1. **DATA ACQUISITION AND CLEANING**

The data used for this project comprises of three sources:

1) List of Neighbourhood in Auckland

2) Latitude and Longitude of the desired neighbourhoods.

3) Venue data obtained from Foursquare app

Here the scope of this project is confined to the city of Auckland, New Zealand.

The first data is scraped from a Wikipedia page using the Beautiful Soup library in python. With the help of this library, we can extract the data in the tabular format as shown in the website. (Website- https://en.wikipedia.org/wiki/List\_of\_suburbs\_of\_Auckland) This is a list of 64 suburbs in the Auckland metropolitan area, New Zealand, surrounding the Auckland Central Business District. They are broadly grouped into the local government areas that existed from 1989 to 2010.

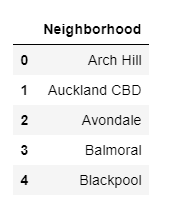


Figure 2.1: List of first five Neighbourhoods after scrapping

Latitude and longitude coordinates of those neighbourhoods are required in order to plot the map and also to get the venue data. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give the latitude and longitude coordinates of the neighbourhoods.



Figure 2.2: Latitude and Longitude details are added using Geocoder

Venue data, particularly data related to fruits and vegetable stores. We will use this data to perform clustering on the neighbourhoods. Foursquare API is used to get the venue data for the neighbourhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide different categories of the venue data, we are particularly interested in the category in order to help us to solve the business problem put forward.



Figure 2.3 Venue Category added using Foursquare API