```
!apt-get install tree
```

```
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
tree is already the newest version (2.0.2-1).
0 upgraded, 0 newly installed, 0 to remove and 18 not upgraded.
```

preprocesses of dataset

Double-click (or enter) to edit

import libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_squared_error
```

load the dataset

```
df=pd.read_csv("/content/Electricity (1) (1).csv")
```

```
<ipython-input-79-686a806222b2>:1: DtypeWarning: Columns (9,10,11,14,15,16,17) have mixed types. Specify dtype option on import or set l
  df=pd.read_csv("/content/Electricity (1) (1).csv")
```

```
df
```

	DateTime	Holiday	HolidayFlag	DayOfWeek	WeekOfYear	Day	Month	Year	PeriodOfDay	ForecastWindProduction	SystemLoadEA	SMPEA
0	01/11/2011 00:00	None	0	1	44	1	11	2011	0	315.31	3388.77	49.26
1	01/11/2011 00:30	None	0	1	44	1	11	2011	1	321.80	3196.66	49.26
2	01/11/2011 01:00	None	0	1	44	1	11	2011	2	328.57	3060.71	49.10
3	01/11/2011 01:30	None	0	1	44	1	11	2011	3	335.60	2945.56	48.04
4	01/11/2011 02:00	None	0	1	44	1	11	2011	4	342.90	2849.34	33.75
...	...	...	...	...	...	...	...	...	...	...	...	...
38009	31/12/2013 21:30	New Year's Eve	1	1	1	31	12	2013	43	1179.14	3932.22	34.51
38010	31/12/2013 22:00	New Year's Eve	1	1	1	31	12	2013	44	1152.01	3821.44	33.83
38011	31/12/2013 22:30	New Year's Eve	1	1	1	31	12	2013	45	1123.67	3724.21	31.75
38012	31/12/2013 23:00	New Year's Eve	1	1	1	31	12	2013	46	1094.24	3638.16	33.83
38013	31/12/2013 23:30	New Year's Eve	1	1	1	31	12	2013	47	1064.0	3624.25	33.83

38014 rows × 18 columns

read the head

read the head

```
df.head
```

```
<bound method NDFrame.head of               DateTime         Holiday  HolidayFlag  DayOfWeek  WeekOfYear  \
0      01/11/2011 00:00            None           0          1          44
1      01/11/2011 00:30            None           0          1          44
2      01/11/2011 01:00            None           0          1          44
3      01/11/2011 01:30            None           0          1          44
4      01/11/2011 02:00            None           0          1          44
...                 ...             ...         ...        ...         ...
38009  31/12/2013 21:30  New Year's Eve           1          1           1
38010  31/12/2013 22:00  New Year's Eve           1          1           1
38011  31/12/2013 22:30  New Year's Eve           1          1           1
38012  31/12/2013 23:00  New Year's Eve           1          1           1
38013  31/12/2013 23:30  New Year's Eve           1          1           1

       Day  Month  Year  PeriodOfDay ForecastWindProduction SystemLoadEA  \
0        1     11  2011           0                315.31      3388.77
1        1     11  2011           1                321.80      3196.66
2        1     11  2011           2                328.57      3060.71
3        1     11  2011           3                335.60      2945.56
4        1     11  2011           4                342.90      2849.34
...    ...    ...   ...         ...                   ...         ...
38009   31     12  2013          43               1179.14      3932.22
38010   31     12  2013          44               1152.01      3821.44
38011   31     12  2013          45               1123.67      3724.21
38012   31     12  2013          46               1094.24      3638.16
38013   31     12  2013          47                1064.0      3624.25

       SMPEA ORKTemperature ORKWindspeed CO2Intensity ActualWindProduction  \
0      49.26           6.00         9.30       600.71               356.00
1      49.26           6.00        11.10       605.42               317.00
2      49.10           5.00        11.10       589.97               311.00
3      48.04           6.00         9.30       585.94               313.00
4      33.75           6.00        11.10       571.52               346.00
...      ...            ...         ...          ...                  ...
38009  34.51           6.00        22.20       285.31                812.0
38010  33.83           5.00        24.10       278.31                852.0
38011  31.75           4.00        20.40       280.91                962.0
38012  33.83           5.00        14.80       302.46                950.0
38013  33.83           5.00        16.70       308.01               1020.0

      SystemLoadEP2 SMPEP2
0           3159.60  54.32
1           2973.01  54.23
2           2834.00  54.23
3           2725.99  53.47
4           2655.64  39.87
...             ...    ...
38009       3692.95  42.45
38010        3571.0  33.83
38011       3460.29  31.75
38012       3563.99   50.6
38013       3517.08   34.9

[38014 rows x 18 columns]>
```

read the last 10 columns

```
df.tail
```

```
<bound method NDFrame.tail of               DateTime         Holiday  HolidayFlag  DayOfWeek  WeekOfYear  \
0      01/11/2011 00:00            None           0          1          44
1      01/11/2011 00:30            None           0          1          44
2      01/11/2011 01:00            None           0          1          44
3      01/11/2011 01:30            None           0          1          44
4      01/11/2011 02:00            None           0          1          44
...                 ...             ...         ...        ...         ...
38009  31/12/2013 21:30  New Year's Eve           1          1           1
38010  31/12/2013 22:00  New Year's Eve           1          1           1
38011  31/12/2013 22:30  New Year's Eve           1          1           1
38012  31/12/2013 23:00  New Year's Eve           1          1           1
38013  31/12/2013 23:30  New Year's Eve           1          1           1

       Day  Month  Year  PeriodOfDay ForecastWindProduction SystemLoadEA  \
0        1     11  2011           0                315.31      3388.77
1        1     11  2011           1                321.80      3196.66
2        1     11  2011           2                328.57      3060.71
3        1     11  2011           3                335.60      2945.56
4        1     11  2011           4                342.90      2849.34
...    ...    ...   ...         ...                   ...         ...
```

```
38009   31     12  2013           43               1179.14      3932.22
38010   31     12  2013           44               1152.01      3821.44
38011   31     12  2013           45               1123.67      3724.21
38012   31     12  2013           46               1094.24      3638.16
38013   31     12  2013           47                1064.0      3624.25

       SMPEA ORKTemperature ORKWindspeed CO2Intensity ActualWindProduction  \
0      49.26           6.00         9.30       600.71               356.00
1      49.26           6.00        11.10       605.42               317.00
2      49.10           5.00        11.10       589.97               311.00
3      48.04           6.00         9.30       585.94               313.00
4      33.75           6.00        11.10       571.52               346.00
...      ...            ...          ...          ...                  ...
38009  34.51           6.00        22.20       285.31                812.0
38010  33.83           5.00        24.10       278.31                852.0
38011  31.75           4.00        20.40       280.91                962.0
38012  33.83           5.00        14.80       302.46                950.0
38013  33.83           5.00        16.70       308.01               1020.0

      SystemLoadEP2 SMPEP2
0           3159.60  54.32
1           2973.01  54.23
2           2834.00  54.23
3           2725.99  53.47
4           2655.64  39.87
...             ...    ...
38009       3692.95  42.45
38010        3571.0  33.83
38011       3460.29  31.75
38012       3563.99   50.6
38013       3517.08   34.9

[38014 rows x 18 columns]>
```

```
df.shape
```

```
(38014, 18)
```

```
df.describe
```

```
<bound method NDFrame.describe of               DateTime         Holiday  HolidayFlag  DayOfWeek  WeekOfYear  \
0      01/11/2011 00:00            None           0          1          44
1      01/11/2011 00:30            None           0          1          44
2      01/11/2011 01:00            None           0          1          44
3      01/11/2011 01:30            None           0          1          44
4      01/11/2011 02:00            None           0          1          44
...                 ...             ...         ...        ...         ...
38009  31/12/2013 21:30  New Year's Eve           1          1           1
38010  31/12/2013 22:00  New Year's Eve           1          1           1
38011  31/12/2013 22:30  New Year's Eve           1          1           1
38012  31/12/2013 23:00  New Year's Eve           1          1           1
38013  31/12/2013 23:30  New Year's Eve           1          1           1

       Day  Month  Year  PeriodOfDay ForecastWindProduction SystemLoadEA  \
0        1     11  2011            0                315.31      3388.77
1        1     11  2011            1                321.80      3196.66
2        1     11  2011            2                328.57      3060.71
3        1     11  2011            3                335.60      2945.56
4        1     11  2011            4                342.90      2849.34
...    ...    ...   ...          ...                   ...         ...
38009   31     12  2013           43               1179.14      3932.22
38010   31     12  2013           44               1152.01      3821.44
38011   31     12  2013           45               1123.67      3724.21
38012   31     12  2013           46               1094.24      3638.16
38013   31     12  2013           47                1064.0      3624.25

       SMPEA ORKTemperature ORKWindspeed CO2Intensity ActualWindProduction  \
0      49.26           6.00         9.30       600.71               356.00
1      49.26           6.00        11.10       605.42               317.00
2      49.10           5.00        11.10       589.97               311.00
3      48.04           6.00         9.30       585.94               313.00
4      33.75           6.00        11.10       571.52               346.00
...      ...            ...          ...          ...                  ...
38009  34.51           6.00        22.20       285.31                812.0
38010  33.83           5.00        24.10       278.31                852.0
38011  31.75           4.00        20.40       280.91                962.0
38012  33.83           5.00        14.80       302.46                950.0
38013  33.83           5.00        16.70       308.01               1020.0

      SystemLoadEP2 SMPEP2
0           3159.60  54.32
1           2973.01  54.23
```

```
2          2834.00  54.23
3          2725.99  53.47
4          2655.64  39.87
...            ...    ...
38009      3692.95  42.45
38010       3571.0  33.83
38011      3460.29  31.75
38012      3563.99   50.6
38013      3517.08   34.9

[38014 rows x 18 columns]>
```

```
df.columns
```

```
Index(['DateTime', 'Holiday', 'HolidayFlag', 'DayOfWeek', 'WeekOfYear', 'Day',
       'Month', 'Year', 'PeriodOfDay', 'ForecastWindProduction',
       'SystemLoadEA', 'SMPEA', 'ORKTemperature', 'ORKWindspeed',
       'CO2Intensity', 'ActualWindProduction', 'SystemLoadEP2', 'SMPEP2'],
      dtype='object')
```

check the null values

```
df.isnull
```

```
<bound method DataFrame.isnull of               DateTime         Holiday  HolidayFlag  DayOfWeek  WeekOfYear  \
0      01/11/2011 00:00            None            0          1          44
1      01/11/2011 00:30            None            0          1          44
2      01/11/2011 01:00            None            0          1          44
3      01/11/2011 01:30            None            0          1          44
4      01/11/2011 02:00            None            0          1          44
...                 ...             ...          ...        ...         ...
38009  31/12/2013 21:30  New Year's Eve            1          1           1
38010  31/12/2013 22:00  New Year's Eve            1          1           1
38011  31/12/2013 22:30  New Year's Eve            1          1           1
38012  31/12/2013 23:00  New Year's Eve            1          1           1
38013  31/12/2013 23:30  New Year's Eve            1          1           1

       Day  Month  Year  PeriodOfDay ForecastWindProduction SystemLoadEA  \
0        1     11  2011            0                315.31      3388.77
1        1     11  2011            1                321.80      3196.66
2        1     11  2011            2                328.57      3060.71
3        1     11  2011            3                335.60      2945.56
4        1     11  2011            4                342.90      2849.34
...    ...    ...   ...          ...                   ...          ...
38009   31     12  2013           43               1179.14      3932.22
38010   31     12  2013           44               1152.01      3821.44
38011   31     12  2013           45               1123.67      3724.21
38012   31     12  2013           46               1094.24      3638.16
38013   31     12  2013           47                1064.0      3624.25

       SMPEA ORKTemperature ORKWindspeed CO2Intensity ActualWindProduction  \
0      49.26           6.00         9.30       600.71               356.00
1      49.26           6.00        11.10       605.42               317.00
2      49.10           5.00        11.10       589.97               311.00
3      48.04           6.00         9.30       585.94               313.00
4      33.75           6.00        11.10       571.52               346.00
...      ...            ...          ...          ...                  ...
38009  34.51           6.00        22.20       285.31                812.0
38010  33.83           5.00        24.10       278.31                852.0
38011  31.75           4.00        20.40       280.91                962.0
38012  33.83           5.00        14.80       302.46                950.0
38013  33.83           5.00        16.70       308.01               1020.0

      SystemLoadEP2 SMPEP2
0           3159.60  54.32
1           2973.01  54.23
2           2834.00  54.23
3           2725.99  53.47
4           2655.64  39.87
...             ...    ...
38009       3692.95  42.45
38010        3571.0  33.83
38011       3460.29  31.75
38012       3563.99   50.6
38013       3517.08   34.9

[38014 rows x 18 columns]>
```

```
df.isnull().sum
```

```
<bound method NDFrame._add_numeric_operations.<locals>.sum of       DateTime  Holiday  HolidayFlag  DayOfWeek  WeekOfYear    Day
Month  \
0         False    False        False      False       False  False  False
1         False    False        False      False       False  False  False
2         False    False        False      False       False  False  False
3         False    False        False      False       False  False  False
4         False    False        False      False       False  False  False
...         ...      ...          ...        ...         ...    ...    ...
38009     False    False        False      False       False  False  False
38010     False    False        False      False       False  False  False
38011     False    False        False      False       False  False  False
38012     False    False        False      False       False  False  False
38013     False    False        False      False       False  False  False

        Year  PeriodOfDay  ForecastWindProduction  SystemLoadEA  SMPEA  \
0      False        False                   False         False  False
1      False        False                   False         False  False
2      False        False                   False         False  False
3      False        False                   False         False  False
4      False        False                   False         False  False
...      ...          ...                     ...           ...    ...
38009  False        False                   False         False  False
38010  False        False                   False         False  False
38011  False        False                   False         False  False
38012  False        False                   False         False  False
38013  False        False                   False         False  False

       ORKTemperature  ORKWindspeed  CO2Intensity  ActualWindProduction  \
0               False         False         False                 False
1               False         False         False                 False
2               False         False         False                 False
3               False         False         False                 False
4               False         False         False                 False
...               ...           ...           ...                   ...
38009           False         False         False                 False
38010           False         False         False                 False
38011           False         False         False                 False
38012           False         False         False                 False
38013           False         False         False                 False

       SystemLoadEP2  SMPEP2
0              False   False
1              False   False
2              False   False
3              False   False
4              False   False
...              ...     ...
38009          False   False
38010          False   False
38011          False   False
38012          False   False
38013          False   False

[38014 rows x 18 columns]>
```

```
df.dropna()
```

	WeekOfYear	Day	Month	Year	PeriodOfDay	ForecastWindProduction	SystemLoadEA	SMPEA
	44	1	11	2011	0	315.31	3388.77	49.26
	44	1	11	2011	1	321.80	3196.66	49.26
	44	1	11	2011	2	328.57	3060.71	49.10
	44	1	11	2011	3	335.60	2945.56	48.04
	44	1	11	2011	4	342.90	2849.34	33.75

```
df.fillna
```

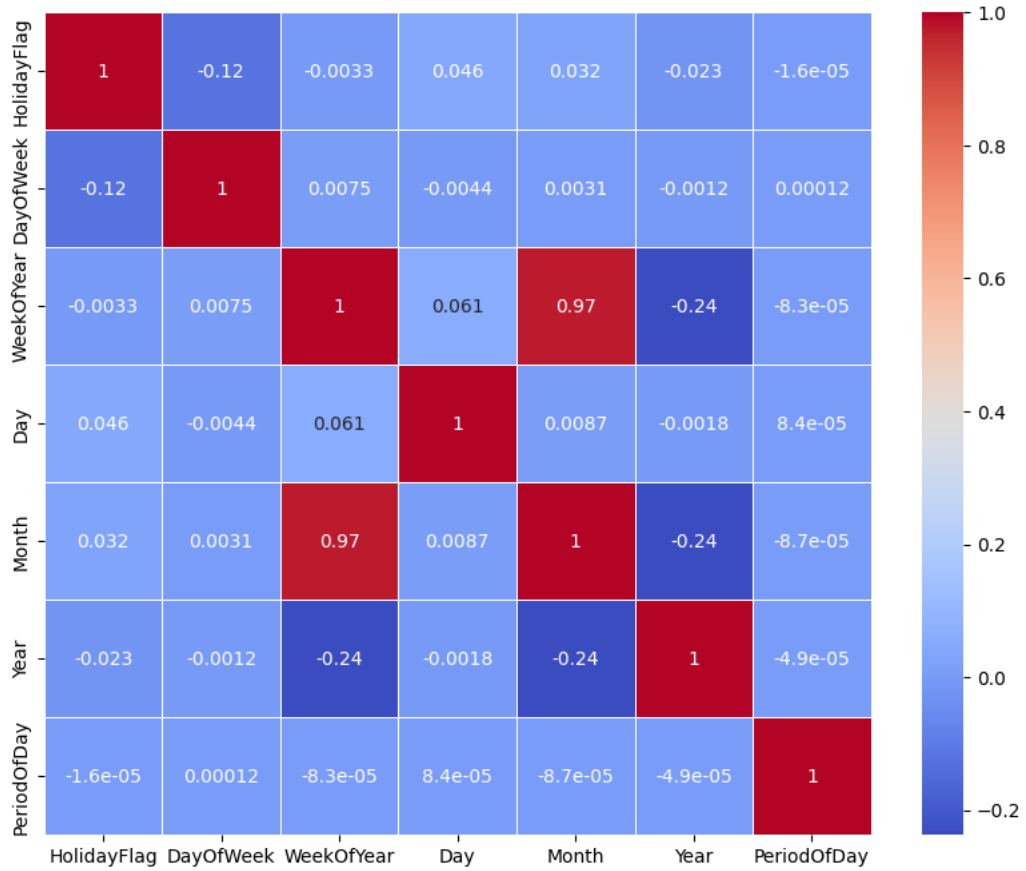
```
<bound method DataFrame.fillna of               DateTime         Holiday  HolidayFlag  DayOfWeek  WeekOfYear  \
0      01/11/2011 00:00            None           0          1          44
1      01/11/2011 00:30            None           0          1          44
2      01/11/2011 01:00            None           0          1          44
3      01/11/2011 01:30            None           0          1          44
4      01/11/2011 02:00            None           0          1          44
...                 ...             ...         ...        ...         ...
38009  31/12/2013 21:30  New Year's Eve           1          1           1
38010  31/12/2013 22:00  New Year's Eve           1          1           1
38011  31/12/2013 22:30  New Year's Eve           1          1           1
38012  31/12/2013 23:00  New Year's Eve           1          1           1
38013  31/12/2013 23:30  New Year's Eve           1          1           1

       Day  Month  Year  PeriodOfDay ForecastWindProduction SystemLoadEA  \
0        1     11  2011            0                315.31      3388.77
1        1     11  2011            1                321.80      3196.66
2        1     11  2011            2                328.57      3060.71
3        1     11  2011            3                335.60      2945.56
4        1     11  2011            4                342.90      2849.34
...    ...    ...   ...          ...                   ...         ...
38009   31     12  2013           43               1179.14      3932.22
38010   31     12  2013           44               1152.01      3821.44
38011   31     12  2013           45               1123.67      3724.21
38012   31     12  2013           46               1094.24      3638.16
38013   31     12  2013           47                1064.0      3624.25

       SMPEA ORKTemperature ORKWindspeed CO2Intensity ActualWindProduction  \
0      49.26           6.00         9.30       600.71               356.00
1      49.26           6.00        11.10       605.42               317.00
2      49.10           5.00        11.10       589.97               311.00
3      48.04           6.00         9.30       585.94               313.00
4      33.75           6.00        11.10       571.52               346.00
...      ...            ...          ...          ...                  ...
38009  34.51           6.00        22.20       285.31                812.0
38010  33.83           5.00        24.10       278.31                852.0
38011  31.75           4.00        20.40       280.91                962.0
38012  33.83           5.00        14.80       302.46                950.0
38013  33.83           5.00        16.70       308.01               1020.0

      SystemLoadEP2 SMPEP2
0           3159.60  54.32
1           2973.01  54.23
2           2834.00  54.23
3           2725.99  53.47
4           2655.64  39.87
...             ...    ...
38009       3692.95  42.45
38010        3571.0  33.83
38011       3460.29  31.75
38012       3563.99   50.6
38013       3517.08   34.9

[38014 rows x 18 columns]>
```

Feature engineering

label encoding

heatmap

```
<ipython-input-89-0ef81c6e1182>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version
  corr=df.corr()
```



convert float to numeric

```
def clean_column(column):


    return column

# Apply the clean_column function to the affected columns
columns_to_clean = ['ForecastWindProduction', 'SMPEA','SystemLoadEA','SystemLoadEP2','ORKWindspeed','CO2Intensity','SMPEP2']  # Replace with t
for column in columns_to_clean:
    df[column] = clean_column(df[column])

# Fill NaN values with the mean or other appropriate strategies
df.fillna(df.mean(), inplace=True)
```

```
<ipython-input-119-7dc0e84b08d9>:12: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future versi
  df.fillna(df.mean(), inplace=True)
```

specify target variabes and feature variables

```
target_variable = 'SystemLoadEA'  # Replace with your target variable
feature_columns = ['ForecastWindProduction', 'CO2Intensity', 'ActualWindProduction', 'SystemLoadEP2', 'SMPEP2']
```

split the dataset

```
from sklearn.model_selection import train_test_split


from pandas.core.groupby.generic import DataFrameGroupBy


# Define training and testing sets
X = df[feature_columns]
y = df[target_variable]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)



df["ForecastWindProduction"] = pd.to_numeric(df["ForecastWindProduction"], errors= 'coerce')
df["SystemLoadEA"] = pd.to_numeric(df["SystemLoadEA"], errors= 'coerce')
df["SMPEA"] = pd.to_numeric(df["SMPEA"], errors= 'coerce')
df["ORKTemperature"] = pd.to_numeric(df["ORKTemperature"], errors= 'coerce')
df["ORKWindspeed"] = pd.to_numeric(df["ORKWindspeed"], errors= 'coerce')
df["CO2Intensity"] = pd.to_numeric(df["CO2Intensity"], errors= 'coerce')
df["ActualWindProduction"] = pd.to_numeric(df["ActualWindProduction"], errors= 'coerce')
df["SystemLoadEP2"] = pd.to_numeric(df["SystemLoadEP2"], errors= 'coerce')
df["SMPEP2"] = pd.to_numeric(df["SMPEP2"], errors= 'coerce')
```

implement random forest algorithmn

```
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()
model.fit(X_train,y_train)
```

▾ RandomForestRegressor

RandomForestRegressor()

sample features

```
features = np.array([[10, 12, 54.10, 4241.05, 49.56]])
model.predict(features)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but RandomForestRegressor
  warnings.warn(
array([4284.7246])
```