Project: Measure Energy Consumption

# Phase 2: Transformation and Implementation Introduction to Energy Consumption Measurement and Prediction:

Energy consumption measurement and prediction involves several methods that is a critical aspect of modern resource management and sustainability efforts.The ability to accurately measure and forecast energy consumption patterns is essential for various sectors, including residential, commercial, and industrial, as it enables organizations and individuals to make informed decisions about energy use and optimize their operations. This approaches enable accurate tracking and forecasting of energy usage in various settings.This project aims to leverage innovative techniques to improve the precision and robustness of energy consumption measurement and prediction.

# Project Plan:

In this project, we emphasize the application of innovative techniques to enhance the accuracy and reliability of energy consumption measurement and prediction. These innovative techniques include:

**Machine Learning Models:** Traditional machine learning models, such as Random Forests, Support Vector Machines (SVMs), and Gradient Boosting, provide a strong foundation for predicting energy consumption. These models can be employed in conjunction with the other techniques to achieve robust and reliable predictions.

# Data Collection:

* Begin by downloading the dataset from the Kaggle dataset link [(https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption).](https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption) Ensure that the dataset is in a format that can be readily used for analysis.

## Data Cleaning and Missing Value Handling:

* Check the dataset for missing values and anomalies. It's crucial to address these issues to ensure the accuracy of the analysis.
* Handle missing data using appropriate techniques such as:
* Imputation: Filling missing values with a suitable estimate (e.g., mean, median, forward- fill, backward-fill).
* Interpolation: If the dataset has a time component, consider interpolating missing values based on surrounding time points.
* Removal: If data is severely corrupted or missing, you might need to exclude affected records.
* Outlier Handling: Identify and address any outliers that might skew the analysis.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | count | mean | std | min | 25% | 50% | 75% | max |
| AEP | 121273.0 | 15499.513  717 | 2591.3990  65 | 9581.0 | 13630.0 | 15310.0 | 17200.00 | 25695.0 |
| COMED | 66497.0 | 11420.152  112 | 2304.1395  17 | 7237.0 | 9780.0 | 11152.0 | 12510.00 | 23753.0 |
| DAYTON | 121275.0 | 2037.8511  40 | 393.40315  3 | 982.0 | 1749.0 | 2009.0 | 2279.00 | 3746.0 |
| DEOK | 57739.0 | 3105.0964  86 | 599.85902  6 | 907.0 | 2687.0 | 3013.0 | 3449.00 | 5445.0 |
| DOM | 116189.0 | 10949.203  625 | 2413.9465  69 | 1253.0 | 9322.0 | 10501.0 | 12378.00 | 21651.0 |
| DUQ | 119068.0 | 1658.8202  96 | 301.74064  0 | 1014.0 | 1444.0 | 1630.0 | 1819.00 | 3054.0 |
| EKPC | 45334.0 | 1464.2184  23 | 378.86840  4 | 514.0 | 1185.0 | 1386.0 | 1699.00 | 3490.0 |
| FE | 62874.0 | 7792.1590  64 | 1331.2680  06 | 0.0 | 6807.0 | 7700.0 | 8556.00 | 14032.0 |
| NI | 58450.0 | 11701.682  943 | 2371.4987  01 | 7003.0 | 9954.0 | 11521.0 | 12896.75 | 23631.0 |
| PJME | 145366.0 | 32080.222  831 | 6464.0121  66 | 14544.0 | 27573.0 | 31421.0 | 35650.00 | 62009.0 |
| PJMW | 143206.0 | 5602.3750  89 | 979.14287  2 | 487.0 | 4907.0 | 5530.0 | 6252.00 | 9594.0 |
| PJM\_ Load | 32896.0 | 29766.427  408 | 5849.7699  54 | 17461.0 | 25473.0 | 29655.0 | 33073.25 | 54030.0 |

Model Selection:

In this phase, we need to define the models to be implemented for energy consumption measurement and prediction. We will explore a combination of ensemble methods and deep learning architectures to harness their respective strengths.

Ensemble Methods:

Ensemble methods, such as Random Forest, Gradient Boosting (e.g., XG Boost or Light GBM), and stacking, will be employed. These models excel at capturing complex relationships in the data by combining the predictions of multiple base models.

Deep Learning Architectures:

Deep learning models, particularly Long Short-Term Memory (LSTM) networks, will be used for their ability to model temporal dependencies in time series data. LSTM networks are suitable for capturing long-term dependencies and patterns in energy consumption.

Deep Learning (LSTM):

* Number of LSTM units in layers.
* Learning rate for optimization.
* Batch size for training.
* Number of training epochs.
* Dropout rates to prevent overfitting.

Model Evaluation:

To ensure the models' performance is rigorously assessed, we will use appropriate evaluation metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE):

* Mean Absolute Error (MAE):

The MAE measures the average absolute difference between the predicted and actual values. It provides a straightforward interpretation of the model's accuracy.

### Formula: MAE = (1/n) \* Σ |predicted - actual|

* Mean Squared Error (MSE):

MSE calculates the average of the squared differences between predictions and actual values. It penalizes larger errors more than MAE.

### Formula: MSE = (1/n) \* Σ (predicted - actual) ^2

* Root Mean Squared Error (RMSE):

RMSE is the square root of MSE and provides an error measure in the same units as the target variable. It is a common metric for assessing prediction accuracy.

**Formula: RMSE = √(MSE)**

# Model Validation and Selection:

Selecting the Best-Performing Models:

* + Choose the models that achieve the lowest MAE, MSE, and RMSE values, indicating the best predictive accuracy.
  + Consider additional factors such as computational complexity, model interpretability, and the specific needs of the project when making the final selection.

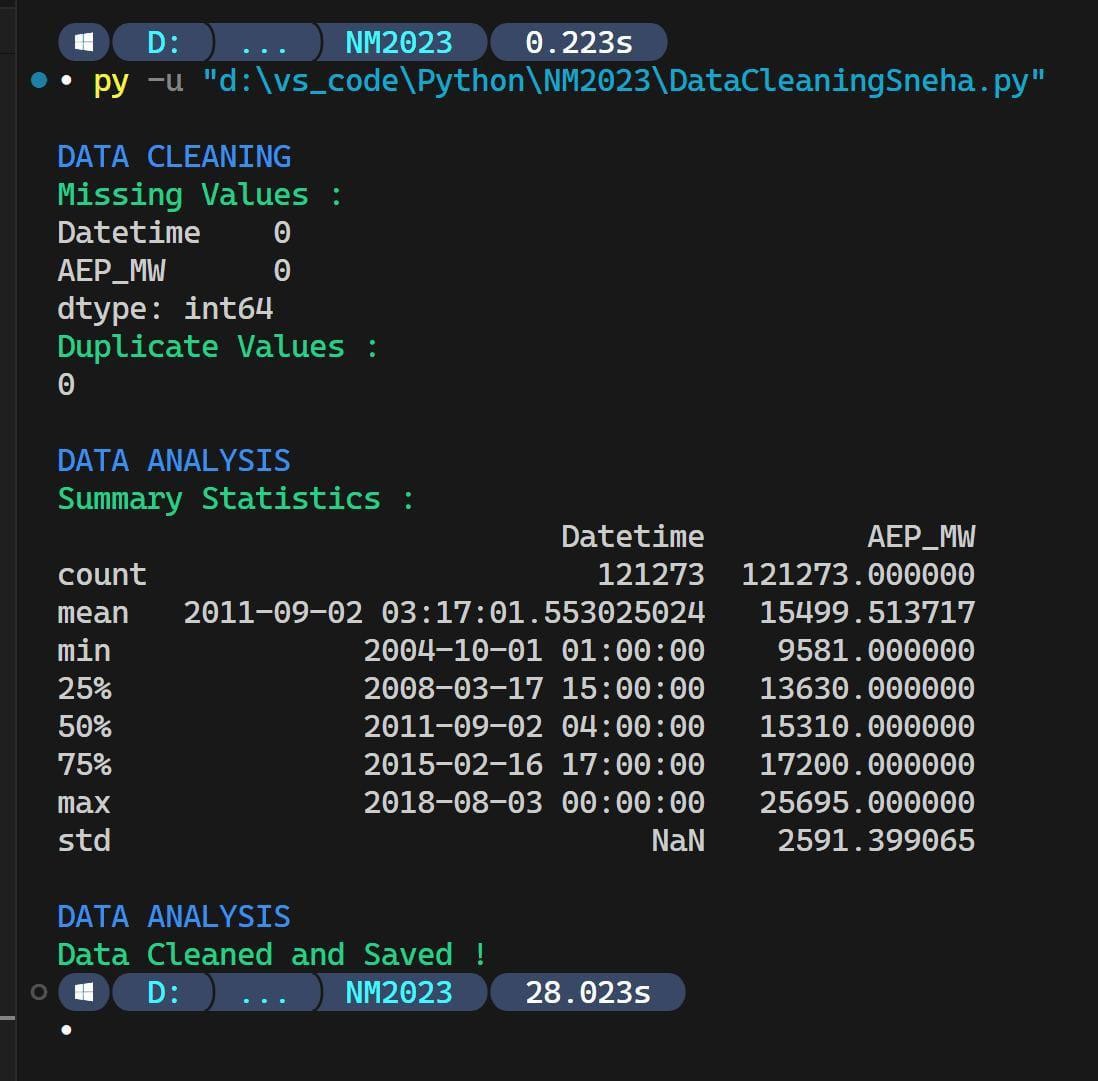
## Validation Sets:

It's crucial to perform this evaluation on validation sets, not just the training data. The validation set serves as an independent dataset that hasn't been seen by the model during training.

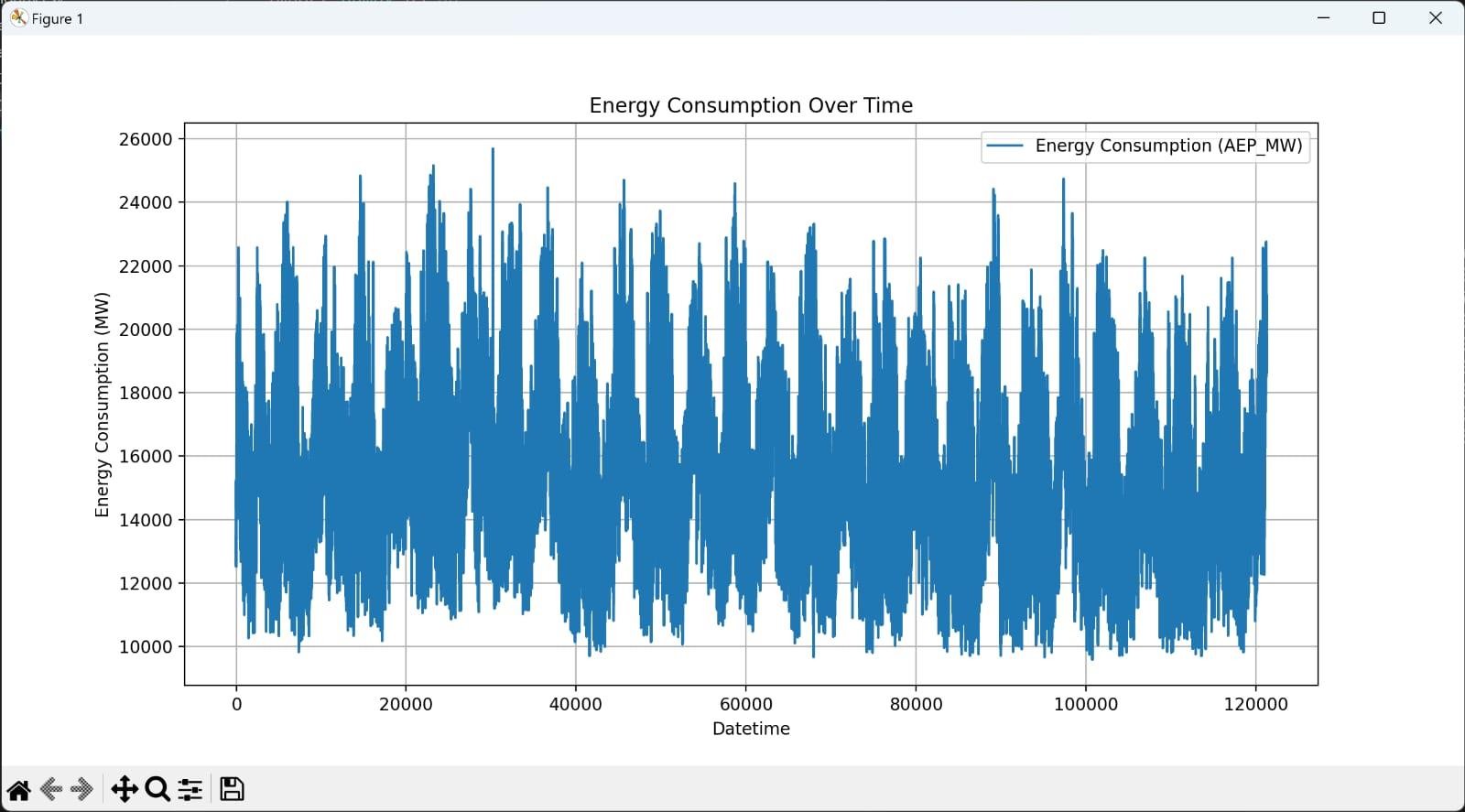
# Program:



Output:



# Graphical representation:



Conclusion & Future work:

⦁ Hourly energy consumption data analysis is crucial for understanding patterns and making informed decisions about energy management.

⦁ The conclusion is patterns can help identify areas where energy efficiency measures can be implemented, reducing energy waste during low demand periods.

⦁ The future work of this project is this project is a dynamic field with significant potential for improving energy efficiency, reducing costs and contributing to a more sustainable and reliable energy future.