A Property buyers guide in the coastal city of Udupi

Statistical research and insights on trends in property market

Aston Business School

Contents

1. Executive summary

This aim of this report is to successfully aid someone who is looking into understanding the housing market in the city of Udupi.

The various ways in which this report would be helpful is by providing key insights into the various dependent characteristics which influence the pricing pattern. Initially an insight into the relationship between the total number of bedrooms (BHK) and price is analysed. An important insight found here was with respect to the rapid increase in the pricing when total BHK's increased from 4 to 5. This can be very useful when someone is considering buying a residence with a big family, where they can save some money by compromising to 4BHK from 5BHK if possible, to do so. Another important trend which is highlighted in this report is regarding the availability of options for based on different locations. It was understood that Manipal being a hub for students has the highest concentration of available residences for sale which also come with the highest furnished & semi furnished options, this fact can be capitalized by buyer who is looking to buy residence as an investment opportunity since these residences which are already furnished can be rented out to students and can be a source of continuous income, further analysis was conducted to understand the nature of pricing and area, as expected they seem to be related to each other where increase in area led to an increase in price. However, another interesting insight found here was that of the erratic nature of area availability and pricing of 'Apartment' type of property. Anyone having unique requirement such as need for very high carpet area for tuition activities, musical instruments, antique display can look for apartments mainly rather than independent house as our data suggests availability of these kinds of properties.

Rich insights were also found where statistics was interpreted through groupings of different combination of BHK. Interestingly we found that 3 and 4 BHK are having almost the same average price. This is especially useful when someone is considering buying a 3BHK and can spend only slightly more to obtain a 4BHK house.

Furthermore, some hunches were cleared and confirmed by performing statistical hypothesis testing. First to ensure the sanctity of the data, i.e., to prove the dataset considered for this report in fact represents the whole of Udupi. Secondly as claimed above where we found mean difference very less between 3BHK and 4BHK. Both statements were proved to be statistically right and hence the report can strongly advice with proof that the insights provided by this report are in fact true.

Finally the report ends with a calculator tool for keen buyers who already know their requirement with regard to the expected carpet area and the number of BHK, find the expected price they would end up paying for their specific requirement.

1. Introduction

This report targets to understand the housing market trends in the coastal district of Udupi, located in Karnataka State in India. Multiple statistical analysis is conducted in this report to give more insights and patterns on the data. The data was collected from an online website for housing market (99acres.com) and around 5 key characteristics was collected and considered for this analysis which are

1.1. Characteristic selection

Variable	Description	UOM	Reason for selection
Price	Price of establishment	INR	Main constraint for considering a house by people, dependent variable
Area	Total carpet area	Sq. ft	Price and Area are generally found to be increasing proportional to each other
BHK	Total bedroom hall and kitchen	Each	This is one of the important search criteria used by people
Location	Subdivisions inside main city Udupi	None	Convenience is depicted by location usually with regards to amenities
Property type	The type of property being sold	None	This is a small influencing factor which can help in deciding a housing

1.2. Sampling method

The sample dataset has been derived from the population using random sampling. This method was employed by thoroughly shuffling the data and selecting 100 samples out of 196. This method has been employed since the chances of sampling bias is very less, and the simplicity of the sampling. However due to the low population that was collected, the chances of bias cannot be eliminated.

1.3. Outliers

The outliers are handled by calculating the z-score statistics, the rows that were having z score greater than 3 were dropped from the data set and ultimately left with 96 data points

2. Visualization

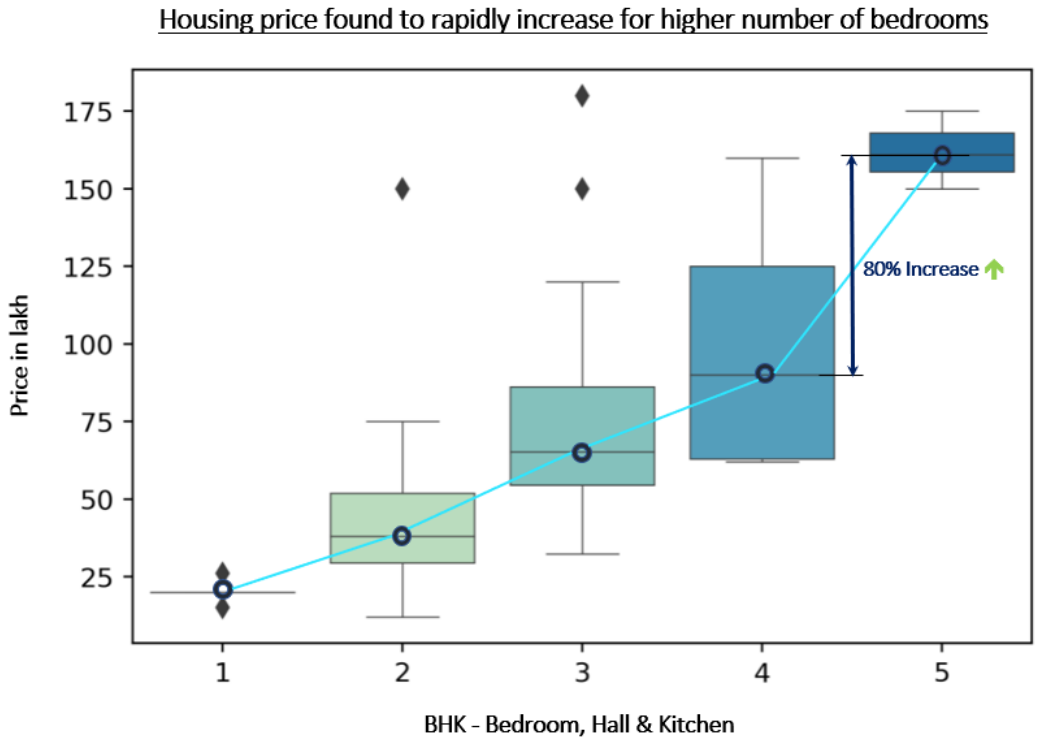
2.1. Impact of BHK selection on price

Introduction: The below graph shows a box and whisker plot to represent the categorical variable BHK's against quantitative variable Price. The box and whisker plot is a highly information dense plot which provides an overview on 5 key summary statistics that is minimum (Whisker lower edge), maximum (Whisker upper edge), Quartile 1(Box lower edge), Quartile 3(Box upper edge).

Insight: Some of the main observations that can be noticed from this graph are

- A linear increase in the price of the residence proportional to the price in lakhs
- A rapid increase of price from 4BHK to 5BHK
- High interquartile range for 4BHK i.e., high fluctuation of prices for 4BHK

- Low interquartile range for 1BHK i.e., relatively same price range for 1BHK



Design rules compliance: The area represented under the box for each BHK is **proportional** to the increase in price. The colouring scheme used has an **increasing darkness and saturation** for higher BHK's. Key **highlight** has been shown in the data regarding rapid increase of price range.

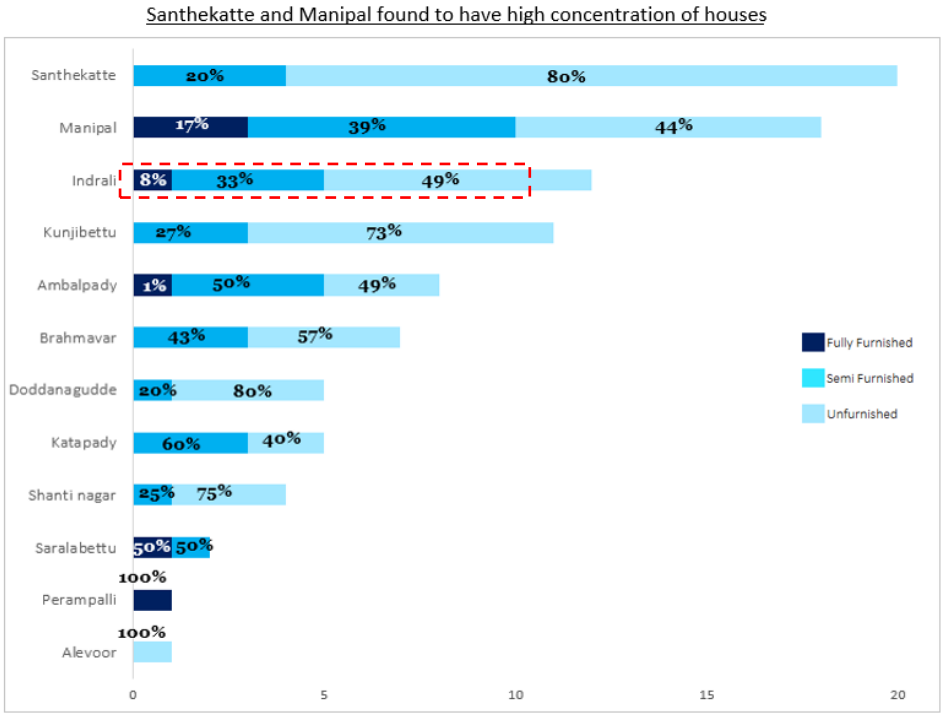
Trend of the data has been clearly shown by connecting the medians of each of the boxplot.

2.2. Study of trends in location

Introduction: The graph depicted below is a bar chart which shows the concentration of available residences for sale. It is a multidimensional graph with the 2nd dimension portrayed as stacks in the bars. The size of the bar is directly proportional to the frequency of occurrence of a residence in that location.

Insight: Some notable observations from the graph are as below

- Santhekatte found to have highest concentration of properties for sale followed by Manipal, this is due to the high population concentration in these areas
- Alevoor found to have lowest concentration of properties for sale followed by Perampalli
- Fully furnished houses are less frequently found in each location, this is since most of these residences available are new apartment offerings.
- Manipal has high concentration of fully furnished and semi furnished offerings due to the reason that it is a student hub with 2 universities, usually these properties are bought and put out for rent.



Design rules compliance: Decreasing order of concentration is shown for ease of understanding for the reader. Colour saturation has been to show increasing level of furnishing. Since stacked columns have the effect of being hard to interpret during side-by-side comparison, the composition percentages are shown for each stack for better understanding and interpretation by the reader.

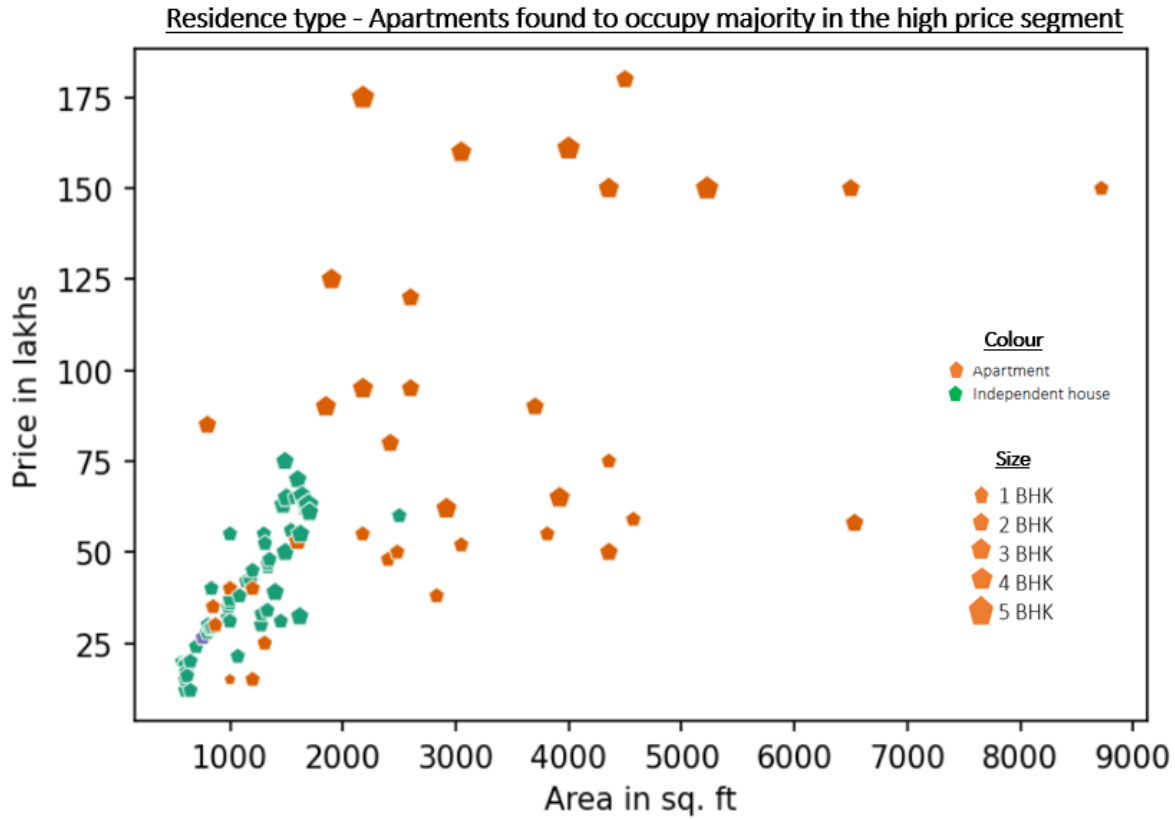
2.3. Price change pattern with respect to area

Introduction: The graph depicted below is a scatterplot, this type of graph is usually used to understand the relation of a dependent variable (Price) with an independent variable(Area), Scatterplot can be utilized to show a wide range of dimensions by manipulating the size, colour and shape, thus a rich information dense visual can be obtained by employing scatterplot.

Insight:

- There seems to be a positive relationship between area and price as we can see the uphill pattern between price value (y variable) which tends to increase with increase in the area (x variable) which is normally expected as properties are generally rated with respect to price per square foot in most of the cases
- Apartments are found to be in the higher segment of price and area, this shows an trend recently seen where apartment construction is seen to be skyrocketing in the suburbs of Udupi.
- It can be seen higher number of bedrooms is mainly found in the outer region in the scatter plot, showing extreme ends
- Majority of the independent house seems to be lying in the price range of 25 to 75 lakh and area under 2000 square feet. This confirms the decreasing popularity of

owning independent houses and population moving to buy apartments and leaving them on rent and also the high maintenance dependency and cost related to owning independent house.



Design rule compliance: Clear **labelling** of the dimensions are shown in the graph, visual **contrast** provided with regards to the **colour** and **size** of the scatter bubbles which can help the reader interpret quickly to develop ideas and insights on the data.

3. Descriptive statistics

3.1. Quantitative variables

	Price	Area	
Category	Quantitative	Quantitative	Central tendency
Sub category	Continuous	Continuous	
Mean	5648627.66	1892.10	
Median	4800000	1425	
Mode	5500000	1000	
Range	16800000	8160	Dispersion
Inter Quartile Rang	3200000	1344.5	
Variance	1.5087E+13	2165570.819	
Standard deviation	3884161.73	1471.59	
Kurtosis	2.24	5.56	

3.2. Categorical variables

	BHK	Location	Property type
Category	Categorical	Categorical	Categorical
Sub category	Ordinal	Nominal	Nominal
Mode	2	Santhekatte	Apartment

* The average price which a customer can expect to spend when looking into the housing data is 56,48,627.66 INR.

* The most frequently occurring price range for a property is 5500000 INR which is not much far from the mean.

* Average carpet area per house which is currently present in the market is 1892.1 square feet

* The most frequently occurring category in BHK is 2, which is naturally the most preferred by customers for a family of 3 - 4

3.3. Summary statistics for - Categorical split – based on BHK

			Central tendency			Dispersion			
Grouping criteria	**Variables**	**Description**	**Mean**	**Median**	**Mode**	**Range**	**Inter Quartile Range**	**Variance**	**Standard deviation**
BHK = 1	Price	Price of establishment	2020000 (1)	2000000	2000000	1100000	3200000 (2)	1.216E+11	348711.9155
	Area	Total carpet area	702.00	600	600	440	150	26416	162.53
BHK = 2	Price	Price of establishment	4057491.228	3800000	5500000	13800000	2250000	4.3561E+12	2087115.739
	Area	Total carpet area	1547.947368	1547.947368	1000	8115	8115	8115	1299.997023
BHK = 3	Price	Price of establishment	7684700	6530000	6530000	14766000	14766000	1.2725E+13	3567142.233
	Area	Total carpet area	2557.4	1632.5	1600	5734	1378	2693462.84	1641.177273
BHK = 4	Price	Price of establishment	9700000	9000000	6300000	9800000	6200000	1.3529E+13	1.35289E+13 (3)
	Area	Total carpet area	2617.888889	2178	1695	2661	1199	888268.321	942.4798783
BHK = 5	Price	Price of establishment	16200000	16100000	15000000	2500000	1250000	1.0467E+12	1023067.284
	Area	Total carpet area	3801.666667	4000	2178	3049	1524.5	1524.5	1252.624534

1. It can be clearly observed and interpreted that with increase in BHK the mean price of the property is also increasing. The reason this is happening can be correlated to the fact that with increase in number of bedrooms will also increase the total area of the property which would obviously increase the pricing of that property. An interesting fact to notice here is the average price almost doubles from one BHK to the next, except for 4BHK whose average price increases only by approximately 20 lakhs.
2. As the measure of dispersion, we see the variance statistics of 1 BHK price. This is the lowest variance out of all the BHK categories, indicative of the data points concentrated around the mean statistically speaking, however this means that someone looking to buy a 1BHK would find relatively stable pricing range in most of the location. This is mainly because of the reason that there are relatively low number of offerings of 1BHK in Udupi.
3. Standard deviation is also the measure of dispersion, where highlighted in the table, the highest standard deviation out of all the BHK categories is considered. This measure being high is indicative of the fact that there is a huge disparity in the available options for 3BHK properties. Someone who is considering buying a property should carefully consider all options as there might be cheaper ones available.

4. Statistical inference

4.1. Confidence interval (95%)

Price of property: For someone looking to buy a house in Udupi it can be said they would most probably (95% chances) end up finding a house in the range of 48,53,074.36 INR to 64,44,180.959 INR, this is derived from the.

Carpet area: It can be said with some degree of certainty that most of the residences (95%) are going to have their carpet areas in the range of 1590.69 Sq. ft and 2193.51 Sq. ft.

Some assumptions that come into our mind based on the data we have seen until now and relating with the population

Hypothesis 1: The comparison of mean price of property of the representative sample data to the population mean property price, are they equal?

Hypothesis 2: There is no significance difference between mean price of 3BHK and 4BHK houses. Based on the descriptive statistic observed for categorical split in section 3.

To confirm if these above hypotheses are true, we employ the usage of appropriate statistical tests

4.2. One sample t-test for Hypothesis 1

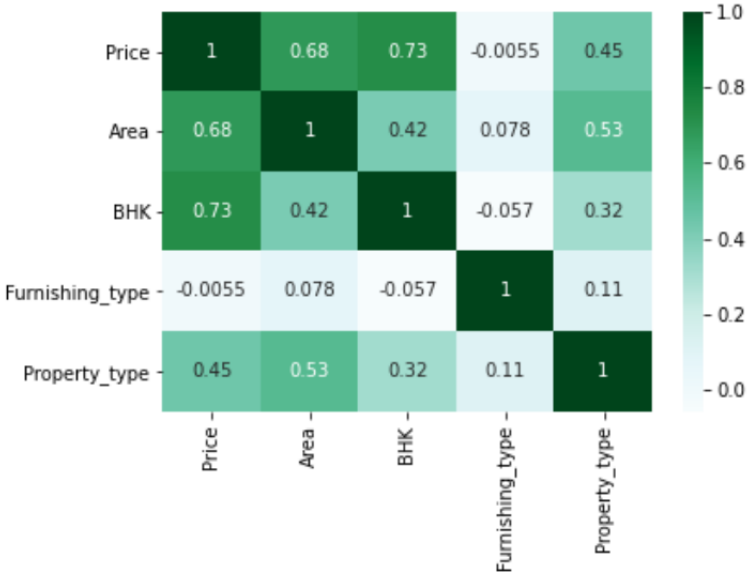
This test is used to determine if mean of a population is statistically different from a known value, perfectly fitting requirement of hypothesis 1. The population mean of 52,48,263(average cost per sq. mt * mean carpet area from dataset) is obtained from the site (https://www.numbeo.com/cost-of-living/in/Udupi-India) The test was run in python and the obtained p-value was – 32% which is much higher than the significance threshold of 5% confirming the hypothesis that the population mean is not different from sample dataset mean.

4.3. Two sample t- test for hypothesis 2

This test is used to determine whether unknown population means of two groups are equal or not. Which is appropriate for the testing and confirming our hypothesis 2. The test was run in python and the obtained p-value was 18.98% which is higher than the significance threshold of 5%. This confirms our hypothesis that both datasets have equal means. This is critical information for a buyer since they can almost always opt for 4BHK instead of 3BHK as there is no significant difference in their average prices.

5. Correlation and regression analysis

5.1. Correlation



Correlation coefficient as the name suggests quantifies the relation between 2 variables by assigning a value, values are assigned from -1 to +1 with -1 being significantly negatively correlated and +1 being significantly positively correlated and 0 - no correlation. The table considers the variables – Price, Area, BHK, furnishing type and property type.

Observations

- Price & BHK are seen to be having the highest correlation out of all the combinations meaning a change in the value of BHK would also change the price of the property which makes sense
- It is evident from the heatmap that there seems to be a significant correlation between Price & Area as expected and observed in our analysis above.
- Comparing our dependent variable Price with the variable furnishing type, there seems to be no real correlation between these 2 since the value of coefficient is -0.0055

5.2. Regression analysis

Initial model with all variables considered

Dependent variable: Price

Independent variables: Area, BHK, Property type, Furnishing type

Dep. Variable:		Price		**R-squared:**	0.706	
Model:		OLS		**Adj. R-squared:**	0.693	
	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-2.64e+06	1.2e+06	-2.198	0.031	-5.03e+06	-2.54e+05
Area	1154.7167	188.193	6.136	0.000	780.781	1528.653
BHK	2.371e+06	2.89e+05	8.210	0.000	1.8e+06	2.94e+06
Furnishing_type	-9.484e+04	3.56e+05	-0.266	0.791	-8.02e+05	6.13e+05
Property_type	3.771e+05	4.92e+05	0.766	0.446	-6.01e+05	1.36e+06

* We can observe p value for Furnishing type and property type being more than significance threshold of 0.05 making it insignificant, hence we need to eliminate these variables from the model

* The adjusted R^2 value of this model is 0.693 which is of acceptable range

Final model after eliminating the p values greater than 0.05

Dependent variable: Price

Independent variables: Area, BHK

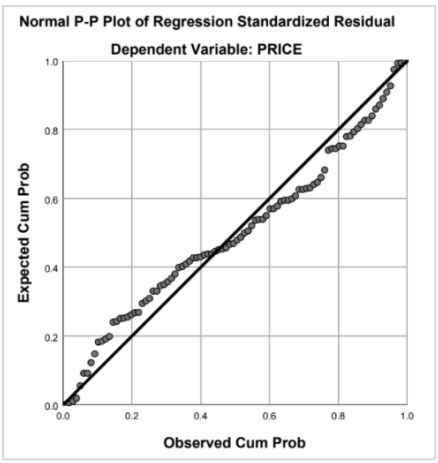
Dep. Variable:		Price		**R-squared:**	0.704	
Model:		OLS		**Adj. R-squared:**	0.698	
	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-2.538e+06	6.66e+05	-3.812	0.000	-3.86e+06	-1.22e+06
Area	1215.9710	165.745	7.336	0.000	886.738	1545.204
BHK	2.406e+06	2.83e+05	8.510	0.000	1.84e+06	2.97e+06

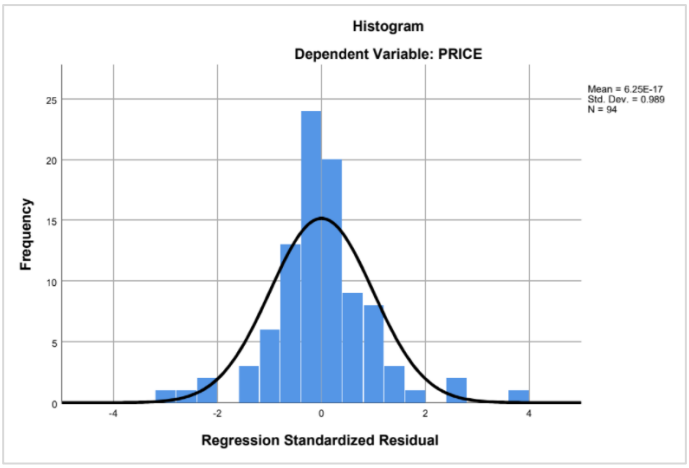
* This can be considered as the final model for our regression

* The adjusted R^2 value of this model is 0.698 which is an improvement over the initial model.

* This can be considered as a parsimonious model

Adequacy analysis





Linearity: The standardized residuals follow almost a linear pattern. Hence this assumption shall stay true

Homoscedasticity: The error terms are independent, there is no grouping or pattern found.

Independence of errors: Distance of residuals in the graph in relation to zero remains more or less the same, this assumption holds

Normality: The histogram shows clearly normal distribution of the standardized residuals.

Multicollinearity: Area & BHK have a correlation coefficient of 0.42 based on the previously discussed correlation heatmap, hence there is no issue of multicollinearity found.

Using the model

The coefficient values from the model is extracted to obtain the property prediction using below formula

Predicted_value = $-2.538 \times 10^6 + 1215.9710$ **Area** $+ 2.406 \times 10^6$ **BHK**

Example from the considered dataset:

Ex1: Area – 760, BHK – 2, Actual value – **2650000**, Predicted value - **3198137.96**

Difference **- 548137.96**

Ex2: Area – 1487, BHK – 3, Actual value - **7500000**, Predicted value - **6488148.877**

Difference - **1011851.12**

The model comes close to the actual values with deviation of 5Lac to 10Lac as observed above, which can be acceptable since it is under 10-20% deviation from the actual.

The unit square feet impact of increasing the area would be an increase in the price by **1215.97 INR**. The unit impact by increasing BHK by 1 would be **24,06,000 INR**.