Dear Lead Data Scientist,

Our current task is to test the hypothesis that the customer churn is driven by the customers’ price sensitivities. From the perspective of data science, this question can be rephrased as “if price is an important feature to predict customer churn”. The major strategy is to build a classification model to predict whether a customer churned or not.

Steps to move forward:

* Step one: Data collection. Potential data we would need as follows:
  + Customer profile: features about customer’s profile in the last 5 years, such as industry, category, location, the date of joining our service, etc.
  + Consumption behavior data: such as unit price in each bill interval, electricity consumption, whether churn or not
  + Other reference: unit price of electricity for different types of clients in last 5 years
* Step two: Data cleaning and feature engineering.
* Step three: Modeling.
  + Build several classification models including but not limited to logic regression, random forest, and XGBoost. And we would use ROC-AUC as a major evaluation metric to pick the best model among others.
* Step four: Model explanation.
  + Based on the best model we chose, we are able to rank the feature importance.
  + To measure if price is a leading factor of customer churn and its magnitude.
* Step five: Test the discounting strategy.
  + Run an experiment to check if 20% discount would improve current situation.

Best regards,

Chuanlu