WEATHER DATA ANALYSIS

VEDANSHI PATHAK 22CD3033

Overview of the Project

Exploratory Data Analysis is an approach to analyze data

- To summarize the main characteristics of data, and better understand the data set.

- It also allows us to quickly interpret the data and adjust different variables to see their effect.

- The three main steps to get a perfect EDA are extracting the data from an authorized source,

- cleaning and processing the data, and performing data visualization on the cleaned data set.

Overview of the Project

- Many sites provide historical data on many meteorological parameters such as pressure, temperature, humidity, wind speed, visibility, etc.

We are going to analyze the Weather data set.
- Our main aim is to perform data cleaning, data normalizing, testing the hypothesis, and deriving appropriate insights.

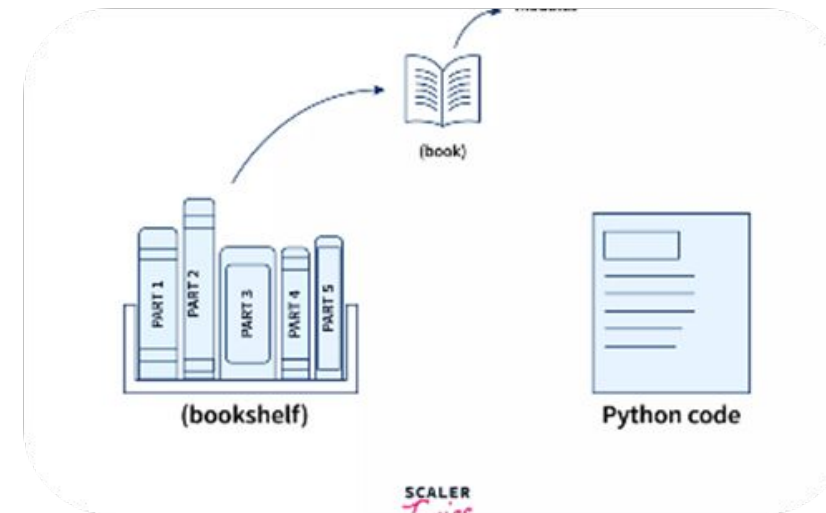- Our dataset is extracted from Kaggle.com
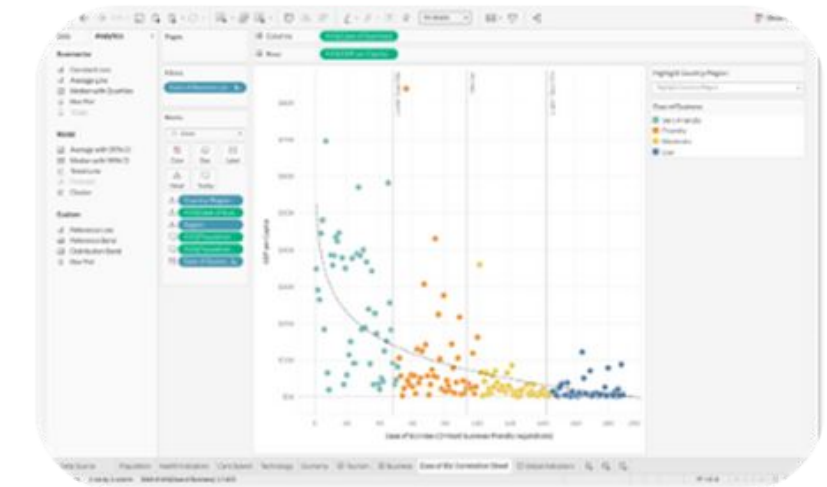
Tools used for this Project:

MS Excel

These spreadsheet tools are user-friendly and commonly used for basic data analysis and visualization tasks.

Python Libraries

Pandas, NumPy, and Matplotlib are libraries used for data preprocessing and analysis. They provide more advanced functionality

Tableau

A visualization tool that simplifies complex data into interactive dashboards and visualizations.

The Importance of Weather Data Analysis

1 **PREDICTIVE MODELING** 🌡️
Climate models help predict future weather patterns, enabling us to make better decisions for agriculture, disaster management, and urban planning.

2 **IMPROVED SAFETY**
By understanding weather patterns, individuals and organizations can make informed decisions, minimizing the impact of severe weather conditions.

3 **ECONOMIC IMPACT** 💰
Weather events have a major impact on various industries, such as aviation, energy, and tourism. Analyzing weather data helps these industries predict and mitigate the economic impact.

Data Collection

Sources of Weather Data

Weather data can be obtained from a variety of sources, such as national weather agencies, private companies, and citizen science initiatives

Methods of Data Collection

Data can be collected using weather stations, satellites, drones, or crowd-sourced data collection methods, such as personal weather stations and smartphone apps.

Data Preprocessing

Before analysis, data must be cleaned, formatted, and prepared. This involves handling missing data, removing outliers, and normalizing data.

Data Preparation

Data Cleaning

Raw data needs to be processed to remove or correct erroneous data that may affect analysis results, such as repeated or incomplete measures.

Data Normalization

To minimize variation and make data easier to compare, normalization is used to adjust differences in scale, size, or other characteristics.

Exploratory data analysis

Data visualization tools, such as graphs and charts, allow users to easily understand trends and patterns in data and make predictions.

Visualization Techniques

Statistical Analysis

Statistical methods enable us to identify patterns and trends in data, assess the accuracy of predictions, and make data-driven conclusions.

Predictive Modeling

Machine learning algorithms and predictive models enable us to forecast future weather conditions and their impact.
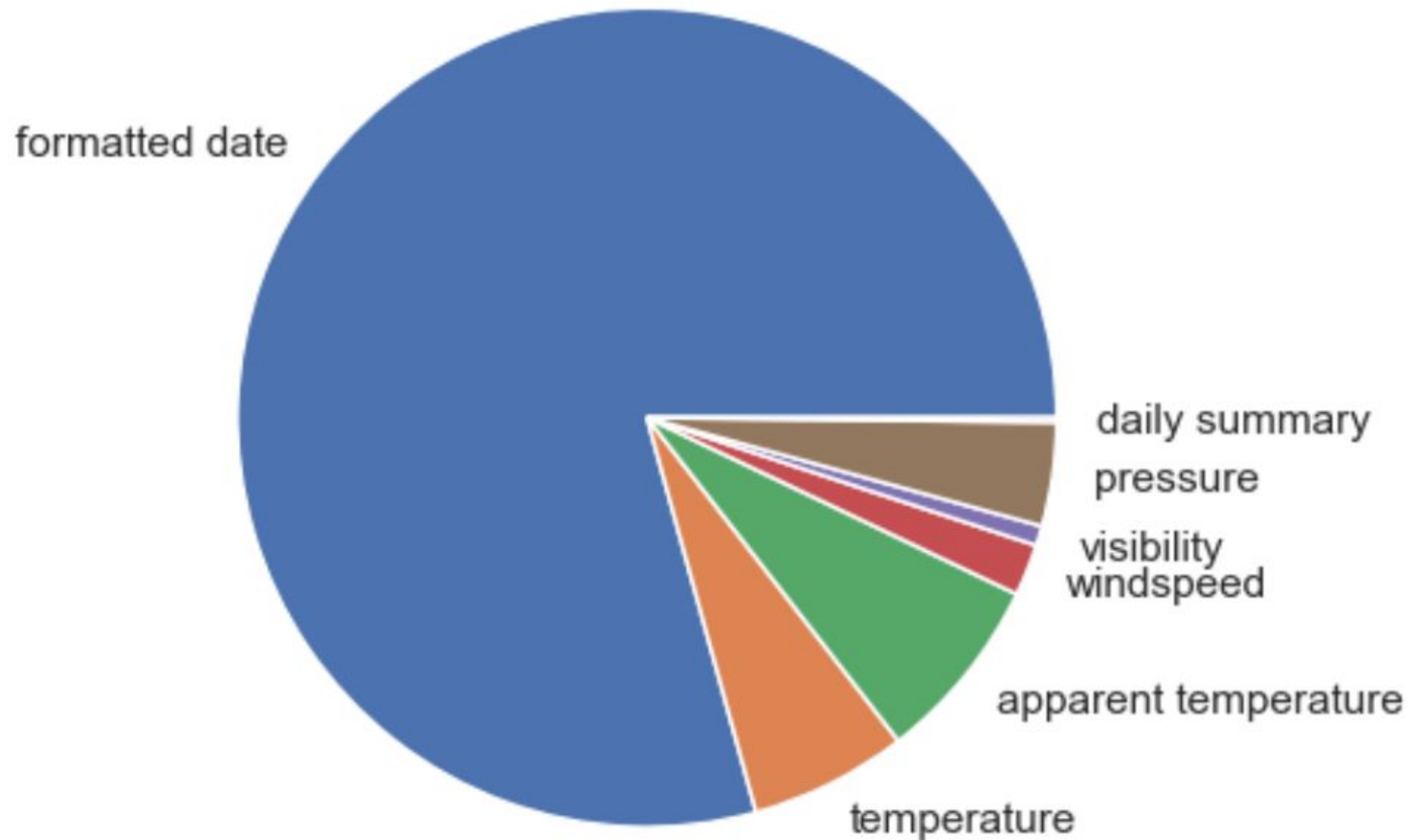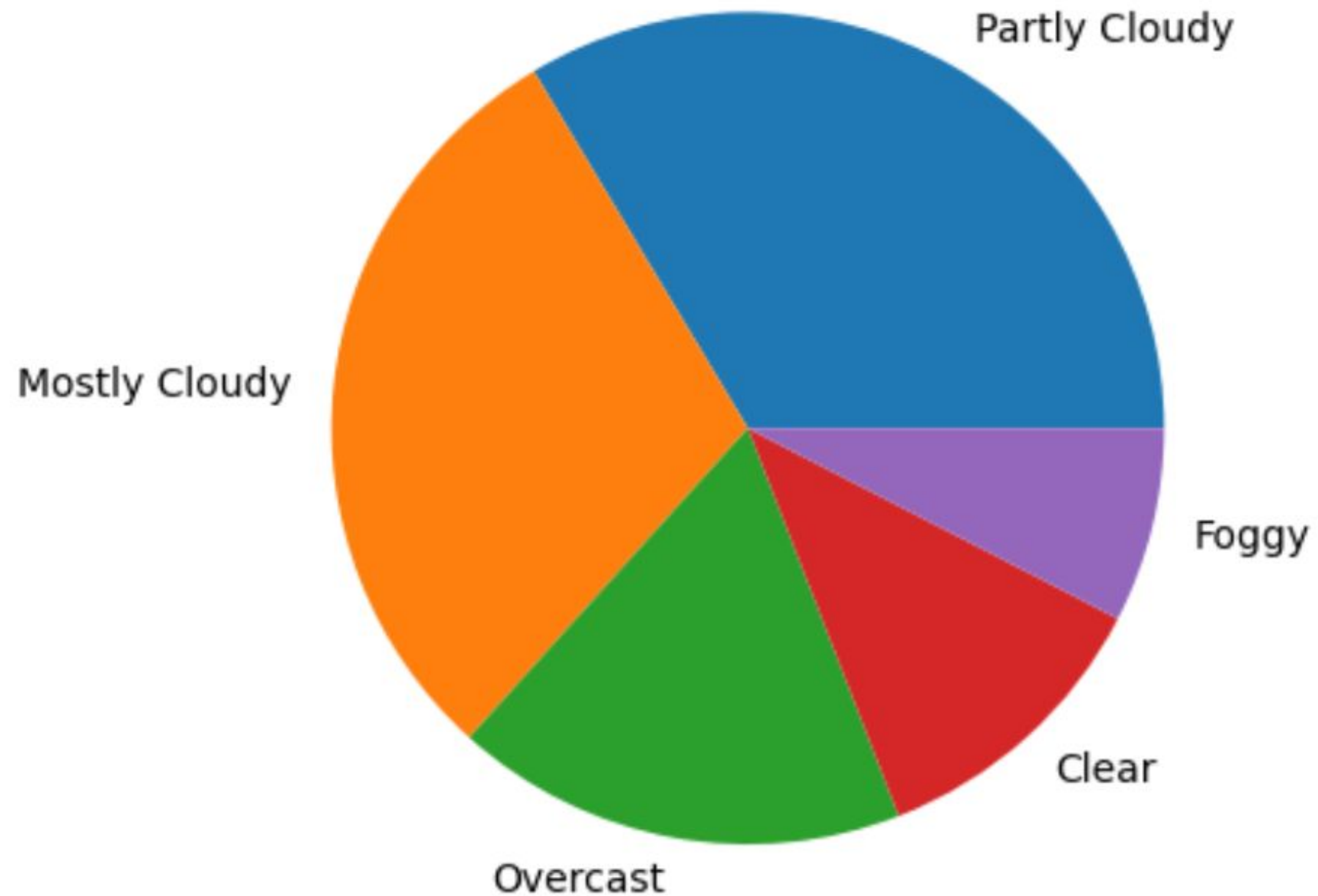
VISUALIZATION

Insights

The previous PIE CHART tells you about the summary of the given dataset. It provides you with the count for the given data . Here in this case, formatted data is more in number and the visibility count is the least.

Representation of summary using pie chart

Insights

This PIE CHART shows you which weather condition occurred the most in the given dataset.

Word Cloud

Insights

The word cloud visually represents the frequency of Summary, with more frequently occurring Summary appearing larger in the cloud.

Heat Map

**Representation of (Temp, Humidity, Visibility, Wind Speed, Apparent**

Correlation Heatmap of the First 8 Columns in df_w1
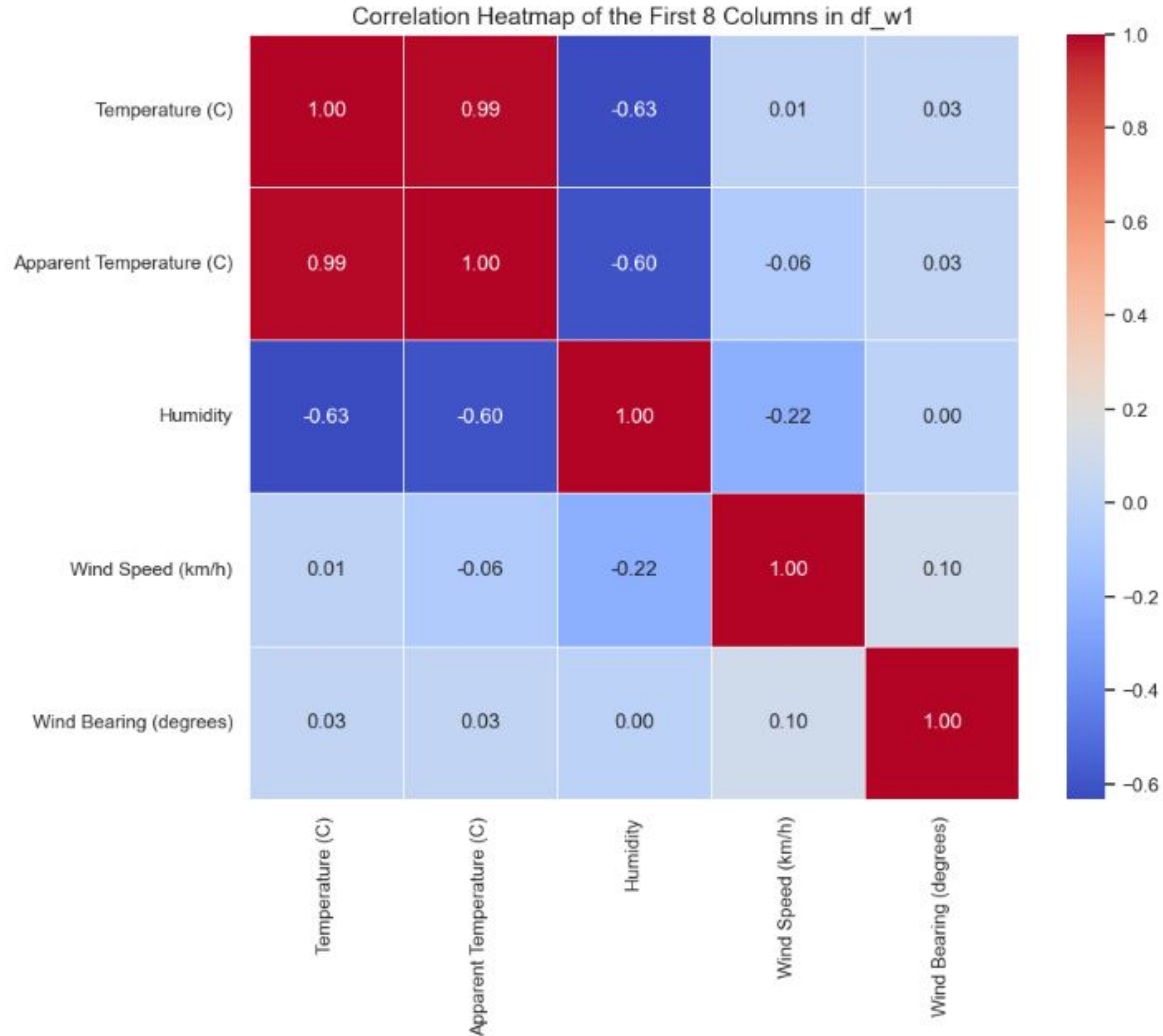
Insights

The given HEATMAP explains you about the the correlation between the different parameters. As we can see here , the red color shows the maximum correlation between the parameters while the dark blue part shows the least correlation.
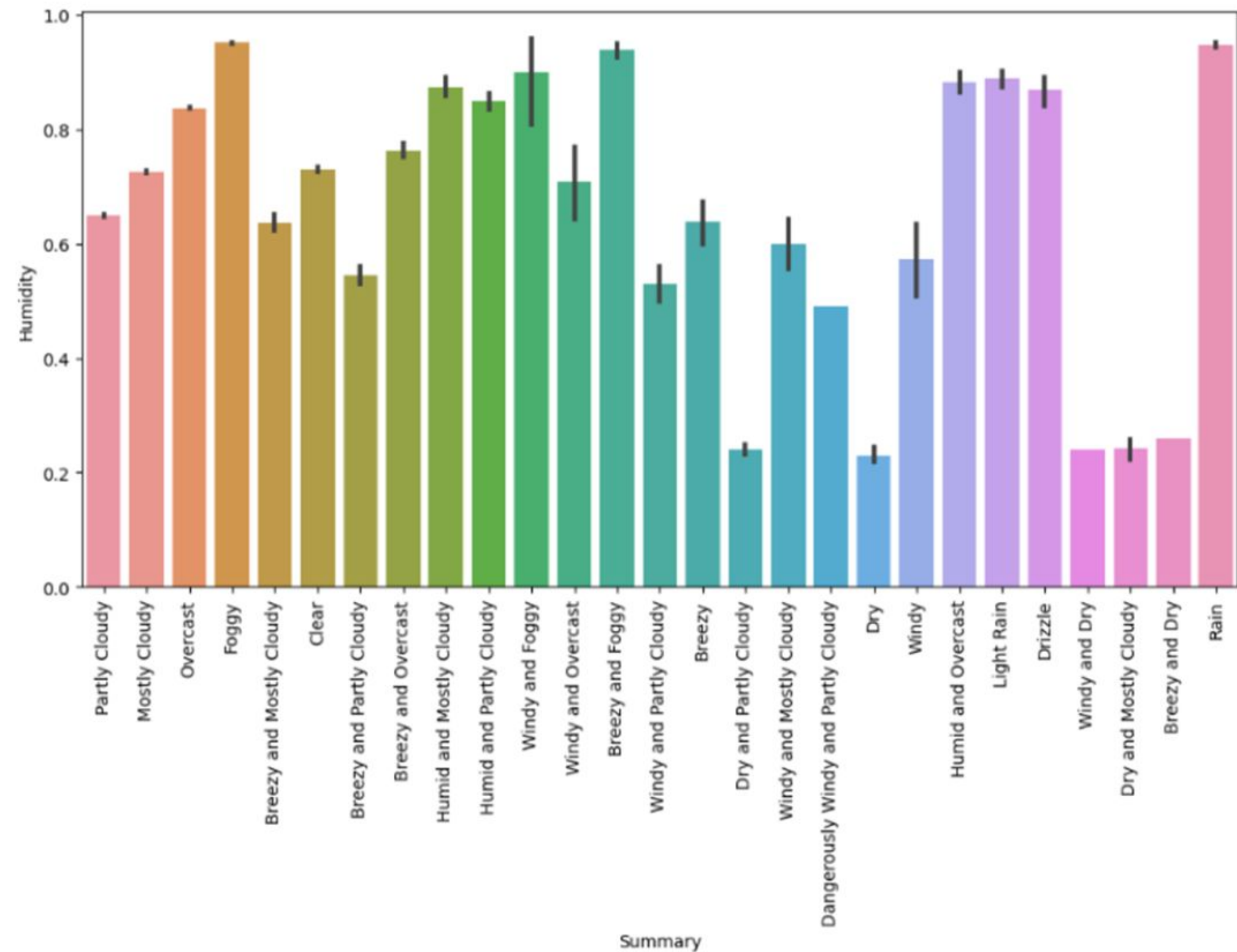
Bar Chart

Representation of Temp VS summary graph
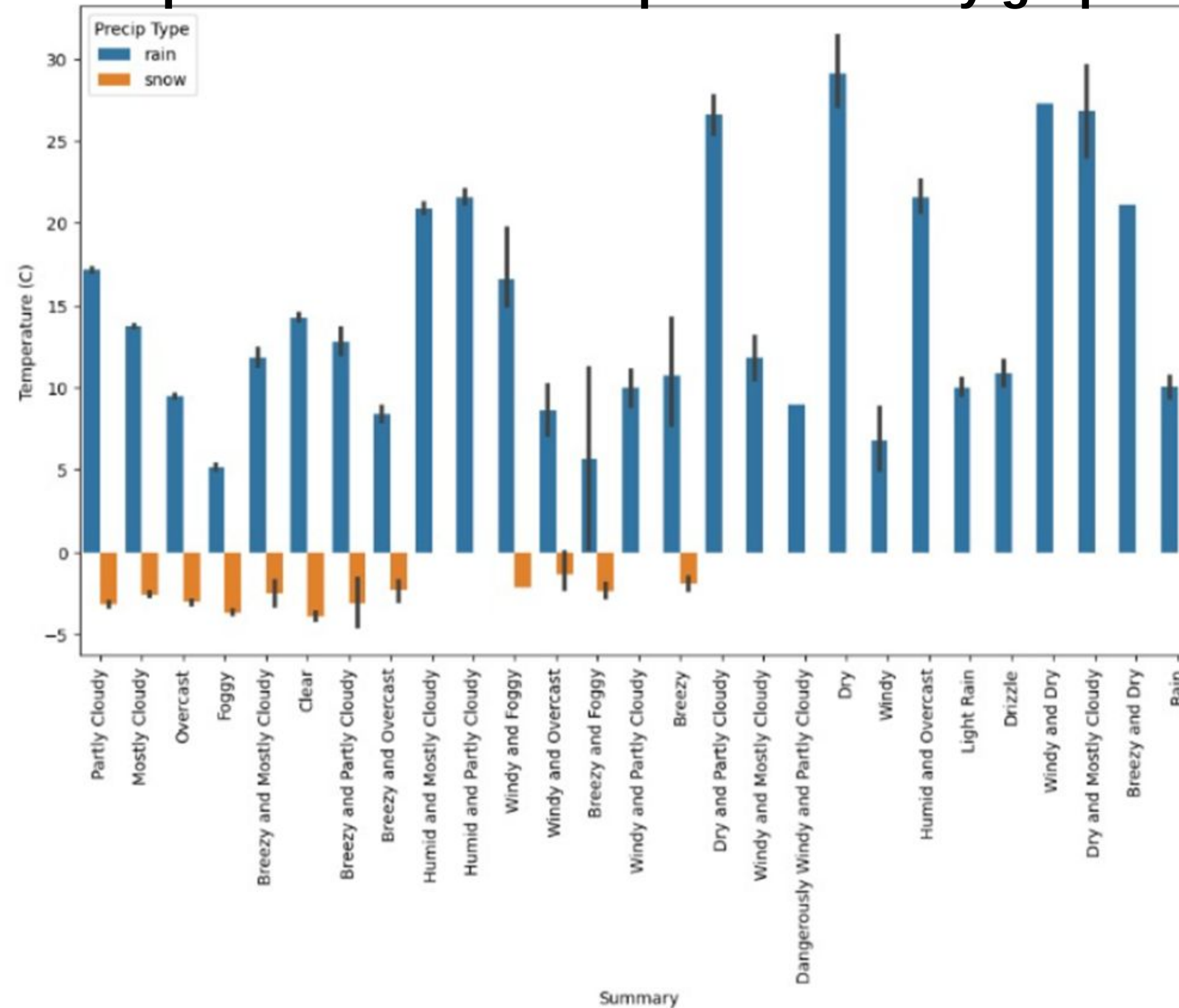
Insights

The BAR CHART represents the relation between the summary and average temperature along with the precipitation type i.e. snow or rain.

Representation of Humidity VS Summary using Bar plot
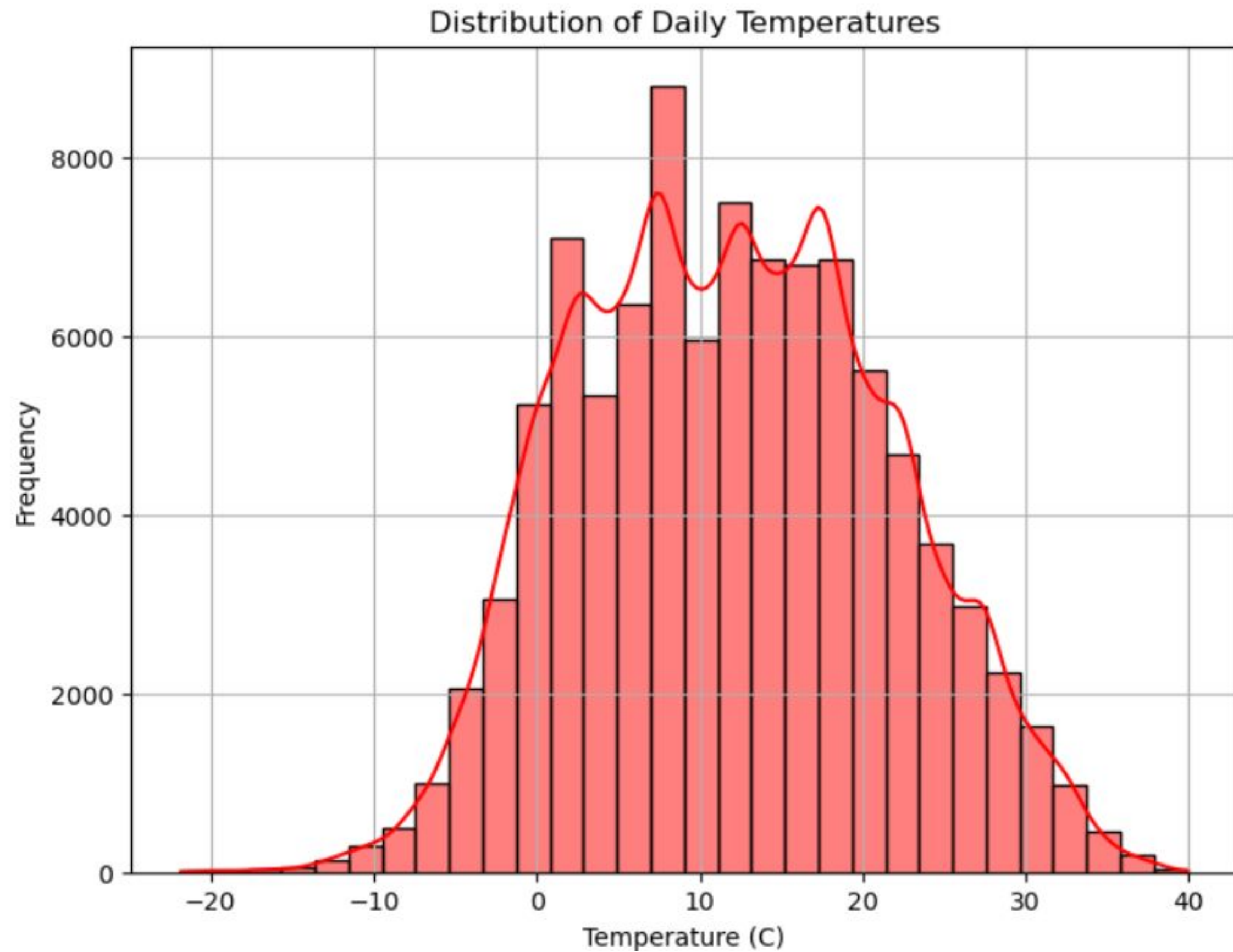
Insights

The visualization used here is the BAR PLOT which shows you how the weather condition is related to the humidity in the given data set. Here during RAIN, we observed the maximum humidity.

Histogram

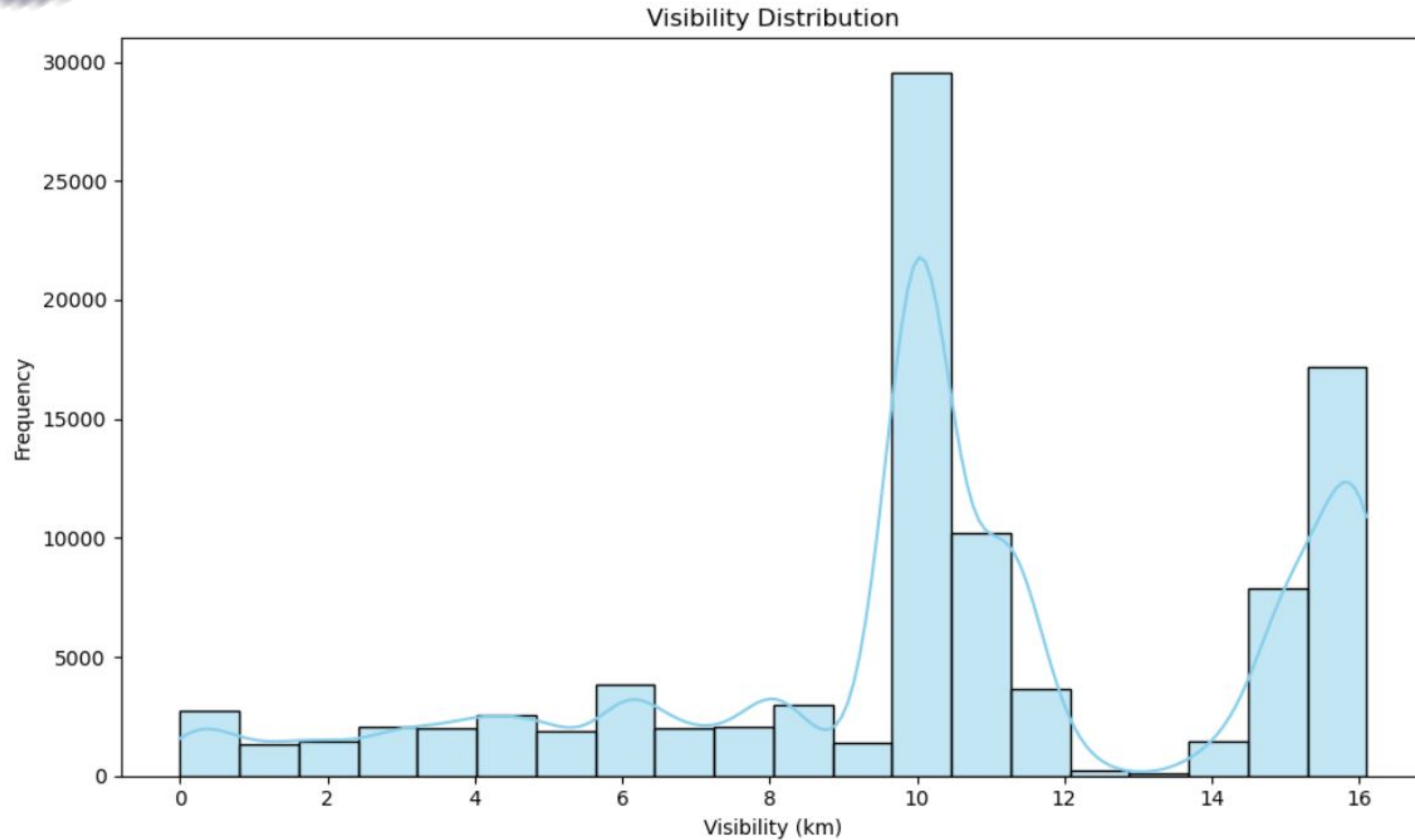Distribution of Daily Temperatures

Insights

The given visualization the distribution of daily temperatures using histograms or kernel density plots with the provided dataset.

Kernel density checks for the estimated probability density

Histogram
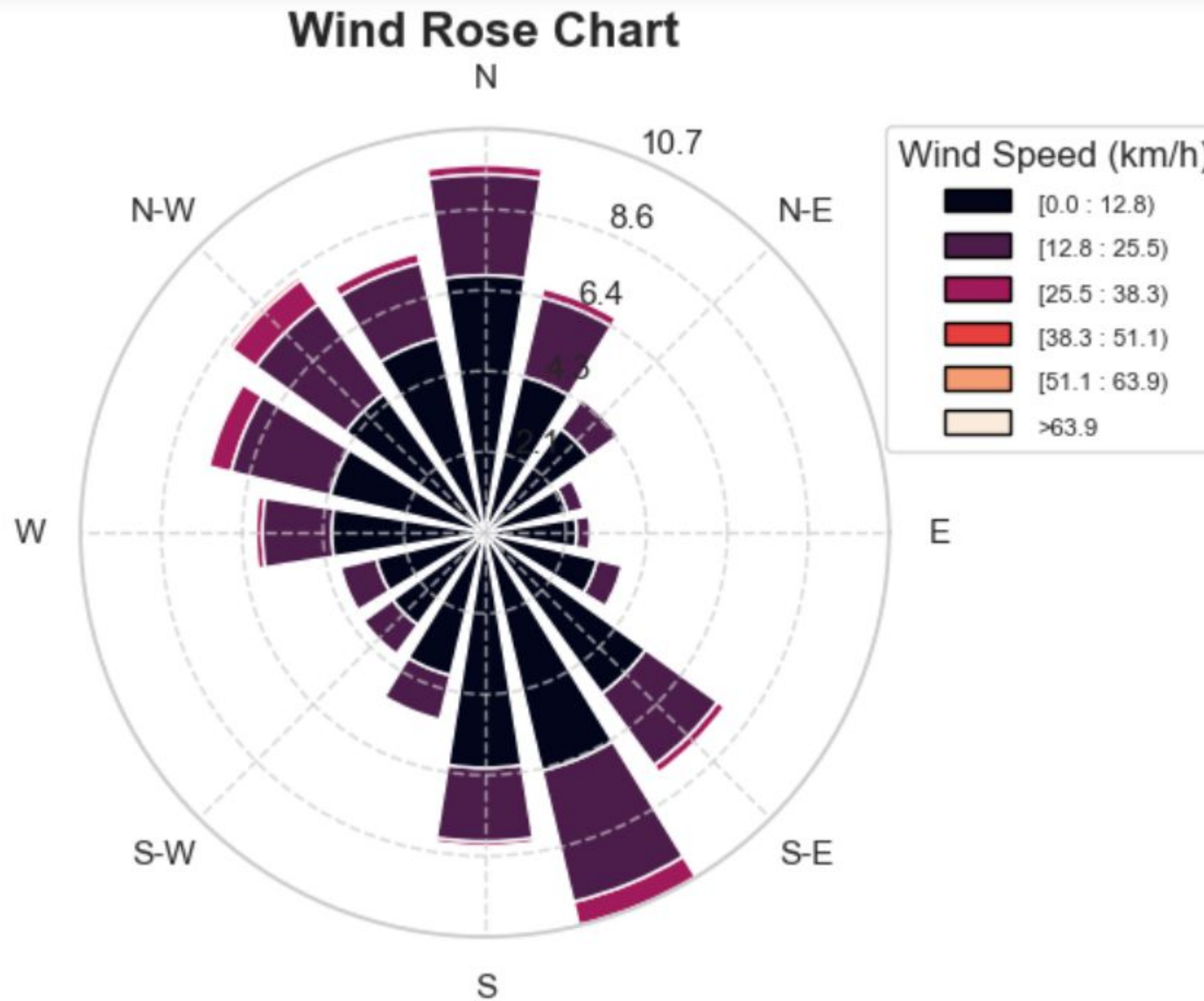

Visibility Distribution

Insights

Visibility is an essential weather parameter. Analyzing its distribution helps meteorologists understand typical visibility patterns, detect anomalies, and predict potential weather-related events. Here we are doing the same using HISTOGRAM.

Wind Rose Chart

Insights

A wind rose depicts the distribution of wind direction and speed at a location over a period of time. The length of each bar represents the percent of time the wind blows from that direction. Each bar is subdivided with colors to show wind speed ranges associated with each direction.

TABLEAU
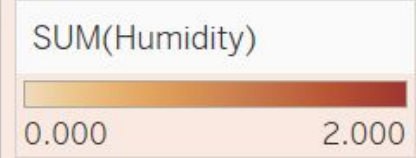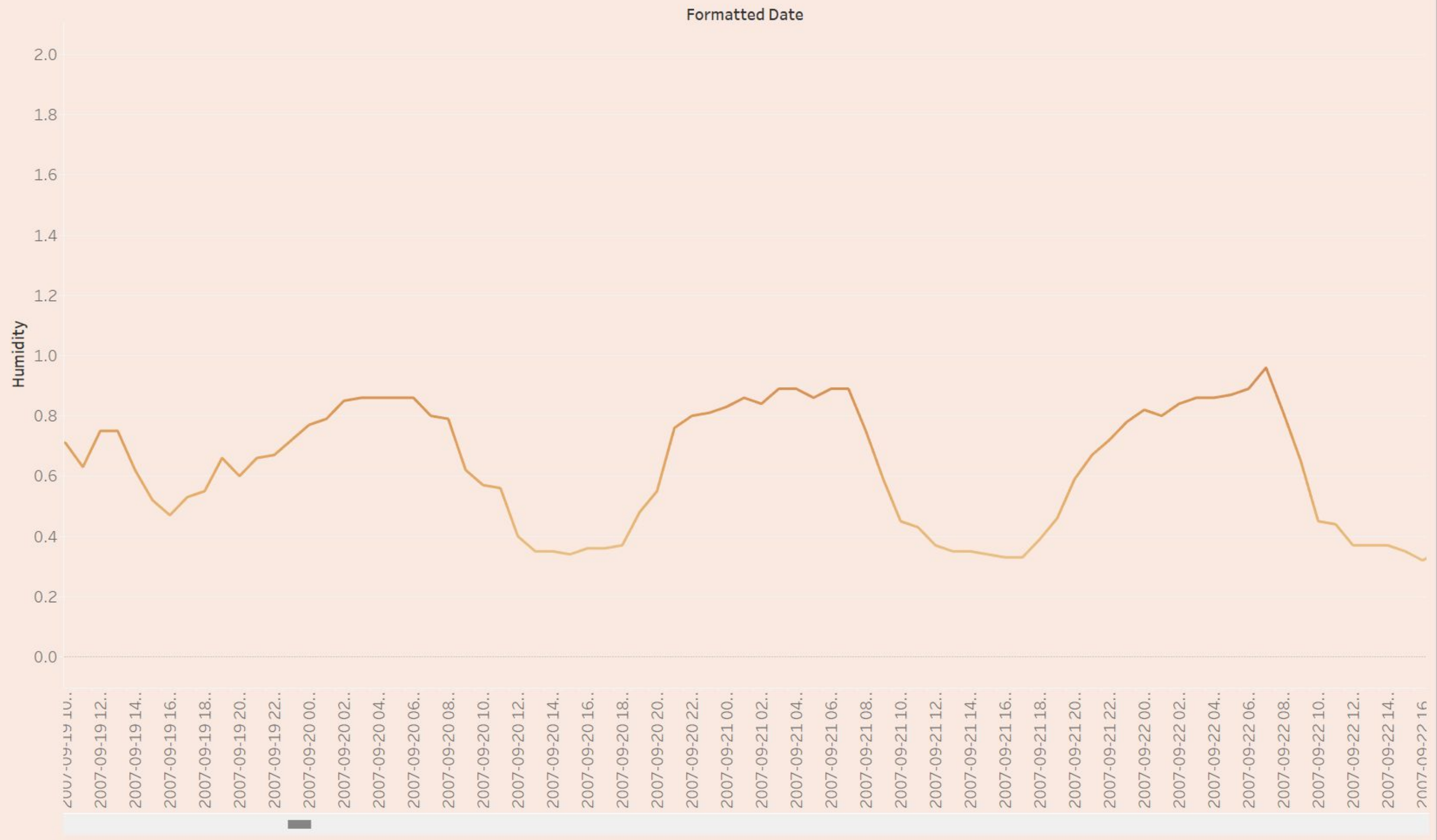
Insights

The LINE CHART here shows the variation in humidity over the formatted date. Here the data is taken for the year 2006 and the variation is showed for every hour.

Weather condition with max AVG temp

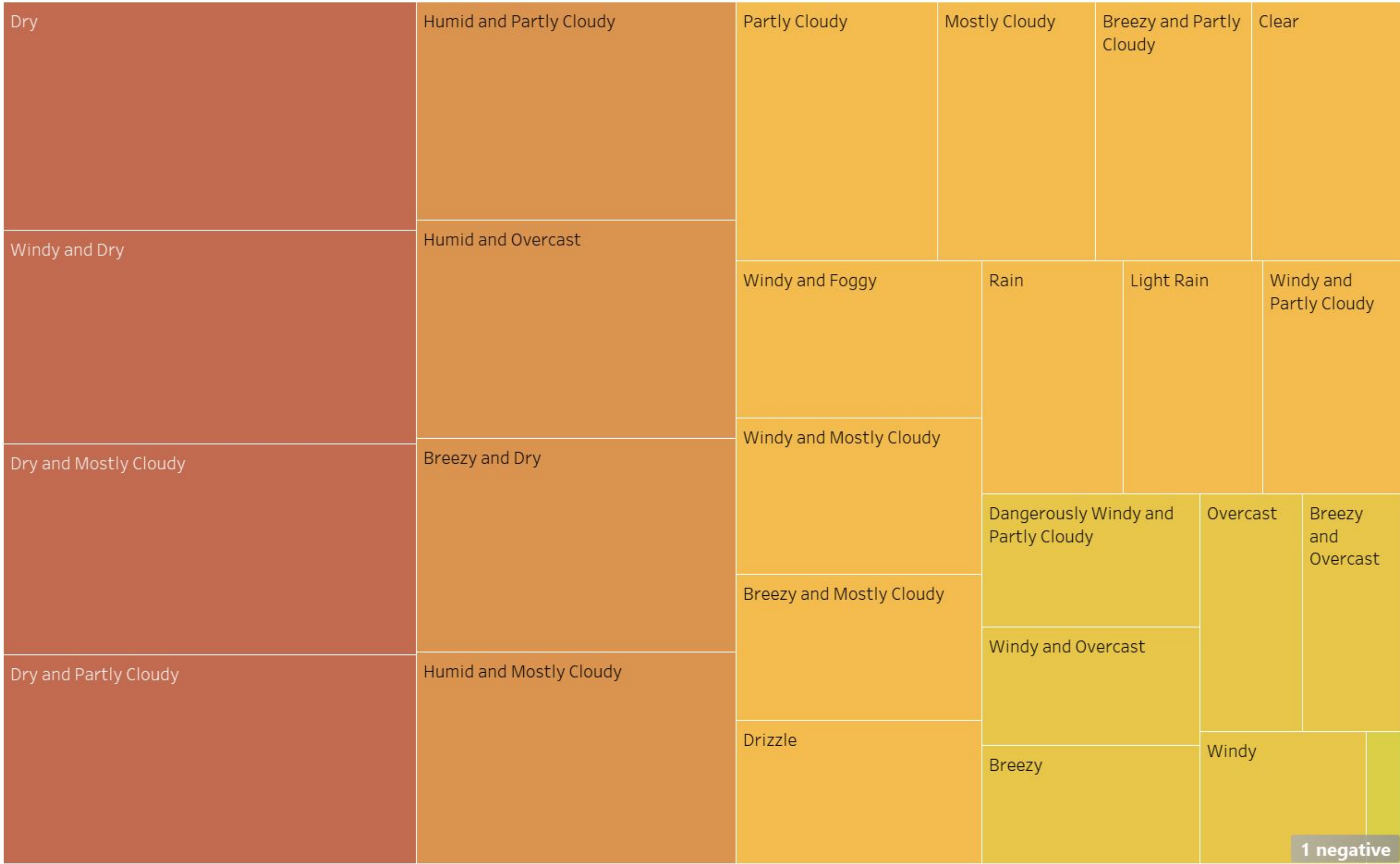| Dry	Humid and Partly Cloudy	Partly Cloudy	Mostly Cloudy	Breezy and Partly Cloudy	Clear

Windy and Dry

Humid and Overcast

Windy and Foggy	Rain	Light Rain	Windy and Partly Cloudy

Windy and Mostly Cloudy

Dry and Mostly Cloudy

Breezy and Dry

Dangerously Windy and Partly Cloudy	Overcast	Breezy and Overcast

Breezy and Mostly Cloudy

Windy and Overcast

Dry and Partly Cloudy

Humid and Mostly Cloudy

Drizzle

Breezy

Windy

1 negative

Summary
- ☑ (All)
- ☑ Breezy
- ☑ Breezy and Dry
- ☑ Breezy and Foggy
- ☑ Breezy and Mostly C...
- ☑ Breezy and Overcast
- ☑ Breezy and Partly Cl...
- ☑ Clear
- ☑ Dangerously Windy ...
- ☑ Drizzle
- ☑ Dry
- ☑ Dry and Mostly Clou...
- ☑ Dry and Partly Cloudy
- ☑ Foggy
- ☑ Humid and Mostly C...
- ☑ Humid and Overcast
- ☑ Humid and Partly Cl...
- ☑ Light Rain
- ☑ Mostly Cloudy
- ☑ Overcast
- ☑ Partly Cloudy
- ☑ Rain
- ☑ Windy
- ☑ Windy and Dry
- ☑ Windy and Foggy
- ☑ Windy and Mostly Cl...
- ☑ Windy and Overcast
- ☑ Windy and Partly Cl...

AVG(Temperature (C))

-3.23   29.08

Care Spend | Economy | ⊞ Tourism | ⊞ Business | **max avg temp** | max visibility | avg_wind | ⊞ Dashboard 3 | Sheet 12 | Sheet 12 (2) | Sheet 14 | Sheet 15
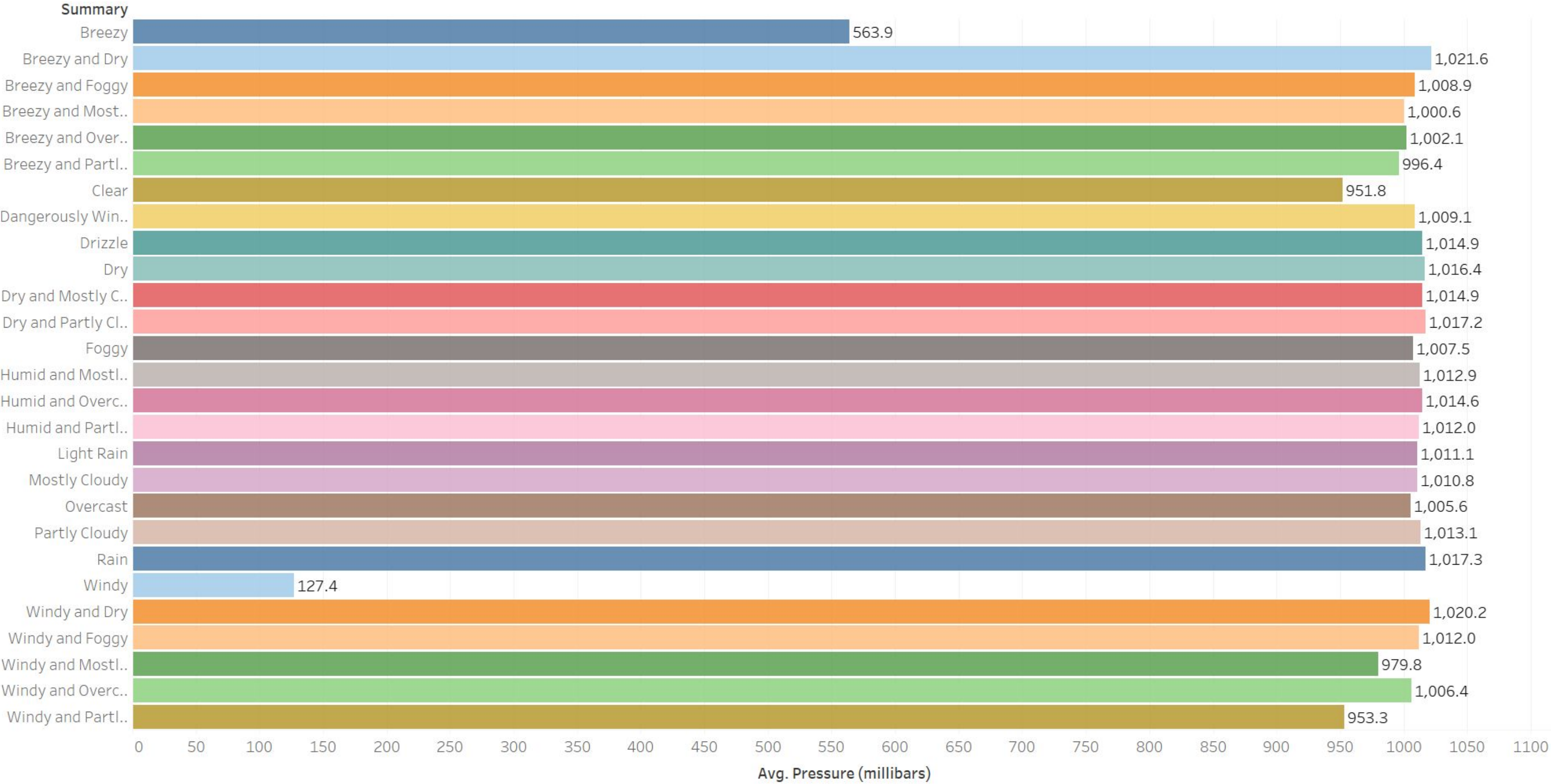
Insights

The visualization here used is TREEMAP which is used to display hierarchical (tree-structured) data. Here in this case the DRY weather condition has the highest AVERAGE TEMPERATURE and is therefore placed in a bigger rectangle.

SUMMARY VS MEAN PRESSURE

Summary

Summary	Avg. Pressure (millibars)
Breezy	563.9
Breezy and Dry	1,021.6
Breezy and Foggy	1,008.9
Breezy and Most..	1,000.6
Breezy and Over..	1,002.1
Breezy and Partl..	996.4
Clear	951.8
Dangerously Win..	1,009.1
Drizzle	1,014.9
Dry	1,016.4
Dry and Mostly C..	1,014.9
Dry and Partly Cl..	1,017.2
Foggy	1,007.5
Humid and Mostl..	1,012.9
Humid and Overc..	1,014.6
Humid and Partl..	1,012.0
Light Rain	1,011.1
Mostly Cloudy	1,010.8
Overcast	1,005.6
Partly Cloudy	1,013.1
Rain	1,017.3
Windy	127.4
Windy and Dry	1,020.2
Windy and Foggy	1,012.0
Windy and Mostl..	979.8
Windy and Overc..	1,006.4
Windy and Partl..	953.3

Avg. Pressure (millibars)

Care Spend | Economy | ⊞ Tourism | ⊞ Business | max avg temp | max visibility | avg_wind | ⊞ Dashboard 3 | Sheet 12 | Sheet 12 (2) | **Sheet 14**
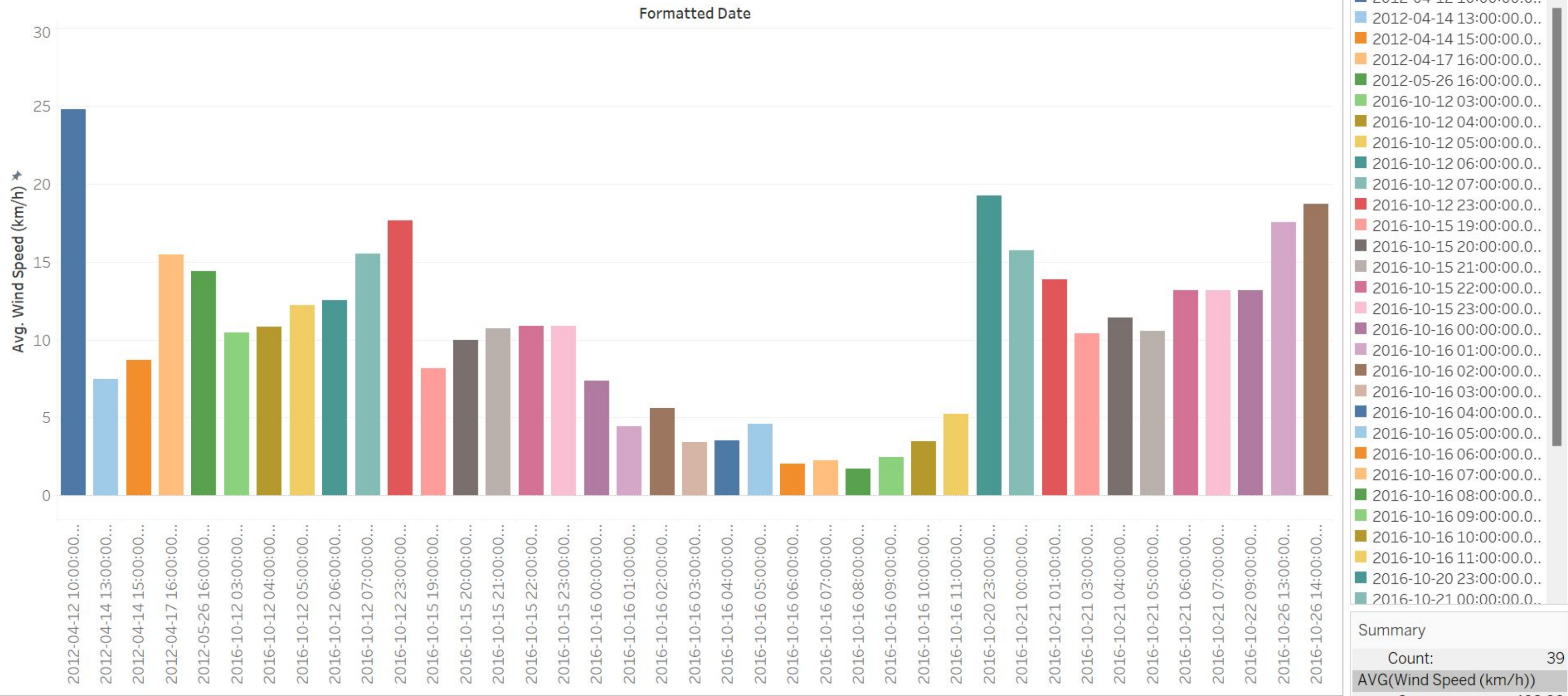
Insights

The chart compares the average atmospheric pressure for different weather summaries using a bar chart. This can reveal if certain weather conditions are associated with specific pressure patterns.

Insights

The bar plot here represents the AVERAGE of WIND SPEED (km/hr) for each FORMATTED DATE. The colors show the details about the formatted date. The data is filtered on SUMMARY i.e. DRIZZLE.

Applications

Agriculture and Farming 🌾

Weather data analysis can help farmers optimize crop yields, predict weather patterns, and minimize risks associated with crop failure.
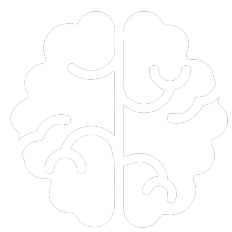
Aviation

Weather plays a critical role in aviation safety, and weather data analysis helps pilots and air traffic control teams make informed decisions to ensure safe and efficient travel.

Disaster Response and Management

Weather data analysis helps emergency responders prepare for and mitigate the impact of natural disasters, such as hurricanes, floods, and forest fires.

Conclusion

- Global warming is increasing the frequency and intensity of some types of extreme weather.
- From the analysis, we have gathered useful insights over the past 10 years.
- For example, warming is causing more rain to fall in heavy downpours.
- There are also longer dry periods between rainfalls. This, coupled with more evaporation due to higher temperatures.
- Wet places have generally become wetter, while dry places have become drier.