Telecom Churn Case Study

Submitted by

Subhranil Sengupta

Steve Thomas

Suchitra Gupta

- In telecom industry, customers are able to choose multiple service providers and actively switch from one operator to another. In this highly competitive market, telecommunications industries experiences an average of 15-25% churn rate. Given the fact it costs 5-20times more to acquire a new customer than to retain an existing one, customer retention has now become more important than customer acquisition.

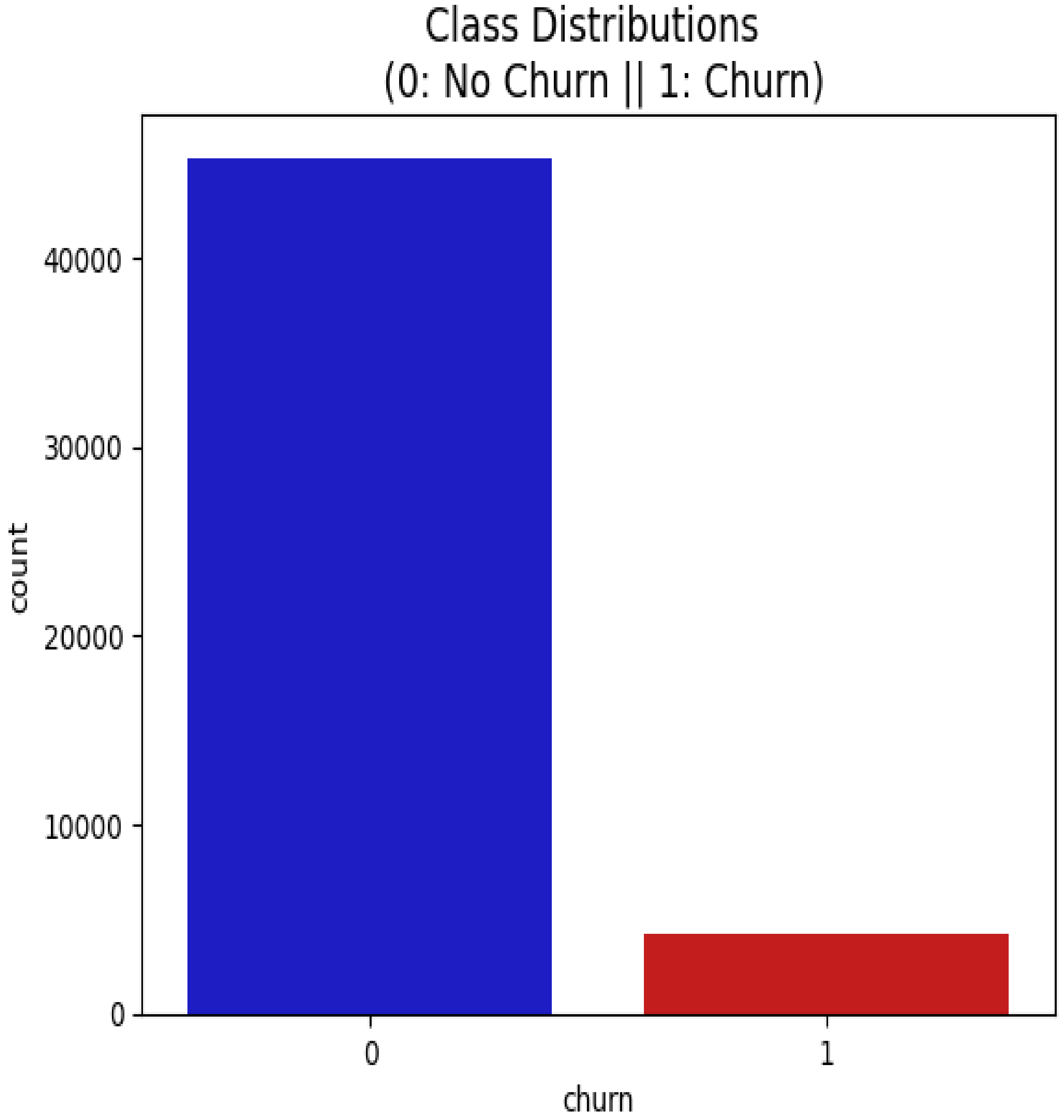
- For many incumbent operators, retaining high profitable customers is the major business goal.

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

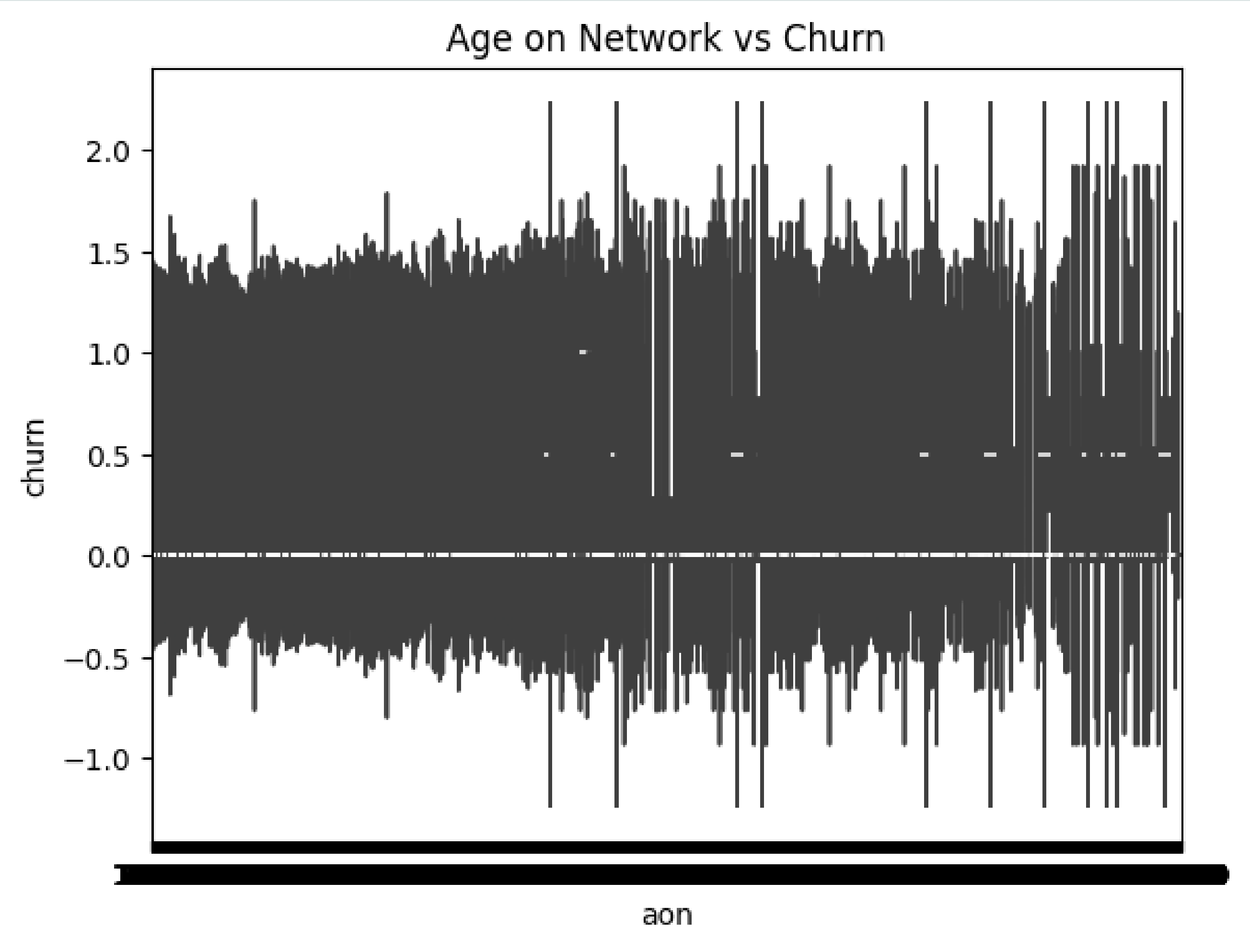
- In this project we will analyse customer-level-data of telecom company, build predictive models to identify customers at high risk of churns and identify main risk of churns.

Project Objective

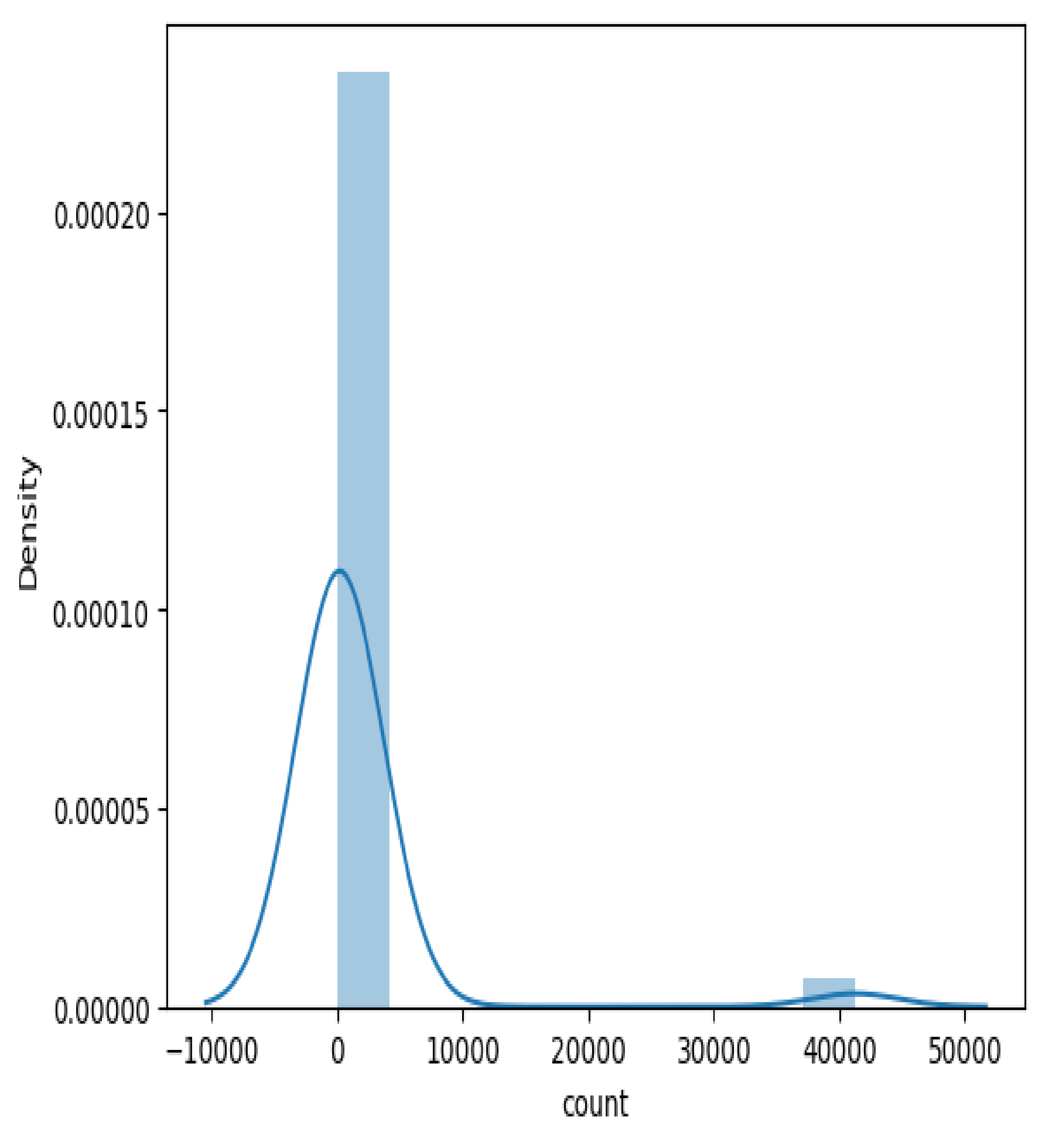
- To predict customer churn.
- Highlighting main factors influencing customer churn.
- Use various ML algorithms to build predictive models and performance of these models with accuracy.
- Finding out best model for our business case and providing executive suggestions.



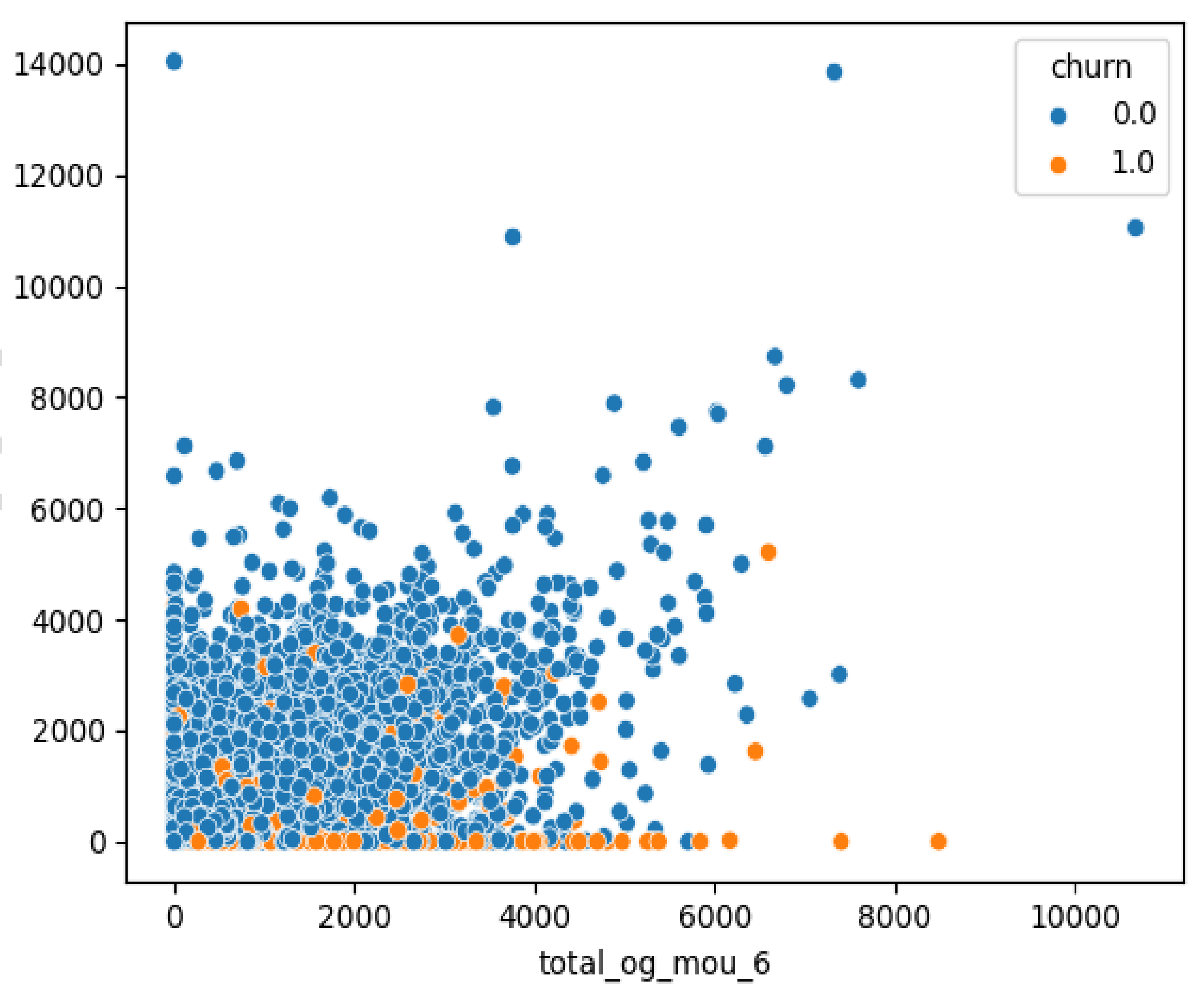
From the image we could gather that No churn customers are ranking high as compared to churn customers



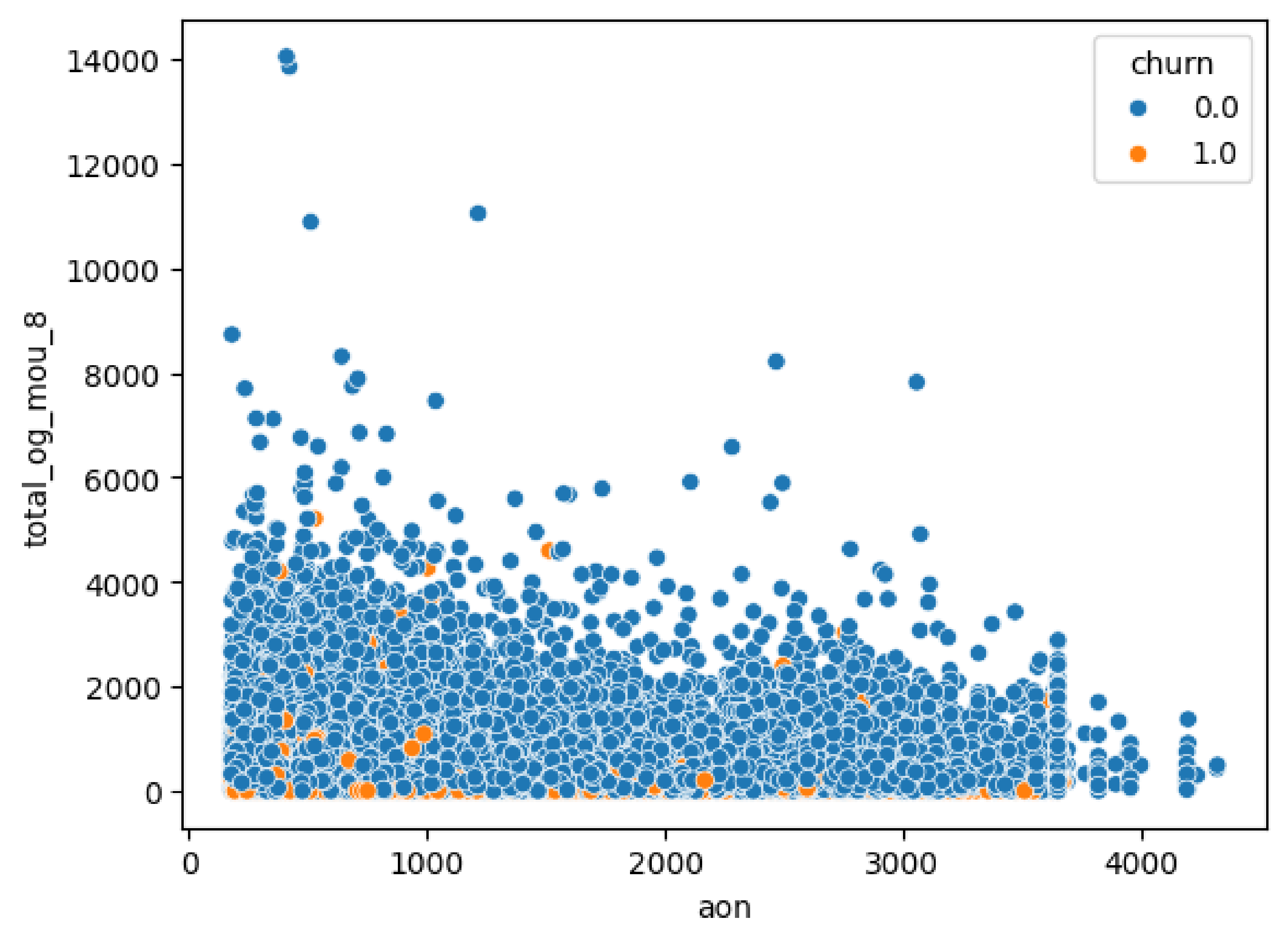
The customer with lesser "aon" are likely to churn as compared to customer with high "aon"



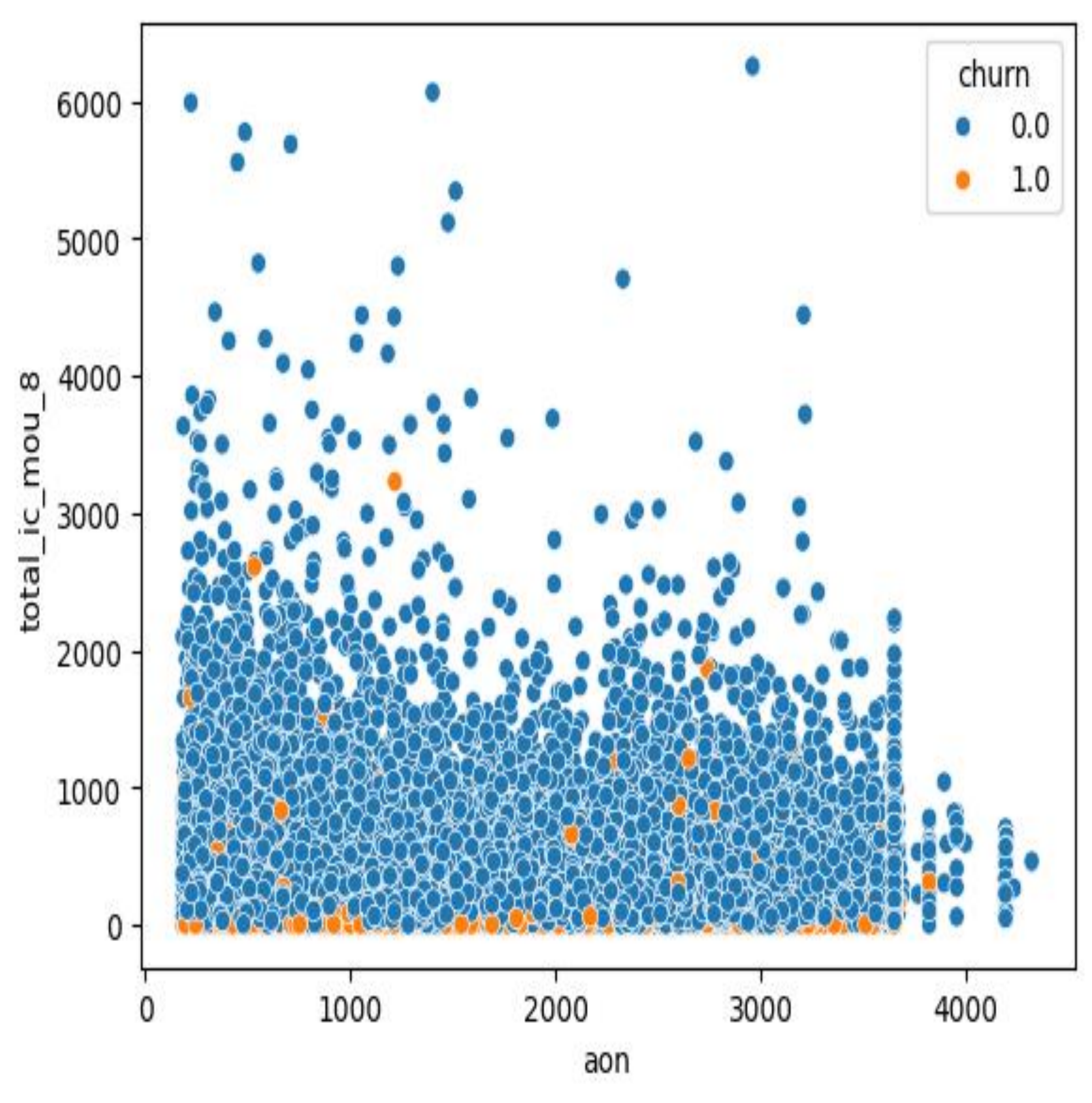
As we can see the that it is not skewed but highly imbalanced. The no.of non churners are very high.



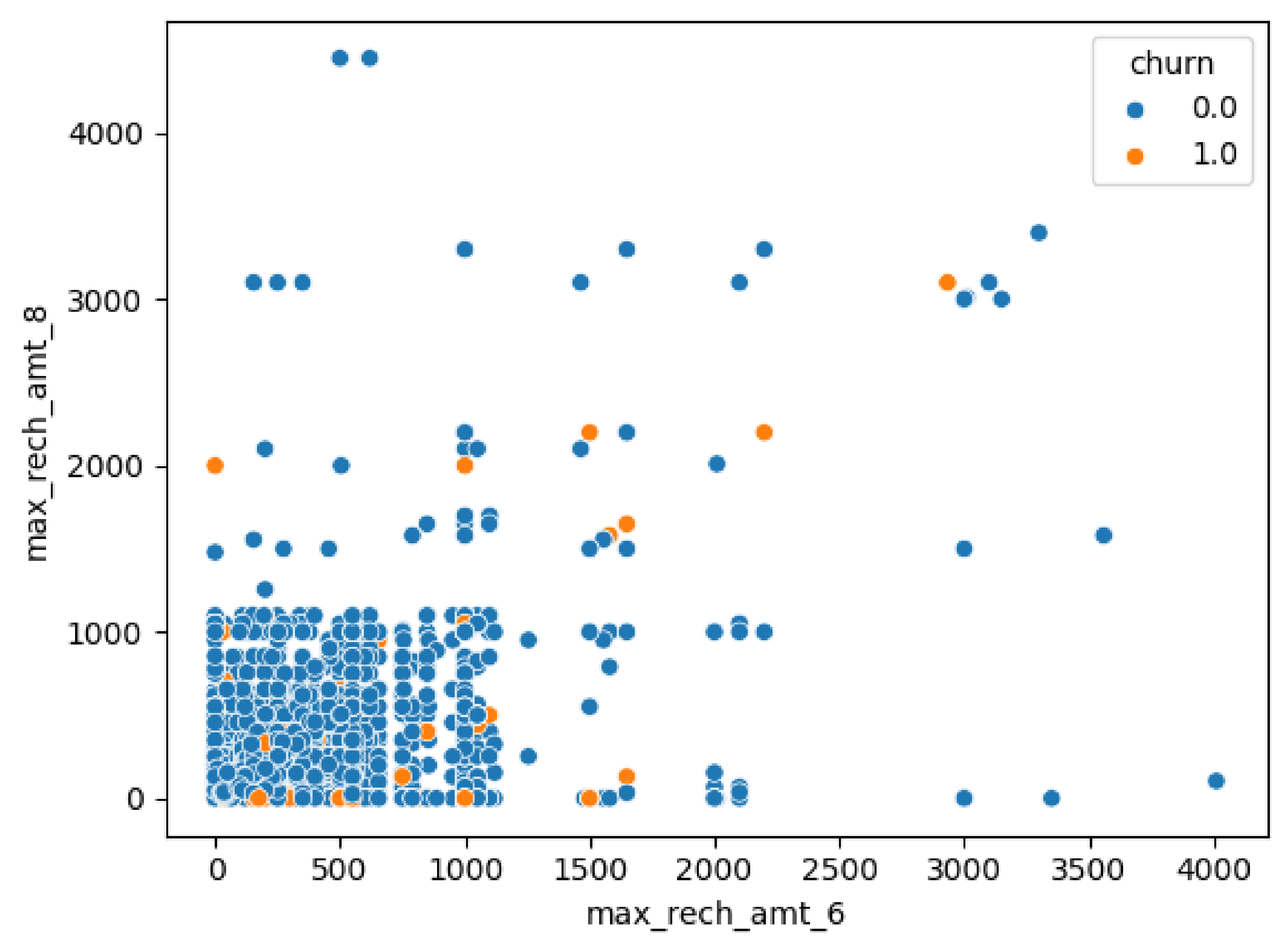
As we can see the customer with lower "total_og_mou" of 6th and 8th month are likely to churn.



The customer with lesser "total_og_mou_8" and "aon" are likely to churn.



- *The customers with less "total_ic_mou_8" are likely to churn irrespective of "aon".*
- *Customers with "total_ic_mou_8">2000 are very less likely to churn.*

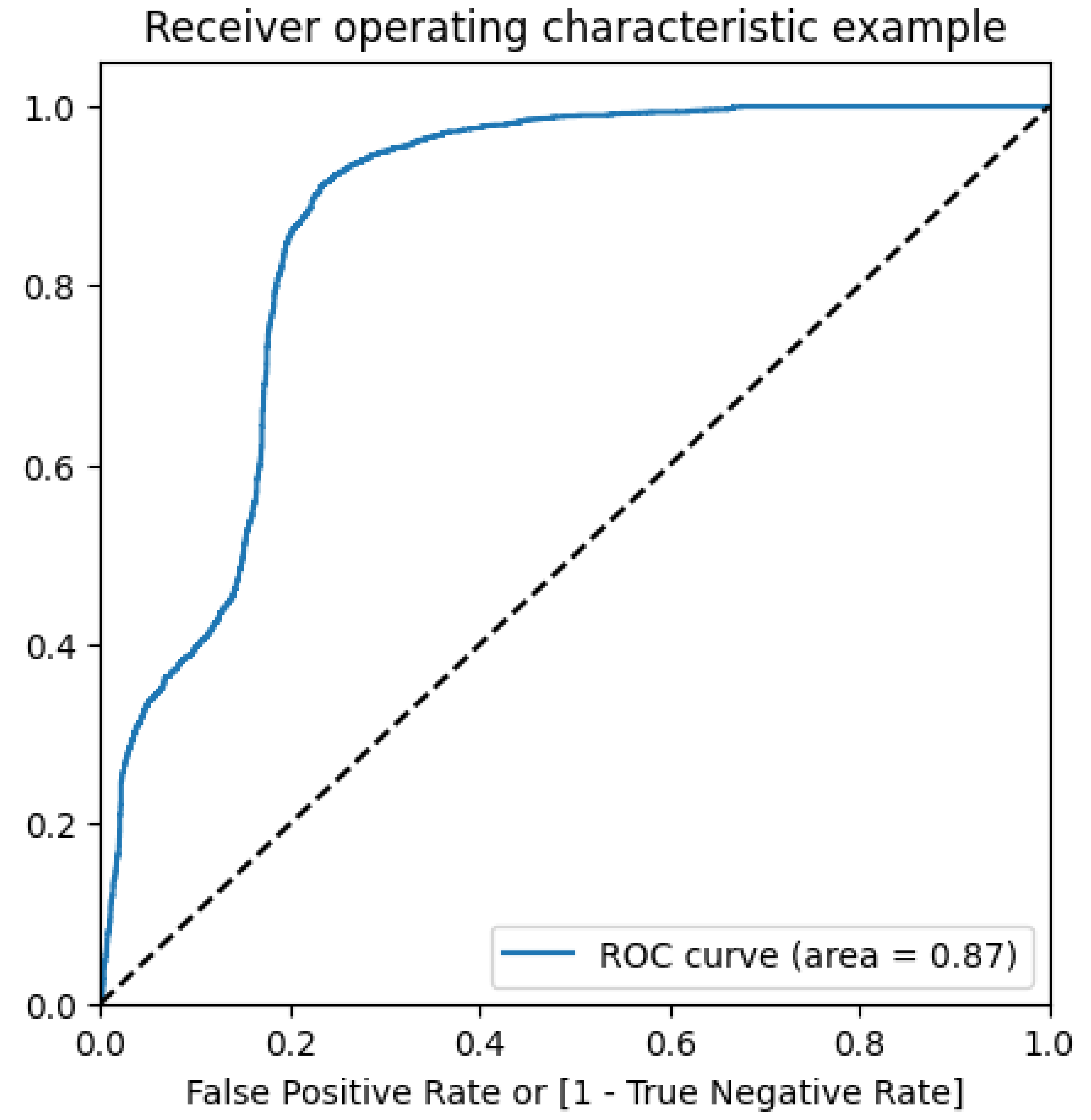


The customers with "max_rech_amt_8 and 6" >1000 are less likely to churn.

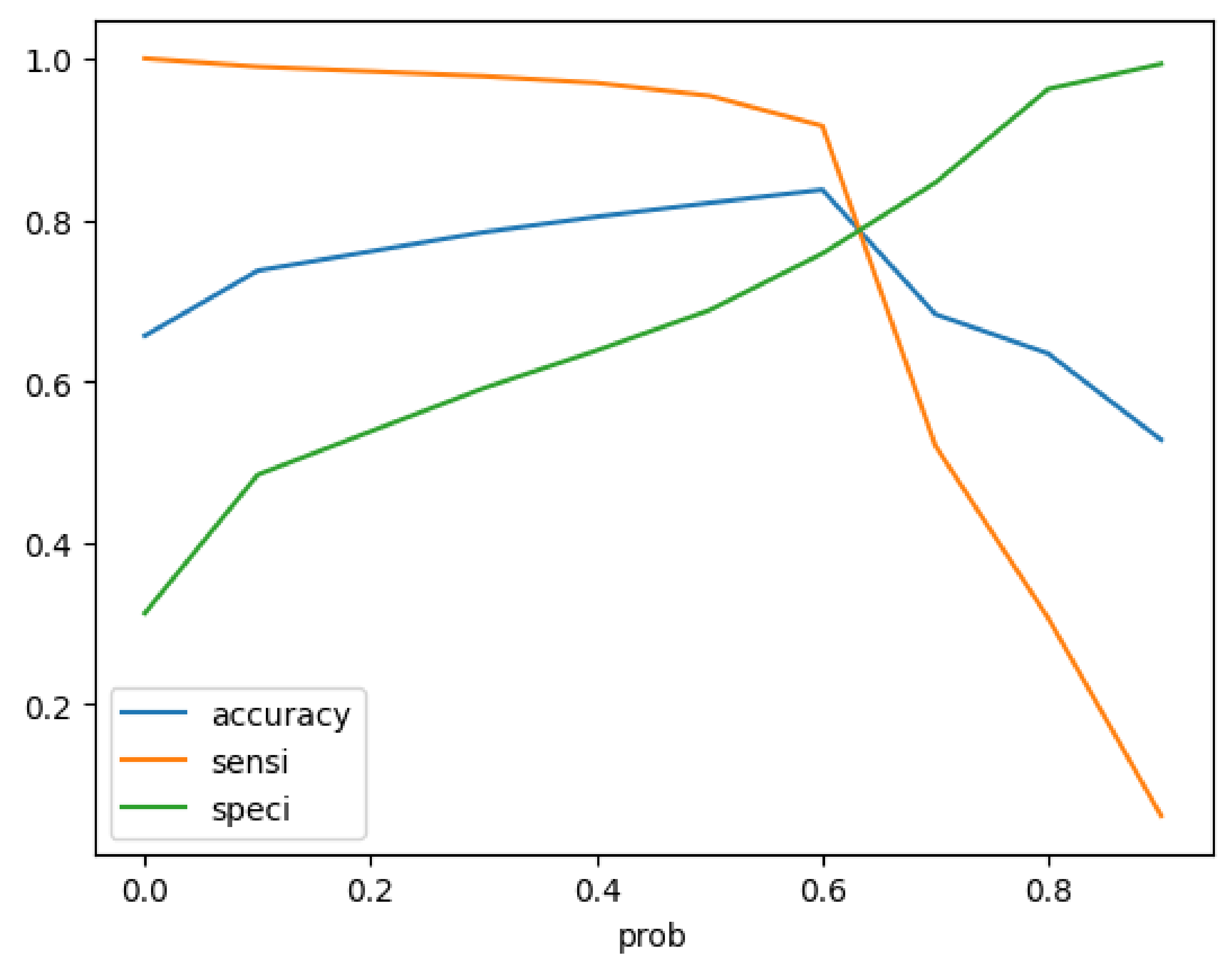
```python
x_train_sm = sm.add_constant(x_train[col])
logm2 = sm.GLM(y_train, x_train_sm, family = sm.families.Binomial())
res = logm2.fit()
print(res.summary())
```

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                  churn   No. Observations:                 6272
Model:                            GLM   Df Residuals:                     6245
Model Family:                Binomial   Df Model:                           26
Link Function:                  Logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                    nan
Date:                Sun, 07 Apr 2024   Deviance:                       5429.9
Time:                        11:03:12   Pearson chi2:                 7.25e+07
No. Iterations:                   100   Pseudo R-squ. (CS):                nan
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------
const                 -0.7956      0.257     -3.101      0.002      -1.298      -0.293
roam_og_mou_6          0.0008      0.000      2.284      0.022       0.000       0.001
roam_og_mou_8          0.0051      0.001      7.579      0.000       0.004       0.006
loc_og_t2t_mou_6    4.043e-05      0.000      0.222      0.825      -0.000       0.000
loc_og_t2f_mou_8      -0.0104      0.004     -2.478      0.013      -0.019      -0.002
loc_og_t2c_mou_6      -0.0021      0.007     -0.320      0.749      -0.015       0.011
loc_og_t2c_mou_7       0.0088      0.006      1.457      0.145      -0.003       0.021
std_og_t2f_mou_6       0.0024      0.005      0.518      0.604      -0.007       0.012
std_og_t2f_mou_8      -0.0131      0.007     -1.915      0.055      -0.026       0.000
isd_og_mou_6           0.0241      0.009      2.831      0.005       0.007       0.041
isd_og_mou_7           0.0233      0.012      1.989      0.047       0.000       0.046
isd_og_mou_8          -0.0061      0.004     -1.369      0.171      -0.015       0.003
spl_og_mou_9          -0.0322      0.006     -5.099      0.000      -0.045      -0.020
loc_ic_t2m_mou_9      -0.0047      0.000     -9.905      0.000      -0.006      -0.004
loc_ic_t2f_mou_8      -0.0056      0.002     -3.402      0.001      -0.009      -0.002
std_ic_t2t_mou_8      -0.0039      0.001     -2.639      0.008      -0.007      -0.001
std_ic_t2f_mou_8      -0.0080      0.005     -1.563      0.118      -0.018       0.002
isd_ic_mou_7           0.0006      0.001      0.693      0.489      -0.001       0.002
```

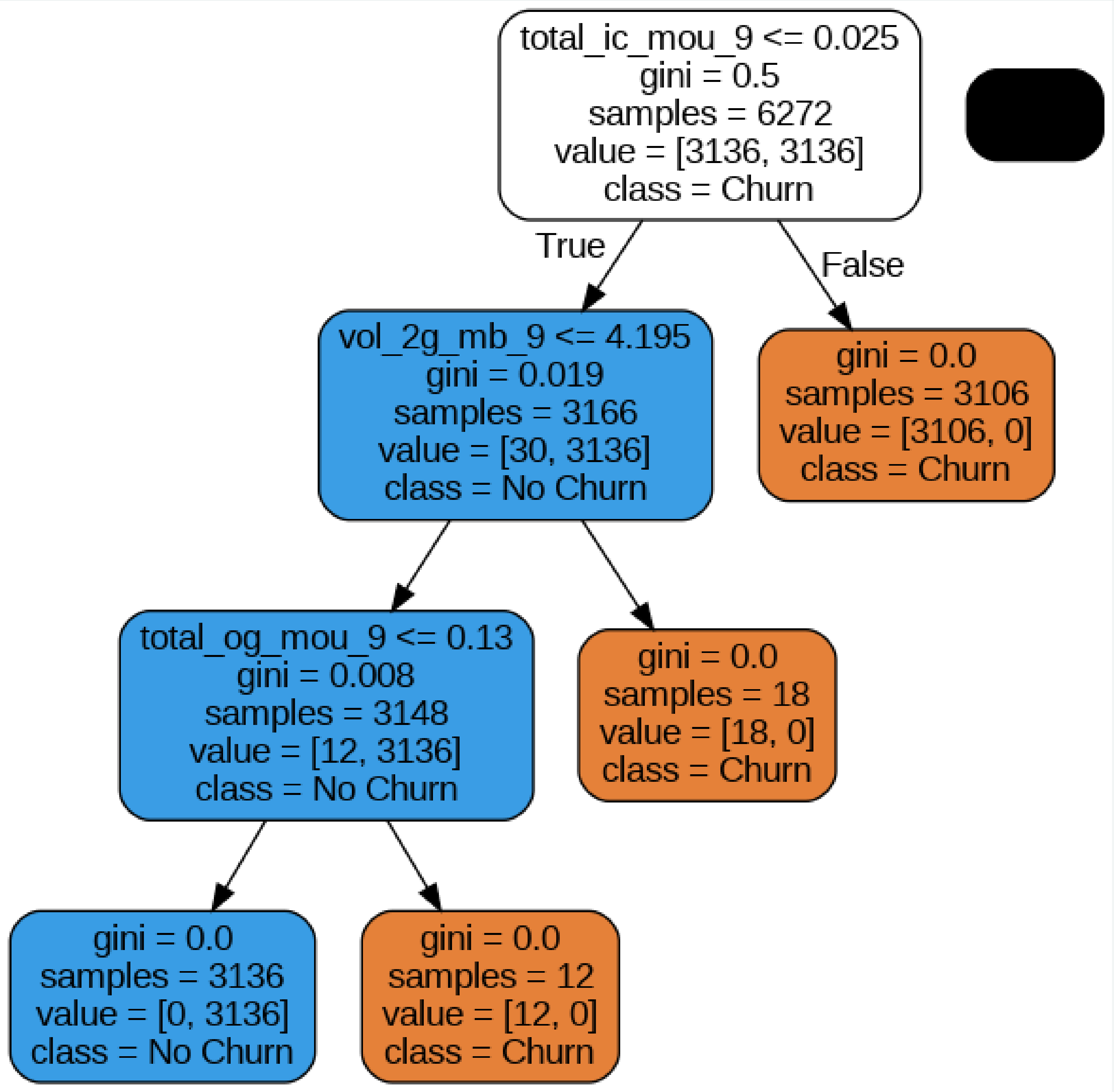
Final Logistic Model where p-value for all predictive variables coming down to zero



Plotting the ROC Curve



Optimal cutoff is 0.55 so we can keep it.



With decision Tree we are getting 90% accuracy

Important Features

- **total_ic_mou_9: Total incoming call minutes of usage in September (9th month).**
- **total_og_mou_9: Total outgoing call minutes of usage in September (9th month).**
- **arpu_9: Average revenue per user in September (9th month).**
- **total_rech_amt_9: Total recharge amount in September (9th month).**
- **roam_og_mou_9: Roaming outgoing call minutes of usage in September (9th month).**
- **og_others_9: Outgoing call minutes to other operators in September (9th month).**
- **std_ic_t2t_mou_9: Standard incoming call minutes of usage within the same operator network in September (9th month).**
- **loc_ic_t2m_mou_9: Local incoming call minutes of usage from other operators in September (9th month).**
- **roam_ic_mou_9: Roaming incoming call minutes of usage in September (9th month).**
- **isd_og_mou_9: ISD outgoing call minutes of usage in September (9th month).**

Important Features

- max_rech_amt_9: Maximum recharge amount in September (9th month).
- std_og_t2m_mou_9: Standard outgoing call minutes of usage within the same operator network in September (9th month).
- loc_og_t2t_mou_9: Local outgoing call minutes of usage within the same operator network in September (9th month).
- loc_ic_mou_9: Local incoming call minutes of usage in September (9th month).
- isd_ic_mou_9: ISD incoming call minutes of usage in September (9th month).
- arpu_8: Average revenue per user in August (8th month).
- total_og_mou_8: Total outgoing call minutes of usage in August (8th month).
- loc_og_mou_9: Local outgoing call minutes of usage in September (9th month).
- loc_ic_t2t_mou_9: Local incoming call minutes of usage within the same operator network in September (9th month).
- roam_og_mou_8: Roaming outgoing call minutes of usage in August (8th month).

Recommendation: Positive Impact

- **sachet_2g_7:** Customers who have availed service schemes with validity smaller than a month for 2G network in July (7th month) are more likely to churn.
- **count_rech_2g_8:** The count of recharges for 2G network in August (8th month) positively influences churn prediction. More recharges might indicate dissatisfaction or uncertainty leading to churn.
- **isd_og_mou_6 and isd_og_mou_7:** ISD outgoing minutes of usage in June (6th month) and July (7th month) respectively contribute positively to churn prediction. Customers making more ISD calls might have higher churn rates.
- **arpu_3g_9:** Average revenue per user for 3G network in September (9th month) positively influences churn prediction.

Recommendation: Negative Impact

- **vol_3g_mb_9 and and vol_2g_mb_9:** Volume of mobile internet usage in MB for 3G and 2G networks respectively in September (9th month) have a strong negative impact on churn prediction. Higher usage might indicate customer engagement and satisfaction, thus reducing churn likelihood.
- **spl_og_mou_9:** Special outgoing call minutes of usage in September (9th month). Lower usage of these special calls is correlated with higher churn probability.
- **last_day_rch_amt_9:** The recharge amount on the last day of September (9th month). A higher recharge amount on the last day is associated with a lower probability of churn.

Suggestions to handle Churn customers

- Churners should show higher roaming usage.
- Network operators should investigate their roaming tariffs and quality of service.
- New campaigns that target roaming customers can be rolled out.
- Like
- Discounted roaming rates during particular hours of day.
- Free monthly roaming on minutes of usage of voice calls depending on users past roaming usage history.