Dear [LDS],

In order to test the hypothesis that the churn is driven by price sensitive, we need to build a model on prediction of the probability of customer churn, and derive the effect of prices on churn rates. We will need following data:

Dataset:

1. Customer Data - which should include characteristics of each client, for example, industry, historical electricity consumption, date joined as customer …
2. Churn data - which could contain if the client has churned
3. Historical price data - which should include the prices which the client got charged to each customer for both electricity and gas at granular time intervals

Once we have the data, we would need to engineer features based on the data that we obtain, and build a binary classification model (eg, Logistics Regression, Random Forest, Gradient Boosted Machines to name a few), picking the most appropriate model based on the tradeoff between the complexity, the explainability, and the accuracy of the models. Based on the model we picked, we would be able to understand the direction and magnitude of the impact of price on customer churn, as well as the relative importance of prices compared to other factors. Furthermore, the model could size the business impact of the clients proposed discounting strategy.

Regards,
Murong (Sophie) Cui