**Air Quality Assessment**

You are a data analyst at a company that monitors air quality in different cities. The company has collected a large amount of data on various pollutants and environmental factors, and you have been tasked with managing and analyzing this data using SQL and Python.

This dataset contains measurements of air quality in a specific city, including data on concentrations of various pollutants and meteorological variables. [Air Quality Data Set](https://archive.ics.uci.edu/ml/datasets/Air+Quality).

1. **SQL Assessment**

The objective of this task is to assess your ability to create a database and a table to store air quality data, insert records into the table, perform SQL queries to obtain specific information from the dataset, and connect to the database from a Python script to consume the data for analysis.

**Questions to Solve**

1. **Create a Database and Table** You must create a database called **AirQualityDB** and a table called **AirQuality** with the appropriate structure to store the provided data.
2. **Insert Data into the AirQuality Table** Insert data into the **AirQuality** table using the information from the **AirQualityUCI** file.
3. **Query to Identify the Worst Air Quality** Perform a query that selects the record with the worst air quality. For this, consider the highest concentrations of CO(GT), C6H6(GT), and NO2(GT).
4. **Query for Average and Variability** Perform a query to obtain the average and standard deviation of NO2(GT) for all recorded hours.
5. **Query with Specific Conditions** Perform a query that selects all records where the value of CO(GT) is greater than 2.0 and the temperature (T) is less than 15.0.
6. **Write a Python Script** Write a Python script that connects to the **AirQualityDB** database, performs a query to select all records, and displays the results in a panda Data Frame. After doing this, answer the following questions.
7. **Python assessment**

**Problem Description**

The objective is to predict pollution levels and classify the days as "Good Air," "Moderate," and "Bad" based on pollutant concentrations using the data provided in the SQL assessment and the database created.

**Questions to Solve**

1. **Data Exploration:**
   * Perform an exploratory data analysis (EDA). What general patterns can you observe in the pollutant concentrations?
   * Are there any outliers or missing values? How would you handle them?
2. **Data Preprocessing:**
   * Standardize and normalize the variables. Justify why this is important in this case.
   * Perform feature engineering to create new variables that may be useful in the analysis.
   * Perform any additional transformations you consider necessary to prepare the data for modeling.
3. **Predictive Analysis:**
   * Build a regression model to predict NO2 concentrations. Evaluate the model's performance using appropriate metrics.
   * Build a classification model to predict air quality (Good Air, Moderate, Bad). Evaluate the model's performance.
4. **Clustering Analysis:**
   * Apply clustering techniques to identify patterns in the air pollution data. How many clusters are appropriate? Justify your choice.
   * Analyze and interpret the obtained clusters. What useful information can you extract from them?
5. **Validation and Evaluation:**
   * Perform cross-validation of the predictive models. How do the performance metrics vary in each fold?
   * Discuss any limitations of the models and the analysis performed. What would you improve if you had more time or resources?
   * Propose possible improvements for the model or the analysis process.
6. **Documentation:**
   * Document the entire process in a Jupyter notebook and make it available on GitHub.
   * Ensure to clearly describe each step you took and your conclusions.