**Analyzing datasets and identifying any missing features or additional ones**

**Existing weather features:**

Light rain, heavy rain, snow, wind, precipitations, fog

**Additional Features**

Fog => Visibility

Humidity => Precipitation

Wind => wind direction

**For DATE Column**

* Extract days, month
* Extract by day-of-week
* Deduce SEASON

**Calculations/Predictions**

* ***Count occurrences of different weather conditions => see which type is more common***
* ***Compare weather across different months and season => calculate mean***
* Average daily wind
* To verify the datatypes used for easy manipulations
* Final average month temperature and annual temperature
* ***Graph plot to analyze the data for any gaps or superfluous data***
* *Check for which year’s data may be missing*
* Calculate how much it rains per year
* Maximum temperature each month

**Ratio that can be calculated:**

How different is the day’s temperature to the average temperature

The interval between max temp and min temp

**ADDITIONAL SNIPETS THAT CAN BE USED THROUGHOUT THE REPORT:**

... how to use a set of ML-based models and large amounts of data to make informed weather or climate-related predictions within hours

1. they can be an end-to-end data analysis project spanning different temporal in weather, while still providing a foundation for forecasting other weather conditions.
2. machine learning-based tools ready to be used operationally and tuned for various applications the user might have in mind
3. Can be used as a tool to tackle deep fundamental scientific questions weather prediction including, for example, tuning such tools for a seamless prediction framework

*Reference:* https://www.climatechange.ai/blog/2024-02-07-forecast-tutorials

Comments on why using humidity:

Humidity can help explain certain weather phenomena. For example, high humidity levels often precede or accompany precipitation. Even though the dataset has a precipitation column, humidity could provide more insight into the atmospheric conditions leading to rain or fog.

Humidity, temperature, and wind are all interrelated. For instance, the combination of high humidity and high temperature can make the weather feel hotter due to the heat index, and variations in humidity levels can influence how temperature changes are experienced.

Analyze the dataset's dimensionality to determine its suitability for the

study.

With only six columns, the dataset is considered low-dimensional.

Advantages:

Ease of Modeling: Fewer features mean simpler models with lower risk of overfitting, especially if the sample size is large relative to the number of features.

Interpretability: It is easier to interpret relationships between a small set of variables.

Potential Limitations:

Feature Completeness: The current features capture basic weather information. However, if the study requires a deep analysis (e.g., predicting subtle weather patterns or capturing micro-climatic effects), additional variables like humidity, atmospheric pressure, or wind direction might be necessary.

Derived Features: The date column can be leveraged to create extra temporal features. While this increases dimensionality, it can be beneficial if seasonal or trend patterns are important for the study.

Suitability:

If the study focuses on general weather trends, seasonal analysis, or simple forecasting, these six features may be entirely adequate—especially when enhanced with derived features from the date.

The dataset’s dimensionality is relatively low, which makes it straightforward for exploration analysis and initial modeling. Its suitability ultimately depends on the study’s goals. For basic weather analysis or trend detection, these features (especially when enhanced with engineered temporal attributes) may be sufficient. However, if the study demands a more detailed or complex understanding of weather phenomena, you might need to augment the dataset with additional relevant variables.