Hi AD,

In order to test the hypothesis of the churning, we need data of our clients and build predictive model to forecast the probability of churning. Furthermore, we use the model to explore the best discount rate to keep our clients stay.

The customer data could include customers' historical behaviors (Churn data) and some personal info (Customer data and price data).

1. Churn data – This attribute will indicate if the customer has churned or not.
2. Customer data – Personal info such as age, plan, commercial or residential, etc.
3. Price data- Shows the prices that client charged to customers.

After obtaining the data, we need to engineer the data, including cleaning, binning, dimensional reduction, etc. We can also visualize the data to do some explorations for finding patterns between features. Then we can build predictive model for a binary classification task. The potential model could be naïve bayes, SVM, Decision Tree, Random Forest. Some ensemble learning models except Random Forest can also be tried in order to get a better accuracy but losing the ability of interpretation as a cost. The final selection will be based on the trade-off between explanability, complexity and performance. On evaluation of the performance, True Positive would be our priority because we do not want any customers of us churn, therefore the sensitivity (precision) will be the evaluation metric. Based on the model, we can also size the business impact of the proposed discount strategy.

Regards,
Changhui He