

Residential Multistory Apartment Sale Price Analysis for Bangalore – India

Jothi Nadarajan

May-2019

Table of Contents

Residential Multistory Apartment Sale Price Analysis for Bangalore – India	1
1. Introduction	3
Problem Description & Background	3
2. Data Description	3
a. Data sources	3
b. Data cleaning	4
3. Methodology	4
i. Fetching Location Co-ordinates of Localities	4
ii. Retrieving the Venue details for each Localities	5
4. Results	5
i. Frequency distribution of Price Range	5
ii. Visualization of Localities in Bangalore with avg. sales Price	6
iii. Sale Price Vs Venues	7
iv. Clustering & Visualization using Choropleth Map	8
5. Discussion section	10
6. Conclusion	10
7. References:	10

1. Introduction

Problem Description & Background

Bangalore (also called as Bengaluru) is one of the metropolitan city in India. It has a population of over ten million, making it a megacity and the third-most populous city and fifth-most populous urban agglomeration in India, located in an area of **741 km²** of the metropolis. The population density of the city of Bangalore is **4,381 persons per square kilometer**. The population growth has extended by 47% in only ten years in light of creating open entryways for employment and growth that are attracting individuals from everywhere throughout the nation. Bangalore is also referred as "Silicon Valley of India" (or "IT capital of India") [1].

A demographically diverse city, Bangalore is the second fastest-growing major metropolis in India. Bangalore suffers from the same major urbanization problems seen in many fast-growing cities in developing countries. The encouragement of high-tech industry in Bangalore, has not favored local employment development, but has instead increased land values and forced out small enterprise. As from the figures, Bangalore has high population and population density. When we think of the city residents, they may want to choose the regions where real estate values are lower. At the same time, they may want to choose the locality according to the social places density. However, it is difficult to obtain information that will guide investors in this direction. As city has high population density and land prices are very expensive, residents prefer to buy multistory residential apartments.

Considering all these problems, we can create a map and charts to understand the residential apartment prices in Bangalore and see how each locality is clustered according to the venue density.

2. Data Description

I have found data from following sources and used for this analysis.

a. Data sources

- **Foursquare:** I used Foursquare API to get most common venues for the given Locality of Bangalore [2].
- **JSON File:** I found the Ward-level Administrative Divisions of the Bangalore from Github Repository. The .json file has coordinates of the city and sub urban areas. I cleaned the data and reduced it to city of Bangalore, which I used it to create choropleth map of Residential Apartment Sales Price of Bangalore [3].
- **Real-estate sites:** There are not too many public data related to demographic and real estate prices for the city of Bangalore. I collected latest per square feet Residential Apartment Sale Price Average for each Locality of Bangalore from housing retail web pages. Then I have created data tables using the values got from the retail web pages [4].

b. Data cleaning

Data downloaded or scraped from multiple sources, were combined into one table. I decided to use the average sale Price of Residential Apartments of this year (Q1 2019). I had challenge with name of localities. Some names are similar or the same localities had multiple names. Since I was using json file to plot Choropleth map the names in the file has to match with the names in my data set. For example, data set had Locality name 'K. R Puram' which is also called as 'Krishnaraja Puram'. Therefore, I have to do cleanup to ensure the names are matching in data set and the .json file.

As data collected from multiple retail websites, I found duplicate entries. So removed all duplicate entries. I have also fixed the data format for Avg. Sale Price. Average sale price was having different formats and representation as it was collected from multiple websites.

After fixing these problems, I checked for outliers in the data. I found there were some extreme outliers, mostly caused by localities with same/similar names in or around Bangalore. I have cleaned up them.

3. Methodology

i. Fetching Location Co-ordinates of Localities

I used GitHub repository as a database in my study. Master data contains the main components *Locality*, *Average House Price*, *Latitude* and *Longitude* information's of the city. Price indicates the Average sale price of residential apartment, per Square feet in Rupees. I have used geocoding web services to get the latitude and longitude values for each locality in Bangalore. Here is the sample data.

	Locality	price	Latitude	Longitude
0	Vijayanagar	5,686	12.971889	77.545789
1	Whitefield	5,866	12.969637	77.749745
2	Yelahanka	5,064	13.100698	77.596345
3	Yelahanka New Town	4,895	13.097804	77.581189
4	Yemalur	7,173	12.945464	77.680550
5	Yeshwanthpura	8,025	13.025609	77.554107
6	Basavanagudi	10,108	12.941726	77.575502
7	Begur	4,668	12.863389	77.613011
8	Begur Road	4,844	12.874601	77.625511
9	Bellandur	6,443	12.979120	77.591300

Fig1. Dataset with Sale Price & Locality information

ii. Retrieving the Venue details for each Localities

I utilized the Foursquare API to explore the localities and segment them. I have set the limit as **100 venues** and the radius **750 meter** for each locality from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API. Foursquare returned 1125 venues for the given set of localities.

	Neighborhood	Price	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	A Narayanapura	8490	12.993194	77.670377	McDonald's	12.993667	77.666576	Burger Joint
1	A Narayanapura	8490	12.993194	77.670377	Domino's Pizza	12.995288	77.666318	Pizza Place
2	A Narayanapura	8490	12.993194	77.670377	Nilgiris	12.993644	77.667203	Department Store
3	A Narayanapura	8490	12.993194	77.670377	MedPlus	12.993617	77.666853	Pharmacy
4	A Narayanapura	8490	12.993194	77.670377	Hot Chips	12.993521	77.666746	Snack Place

Fig2. Dataset containing Locality & Venue details

There are 175 unique categories returned by Foursquare, and then I created a table to show the top ten venues for each locality.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 A Narayanapura	Bus Stop	Pharmacy	Snack Place	Department Store	Dessert Shop	Burger Joint	Pizza Place	Bus Station	Yoga Studio	Farmers Market
1 Adugodli	Playground	Ice Cream Shop	Bus Station	Athletics & Sports	Café	Diner	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store
2 Amruthalli	Shoe Store	Yoga Studio	Coffee Shop	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
3 AnjanaPura	Pool	Coffee Shop	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Event Space	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop
4 Ashok Nagar	Café	Indian Restaurant	Pub	Hotel	Clothing Store	Donut Shop	Bar	Shopping Mall	Ice Cream Shop	Department Store

Fig3. Bangalore Localities with most common Venues

4. Results

i. Frequency distribution of Price Range

We can see the frequency of average residential apartment sales prices in different ranges using a Histogram.

Histogram of Residential Property Rates of Bangalore Multistorey Apartment for Jan-Mar 2019

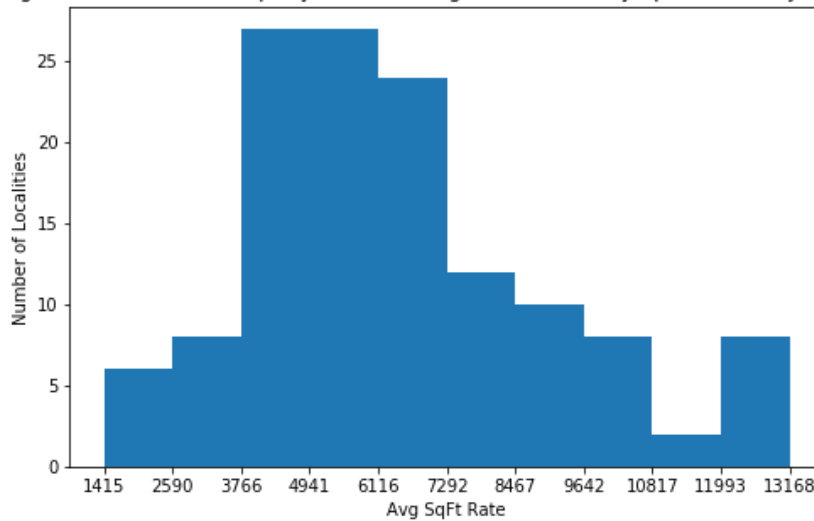


Fig4. Frequency Distribution of Sale Price

From the Histogram, we can see that most of the localities had average sales price in Mid-levels, between 3500 to 6000 per Sq. Feet. Most of the resident's preference will be of this range. Let's define the price ranges in Bangalore for residential apartments based on the histogram data.

- 3500 : "Low Level"
- 3500–6000 : "Mid-1 Level"
- 6000–8500 : "Mid-2 Level"
- 8500–10500 : "High-1 Level"
- > 10500 : "High-2 Level"

ii. Visualization of Localities in Bangalore with avg. sales Price

I used python folium library to visualize geographic details of Bangalore and I created a map of Bangalore with Localities superimposed on top. I used latitude and longitude values to get the visual as below and added markers to show the average sales Price of the locality.

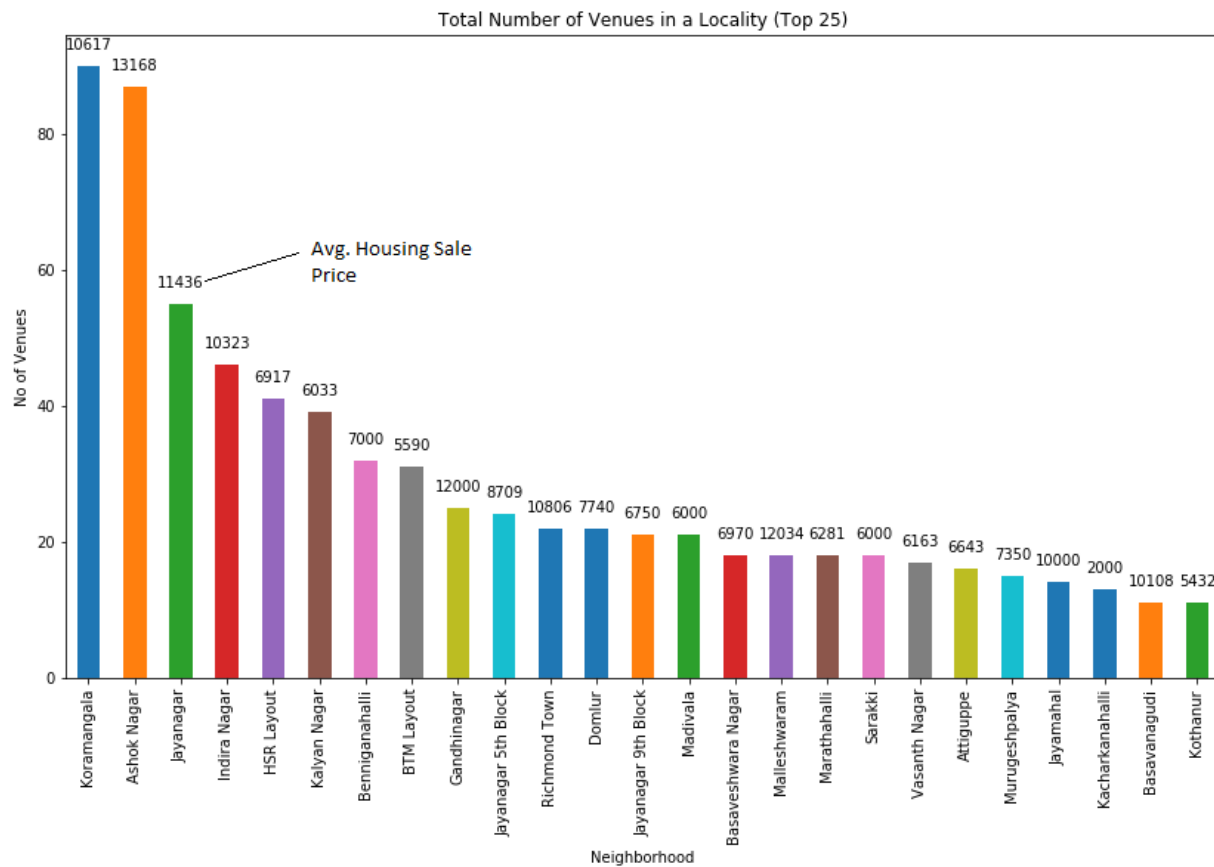


Fig6. Bar Chart to show relation between Venues and Sale Price

iv. Clustering & Visualization using Choropleth Map

We can cluster the localities based on the venue details. I used unsupervised learning **K-means** algorithm to cluster the localities. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I will run K-Means to cluster the Localities into 6 clusters because when I analyzed the K-Means with elbow method it gave me value **6** for optimum k of the K-Means.

Let us visualize the clusters with Choropleth Map with locality names and cluster number added as markers to the map.

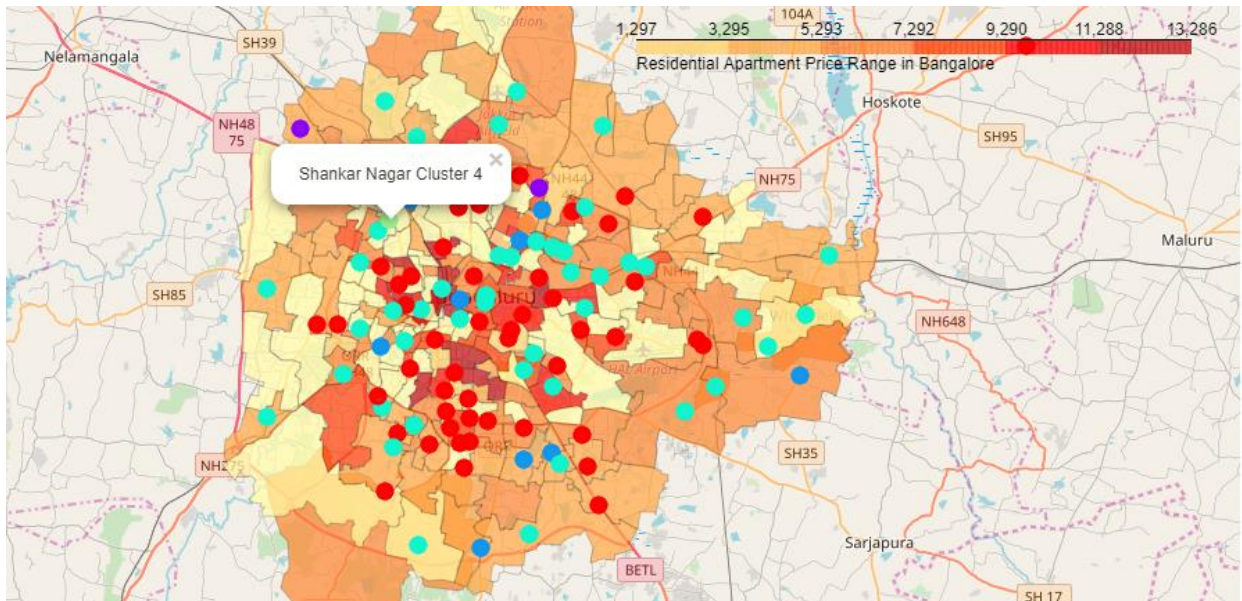


Fig7. Choropleth Map of Bangalore with Clusters & Sale Price

After analyzing the top venues details in each cluster, I have named them as below,

- Cluster 0 – Accommodation & Food Area
- Cluster 1 – Shopping & Cafe Zone (Market, electronics Store & Coffee shops)
- Cluster 2 – Indian Restaurants (High number of Indian Restaurants)
- Cluster 3 – Multiple Social Venues
- Cluster 4 – Market Place
- Cluster 5 – Food Zone

Fig8. Bar chart with Clusters & Most common Venues

5. Discussion section

Bangalore is a big city with a high population density in a narrow area. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropole.

I used the Kmeans algorithm as part of this clustering study with optimum k value of 6. However, I have used only single co-ordinates for each locality to get venue details. For more detailed and accurate guidance, the data set can be expanded by passing more co-ordinates to get venue details.

I have performed data analysis with the information of Localities and residential apartment sale price averages as a static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages or web scraped.

6. Conclusion

In this study, I have analyzed the Residential apartment average sale price of Bangalore localities with its Venues and most happening places. I have identified location co-ordinates using Geolocator option. Top ten venues of localities are analyzed and Clustering done using K Means unsupervised Algorithm. I have also visualized the total venues and the average sale price of each locality. This may help the residents to choose locality of their choice based on venues and average price.

This study focused on the Average Sale price and Venues for each Localities in Bangalore City. This study can be extended by adding more parameters like water availability, pollution levels, transport etc. for each locality, which are the primary factors that any resident will consider before buying a property in Bangalore City.

7. References:

- [1] [Bangalore — Wikipedia](#)
- [2] [Foursquare API](#)
- [3] [JSON File](#)
- [4] [Residential Property Rates-Trends in Bangalore 2019](#), [Property Rates & Trends](#), [Property Rates for Buy in Bangalore](#).