

Summary

One of the key findings in our data exploratory is that only around 10% of customers have churned. Further, Consumption data was positively skewed to the right with many extreme outliers which need to be resolved prior to feature engineering, even Forecast data had some positive skewness to the right, though some forecast fields were unimodal as well as bimodal (one peak and two peaks respectively). Moreover, when analysing the correlation between features, it seemed that mostly there isn't a strong correlation between them, with a few exceptions that may be considered moderate correlation. So far, there is a weak correlation between price sensitivity and churn. In brief, feature engineering is going to be essential in further analysing and investigating the hypothesis. One suggestion for data augmentation can be top competitors' prices data, this would give us insight allow comparison of average prices at different periods, which eventually may help us understand the motive behind customer churn.