

# MEASURE FOR ENERGY CONSUMPTION

## PHASE3:DEVELOPMENT PART 1

### INPUT :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings

warnings.filterwarnings("ignore", category=UserWarning)

RED = "\033[91m"
GREEN = "\033[92m"
YELLOW = "\033[93m"
BLUE = "\033[94m"
RESET = "\033[0m"

df = pd.read_csv("/kaggle/input/hourly-energy-consumption/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
print(BLUE + "\nDATA VISUALIZATION" + RESET)
# --- Line plot
print(GREEN + "LinePlot : " + RESET)
plt.figure(figsize=(10, 6))
sns.lineplot(data=df, x="Datetime", y="AEP_MW")
plt.xlabel("Datetime")
plt.ylabel("Energy Consumption (MW)")
plt.title("Energy Consumption Over Year")
plt.grid()
plt.show()
# --- Histogram
print(GREEN + "Histogram : " + RESET)
plt.figure(figsize=(10, 6))
plt.hist(
    df["AEP_MW"],
    bins=100,
    histtype="barstacked",
    edgecolor="white",
)
plt.xlabel("AEP_MW")
plt.ylabel("Frequency")
plt.title("Histogram of MEGAWATT USAGE")
plt.show()

```

## RESULT :

### DATA CLEANING

Missing Values :

Datetime 0

AEP\_MW 0

dtype: int64

Duplicate Values :

0

### DATA ANALYSIS

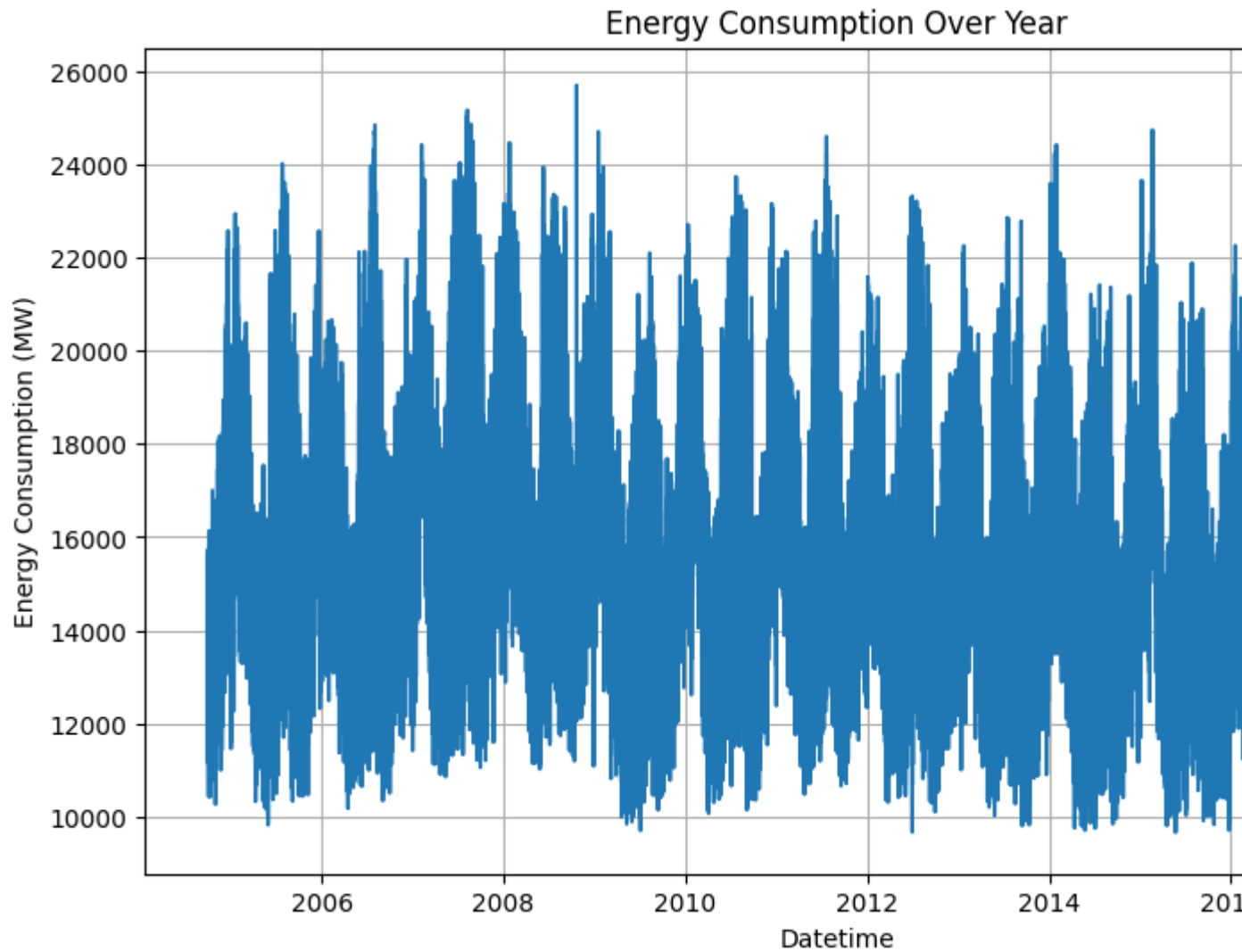
Summary Statistics :

	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 VISUALIZATION

LinePlot :



Histogram :

Histogram of MEGAWATT USAGE

