DIVE INTO CODE

MACHINE LEARNING GRADUATION ASSIGNMENT

PROJECT

ELECTRICITY POWER CONSUMPTION

SELF INTRODUCTION

NAME: MOSES ALIEU BANGURA

UNIVERSITY: UNIVERSITY OF MANAGEMENT AND TECHNOLOGY

DEPARTMENT: COMPUTER SCIENCE

LEVEL: YEAR 3

COURSE: DIVE INTO CODE (DIC)

PROJECT: ELECTRICITY POWER CONSUMPTION

DATASET CODE OF ELECTRICITY POWER CONSUMPTION

In 1

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/electricity-consumption/train.csv
/kaggle/input/electricity-consumption/test.csv
```

In 2

```
!pip install DataScienceHelper
```

```
Collecting DataScienceHelper
  Downloading datasciencehelper-1.5.2.tar.gz (6.4 kB)
Building wheels for collected packages: DataScienceHelper
  Building wheel for DataScienceHelper (setup.py) ... done
  Created wheel for DataScienceHelper: filename=datasciencehelper-1.5.2-py3-none-any.whl size=5727
sha256=6aaeddf275658cdbca7b7f0564ea1b409583a584d2286b32b46d052c0b3399ae
  Stored in directory: /root/.cache/pip/wheels/49/5e/d2/84a664218a270ce173c1d02086f556367a00002afe0
aae6409
Successfully built DataScienceHelper
Installing collected packages: DataScienceHelper
Successfully installed DataScienceHelper-1.5.2
```

In 3

```
!pip install --upgrade pip
```

```
Collecting pip
  Downloading pip-20.2.2-py2.py3-none-any.whl (1.5 MB)
     |████████████████████████████████| 1.5 MB 402 kB/s eta 0:00:01
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 20.2.1
    Uninstalling pip-20.2.1:
      Successfully uninstalled pip-20.2.1
Successfully installed pip-20.2.2
```

In 4

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import DataScienceHelper as dsh
Import plotly.express as px
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

%matplotlib inline
```

In 5

```
import time
from datetime import datetime
import re
from math import *
```

In 6

```
data = pd.read_csv("/kaggle/input/electricity-consumption/train.csv")
data.head()
```

Out 6

	ID	datetime	temperature	var1	pressure	windspeed	var2	electricity_consumption
0	0	2013-07-01 00:00:00	-11.4	-17.1	1003.0	571.910	A	216.0
1	1	2013-07-01 01:00:00	-12.1	-19.3	996.0	575.040	A	210.0
2	2	2013-07-01 02:00:00	-12.9	-20.0	1000.0	578.435	A	225.0
3	3	2013-07-01 03:00:00	-11.4	-17.1	995.0	582.580	A	216.0
4	4	2013-07-01 04:00:00	-11.4	-19.3	1005.0	586.600	A	222.0

In 7 `data.tail()`

Out 7

	ID	datetime	temperature	var1	pressure	windspeed	var2	electricity_consumption	
	26491	34891	2017-06-23 19:00:00	-0.7	-15.0	1009.0	51.685	A	225.0
	26492	34892	2017-06-23 20:00:00	-2.9	-11.4	1005.0	56.105	A	213.0
	26493	34893	2017-06-23 21:00:00	-1.4	-12.9	995.0	61.275	A	213.0
	26494	34894	2017-06-23 22:00:00	-2.9	-11.4	996.0	67.210	A	210.0
	26495	34895	2017-06-23 23:00:00	-2.1	-11.4	1009.0	71.880	A	210.0

In 8 `data.isnull().sum()`

Out 8

```
ID                      0
Datetime                0
temperature             0
var1                    0
pressure                0
windspeed               0
var2                    0
electricity_consumption 0
dtype: int64
```

In 9 data.describe()

Out 9

	ID	temperature	var1	pressure	windspeed	electricity_consumption	
count	26496.000000	26496.000000	26496.000000	26496.000000	26496.000000	26496.000000	
mean	17455.500000	5.098989	-1.916233	986.450615	23.959956	298.359601	
std	10122.873673	8.682860	10.424860	12.002647	48.280321	108.020555	
min	0.000000	-17.100000	-32.900000	953.000000	1.075000	174.000000	
25%	8717.750000	-2.900000	-10.700000	978.000000	3.155000	219.000000	
50%	17435.500000	6.400000	-1.400000	986.000000	6.545000	267.000000	
75%	26177.250000	12.100000	7.900000	995.000000	22.260000	342.000000	
max	34895.000000	23.600000	18.600000	1024.000000	586.600000	1386.000000	

In 10 data.count()

Out 10

```
ID 26496 datetime                26496
temperature                      26496
var1                             26496
pressure                         26496
windspeed                        26496
var2                             26496
electricity_consumption          26496
dtype: int64
```

In 11 data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26496 entries, 0 to 26495
Data columns (total 8 columns):
 #     Column                 Non-Null Count      Dtype
---     ------ ---            -------------- ------      ---------
0      ID                     26496 non-null      int64
1      datetime               26496 non-null      object
2      temperature            26496 non-null      float64
3      var1                   26496 non-null      float64
4      pressure               26496 non-null      float64
5      windspeed              26496 non-null      float64
6      var2                   26496 non-null      object
7    electricity_consumption   26496 non-null       float64
dtypes: float64(5), int64(1), object(2)
memory usage: 1.6+ MB
```

In 12 `data.memory_usage()`

Out 12

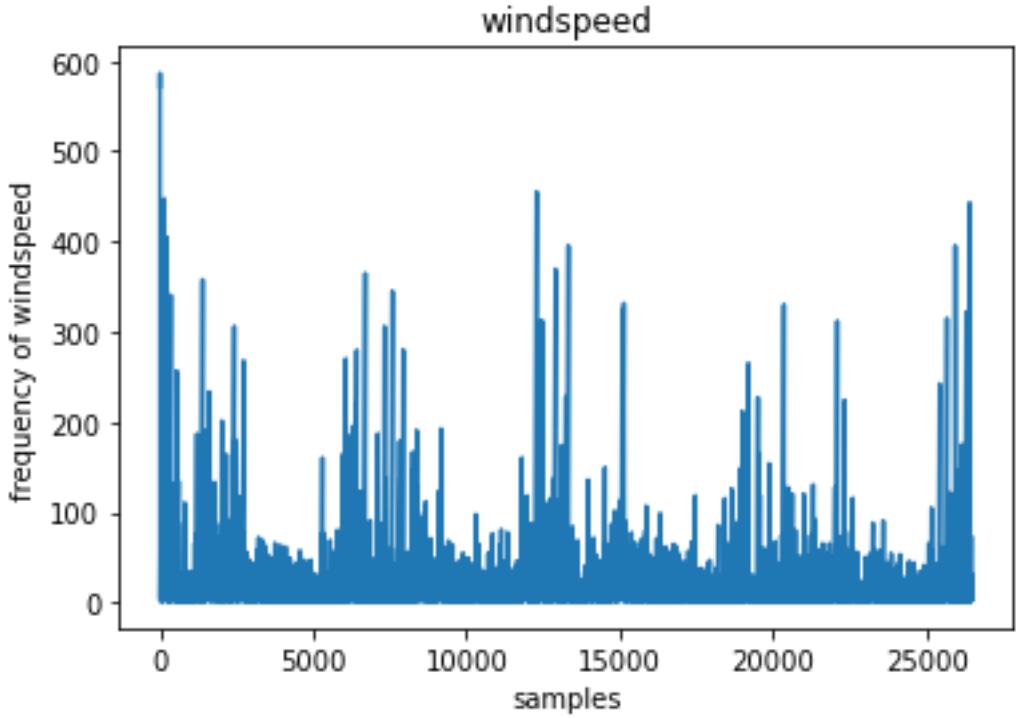
```
Index                    128
ID                    211968
datetime              211968
temperature           211968
var1                  211968
pressure              211968
windspeed             211968
var2                  211968
electricity_consumption  211968
dtype: int64
```

In 13 `data.windspeed.value_counts()`

Out 13

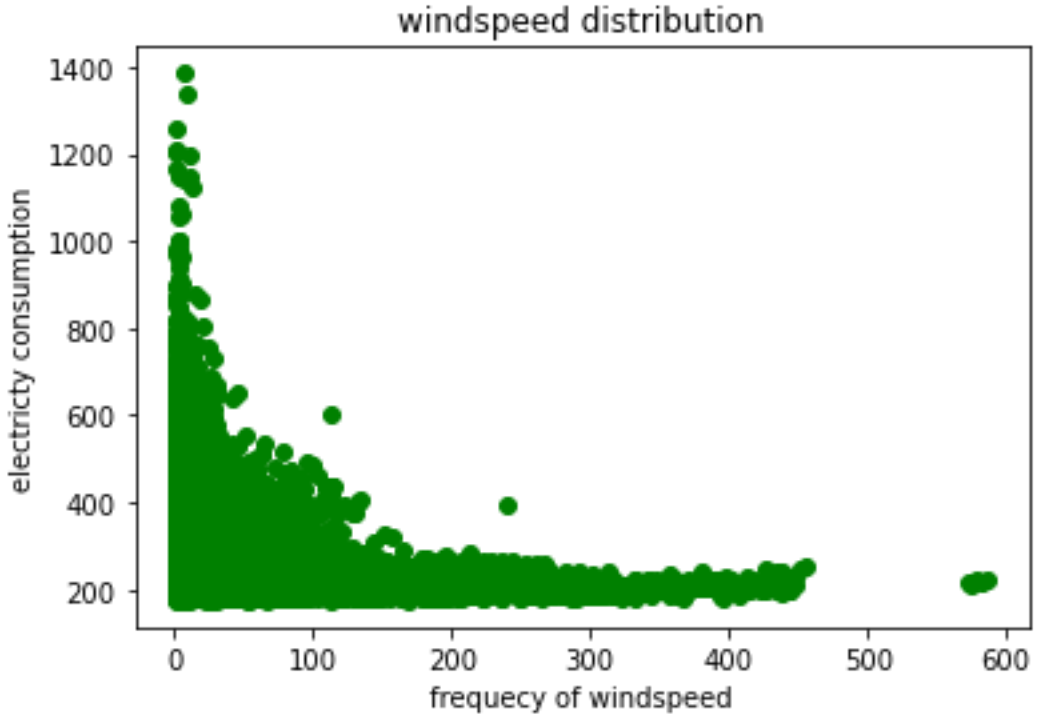
```
2.265      380
1.890      369
2.015      359
2.390      354
2.140      347
           ...
318.210      1
123.435      1
282.485      1
160.465      1
27.825       1
Name: windspeed, Length: 5603, dtype: int64
```

In 14
```
plott = data.windspeed
plt.plot(plott)
plt.xlabel("samples")
plt.ylabel("frequency of windspeed")
plt.title("windspeed")
plt.show()
```



In 15

```
plt.scatter(data.windspeed,data.electricity_consumption,c='green')
plt.xlabel("frequecy of windspeed")
plt.ylabel("electricty consumption")
plt.title("windspeed distribution")
plt.show()
```



In16

```
average = round(data.windspeed.mean(),3)
max_windspeed = round(max(data.windspeed),3)
min_windspeed = round(min(data.windspeed),3)
print(f'The average windspeed is : {average} ')
print(f'The maximum windspeed is : {max_windspeed}')
print(f'The minimum windspeed is : {min_windspeed}')
```

```
The average windspeed is : 23.96
The maximum windspeed is : 586.6
The minimum windspeed is : 1.075
```

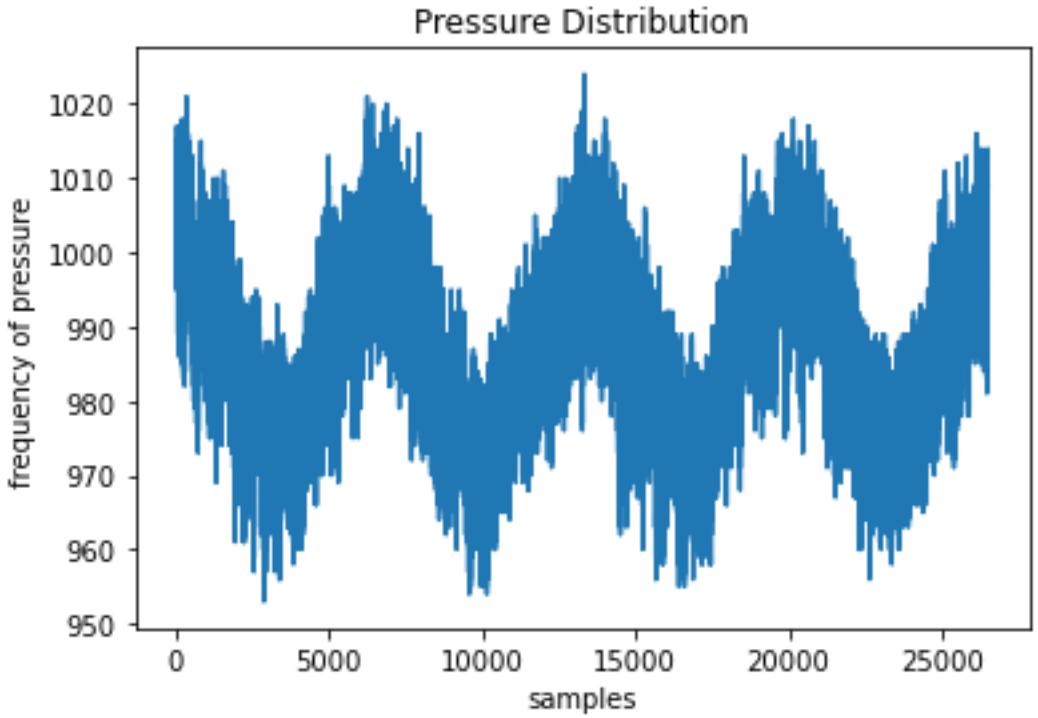
In17

```
avg_pressure = round(data.pressure.mean(),3)
max_pressure = round(data.pressure.max(),3)
min_pressure = round(data.pressure.min(),3)
print(f'The average pressure is : {avg_pressure}')
print(f'The maximum pressure is : {max_pressure}')
print(f'The minimum pressure is : {min_pressure}')
```

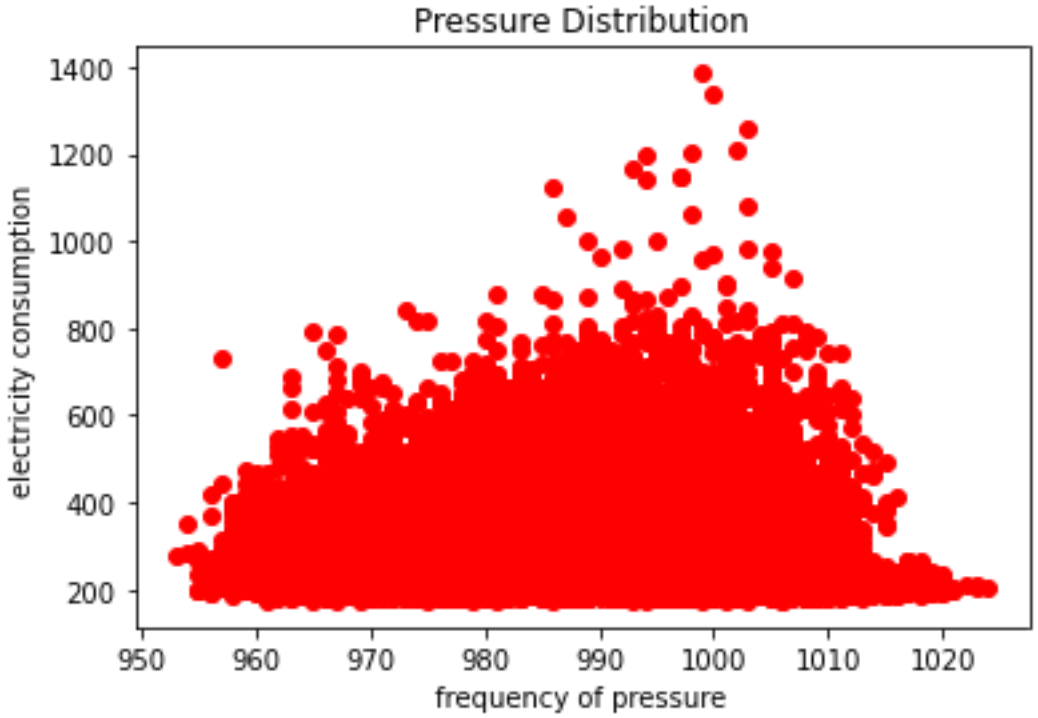
```
The average pressure is : 986.451
The maximum pressure is : 1024.0
The minimum pressure is : 953.0
```

In18

```
plt.plot(data.pressure)
plt.xlabel("samples")
plt.ylabel("frequency of pressure")
plt.title("Pressure Distribution")
plt.show()
```

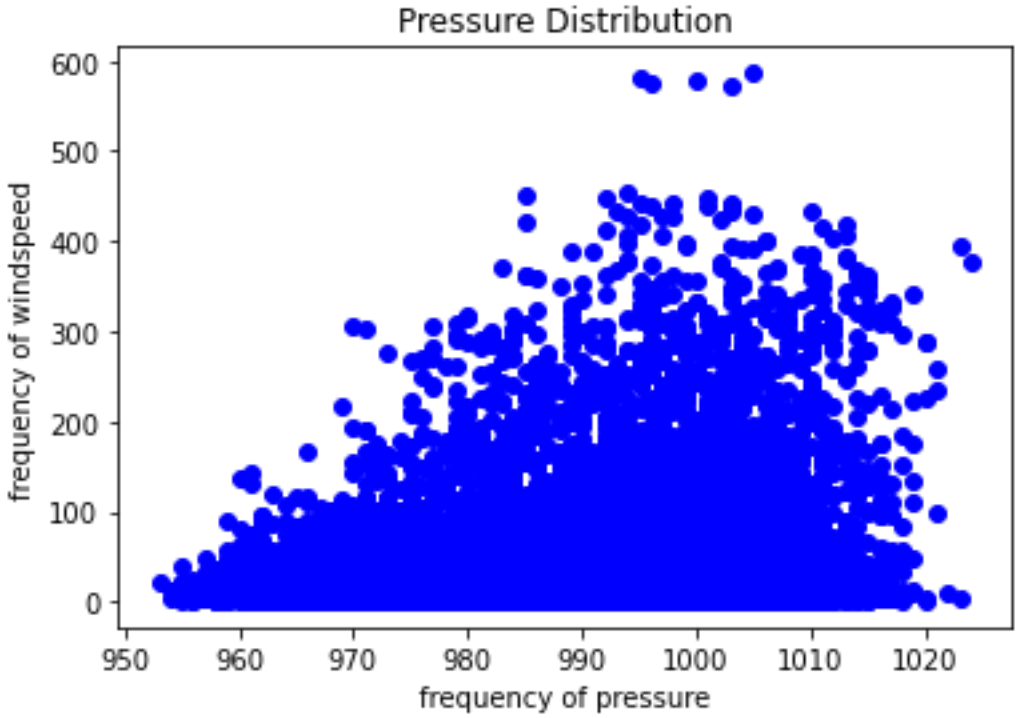


In19
```
plt.scatter(data.pressure,data.electricity_consumption,c='red')
plt.xlabel("frequency of pressure")
plt.ylabel("electricity consumption")
plt.title("Pressure Distribution")
plt.show()
```



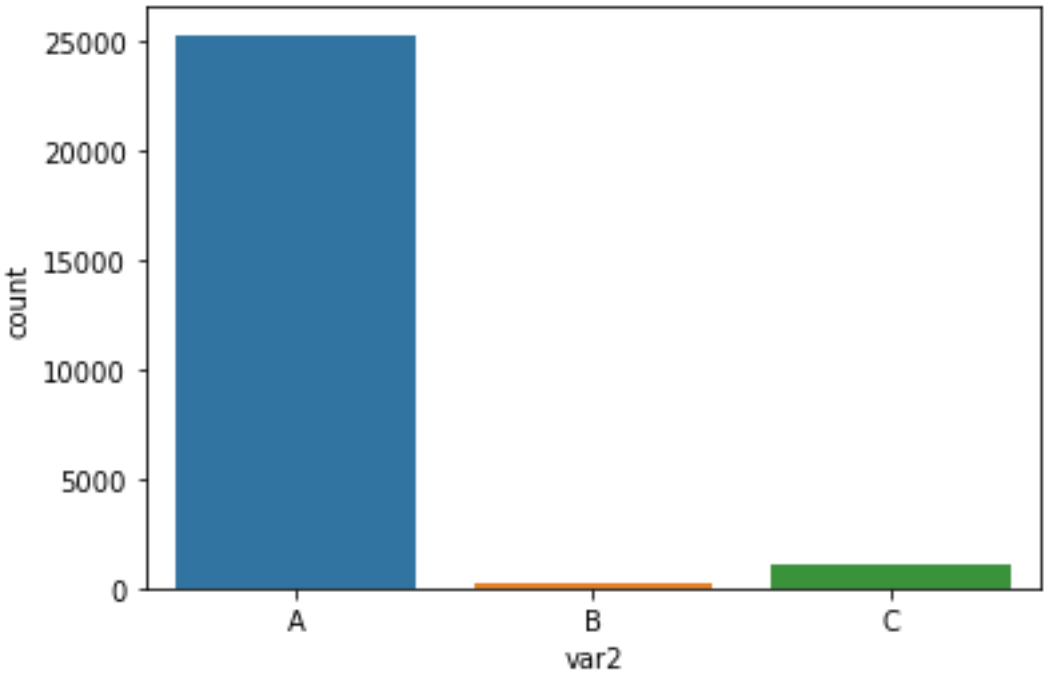
In20

```
plt.scatter(data.pressure,data.windspeed,c='blue')
plt.xlabel("frequency of pressure")
plt.ylabel("frequency of windspeed")
plt.title("Pressure Distribution")
plt.show()
```



In 21 sns.countplot(x='var2',data = data)

Out 21 <matplotlib.axes._subplots.AxesSubplot at 0x7f66e65e2f90>

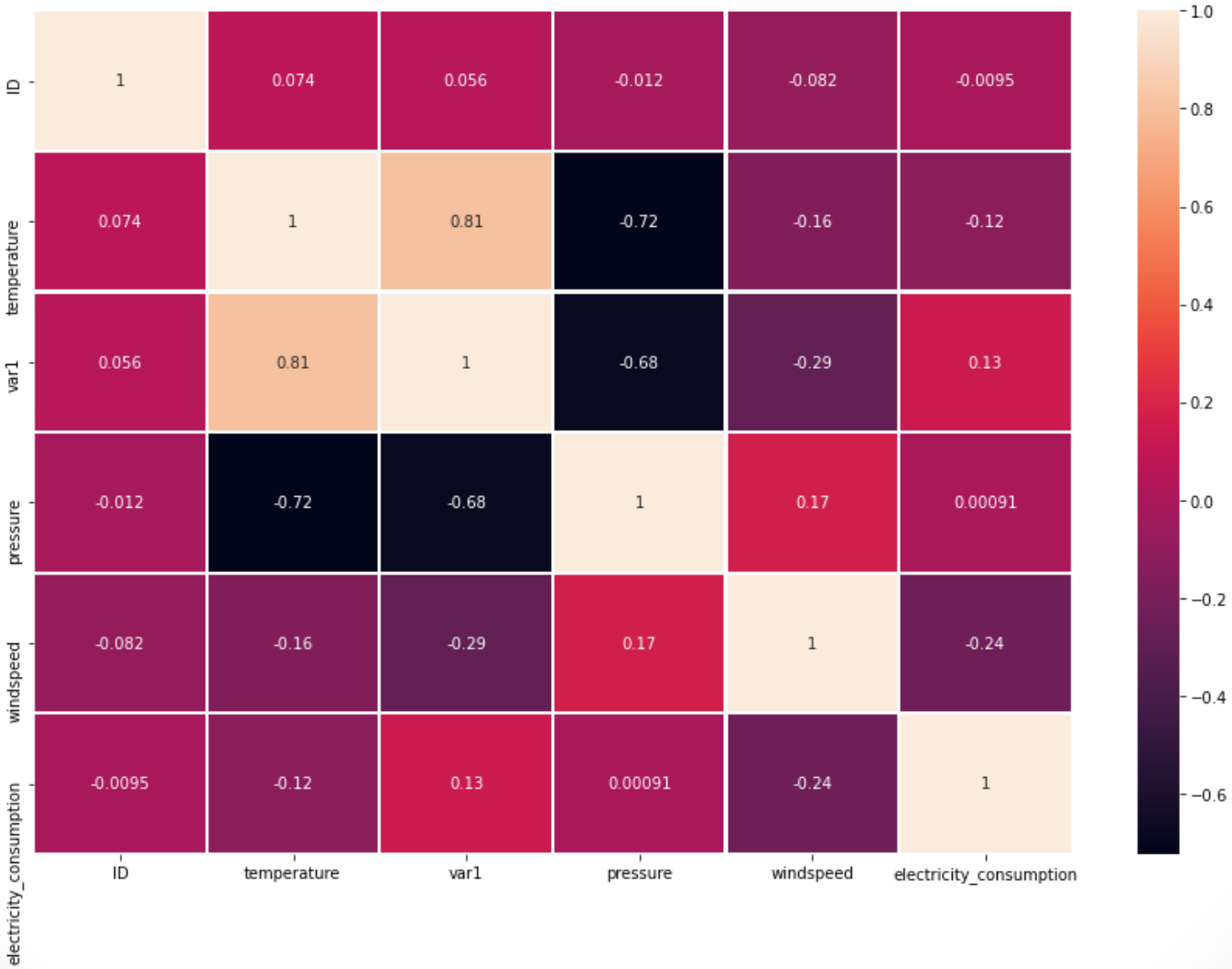


In 43

```
fig,ax = plt.subplots(figsize = (15,10))
corr = data.corr()
sns.heatmap(corr,xticklabels = corr.columns,annot = True,yticklabels = corr.columns,linewidth =1.2)
```

Out 43

<matplotlib.axes._subplots.AxesSubplot at 0x7f66e6c07e50>



In 23
```
corr[abs(corr['electricity_consumption']) > 0.1]['electricity_consumption']
```

Out 23
```
temperature                -0.117254
var1                        0.133914
windspeed                  -0.238883
electricity_consumption     1.000000
Name: electricity_consumption, dtype: float64
```

In 24
```
data.var1.value_counts()
```

Out 24
```
10.0    836
8.6     809
10.7    797
7.9     779
9.3     770
       ...
-29.3     3
17.9      2
-32.9     1
18.6      1
-32.1     1
Name: var1, Length: 71, dtype: int64
```

In 25 `data.var2.value_counts()`

Out 25
```
A    25239
C     1040
B      217
Name: var2, dtype: int64
```

Reference

This is what i have gathered so far for my Graduation Assignment project title ELECTRICITY POWER CONSUMPTION, also gathered some assistance material on KAGGLE to add some important materials on my project

DIVE INTO CODE

MACHINE LEARNING GRADUATION ASSIGNMENT

END OF SESSION