

Data Collection and Preprocessing Phase

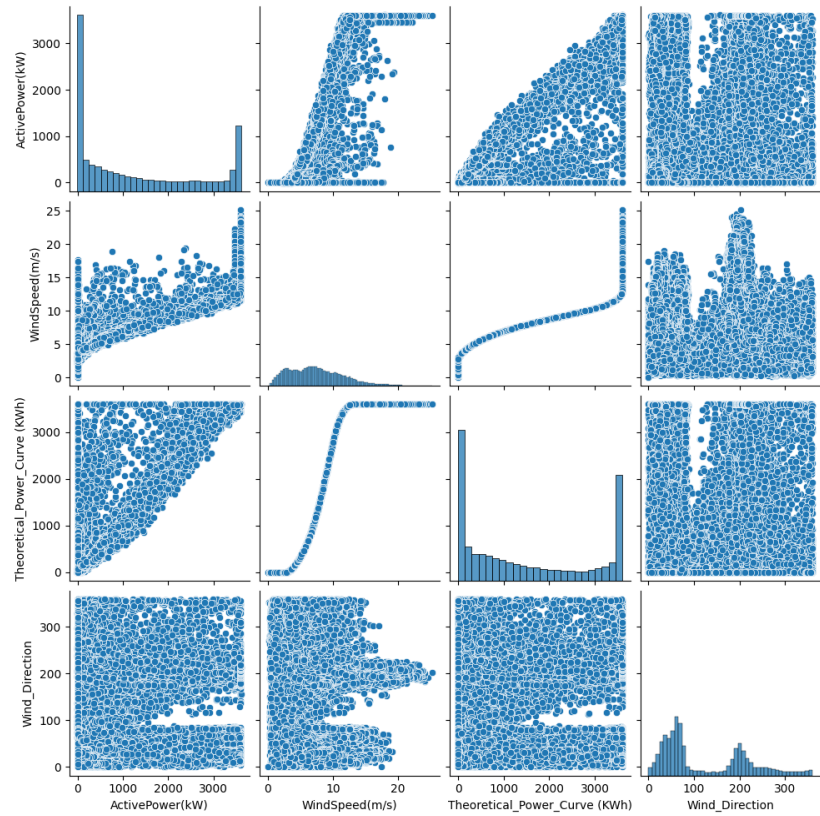
Date	18 July 2024
Team ID	XXXXXX
Project Title	Predicting The Energy Output Of Wind Turbine Based On Weather Condition
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description																																													
Data Overview	Dimenstion: 50530 rows x 5 columns Description statistics:																																													
	<table><tr><th></th><th>ActivePower(kW)</th><th>WindSpeed(m/s)</th><th>Theoretical_Power_Curve (KWh)</th><th>Wind_Direction</th></tr><tr><td>count</td><td>50530.000000</td><td>50530.000000</td><td>50530.000000</td><td>50530.000000</td></tr><tr><td>mean</td><td>1307.684332</td><td>7.557952</td><td>1492.175463</td><td>123.687559</td></tr><tr><td>std</td><td>1312.459242</td><td>4.227166</td><td>1368.018238</td><td>93.443736</td></tr><tr><td>min</td><td>-2.471405</td><td>0.000000</td><td>0.000000</td><td>0.000000</td></tr><tr><td>25%</td><td>50.677890</td><td>4.201395</td><td>161.328167</td><td>49.315437</td></tr><tr><td>50%</td><td>825.838074</td><td>7.104594</td><td>1063.776283</td><td>73.712978</td></tr><tr><td>75%</td><td>2482.507568</td><td>10.300020</td><td>2964.972462</td><td>201.696720</td></tr><tr><td>max</td><td>3618.732910</td><td>25.206011</td><td>3600.000000</td><td>359.997589</td></tr></table>		ActivePower(kW)	WindSpeed(m/s)	Theoretical_Power_Curve (KWh)	Wind_Direction	count	50530.000000	50530.000000	50530.000000	50530.000000	mean	1307.684332	7.557952	1492.175463	123.687559	std	1312.459242	4.227166	1368.018238	93.443736	min	-2.471405	0.000000	0.000000	0.000000	25%	50.677890	4.201395	161.328167	49.315437	50%	825.838074	7.104594	1063.776283	73.712978	75%	2482.507568	10.300020	2964.972462	201.696720	max	3618.732910	25.206011	3600.000000	359.997589
		ActivePower(kW)	WindSpeed(m/s)	Theoretical_Power_Curve (KWh)	Wind_Direction																																									
	count	50530.000000	50530.000000	50530.000000	50530.000000																																									
	mean	1307.684332	7.557952	1492.175463	123.687559																																									
	std	1312.459242	4.227166	1368.018238	93.443736																																									
	min	-2.471405	0.000000	0.000000	0.000000																																									
	25%	50.677890	4.201395	161.328167	49.315437																																									
	50%	825.838074	7.104594	1063.776283	73.712978																																									
	75%	2482.507568	10.300020	2964.972462	201.696720																																									
max	3618.732910	25.206011	3600.000000	359.997589																																										
Univariate Analysis	-																																													

Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies

-

Data Preprocessing Code Screenshots

Loading Data

```
path = "Data\T1.csv"
df = pd.read_csv(path)
df.head()
```

✓ 0.0s

<>:1: SyntaxWarning: invalid escape sequence '\T'
<>:1: SyntaxWarning: invalid escape sequence '\T'
C:\Users\rudra\AppData\Local\Temp\ipykernel_17852\3410472532.py:1: SyntaxWarning: invalid escape sequence '\T'
path = "Data\T1.csv"

	Date/Time	LV ActivePower (kW)	Wind Speed (m/s)	Theoretical_Power_Curve (KWh)	Wind Direction (°)
0	01 01 2018 00:00	380.047791	5.311336	416.328908	259.994904
1	01 01 2018 00:10	453.769196	5.672167	519.917511	268.641113
2	01 01 2018 00:20	306.376587	5.216037	390.900016	272.564789
3	01 01 2018 00:30	419.645905	5.659674	516.127569	271.258087
4	01 01 2018 00:40	380.650696	5.577941	491.702972	265.674286

Handling Missing Data

```
df.isnull().any()
```

✓ 0.0s

Time	False
ActivePower(kW)	False
WindSpeed(m/s)	False
Theoretical_Power_Curve (KWh)	False
Wind_Direction	False
dtype: bool	

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

✓ 0.0s

Time	0
ActivePower(kW)	0
WindSpeed(m/s)	0
Theoretical_Power_Curve (KWh)	0
Wind_Direction	0
dtype: int64	

Data Transformation	<pre>names = x.columns from sklearn.preprocessing import MinMaxScaler scale = MinMaxScaler() x_scaled = scale.fit_transform(x) x = pd.DataFrame(x_scaled, columns=names) x.head()</pre> <p>✓ 0.0s</p> <table><thead><tr><th></th><th>Theoretical_Power_Curve (KWh)</th><th>WindSpeed(m/s)</th></tr></thead><tbody><tr><td>0</td><td>0.115647</td><td>0.210717</td></tr><tr><td>1</td><td>0.144422</td><td>0.225032</td></tr><tr><td>2</td><td>0.108583</td><td>0.206936</td></tr><tr><td>3</td><td>0.143369</td><td>0.224537</td></tr><tr><td>4</td><td>0.136584</td><td>0.221294</td></tr></tbody></table>		Theoretical_Power_Curve (KWh)	WindSpeed(m/s)	0	0.115647	0.210717	1	0.144422	0.225032	2	0.108583	0.206936	3	0.143369	0.224537	4	0.136584	0.221294
	Theoretical_Power_Curve (KWh)	WindSpeed(m/s)																	
0	0.115647	0.210717																	
1	0.144422	0.225032																	
2	0.108583	0.206936																	
3	0.143369	0.224537																	
4	0.136584	0.221294																	
Feature Engineering	-																		
Save Processed Data	-																		