Telecom Churn Case Study

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.

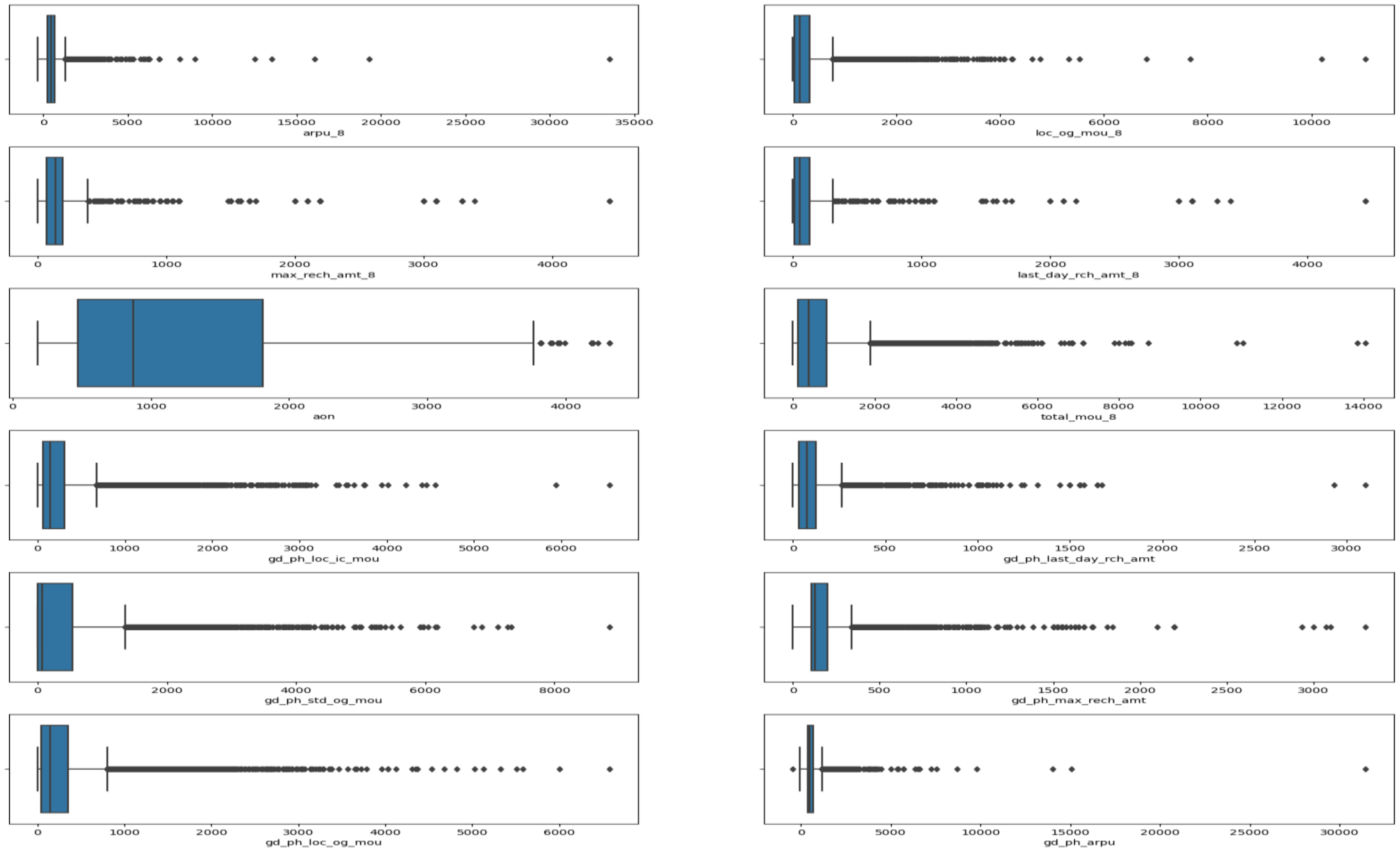
Project Objective

- ❖ To predict Customer Churn
- ❖ Highlighting the main variables/factors influencing Customer Churn
- ❖ Use various ML algorithms to build prediction models, evaluatethe accuracy and performance of these models.
- ❖ Finding out 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

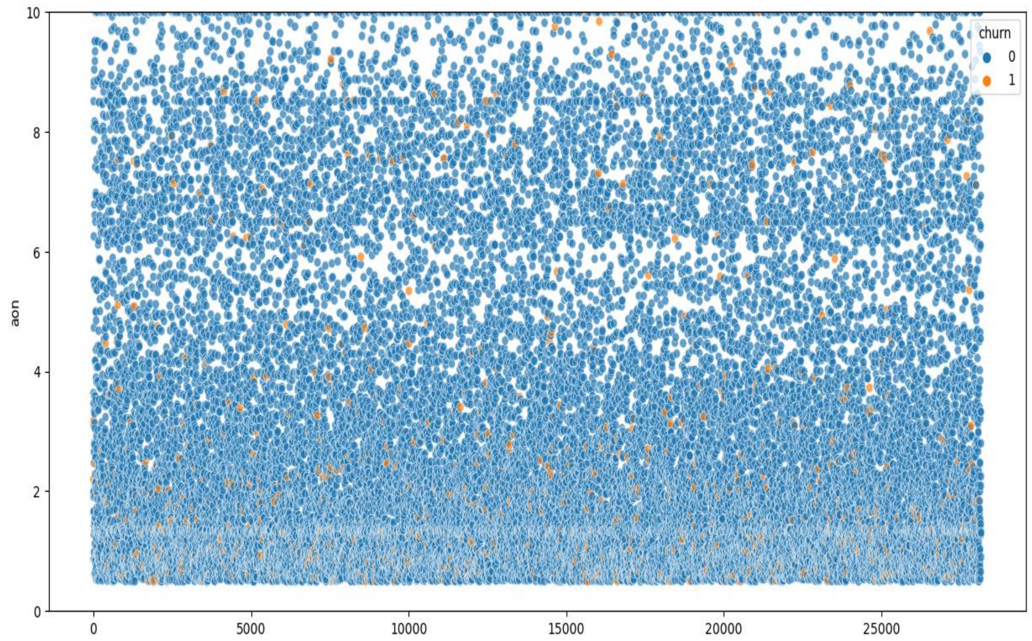
EDA



From plots we can define following upper limits to the variables

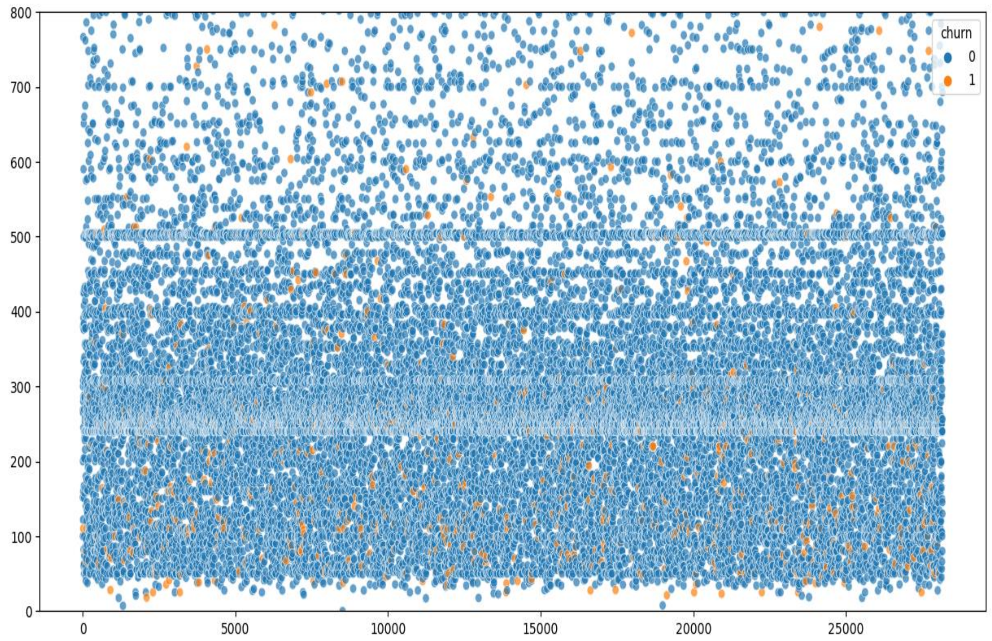
Feature	Value
arpu_8	7000
loc_og_mou_8	4000
max_rech_amt_8	1000
last_day_rch_amt_8	1000
aon	3000
total_mou_8	4000
gd_ph_loc_ic_mou	3000
gd_ph_last_day_rch_amt	1000
gd_ph_std_og_mou	4000
gd_ph_max_rech_amt	1500
gd_ph_loc_og_mou	3000
gd_ph_arpu	7000

Churn based on tenure



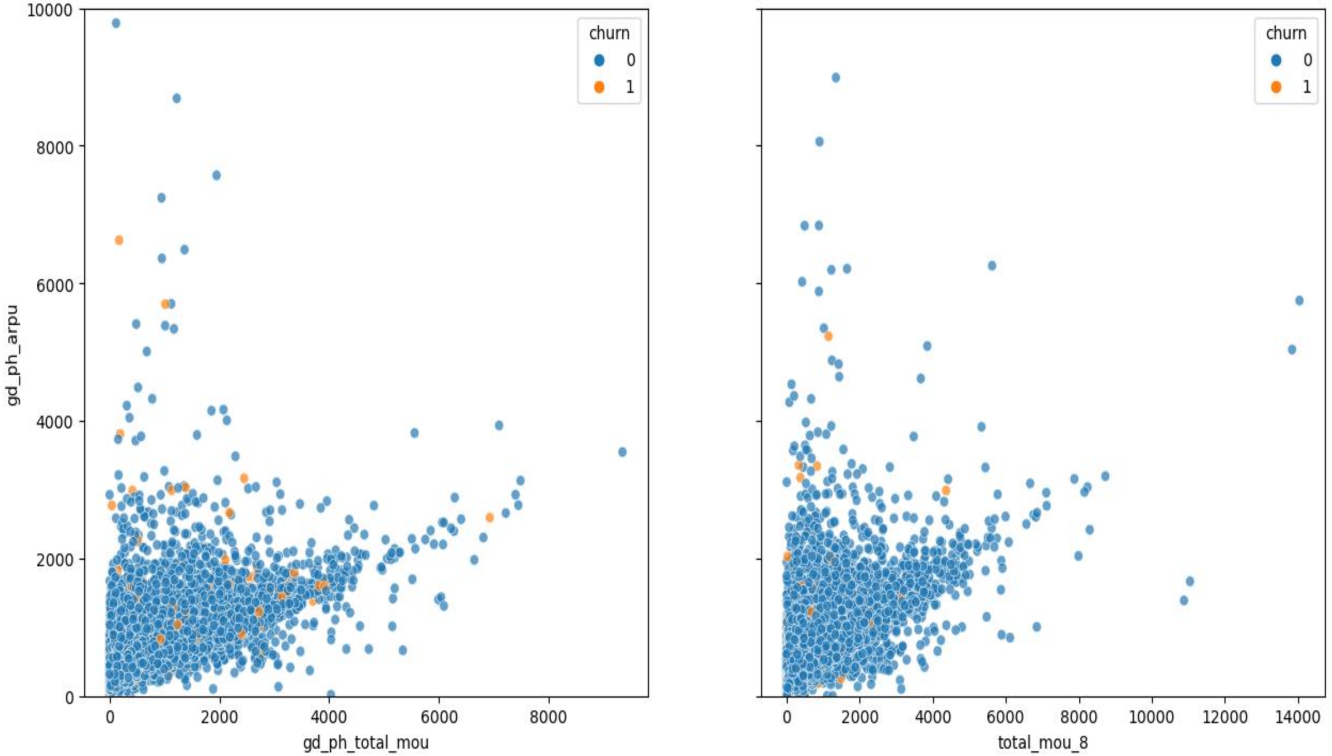
Insights:
As we can see that most of the churners have a tenure less than 4 years

Effect of max recharge amount on churn



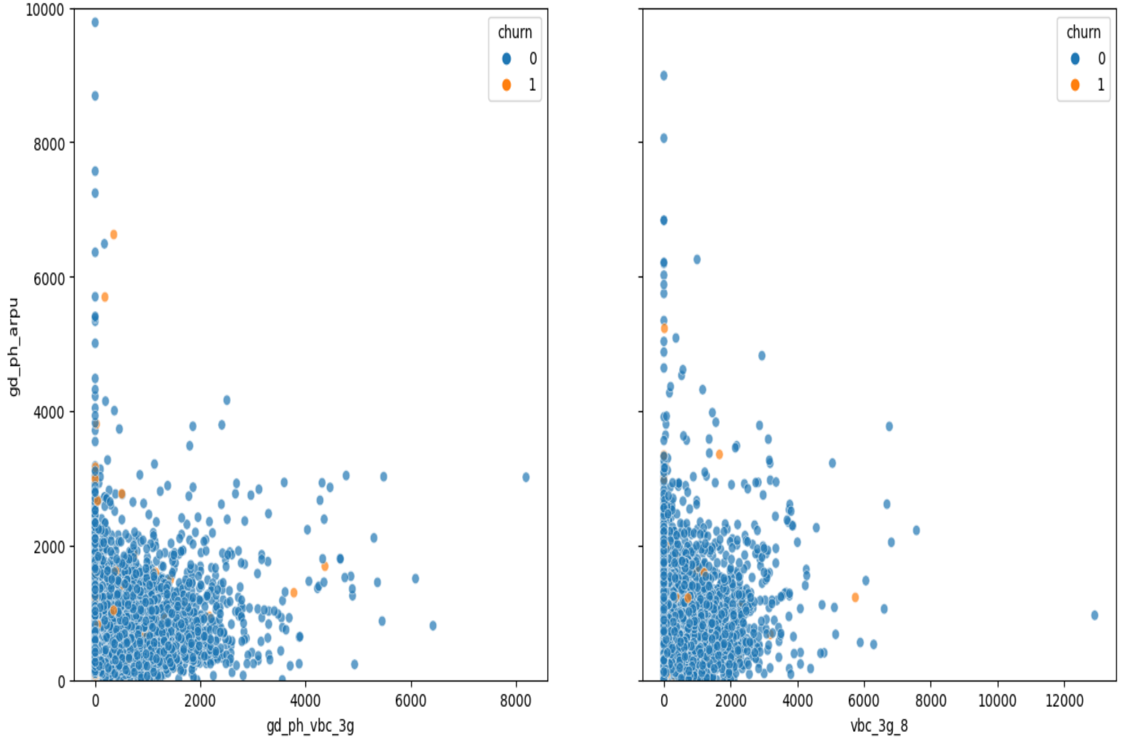
Insights:
As we can observe users having the max recharge amount less tha 250 churned more.

VBC effects on revenue



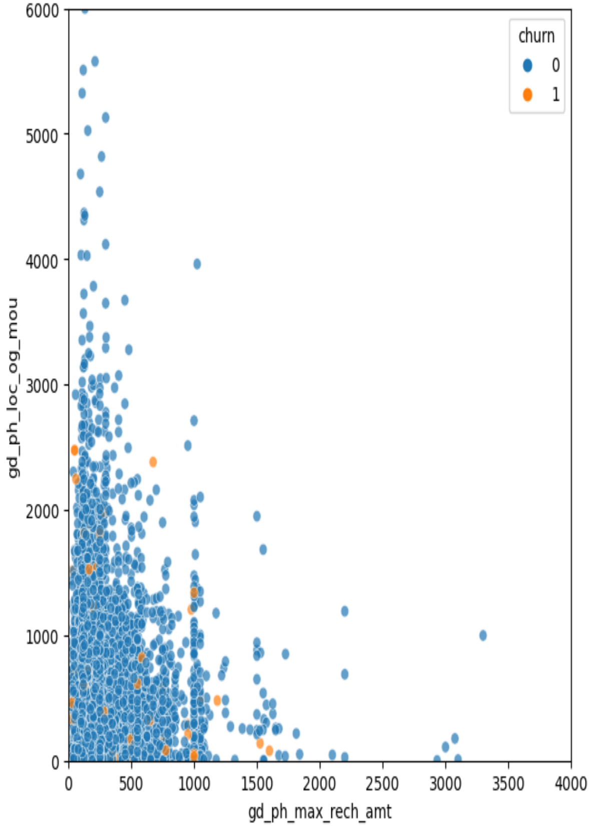
Insights:
As we can observe that MOU is dropping significantly for churners in action phase which hitting the revenue generation. But then also revenue is higher in that part which indicates that the users are taking other services which increasing the revenue generation.

total_mou effects on revenue



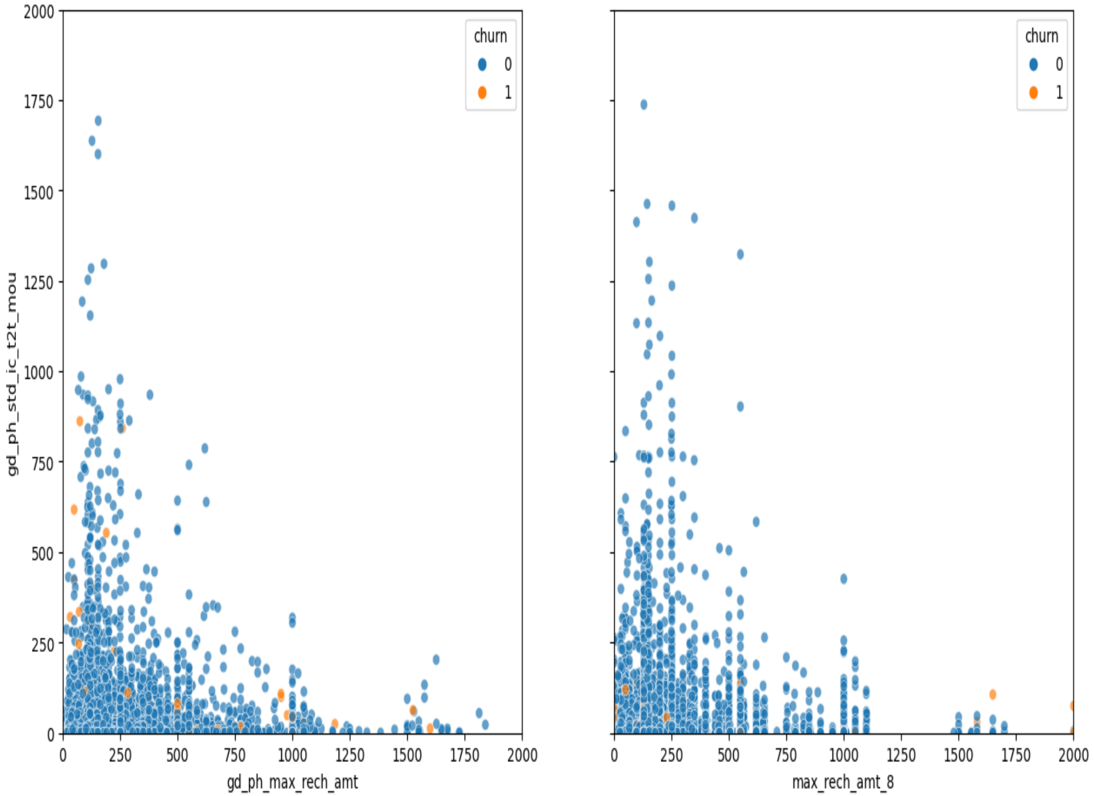
Insights:
As we can see users using less amount of VBC generating high revenue churned and also revenue is higher from less consumption part.

Recharge amount vs local outgoing calls



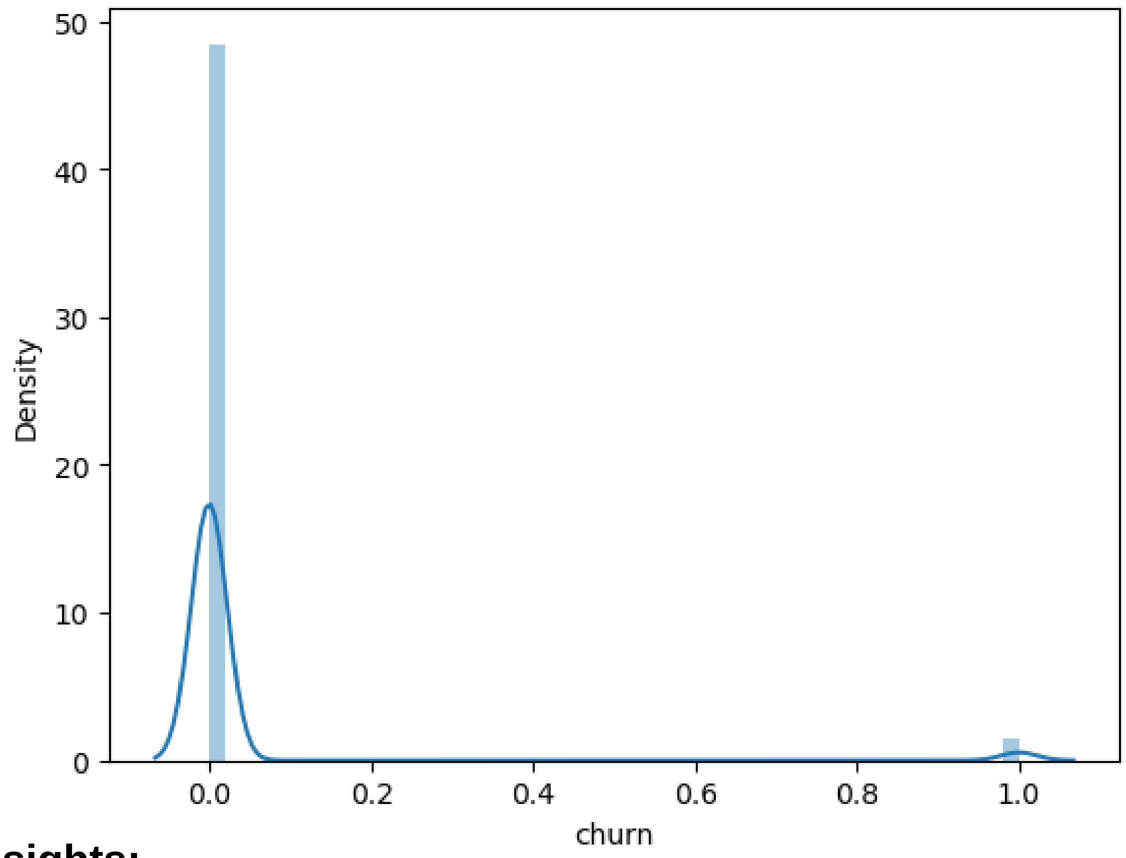
Insights:
As we can see users recharging with high amounts using less local services in compare to users recharging with less amount. And users having max recharge amount as well as local out going were very less even in the good phase churned more.

Same service provider vs the recharge amount



Insights:
As we can observe users having max recharge amount on the higher end and low incoming call mou during the good pahse churned more

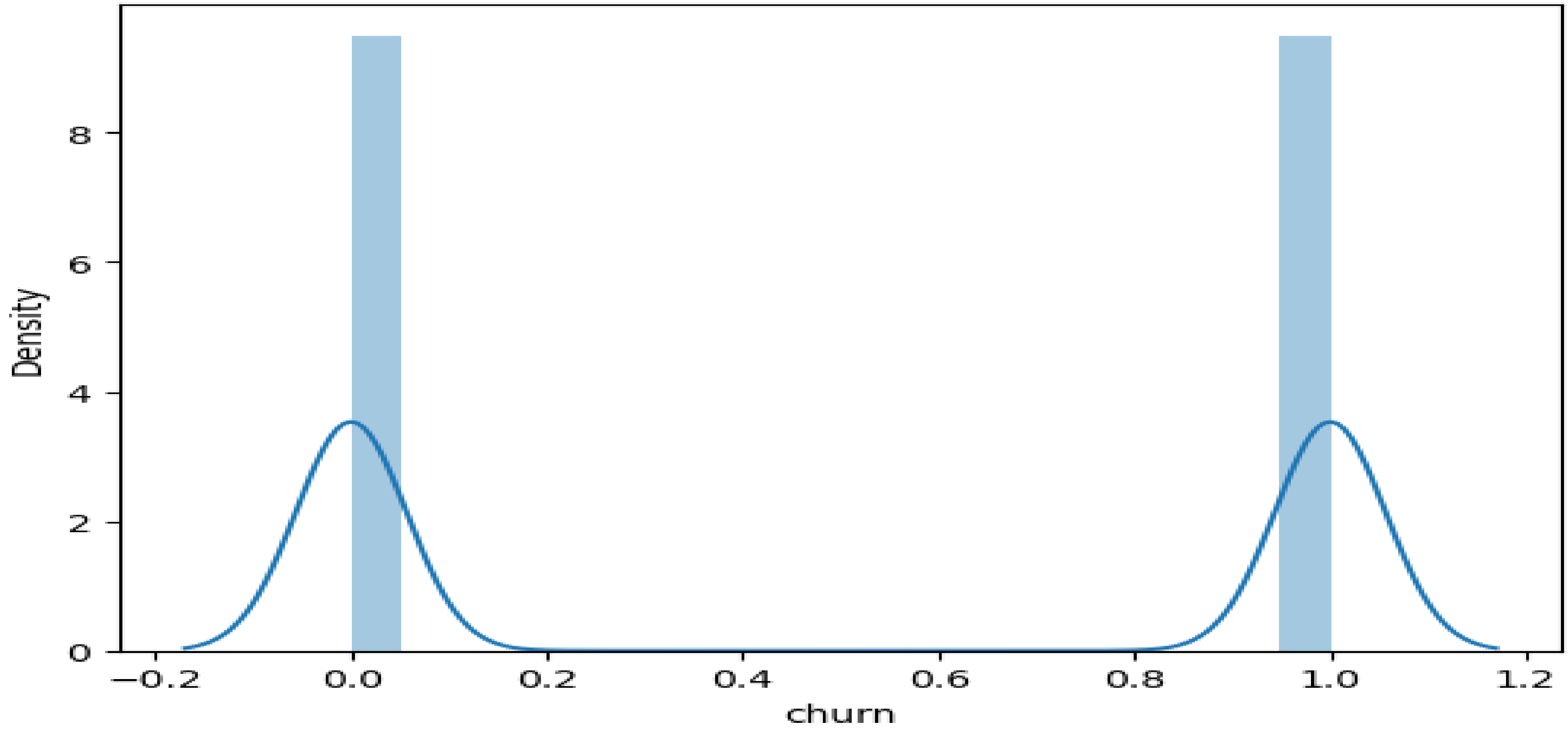
Distribution of target variable



Insights:

As we can see that it is not skewed but highly imbalanced. The number of non churners are vwery high. so we will handle this using SMOTE.

Handling class imbalance using SMOTE



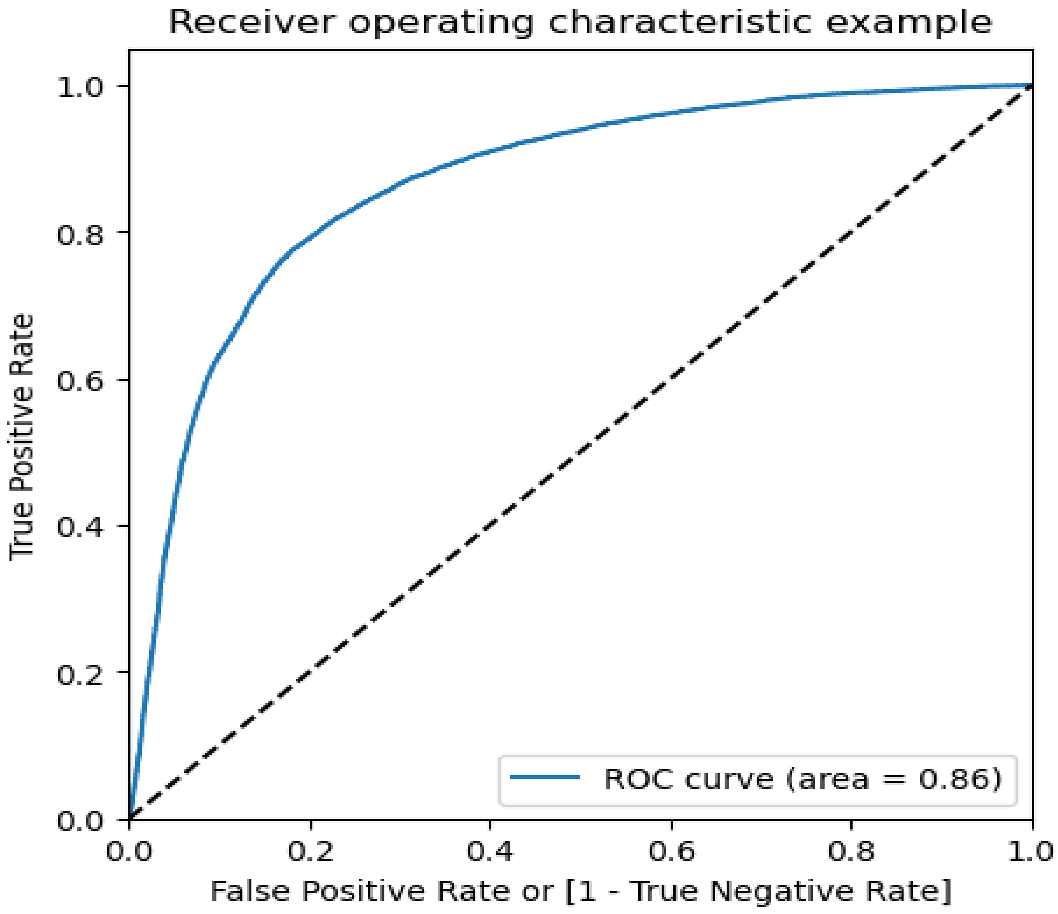
Model Building

1. Logistic Regression using RFE

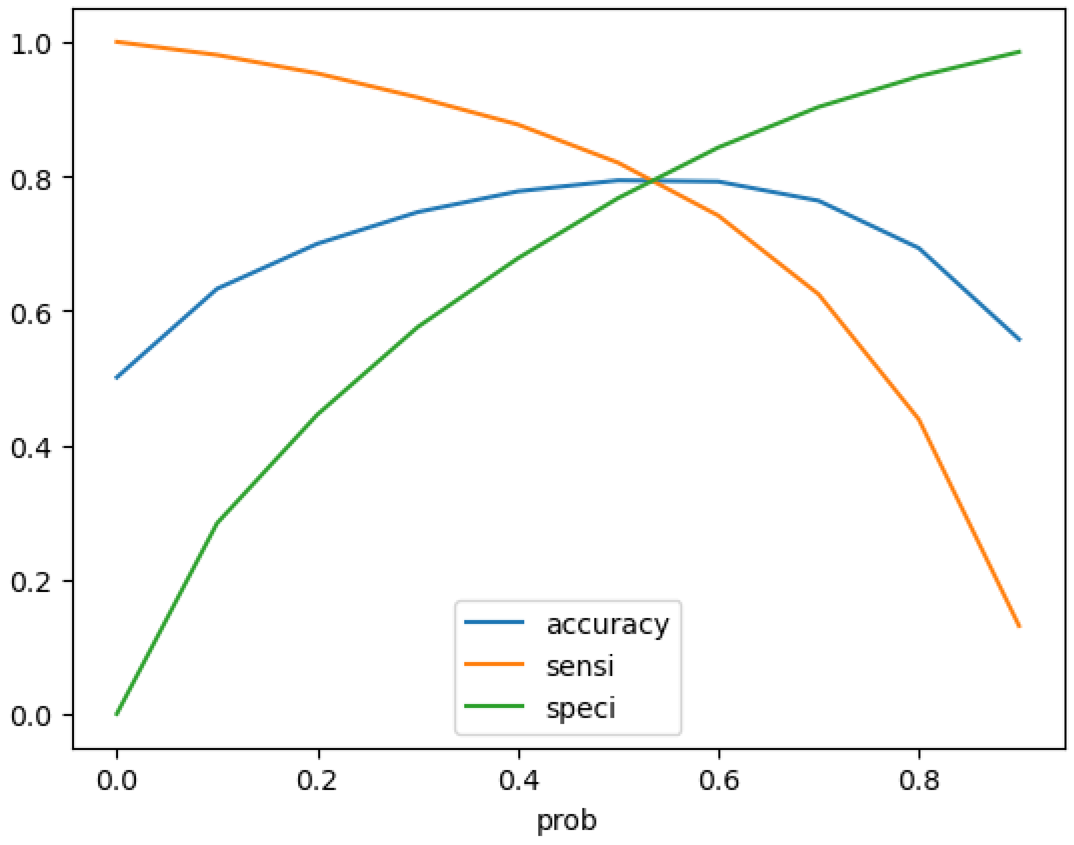
Dep. Variable:	churn	No. Observations:	38213
Model:	GLM	Df Residuals:	38187
Model Family:	Binomial	Df Model:	25
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-17764.
Date:	Sun, 03 Dec 2023	Deviance:	35528.
Time:	21:09:54	Pearson chi2:	1.92e+05
No. Iterations:	6	Pseudo R-squ. (CS):	0.3665
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.3573	0.021	-63.458	0.000	-1.399	-1.315
arpu_8	0.3533	0.033	10.825	0.000	0.289	0.417
roam_ic_mou_8	-0.3624	0.026	-14.202	0.000	-0.412	-0.312
loc_og_mou_8	-0.2828	0.047	-6.008	0.000	-0.375	-0.191
loc_ic_mou_8	-1.7448	0.058	-30.105	0.000	-1.858	-1.631
std_ic_t2t_mou_8	-0.3962	0.042	-9.417	0.000	-0.479	-0.314
spl_ic_mou_8	-0.2286	0.021	-10.804	0.000	-0.270	-0.187
total_rech_num_8	-0.5703	0.032	-17.630	0.000	-0.634	-0.507
max_rech_amt_8	0.2382	0.022	10.779	0.000	0.195	0.282
last_day_rch_amt_8	-0.5497	0.021	-26.072	0.000	-0.591	-0.508
vol_2g_mb_8	-0.2671	0.030	-8.989	0.000	-0.325	-0.209
monthly_2g_8	-0.6972	0.025	-27.787	0.000	-0.746	-0.648
sachet_2g_8	-0.4703	0.023	-20.526	0.000	-0.515	-0.425
monthly_3g_8	-0.9591	0.036	-26.835	0.000	-1.029	-0.889
sachet_3g_8	-0.4200	0.035	-11.884	0.000	-0.489	-0.351
aon	-0.3985	0.016	-24.794	0.000	-0.430	-0.367
total_mou_8	-0.8328	0.037	-22.587	0.000	-0.905	-0.761
gd_ph_total_mou	-0.8290	0.203	-4.078	0.000	-1.227	-0.431
gd_ph_std_og_mou	1.0200	0.179	5.697	0.000	0.669	1.371
gd_ph_sachet_3g	0.2044	0.022	9.505	0.000	0.162	0.247
gd_ph_vol_2g_mb	0.2244	0.020	11.169	0.000	0.185	0.264
gd_ph_monthly_3g	0.2872	0.023	12.505	0.000	0.242	0.332
gd_ph_loc_og_mou	0.7534	0.113	6.663	0.000	0.532	0.975
gd_ph_roam_og_mou	0.3068	0.033	9.295	0.000	0.242	0.371

ROC Plotting



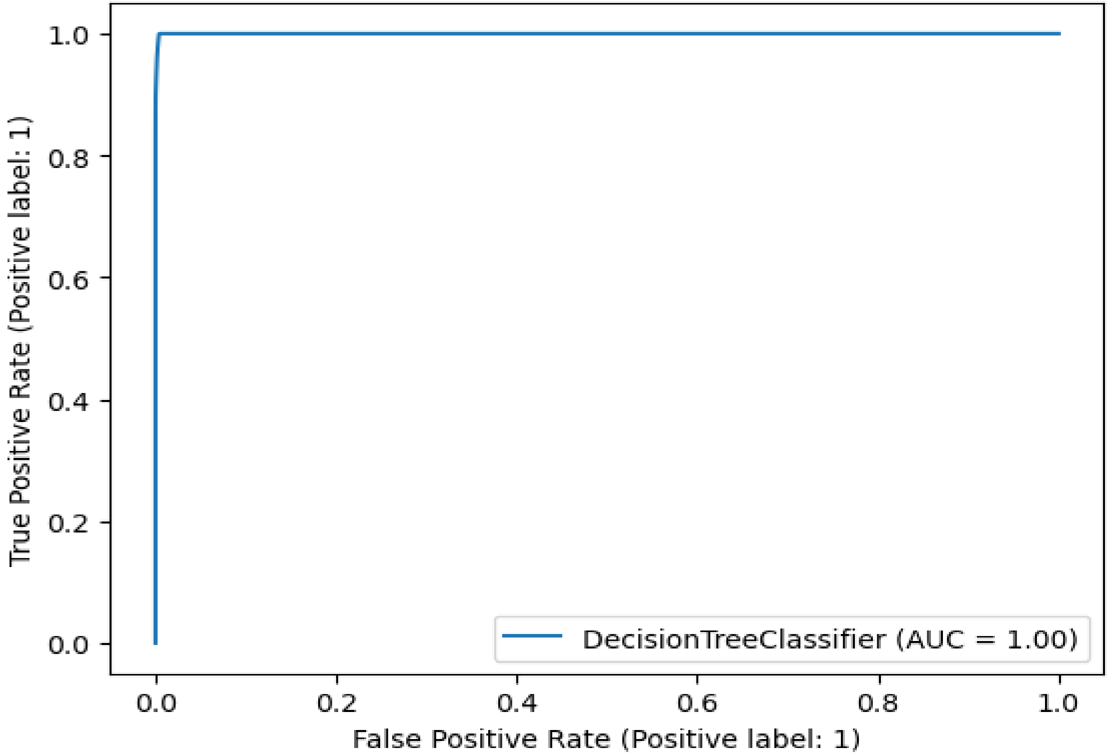
Optimal Cutoff



As we can see optimal cutoff is 0.5 so we will keep it

Decision Tree

ROC curve using hyperparameter tunning

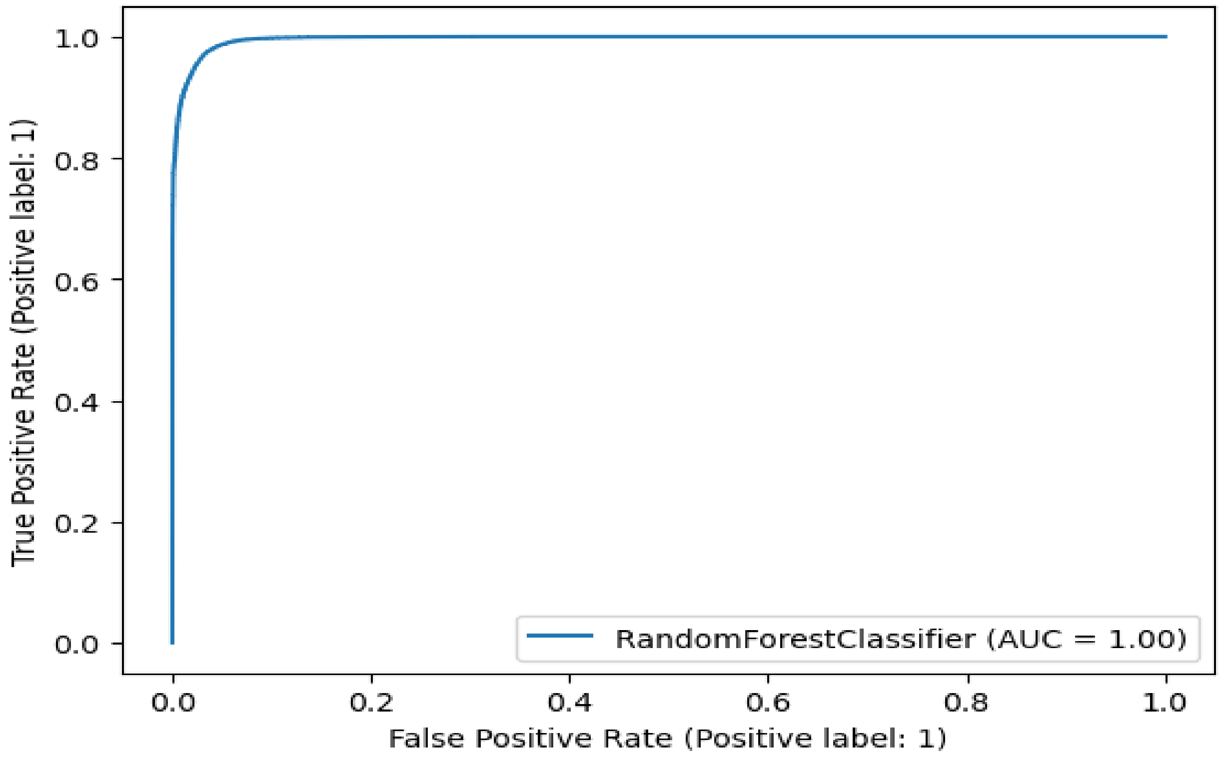


Insights:

With Decision Tree, we are getting 90% accuracy.

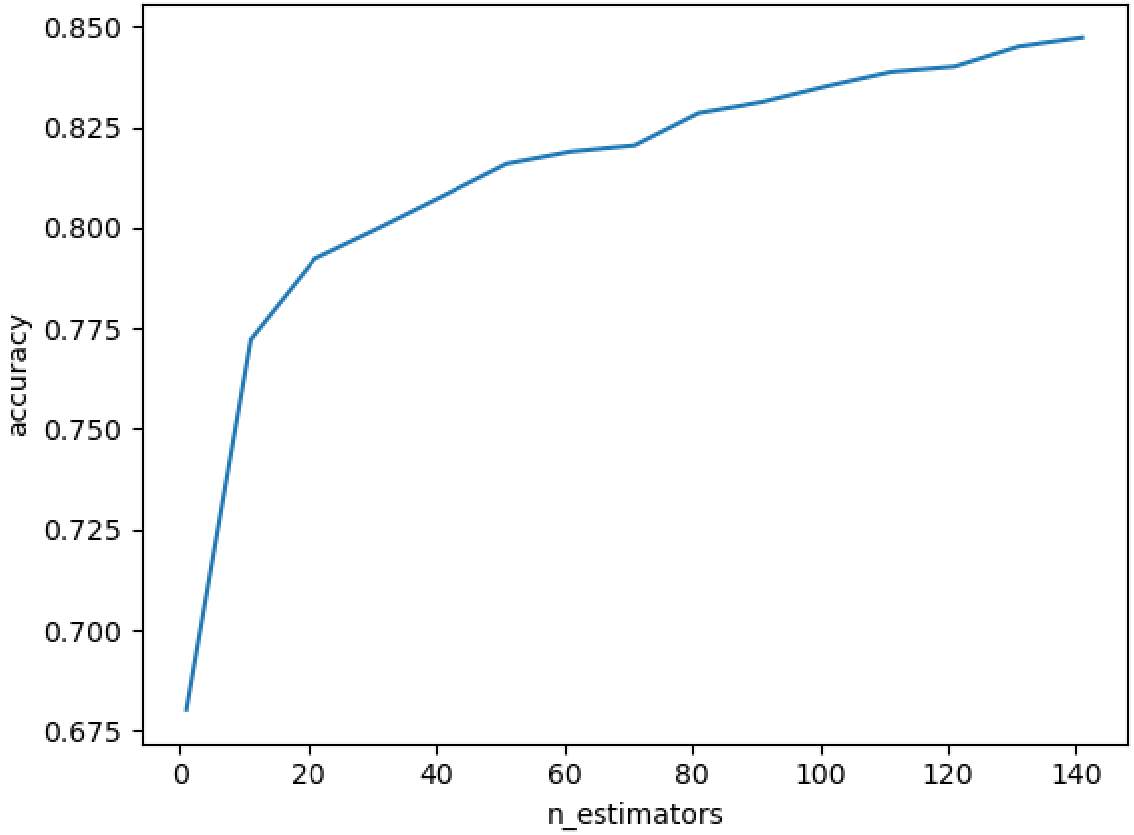
Random Forest

ROC curve using hyperparameter tunning



Insights:
With Random Forest, we are getting 94% accuracy

ADABOOST



Insights:
With ADABOOST, we are getting 94% accuracy.

Conclusion:

•We will consider accuracy to check as giving an offer to an user not who going to churn will cost less as compared to loosing a customer and bring new customer, we need to have high rate of correctly identifying the true positives.

•And as Random Foret and ADABOOST both have same accuracy of 94% but we will consider Random Forest as it is more robust.

Final model

```
Report on train data
              precision    recall  f1-score   support

           0       0.98      0.96      0.97     19080
           1       0.96      0.98      0.97     19133

    accuracy                           0.97     38213
   macro avg       0.97      0.97      0.97     38213
weighted avg       0.97      0.97      0.97     38213

Report on test data
              precision    recall  f1-score   support

           0       0.95      0.92      0.94      8215
           1       0.92      0.96      0.94      8162

    accuracy                           0.94     16377
   macro avg       0.94      0.94      0.94     16377
weighted avg       0.94      0.94      0.94     16377
```

Suggestions to handle customer churn

Top 10 Predictors to handle customer churn

loc_og_mou_8	1.282065
const	1.192894
total_rech_num_8	0.945401
monthly_3g_8	0.877368
monthly_2g_8	0.687312
gd_ph_loc_og_mou	0.649594
gd_ph_total_rech_num	0.632090
last_day_rch_amt_8	0.548943
std_ic_t2t_mou_8	0.517678
sachet_2g_8	0.441314
aon	0.39376

Some strategies to manage churns :

1. Churners show higher roaming usage than non churners.
2. Network operator should investigate their roaming tariffs and quality of services.
3. It may be a reason that roaming tariffs offered are less competitive than their competitors.
4. It may be a reason that customer is not getting good quality of services while roaming. In such case, quality of service guarantees with roaming partners and network quality needs to be investigated.
5. 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

Thank You