



Business Understanding & Hypothesis Framing

Author : **Aditya Damar Jati**

BCG Data Science & Analytics Intern

Background Problems

PowerCo is a major gas and electricity utility that supplies to corporate, SME (Small & Medium Enterprise), and residential customers. The power-liberalization of the energy market in Europe has led to significant customer churn, especially in the SME segment. They have partnered with BCG to help diagnose the source of churning SME customers.

A fair hypothesis is that price changes affect customer churn. Therefore, it is helpful to know which customers are more (or less) likely to churn at their current price, for which a good predictive model could be useful.

Moreover, for those customers that are at risk of churning, a discount might incentivize them to stay with our client. The head of the SME division is considering a 20% discount that is considered large enough to dissuade almost anyone from churning (especially those for whom price is the primary concern).

The Associate Director (AD) held an initial team meeting to discuss various hypotheses, including churn due to price sensitivity. After discussion with your team, you have been asked to go deeper on the hypothesis that the churn is driven by the customers' price sensitivities.

Your AD wants an email with your thoughts on how the team should go about testing this hypothesis.

The client plans to use the predictive model on the 1st working day of every month to indicate to which customers the 20% discount should be offered.



Task

Main task

Understand what is going on with the client and to think about how we would approach this problem and test the specific hypothesis.

Objective

1. Formulate the hypothesis as a data science problem
2. Lay out the major steps needed to test this hypothesis.
3. Communicate your thoughts and findings in an email to your AD as the solution
4. Focusing on the data that we would need from the client and the analytical models we would use to test such a hypothesis.

Hint:

Remember what the key factors are for a customer deciding to stay with or switch providers Think of data sources and fields that could be used to explore the contribution of various factors to a customer's possible action Ideally, what would a data frame of your choice look like – what should each column and row represent? What kind of exploratory analyses on the relevant fields can give more insights about the customer's churn behavior?

Estimated time for task completion: 1 hour depending on your learning style.

PowerCo Problem Context

Context

1. Client is a major gas and electricity utility (Supplies to corporate, SME and residential customers)
2. Significant churn problem (Driven by power-liberalization of the energy market in Europe and Problem is largest in the SME segment)
3. Has engaged BCG to help with the SME segment

Clients Hypothesis

1. It is possible to predict customers likely to churn using a predictive model
 2. Hypothesis that churn is driven by customer price sensitivity Client wants to try discounting strategy : SME division head suggests that offering customers at high propensity to churn a 20% discount might be effective
-

Solution

Hi [AD],

In order to test the hypothesis of whether churn is driven by the customer's price sensitivity, we would need to model churn probabilities of customers, and derive the effect of prices on churn rates. We would need the following data to be able to build the models.

Data needed:

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

Once we have the data, the work plan would be:

1. We would need to define what price sensitivity is and calculate it.
2. We would need to engineer features based on the data that we obtain, and build a binary classification model (e.g. Logistic Regression, Random Forest, Gradient Boosted Machines to name a few),
3. The best model would be picked based on the tradeoff between the complexity, the explainability, and the accuracy of the models.
4. We would subsequently dive deeper into why and how price changes impact churn.
5. Last but not least, the model would allow us to size the business impact of the client's proposed discounting strategy.

Regards,

Aditya Damar Jati