BRD for Housing Price Prediction using Machine Learning - Gowtham

Objective:

Predict housing prices based on multiple factors like location, size, and amenities using advanced regression techniques.

Tasks:

1. **Data Collection:**
    - Use publicly available housing price datasets (e.g., from **Zillow** or **Kaggle**) or company-provided data.
2. **Data Preprocessing:**
    - Clean the dataset using **Pandas** (handle missing data, normalize features, and handle categorical variables using **One-Hot Encoding**).
    - Feature engineering to create additional features such as price per square foot or neighborhood ranking.
3. **Regression Model Development:**
    - Build and evaluate regression models like **Linear Regression, Ridge Regression, Lasso, and XGBoost** using **Scikit-learn**.
    - Use cross-validation to select the best-performing model.
4. **Model Optimization:**
    - Fine-tune the model by selecting the best hyperparameters and optimizing for metrics like **RMSE** and **MAE**.

5. **Feature Importance:**

    - Analyze feature importance using tools like **LIME** or **SHAP** to understand the impact of each factor on the housing prices.

6. **Visualization:**

    - Create visualizations (scatter plots, bar plots, heatmaps) showing relationships between different features and house prices.

7. **Tableau Dashboard:**

    - Develop a Tableau dashboard showcasing predicted house prices, actual vs predicted, and important features affecting the prices.

Deliverables:

- Python code for housing price prediction.
- Visualizations using Seaborn and Matplotlib.
- Tableau dashboard showcasing price predictions and insights.