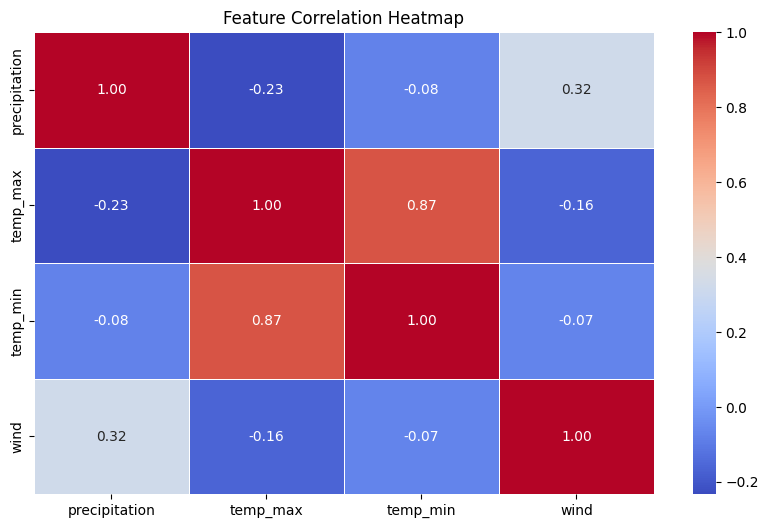
***ANALYSIS OF EDA***

1. ***Correlation Heatmap***

**Type of Analysis:** **Multivariate Analysis** (Multiple numerical features analyzed together)

**Goal**:

* Identify **relationships between numerical features** in the dataset.
* Determine **which variables most affect each other**.
* Help **feature selection** for machine learning.

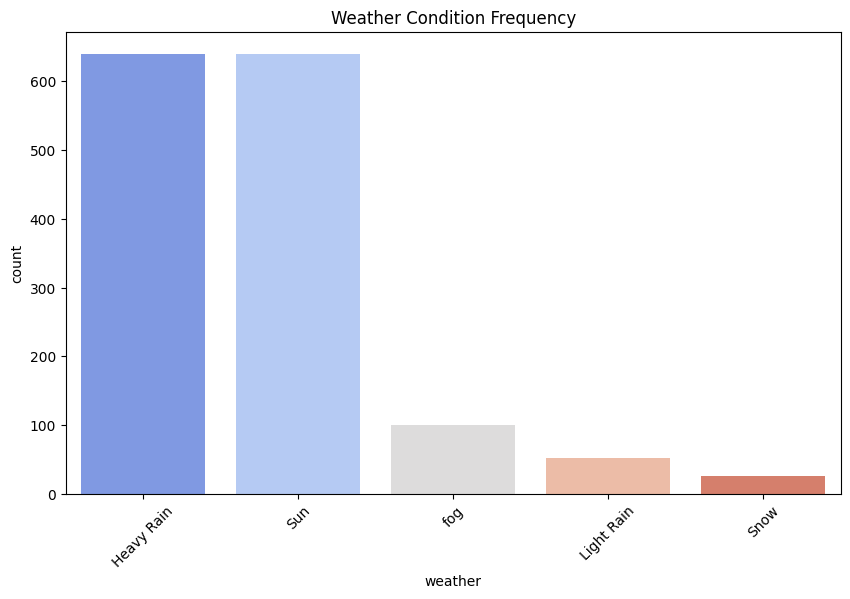
**Trends observed**:

* **Strong positive correlation (0.87) between temp\_max and temp\_min** → Hot days mostly have hot nights.
* **Moderate positive correlation (0.32) between wind speed and precipitation** → Heavy rainfalls are followed by strong winds, but not always.
* **Weak negative correlation (-0.23) between precipitation and max temperature** → Rainfall happens more frequently on cooler days but is not temperature driven.

**Deduction**:

* **Temperature trends are predictable**, but **rainfall is harder to forecast** as it depends on multiple variables.
* **Wind speed and rain have a moderate correlation**, but wind alone does not always lead to heavy rain.
* **Feature selection:** temp\_max, temp\_min, and precipitation are good predictors, and wind has lesser influence.

1. ***Weather Condition Frequency***



**Type of Analysis:** **Univariate Analysis** (Single categorical variable: weather)

**Goal:**

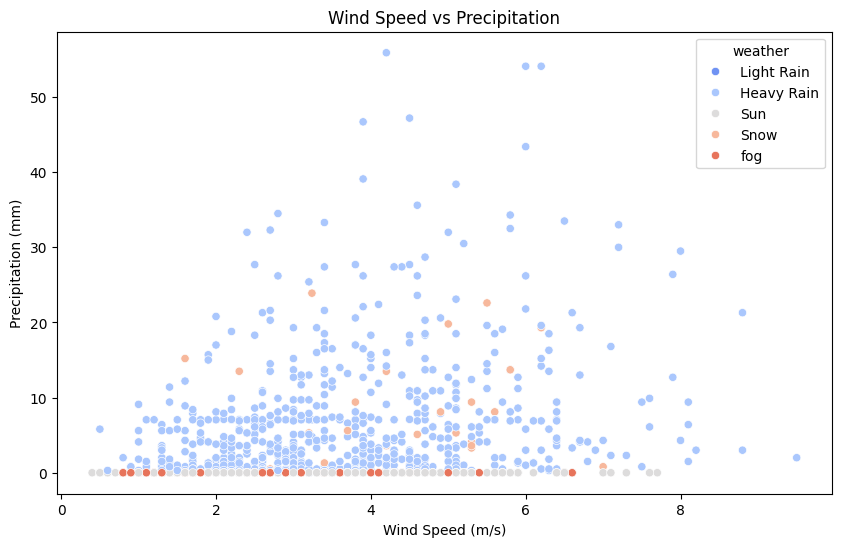
* **Comprehend the distribution of various weather types** within the dataset.
* Determine **the most common and rarest weather conditions**.

**Trends Observed:**

* **Light Rain is the most frequent occurring weather condition**.
* **Heavy Rain is less frequent but still appears regularly.**
* **Sunny days are frequent but happen in specific seasonal patterns.**
* **Fog and snow show up the least and hence are rare occurrences**.

**Deduction**:

* **London experiences a high frequency of light rain** instead of extreme weather.
* **Machine learning models may struggle to distinguish between light and heavy rain**, as they are very similar.
* **Rare weather conditions (fog, snow) may have fewer training samples**, affecting classification model performance.

1. ***Wind vs. Precipitation (Color-Coded by Weather Type)***

**Type of Analysis:** **Bivariate Analysis** (Wind Speed & Precipitation)

**Goal:**

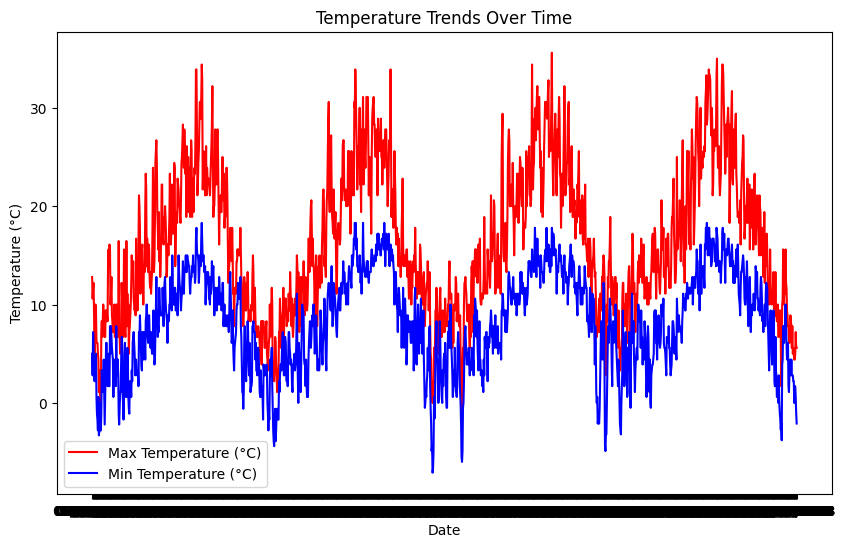
* Examine the link **between wind speed and precipitation**.
* Determine if **certain weather types are associated with high wind speeds and heavy rainfall**.

**Trends Observed**:

* A weak relationship is observed between wind speed and precipitation, suggesting they are **not strongly correlated** in this dataset.
* Most data points show **low wind speeds (0-15 km/h) and low precipitation rates (0-5 mm)**.
* There are certain points of extreme with **heavy precipitation (above 10 mm) and relatively high wind velocities (above 20 km/h)** exist.
* These could correspond to **strong weather conditions**, such as storms or heavy rain accompanied by gusty winds.
* A few scattered points have **very low wind velocities but heavy rainfall**, possibly indicating **calm rain events** or measurement inconsistencies.

**Deduction:**

* The scatter plot **does not have a significant trend** of wind speed versus precipitation. This suggests that **high wind speeds are not always associated with more precipitation** and conversely.
* A few **high-precipitation and high-wind-speed points** likely represent storms or heavy rainfall events.
* Some **low-wind yet high-precipitation** points suggest calm rain events.
* The presence of outliers could **skew predictive models**, especially if using regression techniques. A classification model should **account for these extreme events separately** rather than treating them as noise.

1. ***Time Series Plot: Max and Min Temperature Over Time***

**Type of Analysis:** **Bivariate Analysis** (Date vs. Temperature)

**Goal:**

* To analyze temperature fluctuations over time.
* Identify trends, seasonal variations, and possible anomalies.

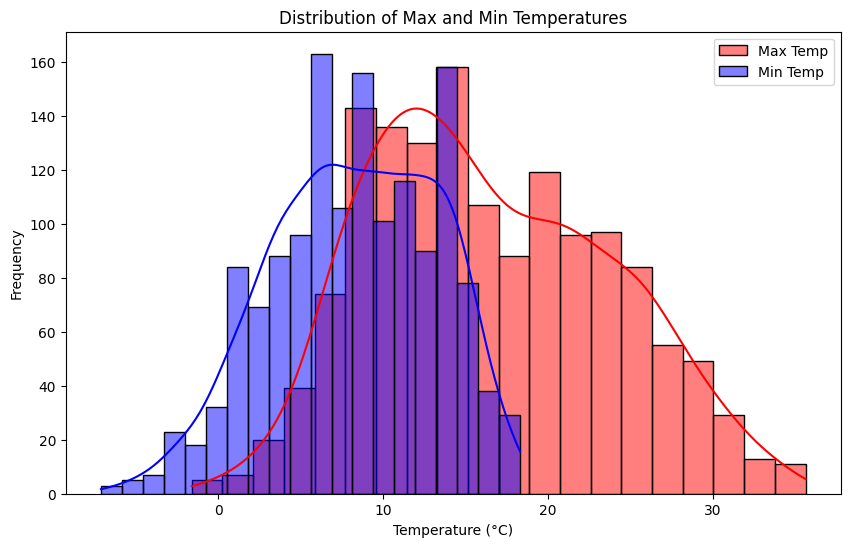
**Trends Observed**:

* Max and min temperature change daily, with obvious peaks and drops.
* Possible **seasonal variations** may exist, depending on data range.
* Some **outliers** (unusually high or low temperatures) may indicate extreme weather conditions.

**Deduction:**

* **Temperature follows expected seasonal changes** (increasing and decreasing trends).
* **Outliers indicate extreme temperature events**, worth investigating.
* **Trendlines (moving averages) may enhance clarity** and assist in identifying patterns like heatwaves or cold spells.

***5) Distribution of Max and Min Temperatures***



**Type of Analysis: Univariate Analysis** (Max and Min Temperature separately)

**Goal:**

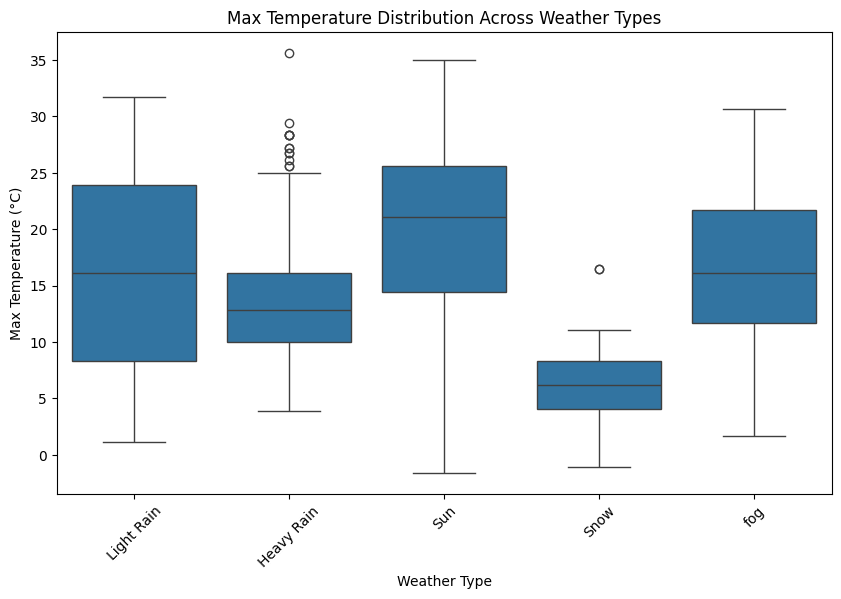
* To understand how temperatures are distributed.
* To find the **most common** temperature ranges.
* Detect **skewness** or anomalies in the data.

**Trends Observed:**

* The max temperature follows an approximately **normal distribution**, with most values grouped around a central range.
* The min temperature distribution may be somewhat skewed.
* The two distributions **may overlap**, but max temperatures tend to be higher.

**Deduction:**

* **The dataset has a reasonable temperature spread** with no extreme skewness.
* **Most temperature values fall within expected climate conditions** of the location.
* **Outliers in the tail regions** could represent heatwaves or cold snaps.
* Adding mean/median labels could provide better insights into the central tendency.

***6) Max Temperature Distribution Across Weather Types***

**Type of Analysis:** **Bivariate Analysis** (Categorical: Weather, Numerical: Max Temperature)

**Goal:**

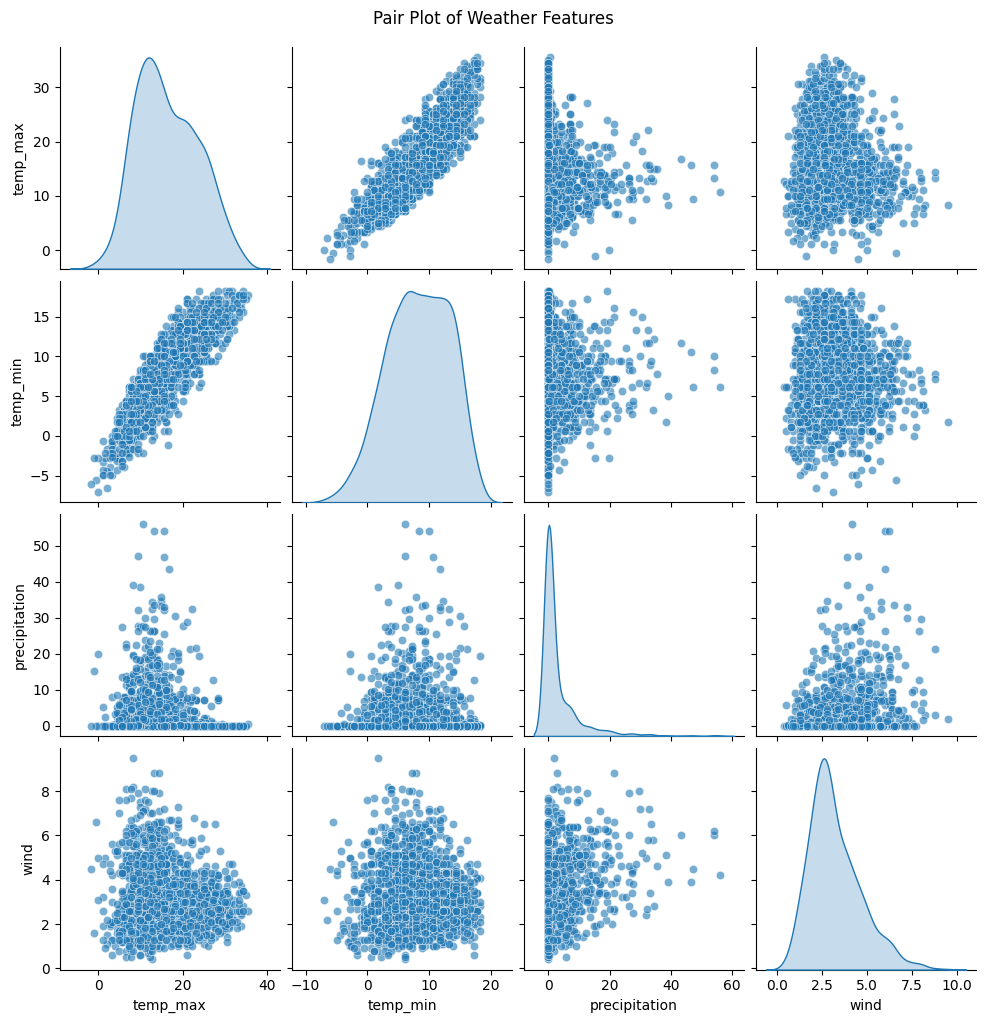
* To compare how different weather types affect max temperature.
* Identify **outliers** that signify unusual temperatures under certain weather conditions.

**Trends Observed:**

* Some weather types have a **wider spread of max temperatures** than others.
* Certain weather types, like **clear/sunny days, generally have higher max temperatures**.
* **Outliers in certain weather categories** suggest unexpected temperature events.

**Deduction:**

* Weather significantly influences max temperature.
* Extreme temperature values (outliers) may indicate data recording errors or rare weather events.
* **Sorting weather types by median temperature would enhance clarity.**
* Overlaying swarm plot would show the exact distribution of values.

***7) Relationships Between Weather Variables***

**Type of Analysis: Multivariate Analysis** (Multiple numerical variables)

**Goal:**

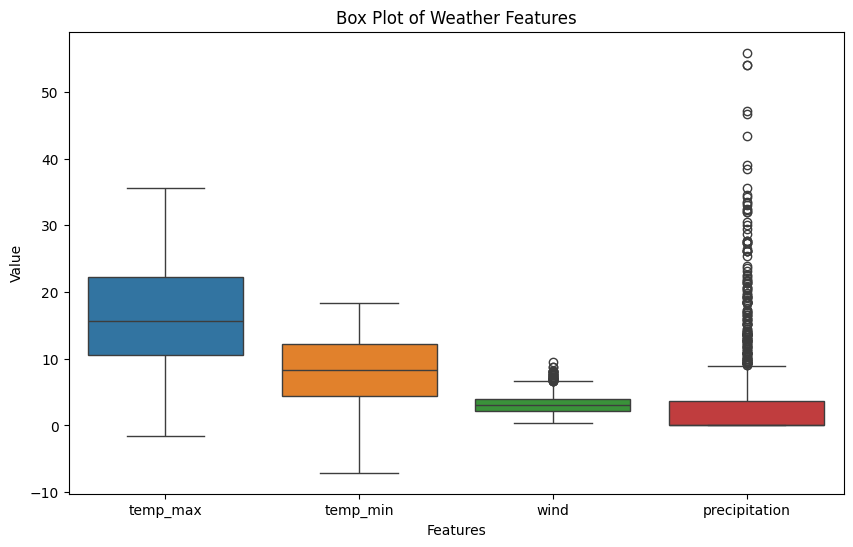
* To explore **relationships between numerical weather variables** (e.g., temperature, wind, precipitation).
* Detect correlations and potential **patterns useful for predictive modeling**.

**Trends Observed:**

* **Temp\_max and temp\_min is strongly correlated** (expected).
* **Wind speed does not strongly correlate with temperature**, but extreme winds might coincide with extreme temperatures.
* **Precipitation seems independent of temperature**, suggesting rain events occur at varied temperatures.

**Deduction:**

* **Temp\_max and temp\_min move together**, confirming seasonal trends.
* Wind and precipitation do not show strong patterns with temperature.
* Further analysis is needed on extreme wind speeds and their impact.

***8) Temperature, Wind, and Precipitation Distribution***

**Type of Analysis: Univariate Analysis** (for each variable)

**Goal**:

* To **detect outliers** in temperature, wind, or precipitation.
* To compare the distribution of numerical weather variables.

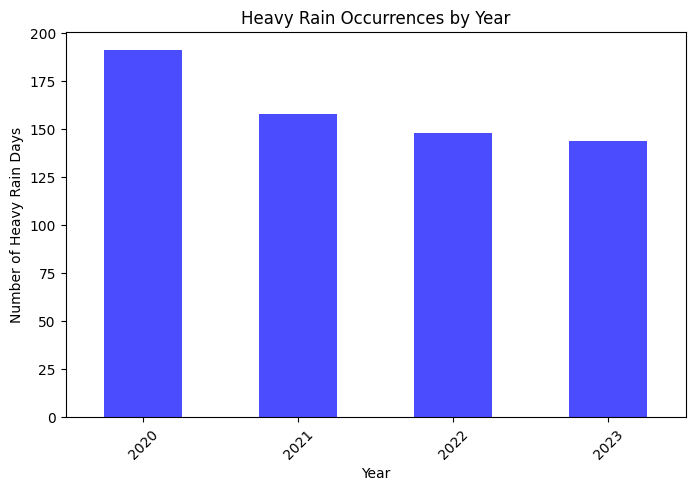
**Trends Observed:**

* **Max & Min Temperature**: Likely have some **outliers** (extremely high or low values).
* **Wind Speed**: May have **a few extreme values**, possibly from storms or high-wind events.
* **Precipitation**: Expected to show **skewed distribution**, with rare heavy rainfall events appearing as outliers.

**Deduction:**

* Temperature distributions are mostly consistent, but outliers indicate possible heatwaves or cold spells.
* **Wind data may have occasional extreme events** (e.g., storms).
* **Precipitation shows extreme outliers**, confirming that heavy rain is a rare but impactful event.
* **Further investigation is needed into extreme precipitation events** (e.g., did they coincide with storms?).

***9) Heavy Rain Occurrences per Year***

******

**Type of Analysis: Univariate Time-Series Analysis**

**Goal**:

* To analyze **how often heavy rain occurs** on a yearly basis.
* Identify **trends** in heavy rain frequency over time.

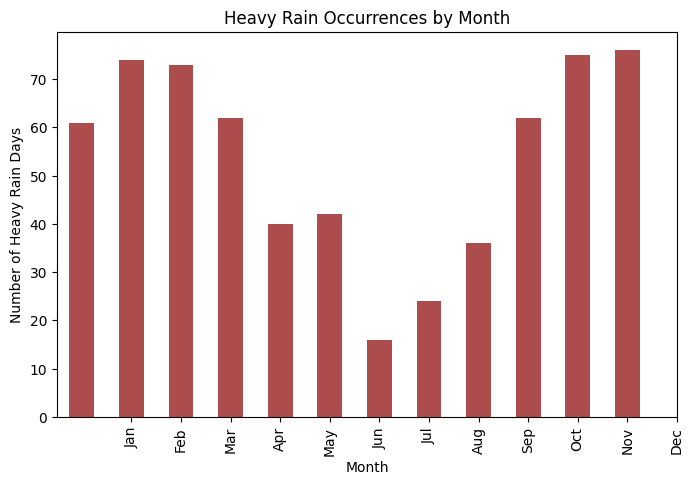
**Trends Observed:**

* The **number of heavy rain days fluctuates** from year to year.
* **Trends** (increasing or decreasing) may indicate climate shifts or data inconsistencies.

**Deduction:**

* If there's an increasing trend, it could suggest climate change effects or shifting weather patterns.
* If the trend is decreasing, it may indicate drier years or variations in reporting.
* Further analysis could explore correlations with global climate data.

***10) Heavy Rain Occurrences per Month***



**Type of Analysis: Univariate Seasonal Analysis**

**Goal:**

* To determine **which months, experience the heaviest rain**.
* Identify **seasonal trends in heavy rainfall**.

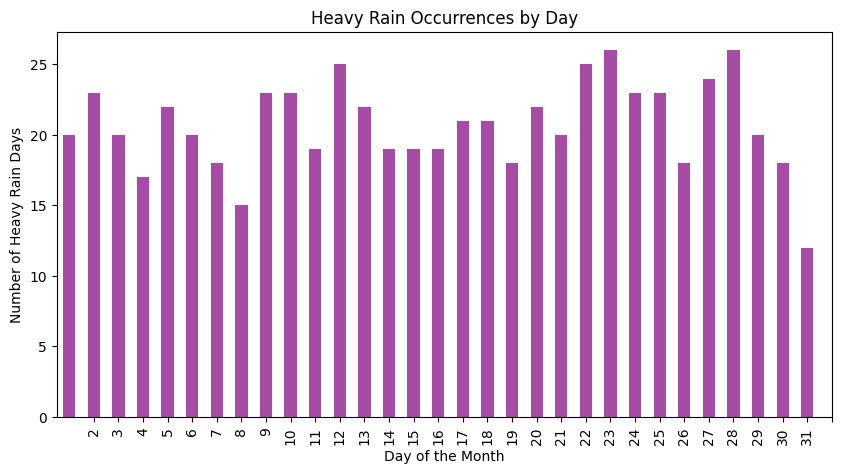
**Trends Observed:**

* **Certain months may show consistently higher occurrences** of heavy rain (e.g., monsoon season).
* Other months may show **little to no heavy rain**, confirming seasonal trends.

**Deduction:**

* If heavy rain is concentrated in certain months, it confirms clear seasonal trends.
* Months with little to no heavy rain suggest a distinct dry season.
* Comparing with wind data may indicate if storms are linked to heavy rain.

***11) Heavy Rain Occurrences by Day of the Month***



**Type of Analysis: Univariate Analysis**

**Goal**:

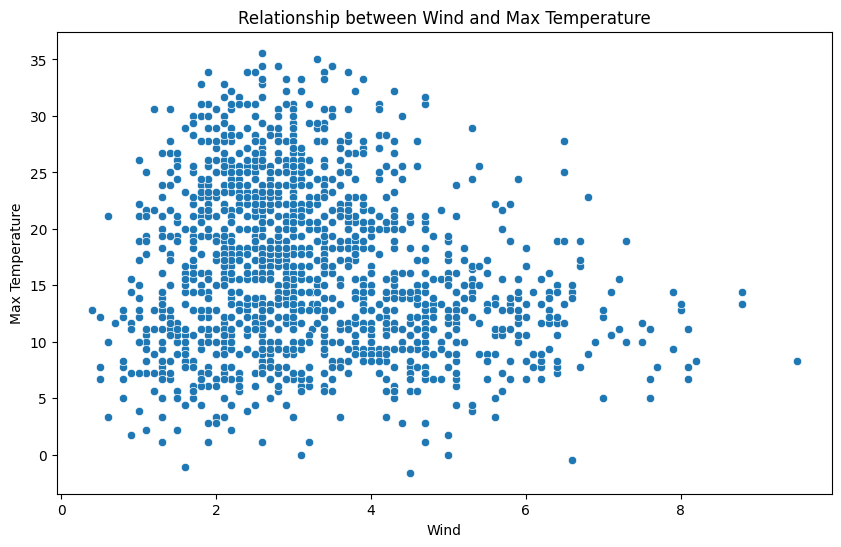
* To check if heavy rain occurs **randomly or follows a pattern** within each month.

**Trends Observed:**

* A uniform distribution would mean heavy rain is **random**.
* Peaks on certain days might indicate **a cyclical or meteorological pattern** (e.g., linked to weather fronts).

**Deduction:**

* **If heavy rain days are evenly spread, it suggests no specific pattern.**
* If certain days have more heavy rain, further meteorological analysis is needed.

***12) Wind vs. Max Temperature***

**Type of Analysis: Bivariate Analysis**

**Goal:**

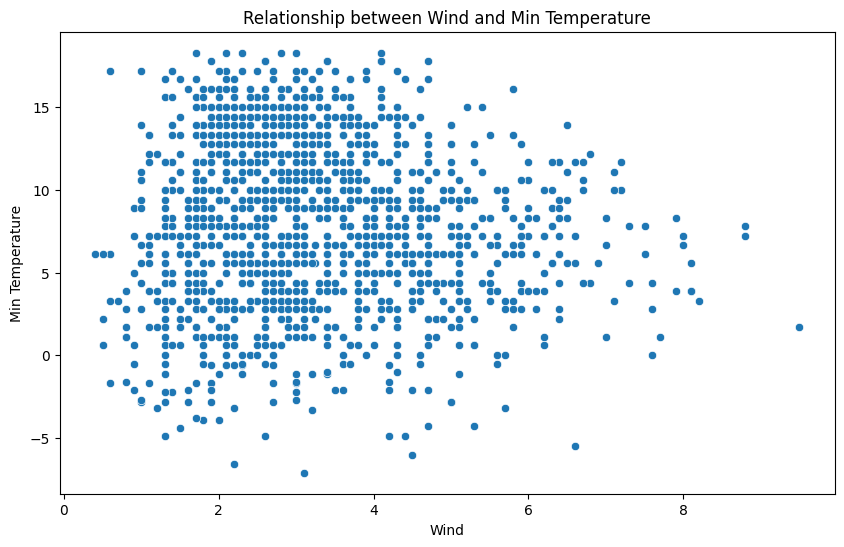
* To determine if wind speed influences **maximum temperature.**
* Identify **any correlation** (positive, negative, or none).

**Trends Observed:**

* If a **negative correlation** exists, higher wind speeds might be associated with lower max temperatures.
* If **no clear pattern** is visible, wind speed may not significantly affect max temperature.

**Deduction:**

* If temperatures drop as wind increases, wind could have a cooling effect.
* If no relationship exists, temperature fluctuations are likely due to other factors (e.g., sunlight, cloud cover).

***13) Wind vs. Min Temperature***

**Type of Analysis: Bivariate Analysis**

**Goal:**

* To analyze whether **wind speed affects minimum temperatures** (cooling or warming effect at night).

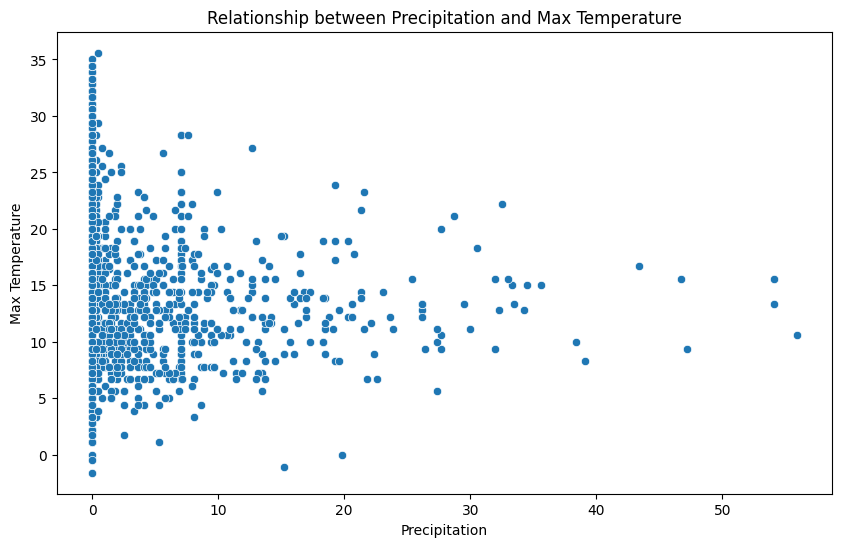
**Trends Observed:**

* If a **negative correlation** is found, stronger winds might contribute to **colder nights**.
* A **positive correlation** could suggest that wind traps heat near the surface, preventing extreme cooling.

**Deduction:**

* **If stronger winds are linked to colder nights, wind likely enhances cooling.**
* If there’s no pattern, nighttime temperatures might be influenced more by cloud cover or humidity.

***14) Precipitation vs. Max Temperature***

******

**Type of Analysis: Bivariate Analysis**

**Goal:**

* To investigate how **rainfall affects daily high temperatures.**
* Determine if rainy days tend to have **cooler max temperatures.**

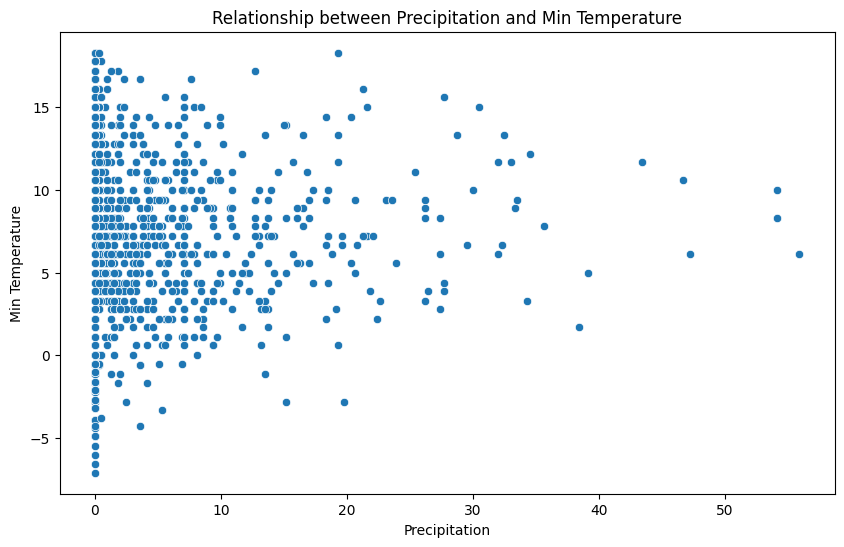
**Trends Observed:**

* If **heavier precipitation correlates with lower temperatures**, it suggests rain brings cooling.
* If no strong correlation is observed, temperature may be driven by other factors.

**Deduction:**

* Rainy days often reduce max temperature due to cloud cover and evaporative cooling.
* If no strong relationship exists, localized weather patterns may play a role.

***15) Precipitation vs. Min Temperature***

******

**Type of Analysis: Bivariate Analysis**

**Goal:**

* To assess if **rain impacts nighttime temperatures** (does rain trap heat at night?).

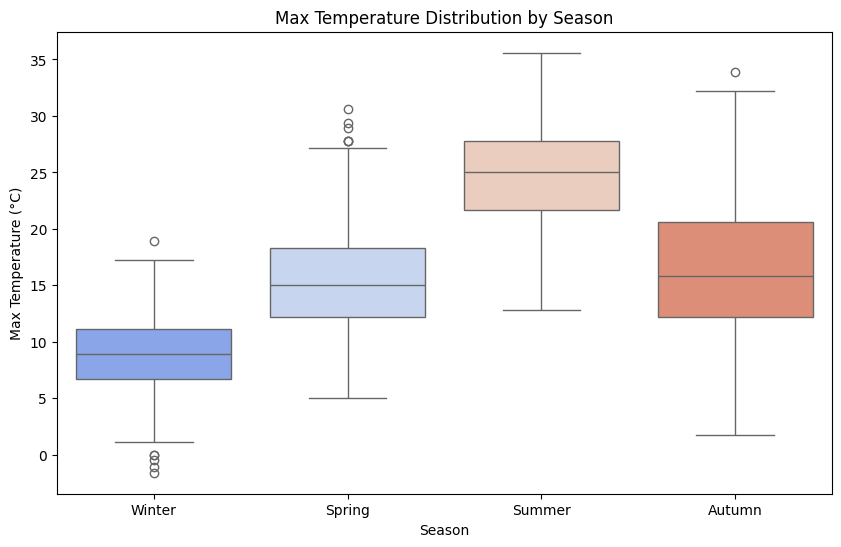
**Trends Observed:**

* If rain correlates with **warmer nights**, cloud cover may act as insulation.
* If a **negative correlation** is found, rain could contribute to **cooler temperatures** through evaporative cooling.

**Deduction:**

* Rainy nights may prevent extreme cooling by trapping heat.
* If there’s a drop in temperature on rainy nights, evaporative cooling might be a dominant factor.

***16) Max Temperature Distribution by Season***

****

**Type of Analysis: Bivariate Analysis** (Season vs. Max Temperature)

**Goal:**

* To analyze how **maximum temperature varies** across the four seasons.
* Identify **seasonal extremes** and **temperature ranges.**

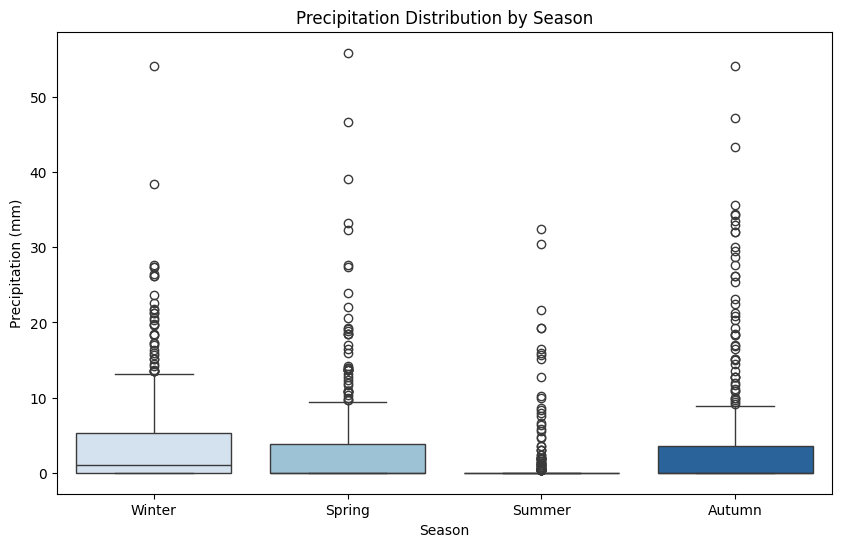
**Trends Observed:**

* **Summer** shows the highest median max temperatures.
* **Winter** has the lowest median temperatures, with a smaller spread.
* **Spring & Autumn** exhibit intermediate temperatures, with slightly overlapping distributions.
* **Outliers** (extremely high/low max temperatures) are present in all seasons.

**Deduction:**

* Summer has the highest max temperatures, as expected.
* Autumn & Spring transition smoothly between temperature extremes.
* The presence of extreme outliers suggests occasional heatwaves or cold spells.

***17) Precipitation Distribution by Season***

******

**Type of Analysis: Bivariate Analysis** (Season vs. Precipitation)

**Goal:**

* To determine **which seasons, receive the most precipitation** and observe variations.

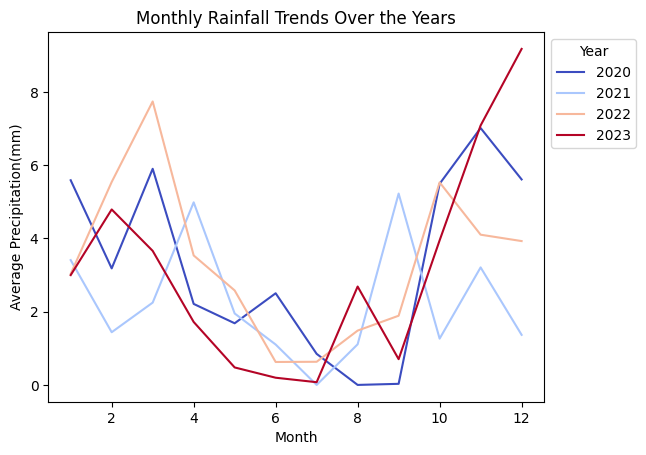
**Trends Observed:**

* **Winter and Autumn** have more rainfall.
* **Summer** generally has lower precipitation but may contain extreme outliers (storms).
* **Spring is comprised of average rainfall**, acting as a transition period.
* Presence of **outliers in all seasons** (e.g., extreme rain events).

**Deduction:**

* Winter & Autumn are the wettest seasons, maybe due to heavy rain events.
* **Summer may experience occasional heavy rainfall due to thunderstorms.**
* Outliers signal extreme weather events such as storms or flash floods.

***18) Monthly Rainfall Trends Over the Years***



**Type of Analysis: Multivariate Analysis** (Month vs. Year vs. Precipitation)

**Goal:**

* To analyze how **average monthly rainfall changes over several years.**
* Identify long-term trends in precipitation.

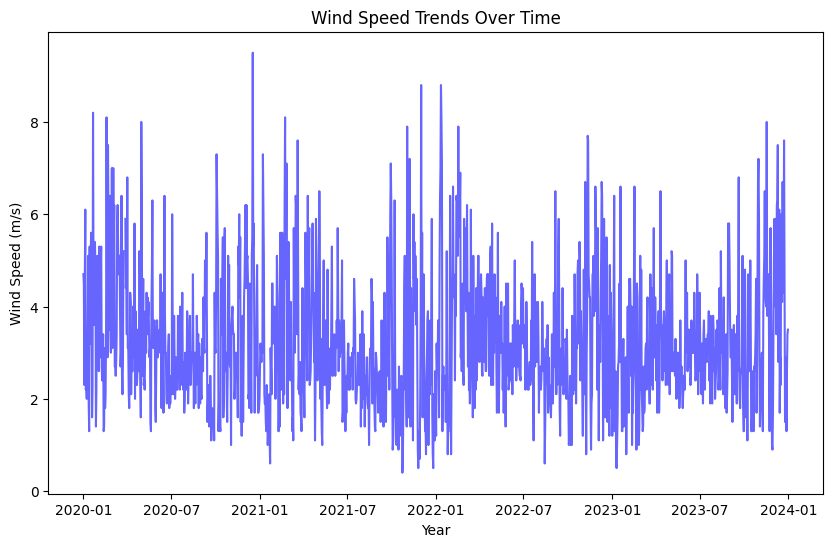
**Trends Observed:**

* Steady rainfall in some months, with variation from year to year in other months**.**
* Certain years might have seen **abnormal rainfall trends** (e.g., exceptionally wet, or dry periods).
* **Long-term climate trends** may emerge if some months show a persistent increase/decrease in rainfall.

**Deduction:**

* **If some months are showing increasing rainfall over the years, climate change or shifting weather patterns may be the reason.**
* Wet and dry periods in different years represent natural variability.

***19) Wind Speed Trends Over Time***

******

**Type of Analysis: Univariate Analysis** (Wind Speed Over Time)

**Goal:**

* To track the variation **wind speed** over the period of the dataset.

**Trends Observed:**

* **Fluctuations in wind speed** can be seen over time.
* Certain periods might show **higher average wind speeds**, potentially linked to storms.
* If a **long-term increase or decrease is observed**, it may indicate climatic changes.

**Deduction:**

* **If wind speeds increase over time, it may indicate that storms happen more frequently, or atmospheric patterns shift.**
* Abnormally low or high wind speeds over a period may be due to extreme weather conditions.