Hi Iman,

To test the hypothesis that pricing leads to customer churn and that customer churn could be modelled using a predictive model, the following is the data needed, modelling approach to follow and the possible outcomes expected in the end.

Potential Data

- Customer data - data about the client's customers; demographic, utilities subscribed, etc
- Pricing data - pricing data on gas and electricity utilities
- Churn data - churn data identified customers who churned

Once we've collected the needed data, we will conduct Exploratory Data Analysis to gain insights on any correlations among churn, pricing and customer data. This stage will potentially result in the creation of new features that will help us model the hypothesis. The resulting data will be fit on the following classification models (Decision Trees, Logistic Regression, SVM, Xgboost, etc) to model customer churn.

The potential outcome of the process will be 1) a predictive model that models customer churn 2) a dashboard explaining the model (Explainable AI) and 3) a discounting strategy for customers with a high propensity to churn.

Regards,
David Masupa