**04 read EIA Nuke Documentation**

**Main Purpose**

Process EIA nuclear data by reading, cleaning, transforming, and analyzing the data from the EIA Nuke datasets. The script merges capacity and outage data, performs internal consistency checks, and visualizes the results. This Python script serves as a data processing tool, centralizing the reading, cleaning, transformation, and initial analysis of EIA nuclear datasets. It sets up a consistent directory structure and file format (CSV) for further analysis.

**Functions**

1. **Environment Setup**
   1. Import necessary libraries: pandas, numpy, matplotlib, downloads.globals\_regular, datetime, and os.
2. **Data Reading**
   1. Read CSV files containing nuclear capacity and outage data from specified directories.
3. **Data Cleaning**
   1. Clean the facilityName column by stripping whitespace.
   2. Remove dashes from the period column and convert it to an integer type.
4. **Data Merging**
   1. Merge the capacity and outage DataFrames on the period and facilityName columns using an inner join.
5. **Data Filtering**
   1. Filter the DataFrame to exclude records with a period greater than or equal to 20220101.
   2. Exclude records where facilityName is "U.S. nuclear".
6. **Region Assignment**
   1. Define a function to assign regions based on the facilityName using a predefined mapping.
7. **Time Series Preparation**
   1. Calculate the UTC date from the period column and create a new column for hours by repeating the DataFrame index.
8. **Generation Calculation**
   1. Calculate the nuclear generation by subtracting the outage from the capacity.
   2. Rename the facilityName column to unitid and add a new column for PLANT.
9. **Data Export**
   1. Save the processed DataFrame to a CSV file for further analysis.
10. **Internal Consistency Check**
    1. Define a function to check the consistency of the processed data against EIA930 data by grouping and merging datasets.
11. **Data Loading**
    1. Load EIA930 data and preprocess it to create a new column for interconnections.
12. **Consistency Checks**
    1. Perform consistency checks by interconnection and region using the defined function.
13. **Visualization**
    1. Generate scatter plots to compare nuclear generation data from the processed dataset and EIA930 data.

**Core Issues and Potential Improvements**

1. **Error Handling**
   1. Implement try-except blocks to handle potential errors during file reading, writing, and data processing.
2. **Modularity**
   1. Refactor repeated operations (like data filtering and merging) into separate functions for better code organization.
3. **Progress Tracking**
   1. Add progress indicators or logging for long-running operations to improve user experience.
4. **Parameterization**
   1. Allow user input for date ranges and facility names to make the script more flexible and user-friendly.