To LDS@BCG,

Re: Thoughts on how to proceed to test the PowerCo Company Churn hypothesis case.

Problem Context: The power-liberalization of the energy market in Europe has led to significant customer churn, especially in the SME segment.

Hypothesis: Churn is driven by the customers' price sensitivities and possible to predict customers likely to churn using a predictive model.

Task: To deep dive on the hypothesis that the churn is driven by the customers' price sensitivities.

I went through the client's concerns and also read a few articles to know more about:

- The power-liberalization of the energy market in Europe
- Price sensitivities and its economic factors

From business perspective it seems that power-liberalization of the energy market can cause a huge effect on any "Energy" based industry. Liberalizing energy market allows both industrial and domestic consumers to freely choose their suppliers and thus introduce a highly competitive market. The overall aim of liberalization is to increase efficiency through the pressure of competition. Greater efficiency leads to lower costs and hence prices, than would otherwise have been the case. This improves industrial competitiveness which is crucial for companies competing in an increasingly global market. Thus to retain its customers, former monopolistic power company has to lower their prices and also provide a higher quality of service.

The client desires to predict the customers who are likely to churn out and offer them discounts. From data science/mathematical modelling approach the problem can be translated as:

Given customer's data and market prices as "input", predict as "output" whether a certain customer will churn or not. Further, the client wants a particular hypothesis to be tested. Following points are important to be noted considering a model solution:

1. What does "**churn**" mean? A customer who hasn't purchased gas for the last 2 weeks is not the same as someone who hasn't purchased gas since the last 4 months. For electricity it may differ drastically based on the consumer (domestic/industrial). So, we need to select a time frame after which we can conclude that the customer is lost.
2. Based on the data provided by the client, we can go for either Supervised learning or Unsupervised learning. If the customer data provided to us is of before the power-liberalization, then it won't contain data about which customers churned out and thus have to go for Unsupervised learning. But if the data does contain about customers churned or not, then it would be a binary classification problem.
3. The metric we should be focusing on to evaluate our model should be Recall (False Negatives) because the client can gladly pay 20% discount to a customer which our model predicted would

churn out but didn't. But client cannot afford to lose a customer because our model predicted the customer would not churn out but in reality, they did and thus loosing a valuable customer & huge chunk of money.

4. If we get recent data consisting of customer data who churned out would be great. This would help us go for suitable classification algorithms like logistic regression or K-Nearest Neighbors or Decision Trees or SVMs but before that we will have to apply some Undersampling and Oversampling techniques like One-Side Selection (OSS) or SMOTE + Tomek Links because the data we will be getting will be highly imbalanced.

So, once we have the access to data, we can go for suitable modelling approaches.

Yours sincerely,

Subham Surana