```
#Dealing with a lot of missing data.
df.isna().sum()
```

```
Unnamed: 0                         0
ActivePower                    23474
AmbientTemperatue              24407
BearingShaftTemperature        55706
Blade1PitchAngle               76228
Blade2PitchAngle               76333
Blade3PitchAngle               76333
ControlBoxTemperature          56064
GearboxBearingTemperature      55684
GearboxOilTemperature          55786
GeneratorRPM                   55929
GeneratorWinding1Temperature   55797
GeneratorWinding2Temperature   55775
HubTemperature                 55818
MainBoxTemperature             55717
NacellePosition                45946
ReactivePower                  23476
RotorRPM                       56097
TurbineStatus                  55316
WTG                                0
WindDirection                  45946
WindSpeed                      23629
dtype: int64
```

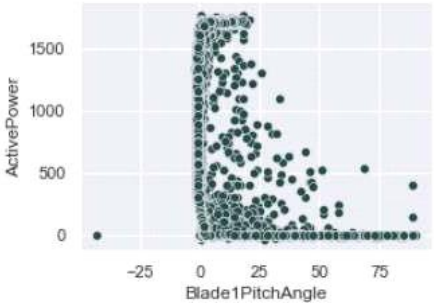
```
df.info(verbose=True)
```

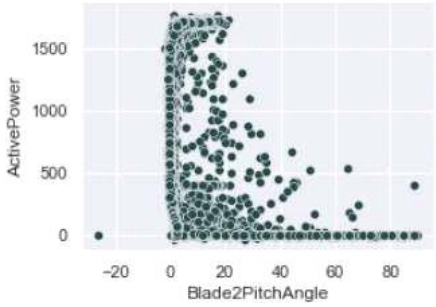
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 118224 entries, 0 to 118223
Data columns (total 22 columns):
 #   Column                         Non-Null Count   Dtype  
---  ------                         --------------   -----  
 0   Unnamed: 0                     118224 non-null  object 
 1   ActivePower                    94750 non-null   float64
 2   AmbientTemperatue              93817 non-null   float64
 3   BearingShaftTemperature        62518 non-null   float64
 4   Blade1PitchAngle               41996 non-null   float64
 5   Blade2PitchAngle               41891 non-null   float64
 6   Blade3PitchAngle               41891 non-null   float64
 7   ControlBoxTemperature          62160 non-null   float64
 8   GearboxBearingTemperature      62540 non-null   float64
 9   GearboxOilTemperature          62438 non-null   float64
 10  GeneratorRPM                   62295 non-null   float64
 11  GeneratorWinding1Temperature   62427 non-null   float64
 12  GeneratorWinding2Temperature   62449 non-null   float64
 13  HubTemperature                 62406 non-null   float64
 14  MainBoxTemperature             62507 non-null   float64
 15  NacellePosition                72278 non-null   float64
 16  ReactivePower                  94748 non-null   float64
 17  RotorRPM                       62127 non-null   float64
 18  TurbineStatus                  62908 non-null   float64
 19  WTG                            118224 non-null  object 
 20  WindDirection                  72278 non-null   float64
 21  WindSpeed                      94595 non-null   float64
```

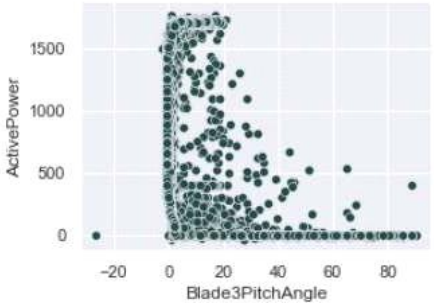
```
#filling missing data with the median
df.fillna(df.median(),inplace = True)
```

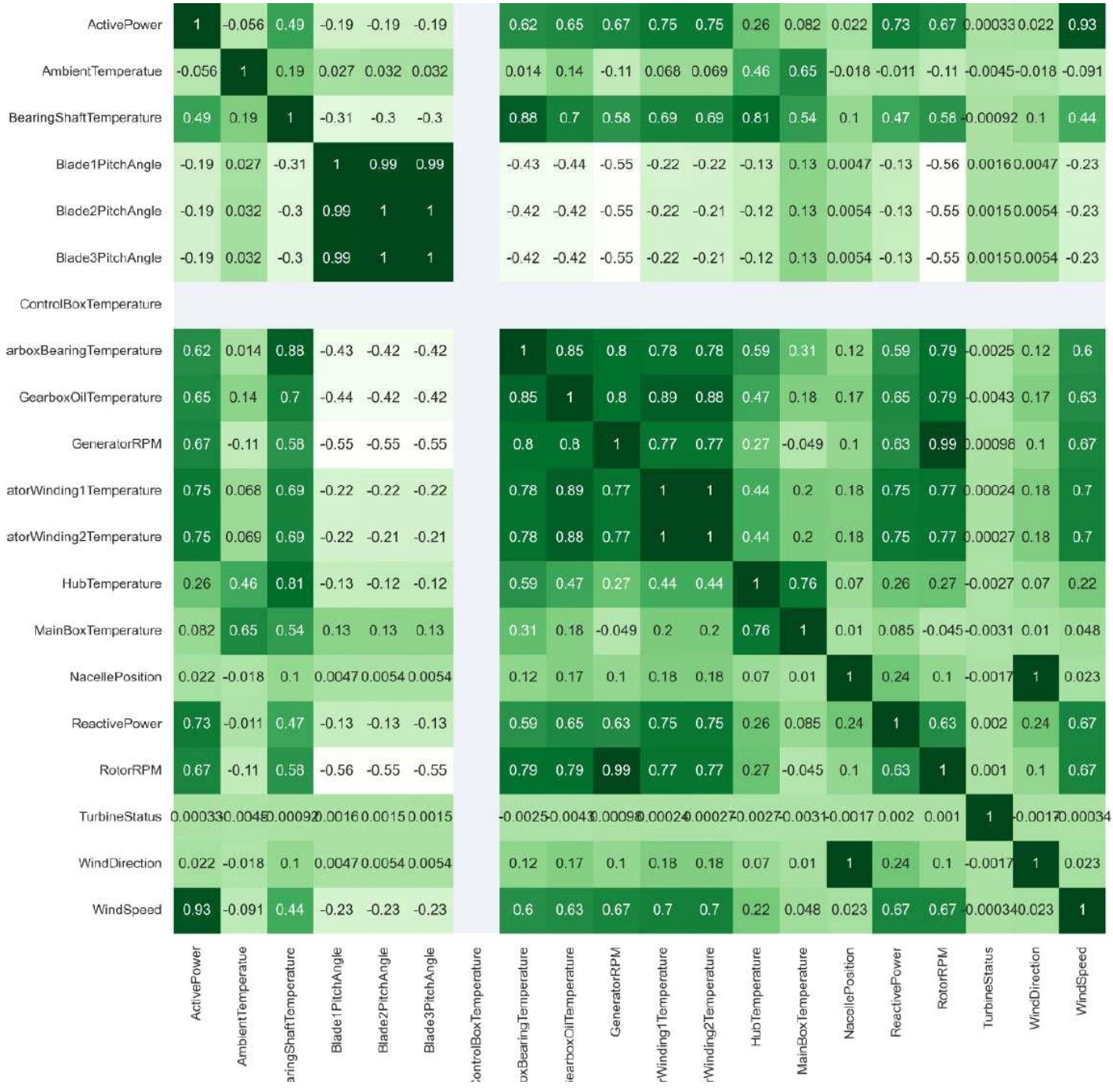
```
df.describe()
```

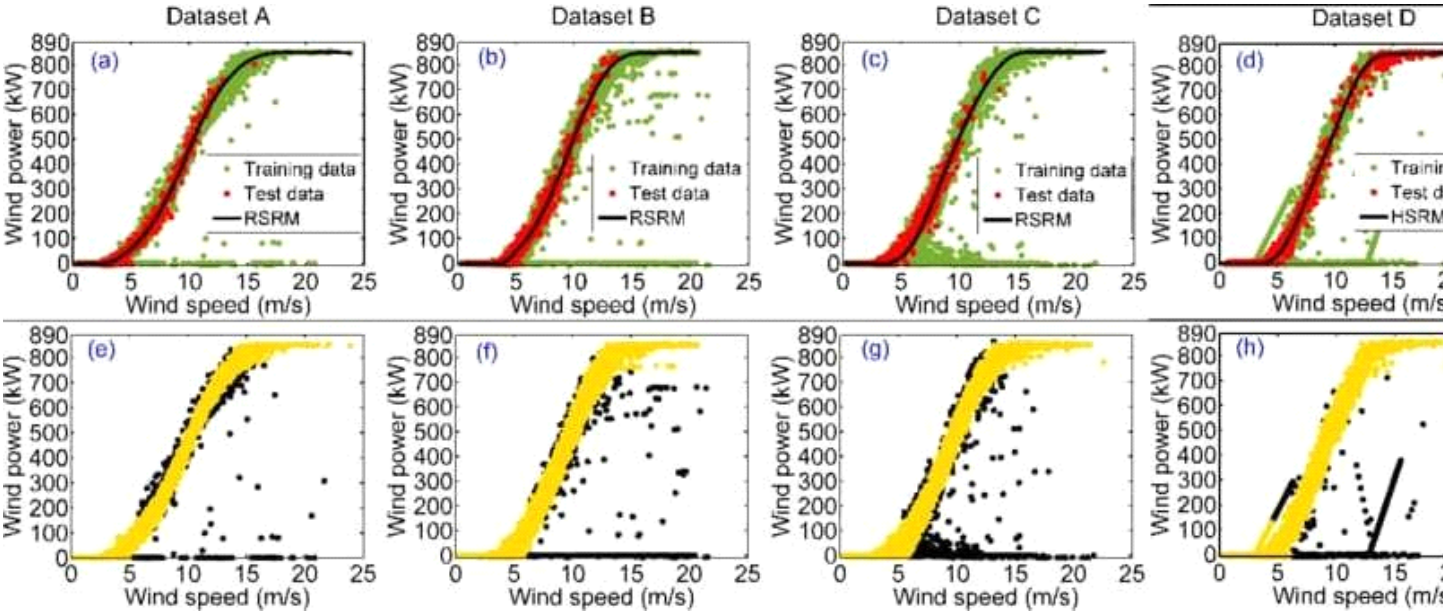
	ActivePower	AmbientTemperatue	BearingShaftTemperature	Blade1PitchAngle	Blade2PitchAngle	Blade3PitchAngle	ControlBoxTemperature	Gearb
count	118224.000000	118224.000000	118224.000000	118224.000000	118224.000000	118224.000000	118224.0	
mean	576.131539	28.685033	42.963394	3.717605	4.130284	4.130284	0.0	
std	554.004562	3.896064	4.032802	13.093631	12.834936	12.834936	0.0	
min	-38.524659	0.000000	0.000000	-43.156734	-26.443415	-26.443415	0.0	
25%	149.921738	26.350811	42.512424	0.394399	0.888977	0.888977	0.0	
50%	402.654893	28.340541	42.910877	0.394399	0.888977	0.888977	0.0	
75%	838.515253	30.714347	43.291577	0.394399	0.888977	0.888977	0.0	
max	1779.032433	42.405597	55.088655	90.143610	90.017830	90.017830	0.0	








ActivePower 1 -0.056 0.49 -0.19 -0.19 -0.19 0.62 0.65 0.67 0.75 0.75 0.26 0.082 0.022 0.73 0.67 0.00033 0.022 0.93
AmbientTemperatue -0.056 1 0.19 0.027 0.032 0.032 0.014 0.14 -0.11 0.068 0.069 0.46 0.65 -0.018 -0.011 -0.11 -0.0045 -0.018 -0.091
BearingShaftTemperature 0.49 0.19 1 -0.31 -0.3 -0.3 0.88 0.7 0.58 0.69 0.69 0.81 0.54 0.1 0.47 0.58 -0.00092 0.1 0.44
Blade1PitchAngle -0.19 0.027 -0.31 1 0.99 0.99 -0.43 -0.44 -0.55 -0.22 -0.22 -0.13 0.13 0.0047 -0.13 -0.56 0.0016 0.0047 -0.23
Blade2PitchAngle -0.19 0.032 -0.3 0.99 1 1 -0.42 -0.42 -0.55 -0.22 -0.21 -0.12 0.13 0.0054 -0.13 -0.55 0.0015 0.0054 -0.23
Blade3PitchAngle -0.19 0.032 -0.3 0.99 1 1 -0.42 -0.42 -0.55 -0.22 -0.21 -0.12 0.13 0.0054 -0.13 -0.55 0.0015 0.0054 -0.23
ControlBoxTemperature
arboxBearingTemperature 0.62 0.014 0.88 -0.43 -0.42 -0.42 1 0.85 0.8 0.78 0.78 0.59 0.31 0.12 0.59 0.79 -0.0025 0.12 0.6
GearboxOilTemperature 0.65 0.14 0.7 -0.44 -0.42 -0.42 0.85 1 0.8 0.89 0.88 0.47 0.18 0.17 0.65 0.79 -0.0043 0.17 0.63
GeneratorRPM 0.67 -0.11 0.58 -0.55 -0.55 -0.55 0.8 0.8 1 0.77 0.77 0.27 -0.049 0.1 0.63 0.99 0.00098 0.1 0.67
atorWinding1Temperature 0.75 0.068 0.69 -0.22 -0.22 -0.22 0.78 0.89 0.77 1 1 0.44 0.2 0.18 0.75 0.77 0.00024 0.18 0.7
atorWinding2Temperature 0.75 0.069 0.69 -0.22 -0.21 -0.21 0.78 0.88 0.77 1 1 0.44 0.2 0.18 0.75 0.77 0.00027 0.18 0.7
HubTemperature 0.26 0.46 0.81 -0.13 -0.12 -0.12 0.59 0.47 0.27 0.44 0.44 1 0.76 0.07 0.26 0.27 -0.0027 0.07 0.22
MainBoxTemperature 0.082 0.65 0.54 0.13 0.13 0.13 0.31 0.18 -0.049 0.2 0.2 0.76 1 0.01 0.085 -0.045 -0.0031 0.01 0.048
NacellePosition 0.022 -0.018 0.1 0.0047 0.0054 0.0054 0.12 0.17 0.1 0.18 0.18 0.07 0.01 1 0.24 0.1 -0.0017 1 0.023
ReactivePower 0.73 -0.011 0.47 -0.13 -0.13 -0.13 0.59 0.65 0.63 0.75 0.75 0.26 0.085 0.24 1 0.63 0.002 0.24 0.67
RotorRPM 0.67 -0.11 0.58 -0.56 -0.55 -0.55 0.79 0.79 0.99 0.77 0.77 0.27 -0.045 0.1 0.63 1 0.001 0.1 0.67
TurbineStatus 0.00033 -0.0045 -0.00092 0.0016 0.0015 0.0015 -0.0025 -0.0043 0.00098 0.00024 0.00027 -0.0027 -0.0031 -0.0017 0.002 0.001 1 -0.0017 0.00034
WindDirection 0.022 -0.018 0.1 0.0047 0.0054 0.0054 0.12 0.17 0.1 0.18 0.18 0.07 0.01 1 0.24 0.1 -0.0017 1 0.023
WindSpeed 0.93 -0.091 0.44 -0.23 -0.23 -0.23 0.6 0.63 0.67 0.7 0.7 0.22 0.048 0.023 0.67 0.67 0.00034 0.023 1
ActivePower
AmbientTemperatue
aringShaftTemperature
Blade1PitchAngle
Blade2PitchAngle
Blade3PitchAngle
ontrolBoxTemperature
oxBearingTemperature
iearboxOilTemperature
GeneratorRPM
rWinding1Temperature
rWinding2Temperature
HubTemperature
MainBoxTemperature
NacellePosition
ReactivePower
RotorRPM
TurbineStatus
WindDirection
WindSpeed




```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dataset= pd.read_csv('wind turbine.csv')
X= dataset.iloc[:,:-1].values
Y= dataset.loc[:,:-1].values
print X
print Y
```