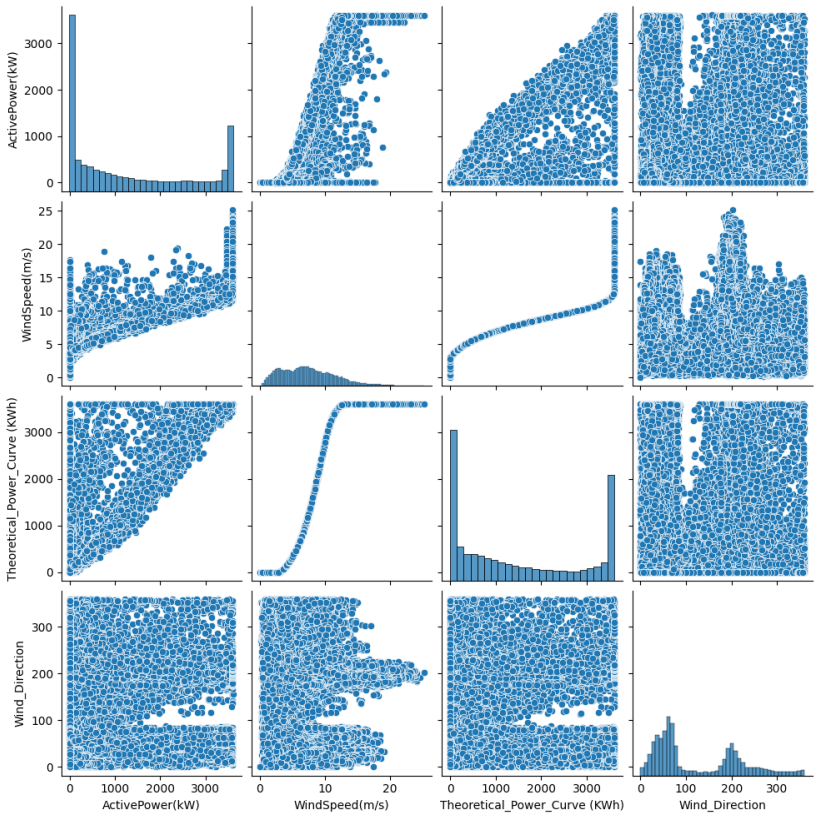
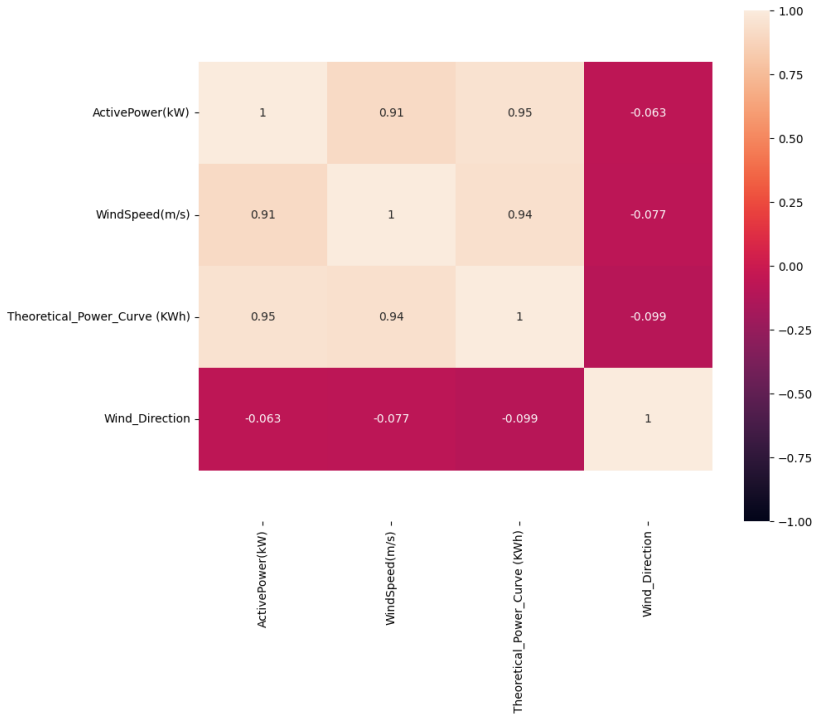
Data Collection and Preprocessing Phase

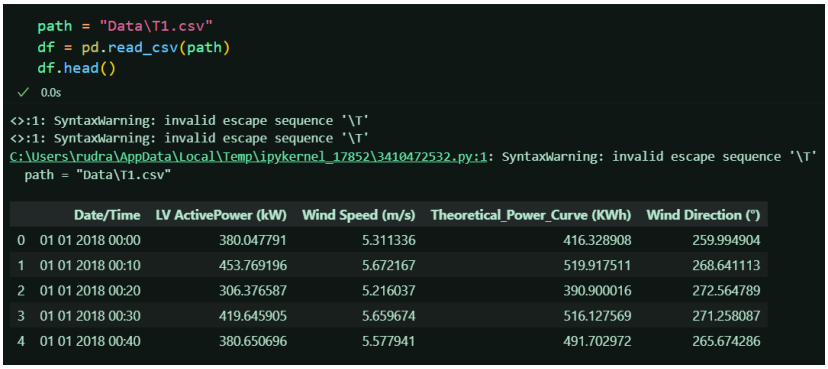
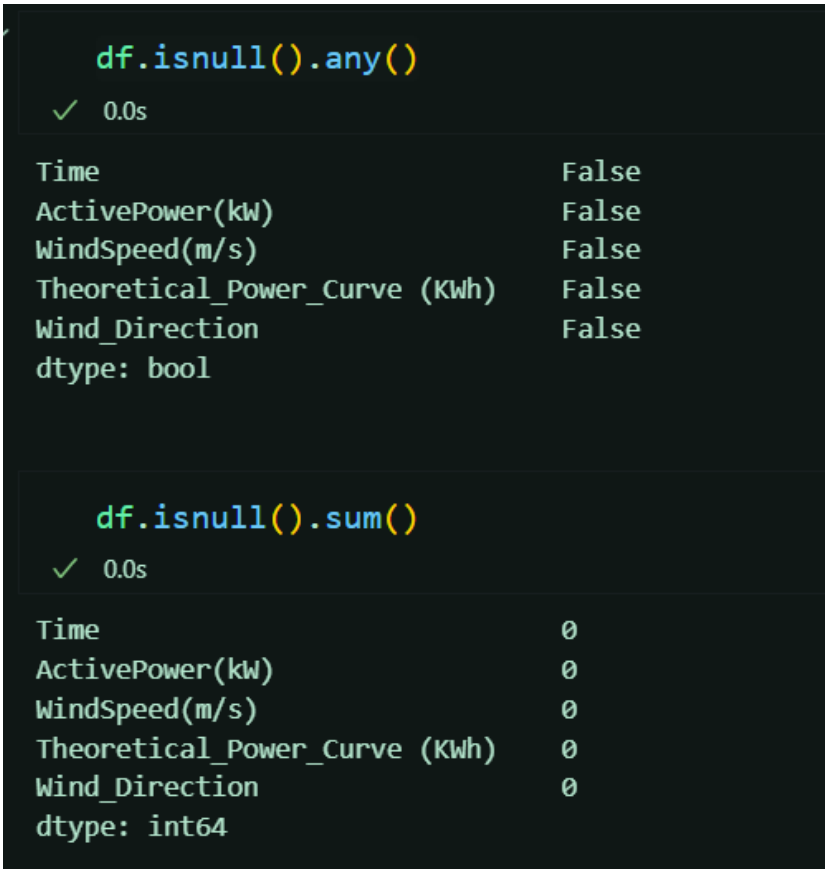
Date	18 July 2024
Team ID	xxxxxx
Project Title	Predicting The Energy Output Of Wind Turbine Based On Weather Condition
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

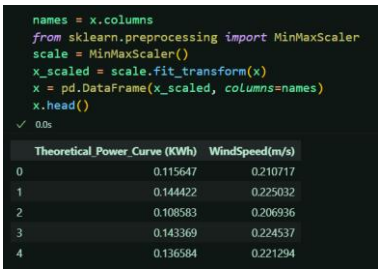
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	Dimenstion: 50530 rows x 5 columns Description statistics: 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	
Handling Missing Data	

	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

Data Transformation	
Feature Engineering	-
Save Processed Data	-