BCG x Data Scientist Intern

Client: PowrCo Utility Industry, Specifically gas, and Electricity.

Problem: (Customer Churn)Powerco's client is switching to another Service Provider.

TASK: 1 - **Business Understanding & Hypothesis Framing**

Task Overview

What you'll learn

- Meet your client **PowerCo** - a major gas and electricity utility that 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

Before we begin

Here are some key facts about this program:

- You are working as a Junior Data Scientist at BCG GAMMA
- You work within a larger team, where each member has a different role and level of responsibility.
- Your team has been assigned a new project for a client called **Powerco**
- As a Junior Data Scientist, this is an opportunity to get valuable hands-on experience. You'll work on your programming, problem-solving, and communication skills - all essential skills for life on the job as a data scientist!

What is a Data Scientist?

What exactly is a Data Scientist?

- A Data Scientist works with data to try and deliver value to a business.
- But how is that achieved?
  - Data Science can sometimes overlap with domains such as Data Analysts, Data Engineers, and DevOps.

- How much the Data Scientists role overlaps with these other roles depends on the company, but in essence, a Data Scientist is generally focused on modeling data to be able to accurately predict an outcome, for example, predicting how likely customers are to leave.
- 4 core skills are used by a Data Scientist: statistics, mathematics, programming, and communication.

Q. What are the 4 core skills of a Data Scientist?

Ans: Statistics, mathematics, programming, communication

Key roles and responsibilities of a Data Scientist at BCG GAMMA

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

Business understanding & problem framing: what is the context of this problem and why are they trying to solve it?
Exploratory data analysis & data cleaning: what data are we working with, what does it look like and how can we make it better.
Feature engineering: can we enrich this dataset using our own expertise or third party information?
Modeling and evaluation: can we use this dataset to accurately make predictions? If so, are they reliable?
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 GAMMA 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 is 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?

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?

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]

TASK: 2 - **Exploratory Data Analysis**

What you'll learn

- How to investigate whether price sensitivity is the most influential factor for a customer churning
- How to use frameworks to conduct exploratory data analysis

What you'll do

- Use Python to analyze client data
- Create data visualizations to help you interpret key trends

Your AD is giving you more responsibility!

Well done for your initial understanding of the case with PowerCo. After reviewing your project plan, the AD would like you lead on the Data Science deliverables for the rest of the project.

The AD would like you to investigate whether price sensitivity is the most influential factor for a customer churning, and if not, to what extent does price sensitivity influence churn.

Before we begin on this task, what exactly is price sensitivity?

What is price sensitivity?

Price sensitivity is **the degree to which demand changes when the cost of a product or service changes.**

In the context of PowerCo, the “demand” refers to the demand for energy consumption.

Price sensitivity is commonly measured using the price elasticity of demand, which states that some consumers won't pay more if a lower-priced option is available.

What is the price elasticity of demand?

Price elasticity of demand is a measurement of the change in consumption of a product in relation to a change in its price

Complete the quick knowledge check and move on to your exploratory data analysis.

Q. Which of the following statements best describes price sensitivity?

Ans: Price sensitivity is the degree to which demand changes when the cost of a product or service changes

Q. Which of the following statements best describes the price elasticity of demand?

Ans: A measurement of the change in consumption of a product in relation to a change in its price

Exploratory data analysis

The client has sent over 2 datasets and it is your responsibility to perform some exploratory data analysis.

What is exploratory data analysis?

Exploratory data analysis (EDA) is a technique used by a Data Scientist to gain a holistic understanding of the data that they are working with.

It is mainly based on using statistical techniques (such as descriptive statistics) and visualizations to gain a deeper understanding of the statistical properties that the data holds.

Complete the quick knowledge check on the next step and let's get started.

Q. Which of the following is typically not a technique that a Data Scientist would use during exploratory data analysis?

Ans: Deep learning.

Let's get familiar with the data

As a Data Scientist at BCG, there will be occasions when you need to analyse data or investigate an issue and you are not provided strict instructions or guidance. You may be thinking, where do I start?

It is a highly valuable skill to begin to learn how to investigate a problem independently. A great way to learn this skill is to build a framework for analysis that works for you.

In this step, you'll need to analyse client data sets using Python and upload your work as a Jupyter notebook. We'll show you an example answer on the next step, but we encourage you to give it a go first!

The client has sent over 3 data sets (shown below):

1. Historical customer data: Customer data such as usage, sign up date, forecasted usage etc
2. Historical pricing data: variable and fixed pricing data etc
3. Churn indicator: whether each customer has churned or not

You need to analyze the following using Python:

- The data types of each column
- Descriptive statistics of the dataset
- Distributions of columns

Estelle has provided a starter Jupyter notebook has been provided for you to use as a template to complete your work.

Here are some tips to help you:

Let's take a look at the 3 data sets PowerCo. has sent over:

- A first good step is to review the data to make sense of the columns...
- **Hint:** Look *at data types of column to gain a better understanding of what the columns mean. This is why data description documents are important - they describe exactly what the columns represent.*
  Once you understand the columns in the dataset. Now you want to look at how the values in the data vary...
- ***Hint:*** *this is why reporting descriptive statistics is useful because it'll tell you some basic statistical properties of the columns in the data. It will also tell you how many values feature within a column, e.g. does a column only have 1 unique value or 100? This is useful to know because you can then start to build a picture of what this data represents.*
  You now understand how the values vary and what the data represents - next up, it can be useful to visualize some of this...
  - ***Hint:*** *Not all visualizations are useful. Keep the visualizations simple and always keep in mind what you're trying to show. E.g. if you want to see how the distribution of a*

column looks and that column has 1000 unique values, using a pie chart would not be good because it would become too crowded! If the values are numeric, a distribution plot would be more appropriate.

Hint: *make sure to use the starter Jupyter notebook provided, as this will show you some example visualizations and sample code to use!*
At this stage, you should now have a clearer understanding of what the data is and how it looks. This framework is not exhaustive, but it shows how you could start to build your own framework for analysing data.

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.

We've also provided an explanation for you to read through as you review the Jupyter Notebook.

Explanation

Getting set up - This task is focused on exploratory data analysis of the client and price data provided:

- The first thing you should do is download the provided Jupyter notebook and the CSV datasets.
- To run the notebook, you need to make sure that you provide the path for the CSV files so that you can load the data.
- By running the cells that exist within the notebook from Estelle, this will show you what the two datasets look like, it will provide you with code to produce descriptive statistics and it will also give some examples and sample code on how to visualize the data.

Analysis - Once you've run the cells provided, it was your job to build on this exploratory analysis:

- The visualization provided by Estelle shows how many companies churned vs. how many companies did not churn. We can see from this that the churn rate is approximately 10%. This is actually a very good churn rate, the closer the rate is to 0%, the better.
- The next series of visualizations were created in an attempt to try and dive deeper into how churn changes based on other factors (using other columns). This is useful for us to investigate because it may help us to understand factors that drive churn.
- In the notebook we visualize churn vs. sales channel, contract type, number of products, number of years and origin/contract offer.
- For example:

    - We see that for sales channel, there are some sales channels that yield customers churning but there are also other sales channels that have no customers churning.
    - For contract type, we see quite an even split for customers churning. This is interesting because this may suggest that contract type is not a driving factor towards churn rate.
- Additionally, for some columns their distributions with churn rate included. This is useful for us to understand because based on the distribution of a column, this could affect our feature engineering later.
- We look at the distribution of consumption, subscribed power and forecast in the notebook.
- For example:
    - We notice that the distribution of consumption is very skewed, this is called a positive skew since it is biased towards lower values on the x axis.
    - This is interesting because you may decide to treat this column to reduce the skewness later on during feature engineering. But also because we may want to visualize if there are any outliers within this column.
    - To investigate outliers, we use a boxplot. From the boxplot we can see that with the column as it is there are definitely some outliers. Once again this is interesting because we may choose to remove some of these outliers later.

TASK: 3 - **Feature Engineering & Modelling**

Task Overview

What you'll learn

- How feature engineering can be used to test hypotheses
- How to build features to analyse the data for PowerCo

What you'll do

- Use Python to build a new feature for your analysis

Now it’s time for feature engineering

Well done for your analysis on the influence of price sensitivity relative to churn!

Estelle reviewed your work with the AD and Estelle has come up with an idea to enrich the dataset when trying to predict churn:

- *“I think that the **difference between off-peak prices in December and January the preceding year** could be a significant feature when predicting churn”*

As the Data Scientist on the team, you need to investigate this question. So, in this task you’ll be responsible for completing feature eingeering for the dataset.

Before we start on this task, let's explain what feature engineering is...

What is feature engineering?

Feature engineering refers to:

- **Addition**
- **Deletion**
- **Combination**
- **Mutation**

of your data set to improve machine learning model training, leading to better performance and greater accuracy.

In context of this task, feature engineering refers to the engineering of the price and client data to create new columns that will help us to predict churn more accurately.

Effective feature engineering is based on sound knowledge of the business problem and the available data sources.

Q. Which of the following techniques would not fall under feature engineering?

Ans: Inspection

Creating the new features

Estelle has done some further cleaning of the data and provided you with a new CSV file to complete our work from named "clean_data_after_eda.csv". Be sure to use this data for your work on this task.

To help you understand her idea about the new feature to create, Estelle has also provided a starter notebook which will help you to create the feature she described. This notebook is called "feature_engineering.ipynb". Use this as a template to start your feature engineering.

We'll show you some tips on the next step before you start your task, but make sure you have the relevant files downloaded before moving on!

- Download the new CSV data and Jupyter notebook
- Run the cells in the Jupyter notebook to create Estelle's suggested feature
- We'll continue working in this Jupyter notebook to create some more columns, it's time to get creative!

Here’s what you need to think about before you submit your work

Your task is to create new features for your analysis and upload your completed python file. We'll show you an example answer on the next step, but we encourage you to give it a go first! Below are some tips on how to get started.

As before, a good way to quickly learn how to effectively feature engineer is to build a framework to follow. Below is an example of how you could attempt this task:

First - can we remove any of the columns in the datasets?

- There will almost always be columns in a dataset that can be removed, perhaps because they are not relevant to the analysis, or they only have 1 unique value.

Second - can we expand the datasets and use existing columns to create new features?

- For example, if you have “date” columns, in their raw form they are not so useful. But if you were to extract month, day of month, day of year and year into individual columns, these could be more useful.

Third - can we combine some columns together to create “better” columns?

- How do we *define* a “better” column and how do we *know which* columns to combine?
  - We’re trying to accurately predict churn - so a “better” column could be a column that improves the accuracy of the model.
  - And which columns to combine? This can sometimes be a matter of experimenting until you find something useful, or you may notice that 2 columns share very similar information so you want to combine them.

Finally - can we combine these datasets and if so, how?

- To combine datasets, you need a column that features in both datasets that share the same values to join them on.

At this stage, your data could look vastly different, or may have just some subtle differences to how it was before.

You will be done with this task when you’re happy with the new set of features that you’ve created and you think you’re ready to build a predictive model to see which of these features are useful for predicting churn. Upload your python file and move onto the example answer.

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.

Explanation:

Set up:

- This task is focused on feature engineering, Estelle has provided a CSV dataset for you to use as a base for this task.
- You should download the notebook and CSV and start by running the cells within the notebook.
- Estelle has also provided some insight into a feature that would be interesting to add to the data. By running the cells in the notebook, this will create the feature that she described for you and provide a foundation for you to begin your own feature engineering.

Here is some context around the additional features that have been engineered in the notebook, to help you in the future:

- Firstly we have the average price changes across periods. This is a measure of the average price change by company between peak, mid-peak and off peak periods.
- We then take this idea one step further by creating another similar feature but instead of looking at the average price difference, we look at the maximum price difference across periods and months. This gives another way to look at the price changes across months.
- The reason why these 2 features could be useful is because they are another way of representing the variance of prices throughout the year. Imagine, if your utilities bill massively increased over winter, as a consumer you'd be annoyed and want to find a better deal!
- After this we continue feature engineering with some more concepts, including transformation of columns.
- To make predictions with a statistical or machine learning algorithm, all of the data must be converted to numeric data types.
- Therefore, we convert date into months and remove the raw date column, as we cannot use it in its original form.
- We also convert boolean columns into binary values.
- And we convert categorical columns into dummy variables. A dummy variable is a binary flag that indicates when a row matches the value from the categorical column that it was created from.
- As we saw during exploratory data analysis, the distribution of some columns was skewed. This is important to identify because when modeling data for prediction, based on the technique or algorithm that we use, there are sometimes assumptions within the data that we should follow.

- One common assumption is that the columns within the data are normally distributed. Hence, if we find that columns are not normally distributed, we should treat these columns to try and transform them into a distribution that is more normal.

- Therefore, the next thing we do is transform some columns to have a closer to normal distribution. We do this using the logarithm function. As you can see from the visualisations, the newly transformed columns are much closer to a normal distribution than what they were earlier.
- Finally, we plot correlations of all the columns to see if we can identify any columns to remove. Columns that have very high correlations indicate an area to look out for. In this case, you may want to remove one of the columns, since they are likely both holding very similar information.

TASK: 2 - Findings & Recommendations

Task Overview

What you'll learn

- How predictive modelling can be used to indicate churn risk
- How to communicate your insights with clients

What you'll do

- Build a predictive model for churn using a random forest technique
- Write an executive summary with your findings

We're now ready to begin predicting churn!

Now that you have a dataset of cleaned and engineered features, it is time to build a predictive model to see how well these features are able to predict a customer churning.

Estelle has informed you that a classification model would be best for this task, and has suggested that you try the Random Forest classifier.

What is classification?

When you are trying to predict an outcome, the result that you're trying to predict can either be:

- A continuous number, e.g. an employees salary
- Or a discrete value, e.g. a job title

In our example, we are trying to predict whether or not a client will churn, so it will only ever been 1 of 2 values (True/False, 1/0, etc...).

If the outcome that you're trying to predict has a fixed number of discrete values, this is a classification problem, as you are trying to "classify" the observations in the data. If the outcome is a continuous number, this is a regression problem. We will not cover regression problems in this task.

And how does a Random Forest work?

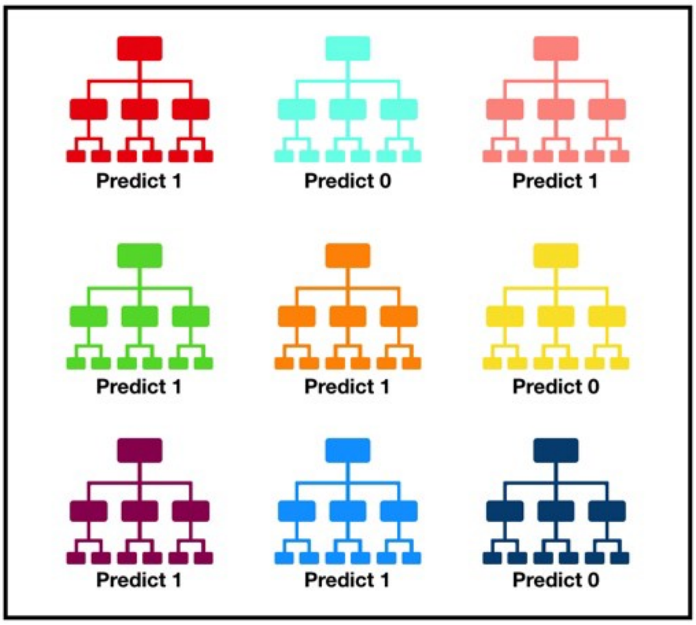
A random forest is a supervised learning algorithm which means that you must provide the algorithm with a set of features, as well as the outcome that you're trying to predict, in our case churn.

The way it makes predictions is by building a set of decision trees on different samples of the data and by taking a majority vote to decide what prediction to make.

To visualize this, the image below shows 9 decision trees and they are all trying to predict an outcome which is either a 1 or a 0 (similar to our case, where if someone has churned you see a 1, and if they haven't you see a 0).

The random forest would look at all the predictions generated from the 9 trees. You can see that 6 trees have predicted 1 and 3 have predicted 0. Therefore, the random forest would take the majority vote and present it's prediction as equal to 1.

If you wish to learn more on how this algorithm works, you can read more here.



Tally: Six 1s and Three 0s
Prediction: 1

Q. Is our task of predicting churn a classification or a regression problem?

Ans: Classification

Q. How does a random forest classifier make its predictions?

Ans: Takes the majority vote of its decision trees

Outline for making your predictions

It is your task to:

- Train a random forest classifier to predict churn
- Evaluate the predictions using evaluation metrics to demonstrate how accurately the model has performed

Estelle has provided a Jupyter notebook to get started. You should use this as a template to complete the work for this task.

Furthermore after the previous task of feature engineering, Estelle conducted a further review and has provided you with a final dataset to use this for this task, named "data_for_predictions.csv". Be sure to use this dataset for this task.

You will notice that within the notebook, Estelle has imported various packages that would be used. One of them is named "scikit-learn". This is an open source machine learning package and will be the source of the random forest model, as well as other things, that we use.

For more information on how to use the Random Forest classifier in scikit-learn, see the documentation site here

The outputs of your work will be shared with the AD and Estelle has given you a few points to include within the notebook:

- Why did you choose the evaluation metrics that you used? Please elaborate on your choices.
- Do you think that the model performance is satisfactory? Give justification for your answer.
- Make sure that your work is presented clearly with comments and explanations

When you are happy with your notebook, you should submit the notebook below. We'll show you an example answer on the next step!

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.

Explanation:

This final task is focused on building the predictive model using the CSV file that Estelle has shared.

- This CSV file contains a set of cleaned and engineered features so that you can focus purely on training your predictive model.
- You should download the Jupyter notebook and CSV file and run the cells provided in the notebook.
- These cells will load the data and create train and test samples of the data.
- It is important to split your data into train and test samples so then you can measure how well the trained model performs on an unseen set of data.
- This is a massively important thing to do when building a predictive model, otherwise you will have no way of measuring how well your model is able to predict churn for new customers!
- The code in the notebook provides you with skeleton code to create the random forest classifier, but it is your job to fill in the details of the code by using the documentation site provided.
- By adding in values for parameters within the random forest and by fitting the model on the training data, you will have a trained model to predict churn!

Now the most important part, evaluation of the model:

- It is left for you to decide how to evaluate the performance of the model. In general, you want to use metrics that reflect honestly how well the model has performed.
- In the notebook we use 3 metrics, accuracy, precision and recall.
- The reason why we are using these three metrics is because a simple accuracy measure (what percentage did I predict correctly) is not always a good measure to use.
- To give an example, let's say you're predicting heart failures with patients in a hospital and there were 100 patients out of 1000 that did have a heart failure.
- If you predicted 80 out of 100 (80%) of the patients that did have a heart failure correctly, you might think that you've done well! However, this also means that you predicted 20 wrong and what may the implications of predicting these remaining 20 patients wrong? Maybe they miss out on getting vital treatment to save their lives.
- As well as this, what about the impact of predicting negative cases as positive (people not having heart failure being predicted that they did), maybe a high number of false positives means that resources get used up on the wrong people and a lot of time is wasted when they could have been helping the real heart failure sufferers.
- This is just an example, but it illustrates why other performance metrics are necessary such as precision and recall, which are good measures to use in a classification scenario like this.
- After calculating the 3 metrics, we can see that we're able to accurately identify clients that do not churn, but not so accurately identify clients that will churn. Our model is predicting a high percentage of clients to not churn, when in fact they did!
- This tells me that the current set of columns are not a good set of features to predict churn. As the data scientist, it would normally be my job to go back and try to engineer a

set of features that is able to predict churn more accurately.

Finally, we produce a feature importance chart to visualise which features were indeed useful within the model and which ones weren't.

- We can see that net margin and consumption over 12 months were important, to name a few.
- However the price sensitivity features are scattered around and do not shine through as a main driver for churn in their current form.

Finally, let's create a quick summary for the client

Before we finish up, the client wants a quick update on the project progress. Your AD wants you to draft an abstract (executive summary) of your findings so far.

Here is your task:

Develop an abstract slide synthesizing all the findings from the project so far, keeping in mind that this will be for the key stakeholders meeting which the Head of the SME division, as well as other various stakeholders, will be attending.

Note: a steering committee meeting is a meeting where the BCG team presents key findings and recommendations (and/or project progress) to key client stakeholders.

Please use the template below and submit your summary slide in PDF format. We'll show you an example answer on the next step

A few things to think about for this abstract include:

- What is the most important number or metric to share with the client?
- What impact would the model have on the client's bottom line?

Please note, there are multiple ways to approach the task and that the sample answer is just one way to do it.

If you are stuck:

- What do you think the client wants to hear? How much detail should you go into, especially with the technical details of your work?
- Always test what you write with the "so what?" test, i.e. sharing a fact, even an interesting one, only matters if the client can actually do something useful with it. E.g. 60% of your customers are from City A is pointless, but customers in City A should be prioritized for

giving discount as they are among your most valuable ones, if true, is an actionable finding.

Resume Snippet

Take your updated resume and apply for Jobs , Programs , Events and Talent Networks

BCG Data Science Job Simulation on Forage - March 2024

- Completed a customer churn analysis simulation for XYZ Analytics, demonstrating advanced data analytics skills, identifying essential client data and outlining a strategic investigation approach.
- Conducted efficient data analysis using Python, including Pandas and NumPy. Employed data visualization techniques for insightful trend interpretation.
- Completed the engineering and optimization of a random forest model, achieving an 85% accuracy rate in predicting customer churn.
- Completed a concise executive summary for the Associate Director, delivering actionable insights for informed decision-making based on the analysis.

Use our resume guidance here

Interview Tip

In a typical interview you'll be asked "why are you interested in this role?" or "why are you interested in working at our company?". Use this interview tip to explain why you want the job.

"Why are you interested in this role?"

I recently participated in BCG's Data Science job simulation on the Forage platform, and it was incredibly useful to understand what it might be like to participate on a data science and analytics team at BCG.

I worked on a project to create a customer churn analysis simulation using Python. This project built my data science skills in a real-world context, as well as my presentation skills through creating an executive summary of my findings for the team.

Doing this program confirmed that I really enjoy working on data science projects and I'm excited to apply these skills on a data science team at a company like BCG.