House Price Prediction App

This is a **House Price Prediction** application built using **Random Forest Regressor** and **Streamlit** for an interactive user interface. The app predicts the price of a house based on various features like bedrooms, bathrooms, area, year built, city, and more.

➢ Features

- Input all relevant house features:

- Bedrooms, Bathrooms, Living Area (sqft), Lot Size (sqft), Floors

- Waterfront, View, Condition

- Above Ground Area, Basement Area

- Year Built, Year Renovated

- Days since Earliest Sale

- City (select from dropdown to avoid errors)

- Predicts **house price** instantly

- Displays **Top 5 Feature Importance** for better understanding

- Uses **trained Random Forest model** (`house_price_model.pkl`)

- Uses **city encoder** (`city_encoder.pkl`) to handle categorical city data

➢ Folder Structure

- HousePriceProject/
  - app.py # Streamlit app
  - house_price_model.pkl # Trained Random Forest model
  - city_encoder.pkl # Saved Label Encoder for cities
  - data.csv # (optional, reference dataset)
  - README.md # Project description

- **How It Works**

1. Data Cleaning & Preprocessing
   - Duplicate and missing rows are removed.
   - Extreme outliers in price are removed (1% - 99% quantiles).
   - Categorical features (City) are label encoded.

2. Feature Engineering
   - Date converted to numeric (`days since earliest sale`).
   - Selected numeric features include bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfront, view, condition, sqft_above, sqft_basement, year built, year renovated, and date_numeric.
   - City is transformed using the saved encoder.

3. Prediction
   - User inputs all house features in the Streamlit UI.
   - City selected from dropdown ensures no unseen city error.
   - Model predicts price using **Random Forest Regressor**.
   - Top 5 features contributing to the prediction are displayed in a bar chart.

- **How to Run**

1. Open terminal in the project folder
2. Install required Python libraries:
   Pip install streamlit pandas numpy scikit-learn matplotlib joblib
3. Run the app:
   Streamlit run app.py
4. Browser will open → Enter house details → Click **Predict Price** → Predicted price and top Features will be displayed.

Notes

- **City dropdown:** Only cities present in the training dataset can be selected.
- **Numeric inputs:** Bedrooms, Bathrooms, and other numeric fields are set to proper integer increments.
- **Model & Encoder Files:** Make sure `house_price_model.pkl` and `city_encoder.pkl` are in the same folder as `app.py`.

Author

Mehak
Thapar Institute of Engineering and Technology, Patiala