STAT40830 ADV Data Programming with R

Pushkar M Telavane & 24223351

Introduction :

The report explores a housing price dataset containing 545 observations with 6 numeric and 7 categorical variables. The dataset belongs to a city in India with housing rates influenced by the area and other factors. With determining the relation of price with respect to the other factors, will help the builders to understand the on-going trend and buyers perspective while purchasing a property.

Loading the house price dataset

The dataset includes house price, area, number of bedrooms and bathrooms with different furnishing features.

Data Preparation:

Checking for any missing values as well converting the columns to factors for better exploration.

Exploratory Data Analysis:

The plot shows the variance of price raises w.r.t to area. The red line indicating a linear regression fit.

Area vs Price
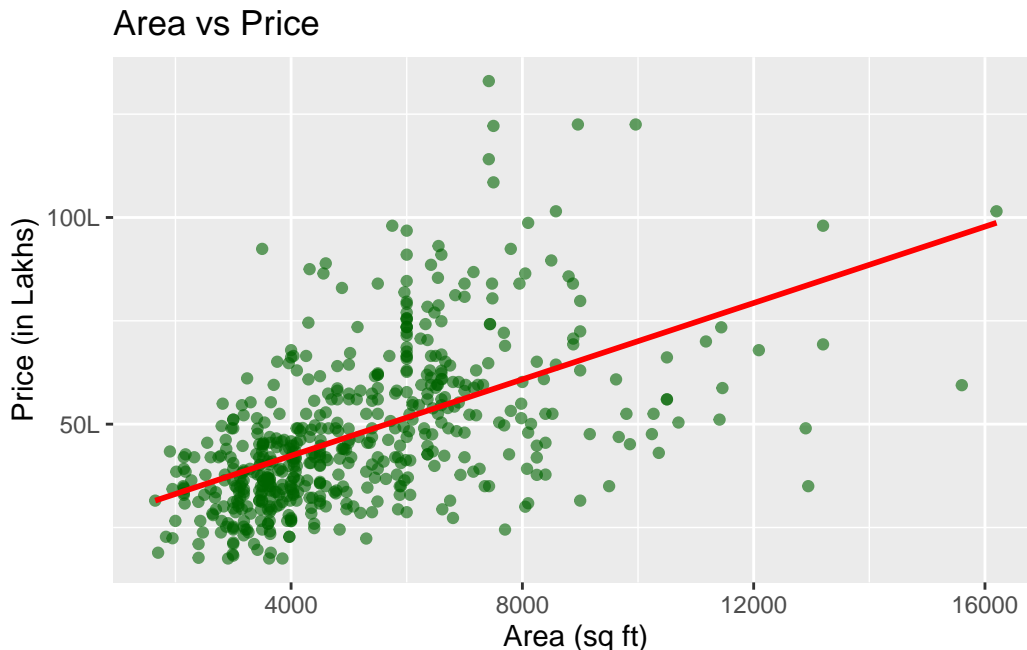
Results : It is quite evident tat area plays a major factor in variation of the housing prices with red line representing a positive correlation. A major chuck of price w.r.t to are falls under the range of 3000 to 8000 sq ft per flat as many flats in India are majorly two or three bhks. Few outliers are observed beyond the range of 12000 sq ft maybe it would be premium penthouse or bungalows. Also, most of the house properties ranging from 20 to 80 lakhs as occupied by the low or middle class persons.

Data Visualization :

Price by Air Conditioning

Price by Furnishing Status
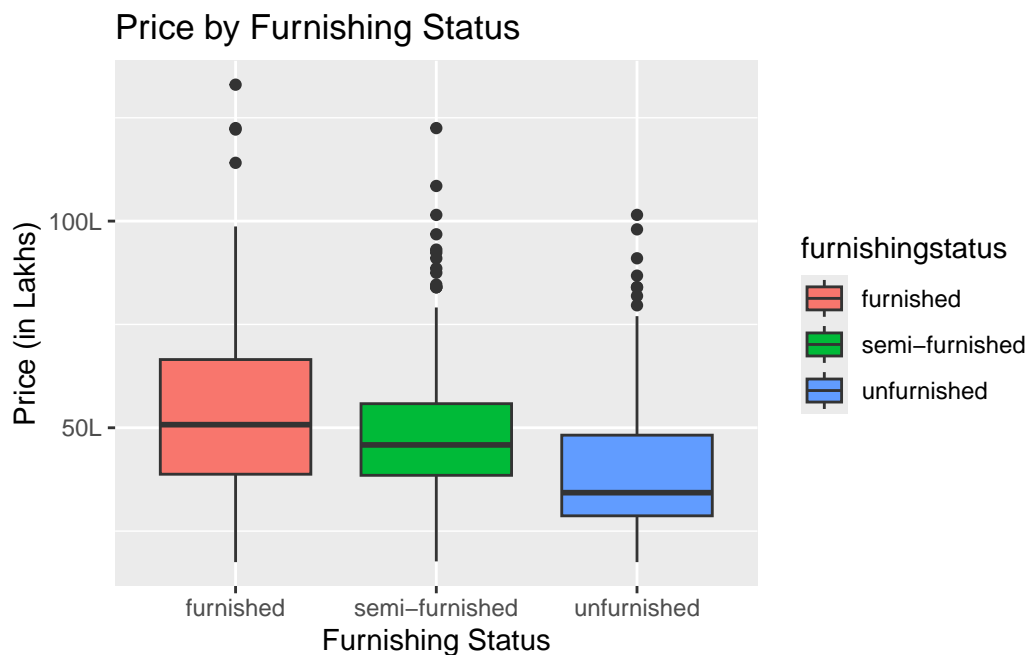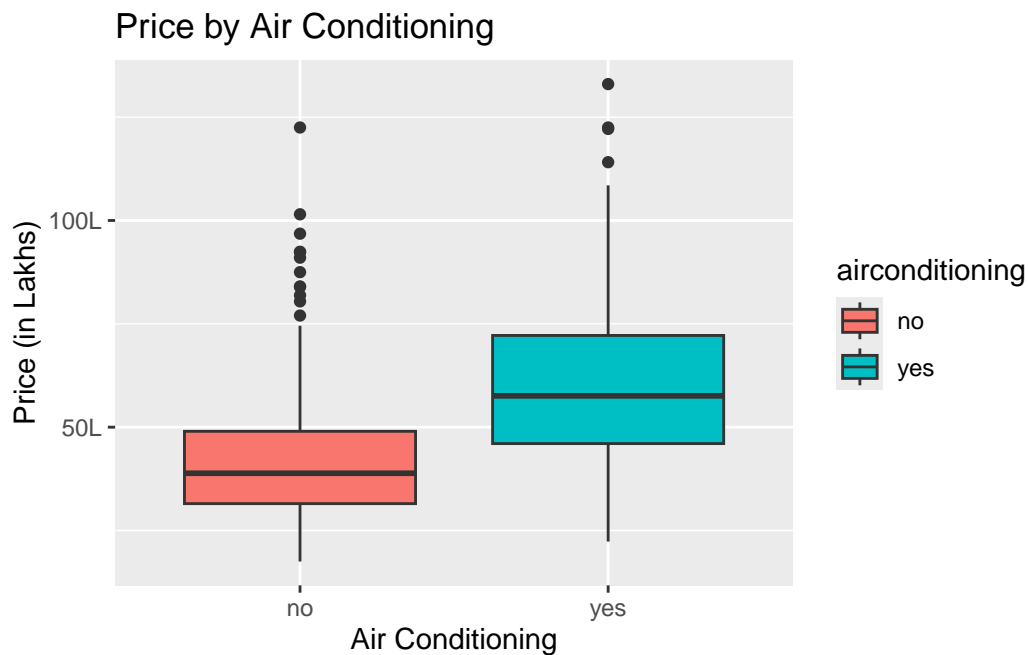
Plot 1 : The homes with air Conditioner have the higher prices as the setup and installation and maintenance is quite expensive. Both the options have outliers but the no option has more comparatively. In recent times, people prefer the homes with air conditioning as the box variation is more with longer whiskers. Ranging from 30 to 75 lakhs aprrox the range for the price of home.

Plot 2 : Furnishing the house or apartment increases the value totally but the budget constraint buyers may choose the unfurnished or semi-furnished as it lowers the cost significantly. We can observe a significant outliers throughout the options of furnishing status. Furnished homes has the widest IQR as nowadays the buyers prefers the most of the worked estate.

Correlation Analysis:
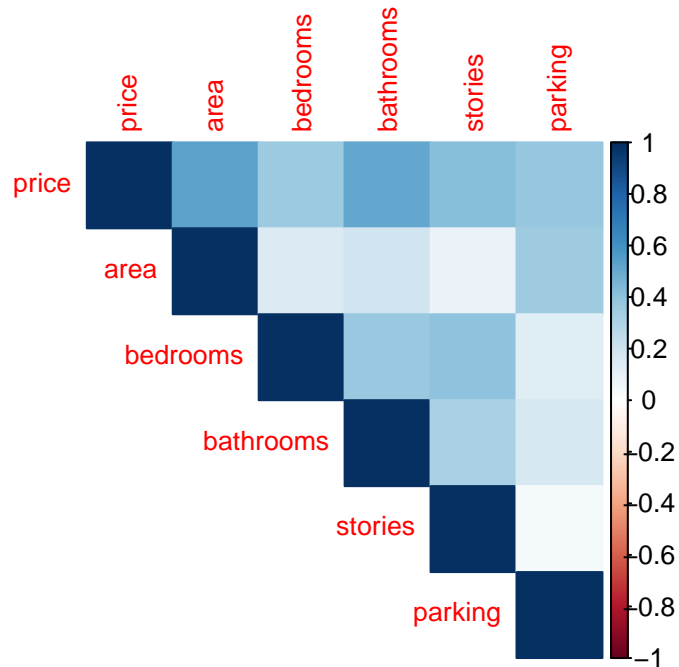
Results :

- The price is mostly influenced by the factors of **area, bathrooms, stories** and **parking** respectively.
- While, the bedrooms have a **weaker influence** as compared to other variables.
- We can observe , their are **no strong negative correlation** , as all of the factors contribute to a certain extend to the property price.