KARIMGANJ Weather Dataset EDA

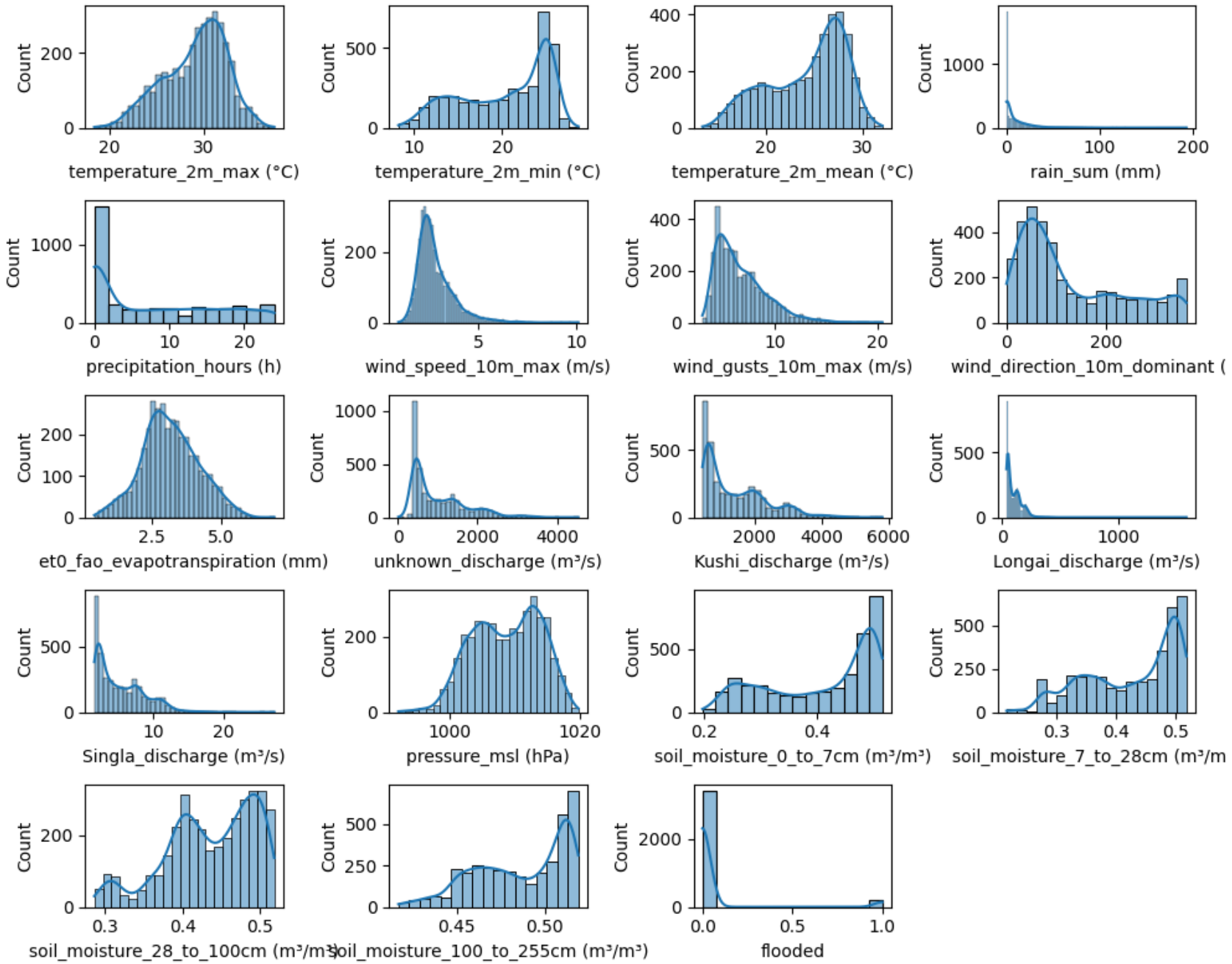
Dataset Information

```
(3654, 19)
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 3654 entries, 2015-01-01 to 2025-01-01
Data columns (total 19 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   temperature_2m_max (°C)             3654 non-null   float64
 1   temperature_2m_min (°C)             3654 non-null   float64
 2   temperature_2m_mean (°C)            3654 non-null   float64
 3   rain_sum (mm)                       3654 non-null   float64
 4   precipitation_hours (h)             3654 non-null   int64
 5   wind_speed_10m_max (m/s)            3654 non-null   float64
 6   wind_gusts_10m_max (m/s)            3654 non-null   float64
 7   wind_direction_10m_dominant (°)     3654 non-null   int64
 8   et0_fao_evapotranspiration (mm)     3654 non-null   float64
 9   unknown_discharge (m³/s)            3654 non-null   float64
 10  Kushi_discharge (m³/s)              3654 non-null   float64
 11  Longai_discharge (m³/s)             3654 non-null   float64
 12  Singla_discharge (m³/s)             3654 non-null   float64
 13  pressure_msl (hPa)                  3654 non-null   float64
 14  soil_moisture_0_to_7cm (m³/m³)      3654 non-null   float64
 15  soil_moisture_7_to_28cm (m³/m³)     3654 non-null   float64
 16  soil_moisture_28_to_100cm (m³/m³)   3654 non-null   float64
 17  soil_moisture_100_to_255cm (m³/m³)  3654 non-null   float64
 18  flooded                             3654 non-null   int64
dtypes: float64(16), int64(3)
memory usage: 570.9 KB
```

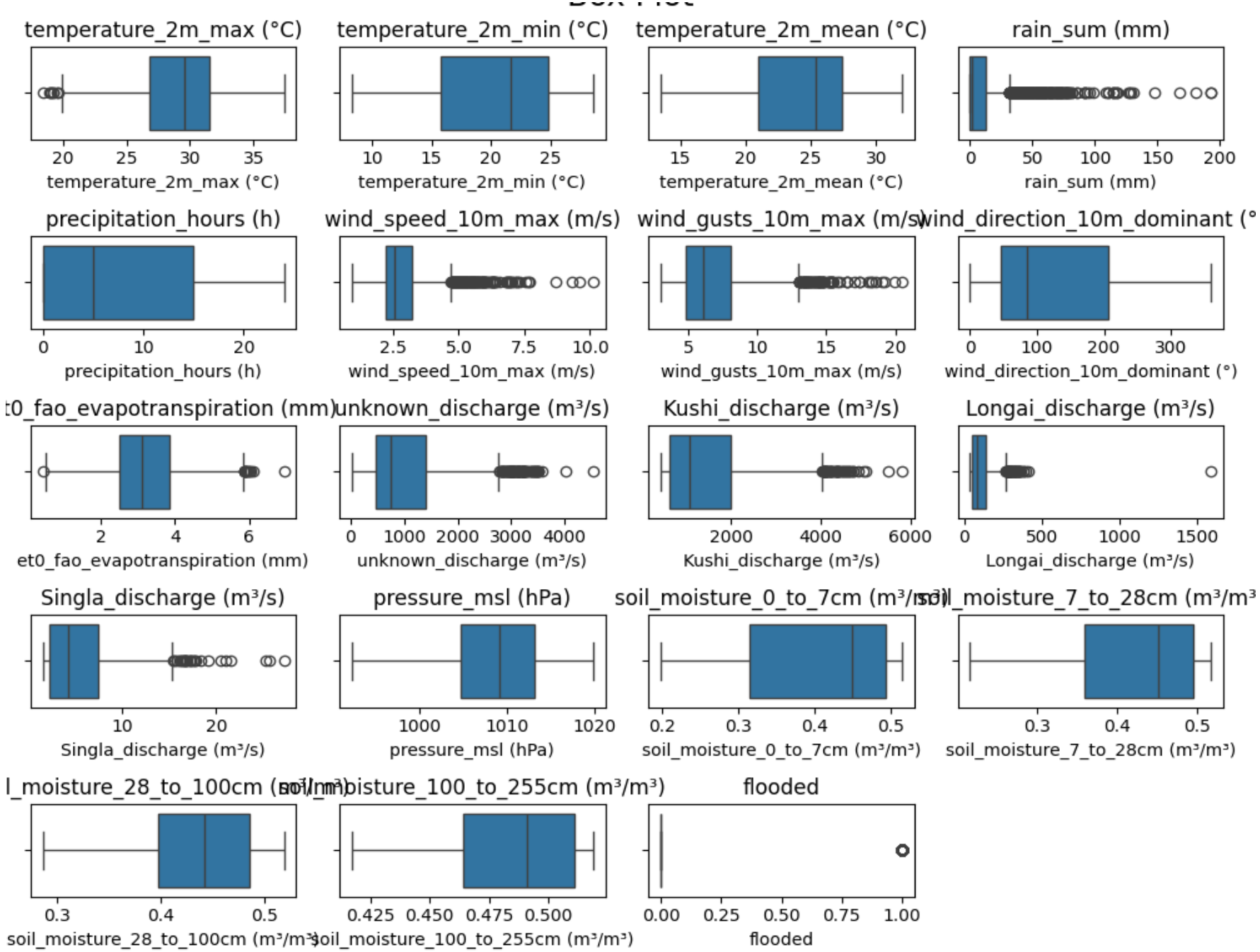
Dataset Statistics

	count	mean	std	min	25%	50%	75%	max
temperature_2m_max (°C)	3654.0	29.057608	3.295531	18.400	26.80000	29.6000	31.500	37.400
temperature_2m_min (°C)	3654.0	20.290011	4.989983	8.300	15.80000	21.7000	24.800	28.600
temperature_2m_mean (°C)	3654.0	24.220662	3.958377	13.500	21.00000	25.4000	27.400	32.000
rain_sum (mm)	3654.0	9.425698	16.915257	0.000	0.00000	1.7000	12.600	193.100
precipitation_hours (h)	3654.0	7.788451	8.283795	0.000	0.00000	5.0000	15.000	24.000
wind_speed_10m_max (m/s)	3654.0	2.803503	0.921590	0.920	2.20000	2.5600	3.200	10.120
wind_gusts_10m_max (m/s)	3654.0	6.740887	2.453858	3.000	4.80000	6.1000	8.100	20.500
wind_direction_10m_dominant (°)	3654.0	128.991516	105.244328	0.000	46.00000	86.0000	207.000	360.000
et0_fao_evapotranspiration (mm)	3654.0	3.183656	1.017729	0.440	2.51000	3.1150	3.850	6.930
unknown_discharge (m³/s)	3654.0	1018.786582	675.769568	20.430	470.25000	738.6100	1398.590	4527.970
Kushi_discharge (m³/s)	3654.0	1405.743941	917.463724	438.540	621.30000	1079.1900	1985.100	5820.030
Longai_discharge (m³/s)	3654.0	100.429540	64.971843	37.020	48.66000	79.7200	136.400	1591.210
Singla_discharge (m³/s)	3654.0	5.252110	3.381358	1.590	2.28000	4.2800	7.530	27.180
pressure_msl (hPa)	3654.0	1008.789381	5.122606	992.200	1004.60000	1009.1000	1013.100	1019.800
soil_moisture_0_to_7cm (m³/m³)	3654.0	0.407889	0.095364	0.198	0.31525	0.4490	0.493	0.515
soil_moisture_7_to_28cm (m³/m³)	3654.0	0.426486	0.076565	0.217	0.36000	0.4515	0.496	0.518
soil_moisture_28_to_100cm (m³/m³)	3654.0	0.435591	0.057907	0.286	0.39800	0.4420	0.486	0.519
soil_moisture_100_to_255cm (m³/m³)	3654.0	0.486138	0.026042	0.417	0.46400	0.4910	0.511	0.519
flooded	3654.0	0.058566	0.234843	0.000	0.00000	0.0000	0.000	1.000

Pair Plot


Count
temperature_2m_max (°C)
temperature_2m_min (°C)
temperature_2m_mean (°C)
rain_sum (mm)
precipitation_hours (h)
wind_speed_10m_max (m/s)
wind_gusts_10m_max (m/s)
wind_direction_10m_dominant (
et0_fao_evapotranspiration (mm)
unknown_discharge (m³/s)
Kushi_discharge (m³/s)
Longai_discharge (m³/s)
Singla_discharge (m³/s)
pressure_msl (hPa)
soil_moisture_0_to_7cm (m³/m³)
soil_moisture_7_to_28cm (m³/m
soil_moisture_28_to_100cm (m³/m³)
soil_moisture_100_to_255cm (m³/m³)
flooded

Box Plot


temperature_2m_max (°C)
temperature_2m_min (°C)
temperature_2m_mean (°C)
rain_sum (mm)
precipitation_hours (h)
wind_speed_10m_max (m/s)
wind_gusts_10m_max (m/s)
wind_direction_10m_dominant (°
et0_fao_evapotranspiration (mm)
unknown_discharge (m³/s)
Kushi_discharge (m³/s)
Longai_discharge (m³/s)
Singla_discharge (m³/s)
pressure_msl (hPa)
soil_moisture_0_to_7cm (m³/m³)
soil_moisture_7_to_28cm (m³/m³
soil_moisture_28_to_100cm (m³/m³)
soil_moisture_100_to_255cm (m³/m³)
flooded

Outliers and Skewness Analysis

Outlier Detection Summary Using IQR Method

Feature	Number of Outliers
rain_sum (mm)	277
flooded	214
wind_speed_10m_max (m/s)	161
unknown_discharge (m³/s)	80
wind_gusts_10m_max (m/s)	79
Kushi_discharge (m³/s)	44
Longai_discharge (m³/s)	42
Singla_discharge (m³/s)	26
et0_fao_evapotranspiration (mm)	12
temperature_2m_max (°C)	8
precipitation_hours (h)	0
temperature_2m_mean (°C)	0
wind_direction_10m_dominant (°)	0
temperature_2m_min (°C)	0
pressure_msl (hPa)	0
soil_moisture_0_to_7cm (m³/m³)	0
soil_moisture_7_to_28cm (m³/m³)	0
soil_moisture_28_to_100cm (m³/m³)	0
soil_moisture_100_to_255cm (m³/m³)	0

Skewness Analysis of Numerical Features

```
temperature_2m_max (°C)                -0.461317
temperature_2m_min (°C)                -0.544052
temperature_2m_mean (°C)               -0.575264
rain_sum (mm)                           3.763007
precipitation_hours (h)                 0.595463
wind_speed_10m_max (m/s)                1.796237
wind_gusts_10m_max (m/s)                1.259856
wind_direction_10m_dominant (°)         0.821181
et0_fao_evapotranspiration (mm)         0.142698
unknown_discharge (m³/s)                1.219477
Kushi_discharge (m³/s)                  1.062425
Longai_discharge (m³/s)                 4.194884
Singla_discharge (m³/s)                 1.082289
pressure_msl (hPa)                     -0.193027
soil_moisture_0_to_7cm (m³/m³)         -0.582044
soil_moisture_7_to_28cm (m³/m³)        -0.574779
soil_moisture_28_to_100cm (m³/m³)      -0.602612
soil_moisture_100_to_255cm (m³/m³)     -0.491933
flooded                                 3.761461
dtype: float64
```

Outliers Analysis

- ✓ rain_sum (mm): 277 outliers - This indicates a significant number of days with unusually high rainfall amounts. This is expected in climate data, as rainfall can vary greatly, buit can also signify moonson period.
- ✓ wind_speed_10m_max (m/s): 161 outliers - A substantial number of days had exceptionally high maximum wind speeds. This could be due to storms or other weather patterns.
- ✓ wind_gusts_10m_max (m/s): 79 outliers - A notable number of days experienced very strong wind gusts. These are often associated with severe weather that could aid the occurrence of flood.
- ✓ Longai_discharge (m^3/s): 42 outliers - Similar to the Kushi river, the Longai river discharge also has a moderate number of outlier days.
- ✓ et0_fao_evapotranspiration (mm): 12 outliers - This indicates some days with unusually high evapotranspiration rates, possibly influenced by temperature, wind, or solar radiation.

Skewness Analysis

Highly Positively Skewed (Skewness > 1):

- rain_sum (mm) (3.763007): Rainfall data is often positively skewed. This means there are more days with lower rainfall amounts and fewer days with very high rainfall.
- wind_speed_10m_max (m/s) (1.796237): Wind speed is also frequently positively skewed, with more days having moderate wind speeds and fewer days with very high wind speeds.
- Longai_discharge (m^3/s) (4.194884): This has the highest positive skewness, indicating a strong right skew and potential for extreme discharge events which could be responsible for flood occurrence in Karimganj

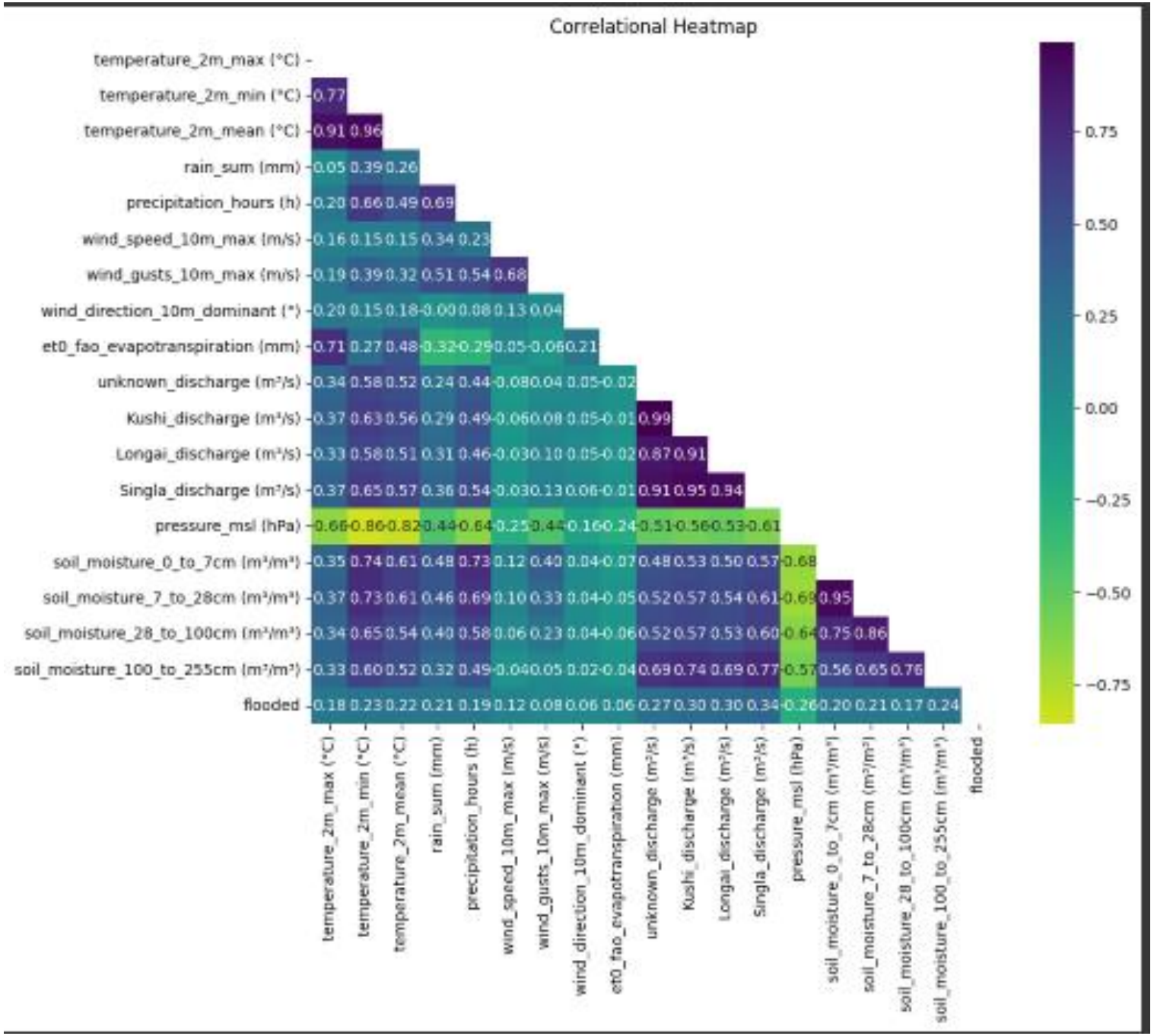
Moderately Positively Skewed (0 < Skewness < 1):

- precipitation_hours (h) (0.595463): There's a slight positive skew in the number of precipitation hours.
- wind_direction_10m_dominant (°) (0.821181): Wind direction might have a slight preference towards certain directions, creating a moderate positive skew.
- et0_fao_evapotranspiration (mm) (0.142698): This has a very slight positive skew.

Slightly Negatively Skewed (-1 < Skewness < 0):

- temperature_2m_max (°C) (-0.461317): There's a slight tendency towards lower maximum temperatures.
- temperature_2m_min (°C) (-0.544052): Similar to maximum temperature, minimum temperatures also have a slight negative skew.
- temperature_2m_mean (°C) (-0.575264): The mean temperature shows a slightly stronger negative skew.
- pressure_msl (hPa) (-0.193027): Air pressure has a slight negative skew.
- soil_moisture_28_to_100cm (m^3/m^3) (-0.602612): This soil moisture layer has the strongest negative skew among the soil moisture features.
- soil_moisture_100_to_255cm (m^3/m^3) (-0.491933): A slight negative skew for this deeper soil moisture layer.

Correlation Analysis



Correlation Analysis

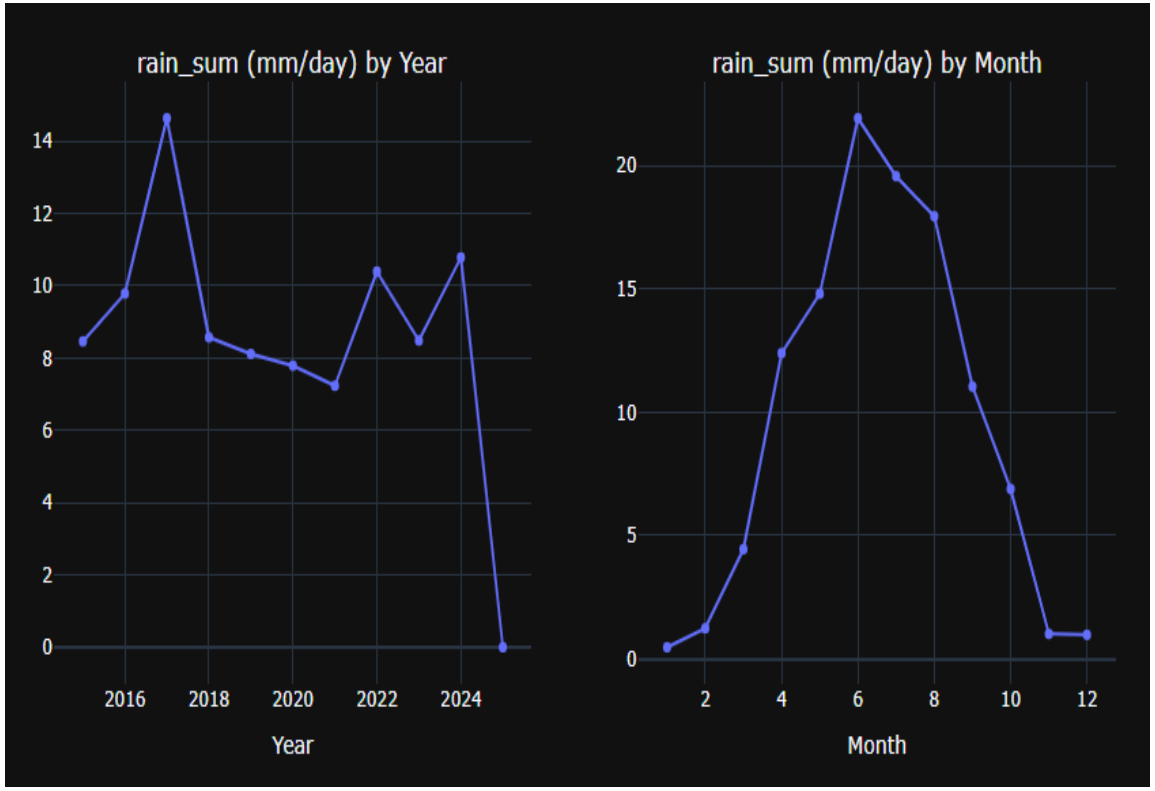
Strongest Positive Correlated Features

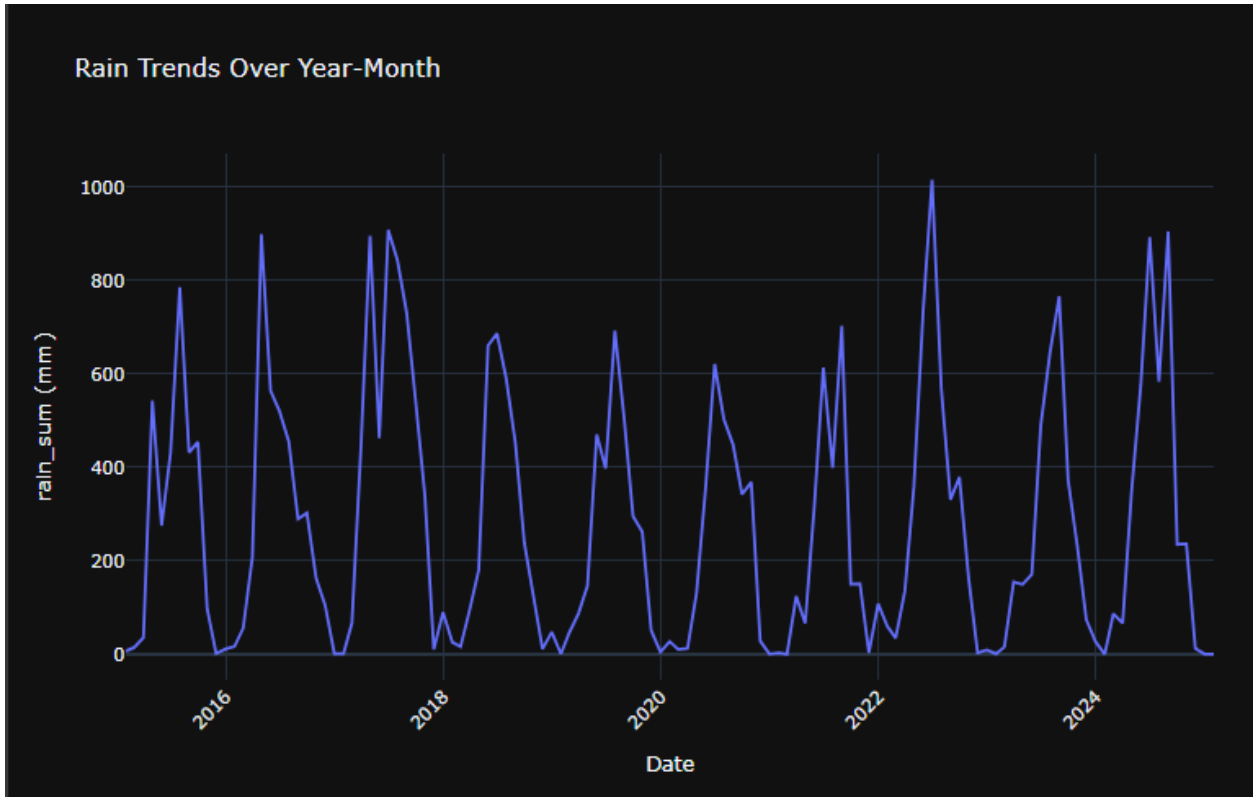
	Feature1	Feature2	Correlation
4	unknown_discharge (m³/s)	Kushi_discharge (m³/s)	0.991719
1	temperature_2m_min (°C)	temperature_2m_mean (°C)	0.958794
9	Kushi_discharge (m³/s)	Singla_discharge (m³/s)	0.951603
16	soil_moisture_0_to_7cm (m³/m³)	soil_moisture_7_to_28cm (m³/m³)	0.945590
12	Longai_discharge (m³/s)	Singla_discharge (m³/s)	0.937195
6	unknown_discharge (m³/s)	Singla_discharge (m³/s)	0.911278
8	Kushi_discharge (m³/s)	Longai_discharge (m³/s)	0.907565
0	temperature_2m_max (°C)	temperature_2m_mean (°C)	0.905047
5	unknown_discharge (m³/s)	Longai_discharge (m³/s)	0.872298
18	soil_moisture_7_to_28cm (m³/m³)	soil_moisture_28_to_100cm (m³/m³)	0.858629

Strongest Negative Correlated Features

	Feature1	Feature2	Correlation
1	temperature_2m_min (°C)	pressure_msl (hPa)	-0.858222
2	temperature_2m_mean (°C)	pressure_msl (hPa)	-0.820842
17	pressure_msl (hPa)	soil_moisture_7_to_28cm (m³/m³)	-0.694662
16	pressure_msl (hPa)	soil_moisture_0_to_7cm (m³/m³)	-0.683415
0	temperature_2m_max (°C)	pressure_msl (hPa)	-0.658201
3	precipitation_hours (h)	pressure_msl (hPa)	-0.642510
18	pressure_msl (hPa)	soil_moisture_28_to_100cm (m³/m³)	-0.635212
7	Singla_discharge (m³/s)	pressure_msl (hPa)	-0.611522
19	pressure_msl (hPa)	soil_moisture_100_to_255cm (m³/m³)	-0.573872
5	Kushi_discharge (m³/s)	pressure_msl (hPa)	-0.562477
6	Longai_discharge (m³/s)	pressure_msl (hPa)	-0.529452
4	unknown_discharge (m³/s)	pressure_msl (hPa)	-0.514078

Rainfall Trend

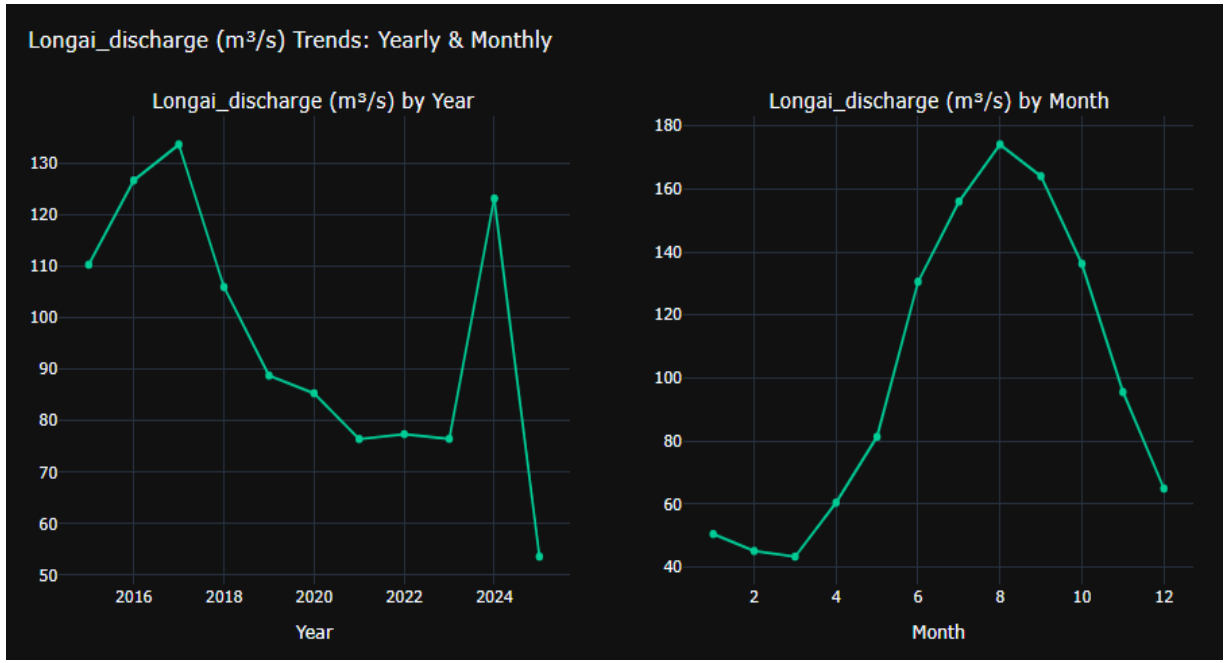


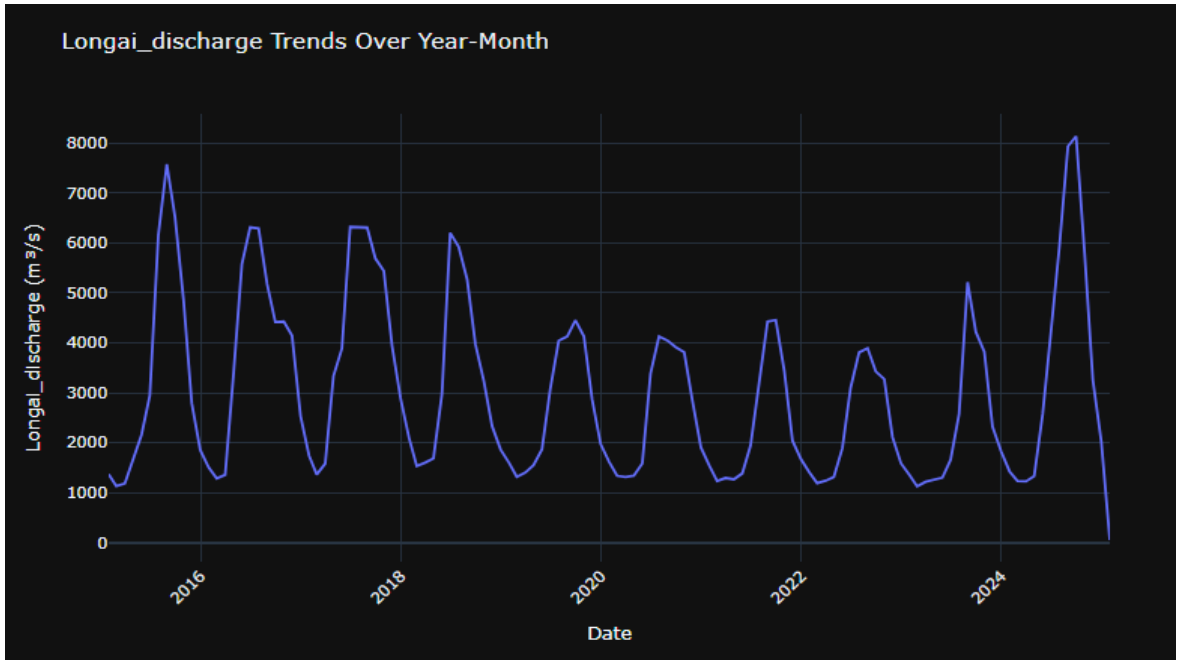


- ✓ **Yearly Trend** The mean rainfall was significantly high in 2017 and a sharp drop in 2018 through 2021, but then it increased in 2022.
- ✓ **Monthly Trend:** The mean rainfall shows an upward trend with peak period in June. The mean rainfall is seen to start dropping back from July to December

- ✓ The rainfall exhibits strong seasonal variations with a repeating annual cycle.
- ✓ There are periodic peaks (cooler months) and troughs (warmer months) consistently across the years.

Longai River Discharge Trend





- ✓ **Yearly Trend:** The mean yearly discharge of the river peaked in 2017 and dropped till 2023. It was peaked again in 2024.
- ✓ **Monthly Trend:** The average daily monthly discharge follows a clear seasonal pattern, peaking between June and August.

- ✓ The Longai river discharge exhibits strong seasonal variations with a repeating annual cycle.
- ✓ While seasonal fluctuations remain prominent, there seems to be an overall downward trend until 2024.

Soil Moisture between 100 to 255cm deep