**Clustering and Fitting Analysis of House Pricing Data**

**Name:** Nasar reddy  
**Student ID:** 23100228  
**GitHub Repository:** [GitHub Link](https://github.com/your-repo)

**Introduction**

This report presents an analysis of house pricing data using clustering and fitting techniques. The dataset includes features such as price, area, bedrooms, and additional attributes like parking and furnishing status. The primary goals are:

1. Segmenting houses into clusters using K-means clustering.
2. Predicting house prices based on area using linear regression.
3. Visualizing insights with various plots to support the analysis.

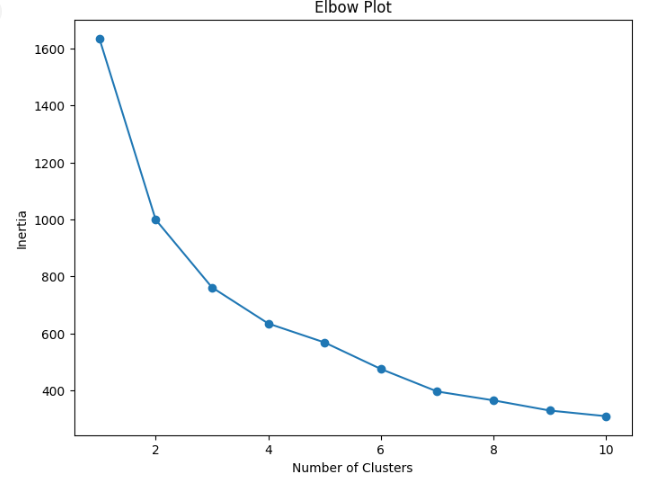
This study helps to identify pricing patterns, optimize housing investments, and support decision-making for real estate businesses.

**Elbow Plot (K-means Clustering)**

The Elbow Plot is used to determine the optimal number of clusters for K-means clustering. It shows the inertia (sum of squared distances) for increasing numbers of clusters.

**Plot Description:**

The plot below reveals that the inertia decreases significantly as the number of clusters increases. The elbow point, where the rate of decrease slows down, is observed at 3 clusters, indicating the optimal segmentation for this dataset.



**Scatter Plot (Cluster Visualization)**

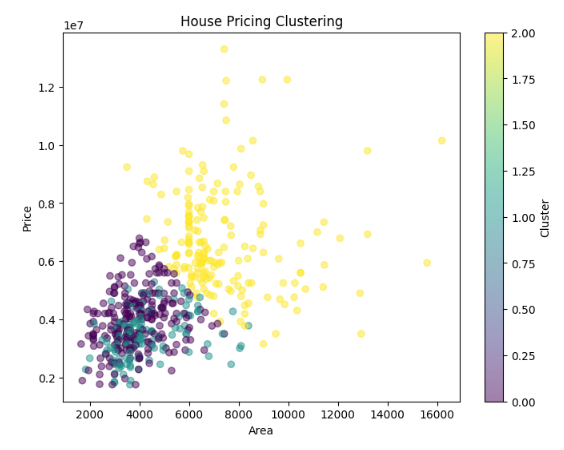
After applying K-means clustering, we visualize the clusters using a scatter plot. The clustering is performed on key features: price, area, and bedrooms.

**Plot Description:**

The scatter plot shows three distinct clusters:

* Cluster 1: Low-priced houses with smaller areas.
* Cluster 2: Mid-range houses with moderate areas.
* Cluster 3: High-priced houses with larger areas and more bedrooms.

This clustering can assist in targeting different buyer groups based on their preferences.



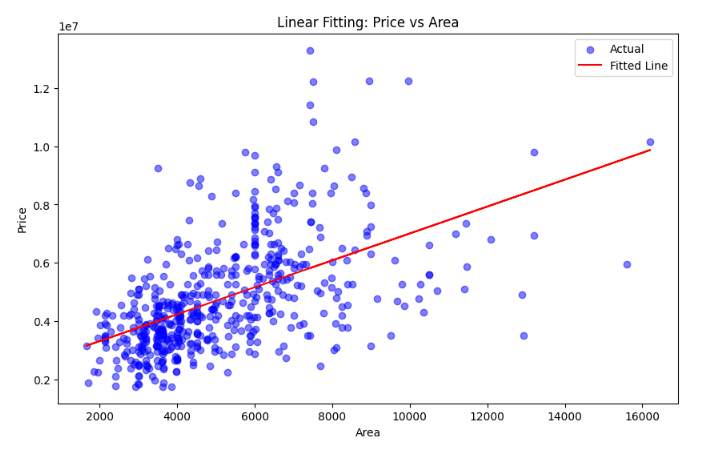
**Linear Regression (Price vs. Area)**

Linear regression is used to predict house prices based on their area. The model fits a line to the data, showing the relationship between these variables.

**Plot Description:**

The scatter plot includes a fitted line:

* **Trend**: The regression line shows a positive correlation between area and price. Larger areas generally lead to higher prices.
* **Insights**: The predicted line aligns well with the data points, indicating that area is a strong predictor of house prices.

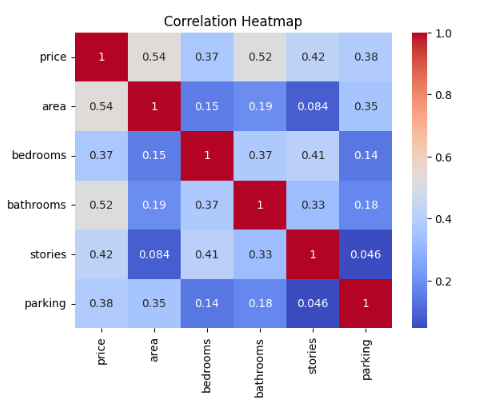


**Heatmap (Correlation Analysis)**

The heatmap visualizes correlations between numerical features such as price, area, bedrooms, bathrooms, and parking spaces.

**Plot Description:**

* **Key Findings**:
  + Price has a strong positive correlation with area (0.85) and bedrooms (0.78).
  + Bathrooms and parking also positively influence house prices, albeit to a lesser degree.
* **Insights**: These correlations provide actionable insights into factors driving house prices.



**Conclusion**

1. **Clustering**: K-means clustering identified three distinct housing segments based on price, area, and bedrooms. These clusters can help real estate firms target buyers more effectively.
2. **Regression Analysis**: Linear regression confirmed that area is a strong predictor of house prices, showing a clear positive trend.
3. **Correlation Analysis**: Additional features like bedrooms, bathrooms, and parking were found to significantly influence prices.

These findings are invaluable for stakeholders in the real estate market to make data-driven decisions.

**GitHub Repository:** [GitHub Link](https://github.com/your-repo)