# **Battle Of the Neighborhoods**

# **Background**

Toronto is a very large city with a lot of different neighborhoods. Because of the large population it is an ideal location to open a restaurant. But there are many neighborhoods in Toronto. To determine the best restaurant neighborhood, we will use data science to recommend the best neighborhood for a restaurant.

# **Objective**

The object of this project is to find the ideal neighborhood to open a restaurant in Toronto. We must also use Foursquare data.

# **Target Audience**

Corporation or individual interested in investing or opening a restaurant in the Toronto Area.

# **Data Description**

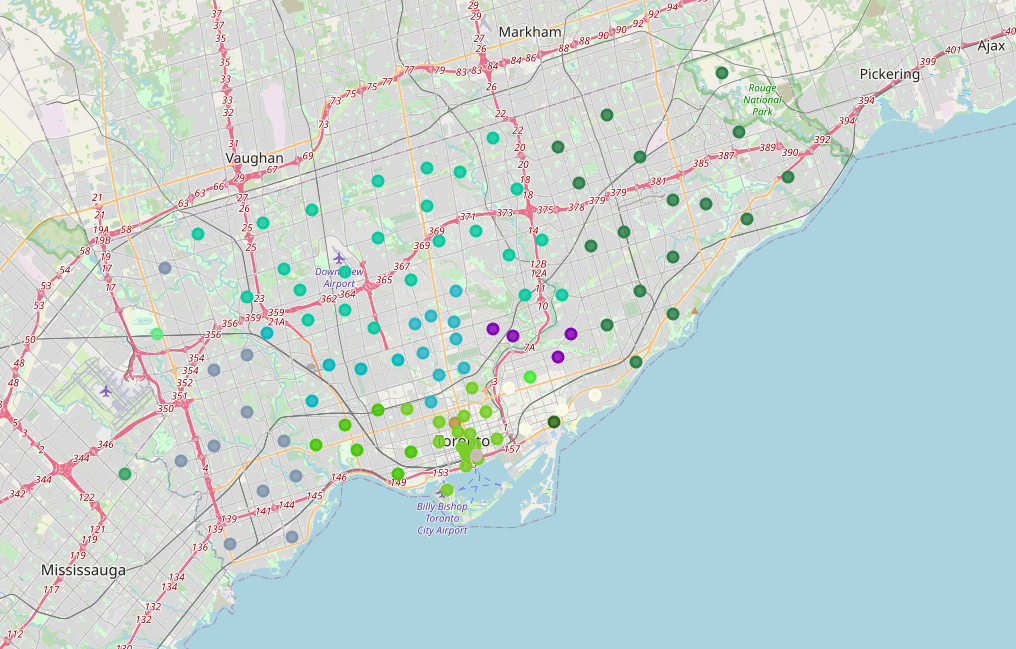
We will be using the Toronto city data scrapped in week3 and the foursquare API.

We will scrape the data from Wikipedia and clean the data (removing the not assigned boroughs). We will then add latitudes and longitude data to the neighborhoods using the geospatial dataset and postal codes. We will then retrieve venue data from the Foursquare API for each neighborhood. And Recommend the location based on number of venues present in the neighborhoods

What our data looks like:



What our map looks like:



# **Data Sources**

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

<https://cocl.us/Geospatial_data>

<https://foursquare.com/developers/apps>

**Methodology**

After scraping the data from wikipedia and merging it with the geo data and then fetching and merging the venue category data. We got the table below



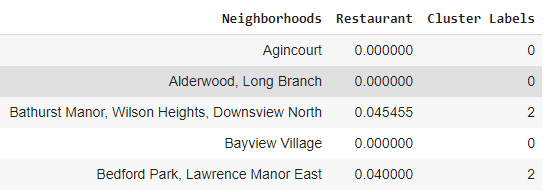
Afterwards we wanted to see how many average restaurants were in each neighborhood so we grouped the rows by neighborhood and took frequency of each venue category in that neighborhood.

Achieving the table below with frequency of restaurants in each neighborhood



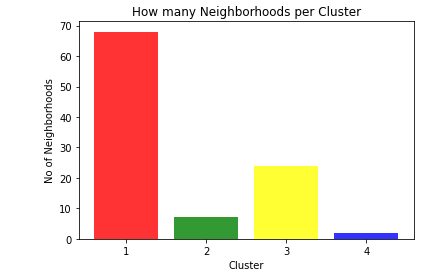
**k-Means Clustering**

We then clustered the neighborhoods based on the frequency of italian restaurants to find the best area for the business. And getting the table below

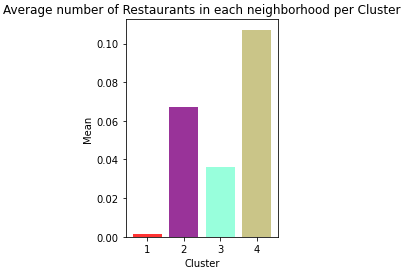


Results

We then found out how many neighborhoods per cluster were present.



Cluster 1 had 68, cluster 2 had 24 and cluster 3 had 7 and cluster 4 had 2. We then found out the average number of restaurants per neighborhood in each cluster.



**Conclusion**

From the above stats we can see cluster 1 would be the best place for a new restaurant with cluster 3 being the second best and cluster 2 being 3rd best and cluster 4 being the worst. Thus it is recommended to open the restaurant in cluster 1.