

## Summary of Key Findings

The given datasets, `client_data.csv` and `price_data.csv`, provide comprehensive details of various attributes. The `client_data.csv` comprises 14,606 entries with 26 columns that encapsulate information about the clients, their usage patterns, forecasted values, churn and more. On the other hand, the `price_data.csv` dataset has significantly larger entries (193,002) which primarily revolve around the variations in price over different periods.

The exploratory data analysis (EDA) has provided several key findings. The client dataset consists of a mix of numerical and categorical variables, representing various parameters such as electricity consumption, forecasted consumption, number of active products, net margin, number of years of activity, power capacity and churn status. The price dataset contains date-based price variation records for different peak times.

Overall, the churn rate is approximately 10%, and there are a few null entries in both datasets that might need attention before further data processing. The data appears to be varied and could potentially offer meaningful insights into client behavior, churn likelihood, and their relation to pricing models.

## Suggestions for Data Augmentation

The current datasets provide a wealth of information. However, additional data could help improve the quality of our analysis and predictive models:

1. **Demographic Data:** Understanding the client's demographics, such as age, income, and location, can provide deeper insights into client behavior.
2. **Time-Series Data:** More detailed time-series data can help analyze seasonal patterns, trends, and cyclic behavior.
3. **Product Data:** Detailed information about the products clients are using might reveal patterns correlating with churn rates.

In terms of **open-source datasets**, we might want to consider datasets that provide insights into the energy consumption habits of consumers. These datasets could provide a broader context for understanding our own dataset. Some potential datasets include:

**The UCI Machine Learning Repository:** Specifically, the "ElectricityLoadDiagrams" dataset could offer valuable context for the consumption details we have.

**Data.gov:** A wealth of U.S. energy-related datasets are available here, including consumer usage, costs, and more.

**The European Union Open Data Portal:** This includes data related to energy consumption and pricing throughout the EU.