Title: Predicting House Prices in the USA Using Machine Learning Models

Project Overview
This project aims to predict residential house prices across the United States using a large scale real estate dataset. The primary goal is to build machine learning models that accurately estimate housing prices based on these features, providing insights for buyers, sellers, and real estate agents.

This project aims to answer several questions such as: What are the main factors that affect house prices in the United States? How do features like the number of bedrooms, bathrooms, lot size, house size, and location influence the overall price of a house?

Data Source
The dataset, sourced from Kaggle and includes over 2 million records and covers essential property characteristics.

Dataset Link: USA Real Estate Dataset - Kaggle
https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset

Dataset Summary
The dataset contains:
- 2,226,382 entries.
- Key variables: price, bed, bath, acre_lot, house_size, city, zip_code, prev_sold_date, brokered_by, status.

Modeling Approach and Analysis
Exploratory Data Analysis (EDA)
- Distribution analysis of housing prices
- Relationship between price and number of beds, baths, lot size, and house size
- Price trends by city and zip code

Regression Model
- **Goal:** Predict house price
- **Input Features:** bed, bath, acre_lot, house_size, city, status, brokered_by, zip_code

Models to Use:
  I am planning to use the models mentioned below for predicting house prices. But, the final choice of the best model will depend on the evaluation results, such as prediction accuracy and error metrics.

  o Random Forest Regressor
  o Gradient Boosting (XGBoost / LightGBM)

- o Linear Regression (as baseline)

- **Metrics:**
    - o Root Mean Squared Error (RMSE)
    - o Mean Absolute Error (MAE)
    - o R-squared (R² Score)

Power BI / Tableau Dashboard
Visuals:
- Map View: Median house prices by city/state
- Filters: Number of beds, city, zip code, house size range
- Charts:
    - o Avg price by number of beds and baths
    - o Price trend by city
    - o Top 10 most expensive cities

Streamlit Web App
User Inputs:
- Number of beds and baths
- House size
- City and zip code
- Lot size and status

Model Output:
- Predicted house price

Conclusion
This project will demonstrate how real estate data can be used to build accurate, predictive pricing models. It will help buyers and sellers make informed decisions and assist realtors with better market insights.