Telecom Churn Case Study

SUBMITTED BY

RAGHAV VIR KAKKAR

Problem Statement

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, **customer retention** has now become even more important than customer acquisition.

For many incumbent operators, *retaining high profitable customers is the number one business goal*.

To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**

In this project, we will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

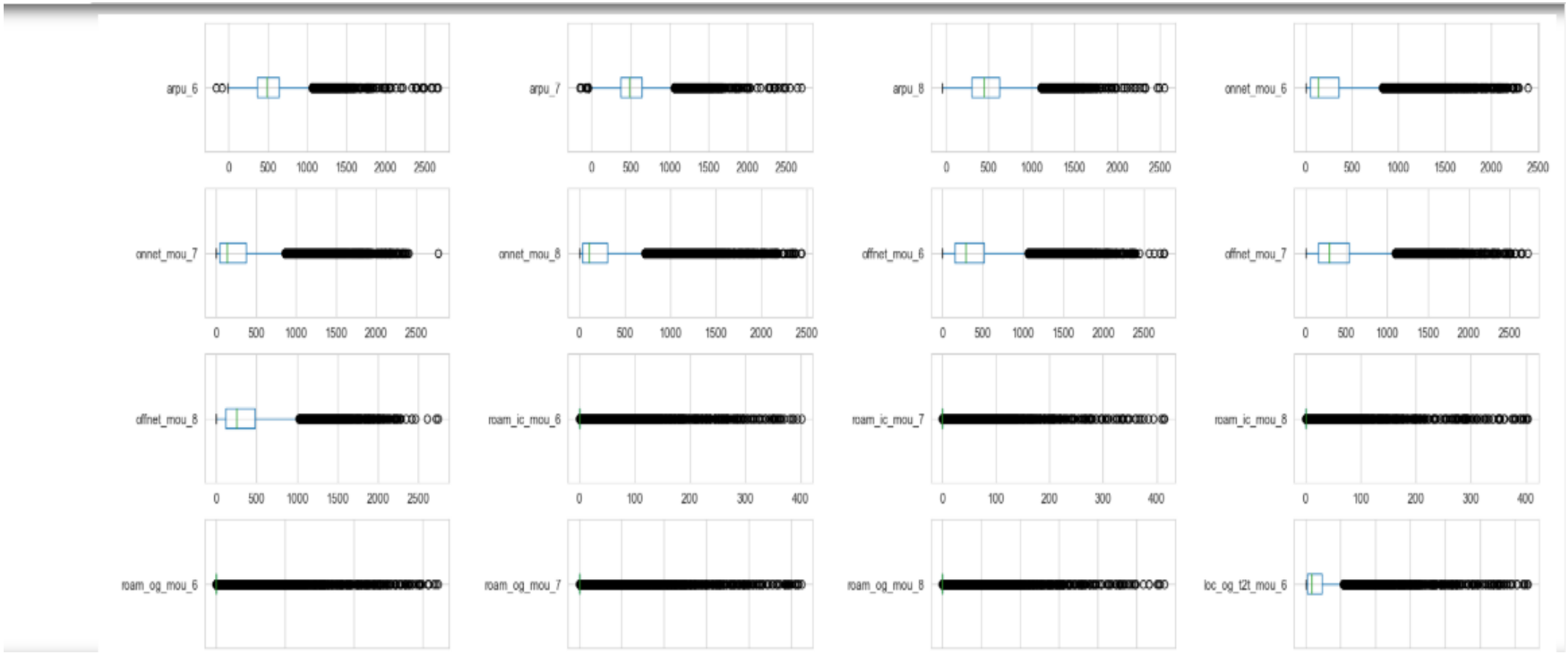
Project Objective

- To Predict Customer Churn
- Highlighting the main variables/factors influencing Customer Churn
- Use various ML algorithms to build prediction models, and evaluate the accuracy and performance of these models.
- Finding the best model for our business case and providing executive suggestions.

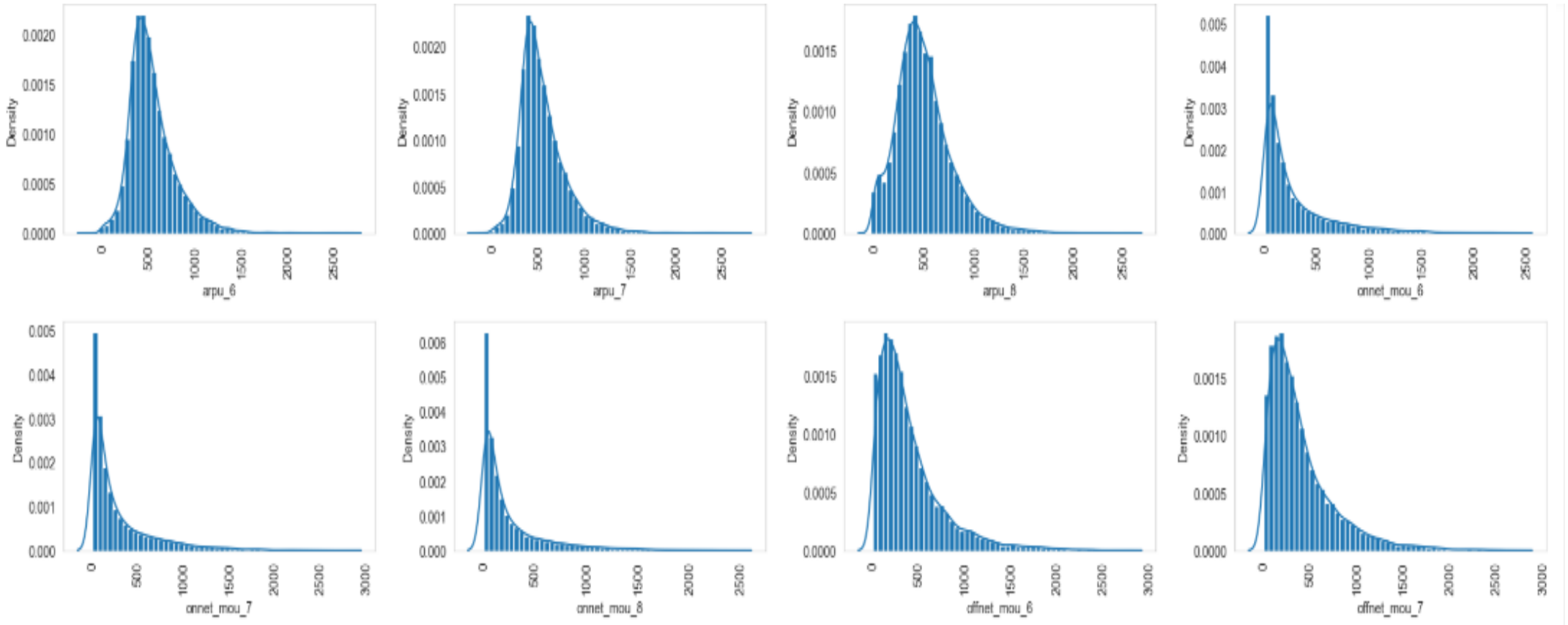
Model Building Steps

- Data collection
- Data preparation
- Perform EDA
- Feature selection
- Building models
- Validate and measure models performance
- Improve models performances
- Executive models for prediction
- Select best fit model for our business problem

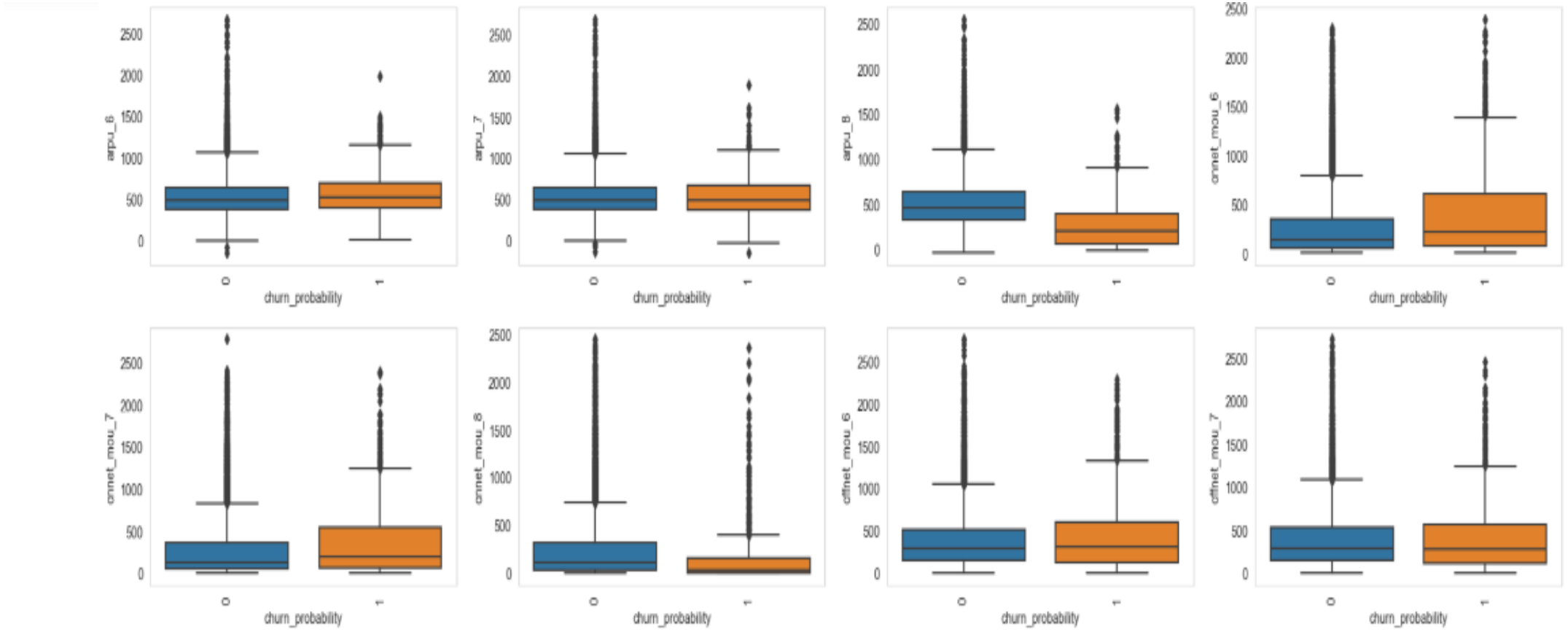
EDA


arpu_6
arpu_7
arpu_8
onnet_mou_6
onnet_mou_7
onnet_mou_8
offnet_mou_6
offnet_mou_7
offnet_mou_8
roam_ic_mou_6
roam_ic_mou_7
roam_ic_mou_8
roam_og_mou_6
roam_og_mou_7
roam_og_mou_8
loc_og_t2t_mou_6
0 500 1000 1500 2000 2500
0 100 200 300 400

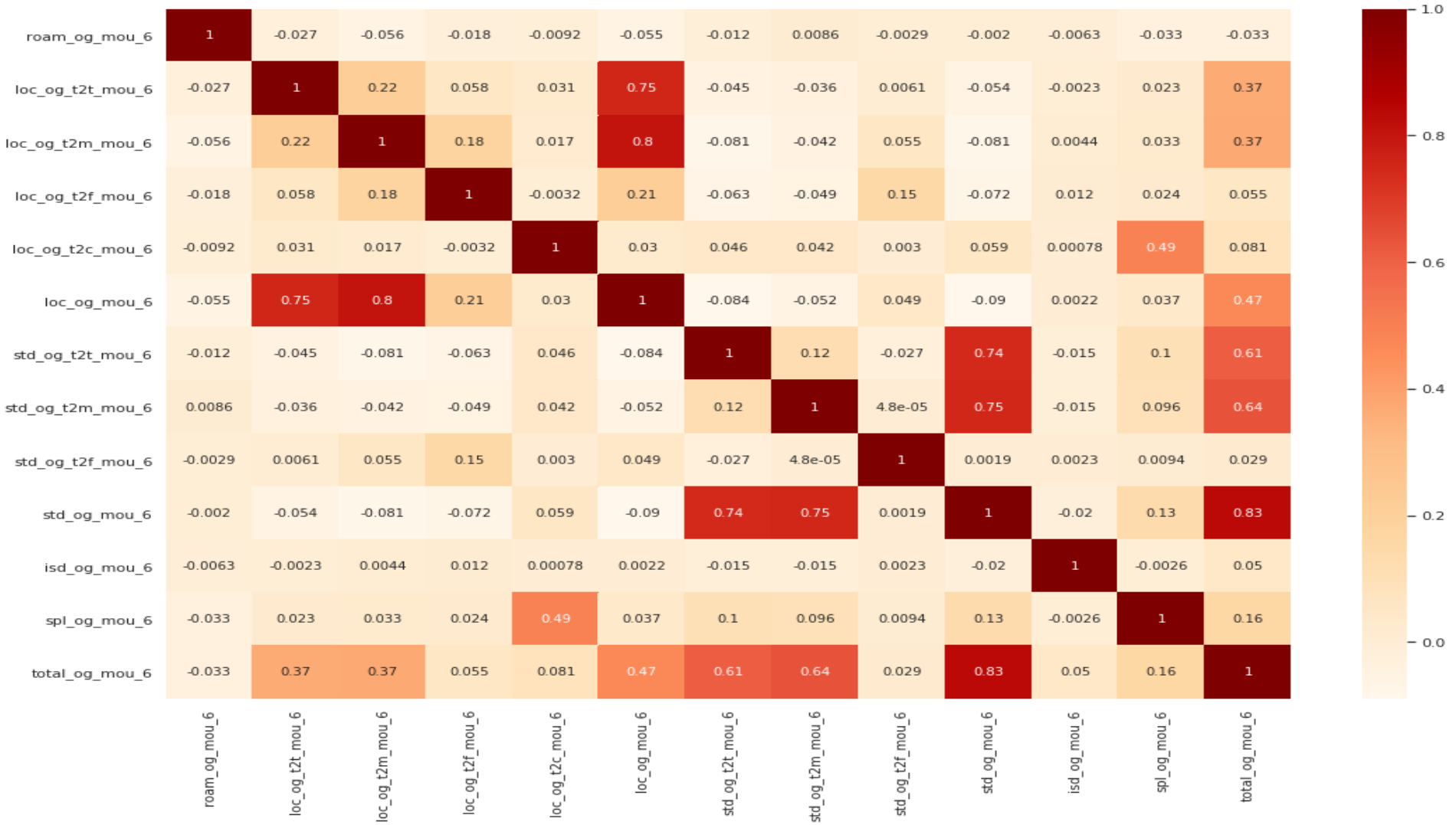
EDA


Density
arpu_6
arpu_7
arpu_8
onnet_mou_6
onnet_mou_7
onnet_mou_8
offnet_mou_6
offnet_mou_7

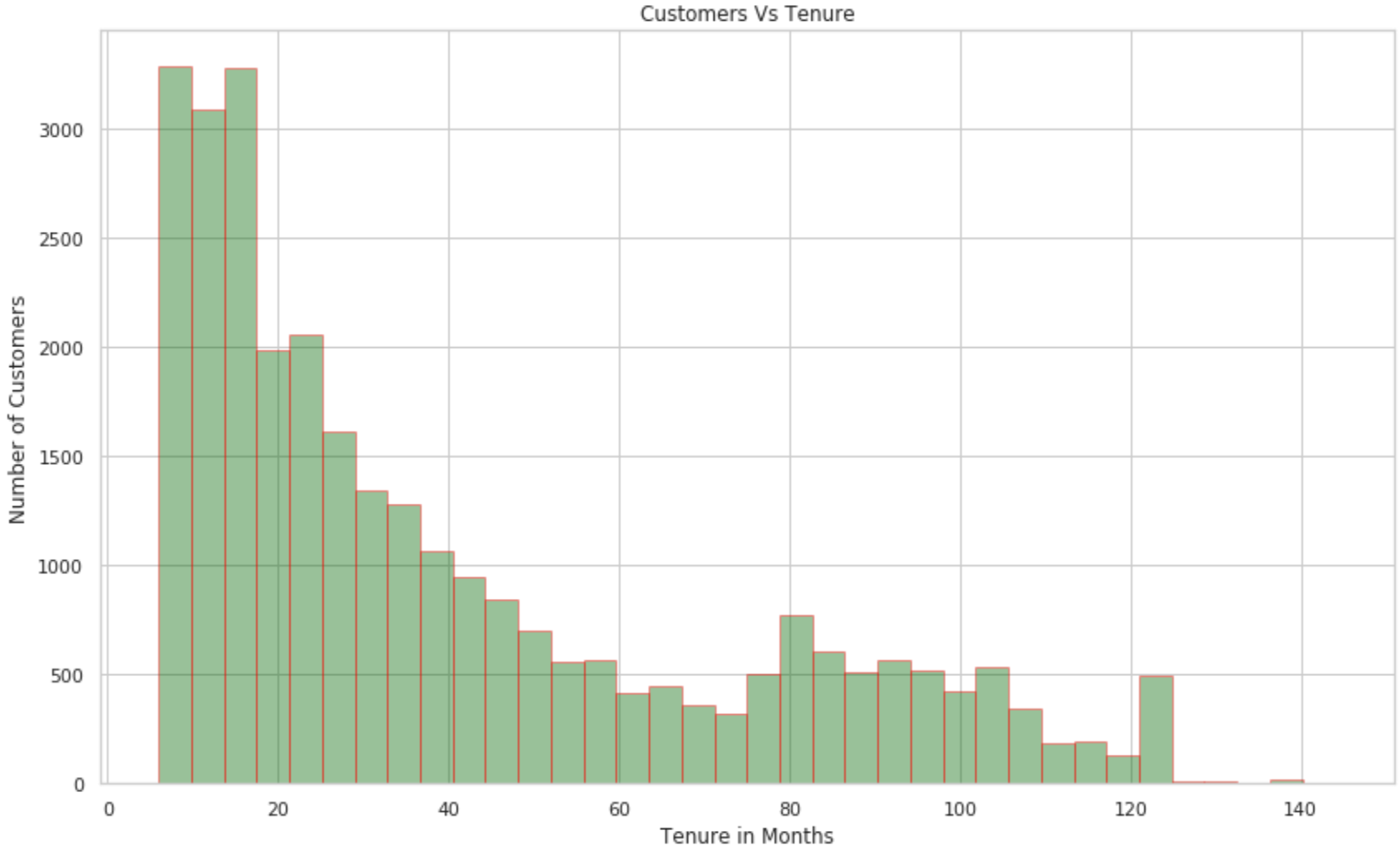
EDA


arpu_6
arpu_7
arpu_8
onnet_mou_6
onnet_mou_7
onnet_mou_8
offnet_mou_6
offnet_mou_7
0
500
1000
1500
2000
2500
churn_probability
0
1

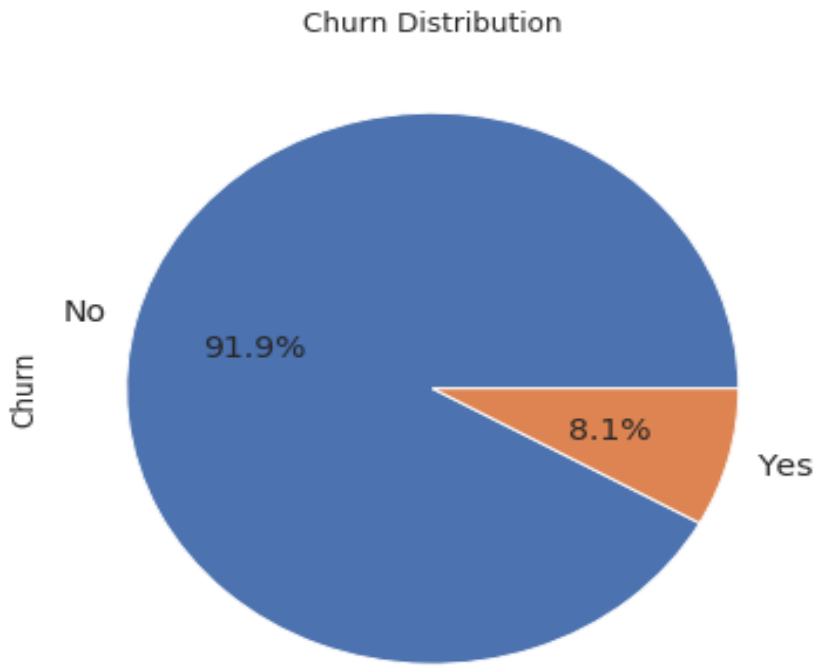
Heat Map


roam_og_mou_6 1 -0.027 -0.056 -0.018 -0.0092 -0.055 -0.012 0.0086 -0.0029 -0.002 -0.0063 -0.033 -0.033
loc_og_t2t_mou_6 -0.027 1 0.22 0.058 0.031 0.75 -0.045 -0.036 0.0061 -0.054 -0.0023 0.023 0.37
loc_og_t2m_mou_6 -0.056 0.22 1 0.18 0.017 0.8 -0.081 -0.042 0.055 -0.081 0.0044 0.033 0.37
loc_og_t2f_mou_6 -0.018 0.058 0.18 1 -0.0032 0.21 -0.063 -0.049 0.15 -0.072 0.012 0.024 0.055
loc_og_t2c_mou_6 -0.0092 0.031 0.017 -0.0032 1 0.03 0.046 0.042 0.003 0.059 0.00078 0.49 0.081
loc_og_mou_6 -0.055 0.75 0.8 0.21 0.03 1 -0.084 -0.052 0.049 -0.09 0.0022 0.037 0.47
std_og_t2t_mou_6 -0.012 -0.045 -0.081 -0.063 0.046 -0.084 1 0.12 -0.027 0.74 -0.015 0.1 0.61
std_og_t2m_mou_6 0.0086 -0.036 -0.042 -0.049 0.042 -0.052 0.12 1 4.8e-05 0.75 -0.015 0.096 0.64
std_og_t2f_mou_6 -0.0029 0.0061 0.055 0.15 0.003 0.049 -0.027 4.8e-05 1 0.0019 0.0023 0.0094 0.029
std_og_mou_6 -0.002 -0.054 -0.081 -0.072 0.059 -0.09 0.74 0.75 0.0019 1 -0.02 0.13 0.83
isd_og_mou_6 -0.0063 -0.0023 0.0044 0.012 0.00078 0.0022 -0.015 -0.015 0.0023 -0.02 1 -0.0026 0.05
spl_og_mou_6 -0.033 0.023 0.033 0.024 0.49 0.037 0.1 0.096 0.0094 0.13 -0.0026 1 0.16
total_og_mou_6 -0.033 0.37 0.37 0.055 0.081 0.47 0.61 0.64 0.029 0.83 0.05 0.16 1
roam_og_mou_6
loc_og_t2t_mou_6
loc_og_t2m_mou_6
loc_og_t2f_mou_6
loc_og_t2c_mou_6
loc_og_mou_6
std_og_t2t_mou_6
std_og_t2m_mou_6
std_og_t2f_mou_6
std_og_mou_6
isd_og_mou_6
spl_og_mou_6
total_og_mou_6
1.0
0.8
0.6
0.4
0.2
0.0

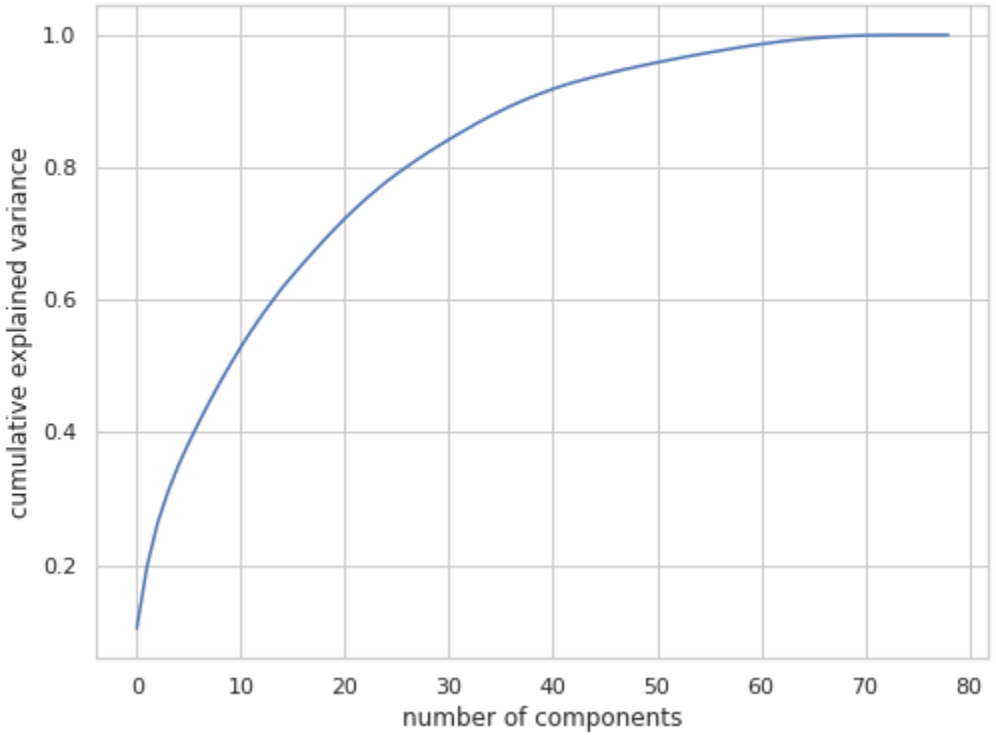
Customer Vs Tenure

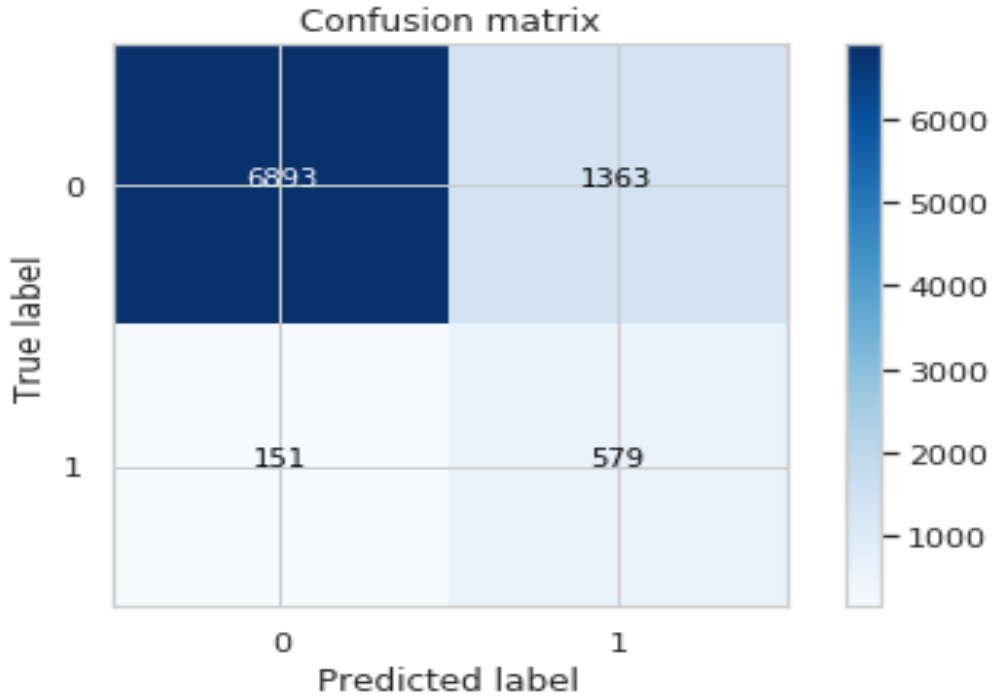


Churn Distribution

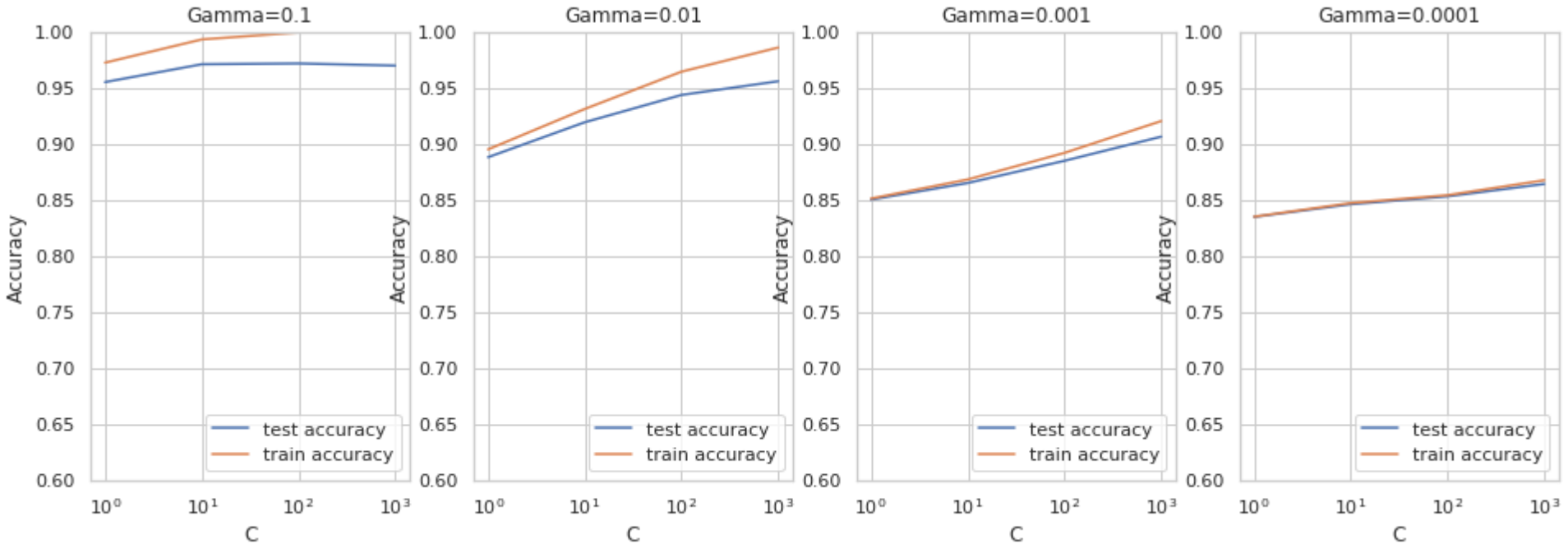


Confusion Matrix

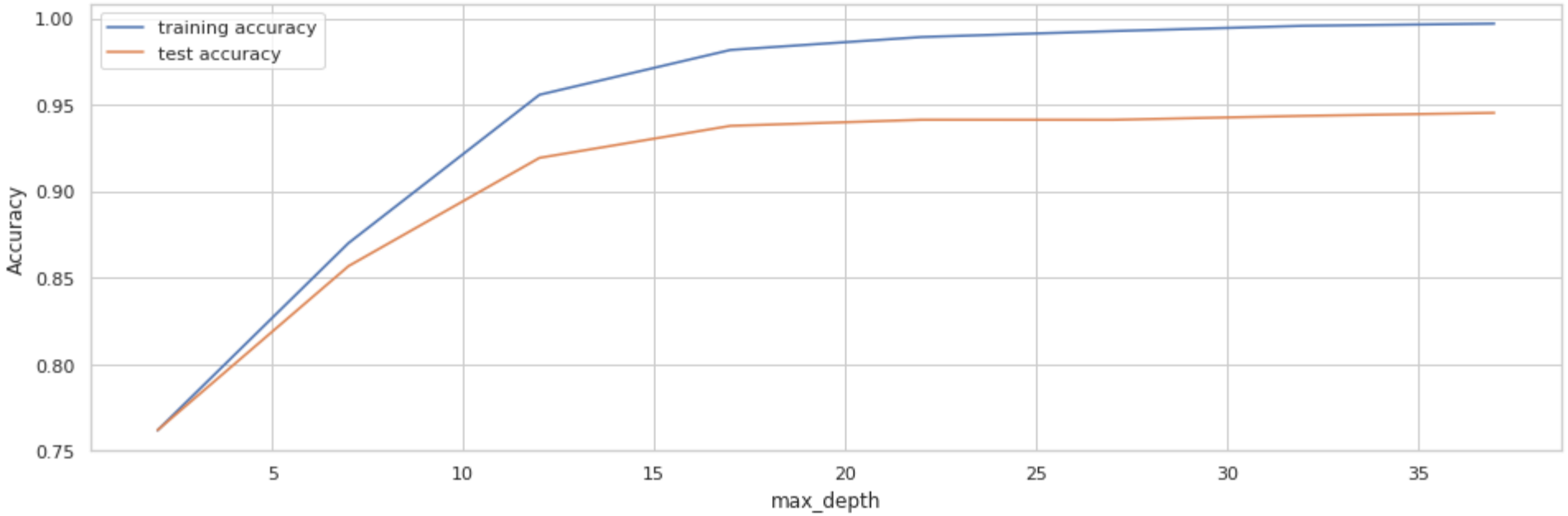




Training & Test Accuracy



Training & Test Accuracy



Conclusion

SVM with tuned hyperparameters produce best result on this dataset with 0.92 accuracy.

Random forest also produce good accuracy with 0.91 (default overfit model) and 0.90 with tuned hyperparameters.

XGBoost also produce apt accuracy of 0.86 (default overfit model) and 0.85 with tuned hyperparameters.

As per our analysis SVM and Random forest produce best accuracy and models can be selected to predict churn data for future dataset or production.

Thank You