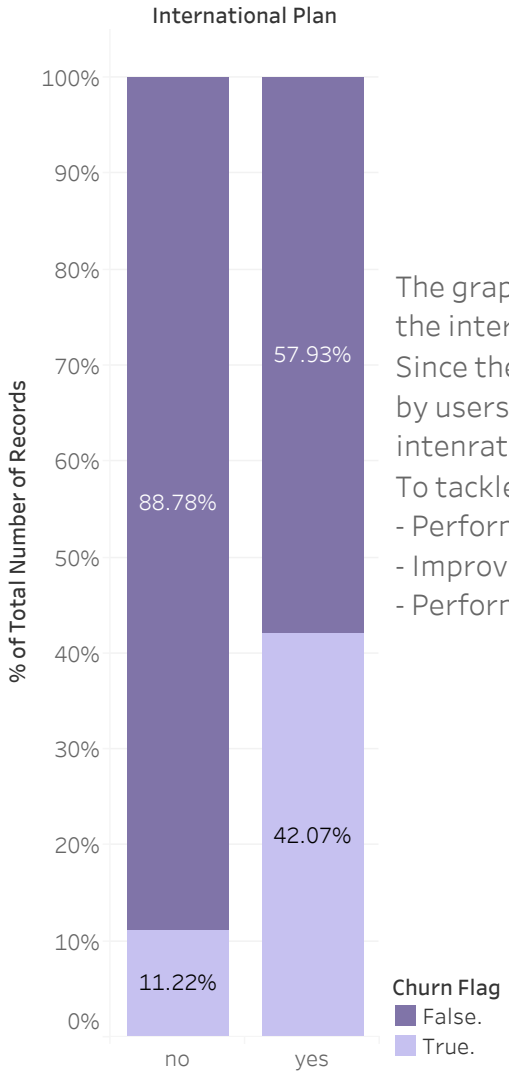
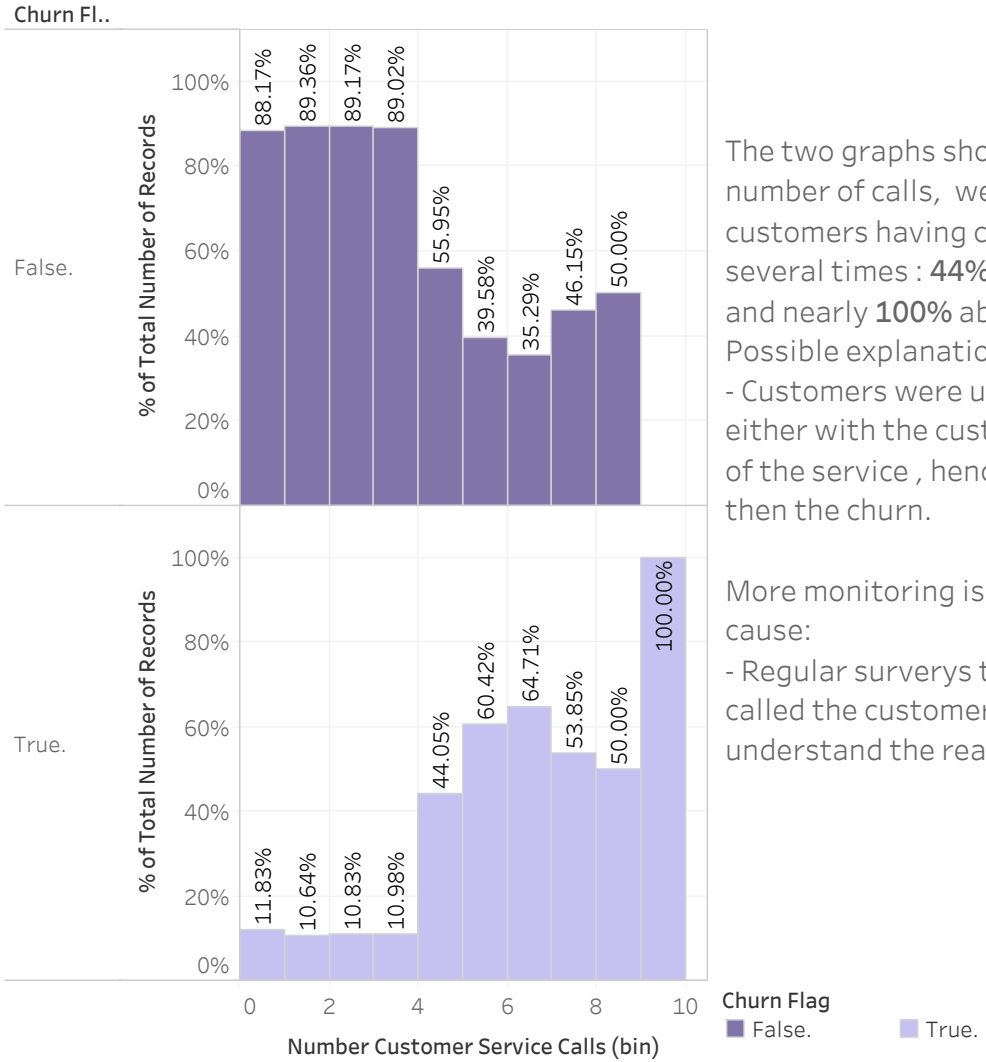
International plan



The graph on the left shows that the churn rate is quite high for users of the international plan : **42.07%** against **11.22%** for non subscribers.
Since the Telecom market is quite saturated, this churn might be explained by users moving to other operators with a cheaper/better plan for intenrational calls.
To tackle this issue, the international plan could be improved by:
- Performing a study what the other competitors are offering
- Improving the plan and provide a better deal
- Performing a marketing compaign to get more subscribers

customer calls



The two graphs show the ratio of churners for each number of calls, we see a high churn among customers having called the customer service several times : **44%** for more thant 4 service calls and nearly **100%** above 9 calls.
Possible explanation for this :
- Customers were unhappy, might have had issues either with the customer service call or the quality of the service , hence the high number of calls, and then the churn.

More monitoring is needed to identify the root cause:
- Regular surverys targetting customers having called the customer service more than 4 times to understand the reason these customers leave.

churn/state

State	Churn Flag False.	Churn Flag True.
California	73.08%	26.92%
New Jersey	75.00%	25.00%
Washington	75.51%	24.49%
Texas	77.59%	22.41%
Montana	78.79%	21.21%
Maryland	79.41%	20.59%
Nevada	81.11%	18.89%
Maine	81.55%	18.45%
Kansas	81.82%	18.18%
Oklahoma	82.22%	17.78%
South Carolina	82.42%	17.58%
New York	82.46%	17.54%
Michigan	82.52%	17.48%
Arkansas	82.61%	17.39%
Mississippi	82.83%	17.17%
Massachusetts	83.50%	16.50%
Connecticut	83.84%	16.16%
Minnesota	84.00%	16.00%
Delaware	84.04%	15.96%
Oregon	84.21%	15.79%
Tennessee	84.27%	15.73%
Kentucky	84.85%	15.15%
Indiana	85.71%	14.29%
Utah	85.71%	14.29%
South Dakota	85.88%	14.12%
West Virginia	86.08%	13.92%
Florida	86.67%	13.33%
North Carolina	86.81%	13.19%
Ohio	87.07%	12.93%
Missouri	87.10%	12.90%

% of Total Number of Records

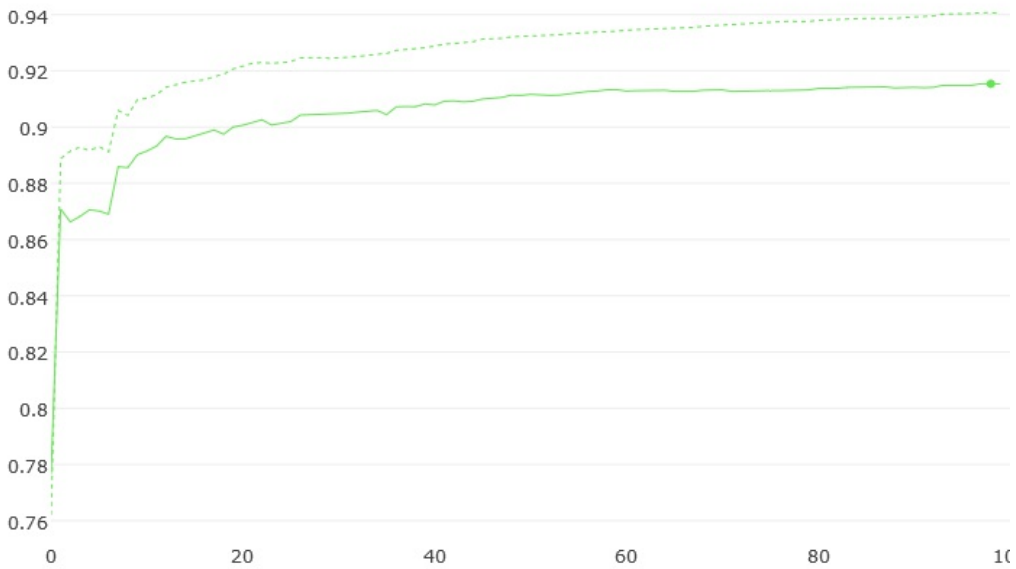
5.08%  94.92%

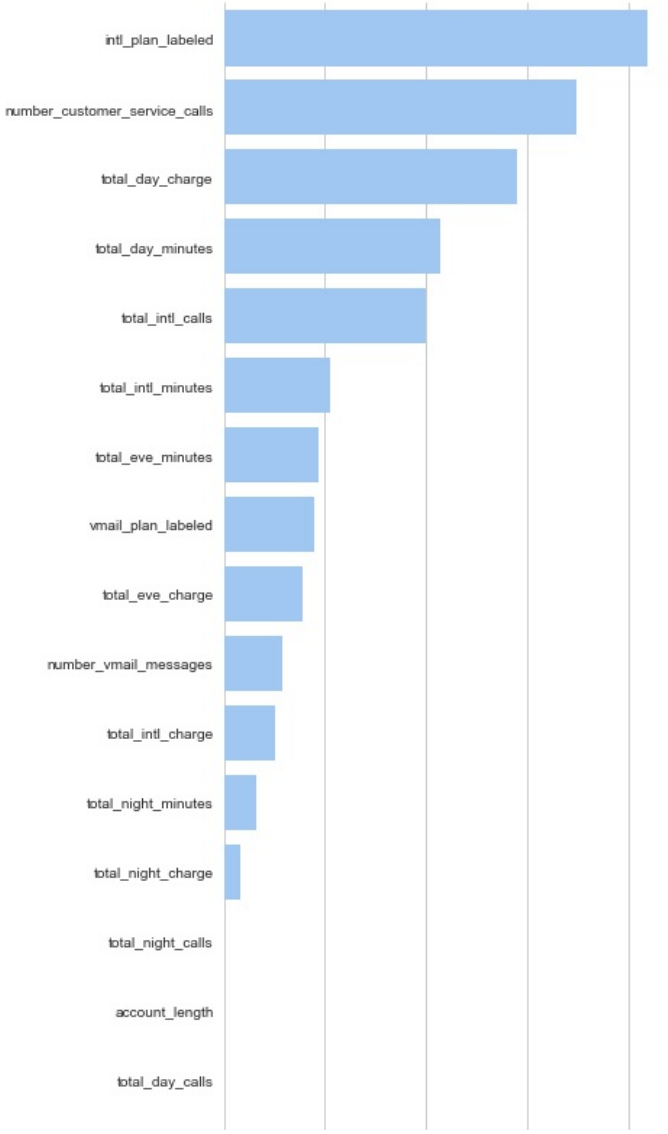
If we look at the churn rate by state, we see that **California**, **New Jersey**, and **Washington** have the highest churn rate (more than **24%**).
ALthough a further analysis is needed to get more visibility about the market (how many competitors, population size to assess the market saturation and services comparison with other operators), It could help to concentrate a targetting compaign on those states to get more subscribers.

In order to predict the probability of a customer to churn, I used a Gradient boosting model (Catboost)
All features were used in the model with the exception :
- States , Area code, Phone number

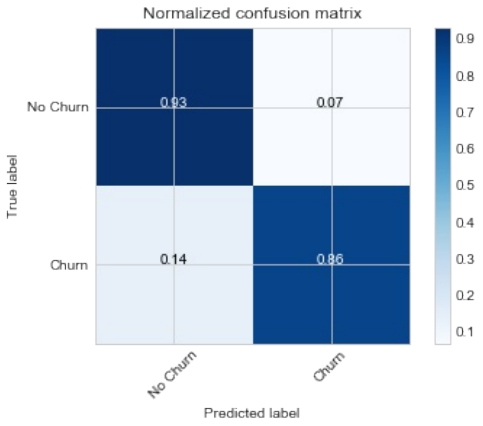
Metric used:
Since the classes are imbalanced, I used AUC metric (area under the ROC curve)

In the bellow graph, the line in dots represents the training set AUC over the interations, the continued line belongs to the validation set





The graph on the left shows the importance of the features of the trained model.
It is interesting to see that the international plan and service calls are strong predictors for the churn as the descriptive has suggested earlier.
On the contrary, how many calls the customer has made seems to be irrelevant



The model output is probability of the customer churning, in order to compute prediction, a threshold should be determined.
By default it is usually set at 0.5, however depending on the usecase, we might want to change it to give more importance to precision over recall or the opposit.

The graphs show the confusion matrix of the trained model, below is the number of TP,FP,TN,FN .
On the left the numbers are normalised by number of churners/non churners

In this usecase, a churn of a customer is costly, even more than applying a retention strategy on a customer who is unlikely to churn.
Therefore, we want to have a high true positive rate, and low false negative rate to select most of the customers who are likely to churn.

After testing with different threadsholds, I settled with **0.25** which yielded a **86%** of TP and **14%** of FN.