DISCOVERING THE IDEAL LOCALITY FOR RENTING A HOUSE IN BANGALORE

Ashish Arun

03 May 2020

TABLE OF CONTENTS

1.INTRODUCTION

1.1 Background

Based on the data released from Office of the Registrar General & Census Commissioner in 2019 about 50.9% of the population in Bangalore are migrants. That is, these people are leaving their hometowns and starts living in Bangalore. One example of these people would be employees of the IT industry as Bangalore is known as the 'Silicon Valley of India'. IT sector has a fast-changing nature and it is not uncommon for IT employees to shift cities because of their careers. So, majority of IT employees would prefer renting a house rather than owning one. So is the case of students, since Bangalore has some of the premier institutes in the world like IISC, students from all over the world moves in to Bangalore. These changes are well reflected on the houses in Bangalore. In the recent years, single storeyed houses are renovated to multi-storeyed buildings due to the influx of migrant population. Because of that, houseowners or landlords has opportunities for an additional source of income.

1.2 Problem

Cities are expanding their area rapidly, everywhere change is being brought to the surroundings. So, for a person to identify similarities between localities is hard. It is in essence impossible if that person is new to the city. Even for natives, they might not have memorized all areas of the city or there is a good chance that their memory of a least visited locality might be outdated. So, for an individual trying to identify an ideal locality is very much time consuming, one literally needs days to group similar localities. For migrants it may take weeks. Rents for a house has various attributes, but most of the time the major impact on the price is the Locality (Neighborhood).So, grouping localities based on the similarity becomes significant to understand the options available for renting a house. Since people have different reasons to satisfy by renting a house, knowing the different average rent price for a locality inside of a group of similar localities becomes significant.

1.3 Interest

Most migrants to Bangalore would prefer to rent a house. So, identifying an ideal locality which has a rent within her budget is significant for the tenant. Compromising their requirements due to budget constraints is common. This project aims to introduce people to alternate options for a specific type of locality with less compromises. Also, it will help the landlords to identify similar localities and fix a fair price based on the average prices of the similar locality. Over pricing can lead to the property being not rented. Home owners can also identify which group their locality is in and plan for renovation accordingly. Similarly, construction companies can also leverage this information. These are the main target audience of this project. And of course Data science enthusiasts will also be inside target audience.

2. DATA

2.1 Data Requirement

Since this project is aimed to group localities in Bangalore as mutually exclusive clusters based on their similarity, there should be a dataset of the type of buildings/venues on every single locality. List of Locality names and their geographical information is the next requirement. Also we need average rents in these localities, so a dataset on rent prices is also required.

2.2 Data Sources

1. Foursqaure.com:- Foursquare.com provides location based data on major places around the world. It has abundance of crowd sourced data on places, venues and even reviews(tips) on them.
2. Wikipedia:- List of Localities in Bangalore is obtained from this Wikipedia list.
3. 99acres.com:- It is one of India's leading real-estate website in which people can buy, sell, rent properties.
4. Geocode by Awesome Table:- It is an addon in Google sheets which finds the Latitude and Longitude of an Address.

2.3 Data Collection

To obtain data on rent of houses in Bangalore, I scraped the website 99acres.com using Beautiful Soup. Also I used the wikipedia list to get the names of Localities in Bangalore. Then I Geocoded this data to obtain respective latitudes and longitudes of localities. Foursquare provides an API for developers to extract information from their database. In this project we use the explore feature of Foursquare API which returns an user defined number of nearby venues with in a radius from a geographical location. For this project I set the limit for number of venues as 75 and a radius of 1.5 km.

2.4 Data Cleaning and Data Preparation

The scraped rent data had 4060 records of houses available for rent. Attributes like price and area was in object datatype. For example Price –'15,000', Area-'9,000'. These issues were fixed. In this dataset there was about 957 localities, obviously the localities mentioned in the website is very much detailed . However for the sake of simplicity I only took the records where the locality name was matching with the Wikipedia list. This truncated the rent dataset to 656 rows with 46 unique localities. The attributes of the data set were:-

Title	Locality	Building/ Street	Price	Area	Number_of_Bedrooms	Number_of_Bathrooms	Description

Since we are only displaying average price of a locality, localities which had less than 5 records were dropped. The resultant rent data had 34 unique localities. Since this is not a predictive modelling project this should not cause us any issues. If the need to improve the accuracy of average price occurs, all the dropped rows can be used in the future.

Now we needed the venues data from Foursquare. So every Locality was geocoded, i.e. respective latitude and longitude was added. And this details were passed to the Foursquare API. The resultant data had top 75 venues in a radius of 1.5 Km, and its attributes are :-

Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category

There was 156 unique venue categories like ATM, Bus Station, Park, Lake etc. Frequency of each category for every Locality was calculated. Clustering algorithm will cluster localities based on this frequency data. And corresponding clusters will be obtained. Merging these clusters with average prices in the rent data for each locality is the required output.

3.METHODOLOGY

3.1 Extrapolatory Data Analysis

3.1.1 Relationship between Price and Number of Bedrooms

Logically, the rent of property will increase with the number of bedrooms. This relationship is visible in this data set too(Figure 1). However, there is some overlap between 4 bedroom houses and 5 Bedroom houses. This suggests that some other attribute has more impact on the Price. All over the project Price is in INR(₹).

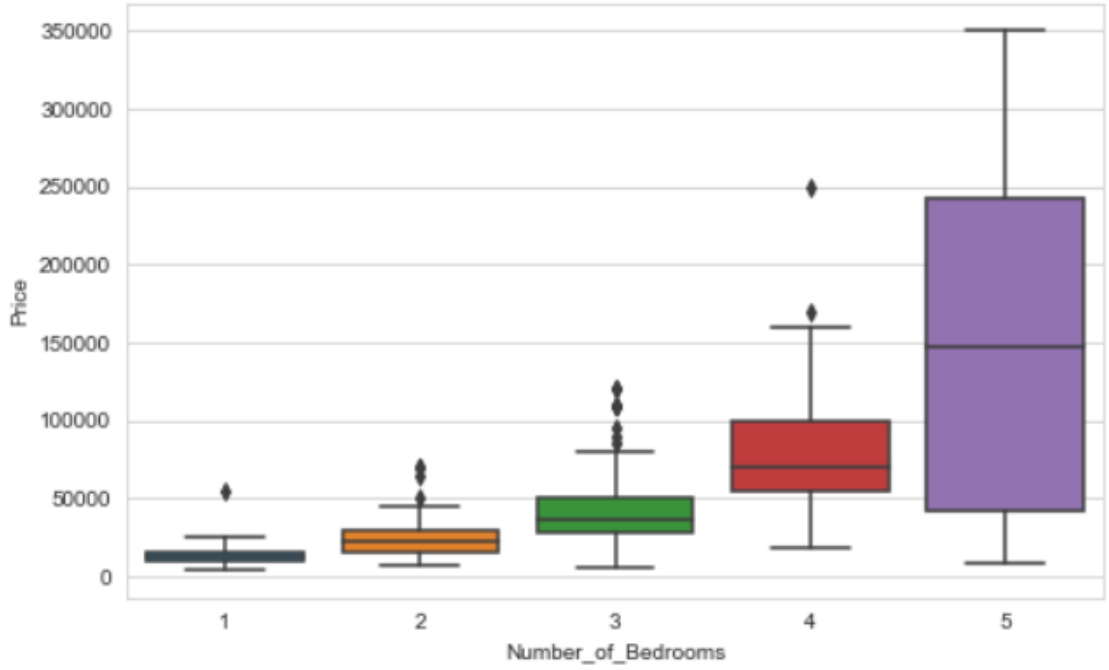

Figure 1: Box plot of price against Number of bedrooms

3.1.2 Relationship between Price and Number of Bathrooms

Evidently the price is increasing with the number of bathrooms (Figure 2). However the number of bathrooms has not impacted the price as the number of bedrooms. Also the medians for 6 and 7 bathroom houses is almost the same price.

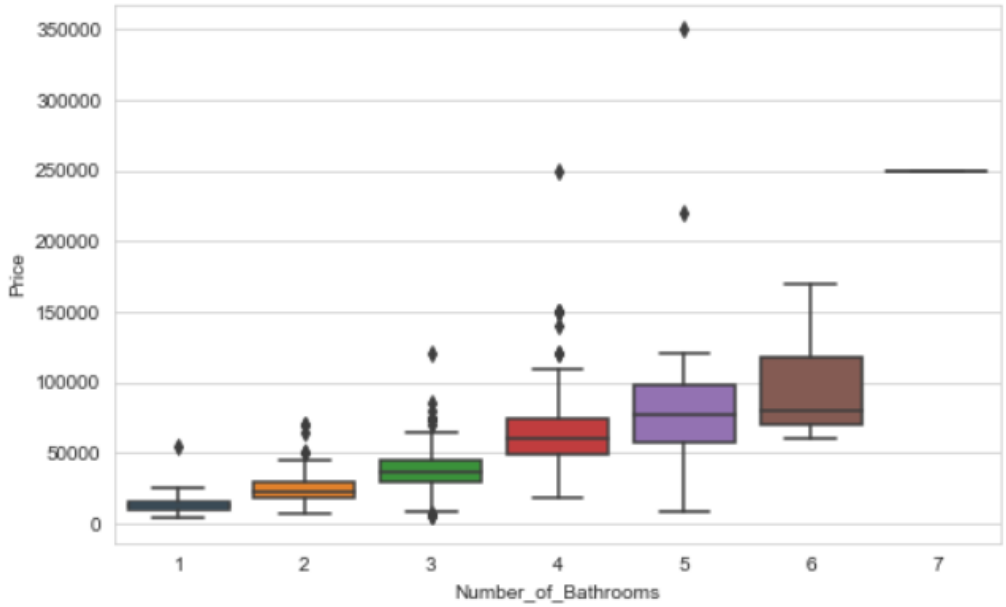

Figure 2: Box plot of Price against Number of Bathrooms

3.1.3 Relationship between Price and Area

As expected price does have a linear correlation with area of the property in sqft. (Figure 3) But it is not clear that is this the attribute that impacted the overlap of prices in the above figure, since the correlation looks weak.

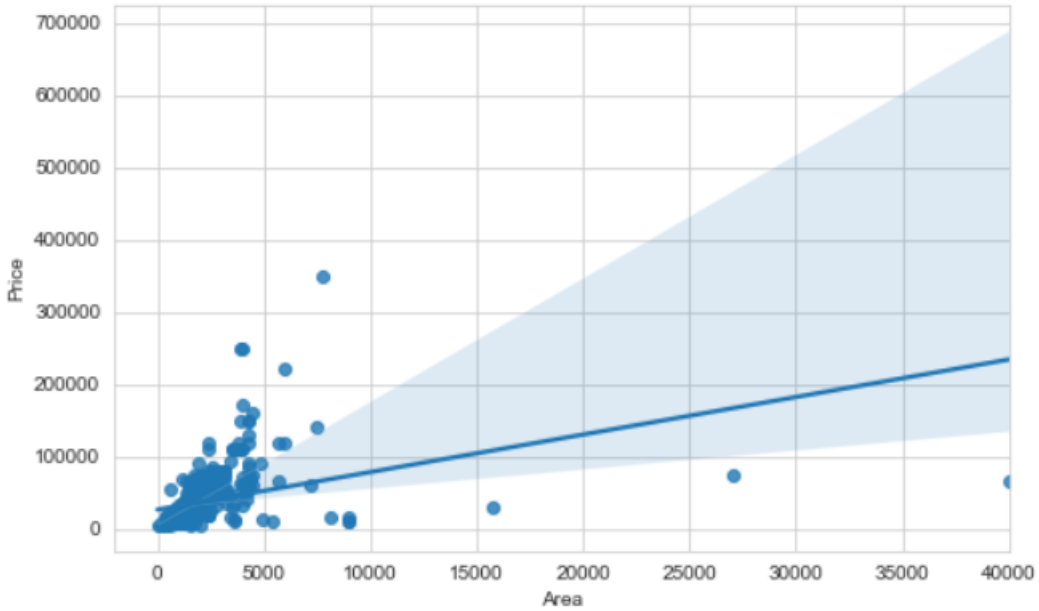

Figure 3: Regression Plot of price against Area

3.1.4 Relationship between Average Price and Locality

Since looking at each individual feature is complex, here we are taking averages of price for each Locality.As per the figure(Figure 4) price does have variations respective to locality. Hebbal has the most average price of all the locality any one who knows Bangalore will agree with this as starting from 2011 Hebbal's Real estate value has gone up. Okay so we can conclude locality has an effect on price.

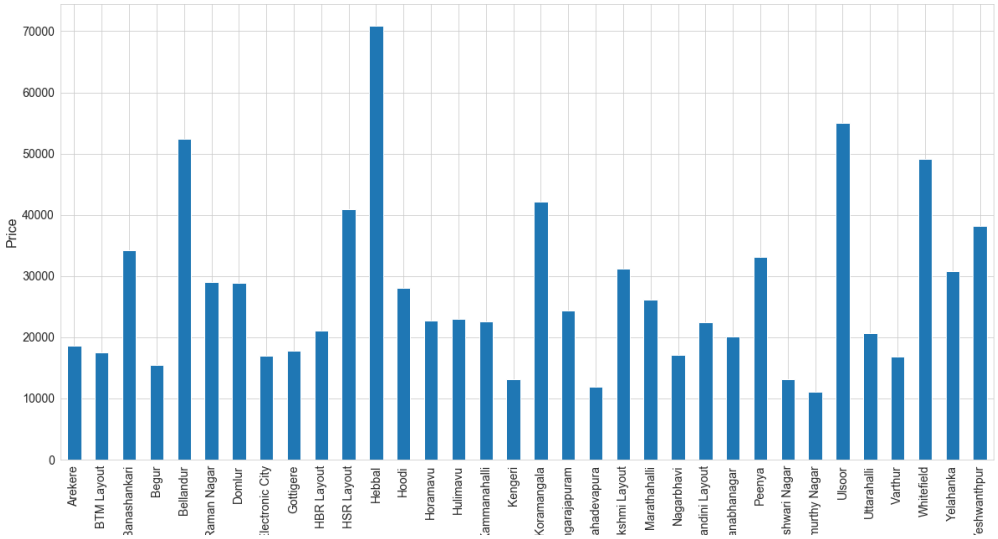

Figure 4: Bar Plot of Average price for Localities

3.1.5 Relationship between Average Number of Bedrooms and Locality

Since we know that number of bedrooms have an effect on the price we would expect the higher the average number of bedrooms is for a locality higher the price. However Figure 5 shows that it is not entirely true.

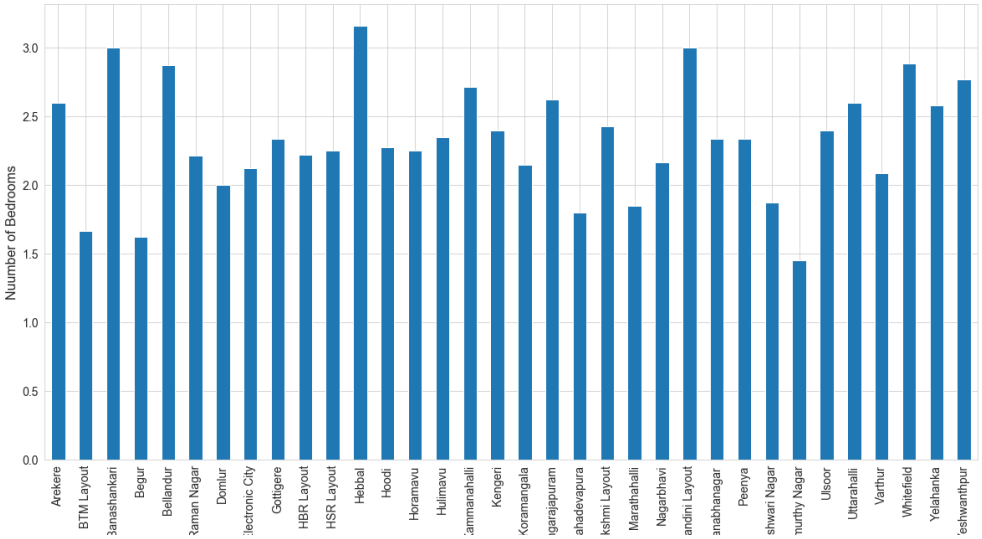

Figure 5 : Bar Plot of Average Number of Bedrooms for Localities

3.1.6 Relationship between Average Number of Bathrooms and Locality

Similar to the case in number of bathrooms, localities with higher prices are not always the ones with highest number of bathrooms.(Figure 6)

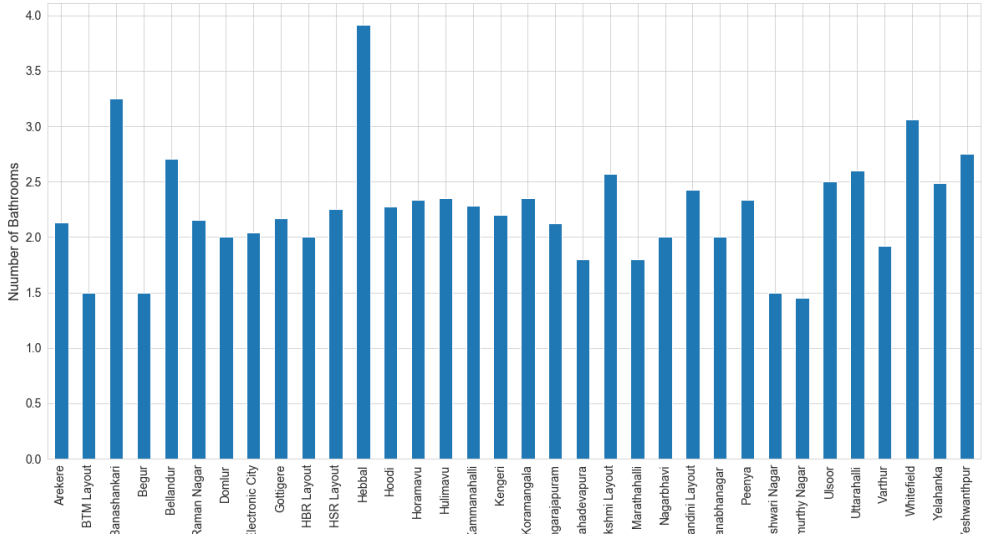

Figure 6 : Bar Plot of Average Number of Bathrooms for Localities

3.1.7 Relationship between Average Area and Locality

Average area is the average indoor area in sqft for a locality. Looking at Figure 7 It is very clear the Higher priced Localities are not the ones with highest area. We saw before area had a weak linear correlation. This is a string indication that it is the **locality** which impacts the price of a property more than any other attributes.

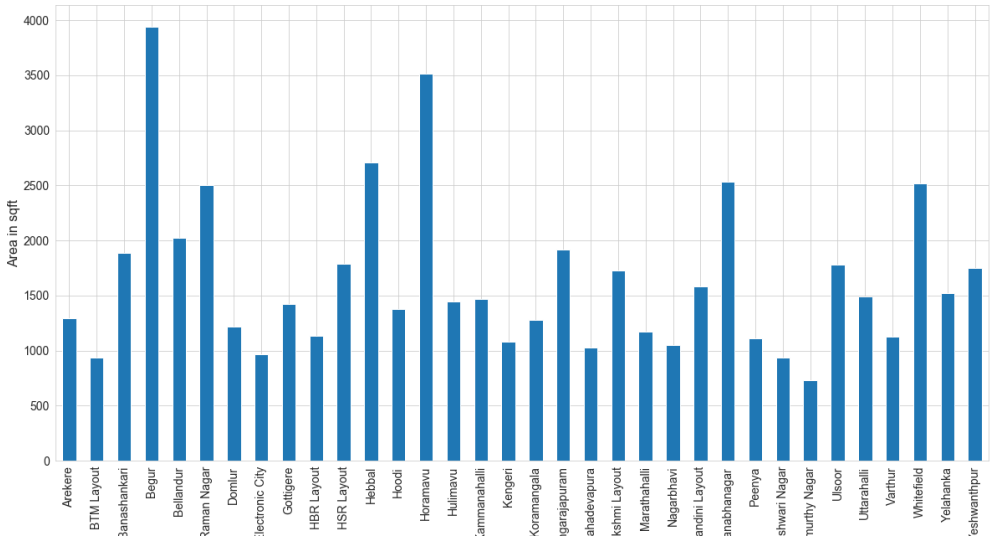

Figure 7 : Bar Plot of Average Area for Localities

3.2 Cluster Modelling

On the EDA section we have established the fact that locality has a bigger impact on the price of a property. So if we clustered similar localities of Bangalore, and know the average price for each cluster a migrant could pick out the best locality to find a house for rent without breaking her budget. Similarity of localities is done by assessing the similarty for nearby venues.

I have used Kmeans clustering model over DBSCAN because DBSCAN clusters based on density which is not very appropriate for this project. The frequency of venue categories occurring from 156 unique categories was the input data of the model. The ideal number of clusters:12 was found by trail and error.

4 RESULTS AND DISCUSSION

4.1 Resultant Clusters

Cluster 0: Cluster 0 is having the oldest localities in Bangalore. These localities are the nearest to the city center.

Locality	Price
Nagarbhavi	17066.666667
Domlur	28825.000000
Banashankari	34262.500000
Koramangala	42100.000000
Ulsoor	54950.00000

Cluster 1: Kengeri is a surrounded by farm areas unlike any other locality so it is clustered alone.

Locality	Price
Kengeri	13170.0

Cluster 2: Yelahanka is more like a satellite town to Bangalore therefore it is a lone member cluster.

Locality	Price
Yelahanka	30811.290

Cluster 3: Hormavu is an emerging locality.

Locality	Price
Horamavu	22683.333333

Cluster 4: This is the main residential cluster of Banglore.

Locality	Price
Ramamurthy Nagar	11063.636364
Rajarajeshwari Nagar	13123.750000
BTM Layout	17483.333333
Kammanahalli	22571.428571
Marathahalli	26107.692308
CV Raman Nagar	29063.636364
Hebbal	70847.368421

Cluster 5: Gottigere is a lakeside emerging locality.

Locality	Price
Gottigere	17850.0

Cluster 6: It is a park locality.

Locality	Price
Nandini Layout	22414.285

Cluster 7: These are border localities of Bangalore.

Locality	Price
Begur	15487.500000
Electronic City	17035.416667
Hoodi	27990.909091
Yeshwanthpur	38117.857143
Bellandur	52435.416667

Cluster 8: Peenya is an industrial Locality.

Locality	Price
Peenya	33166.66

Cluster 9: This is the suburban cluster.

Locality	Price
Mahadevapura	11950.000000
Arekere	18630.000000
Padmanabhanagar	20100.000000
HBR Layout	21022.222222
Hulimavu	23040.400000
Lingarajapuram	24425.000000
Mahalakshmi Layout	31250.000000
HSR Layout	40906.250000
Whitefield	49082.885965

Cluster 10: It is still emerging.

Locality	Price
Uttarahalli	20700.0

Cluster 11: Locality with educational institutions, IT parks.

Locality	Price
Varthur	16808.333333

4.2 Discussions

I would suggest areas like Ramamurthy Nagar, Rajarajeshwari Nagar in the main residential area of Bangalore as the cheapest. We see that in every cluster of localities there are cheaper options. And a very good list for alternative localities are also generated. A person can now choose a locality according to their needs with less compromise. Like for families, Suburban residential areas will be preferable. Students can look at the areas near their institutes. Now only 34 localities been classified. In future this model can be updated with more data and localities.

Also Localities which is in the border cluster Electronic city is near the National Highways which is suitable for people who travel interstate(migrants). In the end it is up to an individual's needs. This project gives people easy, and cheaper alternative options. Home owners in emerging localities like Gottigere should plan for renovating homes to multi storeyed buildings as the migratory residents numbers will grow more in the subsequent years. Landlords in the Localities like CV Raman Nagar, Marathahalli can take note of Hebbal and expect a higher rent in the near future.

5 CONCLUSION

This project was aimed to give an idea about rent price patterns in the city of Bangalore. It was identified that the Locality is the main impactive attribute for the price of a property. Also it was successful to provide alternatives and insights to migrating individuals, current residents in Bangalore on these rent patterns. 34 localities in Bangalore was clustered into 12 mutually exclusive clusters. The clustering model was very accurate as it picked out the Oldest localities, residential localities ,suburbs separately. Also emerging localities were clustered into lone member clusters as these places is still developing in their own ways. Industrial, Park, Lake areas were correctly identified. And average rent of these localities were also representative of the real data.