# **Data Description**

Data for this project was collected from various sources which are listed below:

* Wikipedia: The list of neighbourhoods in Bengaluru was web scraped from Wikipedia. A few more neighbourhoods were also added as the list gathered missed some localities in Bengaluru. The attributes of this dataset are explained in Table 1.

**Table 1: Neighbourhoods from Wikipedia**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO. | ATTRIBUTE | TYPE | DESCRIPTION |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Area | String | The area the neighbourhood belongs to – north, south, east, west, central, etc. |

* Google: Initially the library ‘Nominatim’ was used to fetch the geographical co-ordinates of neighbourhoods, but it was observed that some results were incorrect. Geographical co-ordinates of a few neighbourhoods were checked from Google and were found to be more accurate. Subsequently, a small process was developed in the RPA tool UiPath which picked neighbourhood names from a CSV file, fetched their respective co-ordinates from Google search and populated them back in the file. The attributes of this dataset are explained in Table 2.

**Table 2: Bengaluru neighbourhoods (Master table)**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO. | ATTRIBUTE | TYPE | DESCRIPTION |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Area | String | The area the neighbourhood belongs to – north, south, east, west, central, etc. |
| 3. | Latitude |  | Latitude of the neighbourhood. |
| 4. | Longitude |  | Longitude of the neighbourhood. |

* Foursquare: Details of venues in the neighbourhoods were gathered using Foursquare APIs. The attributes of this dataset are listed in Table 3.

**Table 3: Venue data from Foursquare**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO. | ATTRIBUTE | TYPE | DESCRIPTION |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Area | String | The area the neighbourhood belongs to – north, south, east, west, central, etc. |
| 3. | Neighbourhood Latitude | String | The latitude of the neighbourhood. |
| 4. | Neighbourhood Longitude | String | The longitude of the neighbourhood. |
| 5. | Venue | String | The name of the venue. |
| 6. | Venue Latitude | String | The latitude of the venue. |
| 7. | Venue Longitude | String | The longitude of the venue. |

* Makaan.com: The dataset of cost of buying an apartment in various neighbourhoods was web scraped from the website Makaan.com. The attributes of this dataset gathered are listed in Table 4.

**Table 4: Real estate data from Makaan.com**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.NO.** | **ATTRIBUTE** | **TYPE** | **DESCRIPTION** |
| 1. | Neighbourhood | String | The name of the neighbourhood. |
| 2. | Price Range per sqft | String | The price range of buying an apartment per sqft. |
| 3. | Average Price per sqft | String | The average price of buying an apartment per sqft. |

* 99acres.com: This website to fill in the values missing in the dataset prepared from Makaan.com.