Dear AD,

I hope this message finds you well. I appreciate the opportunity to delve deeper into the hypothesis that customer churn is driven by price sensitivity for PowerCo's SME customers. To test this hypothesis, we need a structured approach that involves data collection, analysis, and model development. The following are my strategy:

1. Data Collection: To test the hypothesis effectively, we need access to relevant data from PowerCo. Here are the key data sources and fields that would be crucial for our analysis:
   ● Customer Data: Collect comprehensive information about the SME customers, including their contract details, location, company size, industry, and historical usage data.
   ● Billing Data: This should include historical billing records, payment history, and any past discounts or promotions offered to each customer.
   ● Churn Data: A dataset that specifies which SME customers have churned (switched to a different provider) over a given period.

2. Data Frame Structure: The ideal data frame for our analysis should have the following columns:
   ● Customer ID
   ● Customer Characteristics: Location, company size, industry, etc.
   ● Historical Usage Data: Electricity and gas consumption patterns.
   ● Billing Data: Previous bills, payment history, applied discounts.
   ● Competitor Information: Presence of competitors in the service area, pricing data.
   ● Churn Status: Binary variable indicating churn or non-churn (1 for churn, 0 for non-churn).
Each row represents a unique SME customer.

3. Exploratory Data Analysis: Before building predictive models, we need to gain a deep understanding of the data and the factors influencing customer churn. We can perform various exploratory analyses, including:
   ● Descriptive Statistics: Summarize key variables to identify trends and patterns.
   ● Correlation Analysis: Determine the relationships between customer characteristics, historical usage, billing, and churn.
   ● Data Visualization: Utilize data visualization techniques such as histograms, box plots, and heatmaps to identify outliers and distributions.
   ● Feature Engineering: Create new features based on insights gained from the data exploration.

4. Model Development: To test the hypothesis, we can develop predictive models using techniques such as logistic regression, decision trees, or random forests. The target variable will

be churn status (1 or 0), and the predictors will include customer characteristics, historical usage data, billing information, and competitor-related variables.

We will also evaluate model performance using metrics like accuracy, precision, recall, and F1-score. This will help us identify which factors, including price sensitivity, contribute most to customer churn.

5. Hypothesis Testing: To specifically test the impact of price sensitivity on churn, we can conduct hypothesis tests, such as t-tests or ANOVA, to compare churn rates between groups of customers with different price sensitivity levels. We may derive a price sensitivity metric based on historical billing data and use it to categorize customers into high, medium, and low price sensitivity groups.

6. Model Deployment: Once the predictive model is developed and validated, we can deploy it to identify at-risk customers. On the 1st working day of each month, the model can provide a list of customers who are likely to churn. We can then offer a 20% discount to these customers as a retention strategy.

In conclusion, to test the hypothesis that customer churn is driven by price sensitivity, we need to collect relevant data, perform comprehensive exploratory analysis, develop predictive models, and conduct hypothesis testing. By following this structured approach, we can provide valuable insights and practical recommendations to reduce SME customer churn.

I look forward to your feedback and further guidance on how to proceed with this important project.

Best regards,
Jammal