Data Engineering Student
Final Project

Group Project 6 - Members

Mr. Merua Goodluck Ozoemela (Leader)

Arch Taiwo Olabode Gabriel

Mr. Tayo Oyeleke

TITLE: Smart Home Energy Monitoring (IoT/Energy)

OBJECTIVE: To measure household electricity power consumption

ER diagram or Schema (if applicable)

We only had 1 table so there is no need for Entity-Relationship diagram, however, we'll explain the dataset used for this project work.

Dataset Context: Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available.

Data Set Characteristics:
Multivariate, Time-Series

Data Set Information:

This dataset contains 2075259 measurements gathered between December 2006 and November 2010 (47 months).

Attribute Information:

1. date: Date in format dd/mm/yyyy

2. time: time in format hh:mm:ss

3. global_active_power: household global minute-averaged active power (in kilowatt)

4. global_reactive_power: household global minute-averaged reactive power (in kilowatt)

5. voltage: minute-averaged voltage (in volt)

6. global_intensity: household global minute-averaged current intensity (in ampere)

7. sub_metering_1: energy sub-metering No. 1 (in watt-hour of active energy). It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas powered).

8. sub_metering_2: energy sub-metering No. 2 (in watt-hour of active energy). It corresponds to the laundry room, containing a washing-machine, a tumble-drier, a refrigerator and a light.

9. sub_metering_3: energy sub-metering No. 3 (in watt-hour of active energy). It corresponds to an electric water-heater and an air-conditioner.

Data Source: Kaggle Household Electricity Power Consumption Data

SQL queries used (with explanation)

First and foremost, we created the Database with the SQL CODE

```
CREATE DATABASE First_ETL;
```

Followed by the table creation

```
CREATE TABLE electric_consumption (
Date DATE,
Time TIME,
Global_active_power FLOAT,
Global_reactive_power FLOAT,
Voltage FLOAT,
Global_intensity FLOAT,
Sub_metering_1 FLOAT,
Sub_metering_2 FLOAT,
Sub_metering_3 FLOAT
);
```

The dataset was loaded into Postgres and view with the SELECT statement

```
SELECT * FROM electric_consumption;
```

The first query was to fetch the **" Daily average global active power"**

```
SELECT date, AVG(global_active_power) AS avg_power
FROM electric_consumption
GROUP BY date
```

```
ORDER BY date;
```

The second query was to retrieve **''the Hourly power usage trend"**

```
SELECT EXTRACT(HOUR FROM time) AS hour,
    AVG(global_active_power) AS avg_power
FROM electric_consumption
GROUP BY hour
ORDER BY hour;
```

The third query was to fetch **"Monthly power usage"**

```
SELECT EXTRACT(MONTH FROM date) AS month,
    AVG(global_active_power) AS avg_power
FROM electric_consumption
GROUP BY month
ORDER BY month;
```

The fourth query was to fetch "Sub-metering contribution"

```
SELECT date,
    SUM(sub_metering_1) AS sm1,
    SUM(sub_metering_2) AS sm2,
    SUM(sub_metering_3) AS sm3
FROM electric_consumption
GROUP BY date
ORDER BY date;
```

The fifth query was to retrieve **"Power factor (efficiency measure)"**

```
SELECT date, time,
```

```
    global_active_power /
    SQRT(POWER(global_active_power,2) + POWER(global_reactive_power,2)) AS power_factor
FROM electric_consumption
        limit 100;
```

Summary of Results and Insights

We primarily analyzed five key essential areas in this electric power consumption dataset, the key areas analyzed include the following: Daily average global active power, Hourly power usage trend, Monthly power usage, Sub-metering contribution and Power factor (efficiency measure).

i. Daily average global active power: This shows the trend of the actual electric energy consumed daily by the household and from our graph, we saw many spikes and fluctuations like a cubic graph.

Insights: The electric company needs to stabilize the global active power to avoid the frequent surge in electric power that can lead to burnout of electric gadgets.

ii. Hourly power usage trend: This shows the hourly consumption rate of electric power in the household. We saw a surge from 6am - 7am, when the various household are awake preparing for their chores quickly to meet-up their various morning engagements and rendezvous. Later we saw another surge from 5pm – 9pm finally and drops because the households are resting and sleeping and power usage drops and we see this cycle continues every day.

Insights: The Power holding company needs to make sure that power is regulated during these rush-hours.

iii. Monthly power usage: It will measure which months of the year that power is utilize more and we saw from the graph that December was the highest followed by January because many people travel down to their homes for the Yuletide and economic activities peaks at this month, it impacts greatly on power.

Insights: The Power holding company needs to be at their best during this period to regulate power utilization and make sure adequate power is supplied to dwellers and visitors during these two peak periods.

iv. Sub-metering Contribution: It measures the power reading of each meter and we know that sub-metering 1 reads for dishwasher, an oven and a microwave, sub-metering 2 reads for laundry room, washing-machine, tumble-drier, refrigerator and light and finally sub-metering 3 reads for electric water-heater and air-conditioner (which are power consuming). From the graph we saw sub-metering 3 over-shadowed others because it has the highest energy contribution followed by sub-metering 2 and lastly sub-metering 1 based on their power contribution.

Insights: Those using heavy electric gadgets should subscribe to sub-metering 3 and moderate power consumption to sub-metering 2 and less power consumption to sub-metering 1.

v. Power Factor: The power factor measures the efficiency of the power supplied and a ratio of 0.90+ is very good and 0.99 and 1 excellent, which was more of the cases. Although in the graph, we had slight drops but overall it was great like a heart pulp maintain its rhythmic pulse rate as can be seen from the graph.

Insights: The places of little drops should still be looked into so that we won't think everything is in order while is already out of control just like what a small variation can do to production plant, so they should still investigate more into it to cater for this slight changes.

Challenges encountered and how they were resolved.

The first challenge, we encountered was extracting the dataset after it had been downloaded from Kaggle because initially we thought it was a JSON document not knowing it was a Comma Separated Value (CSV) file. The dataset was a special case of CSV file because it was separated by a semi-colon instead of a comma by default.

Another challenge was loading the dataset from Python to Postgres, we suffered in this regard.

Finally, it was the visualization using Grafana was another herculean task, we discovered graph has challenge of when the data was outside the time range, at this instance Grafana breaks-down and just brings visualizations' that summarizes the dataset and only Excel can be used to salvage this case-scenario.