Evaluation Report: Linear Regression Model for Median house value Prediction

Objective: Predict median house value based on available features (median income, total bedrooms, total rooms,…etc) using a Linear Regression model.

1. Model Evaluation Metrics:

- **Mean Squared Error (MSE)**:
  The MSE is a measure of how well the model's predictions match the actual median house value s. A lower MSE indicates better performance. The value is lower compared to the previous model, indicating improved accuracy.
    - **MSE**: **5,059,928,371.17**
    - This MSE value suggests that the model's predictions are closer to the actual values, but there's still some room for improvement.
- **Root Mean Squared Error (RMSE)**:
  The RMSE gives us an idea of the magnitude of the error in the same units as the target variable (median house value). A lower RMSE indicates better predictive accuracy.
    - **RMSE**: **71,133.17**
    - The RMSE of $71,133 indicates that, on average, the model's predictions are off by approximately $71,133. This is a much smaller error compared to the previous RMSE of $233,418, which shows improvement in prediction accuracy.
- **Mean Absolute Error (MAE)**:
  The MAE measures the average magnitude of errors in the predictions. It gives a direct interpretation of the average error made by the model without considering the direction of the error.
    - **MAE**: **51,810.48**
    - The MAE of $51,810 indicates that, on average, the model is off by about $51,810 in predicting median house value s. This is an improvement from the previous MAE of $183,592, showing that the model's accuracy has significantly improved.
- **R-squared (R^2)**:
  The R^2 value indicates how well the model explains the variance in the target variable (median house value s). A higher R^2 means the model fits the data better. A value closer to 1 indicates a good fit, while a value closer to 0 indicates that the model does not explain much of the variance.
    - **R^2**: **0.6139**
    - The R^2 value of **0.6139** suggests that the model explains approximately 61.39% of the variance in median house value s. This is a significant improvement over the previous negative R^2 value, indicating a more meaningful fit and predictive power.

2. Analysis:

- The **MSE** and **RMSE** values are significantly lower than before, indicating that the model is now making more accurate predictions.
- The **MAE** value has also decreased, showing that the model's average error has been reduced.

- The **R^2** value of **0.6139** indicates a reasonable fit, meaning that the model explains over half of the variance in median house value s, which is a significant improvement compared to the previous negative R^2.

3. Recommendations:

- **Model Complexity**: While the linear regression model has shown improvements, more complex models (**Random Forest** or **Gradient Boosting** algorithms like **XGBoost**) may further enhance performance.
- **Regularization**: consider applying regularization techniques like **Ridge** or **Lasso Regression** to prevent the model from learning noise in the data.
- **Hyperparameter Tuning**: Further tune hyperparameters using methods (**Grid Search** or **Random Search**)to optimize model performance.
- **Cross-validation**: Use **k-fold cross-validation** to assess model performance more robustly and prevent overfitting or underfitting.

4. Conclusion:

The Linear Regression model shows a significant improvement in predicting median house value s, as indicated by the lower MSE, RMSE, and MAE, and the positive R^2 value of **0.6139**. The model is now able to explain a substantial portion of the variance in median house value s, with a much lower error margin compared to the initial iteration. However, there is still room for improvement, particularly in reducing the prediction error and further increasing the explanatory power of the model.