Dear AD,

Following our discussion on the hypothesis that churn is driven by customer price sensitivity, I have formulated the problem as a data science task and laid out the major steps needed to test this hypothesis.

Hypothesis: The churn rate of Power Co’s SME customers is driven by their price sensitivity.

Problem Formulation: We aim to develop a predictive model that can identify which SME customers are likely to churn based on their current price and price sensitivity. The model should help company determine which customers should receive the 20% discount to incentivize them to stay.

Proposed Approach:

Testing the hypothesis that churn is driven by customer price sensitivity requires a data-driven approach that involves collecting historical customer data, conducting EDA, feature engineering, developing a predictive model, evaluating its performance, and deploying it for operational use. By following these steps, we can help company to reduce customer churn and improve customer retention.

* **Data Collection:** We would need historical customer data, including demographics, usage patterns, billing information, and customer feedback. We would also need information on prices charged to each customer and whether they churned or not.
* **Exploratory Data Analysis (EDA):** We would conduct EDA to understand the distribution of customer characteristics, identify patterns in customer behavior, and assess the relationship between prices charged and churn rate. This step would help us identify potential confounding factors that may impact customer churn.
* **Feature Engineering:** Based on our EDA, we would create new features that capture customer behavior, such as the length of time they have been a customer, their usage patterns, and their payment history. We would also create a feature that captures the customer's price sensitivity based on their willingness to pay, which we can estimate from customer feedback or surveys.
* **Model Development:** We would use a binary classification model, such as logistic regression or random forest, to predict which customers are likely to churn. The model would use the customer features and their price sensitivity as inputs. We would train the model on historical data and validate its performance using cross-validation techniques.
* **Model Evaluation:** We would evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score. We would also conduct sensitivity analysis to assess the impact of changing the discount threshold on customer retention rates.
* **Model Deployment:** We would deploy the model on the 1st working day of every month to identify which customers are at risk of churning and offer them the 20% discount to incentivize them to stay.

Please let me know if you have any questions or concerns.

Best regards,

Vishita