**The Battle of Neighborhoods in Bangalore**

**Sreenivasulu Parimi**

**23 March 2020**

**Introduction**

**Background**

**Bangalore**, officially known as **Bengaluru**, is the capital of the Indian state of Karnataka. It has a population of over 13 million, making it a megacity and the third-most populous city and fifth-most populous urban agglomeration in India. It is situated in the southeastern part of the state on the Deccan Plateau at an elevation of over 900 m (3,000 ft) above sea level, which is the highest among India's major cities. It is multi-ethnic, multi-religious, and cosmopolitan character. Bengaluru is sometimes referred to as the **Silicon Valley of India** (or **IT capital of India**) because of its role as the nation's leading information technology (IT) exporter. A demographically diverse city, it is the second fastest-growing major metropolis in India. It has one of the most highly educated workforces in the world. It is home to many educational and research institutions.

**Problem Statement**

BDA projects that by 2031, the city will have 20.3 million residents. Bangalore is proving to be a shining spot for real estate investors. Due to good market drivers, IT/ITes sector, which ensures strong macro -economic dynamics, most home buyers find Bengaluru a perfect place to settle and retire. It is not a surprise then that the real estate demand in the city arises majorly from the migrant population.

Real estate requires large sums of capital and is not as liquid as other investment avenues like mutual funds, stocks and bonds. Real estate investment in Bangalore is very tricky, not all investments are in profit. **Locality** is the most important factor for profitability in real estate investment. Proximity to amenities, peaceful conforming areas, neighborhood status, scenic views, etc. are major factors for **residential** property valuations. While proximity to markets, warehouses, transport hubs, freeways, tax-exempt areas, etc. play an important role for **commercial** property valuations.

So, real estate investor should invest only if he/she is very sure about the commercial developments in that location and proximity to all amenities as its the backbone for one’s rentals and sales. At the same time, the investment should be affordable.

Its very challenge to a new real estate investor to find such lucrative area in such a big city. By looking at the above pros and cons of real estate investment, as a data scientist, provide a solution which should help to the investor to find the desirable place.

**Target Audience**

Individual or corporate real estate investors who are interested in growing Bangalore’s real estate and explore its neighborhoods & common venues around each neighborhood.

**Data Acquisition and Preprocessing**

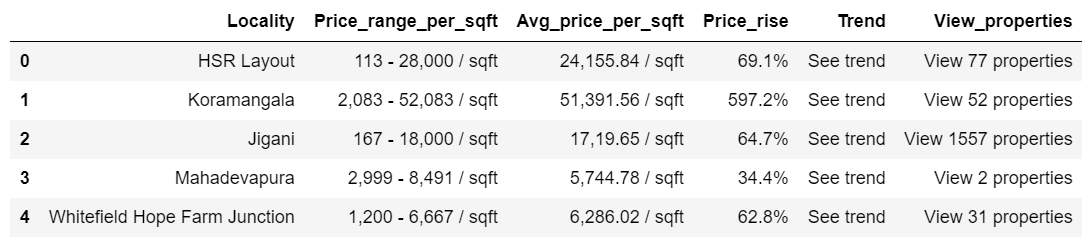
**Data Acquisition**

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

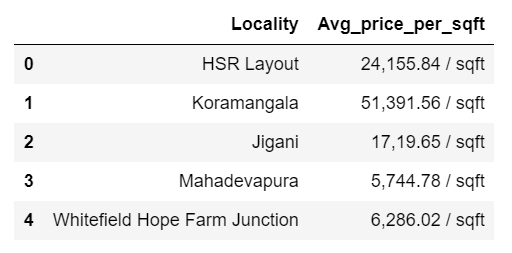
1. Collected the **real estate prices** data from Makaan.com which is an online real estate portal in India. The data has the following information
   1. **Locality Name** - various localities in Bangalore
   2. **Price range per sqft** - Price range per square feet in INR
   3. **Avg price per sqft** - 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**

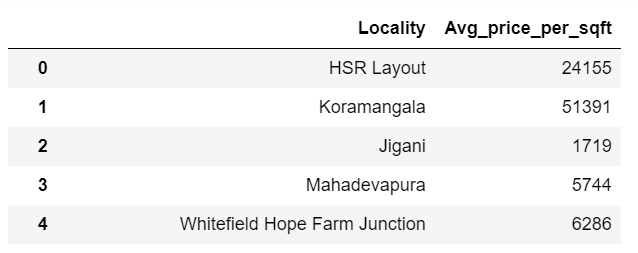
Used web scraping technique to collect the latest real estate prices data in Bangalore from makaan.com web portal. The raw data has multi-index columns. After Transformed the multi-index to single index columns, the data is shown below. The data shape is (4993, 6).



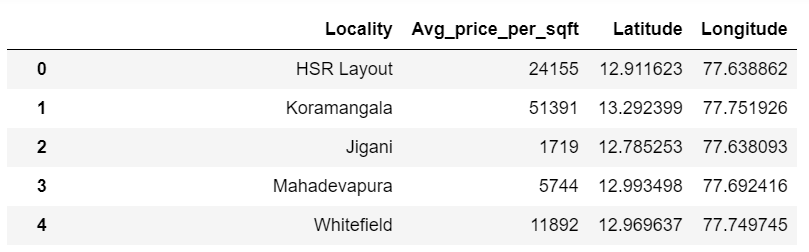
The price data is not available for most of the localities in the data. Once dropped the observations for which price is not available and un-wanted features/columns, the data set looks as follows with shape (869, 2)



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** (559, 4)

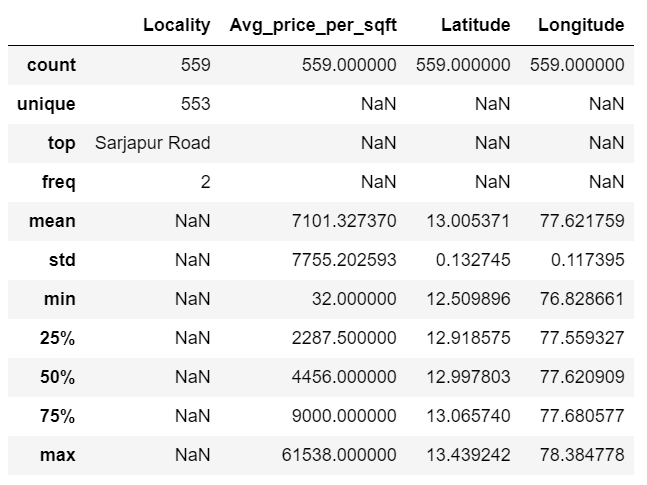


The above final data set is used to obtain venues for each locality using **Foursquare** API.

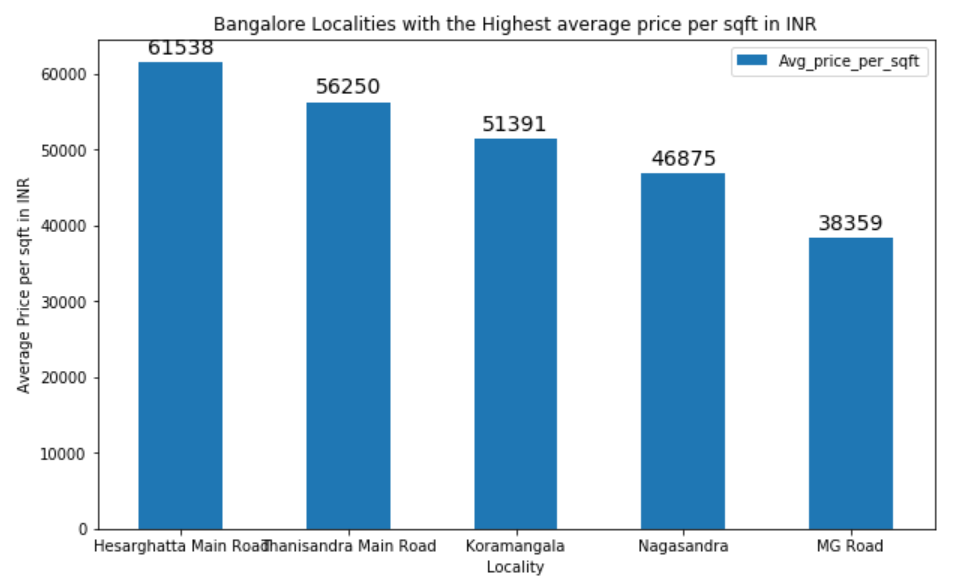
**Methodology**

**Data Exploration**

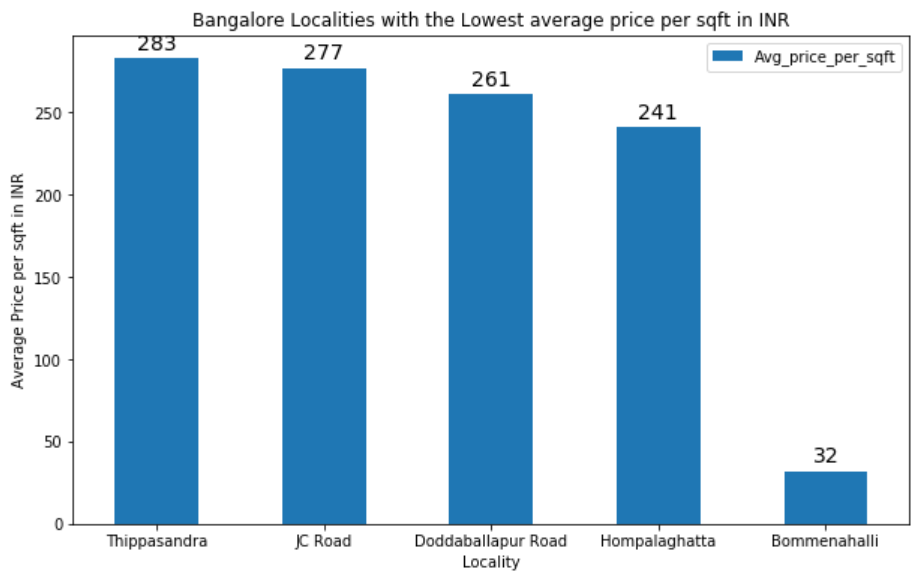
**Statistical Summary**: - Used pandas dataframe’s describe function to obtain the summary of the Bangalore’s real estate price data. From this, we can say that maximum INR 61538 and minimum INR 32 whereas mean INR 7101 average price per sqft.



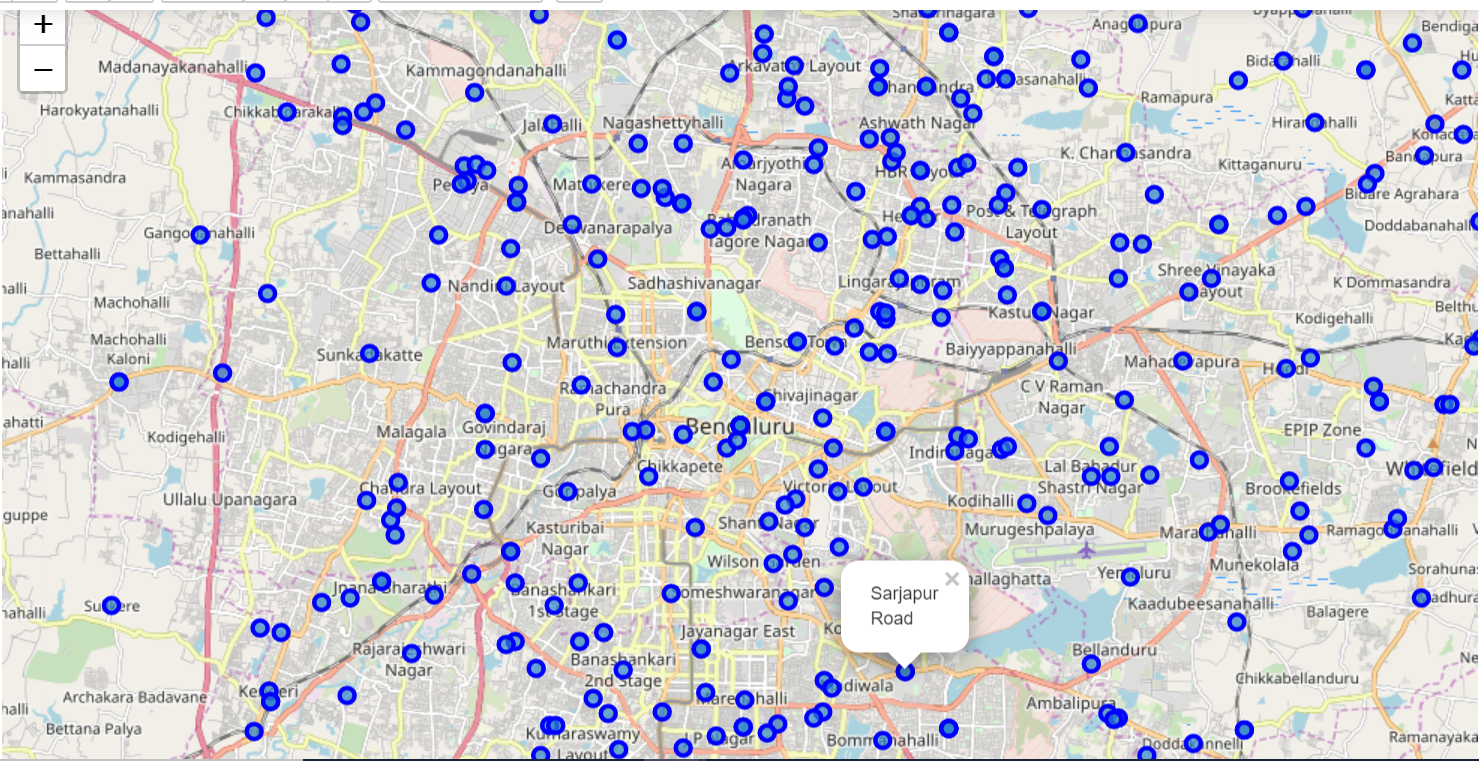
**Localities with the Highest Average Price per sqft**: - Depicted bar graph to visualize the top 5 localities with the highest average price per sqft. We can see that the areas with highly commercialized and proximity to best amenities demand premium prices.



**Localities with the Lowest Average Price per sqft**: - Depicted bar graph to visualize the bottom 5 localities with the lowest average price per sqft. It says that the areas with not commercialized and slum demand low prices.



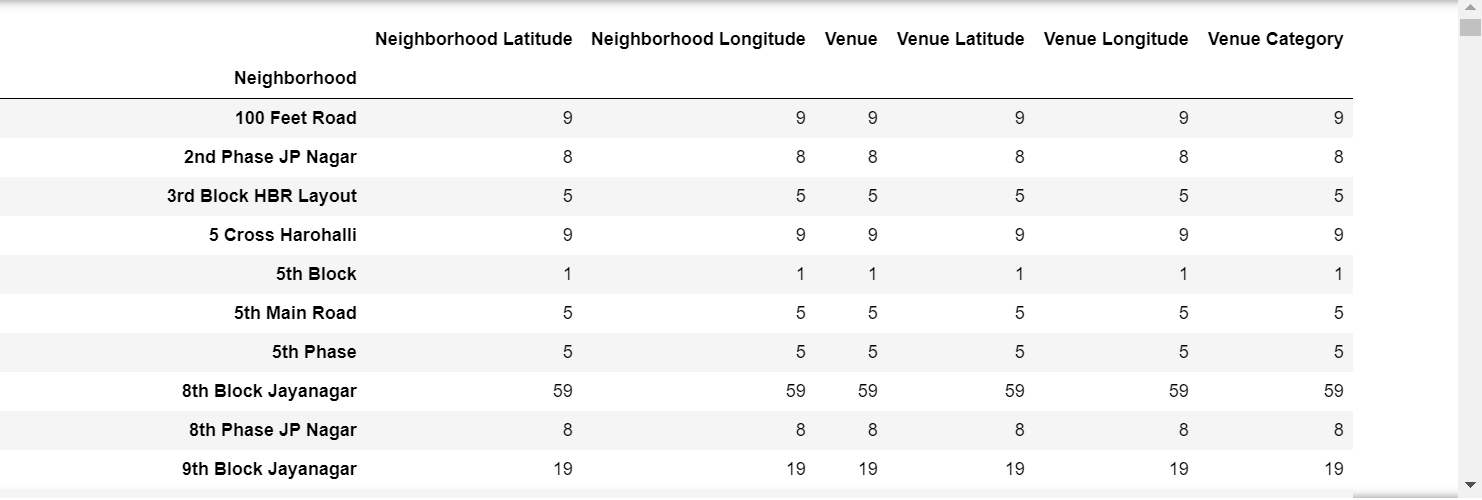
**Map of Bangalore with its Localities/Neighborhoods**: - To visualize Bangalore’s geographic details and its neighborhoods, used python **Folium** library. First created a map of Bangalore and then superimposed its localities by using their coordinates.



**Explore neighborhoods of Bangalore**: - To explore the neighborhoods and segment them, used **Foursquare** API. Here, listed the list of venues, category, latitude and longitude which were returned from the Foursquare API when **limit** (no of venues return by Foursquare) as 100 and **radius** as 500 meters (distance for each neighborhood from their latitude & longitude).



No of venues returned by Foursquare AIP for each neighborhood are depicted here.



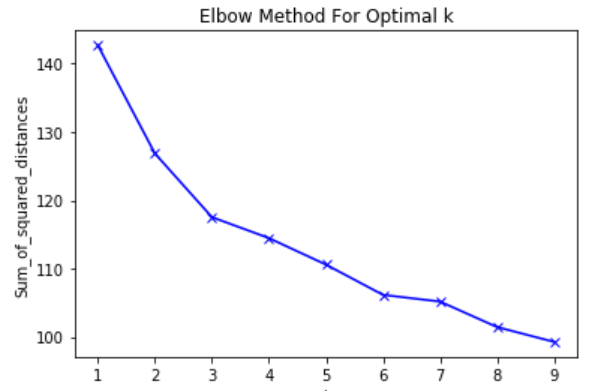
**Analyze each neighborhood**: - Finally generated a new data set, which shows the top 10 most common visited venues for each neighborhood in Bangalore, with the help of Foursquare API.



**Machine Learning**

Chosen, frequently used un-supervised machine learning, K-means algorithm to cluster neighborhoods since there are some common venue categories in the neighborhoods. K-means is used to cluster or segment similar objects/items.

Ran K-means to cluster the neighborhoods into 3 clusters as elbow method return the optimal number of clusters as 3 for this data. Elbow method is one of the mostly used techniques to find the best number of clusters in K-means.

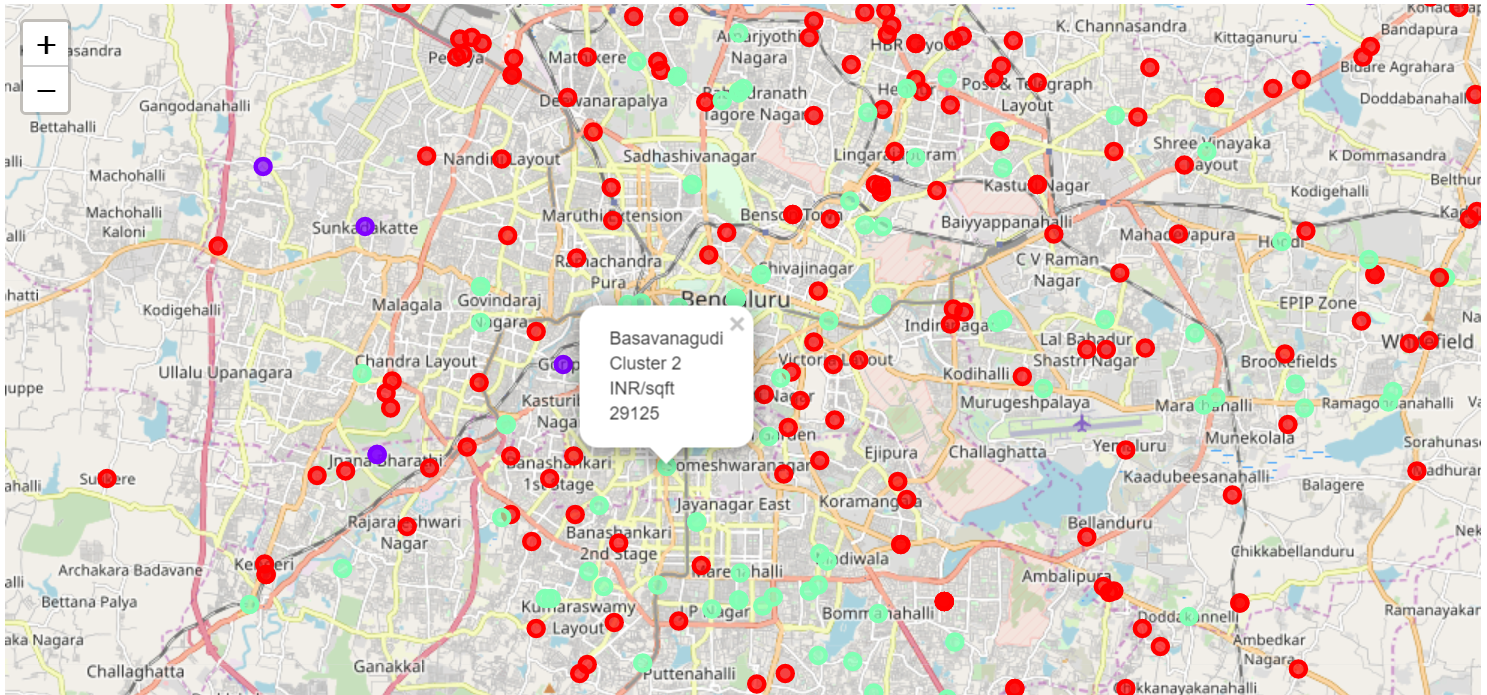


**Results**

The following picture show the resulted data frame that includes cluster as well as the top 10 venues for each neighborhood after ran the K-means with 3 clusters.

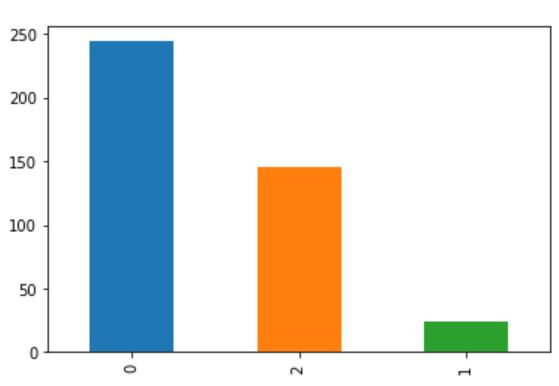


Visualized the cluster map below.



1. Red - Cluster 0
2. Blue - Cluster 1
3. Green - Cluster 2

There are 244, 145 and 24 neighborhoods in Cluster 0, Cluster 2 and Cluster 1 respectively. The visualization for the same is here.

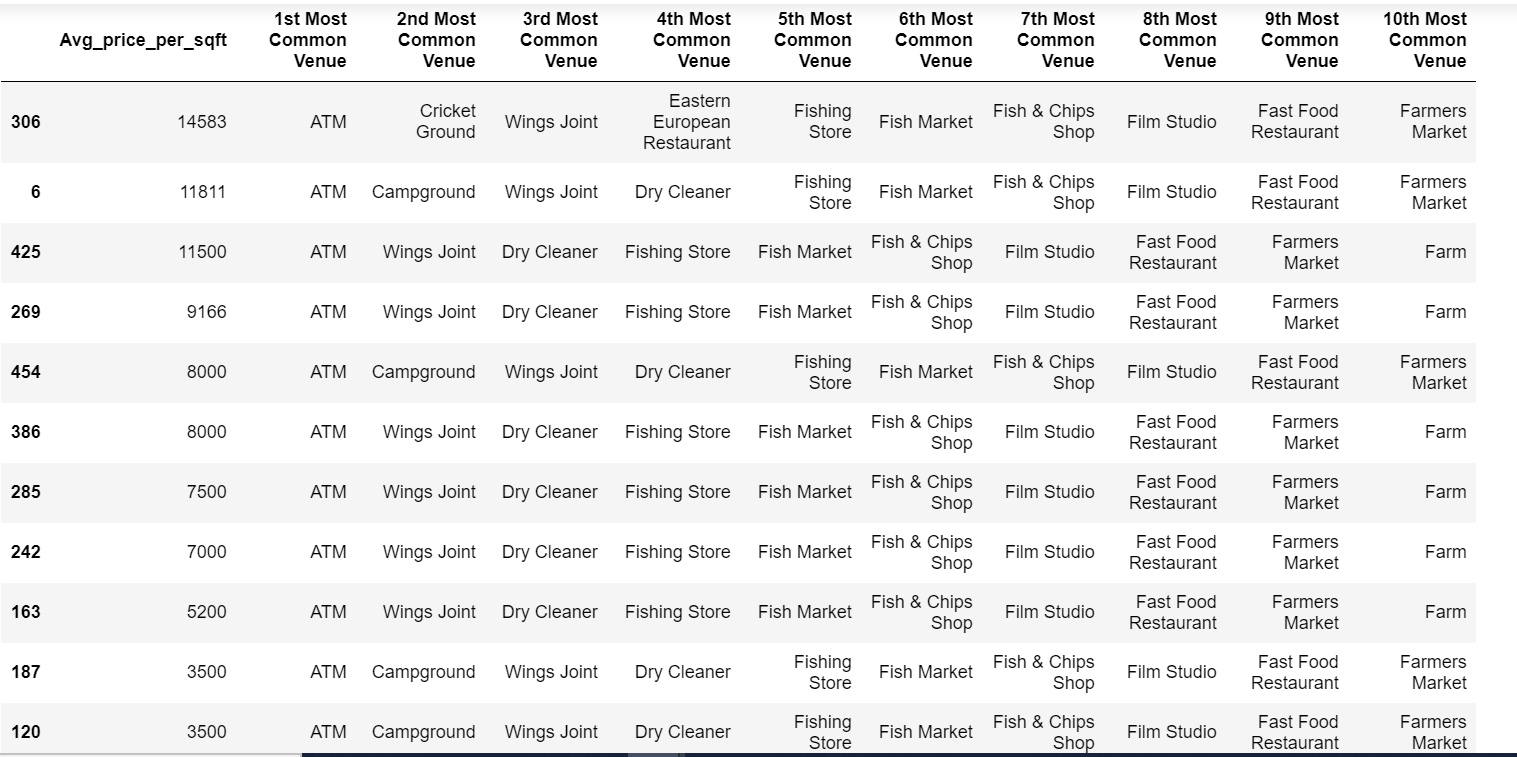


**Examine the clusters**: - Looking into the below tables, we can see that most common venues in cluster 1 neighborhoods are restaurants & cafes and cluster 2 neighborhoods are ATM & Wings Joint whereas cluster 3 restaurants & brewery.

**Cluster 1**



**Cluster 2**



**Cluster 3**



All the above analysis resulted that the real estate prices are high in localities where there is more proximity to amenities, peaceful conforming areas, neighborhood status, scenic views, transport hubs, markets, sport facilities and restaurants & bars etc.

**Discussion**

The purpose of the project is to assist people who want to relocate/invest in the localities of Bangalore based on the good neighborhoods and prices. For example, if a person who is looking for proximity to amenities such as restaurants, departmental store, transport hubs etc, he/she can look at cluster 3.

**Conclusion**

We conclude that this helps people to easily interpret where to live and invest with desirable amenities and affordable price in the neighborhoods of Bangalore.