EDA Summary

Findings

● 9.7% of customers have churned

● Client dataset is an imbalanced set, during the modelling this imbalance will lead to a large number of false negatives.

● Consumption data is highly skewed and must be treated before modelling

● There are outliers present in the data and these must be treated before modelling

● Price sensitivity has a low correlation with churn

Suggestions

● Consumption data can be predictor variables

● Client feedback - a track record of any complaints, calls or feedback provided by the client to PowerCo might reveal if a client is likely to churn