**10 Create List of Plants and Units in CEMS Documentation**

**Main Purpose**

Process emissions data from the Continuous Emissions Monitoring System (CEMS) for the years 2019-2022 by reading, cleaning, transforming, and consolidating the data. The script generates a list of plants and their corresponding units, saving the results in a structured format for further analysis. This Python script serves as a data processing tool, centralizing the reading, cleaning, transformation, and aggregation of CEMS datasets.

**Functions**

1. **Environment Setup**
   1. Import necessary libraries: pandas, os, glob, and downloads.globals\_regular.
   2. Define global variables for directory paths, including CEMS data directories and temp directories
2. **State File Processing**
   1. Define a function process\_state\_file(state, year) to read and process monthly emissions data for a given state and year.
   2. Read the initial CSV file for January and save it as a pickle file.
   3. Loop through the months from February to December, reading each month's data, concatenating it with the previous data, and saving the combined DataFrame.
   4. Remove temporary files after processing.
3. **Yearly Processing**
   1. Define a function process\_year(year) to process emissions data for all states in a given year.
   2. Loop through a predefined list of states, calling process\_state\_file for each state.
   3. Rename columns for consistency and clarity.
   4. Calculate the maximum load for each plant and filter out records with zero max load
   5. Save the processed DataFrame for each state into a consolidated pickle file.
4. **Processing for 2022**
   1. Define a function process\_2022() to handle emissions data specifically for the year 2022.
   2. Read and process emissions data for each state, renaming columns and filtering as done in previous years.
   3. Save the consolidated DataFrame for 2022.
5. **Plant List Creation**
   1. Define a function make\_plant\_list() to create a list of unique plants from the processed data for each year.
   2. Drop the unitid column and save the unique plant list for each year in a structured format.
6. **Main Execution Block**
   1. In the \_\_main\_\_ block, call process\_year() for the years 2019, 2020, and 2021, followed by process\_2022() and make\_plant\_list() to execute the entire processing workflow.