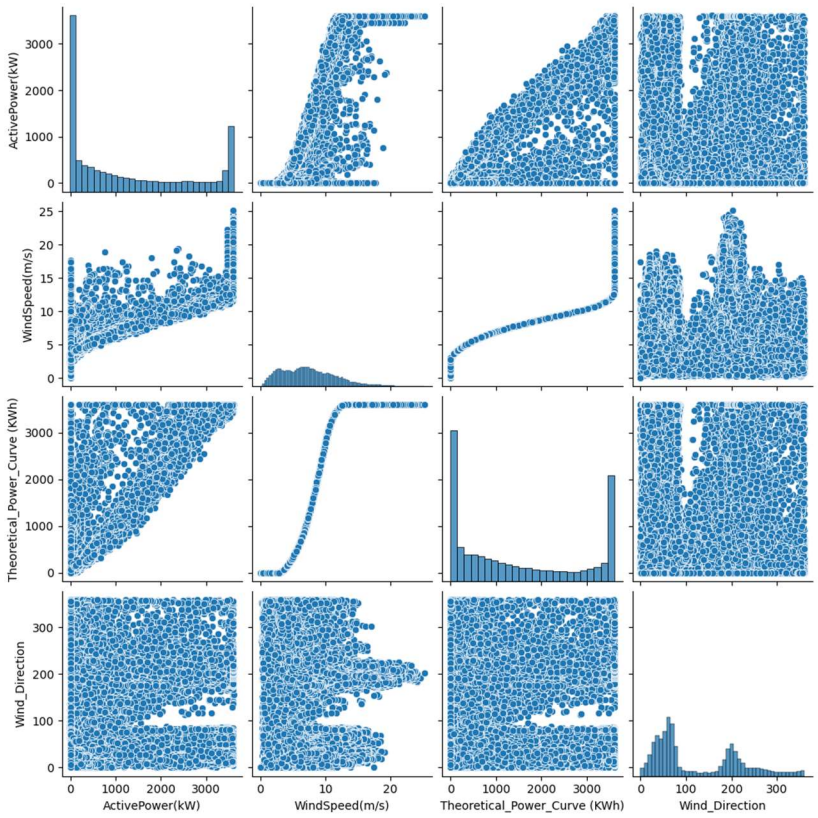
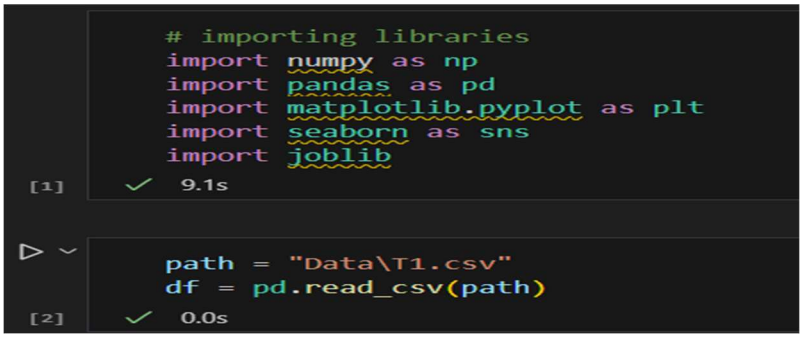
Data Collection and Preprocessing Phase

Date	15 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	Basic statistics is include of the mean targeted column ActivePower(kW) is 1307.68, dimensions of this dataset is (50530, 5).
Univariate Analysis	Mean of perticular column in dataset : 1. ActivePower(kW) : 1307.68 2. WindSpeed(m/s) : 7.55 3. Theoretical_Power_Curve (KWh) : 1492.17 4. Wind_Direction : 123.68
Bivariate Analysis	Relationships between two variables 1.Theoretical_Power_Curve(KWh) and ActivePower(kW):0.94 2. WindSpeed(m/s) and ActivePower(kW) : 0.912774

Outliers and Anomalies	
Data Preprocessing Code Screenshots	
Loading Data	
Handling Missing Data	There is no null values.

<table>
<tr>
<td></td>
<td>

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

[9] ✓ 0.0s

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

</td>
</tr>
<tr>
<td>Data Transformation</td>
<td>

```
# Standard Scalling  - mean=0 and std=1
# Minmax Scalling  - Values after scalling are between 0
names = x.columns
from sklearn.preprocessing import MinMaxScaler # import
scale = MinMaxScaler() # here also you want to change in
x_scaled = scale.fit_transform(x)
x = pd.DataFrame(x_scaled, columns=names)
x.head()
```

[12]

	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

</td>
</tr>
<tr>
<td>Feature Engineering</td>
<td>Code for modifying existing ones.

```
df.rename(columns={"Date/Time": "Time",
                   'LV ActivePower (kW)': 'ActivePower(kW)',
                   'Wind Speed (m/s)': 'WindSpeed(m/s)',
                   'Wind Direction (°)': 'Wind_Direction'},
          inplace=True)
```

[6] ✓ 0.0s

</td>
</tr>
<tr>
<td>Save Processed Data</td>
<td>

```
y = df[['ActivePower(kW)']]
x = df[['Theoretical_Power_Curve (KWh)','WindSpeed(m/s)']]
```

[13] ✓ 0.0s

</td>
</tr>
</table>