Task 3: Linear Regression - Housing Dataset Analysis

Task 3: Linear Regression - Housing Dataset Analysis

Objective:

Implement and understand simple and multiple linear regression.

Dataset Summary:

- Total entries: 545

- Features: 12 + 1 target variable (price)

Modeling:

- Linear Regression model trained using scikit-learn

- Train-Test Split: 80% training, 20% testing

Evaluation Metrics:

- Mean Absolute Error (MAE): 970,043.40

- Mean Squared Error (MSE): 1,754,318,687,330.66

- R-squared (R^2) Score: 0.6529

Model Interpretation:

- Coefficients:

  - area: 235.97

  - bedrooms: 76,778.70

  - bathrooms: 1,094,444.79

- stories: 407,476.59
- parking: 224,841.91
- mainroad_yes: 367,919.95
- guestroom_yes: 231,610.04
- basement_yes: 390,251.18
- hotwaterheating_yes: 684,649.89
- airconditioning_yes: 791,426.74
- prefarea_yes: 629,890.57
- furnishingstatus_semi-furnished: -126,881.82
- furnishingstatus_unfurnished: -413,645.06

Conclusion:

- The model shows how each feature influences the house price.
- The R^2 score of 0.6529 indicates the proportion of variance explained by the model.
- The regression line plot for 'area vs price' is attached for visual analysis.

Simple Linear Regression Plot: Area vs Price