Predicting House Prices using Linear Regression

Data Analytics Internship

Anisha More

Problem Statement

- Predict house prices
- Use numerical features
- Apply Linear Regression
- Evaluate model performance

Dataset Overview

1460

Total Records

81

Total Features

columns

Target Variable: SalePrice

Mix of numerical & categorical features

Data Cleaning & Preparation

Handled missing values

Selected important numerical features

Removed unnecessary columns

Train-Test Split (80:20)

Feature Selection

Important Features:

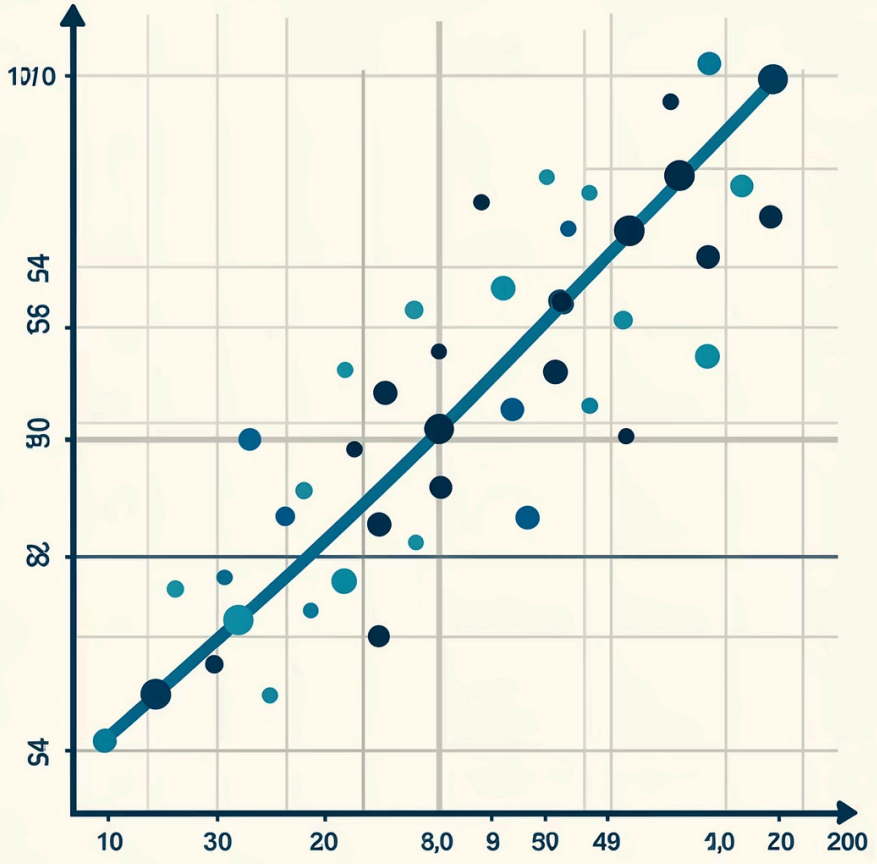
- OverallQual
- GrLivArea
- GarageCars
- TotalBsmtSF
- YearBuilt

Model Training

Algorithm: Linear Regression

Library: Scikit-learn

- Supervised Learning
- Continuous Prediction



Model Evaluation

R^2 Score: 0.81

Mean Squared Error: 1.4×10^{10}

Good fit between predicted & actual values

Visualization

- Scatter Plot: Actual vs Predicted
- Positive Linear Relationship
- Points close to regression line

Key Insights

- Overall Quality has highest impact
- Larger living area increases price
- Newer houses priced higher
- Garage capacity affects value

Conclusion

- Successfully built regression model
- Achieved 81% accuracy (R^2)
- Identified key price-driving factors
- Useful for real estate valuation