

DATASET SUMMARY

- Number of customers – 200,000
- Time period – 2024
- Variables included:
 - customer_id
 - disco
 - tariff_band
 - billing_month
 - kwh
 - price_ngn_kwh
 - amount_billed_ngn
 - amount_paid_ngn
 - paid_on_time
 - arrears_ngn
 - The newly created variables: payment ratio & Owes_money

TOOLS USED:

This analysis was carried out in a Jupyter Notebook environment using Python.

Libraries used are:

- Pandas
- Matplotlib/Seaborn
- Datetime function

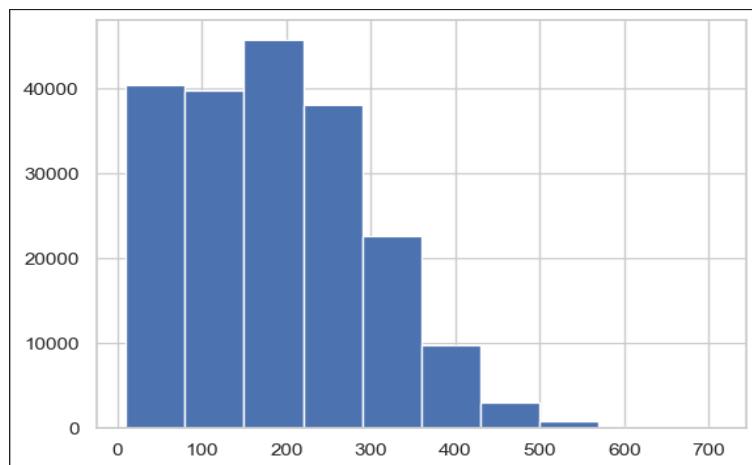
DATA CLEANING

- The data type for all columns was checked to ensure they were suitable for analysis, and all of them were correct. The billing_month was converted to a datetime to allow time-based analysis. Also, there was no invalid input in the numeric and categorical columns. Cleaned the disco column by stripping white spaces.
- There were no duplicate or missing values.

- Logical validation checks were carried out to see if there was any negative energy usage, billing, or payment; none were found. There was also no case of overpayment.
- New features were created to make the analysis easier to interpret. These include a payment ratio (showing what percentage of the bill each customer paid) and an “owes money” flag, which is a Boolean indicator showing whether a customer still has an outstanding balance. These additions help provide clearer insight into customer payment behavior.

Exploratory Data Analysis

CONSUMPTION PATTERN

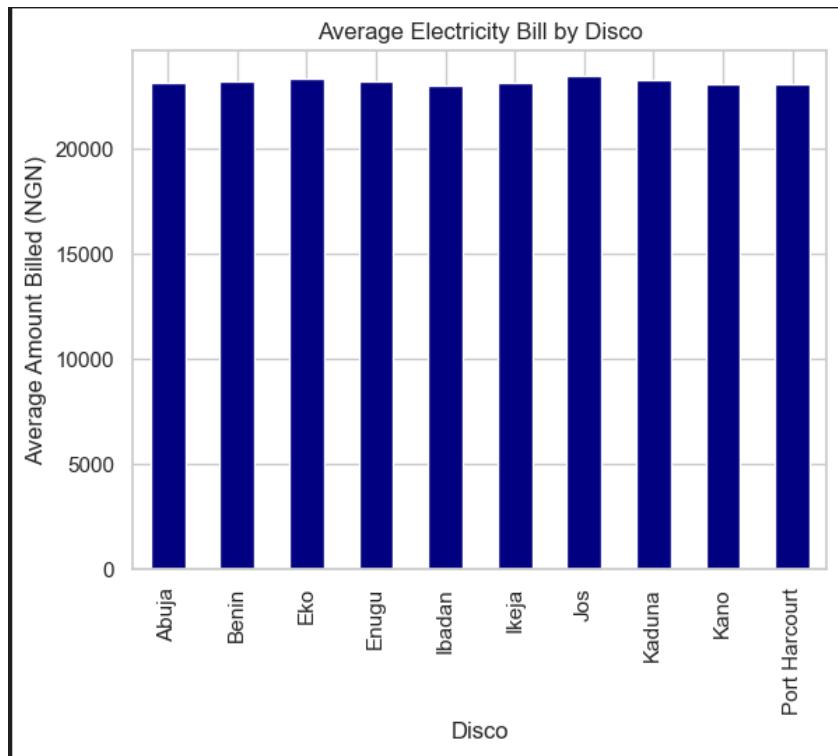


The electricity consumption, from the chart above, is right-skewed, showing that most of the customers consume moderate to low amounts of power, while a small number have high consumption.

BILLING AND PAYMENT BEHAVIOUR

As shown by the chart below, on average, billed and payment amounts are very close; there are still some variations across distribution companies.

The payment ratio shows that some Discos have better payment compliance than others.



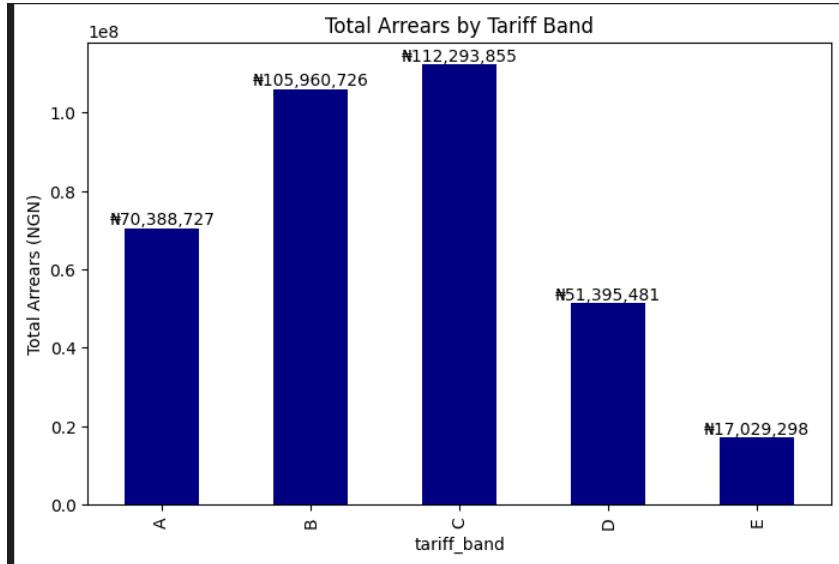
TARIFF BAND RISK ANALYSIS

Tariff Band C has the highest total arrears (≈ 112 million NGN). Followed by **B, A, D, and E**.

This could mean that there are **more people** in **Tariff Band C**, or/and they tend to use **more electricity**, thus owe more.

Tariff Band E has the lowest arrears (≈ 17 million NGN). This could mean there are **few people** in it, or/and they tend to **pay on time**.

Although **Tariff Band C** records the highest total arrears, differences in average arrears and customer payment behavior across bands suggest that risk is not driven by customer volume alone.



KEY INSIGHTS

- Electricity consumption is unevenly distributed, with a small number of high-usage customers.
- Billing amounts are relatively consistent across Discos.
- Tariff Band C accounts for the largest share of arrears, partly due to customer concentration.
- Energy usage alone does not explain outstanding arrears.
- Payment behavior varies more strongly by tariff band than by consumption level.

CODE REFERENCE

https://github.com/mayomilatasha123-debug/nigeria_energy_analysis