Energy Consumption Prediction – Final Report

1. Approach to the Problem

We aimed to predict equipment energy consumption using multi-zone temperature, humidity, and weather features. The workflow included:

- Data cleaning and handling of null values.
- Converting timestamp and engineering time-based features.
- Dropping irrelevant columns and low-importance features.
- Using Random Forest Regressor with hyperparameter tuning for better accuracy.

---

2. Key Insights from the Data

- Several columns had substantial missing values, filled using column-wise means.
- Some columns had incorrect data types (e.g., numbers stored as strings); these were converted using `pd.to_numeric()`.
- Timestamp was useful in generating `hour`, `day`, `month`, and `is_weekend`, which showed modest correlation with energy consumption.
- Outdoor humidity and zone temperatures were among the top correlated and important features.

---

3. Model Performance Evaluation

Using **Random Forest Regressor**:

- **R^2 Score**: High (Exact value visible in notebook output)
- **RMSE** and **MAE**: Reasonably low, indicating good prediction accuracy.
- Feature importance highlighted outdoor and zone conditions as strong influencers.

Optional tuning and XGBoost models can be used to further refine performance.

4. Recommendations for Reducing Equipment Energy Consumption

- **Optimize based on high-impact features**: Focus on improving outdoor/zone temperature regulation and ventilation efficiency.
- **Smart scheduling**: Since time-related features like hour and weekday were useful, automate energy-heavy operations to off-peak hours or cooler times.
- **Anomaly detection**: Use this model to flag unusual consumption patterns for proactive maintenance.
- **Data-driven policy**: Regularly update the model with recent data for ongoing improvements.