**Subject: Testing the Hypothesis - Predictive Model for Discount Offer Strategy**

Dear [Assistant Director],

I hope this message finds you well. I wanted to share my thoughts on how we can approach testing the hypothesis for the predictive model and discount offer strategy as requested by the client. The client intends to use the model on the 1st working day of each month to identify customers to whom they should offer a 20% discount.

**Hypothesis:** Customer churn in the SME segment is influenced by price sensitivity.

**Hypothesis Formulation:**

The hypothesis to be tested is whether offering a 20% discount to customers identified by our predictive model will effectively reduce churn rates within the SME segment. This hypothesis implies that customer churn is driven by price sensitivity, and offering discounts to price-sensitive customers will encourage them to stay with the service.

Major Steps to Test the Hypothesis:

**1. Data Collection:**

We need to gather historical customer data from the client, including but not limited to:

- Customer demographics (age, location, industry, etc.)

- Historical usage patterns (energy consumption)

- Billing data (payment history, bill amounts)

- Contract duration and terms

- Interaction history (complaints, inquiries, etc.)

- Historical price changes and discounts offered

- Churn history (customers who left and those who stayed)

**2. Data Preprocessing:**

Clean and preprocess the data to handle missing values, outliers, and format it in a way suitable for analysis. Feature engineering is crucial to create variables that represent customer price sensitivity, such as price change ratios and usage fluctuation.

**3. Exploratory Data Analysis (EDA):**

Conduct EDA to gain insights into the data and understand what drives churn. Explore relationships between variables and identify patterns. Key questions to answer include:

- What are the common characteristics of customers who have churned?

- Are price-sensitive customers more likely to churn?

- Do customer interactions impact churn behaviour?

**4. Predictive Model Development:**

Develop a predictive model, such as logistic regression, decision trees, random forests, or a machine learning algorithm, to identify customers with a high propensity to churn. The target variable is whether a customer will churn (1) or not (0).

**5. Model Evaluation:**

Evaluate the predictive model's performance using metrics like accuracy, precision, recall, F1-score, and ROC AUC. Given the business context, we should pay special attention to minimizing false negatives.

**6. Discount Offer Simulation:**

Using the predictive model, simulate the impact of offering a 20% discount to customers with a high churn propensity. This can be done on a monthly basis, as per the client's plan, and we can measure the effects on churn rates and financial implications.

**7.Hypothesis Testing:**

We will test the hypothesis that price sensitivity influences churn by:

1.Performing statistical tests (e.g., correlation analysis) to determine the relationship between price changes and churn.

2.Conducting feature importance analysis within the predictive models to assess the significance of price-related features.

**8. Reporting and Recommendations:**

Present the findings and recommendations to the client. This should include the effectiveness of the predictive model, insights into the factors driving churn, and the potential impact of the discount strategy. Provide actionable recommendations on how to proceed.

**Ideal Data Frame:**

An ideal data frame for this analysis would have rows representing individual customers and columns representing customer attributes, historical data, and model-derived features. It should include a binary column indicating churn (1 for churned, 0 for retained).

**Exploratory Analyses:**

In exploratory analysis, we can perform various tests and visualizations, such as:

- T-tests to compare means of price-sensitive and non-price-sensitive customers.

- Correlation analysis to identify factors influencing churn.

- Box plots to visualize the distribution of key variables among churned and retained customers.

I would be happy to discuss this further and start working on this project as soon as we have access to the necessary data. Please let me know if you have any specific instructions or additional considerations.

Best regards,

[Nida Attaraut]