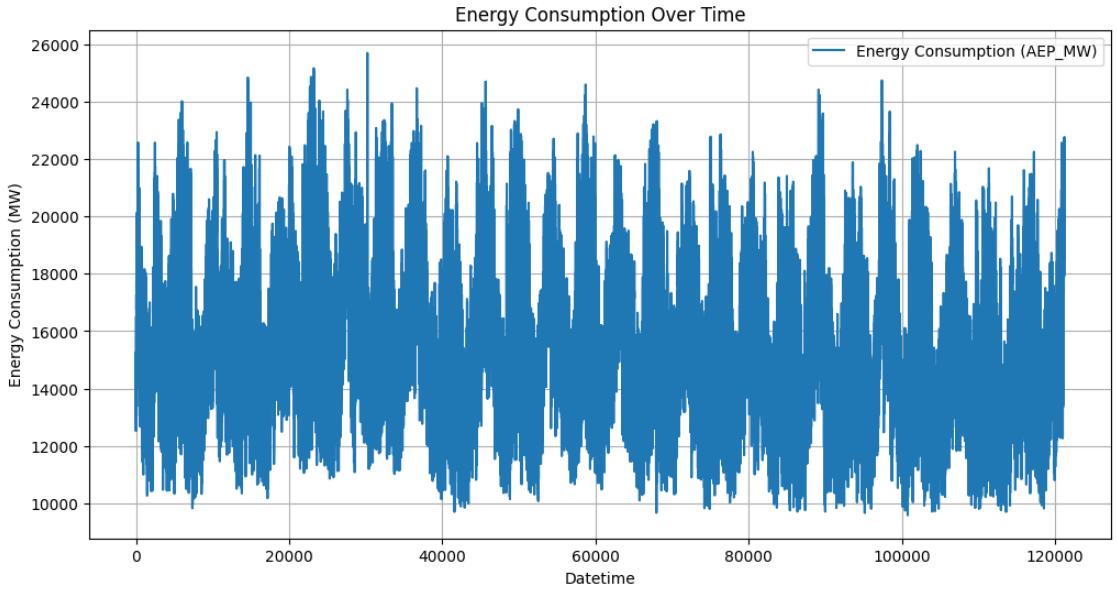
[7]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

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

# Load the dataset
df = pd.read_csv("/kaggle/input/hourly-energy-consumption/AEP_hourly.csv")

# Convert the "Datetime" column to datetime data type
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")
```

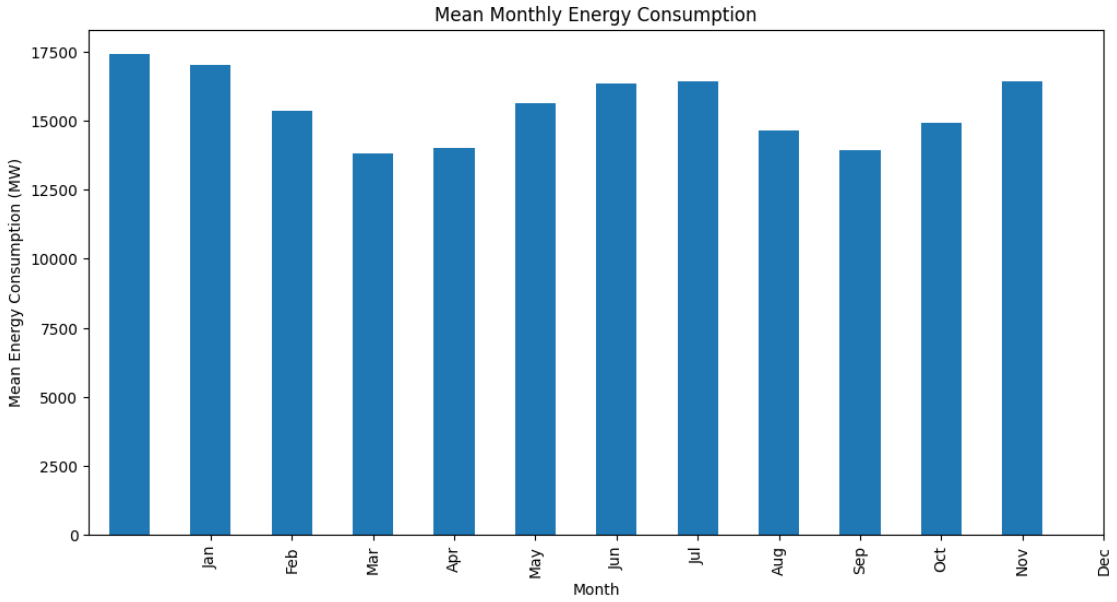
```
plt.ylabel("Energy Consumption (MW)")
plt.title("Energy Consumption Over Time")
plt.grid()
plt.legend()

# Group data by month and calculate the monthly mean energy consumption
monthly_energy_consumption = df.groupby(df["Datetime"].dt.month)["AEP_MW"].
 ↪mean()

# Data Visualization: Bar plot for monthly energy consumption
plt.figure(figsize=(12, 6))
monthly_energy_consumption.plot(kind="bar")
plt.xlabel("Month")
plt.ylabel("Mean Energy Consumption (MW)")
plt.title("Mean Monthly Energy Consumption")

# Rename the x-axis labels to display month names
plt.xticks(range(1, 13), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul',␣
 ↪'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])

plt.show()

# SAVING THE FILE
df.to_csv("/kaggle/working/cleaned_AEP_hourly.csv", index=False)
print(BLUE + "\nDATA ANALYSIS" + RESET)
print(GREEN + "Data Cleaned and Saved !" + RESET)
```

```
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 ANALYSIS

Data Cleaned and Saved !