

# Telecom Churn Case Study

Anurag Bhardwaj





## PROBLEM STATEMENT

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 customer's at high risk of churn and identify the main indicators.

---



1

Reading Understanding and Visualizing Data

Preparing Data for Modelling

2

3

Model Building

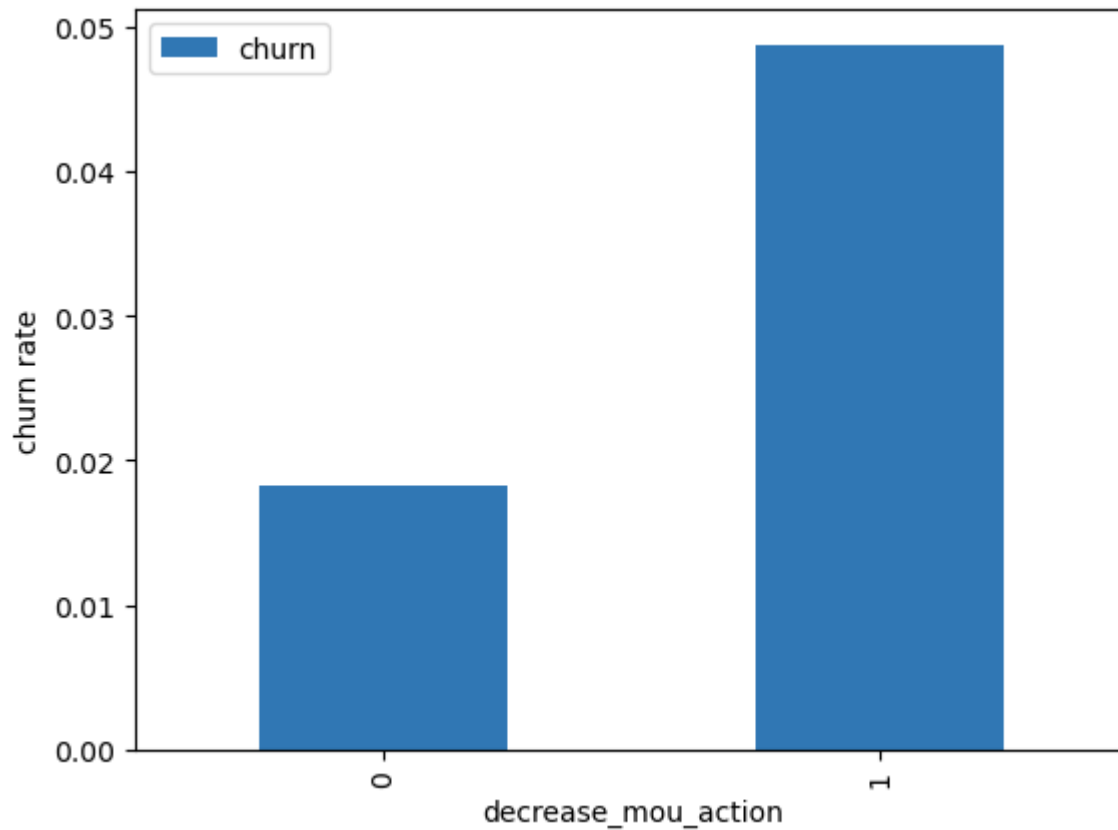
Create Test Models

4

5

Evaluate the Model

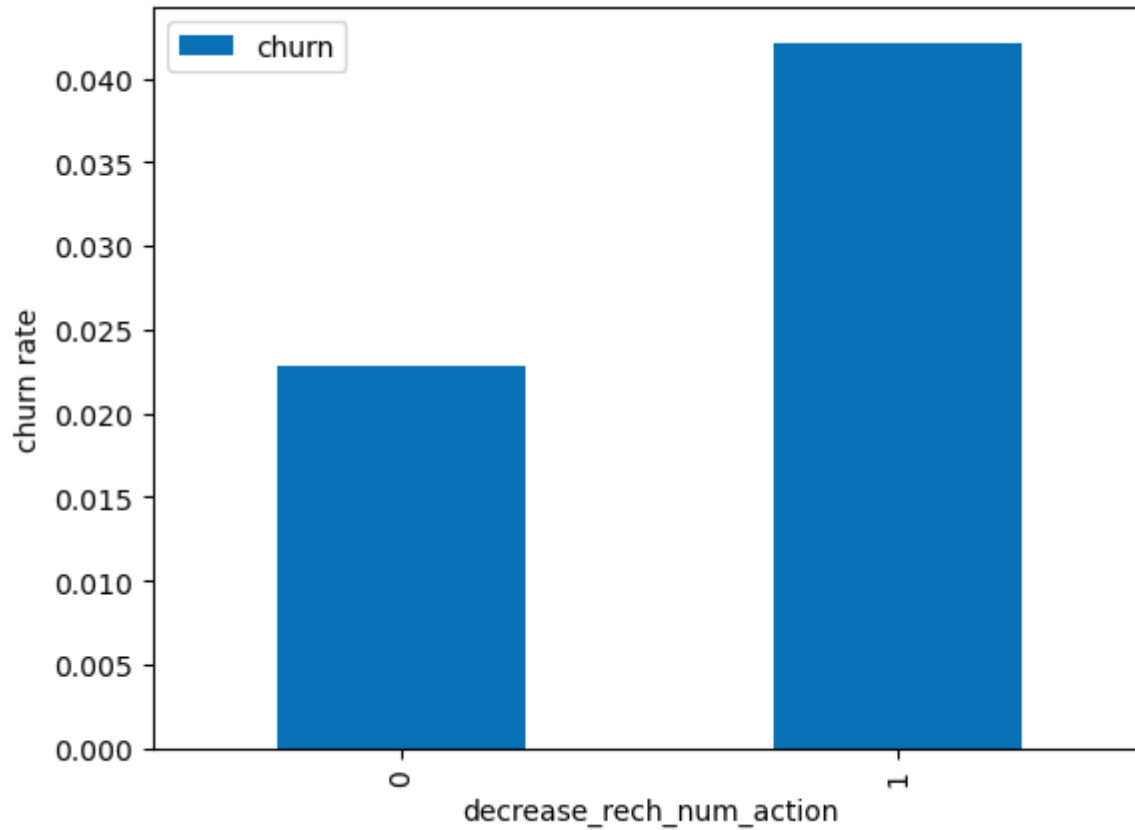
# Univariate Analysis



Churn Rate per Decrease MOU  
Action

Churn rate is more for the customer's  
whose MOU decreased in the action  
phase than the good phase.

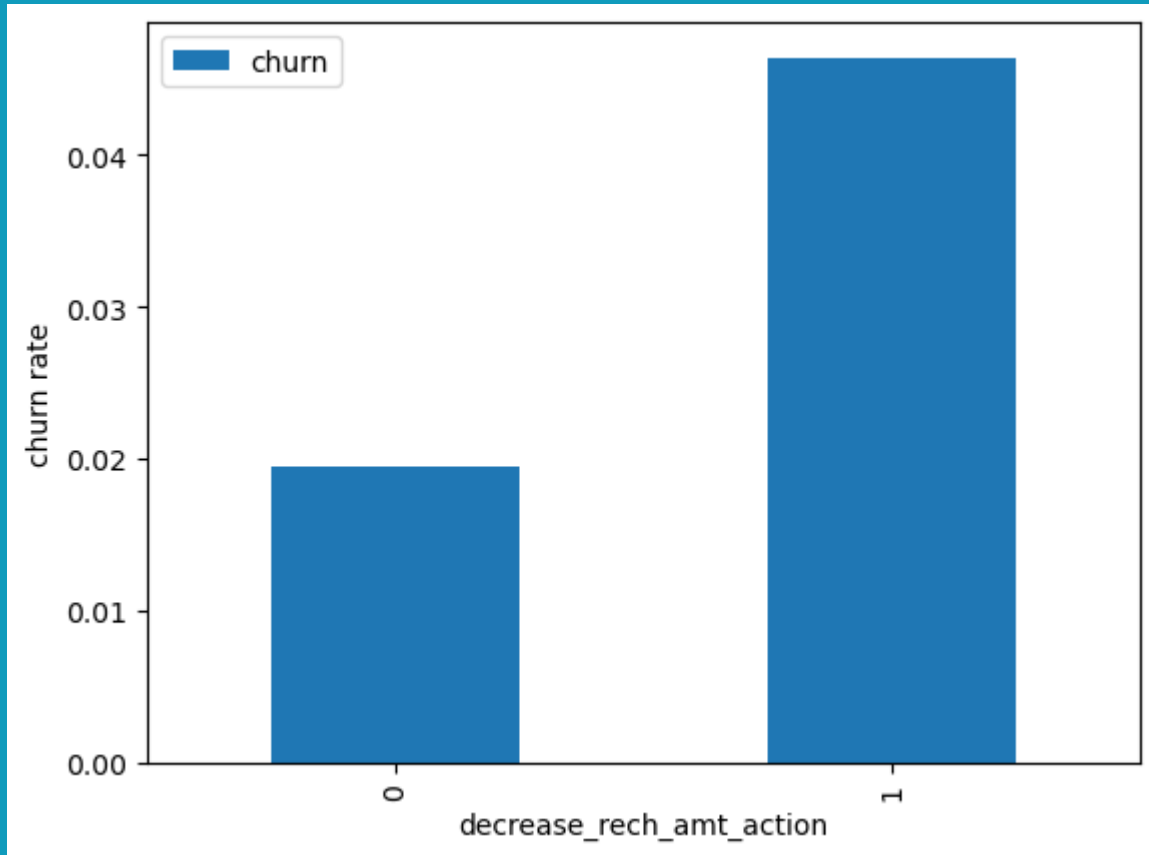
# Univariate Analysis



Churn Rate per recharge in action month

1. Churn rate is more for the customer's whose number of recharge in the action phase is lesser than the number in good phase.

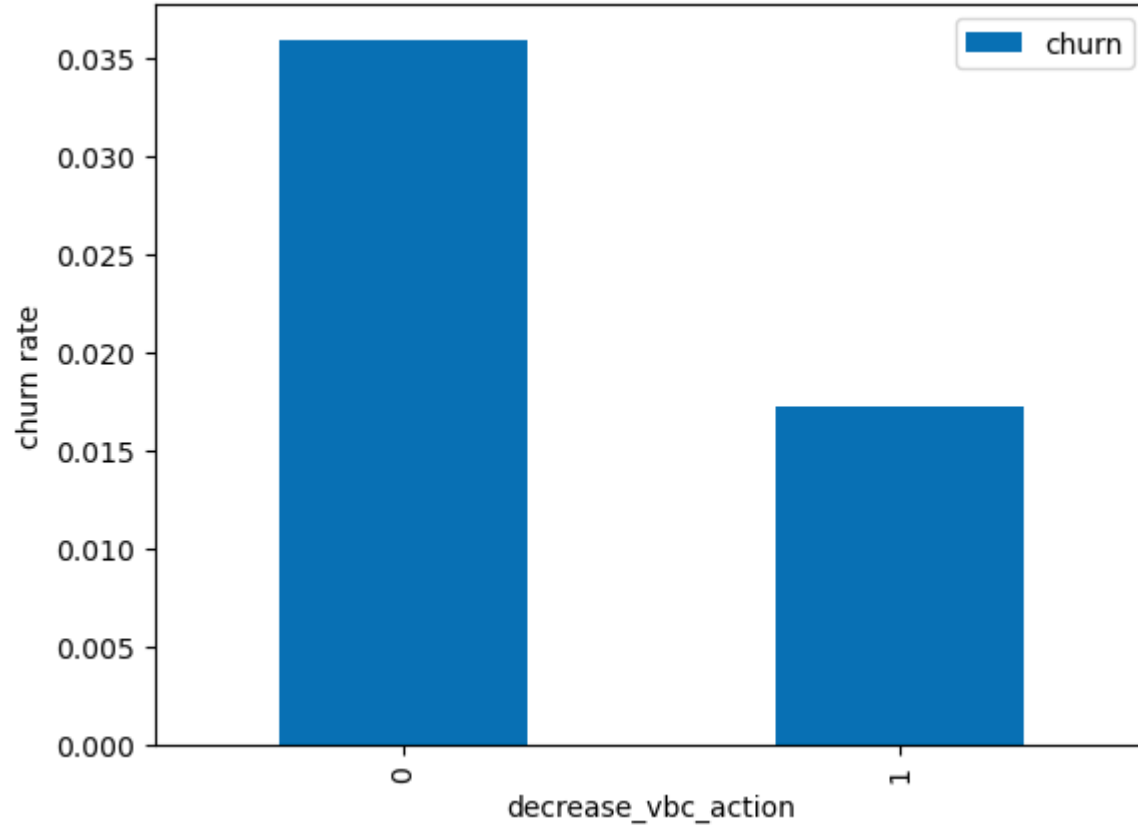
## Univariate Analysis



Churn Rate per recharge in action month

Here, we also see the same behavior. Churn rate is more for the customer's whose number of recharge in the action phase is lesser than the number in good phase.

## Univariate Analysis

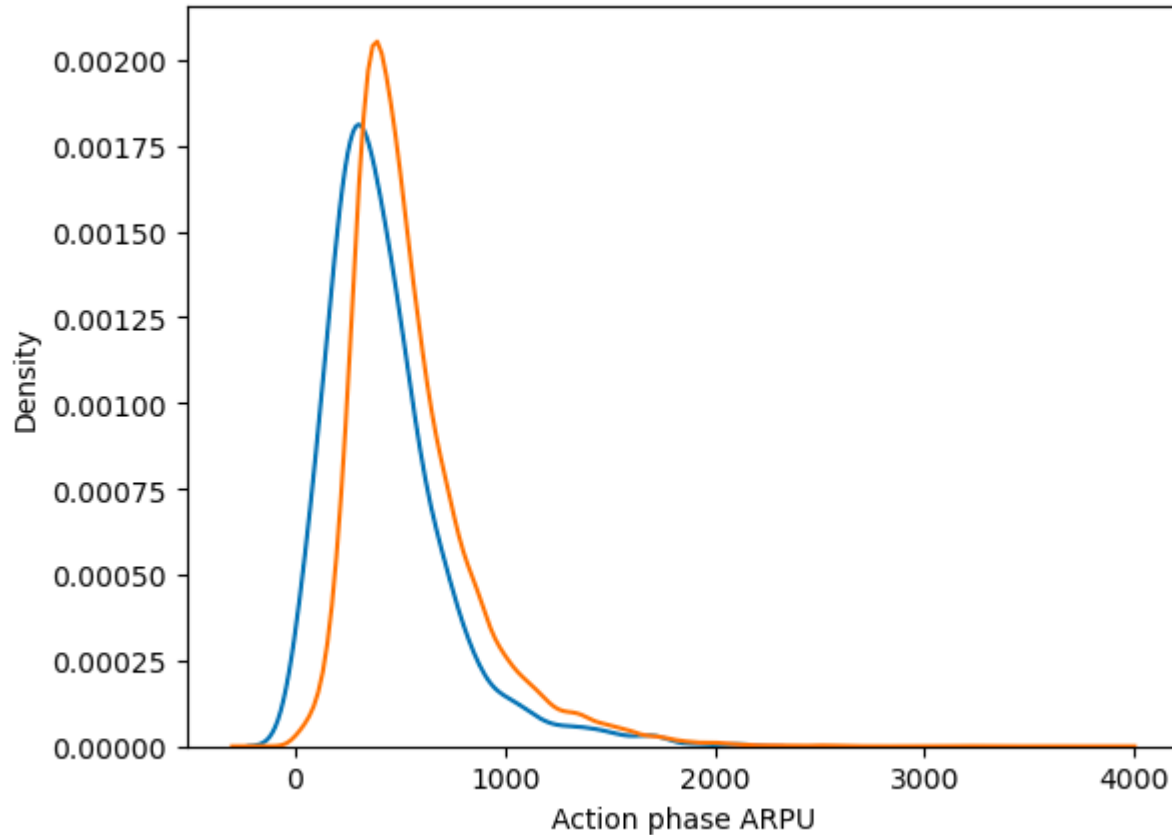


Volume Based Cost in action month:

Churn rate is more for the customer's whose volume based cost in action month is increased. This means the customer's do not do the monthly recharge more when they are in the action phase.

# Univariate Analysis

```
[Text(0.5, 0, 'Action phase ARPU')]
```



Average Revenue per customer –  
Churn and non-churn

ARPU for the churned customer is  
mostly densed on the 0-900.

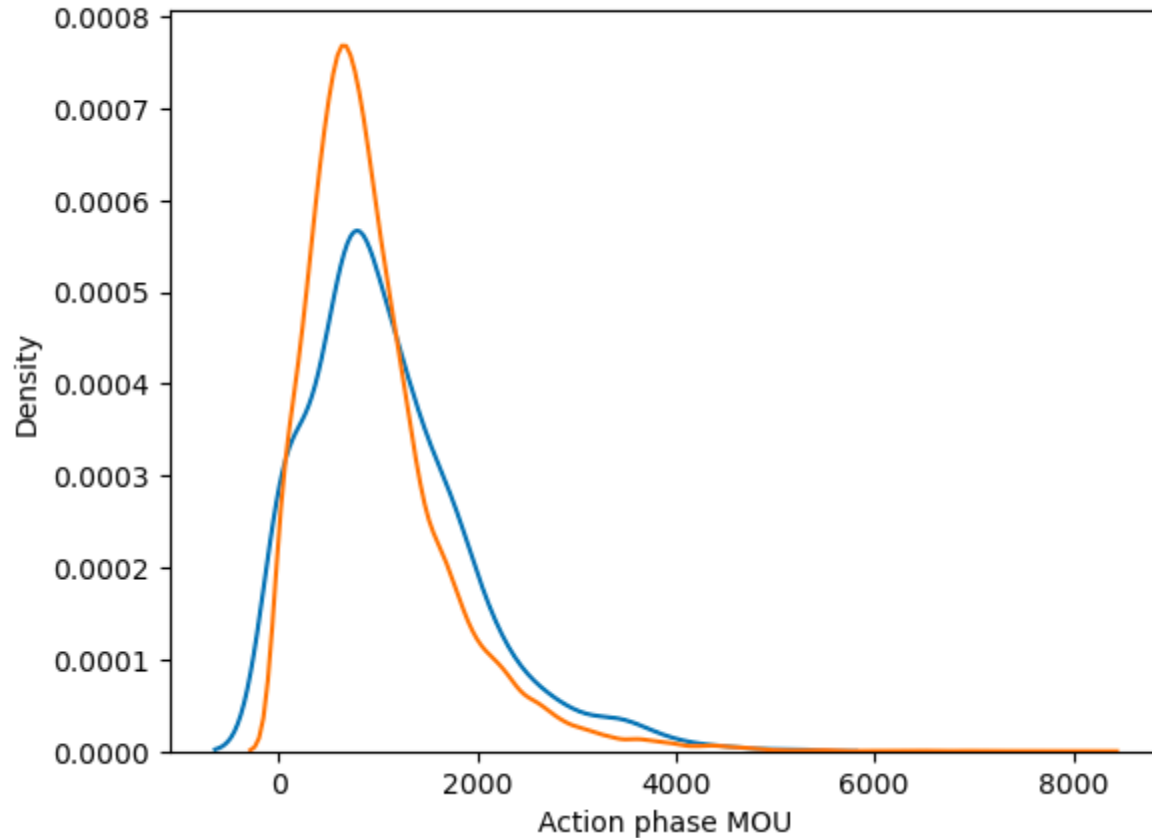
The higher ARPU customer's are less  
likely to be churned.

ARPU for not churned is mostly  
densed on the 0 to 1000.



# Univariate Analysis

```
[Text(0.5, 0, 'Action phase MOU')]
```

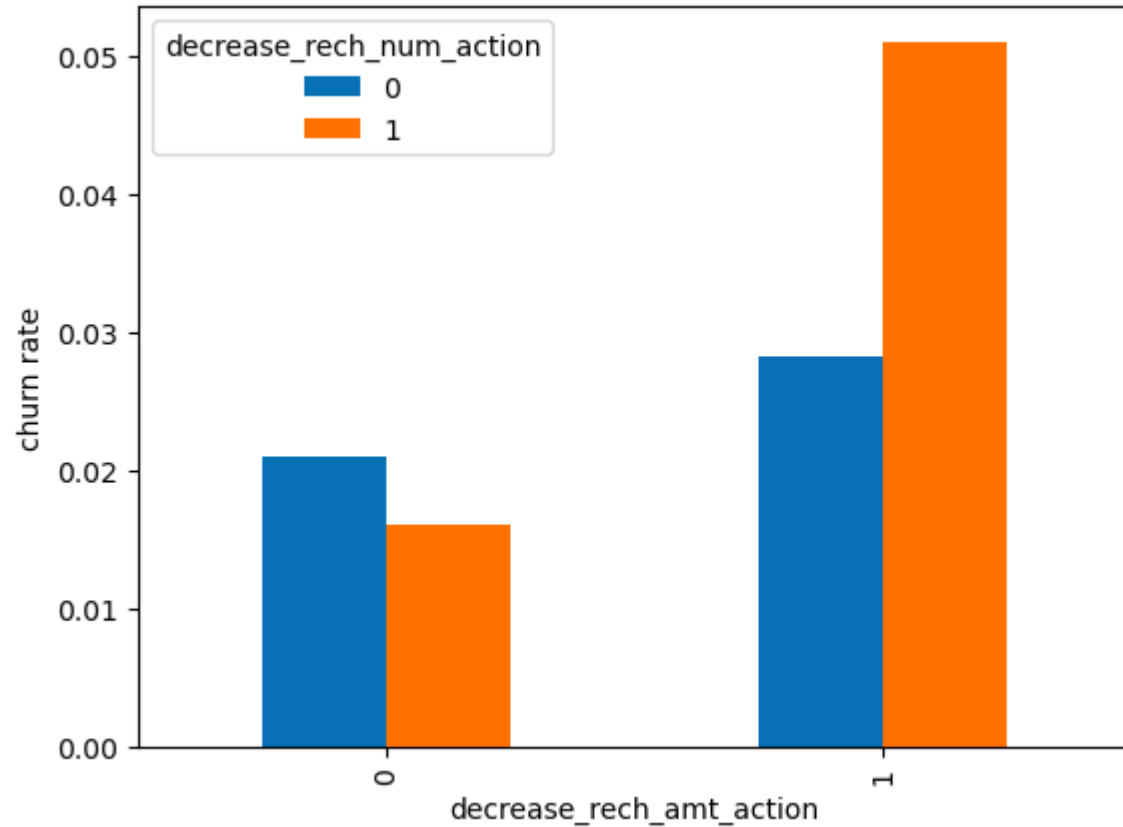


Minutes of Usage in Action Phase

MOU of the churned customer is mostly populated on the 0 to 2500 range.

Higher the MOU, lesser the churn probability.

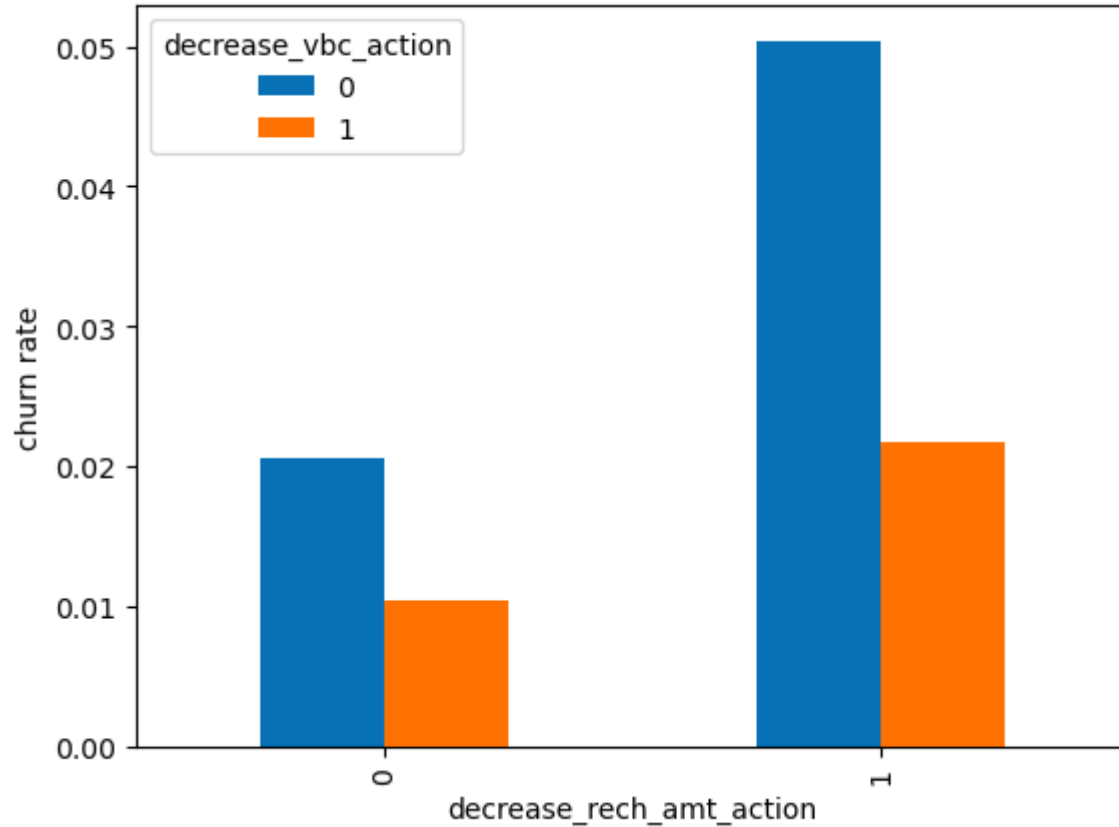
## Bivariate Analysis



Analysis of churn rate by the decreasing recharge amount and number of recharge in the action phase

We can see from the plot, that the churn rate is more for the customer's whose recharge amount as well as number of recharge have decreased in the action phase than the good phase.

## Bivariate Analysis

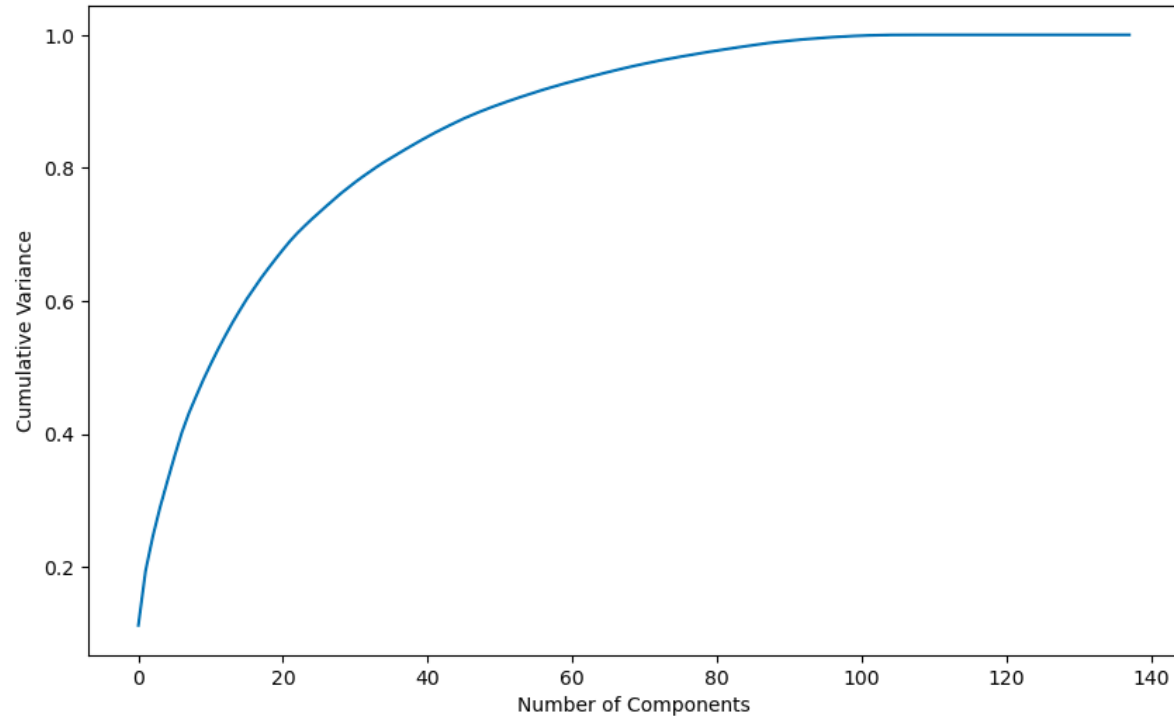


churn rate by the decreasing recharge amount  
and volume based cost in the action phase

Churn rate is more for the customer's  
whose recharge amount is decreased  
along with the volume based cost is  
increased in the action month.

## Test Train – Model with PCA

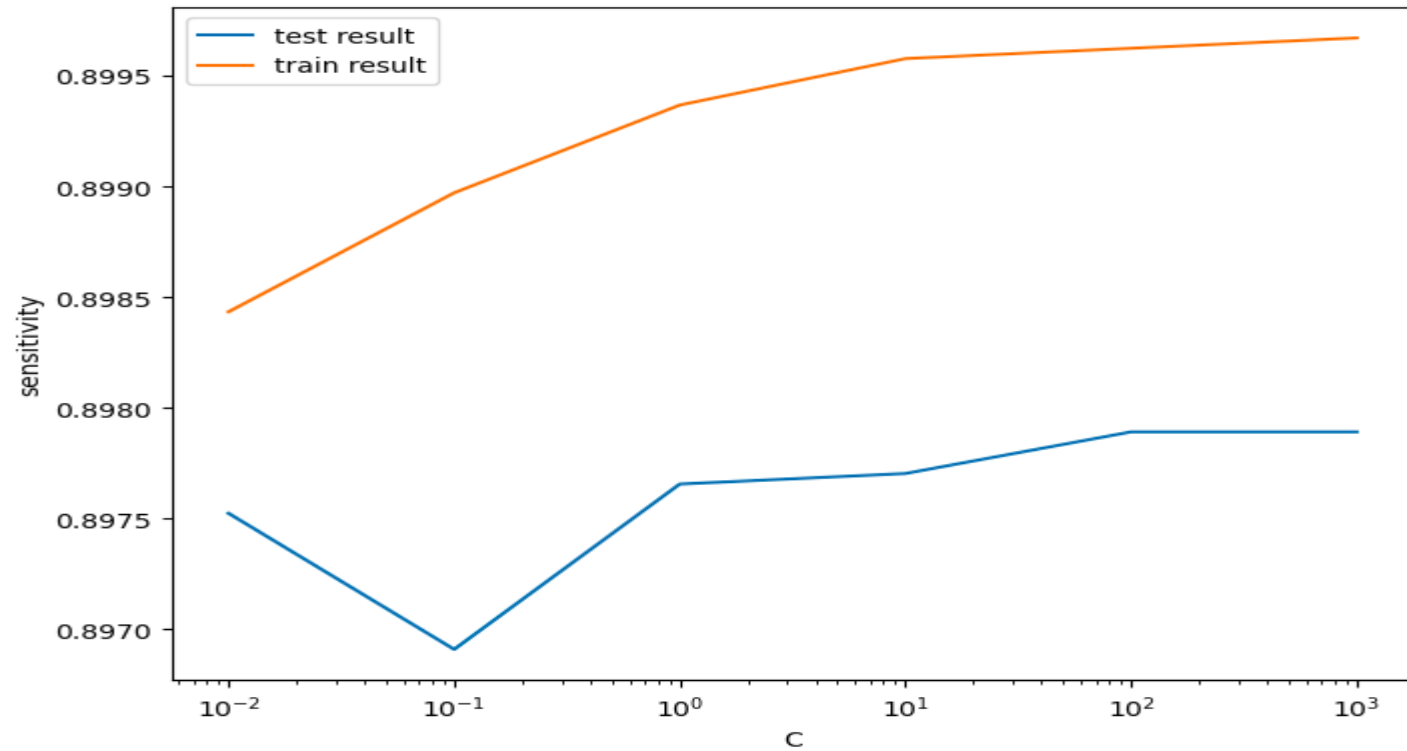
Text(0, 0.5, 'Cumulative Variance')



## Test-Train with PCA

We can see that 60 components explain almost more than 90% variance of the data. So, we will perform PCA with 60 components.

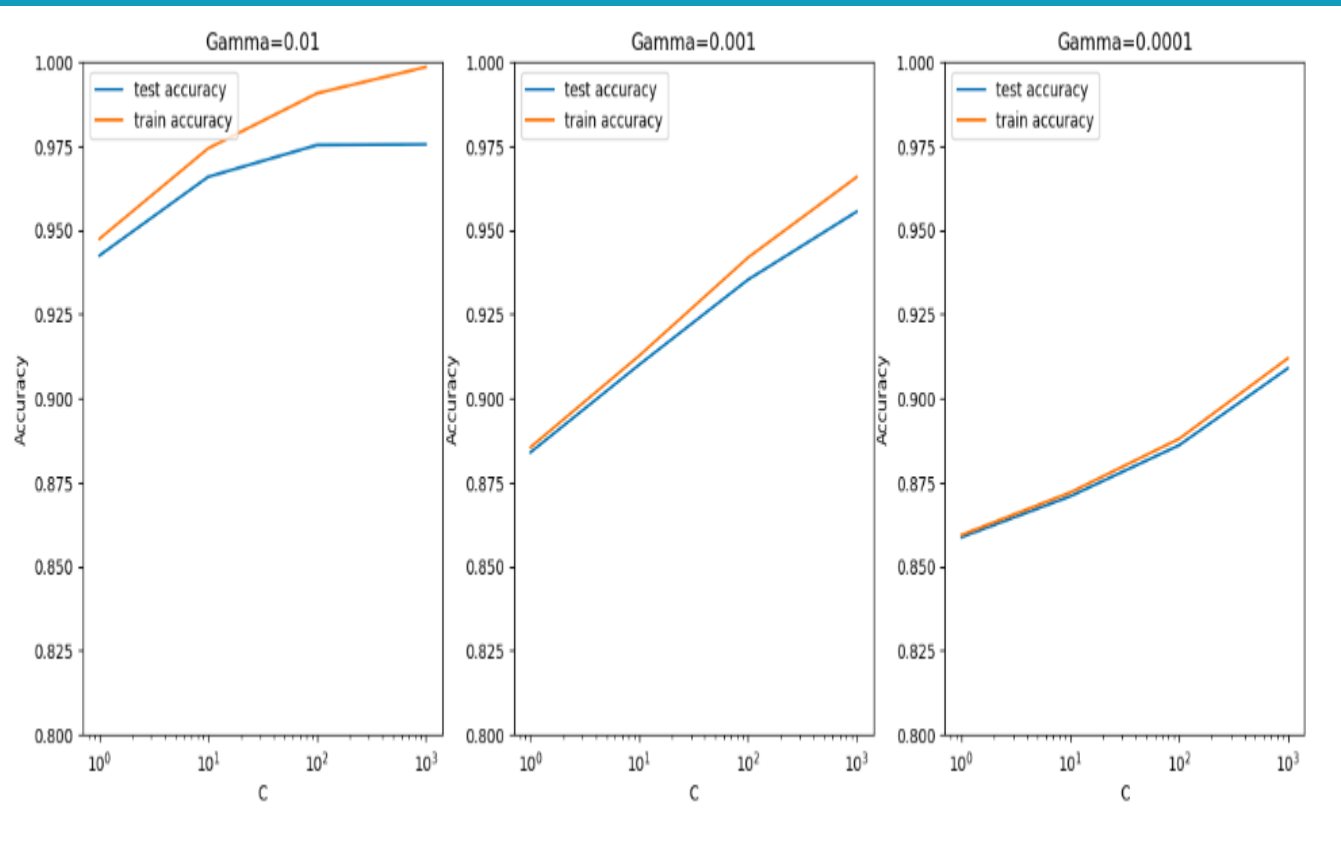
## Logistic regression with PCA



Logistic regression with PCA

The highest test sensitivity is  
0.8978916608693863 at  $C = 100$

# Support Vector Mechanism with PCA

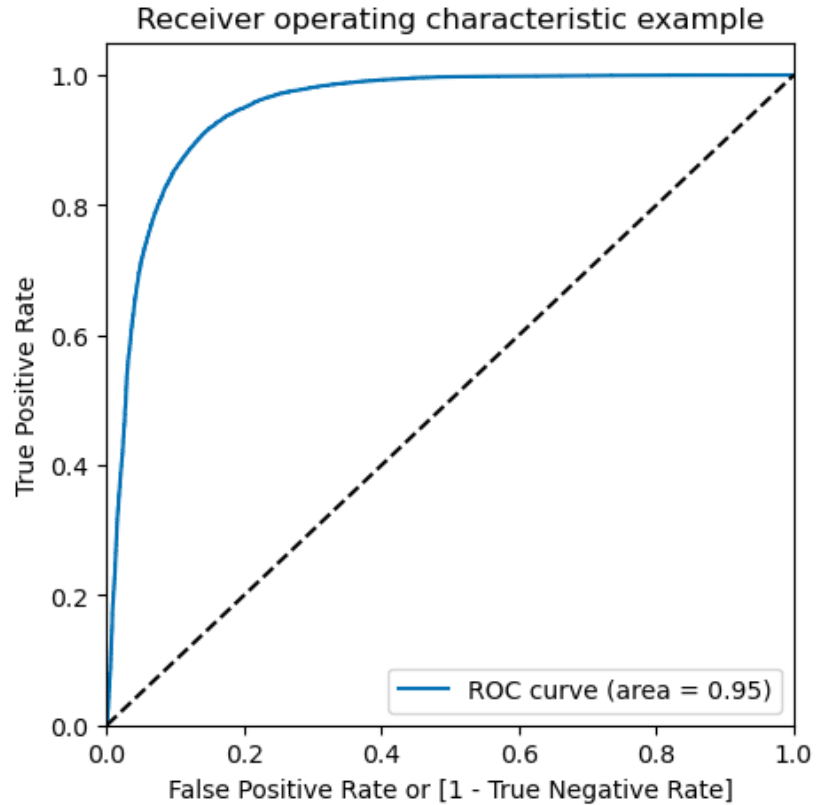


SVM with PCA

The best test score is 0.9754959911159373  
corresponding to hyperparameters {'C':  
1000, 'gamma': 0.01}



## Plotting the ROC Curve

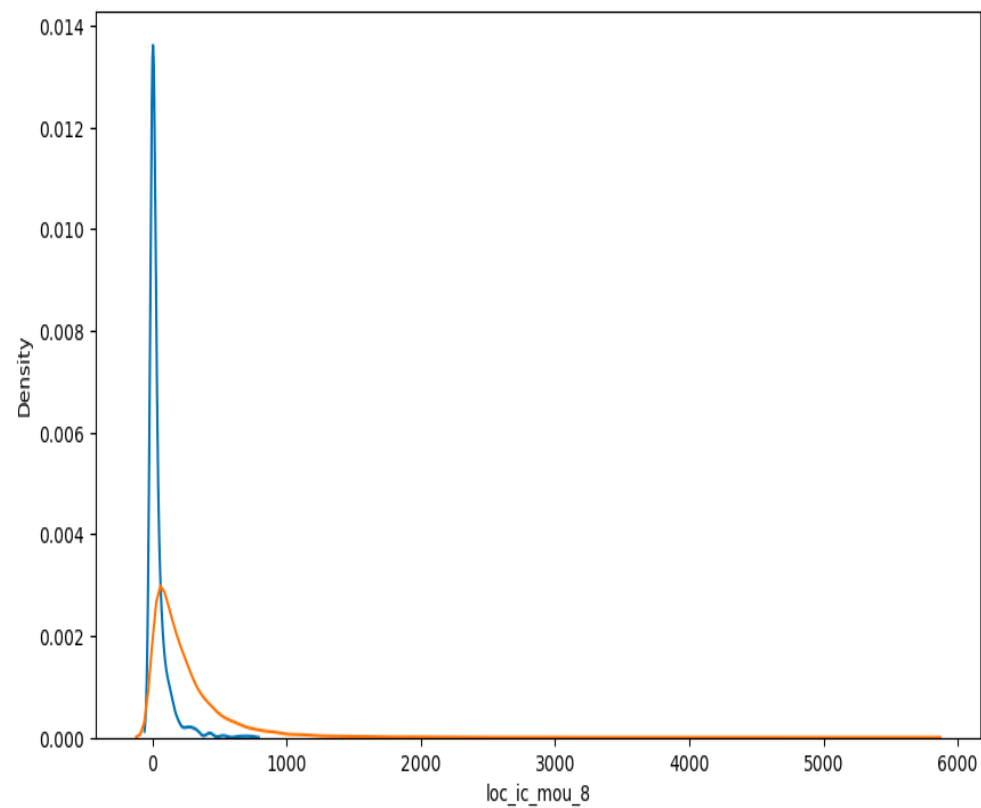


## ROC Curve

We can see the area of ROC curve is closure to 1, which is the GINI Model

*We can see the area of the ROC curve is closer to 1, which is the Gini of the model.*

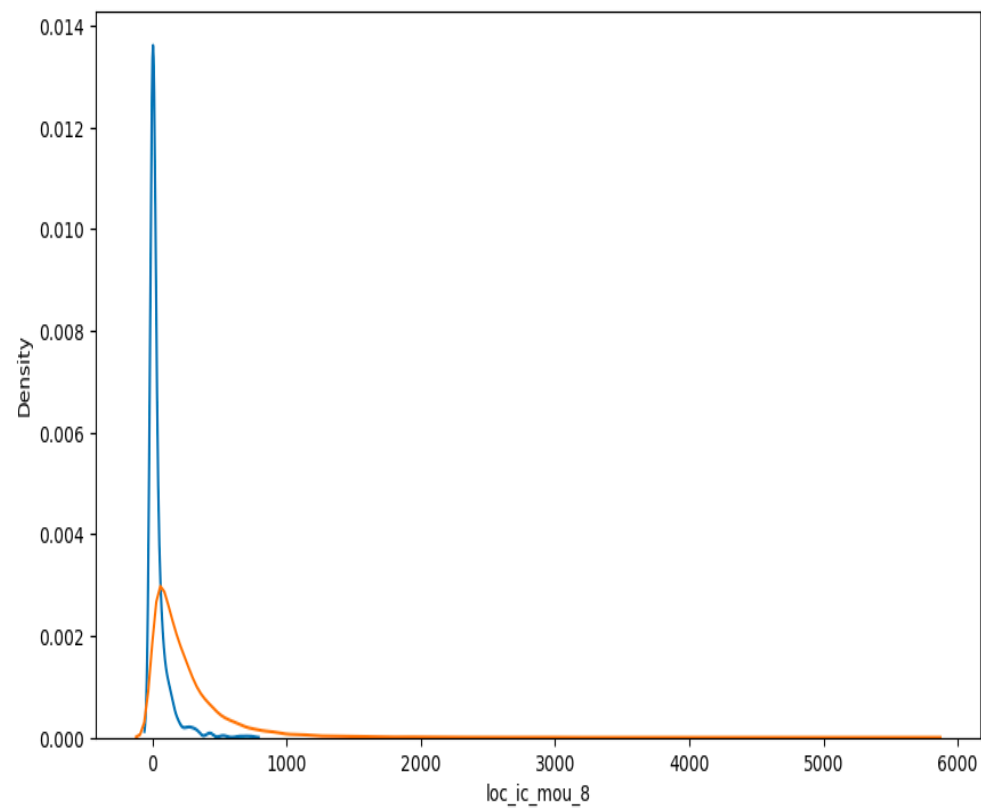
**Plots of important predictors for churn and non  
churn customers**



**Interpretation:**

We can see that for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.

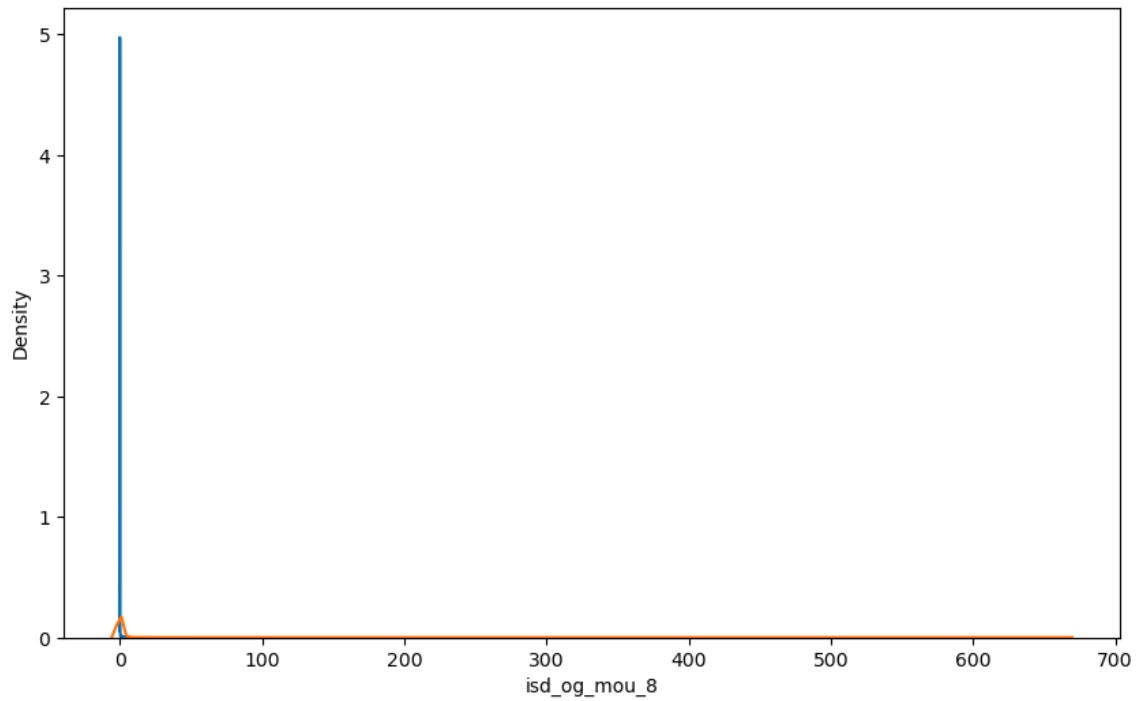
**Plots of important predictors for churn and non  
churn customers**



**Interpretation:**

We can see that for the churn customers the minutes of usage for the month of August is mostly populated on the lower side than the non churn customers.

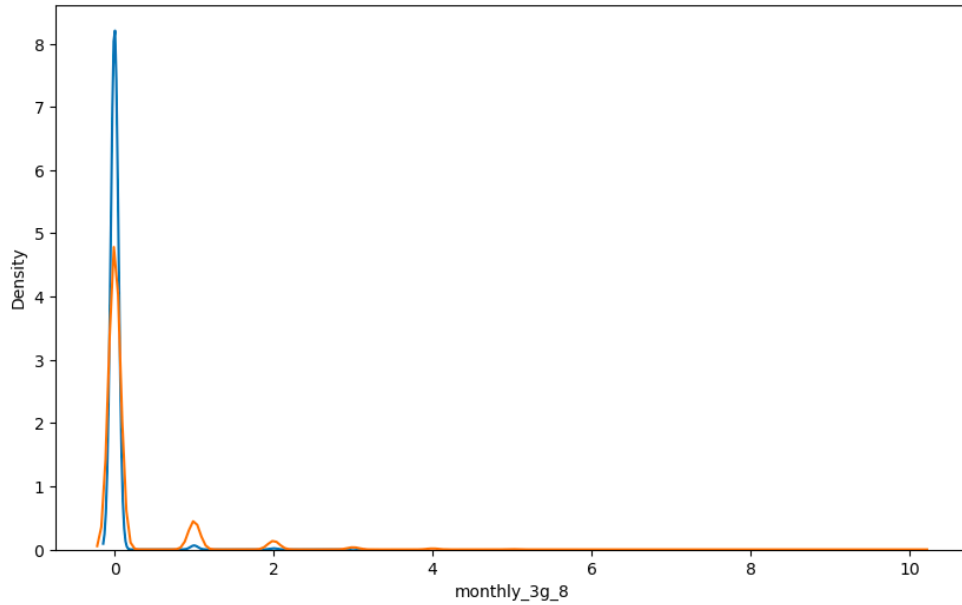
## Plotting the ROC Curve



## ROC Curve

We can see that the ISD outgoing minutes of usage for the month of August for churn customers is densed approximately to zero. On the onther hand for the non churn customers it is little more than the churn customers.

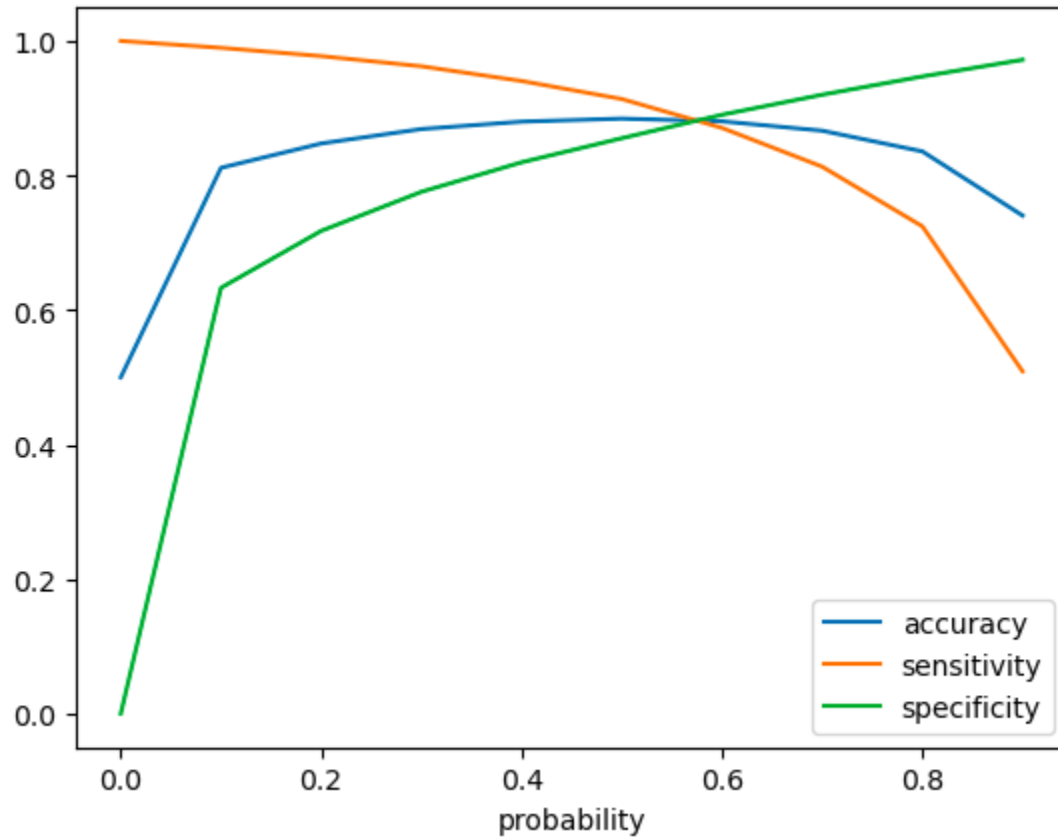
## Plotting the ROC Curve



## ROC Curve

The number of monthly 3g data for August for the churn customers are very much populated around 1, whereas for non-churn customers it is spread across various numbers.

## Plotting the ROC Curve



### Interpretation:

+ Accuracy - Becomes stable around 0.6

+ Sensitivity - Decreases with the increased probability.

+ Specificity - Increases with the increasing probability.

=> At point 0.6 where the three parameters cut each other, we can see that there is a balance between sensitivity and specificity with a good accuracy.

=> Here we are intended to achieve better sensitivity than accuracy and specificity. Though as per the above curve, we should take 0.6 as the optimum probability cutoff, we are taking \*0.5\* for achieving higher sensitivity, which is our main goal.





THANK YOU

