# **Weather dataset**

The dataset spans from January 1, 1981, to December 2023, providing a time series representation of synthetic weather conditions. The dataset aims to capture key aspects of daily weather patterns, allowing for exploration and analysis of temporal trends, seasonal variations, and potential correlations among the features. Missing values are intentionally introduced to simulate the occasional gaps or errors in real-world observational data. The dataset is designed for educational purposes and can serve as a starting point for practicing time series analysis and data manipulation techniques.

The synthetic time series weather dataset includes the following features, measured in specific metrics:

* Date: The timestamp indicating the date of each record.
* Temperature: The ambient air temperature, measured in degrees Celsius. The temperature values are generated with a mean of 20°C and a standard deviation of 5°C, simulating daily temperature variations.
* Humidity: The relative humidity, measured as a percentage. Humidity values are generated with a mean of 70% and a standard deviation of 10%, capturing typical humidity fluctuations.
* Wind Speed: The speed of the wind, measured in kilometers per hour (km/h). Wind speed values are generated with a mean of 10 km/h and a standard deviation of 5 km/h, simulating variations in wind intensity.
* Atmospheric Pressure: The atmospheric pressure, measured in hectopascals (hPa). Atmospheric pressure values are generated with a mean of 1010 hPa and a standard deviation of 5 hPa, representing variations in atmospheric pressure.

**Possible research questions**

These research questions encompass a range of descriptive and predictive analyses, offering opportunities to explore patterns, develop models, and gain insights into the dynamics of the time series weather dataset.

**Descriptive Analysis**

1. What are the long-term temporal trends in temperature, humidity, wind speed, and atmospheric pressure over the years?
2. Can we identify and describe seasonal patterns in temperature, humidity, and wind speed?
3. How are missing values distributed across different features, and is there a temporal pattern in their occurrence?
4. What is the correlation between temperature, humidity, wind speed, and atmospheric pressure?
5. Can we identify and analyze extreme weather events, such as temperature spikes or unusually high wind speeds?

**Predictive Analysis**

1. Can we build a predictive model to forecast temperature for the next week based on historical data?
2. How accurately can we predict wind speed using historical information, and are there leading indicators?
3. Can we develop a model to predict seasonal variations in humidity over the next few months?
4. How does the presence of missing data impact the accuracy of predictive models, and can it be mitigated?
5. How do different predictive models perform in forecasting temperature, and what factors influence their accuracy?
6. Can we develop a model to detect anomalous patterns in atmospheric pressure and investigate their causes?
7. To what extent do weather conditions on previous days influence the current day's temperature and humidity?
8. How does the seasonality of the data impact the predictability of wind speed?
9. Are there spatial variations in weather patterns, and can we predict these variations across different locations?
10. How well can predictive models account for uncertainties, and what factors contribute to prediction uncertainty?