📜 Power Consumption Regression Analysis Report

Prepared by: [Solomon-Friday Ebuka]

Date: [30/10/2025]

Tool Used: Python (Pandas, Scikit-learn, Matplotlib)

1. Summary

This report presents the findings from a regression analysis conducted on a Power Consumption dataset.

The objective was to predict power usage based on available numeric factors and to evaluate the accuracy of the model's predictions.

A Linear Regression model was implemented as a baseline approach. The model achieved an R^2 score of 0.639, indicating that approximately 63.9% of the variation in power consumption can be explained by the selected input features.

2. Objective

The primary goal of this analysis was to:

- Identify the relationship between power consumption and other influencing factors in the dataset.
- Develop a regression model capable of estimating power usage.
- Evaluate the model's predictive accuracy using standard statistical metrics.

3. Data Overview

Dataset Name: powerconsumption.csv

Data Source: Internal Measurement Logs (assumed)

Number of Observations: Displayed in analysis output

Preprocessing Performed:

Filled missing numeric values with column means.

Retained only numeric variables for analysis.

Assumed the last numeric column represents Power Consumption (Target Variable).

4. Methodology

The analysis followed these structured steps:

Step	Description
Data Preparation:	Cleaned missing data and selected numeric columns.
Feature Selection:	Chose all numeric columns (except last) as predictors (X); last column as target (y).
Model Selection	Applied: Linear Regression from Scikit-learn.
Data Split:	80% training, 20% testing (randomized).
Evaluation Metrics:	Calculated MAE, MSE, RMSE, and R^2.
Visualization:	Created scatter plots to compare actual vs predicted values.

5. Model Performance

Metric Description   Result respectively:

MAE (Mean Absolute Error): Average absolute deviation between predictions and actuals

=3,086.29

MSE (Mean Squared Error): Average squared difference (penalizes large errors).

=15,740,099.14

RMSE (Root Mean Squared Error): Standard deviation of prediction errors.

=3,967.38

R^2 Score: Proportion of variance explained by the model.

=0.639

6. Interpretation

The model explains 63.9% of observed variation in power consumption — a moderate fit, indicating meaningful but incomplete predictive capability.

The MAE of 3,086 suggests that on average, predictions deviate by roughly 3,000 units from actual values.

Depending on the scale of power usage (e.g., 20,000+ units), this may represent an acceptable margin of error for operational and dependable forecasting.

7. Visualization Insights

Actual vs Predicted Power Consumption

The scatter plot comparing predicted values against actual values indicates that most predictions align near the ideal diagonal line — confirming that the model captures the general pattern, though with noticeable variance at higher consumption levels.

8. Recommendations

Based on the findings:

Enhance Predictive Features:

Incorporate additional variables (e.g., weather patterns, time of day, regional demand) to strengthen model accuracy.

Address Data Outliers:

Review and handle extreme consumption spikes, which may distort model performance.

Explore Non-Linear Models:

Test advanced algorithms (e.g., Random Forest, Gradient Boosting) to capture complex, non-linear relationships.

Feature Scaling:

Normalize numeric inputs to reduce bias from variables with different scales.

10. Conclusion

The regression analysis successfully demonstrates that linear models can moderately predict power consumption trends using numeric inputs.

While the baseline model performs reasonably well (R^2 = 0.639), additional feature engineering, data scaling, and non-linear models are recommended to improve precision.

This report provides a foundational analytical framework for future forecasting, operational optimization, and energy management initiatives.

11. Summary Table

Section	Key Point
Objective	Predict power consumption
Algorithm	Linear Regression
Data Split	80/20
R^2 Score	0.639
MAE	3,086.29
Recommendation	Add features, scale data, test non-linear models