Dear Sir/Ma’am,

In order to test the hypothesis presented by PowerCo, of whether churn is driven by customers’ price sensitivity we need to model churn probabilities of customers, and derive the effect of prices on churn rate. This will be conducted as an A/B test. We need the following data to be able to build the following models.

1. Customer data – which should include characteristics of each client such as industry, size, historical electricity consumption, date joined as a customer.
2. Churn data – which should indicate if the 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, we will need to clean the data and engineer features based on the data we have obtained to build a binary classification model, (Logistic Regression, Random Forest, Gradient Boosted Machines to name a few), picking the most appropriate model based on the tradeoff between the complexity, the expandability, the accuracy of the models. Based on the model picked, we would be able to understand the direction and magnitude of the impact of prices on churn rates, as well as the relative importance of prices compared to other factors. Furthermore, the model would allow us to size the business impact of the client’s proposed discounting strategy.