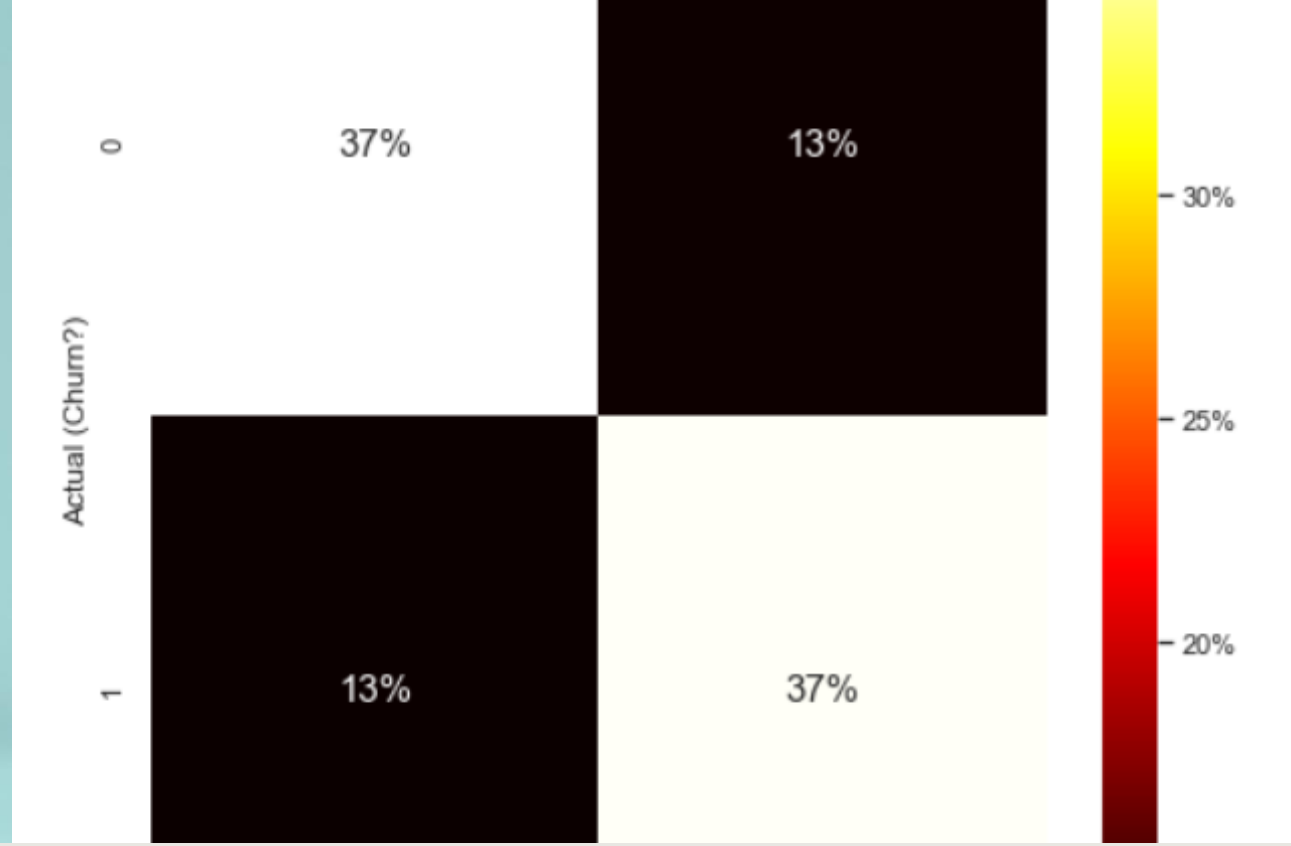
The hypothesis, when the label is ‘0’ null-hypothesis (since it’s the majority class it doesn’t make much sense) and when the label is ‘1’ null hypothesis is false

Suggestion:

- In this task, there is a feature called ‘senior citizen’, it has more 0’s.
- Collecting more unbiased data.



Business question: Can we predict if a customer account will churn?

Though the causes of imbalanced models are many, according to my understanding the high possibility is because of biased/skewed data set. In other way, biased sampling and measurement errors.

In this example, the challenge was the class has ‘0’ majority and ‘1’ minority in which the majority has more than 5000 rows and the minority has less than 2000 rows.

The plot above shows the confusion matrix output after oversampled with the technique SMOTE (Synthetic Minority oversampling technique)