**Data Acquisition and Preprocessing**

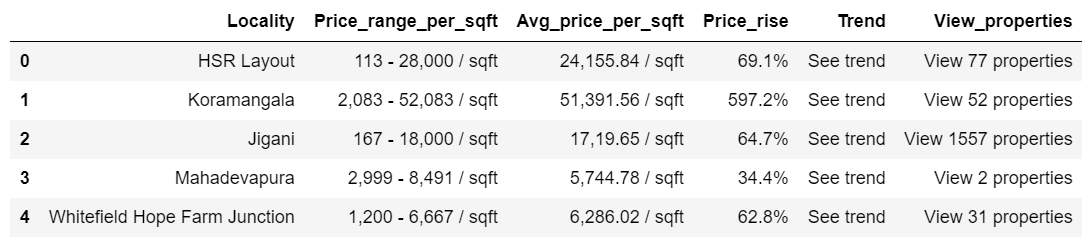
**Data Acquisition**

To solve the above problem, collected the following data along with **Foursquare** location data.

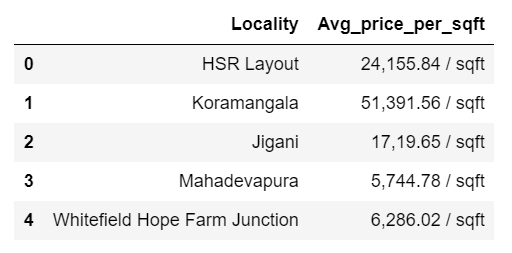
1. Collected the **real estate prices** data from Makaan.com which is an online real estate portal in India. The data has the following information
   1. **Locality Name** - various localities in Bangalore
   2. **Price range per sqft** - Price range per square feet in INR
   3. **Avg price per sqft** - Average price per square feet in INR
   4. **Price rise** - Price rise in percentage
   5. **Trend** - Price trend of each locality
2. Acquired most common **venues** data of a given locality of Bangalore via **Foursquare** API

**Data Preprocessing**

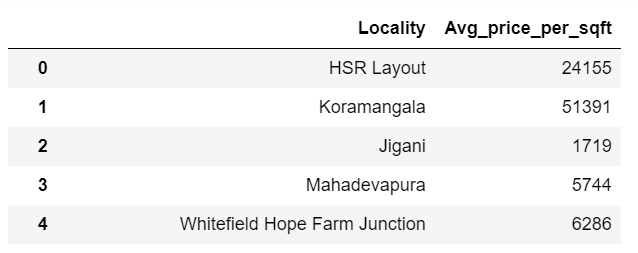
Used web scraping technique to collect the latest real estate prices data in Bangalore from makaan.com web portal. The raw data has multi-index columns. After Transformed the multi-index to single index columns, the data is shown below.



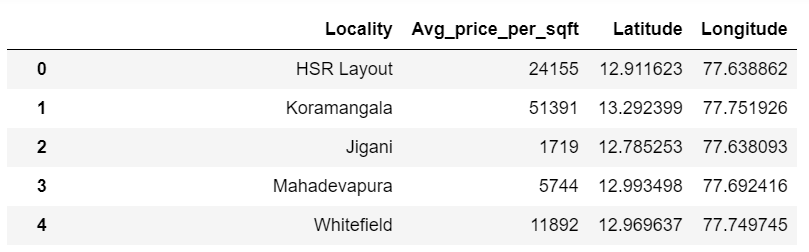
The data shape is (4993, 6). The price data is not available for most of the localities in the data. Once dropped the observations for which price is not available and un-wanted features/columns, the data set looks as follows with shape (869, 2)



The Average price column has string data type. So, we need to convert it into integer/float. After removed the unnecessary string such as **‘ / sqft’** and converted the data type to integer, the data is as follows.



Used **geopy** to obtain coordinates of the localities in Bangalore. Some of the localities, the geopy did not find coordinates. So, as a final data set, took the localities which have proper coordinates. The **final data** set looks as follows with coordinates and **shape** (570, 4)



The above final data set is used to obtain venues for each locality using **Foursquare** API.