Synthetic Dataset Generation

To forecast grid collapses in Nigeria's electricity grid from January 1, 2010, to May 31, 2025, we generated a synthetic dataset capturing real-world grid characteristics and historical collapse events. The dataset was produced at three temporal resolutions—6-hourly ($\sim 22\,520$ rows), daily (~ 5630 rows), and weekly (
(~ 804 rows)—to evaluate granularity for time-series forecasting. Each dataset includes 14 columns: `Timestamp`, `Voltage (V)`, `Current (A)`, `Power Consumption (kW)`, `Reactive Power (kVAR)`, `Power Factor`, `Grid Supply (kW)`, `Grid Frequency (Hz)`, `Transformer Fault`, `Overload Condition`, `Line Trip Events`, `Generation Capacity (MW)`, `Temperature (C)`, `Humidity (%)`, and `Grid Collapse Events`.

The dataset incorporates 246 grid collapses, with specific events assigned to December 11, 2024 (line tripping, vandalism), and January 11, February 12, and March 7, 2025 (line tripping, overloads). Remaining collapses were distributed randomly per year, as shown in Table 1.

Table 1: Annual Grid Collapse Counts (2010–2025)

Year	**Collapses**
2010	42
2011	19
2012	24
2013	24
2014	13
2015	10
2016	28
2017	24
2018	13
2019	10
2020	4
2021	4
2022	12
2023	12
2024	4
2025	3
Total	**246**

Key parameters were generated as follows:

- **Independent Parameters**: `Voltage (V)` ($\sim 330\,000$ V, ± 5000 V), `Power Factor` (0.8–0.95), `Grid Supply (kW)` ($\sim 4\,000\,000$ kW, $\pm 400\,000$ kW), `Grid Frequency (Hz)` (50 Hz, ± 0.2 Hz), and `Generation Capacity (MW)` (~ 4000 MW, ± 500 MW) were sampled from normal or uniform distributions.

- **Derived Parameters**:
  - `Power Consumption (kW)`: Derived as `Grid Supply` × normal deviation ($\sim 1.05 \pm 0.1$), increased by 1.1–1.3 during collapses.
  - `Current (A)`: Calculated as $I = \frac{P \cdot 1000}{V \cdot PF \cdot \sqrt{3}}$.
  - `Reactive Power (kVAR)`: Computed as $Q = P \cdot \tan(\cos^{-1}(PF))$.
  - `Overload Condition`: Set to 1 if `Power Consumption` > `Grid Supply`× 1.1.
- **Environmental Parameters**: `Temperature (C)` ($\sim 28\,^{\circ}$C wet season, $\sim 32\,^{\circ}$C dry season) and `Humidity (%)` ($\sim 70\%$ wet, $\sim 35\%$ dry) were seasonally adjusted.
- **Faults and Events**: `Transformer Fault` (2% probability) and `Line Trip Events` (5% probability) were randomly assigned, with higher likelihood (50% and 60%, respectively) during collapses.

Table 2 summarizes the dataset characteristics. The 6-hourly dataset was selected for its fine granularity, ideal for forecasting grid collapses using time-series models.

Table 2: Dataset Characteristics for Different Sampling Frequencies

Sampling	**Rows**	**Collapse Proportion (%)**
6-hourly	$\sim 22\,520$	~ 1.09
Daily	~ 5630	~ 4.37
Weekly	~ 804	~ 30.60