CLASSIFICATION PROBLEM: LOGISTIC REGRESSION ON A TELECOM CUSTOMER

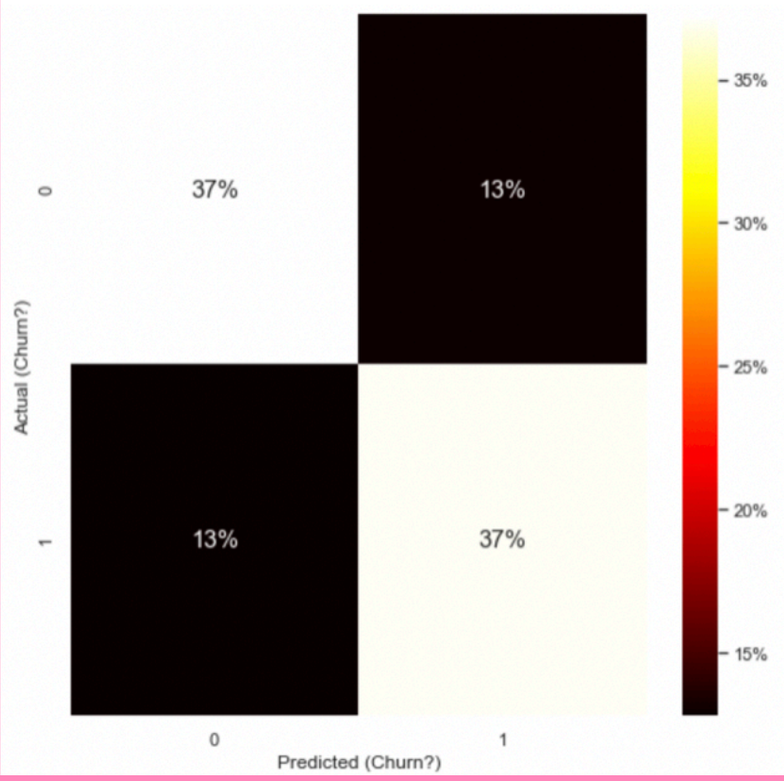
H0 : We cannot predict if a customer is more likely to churn or not.

Ha : We can reliably predict on whether a customer is more likely to churn or not.

BUSINESS IMPACT

It would be very valuable to find a model that predicts whether a customer is more likely to churn or not, as we will have an approach on how the customers will behave and **define strategic actions to retain them.**

ACCURACY OF THE MODEL



NEXT STEPS

- Collecting more data from telecom customers: income… and also collect data from previous years so we have more data points
- Feature selection: we could add also the contract column and the rest of the columns after doing some dummies
- Data wrangling, pre-processing:
  - Cleaning: looking for proportion of null values and decide whether it would be useful to get rid of some rows. Maybe delete customerID column as we don't need it.
  - We already applied SMOTE to handle the imbalance of data. Maybe we could also apply TOMEK and compare which would perform better
- Statistical tools: Do some visualizations: churn rate by contract, churn rate by monthly charges, by tenure… and look for patterns, tendencies…