Task 1

Problem Statement

Here is the background information on your task

PowerCo is a major gas and electricity utility that supplies to corporate, SME (Small & 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 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 (or less) 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 incentivize them to stay with our client. The head of the SME division is considering a 20% discount that is considered large enough to dissuade almost anyone from churning (especially those for whom price is the primary concern).

The Associate Director (AD) held an initial team meeting to discuss various hypotheses, including churn due to price sensitivity. After discussion with your team, you have been asked to go deeper on the hypothesis that the churn is driven by the customers' price sensitivities.

Your AD wants an email with your thoughts on how the team should go about testing this hypothesis.

The client plans to use the predictive model on the 1st working day of every month to indicate to which customers the 20% discount should be offered.

Here is your task

Your first task today is to understand what is going on with the client and to 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. Communicate your thoughts and findings in an email to your AD, focusing on the data that you would need from the client and the analytical models you would use to test such a hypothesis.

We would suggest spending no more than one hour on this task.

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

If you are stuck:

1. Remember what the key factors are for a customer deciding to stay with or switch providers
2. Think of data sources and fields that could be used to explore the contribution of various factors to a customer's possible action

3. Ideally, what would a data frame of your choice look like – what should each column and row represent?
4. What kind of exploratory analyses on the relevant fields can give more insights about the customer's churn behavior?

Estimated time for task completion: 1 hour depending on your learning style.

Answer

Dear Associate Director,

I have thought about the problem at hand and how we can go about testing the hypothesis that churn is driven by customers' price sensitivities.

To test this hypothesis, the first step would be to gather data on the customers that have churned, as well as a representative sample of customers that have not churned. This data should include information on the customers' price plan, as well as any other relevant factors that may influence churns, such as their contract length, usage levels, and satisfaction with the service.

Once we have this data, we can use a variety of analytical techniques to explore the relationship between price sensitivity and churn. One approach could be to use logistic regression to model the likelihood of churn as a function of the customer's price plan and other relevant factors. This would allow us to determine whether customers on higher-priced plans are more likely to churn and whether this relationship is statistically significant.

Alternatively, we could use a machine learning algorithm, such as a decision tree or random forest, to identify the most important drivers of churn. This would involve the following steps:

1. Split the data into training and test sets.
2. Pre-process the data, such as by handling missing values and scaling numeric features.
3. Train the machine learning model on the training data.
4. Evaluate the model's performance on the test data.
5. Use the model to predict churn for the full dataset.
6. Analyze the model's predictions to identify the most important factors driving churn.

Overall, the key to testing this hypothesis will be to gather a rich and representative dataset, and to use a range of analytical techniques to explore the relationships between price sensitivity and churn.

In addition to the above steps, we could also consider using techniques such as customer segmentation or survival analysis to further investigate the relationship between price sensitivity and churn. This could provide additional insights into how different types of customers respond to price changes and the factors that influence their decision to stay with or switch providers.

I hope this gives you a sense of how we can approach this problem. Let me know if you have any further questions or if there is anything else I can do to help.

Best regards,

Vaishali Gupta