House Price Prediction Project Report

# 1. Introduction

House price prediction involves estimating the selling price of a house using historical data and relevant features such as location, year built, condition, and other physical characteristics. This project aims to apply data analysis techniques and machine learning to identify key factors affecting house prices and develop a predictive model.

# 2. Methodology

- Tools Used: Python (Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn), Power BI, Excel  
- Process Flow:  
 1. Data Collection & Exploration  
 2. Data Cleaning (handling missing values, outliers)  
 3. Feature Engineering  
 4. Data Visualization  
 5. Model Development  
 6. Insights and Interpretation

# 3. Requirement Analysis

- Dataset: Contains multiple features related to houses (e.g., year built, lot area, neighborhood, etc.)  
- Target Variable: SalePrice  
- Goals:  
 - Analyze features affecting sale price  
 - Build an interpretable predictive model  
 - Generate actionable business insights

# 4. Other Parameters Depending Upon the Project

- Missing Values: Identified and handled via imputation (mean, mode, or removal)  
- Outlier Detection: Used box plots and IQR  
- Feature Types:  
 - Continuous (e.g., LotArea, GrLivArea)  
 - Discrete (e.g., OverallQual)  
 - Categorical (e.g., Neighborhood, HouseStyle)  
 - Temporal (e.g., YearBuilt, YrSold)

# 5. All Visualizations (from Dashboards and Notebook)

Key Charts and Graphs:  
- Histogram: Distribution of SalePrice  
- Heatmap: Correlation matrix of numerical features  
- Boxplots: Relationship between OverallQual, YearBuilt vs SalePrice  
- Scatterplots: GrLivArea vs SalePrice  
- Bar Charts: Categorical variables like Neighborhood vs Average SalePrice  
- Power BI Dashboards: Sales trends, top features, filters by location/year

# 6. Insights from the Analysis

- Strong Correlations:  
 - OverallQual, GrLivArea, GarageCars, TotalBsmtSF show strong positive correlation with SalePrice.  
- Categorical Impact:  
 - Neighborhood and HouseStyle significantly influence price.  
- Temporal Influence:  
 - YearBuilt and RemodAdd impact value—newer houses tend to have higher prices.  
- Feature Engineering Enhanced Prediction:  
 - Log transformation of skewed data improved model accuracy.  
 - Grouping rare categories in categorical features helped reduce noise.