Data Collection and Preprocessing Phase

Date	20 June 2025
Project Title	Rising Waters: A Machine Learning Approach to Flood Prediction
Team ID	SWTID1749705847
Maximum Marks	6 Marks

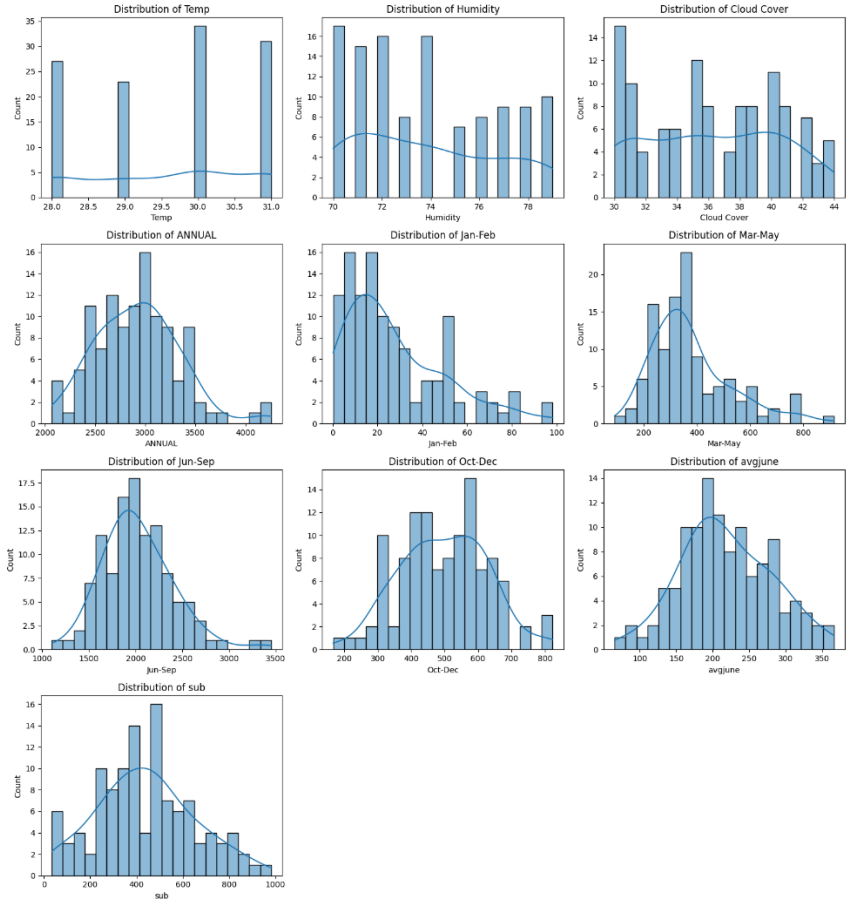
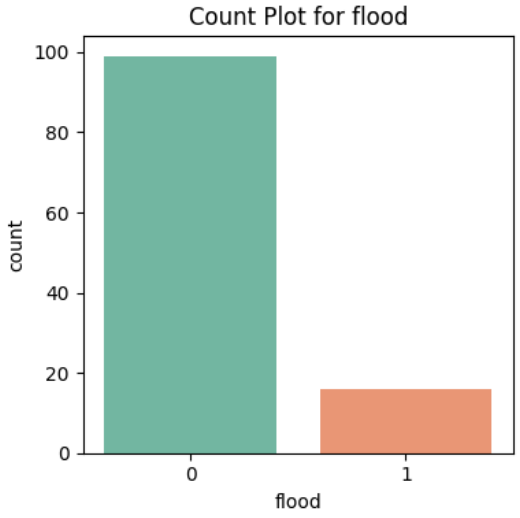
Data Exploration and Preprocessing Report

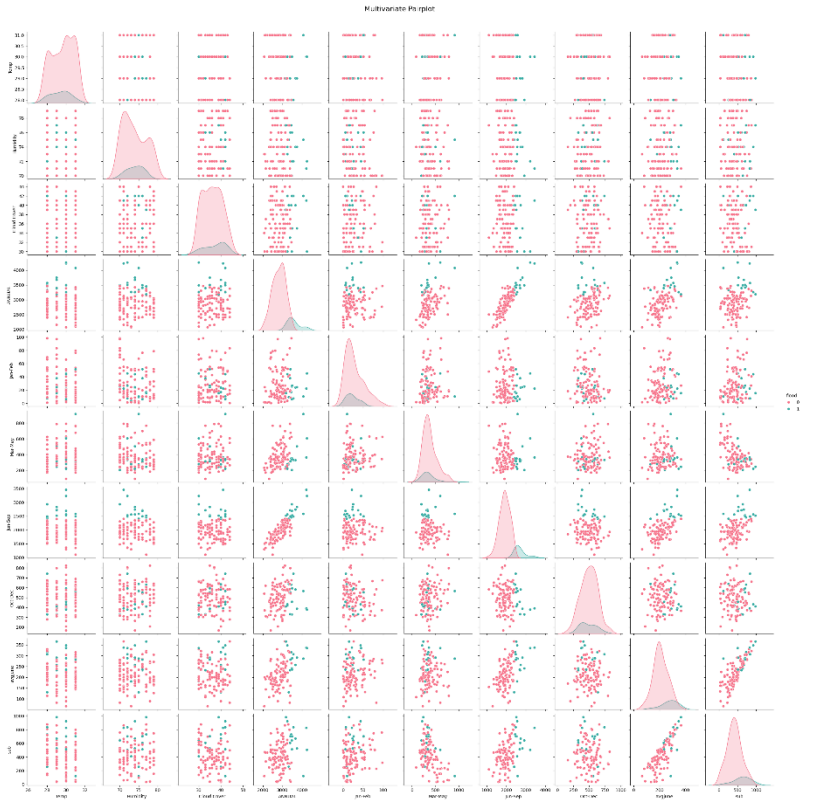
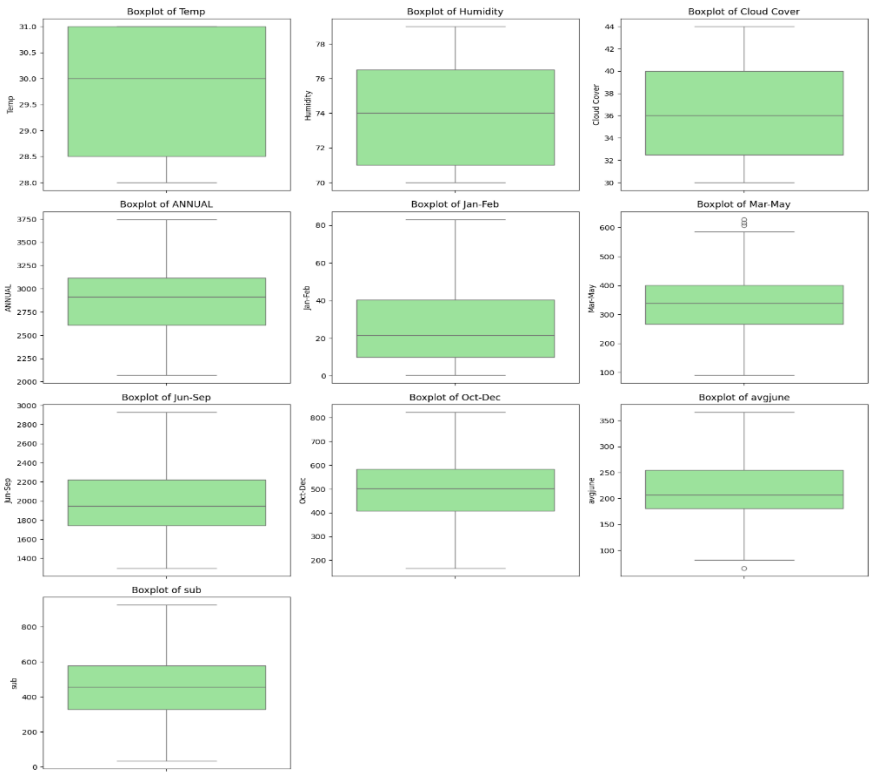
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature scaling. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong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

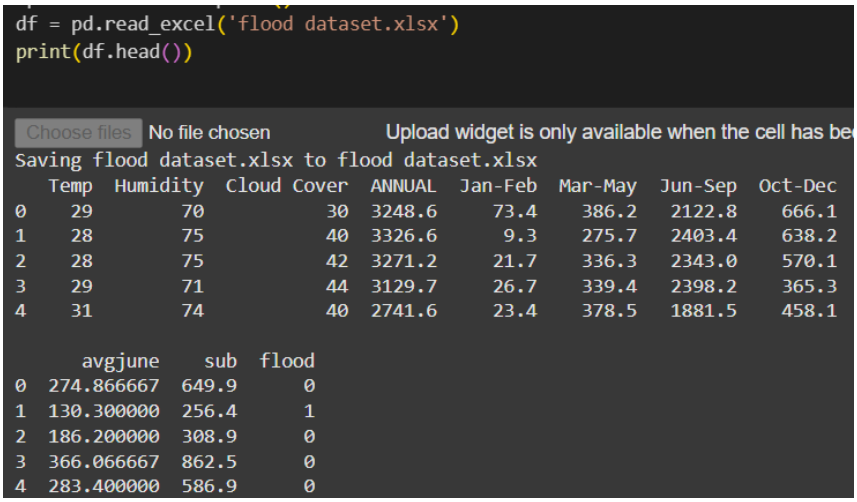
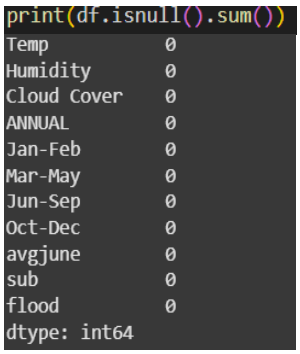
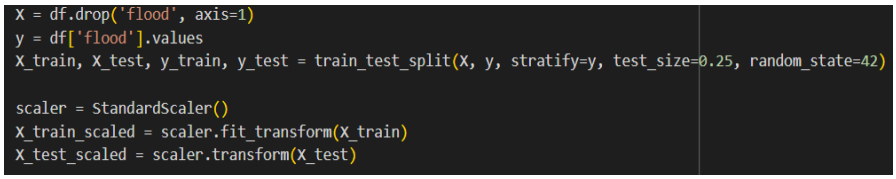
Section	Description
Data Overview	Dimension: 116 rows × 11 columns Descriptive statistics:

	Temp	Humidity	Cloud Cover	ANNUAL	Jan-Feb	\
count	115.000000	115.000000	115.000000	115.000000	115.000000	
mean	29.600000	73.852174	36.286957	2925.487826	27.739130	
std	1.122341	2.947623	4.330158	422.112193	22.361032	
min	28.000000	70.000000	30.000000	2068.800000	0.300000	
25%	29.000000	71.000000	32.500000	2627.900000	10.250000	
50%	30.000000	74.000000	36.000000	2937.500000	20.500000	
75%	31.000000	76.000000	40.000000	3164.100000	41.600000	
max	31.000000	79.000000	44.000000	4257.800000	98.100000	

	Mar-May	Jun-Sep	Oct-Dec	avgjune	sub	flood
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	377.253913	2022.840870	497.636522	218.100870	439.801739	0.139130
std	151.091850	386.254397	129.860643	62.547597	210.438813	0.347597
min	89.900000	1104.300000	166.600000	65.600000	34.200000	0.000000
25%	276.750000	1768.850000	407.450000	179.666667	295.000000	0.000000
50%	342.000000	1948.700000	501.500000	211.033333	430.600000	0.000000
75%	442.300000	2242.900000	584.550000	263.833333	577.650000	0.000000
max	915.200000	3451.300000	823.300000	366.066667	982.700000	1.000000

Univariate Analysis	
Categorical Count Plot	

Multivariate Analysis	
Outliners and Anomalies	

Data Preprocessing Code Screenshots	
Loading Data	
Handling Missing Data	
Data Transformation	
Feature Scaling	Attached the codes in final submission
Save Processed Data	-