House Price Prediction – Project Summary

By Niher Ranjan Halder

Table of Contents

1. Project Overview

This project focused on predicting house prices using structured tabular data with over 80 features such as area, location, amenities, and house condition. The objective was to build a robust regression model that could estimate prices with strong generalization.

2. Tools & Technologies Used

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- scikit-learn
- XGBoost
- Jupyter Notebook

3. Data Summary

The dataset included approximately 1500 observations with 80+ features.
Key features included:

- Lot size
- Year built
- Overall quality
- Neighborhood
- Number of rooms

Target variable: `SalePrice`

4. Data Preprocessing & EDA

- Handled missing values using mean/median and domain-based methods
- Applied label encoding and one-hot encoding for categorical features
- Correlation analysis and heatmaps were used to identify important variables
- Visualized price distributions, outliers, and relationships using boxplots and bar charts

5. Model Building & Tuning

Used **XGBoost Regressor** due to its:

- High accuracy
- Built-in regularization
- Suitability for structured/tabular data

Tuning:

- Applied GridSearchCV and EarlyStopping
- Evaluated using RMSE on validation data

6. Model Evaluation & Feature Importance

- RMSE score on validation set used as main metric
- Plotted residuals and prediction error distribution
- Top contributing features:
  - `OverallQual`
  - `GrLivArea`
  - `GarageCars`
  - `TotalBsmtSF`
  - `YearBuilt`

Visualized feature importances from the trained XGBoost model.

7. Business Use Cases

This model can be applied in:

- Real estate pricing systems
- Investment analysis platforms
- Buyer-side recommendation engines
- Mortgage and insurance risk assessment tools

8. Conclusion & Learnings

This project strengthened my practical understanding of:

- End-to-end tabular data pipelines
- Handling real-world missing/categorical data
- Regression with XGBoost
- Model interpretation for business insight

It also improved my ability to explain ML results clearly for clients and real-world stakeholders.