Subject: Testing Hypothesis on Customer Churn Due to Price Sensitivity - Task Approach and Data Requirements

Dear Associate Director,

I hope this email finds you well. As per our recent team meeting, I have delved deeper into the hypothesis that customer churn at PowerCo may be driven by price sensitivities. I have formulated a data science problem and outlined the major steps needed to test this hypothesis effectively. Below are my thoughts and findings:

Hypothesis Formulation:
The hypothesis we aim to test is whether price changes affect customer churn at PowerCo, especially in the SME segment. We want to identify which customers are more (or less) likely to churn at their current price and whether offering a 20% discount would incentivize at-risk customers to stay with our client.

Approach:
To test this hypothesis, we can follow the following steps:

Data Collection:
We need historical data of gas and electricity utility customers, specifically from the SME segment, which includes information about pricing, customer profiles, churn status, and other relevant variables.

Data Preprocessing:
Cleanse and preprocess the data to handle missing values, outliers, and ensure data integrity.

Feature Engineering:
Create relevant features such as customer tenure, usage patterns, historical churn behavior, and pricing metrics to enrich the dataset.

Exploratory Data Analysis (EDA):
Perform EDA to gain insights into the factors influencing customer churn. Analyze relationships between churn, pricing, and other variables to identify patterns.

Hypothesis Testing:
Conduct statistical tests to determine if there is a significant correlation between price changes and customer churn. We can use methods like t-tests, ANOVA, or correlation analysis.

Predictive Modeling:
Build a predictive model using machine learning algorithms such as logistic regression, decision trees, or random forests to predict customer churn based on pricing and other relevant features.

Model Evaluation:
Evaluate the model's performance using metrics like accuracy, precision, recall, and F1-score to assess its predictive power.

Offer Recommendation:
Identify customers who are at risk of churning and assess whether offering a 20% discount would likely dissuade them from switching providers.

Data Requirements:
To conduct this analysis, we would require the following data from PowerCo:

Customer Information:
Customer ID or unique identifier
Customer type (corporate, SME, residential)
Historical churn status (if available)
Pricing Information:
Pricing history, including any recent price changes
Pricing plans offered to SME customers
Usage Patterns:
Historical usage data (if available)
Customer Feedback:
Any customer feedback or complaints related to pricing
Customer Interaction Data:
Customer inquiries or interactions regarding price changes or discounts
Timeframe:
I anticipate that this analysis could be completed within the next few weeks, allowing us to present our findings to you by [Date].

Please let me know if you require any additional information or have any specific preferences regarding the analytical models we should use. I am eager to discuss this further with you in our upcoming meeting.

Thank you for entrusting me with this task. I am committed to delivering valuable insights to help PowerCo address the issue of customer churn effectively.

Best regards,
Sumant Kale
Data Scientist
BCG Gamma