

Data Collection and Preprocessing Phase

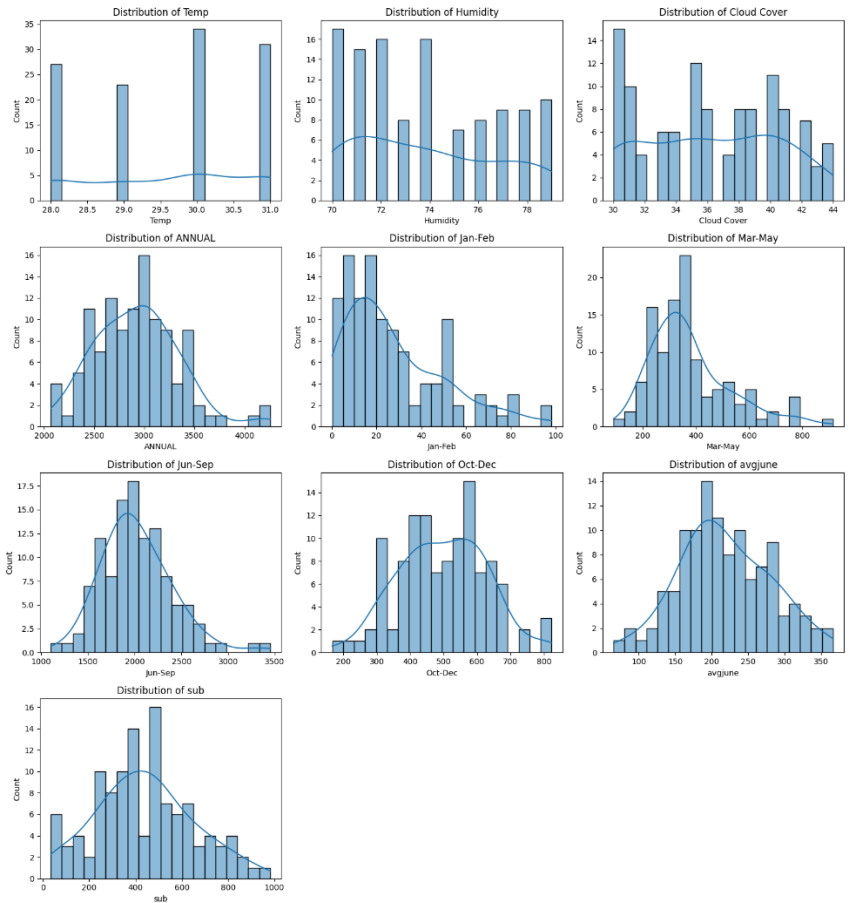
Date	20 June 2025
Project Title	Rising Waters: A Machine Learning Approach to Flood Prediction
Team ID	SWTID1749705847
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

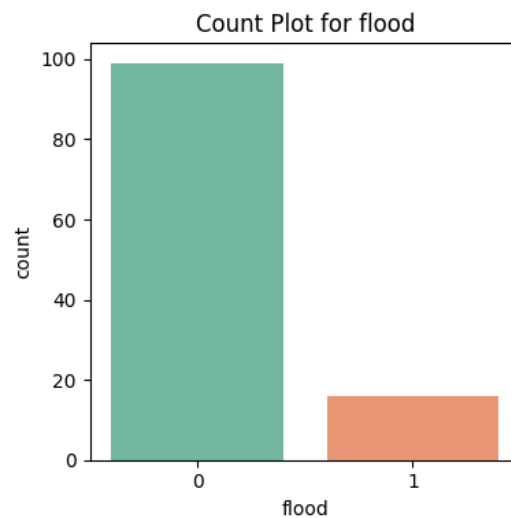
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature scaling. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong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

Section	Description																																																															
Data Overview	<u>Dimension:</u> 116 rows × 11 columns																																																															
	<u>Descriptive statistics:</u>																																																															
	<table><tr><td></td><td>Temp</td><td>Humidity</td><td>Cloud Cover</td><td>ANNUAL</td><td>Jan-Feb</td><td>\</td></tr><tr><td>count</td><td>115.000000</td><td>115.000000</td><td>115.000000</td><td>115.000000</td><td>115.000000</td><td></td></tr><tr><td>mean</td><td>29.600000</td><td>73.852174</td><td>36.286957</td><td>2925.487826</td><td>27.739130</td><td></td></tr><tr><td>std</td><td>1.122341</td><td>2.947623</td><td>4.330158</td><td>422.112193</td><td>22.361032</td><td></td></tr><tr><td>min</td><td>28.000000</td><td>70.000000</td><td>30.000000</td><td>2068.800000</td><td>0.300000</td><td></td></tr><tr><td>25%</td><td>29.000000</td><td>71.000000</td><td>32.500000</td><td>2627.900000</td><td>10.250000</td><td></td></tr><tr><td>50%</td><td>30.000000</td><td>74.000000</td><td>36.000000</td><td>2937.500000</td><td>20.500000</td><td></td></tr><tr><td>75%</td><td>31.000000</td><td>76.000000</td><td>40.000000</td><td>3164.100000</td><td>41.600000</td><td></td></tr><tr><td>max</td><td>31.000000</td><td>79.000000</td><td>44.000000</td><td>4257.800000</td><td>98.100000</td><td></td></tr></table>		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	
		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																																																											
<table><tr><td></td><td>Mar-May</td><td>Jun-Sep</td><td>Oct-Dec</td><td>avgjune</td><td>sub</td><td>flood</td></tr><tr><td>count</td><td>115.000000</td><td>115.000000</td><td>115.000000</td><td>115.000000</td><td>115.000000</td><td>115.000000</td></tr><tr><td>mean</td><td>377.253913</td><td>2022.840870</td><td>497.636522</td><td>218.100870</td><td>439.801739</td><td>0.139130</td></tr><tr><td>std</td><td>151.091850</td><td>386.254397</td><td>129.860643</td><td>62.547597</td><td>210.438813</td><td>0.347597</td></tr><tr><td>min</td><td>89.900000</td><td>1104.300000</td><td>166.600000</td><td>65.600000</td><td>34.200000</td><td>0.000000</td></tr><tr><td>25%</td><td>276.750000</td><td>1768.850000</td><td>407.450000</td><td>179.666667</td><td>295.000000</td><td>0.000000</td></tr><tr><td>50%</td><td>342.000000</td><td>1948.700000</td><td>501.500000</td><td>211.033333</td><td>430.600000</td><td>0.000000</td></tr><tr><td>75%</td><td>442.300000</td><td>2242.900000</td><td>584.550000</td><td>263.833333</td><td>577.650000</td><td>0.000000</td></tr><tr><td>max</td><td>915.200000</td><td>3451.300000</td><td>823.300000</td><td>366.066667</td><td>982.700000</td><td>1.000000</td></tr></table>		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	
	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																																																										

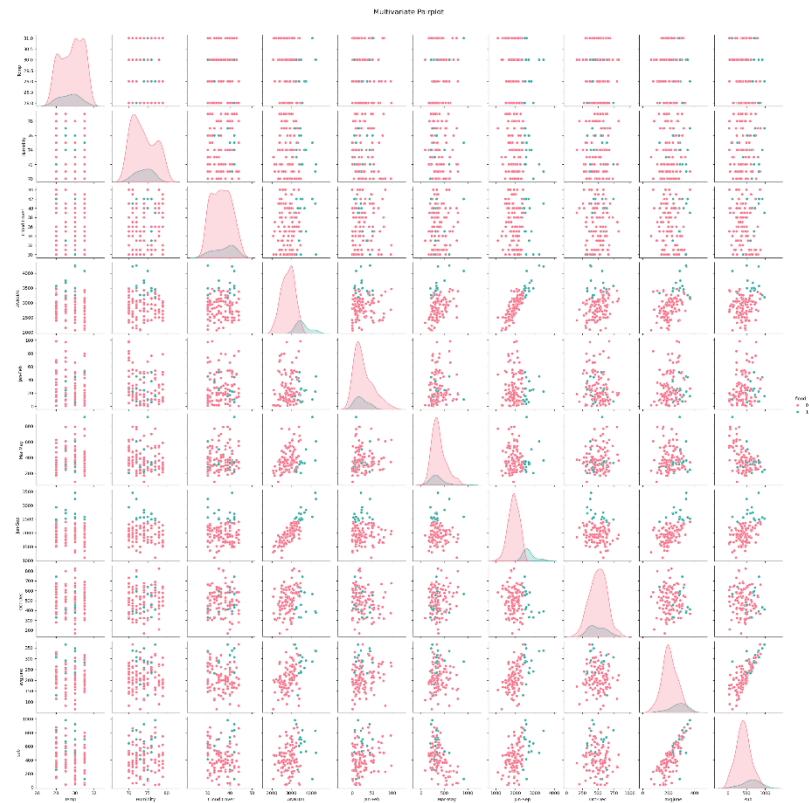
Univariate Analysis



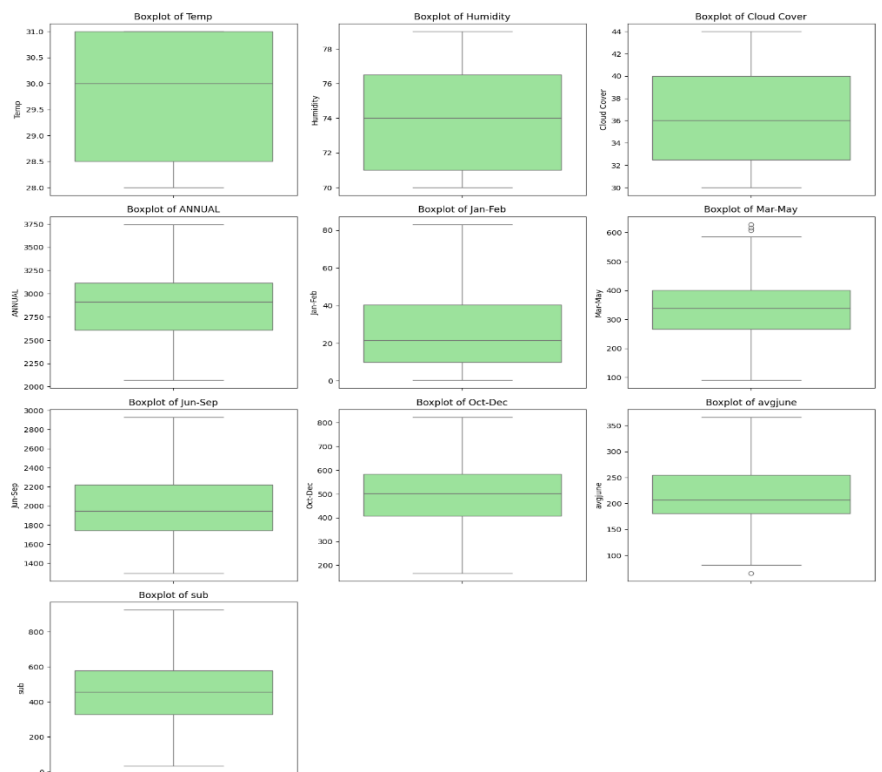
Categorical Count Plot



Multivariate Analysis



Outliners and Anomalies



Data Preprocessing Code Screenshots

Loading Data

```
df = pd.read_excel('flood dataset.xlsx')
print(df.head())
```

Choose files No file chosen Upload widget is only available when the cell has been edited

Saving flood dataset.xlsx to flood dataset.xlsx

	Temp	Humidity	Cloud Cover	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
0	29	70	30	3248.6	73.4	386.2	2122.8	666.1
1	28	75	40	3326.6	9.3	275.7	2403.4	638.2
2	28	75	42	3271.2	21.7	336.3	2343.0	570.1
3	29	71	44	3129.7	26.7	339.4	2398.2	365.3
4	31	74	40	2741.6	23.4	378.5	1881.5	458.1

	avgjune	sub	flood
0	274.866667	649.9	0
1	130.300000	256.4	1
2	186.200000	308.9	0
3	366.066667	862.5	0
4	283.400000	586.9	0

Handling Missing Data

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

```
Temp      0
Humidity  0
Cloud Cover 0
ANNUAL    0
Jan-Feb   0
Mar-May   0
Jun-Sep   0
Oct-Dec   0
avgjune   0
sub       0
flood     0
dtype: int64
```

Data Transformation

```
X = df.drop('flood', axis=1)
y = df['flood'].values
X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y, test_size=0.25, random_state=42)

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Feature Scaling

Attached the codes in final submission

Save Processed Data

-