Data Appendix

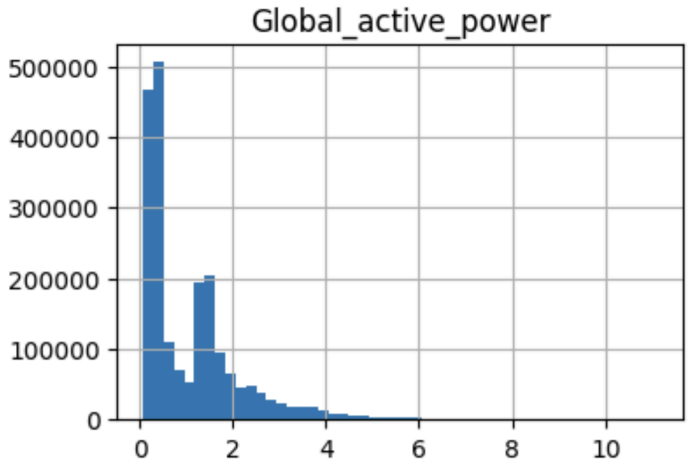
Analysis Data File: Power Consumption Analysis

Unit of Observation:
Each row in this dataset represents a single observation of power consumption for a household, recorded at one-minute intervals over 47 months. The dataset includes various features related to power consumption, voltage, and energy sub-metering, with some instances containing missing values.

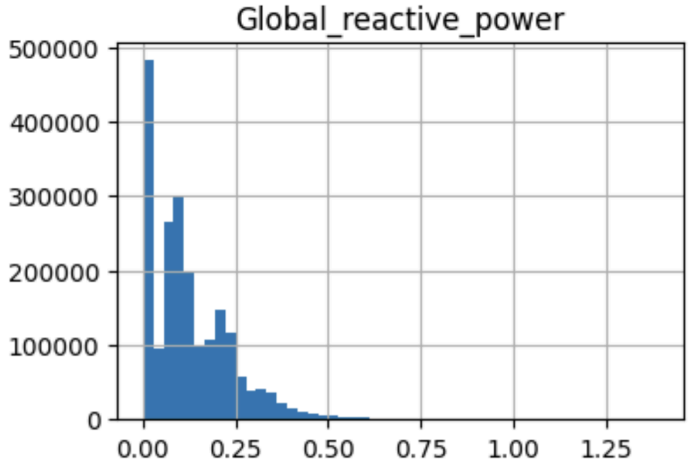
Overview of Analysis Data File:
The cleaned dataset, `power_data_cleaned`, is derived from the original file by forward-filling missing values from the previous valid row and resetting the index. This cleaned dataset serves as the basis for subsequent analysis.

Variables in the Analysis Data File:

1. **Global Active Power**
    - **Definition:** The total active power consumed by the household in kilowatts.
    - **Type:** Quantitative
    - **Summary Statistics (units kW):**
        - Count: 178852
        - Mean: 1.0916 kW
        - Standard Deviation: 1.0573
        - Min: 0.122000
        - 25th: 0.320000
        - Median: 1.264000
        - 75th percentiles: 2.104000
        - Max : 10.670000
    - **Visualization:** Histogram illustrating the distribution of global active power usage.

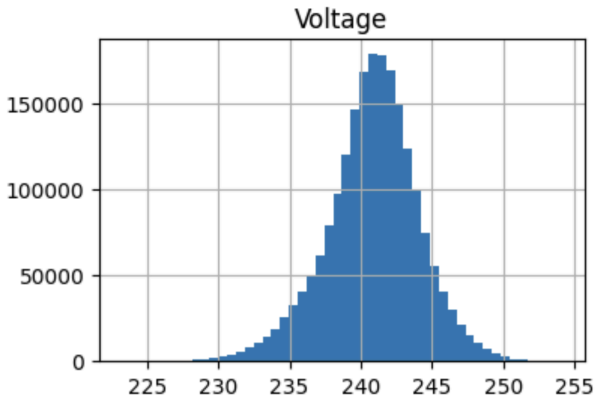


2. **Global Reactive Power**
    - **Definition:** The reactive power consumed in the household, representing non-productive power consumption.
    - **Type:** Quantitative
    - **Summary Statistics (units kW):**
        - Count: 178852
        - Mean: 0.121508 kW
        - Standard Deviation: 1.0573
        - Min: 0.122000
        - 25th: 0.320000
        - Median: 0.106000
        - 75th percentiles: 0.188000
        - Max : 0.874000
    - **Visualization:** Histogram showing the distribution of global reactive power usage.

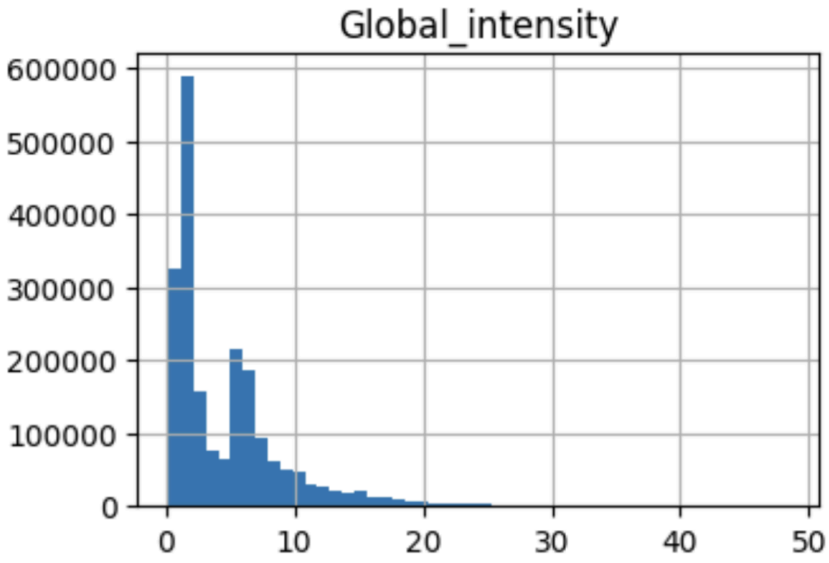


3. **Voltage**

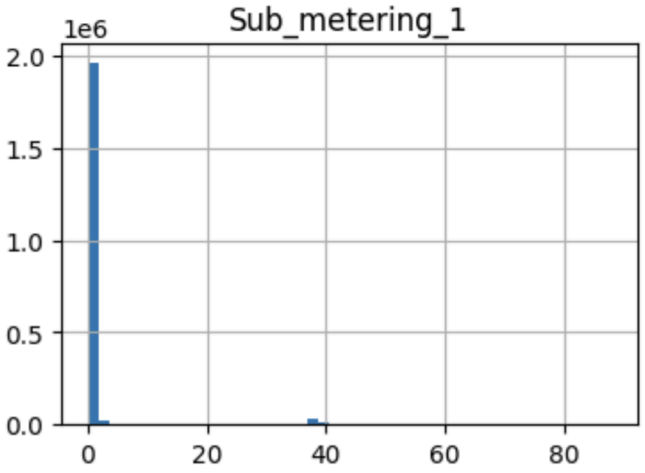
- **Definition:** Voltage measured at the household's distribution panel.
- **Type:** Quantitative
- **Summary Statistics(units V):**
  - Count: 178852
  - Mean: 240.720369 V
  - Standard Deviation: 3.141472
  - Min: 223.490000
  - 25th: 238.780000
  - Median: 240.880000
  - 75th percentiles: 242.830000
  - Max : 251.700000
- **Visualization:** Line plot showing voltage fluctuations over time.



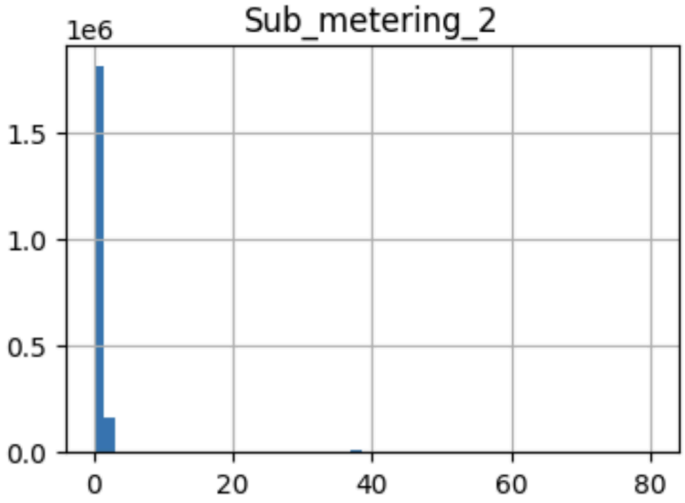
4. **Global Intensity**
   - **Definition:** Total current intensity of the household in amperes.
   - **Type:** Quantitative
   - **Summary Statistics (units A):**
     - Count: 178852
     - Mean: 5.923020 A
     - Standard Deviation: 5.478759
     - Min: 0.600000
     - 25th: 1.400000
     - Median: 5.200000
     - 75th percentiles: 8.800000
     - Max : 46.400000
   - **Visualization:** Histogram illustrating the distribution of global intensity.



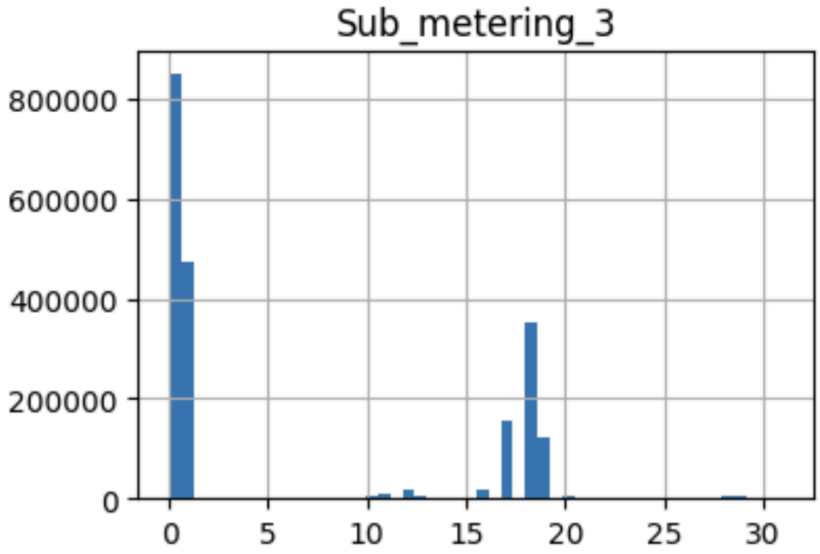
5. **Sub-Metering 1**
    - **Definition:** Energy sub-metering for kitchen appliances.
    - **Type:** Quantitative
    - **Summary Statistics (units W):**
        - Count: 178852
        - Mean: 1.205349 kW
        - Standard Deviation: 6.412824
        - Min: 0
        - 25th: 0
        - Median: 0
        - 75th percentiles: 0
        - Max : 77.000000
    - **Visualization:** Histogram showing the distribution of sub-metering 1 values.



6. **Sub-Metering 2**
    - **Definition:** Energy sub-metering for laundry appliances.
    - **Type:** Quantitative
    - **Summary Statistics (units W):**
        - Count: 178852
        - Mean: `1.829585 W`
        - Standard Deviation: `7.274617`
        - Min: 0
        - 25th: 0
        - Median: 0
        - 75th percentiles: 1
        - Max : 78.000000
    - **Visualization:** Histogram showing the distribution of sub-metering 2 values.



7. **Sub-Metering 3**
    - **Definition:** Energy sub-metering for climate control systems.
    - **Type:** Quantitative
    - **Summary Statistics (units W):**
        - Count: 178851
        - Mean: 6.611034 W
        - Standard Deviation:8.487417
        - Min: 0
        - 25th: 0
        - Median: 0
        - 75th percentiles: 17.000000
        - Max : 20.000000
    - **Visualization:** Histogram showing the distribution of sub-metering 3 values.



Data Cleaning and Preprocessing:
Rows with missing values were forward filled to ensure consistency (used previous valid row). Relevant data transformations, including resampling for time-series analysis and standardization, were performed to prepare the data for statistical analysis.

Visualization and Summary:
Various plots, including histograms, line plots, and time-series graphs, were generated to understand the distribution and trends across key variables like power consumption, voltage, and sub-metering categories. These visualizations are available in the Data Appendix for further reference.