Dear AD,

There are three major questions related to the customers churn problem of PowerCo. They are the followings:

1. Is there any significant relationship between price sensitivity and customer churn?
2. Will offering a 20% discount rate to customers predicted to churn significantly reduce churn rates?
3. Are there any other factors that affect the churn rates other than the price.

To gain insights of these questions, we may require a dataset that takes customers ‘id numbers as row indices and customers’ independent variables as column indices. Such independent variables include a customer’s basic information (age, gender, location, etc.), historical data in gas and electricity consumption over time, number of the customer’s complaints during a certain period, and the price plan offered to each customer (including whether the customer being given a discount). The last column of the dataset will be the dependent variable indicating whether the customer has churned or not within a specific time frame. The dependent variable should be binary, where ‘0’ stands for no churn and ‘1’ stands for churn.

After we receive such dataset, we can do formal data analysis and answer the questions. For the first question, we formulate a null hypothesis that there is no relationship between price sensitivity and customer churn. An alternative hypothesis is then there is significant relationship between price sensitivity and customer churn. In formal statistical testing, we first choose an appropriate statistical test. Since the dependent variable is binary, and other independent variables are likely to be continuous. We could use logistic regression. Let us denote the coefficient of price in our logistic regression model as β. Mathematically, the hypothesis indicated above is equivalent to a null hypothesis that β = 0 and an alternative hypothesis β ≠ 0. To verify that a certain coefficient in the logistic regression model is 0, we use t-statistics. The test statistics is given by t = β/SE(β), where SE stands for the standard error. It is then standard that we choose a significance level and calculate the p-value of our test. If the p-value is less than the significance level we set, we reject the null and we can say that we have sufficient evidence that price sensitivity has significant relationship to churn.

Similar to the analysis done above, for the second question, we could formulate a null hypothesis that offering a 20% discount rate to customers predicted to churn has no significant impact to churn rates. And an alternative hypothesis that offering a 20% discount rate to customers predicted to churn will significantly reduce the churn rates. It is worth-noting that this is a one-side test. We could also use the logistic regression model indicated above to analyse the problem. By noting that both whether a customer being offered a discount and whether a customer has churned are binary, we could also use Chi-squared test.

For the third question, we will be looking for other independent variables that has large coefficients in our logistic regression model. For example, number of complaints per month may be indictive for a customer that has churned.

Hope this helps with analysing the problem.

Best regards,

Yikang Li