# Data Description

## a. Data Sources

1. Wiki page having the details of the postal codes, Borough and Neighborhood of Canada was used to extract the data and read into data frame using pandas. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. Google API was used to gather the coordinates for each of the postal code

## b. Data Cleaning

Before processing the gathered data, it was cleaned thoroughly to remove all the non-required information and convert the data ready for use.

1. Initially all the missing column values of the column 'Borough' were removed as they were of no significance.
2. Neighbourhoods having same postal codes were combined into one separated by commas.
3. For each of the postal code respective Latitude and Longitude coordinates were added using Google API.
4. Finally, the postal codes column was dropped as it was irrelevant in the further analysis.

## c. Data Information

Below is the description of the data columns that are used for the analysis:

| Postal code | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|

1. **Postal code**: It defines the particular Neighbourhood in a city, it is required for calculating the latitude and longitude coordinates.
2. **Borough**: A town or a district which is an administrative unit, for every Borough there are multiple neighbourhoods
3. **Neighbourhoods**: Basically, a community within a city, there may be multiple neighbourhoods for a particular postal code.
4. **Latitude and Longitude**: The geospatial coordinates defining a particular location, it is required for Foursquare API for analysis and information gathering.

After cleaning the data our ready to use data looked like:

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Scarborough | Malvern , Rouge | 43.806686 | -79.194353 |
| 1 | Scarborough | Rouge Hill , Port Union , Highland Creek | 43.784535 | -79.160497 |
| 2 | Scarborough | Guildwood , Morningside , West Hill | 43.763573 | -79.188711 |
| 3 | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | Scarborough | Cedarbrae | 43.773136 | -79.239476 |