**The Battle of Neighborhood: Data Description**

We are going to use the dataset consisting of postal codes, Borough and Neighborhood of North York which is scrapped from the Wikipedia page mentioned below:

Data Link: [https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M,](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Most time consuming part of any data science project is collection of data and data cleaning, Some of the data can be gathered with the help of web-scrapping and some of the data can be obtained from the data providers but it adds some cost to the project. By considering the limitations related to the cost, we are using API’s and other data sources which are available for free, as free membership offers limitation of usage, we are limiting the scope of project.

The data gathered from Foursquare API for the North York Borough will be used to cluster the neighborhood according to the similarities by using K-means clustering algorithm, Foursquare provides the location data within the area of interest such as venue name, locations, business etc.

The Data obtained from web-scraping is analyzed and processed and further meta data is added to the original data to get insights of the data. Some of the attributes of the data are:

1. Postal codes
2. Neighborhood’s
3. Neighborhood’s latitude
4. Neighborhood’s longitude
5. Venue
6. Venue’s latitude
7. Venue’s longitude
8. Venue Category
9. Nearby Schools
10. Nearby Hospitals