You need to know where the gyms are located in Toronto. To do this, you must know the Toronto neighborhoods, their respective geographical locations and the gyms that are in them.

The necessary data as well as from where it will be obtained are detailed below:

- Neighborhoods in Canada
  ◦ Data for districts and neighborhoods in Canada
  ◦ They are obtained by scrapping the page

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and

- Geospatial location of neighborhoods in Canada
  ◦ Latitude and longitude of neighborhoods in Canada
  ◦ Obtained from http://cocl.us/Geospatial_data
- Venue data
  ◦ Venues in each neighborhood.
  ◦ Obtained through the Foursquare API.

Let's now look at the data in a little more detail.

**1.** Postal codes in Canada:

Provide postal code, borough, and neighborhoods for Canada.

| Postal Code | Borough | Neighbourhood |
|---|---|---|
| M1A | Not assigned | Not assigned |
| M2A | Not assigned | Not assigned |
| M3A | North York | Parkwoods |
| M4A | North York | Victoria Village |
| M5A | Downtown Toronto | Regent Park, Harbourfront |
| M6A | North York | Lawrence Manor, Lawrence Heights |
| M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government |
| M8A | Not assigned | Not assigned |
| M9A | Etobicoke | Islington Avenue, Humber Valley Village |
| M1B | Scarborough | Malvern, Rouge |
| M2B | Not assigned | Not assigned |
| M3B | North York | Don Mills |

Figure 1. Postal Codes of Canada

**2.** Geoespatial data

Provides latitude and longitude for Canadian postal codes.

| | A | B | C |
|---|---|---|---|
| 1 | Postal Cod | Latitude | Longitude |
| 2 | M1B | 43.80669 | -79.1944 |
| 3 | M1C | 43.78454 | -79.1605 |
| 4 | M1E | 43.76357 | -79.1887 |
| 5 | M1G | 43.77099 | -79.2169 |
| 6 | M1H | 43.77314 | -79.2395 |
| 7 | M1J | 43.74473 | -79.2395 |
| 8 | M1K | 43.72793 | -79.262 |
| 9 | M1L | 43.71111 | -79.2846 |
| 10 | M1M | 43.71632 | -79.2395 |
| 11 | M1N | 43.69266 | -79.2648 |

Figure 2. Geospatial data.

**3.** Foursquare API

Provides data about a venue. Depending on the type of account, what and how much information can be obtained. In this project we are mainly interested in the location as well as the type of venue.

Canada postal code data will be combined with geospatial location data to have latitude and longitude for each neighborhood, and then filtered to stay with Toronto only. Then the Foursquare API will be used to obtain the venues in the different Toronto neighborhoods and finally only those that correspond to gyms will be selected. Based on this we will obtain the clusters using ML algorithms and we will analyze them in order to provide a recommendation.