

## **2.0 Data Description:**

As we need to explore, segment, and cluster the neighborhoods in the city of Toronto, the Toronto neighborhoods data is key for this project. Unfortunately, the data is for the Toronto neighborhood data is not widely available on the Internet in the structured format, hence we need to scrap it through an existing Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. The data should contain the coordinates for each of the neighborhood in Toronto that will help us to further obtaining more information critical for this project. We will also like to obtain the key information like below; such as number of residences information for each neighborhood which is one of the key factors for the neighborhood of choice in the final report. The data needs to be clean up and eventually in a structured format like the example below.

1. Neighborhood Name
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Number of residences in each neighborhood.

## **2.1 Data Features:**

We will be leveraging on features in a reliable location information provider such as the Foursquare.com to explore the various types of venues and its categories available in each neighborhood. We will also need to understand the type of these venues nearby (i.e. within 500M) in each of the respective neighborhood. The information obtained per neighborhood will be as such like below and has to be in a structured format so to allow for further computation:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue Name
5. Venue Category
6. Venue Latitude
7. Venue Longitude