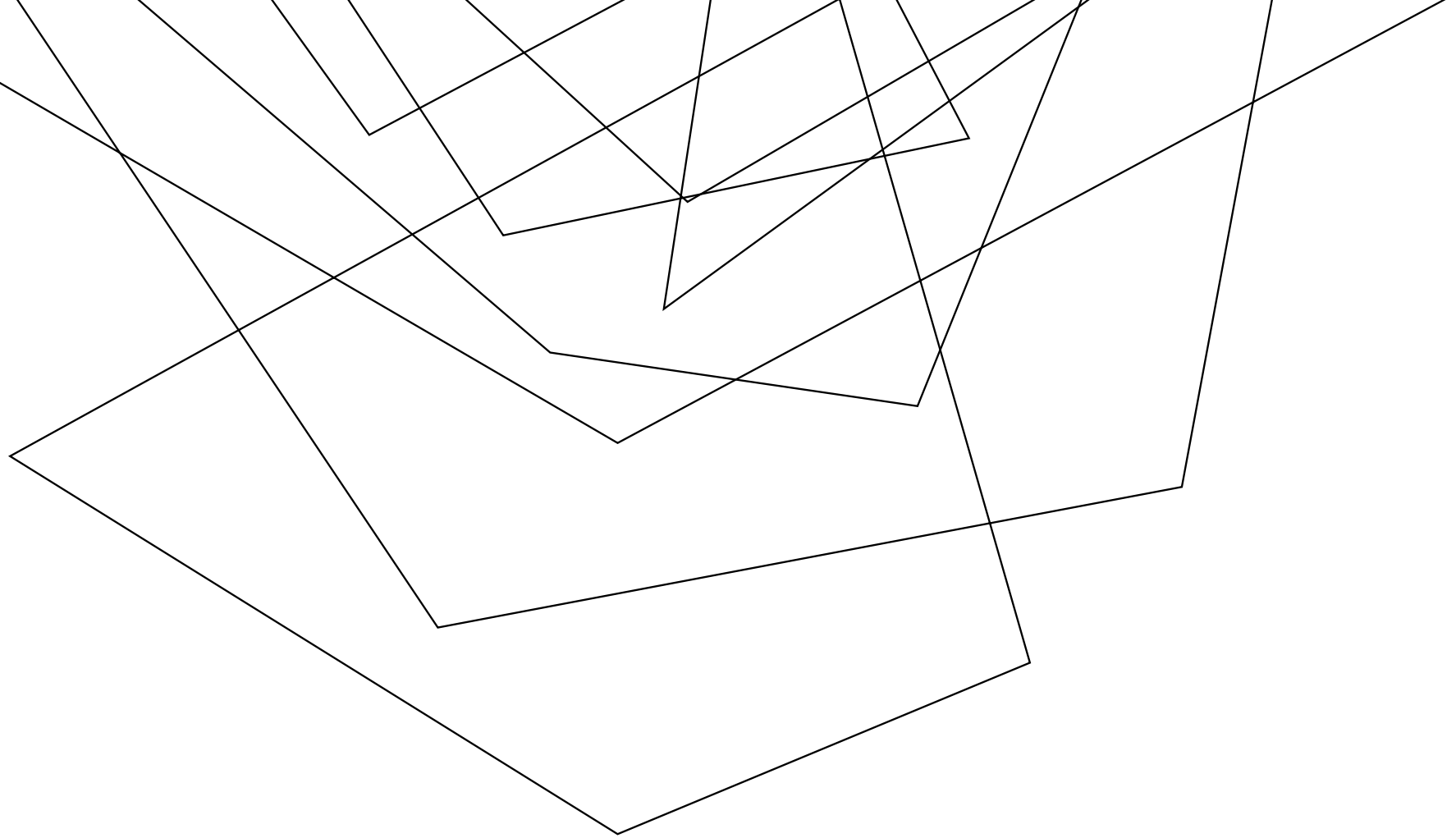


TELECOM CHURN CASE STUDY

Pompy Mukhopadhyay

E Hemanth Kumar

Preeti Gupta

AGENDA

Introduction

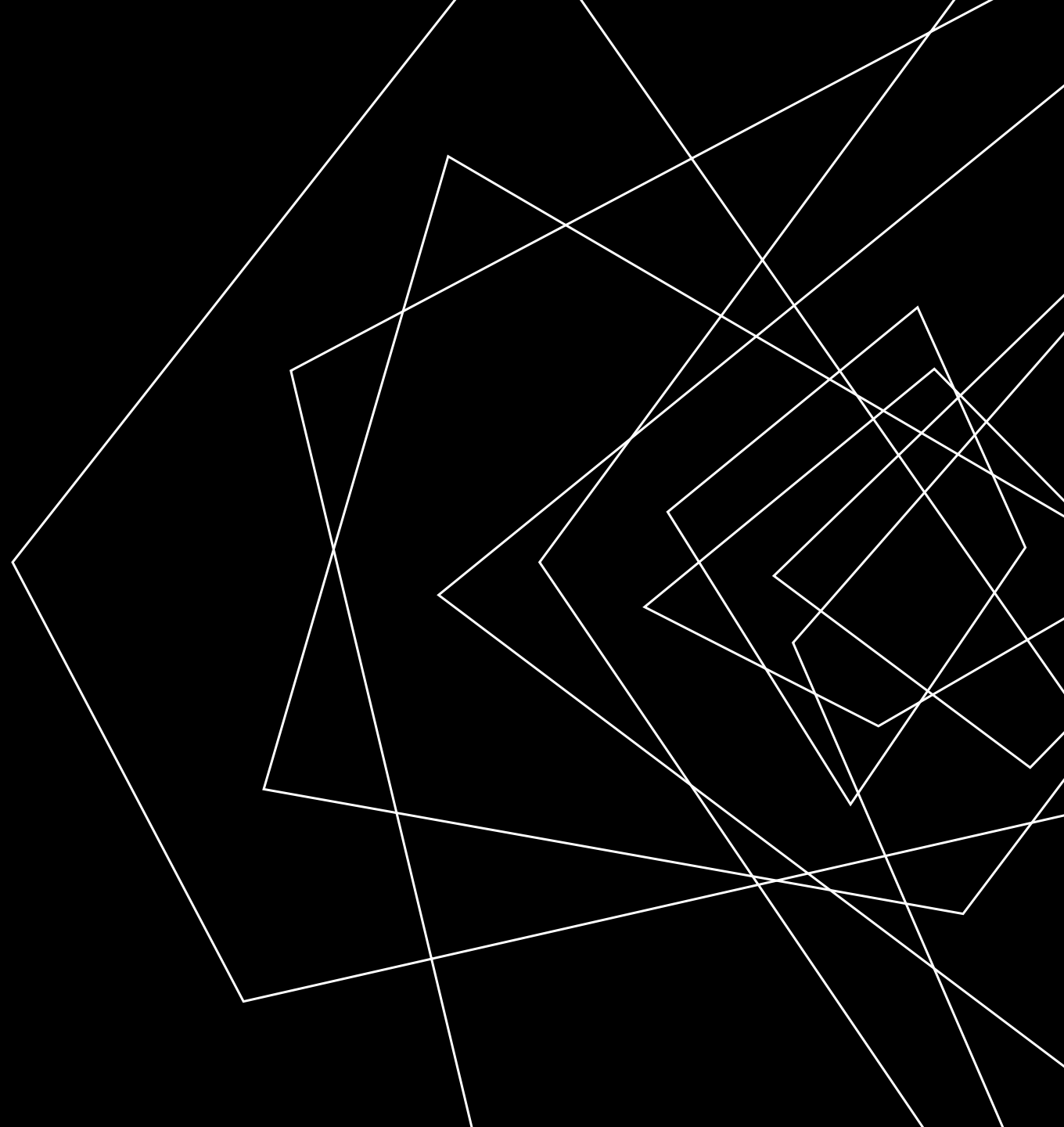
Problem definition

Cleaning & EDA

Modelling

Random forest

Summary

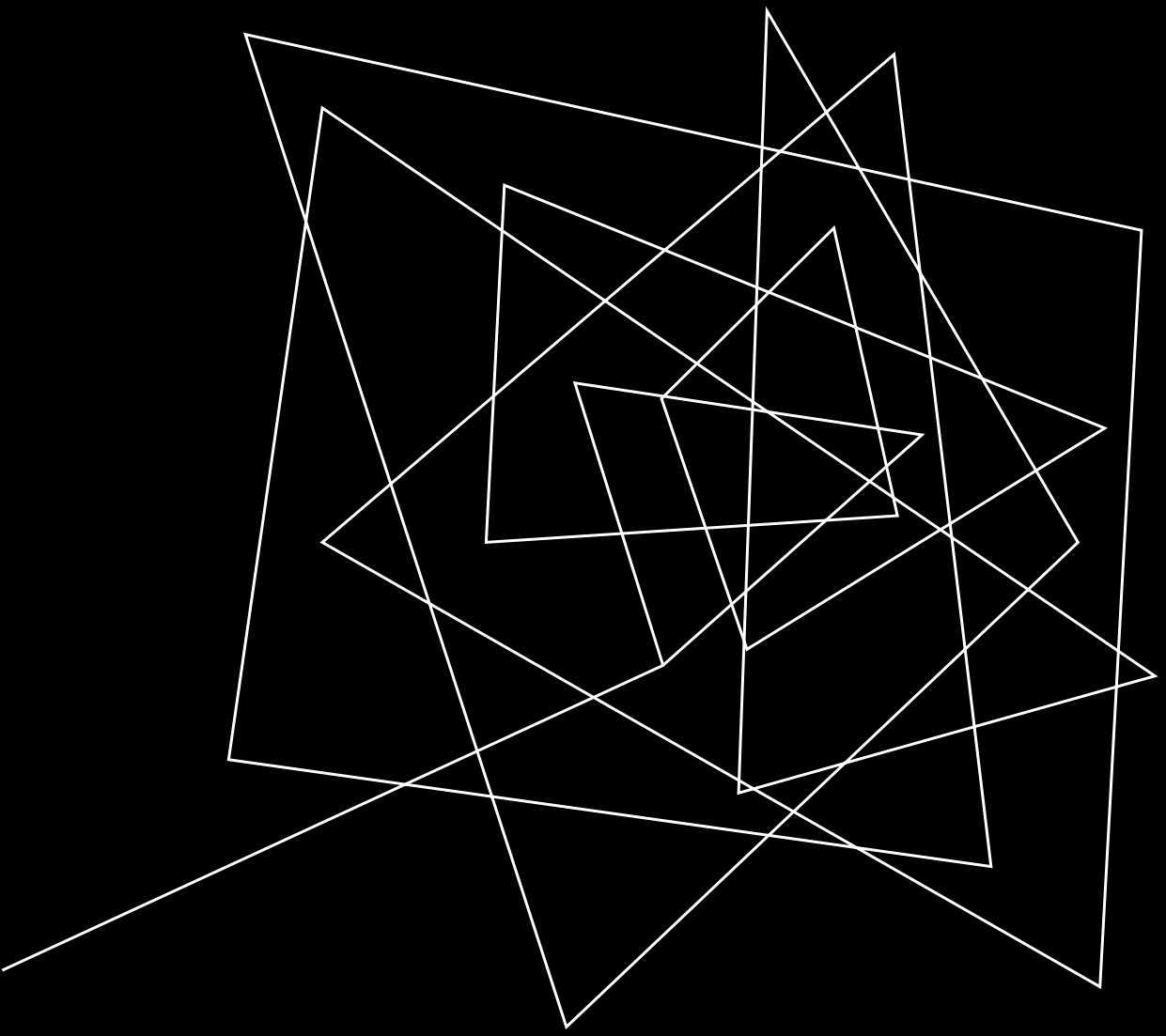


INTRODUCTION

The rivalry in the telecom industry in India and south Asia has greatly risen as a result of the numerous rising telecom beasts. Customers have a large selection of service providers to choose from, thus there are many chances that they will stop using the services of the current service providers. In the telecom sector, this is referred to as Churn. As a result, it has a direct impact on a company's revenue contribution.

Therefore, maintaining consumer satisfaction with the services becomes even more crucial. It is less expensive to keep an existing customer than to find new ones. Therefore, it is crucial for a business to concentrate on offering deals and offering quality services.

PROBLEM DEFINITION



“To predict the customers who are going to churn”

- In essence, it is a binary classification problem, and the outcome will be 1 if the problem is solved successfully.
- The cost of a misclassification, particularly false negatives, can be very substantial because, if a client is forecasted to leave but actually stay, it will have a direct impact on revenue loss.
- Output probabilities are crucial to compute since the probabilistic output may help determine the actual chances of being churned.
- The time latency of prediction is not a major concern because this model does not deal with immediate output generation.
- As we need to know what aspects contribute to consumer pleasure, interpretability is also crucial. But it won't directly benefit us.

What does each feature actually mean?

1)**Mobile Number of each user**

2)**Circle Id:** 109 only one region's data is there

3)**Dates:** Various recharge dates are available month wise.

4)**ARPU**: Average Revenue Per user

5)**MOU**: Time in minutes of user

6)**Roaming, ISD, STD, Local, Special, and other Recharge information**

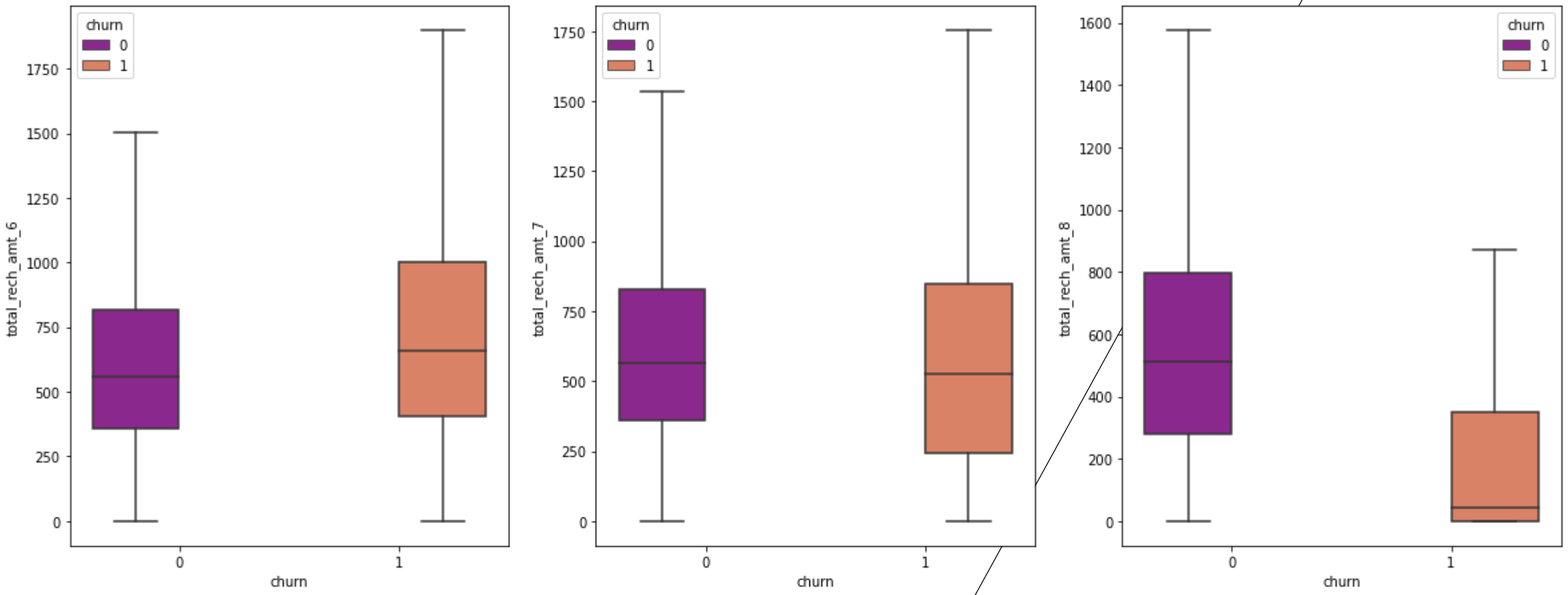
7)**Recharge amounts of various months, types, Maximum amount, & Last Day Recharge Amount, etc.**

8)**Data usage information of different types like Night pack, FB, etc.**

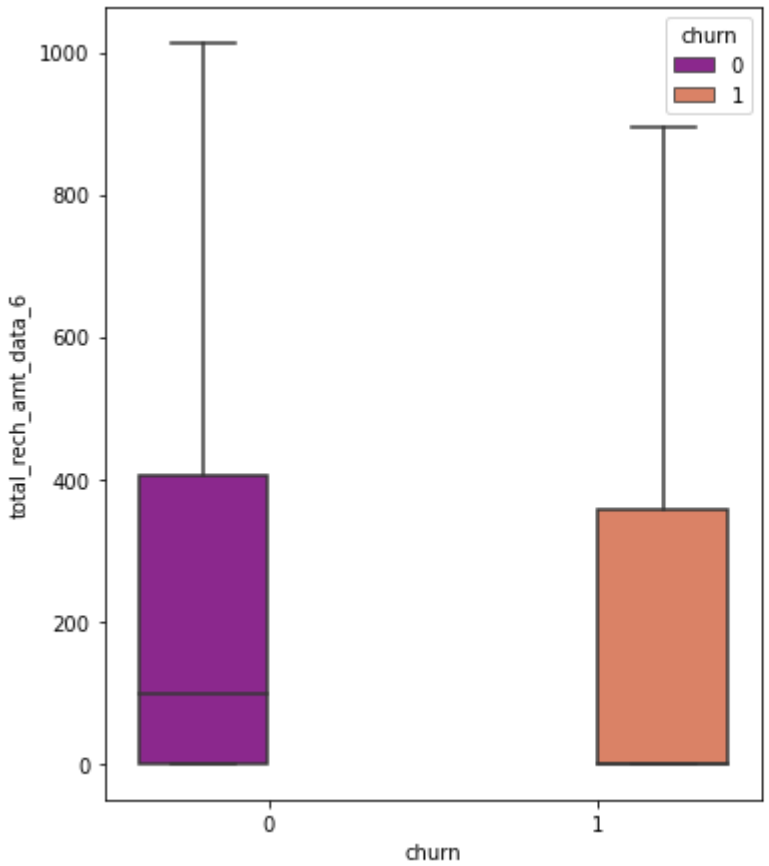
9)**AON: Age of Network**

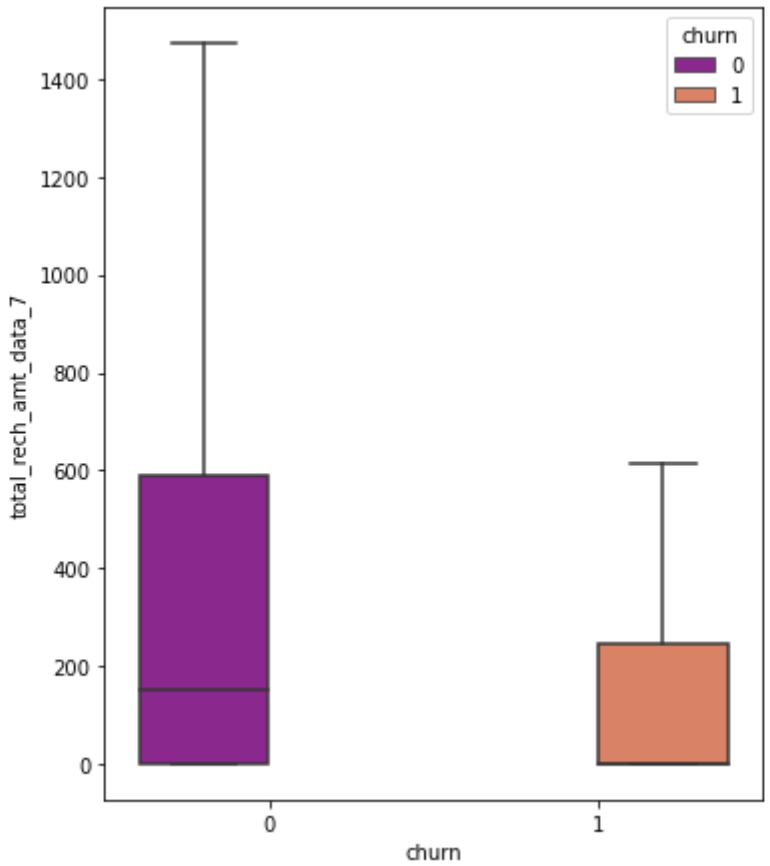
DATA CLEANING AND EXPLORATORY DATA ANALYSIS

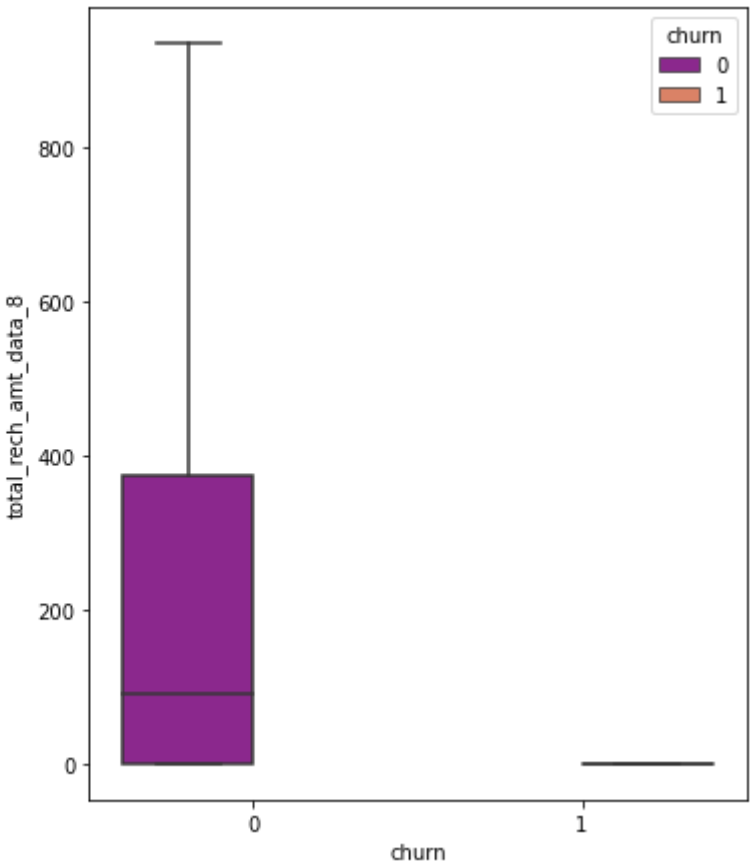
We can see a drop in the total recharge amount for churned customers in the 8th Month (Action Phase)



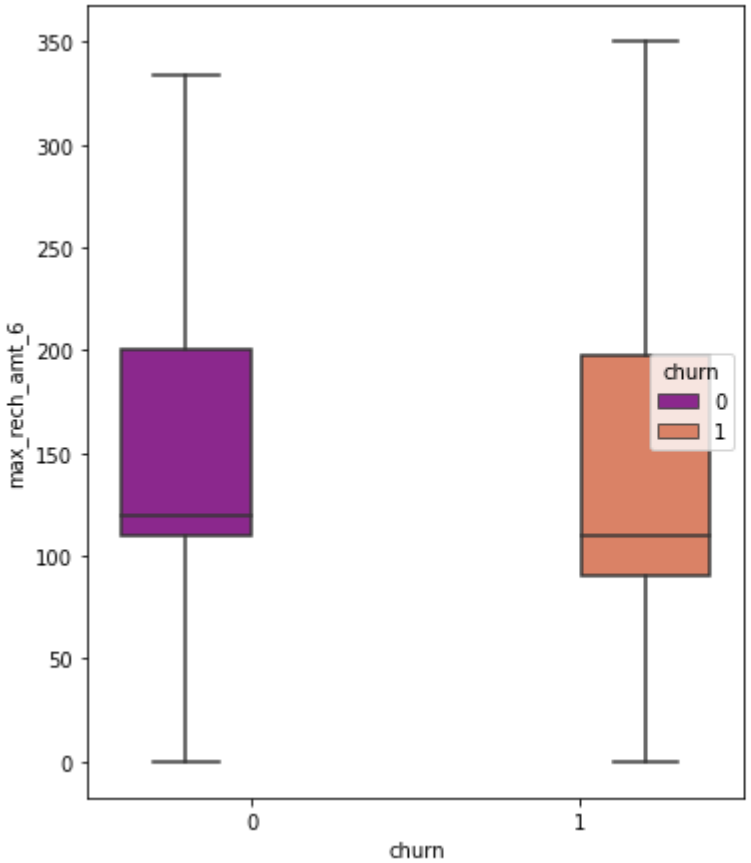
WE CAN SEE THAT THERE IS A HUGE DROP IN TOTAL RECHARGE AMOUNT FOR DATA IN THE 8TH MONTH (ACTION PHASE) FOR CHURNED CUSTOMERS

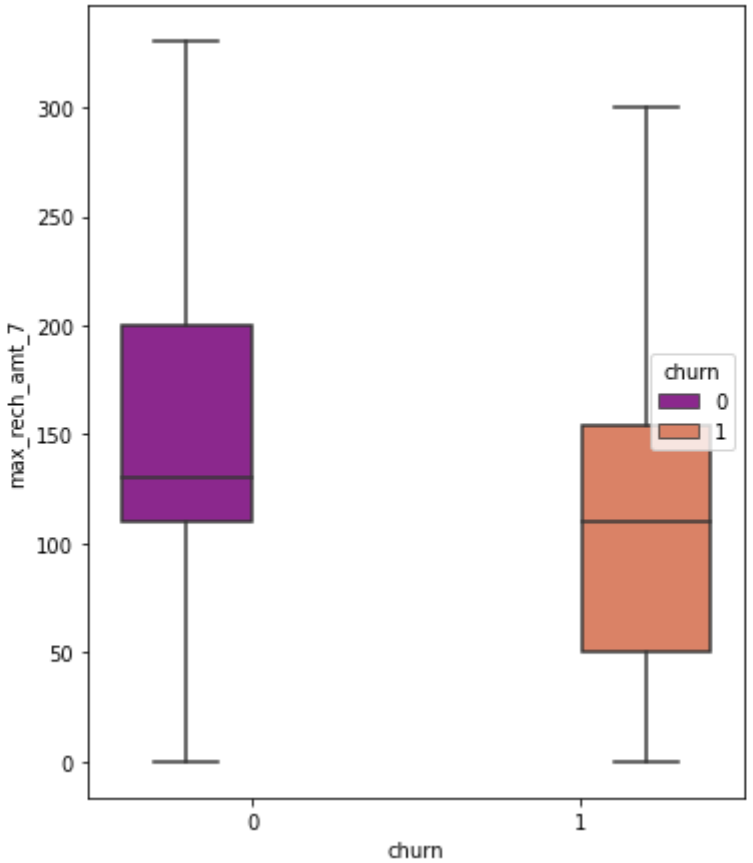


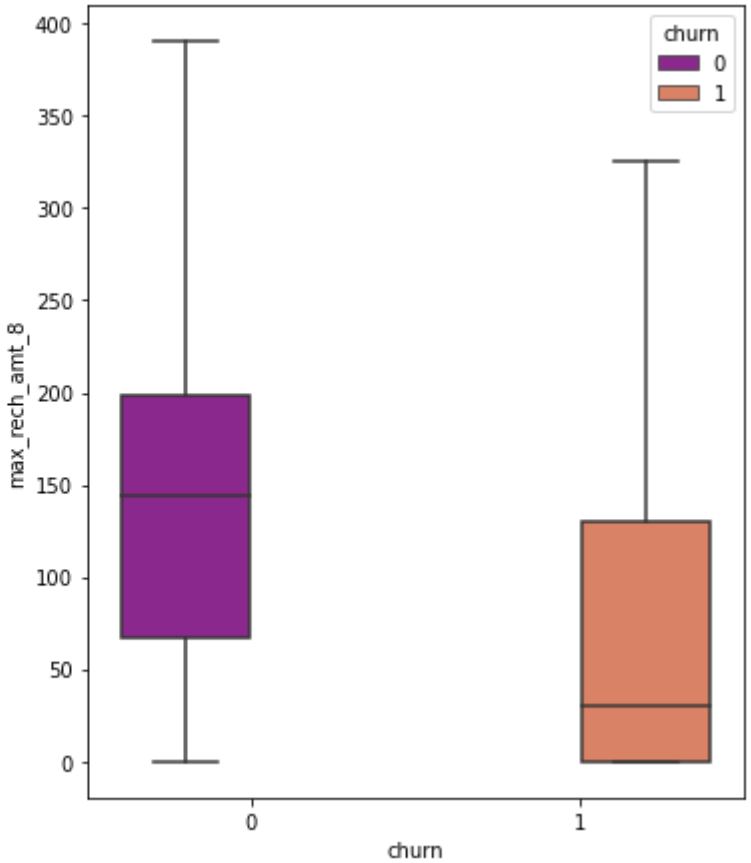




WE CAN SEE THAT THERE IS A HUGE DROP IN MAXIMUM RECHARGE AMOUNT FOR DATA IN THE 8TH MONTH (ACTION PHASE) FOR CHURNED CUSTOMERS.





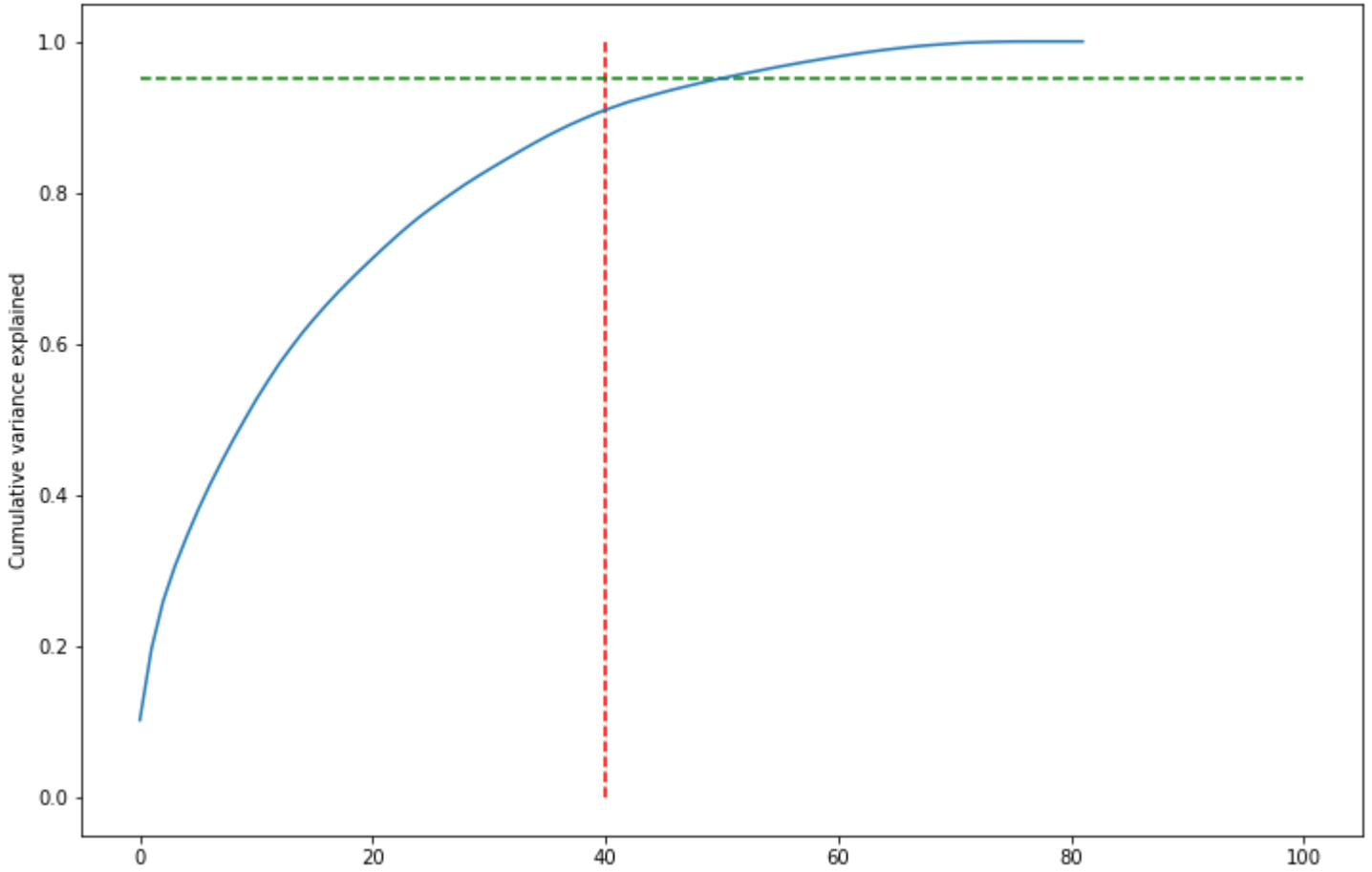


MODELLING

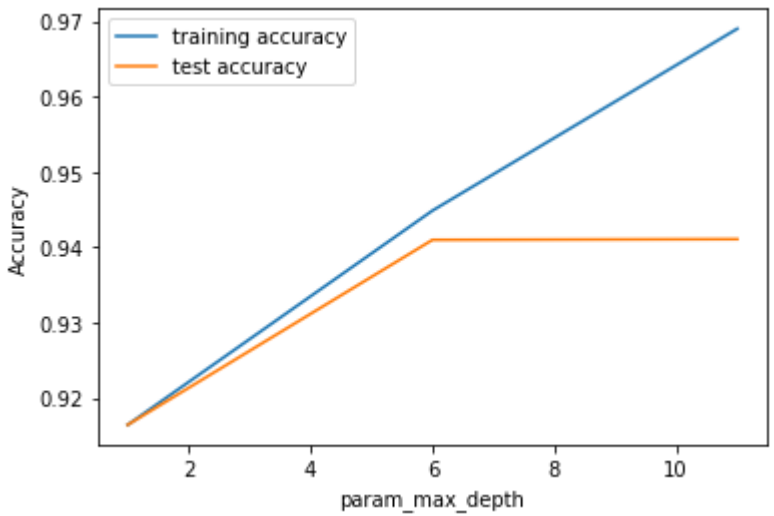
- Scree plot for the explained variance

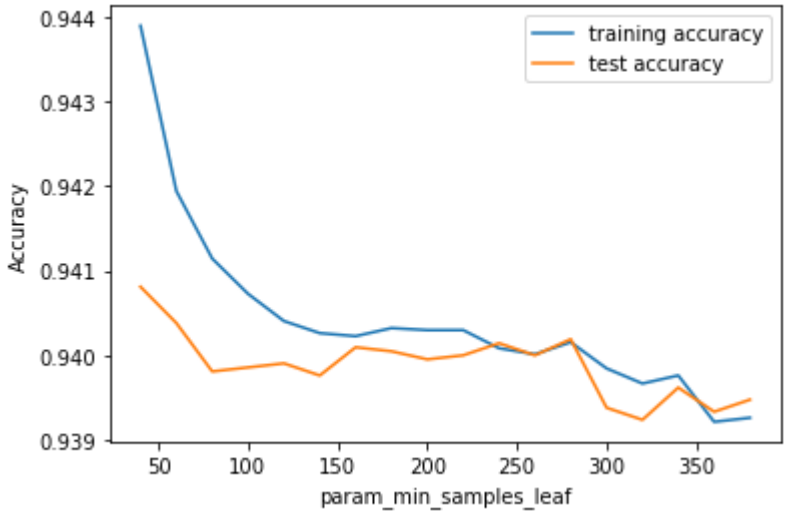
Looks like 45 components are enough to describe 95% of the variance in the data set.

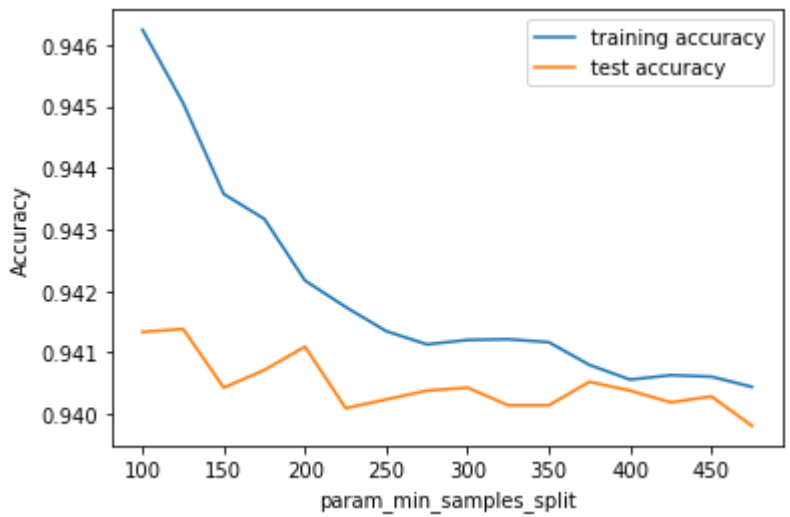
We'll choose 45 components for our modeling



RANDOM FOREST







Analysis:

We can see that as we increase the value of max_depth, both train and test scores increase till a point, but after that test score become stagnant. The ensemble tries to overfit as we increase the max_depth. Thus, controlling the depth of the constituent trees will help reduce overfitting in the forest. 6 value has peek convergents and can be used for grid view search.

We can see that the model starts to overfit as value is decrease the value of min_samples_leaf. 100 seems to be a good range and that will be used in grid search.

Score almost remain the same with very low dip throught the range. We will use 1000 for grid view search.

SUMMARY

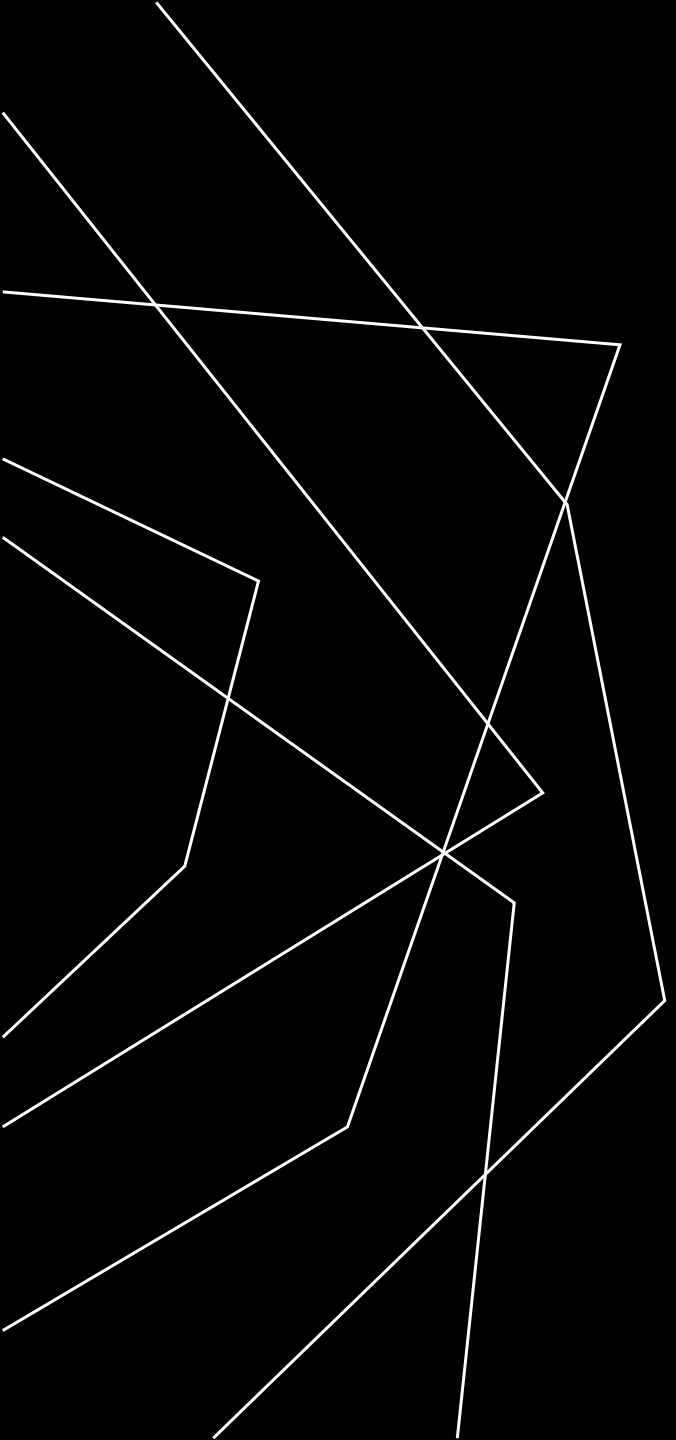
Business Insights

#Less number of high value customer are crunching but for last 6 month no new high valued customer is onboarded which is concerning, and company should concentrate on that aspect.

#Customers with less than 4 years of tenure are more likely to churn and company should concentrate more on that segment by rolling out new schemes to that group.

#Average revenue per user seems to be most important feature in determining churn prediction.

#Incoming and Outgoing Calls on roaming for 8th month are strong indicators of churn behaviour



THANK YOU

Pompy Mukhopadhyay

E Hemanth Kumar

Preeti Gupta