# **Data Collection & Processing**

## **Task Summary**

Collect and structure a **time-series electricity consumption dataset** for AI-powered energy theft and leak detection .

* Dataset includes meter ID, timestamp (UTC), and cumulative energy consumption.
* Saved as a **CSV file** in /data/.

## **Expected Outputs**

* File: data/sample\_consumption\_3meters.csv
* 3 meters, 5 days, hourly readings (~217 rows)
* Timestamps in UTC, meter IDs anonymized

## **Processing Steps**

1. Validate schema: check for presence of all required fields
2. Remove duplicates and deal with missing readings.
3. Resample to hourly or daily intervals.
4. Output features: consumption delta, rolling statistics, anomaly scores.
5. Label data using known theft/leak events (if available).

## **Quality Checks**

* Completeness (>95% of expected readings present)
* Unique (meter\_id, timestamp) entries
* Valid value ranges (e.g., voltage, cumulative energy)
* Temporal consistency (cumulative readings should not decrease)

**Background & Relevance**

The proposed AI-powered energy theft and leak detection solution utilizes cutting-edge analytics, such as machine learning, to analyze electricity consumption patterns and equipment behavior in real time. According to the time-of-use consuming record of meters and sensors' inputs, the system is able to identify incredibly high peaks, tiny dips, and irregular impression of certain consumption trends that may point to theft or technical leakage. It also keeps an eye on the infrastructure for signs of oil leaks in transformers, pipelines, and other heavy-duty assets ahead of potential big damage or loss. Simultaneously addressing energy anomalies and petroleum-related risks, it’s more effective and faster than a single scanning method in the process of trouble detection; Thus, It effectively reduces economic losses for utilities, improves grid security, and ensures fair billing for customers.