**20 Create Plant All Data Documentation**

**Main Purpose**

Consolidate and process emissions data from various sources to create a comprehensive master dataset of plants for the years 2019-2022. This script integrates data from the Continuous Emissions Monitoring System (CEMS), EPA crosswalks, and EIA 860 information, including plant identifiers, geographic information, and pollution damage estimates. The final dataset is saved in a structured format for further analysis.

**Functions**

1. **Environment Setup**
   1. Import necessary libraries: pandas, numpy, and Path from pathlib.
   2. Import configuration settings from downloads.globals\_regular.
   3. Define global variables for directory paths, including raw data, processed data, temporary files, and CEMS data directories.
2. **Balancing Authority Region Crosswalk**
   1. Read the Balancing Authority (BA) to region crosswalk from an Excel file.
   2. Rename relevant columns for clarity and consistency.
   3. Save the crosswalk data to a Stata file for further use.
3. **CEMS Plant List Creation**
   1. Create a master list of all plants in CEMS data by concatenating data from 2019 to 2022.
   2. Save the combined plant list to a Stata file.
4. **EPA Crosswalk Processing**
   1. Read the EPA crosswalk file containing plant information, including state and city.
   2. Convert the ORIS code to numeric and drop any rows with missing Plant Codes.
   3. Retain relevant columns and remove duplicates, saving the cleaned data to a temporary Stata file.
5. **EIA 860 Plant Information**
   1. Read the EIA 860 plant information from an Excel file.
   2. Select relevant columns for plant details and save the data to a temporary Stata file.
6. **Final Data Merging and Cleanup**
   1. Load the master CEMS plant list and drop unnecessary columns.
   2. Merge the final dataset with additional datasets containing plant location and pollution damage information.
   3. Rename columns for consistency and perform specific data cleaning tasks.
7. **Interconnection Assignment**
   1. Add an INTERCON column to categorize plants based on their NERC region.
   2. Assign default values and update based on specific conditions.
8. **Data Standardization and Saving**
   1. Perform any additional data cleaning and standardization steps.
   2. Save the final consolidated dataset to a Stata file for future analysis.