TELCOM CHURN CASE STUDY

IIIT-BANGALORE(BA)

Presented by:-
Deeksha Shrivastava
Devyani Rathod
Dharshan Jain
EPGP DSC 59 Batch

INTRODUCTION

- **Churn** is a problem of Telecom Companies where customers are able to choose from multiple service providers and actively switch from one operator to another. It is difficult to acquire new customers than keeping the existing one from leaving.
- There are two main models of payment in the telecom industry - **postpaid** (customers pay a monthly/annual bill after using the services) and **prepaid** (customers pay/recharge with a certain amount in advance and then use the services).
- This project is based on prepaid churn. **Churn prediction** is usually more critical (and non-trivial) for prepaid customers as it is hard to know whether someone has actually churned or is simply not using the services temporarily.
- This project is based on **Usage-based churn**: Customers who have not done any usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.

Problem Statement

- 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.
- Retaining high profitale customers is the number one business goal.

Business Objective

- Telecom companies need to **predict which customers are at high risk of churn.**
- 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.
- The dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively.
- **Business objective** is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behaviour during churn will be helpful.

Data Preparation

- Filter high-value customers : Those who have recharged with an amount more than or equal to X, where X is the **70th percentile** of the average recharge amount in the first two months (the good phase).
- Tag churners (churn=1, else 0) based on the fourth month : Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase
- Remove all the attributes of the churn phase

Analysis Approach

- Loading and Inspecting the Data
- Data Quality Check and Data Cleaning

  > checking duplicates,

  > checking null values,

  > dropping unique columns,

  > imputing missing values
- Data Understanding and Imbalance percentage
- Exploratory Data Analysis (EDA)

  > Univariate Data Analysis – checking value count,vizualise data distribution of variable,

  > Bivariate Data Analysis - correlation coefficient, pattern between the variables

Analysis Approach

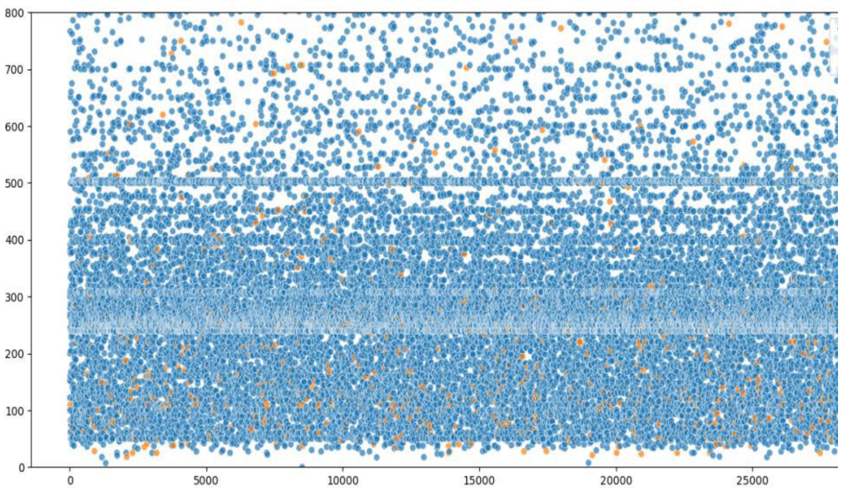
- Data Preparation –Standardization, Handling Class Imbalance, Principal Component Analysis(PCA)
- Test-Train Split
- Feature Scaling-checking correlation matrix
- Model Building- using stats model, feature selection using RFE , using manual feature selection
- Check VIF.
- Making predictions on train set
- Model Evaluation – confusion matrix, accuracy, sensitivity, specificity, recall, precision, precision and recall tradeoff
- Plotting the ROC Curve
- Finding Optimal Cutoff Point
- Making predictions on the test set
- Selecting the best classification model – Logistic Regression, Decision Tree, Random Forest
- Final Observation

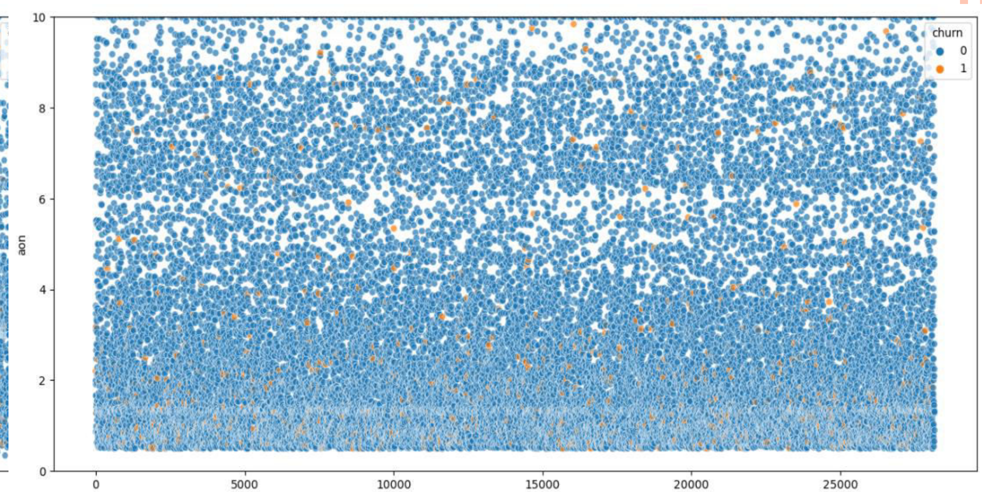
Variables contributing in Telecom Churn

- loc_ic_mou_8
- gd_ph_loc_ic_mou
- monthly_3g_8
- total_rech_num_8
- monthly_2g_8
- last_day_rch_amt_8
- gd_ph_total_rech_num
- loc_og_mou_8
- std_ic_t2t_mou_8
- sachet_2g_8

Exploratory Data Analysis

- Univariate Analysis

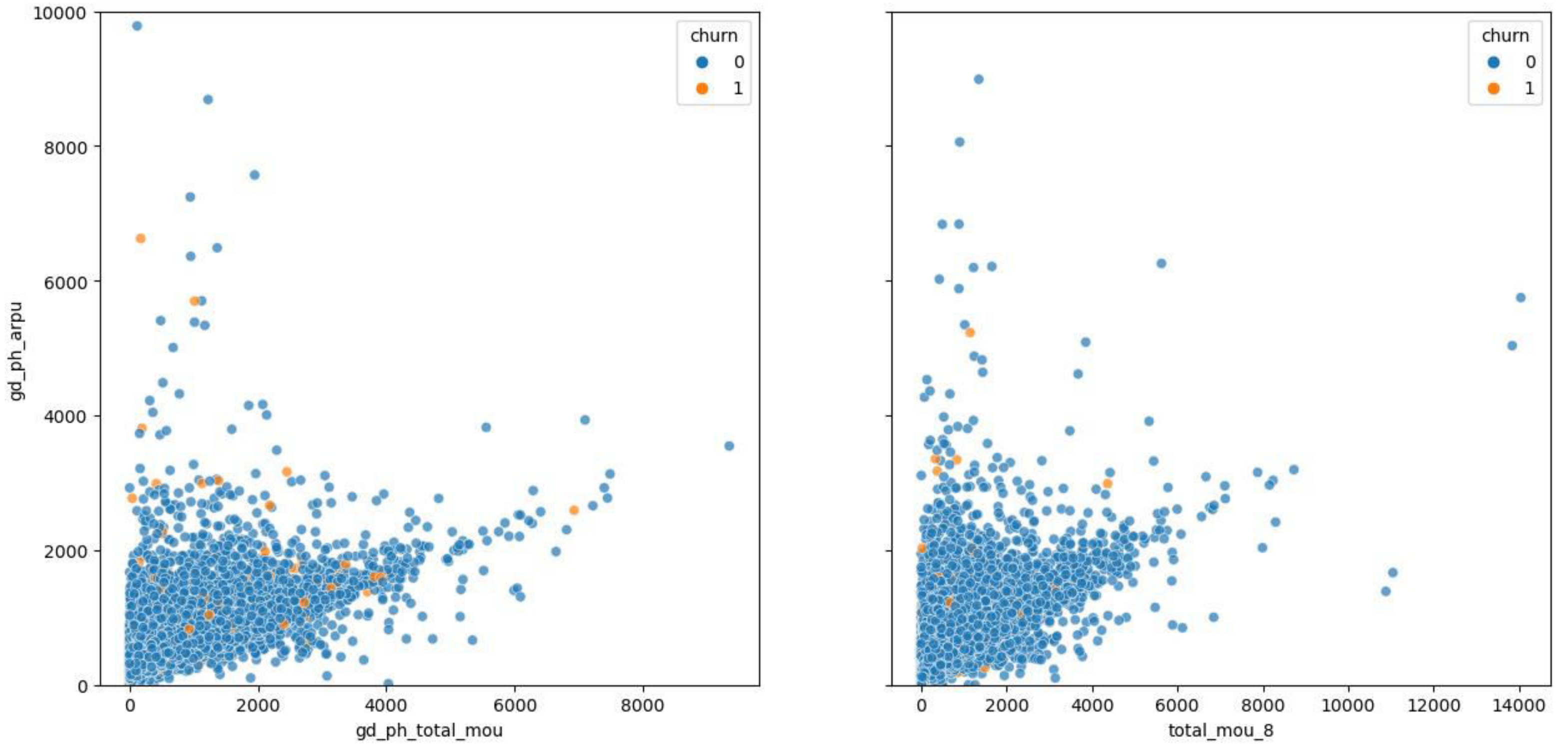




Effect of max recharge amount on churn : We can see that users who had max recharge amount is churned more.

Churn based on tenure :
There is no clear pattern visible but we can notice that the majority of churners had a tenure<4 years.

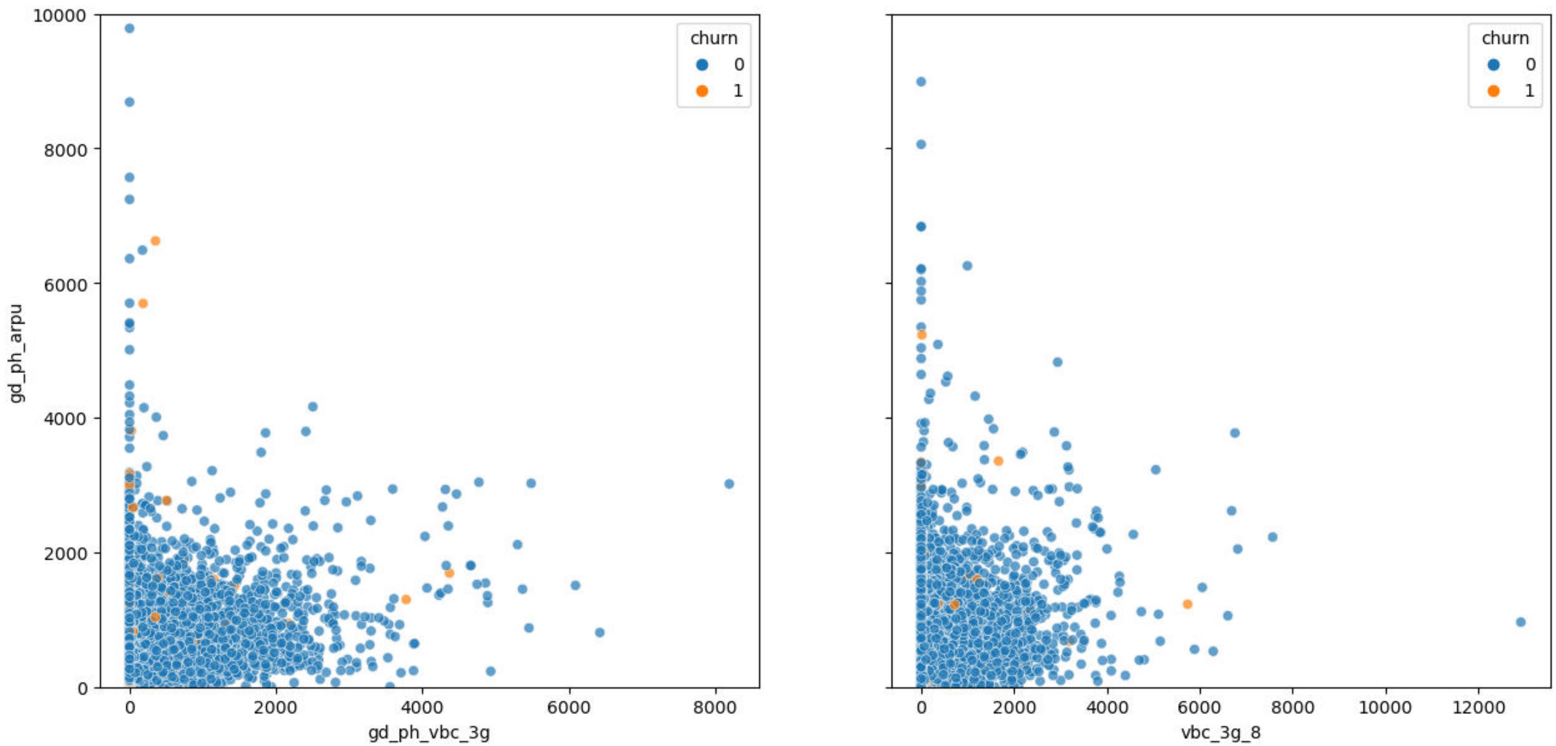
Multivariate analysis



<u>Effect of VBC on revenue</u>

We can clearly see that MOU is dropped significantly for the churners in the action phase i.e 8th month, thus hitting the average revenue generated from them It is also interesting that though the MOU is between 0-2000, the revenue is highest in that region that tells us these users had other services that were boosting the revenue

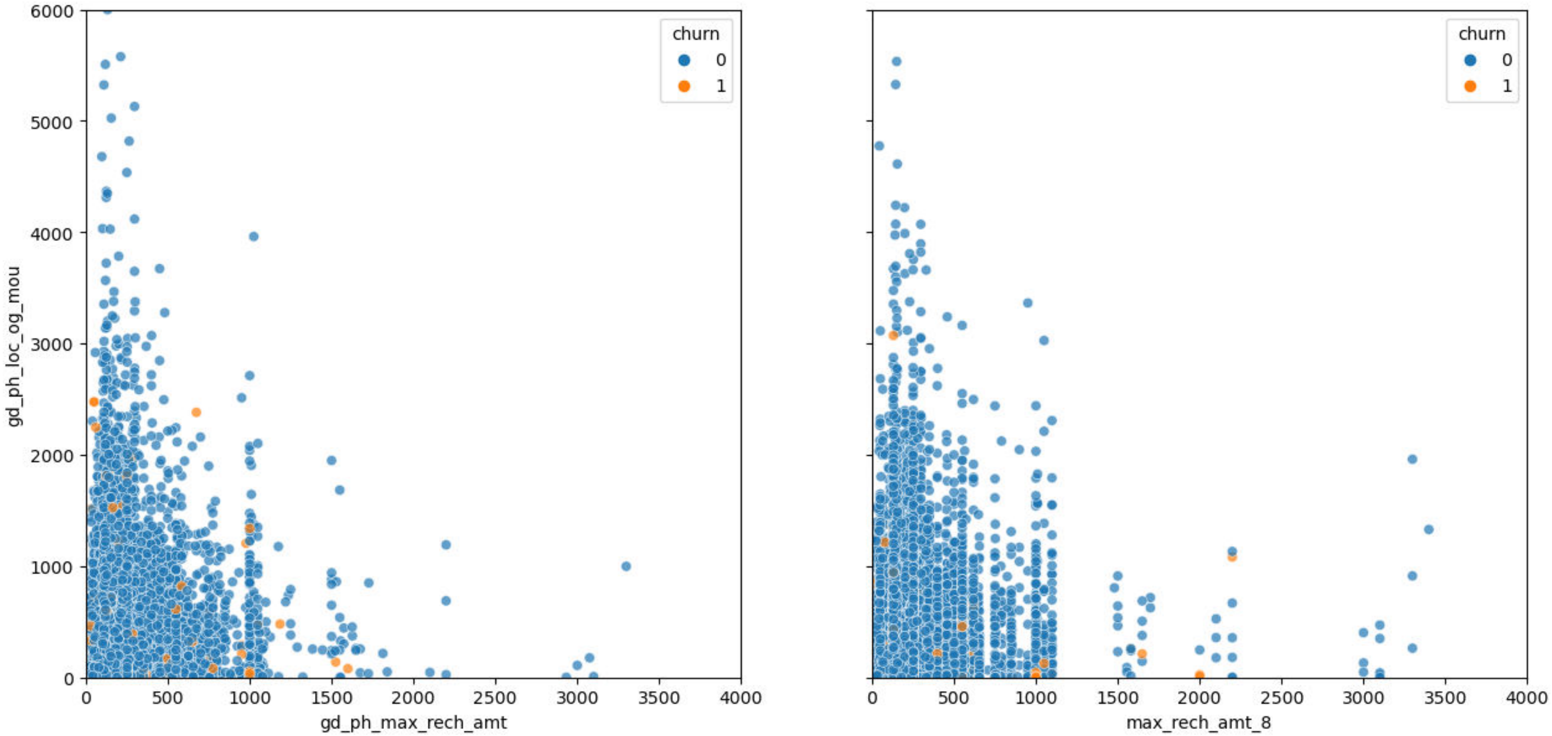
MULTIVARIATE ANALYSIS



Effect of total_mou on revenue

We can see that the users who were using very less amount of VBC data and yet were generating high revenue churned Yet again we see that the revenue is higher towards the lesser consumption side

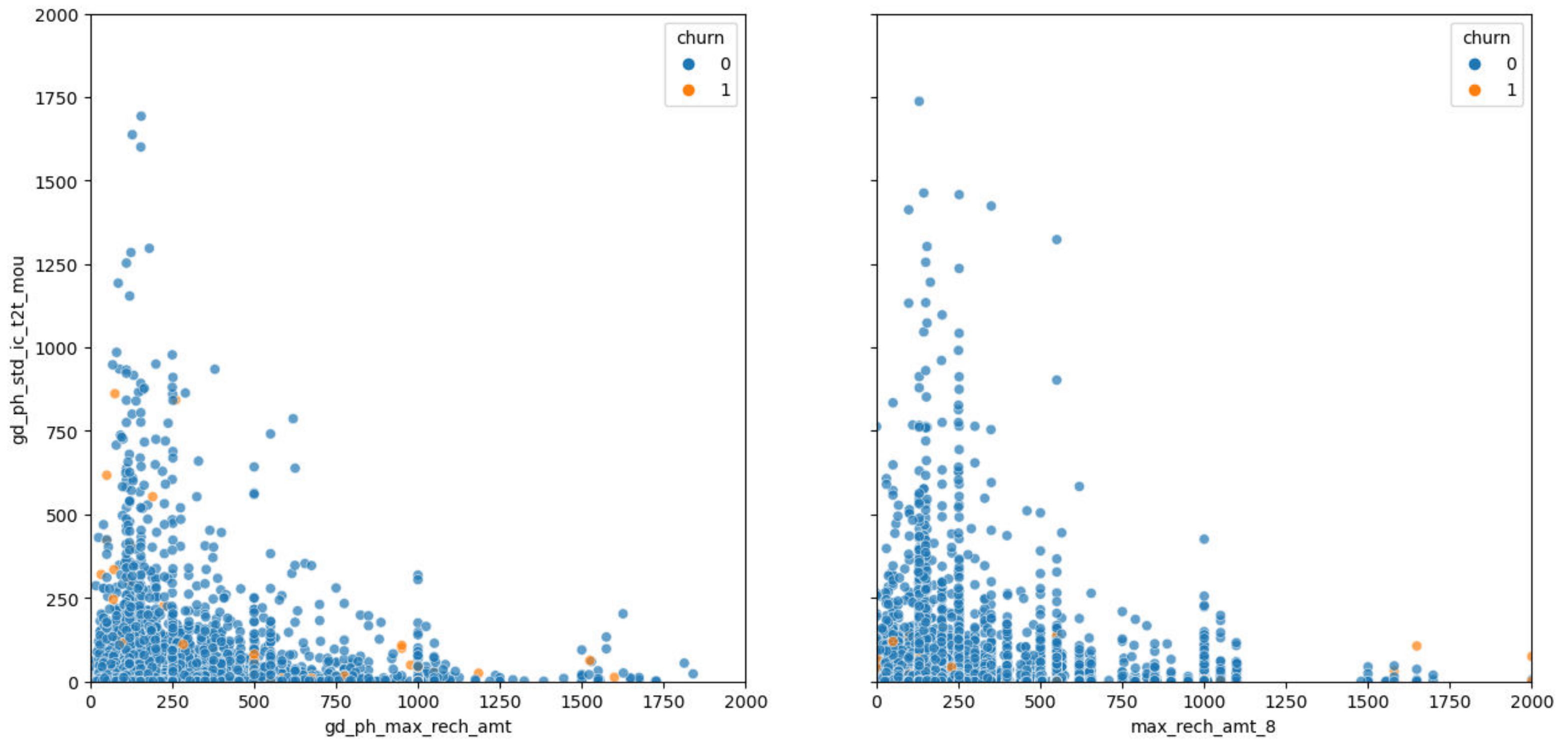
Multivariate analysis



<u>Relation between recharge amount and local outgoing calls</u>

Users who were recharging with high amounts were using the service for local uses less as compared to user who did lesser amounts of recharge Intuitively people whose max recharge amount as well as local out going were very less even in the good phase churned more

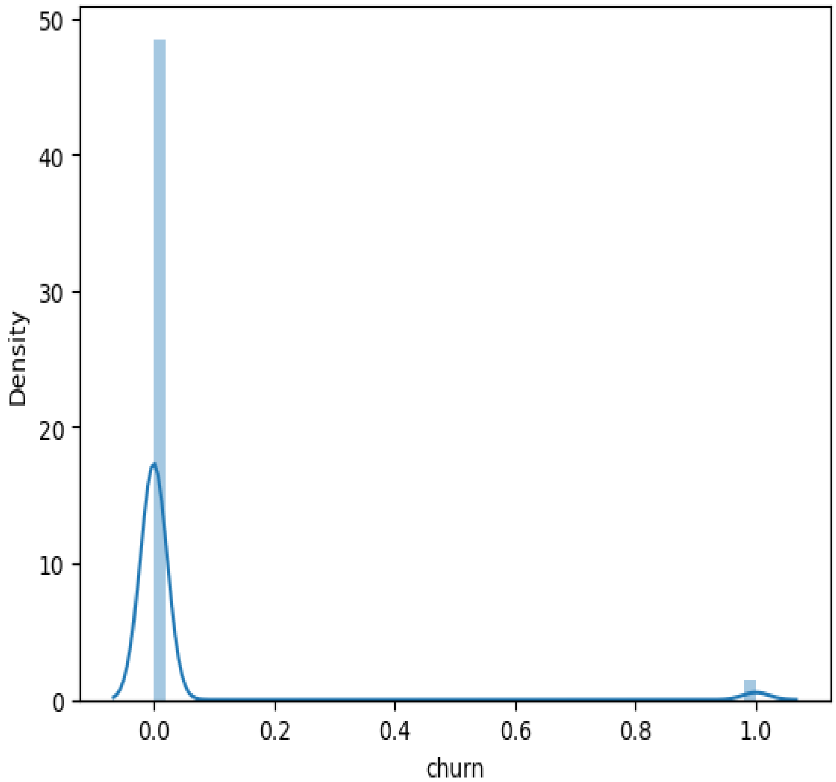
Multivariate analysis

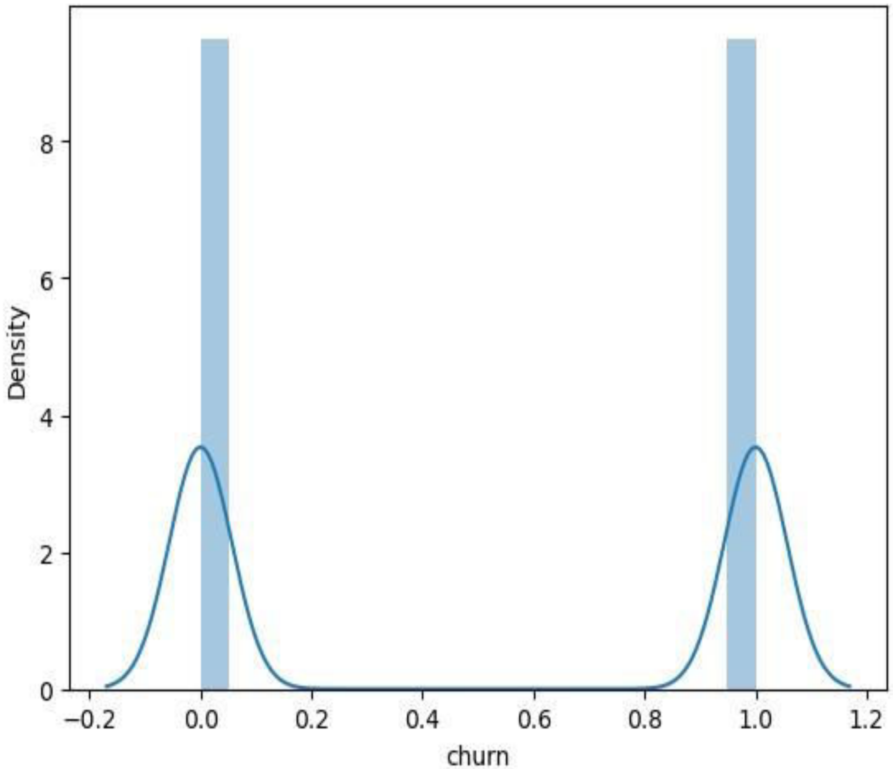


Incoming from the same service provider vs the recharge amount

Users who have max recharge amount on the higher end and still have low incoming call mou during the good phase, churned out more

Handling Class Imbalance

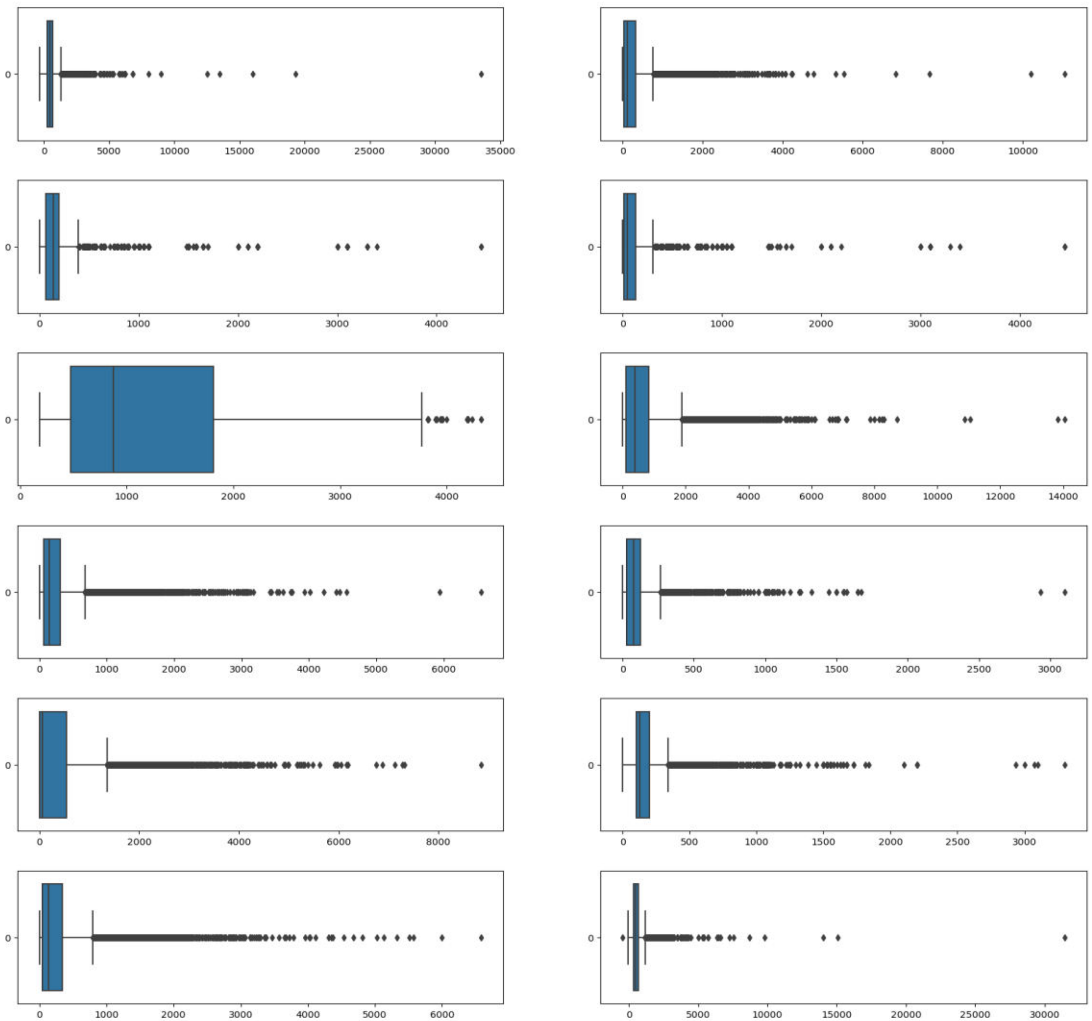




Though the varible is not skewed it is higly imbalanced, the number of non-churners in the dataset is more.
We handled this imbalance using SMOTE algorithm

Now the class is balanced and the target variable is not skewed

Observing Outliers Using Box



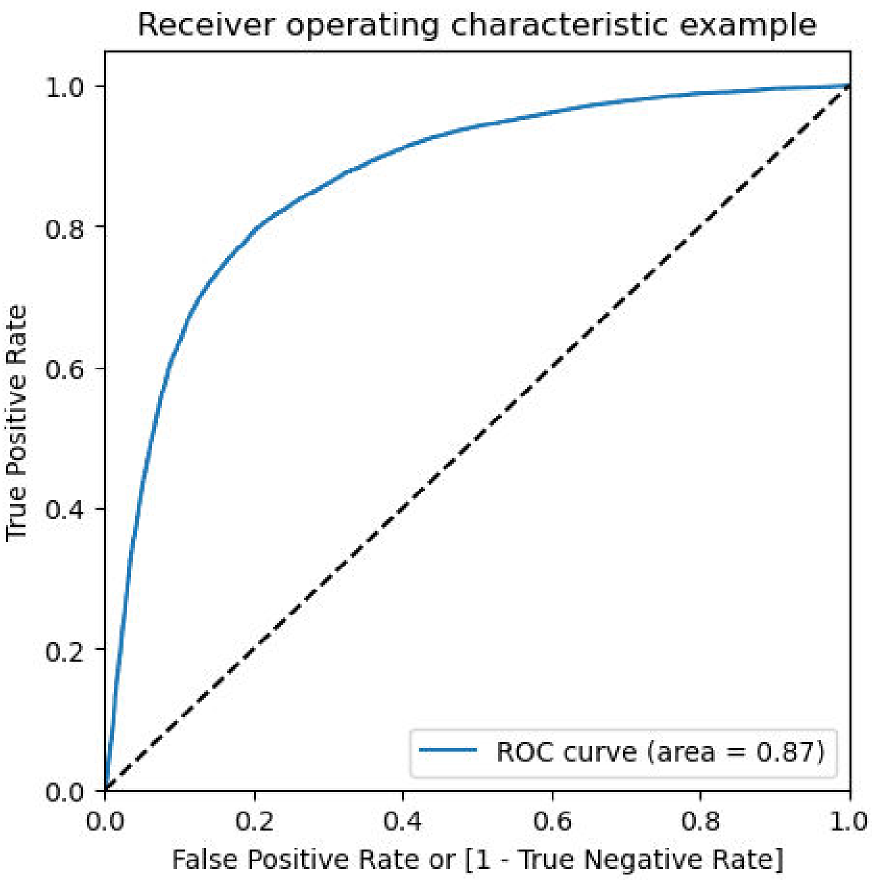
Observations:

From the above plots we can define following upper limits to the suspected 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

We will make these changes post exploration of other features

Model Evaluation

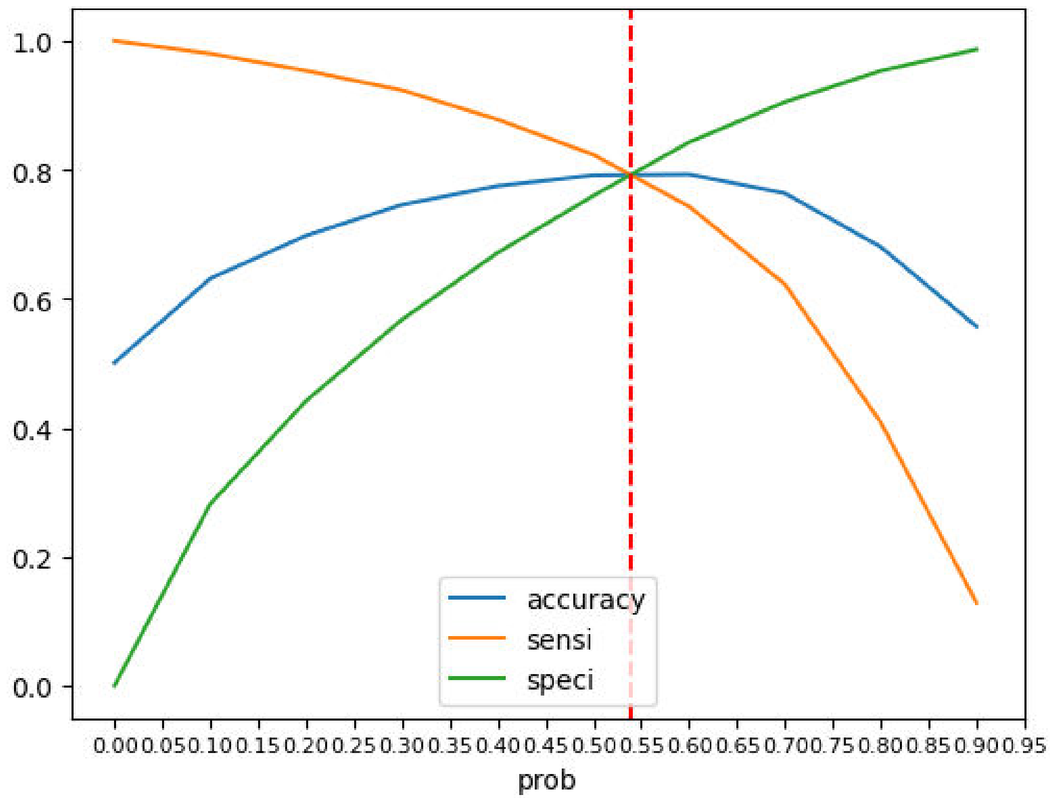


<u>ROC Curve</u>

ROC curve value should be close to 1 and we are getting value as 0.87. This shows it is a good predictive model.

Model Evaluation

Optimal Cut off Point



- Optimal cutoff probability is that prob where we get balanced sensitivity and specificity.
- From the plotted curve, we have found 0.539 as the best optimal cutoff point

Observations

- Train Data Set
- ***Accuracy = 79%***
- ***Sensitivity = 82%***
- ***Specificity = 76%***
- ***Precison = 77%***
- ***Recall = 82%***
- ***ROC Curve = 0.87***
- ***Optimal Cutoff = 0.539***

- ***Decision Tree: Accuracy = 87.8%, AUC = 0.93***
- ***Random Forest: Accuracy = 87.8%, AUC = 0.96***

CONCLUSION

- The customers whose recharge amount<200 in good phase, should be focused more as they are more likely to churn.
- MINUTES OF Usage is an important factor which clearly indicates the churn probability. If customers decrease the usage, they are more likely to be churned.
- Customers with poor network facility are more likely to shift from one telecom operator to another.
- Customers who have decreased their monthly usage for August is more likely to be churned.
- The key parameter to measure here is sensitivity as we had to measure the customers who churned correctly not predecting those who havent churned.

- **BUSINSS RECOMMENDATIONS :**

- We should target the customers whose MOU of incoming local calls and outgoing ISD calls are less in the action phase.
- We should target the customers whose outgoing_others and incoming_others in August are less.
- The customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be the good target to provide offer.
- Customers with monthly 3G recharge and monthly 2G data usage for August should be focused more as they are more probable to be churned.

Thank You