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

DATA PROFILE (Attempt no.3)

‘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’
- power_plant_df

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. These

make the datasets immediately relevant for developing models to predict the energy efficiency of electricity generation technologies. For future research, 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.

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?

Fuel Types and Energy Sources

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

State-Level and Regional Analysis

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

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?