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

1. Business Understanding and Hypothesis Framing
   * Background: PowerCo, an electricity and gas utility has faced a significant customer churn in Europe especially for SME (Small & Median Enterprise)
   * Hypothesis: price chance affects the customer churn
   * 20% discount for the customers that are at risk of churning
2. Exploratory Data Analysis
   * Dataset understanding
     + Imbalance data: 9.7% clients has churned
     + Churn customers
       - Related to 5 specific channels including MISSING values which has a churn rate of 7.6%
       - Related to small number of products (≤5)
       - Related to the number of years they involved in (with 2y most followed by 3 and 4y, 5y has a churn rate 3% lower than 4 -> if customers stay more than 4y, less likely to churn)
       - Related to 4 specific original service code including MISSING
       - Related to if consume gas and electricity together (who consume both was less likely to churn compared to consuming electricity only 8.2% < 10.1%)
     + Consumption, forecasting, subscribed power data is highly right skewed
     + Outliers: margin, consumption
   * Verify the hypothesis of price sensitivity being to some extent correlated with churn
     + Define price sensitivity
       - Mean average price of last year/ 6m/ 3m
     + Finding: price sensitivity has low correlation with churn -> need to increase the prediction power
3. Feature Engineering and Modelling
   * Feature Engineering
     + Improve prediction power of price sensitivity
       - Difference between off-peak price in 12m
       - Average price across period (off-peak, peak, mid-peak)
       - Max price changes across periods and months
     + Others features
       - Tenure
       - Months since activation/to end/since modification/ since renewal
       - Remove small number of *channel\_sales*, *origin\_up*
       - Transforming skewed data (log +1)
     + Correlations
       - Removed similar variables
   * Modelling
     + Random Forest (ensemble algorithms)
     + Redefine threshold, improve precision from 79.31% to 94.44%
     + Evaluation
       - margin and consumption over 12m are top 2 important variables
       - forecasting on next 12m performance are also important drivers
       - time is also sensitive
       - the price factors are scatter around but not the most important factors
   * Discount Impacts
     + 20% discount to high propensity to churn customers (assume all will accept)
     + Based on the forecast revenue for customers
       - 1. No discount offered
         * *Basecase\_revenue* for next 12m based on forecast consumption and price
         * If churning: *basecase\_revenue\_after\_churn = basecase\_revenue\*(1-0.9195\*churn)*

0.9195 came from the average churn date after the first data calculation

* + - * 2. Discount is offered based on the probability cutoff who should receive 20% off
        + Choose the cutoff

Receive discount: 0.8\* *basecase\_revenue*

Counterbalancing: choosing from TP and FP

TP: revenue retention vs. no-discount scenario(churn)

0.8\* *revenue* vs *0*

FP: reduced revenue from giving discount while they would not churn in fact (not actual cost difference)

*0.2\*revenue*

Optimal

Considering do not offer a discount to the customers which are not very profitable -> retaining high value customers

Results: even lower than giving discount to everyone

Using forecast rather than actual churn

Actual outcomes: “backtesting”

Replace the churn outcomes with the predicted probability of churn: “model-dependent”

* + - * + Reduced lost profits by 9.48%

1. Findings and Recommendations
   * Churn is indeed high in the SME division
     + 9.7% across 14606 customers
   * Predictive model is able to predict churn but the main driver is not customer price sensitivity
     + Yearly consumption, forecasted consumption and net margin are the 3 largest drivers
   * Discount strategy of 20% is effective but only if targeted appropriately
     + Offer discount to only to high-value customers with high churn probability

Developed a customer churn and price volatility prediction model for Powerco Energy via Python

Verified the hypothesis of price sensitivity is correlated with churn and increased the prediction power by adding price-related variables

Performed variable selection and predict churn using random forest, improved precision from 79.31% to 94.44%

Segregated high propensity customers with high churn probability and offered 20% discount, reducing lost profits by 9.48%