**Task 1**:

Your first task today is to understand what is going on with the client and to think about how you would approach this problem and test the specific hypothesis.

You must formulate the hypothesis as a data science problem and lay out the major steps needed to test this hypothesis. Communicate your thoughts and findings in an email to your AD, focusing on the data that you would need from the client and the analytical models you would use to test such a hypothesis.

**Solution:**

Subject: Analysis Plan for Testing the Hypothesis of Churn Driven by Price Sensitivity

Dear AD,

I hope this email finds you well. I'm excited to outline our plan to test this hypothesis for PowerCo's SME segment.

**Hypothesis Formulation**:

Our data science problem is to determine whether customer churn is influenced by price changes. Specifically, we want to understand if certain customers are more or less likely to churn at their current price. To achieve this, we will build a predictive model that can identify the customers at risk of churning and assess how a 20% discount might affect their likelihood of staying with PowerCo.

**Analytical Approach**:

To test the hypothesis, we propose the following major steps:

1. ***Data Collection***:

We will collaborate with PowerCo to gather relevant data on customer characteristics, historical energy consumption patterns, billing information, and any past churn records. Additionally, we need data on the price changes that occurred during the power liberalization period. The dataset should cover a sufficient time period to capture customer behaviors under different price conditions.

1. ***Feature Engineering:***

Once we have the data, we will perform feature engineering to extract relevant insights from the raw dataset. We'll consider various features such as customer demographics, contract type (fixed vs. variable), contract duration, historical consumption patterns, and payment history. We will also calculate metrics like customer tenure and price change percentages for individual customers.

1. ***Labeling Churn***:

Next, we need to define the churn label. For this analysis, we can define a customer as churned if they terminate their contract with PowerCo within a specific period (e.g., three or four months) after a price change. Customers who remain with the company during this period will be considered non-churned.

1. ***Model Development***:

We will employ machine learning techniques to develop a predictive model for customer churn. Considering that churn prediction is a binary classification problem, we can explore models such as Logistic Regression, Random Forest, Gradient Boosting, or Neural Networks. The dataset will be split into training and testing sets to evaluate model performance accurately.

1. ***Model Evaluation and Interpretation***:

Once the model is trained, we will evaluate its performance using appropriate metrics like accuracy, precision, recall, and F1 score. Additionally, we will analyze the model's feature importance to understand which factors contribute most significantly to churn.

1. ***Identifying At-Risk Customers:***

With a validated model in hand, we can apply it to predict churns for all active SME customers at the beginning of each month. This will allow us to identify the customers most at risk of churning due to price sensitivity.

1. ***Discount Incentive Strategy***:

For the at-risk customers, we will simulate the impact of offering a 20% discount. We can estimate how many customers would be incentivized to stay with PowerCo by comparing the predicted churn probabilities before and after the discount.

1. ***Monitoring and Refinement***:

It's crucial to continuously monitor the model's performance and retrain it periodically to adapt to changing customer behaviors and market dynamics.

**Data Required:**

To proceed with the analysis, we will need the following data from PowerCo:

***Customer Information***: Demographics, contract type, contract duration, payment history, etc.

***Consumption Data***: Historical energy usage patterns for each customer.

***Billing Information***: Pricing details and any past price changes.

***Churn Data***: Records of customers who churned and the corresponding time of churn.

***Timeline***: We understand the importance of timely results for PowerCo. We aim to complete the initial analysis and build a predictive model within 12 months. Following that, we will provide regular updates and insights at the start of each month to support PowerCo's discount incentive strategy.

By following this plan, I am confident we can gain valuable insights into the relationship between price sensitivity and customer churn for PowerCo's SME segment.

If you have any questions or suggestions, please feel free to reach out to us. We look forward to starting the analysis and contributing to PowerCo's efforts to retain its valued customers.

Best,

Manisha Kaila