Hi AD,

In order to test the hypothesis of whether churn is driven by customer price sensitivity, we would need to model churn probabilities of customers, and find the effect of prices on churn rates.

We would need the following data in order to build the models-

Customer Data- includes characteristics of each client. Eg- industry, historical electricity consumption, date joined etc.
Churn Data- indicating whether customer has churned.
Historical price data- indicating the prices the client charges to each customer for both electricity and gas.

I intend to solve the problem by the following steps:

 Step 1: Test the hypothesis-
The hypothesis can be tested by analyzing the historical data about customers' contracts.
- Collect historical data about all customers' contracts.
- Extract data on the customers who transferred to other providers
- Identify whether customers churn and an increase in price happened at the same time.
- Prove the hypothesis

If the customer churn is along with an increase in price, we can conclude the growth in prices affects customer churn. On the contrary, if customers transfer to another providers when there is no change or even a decrease in price, we can say prices do not affect customer churn.

Step 2: Build a model to predict customer churn.
It is a binary classification problem. All customers are labeled with either 1 (switch to another provider) or 0 (stay with the current provide). In order to build a powerful model to predict customer churn, I intend to complete the following tasks:

Task 1: Collect data on all customers, including

- Contract data: starting time and ending time, price, discount, promotions, etc.
- Customer information: name, country, city, profit, revenue, company size, industry, etc.
- Energy usage: energy consumption amount.

If transfer to other providers, this is the target label

Task 2: Understand and clean the data

Task 3: Exploratory data analysis
Discovering data patterns by data visualization

Task 4: Build a machine learning model and evaluate its performance.
- Train a model on the clean dataset got in task 2.

- Use cross-validation to evaluate the model performance.
-

Task 5: Identify whether a discount on price can prevent customer churn.

Apply 20% discount on the prices, then predict the customers' labels again. If the predicted labels (got in task.4) change from 1 to 0, this indicates the discount can stop customer churn. On the contrary, if the predicted labels are still 1, this indicates the discount cannot stop customer churn.

Regards