Data Collection and Preprocessing Phase

Date	22 June 2024
Team ID	739657
Project Title	Prediction Of Full Load Electrical Power Output Of A Base Load Operated Combined Cycle Power Plant Using Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

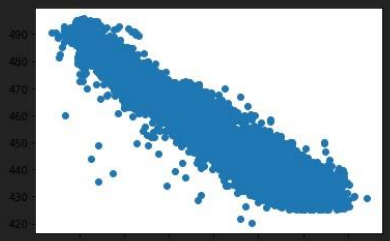
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	Descriptive statistics:
Univariate Analysis	

	AT	V	AP	RH	PE
count	9568.000000	9568.000000	9568.000000	9568.000000	9568.000000
mean	19.651231	54.305804	1013.259078	73.308978	454.365009
std	7.452473	12.707893	5.938784	14.600269	17.066995
min	1.810000	25.360000	992.890000	25.560000	420.260000
25%	13.510000	41.740000	1009.100000	63.327500	439.750000
50%	20.345000	52.080000	1012.940000	74.975000	451.550000
75%	25.720000	66.540000	1017.260000	84.830000	468.430000
max	37.110000	81.560000	1033.300000	100.160000	495.760000

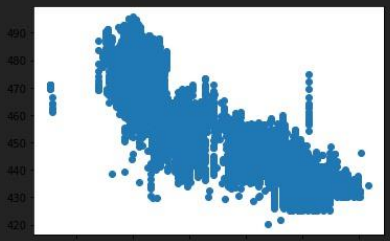
```
# draw a scatter plot diagram
# The x-axis represents AT, and the y-axis represents PE.
plt.scatter(data['AT'],data['PE'])
```

```
<matplotlib.collections.PathCollection at 0x1c612733b80>
```



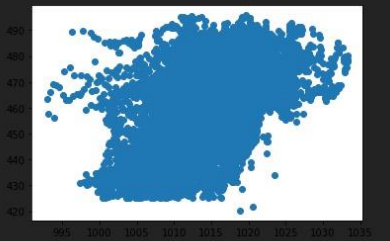

```
# draw a scatter plot diagram
# The x-axis represents V, and the y-axis represents PE.
plt.scatter(data['V'],data['PE'])
```

```
<matplotlib.collections.PathCollection at 0x1c6127f1370>
```

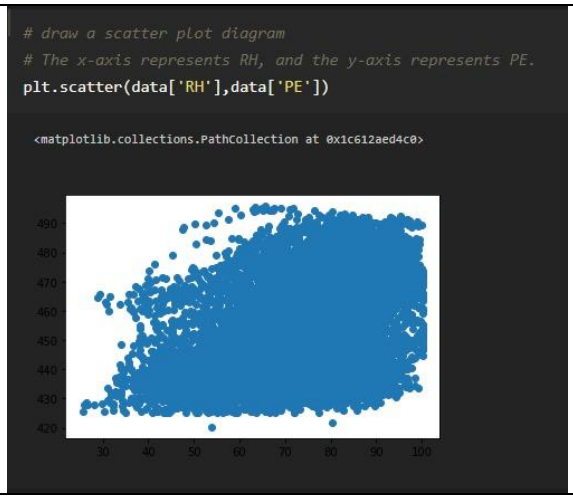
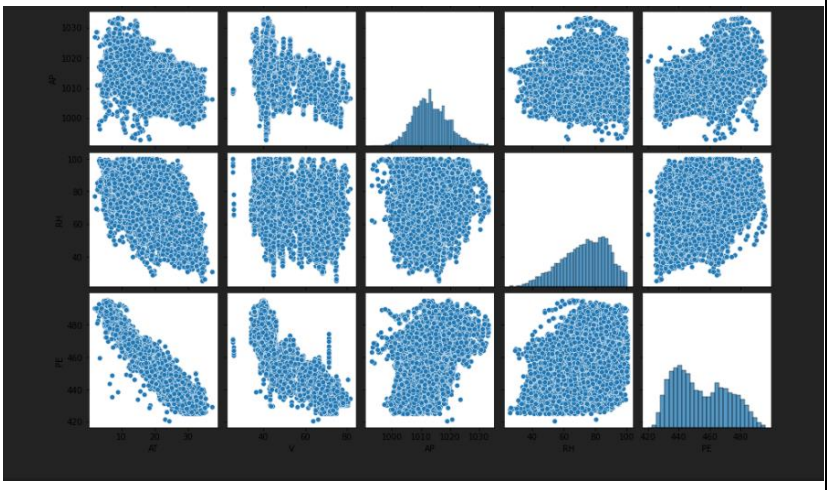


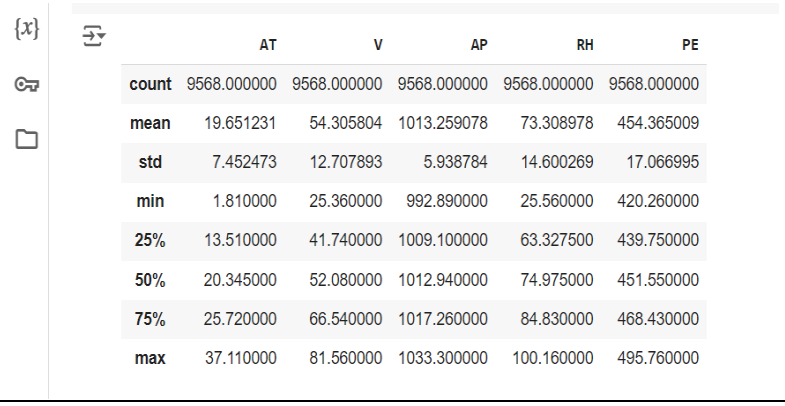

```
# draw a scatter plot diagram
# The x-axis represents AP, and the y-axis represents PE.
plt.scatter(data['AP'],data['PE'])
```

```
<matplotlib.collections.PathCollection at 0x1c61290c370>
```



\

	
Bivariate Analysis	 

Outliers and Anomalies	-
Data Preprocessing Code Screenshots	
Loading Data	
Handling Missing Data	No missing values
Feature Engineering	Attached the codes in final submission.
Save Processed Data	-

	AT	V	AP	RH	PE
count	9568.000000	9568.000000	9568.000000	9568.000000	9568.000000
mean	19.651231	54.305804	1013.259078	73.308978	454.365009
std	7.452473	12.707893	5.938784	14.600269	17.066995
min	1.810000	25.360000	992.890000	25.560000	420.260000
25%	13.510000	41.740000	1009.100000	63.327500	439.750000
50%	20.345000	52.080000	1012.940000	74.975000	451.550000
75%	25.720000	66.540000	1017.260000	84.830000	468.430000
max	37.110000	81.560000	1033.300000	100.160000	495.760000