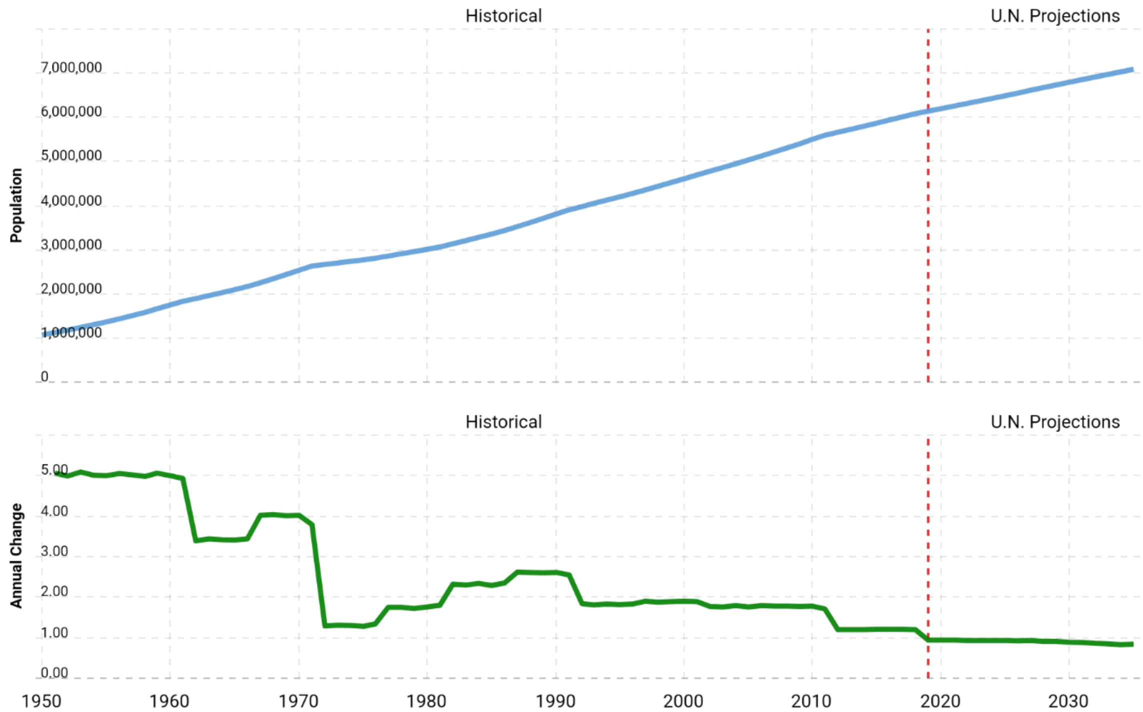
# Segmenting and Clustering Neighborhoods of Boroughs in Toronto

# Introduction

### Description & Discussion of the Background

The demographics of Toronto, Ontario, Canada make Toronto one of the most multicultural and multiracial cities in the world. In 2016, 51.5% of the residents of the city proper belonged to a visible minority group, compared with 49.1% in 2011 and 13.6% in 1981. The 2016 census by Statistics Canada estimated there were 2,731,571 living in Toronto, making it the largest city in Canada, and the fourth most populous municipality in North America. [1]




Toronto's population grew by 4.3% from 2011 to 2016, wirth an annual growth rate of 0.86%. The population of Toronto in 2019 is 6,139,000, a 0.94% increase from 2018. Toronto gas 6 boroughs (districts) are East York, Etobicoke, North York, Old Toronto, Scarborough and York.

This work segmented and clustered two of these boroughs and compared the results from this analysis.

I am an admirer of Canada as a whole, and I look forward for my PhD someday, therefore, I decided to use Toronto in my project. The city is divided into 39 districts in total.

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's northwestern shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen's Park to 400-arce High Park and its trails, sport facilities and zoo. [2]

As you can see from the figures, Toronto is a city with a high population and population density. Being such a crowded city leads the owners of shops and social sharing places in the city where the population is dense.

When we think of it by the investor, we expect from them to prefer the districts where the type of business they want to install is less intense or how as investors they will enter the market in a neighborhood.

If we think of the city residents, they may want to choose the district according to the social places density. However, it is difficult to obtain information that will guide investors in this direction, nowadays. When we consider all these problems, we can create maps and information charts where each district is clustered according to the venue density and the business in these areas.

## Data Description

To consider the problem we can list the data as below:

- The Toronto dataset exists for free on the web, and here is the link to the dataset: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The .json file has postal codes, boroughs and neighborhoods of Toronto. I cleaned the data, removed the boroughs that were not assigned. I also joined neighborhoods with the same postal code.
- To get the latitude and longitudes, here is a link to a csv file that has the geographical coordinates of each postal code: http://cocl.us/Geospatial_data. Afterwards I joined the latitude and longitude to the data frame.
- I used the Foursquare API to explore neighborhoods in Toronto. I used the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters.

## Methodology

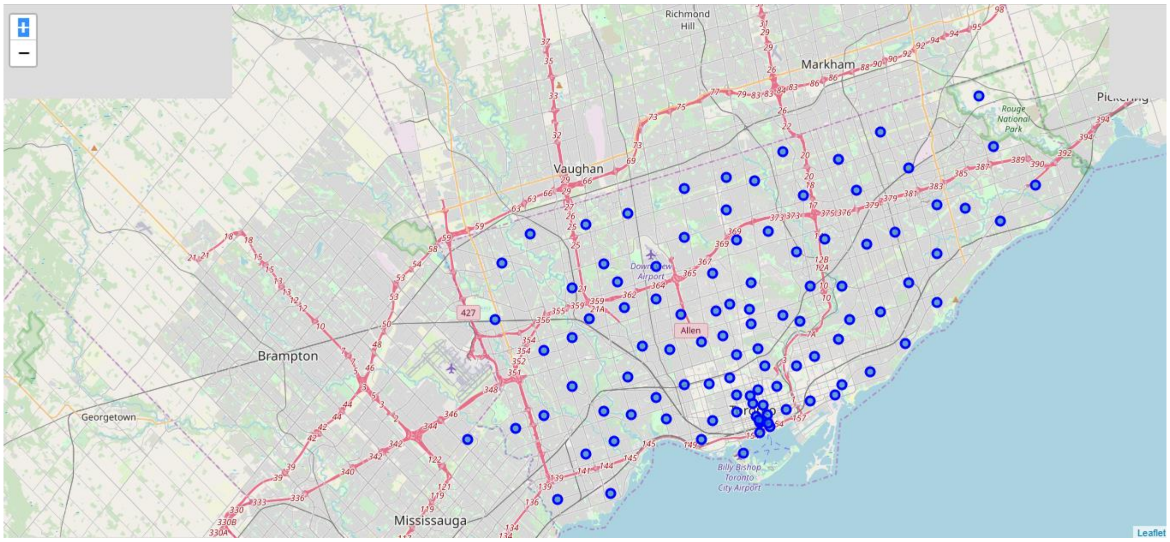
As a database, I used GitHub repository in my study. As seen from below, my master data which has the main components *Postal Code, Borough, neighborhood Latitude* and *Longitude* information of Toronto.

In [21]:
```
# Viewing the first five rows of the data
geo_data_new.head()
```

Out[21]:

| | PostalCode | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Rouge,Malvern | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Highland Creek,Rouge Hill,Port Union | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |

I used python **folium** library to visualize geographic details of Toronto and its boroughs and I created a map of Toronto with boroughs and neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:




Since I want to compare venues in two of the boroughs, Scarborough and Etobicoke.

**Clustering Scarborough:**

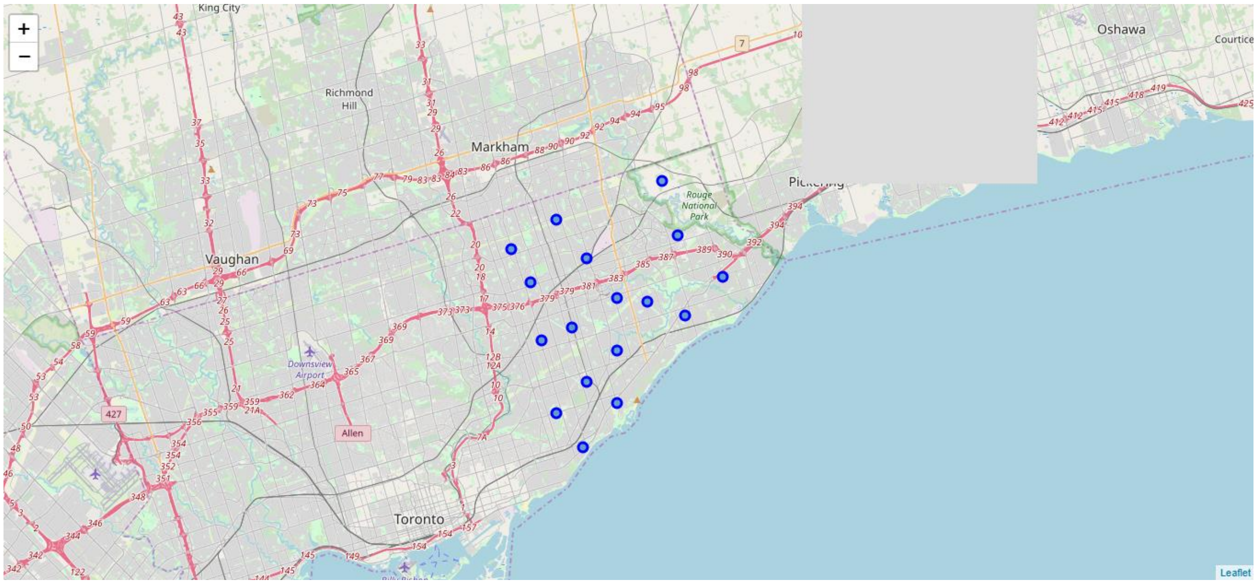
First, using python **folium** library to visualize geographic details of Scarborough and its neighborhoods. I created a map of Scarborough with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:

In [25]:
```
scar_data = geo_data_new[geo_data_new['Borough'] == 'Scarborough'].reset_index(drop=True)
scar_data.head()
```

Out[25]:

| | PostalCode | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Rouge,Malvern | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Highland Creek,Rouge Hill,Port Union | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |

The map is produced forth:




I utilized the Foursquare API to explore Scarborough and segment them. I designed the limit as **100 venues** and the radius **5000 meters** from each borough from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

Out[34]:

| | name | categories | lat | lng |
|---|---|---|---|---|
| 0 | Toronto Pan Am Sports Centre | Athletics & Sports | 43.790623 | -79.193869 |
| 1 | African Rainforest Pavilion | Zoo Exhibit | 43.817725 | -79.183433 |
| 2 | Toronto Zoo | Zoo | 43.820582 | -79.181551 |
| 3 | Polar Bear Exhibit | Zoo | 43.823372 | -79.185145 |
| 4 | Australasia Pavillion | Zoo Exhibit | 43.822563 | -79.183286 |

In summary of this data **100** venues were returned by Foursquare. Here is a merged table of neighborhoods and venues.

```
scar_venues.groupby('Neighborhood').count()
```

Out[38]:

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| Agincourt | 100 | 100 | 100 | 100 | 100 | 100 |
| Agincourt North,L'Amoreaux East,Milliken,Steeles East | 100 | 100 | 100 | 100 | 100 | 100 |
| Birch Cliff,Cliffside West | 100 | 100 | 100 | 100 | 100 | 100 |
| Cedarbrae | 100 | 100 | 100 | 100 | 100 | 100 |
| Clairlea,Golden Mile,Oakridge | 100 | 100 | 100 | 100 | 100 | 100 |
| Clarks Corners,Sullivan,Tam O'Shanter | 100 | 100 | 100 | 100 | 100 | 100 |
| Cliffcrest,Cliffside,Scarborough Village West | 100 | 100 | 100 | 100 | 100 | 100 |
| Dorset Park,Scarborough Town Centre,Wexford Heights | 100 | 100 | 100 | 100 | 100 | 100 |
| East Birchmount Park,Ionview,Kennedy Park | 100 | 100 | 100 | 100 | 100 | 100 |
| Guildwood,Morningside,West Hill | 100 | 100 | 100 | 100 | 100 | 100 |
| Highland Creek,Rouge Hill,Port Union | 82 | 82 | 82 | 82 | 82 | 82 |
| L'Amoreaux West | 100 | 100 | 100 | 100 | 100 | 100 |
| Maryvale,Wexford | 100 | 100 | 100 | 100 | 100 | 100 |
| Rouge,Malvern | 100 | 100 | 100 | 100 | 100 | 100 |
| Scarborough Village | 100 | 100 | 100 | 100 | 100 | 100 |
| Upper Rouge | 100 | 100 | 100 | 100 | 100 | 100 |
| Woburn | 100 | 100 | 100 | 100 | 100 | 100 |

We can see that all neighborhoods how reached the **100** limit of venues. On the other hand; Highland Creek, Rouge Hill and Port Union neighborhoods have **82** venues in our given coordinates with Latitude and Longitude from the table above.

In summary of this table **139** unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for scarborough in below table.

```
neighborhoods_venues_sorted.head()
```

Out[46]:

| | 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 | Chinese Restaurant | Indian Restaurant | Caribbean Restaurant | Coffee Shop | Bakery | Park | Sushi Restaurant | Noodle House | Hakka Restaurant | Supermarket |
| 1 | Agincourt North,L'Amoreaux East,Milliken,Steel... | Chinese Restaurant | Bakery | Caribbean Restaurant | Indian Restaurant | Japanese Restaurant | Vietnamese Restaurant | Bubble Tea Shop | Noodle House | Breakfast Spot | Supermarket |
| 2 | Birch Cliff,Cliffside West | Beach | Coffee Shop | Park | Breakfast Spot | Bakery | Café | BBQ Joint | Gastropub | Ice Cream Shop | Fish & Chips Shop |
| 3 | Cedarbrae | Coffee Shop | Indian Restaurant | Sandwich Place | Fast Food Restaurant | Pizza Place | Caribbean Restaurant | Breakfast Spot | Pharmacy | Park | Chinese Restaurant |
| 4 | Clairlea,Golden Mile,Oakridge | Park | Coffee Shop | Beach | Middle Eastern Restaurant | Café | American Restaurant | Thai Restaurant | Skating Rink | Breakfast Spot | Gym / Fitness Center |

We have some common venue categories in Scarborough. For this reason, I used unsupervised learning **K-means algorithm** to cluster Scarborough. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

First, I will  K-Means to cluster the boroughs into **5** clusters because when I analyze the K-Means with elbow method it ensured me the 5 degree for optimum k of the K-Means.

Here is my merged table with cluster labels for Scarborough.

Out[48]:

| | PostalCode | 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 M Comn Ve |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Rouge,Malvern | 43.806686 | -79.194353 | 0 | Zoo Exhibit | Coffee Shop | Pharmacy | Sandwich Place | Gas Station | Fast Food Restaurant | Pi Pl |
| 1 | M1C | Scarborough | Highland Creek,Rouge Hill,Port Union | 43.784535 | -79.160497 | 0 | Zoo Exhibit | Park | Coffee Shop | Pharmacy | Pizza Place | Beer Store | Break S |
| 2 | M1E | Scarborough | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 | 2 | Coffee Shop | Pharmacy | Sandwich Place | Bank | Fast Food Restaurant | Indian Restaurant | F |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 | 2 | Coffee Shop | Fast Food Restaurant | Pizza Place | Pharmacy | Indian Restaurant | Breakfast Spot | Bu J |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 | 2 | Coffee Shop | Indian Restaurant | Sandwich Place | Fast Food Restaurant | Pizza Place | Caribbean Restaurant | Break S |

## Clustering Etobicoke:

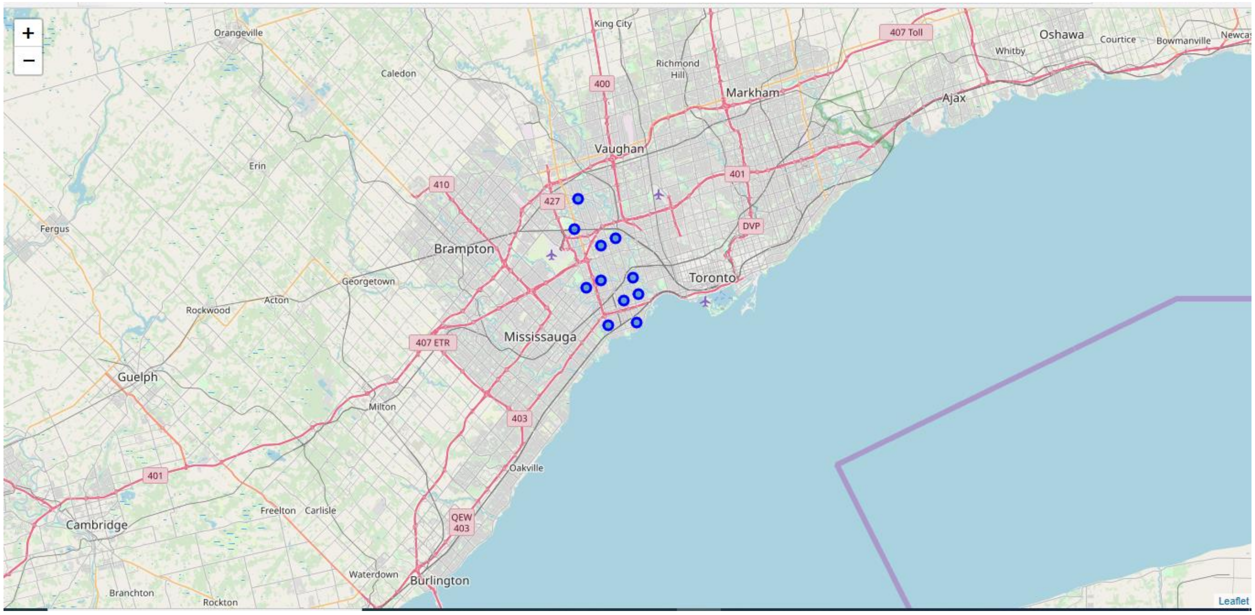
Following the same approach as above, using python **folium** library to visualize geographic details of Etobicoke and its neighborhoods. I created a map of Etobicoke with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:

In [55]:
```
eto_data = geo_data_new[geo_data_new['Borough'] == 'Etobicoke'].reset_index(drop=True)
eto_data.head()
```

Out[55]:

| | PostalCode | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M8V | Etobicoke | Humber Bay Shores,Mimico South,New Toronto | 43.605647 | -79.501321 |
| 1 | M8W | Etobicoke | Alderwood,Long Branch | 43.602414 | -79.543484 |
| 2 | M8X | Etobicoke | The Kingsway,Montgomery Road,Old Mill North | 43.653654 | -79.506944 |
| 3 | M8Y | Etobicoke | Humber Bay,King's Mill Park,Kingsway Park Sout... | 43.636258 | -79.498509 |
| 4 | M8Z | Etobicoke | Kingsway Park South West,Mimico NW,The Queensw... | 43.628841 | -79.520999 |

The map is produced forth:




I utilized the Foursquare API to explore Etobicoke and segment them. I designed the limit as **100 venues** and the radius **10000 meters** from each borough from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

Out[64]:

| | name | categories | lat | lng |
|---|---|---|---|---|
| 0 | Huevos Gourmet | Mexican Restaurant | 43.601188 | -79.503717 |
| 1 | LCBO | Liquor Store | 43.602281 | -79.499302 |
| 2 | Sweet Olenka's | Dessert Shop | 43.601099 | -79.500325 |
| 3 | Kitchen on 6th | Breakfast Spot | 43.601396 | -79.504563 |
| 4 | SanRemo Bakery | Bakery | 43.618542 | -79.499485 |

In summary of this data **100** venues were returned by Foursquare. Here is a merged table of neighborhoods and venues.

In [69]:
```
# Checking how many venues were returned for each neighborhood
eto_venues.groupby('Neighborhood').count()
```

Out[69]:

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| Albion Gardens,Beaumond Heights,Humbergate,Jamestown,Mount Olive,Silverstone,South Steeles,Thistletown | 100 | 100 | 100 | 100 | 100 | 100 |
| Alderwood,Long Branch | 100 | 100 | 100 | 100 | 100 | 100 |
| Bloordale Gardens,Eringate,Markland Wood,Old Burnhamthorpe | 100 | 100 | 100 | 100 | 100 | 100 |
| Cloverdale,Islington,Martin Grove,Princess Gardens,West Deane Park | 100 | 100 | 100 | 100 | 100 | 100 |
| Humber Bay Shores,Mimico South,New Toronto | 100 | 100 | 100 | 100 | 100 | 100 |
| Humber Bay,King's Mill Park,Kingsway Park South East,Mimico NE,Old Mill South,The Queensway East,Royal York South East,Sunnylea | 100 | 100 | 100 | 100 | 100 | 100 |
| Kingsview Village,Martin Grove Gardens,Richview Gardens,St. Phillips | 100 | 100 | 100 | 100 | 100 | 100 |
| Kingsway Park South West,Mimico NW,The Queensway West,Royal York South West,South of Bloor | 100 | 100 | 100 | 100 | 100 | 100 |
| Northwest | 100 | 100 | 100 | 100 | 100 | 100 |
| The Kingsway,Montgomery Road,Old Mill North | 100 | 100 | 100 | 100 | 100 | 100 |
| Westmount | 100 | 100 | 100 | 100 | 100 | 100 |

We can see that all neighborhoods how reached the **100** limit of venues in our given coordinates with Latitude and Longitude from the table above.

In summary of this table 116 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for Etobicoke in below table.

```
neighborhoods_venues_sorted.head()
```

Out[77]:

| | 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 | Albion Gardens,Beaumond Heights,Humbergate,Jam... | Hotel | Steakhouse | Restaurant | Bakery | Italian Restaurant | Sushi Restaurant | Chinese Restaurant | Liquor Store | Pizza Place | Coffee Shop |
| 1 | Alderwood,Long Branch | Coffee Shop | Park | Burger Joint | Seafood Restaurant | Bakery | Furniture / Home Store | Pizza Place | Breakfast Spot | Liquor Store | Grocery Store |
| 2 | Bloordale Gardens,Eringate,Markland Wood,Old B... | Bakery | Grocery Store | Seafood Restaurant | Coffee Shop | Burger Joint | Furniture / Home Store | Japanese Restaurant | Café | Middle Eastern Restaurant | Burrito Place |
| 3 | Cloverdale,Islington,Martin Grove,Princess Gar... | Café | Park | Seafood Restaurant | Grocery Store | Coffee Shop | Bakery | Furniture / Home Store | Ice Cream Shop | Steakhouse | Burger Joint |
| 4 | Humber Bay Shores,Mimico South,New Toronto | Park | Ice Cream Shop | Café | Restaurant | Seafood Restaurant | Bakery | Pizza Place | Burger Joint | Coffee Shop | Italian Restaurant |

We have some common venue categories in Etobicoke. For this reason, I used unsupervised learning **K-means algorithm** to cluster Etobicoke. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

First, I used K-Means to cluster the boroughs into **5** clusters because when I analyze the K-Means with elbow method it ensured me the 5 degree for optimum k of the K-Means.

Here is my merged table with cluster labels for Etobicoke.

```
eto_merged.head() # check the last columns!
```

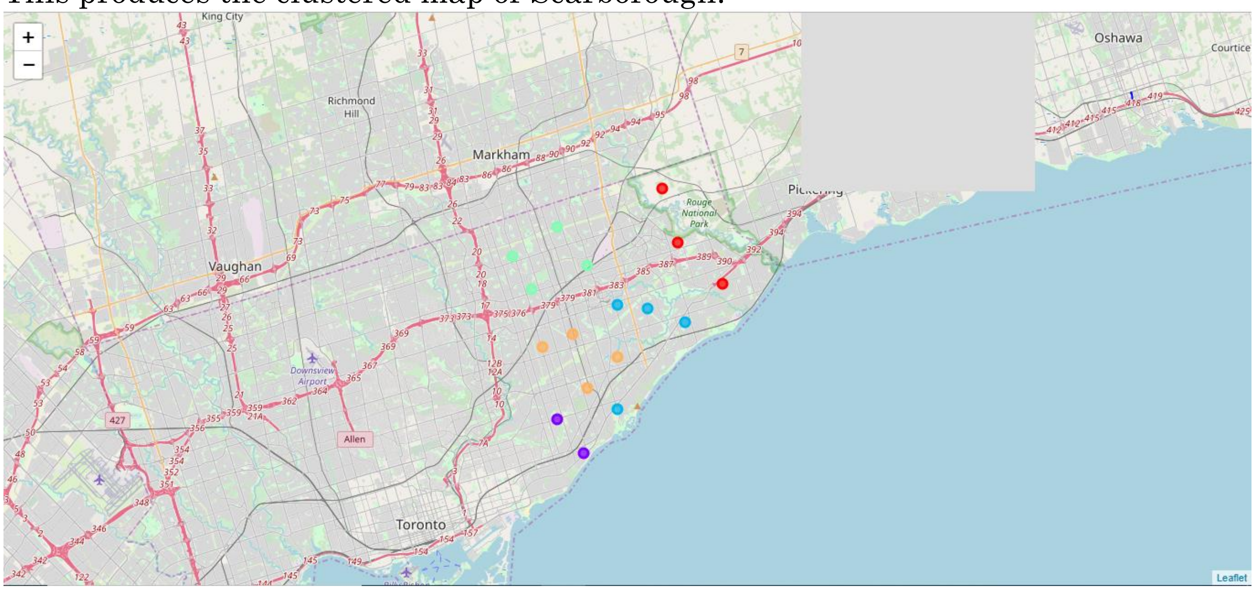
Out[79]:

| | PostalCode | 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 Con V |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M8V | Etobicoke | Humber Bay Shores,Mimico South,New Toronto | 43.605647 | -79.501321 | 2 | Park | Ice Cream Shop | Café | Restaurant | Seafood Restaurant | Bakery | Pizza Place | E |
| 1 | M8W | Etobicoke | Alderwood,Long Branch | 43.602414 | -79.543484 | 4 | Coffee Shop | Park | Burger Joint | Seafood Restaurant | Bakery | Furniture / Home Store | Pizza Place | Bre |
| 2 | M8X | Etobicoke | The Kingsway,Montgomery Road,Old Mill North | 43.653654 | -79.506944 | 0 | Park | Café | Cocktail Bar | Coffee Shop | Bar | Ice Cream Shop | Brewery | Gi |
| 3 | M8Y | Etobicoke | Humber Bay,King's Mill Park,Kingsway Park Sout... | 43.636258 | -79.498509 | 0 | Café | Cocktail Bar | Park | Bakery | Restaurant | Bar | Ice Cream Shop | Se Resta |
| 4 | M8Z | Etobicoke | Kingsway Park South West,Mimico NW,The Queensw... | 43.628841 | -79.520999 | 2 | Park | Café | Coffee Shop | Pizza Place | Seafood Restaurant | Italian Restaurant | Grocery Store | Resta |

## Results

## For Scarborough

This produces the clustered map of Scarborough:




When we examine above map, we can label each cluster as follows:

## Cluster One:

**Cluster 1**

```
In [50]: scar_merged.loc[scar_merged['Cluster Labels'] == 0, scar_merged.columns[[1] + list(range(5, scar_merged.shape[1]))]]
```

Out[50]:

| | 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 | Scarborough | 0 | Zoo Exhibit | Coffee Shop | Pharmacy | Sandwich Place | Gas Station | Fast Food Restaurant | Pizza Place | Burger Joint | Fried Chicken Joint | Bank |
| 1 | Scarborough | 0 | Zoo Exhibit | Park | Coffee Shop | Pharmacy | Pizza Place | Beer Store | Breakfast Spot | Gas Station | Smoothie Shop | Fast Food Restaurant |
| 16 | Scarborough | 0 | Zoo Exhibit | Coffee Shop | Fast Food Restaurant | Sandwich Place | Pizza Place | Pharmacy | Gas Station | Grocery Store | Hakka Restaurant | Burger Joint |

# Cluster Two:

**Cluster 2**

```
In [51]: scar_merged.loc[scar_merged['Cluster Labels'] == 1, scar_merged.columns[[1] + list(range(5, scar_merged.shape[1]))]]
```

Out[51]:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | Scarborough | 1 | Park | Coffee Shop | Beach | Middle Eastern Restaurant | Café | American Restaurant | Thai Restaurant | Skating Rink | Breakfast Spot | Gym / Fitness Center |
| 9 | Scarborough | 1 | Beach | Coffee Shop | Park | Breakfast Spot | Bakery | Café | BBQ Joint | Gastropub | Ice Cream Shop | Fish & Chips Shop |

# Cluster Three:

**Cluster 3**

```
In [52]: scar_merged.loc[scar_merged['Cluster Labels'] == 2, scar_merged.columns[[1] + list(range(5, scar_merged.shape[1]))]]
```

Out[52]:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Scarborough | 2 | Coffee Shop | Pharmacy | Sandwich Place | Bank | Fast Food Restaurant | Indian Restaurant | Park | Pizza Place | Gym | Gas Station |
| 3 | Scarborough | 2 | Coffee Shop | Fast Food Restaurant | Pizza Place | Pharmacy | Indian Restaurant | Breakfast Spot | Burger Joint | Sandwich Place | Caribbean Restaurant | Grocery Store |
| 4 | Scarborough | 2 | Coffee Shop | Indian Restaurant | Sandwich Place | Fast Food Restaurant | Pizza Place | Caribbean Restaurant | Breakfast Spot | Pharmacy | Park | Chinese Restaurant |
| 8 | Scarborough | 2 | Coffee Shop | Pharmacy | Sandwich Place | Park | Burger Joint | Grocery Store | Gym | Bank | Fast Food Restaurant | Pizza Place |

# Cluster Four:

**Cluster 4**

```
In [53]: scar_merged.loc[scar_merged['Cluster Labels'] == 3, scar_merged.columns[[1] + list(range(5, scar_merged.shape[1]))]]
```

Out[53]:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | Scarborough | 3 | Chinese Restaurant | Indian Restaurant | Caribbean Restaurant | Coffee Shop | Bakery | Park | Sushi Restaurant | Noodle House | Hakka Restaurant | Supermarket |
| 13 | Scarborough | 3 | Chinese Restaurant | Caribbean Restaurant | Bakery | Middle Eastern Restaurant | Supermarket | Coffee Shop | Indian Restaurant | Breakfast Spot | Noodle House | Korean Restaurant |
| 14 | Scarborough | 3 | Chinese Restaurant | Bakery | Caribbean Restaurant | Indian Restaurant | Japanese Restaurant | Vietnamese Restaurant | Bubble Tea Shop | Noodle House | Breakfast Spot | Supermarket |
| 15 | Scarborough | 3 | Chinese Restaurant | Caribbean Restaurant | Bubble Tea Shop | Bakery | Japanese Restaurant | Supermarket | Middle Eastern Restaurant | Noodle House | Cantonese Restaurant | Pharmacy |

# Cluster Five:

**Cluster 5**

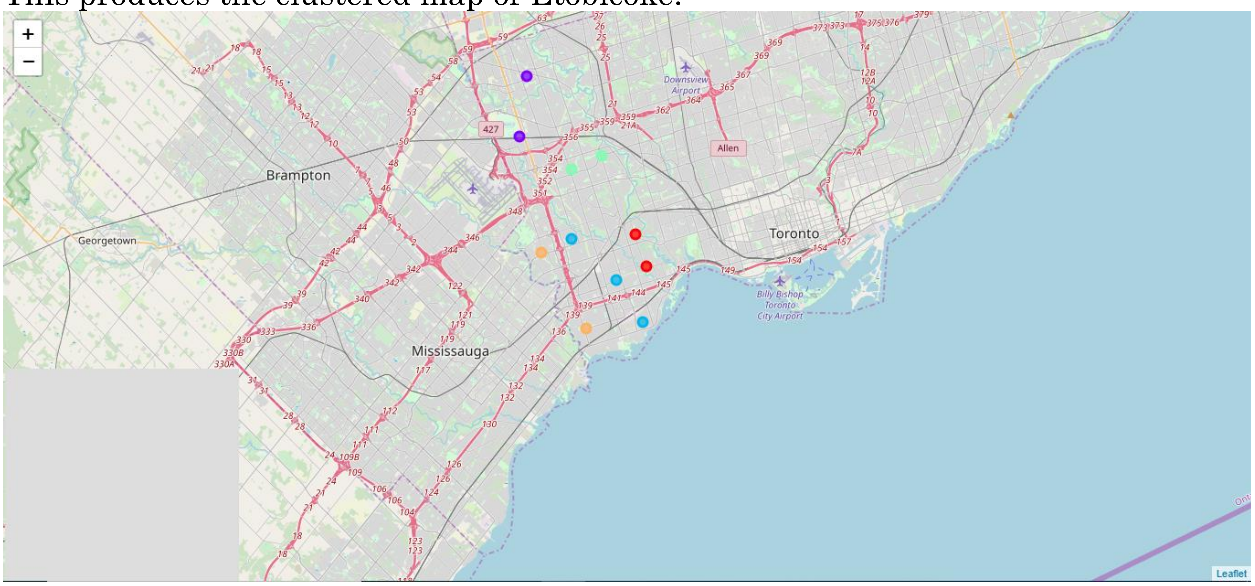
In [54]:
```
scar_merged.loc[scar_merged['Cluster Labels'] == 4, scar_merged.columns[[1] + list(range(5, scar_merged.shape[1]))]]
```

Out[54]:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | Scarborough | 4 | Coffee Shop | Park | Indian Restaurant | Burger Joint | Pharmacy | Fast Food Restaurant | Supermarket | Chinese Restaurant | Pet Store | Burrito Place |
| 6 | Scarborough | 4 | Coffee Shop | Park | Middle Eastern Restaurant | Burger Joint | Indian Restaurant | Gym | Chinese Restaurant | Bakery | Pharmacy | Pet Store |
| 10 | Scarborough | 4 | Coffee Shop | Middle Eastern Restaurant | Indian Restaurant | Caribbean Restaurant | Supermarket | Chinese Restaurant | Pizza Place | Pharmacy | Burger Joint | Bookstore |
| 11 | Scarborough | 4 | Middle Eastern Restaurant | Coffee Shop | Supermarket | Restaurant | Mediterranean Restaurant | Chinese Restaurant | Burrito Place | Caribbean Restaurant | Indian Restaurant | Burger Joint |

## For Etobicoke:

This produces the clustered map of Etobicoke:




When we examine above map, we can label each cluster as follows:

# Cluster One:

**Cluster 1**

In [81]:
```
eto_merged.loc[eto_merged['Cluster Labels'] == 0, eto_merged.columns[[1] + list(range(5, eto_merged.shape[1]))]]
```

Out[81]:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Etobicoke | 0 | Park | Café | Cocktail Bar | Coffee Shop | Bar | Ice Cream Shop | Brewery | Grocery Store | Seafood Restaurant | Japanese Restaurant |
| 3 | Etobicoke | 0 | Café | Cocktail Bar | Park | Bakery | Restaurant | Bar | Ice Cream Shop | Seafood Restaurant | Coffee Shop | Pizza Place |

# Cluster Two:

**Cluster 2**

```
In [82]: eto_merged.loc[eto_merged['Cluster Labels'] == 1, eto_merged.columns[[1] + list(range(5, eto_merged.shape[1]))]]
```

Out[82]:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | Etobicoke | 1 | Hotel | Steakhouse | Restaurant | Bakery | Italian Restaurant | Sushi Restaurant | Chinese Restaurant | Liquor Store | Pizza Place | Coffee Shop |
| 10 | Etobicoke | 1 | Coffee Shop | Hotel | Bakery | Grocery Store | Middle Eastern Restaurant | Chinese Restaurant | Café | Sushi Restaurant | Indian Restaurant | Steakhouse |

# Cluster Three:

**Cluster 3**

```
In [83]: eto_merged.loc[eto_merged['Cluster Labels'] == 2, eto_merged.columns[[1] + list(range(5, eto_merged.shape[1]))]]
```

Out[83]:

| | 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 | Etobicoke | 2 | Park | Ice Cream Shop | Café | Restaurant | Seafood Restaurant | Bakery | Pizza Place | Burger Joint | Coffee Shop | Italian Restaurant |
| 4 | Etobicoke | 2 | Park | Café | Coffee Shop | Pizza Place | Seafood Restaurant | Italian Restaurant | Grocery Store | Restaurant | Bakery | Burger Joint |
| 5 | Etobicoke | 2 | Café | Park | Seafood Restaurant | Grocery Store | Coffee Shop | Bakery | Furniture / Home Store | Ice Cream Shop | Steakhouse | Burger Joint |

# Cluster Four

**Cluster 4**

```
In [84]: eto_merged.loc[eto_merged['Cluster Labels'] == 3, eto_merged.columns[[1] + list(range(5, eto_merged.shape[1]))]]
```

Out[84]:

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | Etobicoke | 3 | Park | Hotel | Steakhouse | Coffee Shop | Grocery Store | Japanese Restaurant | Café | Eastern European Restaurant | Chinese Restaurant | Restaurant |
| 8 | Etobicoke | 3 | Hotel | Coffee Shop | Japanese Restaurant | Grocery Store | Park | Café | Steakhouse | Chinese Restaurant | Middle Eastern Restaurant | Seafood Restaurant |

# Cluster Five:

**Cluster 5**

```
In [85]: eto_merged.loc[eto_merged['Cluster Labels'] == 4, eto_merged.columns[[1] + list(range(5, eto_merged.shape[1]))]]
```

Out[85]:

| | 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 | Etobicoke | 4 | Coffee Shop | Park | Burger Joint | Seafood Restaurant | Bakery | Furniture / Home Store | Pizza Place | Breakfast Spot | Liquor Store | Grocery Store |
| 6 | Etobicoke | 4 | Bakery | Grocery Store | Seafood Restaurant | Coffee Shop | Burger Joint | Furniture / Home Store | Japanese Restaurant | Café | Middle Eastern Restaurant | Burrito Place |

In the final section, I created to examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, one can then assign a name to each cluster from the table which also has the below information for each borough:

- Borough name,
- Cluster name,
- Top 10 number of venues

## Discussion

As I mentioned before, Toronto is a big city with a high population density in a narrow area. The total number of measurements and population densities of the 6 boroughs in total can vary.

As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high-quality results for this metropolis.

I used the Kmeans algorithm as part of this clustering study. When I tested the Elbow method, I set the optimum k value to 5. However, neighbourhoods of two boroughs were compared. For more detailed and accurate guidance, the data set was expanded and the details of the neighborhoods were drilled and the venues in certain radius were produced and clustered for informed decisions.

I also performed data analysis through this information by adding the coordinates of boroughs on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

I ended the study by visualizing the data and clustering information on the Toronto map. In future studies, web or telephone applications can be carried out to direct investors. In addition, more boroughs could be compared to see how their neighborhoods look like.

There were several observations from comparing the neighborhoods of these two boroughs, that is Scarborough and Etobicoke.

Clearly, Scarborough is more populated than Etobicoke, in that Foursquare returned 139 unique categories for Scarborough within a radius of 5000m, while for Etobicoke, it returned 116 unique categories within 10000m radius.

This agrees with a Wikipedia source on the demographics of Toronto [1], that says the neighborhoods in the city of Toronto that experienced the highest increase in population from 2001 to 2011 (higher than 15%). It was noticed that neighborhoods of Scarborough such as Rouge (59.6%), Clairlea-Birchmount (24.0%), and Bendale (21.4%) were higher than the neighborhood of Etobicoke like Islington-City Centre West (20.9%).

Another interesting discovery from these results is the ethnic diversity. According to 2016 census [4], Scarborough had a population of 623,135: White: 26.5%, South Asian: 25.4%, Chinese: 19.0%, Black: 10.8%, Filipino: 8.4% while Etobicoke York had a population of 583,395: White: 48.9%, Black: 15.7%, South Asian: 11.9%, Latin American: 5.6%.

This agrees with the results from the clusters made for Scarborough and Etobicoke. One will notice that there were lots Chinese restaurants venues in the neighborhoods of Scarborough, that is because neighborhoods of Scarborough have the highest number of Chinese in Toronto. One will also notice that there was no one Chinese restaurant in Etobicoke, this is because Chinese do not live in the neighborhoods of Etobicoke.

## Conclusion

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

Arimoro, Olayinka

## References:


- [1] Demographics of Toronto — Wikipedia
- [2] About Toronto Canada. Google search
- [3] Foursquare API
- [4] Toronto, City of (November 14, 2017). Community Council Area Profiles. *City of Toronto,* Retrieved August 21, 2018