Subject: Hypothesis Testing and Approach for SME Customer Churn

Dear Associate Director,

I hope this message finds you well. I wanted to share my thoughts on how we should approach testing the hypothesis that SME customer churn is driven by price sensitivities. This is a critical issue, and I believe a data-driven approach will provide valuable insights.

Hypothesis Formulation:

The hypothesis we aim to test is whether changes in pricing significantly impact SME customer churn. In other words, we want to determine if customers are more likely to churn when faced with price increases and if offering a 20% discount can effectively retain them.

Major Steps to Test the Hypothesis;

1. **Data Collection:**
   - We need historical data on SME customer behavior, including churn status, pricing changes, and other relevant attributes.
   - Data should cover a sufficient time frame to capture different pricing scenarios and customer responses.
   - Ideally, we should have customer-level data to perform a granular analysis.

2. **Data Preprocessing:**
   - Clean and preprocess the data, handling missing values and outliers.
   - Create a target variable (churn indicator) and identify the relevant pricing variables.

3. **Exploratory Data Analysis (EDA):**
   - Conduct EDA to understand the distribution of pricing changes, churn rates, and customer demographics.
   - Analyze price sensitivity by segmenting customers based on historical price changes.
   - Identify any patterns or correlations between price changes and churn rates.

4. **Feature Engineering:**
   - Create relevant features such as price change percentage, customer tenure, industry type, and customer size.
   - Consider seasonality and external factors that may influence churn.

5. **Model Selection:**
   - Choose appropriate predictive models for churn analysis. Logistic regression, decision trees, random forests, or gradient boosting may be suitable candidates.
   - Train the models using historical data, using churn as the target variable.

6. **Model Evaluation:**
   - Assess model performance using metrics like accuracy, precision, recall, and F1-score.

- Use cross-validation to ensure robustness and avoid overfitting.

7. **Interpretation:**
   - Interpret the model results to understand which pricing factors are most influential in predicting churn.
   - Determine the threshold for offering the 20% discount based on the model's predictions.

8. **Validation and Implementation:**
   - Validate the model's predictions on a holdout dataset or through A/B testing.
   - Implement the predictive model to identify customers at risk of churning and recommend the discount on the 1st working day of each month.

9. **Monitoring and Feedback Loop:**
   - Continuously monitor the model's performance and gather feedback from the field.
   - Refine the model and pricing strategy as needed based on real-world results.

Data Requirements:

To proceed with this analysis, we will need historical data on SME customers, including:
- Customer ID
- Churn status (1 for churned, 0 for retained)
- Pricing change history (dates and percentage changes)
- Customer demographics (industry, size, tenure, location)
- Any additional relevant data, such as customer complaints or inquiries

A well-structured data frame would have each row representing a unique SME customer and columns containing the above attributes.

I believe that by following these steps and utilizing data-driven insights, we can effectively test the hypothesis regarding the impact of price sensitivity on churn and make informed decisions about offering discounts to retain at-risk customers.

Please let me know if you would like to discuss this further or if you have any specific preferences regarding the analysis approach.

Best Regards,
Sharon Kiman