Amy Zhang (Instance 332.0323)
Perplexity (Instance ID: Perplexity-20250323)

DATA PROFILE (Attempt no.2)

I. ‘U.S. Power Plants - Electricity Generation & Fuel Consumption (2015-2023)’
II. ‘Cooling System for Power Plants Data (2015-2023)’ + ‘Thermoelectric Data (Cooling Boiler Generator Details) (2015-2023)’

I. ‘U.S. Power Plants - Electricity Generation & Fuel Consumption (2015-2023)’

- merge of individual annual datasets from the US Energy Information Administration (EIA); filtered down to overlapping Plant ID’s with ‘Cooling System for Power Plants Data’
- filtered_power_plant_df.csv

1. SUMMARY

Data Sourcing:
"The U.S. Energy Information Administration (EIA) collects, analyzes, and disseminates independent and impartial energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment." This is an external data source. As government data, it is generally considered reliable.

Data Collection:

- 'US Power Plant Fuel and Generator Dataset (2023)' (df_1_USPR_filtered.csv): A filtered version of Schedule 2, 3, and 4 from the EIA-923 survey form for 2023 (https://www.eia.gov/electricity/data/eia923/). This filtering was performed to include only the plants also present in the water usage dataset (filtered_water_df). The data encompasses power plants in the Continental US and Puerto Rico.
- 'US Energy Storage Plant Data (2023)' (df_3_transformed.csv): Data from the "Energy Storage" sheet within the same EIA-923 survey form. It's important to note that this dataset was not filtered to match the water usage dataset. Instead, it is used to provide broader context around energy storage operations.

According to the EIA, the EIA-923 survey collects detailed monthly and annual electric power data on electricity generation, fuel consumption, fossil fuel stocks, and receipts at the power plant and prime mover level.
The EIA employs quality control measures such as automatic computerized verification of keyed input, subject matter specialist review, and follow-up with nonrespondents. Formulas using the past history of data values are implemented to automatically check data input for

errors, and discrepancies are resolved by contacting respondents. (https://www.eia.gov/electricity/monthly/pdf/technotes.pdf)

Data Contents:

- Power Plant Data (df_1_USPR_filtered.csv): This dataset focuses on the unique Plant IDs (918) also found in the Cooling Systems dataset, representing power plants located in the Continental US and Puerto Rico. It includes key metrics such as fuel consumption and electricity generation at the monthly level. (23,447 rows)

2. LIMITATIONS

- Sampling and Non-Sampling Errors: As acknowledged by the EIA, the data is subject to both sampling and non-sampling errors. Annual census data isn't subject to sampling error, but monthly sample survey data is. Non-sampling errors arise from various sources, including nonresponse, response errors, definitional ambiguities, differences in question interpretation, recording mistakes, and coverage or estimation errors for missing data.
- Potential Biases:
    - Reporting Bias: Plant operators may have incentives to underreport fuel consumption or other metrics if it reflects negatively on their environmental performance, leading to a potential reporting bias.
    - Selection Bias: The dataset may not be fully representative of all power plants in the US, as participation in EIA surveys can be mandatory for some and voluntary for others.
- Energy Storage Data Limitations: The Energy Storage dataset was not filtered based on the water usage dataset. This limits the ability to directly correlate water usage data with energy storage plant operations.

Ethical Considerations:

1. Transparency and Accountability: The EIA data is publicly available, which promotes transparency and accountability in the energy sector. Ensure that the data is interpreted and presented responsibly and without bias.
2. Data Privacy: Protect the identity of individual plants and operators when presenting or sharing the data. Consider using anonymization techniques if needed.
3. Informed Decision-Making: Use the data to inform decisions related to energy policy and water management in a manner that promotes sustainability and considers the needs of all stakeholders.
4. Environmental Justice Considerations:
   The selection and operation of energy facilities can disproportionately impact some communities. These factors are of utmost consideration in our project.

3. DATA RELEVANCE

These datasets provide essential plant metadata—from Census and NERC region details to primary mover types, EIA sector classifications, and electricity generation statistics. This metadata is crucial for contextualizing and analyzing water usage patterns across different types of power generation facilities and in relation to energy storage operations. The datasets allow for a nuanced understanding of the interplay between water, energy, and the environment, which is central to the project's objectives.

II. 'Cooling System for Power Plants (2015-2023)' + 'Thermoelectric Data (Cooler Boiler Generator Details) (2015-2023)'

- water_df_StateCleaned_2015_2023

1. SUMMARY

Data Sourcing: "The U.S. Energy Information Administration (EIA) collects, analyzes, and disseminates independent and impartial energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment."
This is an external data source. It is government data, and as such a trustworthy data source.

Data Collection: 'Cooling System for Power Plants (2015-2023)' is my merge of Schedule 8 (Environmental Data) from the annual survey Form EIA-923 (https://www.eia.gov/electricity/data/eia923/) 'Thermoelectric Data (Cooler Boiler Generator Details (2015-2023) is my merge of annual datasets sharing this name (https://www.eia.gov/electricity/data/water/) According to the website, "The survey Form EIA-923 collects detailed electric power data -- monthly and annually -- on electricity generation, fuel consumption, fossil fuel stocks, and receipts at the power plant and prime mover level." Observations are all at the month level: "Monthly data (M) -approximately 3,034 plants from the monthly survey." Further details as to methodology are as follows: "Automatic, computerized verification of keyed input, review by subject matter specialists, and follow-up with nonrespondents assure quality statistics. To ensure the quality standards established by the EIA, formulas that use the past history of data values in the database have been designed and implemented to check data input for errors automatically. Data values that fall outside the ranges prescribed in the formulas are verified by telephoning respondents to resolve any discrepancies. All survey nonrespondents are identified and contacted." (https://www.eia.gov/electricity/monthly/pdf/technotes.pdf)

Data Contents: The Cooling System data contains 141,287 observations at the Year-Month-Plant Id-Cooling System Id level, while Thermoelectric contains 726,660 observations at the Year-Month-Plant Id-Cooling System Id-Boiler Id-Generator Id level.

Together they contain important Water Metrics– most importantly, Total Withdrawal and Consumption Volume– and details such as Water Type and Water Source.

2. LIMITATIONS

Sampling and Non-Sampling Errors: As acknowledged by the EIA, the data is subject to both sampling and non-sampling errors. While annual census data isn't subject to sampling error, monthly sample survey data is. Non-sampling errors can arise from various sources, including nonresponse, response errors, definitional ambiguities, differences in question interpretation, recording mistakes, and coverage or estimation errors for missing data.

- Incomplete Water Metrics: While total withdrawal, discharge, and consumption volumes are provided, the absence of water rate metrics (e.g., withdrawal rate or flow rate) limits the ability to assess short-term impacts on water bodies.
- Incomplete Fuel Consumption Information: Although fuel consumption data is included, its completeness is limited compared to plant-level aggregated data (as available in the 'US Power Plant Fuel and Generator Dataset'). This may hinder comprehensive water-energy nexus analyses.
- Potential Biases:
    - Reporting Bias: There's a potential for reporting bias, as plant operators may have incentives to underreport water usage if it reflects negatively on their environmental performance.
    - Selection Bias: The dataset may not be representative of all power plants in the US, as participation in the EIA surveys is mandatory for some but voluntary for others. This could lead to a selection bias if plants with certain characteristics are more or less likely to participate.
- Definitional Issues: The dataset may contain ambiguities in the definition of certain terms (e.g., "cooling system"), which could lead to inconsistencies in reporting across different plants.

Ethical Considerations:

- Environmental Justice: The use of water for cooling power plants can have disproportionate impacts on local communities, particularly those that are already vulnerable to water scarcity or pollution. It's essential to consider environmental justice implications when analyzing this data.
- Transparency and Accountability: The EIA data is publicly available, promoting transparency and accountability in the energy sector. However, it's essential to ensure that the data is interpreted and presented in a responsible and unbiased manner.
- Data Privacy: Although the dataset doesn't contain personal information, it's essential to protect the identity of individual plants and operators when presenting or sharing the data. Anonymization techniques may be necessary.

- Informed Decision-Making: The data should be used to inform decision-making related to energy policy and water management in a way that promotes sustainability and considers the needs of all stakeholders.

3. DATA RELEVANCE

One of the main objectives of this project is to understand the impact the current electricity grid has on US water resources; the ecological impact of our current infrastructure. Therefore, water usage related to cooling systems is highly relevant for understanding the operations of our current technologies.

QUESTIONS

General Trends

1. How has the total electricity generation (Net Generation) changed over time (2015–2023)?
2. What are the trends in **fuel consumption** (e.g., natural gas, coal, biomass) across different years and months?
3. Are there seasonal patterns in electricity generation or fuel consumption?

Water Usage and Cooling Systems

4. How do water withdrawal and consumption volumes vary by cooling system type?
5. Which states or regions have the highest water withdrawal and consumption volumes for power plants?
6. Are there correlations between water withdrawal/consumption and fuel types used by power plants?
7. How does water usage intensity (gallons per MWh) vary across different cooling system types or fuel types?

Fuel Types and Energy Sources

8. What is the distribution of primary energy sources (e.g., natural gas, coal, nuclear, renewables) across power plants?
9. How has the share of renewable energy sources (e.g., wind, solar) changed over time?
10. Which fuel types are associated with the highest net generation efficiency (MWh per MMBtu)?

State-Level and Regional Analysis

11. Which states or regions have seen the largest increases or decreases in electricity generation?
12. How does electricity generation vary by census region or NERC region?
13. Are there regional differences in cooling system types or water usage?

Plant-Specific Analysis

14. Which plants have the highest net generation, and what are their primary fuel types?
15. Do combined heat and power (CHP) plants have different efficiency or water usage patterns compared to non-CHP plants?
16. What is the relationship between plant size (total fuel consumption) and water usage?

Environmental Impact

17. How has chlorine usage for cooling systems changed over time, and which plants are the largest users?
18. What is the relationship between water withdrawal/consumption and environmental regulations (e.g., state-level policies)?
19. Are there trends in reducing water intensity rates for cooling systems over time?

Sector-Specific Insights

20. How do residential, commercial, and industrial sectors contribute to electricity demand trends over time?
21. Which sectors rely most heavily on specific energy sources (e.g., natural gas for industrial vs coal for electricity)?

Advanced Comparisons

22. Are there significant differences in efficiency (MWh per MMBtu) across different generator technologies (e.g., steam turbines vs gas turbines)?
23. How do older plants compare to newer plants in terms of efficiency, water usage, and emissions?

Policy Implications

24. What role do combined heat and power plants play in meeting renewable energy targets?
25. How have state-level renewable portfolio standards influenced changes in fuel types used by power plants?