

# Capstone Project - New British or Old British: Toronto vs London

## Applied Data Science Capstone by IBM/Coursera

José Nilton de Santana Júnior

5<sup>th</sup> April 2021

## 1. Introduction

### 1.1 Background

Two major cities in the world. To the one who lives in it and to the ones who wants to move in. Toronto and London are known as cities full of opportunities, full of immigrants and, of course, full of everything that is fun. Restaurants (huge variety of cuisines), bus stops, hotels, coffee shops, shopping malls, stores, culture, history, all the cool things that any human being want in its life.

### 1.2 Interest

Toronto and London are extremely popular tourist and immigration destinations for people all around the world. They are diverse and multicultural and offer a wide variety of experiences that is widely sought after. In this work we tried to group the neighborhoods of London and Toronto to come up with insights to what they look like.

Help tourist and prospective immigrants is the main objective of this work. Depending on the experience and infrastructure that the neighborhoods have to offer, people can choose those two cities for vacation, to immigrate and, if they already live there, to see other neighborhoods that is equal or better to the one they lives and, perhaps, relocate within the city. All the finds will help stakeholders and clients of tourism/immigration's companies to make informed decisions and address concerns they have, including any kind of infrastructure insights in the neighborhoods.

## 2. Data Description

It was used geolocation data for both Toronto and London. As a starting point, anyone can use the postal codes in each city to find out the neighborhoods, boroughs, venues and their most popular venue categories.

### 2.1 London

Data Scrapped from: [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London)

This Wikipedia page has information about all the neighborhoods, and it was limited to London.

1. **Borough:** Name of Neighbourhood
2. **Town:** Name of borough
3. **Post\_Code:** Postal codes for London.
4. **london\_merge:** Latitude and longitude of the Neighborhoods

This Wikipedia didn't have information about the geographical co-ordinates. To solve this issue, it was used ArcGIS API.

## 2.2 Toronto

Data Scrapped  
 from: [https://en.wikipedia.org/w/index.php?title=List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M&oldid=1011037969](https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1011037969)

This Wikipedia page has information about the neighborhoods, and it was limited to Toronto. Geocoder was used to get the co-ordinates.

1. **Postal\_Code:** Postal codes for Toronto
2. **Neighbourhood:** Name of Neighborhoods in Toronto
3. **Borough:** Name of the boroughs
4. **venues\_toronto:** Latitude and longitude of the Neighborhoods.

## 2.3 ArcGIS API

ArcGIS Online lets you to connect locations, people, and data by using interactive maps. Work with smart, data-driven styles and intuitive analysis tools that deliver location intelligence.

It was used ArcGIS to get the geo locations of the neighborhoods of London. The columns below were added to the raw dataset as our data.

1. **Latitude:** Latitude for Neighbourhood
2. **Longitude:** Longitude for Neighbourhood

## 2.4 Foursquare API Data

Data about different venues in different neighborhoods of that specific borough is needed. In order to get that information it was used the "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. This way, the foursquare location platform was used as the sole data source, since all the stated required information was obtained through the API.

We connected to the Foursquare API to get information of venues inside each and all of the neighborhoods, after finding the list of neighborhoods. For each neighbourhood, we chose the radius of 500 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. **Neighbourhood:** Name of the Neighbourhood
2. **Neighbourhood Latitude:** Latitude of the Neighbourhood
3. **Neighbourhood Longitude:** Longitude of the Neighbourhood

4. **Venue** : Name of the Venue
5. **Venue Latitude**: Latitude of Venue
6. **Venue Longitude**: Longitude of Venue
7. **Venue Category**: Category of Venue

After collect all the information about Toronto and London, I had enough data to build the model. The neighborhoods were clustered together, based on similar venue categories. Then, the observations and findings were presented. With all this data, our stakeholders can make decisions and a lot of useful visualization documents can be made to show to the clients.

### 3. Methodology

The model was created using Python. It was started by importing all the required packages.

```
import pandas as pd
import requests
import geocoder (used before)
import numpy as np
import matplotlib.cm as cm
import matplotlib.colors as colors
import folium
from sklearn.cluster import KMeans
```

Package description:

- **Pandas**: Collect, clean and manipulate data in JSON and HTML and then do data analysis;
- **Requests**: Handle http requests;
- **Matplotlib**: Generate and plot maps;
- **Folium**: Generating maps of London and Toronto
- **Sklearn**: Used to import K-means, which is the machine learning model that was used.
- **Geocoder**: Geocoding library for non GIS users that connect locations, people, and data by using interactive maps.
- **Numpy**: Library to work with arrays, linear algebra and big amount of data.

The methodology consists in explore each of the cities individually, plot a map to show their neighborhoods, build a model by clustering all the similar neighborhoods and, finally, plot a new map with those clustered neighborhoods. After that, we analyzed the outputs, gave insights and compared and discussed all the findings.

## 3.1 Data Collection

In this stage, we start to collect all the required data for the cities of London and Toronto. We need data that has the postal codes, neighborhoods and boroughs to each one of the cities.

To collect data for London, the list of areas of London was scrapped at Wikipedia. The 2<sup>nd</sup> table was used, by the following code:

```
london_url = "https://en.wikipedia.org/wiki/List_of_areas_of_London"
wiki_london_url = requests.get(london_url)
wiki_london_data = pd.read_html(wiki_london_url.text)
wiki_london_data = wiki_london_data[1]
wiki_london_data
```

### London's Data:

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
...	...	...	...	...	...	...
526	Woolwich	Greenwich	LONDON	SE18	020	TQ435795
527	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655
528	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815
529	Yeadon	Hillingdon	HAYES	UB4	020	TQ115825
530	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804

To collect data for Toronto, the list of areas of Toronto was scrapped at Wikipedia. The 2<sup>nd</sup> table was used, by the following code:

```
wiki_url = "https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1011037969"
wiki = requests.get(wiki_url)
wiki_data = pd.read_html(wiki.text)
wiki_data = wiki_data[0]
wiki_data
```

## Toronto's Data:

	Postal Code	Borough	Neighbourhood
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
...	...	...	...
175	M5Z	Not assigned	Not assigned
176	M6Z	Not assigned	Not assigned
177	M7Z	Not assigned	Not assigned
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...
179	M9Z	Not assigned	Not assigned

## 3.2 Data Preprocessing

For London, the spaces with underscores in the title were replace and the numbers with square brackets on the column **borough** were removed:

```
wiki_london_data.rename(columns=lambda x: x.strip().replace(" ", "_"), inplace=True)
wiki_london_data
```

	Location	London borough	Post_town	Postcode district	Dial code	OS_grid_ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728
...	...	...	...	...	...	...
526	Woolwich	Greenwich	LONDON	SE18	020	TQ435795
527	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655
528	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815
529	Yeading	Hillingdon	HAYES	UB4	020	TQ115825
530	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804

For Toronto, all borough that were “not assigned” were dropped, and all the data were grouped based on Postal Code:

```
df = wiki_data[wiki_data["Borough"] != "Not assigned"]
df = df.groupby(['Postal Code']).head()
df
```

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
160	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
165	M4Y	Downtown Toronto	Church and Wellesley
168	M7Y	East Toronto	Business reply mail Processing Centre, South C...
169	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

### 3.3 Feature Selection

On both datasets, we only need neighborhood, postal codes, borough and geolocations. The columns were selected considering that:

**Toronto:**

```
df.drop(['index'], axis = 'columns', inplace = True)
df
```

	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto	Church and Wellesley
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...



## London:

```
df_london = wiki_london_data.drop( [ wiki_london_data.columns[0], wiki_london_data.columns[4], wiki_london_data.columns[5] ], axis=1)
```

```
df_london.head()
```

	London borough	Post_town	Postcode district
0	Bexley, Greenwich [7]	LONDON	SE2
1	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4
2	Croydon[8]	CROYDON	CR0
3	Croydon[8]	CROYDON	CR0
4	Bexley	BEXLEY, SIDCUP	DA5, DA14

## 3.4 Feature Engineering

Both datasets also have information related to all the cities. So only neighborhoods/borough that belongs to 'Toronto' and 'London' were selected.

On London, we did not have geolocation data. So, it was needed to search that kind of data somehow. To solve this issue, it was used ArcGIS API to get latitude and longitude of all neighborhoods of London, by creating a function to do that while passing all postal codes of London to get geographical co-ordinates.

In the end, all ray data was merged with the geographical co-ordinates to prepare the dataset to the next process.

```
london_merge = pd.concat([df_london,lat_london.astype(float), long_london.astype(float)], axis=1)
london_merge.columns= [ 'Borough', 'Town', 'Post_Code', 'Latitude', 'Longitude' ]
london_merge
```

	Borough	Town	Post_Code	Latitude	Longitude
0	Bexley, Greenwich	LONDON	SE2	51.49245	0.12127
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4	51.51324	-0.26746
6	City	LONDON	EC3	51.51200	-0.08058
7	Westminster	LONDON	WC2	51.51651	-0.11968
9	Bromley	LONDON	SE20	51.41009	-0.05683
...	...	...	...	...	...
521	Redbridge	LONDON	IG8, E18	51.58977	0.03052
522	Redbridge, Waltham Forest	LONDON, WOODFORD GREEN	IG8	51.50642	-0.12721
525	Barnet	LONDON	N12	51.61592	-0.17674
526	Greenwich	LONDON	SE18	51.48207	0.07143
528	Hammersmith and Fulham	LONDON	W12	51.50645	-0.23691

For Toronto, the library geocoder was used to get the geo-coordinates from a csv file, for each postal code and then, the data was merged just like it was done to London:

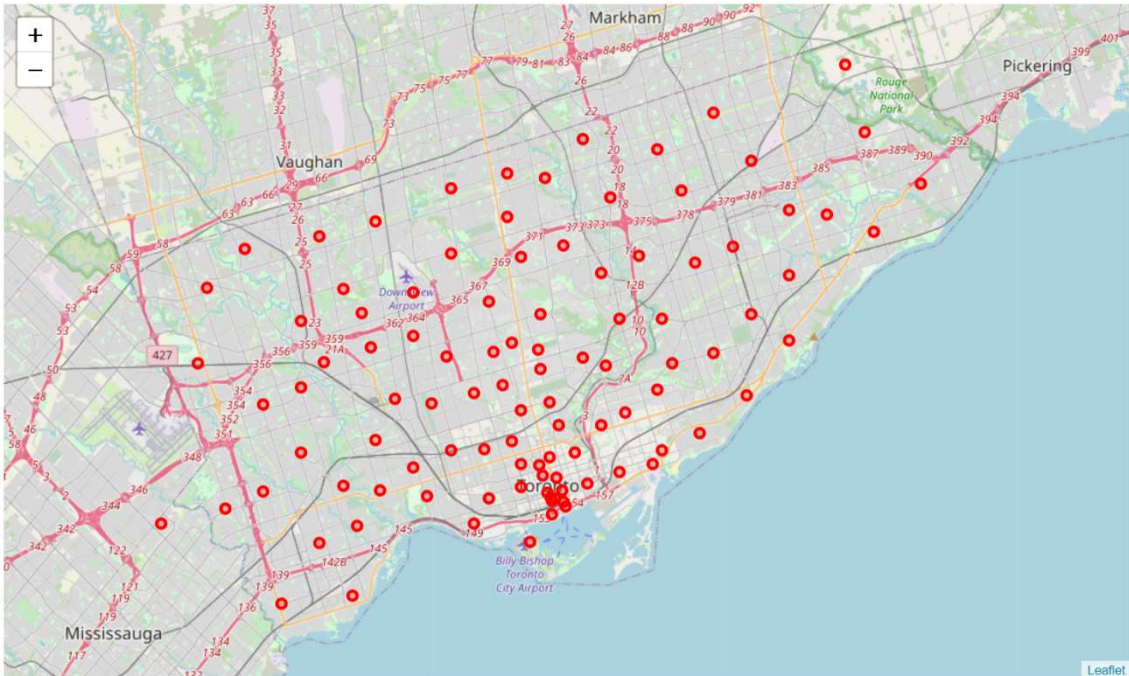
```
df_join = df.join(geo_data.set_index('Postal Code'), on='Postal Code', how='inner')
df_join
```

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
...	...	...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

### 3.5 Visualizing Toronto and London

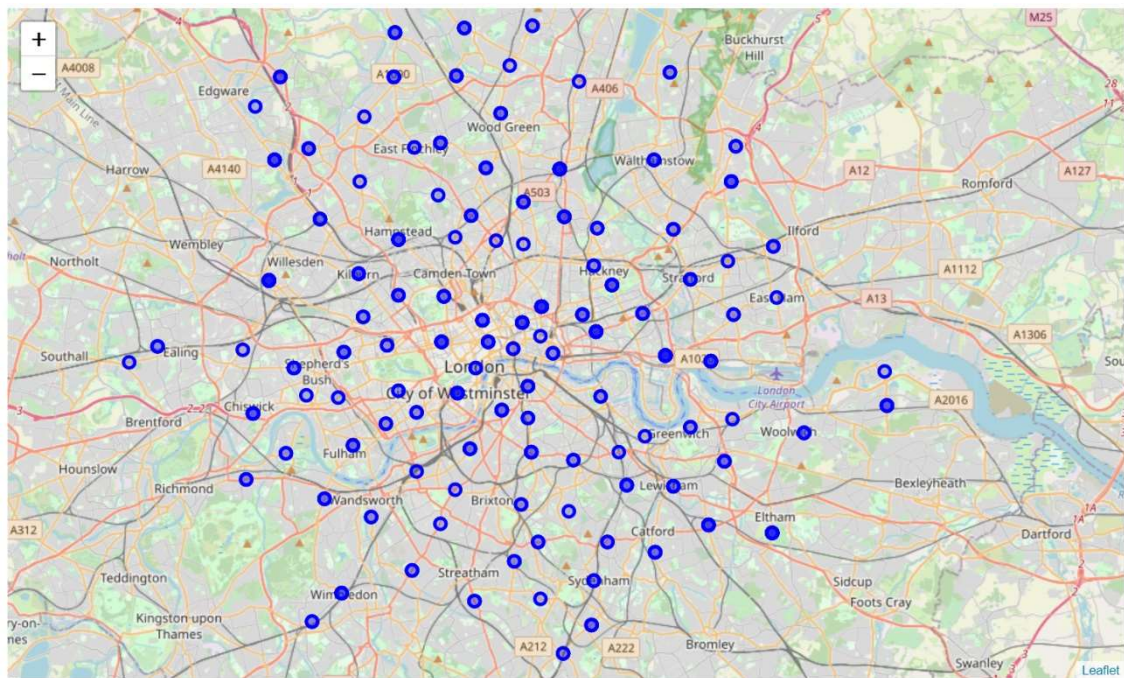
Using Folium package, we can now show a plot of the datasets, showing all the neighborhoods of the two cities.

Toronto:





London:



After visualizing the neighborhoods, it is necessary to find out how each neighborhood is like and what are the common venue and venue categories. It was settled a 500m radius for that.

To do that, it was used **Foursquare**. A function was created with **Foursquare** to collect data that belongs to each neighborhood including: name of the neighborhood, venue category, venue and geo-coordinates.

The output looks like this for London:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Category
0	Bexley, Greenwich	51.49245	0.12127	Lesnes Abbey	Historic Site
1	Bexley, Greenwich	51.49245	0.12127	Sainsbury's	Supermarket
2	Bexley, Greenwich	51.49245	0.12127	Lidl	Supermarket
3	Bexley, Greenwich	51.49245	0.12127	Abbey Wood Railway Station (ABW)	Train Station
4	Bexley, Greenwich	51.49245	0.12127	Bean @ Work	Coffee Shop
5	Bexley, Greenwich	51.49245	0.12127	Platform 1	Platform
6	Ealing, Hammersmith and Fulham	51.51324	-0.26746	Sainsbury's Local	Grocery Store
7	Ealing, Hammersmith and Fulham	51.51324	-0.26746	Acton Main Line Railway Station (AML)	Train Station
8	Ealing, Hammersmith and Fulham	51.51324	-0.26746	Co-op Food	Grocery Store
9	Ealing, Hammersmith and Fulham	51.51324	-0.26746	The Balti House	Indian Restaurant
10	Ealing, Hammersmith and Fulham	51.51324	-0.26746	Springfield Gardens	Park
11	Ealing, Hammersmith and Fulham	51.51324	-0.26746	Harry's Bar Cafe	Breakfast Spot
12	City	51.51200	-0.08058	The Association	Coffee Shop
13	City	51.51200	-0.08058	Curators Coffee Studio	Coffee Shop
14	City	51.51200	-0.08058	Sky Garden	Scenic Lookout

And this for Toronto:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	Park
1	Parkwoods	43.753259	-79.329656	Brookbanks Pool	Pool
2	Parkwoods	43.753259	-79.329656	Variety Store	Food & Drink Shop
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Portugril	Portuguese Restaurant
...	...	...	...	...	...
1325	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Subway	Sandwich Place
1326	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	7-Eleven	Convenience Store
1327	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	McDonald's	Fast Food Restaurant
1328	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Jim & Maria's No Frills	Grocery Store
1329	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	RONA	Hardware Store

## 3.6 One Hot Encoding

To find out what are the different kinds of venue categories in each neighborhood and then, calculate the top 10 common venues to base our similarity, One Hot Encoding is needed. One Hot encoding allows work with categorical datatype of the venue categories in order to convert the categorical data into numeric data.

Use label encoding in this situation, for instance, wouldn't work, since label encoding could cause the ML model to have a bias or a sort of ranking which we are trying to avoid by using One Hot Encoding.

So, One Hot Encoding was performed and the mean of the grouped venue categories for each of the neighborhood was calculated.

London:

	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	...	Vietnamese Restaurant	Warehouse Store	Whisky Bar	Wine Bar
10323	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10324	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10325	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10326	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10327	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0
10328	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10329	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10330	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10331	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10332	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10333	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10334	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10335	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10336	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10337	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10338	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10339	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10340	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10341	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
10342	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

Toronto:

	Neighbourhood	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Turkish Restaurant	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnam Restaurant
0	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Victoria Village	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	Victoria Village	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
5	Victoria Village	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
6	Victoria Village	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
7	Victoria Village	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
8	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
9	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
10	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

3.7 Top Venues in the Neighborhoods

Now, it is time to label and rank the top categories in all neighborhoods. To do that, a function was created to get the top venue categories and label them. But, since there are a huge number of categories, it was only considered the top 10.

Toronto:

	Neighbourhood	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	Chinese Restaurant	Lounge	Yoga Studio	Dessert Shop	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore
1	Alderwood, Long Branch	Pizza Place	Skating Rink	Coffee Shop	Pub	Dance Studio	Sandwich Place	Pharmacy	Gym	Gas Station	Dog Run
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Grocery Store	Supermarket	Bridal Shop	Shopping Mall	Sandwich Place	Restaurant	Pizza Place	Mobile Phone Shop
3	Bayview Village	Chinese Restaurant	Café	Japanese Restaurant	Bank	Yoga Studio	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Department Store
4	Bedford Park, Lawrence Manor East	Sushi Restaurant	Coffee Shop	Sandwich Place	Italian Restaurant	Butcher	Restaurant	Café	Pub	Pizza Place	Grocery Store
5	Berczy Park	Seafood Restaurant	Beer Bar	Cocktail Bar	Farmers Market	Breakfast Spot	Bakery	Basketball Stadium	Japanese Restaurant	Jazz Club	Bistro

London:

	Neighbourhood	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	Barnet	Coffee Shop	Café	Grocery Store	Bus Stop	Pub	Italian Restaurant	Supermarket	Pharmacy	Turkish Restaurant	Gym / Fitness Center
1	Barnet, Brent, Camden	Gym / Fitness Center	Hardware Store	Clothing Store	Supermarket	Zoo Exhibit	Filipino Restaurant	Event Space	Exhibit	Falafel Restaurant	Farmers Market
2	Bexley	Supermarket	Historic Site	Train Station	Coffee Shop	Platform	Park	Construction & Landscaping	Golf Course	Bus Stop	Fish Market
3	Bexley, Greenwich	Bus Stop	Park	Golf Course	Home Service	Construction & Landscaping	Historic Site	Sports Club	Daycare	Food & Drink Shop	Flower Shop
4	Bexley, Greenwich	Supermarket	Platform	Train Station	Historic Site	Coffee Shop	Film Studio	Event Space	Exhibit	Falafel Restaurant	Farmers Market



### 3.8 Model Building – KMeans Clustering

By using KMeans Clustering Machine Learning Algorithm the greatest part arrived: Build the Model! We used KMeans to cluster similar neighborhoods, settling the number of clusters as 5. The model labelled each neighborhood and added the label in the dataset.

After that, the dataset was merged with neighborhood venues, sorting them to add latitude and longitude for each of the neighborhood.

#### London:

	Borough	Town	Post_Code	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
0	Bexley, Greenwich	LONDON	SE2	51.49245	0.12127	3	Supermarket	Platform	Train Station	Historic Site	Coffee Shop	Film Studio	Event Space	Exhibition Space
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4	51.51324	-0.26746	2	Grocery Store	Park	Indian Restaurant	Breakfast Spot	Train Station	Filipino Restaurant	Ethiopian Restaurant	Event Space
6	City	LONDON	EC3	51.51200	-0.08058	1	Hotel	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Pub	Restaurant	Sandwich Place	Wine Bar
7	Westminster	LONDON	WC2	51.51651	-0.11968	1	Coffee Shop	Hotel	Pub	Café	Sandwich Place	Italian Restaurant	Theater	Burgers Joint
9	Bromley	LONDON	SE20	51.41009	-0.05683	1	Supermarket	Convenience Store	Fast Food Restaurant	Hotel	Grocery Store	Park	Indian Restaurant	Bus Stop

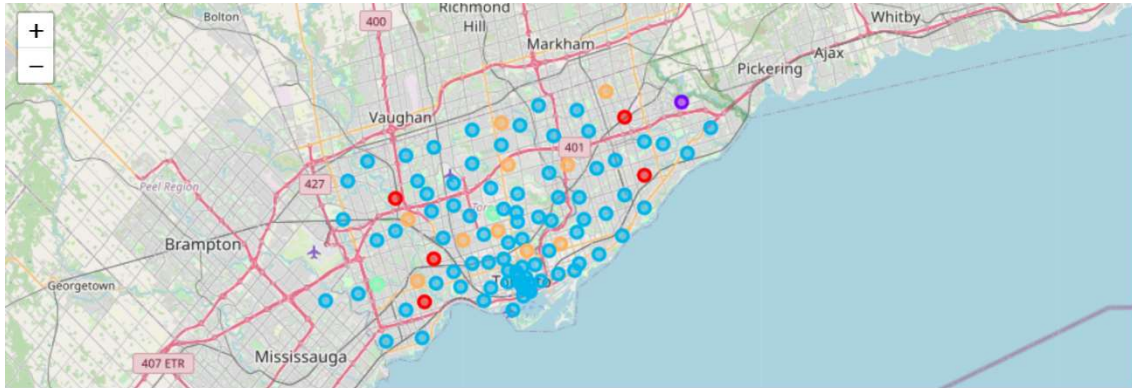
#### Toronto:

	Borough	Neighbourhood	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
	North York	Parkwoods	43.753259	-79.329656	4.0	Pool	Park	Food & Drink Shop	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop
	North York	Victoria Village	43.725882	-79.315572	2.0	Pizza Place	Coffee Shop	Portuguese Restaurant	Hockey Arena	Financial or Legal Service	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant
	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	2.0	Coffee Shop	Park	Bakery	Pub	Café	Breakfast Spot	French Restaurant	Performing Arts Venue	Chocolaterie
	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	2.0	Clothing Store	Furniture / Home Store	Accessories Store	Coffee Shop	Gift Shop	Vietnamese Restaurant	Boutique	Discount Store	Dealership
	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	2.0	Coffee Shop	Sushi Restaurant	Gym	Fried Chicken Joint	Bar	Beer Bar	Smoothie Shop	Sandwich Place	Burgers Joint
	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	4.0	Park	River	Yoga Studio	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Park

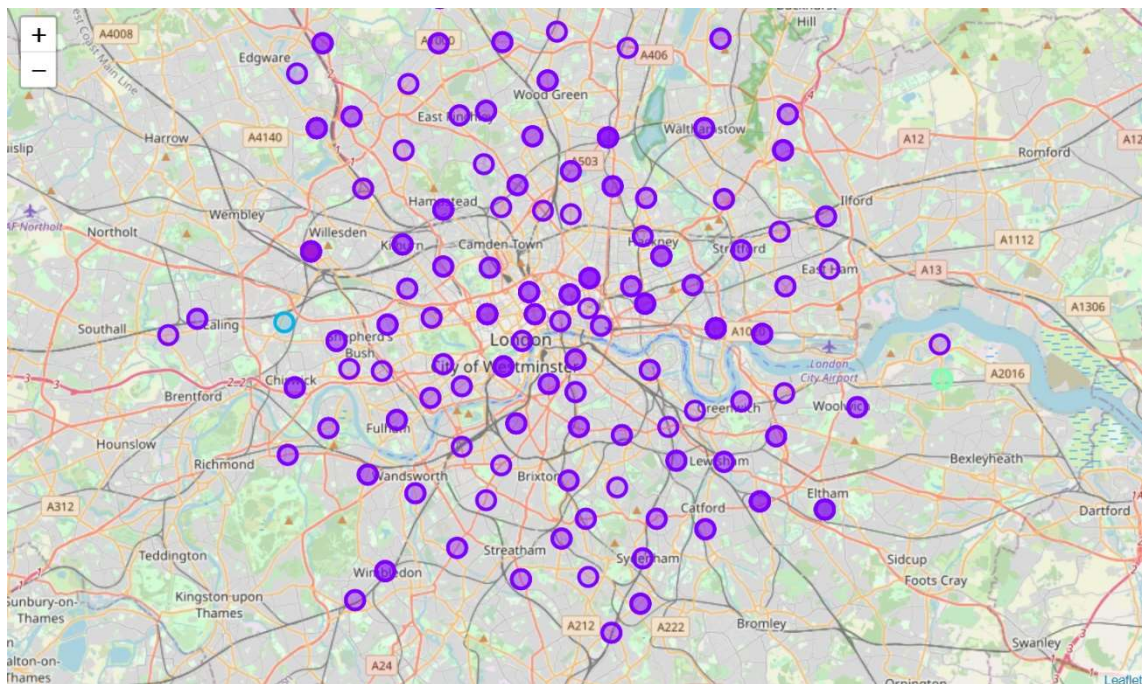
### 3.9 Visualizing all Clustered Neighborhoods

After process and compile the data and build the model, it is time to see all neighborhood clustered! Again, we used Folium to do so:

**Toronto:**



**London:**





### 3.9.1 Analyzing the Clusters

We could examine our clusters by expanding on our code using the `Cluster Labels` column:

#### TORONTO

##### Cluster 1

	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
32	Scarborough	0.0	Playground	College Auditorium	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
57	North York	0.0	Baseball Field	Food Service	Yoga Studio	Department Store	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
63	Toronto/York	0.0	Grocery Store	Breakfast Spot	Caribbean Restaurant	Department Store	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
78	Scarborough	0.0	Latin American Restaurant	Breakfast Spot	Chinese Restaurant	Lounge	Yoga Studio	Dessert Shop	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore
101	Etobicoke	0.0	Breakfast Spot	Baseball Field	Yoga Studio	Dessert Shop	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop

##### Cluster 2

	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
6	Scarborough	1.0	Fast Food Restaurant	Yoga Studio	Falafel Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center

##### Cluster 3

	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	North York	2.0	Pizza Place	Coffee Shop	Portuguese Restaurant	Hockey Arena	Financial or Legal Service	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
2	Downtown Toronto	2.0	Coffee Shop	Park	Bakery	Pub	Café	Breakfast Spot	French Restaurant	Performing Arts Venue	Chocolate Shop	Restaurant
3	North York	2.0	Clothing Store	Furniture / Home Store	Accessories Store	Coffee Shop	Gift Shop	Vietnamese Restaurant	Boutique	Discount Store	Dessert Shop	Dim Sum Restaurant
4	Downtown Toronto	2.0	Coffee Shop	Sushi Restaurant	Gym	Fried Chicken Joint	Bar	Beer Bar	Smoothie Shop	Sandwich Place	Burrito Place	Café
7	North York	2.0	Gym	Coffee Shop	Restaurant	Supermarket	Italian Restaurant	Japanese Restaurant	Discount Store	Dim Sum Restaurant	Clothing Store	Chinese Restaurant
...	...	...	...	...	...	...	...	...	...	...	...	...
96	Downtown Toronto	2.0	Italian Restaurant	Café	Restaurant	Coffee Shop	Bakery	Pet Store	Gastropub	Beer Store	Butcher	Pub
97	Downtown Toronto	2.0	Café	Coffee Shop	Restaurant	Seafood Restaurant	Gastropub	Tea Room	Concert Hall	Pizza Place	Pub	Bookstore
99	Downtown Toronto	2.0	Gay Bar	Indian Restaurant	Steakhouse	Escape Room	Beer Bar	Italian Restaurant	Japanese Restaurant	Bookstore	Breakfast Spot	Bubble Tea Shop
100	East Toronto	2.0	Light Rail Station	Yoga Studio	Auto Workshop	Gym / Fitness Center	Garden Center	Garden	Fast Food Restaurant	Farmers Market	Comic Shop	Park
102	Etobicoke	2.0	Grocery Store	Thrift / Vintage Store	Convenience Store	Discount Store	Burrito Place	Sandwich Place	Burger Joint	Fast Food Restaurant	Supplement Shop	Bakery

Cluster 4

	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
11	Etobicoke	3.0	Home Service	Yoga Studio	Deli / Bodega	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
62	Central Toronto	3.0	Garden	Home Service	Yoga Studio	Deli / Bodega	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run

Cluster 5

	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	North York	4.0	Pool	Park	Food & Drink Shop	Yoga Studio	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
21	York	4.0	Park	Women's Store	Pool	Dance Studio	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
35	East York	4.0	Pizza Place	Park	Convenience Store	Intersection	Deli / Bodega	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run
62	North York	4.0	Park	Yoga Studio	Deli / Bodega	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
64	York	4.0	Park	Jewelry Store	Yoga Studio	Deli / Bodega	Escape Room	Electronics Store	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run

LONDON

Cluster 1

	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
6	City	1	Hotel	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Pub	Restaurant	Sandwich Place	Wine Bar	Garden	French Restaurant
7	Westminster	1	Coffee Shop	Hotel	Pub	Café	Sandwich Place	Italian Restaurant	Theater	Burger Joint	Sushi Restaurant	Bakery
9	Bromley	1	Supermarket	Convenience Store	Fast Food Restaurant	Hotel	Grocery Store	Park	Indian Restaurant	Bus Stop	Gastropub	Bistro
10	Islington	1	Coffee Shop	Pub	Food Truck	Café	Park	Vietnamese Restaurant	Italian Restaurant	Cocktail Bar	Hotel	Gym / Fitness Center
12	Islington	1	Coffee Shop	Pub	Food Truck	Café	Park	Vietnamese Restaurant	Italian Restaurant	Cocktail Bar	Hotel	Gym / Fitness Center
...	...	...	...	...	...	...	...	...	...	...	...	...
521	Redbridge	1	Café	Pub	Coffee Shop	Convenience Store	Bakery	Bar	Grocery Store	Park	Liquor Store	BBQ Joint
522	Redbridge, Waltham Forest	1	Hotel	Café	Monument / Landmark	Plaza	Garden	Theater	Pub	Bakery	Pharmacy	Sandwich Place
525	Barnet	1	Coffee Shop	Café	Grocery Store	Bus Stop	Pub	Italian Restaurant	Supermarket	Pharmacy	Turkish Restaurant	Gym / Fitness Center
526	Greenwich	1	Pub	Grocery Store	Bus Stop	Indian Restaurant	Coffee Shop	Café	Historic Site	Pier	Convenience Store	Construction & Landscaping
528	Hammersmith and Fulham	1	Pub	Coffee Shop	Café	Grocery Store	Gastropub	Park	Thai Restaurant	Bakery	Hotel	Italian Restaurant

## Cluster 2

	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	Ealing, Hammersmith and Fulham	2	Grocery Store	Park	Indian Restaurant	Breakfast Spot	Train Station	Filipino Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Falafel Restaurant

## Cluster 3

	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	Bexley, Greenwich	3	Supermarket	Platform	Train Station	Historic Site	Coffee Shop	Film Studio	Event Space	Exhibit	Falafel Restaurant	Farmers Market
45	Bexley	3	Supermarket	Historic Site	Train Station	Coffee Shop	Platform	Park	Construction & Landscaping	Golf Course	Bus Stop	Fish Market
124	Bexley	3	Supermarket	Historic Site	Train Station	Coffee Shop	Platform	Park	Construction & Landscaping	Golf Course	Bus Stop	Fish Market
291	Bexley	3	Supermarket	Historic Site	Train Station	Coffee Shop	Platform	Park	Construction & Landscaping	Golf Course	Bus Stop	Fish Market
505	Bexley	3	Supermarket	Historic Site	Train Station	Coffee Shop	Platform	Park	Construction & Landscaping	Golf Course	Bus Stop	Fish Market

## Cluster 4

	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
121	Barnet, Brent, Camden	4	Gym / Fitness Center	Hardware Store	Clothing Store	Supermarket	Zoo Exhibit	Filipino Restaurant	Event Space	Exhibit	Falafel Restaurant	Farmers Market

## Cluster 5

	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
356	Brent, Ealing	5	Convenience Store	Fast Food Restaurant	Warehouse Store	Chinese Restaurant	Pharmacy	Zoo Exhibit	Film Studio	Event Space	Exhibit	Falafel Restaurant

# 4. Results and Discussion

Even being smaller than London, Toronto is a super city! It has huge number of cool places to go and it is very supportive, infrastructurally speaking. We can see a huge variety of cuisines, such as Caribbean, Portuguese, Mexican, Chinese, Vietnamese etc. There are a lot of hangout spots: many pubs, coffee shops, parks and Restaurants. On the supportive side, there are a lot of options too: Gym, Parks, Drugstores, Stores, Convenience Store, etc.

Overall, Toronto seems like the relaxing vacation place with a mix fun outdoor places and hangout spots, food spots and a wide variety of cuisines to try out. And if you are an immigrant it seems perfect to build a new life.

You have everything, everywhere in the boroughs and neighborhoods, making Toronto incredibly supportive when you need something, because since you have a lot of supportive

places to go in your neighborhood, you can go fast and don't rely on public/private transportation.

Bigger than Toronto, London seems to be super multicultural, bohemian and supportive. There are a huge variety of cuisines such as Indian, Italian, Turkish and Chinese. London exceeds a lot Toronto in the amount of fun places to go: Restaurants, bars, pubs, bakeries, coffee shops, Fish and Chips shop and Breakfast spots. An interest point: a lot of Hotels too, much more than Toronto. It is impressive because the city seems to be built for fun.

About infrastructure, it has a lot of shopping and supportive places, such as supermarkets, farm markets, train stations, grocery markets, fish markets, clothing stores. The main modes of transport seem to be Buses and trains. For outdoor leisure, the neighborhoods and boroughs also have lots of parks, golf courses, zoo, gyms and Historic sites.

## Conclusion

The main objective of this project was exploring the cities of Toronto and London to see how attractive it is to potential tourists and immigrants. Both cities were explored using their postal codes and then extrapolated the common venues present in each of the neighborhoods, concluding with clustering similar boroughs together.

As result all neighborhoods and boroughs in both the cities have a huge variety of leisure places and experiences to offer. By seeing the variety of cuisines and all different kinds of activities, both cities seem to be prepared to hug anyone, tourist or immigrant, because it gives a lot a feeling of inclusion.

Both Toronto and London seem to offer a perfect vacation stay or an excellent beginning of a new life. They have a lot of places to explore, all kinds of outdoor activities, great infrastructure and a huge variety of culture.

Toronto has the perfect balance. You have everything, everywhere. All neighborhoods are perfectly balanced with a lot of places to go for fun and a lot of supportive places to routine. It is perfect for immigrants and good for tourists.

London exceeds in culture diversity and fun. It seems to be a paradise for a tourist. All neighborhoods and boroughs have every kind of places to go with variety: all kinds of restaurants, bars, cuisines, pubs, hotels. It is impressive. It is probably expensive, because it seems to be a city built for tourism, so, despite of have a good infrastructure too (good public transportation and some supportive places), immigrants should be careful.

In the end, it is up to the stakeholders to decide which experience they would prefer. They should be diligent with the client needs to meet its expectations.