Project Report

Capstone Project - The Battle of Neighborhoods

Lovey Pathak

1. Introduction – Business Problem

Bangalore, also known as Silicon City of India is a hub of 1.5M employees belonging to various IT and ITeS sector. Significant number of employees residing in various parts of bangalore are from various parts of the country. This also has created a hugh demand of rental properties in various Neighborhoods. This project would concentrate on the need to understand the rental market better and understand if average rental prices differ in various neigborhood and if yes then by how much. This analysis report would help owners understand the variables which impact the rental prices and for tenants this analysis would help them identify neighborhood according to their budgets and needs. Through this report, I would like to answer following questions:

- Which are the top neighborhoods in Bangalore where supply of rental properties are large ?
- What are the ranges of rents in these neighborhoods?
- What are the top 10 venues in these locations?
- Is there any relationship between sq-ft area, furnishings, no of bedrooms, bathroom, house facing on the rental prices?
- Which neighborhoods are similar?

2. Data Description

For this project and its objective, I would require following data:

- Property Rental data for Bangalore City, India
- Geolocation data for different neighborhoods identified in the property rental data
- Major venues data in neighborhoods identified above.

Data used for this project will be from the following sources:

- Makaan.com - Use beautiful soup to scrap property listing data. Following variables would be needed for the problem at hand. Variables include
    - Property rentals
    - Area Sqft
    - Number of bedrooms,
    - Bathrooms,
    - House Facing
    - Furnishing
    - Address co-ordinates.
- Foursquare places API - Explore venues and venue categories in the neighborhoods identified using Makaan.com data

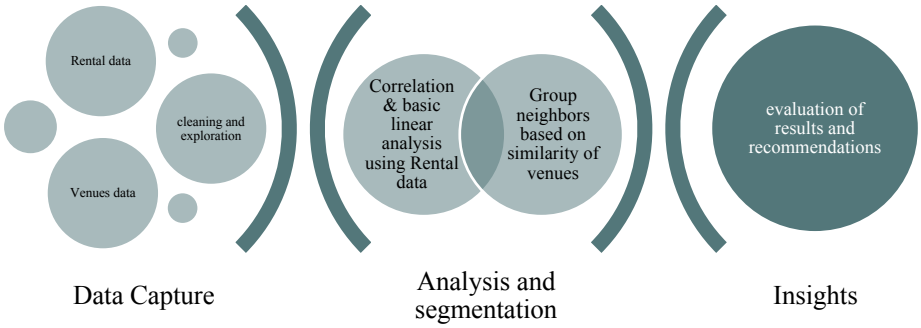
3. Data Collection & Cleaning Process

- Makaan.com web scraping
    - The data was scraped using Beautifulsoup. All the above variables were scraped from 500 pages with 21 rental properties per page. Post collection, dataframe was analyzed for any missing values, wrong addresses, incorrect entries. These were cross checked and cleaned using Python/Excel.
- Foursquare API
    - For each identified neighborhood, geopy module was used to convert an address into latitude and longitude values. Using these added data points, we used FourSquare API to get the trending venues in those locations.
- Data Cleaning
    - Two datasets were prepared using property rental and Venues data. Raw data has many issues which include missing values, wrong entries, exceptionally high rental prices, address corrections, several neighborhoods were grouped together when the addresses belonged to same locality. Post data cleaning, 669 neighborhoods were identified using Rentals data.
    - Venues data pull was clean and minimal data cleaning and alignment was required.

4. Methodology

The objective of analyzing the property rentals data is to get a fair idea about the neighborhoods and variables which may impact the property rental prices. As Bangalore has already seen a significant migration of employees from different parts of the country and it would be helpful if we can get a fair idea about the rental ranges, what kind of apartments are in good supply, which areas of the city are generating lots of property rental advertisements. Visualization, Cluster Analysis and basic regression would be used to derive insights and findings.

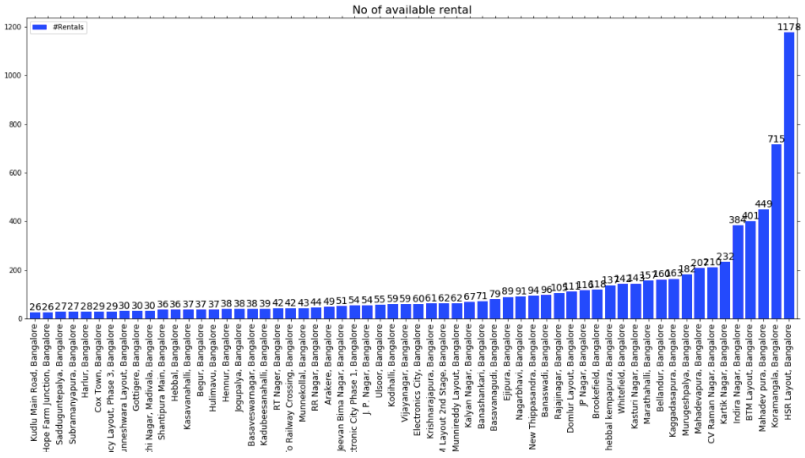
Following process was created to answer the questions identified in the objective statement.



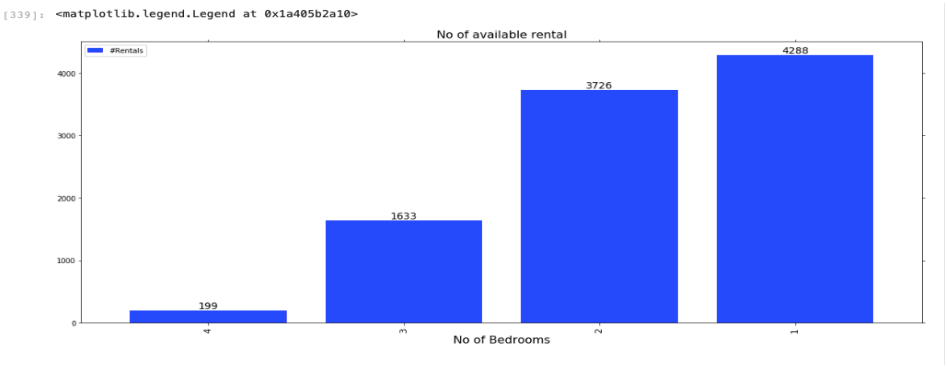
5. Exploration and findings

The purpose of doing exploratory study is to sense how the neighborhoods look like when compared on different metrics.

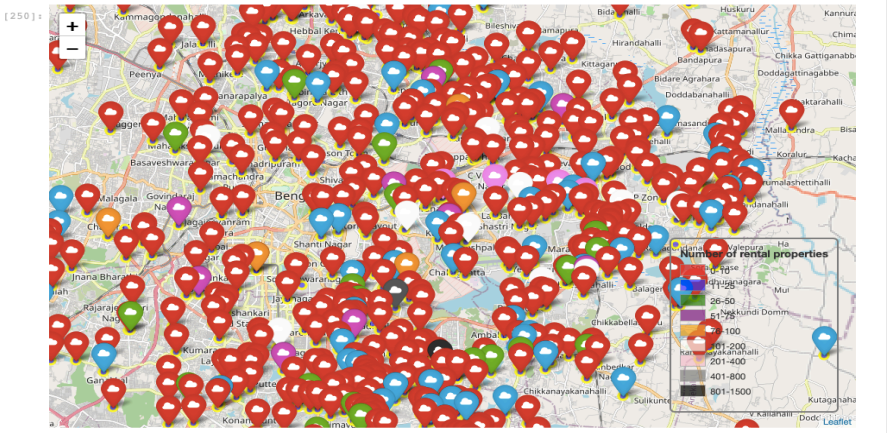
- **Number of Rentals by Neighborhood**: Rentals data clearly shows that Top 5 neighborhoods has the most supply of the rental properties. These neighborhoods are HSR Layout, Koramangala, Mahadevapura, Indiranagar, BTM layout & Kartik Nagar. At a high-level glance, all these neighborhoods are posh locations and have some of the most visited venues.



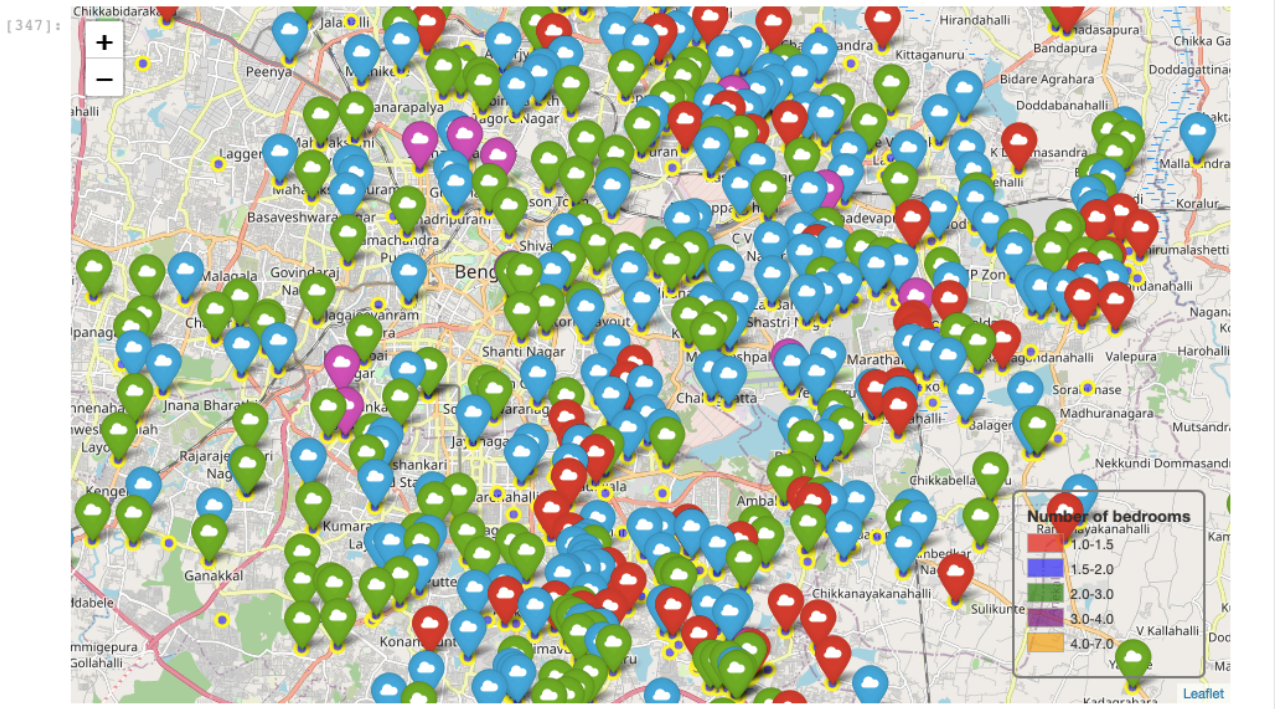
- **<u>Type of Bedrooms Apartment supply</u>**: Most of the demand and Supply is restricted toward 1 & 2BHK apartments and houses.



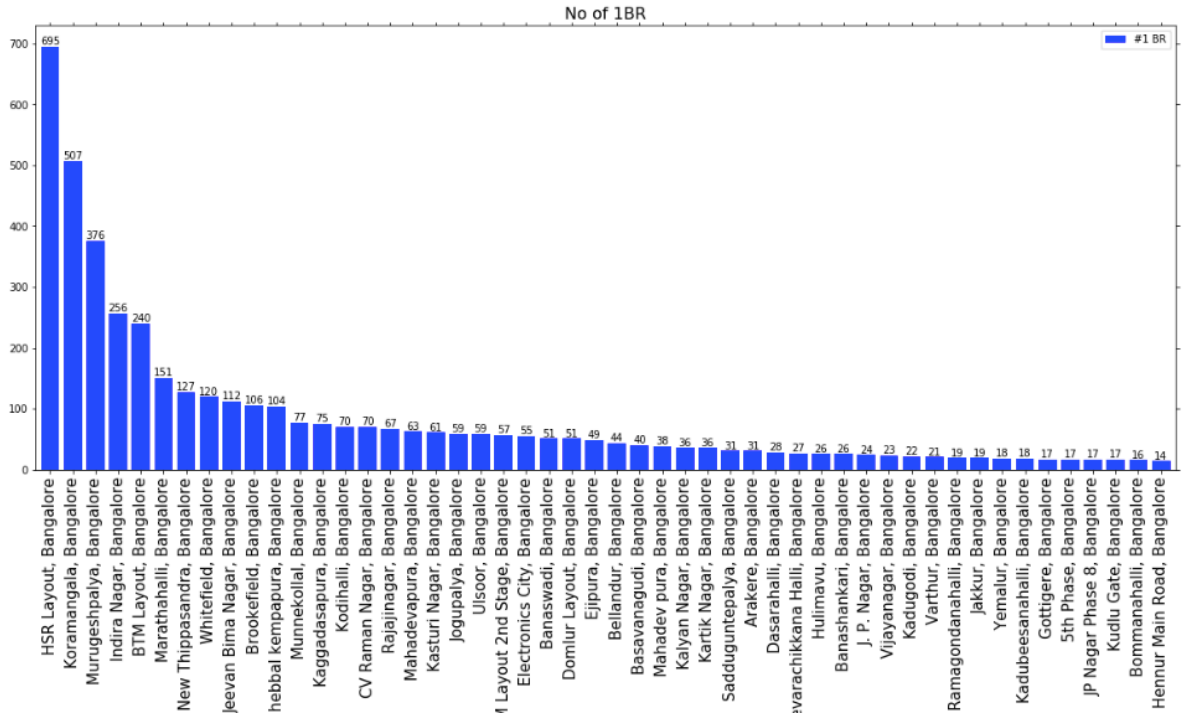
- **<u>Neighborhoods which are showing high supply of 1 & 2 BHK:</u>** As we can see Orange, white and black pointers are very less on this map. These are the neighbors which have a very high supply of apartments. These are the localities which belong to HSR, Koramangala, BTM, Indiranagar etc.



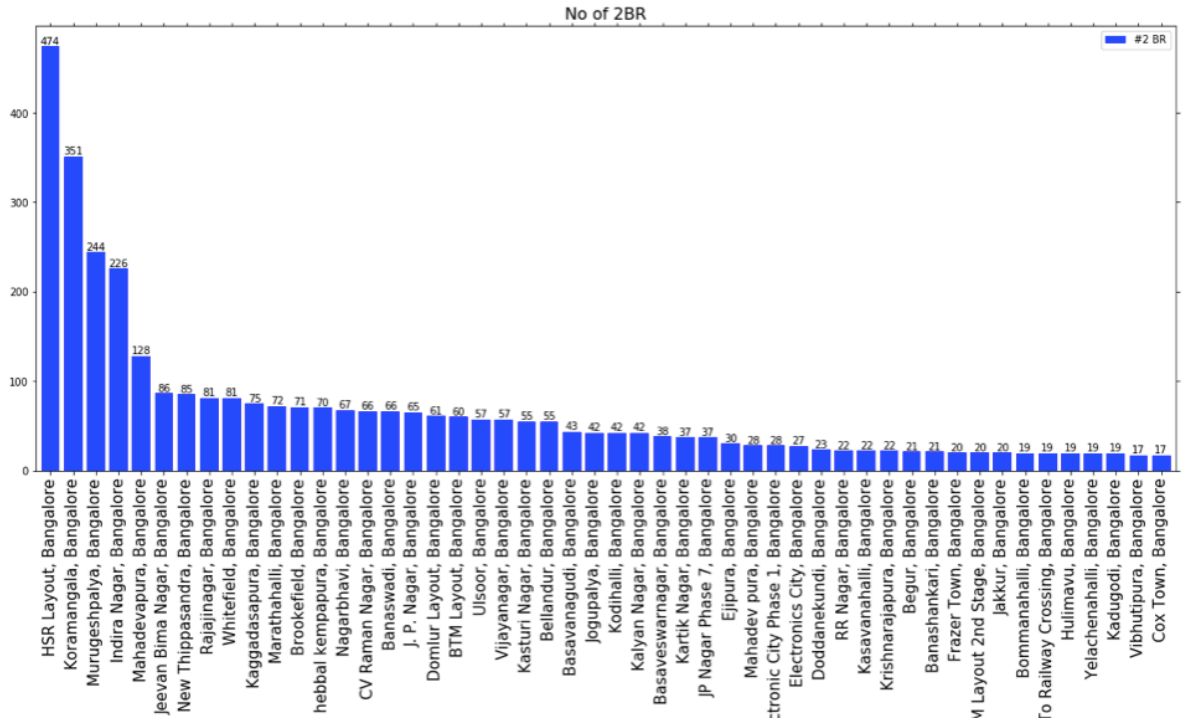
- **<u>Neighborhoods with most 1,2 & 3BHK supply</u>** is shown below.



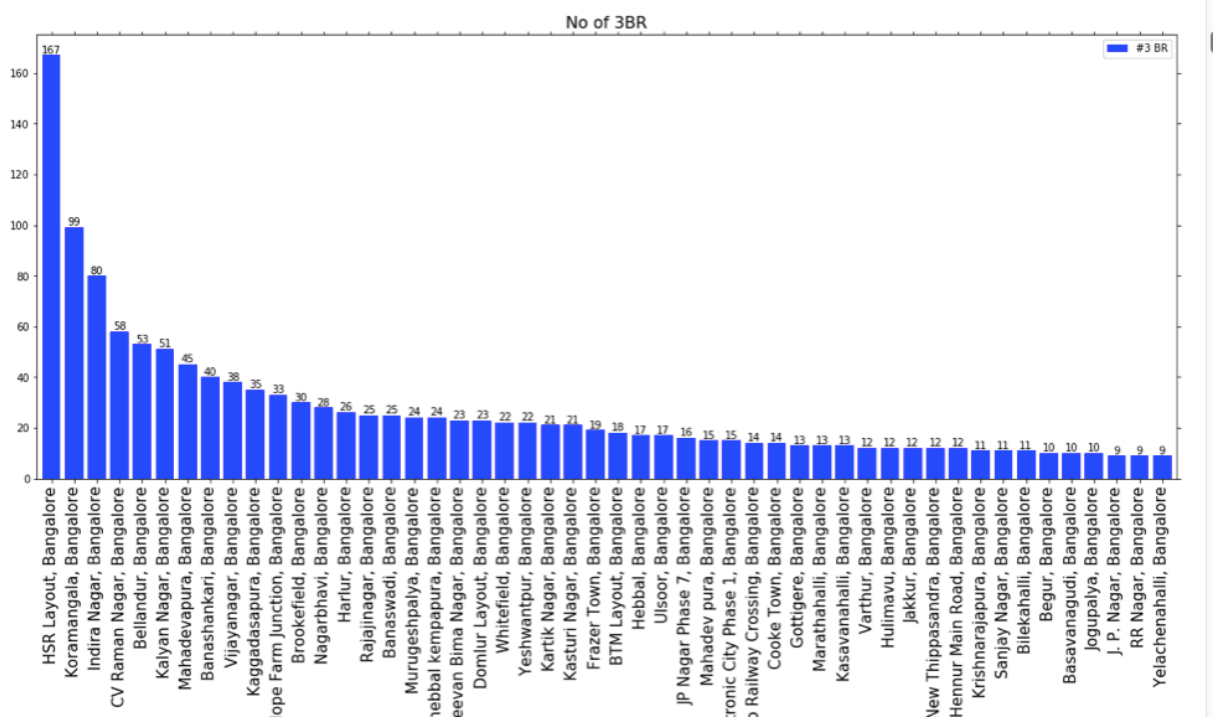
- **<u>1 BHK Neighborhoods</u>**



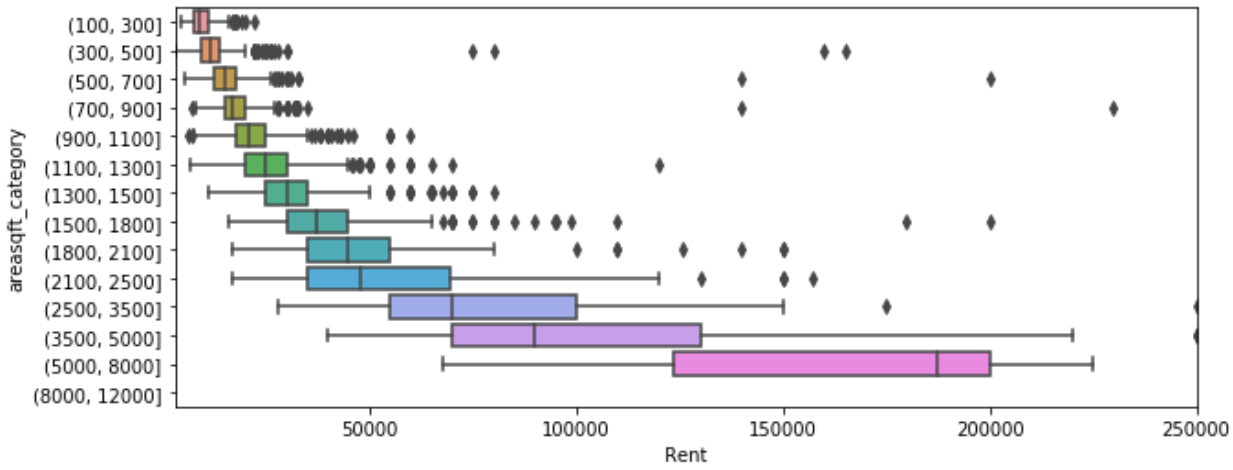
- **2 BHK Neighborhoods**



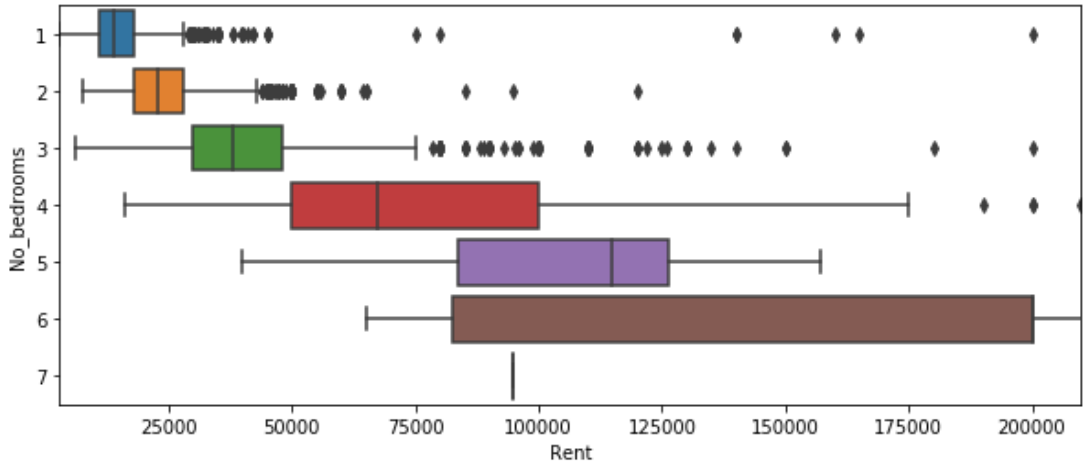
- **3BHK Neighborhoods**



- **Rent distribution by SQFT**



- **Rent distribution by bedrooms**



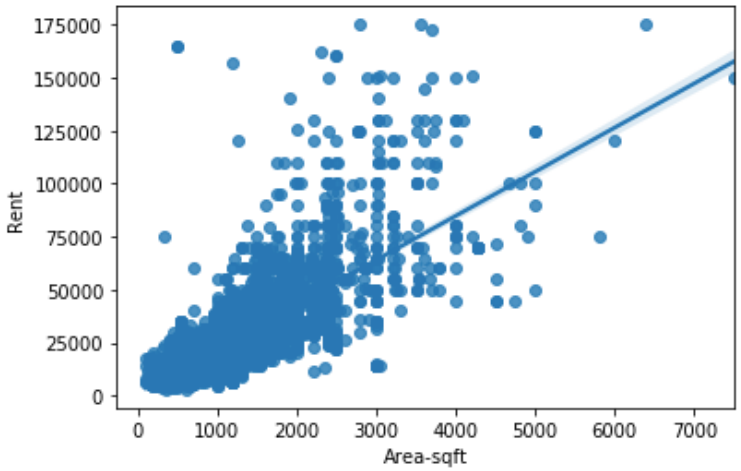
- **Correlation analysis:** I carried out correlation analysis to understand which variables have high, moderate and weak correlation with rental prices.

	Rent
Rent	1.000000
Area-sqft	0.776903
Bedroom	0.667838
Bathroom	0.628858
Address_lat	-0.021800
Address_lng	-0.070507
Furnished	0.173371
Semifurnished	-0.100291
Unfurnished	-0.038409
East facing	0.057608
West facing	0.001380
NorthEast facing	0.029592
South facing	-0.029480
North facing	-0.024244
NorthWest facing	-0.093997
SouthEast facing	0.020607
SouthWest facing	-0.004307

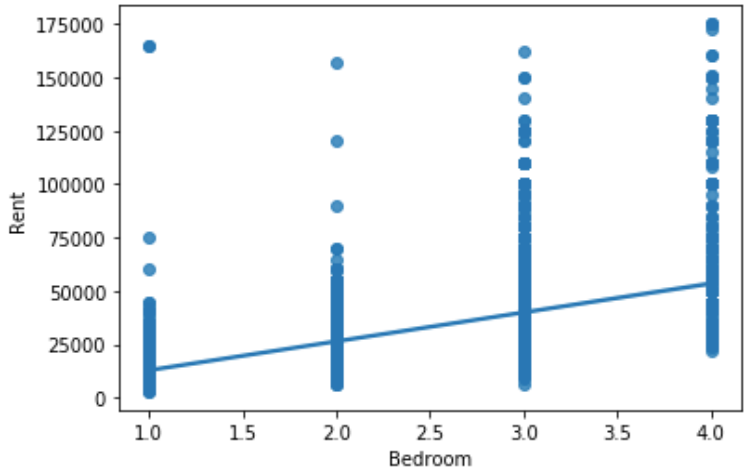
As we can see variables like area of apartment, no of bedrooms and bathroom have high correlations however it is interesting that many other variables are showing very low correlation with rental prices.

Following plots also show positive correlations with the rental prices.

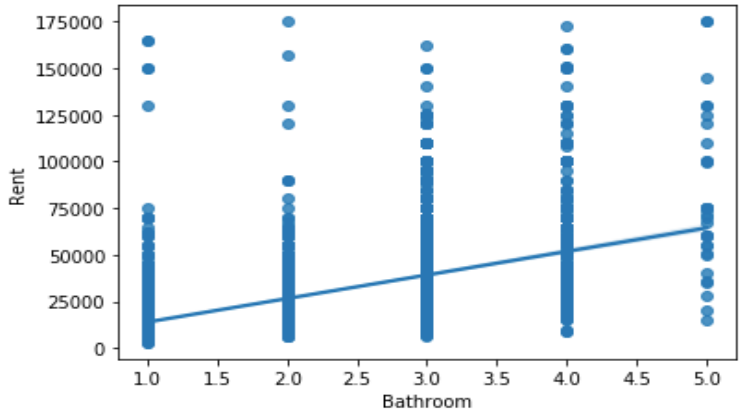
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a3e380d90>
```



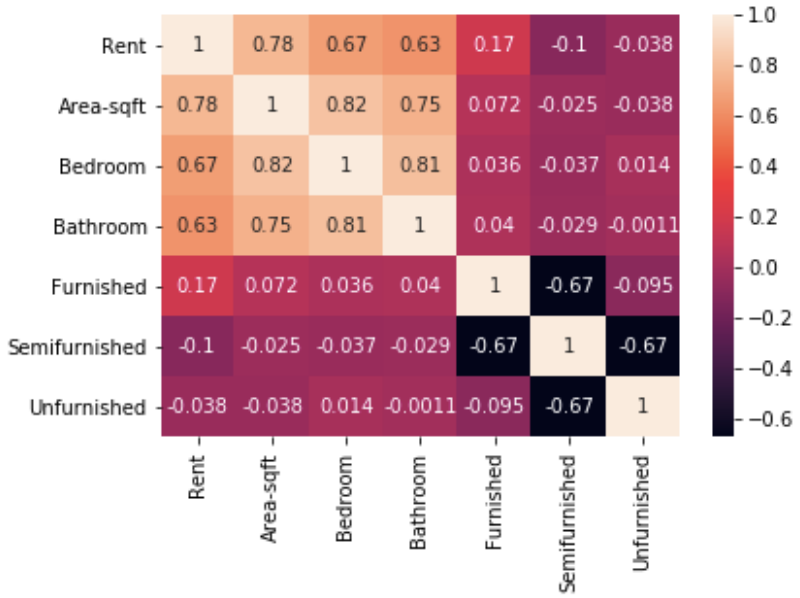

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a43da8fd0>
```



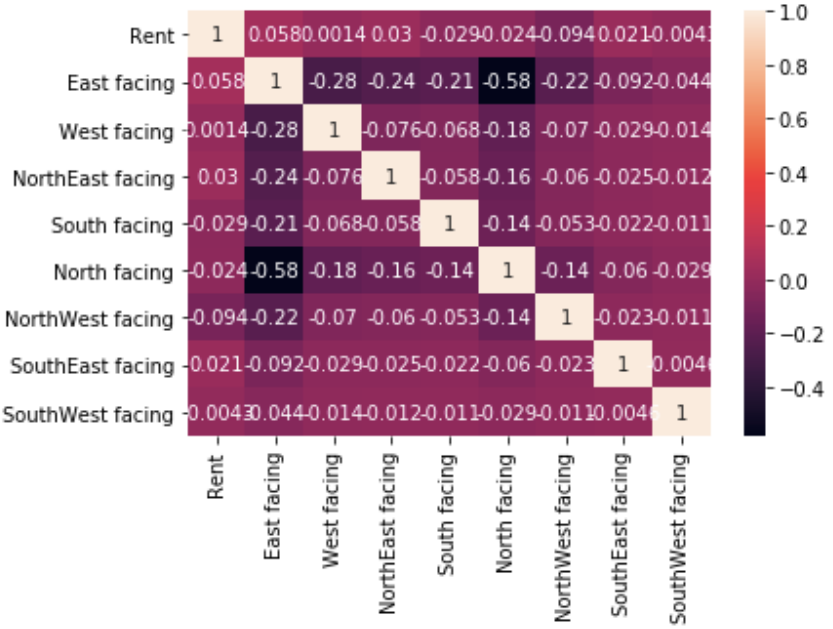
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a43e242d0>
```



```
<matplotlib.axes._subplots.AxesSubplot at 0x1a20d62b10>
```



```
<matplotlib.axes._subplots.AxesSubplot at 0x1a21bff410>
```



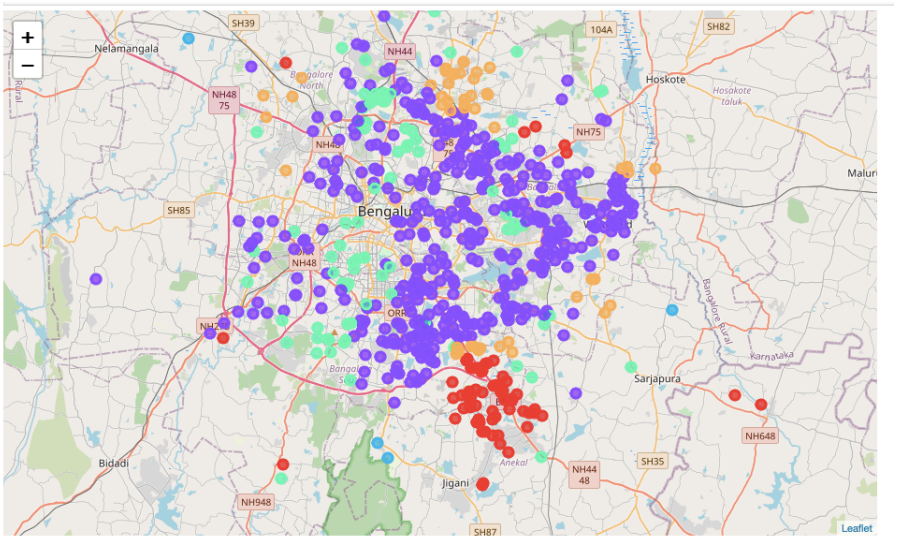
[72]:

Model:	OLS	Adj. R-squared:	0.622
Dependent Variable:	Rent	AIC:	167158.3470
Date:	2020-06-25 23:52	BIC:	167248.9775
No. Observations:	7876	Log-Likelihood:	-83566.
Df Model:	12	F-statistic:	1081.
Df Residuals:	7863	Prob (F-statistic):	0.00
R-squared:	0.623	Scale:	9.6421e+07

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
const	3065.0760	2242.9763	1.3665	0.1718	-1331.7536	7461.9055
Area-sqft	17.4452	0.3378	51.6452	0.0000	16.7830	18.1073
Bedroom	709.7013	281.6076	2.5202	0.0117	157.6756	1261.7269
Bathroom	2083.2811	239.8569	8.6855	0.0000	1613.0977	2553.4644
Semifurnished	-6296.8489	400.2513	-15.7322	0.0000	-7081.4478	-5512.2500
Unfurnished	-7117.4146	535.9808	-13.2792	0.0000	-8168.0795	-6066.7497
East facing	3216.0059	2204.5445	1.4588	0.1447	-1105.4871	7537.4989
West facing	2886.8404	2231.6997	1.2936	0.1959	-1487.8841	7261.5649
NorthEast facing	2762.6356	2246.1494	1.2299	0.2188	-1640.4141	7165.6854
South facing	2395.3933	2254.9287	1.0623	0.2881	-2024.8661	6815.6528
North facing	3076.6779	2208.2293	1.3933	0.1636	-1252.0383	7405.3942
NorthWest facing	-876.0605	2254.8033	-0.3885	0.6976	-5296.0741	3543.9530
SouthEast facing	2269.7041	2484.9435	0.9134	0.3611	-2601.4455	7140.8536

Exploration and Correlation analysis provides clear indications about the rental ranges in various neighbor groups.

Post this exploration of the rental data, I carried out Clustering analysis using K-means on the venue's dataset. Clustering helped identity similar neighbors and a closer look at the clusters reveal a lot about the rental distribution.



6. Results

With all the data and analysis, it is clear that rentals in different neighborhoods different significantly. There are several reasons behind this behavior. Some of them are been captured through this study. Based on the observations, following can be said about the different neighborhood's rental ranges.

There were 5 clusters with following features:

* Rental distribution charts for all clusters are captured in python notebook.

Cluster 1 - This cluster has the following properties:

- Avg. bedroom size: 1.8
- Avg. Rentals : INR 16974
- Avg. Area Sqft : 940 SQFT
- Avg rentals for 1BHK is between 7000 – 12000
- Avg rentals for 2BHK is between 13000 – 20000
- Avg rentals for 3BHK is between 18000-32000

Cluster 2 - This cluster has the following properties:

- Avg. bedroom size: 1.74
- Avg. Rentals : INR 23554
- Avg. Area Sqft : 1036 SQFT
- Avg rentals for 1BHK is between 18,000 – 25,000
- Avg rentals for 2BHK is between 18,000 – 32,000
- Avg rentals for 3BHK is between 22,000 - 50,000

Cluster 3 - This cluster has the following properties:

- Avg. bedroom size: 2.4
- Avg. Rentals : INR 20275
- Avg. Area Sqft : 1123 SQFT
- Avg rentals for 1BHK is between 5000 - 20000
- Avg rentals for 2BHK is between 12,000 – 30,000
- Avg rentals for 3BHK is between 20,000 – 75,000

Cluster 4 - This cluster has the following properties:

- Avg. bedroom size: 2.1

- Avg. Rentals : INR 29191
- Avg. Area Sqft : 1323 SQFT
- Avg rentals for 1BHK is between 7000 – 12000
- Avg rentals for 2BHK is between 13000 – 20000
- Avg rentals for 3BHK is between 18000-32000

Cluster 5 - This cluster has the following properties:

- Avg. bedroom size: 1.8
- Avg. Rentals : INR 20423
- Avg. Area Sqft : 1023 SQFT
- Avg rentals for 1BHK is between 8000 – 14000
- Avg rentals for 2BHK is between 15000 – 30000
- Avg rentals for 3BHK is between 20,000 – 55,000

7. Discussion

The statistical analysis and exploratory study itself was sufficient to make appropriate conclusion about the rental ranges however there is a need to predict rentals in various neighborhood using significant variables. Base model helped conclude significant variables. Also, through base model, we could explain around 60% of the deviation in rental ranges. This would mean that there are multiple other variables which are significant for example, in the current dataset variable like amenities, floor number etc were missing.

This study would definitely help owners and tenants to get a sense of different neighborhood and rental ranges.

8. Conclusion

This study is done at a start level however, It would be interesting to extend this study with more variables and predict accurate rentals for different neighborhoods.