Data Science and Machine Learning Internship
(HDLC Technologies)

Name: *B. Deva SriRama Sai Ganesh*
Email: bandarusaiganesh2003@gmail.com

WEEK - 4 Assignment

Question:

1. Read the csv file named windpower.csv [attached]
2. Print the top 10 rows.
3. Print the bottom 5 rows.
4. Display the structure of dataset with datatypes.
5. Find out Mean, Standard deviation, Min, Max for the column wind speed.
6. Fill the NULL values with the average value of that column.
7. Create a histogram plot for the column windspeed.
8. Create scatter plot between windspeed and Theoretical power curve.
9. Display the rows where windspeed < 5.
10. Extract the rows where wind direction > 200

Solutions:

Python code:

```
!pip install pandas

import pandas as obj
import numpy as np
import matplotlib.pyplot as plt

windpower_data                                                    =
obj.read_csv("C:\\Users\\banda\\Downloads\\windpower.csv")
print(windpower_data)
#print(windpower_data.to_string())

print("\n Displaying the top 10 rows using 'head()': \n")
print(windpower_data.head(10))
```

```
print("\n Displaying the bottom 5 rows using 'tail()': \n")
print(windpower_data.tail())

print("\n Displaying the structure of dataset with datatypes: \n")
print(windpower_data.info())

print("\n The Mean for the column 'Wind Speed (m/s)': \n")
mean_ = windpower_data['Wind Speed (m/s)'].mean()
print(mean_)

print("\n The Standard Deviation for the column 'Wind Speed (m/s)': \n")
standard_deviation = windpower_data['Wind Speed (m/s)'].std()
print("\n", standard_deviation)

print("\n The Min for the column 'Wind Speed (m/s)': \n")
min_ = windpower_data['Wind Speed (m/s)'].min()
print("\n", min_)

print("\n The Max for the column 'Wind Speed (m/s)': \n")
max_ = windpower_data['Wind Speed (m/s)'].max()
print("\n", max_)

average_value = windpower_data['Wind Speed (m/s)'].mean()
print("\n Average value: ", average_value)
windpower_data.fillna(average_value, inplace = True)
print("\n ", windpower_data.head(20))

print("\n A histogram plot for the column 'Wind Speed (m/s)': \n")
#windpower_data['Wind Speed (m/s)'].plot(kind = "hist")

plt.hist(windpower_data['Wind Speed (m/s)'])
plt.show()

print("\n A Scatter plot between the column 'Wind Speed (m/s)' and 'Theoretical_Power_Curve (KWh)': \n")
windpower_data.plot(kind = 'scatter', x = 'Theoretical_Power_Curve (KWh)', y = 'Wind Speed (m/s)' )
plt.show()
```

```
print("\n Displaying the rows where windspeed < 5: \n")
result = windpower_data[windpower_data['Wind Speed (m/s)'] < 5]
print(result)

#for x in windpower_data.index:
#    if windpower_data.loc[x, 'Wind Speed (m/s)'] < 5:
#        print(windpower_data.loc[x])

print("\n Extracting the rows where wind direction > 200: \n")
result = windpower_data[windpower_data['Wind Direction (°)'] > 200]
print(result)

#for x in windpower_data.index:
#    if windpower_data.loc[x, 'Wind Direction (°)'] > 200:
#        print(windpower_data.loc[x])
```

windpower.csv

```
In [1]: !pip install pandas

Requirement already satisfied: pandas in c:\users\banda\anaconda3\lib\site-packages (1.4.4)
Requirement already satisfied: numpy>=1.18.5 in c:\users\banda\anaconda3\lib\site-packages (from pandas) (1.21.5)
Requirement already satisfied: pytz>=2020.1 in c:\users\banda\anaconda3\lib\site-packages (from pandas) (2022.1)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\banda\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\banda\anaconda3\lib\site-packages (from python-dateutil>=2.8.1->pandas) (1.
16.0)
```

1 Ans:

```
In [5]: import pandas as obj
windpower_data = obj.read_csv("C:\\Users\\banda\\Downloads\\windpower.csv")
print(windpower_data)
#print(windpower_data.to_string())
```

Output:

```
             Date/Time  LV ActivePower (kW)  Wind Speed (m/s)  \
0      01 01 2018 00:00           380.047791          5.311336
1      01 01 2018 00:10           453.769196          5.672167
2      01 01 2018 00:20           306.376587          5.216037
3      01 01 2018 00:30           419.645904          5.659674
4      01 01 2018 00:40           380.650696          5.577941
...                 ...                  ...               ...
50525  31 12 2018 23:10          2963.980957         11.404030
50526  31 12 2018 23:20          1684.353027          7.332648
50527  31 12 2018 23:30          2201.106934          8.435358
50528  31 12 2018 23:40          2515.694092          9.421366
50529  31 12 2018 23:50          2820.466064          9.979332

       Theoretical_Power_Curve (KWh)  Wind Direction (°)
0                          416.328908          259.994904
1                          519.917511          268.641113
2                          390.900016          272.564789
3                          516.127569          271.258087
4                          491.702972          265.674286
...                               ...                 ...
50525                     3397.190793           80.502724
50526                     1173.055771           84.062599
50527                     1788.284755           84.742500
50528                     2418.382503           84.297913
50529                     2779.184096           82.274620

[50530 rows x 5 columns]
```

2 Ans:

```
In [10]: print("\n Displaying the top 10 rows using 'head()': \n")
         print(windpower_data.head(10))
```

Output:

```
 Displaying the top 10 rows using 'head()':

            Date/Time  LV ActivePower (kW)  Wind Speed (m/s)  \
0  01 01 2018 00:00           380.047791          5.311336
1  01 01 2018 00:10           453.769196          5.672167
2  01 01 2018 00:20           306.376587          5.216037
3  01 01 2018 00:30           419.645904          5.659674
4  01 01 2018 00:40           380.650696          5.577941
5  01 01 2018 00:50           402.391998          5.604052
6  01 01 2018 01:00           447.605713          5.793008
7  01 01 2018 01:10           387.242188          5.306050
8  01 01 2018 01:20           463.651215          5.584629
9  01 01 2018 01:30           439.725708          5.523228

   Theoretical_Power_Curve (KWh)  Wind Direction (°)
0                     416.328908          259.994904
1                     519.917511          268.641113
2                     390.900016          272.564789
3                     516.127569          271.258087
4                     491.702972          265.674286
5                     499.436385          264.578613
6                     557.372363          266.163605
7                     414.898179          257.949493
8                     493.677652          253.480698
9                     475.706783          258.723785
```

3 Ans:

```
In [11]: print("\n Displaying the bottom 5 rows using 'tail()': \n")
         print(windpower_data.tail())
```

```
 Displaying the bottom 5 rows using 'tail()':

                Date/Time  LV ActivePower (kW)  Wind Speed (m/s)  \
50525  31 12 2018 23:10          2963.980957         11.404030
50526  31 12 2018 23:20          1684.353027          7.332648
50527  31 12 2018 23:30          2201.106934          8.435358
50528  31 12 2018 23:40          2515.694092          9.421366
50529  31 12 2018 23:50          2820.466064          9.979332

       Theoretical_Power_Curve (KWh)  Wind Direction (°)
50525                     3397.190793           80.502724
50526                     1173.055771           84.062599
50527                     1788.284755           84.742500
50528                     2418.382503           84.297913
50529                     2779.184096           82.274620
```

4 Ans:

```
In [12]: print("\n Displaying the structure of dataset with datatypes: \n")
         print(windpower_data.info())
```

```
 Displaying the structure of dataset with datatypes:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50530 entries, 0 to 50529
Data columns (total 5 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Date/Time                       50530 non-null  object
 1   LV ActivePower (kW)             50530 non-null  float64
 2   Wind Speed (m/s)                50528 non-null  float64
 3   Theoretical_Power_Curve (KWh)   50530 non-null  float64
 4   Wind Direction (°)              50530 non-null  float64
dtypes: float64(4), object(1)
memory usage: 1.9+ MB
None
```

5 Ans:

```
In [19]: print("\n The Mean for the column 'Wind Speed (m/s)': \n")
         mean_ = windpower_data['Wind Speed (m/s)'].mean()
         print(mean_)

         print("\n The Standard Deviation for the column 'Wind Speed (m/s)': \n")
         standard_deviation = windpower_data['Wind Speed (m/s)'].std()
         print("\n", standard_deviation)

         print("\n The Min for the column 'Wind Speed (m/s)': \n")
         min_ = windpower_data['Wind Speed (m/s)'].min()
         print("\n", min_)

         print("\n The Max for the column 'Wind Speed (m/s)': \n")
         max_ = windpower_data['Wind Speed (m/s)'].max()
         print("\n", max_)
```

Output:

```
The Mean for the column 'Wind Speed (m/s)':

7.558010703817663

The Standard Deviation for the column 'Wind Speed (m/s)':


4.227238904770305

The Min for the column 'Wind Speed (m/s)':


0.0

The Max for the column 'Wind Speed (m/s)':


25.20601082
```

6 Ans:

```
In [5]: import pandas as obj
        windpower_data = obj.read_csv("C:\\Users\\banda\\Downloads\\windpower.csv")

        average_value = windpower_data['Wind Speed (m/s)'].mean()
        print("\n Average value: ", average_value)
        windpower_data.fillna(average_value, inplace = True)
        print("\n ", windpower_data.head(20))
```

Output:

Average value: 7.558010703817663

	Date/Time	LV ActivePower (kW)	Wind Speed (m/s)	\
0	01 01 2018 00:00	380.047791	5.311336	
1	01 01 2018 00:10	453.769196	5.672167	
2	01 01 2018 00:20	306.376587	5.216037	
3	01 01 2018 00:30	419.645904	5.659674	
4	01 01 2018 00:40	380.650696	5.577941	
5	01 01 2018 00:50	402.391998	5.604052	
6	01 01 2018 01:00	447.605713	5.793008	
7	01 01 2018 01:10	387.242188	5.306050	
8	01 01 2018 01:20	463.651215	5.584629	
9	01 01 2018 01:30	439.725708	5.523228	
10	01 01 2018 01:40	498.181702	7.558011	
11	01 01 2018 01:50	526.816223	5.934199	
12	01 01 2018 02:00	710.587280	6.547414	
13	01 01 2018 02:10	655.194275	6.199746	
14	01 01 2018 02:20	754.762512	6.505383	
15	01 01 2018 02:30	790.173279	6.634116	
16	01 01 2018 02:40	742.985291	6.378913	
17	01 01 2018 02:50	748.229614	6.446653	
18	01 01 2018 03:00	736.647827	6.415083	
19	01 01 2018 03:10	787.246216	7.558011	

	Theoretical_Power_Curve (KWh)	Wind Direction (°)
0	416.328908	259.994904
1	519.917511	268.641113
2	390.900016	272.564789
3	516.127569	271.258087
4	491.702972	265.674286
5	499.436385	264.578613
6	557.372363	266.163605
7	414.898179	257.949493
8	493.677652	253.480698
9	475.706783	258.723785
10	535.841397	251.850998
11	603.014076	265.504700
12	824.662514	274.232910
13	693.472641	266.733185
14	808.098139	266.760406
15	859.459021	270.493195
16	759.434537	266.593292
17	785.281010	265.571808
18	773.172863	261.158691
19	781.771216	257.560211

7 Ans:

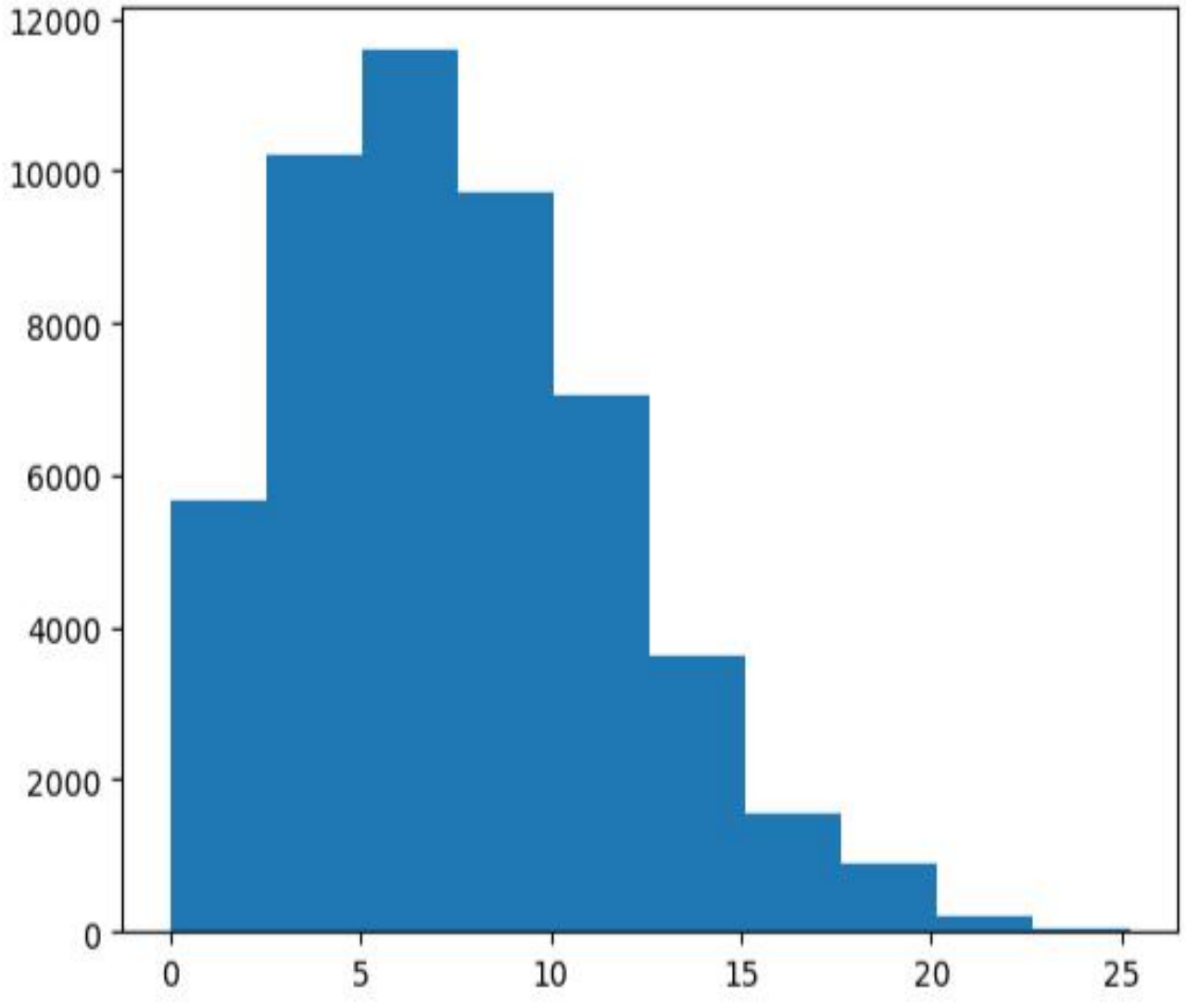
```
In [16]: import numpy as np
         import matplotlib.pyplot as plt
         import pandas as obj
         windpower_data = obj.read_csv("C:\\Users\\banda\\Downloads\\windpower.csv")

         print("\n A histogram plot for the column 'Wind Speed (m/s)': \n")
         #windpower_data['Wind Speed (m/s)'].plot(kind = "hist")

         plt.hist(windpower_data['Wind Speed (m/s)'])
         plt.show()
```

Output:

```
 A histogram plot for the column 'Wind Speed (m/s)':
```



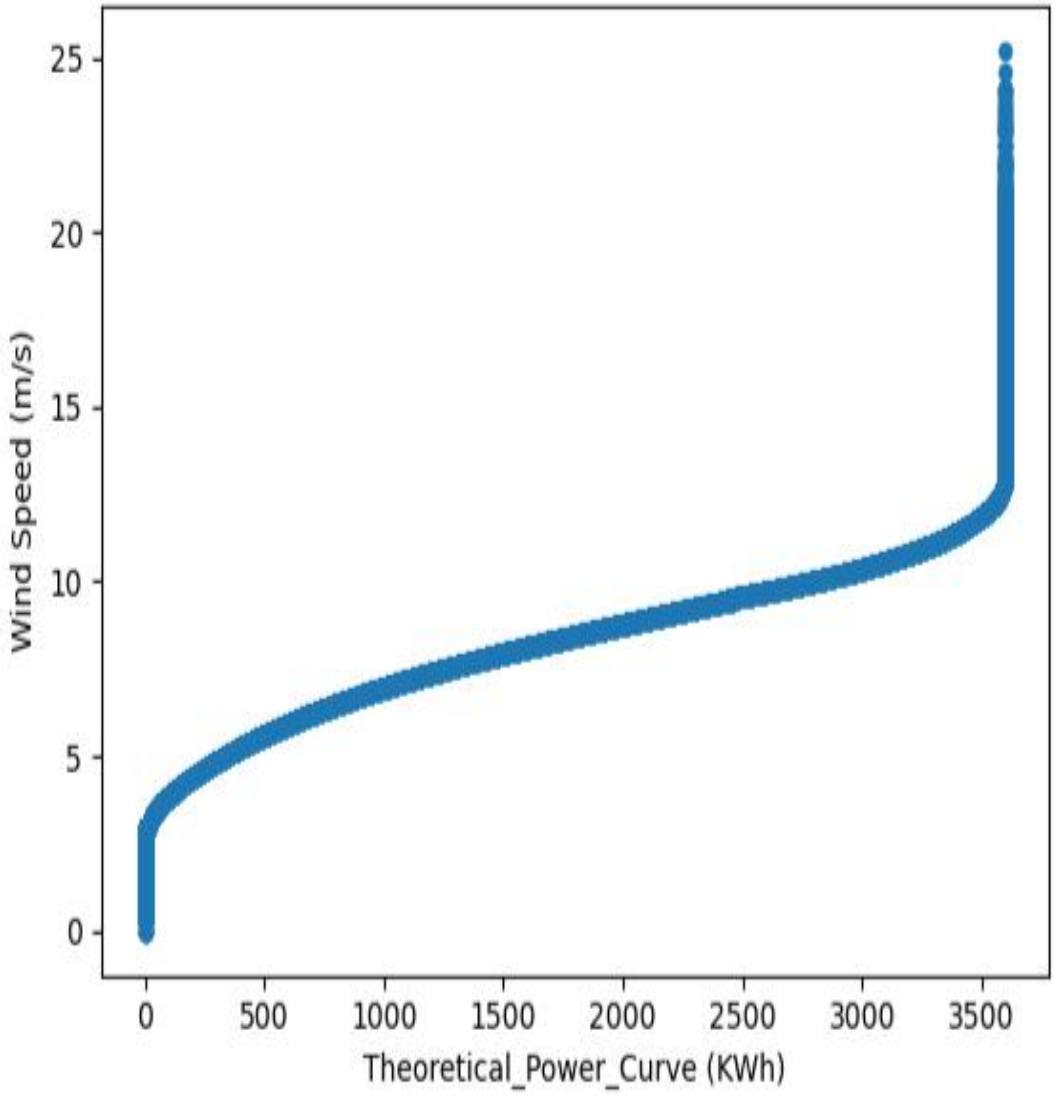
8 Ans:

In [19]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as obj
windpower_data = obj.read_csv("C:\\Users\\banda\\Downloads\\windpower.csv")

print("\n A Scatter plot between the column 'Wind Speed (m/s)' and 'Theoretical_Power_Curve (KWh)': \n")
windpower_data.plot(kind = 'scatter', x = 'Theoretical_Power_Curve (KWh)', y = 'Wind Speed (m/s)' )
plt.show()
```

Output:

A Scatter plot between the column 'Wind Speed (m/s)' and 'Theoretical_Power_Curve (KWh)':



9 Ans:

```
In [28]: import pandas as obj
         windpower_data = obj.read_csv("C:\\Users\\banda\\Downloads\\windpower.csv")

         print("\n Displaying the rows where windspeed < 5: \n")
         result = windpower_data[windpower_data['Wind Speed (m/s)'] < 5]
         print(result)

         #for x in windpower_data.index:
         #    if windpower_data.loc[x, 'Wind Speed (m/s)'] < 5:
         #        print(windpower_data.loc[x])
```

Output:

```
 Displaying the rows where windspeed < 5:

              Date/Time  LV ActivePower (kW)  Wind Speed (m/s)  \
55     01 01 2018 09:10           408.997406          4.977198
64     01 01 2018 10:40           311.050903          4.960732
65     01 01 2018 10:50           230.055496          4.603875
66     01 01 2018 11:00           233.990601          4.554534
67     01 01 2018 11:10           175.592194          4.263629
...                 ...                  ...               ...
50495  31 12 2018 18:10           341.905487          4.556026
50496  31 12 2018 18:20           399.653809          4.798016
50497  31 12 2018 18:30           372.079590          4.795230
50498  31 12 2018 18:40           411.729187          4.860610
50499  31 12 2018 18:50           471.419708          4.994785

       Theoretical_Power_Curve (KWh)  Wind Direction (°)
55                         330.417630          207.997803
64                         326.411025          229.537506
65                         244.316244          231.798492
66                         233.632780          234.105606
67                         173.573663          228.776703
...                               ...                 ...
50495                      233.953613           91.588493
50496                      287.885241           94.329742
50497                      287.241996           98.706261
50498                      302.480001           97.239059
50499                      334.719439           99.136337

[15744 rows x 5 columns]
```

10 Ans:

```
In [31]: import pandas as obj
         windpower_data = obj.read_csv("C:\\Users\\banda\\Downloads\\windpower.csv")

         print("\n Extracting the rows where wind direction > 200: \n")
         result = windpower_data[windpower_data['Wind Direction (°)'] > 200]
         print(result)

         #for x in windpower_data.index:
         #    if windpower_data.loc[x, 'Wind Direction (°)'] > 200:
         #        print(windpower_data.loc[x])
```

Output:

```
 Extracting the rows where wind direction > 200:

              Date/Time  LV ActivePower (kW)  Wind Speed (m/s)  \
0      01 01 2018 00:00           380.047791          5.311336
1      01 01 2018 00:10           453.769196          5.672167
2      01 01 2018 00:20           306.376587          5.216037
3      01 01 2018 00:30           419.645904          5.659674
4      01 01 2018 00:40           380.650696          5.577941
...                 ...                  ...               ...
50421  31 12 2018 05:50             0.000000          0.431046
50422  31 12 2018 06:00             0.000000          0.539925
50424  31 12 2018 06:20             0.000000          0.520940
50425  31 12 2018 06:30             0.000000          0.971962
50429  31 12 2018 07:10             0.000000          0.338600

       Theoretical_Power_Curve (KWh)  Wind Direction (°)
0                          416.328908          259.994904
1                          519.917511          268.641113
2                          390.900016          272.564789
3                          516.127569          271.258087
4                          491.702972          265.674286
...                               ...                 ...
50421                        0.000000          224.736999
50422                        0.000000          316.677307
50424                        0.000000          211.148102
50425                        0.000000          263.404297
50429                        0.000000          319.640289
```