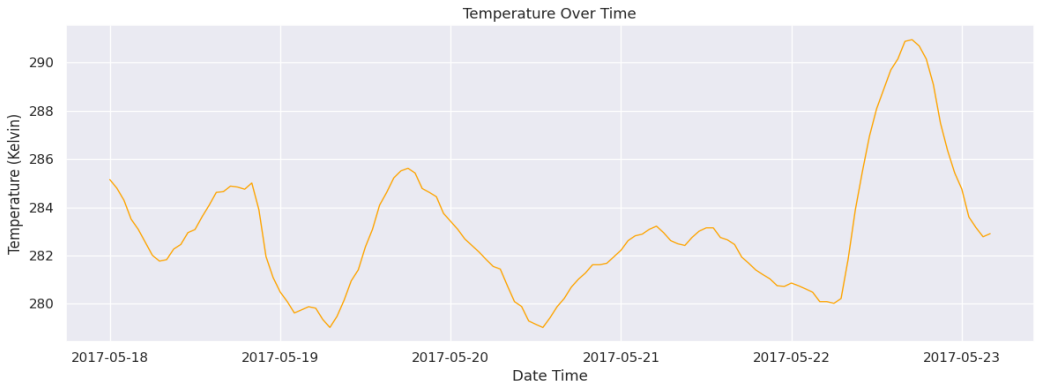
DATA SCIENCE

PROJECT BY:

PAVITHIRA . P

TEMPERATURE TREND OVER TIME



This is a **line chart** showing **temperature over time**, specifically measured in **Kelvin**.

Temperature Over Time

- The chart tracks how temperature fluctuated over a series of days.

Axes Explained

- **X-axis (horizontal):** Time (Date and Time format)
  - From **2017-05-18** to **2017-05-23**.
  - Likely shows temperature readings taken at **hourly intervals** or similar.
- **Y-axis (vertical): Temperature in Kelvin**
  - Ranges from **around 279 K to 291 K**.
  - For reference:
    - 273 K = 0°C
    - 293 K = 20°C

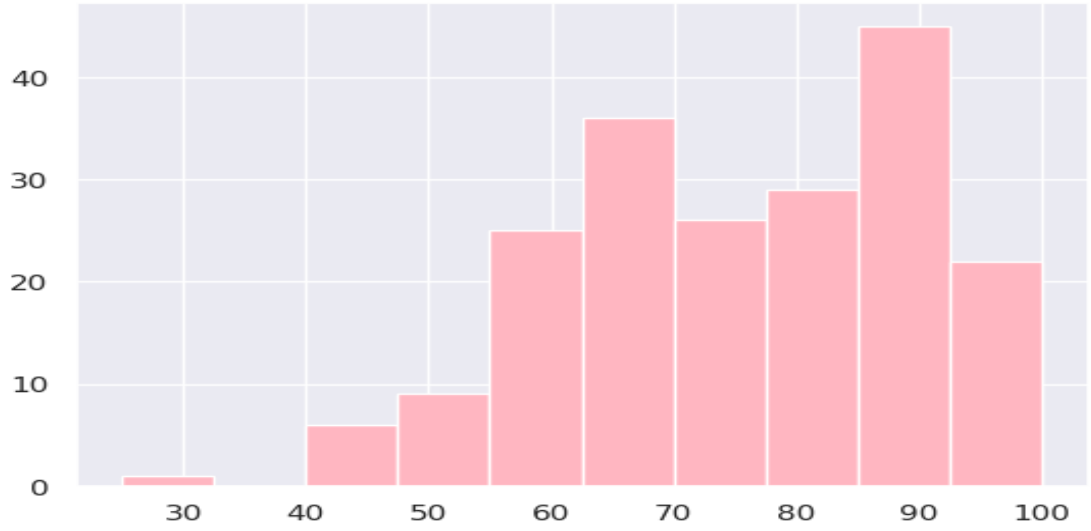
Observations & Trends

1. **Daily Fluctuations:**
   a. You can clearly see **up-and-down patterns each day**, which resemble **daily heating and cooling cycles** (day-night variation).
2. **Lowest Temperature:**

a. Around **279 K** (~6°C), typically during nighttime or early morning.

3. **Highest Temperature:**
    a. Peaks at **just above 291 K** (~18°C) on **May 22**, indicating a relatively warm day.
4. **Sharp Rise on May 22:**
    a. Steep climb in temperature, possibly due to:
        i. Sudden change in weather (e.g., cloud cover cleared)
        ii. Sunny day after several cooler days
5. **General Trend:**
    a. Temperature stays relatively stable from May 18–21, but increases significantly on May 22.

ANALYZING HUMIDITY RANGES



This is a **histogram**, which shows the **distribution of a continuous numerical variable**. Based on the typical appearance and range of values (from ~25 to 100), this could represent something like **temperature**, **humidity**, **exam scores**, or any numerical data grouped into **intervals (bins)**.

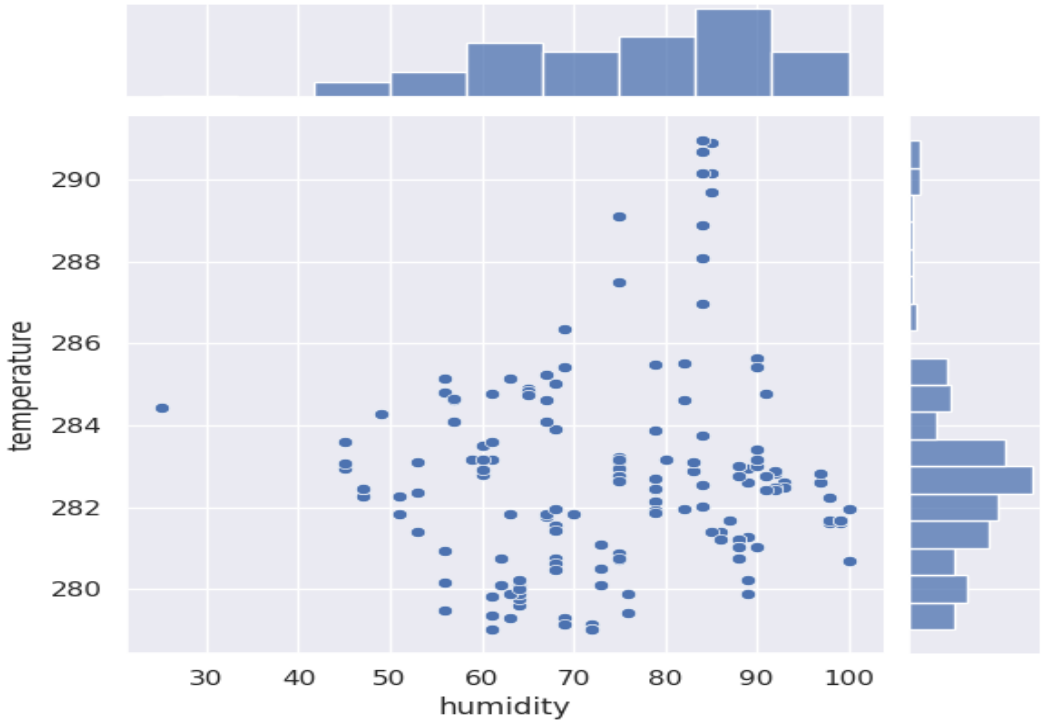
What the Histogram Shows

- **X-axis (horizontal):** Data values grouped into **bins** (intervals).
    - For example: 20–30, 30–40, 40–50, ..., 90–100.
- **Y-axis (vertical): Frequency** (count) of values that fall within each bin.
    - This tells you **how many times** data points appeared in each interval.

Insights from the Histogram

1. **Most Frequent Interval:**
    a. The bin for **90–100** has the **highest count**, around **45** values.
    b. This means a large number of values are concentrated in this high range.
2. **Other High Frequencies:**
    a. Bins **60–70**, **70–80**, and **80–90** also have noticeable heights — around **25 to 35** each.
    b. This suggests the data is **skewed toward the higher end** of the range.
3. **Least Frequent Intervals:**
    a. Bins **30–40** and **40–50** have very low frequencies.
    b. The **lowest bin (20–30)** has barely any data — perhaps just **1 or 2 values**.

HUMIDITY & TEMPERATURE INTERACTION PLOT



Here's a detailed description of the **scatter plot with marginal histograms** showing the relationship between **humidity** and **temperature**:

Axes:

- **X-axis (Humidity):**
  Values range from around **25% to 100%**.

- **Y-axis (Temperature):**
  Values range from approximately **278K to 292K** (in **Kelvin**).

Chart Elements:

Scatter Plot (Center):

- Each dot represents a data point — a pairing of humidity and temperature.
- The scatter shows a **wide spread** of temperature values across different humidity levels.
- Most of the data clusters between:
  - **Humidity:** 60% to 90%
  - **Temperature:** 280K to 285K

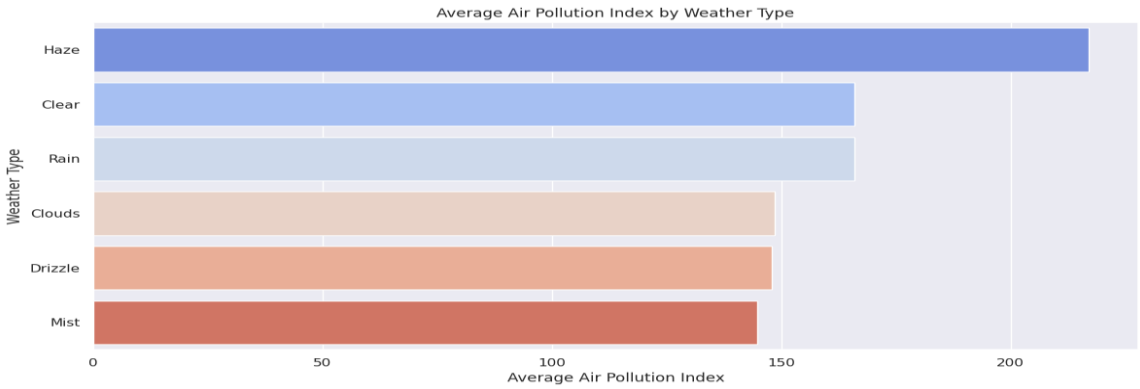
Marginal Histograms (Top and Right):

- These histograms summarize the distribution of each variable:
  - **Top (Humidity):**
  - Shows that most humidity values fall between **70% and 100%**, with peaks around **85%-90%**.
  - **Right (Temperature):**
  - The temperature histogram peaks around **281K to 284K**, indicating this is the most common temperature range.

Interpretation:

- While some clustering is visible (particularly around high humidity and moderate temperatures), the spread of points suggests a **weak or complex relationship**.
- The plot shows **no clear linear correlation** between humidity and temperature.

AVERAGE AIR POLLUTION INDEX ACROSS DIFFERENT WEATHER TYPES



This is a **horizontal bar chart** that shows the **Average Air Pollution Index** for different **weather types**. Let's break it down:

🧭 Axes Explained

1. **Y-axis (vertical):** Weather types
    a. Listed from bottom to top: Mist, Drizzle, Clouds, Rain, Clear, Haze
2. **X-axis (horizontal):** Average Air Pollution Index (likely AQI or a similar metric)
    a. Represents how polluted the air is on average during each weather condition.

🖼️ Insights & Interpretation

🌫️ Haze:

1. **Highest average pollution index** (above 210).
2. Expected, as haze usually results from suspended pollutants in the air.
3. Surprisingly high average pollution, **similar to Rain** (both around 165–170).
4. Could indicate stagnant air on sunny days, allowing pollutants to accumulate.

🌧 **Rain:**

1. Also fairly high pollution, which might seem unexpected.
2. Could be due to temporary pollution before rainfall or urban pollution patterns.

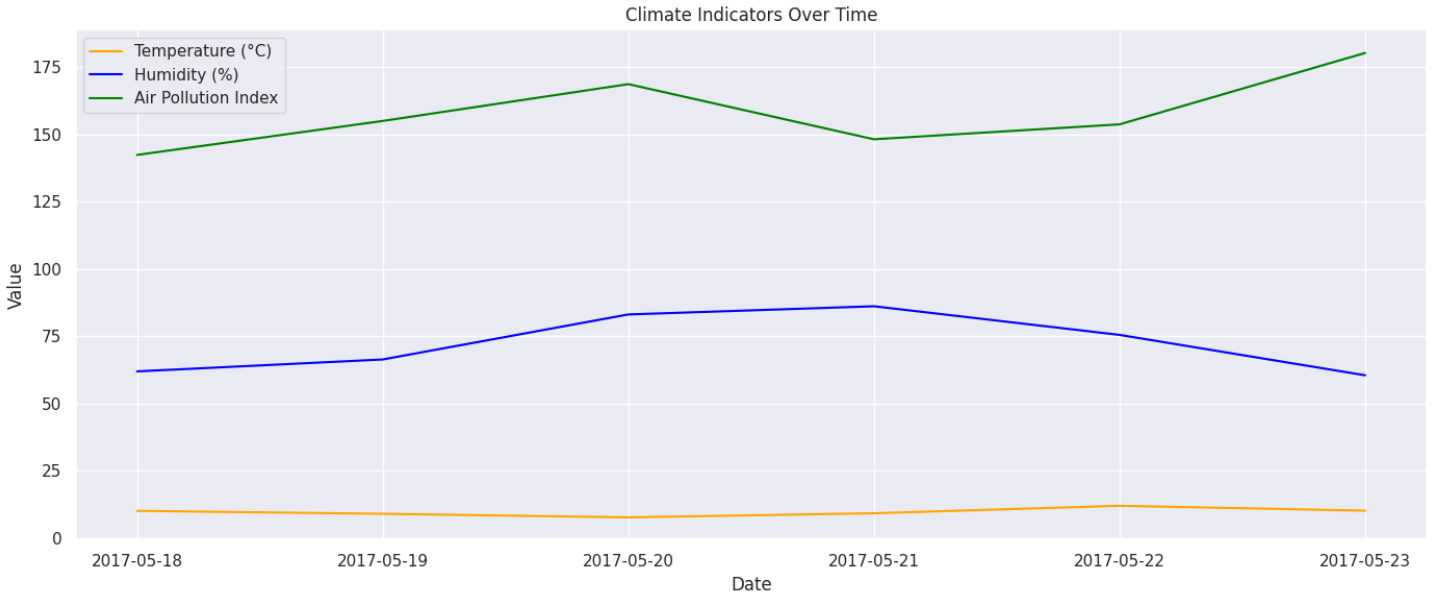
☁ **Clouds /** 🌦 **Drizzle /** 🌫 **Mist:**

1. These three show **similar and lower pollution levels**, around **145–150**.
2. Moisture might help wash pollutants out of the atmosphere or dilute them.

☑ **Key Takeaways**

- **Haze is associated with the worst air quality**, supporting the idea that it's often caused by high pollution levels.**Rain, Clear skies, and Haze** seem to be conditions with **higher pollution**.
- **Misty and drizzly weather** correlates with **cleaner air**, likely due to **humidity reducing dust and particles**.

CLIMATE INDICATOR OVER TIME



Axes:

- **X-axis (Date):**
  Shows dates from **2017-05-18 to 2017-05-23**, representing a 6-day time period.
- **Y-axis (Value):**
  Represents the numerical values of the three climate indicators—Temperature (°C), Humidity (%), and Air Pollution Index.

Trend Observations:

- ***Temperature (Orange Line):***
  - Begins slightly above 10°C on **May 18**.
  - Decreases until **May 20**, reaching its lowest value around 8°C.
  - Gradually increases, peaking at around **13°C on May 22**, and then slightly drops on May 23.
- ***Humidity (Blue Line):***
  - Starts at around **62%** on **May 18**.

- Peaks at **around 87% on May 21**, then decreases sharply to **61% by May 23**.

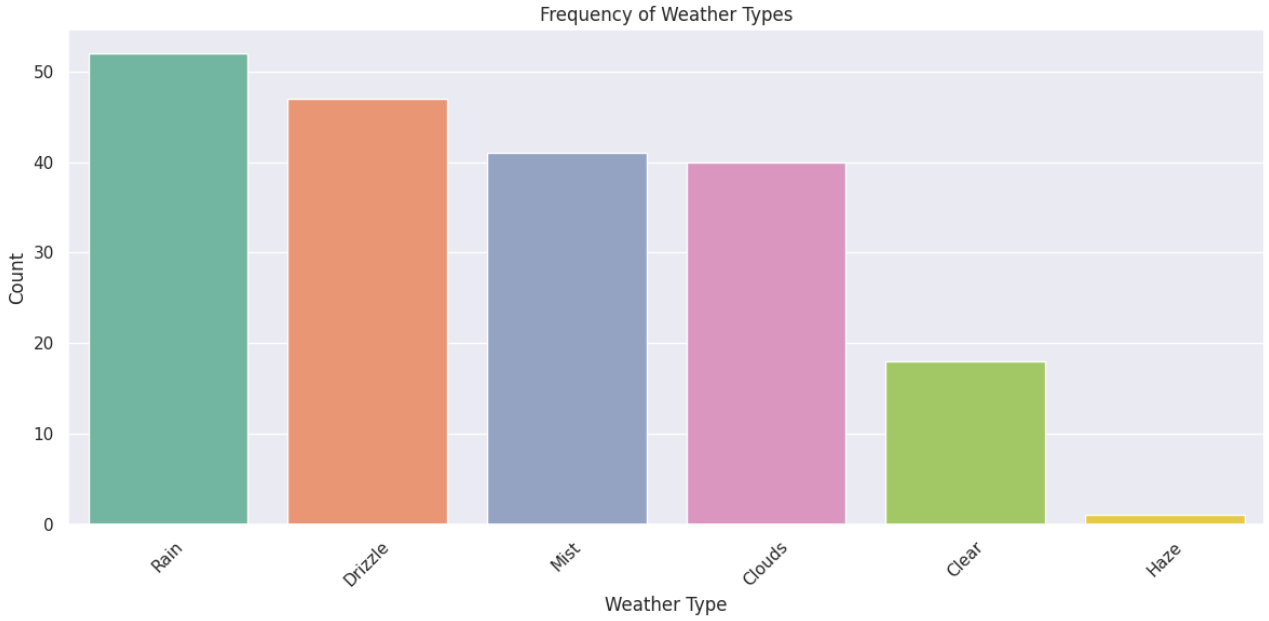
Air Pollution Index (Green Line):

- Starts at **around 142** on **May 18**.
- Gradually increases to **168 on May 20**, then dips slightly on May 21.
- Peaks at **181 on May 23**.

Interpretation:

- There's a **noticeable inverse relationship between temperature and humidity** during certain days: as temperature rises, humidity tends to drop.
- **Air pollution consistently increases** over the 6-day period, suggesting worsening air quality, possibly due to weather conditions or increased emissions.

FREQUENCY OF WEATHER TYPES



What the Chart Shows

1. **X-axis (horizontal):** Different **weather types**.
    a. Rain
    b. Drizzle
    c. Mist
    d. Clouds
    e. Clear
    f. Haze
2. **Y-axis (vertical):** The **count** or **frequency** of how often each weather type occurred.

Insights from the Chart

a. **Most Common Weather:**
    1. **Rain** is the most frequent, appearing **over 50 times**.
    2. **Drizzle** (~47 times), **Mist** (~41 times), and **Clouds** (~40 times) are also frequent.

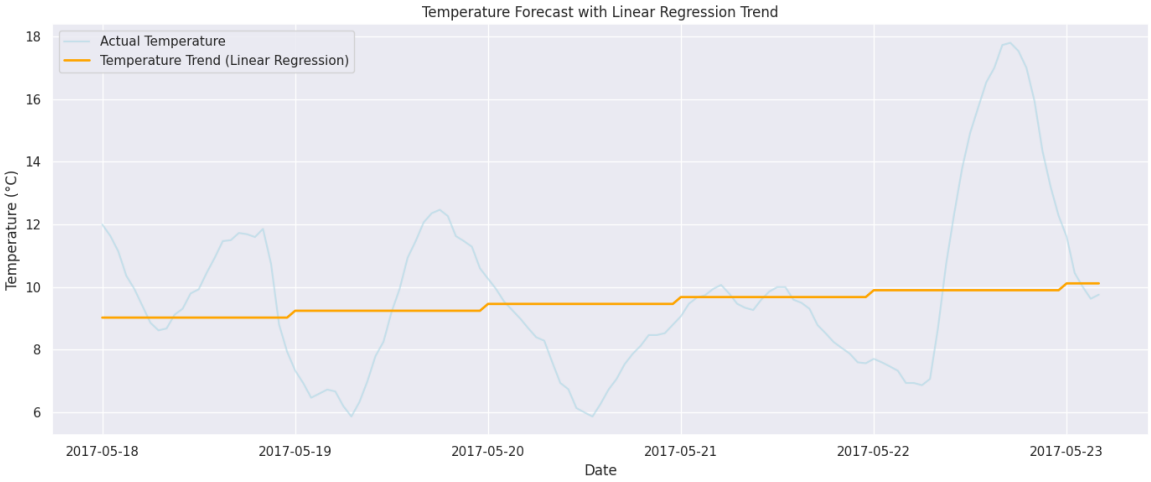
b. **Less Frequent Weather:**

- **Clear** weather occurs far less often — only around **18 times**.
- **Haze** is the least common, showing up **just once or twice**.

Interpretation

1. This suggests the location (or dataset) experiences **predominantly wet and cloudy conditions**.
2. **Sunny or clear days are rare**, and **hazy conditions** are almost negligible.

TEMPERATURE FORECAST WITH LINEAR REGRESSION TREND



Axes:

- **X-axis (Date):**
  Spans from **May 18, 2017, to May 23, 2017**, representing a 6-day period.
- **Y-axis (Temperature °C):**
  Displays temperature values ranging from around **6°C to 18°C**.

Visual Elements and Interpretation:

- ***Actual Temperature (Light Blue Line):***
  - The line fluctuates significantly, showing **hourly or sub-daily variations**.
  - Notable patterns:
    - **Initial Drop:** From around **12°C on May 18**, the temperature drops to **below 7°C on May 19**.
    - **Slight Rise and Fall:** A cyclical pattern follows, with temperatures oscillating between **6°C and 12°C**.
    - **Sharp Spike:** On **May 22**, there's a steep increase, peaking at around **17.8°C**.
    - After the spike, the temperature dips again, stabilizing near **10°C** by **May 23**.
- ***Temperature Trend (Linear Regression - Orange Line):***
  - This line represents a **simplified linear trend** derived from the actual temperature data.

- It increases **gradually and in steps** over time:
  - Starting near **9°C**, it shows small consistent increments, ending just above **10°C**.
  - This stepped appearance suggests **binned regression results**, likely from averaging over discrete periods or applying regression to smoothed segments.

CONCLUSION

- The analysis shows a clustered relationship between humidity and temperature, mostly at moderate levels.
- Air pollution is highest during haze and lowest in misty or cloudy conditions. Weather type alone doesn't determine air quality—other factors also play a role.