Capstone Project - The Battle of Neighborhoods (week 1)

(Best Neighborhoods to settle in, according to some criteria)

Table of contents

B - Data Section: Data understanding

Data Sources

We are going to use **Toronto dataset** scraped from the web. Here is the link: ('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'). This table is consisted of 3 columns: postcodes, boroughs and neighborhoods.

We will also get **Toronto's Postcode location data** by downloading from the web. Here is the link: (http://download.geonames.org/export/zip/). This geospatial table is consisted of 4 columns: postcodes, boroughs, latitudes and longitudes.

Data Collection

By using the **BeautifulSoup package** for web scraping, we will pull out informations of Postal code, boroughs and neighborhoods, from the table scraped on the web.

By using the **Foursquare API**, we will obtain location data and explore venues around.

The borough and neighborhood table joined with location data, present the latitude and the longitude of each neighborhood.

The exploration of venues presents venues near of the neighborhood of Toronto. We will use it to filter gardens, bus station, and medical centers as well. Now we are ready to prepare the data for the modelling section. Before that let's look at the methodology that we are going to use.