**Modeling Churn in Energy Company**

*This data project has been used in the recruitment process for the data science positions at BCG Gamma.*

**Assignment**

**Scenario:**

Our client, PowerCo, is a major utility company providing gas and electricity to corporate, SME, and residential customers. In recent years, post-liberalization of the energy market in Europe, PowerCo has had a growing problem with increasing customer defections above the industry average. Thus, PowerCo has asked BCG to work alongside them to identify the drivers of this problem and to devise and implement a strategy to counter it. The churn issue is most acute in the SME division and thus they want it to be the first priority.

The head of the SME division has asked whether it is possible to predict the customers who are most likely to churn so that they can trial a range of pre-emptive actions. He has a hypothesis that clients are switching to cheaper providers so the first action to be tried will be to offer customers with a high propensity of churning a 20% discount.

**Your task:**

We have scheduled a meeting in one week's time with the head of the SME division in which you will present our findings on the churn issue and your recommendations on how to address it.

You are in charge of building the model and suggesting which commercial actions should be taken as a result of the model's outcome. The client also would like to answer the following questions:

1. What are the most explicative variables for churn,
2. Is there a correlation between subscribed power and consumption,
3. Is there a link between channel sales and churn?

The first stage is to establish the viability of such a model. For training your model you are provided with a dataset that includes features of SME customers in January 2016 as well as the information about whether they have churned by March 2016. In addition to that you have received the prices from 2015 for these customers. Of particular interest for the client is how you frame the problem for training.

Given that this is the first time the client is resorting to predictive modeling, it is beneficial to leverage descriptive statistics and visualization to extract interesting insights from the provided data before diving into the model. Also, while it is not mandatory, you are encouraged to test multiple algorithms. If you do so it will be helpful to describe the tested algorithms in a simple manner.

Using the trained model, you shall “score” customers in the verification data set (provided in the eponymous file) and put them in descending order of the propensity to churn. You should also classify these customers into two classes: those that you predict to churn are to be labeled "1" and the remaining customers should be labeled "0" in the result template.

Finally, the client would like to have a view on whether the 20% discount offer to customers predicted to be churned is a good measure. Given that it is a steep discount bringing their price lower than all competitors we can assume for now that everyone who is offered will accept it. According to regulations they cannot raise the price of someone within a year if they accept the discount. Therefore, offering it excessively is going to hit revenues hard.

**Data Description**

The attached file describes all the data fields which are found in the data. You will notice that the contents of some fields are meaningless text strings. This is due to the "hashing" of text fields for data privacy. While their commercial interpretation is lost as a result of the hashing, they may still have predictive power.

**Project  Materials**

Click here to download the project description and data

See Files \ Week 2