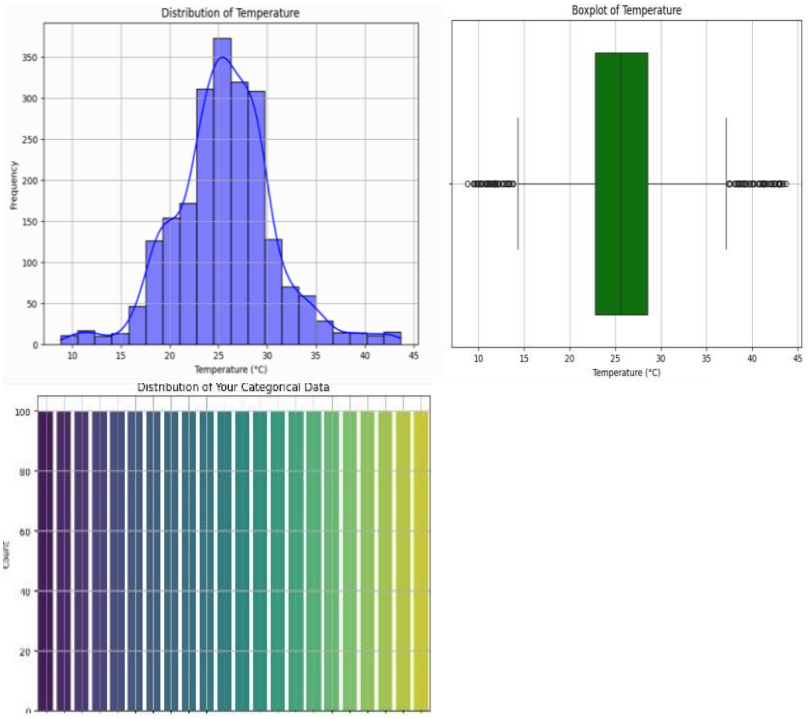
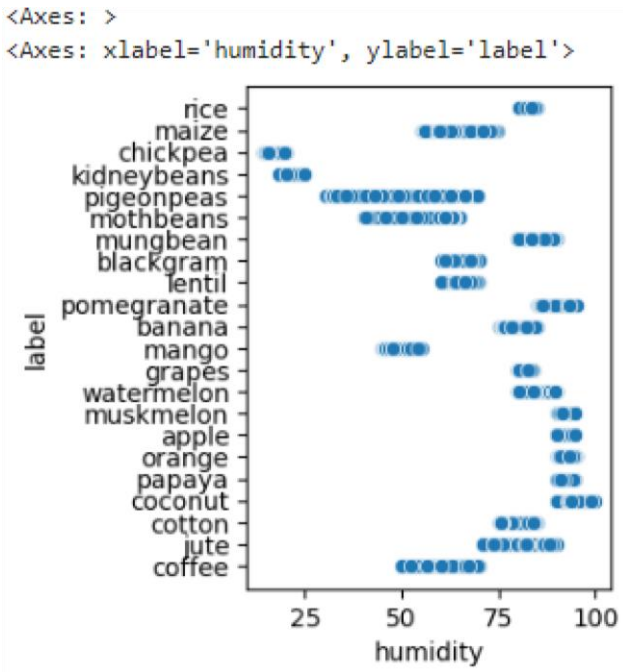
Data Collection and Preprocessing Phase

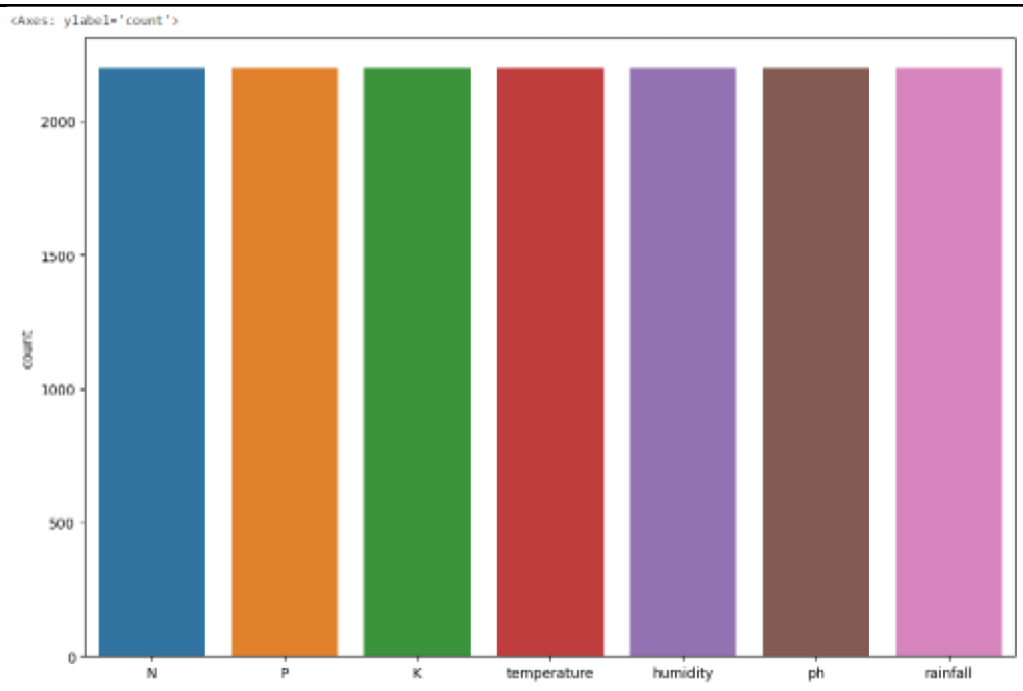
Date	21 June 2024
Team ID	739768
Project Title	Opticrop:Smart Agricultural Prodution Optimization Engine
Maximum Marks	6 Marks

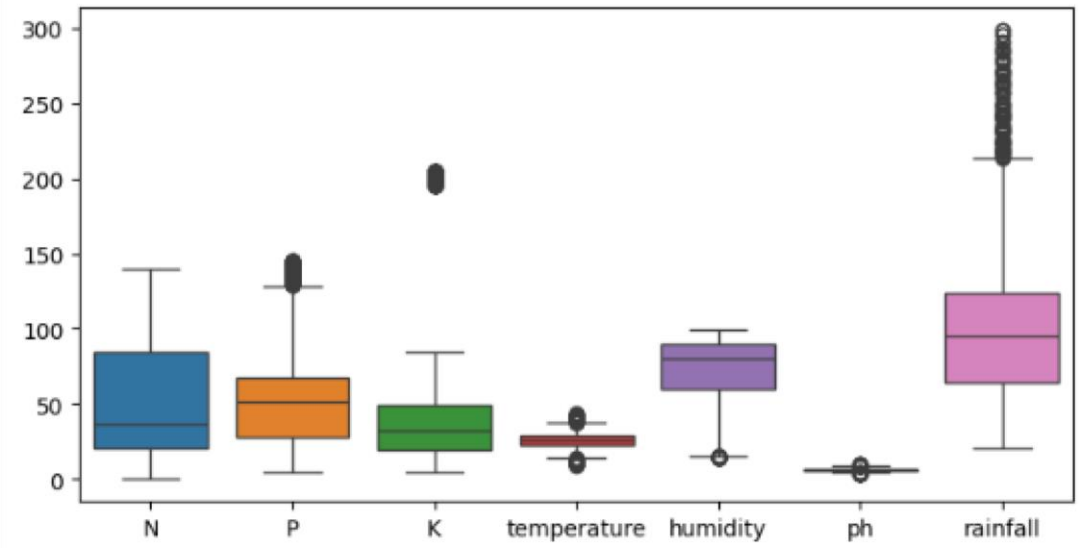
Data Exploration and Preprocessing Report

The purpose of this report is to outline the key steps and findings from the data exploration and preprocessing phase for the Opticrop project. This project aims to develop a Smart Agricultural Production Optimization Engine using data-driven approaches.the data sources are Agricultural sensor networks, satellite imagery, weather stations. The data exploration and preprocessing phases for the Opticrop project have been crucial in ensuring the quality and reliability of input data for subsequent modeling and optimization tasks. By addressing issues such as missing values, outliers, and data inconsistencies, we have prepared a clean and structured dataset ready for machine learning and analytics.

Section	Description
Data Overview	Dimension: 2200 rows × 8 columns Descriptive statistics: `df.describe()` N P K temperature humidity ph rainfall count 2200.000000 2200.000000 2200.000000 2200.000000 2200.000000 2200.000000 2200.000000 mean 50.551818 53.362727 48.149091 25.616244 71.481779 6.469480 103.463655 std 36.917334 32.985883 50.647931 5.063749 22.263812 0.773938 54.958389 min 0.000000 5.000000 5.000000 8.825675 14.258040 3.504752 20.211267 25% 21.000000 28.000000 20.000000 22.769375 60.261953 5.971693 64.551686 50% 37.000000 51.000000 32.000000 25.598693 80.473146 6.425045 94.867624 75% 84.250000 68.000000 49.000000 28.561654 89.948771 6.923643 124.267508 max 140.000000 145.000000 205.000000 43.675493 99.981876 9.935091 298.560117

Univariate Analysis	
Bivariate Analysis	

Multivariate Analysis	

Outliers and Anomalies	`plt.figure(figsize=(8,4))`
	`sns.boxplot(df)`
	
	`<Figure size 800x400 with 0 Axes>`
	`<Axes: >`
	 
	
	`Q1=df['P'].quantile(0.25)`
	`Q3=df['P'].quantile(0.75)`
	`IQR=Q3-Q1`
	`filter=(df['P']>=Q1-1.5*IQR) & (df['P']<=Q3+1.5*IQR)`
	`df=df.loc[filter]`

Data Preprocessing Code Sc reenshots

Loading Data

```
df = pd.read_csv('/content/Crop_recommendation.csv')
df.head()
```

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Handling Missing Data

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

```
N              0
P              0
K              0
temperature    0
humidity       0
ph             0
rainfall       0
label          0
dtype: int64
```

Data Transformation

-

Feature Engineering

Attached the codes in final submission.

Save Processed Data

-