**Project: Measure Energy Consumption**

# **Phase 3:** Development Part 1

## **Introduction to Energy Consumption Measurement and Prediction:**

Energy consumption measurement and prediction 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 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.

# **Program:**

LINK FOR KAGGLE NOTEBOOK DATA CLEANING AND ANALYSING:

<https://www.kaggle.com/code/jeblinaldod/ai-phase-3>

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

RED = "\033[91m"

GREEN = "\033[92m"

YELLOW = "\033[93m"

BLUE = "\033[94m"

RESET = "\033[0m"

df = pd.read\_csv("AEP\_hourly.csv")

df["Datetime"] = pd.to\_datetime(df["Datetime"])

# DATA CLEANING

print(BLUE + "\nDATA CLEANING" + RESET)

# --- Check for missing values

missing\_values = df.isnull().sum()

print(GREEN + "Missing Values : " + RESET)

print(missing\_values)

# --- Handle missing values

df.dropna(inplace=True)

# --- Check for duplicate values

duplicate\_values = df.duplicated().sum()

print(GREEN + "Duplicate Values : " + RESET)

print(duplicate\_values)

# --- Drop duplicate values

df.drop\_duplicates(inplace=True)

# DATA ANALYSIS

print (BLUE + "\nDATA ANALYSIS" + RESET)

# --- Summary Statistics

summary\_stats = df.describe()

print(GREEN + "Summary Statistics : " + RESET)

print(summary\_stats)

# Data Visualization

# Line plot for energy consumption over time

plt.figure(figsize=(12, 6))

plt.plot(df.index, df["AEP\_MW"], label="Energy Consumption (AEP\_MW)")

plt.xlabel("Datetime")

plt.ylabel("Energy Consumption (MW)")

plt.title("Energy Consumption Over Time")

plt.grid()

plt.legend()

plt.show()

# SAVING THE FILE

df.to\_csv("cleaned\_AEP\_hourly.csv", index=False)

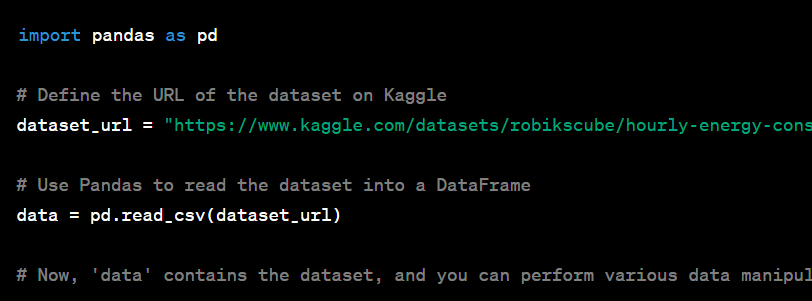
print(BLUE + "\nDATA ANALYSIS" + RESET)

print(GREEN + "Data Cleaned and Saved !" + RESET)

# **Output:**

DATA CLEANING  
Missing Values :   
Datetime 0  
AEP\_MW 0  
dtype: int64  
Duplicate Values :   
0  
  
DATA ANALYSIS  
Summary Statistics :   
 Datetime AEP\_MW  
count 121273 121273.000000  
mean 2011-09-02 03:17:01.553025024 15499.513717  
min 2004-10-01 01:00:00 9581.000000  
25% 2008-03-17 15:00:00 13630.000000  
50% 2011-09-02 04:00:00 15310.000000  
75% 2015-02-16 17:00:00 17200.000000  
max 2018-08-03 00:00:00 25695.000000  
std NaN 2591.399065

## **Data Loading:**

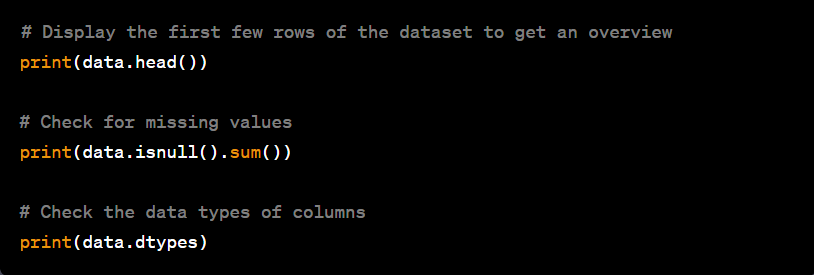


In this example, we're using the pd. read\_ csv function to read a CSV file from the specified Kaggle URL. This function loads the data into a Pandas Data Frame, which is a versatile data structure for data analysis in Python.

## **Data Exploration:**

Start with basic visualizations to understand the dataset:

* + Histograms and density plots to analyze data distributions.
  + Scatter plots to explore relationships between variables.
  + Time series plots to visualize energy consumption trends over time.

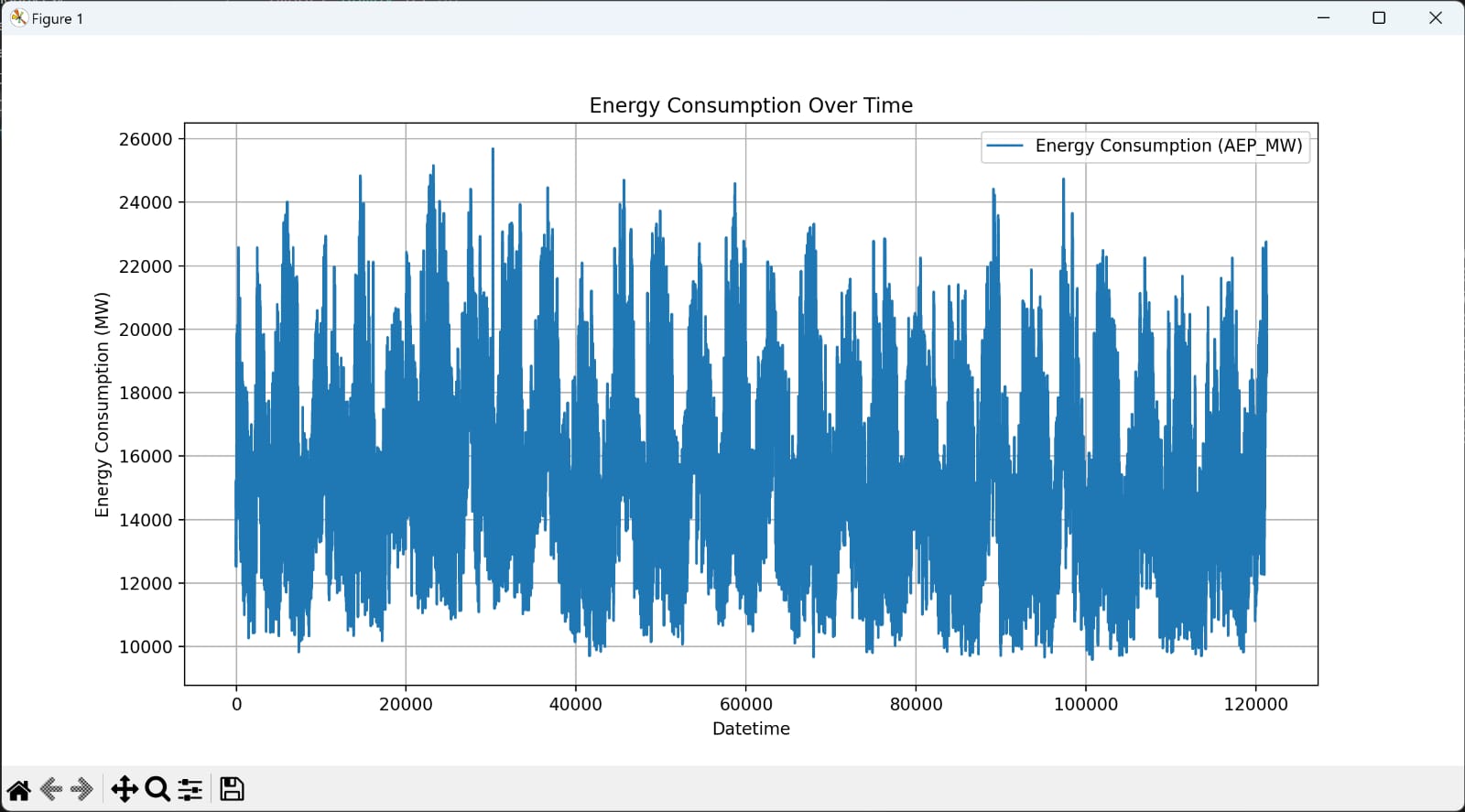


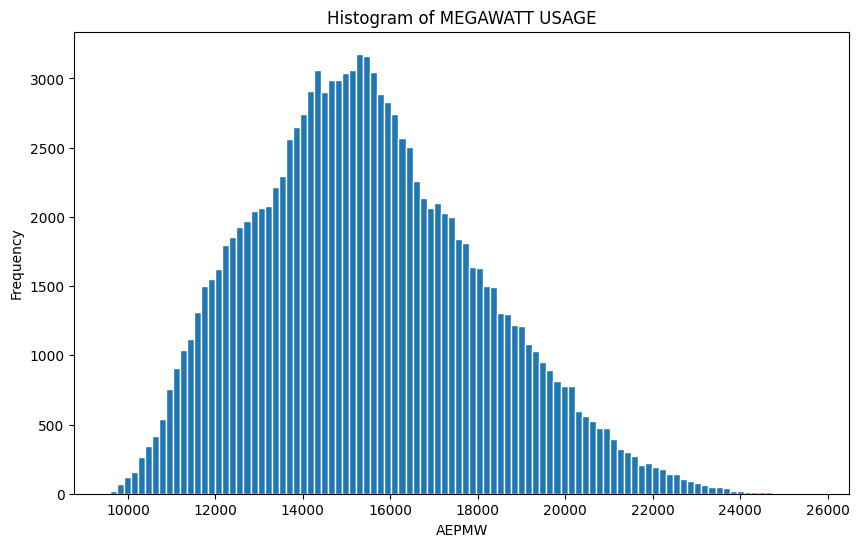
**Visualization:**

Visualization is a powerful tool in data analysis and machine learning project development. It helps in understanding data, identifying patterns, and communicating insights effectively. Here's a general guide on how you can incorporate visualization into the development of energy consumption prediction project:

**Link for the visualization of data:**

<https://www.kaggle.com/code/jeblinaldod/ai-phase-3>





##### **Conclusion & Future work:**

##### The development of measuring energy consumption begins with the data vleaning and analysis.

##### Data cleaning encompasses the treatment of missing values and the removal of duplicate entries, ensuring a complete and unique dataset.

##### The subsequent data analysis unveils valuable insights by computing summary statistics and examining class distribution.

##### The data visualization gives a visual output in terms of plots which are very human understandable and comparable with other data.