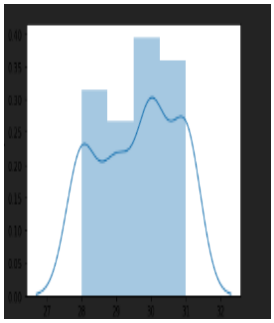
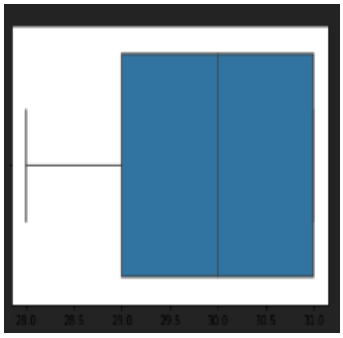
Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	740004
Project Title	Rising Waters:Machine Learning Approch To Flood Prediction
Maximum Marks	6 Marks

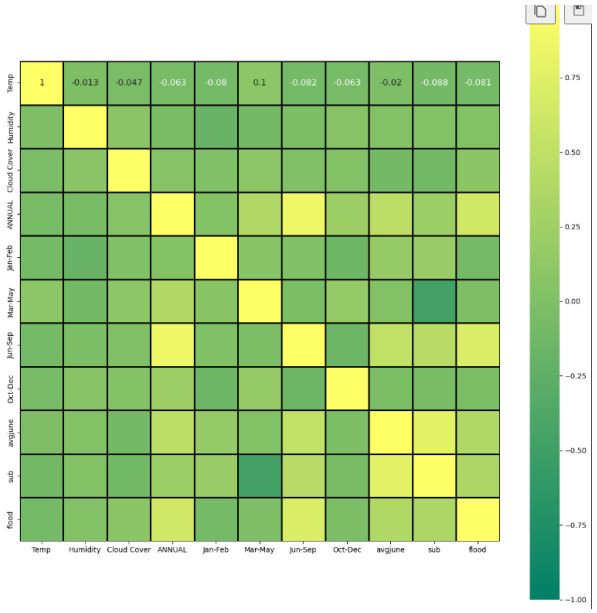
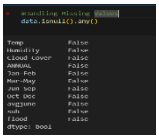
Data Exploration and Preprocessing Template

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	Dimensions 115rows x 9columns
Univariate Analysis	 

	Temp	Humidity	Cloud Cover	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec	avgjune
0	29	70	30.0	3248.6	73.4	386.2	2122.8	666.1	274.866667
1	28	75	40.0	3326.6	9.3	275.7	2403.4	638.2	130.300000
2	28	75	42.0	3271.2	21.7	336.3	2343.0	570.1	186.200000
3	29	71	44.0	3129.7	26.7	339.4	2398.2	365.3	366.066667
4	31	74	40.0	2741.6	23.4	378.5	1881.5	458.1	283.400000
...	...	...	...	...	...	...	...	...	...
110	28	71	30.0	3035.1	66.2	313.5	2209.1	446.3	262.833333
111	29	71	37.0	2151.1	18.3	287.4	1535.6	309.8	143.433333
112	30	74	42.0	3255.4	43.9	218.5	2561.2	431.8	347.566667
113	31	71	31.0	3046.4	14.9	364.5	2164.8	502.1	151.466667
114	28	71	34.0	2600.6	8.9	465.9	1514.7	611.1	187.866667

115 rows × 9 columns

Bivariate Analysis	-
Multivariate Analysis	
Outliers and Anomalies	-
Data Preprocessing Code Screenshots	
Loading Data	`data=pd.read_excel("flood dataset.xlsx")`
Handling Missing Values	

Feature Scaling	`#import Standardscaler`
	`from sklearn.preprocessing import StandardScaler`
	`#create object to Standardscaler class`
	`sc=StandardScaler()`
	`x_train=sc.fit_transform(x_train)`
	`x_test=sc.fit_transform(x_test)`
	
	`#import dump class from joblib`
	`from joblib import dump`
	`dump(sc,"transform.save")`

Feature Engineering	Attached code in final submission.
Save Processed Data	-