**2.0 Data Acquisition and Cleaning**

Data used in this report is acquired from list of postal districts (Urban Redevelopment Authority Singapore 2021) as well as its respective postcode and is categorized into the 5 distinct districts in Singapore (Mapanet/ Divex 2021). These data are compiled into a csv file and a link for the csv file can be seen under **2.1 Data Sources**.

**2.1 Data Sources**

Singapore postal obtained from the various sources mentioned in **2.0 Data Acquisition and Cleaning** are compiled into a csv file for ease of handling. The file is uploaded to my github for references and the link is:

* <https://github.com/edbertfernando/Coursera_Capstone/blob/main/Singapore_Postal.csv>

Another data used in this report is from Foursquare. Foursquare data will be used for segmentation and clustering.

**2.2 Data Cleaning**

Graphical user interface, text, application

Description automatically generatedThe first few rows of the csv data file can be seen as follow:

Grouping Singapore’s general locations into the distinct 5 regions.

Table

Description automatically generated

From the table, Central Singapore has the highest number of locations.

Using Geocode.xyz, I decided to find the latitude and longitude of each of the general location in Singapore. The file is uploaded to my github and the link to the file is as follows:

* <https://github.com/edbertfernando/Coursera_Capstone/blob/main/Singapore%20LongLang.csv>

Combining the two data frames gives the following:

Graphical user interface, application

Description automatically generated

Map

Description automatically generatedFrom the data of latitude and longitude of the different locations in Singapore, a map can be created with pointed area on it.