Dear AD,

After discussing the hypothesis that churn is driven by customers’ price sensitivities, I have formulated a data science problem to test this hypothesis. The problem is:

Given the historical data on customer behavior and pricing, can we predict which SME customers are at the highest risk of churning and would benefit from a 20% discount offer?

To test this hypothesis, we will need access to the following data from the client:

Customer data: This should include information on customer demographics, business size, and historical purchase behavior, such as the length of the customer's relationship with PowerCo, the type and amount of their previous purchases, and their payment history.

Pricing data: This should include information on the pricing plan of each customer, including any changes made over time, and the discount or promotional offers they have received.

Churn data: This should include information on which customers have churned and when they did so.

With these datasets, we can start the following steps:

Data Cleaning and Preparation: This step involves cleaning the data and ensuring that it is suitable for analysis. We will need to remove any duplicates, missing values, or outliers from the dataset.

Exploratory Data Analysis: In this step, we will analyze the data to gain insights into the relationship between customer churn and pricing. This could involve generating summary statistics, visualizations, and correlation analysis.

Feature Engineering: We will create new features that can improve our prediction model’s performance. For example, we can calculate the average customer spending, frequency of purchase, and any changes in customer spending behaviour.

Model Development: We will develop a predictive model that uses historical data to predict which customers are at the highest risk of churning. We can use machine learning algorithms, such as logistic regression or decision trees, to build the model.

Model Evaluation: We will evaluate the performance of our model to ensure that it is accurate and effective. We can use metrics such as AUC, ROC, and precision-recall curves to evaluate the model's performance.

Discount Offer: We can use the predictive model on the 1st working day of every month to indicate which customers are at the highest risk of churning and offer them a 20% discount. We can also evaluate the effectiveness of the discount offer in reducing churn rates.

In summary, testing the hypothesis that churn is driven by customers’ price sensitivities involves accessing and cleaning data on customer demographics, purchase behavior, pricing, and churn. We will then conduct exploratory data analysis, feature engineering, model development, model evaluation, and implement a discount offer to reduce churn rates.

Please let me know if you have any questions or feedback on this approach.

Best regards,

Ayushi Asthana