Subject : Proposed Plan for Testing Churn Hypothesis for PowerCo's SME Segment

Dear Associate Director,

I hope this email finds you in your best of health. After thoroughly examining the churn problem experienced by PowerCo within its SME customer segment, I am putting forth an approach that could help us test the hypothesis suggesting that customer churn is driven by price sensitivity.

The `hypothesis` can be formulated as: "SME customers are churning due to price sensitivity, and offering a 20% discount could reduce this churn.". To verify this, we will need to leverage PowerCo's **historical data**, ideally from the past 2-3 years

Here are the steps I propose:

1. Data Collection

Gather historical data from PowerCo, encompassing the following:

- Customer Details (industry, location, size, etc.)
- Pricing Details (price charged to each customer over time)
- Churn Status (whether the customer churned or not)
- Customer Feedback (customer satisfaction scores, complaints, etc.)
- Customer Usage Data
- Any past discounts and responses to offers

2. Understanding Key Factors for Churn

While price might be one of the key factors influencing churn, we should remember that other factors such as *quality of service, billing transparency, and customer support* can also influence a customer's decision to stay or switch providers. Therefore, any data around these areas will be useful.

3. Ideal Data Framework

Our ideal data frame would include individual rows for each customer and columns representing various factors such as *price, churn status, demographic information, customer feedback scores, and past engagement with discounts or offers*.

4. Data Preparation

The obtained data will need to be cleaned and preprocessed to manage any missing values, normalize variables, and handle outliers.

5. Exploratory Data Analysis

Conduct an initial analysis of the data to find patterns or correlations between churn and these variables, especially price. Various exploratory analyses could include studying the distribution of churn across various pricing levels, examining the churn rate among customers who have previously engaged with discounts, or analyzing the churn rate across different demographic groups.

6. Model Selection and Training

Considering our aim to predict churn `(a binary outcome)`, suitable models might include logistic regression, decision trees, random forest, or gradient boosting machines.

7. Model Evaluation

To validate the performance of the model, we will use cross-validation and consider various metrics like accuracy, precision, recall, F1 score, and AUC-ROC. We will also split the data into a training set and a test set. The model will be trained on the training set and evaluated on the test set to ensure it generalizes well to unseen data.

8. Model Deployment

Once the model is refined and validated, we propose that PowerCo use it on the first working day of every month to identify customers at high risk of churn and to determine to whom the 20% discount should be offered.

Thank you for considering our proposed approach. We are confident that this strategy will provide valuable insights into the SME churn problem and are looking forward to your feedback.

Regards,

Raja Allmdar Tariq Ali