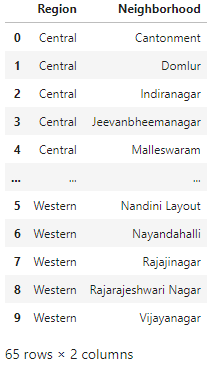
**Applied Data Science Capstone**

**Data :**

We will be dealing with location data for Bangalore. The following are the datasets I have used.

**Data 1: Contains regions in Bangalore and the neighborhoods within them**

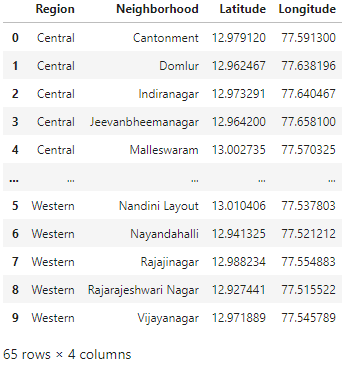
* First we need the various neighborhoods in the different regions in Bangalore.
* I have manually copied the region and neighborhoods into a list and turned it into a dataframe, one for each region.
* Finally I have concatenated the individual region dataframes to arrive at the final dataframe for locations, which looks like this:



Source/Tool: <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore>

**Data 2: Contains the co-ordinates of the neighborhoods above**

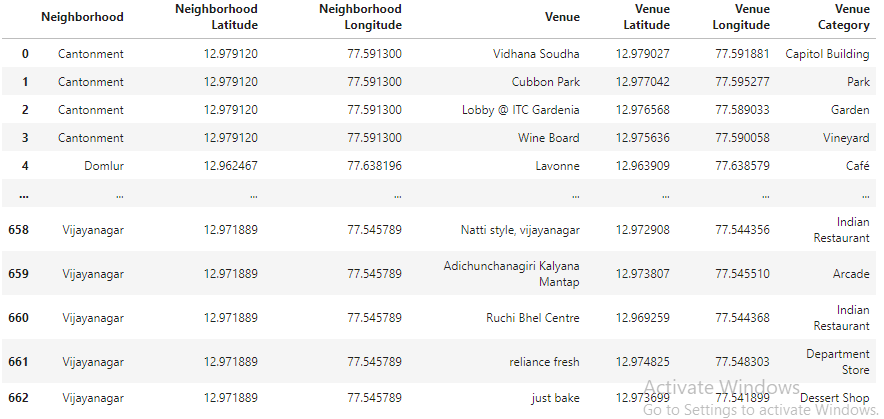
* To get the latitude and longitude values for each neighborhood, I have made use of the python package geocoders in geopy.
* After using geocoders and getting the latitudes and longitudes, our dataframe looks like this:



Source/Tool: Python package Nominatim in geopy.geocoders.

**Data 3: Contains the different venues in each of the neighborhoods**

* To retrieve the venues in all the neighborhoods, I have used the Foursquare API.
* The final dataframe containing the regions, neighborhoods, their coordinates, the venues and the venue coordinates and category is as shown:



Source/Tool: Foursquare API

​