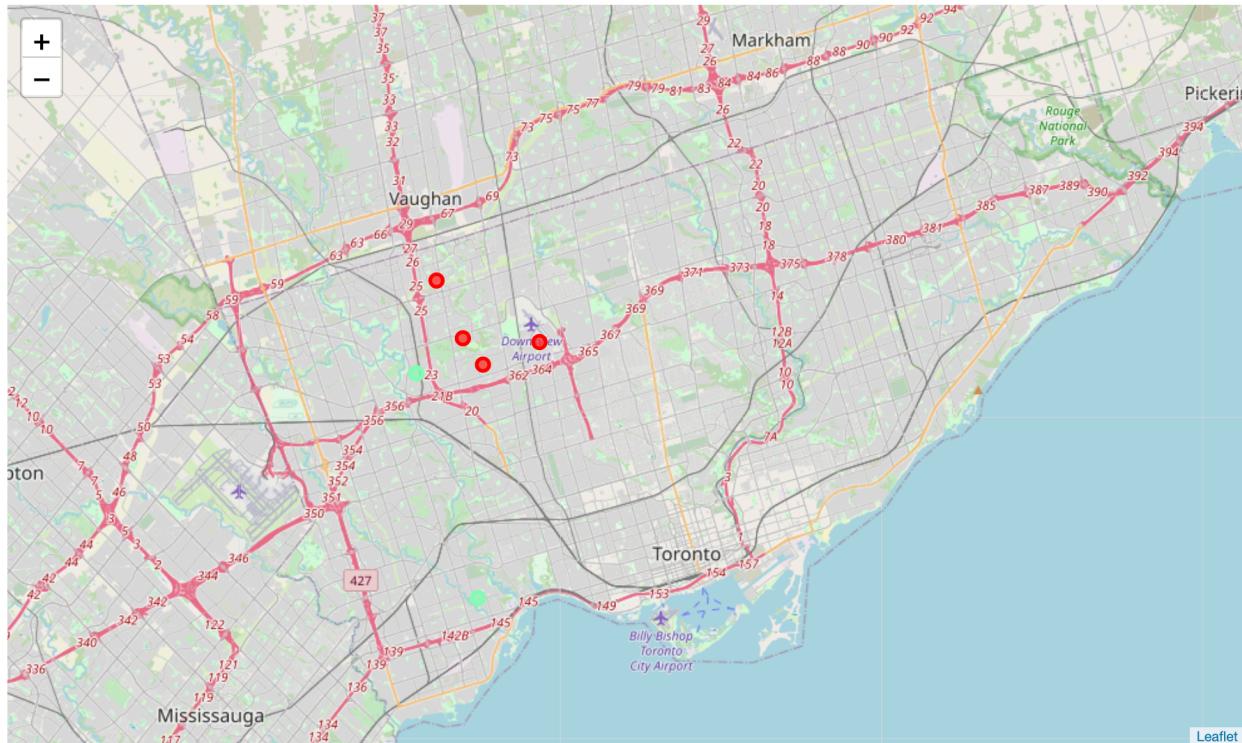
Each neighbourhood was plotted again on the map again and color-coded based on the cluster label. This visually represents all the similar neighbourhoods in GTA.



5. RESULTS

I tested the model by querying all neighborhoods with a ‘Baseball Field’. This resulted in a list of Neighborhoods plotted on a map.

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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
38	M3K	North York	Downsview	43.737473	-79.464763	0	Grocery Store	Park	Gym / Fitness Center	Liquor Store	Shopping Mall	Bank	Food Truck	Athletics & Sports
44	M3L	North York	Downsview	43.739015	-79.506944	0	Grocery Store	Park	Gym / Fitness Center	Liquor Store	Shopping Mall	Bank	Food Truck	Athletics & Sports
51	M3M	North York	Downsview	43.728496	-79.495697	0	Grocery Store	Park	Gym / Fitness Center	Liquor Store	Shopping Mall	Bank	Food Truck	Athletics & Sports
55	M9M	North York	Humberlea, Emery	43.724766	-79.532242	3	Furniture / Home Store	Baseball Field	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
58	M3N	North York	Downsview	43.761631	-79.520999	0	Grocery Store	Park	Gym / Fitness Center	Liquor Store	Shopping Mall	Bank	Food Truck	Athletics & Sports
98	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	3	Baseball Field	Women's Store	Doner Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center



6. CONCLUSION

As seen above, any immigrant moving to Toronto can enter the amenities and get a list of best suited neighborhoods and display their locations on a map. This will help the user understand the types of neighborhoods in Toronto. Moreover, this model could be applied to cities other than Toronto and help anyone who is planning to move to that city.