Project Summary: HOA Fee Prediction Model

Overview:
This project aims to develop a machine learning model that predicts whether a property will have HOA (Homeowners Association) fees and, if applicable, estimates the potential HOA fee amount. This model provides valuable insights for existing homeowners considering joining an HOA and prospective homebuyers looking to understand potential fees before purchasing a property.

Business Problem:
Homeowners and buyers often lack information regarding HOA fees, which makes it challenging to make informed decisions about joining an HOA or purchasing a property. This model addresses the problem by predicting whether a property is likely to have HOA fees based on its features, such as location, property size, and other attributes. For properties predicted to have fees, the model also estimates the potential fee amount. This helps homeowners and buyers assess their financial commitments and make better decisions.

Objective:
The primary objective of the project is to:
1. Predict whether a property will incur HOA fees.
2. Estimate the potential HOA fee amount for properties predicted to have fees.

Data:
The project uses a dataset containing property details such as sold price, square footage, lot size, number of bedrooms and bathrooms, fireplaces, and more. The HOA fees column is used as the target variable for both classification and regression tasks.

Approach:
1. Data Cleaning & Preprocessing: Handled missing values, removed duplicates, performed feature engineering (including log and square root transformations), and treated outliers.
2. Classification Model: Developed a model to classify properties into two categories: "No Fee" or "Has Fee." Techniques like K-Nearest Neighbors (KNN) and Gaussian Naive Bayes were applied to predict whether a property would have HOA fees.
3. Regression Model: For properties predicted to have HOA fees, a regression model was used to estimate the fee amount using features such as location, property size, and year built.
4. Evaluation: The models were evaluated using metrics such as accuracy for classification and RMSE (Root Mean Squared Error) for regression.

Results:
- The KNN classifier model achieved the best accuracy of 93%, successfully predicting whether properties would have HOA fees.
- The KNN regressor model estimated HOA fees with an RMSE of 45.82, indicating that the model can provide a reasonable estimate of HOA fees, though further optimization may be required.

Use Case:
This model is valuable for:
- Homeowners: Helps existing homeowners understand the potential costs if they decide to join a nearby HOA.
- Real Estate Platforms: Can be integrated into online real estate platforms to provide users with additional insights about properties, improving the overall customer experience by offering HOA fee predictions.

Conclusion:
The HOA Fee Prediction Model provides a practical solution to help homeowners and buyers make informed decisions regarding HOA fees. By predicting whether a property will have fees and estimating the fee amount, the model helps reduce uncertainty and supports better financial planning.