**BCG-Data-Science-Analytics-Virtual-Experience-Program**



## Background:

Your client is PowerCo, a major gas and electricity utility company that supplies to corporate, SME (small and 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 and drive 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 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 incentivise them to stay with our client. The head of the SME division is considering a 20% discount as large enough to dissuade almost everyone from churning, especially those for whom price is the primary concern.

## Task 1: Business understanding and hypothesis testing

Your first task is to understand what is going on with the client and think about how you would approach this problem and test the specific hypothesis.

You must formulate the hypothesis as a data science problem and lay out the major steps needed to test this hypothesis, focusing on the data you would need from the client as well as the analytical models you would use to test the hypothesis.

If you are stuck:

* What are the key factors for a customer deciding to stay with or switch providers?
* Data sources and fields that could be used to explore the contribution of various factors to a customer’s potential action
* 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 into churn behaviour?

## Task 2: Exploratory data analysis

The BCG project team thinks that building a churn model to understand whether price sensitivity is the largest driver of churn has potential. The client has sent over some data which includes:

* Historical customer data: customer data such as usage, sign-up date, forecasted usage
* Historical pricing data: fixed and variable pricing data
* Churn indicator: whether or not each customer has churned For task 2, you need to:
* Perform some exploratory data analysis. Look into data types, data statistics, specific parameters, and variable distributions
* Verify the hypothesis of price sensitivity being correlated with churn
* Prepare a half-page summary of key findings and add some suggestions for data augmentation — which other data sources should the client provide you with and which open source datasets might be useful?

## Task 3: Feature engineering and modelling

The team now has a good understanding of the data and feels confident to use the data to further understand the business problem. The team now needs to brainstorm and build out features to uncover signals in the data that could inform the churn model.

Feature engineering is one of the keys to unlocking predictive insight through mathematical modelling. Based on the data that is available and was cleaned, identify what you think could be drivers of churn for our client and build those features to later use in your model.

Your colleague has done some work on engineering the features within the cleaned dataset and has calculated a feature that seems to have predictive power.

For task 3:

* Try to think of ways to improve the feature’s predictive power and elaborate on why you made those choices
* Train a random forest classifier, evaluate the results, and document the advantages and disadvantages of using a random forest for this particular use case

Bonus: how much money could the client save with the use of the model?

## Task 4: Findings and recommendations

The client wants a quick update on the progress of the project.

For task 4, develop an abstract slide synthesising all the findings from the project so far.

A few things to think about for this abstract include:

* What is the most important number or metric to share with the client?
* How much detail should you go into, especially with the technical details of your work?
* What impact would the model have on the client’s bottom line? Always test what you write with the “so what” test