# House Price Prediction Project Report

## 1. Introduction

This project aims to build a machine learning model to predict house prices using the Ames Housing dataset. The dataset provides rich information about various characteristics of residential homes in Ames, Iowa, USA.

## 2. Data Source

The dataset used in this project was obtained from Kaggle: https://www.kaggle.com/datasets/prevek18/ames-housing-dataset?resource=download  
It contains detailed information on 82 variables describing aspects such as the physical characteristics of the house, location, and sale price.

## 3. Project Objectives

- Develop a machine learning model capable of predicting house prices accurately.  
- Understand the relationship between various features and house sale prices.  
- Identify the most influential factors affecting the house prices.  
- Deploy this project as a portfolio piece for job applications in the data field.

## 4. Data Variables

The dataset contains both independent and dependent variables:  
- Dependent variable: SalePrice (the house sale price).  
- Independent variables include:  
 • GrLivArea (Above grade living area)  
 • OverallQual (Overall material and finish quality)  
 • GarageCars (Number of garage cars)  
 • TotalBsmtSF (Total basement area)  
 • 1stFlrSF (First floor square feet)  
 • YearBuilt (Year built)  
 • FullBath (Full bathrooms above grade)

## 5. Model Building Steps

1. Load the dataset using pandas.  
2. Explore and visualize the correlation of variables with SalePrice.  
3. Split the data into training and testing sets.  
4. Handle missing values using median/mode imputation.  
5. Encode categorical variables using OneHotEncoder.  
6. Normalize numerical values using StandardScaler.  
7. Build a pipeline combining preprocessing and a Random Forest Regressor.  
8. Train the model and evaluate its performance using Root Mean Squared Error (RMSE).  
9. Visualize feature importance and correlations.

## 6. Importance and Impact

This model is crucial for real estate valuation and decision-making. By analyzing key features, buyers and sellers can understand what drives property prices. The model also demonstrates essential data science skills, making it a strong portfolio piece for job seekers in the AI and analytics domain.

## 7. Results

The model achieved a low RMSE score on the test data, indicating high prediction accuracy. The most important predictors of house prices were GrLivArea, OverallQual, GarageCars, and TotalBsmtSF.