# Project Iteration 2 Report


## Muhammad Qasim 21L-5246

## Syed Usman Ali 21L-5405

## Section: 7B


## 1. Introduction

This project involves weather data analysis with the goal of predicting precipitation (rainfall). The objective is to do the Exploratory data Analysis to visualise the pre-processed dataset, gain the insights related to dataset, interpreting the relationship between variables and to understand the distribution of each feature. The dataset is sourced from NASA POWER, includes weather conditions recorded over time with the goal of predicting precipitation (rainfall) and contains various weather parameters such as temperature, humidity, and wind speed.

## 2. Univariate Analysis

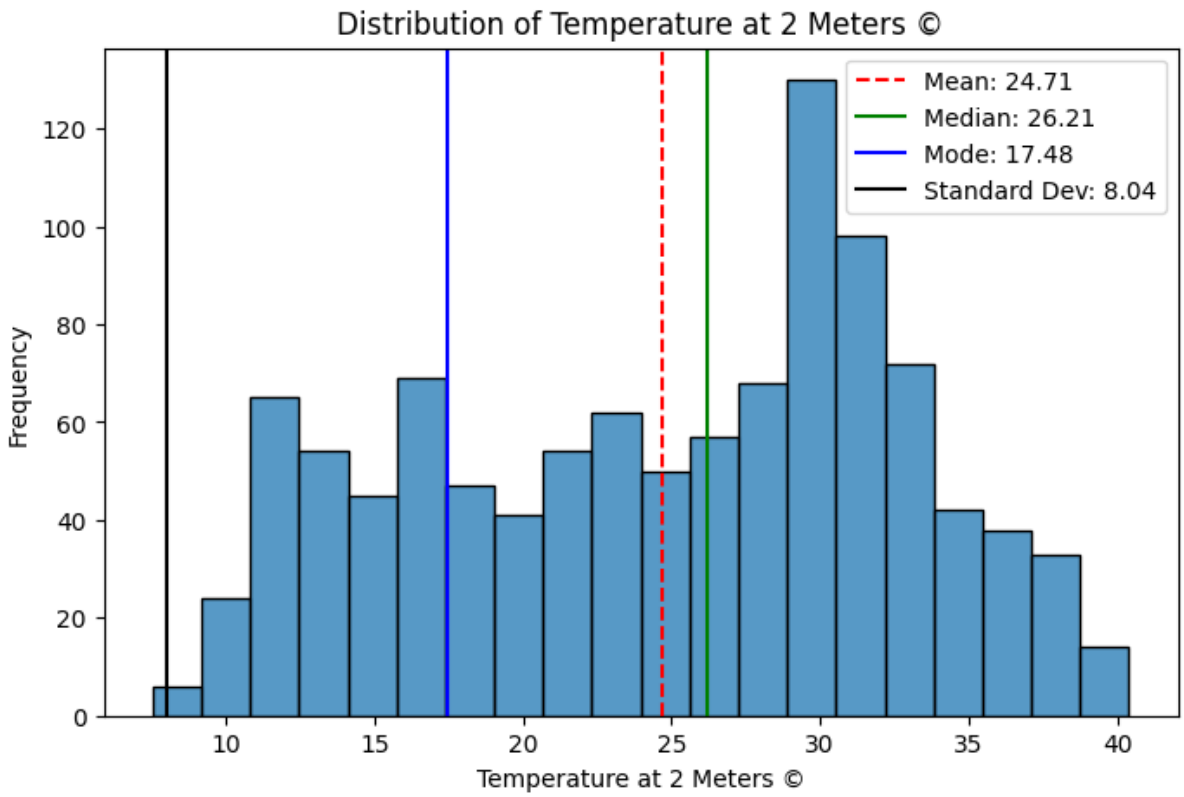
This analysis provided understanding regarding distribution and key statistics such as mean, mode, median etc of each individual continuous variable. The histogram of temperatures at 2 meters above ground suggesting a slightly left skewed feature as the mean (24.71) is pulled towards the left of the distribution. The highest frequency of temperatures lies around 30°C, indicating a clustering of higher temperatures in this range. Temperature values range from 5°C to 40°C, showing broad variability. Histogram of Dew/Frost Point at 2 Meters shows a slightly positive skew, with the mean pulled towards the right of the distribution. It shows a moderately wide spread with a standard deviation of 8.22°C around a mean of 12.82°C. The data ranges widely from around -5°C to 25°C, showing substantial variability in the dew/frost point measurements with

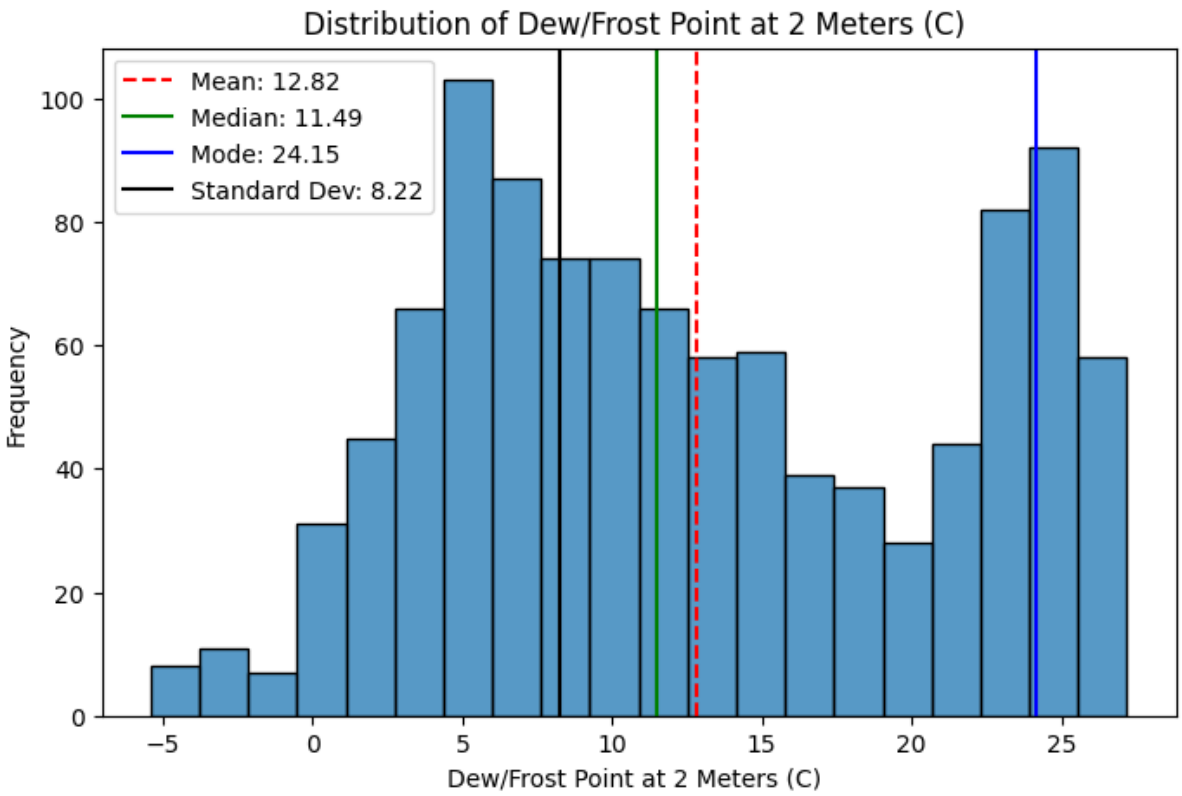
significant number of values concentrated between 5°C to 15°C. The close proximity of the mean, median, and mode of Temperature at 2 Meters Maximum suggests a fairly normal distribution, with a slight skew towards higher temperatures. The standard deviation indicates a moderate spread around the mean, reflecting typical temperature conditions and their variability. In addition to this the histogram also shows that the most common temperature range is between 30°C and 35°C, which aligns with the mode of 30.90°C. Rare occurrences occur on 10°C to 15°C indicating cooler days. Histogram of Temperature at 2 Meters minimum shows that the frequency distribution peaks around 25°C indicating that this is the most common temperature range. The mean temperature is 18.77°C, the median is 19.11°C, and the mode is 6.83°C. The standard deviation is 8.05°C, indicating variability around the mean. The frequency distribution of specific humidity at 2 meters features shows that the most common range is around 5.55 g/kg, indicating typical humidity levels. There are moderate frequencies for 0 to 5 g/kg and 10 to 15 g/kg, representing dry and moderately humid conditions, respectively. Higher humidity levels (15 to 25 g/kg) are less frequent, indicating rarer occurrences of very high humidity. The mean specific humidity is 10.75 g/kg, the median is 8.61 g/kg, and the mode is 5.55 g/kg. The standard deviation is 5.77 g/kg, indicating variability around the mean. The frequency distribution peaks around 50.81% for Relative Humidity at 2 Meters, suggesting this is the most common relative humidity value. The mean relative humidity is 53.68%, the median is 54.50%, and the mode is 50.81%. The standard deviation is 17.94%, indicating variability around the mean. The histogram of Precipitation Corrected shows that the average precipitation is 2.56 mm/day, with a median and mode of 0.00 mm/day, indicating that zero precipitation is the most common occurrence. The standard deviation of 7.74 mm/day suggests significant variability in the data. The frequency distribution reveals a positively skewed pattern, with most data points clustered around

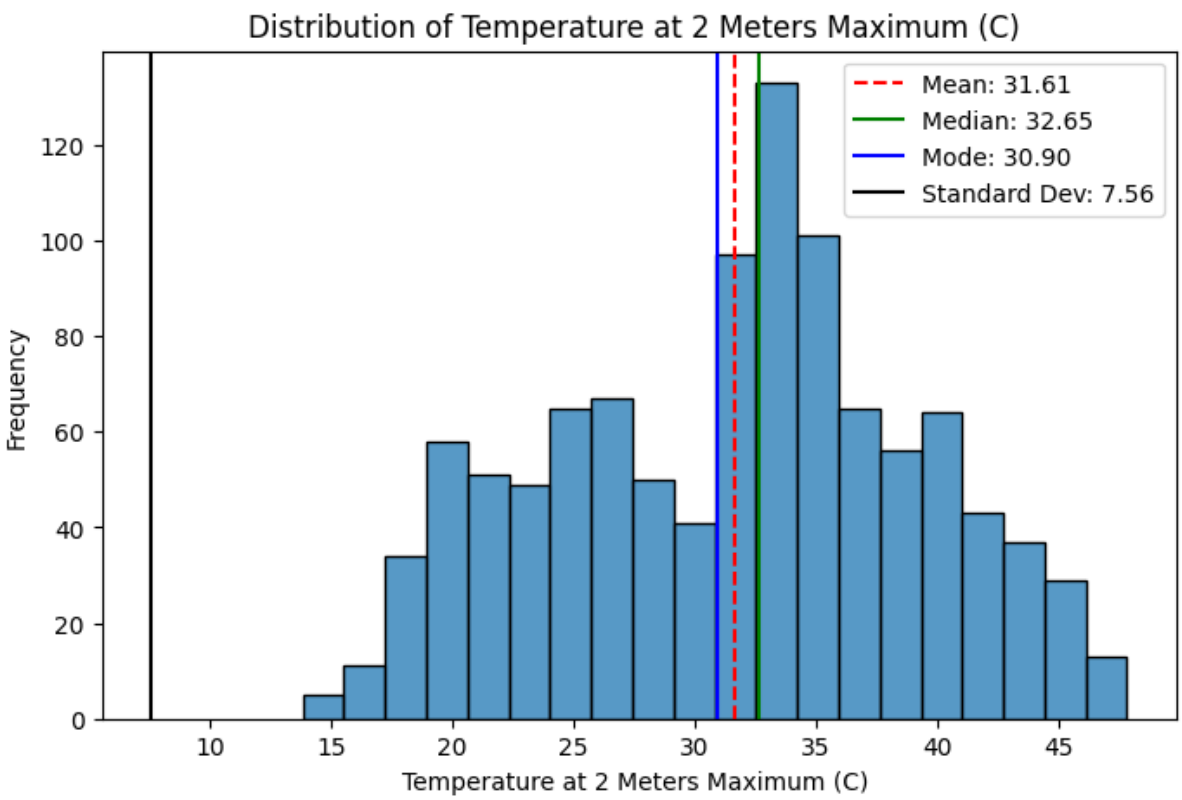
zero and a long tail towards higher values. For the Wind Speed at 10 Meters Feature, the mean wind speed is 2.25 meters per second, with a median of 2.14 meters per second and a mode of 2.09 meters per second, indicating a symmetric distribution. The standard deviation of 0.69 meters per second suggests that the wind speeds are not widely spread out from the mean. The frequency distribution reveals that most values cluster around the mean, indicating an approximately normal distribution. The boxplot of Temperature at 2 Meters shows that the median temperature is around 25°C, indicating the central tendency. The interquartile range (IQR) spans from approximately 20°C to 30°C, showing that 50% of the data falls within this range. The whiskers extend from about 10°C to 40°C, suggesting that temperatures outside the IQR are less frequent but still within this broader range. No outliers are observed in this feature. The boxplot of Dew/Frost Point at 2 Meters shows that the median dew/frost point is around 10°C, indicating the central tendency. The interquartile range (IQR) spans from approximately 5°C to 20°C, showing that 50% of the data falls within this range. The whiskers extend from about -5°C to 25°C, indicating the overall spread of the data, with no outliers present. The boxplot of Temperature at 2 Meters Maximum and Temperature at 2 Meters Minimum shows that the median temperature is around 32.5°C and 20°C, indicating the central tendency. The interquartile range (IQR) spans from approximately 27.5°C to 37.5°C and 15°C to 25°C respectively, showing that 50% of the data falls within this range. The whiskers extend from about 15°C to 45°C and 0°C 32°C, indicating the range of the data, excluding outliers. Furthermore, the boxplot for Specific Humidity at 2 Meters and Relative Humidity at 2 Meters shows that the median specific humidity is around 10 g/kg and 60% respectively, indicating the central tendency. The interquartile range (IQR) spans from approximately 7.5 g/kg to 15 g/kg and 40% to 80% respectively, showing the middle 50% of the data. The whiskers extend from about 5 g/kg to 22.5 g/kg and 20%

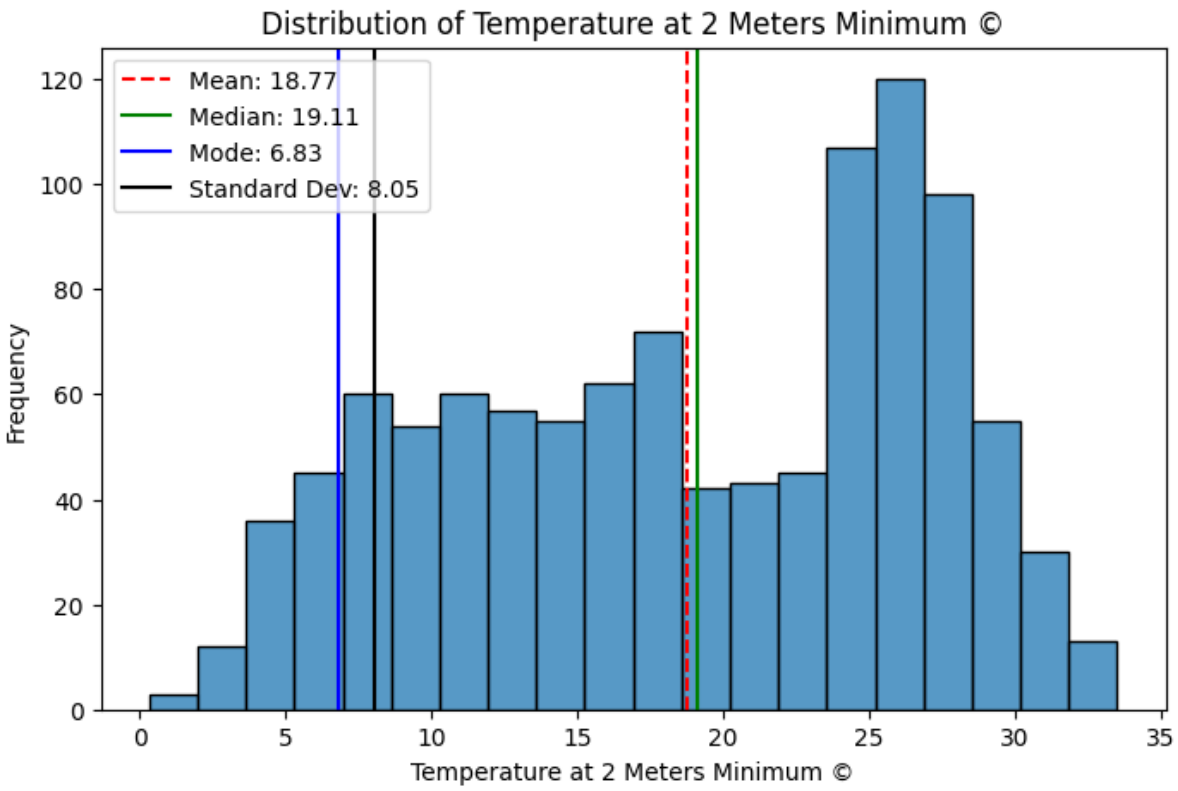
to 90%, indicating the range of the data excluding outliers. The boxplot of Precipitation Corrected and Wind Speed at 10 Meters shows that median value is close to zero, indicating that most days have little to no precipitation and wind speed is around 2.5 meters per second, indicating the central tendency. The presence of numerous outliers in Precipitation, extending up to nearly 100 mm/day, suggests occasional extreme precipitation events. The interquartile range (IQR) for Wind Speed feature spans from approximately 2.0 to 3.0 meters per second, showing the middle 50% of the data. There are also a couple of outliers above 4.0 meters per second, suggesting some unusually high wind speeds.

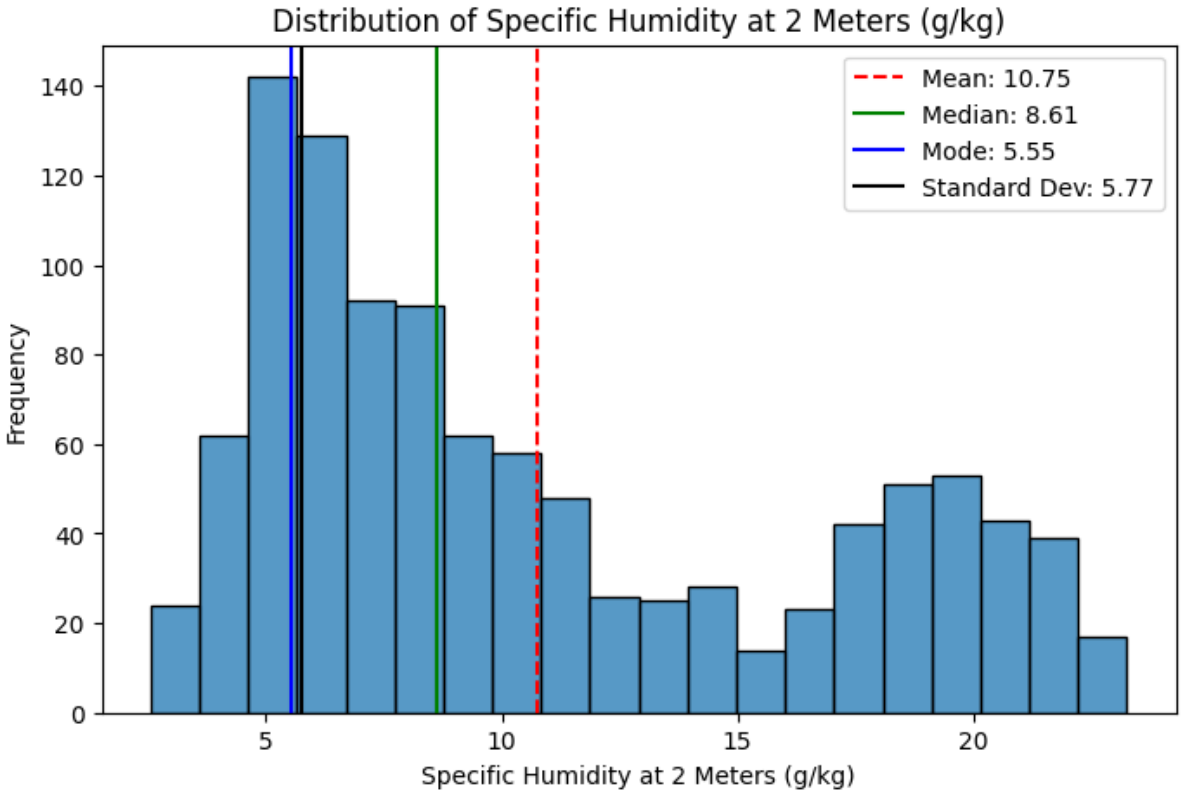
## Histograms

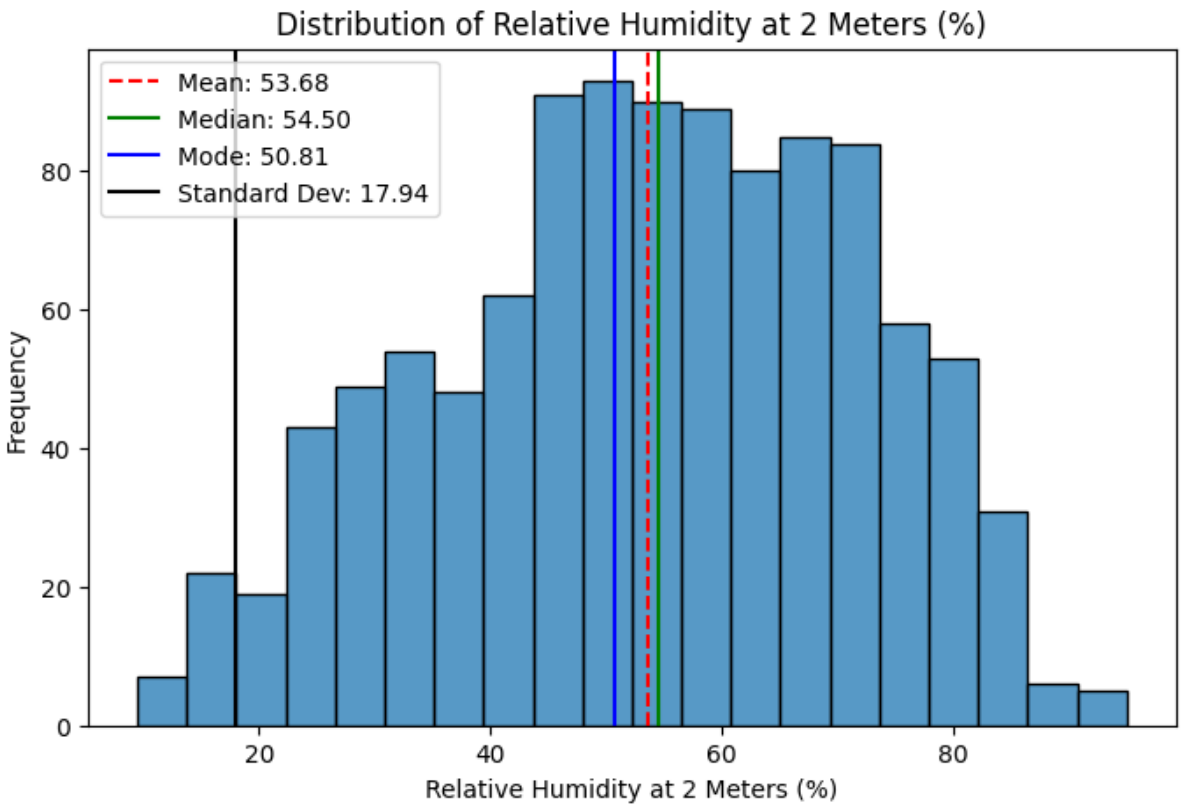


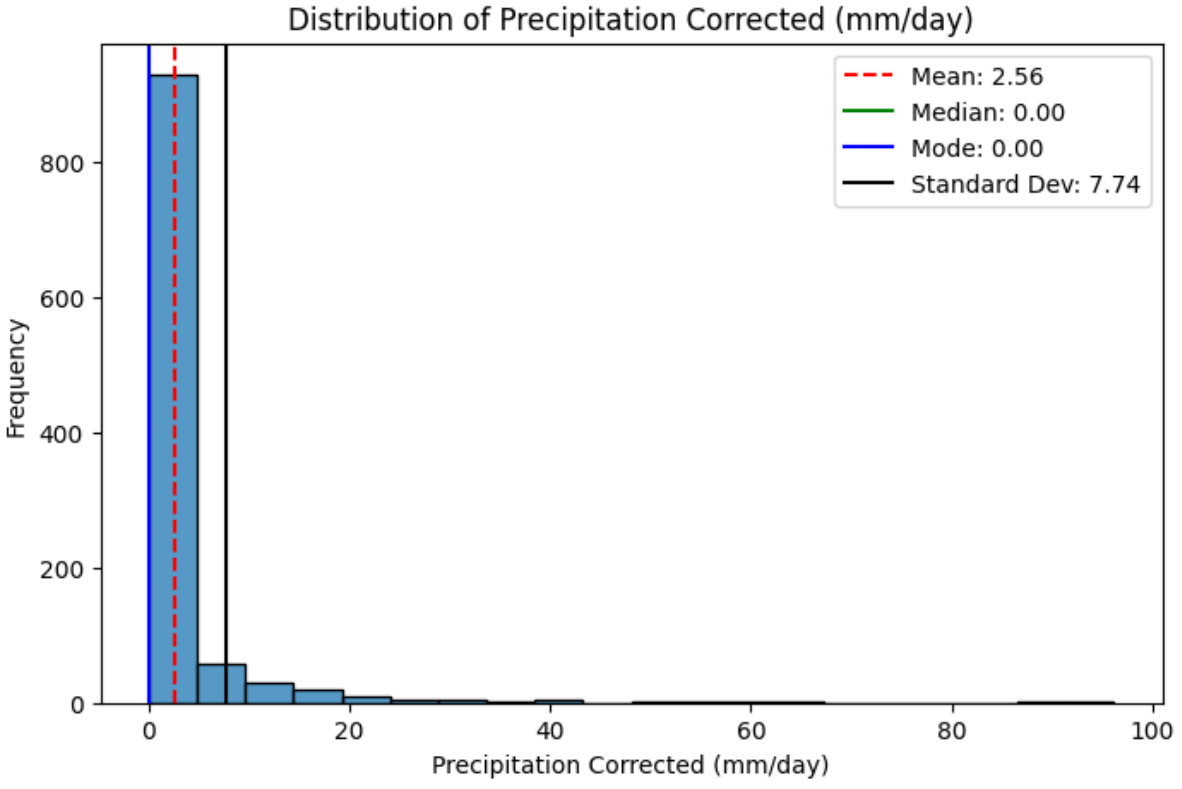

Distribution of Dew/Frost Point at 2 Meters (C)
Mean: 12.82
Median: 11.49
Mode: 24.15
Standard Dev: 8.22
Frequency
Dew/Frost Point at 2 Meters (C)

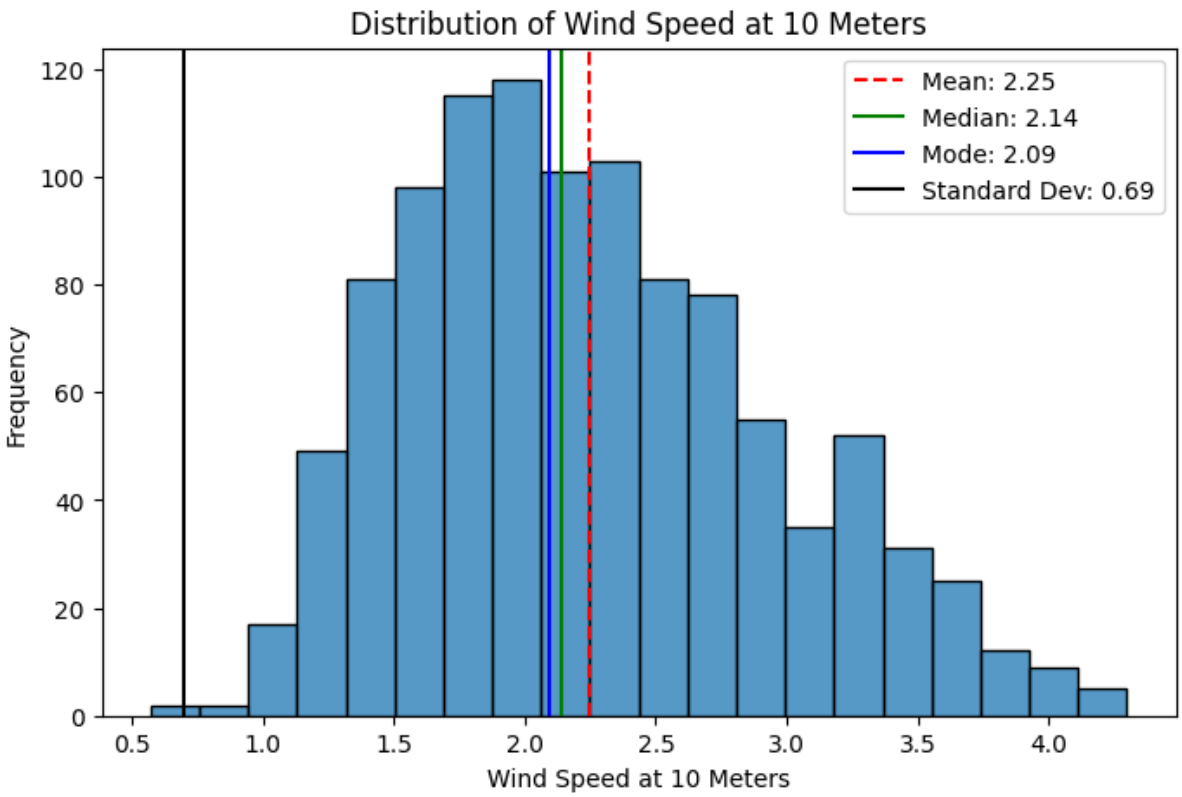

Distribution of Temperature at 2 Meters Maximum (C)
Mean: 31.61
Median: 32.65
Mode: 30.90
Standard Dev: 7.56
Frequency
Temperature at 2 Meters Maximum (C)


Distribution of Temperature at 2 Meters Minimum ©
Mean: 18.77
Median: 19.11
Mode: 6.83
Standard Dev: 8.05
Frequency
Temperature at 2 Meters Minimum ©

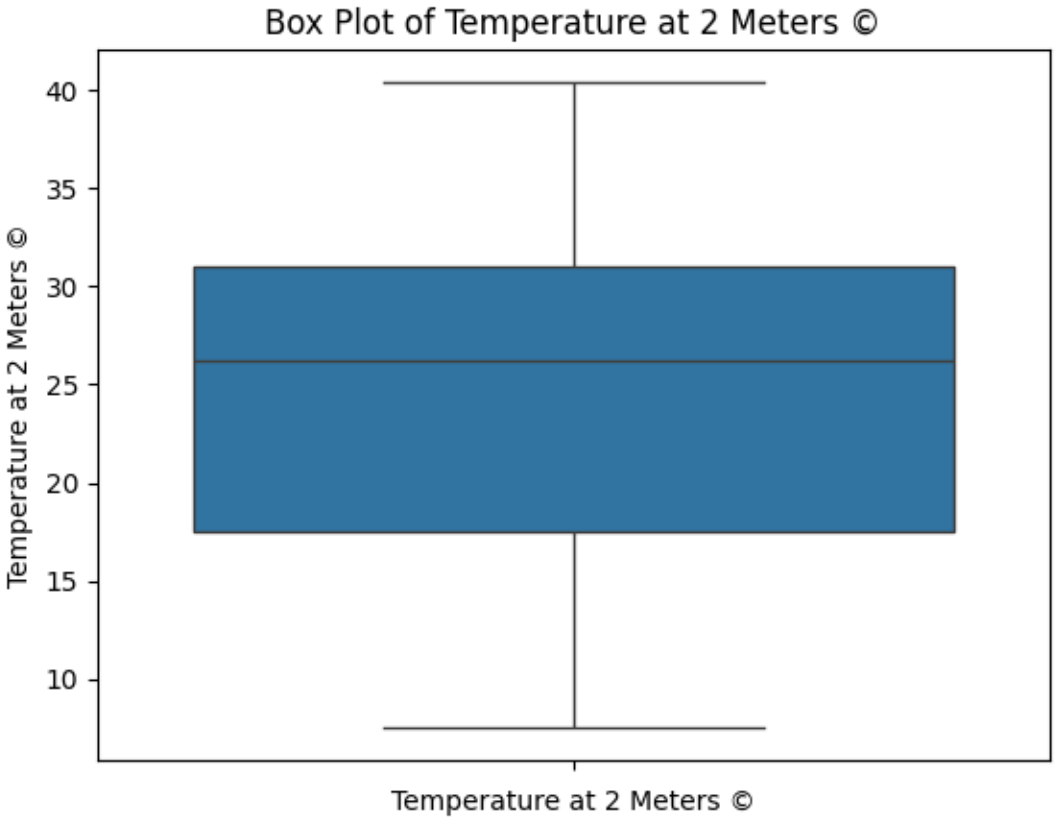

Distribution of Specific Humidity at 2 Meters (g/kg)
Mean: 10.75
Median: 8.61
Mode: 5.55
Standard Dev: 5.77
Frequency
Specific Humidity at 2 Meters (g/kg)

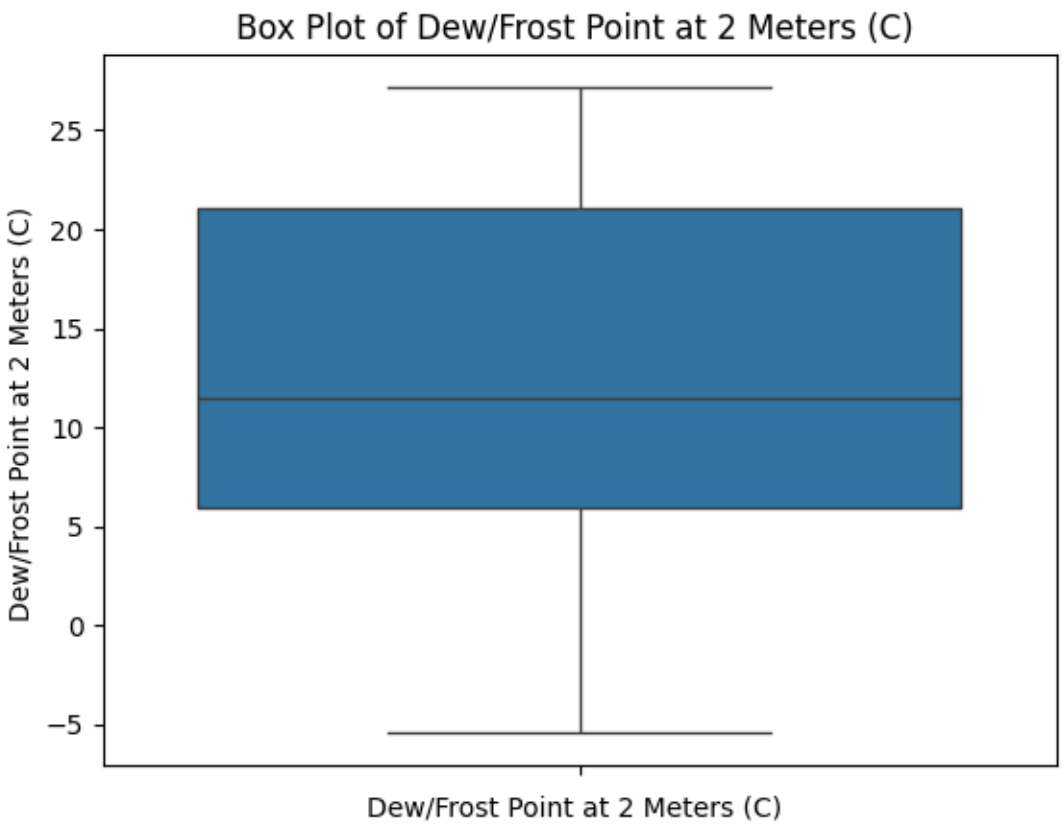

Distribution of Relative Humidity at 2 Meters (%)
Mean: 53.68
Median: 54.50
Mode: 50.81
Standard Dev: 17.94
Frequency
0
20
40
60
80
20
40
60
80
Relative Humidity at 2 Meters (%)

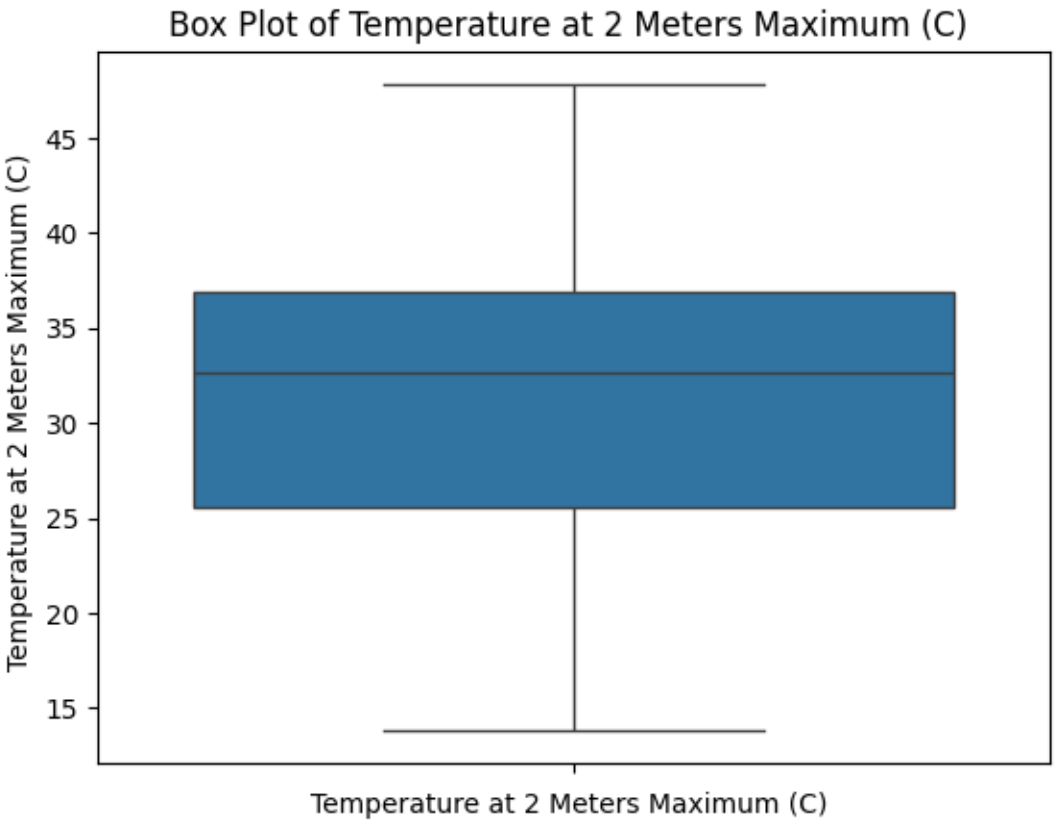

Distribution of Precipitation Corrected (mm/day)
Mean: 2.56
Median: 0.00
Mode: 0.00
Standard Dev: 7.74
Frequency
0
200
400
600
800
0
20
40
60
80
100
Precipitation Corrected (mm/day)

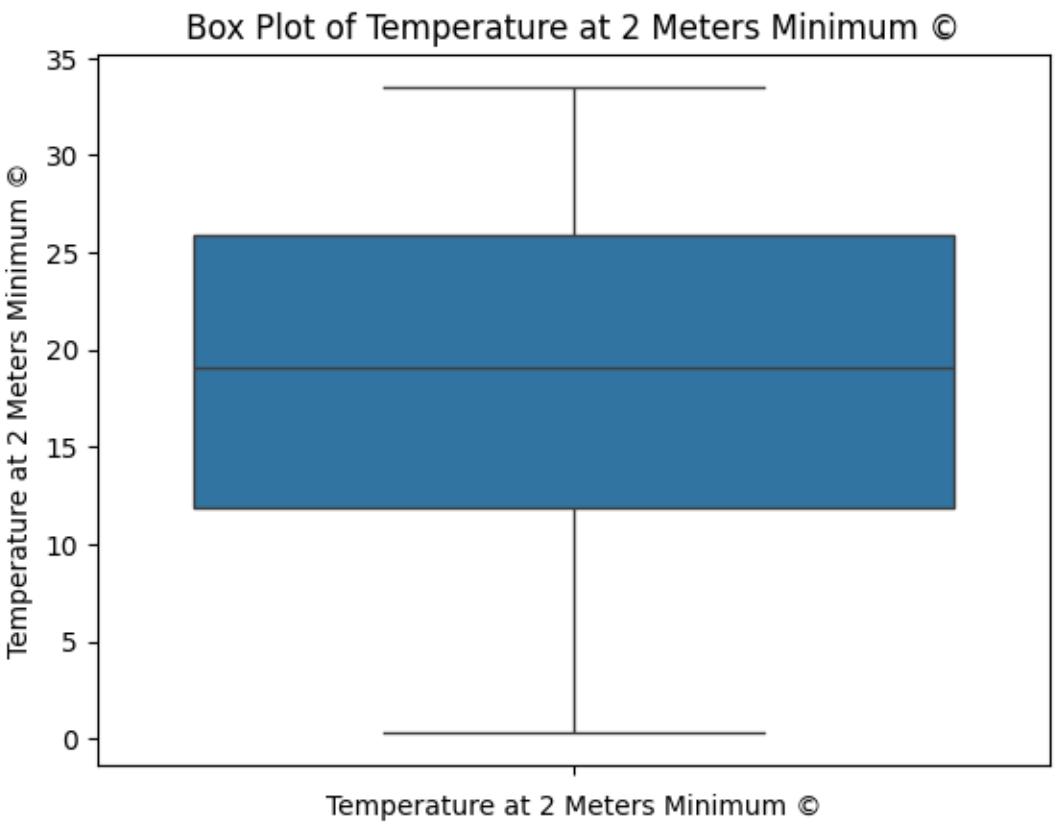

Distribution of Wind Speed at 10 Meters
Mean: 2.25
Median: 2.14
Mode: 2.09
Standard Dev: 0.69
Frequency
Wind Speed at 10 Meters
0
20
40
60
80
100
120
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

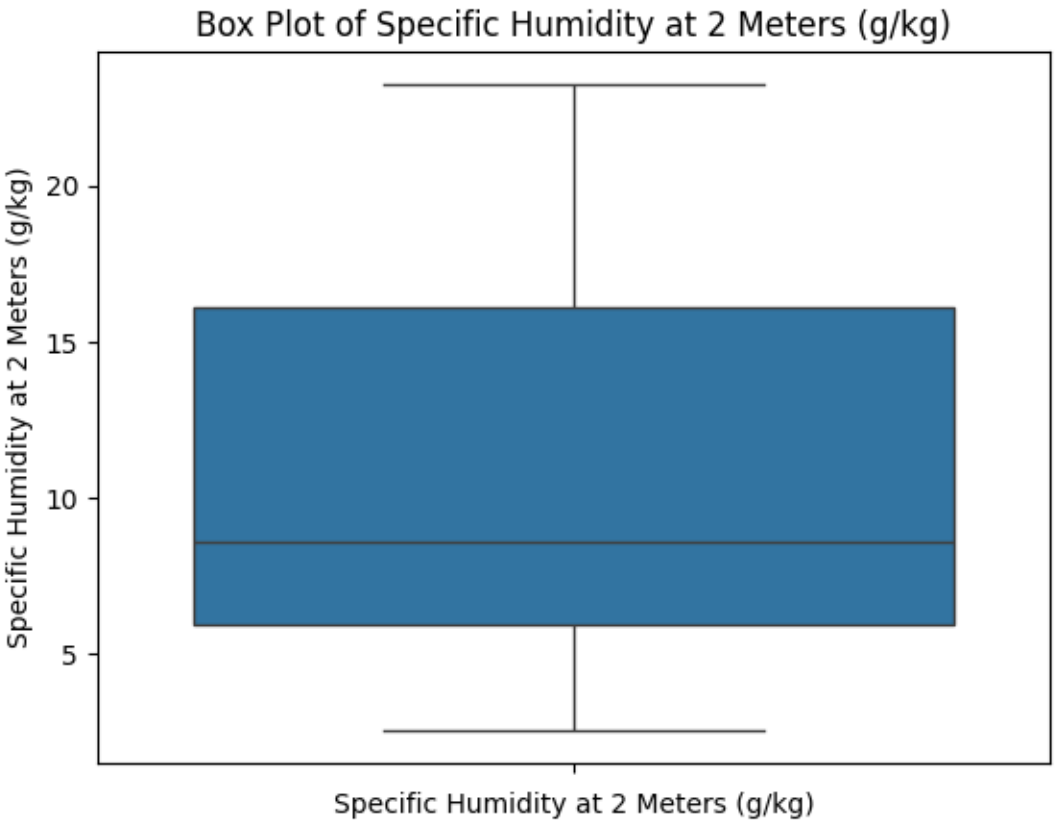
**Boxplots:**

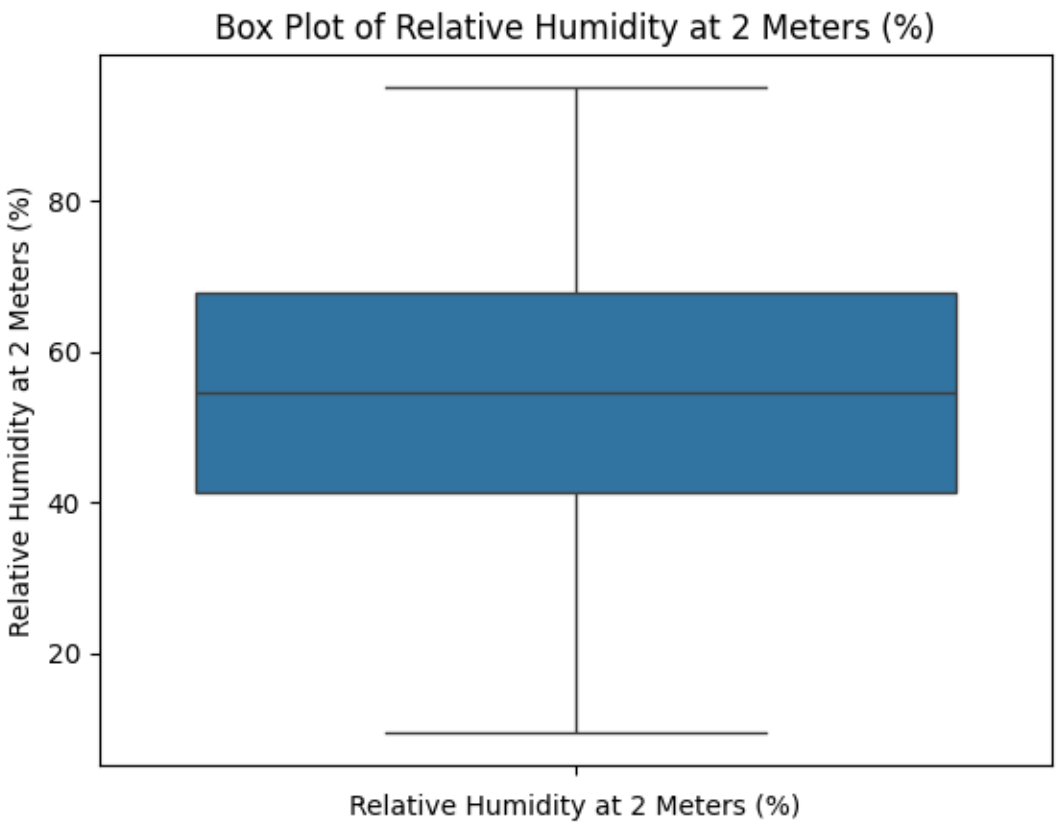

Box Plot of Temperature at 2 Meters ©
Temperature at 2 Meters ©
Temperature at 2 Meters ©
10
15
20
25
30
35
40

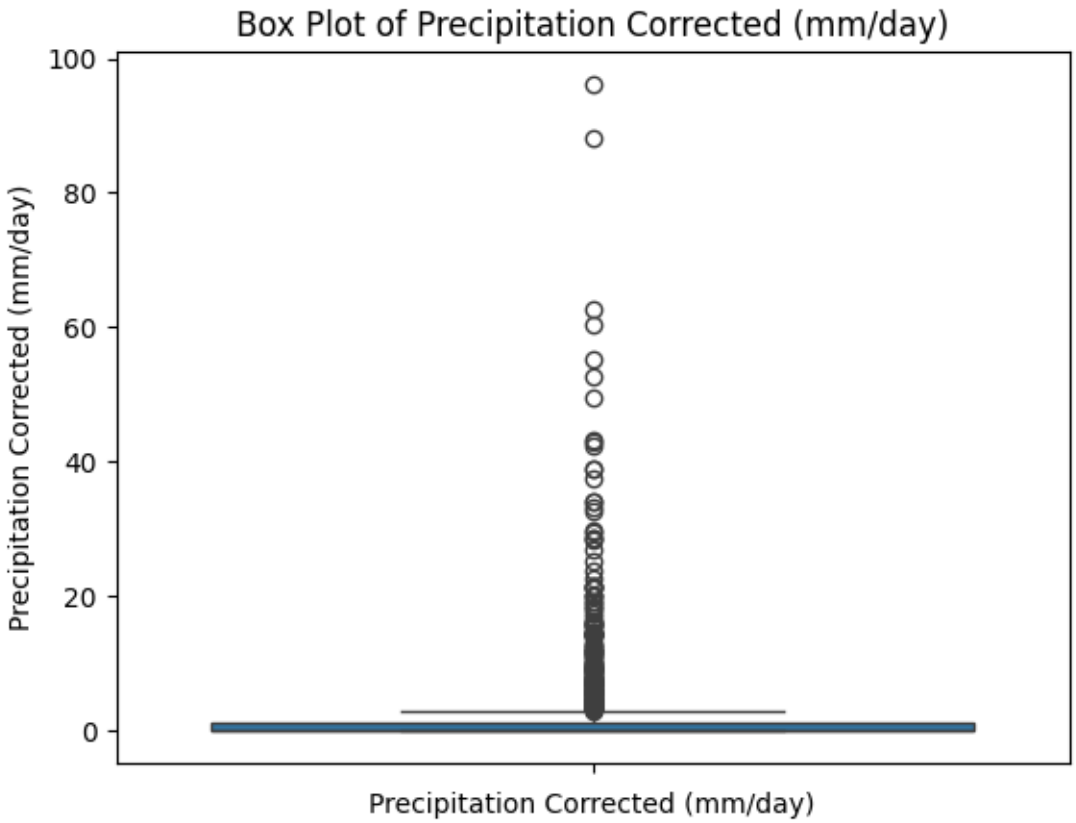

Box Plot of Dew/Frost Point at 2 Meters (C)
Dew/Frost Point at 2 Meters (C)
25
20
15
10
5
0
−5
Dew/Frost Point at 2 Meters (C)

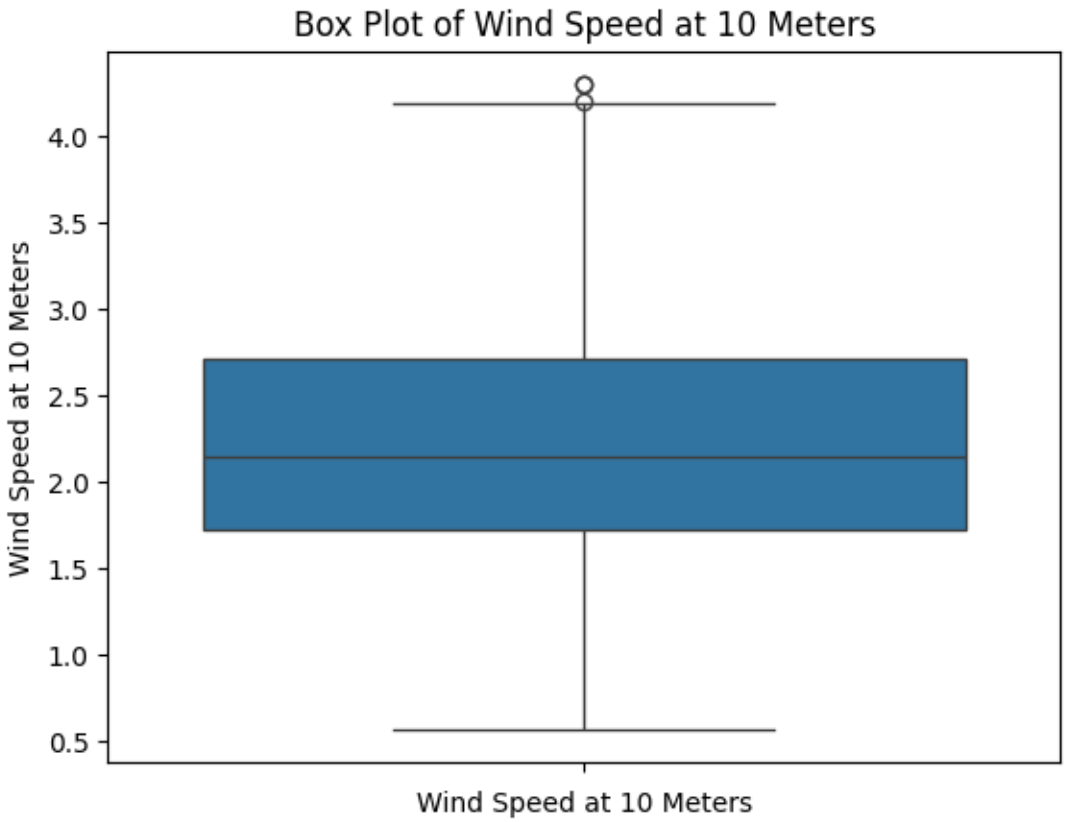

Box Plot of Temperature at 2 Meters Maximum (C)
Temperature at 2 Meters Maximum (C)
45
40
35
30
25
20
15
Temperature at 2 Meters Maximum (C)


Box Plot of Temperature at 2 Meters Minimum ©
Temperature at 2 Meters Minimum ©
35
30
25
20
15
10
5
0
Temperature at 2 Meters Minimum ©


Box Plot of Specific Humidity at 2 Meters (g/kg)
Specific Humidity at 2 Meters (g/kg)
20
15
10
5
Specific Humidity at 2 Meters (g/kg)


Box Plot of Relative Humidity at 2 Meters (%)
Relative Humidity at 2 Meters (%)
80
60
40
20
Relative Humidity at 2 Meters (%)


Box Plot of Precipitation Corrected (mm/day)
Precipitation Corrected (mm/day)
100
80
60
40
20
0
Precipitation Corrected (mm/day)



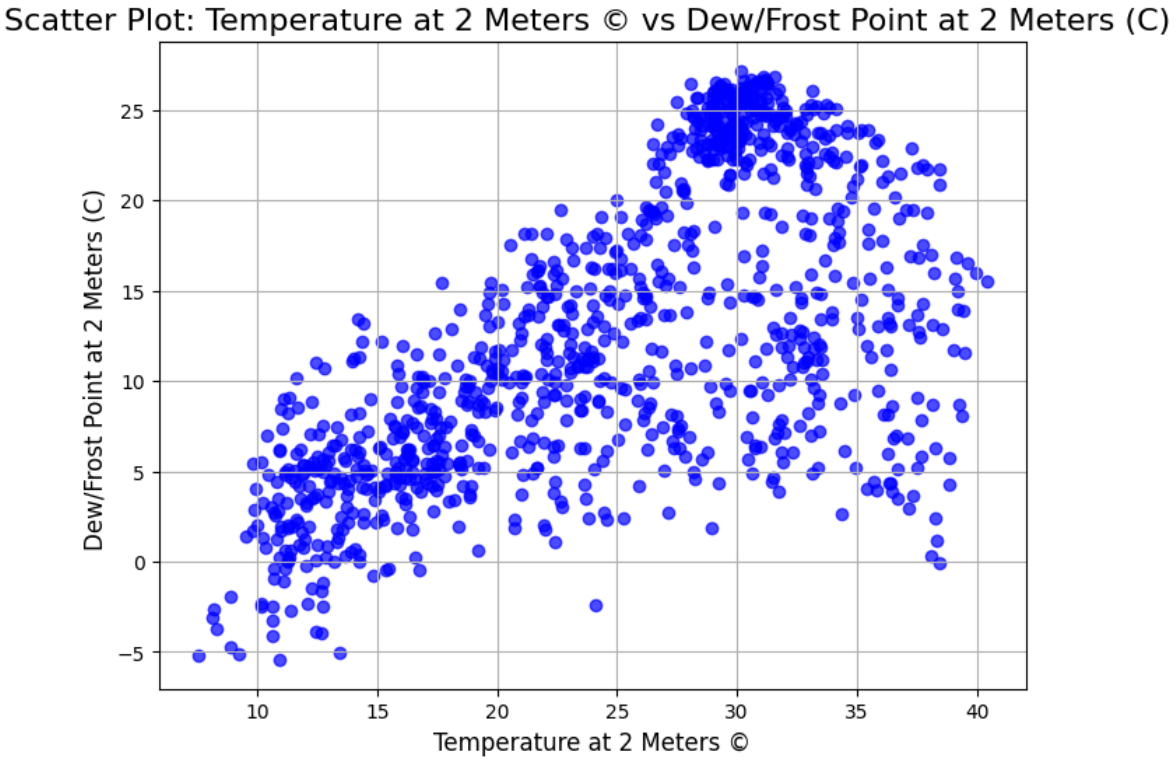

3. **Bivariate and Multivariate Analysis**

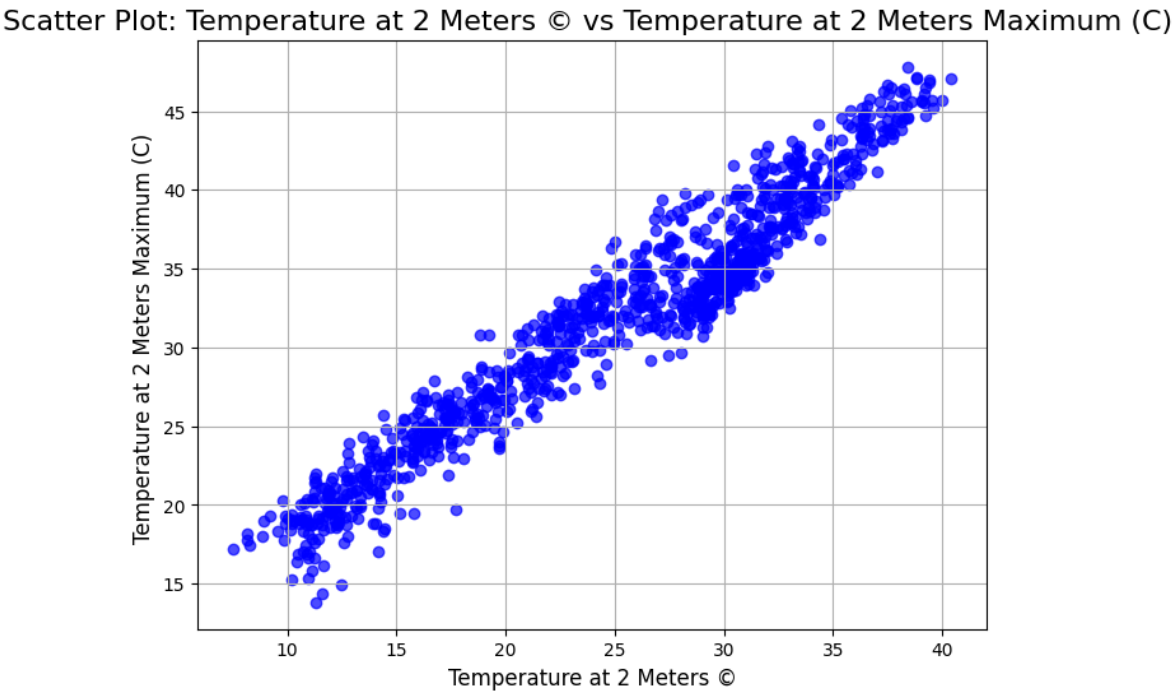
   The scatter plot indicates a positive correlation between Temperature at 2 Meters and Dew/Frost Point at 2 Meters. As the temperature increases, the dew/frost point also tends to increase. There is a noticeable cluster of data points around the temperature of 25°C and dew/frost point of 15°C to 20°C, suggesting that these conditions are common in the dataset. The scatter plot shows a strong positive correlation between Temperature at 2 Meters, Maximum, and Minimum. As the temperature at 2 meters increases, the maximum temperature at 2 meters also tends to increase. This positive correlation implies that higher temperatures at 2 meters are associated with higher maximum temperatures at the same height. Furthermore, the scatter plot indicates a positive correlation between temperature and specific humidity. As the temperature increases, the specific humidity also tends to increase. The data points are more densely packed between temperatures of

20°C to 30°C and specific humidity values of 5 g/kg to 15 g/kg, suggesting that most observations fall within this range. The scatter plot indicates a negative correlation between temperature and relative humidity. As the temperature increases, the relative humidity tends to decrease. For the relation between Temperature vs Precipitation and Wind speed there is very low correlation which can be considered as negligible. The scatter plot between Dew/Frost and Temperature at 2 Meter maximum shows positive correlation but not strong, as confirmed by correlation matrices value. However, the scatter plot of Dew/Frost shows a highly positive correlation with Temperature at 2 Meter minimum and Specific Humidity at 2 Meter i.e. 0.76 and 0.98. The plot with Relative humidity and Precipitation shows positive correlation but not so strong and with Wind Speed the plot shows negative correlation as when Dew/Frost increases wind speed decreases. Scatter plot of Temperature at 2 Meters maximum with Temperature at 2 Meters minimum and Specific humidity shows positive correlation with strong correlation with Temperature at 2 meters minimum. However, with Relative humidity scatter plot shows negative correlation and it can be verified by correlation matrices (-0.44). This feature has positive but quite weak correlation with Precipitation and Wind Speed. Additionally, the plot of Temperature at 2 meters minimum with Specific humidity shows positive correlation i.e. 0.72 but negative correlation with Relative humidity as when one variable increases other decreases. The scatter plot with Precipitation and Wind Speed shows positive relation but quite weak. Now for the Specific Humidity, its scatter plot with Relative humidity and Precipitation Corrected shows positive correlation as the both variables increase together. However, its scatter plot with Wind Speed shows negative correlation as supported by correlation metrices value i.e. -0.13. The scatter plot of Relative Humidity with Precipitation shows positive correlation as both increases together but it shows negative correlation with Wind Speed feature. The

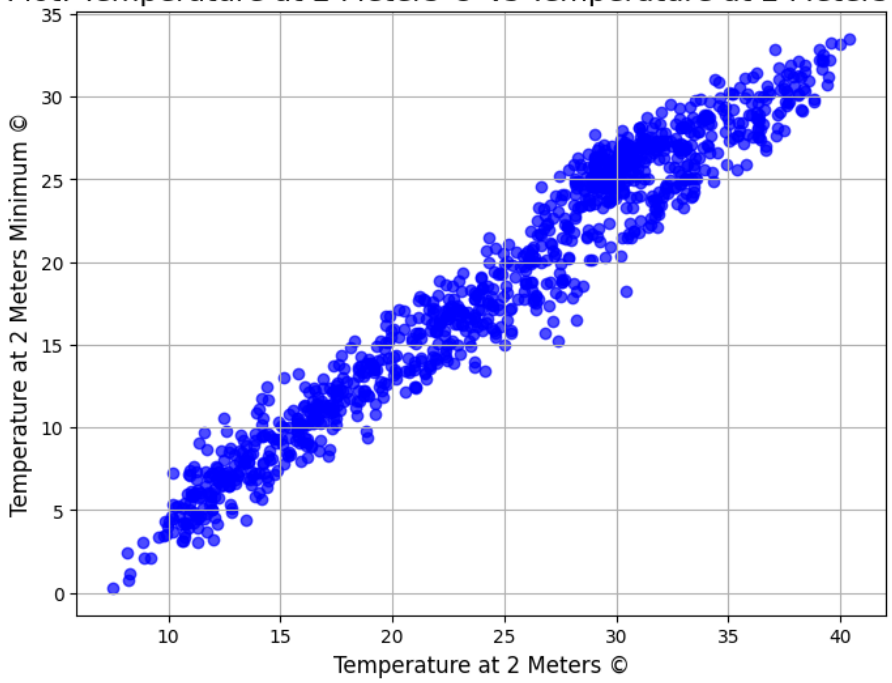
scatter plot of Precipitation Corrected and Wind Speed shows positive correlation but it is not quite strong i.e. 0.11.

## Scatter Plots:

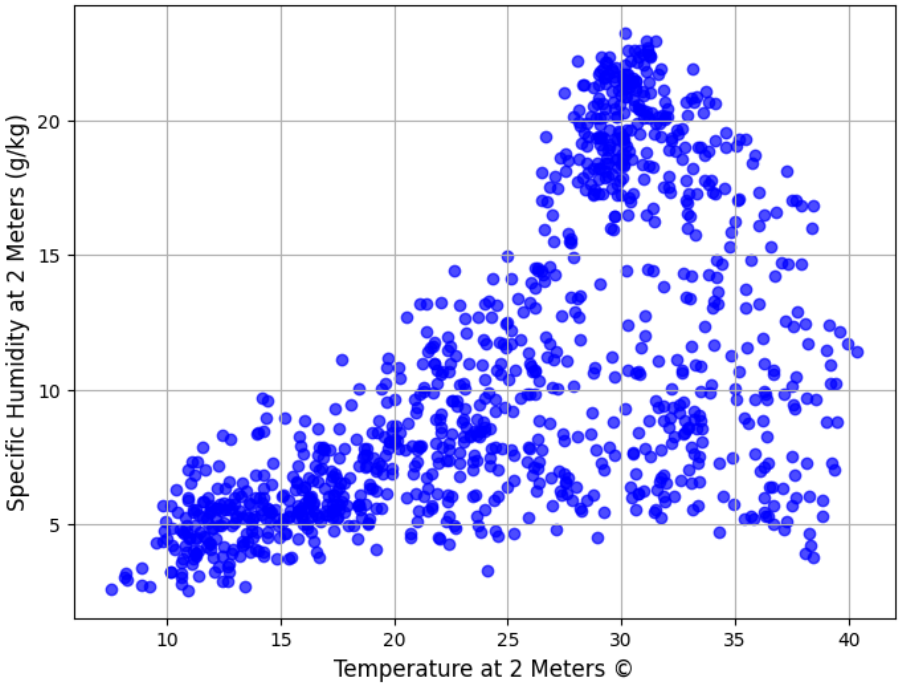


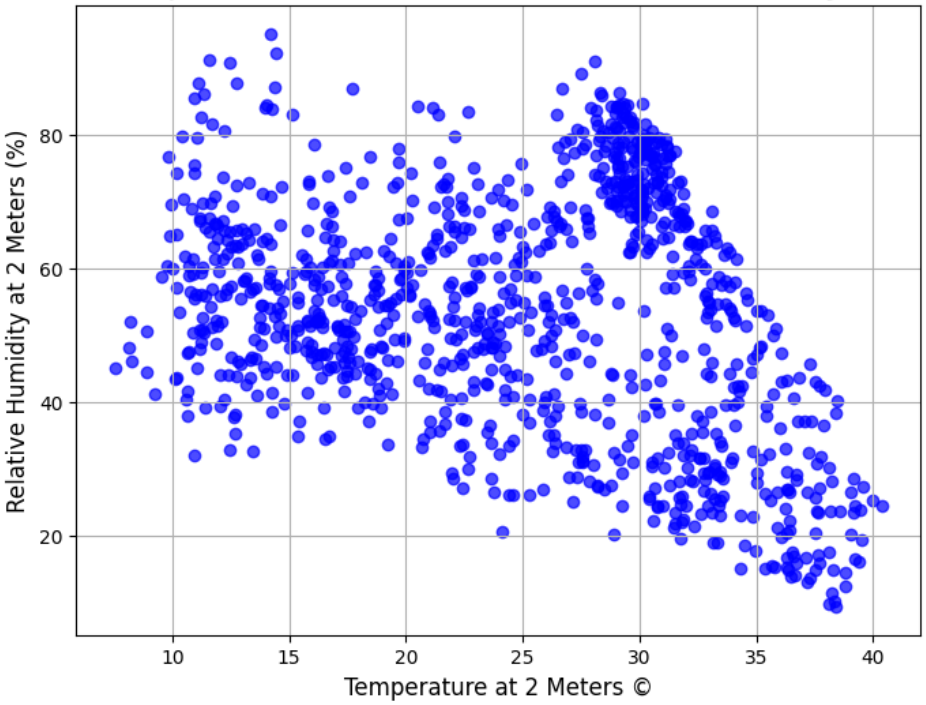
Scatter Plot: Temperature at 2 Meters © vs Temperature at 2 Meters Minimum ©



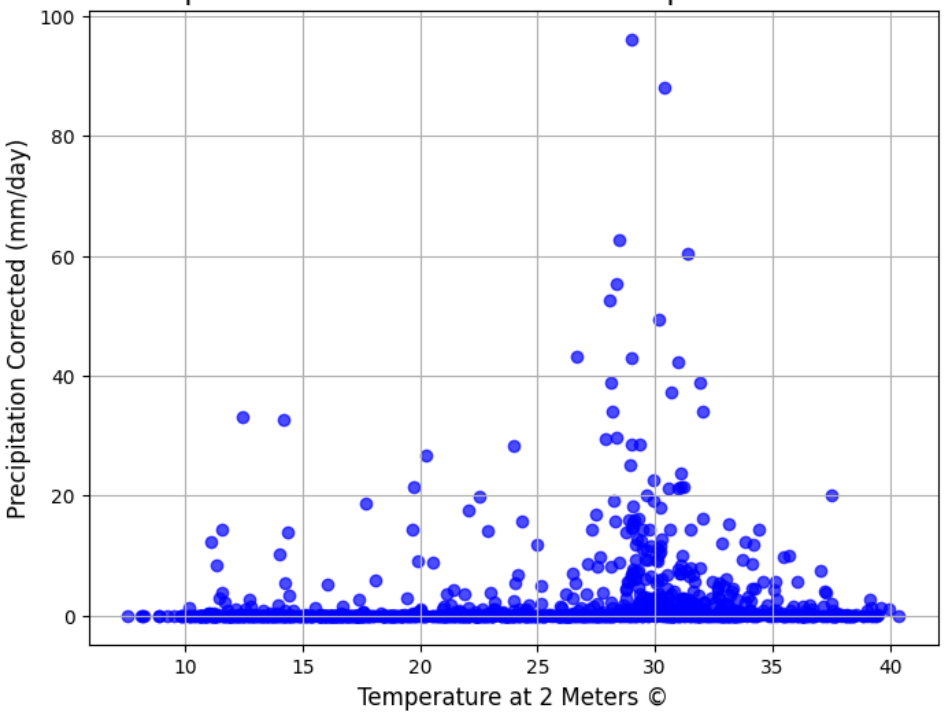

Scatter Plot: Temperature at 2 Meters © vs Specific Humidity at 2 Meters (g/kg)



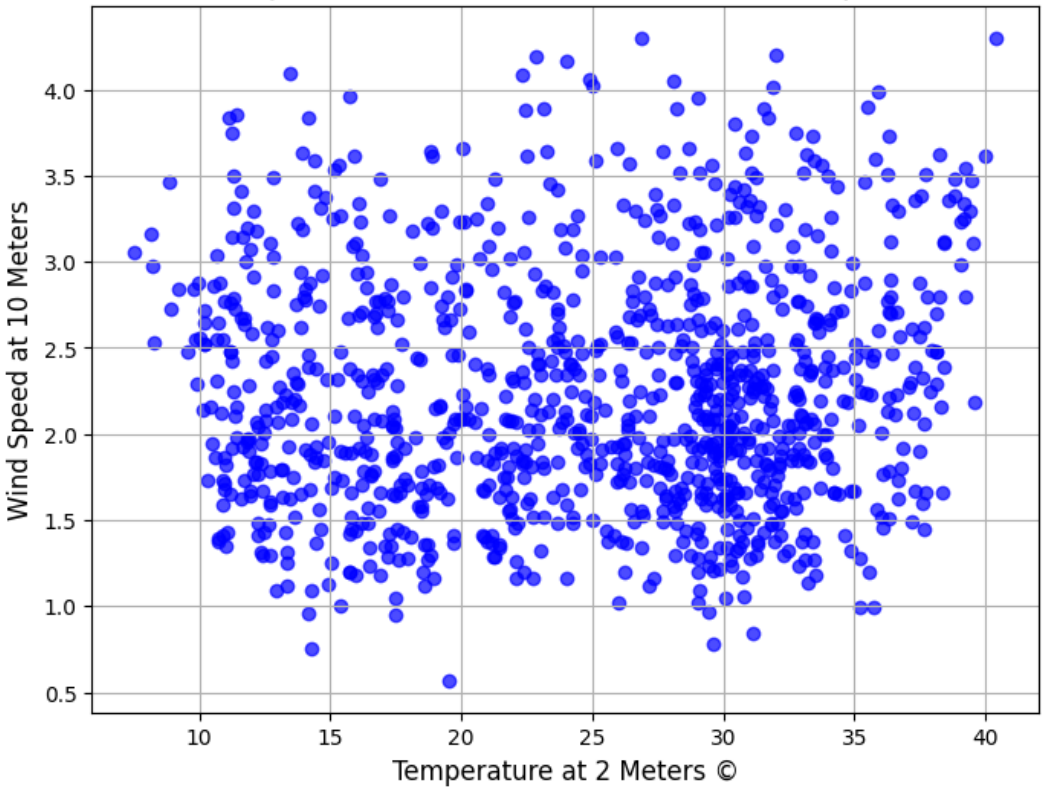
Scatter Plot: Temperature at 2 Meters © vs Relative Humidity at 2 Meters (%)



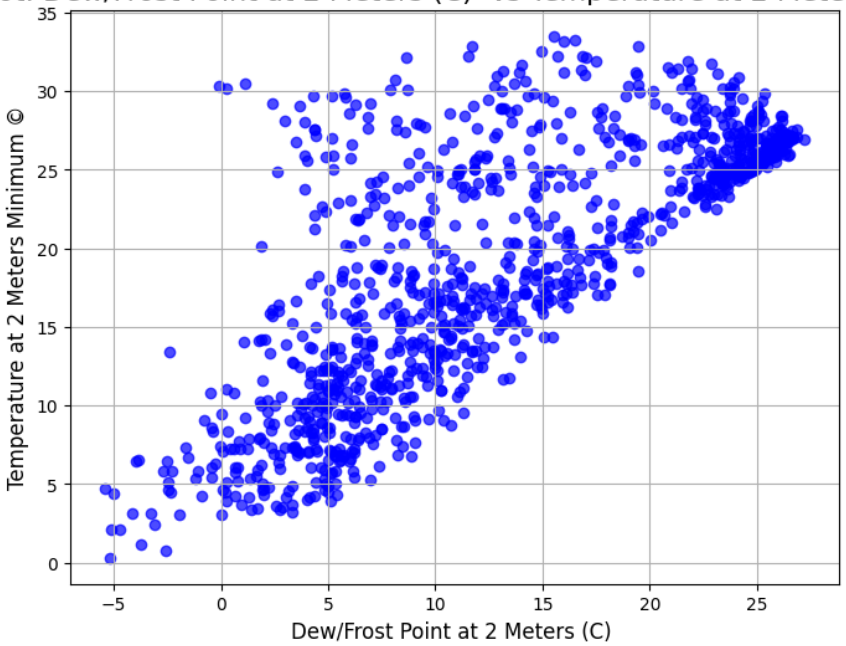

Scatter Plot: Temperature at 2 Meters © vs Precipitation Corrected (mm/day)



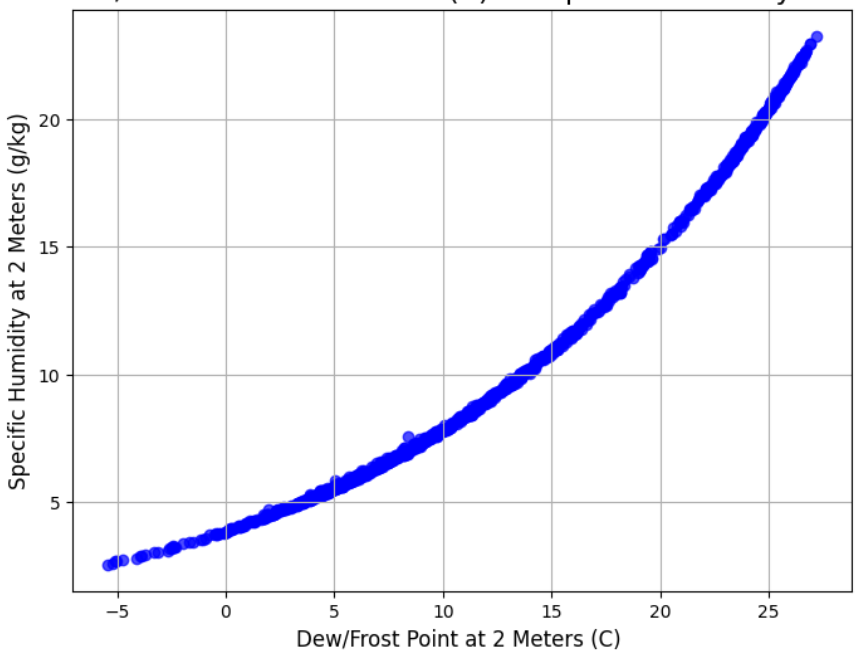
Scatter Plot: Temperature at 2 Meters © vs Wind Speed at 10 Meters



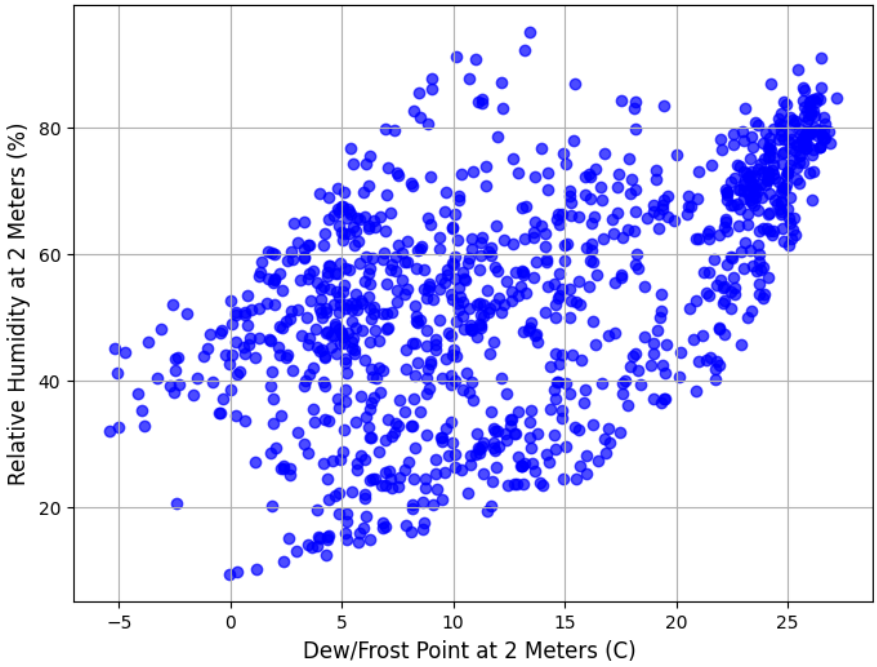

Scatter Plot: Dew/Frost Point at 2 Meters (C)  vs Temperature at 2 Meters Minimum ©



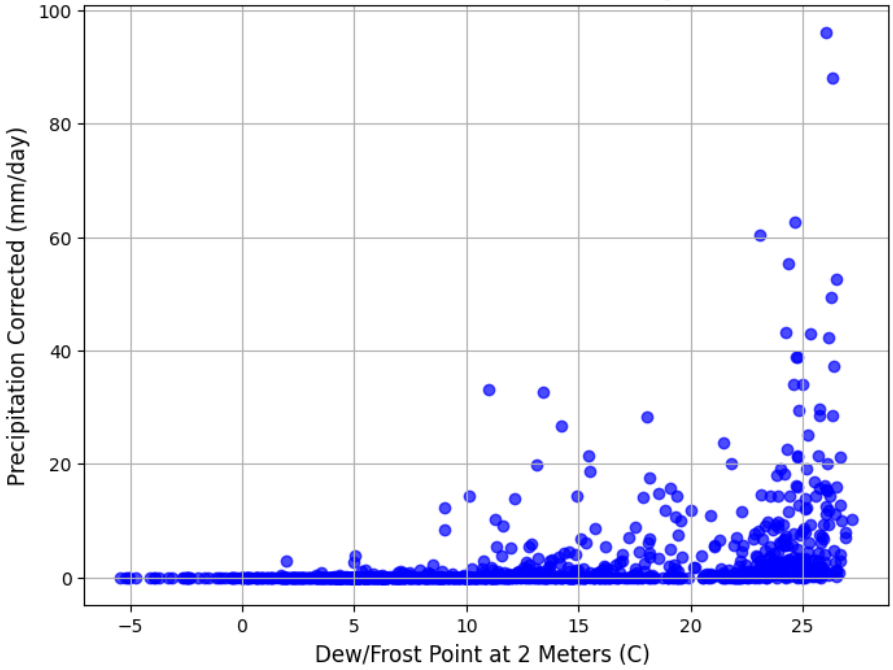
Scatter Plot: Dew/Frost Point at 2 Meters (C)  vs Specific Humidity at 2 Meters (g/kg)



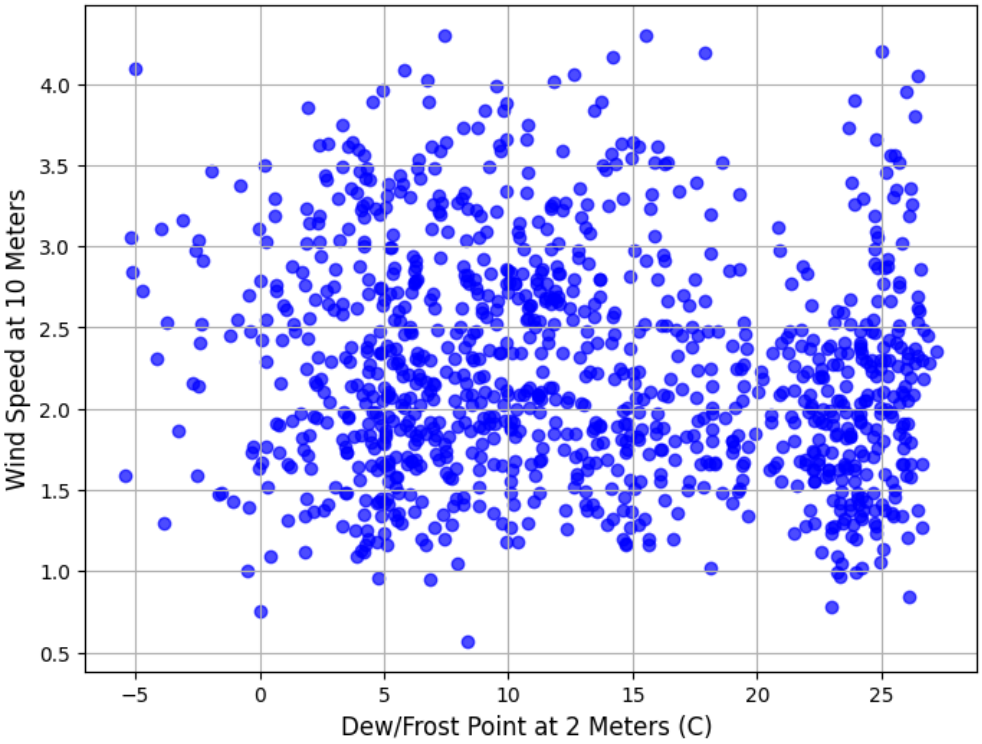

Scatter Plot: Dew/Frost Point at 2 Meters (C)  vs Relative Humidity at 2 Meters (%)



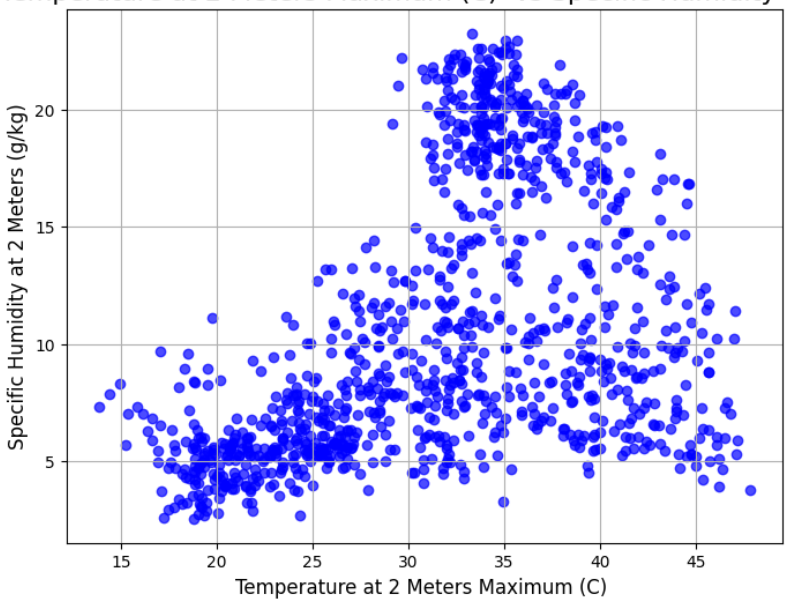
Scatter Plot: Dew/Frost Point at 2 Meters (C)  vs Precipitation Corrected (mm/day)



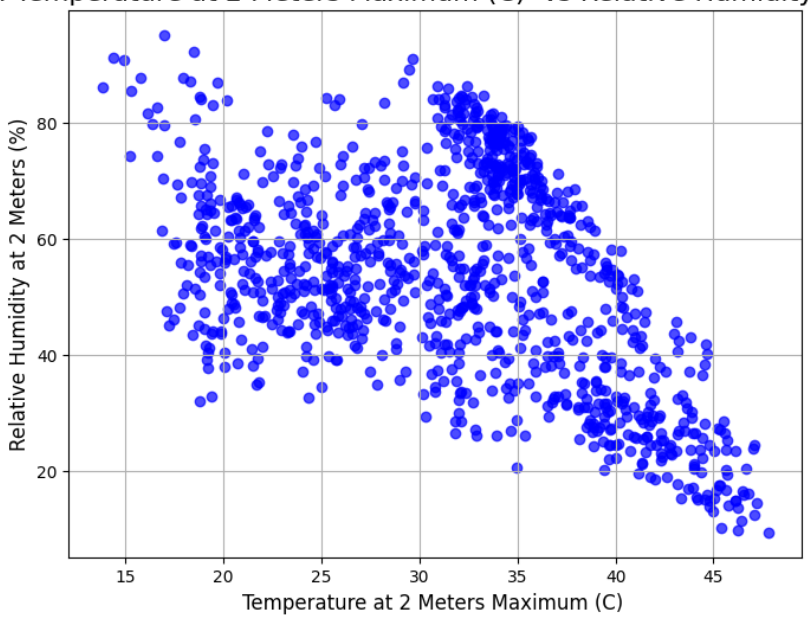

Scatter Plot: Dew/Frost Point at 2 Meters (C)  vs Wind Speed at 10 Meters



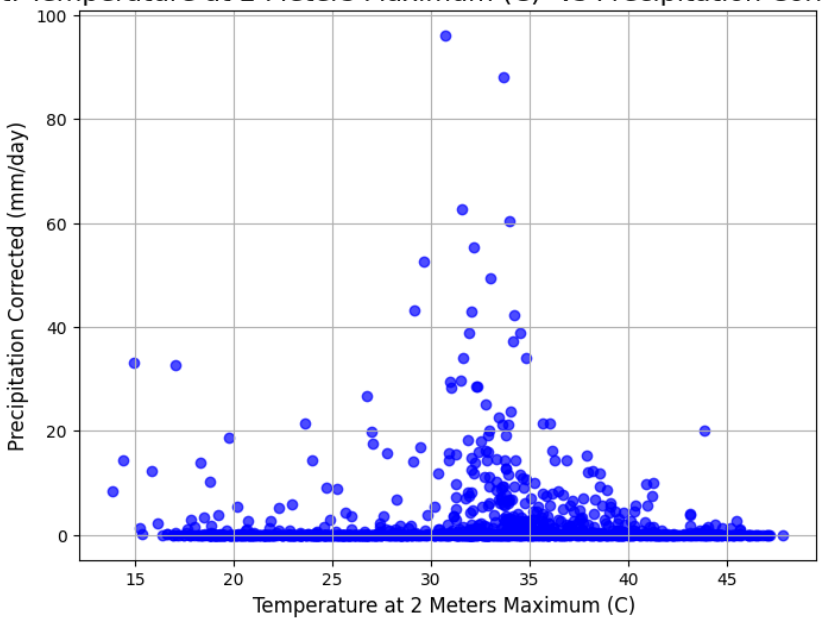
Scatter Plot: Temperature at 2 Meters Maximum (C)  vs Specific Humidity at 2 Meters (g/kg)



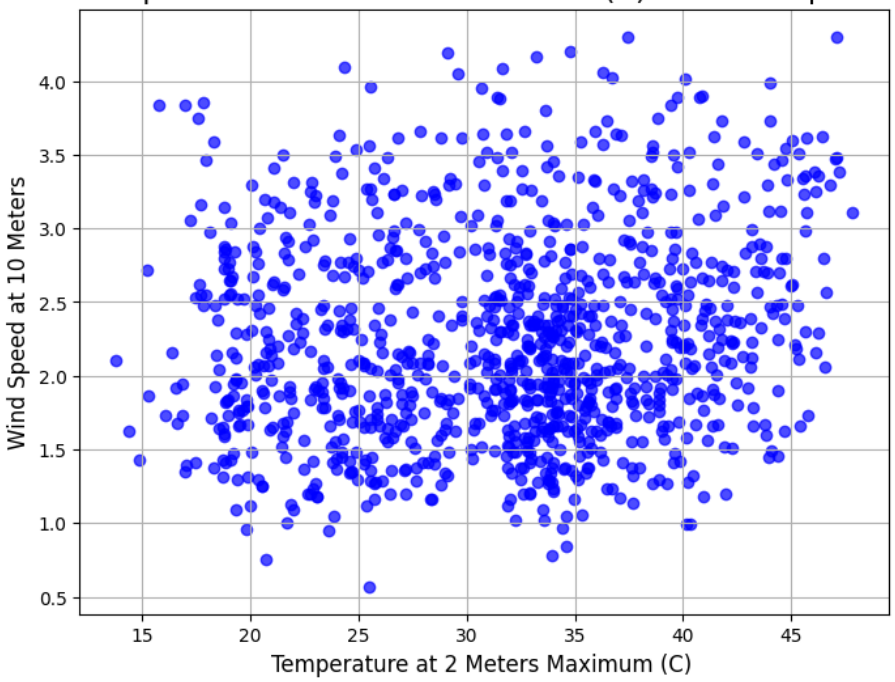

Scatter Plot: Temperature at 2 Meters Maximum (C)  vs Relative Humidity at 2 Meters (%)



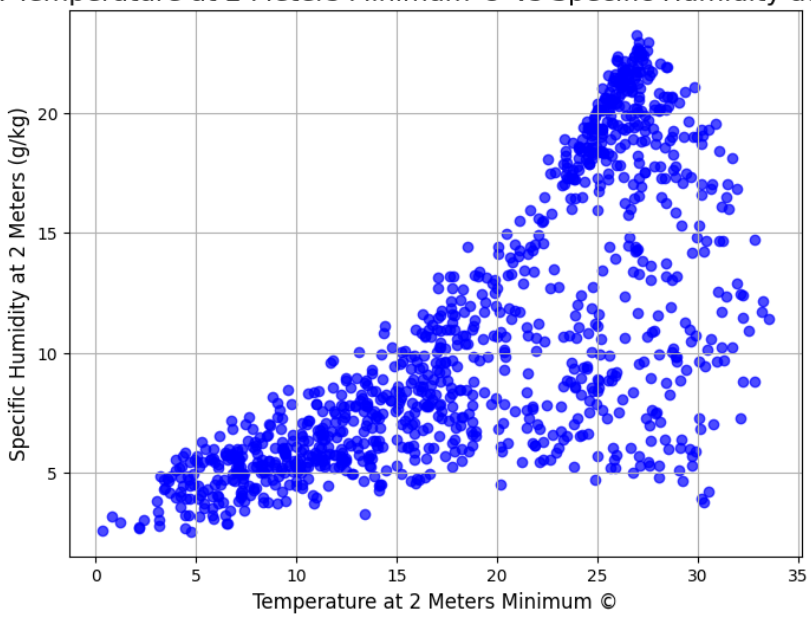
Scatter Plot: Temperature at 2 Meters Maximum (C)  vs Precipitation Corrected (mm/day)



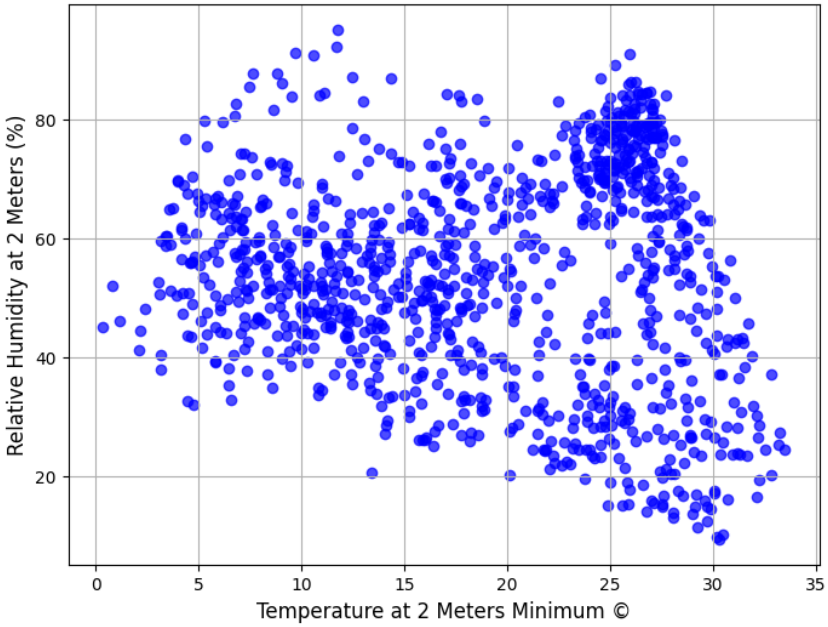

Scatter Plot: Temperature at 2 Meters Maximum (C)  vs Wind Speed at 10 Meters



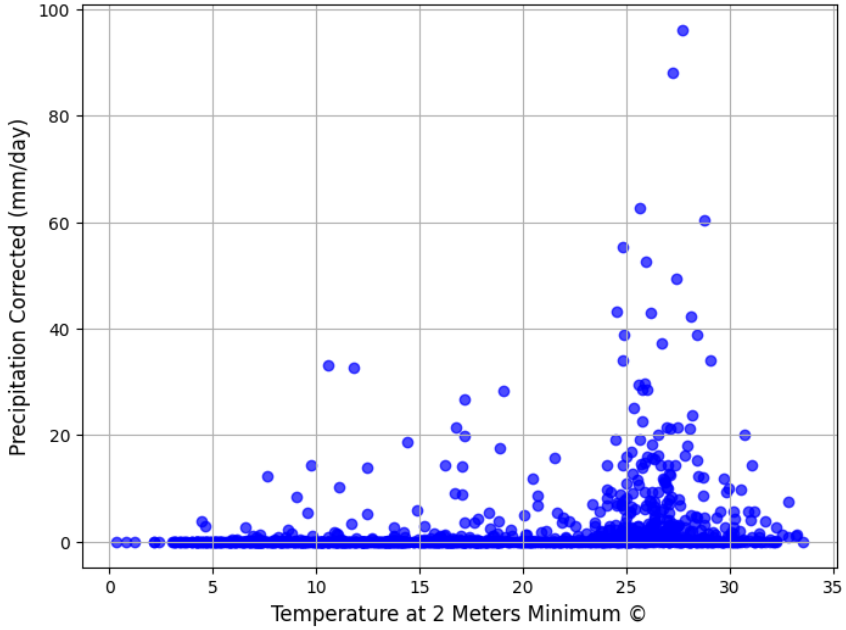
Scatter Plot: Temperature at 2 Meters Minimum © vs Specific Humidity at 2 Meters (g/kg)



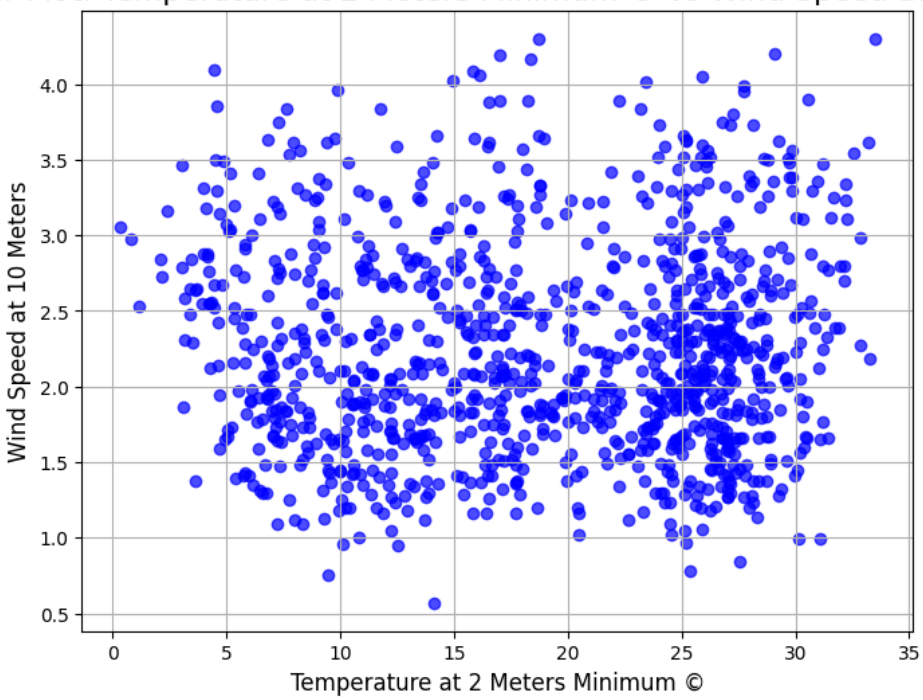

Scatter Plot: Temperature at 2 Meters Minimum © vs Relative Humidity at 2 Meters (%)



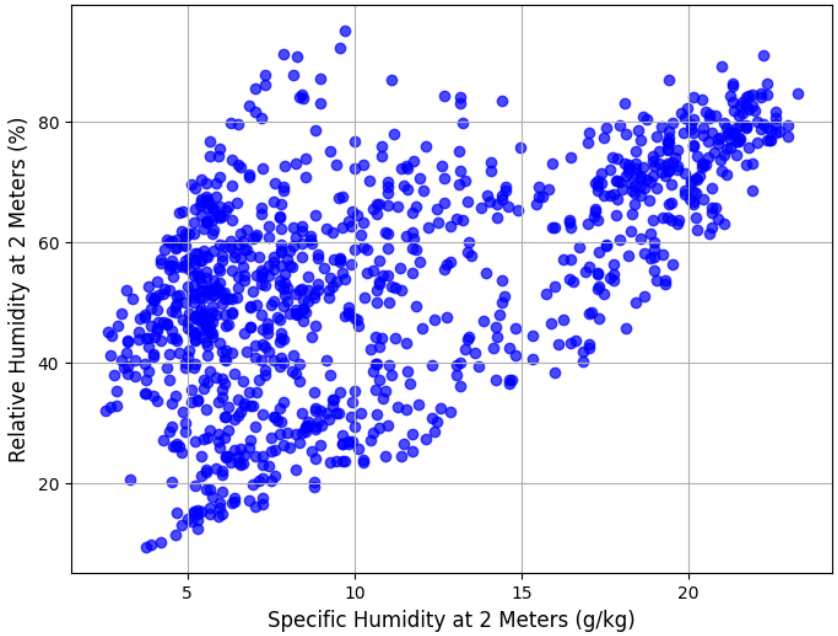
Scatter Plot: Temperature at 2 Meters Minimum © vs Precipitation Corrected (mm/day)



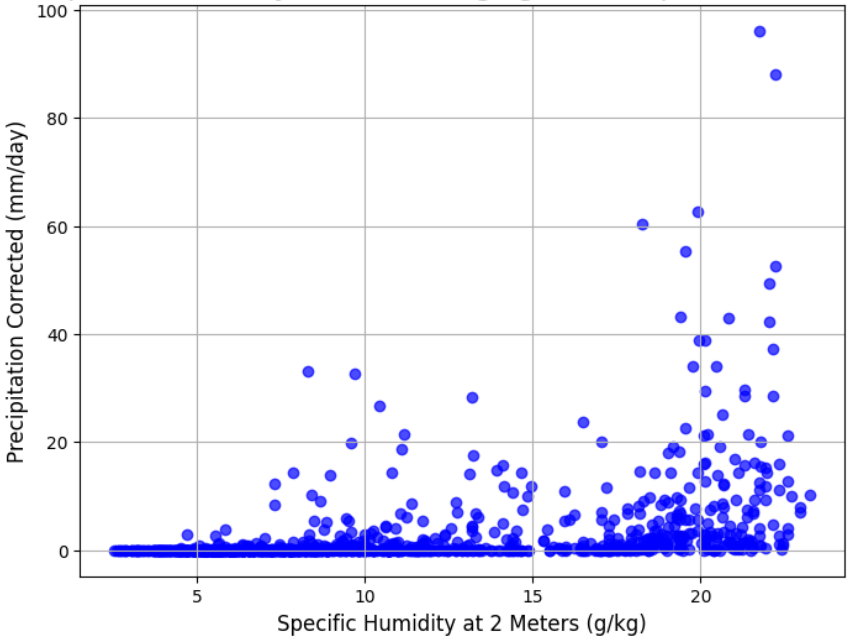

Scatter Plot: Temperature at 2 Meters Minimum © vs Wind Speed at 10 Meters



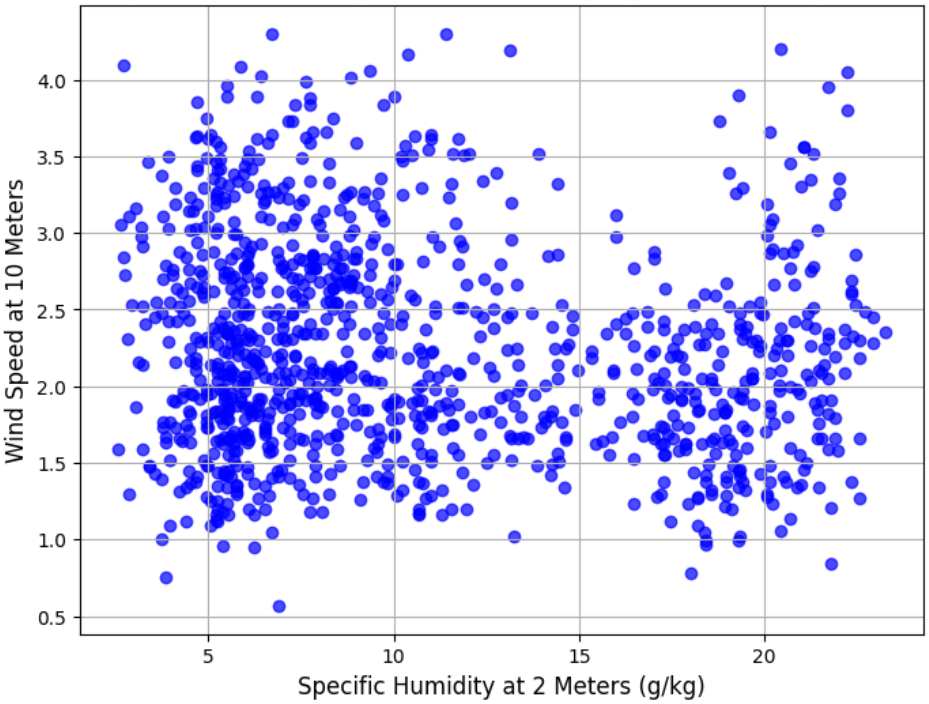
Scatter Plot: Specific Humidity at 2 Meters (g/kg) vs Relative Humidity at 2 Meters (%)



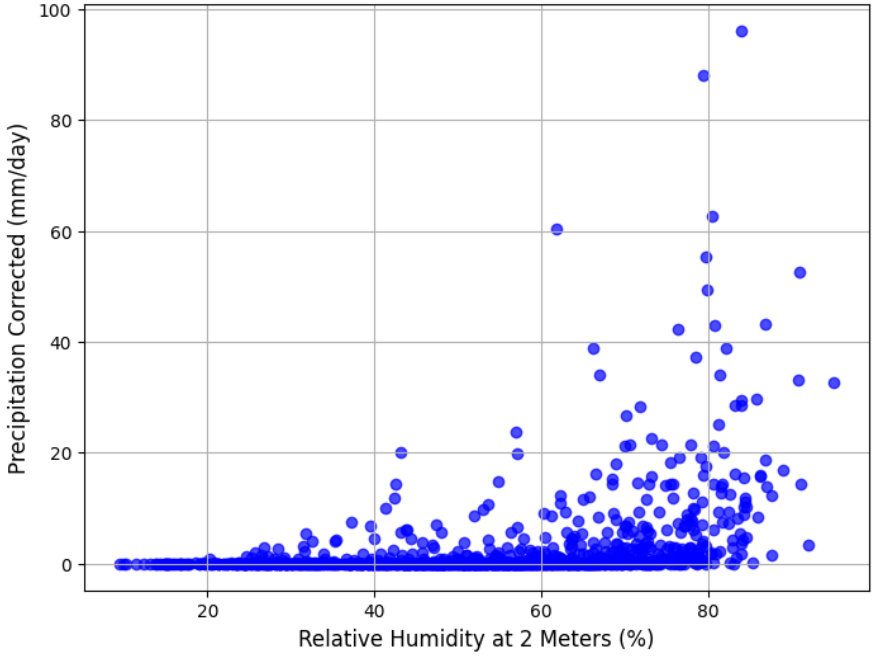

Scatter Plot: Specific Humidity at 2 Meters (g/kg) vs Precipitation Corrected (mm/day)



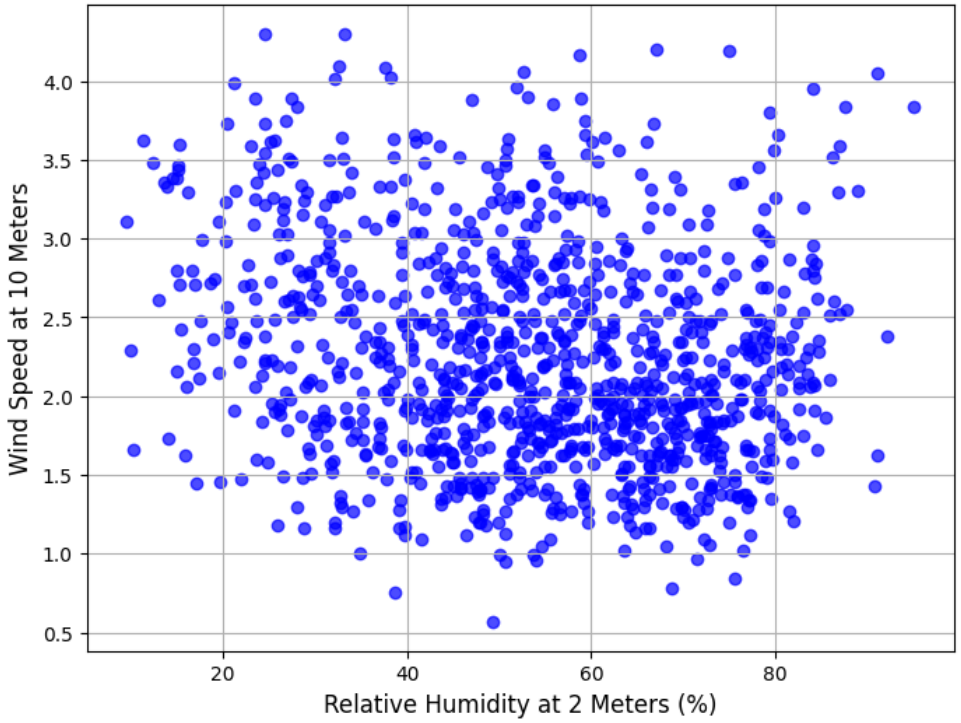
Scatter Plot: Specific Humidity at 2 Meters (g/kg) vs Wind Speed at 10 Meters



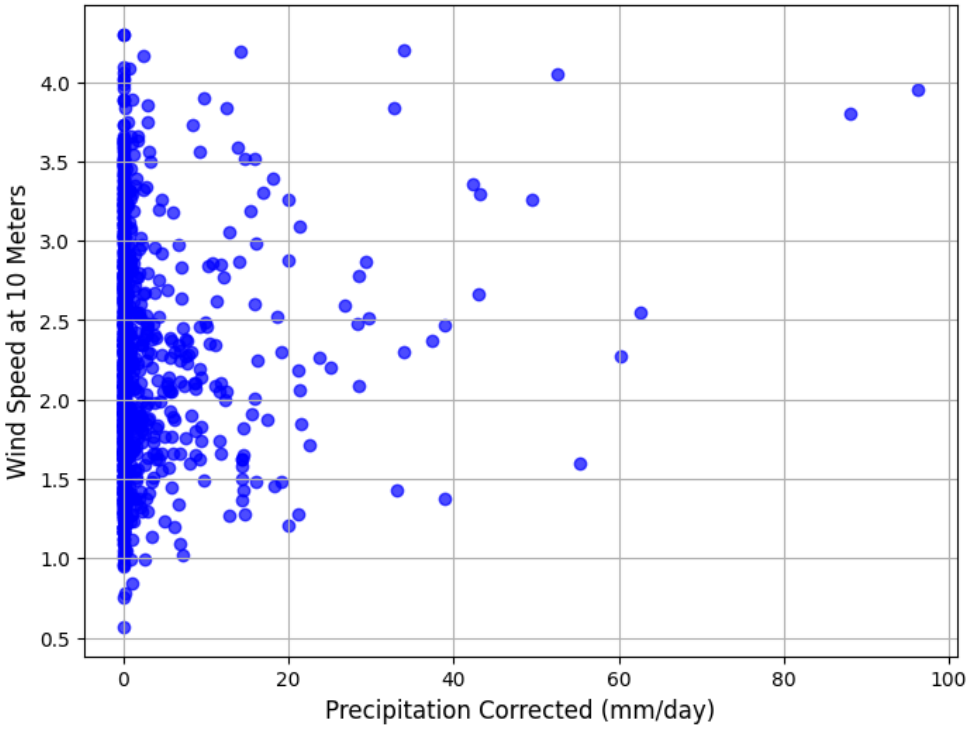

Scatter Plot: Relative Humidity at 2 Meters (%) vs Precipitation Corrected (mm/day)



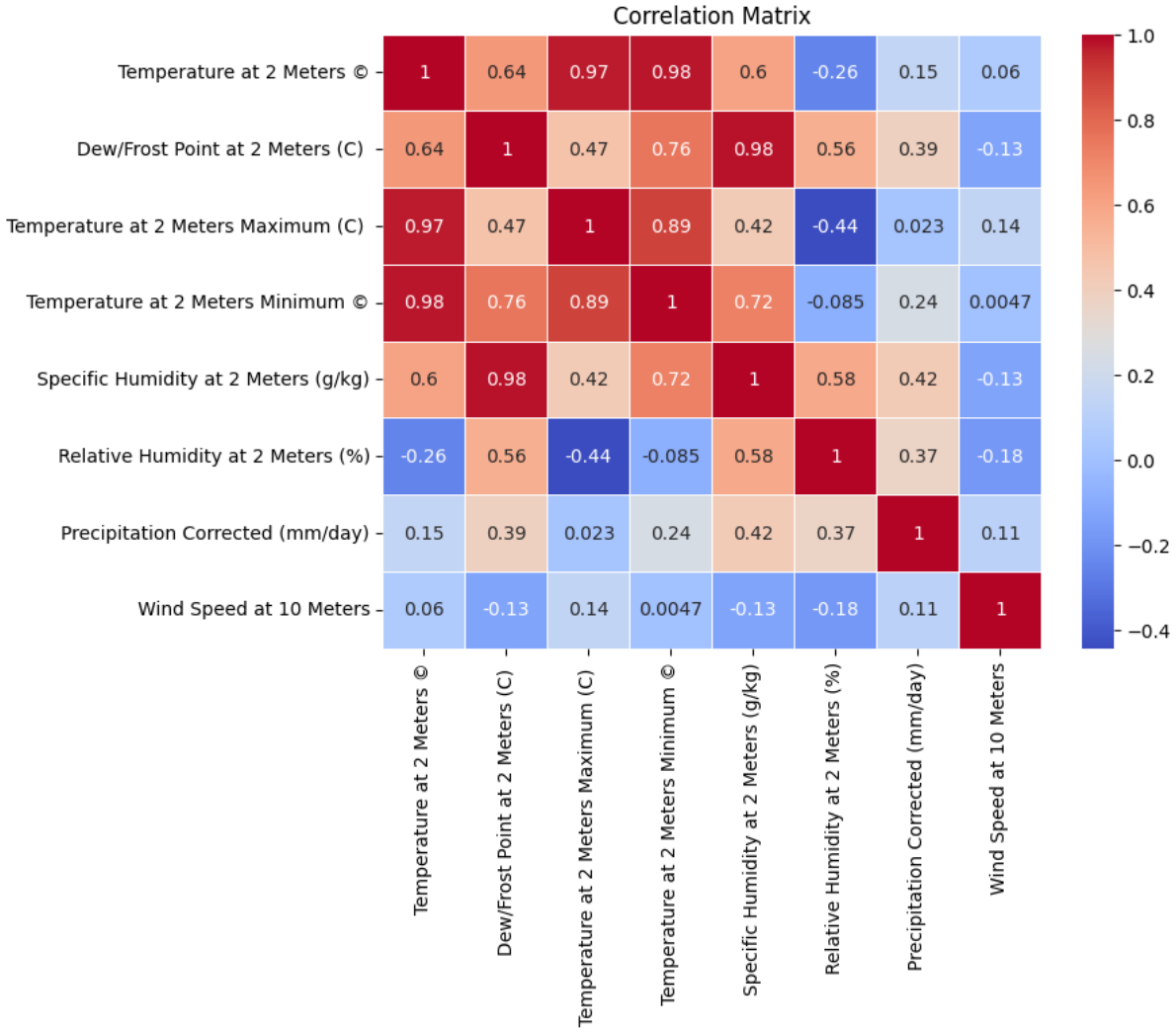
Scatter Plot: Relative Humidity at 2 Meters (%) vs Wind Speed at 10 Meters




Scatter Plot: Precipitation Corrected (mm/day) vs Wind Speed at 10 Meters



**Correlation Matrix:**




## 4. Feature Analysis

The feature importance values indicate how much each feature contributes to predicting the target variable in a model. In this dataset Specific Humidity at 2 Meters feature has the highest importance 0. 423028, indicating a strong relationship with the target variable which is Precipitation Corrected in our case. Dew/Frost Point at 2 Meters (C) is the second most important feature with correlation value of 0. 389070. Relative Humidity

feature shows a moderate importance indicating a direct but less dominant effect on the target compared to specific humidity and dew/frost point. Other features show quite low importance compared to other features. We can conclude that Specific Humidity, Dew/Frost point and Relative Humidity feature importance shows that these can be considered as significant features. On the other hand, features like Temperature at 2 Meters Maximum (C) and Wind Speed at 10 Meters can be considered as irrelevant features as they have very low correlation with the target variable. Temperature at 2 Meters Minimum and Temperature at 2 Meters can be considered as moderate features as they suggested moderate relationship with the target.

**Correlation Analysis w.r.t target:**

```
Specific Humidity at 2 Meters (g/kg)    0.423028
Dew/Frost Point at 2 Meters (C)         0.389070
Relative Humidity at 2 Meters (%)       0.367329
Temperature at 2 Meters Minimum ©       0.242463
Temperature at 2 Meters ©               0.146545
Wind Speed at 10 Meters                 0.112458
Temperature at 2 Meters Maximum (C)     0.023439
Name: Precipitation Corrected (mm/day), dtype: float64
```

## Correlation Matrix/HeatMap: