

Task 1: **Business Understanding & Hypothesis Framing**

Task Overview

What you'll learn

- Meet your client **PowerCo** - a major gas and electricity utility who is concerned about losing customers
- How to interpret the business context
- How to break down the problem before you start your data analysis

What you'll do

- Determine the client data needed for analysis
- Outline the techniques you'll use to investigate your client's problem
- Write an email to your Associate Director summarizing your approach

Key roles and responsibilities of a Data Scientist at BCG X

BCG X is transforming businesses using data science to help companies generate competitive advantage. To do this, we typically follow a 5-step methodology:

1. **Business understanding & problem framing:** what is the context of this problem and why are they trying to solve it?
2. **Exploratory data analysis & data cleaning:** what data are we working with, what does it look like and how can we make it better?

3. **Feature engineering:** can we enrich this dataset using our own expertise or third party information?
4. **Modeling and evaluation:** can we use this dataset to accurately make predictions? If so, are they reliable?
5. **Insights & Recommendations:** how we can communicate the value of these predictions by explaining them in a way that matters to the business?

The tasks in this program will be focused on using different parts of this methodology at different times, so you'll get a taste of the overall process.

It's a really exciting time to be working with BCG X as more clients are needing data to drive key decisions. So, let's check out what case you'll be working on!

The brief from PowerCo

The Associate Director (AD) of the Data Science team held a team meeting to discuss the client brief. You'll be working closely with Estelle Altazin, a senior data scientist on your team.

Here are the key takeaways from the meeting:

- Your client is **PowerCo** - a major gas and electricity utility that supplies to small and medium sized enterprises.
- The energy market has had a lot of change in recent years and there are more options than ever for customers to choose from.
- PowerCo are concerned about their customers leaving for better offers from other energy providers. When a customer leaves to use another service provider, this is called **churn**.
- This is becoming a big issue for PowerCo and they have engaged BCG to help diagnose the reason why their customers are churning.

During the meeting your AD discussed some potential reasons for this churn, one being

how “sensitive” the price is. In other words, how much is price a factor in a customer’s choice to stay with or leave PowerCo?

So, now it’s time for you to investigate this hypothesis.

Your task - we need to understand PowerCo’s problem in detail

First things first, you and Estelle need to understand the problem that PowerCo is facing at a deeper level and plan how you’ll tackle it. If you recall the 5 steps in the Data Science methodology, this is called “business understanding & problem framing”.

Your AD wants you and Estelle to email him by COB today outlining:

1. the data that we’ll need from the client, and
2. the techniques we’ll use to investigate the issue.

Use the text field below to write your email, here’s what you’ll need to include:

You must formulate PowerCo’s issue as a problem using the 5 step data science process and lay out the major steps needed to test it.

1. What do you think are the key reasons for a customer deciding to stay with or switch energy providers? For example: price, is it clean energy, customer service, location etc.
2. What data do you think would be useful in order to investigate these key reasons? E.g. customer purchasing trends over past 5 years, location of business etc.
3. If you were to get this data, how could you analyse or visualize it to test whether these reasons may have an impact on churn?

Task:-

Subject: Plan for Investigating PowerCo's Churn Issue

Dear Associate Director,

Estelle and I have discussed PowerCo's churn problem and would like to outline our approach to investigating this issue, following the 5-step methodology:-

1. Business Understanding & Problem Framing:

- PowerCo is concerned about increasing customer churn, particularly in light of competitive offers from other energy providers. The key hypothesis to test is whether price sensitivity is a major driver of this churn, though other factors may also play a role, such as customer service, contract length, energy source (e.g., clean energy options), and location.

2. Data Needed from PowerCo:

To thoroughly investigate the factors influencing churn, we will need access to several datasets:

- **i) Customer transaction history:** Including consumption patterns, pricing plans, and contract lengths over the past 5 years.
- **ii) Customer demographics and firmographics:** Business type, size, location, and energy preferences.
- **iii) Customer support interactions:** Details of any complaints, service outages, or other issues that could affect customer satisfaction.
- **iv) Churn history:** A list of customers who have left and the timeframes in which they churned.
- **v) Market data:** Competitor pricing, offerings, and any industry trends around energy usage and renewables.
- **vi) Customer feedback or survey data:** Insights into customer satisfaction with PowerCo's services, including their reasons for switching or staying.

3. Exploratory Data Analysis & Data Cleaning:

We'll begin by cleaning and exploring the data to identify patterns and anomalies. Initial steps will include:

- i) Checking for missing or inconsistent data (e.g., incomplete transaction history or incorrect churn dates).
- ii) Visualizing overall churn trends, such as how they vary across price tiers, contract types, or business sizes.
- iii) Conducting correlation analysis to identify relationships between churn and various factors (price, service issues, etc.).

4. Feature Engineering

Once we have clean data, we'll focus on creating features that capture the key reasons customers churn:

- **i) Price sensitivity:** Calculate the average price paid per unit of energy and price fluctuations over time.
- **ii) Contractual details:** Length of the contract, automatic renewal options, and early termination fees.
- **iii) Customer lifetime value (CLV):** To assess the impact of higher-value customers leaving.
- **Interaction metrics:** Frequency of customer service interactions and the nature of issues raised.

We may also enrich the data with third-party information, such as competitor pricing or market trends in the energy industry.

5. Modeling & Evaluation

We'll utilize machine learning models such as logistic regression or random forests to identify which factors most contribute to customer churn. We will evaluate model performance using key metrics like accuracy, precision, and recall to ensure the results are reliable.

6. Insights & Recommendations

Once the model is built, we'll interpret the results to generate actionable insights. For example, if price is the main factor, we can suggest targeted retention strategies for price-sensitive customers. If customer service issues are more prominent, we might recommend improvements in the support process.

Please let us know if you'd like to add anything or adjust this approach.

Best regards,
Sajan Singh Shergill

Example Answer

Great work! Take a look at the example answer below to see how a professional would have attempted this task. Think about what you did well and how you can improve.

Hi [AD],

In order to test the hypothesis of whether churn is driven by the customers' price sensitivity, we would need to model churn probabilities of customers, and derive the effect of prices on churn rates.

We would need the following data to be able to build the models.

1. Customer data - which should include characteristics of each client, for example, industry, historical electricity consumption, date joined as customer etc.
2. Churn data - which should indicate if customer has churned
3. Historical price data – which should indicate the prices the client charges to each customer for both electricity and gas at granular time intervals

Once we have the data, the work plan would be:

1. We need to define what price sensitivity is and calculate it
2. We need to prepare the data and engineer features
3. Then, we can test our hypothesis using a binary classification model (e.g. Logistic Regression, Random Forest, Gradient Boosted Machines to name a few)
4. We would choose a model from one of the tested algorithms based on the model complexity, the explainability, and the accuracy of the models.
5. With the trained model, we would be able to extrapolate the extent to which price sensitivity influences churn

Regards, [Your name]