# Project Overview

The goal of this project is to build model that accurately predicts house prices based on various features such as area, number of rooms, condition, and location. This can help home buyers, real estate agents, or companies make informed decisions.

# Dataset Description

**File Used:** data.csv

**Target Variable:** Price

**Key Features:**

* Area: Size of the house in square feet
* Bedrooms: Number of bedrooms
* Bathrooms: Number of bathrooms
* Floors: Number of floors
* YearBuilt: Year the house was built
* Location: Urban, Suburban, Rural
* Condition: Good, Fair, Poor
* Garage: Yes or No

# Data Preprocessing

* Missing Value Handling
* Used .dropna() to remove rows with missing values
* Feature Engineering
* Created House\_age = 2025 - YearBuilt
* Total\_rooms = Bedrooms + Bathrooms
* Price\_per\_area = Price / Area
* Encoding Categorical Variables
* Used pd.get\_dummies() with drop\_first=True

# 4.Exploratory Data Analysis

* Visualized distributions with histograms and boxplots
* Analyzed correlation with Price using df.corr()
* Found that Area, Floors, and Condition\_Fair had higher positive correlation with Price

# Model Building

* Train-Test Split
* Used train\_test\_split from sklearn with 80% training and 20% test
* Linear Regression Model
* R² Score: 0.7428
* MSE: 20,006,299,283.45
* Random Forest Regressor
* R² Score: 0.9975
* MSE: 152,083,286.91

# Final Model Deployment

* Model saved using joblib: joblib.dump(rf, 'house\_price\_rf\_model.pkl')
* This allows easy loading for future use or integration in applications

# Key Learnings

* Feature engineering and encoding are crucial in real estate data
* Linear models provide a baseline
* Ensemble methods like Random Forest perform significantly better on complex data
* Model serialization enables deployment

# Next Steps

* Build a Streamlit web app for user interaction
* Try XGBoost or LightGBM for further improvement
* Use cross-validation for better generalization