Dear Customer,

In trying to identify if the customer are likely to churn at current price, we need to estimate the probability of the customer churning and the role of the price plays in customer churning.

The data would be collected and used are:
a. The Price Data – The price data should include the price that the each customer paid for electricity and gas at some time intervals
The sample data are likely to but not subject to be:

customer_id	price	Time(Year-Month)
c1	p1	t1
c1	p2	t2
…	…	…
c2	p3	t3
…	…	…
c3	p4	t4
…	…	…

b. The Churn Data – The churn data should include the information that if the customer churned or not. The sample data might look like this:

customer_id	If churn	Time(Year-Month)
c1	No	t1
…	…	…
c2	No	t3
c3	Yes	t4
c3	No	t5
…	…	…

c. The Customer Data – The customer data should include that the characteristic information of the customer, such as the number of people in the unit, address, etc.

When we obtain the data, we firstly clean and pre-process the data for the model preparation. Then, we are going to consider to process a binary analysis through some categorical models such as logistic regression, Support Vector Machines, Random Forest, Naive Bayes, KNN, Stochastic Gradient Descent, etc. Next, evaluate those models by the accuracy of prediction, the interpretability, and flexibility. We also want to implement statistical tests to figure out the importance of the price change in deciding the churn rates.

Best,

Yunhan