IBM Data Science Capstone Project

Week 1

Question 2

# Data Overview

To investigate the neighborhoods in two major cities of the province of Alberta Canada, we need to acquire six sets of data:

1. List of the neighborhoods in Edmonton
   1. This set of data must be retrieved from Wikipedia webpage.
2. Geographical location of each neighborhoods in Edmonton
   1. I use Georecorder package in python to export this set of data.
3. Venue data including the existing venues in each neighborhood in Edmonton
   1. I use Foursquare to retrieve venue data.
4. List of the neighborhoods in Calgary
   1. This set of data must be retrieved from Wikipedia webpage.
5. Geographical location of each neighborhoods in Calgary
   1. I use Georecorder package in python to export this set of data.
6. Venue data including the existing venues in each neighborhood in Calgary
   1. I use Foursquare to retrieve venue data.

## List of Neighborhoods

To have the list of the neighborhoods I have used Wikipedia webpage as a source. I first found the postal codes of neighborhoods in Alberta which is starting with T. Then searched for the list of the postal codes in Canada within the wiki webpage and as shown in Figure 1, I retrieved the list for both Edmonton and Calgary neighborhoods.

Table

Description automatically generated

Figure : A screenshot of list of neighborhoods in different cities in the province of Alberta, Canada.

Since data was in a suitable format for this project, I have scrapped it, cleaned Not Assigned neighborhoods, and exported them to two separate data frames one for Edmonton and the other for Calgary.

Graphical user interface, table

Description automatically generated

Figure : Edmonton and Calgary neighborhood data frames after retrieving and scrapping

Using the Geocoder package, I have added geospatial coordinates of centroids of each neighborhood to the Edmonton and Calgary neighborhood data frames. Figure 3 shows an example of it for Edmonton.

Graphical user interface, text, application, email

Description automatically generated

Figure : Adding geospatial coordinates to the data frame using Geocoder package

Then after, I used Foursquare website to export venues in each neighborhood in Edmonton and Calgary. Figure 4 shows the list of venues by neighborhoods in the city of Edmonton.

Table

Description automatically generated

Figure : Venue data retrieved and stored into a data frame from the Foursquare webpage for the City of Edmonton.