Manufacturing Facility Energy Consumption Analysis

Executive Summary

This report analyses sensor data from a manufacturing facility to identify factors influencing equipment energy consumption and provides recommendations for optimization. Using machine learning models, we've identified key environmental factors that impact energy usage, with potential for up to 57% predictability of consumption patterns.

Approach to the Problem

Data Pre-Processing

- Cleaned dataset by removing negative energy consumption and humidity values
- Imputed missing values using median for numerical features
- Engineered time-based features from timestamps (month, day, hour, day of week)
- Created derived features (temperature and humidity differentials, zone averages)
- Applied outlier detection and clipping for features with >15% outliers
- Split data into training (60%), validation (20%), and test (20%) sets

Modelling Strategy

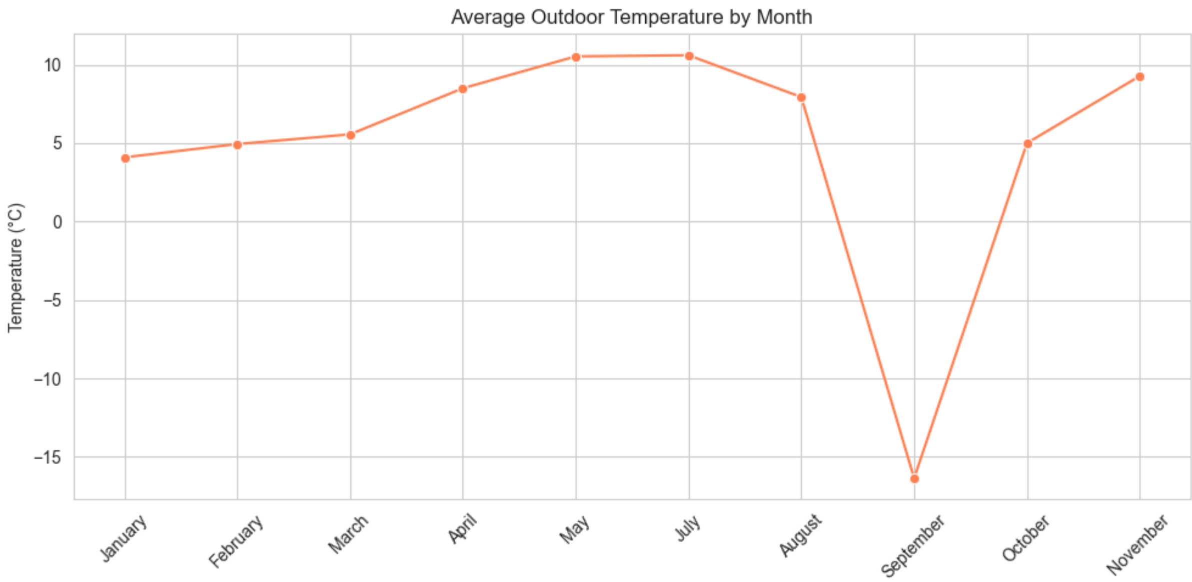
Evaluated multiple regression algorithms to predict equipment energy consumption:

- Linear models (Linear Regression, Ridge)
- Tree-based models (Decision Tree, XGBoost)
- Neural Network with regularization
- Support Vector Regression
- Feature engineering approaches (Polynomial Features, PCA)

Key Insights from the Data

Temporal Patterns

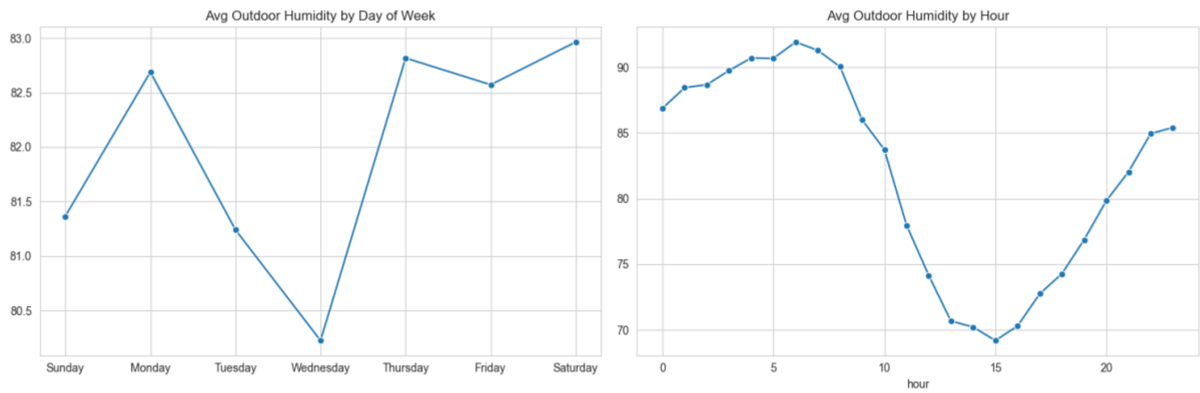
- **Daily patterns**: Energy consumption correlates with outdoor temperature, which peaks between 2-4 PM
- **Weekly variation**: Temperature is coolest on Monday and peaks on Friday
- **Seasonal trends**: Clear temperature progression from January to July, with a sharp drop in September



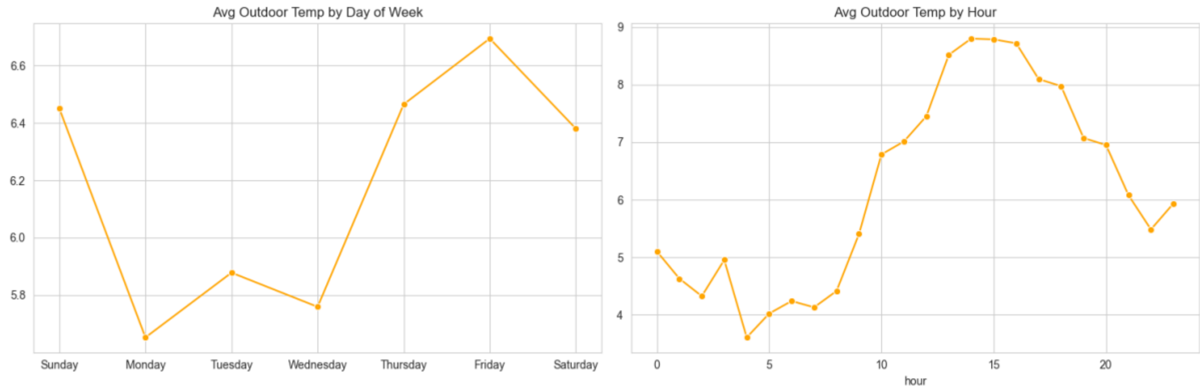
The graph shows a steady temperature increase from January to July, followed by a sharp drop in September and recovery in October, indicating strong seasonal patterns that impact energy consumption.

Environmental Factors

- **Humidity patterns**: Highest in early morning (5-7 AM), lowest in afternoon (2-3 PM)
- **Monthly humidity**: Generally declines from January to July, with a spike in September
- **Temperature-humidity relationship**: Inverse correlation between temperature and humidity.



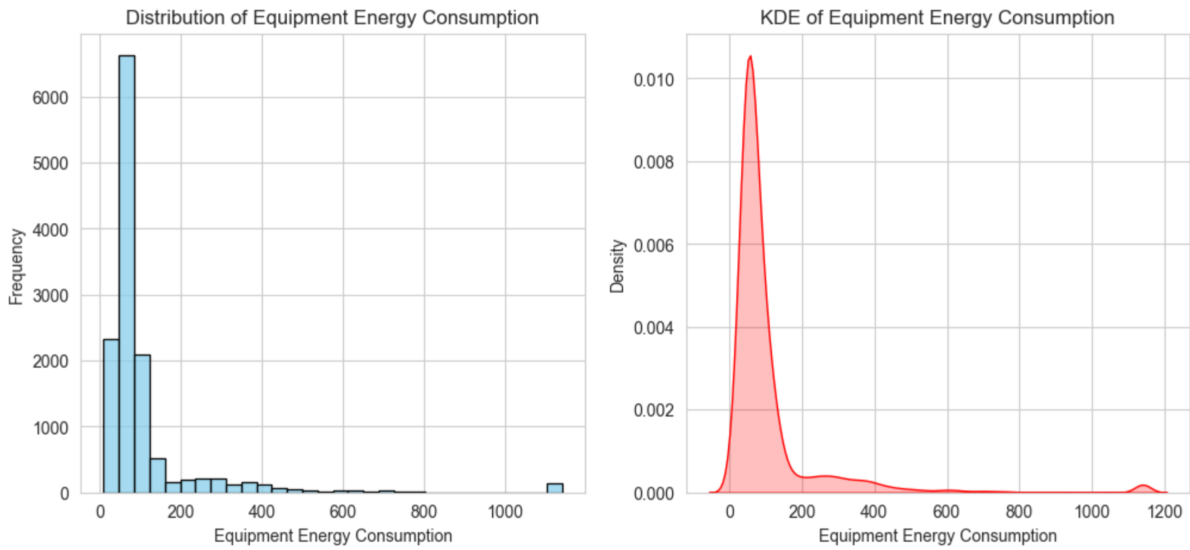
This visualization reveals humidity is highest in early morning (5-7 AM) and lowest in the afternoon (2-3 PM), showing an inverse relationship with temperature patterns throughout the day.



The dual plots demonstrate temperature variation across weekdays (coolest on Monday, warmest on Friday) and throughout the day (steadily rising after 9 AM, peaking between 2-4 PM).

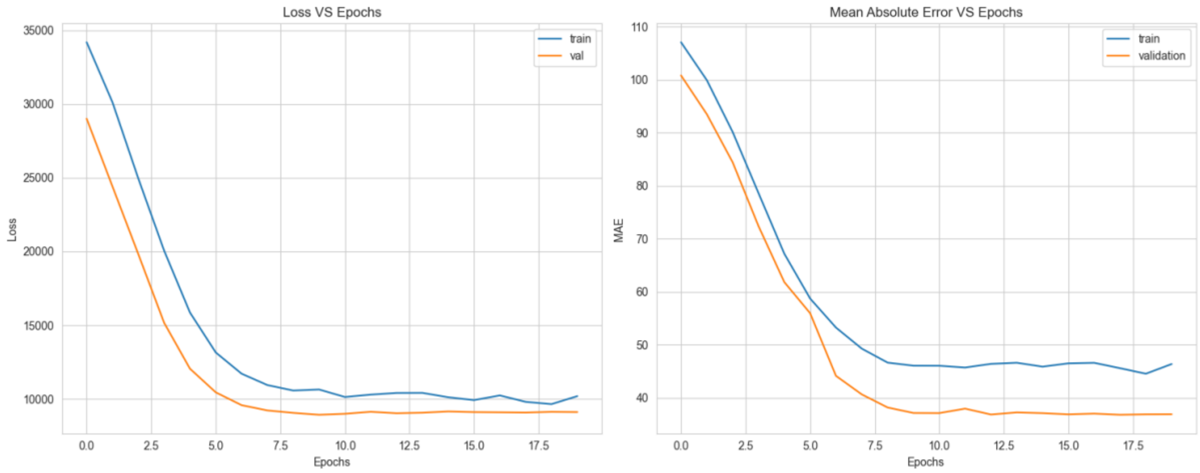
Equipment & Facility Insights

- Energy consumption distribution is right-skewed, with most values below 150 Wh
- Zone temperature and humidity variations across the facility show potential optimization opportunities
- Random variables tested showed no statistically significant impact on energy consumption

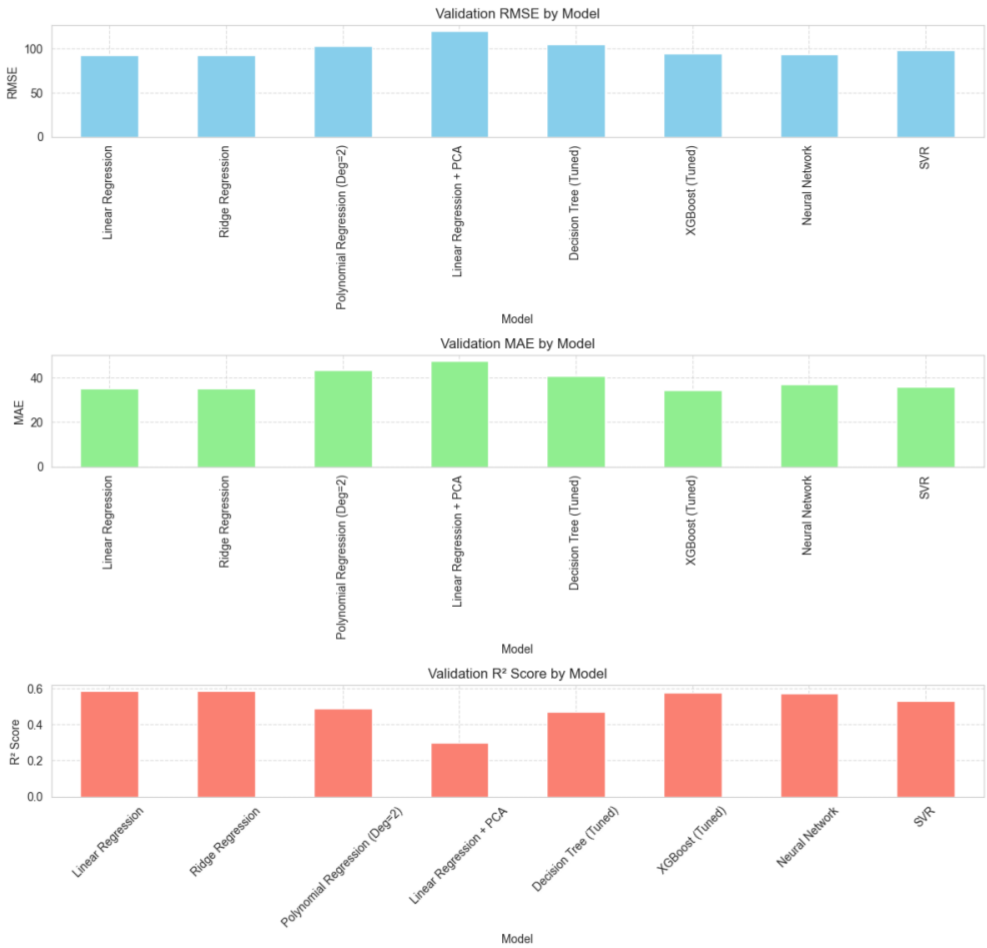


The histogram and KDE plot reveal that energy consumption is right-skewed, with most values concentrated below 150 Wh and a few outliers extending beyond 1000 Wh, suggesting potential optimization opportunities for high-consumption events.

Model Performance Evaluation



The training plots show the neural network's loss and MAE decreasing over epochs, with validation metrics stabilizing, indicating successful training without overfitting. This supports the model's ability to generalize to new data.



Bar charts comparing RMSE, MAE, and R^2 scores across all tested models, clearly showing that Linear Regression and Ridge Regression achieved the best overall performance.

Model	RMSE	MAE	R^2 Score
Linear Regression	92.66	35.26	0.59
Ridge Regression	92.66	35.26	0.59
XGBoost	94.27	34.24	0.58
Neural Network	94.24	36.99	0.58
SVR	98.66	35.78	0.53
Polynomial Regression	103.06	43.23	0.49
Decision Tree	105.39	40.91	0.47
Linear Regression with PCA	120.83	47.70	0.30

Final Test Set Performance (Linear Regression):

- RMSE: 91.68
- MAE: 35.72
- R^2 Score: 0.57

The Linear Regression model provided the best balance of performance and interpretability, explaining about 57% of the variance in equipment energy consumption.

Analysis of Random Variables

A specific hypothesis test was conducted to evaluate whether random_variable1 and random_variable2 had any meaningful impact on energy consumption:

Hypothesis Testing Approach

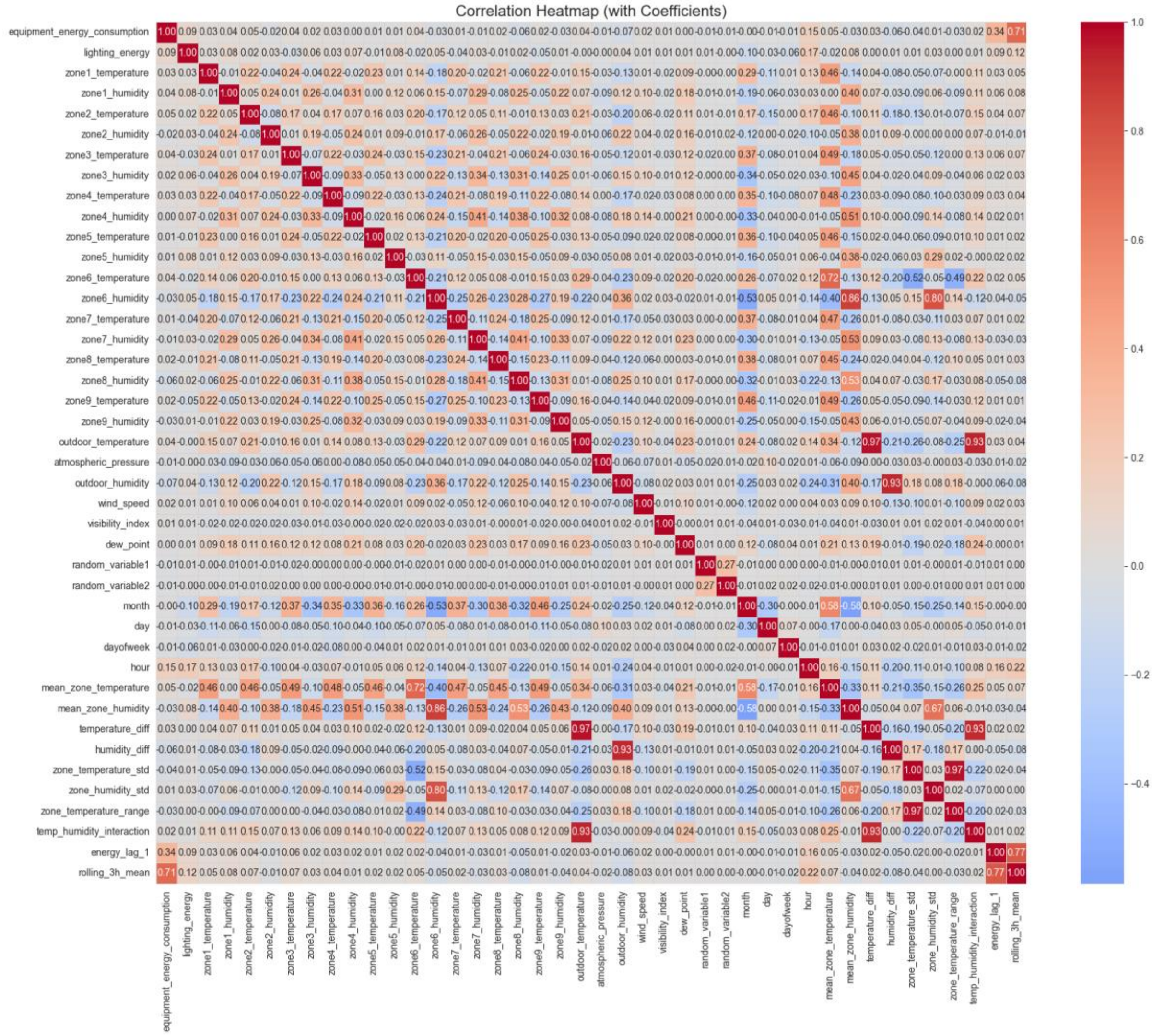
- **Null hypothesis (H_0)**: Regression coefficient of predictor variable (β) = 0 (the variable has no effect)
- **Alternative hypothesis (H_1)**: Regression coefficient (β) $\neq$ 0 (the variable significantly affects energy consumption)
- **Method**: Multiple linear regression with statistical significance testing

Results of Random Variables Analysis

- **R-squared = 0.000**: The model with only these variables explained 0% of variance in equipment energy consumption
- **F-statistic = 1.934, p-value = 0.145**: Overall model not statistically significant ($p > 0.05$)
- **random_variable1 coefficient = -0.0728, p-value = 0.165**: Not statistically significant
- **random_variable2 coefficient = -0.0517, p-value = 0.337**: Not statistically significant

Based on this rigorous statistical analysis, both random variables were removed from the dataset before model training, as they provided no predictive value for energy consumption.

Key Feature Correlations



The correlation heatmap reveals significant relationships between equipment energy consumption and environmental factors. Zone temperatures, outdoor conditions, and time features show the strongest correlations, while the random variables show no significant correlation with the target.

Recommendations for Reducing Equipment Energy Consumption

1. **Optimize HVAC scheduling**: Adjust zone temperatures based on outdoor conditions, especially during afternoon peak hours when outdoor temperatures are highest (2-4 PM).
2. **Implement zone-specific controls**: Different manufacturing zones show varying temperature and humidity patterns. Install zone-specific environmental controls to optimize conditions only where needed.

3. **Seasonal adjustments**: Develop different energy management strategies for summer months (May-July) versus fall/winter periods, as seasonal variations significantly impact energy usage patterns.
4. **Time-based production scheduling**: Schedule energy-intensive production during periods of lower environmental energy demand (early morning or evening).
5. **Weather-responsive operations**: Implement predictive controls that adjust equipment usage based on forecasted weather conditions, particularly focusing on humidity and temperature differentials.
6. **Equipment maintenance scheduling**: Use predictive models to schedule preventative maintenance during periods when energy consumption patterns suggest potential inefficiencies.
7. **Lighting optimization**: Consider coupling lighting systems with equipment operations as the analysis showed potential relationships between lighting energy and equipment energy consumption.
8. **Energy consumption monitoring**: Implement real-time monitoring using the predictive model to identify anomalies that might indicate equipment inefficiencies or failures.

By implementing these recommendations, the facility could potentially reduce equipment energy consumption while maintaining production efficiency. The predictive model can be further refined with additional data collection to improve accuracy and provide more targeted recommendations.