WEATHER DATA ANALYSIS

JUJJAVARAPU SUJAN CHOWDARY

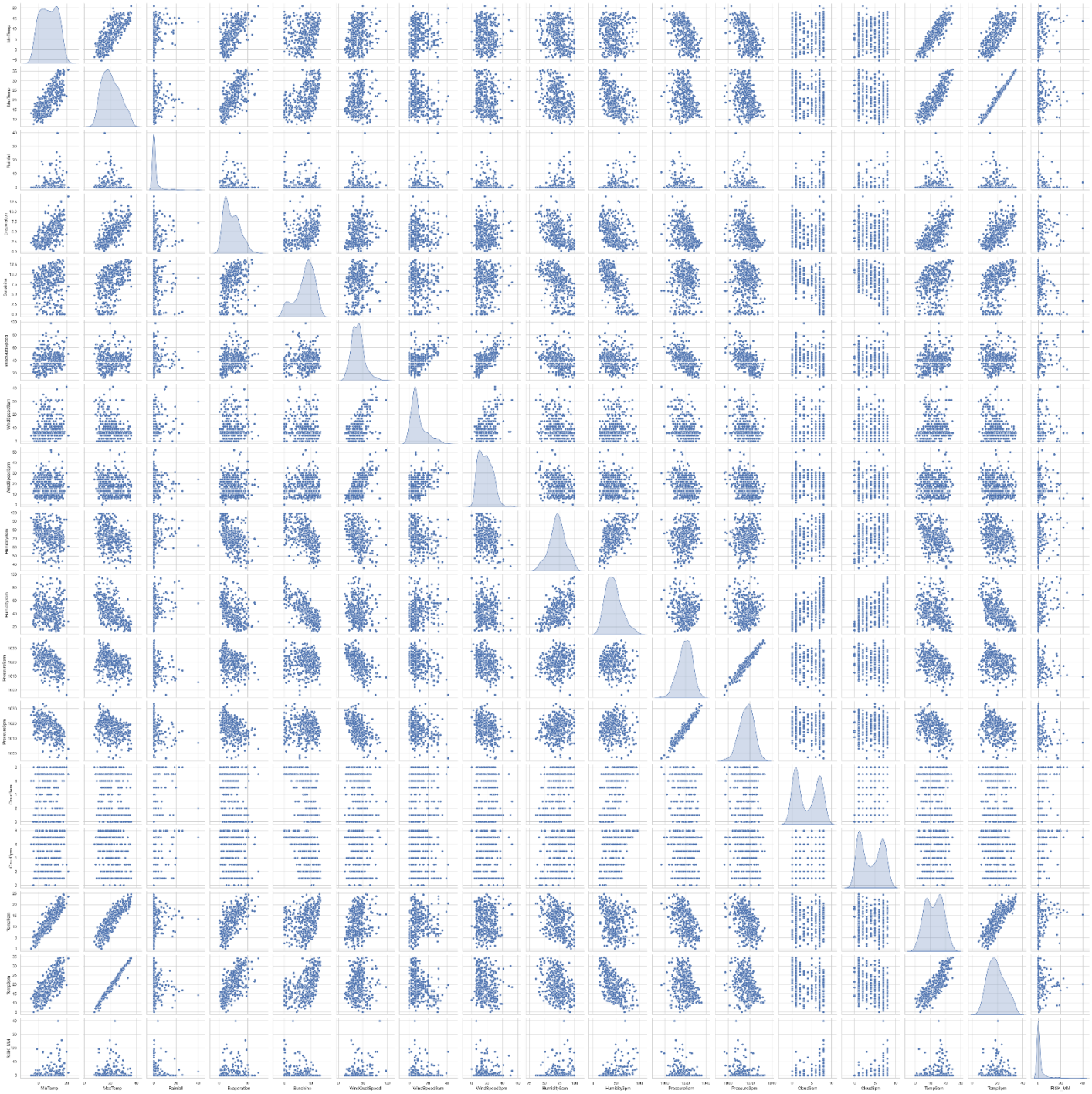
Dataset:

The Dataframe ‘df’ has 366 entries (rows) and 22 columns. The columns are:

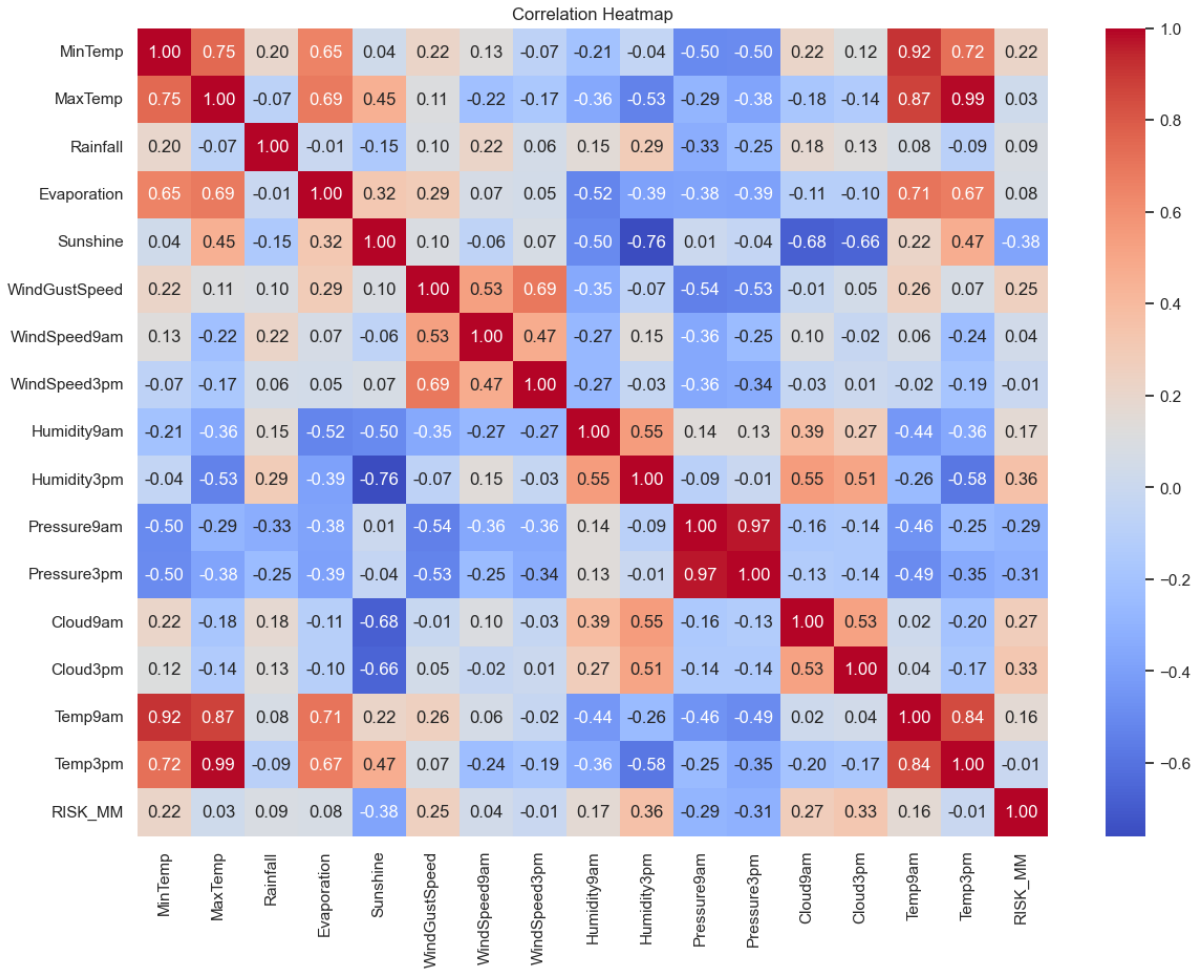
1. MinTemp: Minimum temperature (float64)
2. MaxTemp: Maximum temperature (float64)
3. Rainfall: Amount of rainfall (float64)
4. Evaporation: Amount of evaporation (float64)
5. Sunshine: Amount of sunshine (float64, 3 missing values)
6. WindGustDir: Wind gust direction (object, 3 missing values)
7. WindGustSpeed: Wind gust speed (float64, 2 missing values)
8. WindDir9am: Wind direction at 9am (object, 31 missing values)
9. WindDir3pm: Wind direction at 3pm (object, 1 missing value)
10. WindSpeed9am: Wind speed at 9am (float64, 7 missing values)
11. WindSpeed3pm: Wind speed at 3pm (int64)
12. Humidity9am: Humidity at 9am (int64)
13. Humidity3pm: Humidity at 3pm (int64)
14. Pressure9am: Atmospheric pressure at 9am (float64)
15. Pressure3pm: Atmospheric pressure at 3pm (float64)
16. Cloud9am: Cloudiness at 9am (int64)
17. Cloud3pm: Cloudiness at 3pm (int64)
18. Temp9am: Temperature at 9am (float64)
19. Temp3pm: Temperature at 3pm (float64)
20. RainToday: Whether it rained today (object)
21. RISK_MM: Risk in mm (float64)
22. RainTomorrow: Whether it will rain tomorrow (object)

There are some missing values in the columns Sunshine, WindGustDir, WindGustSpeed, WindDir9am, WindDir3pm, and WindSpeed9am. The data types are float64, int64, and object (likely string).

Pairplot to visualize relationships between numerical variables



Correlation heatmap for numerical variables:



Descriptive Statistics:

Compute and present basic statistics (mean, median, standard deviation) for MinTemp, MaxTemp, Rainfall, and Evaporation.

Here are the statistics for the columns MinTemp, MaxTemp, Rainfall, and Evaporation:

- **MinTemp:**
  - Mean: 7.27
  - Median (50%): 7.45
  - Standard Deviation: 6.03

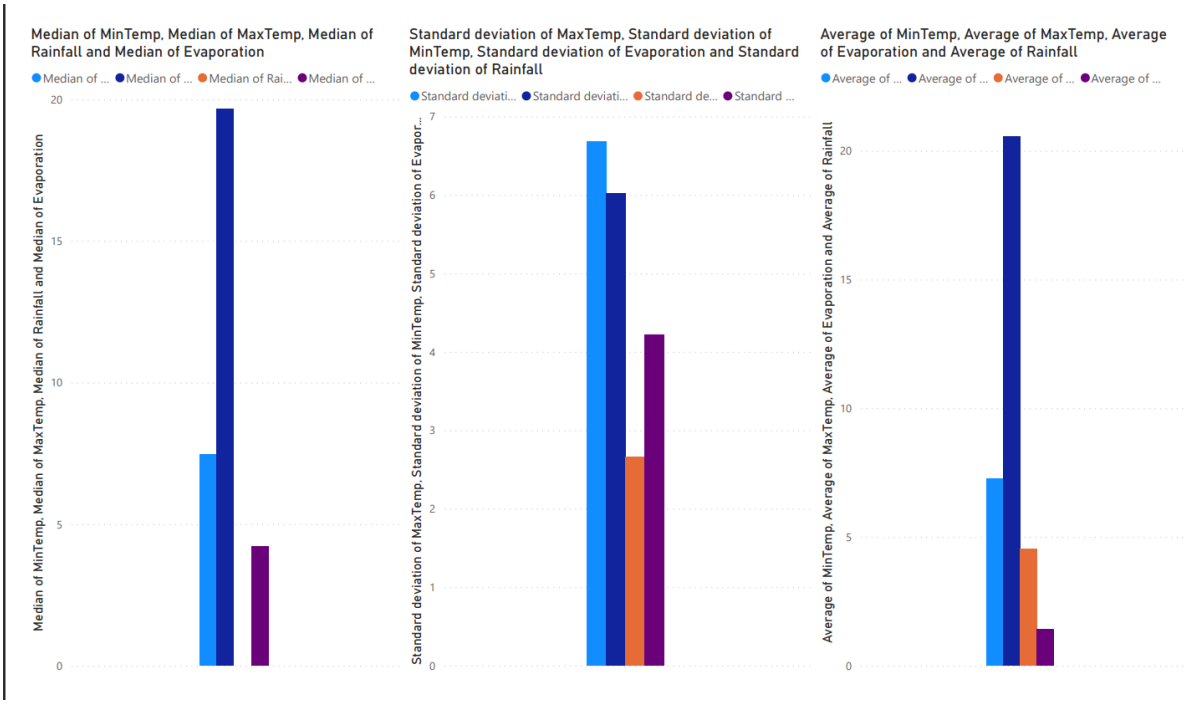
- **MaxTemp:**
  - Mean: 20.55
  - Median (50%): 19.65
  - Standard Deviation: 6.69

```
Statistics for MinTemp:
Mean: 7.265573770491804
Median: 7.45
Standard Deviation: 6.025799834253392

Statistics for MaxTemp:
Mean: 20.550273224043714
Median: 19.65
Standard Deviation: 6.690515669598577

Statistics for Rainfall:
Mean: 1.428415300546448
Median: 0.0
Standard Deviation: 4.225799585804068

Statistics for Evaporation:
Mean: 4.521857923497268
Median: 4.2
Standard Deviation: 2.6693825342212665
```

- **Rainfall:**
  - Mean: 1.43
  - Median (50%): 0.00
  - Standard Deviation: 4.23

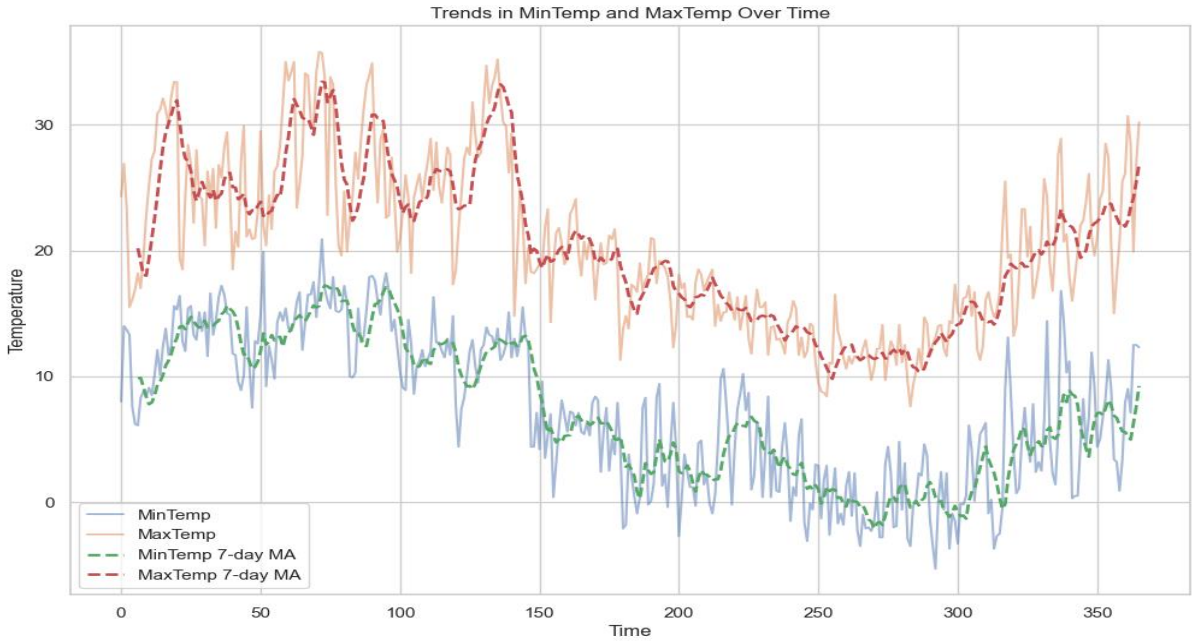
- **Evaporation:**
  - Mean: 4.52
  - Median (50%): 4.20
  - Standard Deviation: 2.67

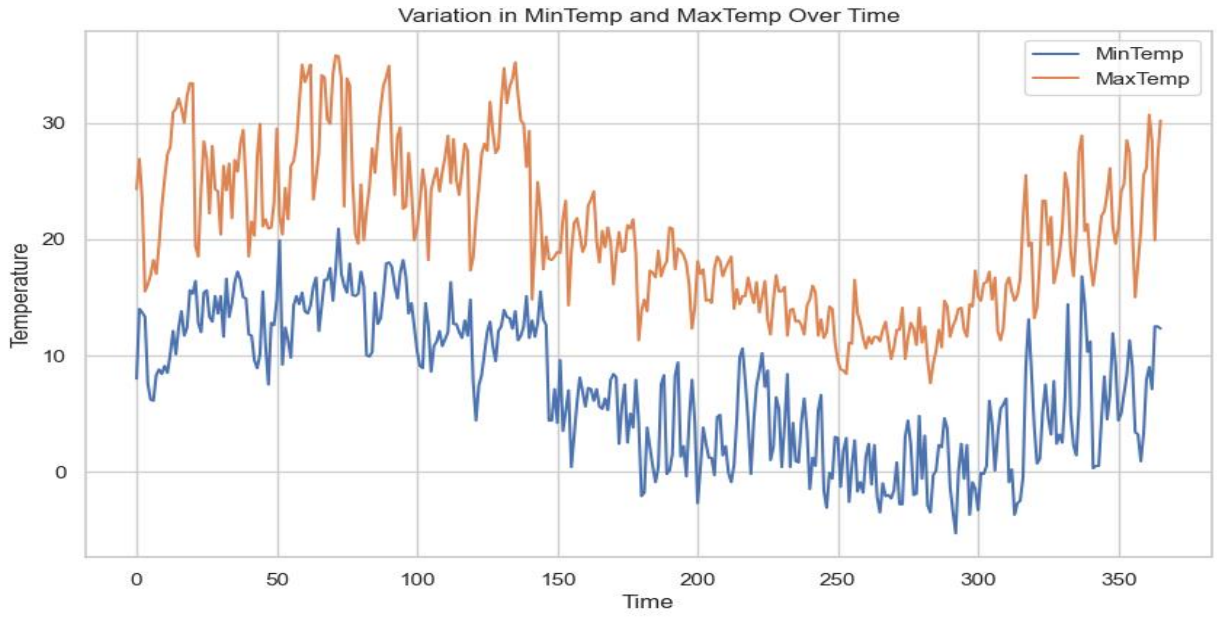
These statistics provide a general understanding of the central tendency and dispersion of the data in these columns.

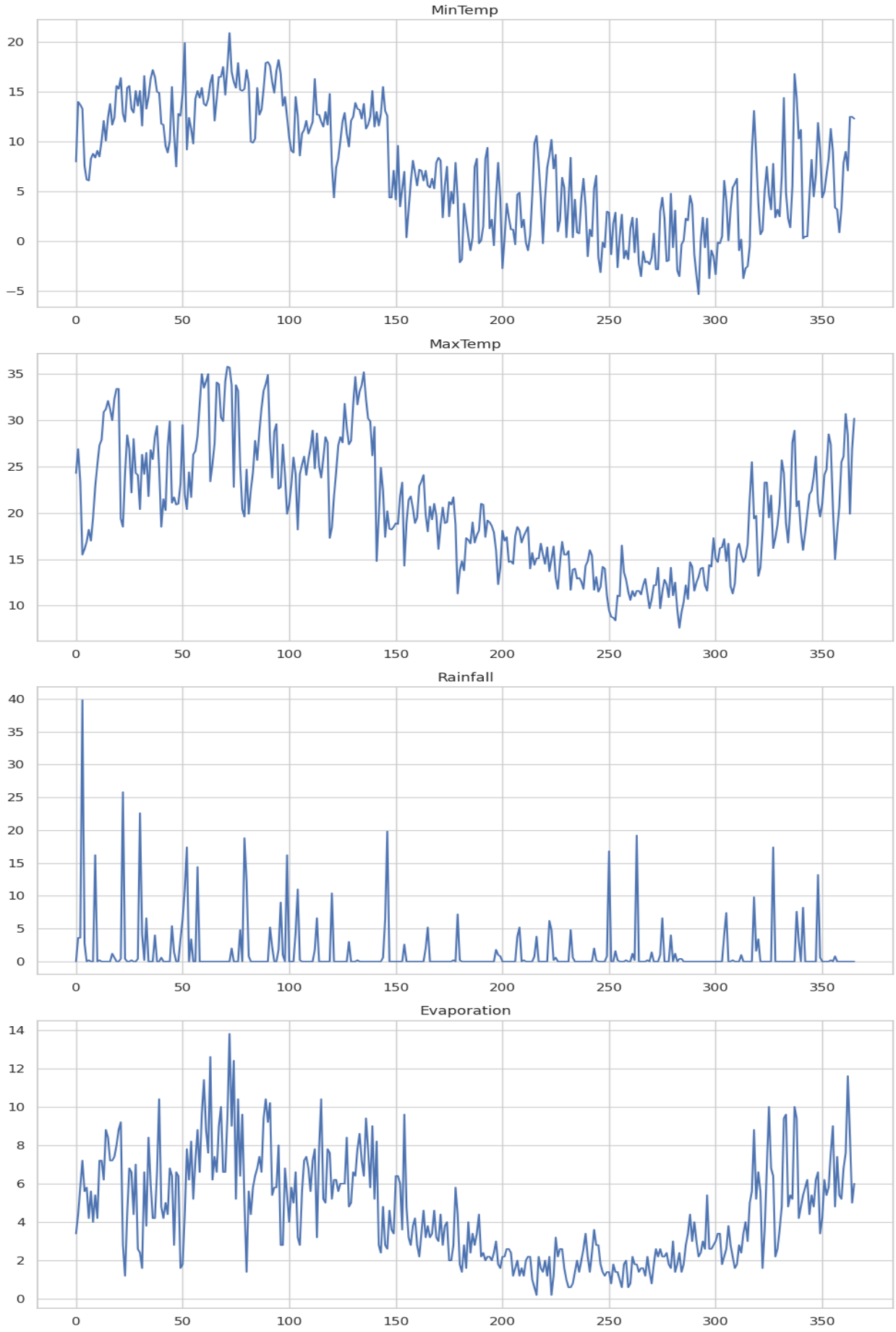


Time Series Visualization:

Create a line chart to show the variations in weather variables over time,identifying trends or patterns.

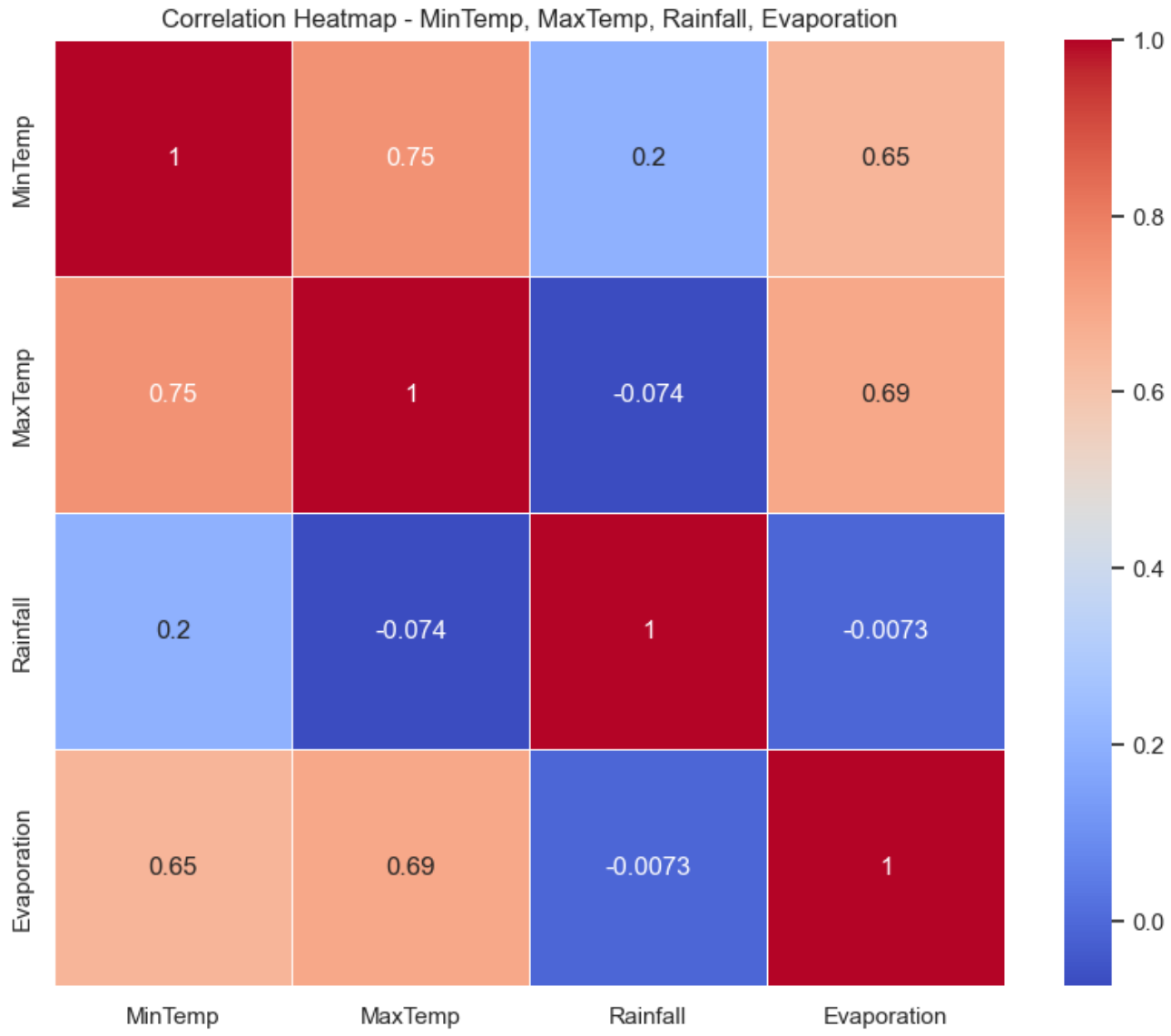





MinTemp
20
15
10
5
0
−5
0
50
100
150
200
250
300
350
MaxTemp
35
30
25
20
15
10
0
50
100
150
200
250
300
350
Rainfall
40
35
30
25
20
15
10
5
0
0
50
100
150
200
250
300
350
Evaporation
14
12
10
8
6
4
2
0
0
50
100
150
200
250
300
350

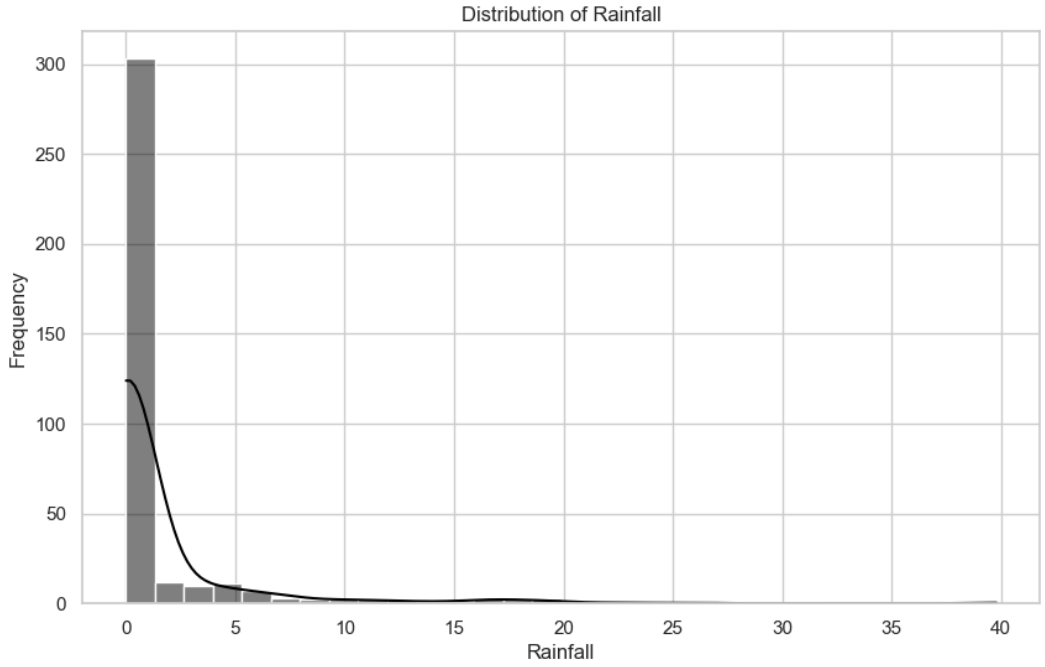
Correlation Analysis:

Calculate and visualize correlations between MinTemp, MaxTemp, Rainfall, and Evaporation using a heatmap.



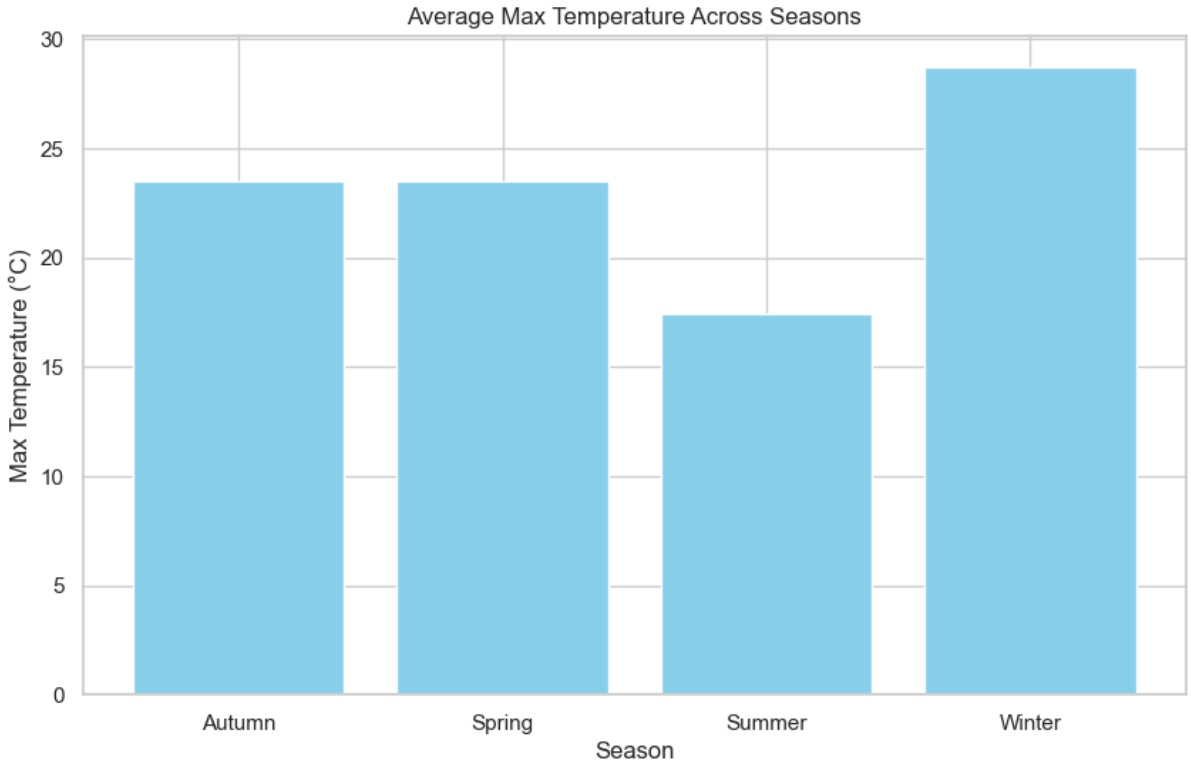
Rainfall Distribution:

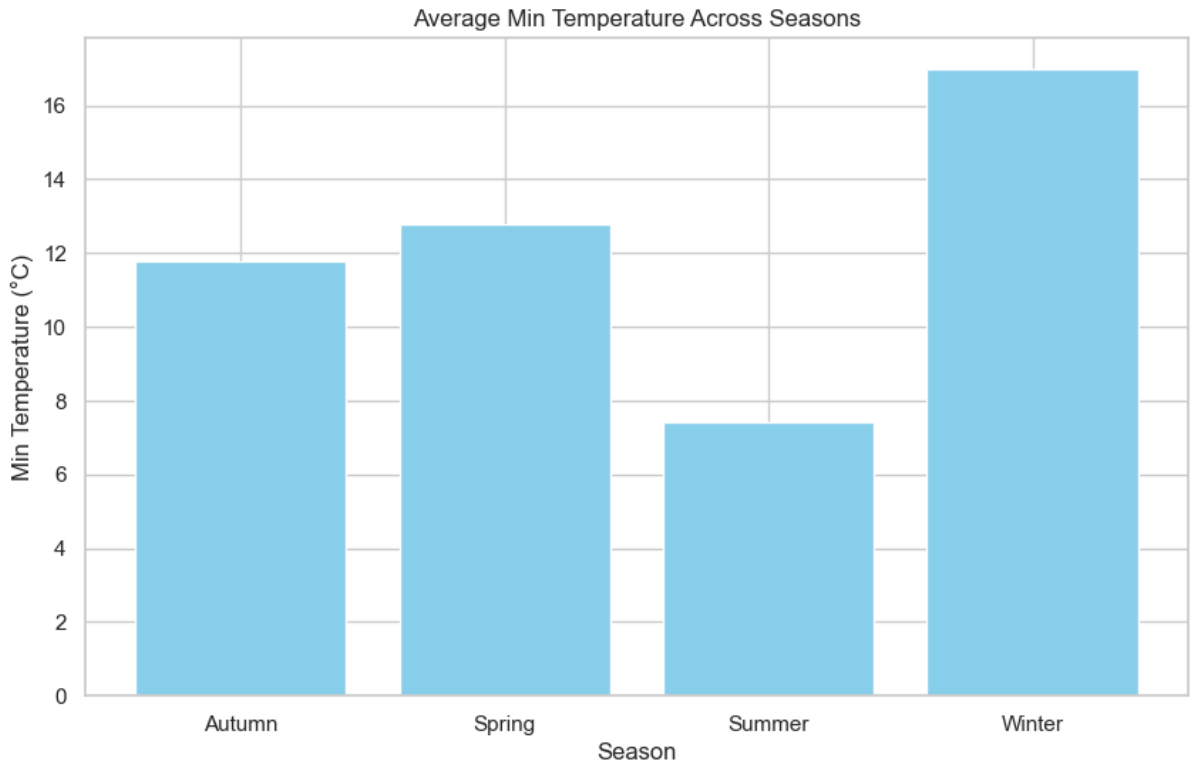
Illustrate the distribution of rainfall through a histogram or kernel density plot, highlighting common levels and outliers.



Seasonal Analysis:

Analyze average values of weather variables across different seasons and visualize seasonal patterns with bar graphs.





NAME: JUJJAVARAPU SUJAN CHOWDARY

EMAIL: chowdarysujan27@gmail.com

PHONE NO: 7780539873

GITHUB LINK: https://github.com/Jsujanchowdary/Weather-Data-Analysis