Location Data Anlaysis of Toronto, Canada to open a Fitness Centre

2. Data Acquisition

The data sources for this project are listed as below:

- To get the list of all the neighborhoods in Toronto, I will be utilizing web scraping technique to extract the content from Wikipedia page of "List of postal codes of Canada: M" (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). The postal code, borough and neighborhood names will be presented in panda dataframe.
- To get the geospatial data of all the neighborhoods that includes geographical coordinates, <Geospatial Coordinates> csv file will be used. Next, the two dataframes will be combined to get the information of postal code, borough, neighborhood names and their coordinates in one dataframe.
- Foursquare API will be used to explore the most common venues in Toronto. Venue information such as names, categories, etc. will be collected. Specifically, I will be exploring the details of neighborhoods that have more offices since it will be good to focus on office workers as potential customers due to exposing more sedentary lifestyle.
- After exploring and screening the potential neighborhoods, I will be extracting the neighborhood profile, including demographic data from the Open Data Portal, provided by the City's Social Policy Analysis & Research Unit from Statistics Canada Census data <https://open.toronto.ca/dataset/neighbourhood-profiles/>.