TIMES SERIES DATASET STEEL INDUSTRY DATASET

In [1]:
```
import matplotlib.pyplot as plt
import pandas as pd


df = pd.read_csv(r'D:\Steel_industry_data.csv')

print(df.head())
```

```
               date  Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
0  01/01/2018 00:15       3.17                                  2.95   
1  01/01/2018 00:30       4.00                                  4.46   
2  01/01/2018 00:45       3.24                                  3.28   
3  01/01/2018 01:00       3.31                                  3.56   
4  01/01/2018 01:15       3.82                                  4.50   

   Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
0                                   0.0        0.0   
1                                   0.0        0.0   
2                                   0.0        0.0   
3                                   0.0        0.0   
4                                   0.0        0.0   

   Lagging_Current_Power_Factor  Leading_Current_Power_Factor   NSM  \
0                         73.21                         100.0   900   
1                         66.77                         100.0  1800   
2                         70.28                         100.0  2700   
3                         68.09                         100.0  3600   
4                         64.72                         100.0  4500   

  WeekStatus Day_of_week   Load_Type  
0    Weekday      Monday  Light_Load  
1    Weekday      Monday  Light_Load  
2    Weekday      Monday  Light_Load  
3    Weekday      Monday  Light_Load  
4    Weekday      Monday  Light_Load  
```

In [2]:
```
import numpy as np
import pandas as pd

sample_df = df.sample(frac=0.1, random_state=1)

print(sample_df.head())
```

```
                   date  Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
5760   02/03/2018 00:15       3.20                                  2.59   
14294  29/05/2018 21:45       3.78                                  0.00   
35035  31/12/2018 23:00       3.85                                  4.86   
30292  12/11/2018 13:15      43.81                                 21.31   
31651  26/11/2018 17:00      65.34                                 28.73   

       Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
5760                                   0.00       0.00   
14294                                 19.76       0.00   
35035                                  0.00       0.00   
30292                                  4.86       0.02   
31651                                  0.00       0.03   

       Lagging_Current_Power_Factor  Leading_Current_Power_Factor    NSM  \
5760                          77.73                        100.00    900   
14294                        100.00                         18.79  78300   
35035                         62.10                        100.00  82800   
30292                         89.93                         99.39  47700   
31651                         91.54                        100.00  61200   

      WeekStatus Day_of_week    Load_Type  
5760     Weekday      Friday   Light_Load  
14294    Weekday     Tuesday  Medium_Load  
35035    Weekday      Monday   Light_Load  
30292    Weekday      Monday  Medium_Load  
31651    Weekday      Monday  Medium_Load  
```

In [3]:

```
df['New_Column'] = df['Usage_kWh'] * 0.5

df.drop('New_Column', axis=1, inplace=True)

print(df.head())
```

```
               date  Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
0  01/01/2018 00:15       3.17                                  2.95
1  01/01/2018 00:30       4.00                                  4.46
2  01/01/2018 00:45       3.24                                  3.28
3  01/01/2018 01:00       3.31                                  3.56
4  01/01/2018 01:15       3.82                                  4.50

   Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
0                                   0.0        0.0
1                                   0.0        0.0
2                                   0.0        0.0
3                                   0.0        0.0
4                                   0.0        0.0

   Lagging_Current_Power_Factor  Leading_Current_Power_Factor   NSM  \
0                         73.21                         100.0   900
1                         66.77                         100.0  1800
2                         70.28                         100.0  2700
3                         68.09                         100.0  3600
4                         64.72                         100.0  4500

  WeekStatus Day_of_week   Load_Type
0    Weekday      Monday  Light_Load
1    Weekday      Monday  Light_Load
2    Weekday      Monday  Light_Load
3    Weekday      Monday  Light_Load
4    Weekday      Monday  Light_Load
```

In [4]:

```
filtered_df = df[df['Usage_kWh'] > 50]

selected_columns_df = df[['date', 'Usage_kWh']]

print(filtered_df.head())
print(selected_columns_df.head())
```

```
                 date  Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
130  02/01/2018 08:45      52.06                                 35.32   
131  02/01/2018 09:00      56.20                                 12.53   
132  02/01/2018 09:15      56.84                                  8.32   
133  02/01/2018 09:30      51.26                                  4.54   
135  02/01/2018 10:00      52.81                                  7.06   

     Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
130                                  0.00        0.0   
131                                  0.11        0.0   
132                                  0.00        0.0   
133                                  0.94        0.0   
135                                  0.54        0.0   

     Lagging_Current_Power_Factor  Leading_Current_Power_Factor    NSM  \
130                         82.75                        100.00  31500   
131                         97.60                        100.00  32400   
132                         98.95                        100.00  33300   
133                         99.61                         99.98  34200   
135                         99.12                         99.99  36000   

    WeekStatus Day_of_week    Load_Type  
130    Weekday     Tuesday   Light_Load  
131    Weekday     Tuesday   Light_Load  
132    Weekday     Tuesday  Medium_Load  
133    Weekday     Tuesday  Medium_Load  
135    Weekday     Tuesday  Medium_Load  
               date  Usage_kWh
0  01/01/2018 00:15       3.17
1  01/01/2018 00:30       4.00
2  01/01/2018 00:45       3.24
3  01/01/2018 01:00       3.31
4  01/01/2018 01:15       3.82
```

In [5]:

```python
df['Usage_kWh'] = df['Usage_kWh'] + 10

comparison_result = df['Usage_kWh'] > 100

df['High_Usage'] = comparison_result

membership_result = df['Day_of_week'].isin(['Monday', 'Tuesday'])

print(df.head())
print(comparison_result.head())
print(membership_result.head())
```

```
               date  Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
0  01/01/2018 00:15      13.17                                  2.95
1  01/01/2018 00:30      14.00                                  4.46
2  01/01/2018 00:45      13.24                                  3.28
3  01/01/2018 01:00      13.31                                  3.56
4  01/01/2018 01:15      13.82                                  4.50

   Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
0                                   0.0        0.0
1                                   0.0        0.0
2                                   0.0        0.0
3                                   0.0        0.0
4                                   0.0        0.0

   Lagging_Current_Power_Factor  Leading_Current_Power_Factor   NSM  \
0                         73.21                         100.0   900
1                         66.77                         100.0  1800
2                         70.28                         100.0  2700
3                         68.09                         100.0  3600
4                         64.72                         100.0  4500

  WeekStatus Day_of_week   Load_Type  High_Usage
0    Weekday      Monday  Light_Load       False
1    Weekday      Monday  Light_Load       False
2    Weekday      Monday  Light_Load       False
3    Weekday      Monday  Light_Load       False
4    Weekday      Monday  Light_Load       False
0    False
1    False
2    False
3    False
4    False
Name: Usage_kWh, dtype: bool
0    True
1    True
2    True
3    True
4    True
Name: Day_of_week, dtype: bool
```

In [6]:
```
mean_usage = df['Usage_kWh'].mean()
sum_usage = df['Usage_kWh'].sum()
variance_usage = df['Usage_kWh'].var()
correlation = df['Usage_kWh'].corr(df['CO2(tCO2)'])

print("Mean Usage:", mean_usage)
print("Sum of Usage:", sum_usage)
print("Variance of Usage:", variance_usage)
print("Correlation between Usage and CO2:", correlation)
```

```
Mean Usage: 37.3868924086758
Sum of Usage: 1310036.71
Variance of Usage: 1118.5265340538886
Correlation between Usage and CO2: 0.9881797716789615
```

In [7]:

```
df.fillna(df.mean(), inplace=True)

df.dropna(inplace=True)

print(df.head())
```

```
               date  Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
0  01/01/2018 00:15      13.17                                  2.95   
1  01/01/2018 00:30      14.00                                  4.46   
2  01/01/2018 00:45      13.24                                  3.28   
3  01/01/2018 01:00      13.31                                  3.56   
4  01/01/2018 01:15      13.82                                  4.50   

   Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
0                                   0.0        0.0   
1                                   0.0        0.0   
2                                   0.0        0.0   
3                                   0.0        0.0   
4                                   0.0        0.0   

   Lagging_Current_Power_Factor  Leading_Current_Power_Factor   NSM  \
0                         73.21                         100.0   900   
1                         66.77                         100.0  1800   
2                         70.28                         100.0  2700   
3                         68.09                         100.0  3600   
4                         64.72                         100.0  4500   

  WeekStatus Day_of_week   Load_Type  High_Usage
0    Weekday      Monday  Light_Load       False
1    Weekday      Monday  Light_Load       False
2    Weekday      Monday  Light_Load       False
3    Weekday      Monday  Light_Load       False
4    Weekday      Monday  Light_Load       False
C:\Users\Administrator\AppData\Local\Temp\ipykernel_4064\1225254314.py:2: FutureWar
ning: The default value of numeric_only in DataFrame.mean is deprecated. In a futur
e version, it will default to False. In addition, specifying 'numeric_only=None' is
deprecated. Select only valid columns or specify the value of numeric_only to silen
ce this warning.
  df.fillna(df.mean(), inplace=True)
```

In [ ]:

In [9]:

```
df['New_Feature'] = df['Usage_kWh'].apply(lambda x: x * 2)

grouped_df = df.groupby('Day_of_week').mean()

print(df.head())
print(grouped_df.head())
```

```
               date  Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
0  01/01/2018 00:15      13.17                                  2.95   
1  01/01/2018 00:30      14.00                                  4.46   
2  01/01/2018 00:45      13.24                                  3.28   
3  01/01/2018 01:00      13.31                                  3.56   
4  01/01/2018 01:15      13.82                                  4.50   

   Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
0                                   0.0        0.0   
1                                   0.0        0.0   
2                                   0.0        0.0   
3                                   0.0        0.0   
4                                   0.0        0.0   

   Lagging_Current_Power_Factor  Leading_Current_Power_Factor   NSM  \
0                         73.21                         100.0   900   
1                         66.77                         100.0  1800   
2                         70.28                         100.0  2700   
3                         68.09                         100.0  3600   
4                         64.72                         100.0  4500   

  WeekStatus Day_of_week   Load_Type  High_Usage  New_Feature  
0    Weekday      Monday  Light_Load       False        26.34  
1    Weekday      Monday  Light_Load       False        28.00  
2    Weekday      Monday  Light_Load       False        26.48  
3    Weekday      Monday  Light_Load       False        26.62  
4    Weekday      Monday  Light_Load       False        27.64  
             Usage_kWh  Lagging_Current_Reactive.Power_kVarh  \
Day_of_week                                                    
Friday       43.195014                             16.103950   
Monday       43.143935                             16.106470   
Saturday     25.919020                              6.309886   
Sunday       17.545633                              3.235633   
Thursday     45.112083                             17.356707   

             Leading_Current_Reactive_Power_kVarh  CO2(tCO2)  \
Day_of_week                                                    
Friday                                   2.618966   0.014339   
Monday                                   2.541812   0.014324   
Saturday                                 6.208910   0.006140   
Sunday                                   7.659093   0.002045   
Thursday                                 2.367344   0.015294   

             Lagging_Current_Power_Factor  Leading_Current_Power_Factor  \
Day_of_week                                                               
Friday                          79.848419                     90.817939   
Monday                          79.618194                     90.648001   
Saturday                        82.226583                     74.348349   
Sunday                          82.171675                     64.022626   
Thursday                        79.561917                     91.823678   

                 NSM  High_Usage  New_Feature  
Day_of_week                                    
Friday       42750.0    0.085938    86.390028  
Monday       42750.0    0.085299    86.287869  
Saturday     42750.0    0.023638    51.838041  
Sunday       42750.0    0.008814    35.091266  
Thursday     42750.0    0.100761    90.224167  
```

```
C:\Users\Administrator\AppData\Local\Temp\ipykernel_4064\429109742.py:5: FutureWarn
ing: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In a
future version, numeric_only will default to False. Either specify numeric_only or
select only columns which should be valid for the function.
  grouped_df = df.groupby('Day_of_week').mean()
```

In [ ]: