Austin House Price Prediction

**Description**:

This project is about predicting the housing price for Austin as it is one of the hottest markets in 2023, and these price predictions show how the market will change over the years. This dataset was sourced from Kaggle. After performing the EDA and preparing the data, I trained the model using different regression techniques like Decision Tree, Random Forest, Linear Regression and then evaluated the best fit with RMSE value.

**Conclusion**:

In this project, I created a model to predict the listing price of homes in Austin, TX using a dataset of recent house listings, including details like location, amenities, and prices. The data was preprocessed, handling missing values, converting boolean columns to integers, and encoding categorical features. Features like living area square footage, lot size, longitude, and latitude were identified as significant predictors through correlation matrix. Model Building and Evaluation:

1. Linear Regression: Best Parameters: The best linear regression model was found with the parameter fit\_intercept set to True. Performance: The Root Mean Squared Error (RMSE) on the test set was approximately 330,854.
2. Random Forest: Best Parameters: The optimal random forest hyperparameters were found to be max\_depth: 15, min\_samples\_split: 5, and n\_estimators: 100. Performance: The Random Forest model achieved an RMSE of approximately 290,847.
3. LightGBM: Performance: The LightGBM model achieved an RMSE of around 281,381. Feature Importance: The top features influencing the model included living area square footage, lot size, longitude, and latitude.

**Overall Evaluation:**

1. The LightGBM model outperformed the Random Forest and Linear Regression model, achieving a lower RMSE.
2. Living area square footage, lot size, and location (longitude, latitude) play a crucial role in predicting housing prices.
3. The Random Forest and LightGBM model, with carefully tuned hyperparameters, seems to be a suitable choice for predicting housing prices in the Austin area. LightGBM is beneficial if top features are mostly categorical.

**Future Steps:**

1. Further optimization of hyperparameters and feature engineering could potentially improve model performance.
2. If the model meets the desired performance, deploy it for real-time predictions and continuous monitoring.

This project provides insights into predicting housing prices in Austin, emphasizing the importance of location and specific property features.