Data Collection and Preprocessing Phase

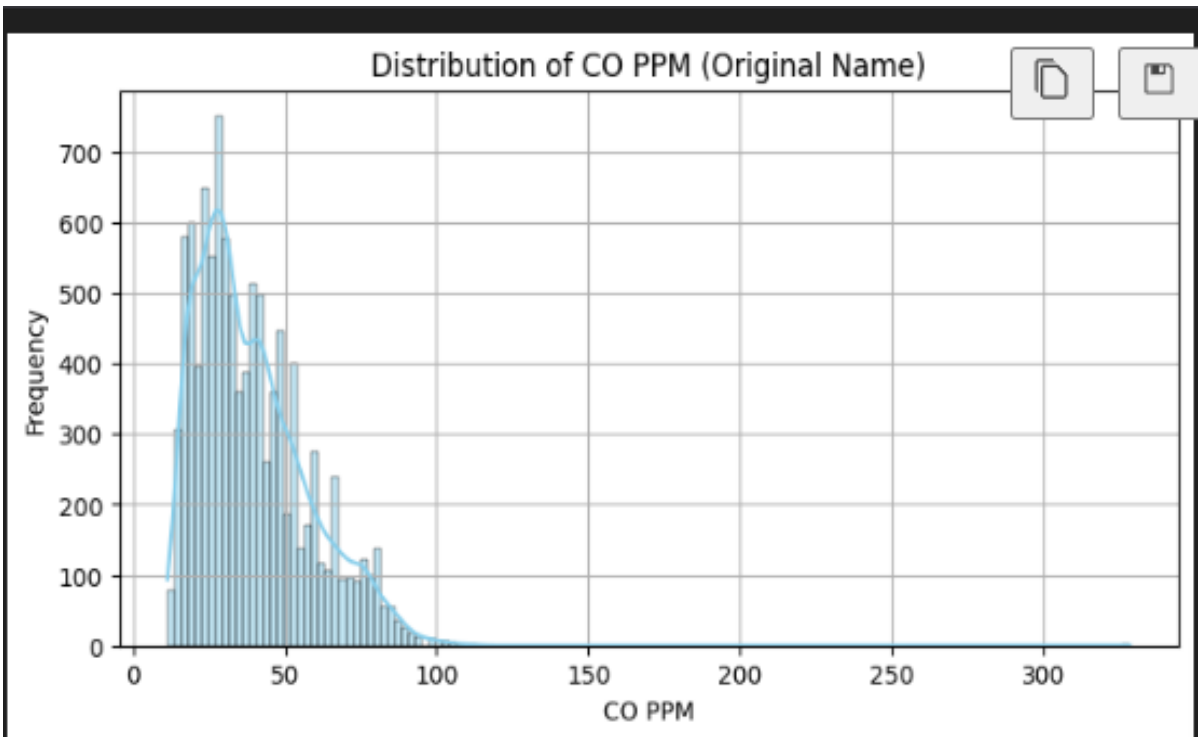
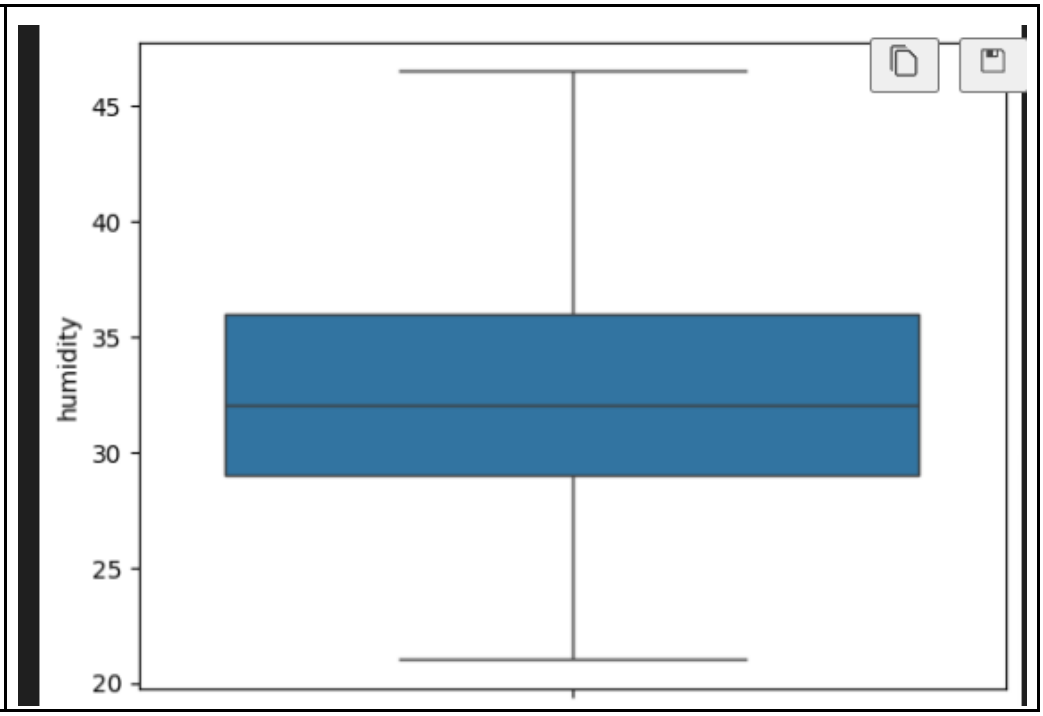
Date	9 JULY 2024
Team ID	739818
Project Title	EcoForecast:AI-powered prediction of carbon monoxide levels
Maximum Marks	6 Marks

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	Dimension: [10308 rows × 8 columns] Descriptive statistics:

	timestamp	Year	Month	Day	Hour	temp	humidity	ppm
0	2023-06-09 10:46:48+05:30	2023	6	9	10	38	38.0	24.01
1	2023-06-09 10:47:49+05:30	2023	6	9	10	38	36.0	22.39
2	2023-06-09 10:48:49+05:30	2023	6	9	10	38	36.0	21.62
3	2023-06-09 10:49:50+05:30	2023	6	9	10	38	36.0	21.62
4	2023-06-09 10:50:50+05:30	2023	6	9	10	38	36.0	21.62
...	...	...	...	...	...	...	...	...
10303	2023-06-17 02:51:13+05:30	2023	6	17	2	28	32.0	25.71
10304	2023-06-17 02:52:13+05:30	2023	6	17	2	28	32.0	25.71
10305	2023-06-17 02:53:14+05:30	2023	6	17	2	28	32.0	25.71
10306	2023-06-17 02:54:14+05:30	2023	6	17	2	28	32.0	28.43
10307	2023-06-17 02:55:15+05:30	2023	6	17	2	28	32.0	30.36

Bivariate Analysis	
Outliers	

Data Preprocessing Code Screenshots	
Loading Data	

	temp	humidity	ppm
count	10308.000000	10308.000000	10308.000000
mean	38.297051	32.617288	39.145906
std	4.053829	5.793688	18.363310
min	28.000000	21.000000	11.270000
25%	35.000000	29.000000	24.850000
50%	39.000000	32.000000	34.520000
75%	41.000000	36.000000	49.670000
max	45.000000	55.000000	328.600000

Data Transformation	0.75 36.0 0.25 29.0 Name: humidity, dtype: float64 36.0 29.0 7.0 46.5 18.5
Feature Engineering	Attached the codes in final submission.