UNDERSTANDING WEATHER DATA

RAISING SEVERAL BASIC STATISTICAL ANALYSIS AND VISUALIZATION DIRECTIONS

ABIODUN OWADOYE

AUGUST 2024

WACREN 2024

SOURCES OF WEATHER DATA

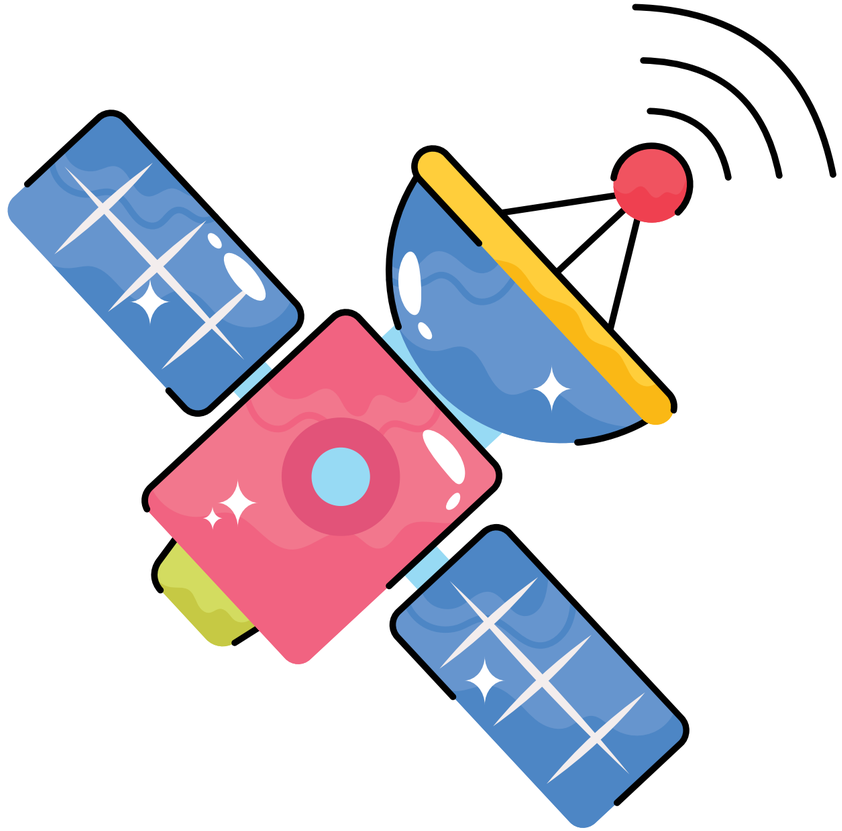
WEATHER STATIONS

- Ground-based stations measuring temperature, humidity, pressure, etc.
- These stations collect data on temperature (how hot or cold it is), humidity (the amount of moisture in the air), atmospheric pressure (the weight of the air), and sometimes other factors like wind speed, direction, and precipitation (rainfall, snowfall
- They provide localized and precise data which is essential for understanding local weather conditions, which can vary significantly from one location to another.

SOURCES OF WEATHER DATA

SATELLITES

- Satellites are spacecraft that orbit the Earth, equipped with advanced sensors and instruments designed to observe and collect data on various aspects of the Earth's atmosphere and surface.
- Unlike ground-based stations, satellites do not physically interact with the atmosphere. Instead, they use sensors to detect and measure energy (such as light or infrared radiation) emitted or reflected by the Earth's surface and atmosphere.
- Weather Patterns: Satellites can monitor large-scale weather systems, such as hurricanes, storm fronts, and jet streams, providing a comprehensive view of how weather systems develop and move over time.



SOURCES OF WEATHER DATA

ONLINE DATABASES

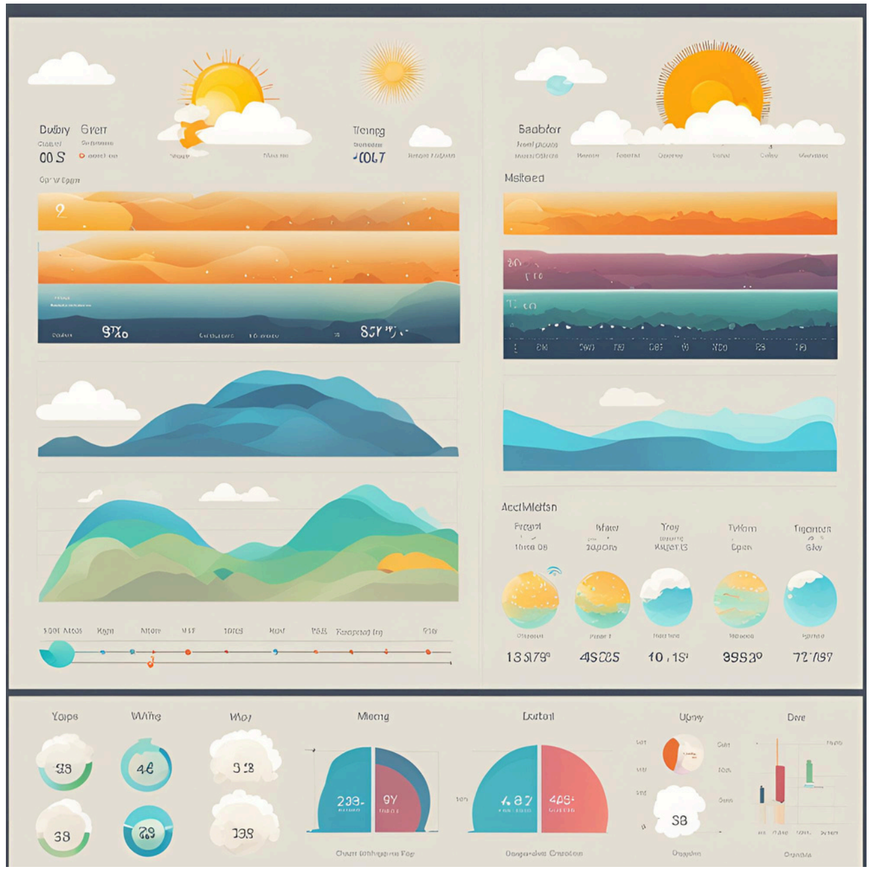
- Online databases are digital platforms that store and provide access to vast amounts of weather data collected from various sources, including weather stations, satellites, radars, and weather models. These databases allow users to retrieve both historical data (past weather records) and real-time data (current weather conditions).
- Examples: NOAA (National Oceanic and Atmospheric Administration), World Weather Online, Weather Underground, and others
- The data available on these platforms can include temperature records, rainfall amounts, wind speeds, atmospheric pressure readings, and satellite images

IMPORTANCE OF WEATHER DATA

- Agriculture
- Disaster Management
- Climate Research
- Aviation and Maritime

WEATHER DATA TYPE

- Most weather data types are numeric. This is because they represent measurable quantities that can be expressed in numbers, allowing for statistical analysis, forecasting, and visualization.



KEY WEATHER FEATURES AND DATA TYPE

Temperature

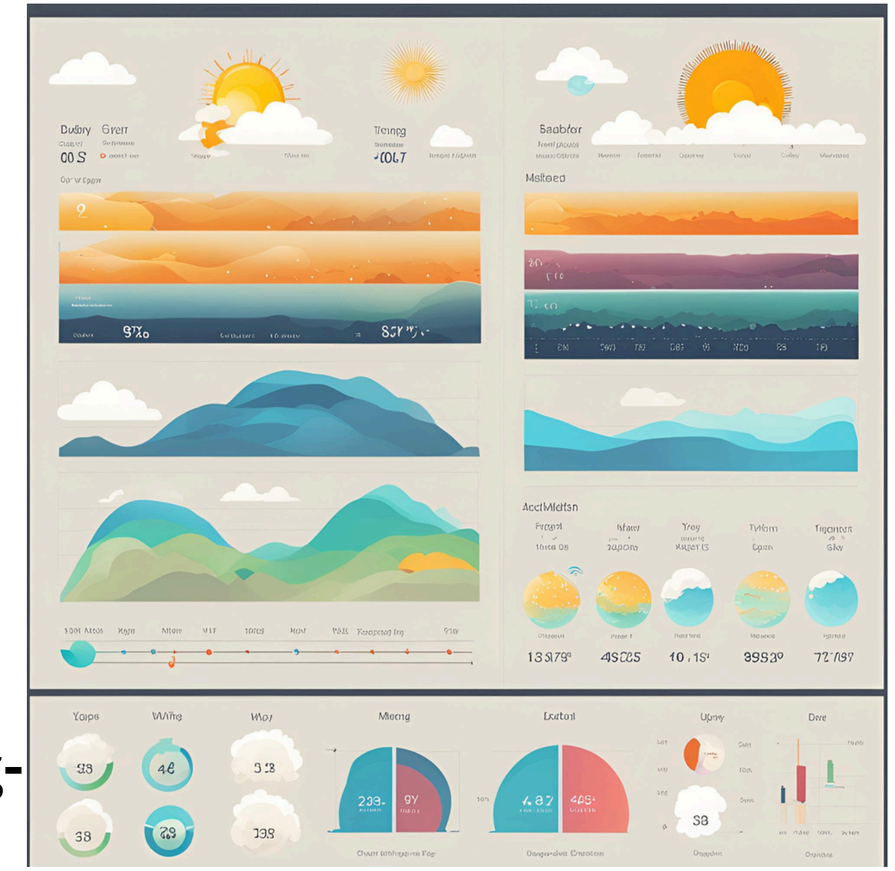
- Description: Measure of warmth or coldness.
- Units: Degrees Celsius (°C) or Fahrenheit (°F).
- Numeric Nature: Typically represented as a floating-point number, e.g., 25.6°C or 78.1°F.

Precipitation

- Description: Amount of rainfall, snowfall, etc.
- Units: Millimeters (mm) or inches.
- Numeric Nature: Represented as a floating-point or integer, e.g., 10.5 mm or 0.42 inches.

Humidity

- Description: Amount of water vapor in the air.
- Units: Relative humidity as a percentage (%).
- Numeric Nature: Represented as an integer or floating-point, e.g., 65%.



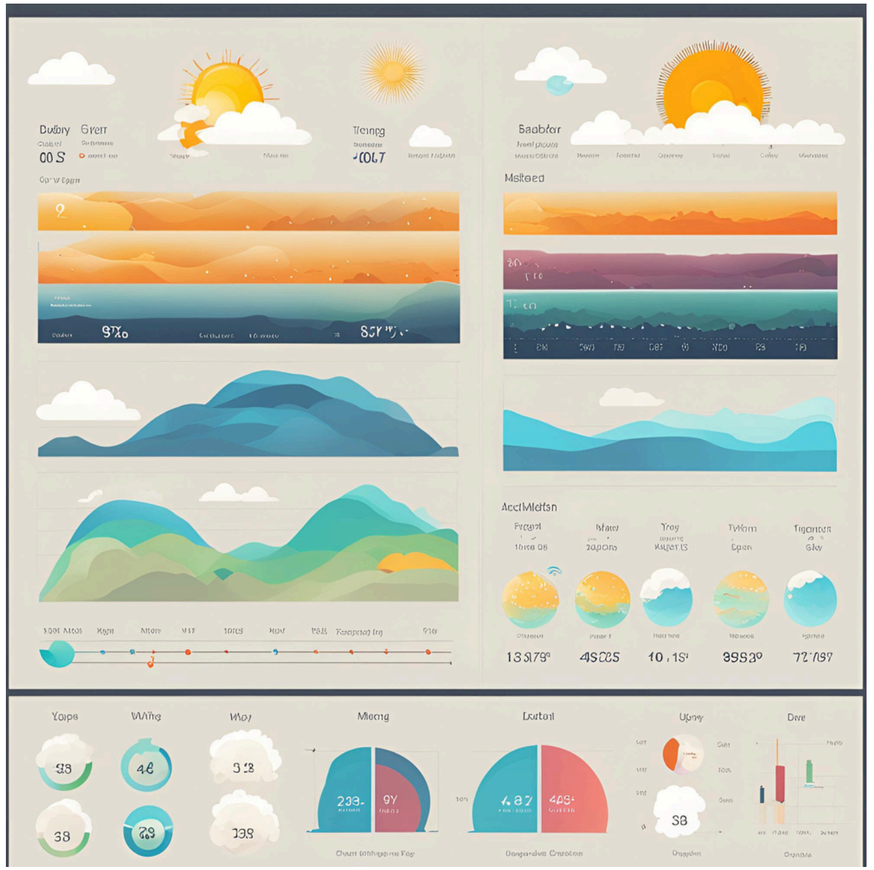
KEY WEATHER FEATURES

Wind Speed

- Units: Meters per second (m/s), kilometers per hour (km/h), or miles per hour (mph)
- Example: 5 m/s, 18 km/h

Wind Direction

- Units: Degrees (°), typically on a 360-degree scale where 0° or 360° represents North
- Example: 90° (East), 270° (West)



KEY WEATHER FEATURES

Solar Radiation

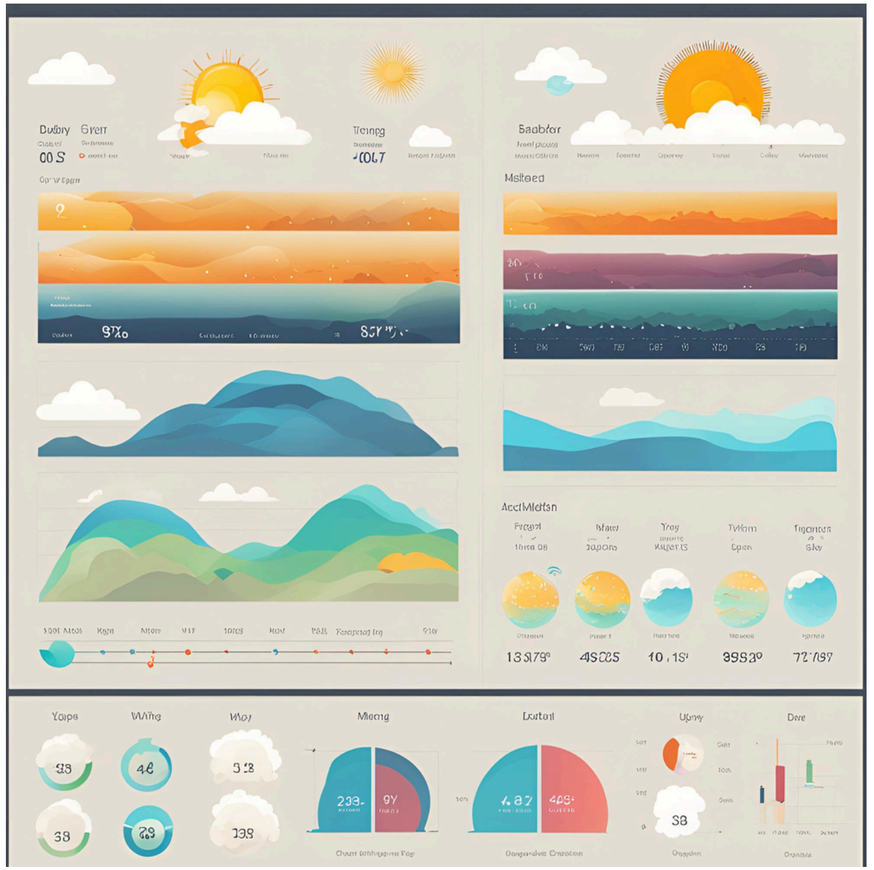
- Units: Watts per square meter (W/m²)
- Example: 800 W/m

Evapotranspiration

- Units: Millimeters per day (mm/day)
- Example: 4 mm/day

Snowfall and Snow Depth

- Units: Centimeters (cm) or inches
- Example: 20 cm of snowfall, 8 inches of snow depth



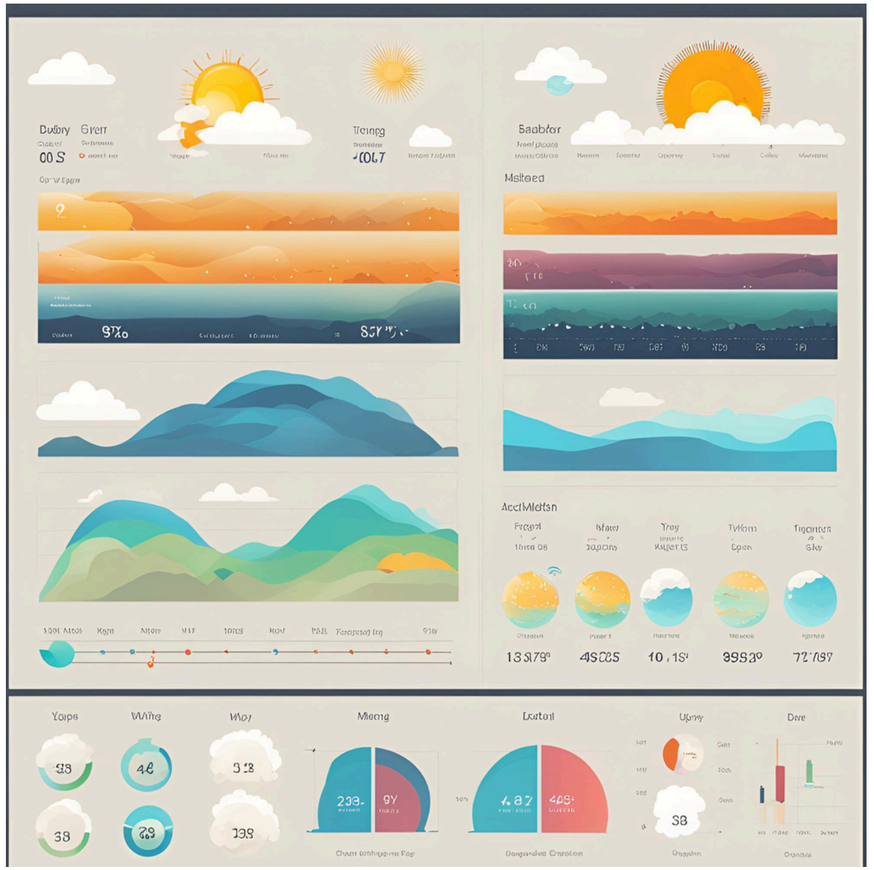
KEY WEATHER FEATURES

Atmospheric Pressure

- **Units: Hectopascals (hPa) or millibars (mb)**
- **Example: 1013 hPa, 1013 mb**

Air Quality Data

- **Units: Micrograms per cubic meter (µg/m³) or parts per million (ppm)**
- **Example: 35 µg/m³ PM2.5, 0.04 ppm CO**



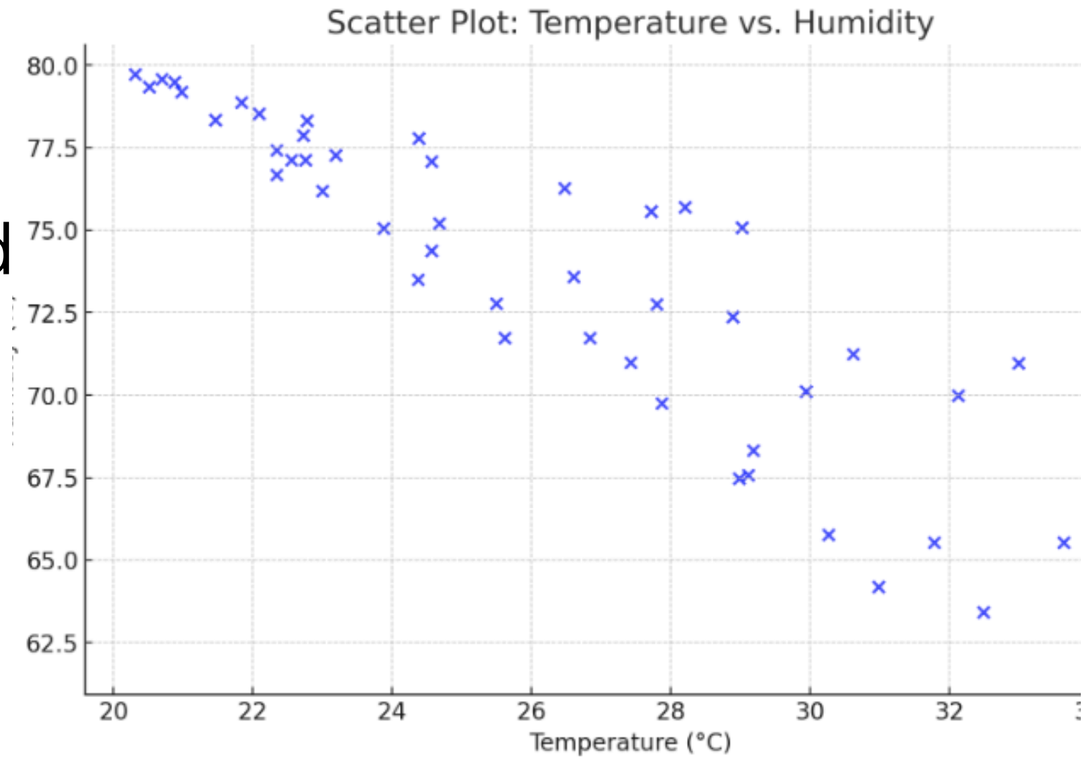
BASIC STATISTICAL ANALYSIS DIRECTIONS

CORRELATION

- correlation measures the degree of relationship between two variables
- Correlation does not imply causation.
- correlation doesn't capture causality but the degree of interrelation
- The correlation between x and y is the same as between y and x.

CORRELATION

- **EXAMPLE OF DATA ANALYSIS QUESTION:**
- How does the temperature influence humidity levels in a specific region during the summer months?
- Context: You want to understand whether higher temperatures in a region during summer are associated with increased or decreased humidity levels
- Scatter Plot is used to show correlation. For this plot correlation coeffient is -0.87



BASIC STATISTICAL ANALYSIS DIRECTIONS

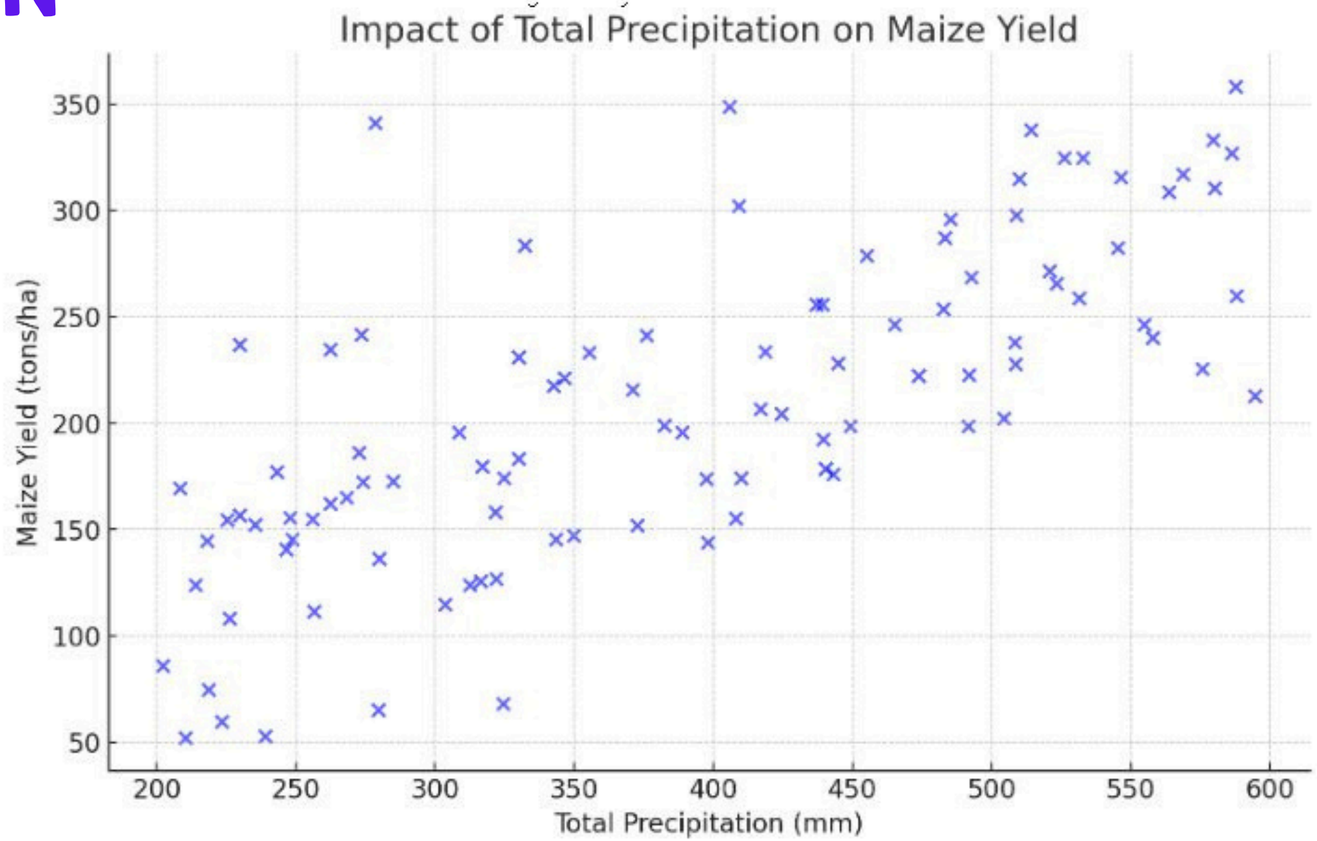
SUMMATION

- **Example Data Analysis Question:**
- How did the total precipitation and cumulative temperature during the growing season (March to May) in Eastern Nigeria impact maize crop growth?

- If the total precipitation was high and cumulative temperatures were within a favorable range, the conditions would likely have been good for maize growth. Conversely, low precipitation or excessively high cumulative temperatures might indicate challenges, such as water stress or heat stress, which could negatively impact crop yield.

BASIC STATISTICAL ANALYSIS DIRECTIONS

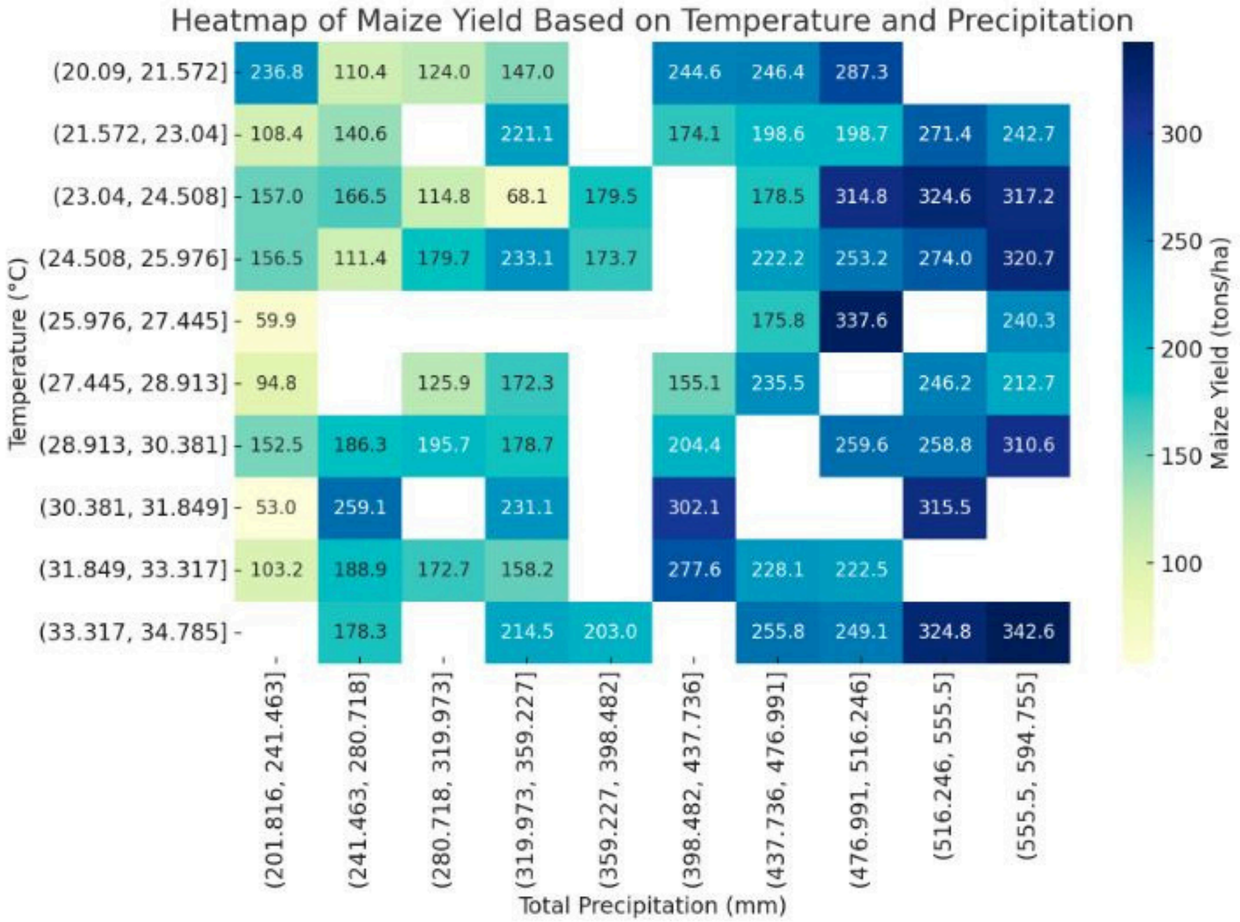
SUMMATION



SCATTER PLOT: THIS PLOT SHOWS THE RELATIONSHIP BETWEEN TOTAL PRECIPITATION AND MAIZE YIELD. AS THE TOTAL PRECIPITATION INCREASES, THERE SEEMS TO BE A POSITIVE TREND IN MAIZE YIELD, ALTHOUGH THERE IS SOME VARIABILITY DUE TO OTHER FACTORS.

BASIC STATISTICAL ANALYSIS DIRECTIONS

SUMMATION



HEATMAP: THIS HEATMAP ILLUSTRATES HOW BOTH TEMPERATURE AND PRECIPITATION JOINTLY AFFECT MAIZE YIELD. HIGHER MAIZE YIELDS ARE GENERALLY ASSOCIATED WITH MODERATE PRECIPITATION AND TEMPERATURES, WHILE EXTREME VALUES IN EITHER PARAMETER MAY LEAD TO REDUCED YIELDS.

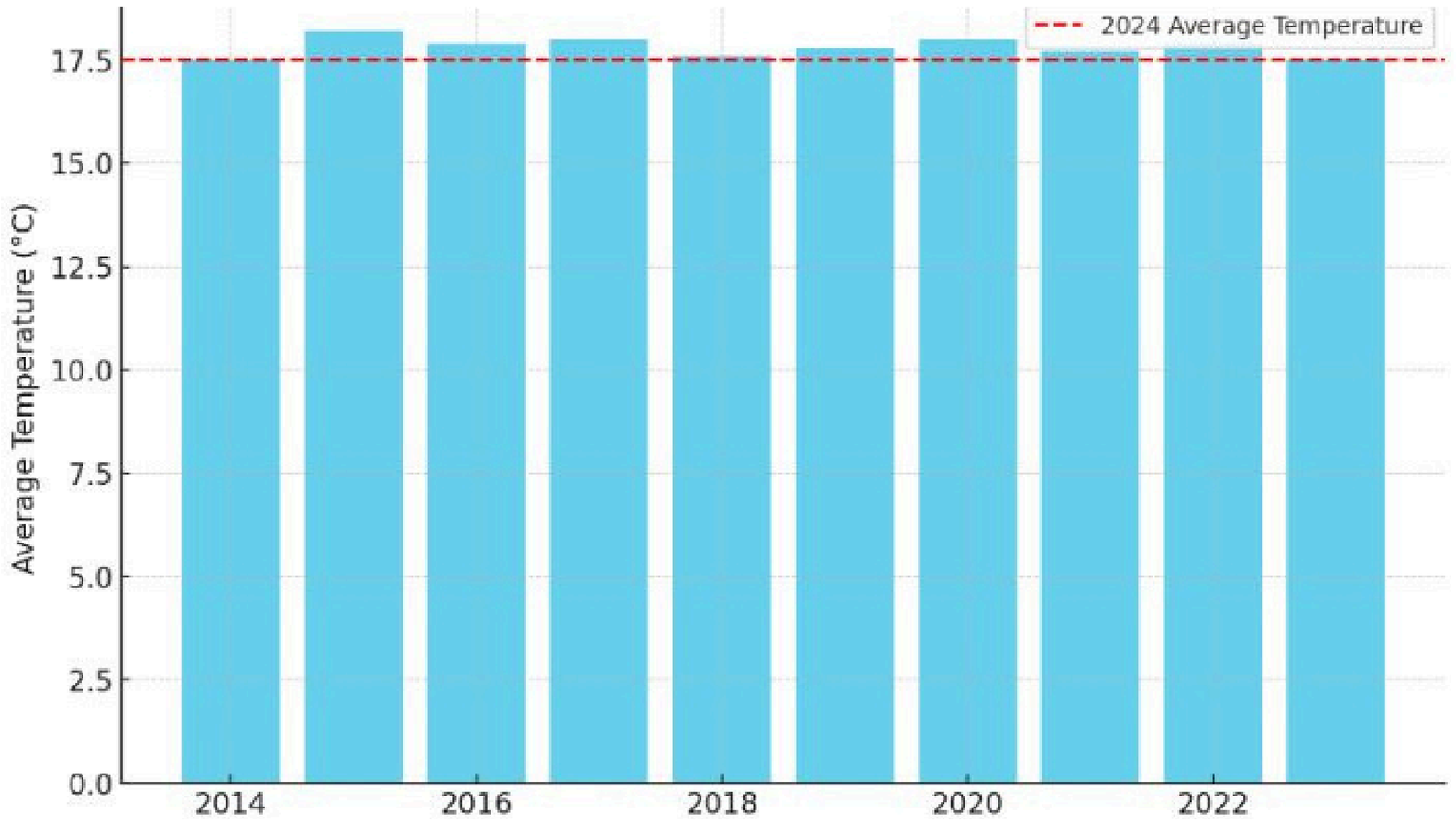
BASIC STATISTICAL ANALYSIS DIRECTIONS

MEAN

- **Example Data Analysis Question:**
- What is the average temperature in Nairobi, Kenya, during the month of July, and how does it compare to the historical average for the same month over the past 10 years?
- **Context:** Nairobi experiences relatively stable temperatures year-round, but slight variations in average temperatures during July could indicate changes in seasonal patterns. By calculating the mean temperature for July this year and comparing it to the historical average, you can determine if there are any significant deviations that might be related to climate change or other factors.

BASIC STATISTICAL ANALYSIS DIRECTIONS

MEAN



HERE IS A BAR CHART DISPLAYING THE AVERAGE TEMPERATURE IN NAIROBI, KENYA, DURING JULY FROM 2014 TO 2024. THE RED DASHED LINE REPRESENTS THE AVERAGE TEMPERATURE FOR JULY 2024, WHICH ALIGNS WITH THE HISTORICAL AVERAGES, INDICATING CONSISTENT TEMPERATURE PATTERNS DURING THIS PERIOD

BASIC STATISTICAL ANALYSIS DIRECTIONS

ANOMALIES

- Anomalies are deviations from the expected or normal pattern in data, often indicating unusual or unexpected behavior.

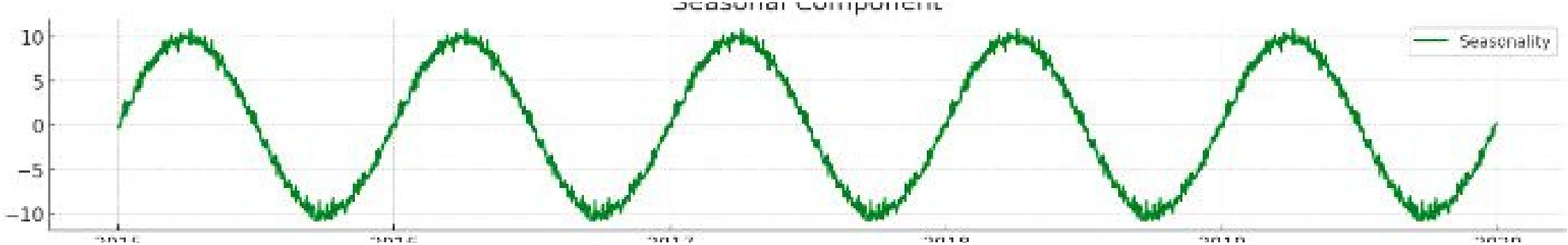
- Detecting anomalies in weather data involves identifying patterns or values that deviate significantly from the norm

BASIC STATISTICAL ANALYSIS DIRECTIONS

ANOMALIES

Weather data often does not follow a normal distribution due to its inherent characteristics and the influence of various factors:

- **Seasonality: Regular, repeating patterns in data that occur at consistent intervals, such as daily, weekly, or yearly cycles.**



THE SEASONAL PLOT REVEALS THE CONSISTENT YEARLY TEMPERATURE CYCLE, WITH WARMER TEMPERATURES IN THE MIDDLE OF THE YEAR (SUMMER) AND COOLER TEMPERATURES AT THE START AND END OF THE YEAR (WINTER).

BASIC STATISTICAL ANALYSIS DIRECTIONS

ANOMALIES

- **Skewness: When the distribution of data is not symmetrical, with a long tail on one side. For example, precipitation data might be heavily skewed because most days have little to no rain, but a few days have heavy rainfall.**

- **Spatial Variability: Differences in weather patterns across different locations. For example, temperature or precipitation might vary significantly from one region to another.**

BASIC STATISTICAL ANALYSIS DIRECTIONS

ANOMALIES

- Removing seasonality in weather data is a crucial step in anomaly detection, as it helps to isolate unusual patterns that are not part of the regular cyclical behavior. Two common approaches for this are Moving Average and Standardization.

BASIC STATISTICAL ANALYSIS DIRECTIONS

ANOMALIES: STATISTICAL METHODS

- How to detect Anomalies
- Moving Average
- Approach: Calculate the moving average for a weather variable (e.g., temperature) and compare actual values against this average.
- Anomalies: Deviations from the moving average that exceed a certain threshold can be considered anomalies.
- Example: A sudden temperature spike in winter could be flagged if it's significantly higher than the moving average for that period.

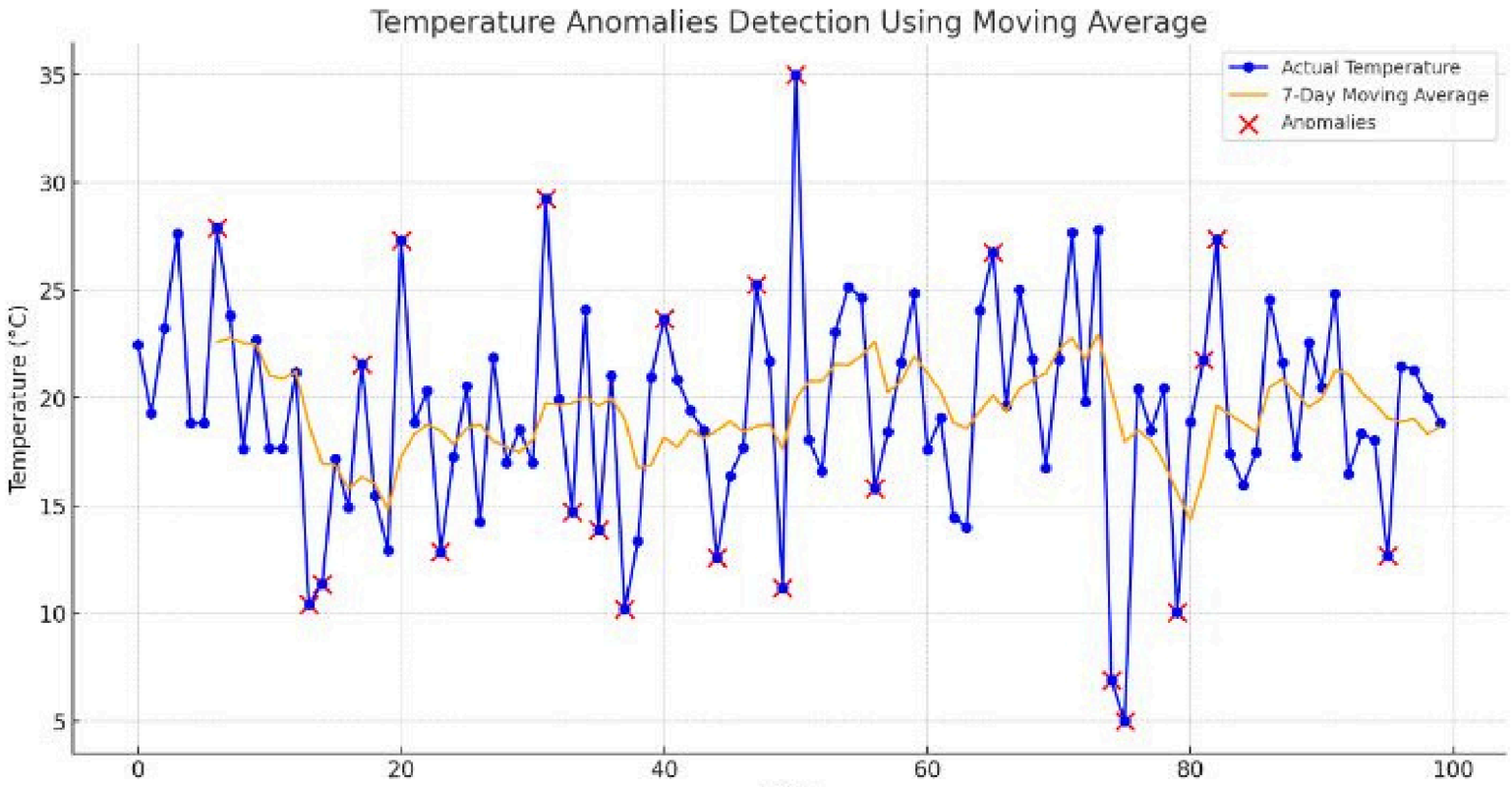
BASIC STATISTICAL ANALYSIS DIRECTIONS

ANOMALIES: MOVING AVERAGE

- Generate Dummy Temperature Data: Create a synthetic time series of daily temperatures over a period (e.g., 100 days).
- Calculate Moving Average: Compute the moving average for the temperature data using a specified window size (e.g., 7 days).
- Identify Anomalies: Flag days where the temperature deviates significantly from the moving average (e.g., by more than 5 degrees).
- Visualize the Data: Plot the actual temperatures, the moving average, and highlight the anomalies.

BASIC STATISTICAL ANALYSIS DIRECTIONS

ANOMALIES: MOVING AVERAGE



Blue Line: The actual daily temperatures over 100 days.

Orange Line: The 7-day moving average of the temperatures.

Red Markers: Days where the temperature significantly deviates from the moving average (by more than 5°C), identified as anomalies.

In this example, the temperature spike on day 50 and the drop on day 75 are flagged as anomalies because they deviate considerably from the moving average.

ANOMALIES:HOW TO DETECT ANOMALIES

STANDARDIZATION

- **To ensure Uniform Scale:** Different seasons can have different ranges for weather variables. For example, temperatures in winter and summer vary significantly. Standardizing seasonal data allows for comparison across different seasons on a uniform scale, making it easier to analyze seasonal trends and anomalies.

ANOMALIES:HOW TO DETECT ANOMALIES

Statistical Methods: Standardization-Z-Score

- **Approach: Standardize the data by subtracting the mean and dividing by the standard deviation to get a Z-score. This shows how many standard deviations a data point is from the mean.**
- **Anomalies: Data points with a Z-score above a certain threshold (e.g., |Z| > 2 or 3) are considered anomalies.**
- **Example: If the average temperature is 25°C with a standard deviation of 5°C, a Z-score of 5 would flag a temperature of 40°C as an anomaly.**

ANOMALIES:HOW TO DETECT ANOMALIES

Statistical Methods: Standardization-Z-Score

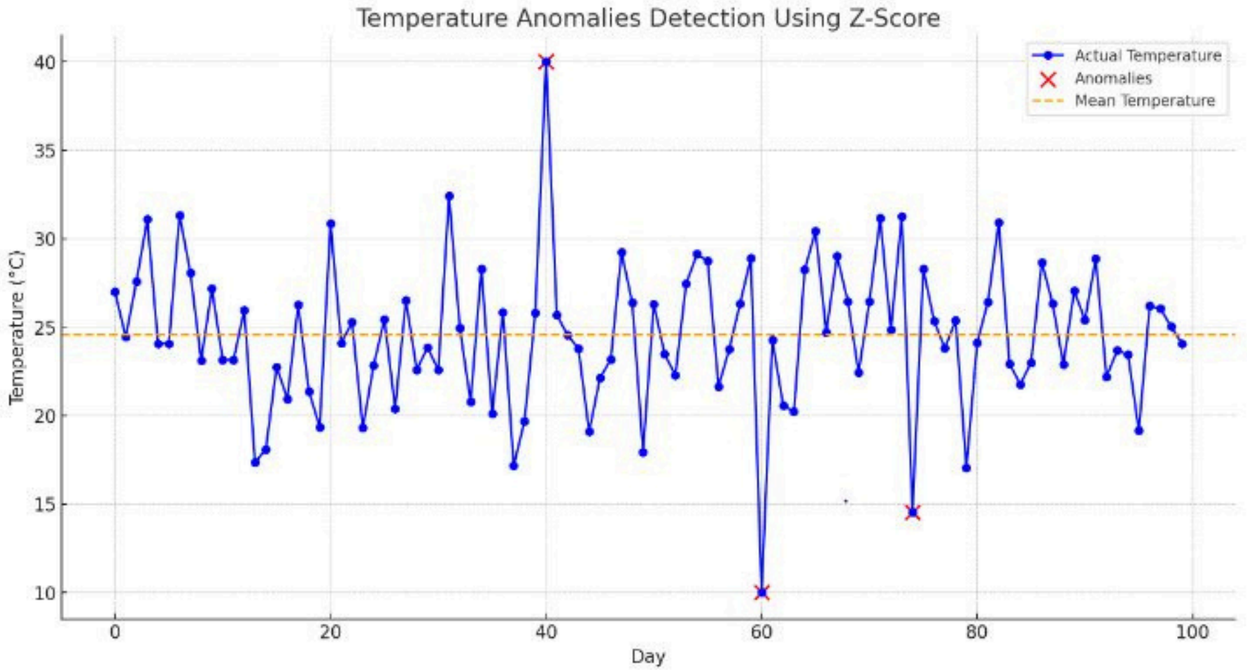
Apply Standardization: Use the following formula to standardize each weather variable X:

$$Z = \frac{X - \mu}{\sigma}$$

where:

- Z is the standardized value,
- X is the original value of the weather variable,
- μ is the mean of the weather variable,
- σ is the standard deviation of the weather variable.

ANOMALIES:HOW TO DETECT ANOMALIES, Z-SCORE



- **BLUE LINE:** THE ACTUAL DAILY TEMPERATURES OVER 100 DAYS.
- **ORANGE LINE:** THE MEAN TEMPERATURE.
- **RED MARKERS:** DAYS WHERE THE Z-SCORE EXCEEDS THE THRESHOLD OF $|Z| > 2$, IDENTIFIED AS ANOMALIES.
- IN THIS EXAMPLE, THE TEMPERATURE SPIKE ON DAY 40 AND THE DROP ON DAY 60 ARE FLAGGED AS ANOMALIES BECAUSE THEIR Z-SCORES INDICATE THAT THEY ARE MORE THAN TWO STANDARD DEVIATIONS AWAY FROM THE MEAN.

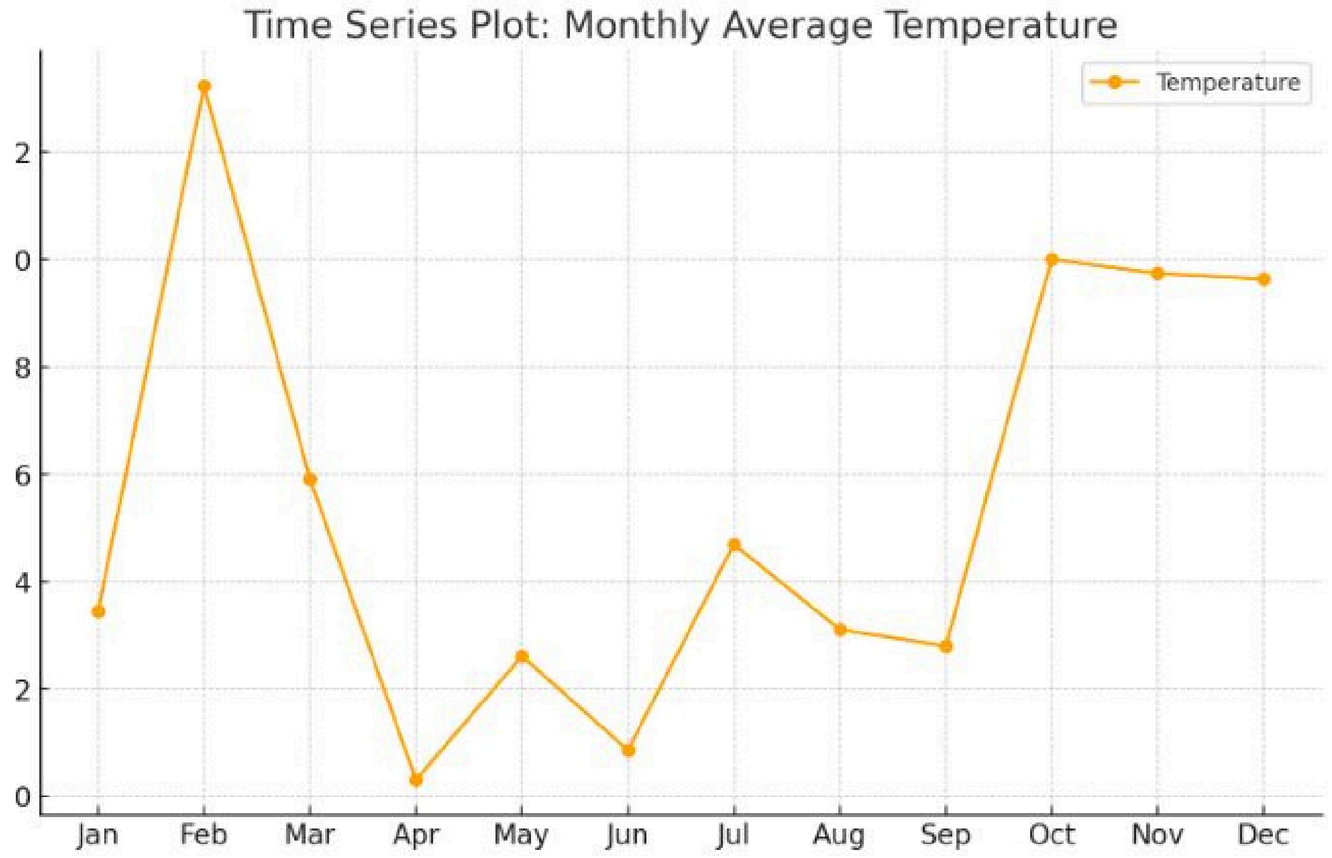
SOME VISUALISATION DIRECTIONS

VISUALIZATION DIRECTIONS

Time Series Plots

- Visualize temperature, precipitation, or wind speed over time.
- Visualize Trends: Time series plots are used to identify patterns, trends, and cycles in weather data over time. For example, they can help observe seasonal variations in temperature or long-term changes in precipitation.
- Detect Anomalies: They also make it easier to spot anomalies or irregularities, such as an unusually hot summer or a sudden spike in wind speed.

VISUALIZATION DIRECTIONS

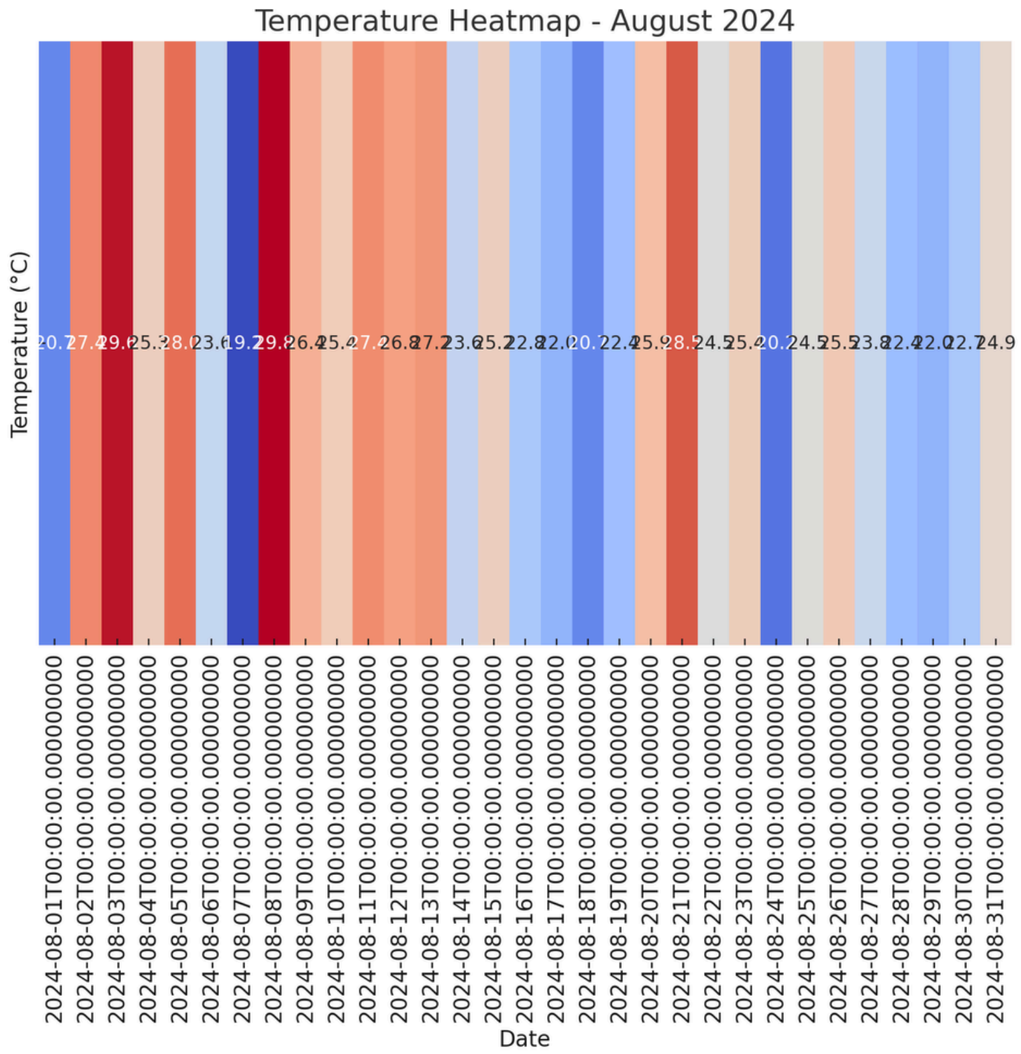


VISUALIZATION DIRECTIONS

HEAT MAP

A heatmap is a great way to visualize temperature or precipitation anomalies over time and across different regions or periods. The colors in the heatmap represent the magnitude of the anomaly, with different colors indicating positive, negative, or neutral deviations from the average.

HEATMAP



HEATMAP FOCUSING SOLELY ON THE TEMPERATURE DATA FOR AUGUST 2024. EACH COLUMN REPRESENTS A DAY IN AUGUST, AND THE COLOR INTENSITY SHOWS THE DAILY TEMPERATURE VARIATIONS

VISUALIZATION DIRECTIONS

- Correlation Matrices
- VISUALIZE RELATIONSHIPS BETWEEN DIFFERENT WEATHER VARIABLES.