Key Findings

Data Set

1) Price_data
   Total no of observations -193002
   Total no of features-8
2) Client_data
   Total no of observations - 14606
   Total no of features- 26
   No missing values

Target variable – churn

1-customers churned (13187)

0-customers retained (1419)

- No missing values in both data sets.
- Non numeric values are present so we need to convert those categorical variables into numeric.
- Target variable is highly imbalanced.
- The distribution of the consumptions is highly right skewed. Outliers presence in consumptions.
- The date's distribution seems does not provide any insight.
- The distribution of the forecast is highly right skewed.