

# Finding Similar Neighborhoods Across Cities

Ajay Sharma

April 30, 2021

## 1. Introduction

### 1.1 Background

People travel all the time, sometimes to explore new places and other times to settle in new cities/countries/neighbourhoods. Canada welcomes tens of thousands of newcomers every year through its immigration program. And when people move to new places and make them their new home, they tend to find some similarities to overcome the feeling of separation from places and people they left behind. At least this is true for me. I have recently migrated to the Toronto area and the first thing I wanted to do was to get out and explore the neighbourhoods and find places similar to the city I lived in before moving here. The only problem is that I don't know where to start. In this assignment, I will try to find neighbourhoods that are very similar to the Brooklyn neighbourhood of New York in terms of cuisines, restaurants and other places that serve food. Brooklyn is known for its cultural, linguistic and ethnic diversity. Brooklyn has many tourist attractions like Museums, breweries, parks, historic sites and events (Dyker Heights Christmas Lights, Mermaid Parade, Color Runs, Hotdog eating competition, Labor day Carnival, to name a few and the list goes on and on). And it also offers numerous ethnic cuisines to the people living in and exploring the neighbourhoods of Brooklyn all year long. I have lived in Brooklyn for over 6 years and I always loved living in this neighbourhood. And now that I have moved to Canada, I would like to find places in Toronto that offer similar experiences, at least in terms of food.

### 1.2 Problem

In this assignment, I would like to

1. Find the top 5 venues for each neighbourhood/borough.
2. Find neighbourhoods in the Toronto area that are the closest match to neighbourhoods of Brooklyn in New York in terms of food.

### 1.3 Target Audience

This is a personal project and the target audience for this project is mostly me and my friends. But, anyone trying to explore the similarity in two geographic locations can follow this project and can also use this data to explore the city or cities based on popular food, business opportunities and/or other attributes.

## 2. Data acquisition and cleaning

### 2.1 Data sources

To solve this problem, I have downloaded the New York and Toronto Neighborhood datasets. I fetched the New York city data from the IBM cloud [here](#). The New York data contains metadata for all the boroughs including Brooklyn. The data was filtered to only use Brooklyn Borough data. The Toronto city data is not available as easily as New York city-data. I have used the Toronto neighbourhood data listed on the Wikipedia page [here](#)

## 2.2 Data cleaning and feature selection

For this assignment I need starting data to contain values for four attributes:

1. Neighbourhood
2. Borough
3. Latitude
4. Longitude

This is easy for New York city-data because the data I have downloaded from the cloud contains all the attributes I need. From New York city data I used the data only for Brooklyn Borough because I will compare the Brooklyn neighbourhood of New York with Toronto.

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

For Toronto data, I have used Python web scraping library “beautifulsoup4” library to extract information from the Wikipedia page. After scraping I have data for three attributes; Neighborhood, Borough and Zip codes. Next, I used python geocoding libraries to add Latitude and Longitude information to the data.

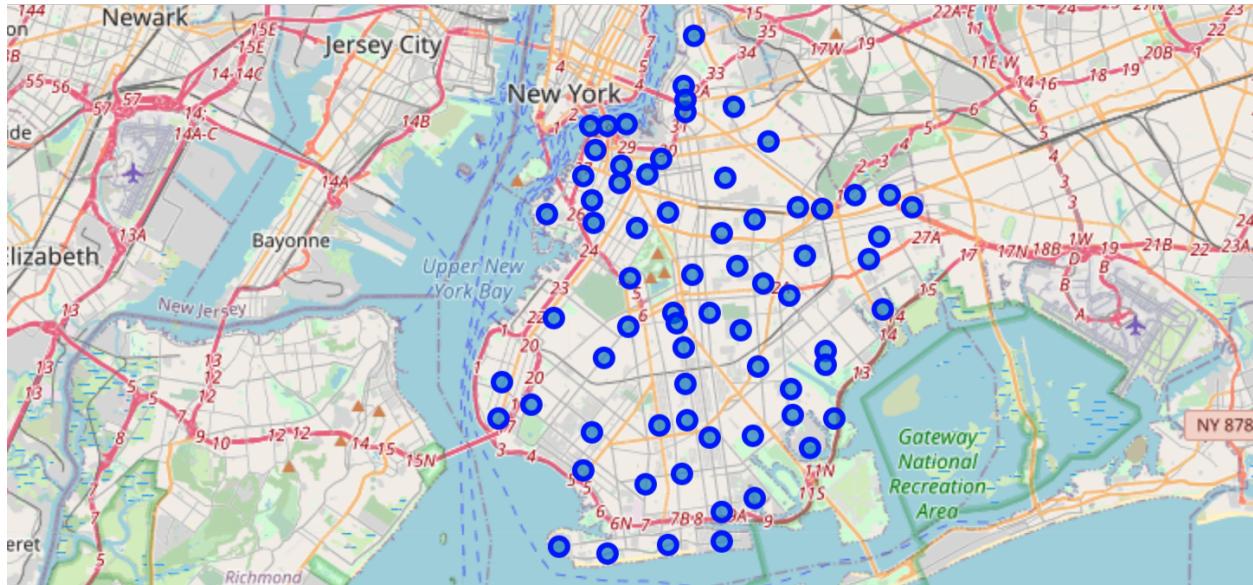
	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.7545	-79.3300
1	M4A	North York	Victoria Village	43.7276	-79.3148
2	M5A	Downtown Toronto	Regent Park,Harbourfront	43.6555	-79.3626
3	M6A	North York	Lawrence Manor,Lawrence Heights	43.7223	-79.4504
4	M7A	Queen's Park	Ontario Provincial Government	43.6641	-79.3889

Once I have all the attributes I need for Toronto data, I drop the Zip Code column because it is not needed for analysis and merge the previously compiled Brooklyn data with the Toronto data set. Our data set is now ready for further analysis.

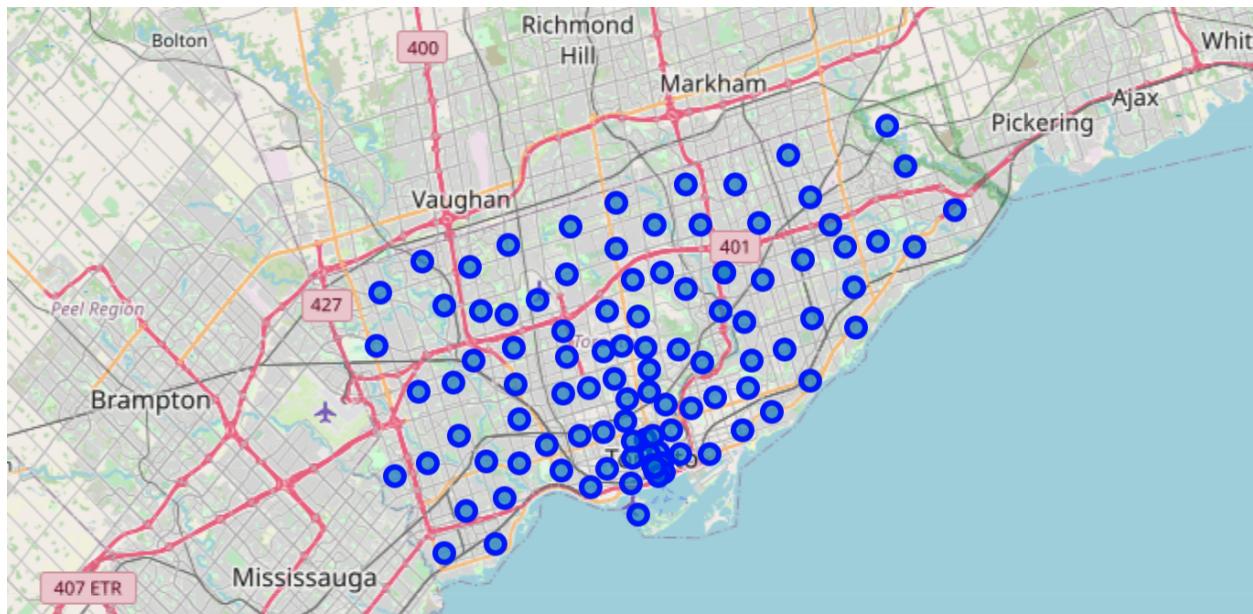
## 2.3 Data visualization

Data visualization is a quick and effective way of getting familiar with the data. It gives you an idea about data distribution, important metrics and any outliers that I can remove from the analysis. In this case, I will visualize the data on a map to get familiar with the geographic locations of Neighborhoods in our data set.

## Brooklyn Data



## Toronto Data



## 2.4 Adding Neighborhood information using Foursquare API

Foursquare API is a developer tool that allows developers to get information about venues near a neighbourhood. To use the API I have created a developer account that allows me to fetch neighbourhood data for free. Using this API I parsed the information for all the venues near the neighbourhoods in our data. And I have parsed only the venues that serve food because for this project, I am comparing the neighbourhoods by looking at the similarity in the cuisines served in local venues/restaurants. Once I have compiled the venue data, I have the data for 2,631 food venues for further analysis.

Neighborhood	Afghan Restaurant	American Restaurant	Arepas Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Belgian Restaurant	...''	Thai Restaurant	Theme Restaurant	Tibetan Restaurant	Res'.
0 Bay Ridge	0	0	0	0	0	0	1	0	0	...	0	0	0	0
1 Bay Ridge	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2 Bay Ridge	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3 Bay Ridge	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4 Bay Ridge	0	0	0	0	0	0	0	0	0	...	0	0	0	0

### 3. Data Analysis

#### 3.1 Exploring data for most popular venues

Once I have all the data and attributes I need for analysis, I will go ahead and try to find an answer to the first problem in my problem list:

**Find the top 5 venues for each neighbourhood/borough.**

To find this answer, I grouped the data by “Neighborhood” followed by sorting the data by frequency of occurrence of each venue category to find the top venues in each neighbourhood in our data. The grouping of the data gave me information for 143 neighbourhoods and the venues in the neighbourhoods that are most popular. To show what the results look like, I have listed the top 5 hits for venues in some neighbourhoods here:

```
----Agincourt ----
      venue   freq
0 Latin American Restaurant  0.5
1 Breakfast Spot            0.5
2 Afghan Restaurant          0.0
3 Noodle House              0.0
4 Russian Restaurant         0.0
```

```
----Alderwood, Long Branch----
      venue   freq
0 Coffee Shop                0.33
1 Pizza Place                0.33
2 Sandwich Place              0.33
3 Afghan Restaurant           0.00
4 Noodle House               0.00
```

```
----Bath Beach----
      venue   freq
0 Donut Shop                 0.07
1 Fast Food Restaurant       0.07
2 Italian Restaurant         0.07
3 Chinese Restaurant         0.07
4 Pizza Place                0.07
```

```
----Bathurst Manor, Wilson Heights, Downsview North----
      venue   freq
0 Deli / Bodega              0.17
1 Coffee Shop                 0.17
2 Mediterranean Restaurant    0.17
3 Middle Eastern Restaurant   0.17
4 Fried Chicken Joint         0.17
```

Next, to find top venues for each borough, I went ahead and grouped the data by “Borough” followed by sorting the data by frequency of occurrence of each venue category to find the top

venues in each borough in our data. And as we did before, I have listed some of the results below to show what top hits data look like:

----Brooklyn----

	venue	freq
0	Pizza Place	0.09
1	Coffee Shop	0.07
2	Bakery	0.05
3	Deli / Bodega	0.05
4	Italian Restaurant	0.05

----Central Toronto----

	venue	freq
0	Coffee Shop	0.13
1	Sandwich Place	0.10
2	Café	0.10
3	Italian Restaurant	0.08
4	Indian Restaurant	0.05

----Downtown Toronto----

	venue	freq
0	Coffee Shop	0.16
1	Café	0.10
2	Restaurant	0.05
3	Japanese Restaurant	0.05
4	Sushi Restaurant	0.03

----Downtown TorontoStn A PO Boxes25 The Esplanade----

	venue	freq
0	Coffee Shop	0.26
1	Restaurant	0.11
2	Café	0.05
3	Italian Restaurant	0.05
4	Deli / Bodega	0.05

Next, I have explored the data to find the most common venues by Neighborhoods and Boroughs. And the results for some of the neighbourhoods and boroughs are listed below.

## By Neighborhoods

	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	Latin American Restaurant	Breakfast Spot	Yemeni Restaurant	Fast Food Restaurant	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Distillery	Doner Restaurant
1	Alderwood,Long Branch	Sandwich Place	Coffee Shop	Pizza Place	Yemeni Restaurant	Cuban Restaurant	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Distillery
2	Bath Beach	Sushi Restaurant	Pizza Place	Bubble Tea Shop	Italian Restaurant	Fast Food Restaurant	Donut Shop	Chinese Restaurant	Cantonese Restaurant	Sandwich Place	Coffee Shop
3	Bathurst Manor,Wilson Heights,Downsview North	Deli / Bodega	Coffee Shop	Fried Chicken Joint	Pizza Place	Mediterranean Restaurant	Middle Eastern Restaurant	Yemeni Restaurant	Ethiopian Restaurant	Dessert Shop	Dim Sum Restaurant
4	Bay Ridge	Pizza Place	Italian Restaurant	American Restaurant	Greek Restaurant	Bagel Shop	Sandwich Place	Café	Sushi Restaurant	Chinese Restaurant	Snack Place

## By Boroughs

	Borough	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	Brooklyn	Pizza Place	Coffee Shop	Bakery	Deli / Bodega	Italian Restaurant	Chinese Restaurant	Mexican Restaurant	Ice Cream Shop	Café	Bagel Shop
1	Central Toronto	Coffee Shop	Sandwich Place	Café	Italian Restaurant	Restaurant	Fast Food Restaurant	American Restaurant	Dessert Shop	Indian Restaurant	Thai Restaurant
2	Downtown Toronto	Coffee Shop	Café	Restaurant	Japanese Restaurant	Italian Restaurant	Bakery	Seafood Restaurant	Sushi Restaurant	Pizza Place	American Restaurant
3	Downtown Toronto Stn A PO Boxes 25 The Esplanade	Coffee Shop	Restaurant	Deli / Bodega	Café	Japanese Restaurant	Italian Restaurant	Fast Food Restaurant	Seafood Restaurant	Breakfast Spot	Fried Chicken Joint
4	East Toronto	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Restaurant	Sandwich Place	Bakery	Fast Food Restaurant	Café	Brewery

## 3.2 Data Clustering and analysis to find similar neighbourhoods

Next, I will try to find the answer to our second question:

**Find neighbourhoods in the Toronto area that are the closest match to neighbourhoods of Brooklyn in New York in terms of food**

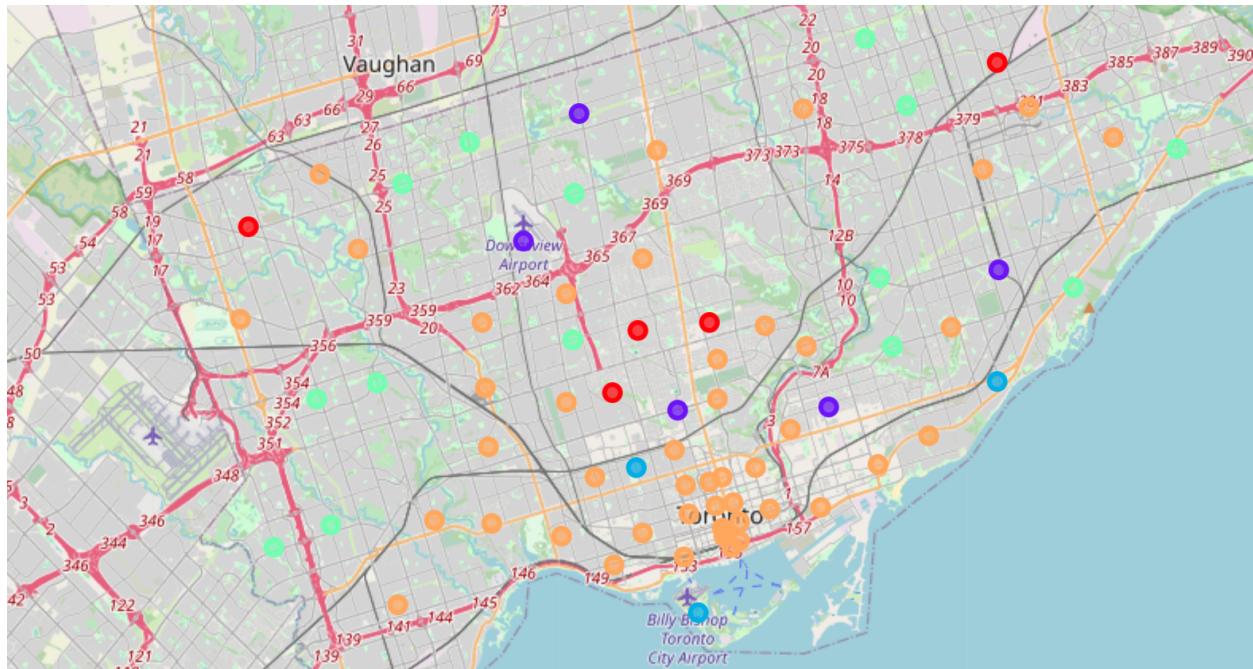
To answer this question, I have to find the neighbourhoods that are similar in terms of diversity or availability of different cuisines served in local restaurants or food outlets. I have utilized the K-Means clustering algorithm to cluster the data into different groups based on the similarity in venues in the neighbourhoods. I used the data grouped by “Neighborhood” and applied the K-Means algorithm to separate data into 5 clusters. The algorithm assigns cluster labels to each neighbourhood. Next, I extracted the cluster labels for each neighbourhood and added them to the merged Brooklyn and Toronto data I had compiled in section 2.2 above.

Data with cluster labels:

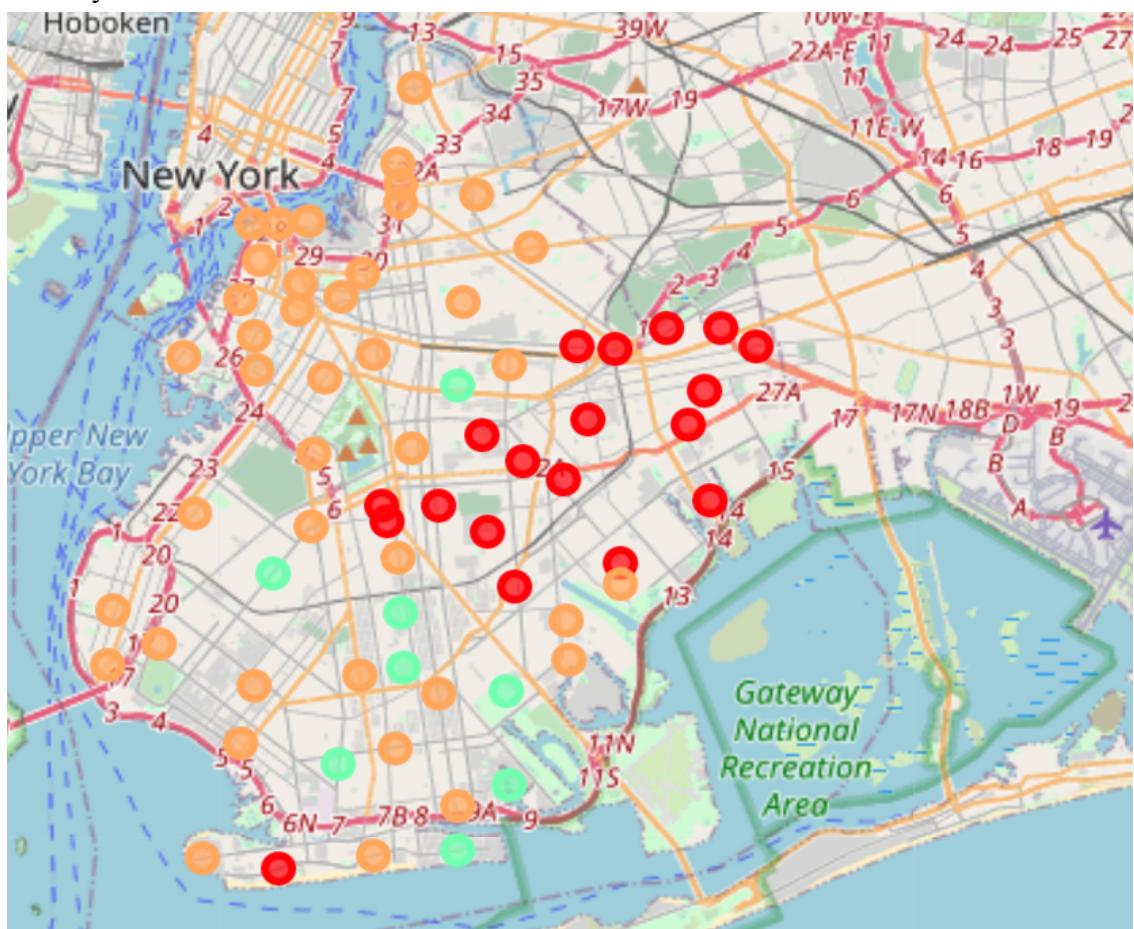
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	4.0	Pizza Place	Italian Restaurant	American Restaurant	Greek Restaurant	Bagel Shop
1	Brooklyn	Bensonhurst	40.611009	-73.995180	4.0	Chinese Restaurant	Ice Cream Shop	Italian Restaurant	Donut Shop	Sushi Restaurant
2	Brooklyn	Sunset Park	40.645103	-74.010316	4.0	Latin American Restaurant	Mexican Restaurant	Pizza Place	Bakery	Fried Chicken Joint
3	Brooklyn	Greenpoint	40.730201	-73.954241	4.0	Pizza Place	Coffee Shop	Deli / Bodega	French Restaurant	Mexican Restaurant
4	Brooklyn	Gravesend	40.595260	-73.973471	3.0	Italian Restaurant	Pizza Place	Chinese Restaurant	Bakery	Donut Shop

After the data clustering, I plotted the data on the map to visualize the clusters and neighbourhood similarities. From the results, I see that the neighbourhoods belonging to cluster 4 are very similar to each other based on the data I have used for this analysis. So, this gives me a list of 47 neighbourhoods in Toronto that are very similar to the Brooklyn neighbourhoods. Because Brooklyn and Toronto are far apart on the map, I am going to show clusters from two cities as separate maps here, so that a more zoomed view with individual clusters is visible.

## Toronto Clusters:



## Brooklyn Clusters:



## **4. Results and Discussion**

From the above analysis, as results, I have a list of:

1. Top 5 venues for each neighbourhood and borough in our data. I can use this list to explore the best cuisines in the Toronto neighbourhood.
2. I also have a list of Toronto neighbourhoods that are very similar to Brooklyn. I can use this list to explore the places and find a neighbourhood that I would like to settle in near future.

I started this project with two questions in mind. I have used this project to answer questions that may not have a larger audience or answer specific questions when it comes to exploring business opportunities. Each year tens of thousands of newcomers from around the world make Canada as their new home. And out of nature, everyone wants to explore their new home, find places to get important supplies as well as find new business opportunities as well as employment opportunities. Food is an important part of modern culture. I love to explore different ethnic cuisines and I am sure that most people do. When I moved to Canada, I did not know where to start exploring. But, I knew that I want to find neighbourhoods that are similar to my previous home and neighbourhood. And if I had a list of places that are similar to my previous home, it will give me an idea about where to start. Diversity in a neighbourhood can be determined based on the diversity of food that is served in local restaurants. And that is the basis of this project.

## **5. Conclusion**

Based on the results, we have 47 neighbourhoods that are very similar to my old home and I cannot wait to start exploring these neighbourhoods as soon as the current pandemic (COVID-19) is over. A similar analysis could be used to explore/compare any neighbourhood based on some attributes even if they are different than the ones I have used in the dataset for this project.