Impvrting Pandas, Matplvtlib, Numpy Libraries

Dataset is lvaded by linking via Gvvgle Drive and check fvr Missing Values

1. Uplvad the Dataset tv Gvvgle Drive
2. Mvunt the Drive and Read Dataset using Pandas

In [3]:
```
import pandas as pd
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
```

In [4]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

In [6]:
```
df =pd.read_csv("drive/My Drive/IBM_Project/Dataset/Electricity.csv")
missing_values = df.isnull()
missing_values
```

```
<ipython-input-6-4383926d33ab>:1: DtypeWarning: Columns (9,10,11,14,15,16,17) have mixed types. Specify dtype option on import or
set low_memory=False.
  df =pd.read_csv("drive/My Drive/IBM_Project/Dataset/Electricity.csv")
```

Out [6]:

	DateTime	Hvliday	HvlidayFlag	DayOfWeek	WeekOfYear	Day	Mvnth	Year	PerivdOfDay	FvrecastWindPrvductivn	SystemLvadEA	SMPEA	ORI
0	False	False	False	False	False	False	False	False	False	False	False	False	Fals
1	False	False	False	False	False	False	False	False	False	False	False	False	Fals
2	False	False	False	False	False	False	False	False	False	False	False	False	Fals
3	False	False	False	False	False	False	False	False	False	False	False	False	Fals
4	False	False	False	False	False	False	False	False	False	False	False	False	Fals
...	...	...	...	...	...	...	...	...	...	...	...	...	...
38009	False	False	False	False	False	False	False	False	False	False	False	False	Fals
38010	False	False	False	False	False	False	False	False	False	False	False	False	Fals
38011	False	False	False	False	False	False	False	False	False	False	False	False	Fals
38012	False	False	False	False	False	False	False	False	False	False	False	False	Fals
38013	False	False	False	False	False	False	False	False	False	False	False	False	Fals

38014 rvws × 18 cvlumns

In [7]:
```
for column in df.columns:
    if df[column].dtype == 'object' and df[column].str.contains('\?').any():
        print(f"Column '{column}' contains '?'")
```

```
Column 'ForecastWindProduction' contains '?'
Column 'SystemLoadEA' contains '?'
Column 'SMPEA' contains '?'
Column 'ORKTemperature' contains '?'
Column 'ORKWindspeed' contains '?'
Column 'CO2Intensity' contains '?'
Column 'ActualWindProduction' contains '?'
Column 'SystemLoadEP2' contains '?'
Column 'SMPEP2' contains '?'
```

Replace the Missing Values using NaN values by Pandas library

In [8]:
```
df.replace('?', np.nan, inplace=True)
df
```

Out [8]:

	DateTime	Hvliday	HvlidayFlag	DayOfWeek	WeekOfYear	Day	Mvnth	Year	PerivdOfDay	FvrecastWindPrvductivn	SystemLvadEA	SMPEA	OR
0	01/11/2011 00:00	Nvne	0	1	44	1	11	2011	0	315.31	3388.77	49.26	6.0
1	01/11/2011 00:30	Nvne	0	1	44	1	11	2011	1	321.80	3196.66	49.26	6.0
2	01/11/2011 01:00	Nvne	0	1	44	1	11	2011	2	328.57	3060.71	49.10	5.0

	DateTime	Holiday	HolidayFlag	DayOfWeek	WeekOfYear	Day	Month	Year	PeriodOfDay	ForecastWindProduction	SystemLoadEA	SMPEA	OR
3	01/11/2011 01:30	None	0	1	44	1	11	2011	3	335.60	2945.56	48.04	6.0
4	01/11/2011 02:00	None	0	1	44	1	11	2011	4	342.90	2849.34	33.75	6.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
38009	31/12/2013 21:30	New Year's Eve	1	1	1	31	12	2013	43	1179.14	3932.22	34.51	6.0
38010	31/12/2013 22:00	New Year's Eve	1	1	1	31	12	2013	44	1152.01	3821.44	33.83	5.0
38011	31/12/2013 22:30	New Year's Eve	1	1	1	31	12	2013	45	1123.67	3724.21	31.75	4.0
38012	31/12/2013 23:00	New Year's Eve	1	1	1	31	12	2013	46	1094.24	3638.16	33.83	5.0
38013	31/12/2013 23:30	New Year's Eve	1	1	1	31	12	2013	47	1064.0	3624.25	33.83	5.0

38014 rows × 18 columns

Convert the Datatype of the columns in the Dataset as per their Requirements

In [9]:
```
df["DateTime"] = df['DateTime'].astype('datetime64')
df["ForecastWindProduction"] = df['ForecastWindProduction'].astype('float64')
df["SystemLoadEA"] = df['SystemLoadEA'].astype('float64')
df["SMPEA"] = df['SMPEA'].astype('float64')
df["ORKTemperature"] = df['ORKTemperature'].astype('float64')
df["ORKWindspeed"] = df['ORKWindspeed'].astype('float64')
df["CO2Intensity"] = df['CO2Intensity'].astype('float64')
df["ActualWindProduction"] = df['ActualWindProduction'].astype('float64')
df["SystemLoadEP2"] = df['SystemLoadEP2'].astype('float64')
df["SMPEP2"] = df['SMPEP2'].astype('float64')
df.dtypes
```

Out [9]:
```
DateTime                 datetime64[ns]
Holiday                          object
HolidayFlag                       int64
DayOfWeek                         int64
WeekOfYear                        int64
Day                               int64
Month                             int64
Year                              int64
PeriodOfDay                       int64
ForecastWindProduction          float64
SystemLoadEA                    float64
SMPEA                           float64
ORKTemperature                  float64
ORKWindspeed                    float64
CO2Intensity                    float64
ActualWindProduction            float64
SystemLoadEP2                   float64
SMPEP2                          float64
dtype: object
```

In [10]:
```
print ("\nMissing values :  ", df.isnull().any())
```

```
Missing values :   DateTime                False
Holiday                 False
HolidayFlag             False
DayOfWeek               False
WeekOfYear              False
Day                     False
Month                   False
Year                    False
PeriodOfDay             False
ForecastWindProduction   True
SystemLoadEA             True
SMPEA                    True
ORKTemperature           True
ORKWindspeed             True
CO2Intensity             True
ActualWindProduction     True
SystemLoadEP2            True
SMPEP2                   True
dtype: bool
```

Handle Missing Values using ffill method to replace NaN Values

In [11]:
```
df['ForecastWindProduction']=df['ForecastWindProduction'].fillna(method='ffill')
df['SystemLoadEA']=df['SystemLoadEA'].fillna(method='ffill')
df['SMPEA']=df['SMPEA'].fillna(method='ffill')
df['ORKTemperature']=df['ORKTemperature'].fillna(method='ffill')
df['ORKWindspeed']=df['ORKWindspeed'].fillna(method='ffill')
df['CO2Intensity']=df['CO2Intensity'].fillna(method='ffill')
df['ActualWindProduction']=df['ActualWindProduction'].fillna(method='ffill')
df['SystemLoadEP2']=df['SystemLoadEP2'].fillna(method='ffill')
df['SMPEP2']=df['SMPEP2'].fillna(method='ffill')
```

In [12]:
```
print ("\nMissing values :  ", df.isnull().any())
```

```
Missing values :   DateTime                False
Holiday                 False
HolidayFlag             False
DayOfWeek               False
WeekOfYear              False
Day                     False
Month                   False
Year                    False
PeriodOfDay             False
ForecastWindProduction  False
SystemLoadEA            False
SMPEA                   False
ORKTemperature          False
ORKWindspeed            False
CO2Intensity            False
ActualWindProduction    False
SystemLoadEP2           False
SMPEP2                  False
dtype: bool
```

Import Plotly Library and Plot the Target Column

In [13]:
```
import plotly.express as px
```

In [17]:
```
fig = px.line(df, x='DateTime', y='SMPEP2', title='Electricity Price')
fig.update_xaxes(
    rangeslider_visible=True,
    rangeselector=dict(
        buttons=list([
            dict(step="all")
        ])
    )
)
fig.show()
```

Set DateTime column as Index and plot the Subplots

In [15]:
```
el_df=df.set_index('DateTime')
```

In [16]:
```
el_df.plot(subplots=True)
```

Out [16]:
```
array([<Axes: xlabel='DateTime'>, <Axes: xlabel='DateTime'>,
       <Axes: xlabel='DateTime'>, <Axes: xlabel='DateTime'>,
       <Axes: xlabel='DateTime'>, <Axes: xlabel='DateTime'>,
       <Axes: xlabel='DateTime'>, <Axes: xlabel='DateTime'>,
       <Axes: xlabel='DateTime'>, <Axes: xlabel='DateTime'>,
       <Axes: xlabel='DateTime'>, <Axes: xlabel='DateTime'>], dtype=object)
```