**Dataset Information**

* **Dataset File**: air\_quality\_data.csv
* **Dataset Description**:

The dataset contains daily air quality measurements collected over a year. It includes:

* **Pollutant Levels**: Measurements of pollutants like PM2.5, PM10, NO2, and O3.
* **No**: Record number.
* **Timestamp of the measurement:** year, month, day, hour  **Pollutant levels:** PM2.5, PM10, SO2, NO2, CO, O3.
* **Weather Data**: TEMP, PRES, DEWP, RAIN, wd, WSPM (temperature, pressure, dew point, rain, wind direction, wind speed).
* **Station**: Monitoring station name.

**Tasks**

**Section 1: Data Cleaning**

1. **Identify Issues**:
   * Identify and document any missing values, outliers, or inconsistencies in the dataset.
   * For example, check for negative pollutant levels or unrealistic weather data.
2. **Handle Missing Values and Outliers**:
   * Handle missing values using appropriate methods (e.g., mean, median, or interpolation).
   * Handle outliers by either removing or adjusting them.
   * Justify your approach in a **Markdown cell**.

**Section 2: Feature Engineering**

1. **Add New Columns**:
   * + Add a new column called air\_quality\_index that calculates a simple air quality index (AQI) using the formula:

**AQI = (PM2.5 + PM10 + NO2 + O3) / 4**

* + - Add a new column called pollution\_level that categorizes the air quality as:
    - "Good" if AQI ≤ 50
    - "Moderate" if 50 < AQI ≤ 100
    - "Unhealthy" if AQI > 100

1. **Additional Features**:
   * Create a new column called season that categorizes the date into seasons (Winter, Spring, Summer, Fall) based on the month.
   * Create a new column called weekday that categorizes days as "Weekday" or "Weekend".

**Section 3: Visualization**

1. **Time-Series Plot**:
   * Create a **time-series plot** of air\_quality\_index over the year.
   * Identify any seasonal patterns or trends in air quality.
2. **Pollutant Comparison**:
   * Create a **box plot** to compare the distribution of each pollutant (PM2.5, PM10, NO2, O3) across different seasons.
   * Write a short explanation of the insights derived from the box plot.
3. **Weather Impact on Air Quality**:
   * Create a **scatter plot** to explore the relationship between temperature and air\_quality\_index.
   * Write a short explanation of the relationship observed.
4. **Pollution Level Analysis**:
   * Create a **bar plot** showing the count of days in each pollution\_level category (Good, Moderate, Unhealthy).
   * Write a short explanation of the insights derived from the bar plot.
5. **Monthly Air Quality Patterns**:
   * Create a **box plot** to visualize the distribution of air\_quality\_index for each month.
   * Write a short explanation of any repeated patterns or anomalies.

**Good Luck!**