MECE Breakdown for Air Pollution Data

1. Data Segmentation (Mutually Exclusive, Collectively Exhaustive)

   a. By Time:

     - Daily pollution levels (e.g., Date column).

     - Hourly pollution variations (e.g., O3 1st Max Hour, CO 1st Max Hour).

     - Seasonal trends (grouping dates by seasons like Winter, Spring, etc.).

   b. By Location:

     - State-level data (e.g., State column: Arizona).

     - County-level data (e.g., County column: Maricopa).

     - City-level data (e.g., City column: Phoenix).

     - Address-specific data (e.g., Address column: exact location).

   c. By Pollutants:

     - Ozone (O3): Metrics like O3 Mean, O3 AQI.

     - Carbon Monoxide (CO): Metrics like CO Mean, CO AQI.

     - Sulfur Dioxide (SO2): Metrics like SO2 Mean, SO2 AQI.

     - Nitrogen Dioxide (NO2): Metrics like NO2 Mean, NO2 AQI.

2. Analysis Scope (Mutually Exclusive, Collectively Exhaustive)

   a. Descriptive Analysis:

     - Identify average pollution levels (mean values for O3, CO, etc.).

     - Find maximum pollution values and the hours they occur.

- Track the daily AQI for each pollutant.

  b. Predictive Analysis:

   - Predict AQI for specific pollutants (e.g., O3 AQI).

   - Forecast daily or hourly pollution levels.

   - Predict the most likely pollutant contributing to poor AQI.

3. Feature Categorization (Mutually Exclusive, Collectively Exhaustive)

  a. Time-Based Features:

   - Date (Date column).

   - Hour (O3 1st Max Hour, CO 1st Max Hour, etc.).

   - Seasonal groupings.

  b. Pollutant Features:

   - Mean levels (O3 Mean, CO Mean, etc.).

   - Max values (O3 1st Max Value, CO 1st Max Value, etc.).

   - AQI values (O3 AQI, CO AQI, etc.).

  c. Geographic Features:

   - State, County, City, and Address.

4. Machine Learning Task Design (Mutually Exclusive, Collectively Exhaustive)

  a. Regression Tasks:

   - Predict AQI for individual pollutants (e.g., O3, CO).

   - Predict the hourly maximum pollution levels.

b. Classification Tasks:

   - Classify AQI into health categories (Good, Moderate, Unhealthy, etc.).

   - Identify regions as high, medium, or low pollution zones.

c. Clustering Tasks:

   - Group locations with similar pollution trends.

   - Identify time periods with similar pollutant behavior.

5. Metrics for Monitoring (Mutually Exclusive, Collectively Exhaustive)

a. Pollutant-Specific Metrics:

   - Daily mean levels (O3 Mean, CO Mean, etc.).

   - Maximum values by pollutant.

   - AQI values by pollutant.

b. Aggregated Metrics:

   - Total AQI as a weighted sum of individual AQIs.

   - Peak pollution hours.