NAME : Ameya
Dabholkar
UID : 2021300023
CLASS : BE COMPS A
BATCH : ADV BATCH F

ADV EXPERIMENT 5

DATASET :

dataset - https://www.kaggle.com/datasets/yasserh/housing-prices-dataset

DATASET DESCRIPTION :

The dataset contains 545 entries and 13 columns, with both numerical and categorical data. Here's a summary of the key columns:

- **Numerical columns**: price, area, bedrooms, bathrooms, stories, parking
- **Categorical columns**: mainroad, guestroom, basement, hotwaterheating, airconditioning, prefarea, furnishingstatus

Here are 5 possible data visualizations that you can implement using R in PowerBI:

1. **Price vs. Area Scatter Plot**:
   - This will help show the relationship between house prices and the area of the houses.
2. **Bedrooms Distribution**:
   - A bar chart showing the count of houses by the number of bedrooms.
3. **Price by Furnishing Status**:
   - A box plot that compares house prices across different furnishing statuses (furnished, semi-furnished, unfurnished).
4. **Parking Spaces by Stories**:
   - A clustered bar chart that shows the distribution of parking spaces across different house stories.
5. **Air Conditioning and Price**:
   - A bar chart comparing the average price of houses with and without air conditioning.

REPORT :

What does the regression plot indicate about the relationship between the area of a property and its price? How do the linear and non-linear regression lines differ in representing this relationship?

Answer:

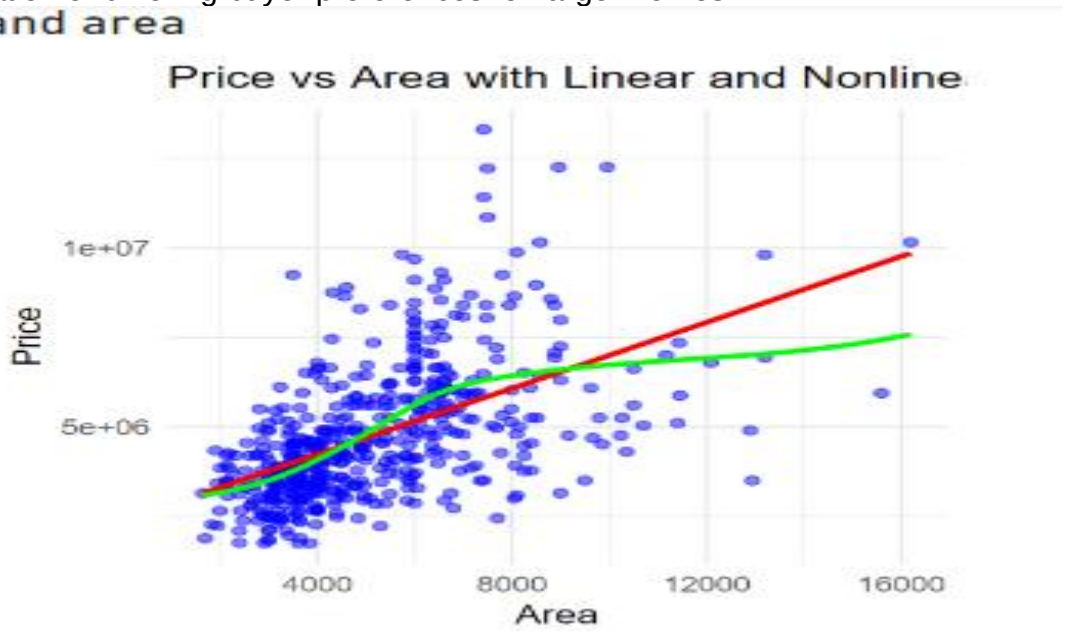
The regression plot clearly shows a positive relationship between area and price, indicating that as the area of a property increases, its price tends to increase as well.

Linear Regression (Red Line): The linear regression line provides a simple overview of the relationship, suggesting a constant rate of increase in price with respect to area. This implies that for

every unit increase in area, the price increases by a consistent amount. However, this model may not capture the complexities of the data, especially in regions with larger areas.

Non-linear Regression (Green Line): The non-linear regression line (possibly a LOESS curve) better captures the variability in the data, particularly at higher area values. It suggests that the rate of price increase may slow down as area increases, which indicates that larger properties might not have a proportional increase in price compared to smaller ones. This non-linear relationship could reflect market saturation or differing buyer preferences for larger homes.

price and area



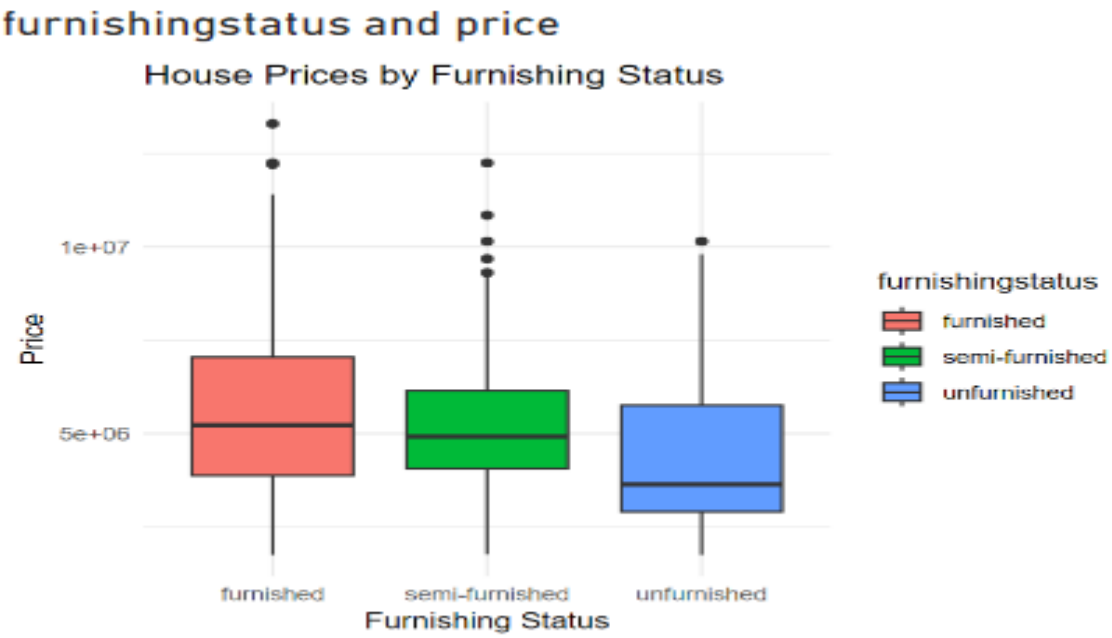
What does the box plot reveal about house prices in relation to their furnishing status? How do the interquartile ranges and median prices compare among the furnished, semi-furnished, and unfurnished categories?

Answer:

The box plot provides a clear visual representation of how house prices vary across different furnishing statuses:

- **Furnished:** The box plot indicates that furnished houses generally have the highest median prices, along with a wide interquartile range. This suggests that furnished homes command higher prices, likely due to the added value of included furnishings and decor.
- **Semi-Furnished:** Semi-furnished houses have a median price that is lower than furnished houses but higher than unfurnished ones. The interquartile range appears narrower compared to furnished houses, suggesting that semi-furnished homes have a more consistent pricing structure, with fewer extreme values.
- **Unfurnished:** The box plot shows that unfurnished houses have the lowest median prices. The range also appears more condensed, indicating that there may be less variation in pricing for unfurnished properties, possibly due to their lower initial investment appeal.

In all categories, there are visible outliers, particularly in the furnished category, indicating some furnished properties have significantly higher prices, which could be attributed to unique features or prime locations.



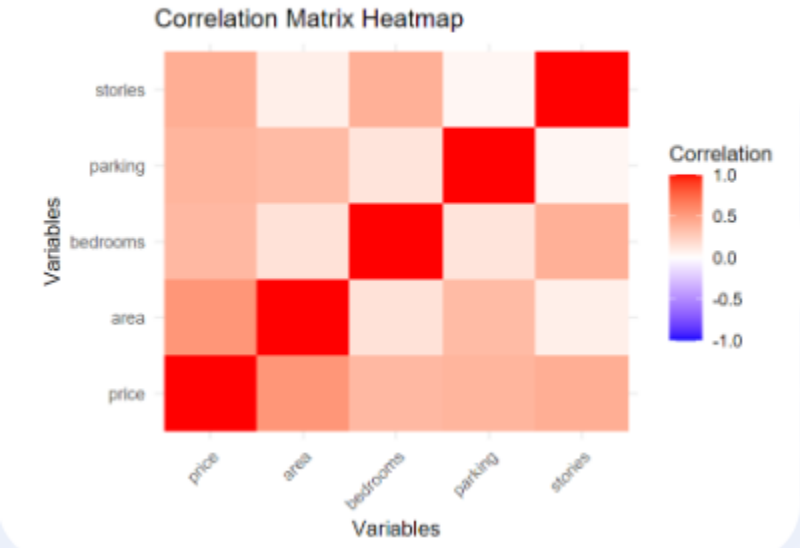
Based on the provided bar graph, what is the relationship between the number of stories in a building and the number of parking spaces required?

Answer: The bar graph illustrates a direct relationship between the number of stories in a building and the number of parking spaces needed. As the number of stories increases, the demand for parking spaces generally rises. This is likely due to the increased number of residents, businesses, and visitors associated with taller buildings. However, the graph also shows some variability in this relationship, suggesting that factors beyond the number of stories, such as building design or local regulations, might influence parking requirements.

Based on the correlation matrix heatmap provided, what are the strongest positive correlations observed among the variables: air conditioning, price, area, bedrooms, parking, and stories?

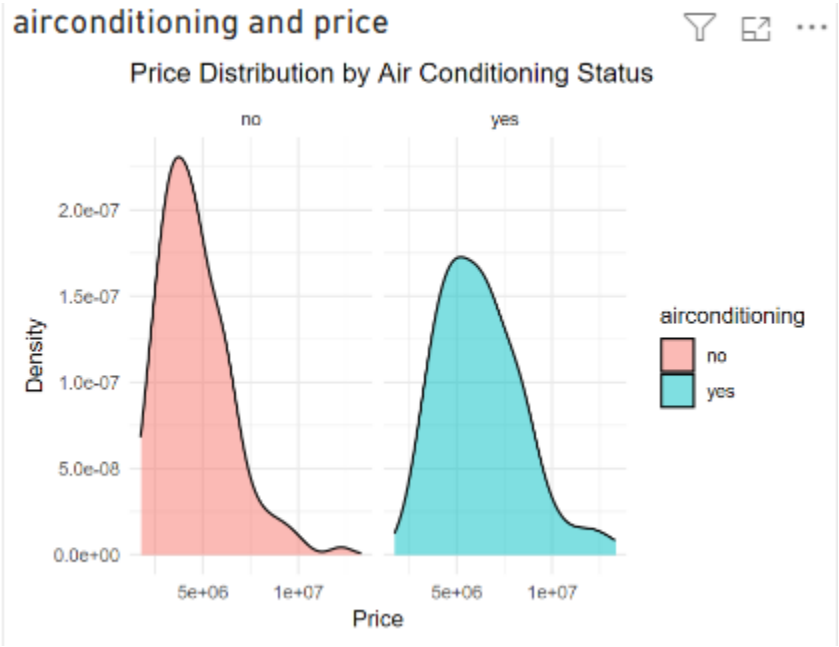
Answer: The heatmap reveals several strong positive correlations between the variables. Notably, there's a very high correlation between the number of stories and the number of parking spaces, indicating that buildings with more stories tend to have more parking facilities. Additionally, a moderate positive correlation exists between price and area, suggesting that larger properties generally have higher prices. Furthermore, there's a weaker positive correlation between the number of bedrooms and the area of a property, implying that properties with more bedrooms tend to be larger. These correlations highlight the interconnectedness of these variables in the context of real estate data.

airconditioning, price, area, bedrooms, parking and stories



Based on the density plot, how does the distribution of property prices differ between properties with air conditioning and those without?

Answer: The density plot illustrates that properties with air conditioning tend to have a higher distribution of prices compared to those without. The peak of the density curve for properties with air conditioning is shifted to the right, indicating that a larger proportion of these properties have higher prices. This suggests that air conditioning is a feature that can significantly increase the value of a property. However, it's important to note that there is still some overlap between the two distributions, indicating that other factors besides air conditioning influence property prices.



CONCLUSION :

- The analysis of house prices in relation to furnishing status, as illustrated by the box plots and regression plots, reveals several key insights into the real estate market dynamics:
- **Impact of Furnishing on Price:** The findings clearly indicate that furnishing status significantly influences house prices. Furnished properties tend to have the highest median prices and greater variability, suggesting they cater to buyers seeking convenience and a complete living solution. In contrast, unfurnished houses, which represent the lowest price point, attract budget-conscious buyers looking for flexibility.
- **Market Segmentation:** The variations in price distributions among furnished, semi-furnished, and unfurnished houses highlight a segmented market. Each category targets different buyer demographics, from high-end consumers interested in furnished properties to those seeking more affordable, unfurnished options.
- **Pricing Strategies:** Sellers can leverage these insights to refine their pricing strategies. By understanding the value attributed to furnishing status, sellers may enhance their properties' appeal—such as by offering semi-furnished homes or adding furnishings to unfurnished properties—to capture a broader audience and maximize their investment.
- **Opportunities for Value Addition:** The analysis suggests potential opportunities for sellers of semi-furnished and unfurnished properties to increase perceived value through strategic enhancements. This could involve providing furnishings that align with market expectations without significantly raising costs.