Formulating the Hypothesis:

Using collected/Historical Data to build a predictive model that can categorize users as long-term customers (Low Churn rate) or not.

Steps Needed to test Hypothesis:

First Approach: predictive machine learning model:

Gather Historical Data that can help point us in the right direction and train a machine learning model. The model will then be able to predict if a customer has a high churn rate or not. Furthermore, after training the model we can extract from it the most relevant data points that affect churn rate. And if 'Discount percentage' is deemed to be valuable. Then we can proceed.

Second Approach: Hypothesis Testing (A/B):

Gather Data with discounts applied for a limited time and measure the difference in churn rate relative to original data (without discounts) to help identify the impact of applying a discount on the churn rate. And if Discounting is deemed affective, we can proceed.

Data Needed:

1. Customer Data – characteristics of each client
2. Pricing Data – the prices the client charges his customers for their services
3. Data to be predicted – if the customer churned or not