My Approach to the Problem

In this project, I worked on predicting equipment energy consumption in a smart factory using sensor data gathered from different zones and environmental conditions. My approach was methodical and involved the following steps:

- **Data Loading & Preprocessing**: I cleaned the dataset, handled missing values, and prepped the features for modeling.
- **Exploratory Data Analysis (EDA)**: I explored the distributions, checked feature relationships, and looked for correlations with the target variable to get a good understanding of the data.
- **Feature Engineering & Selection**: I evaluated the usefulness of each feature—including two random variables—to reduce noise and improve model accuracy.
- **Model Development & Training**: I implemented and trained three regression models: Linear Regression, Random Forest, and Gradient Boosting. I applied hyperparameter tuning and cross-validation to fine-tune performance.
- **Model Evaluation & Testing**: I compared the models using metrics like RMSE, MAE, and R^2, and analyzed feature importance to see which factors most influenced energy consumption.

Key Insights from the Data

- Environmental factors like temperature and humidity across various factory zones had a clear impact on energy consumption.
- The target variable (energy use) was right-skewed, with most readings on the lower end.
- The two random variables provided in the dataset turned out to have negligible correlation with the target, and including them only added noise. Removing them simplified the model without hurting performance.

Model Performance Evaluation

- **Linear Regression** served as a good baseline, but its low R^2 score (~0.01) suggested it wasn't able to capture the complexities of the data.
- **Random Forest Regressor** performed best, with the lowest RMSE (~157) and highest R^2 (~0.07-0.08), thanks to its ability to handle nonlinearities and feature interactions.
- **Gradient Boosting Regressor** followed closely behind, doing better than Linear Regression but slightly below Random Forest.
- Removing the random variables had minimal impact on the scores, supporting their exclusion.

Recommendations for Reducing Energy Consumption

Based on the insights from the model and data analysis, here are some practical suggestions for reducing energy use in the smart factory:

- **Monitor Key Environmental Factors**: Focus on zones where temperature and humidity have the most impact. Tuning environmental controls in these areas can help cut energy costs.
- **Optimize Conditions in High-Impact Zones**: Adjust temperature and humidity settings where the models identified the strongest correlations with high energy use.
- **Use Data for Continuous Improvement**: Keep collecting and analyzing sensor data to improve future models and refine energy strategies.