**Capstone Project - The Battle of Neighborhoods**

**1. Introduction**

**1.1 Background**

As knowing a friend is going to move to live in Toronto, I want to do a explore and make a suggestion for where my friend should live in this city based on all the great amenities and other types of venues that exist in the neighborhood, such as gourmet fast food joints, pharmacies, parks, grad schools and so on.

**1.2 Problem**

The problem is how to identify the best place in Toronto to live.

**1.3 target audience**

Not only people who want to move the family to Toronto but also people who takes a short stay in Toronto for leisure or business would be interested in the result of this project to identify the best place in Toronto to book a hotel or buy/rent a house.

**2. Data acquisition and cleaning**

**2.1 Data sources**

Neighbourhood data are from Wikipedia for Postcode, Borough and Neighbourhood.

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

A screenshot of a cell phone

Description automatically generated

Geospatial Coordinates data are from the CSV file.

<https://cocl.us/Geospatial_data>

A screenshot of a cell phone

Description automatically generated

Amenity data are from Foursquare by API request.

url = 'https://api.foursquare.com/v2/venues/explore?&client\_id={}&client\_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(CLIENT\_ID, \_SECRET, VERSION, lat, lng, radius, LIMIT)

Here is an example to show what information can be extracted from the foursquare data for each neighborhood.

A screenshot of a cell phone

Description automatically generated

Here is an example to show what venue categories can be extracted from the foursquare data.



**2.2 Data cleaning**

Neighbourhood data are available on the web. I got them by using BeautifulSoup and put them into a dataframe. Then read the CSV file with Geospatial Coordinates into another dataframe. As both of the two dataframe have postal code, I can use the postal code as keys to consolidate two dataframes into one dataframe with Neighbourhood and Coordinates. Then I use the Coordinates to call Foursqare API to get Amenity data.

**3. Exploratory Data Analysis**

Once I got the Neighbourhood and Coordinates, I marked those Neighbourhood in the map with blue circles.

A close up of a map

Description automatically generated

Then I used Foursquare API to get the nearby venues within 500 meter area. For this, I created an account on Foursquare API to get Client ID and Client Secret. I used this information to access locations on Foursquare API. I grouped them by neighborhoods and took the mean on the frequency occurrences of each venue category. This is the preprocessed data for Clustering.  
I used K-Means Clustering Method to group different venues in group. I created 4 clusters to show the futures of different neighborhood Clusters.

**A picture containing text, map

Description automatically generated**

I marked the 4 Clusters on the map with circles in different colors.

**4. Result**

The 4 clusters are :  
Cluster 1 - Most Common Venues in this cluster are related to shop, park, Café, Store, etc.

Cluster 2 - Most Common Venues in this cluster are related to Playground, park, Trail, Building, etc.

Cluster 3 - Most Common Venues in this cluster are related to shop, Garden, Yoga Studio, Dive Bar, etc.

Cluster 4 - Most Common Venues in this cluster are related to park, Trail, Health Food Store, Lake, etc.

**5. Conclusions**

As clustered with different futures, the best place to live with family should be cluster 1 supported by the most important amenities for daily life. The other 3 clusters could be also taken into consideration because of diversity of people who prefers to those futures.