Telecom-Churn Case Study

to predict which customers are at high risk of churn.

Group Members:

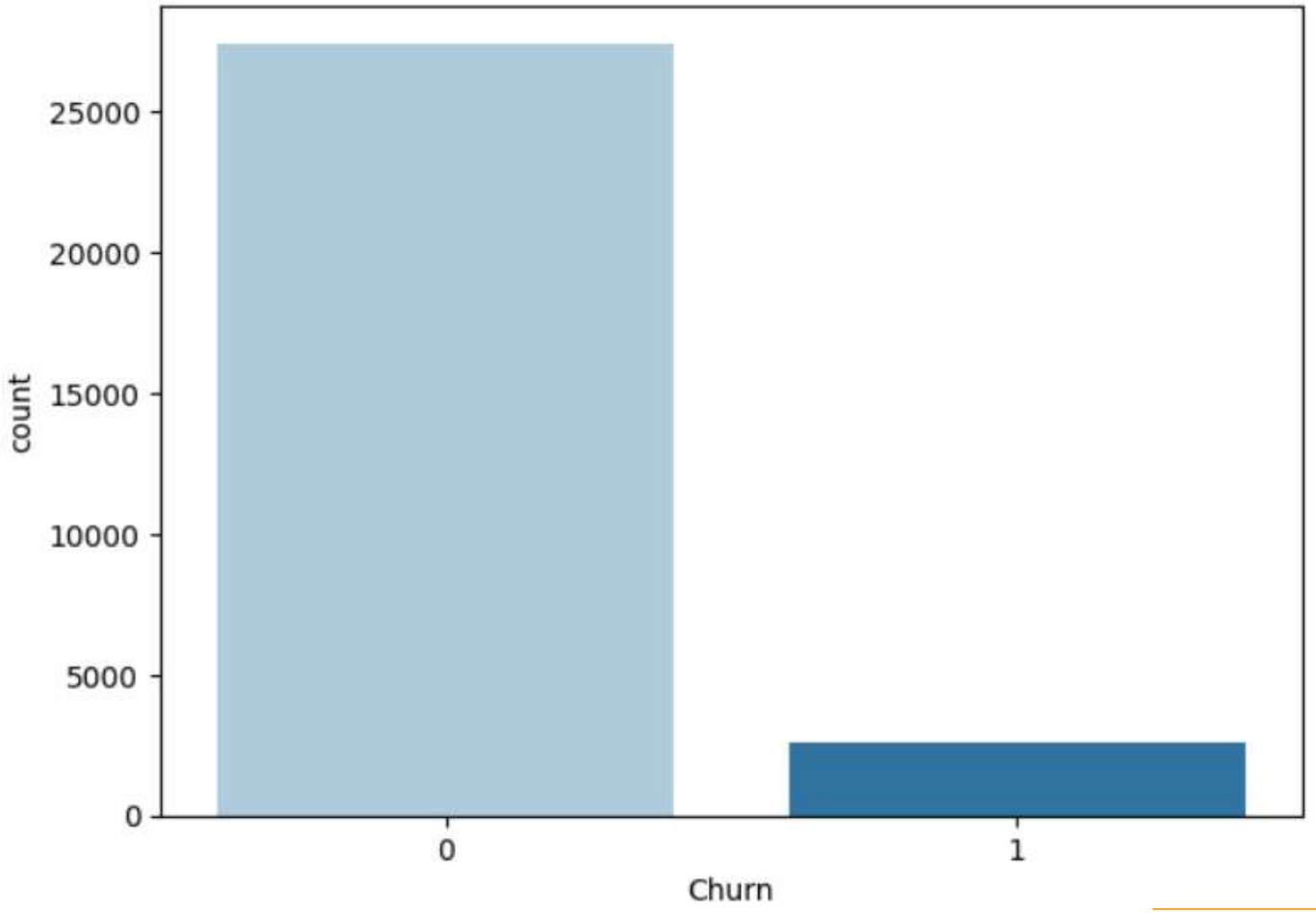
- Simran
- Drupad Muthyala

Problem Statement

- analyze customer-level data of a leading telecom firm, build predictive models to identify high-value customers based on certain metric at high risk of churn and identify the main indicators of churn.

Tagging the Churners- Class Imbalance

- 8.6% class imbalance of customers churned



Outlier Treatment

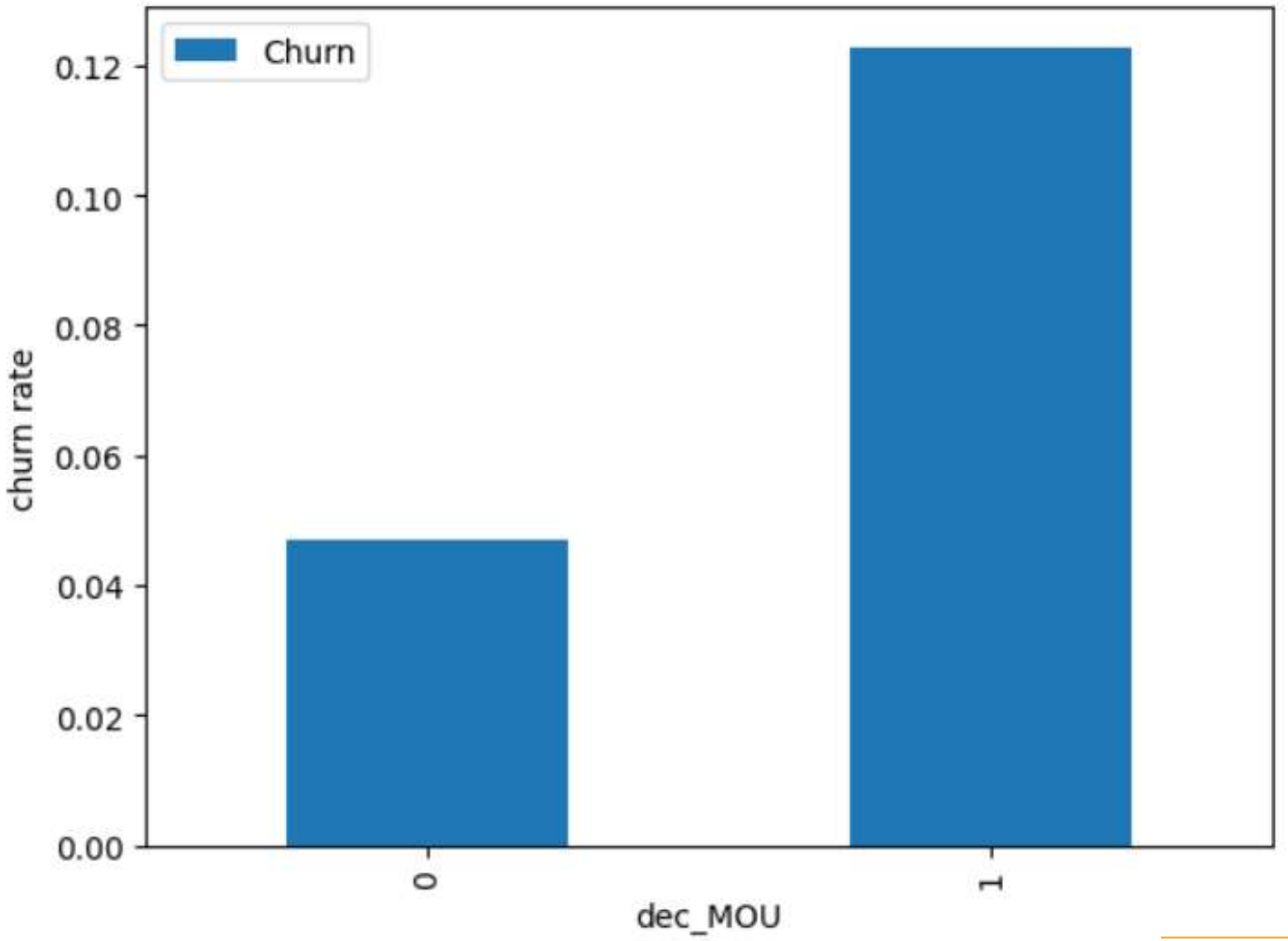
- Removed outliers with less than 10^{th} percentile and more than 90^{th} percentile. And added few columns for further insight on churn in future based on customer comparison to good and action phase.

EDA Analysis

- Univariate analysis
- Bivariate analysis

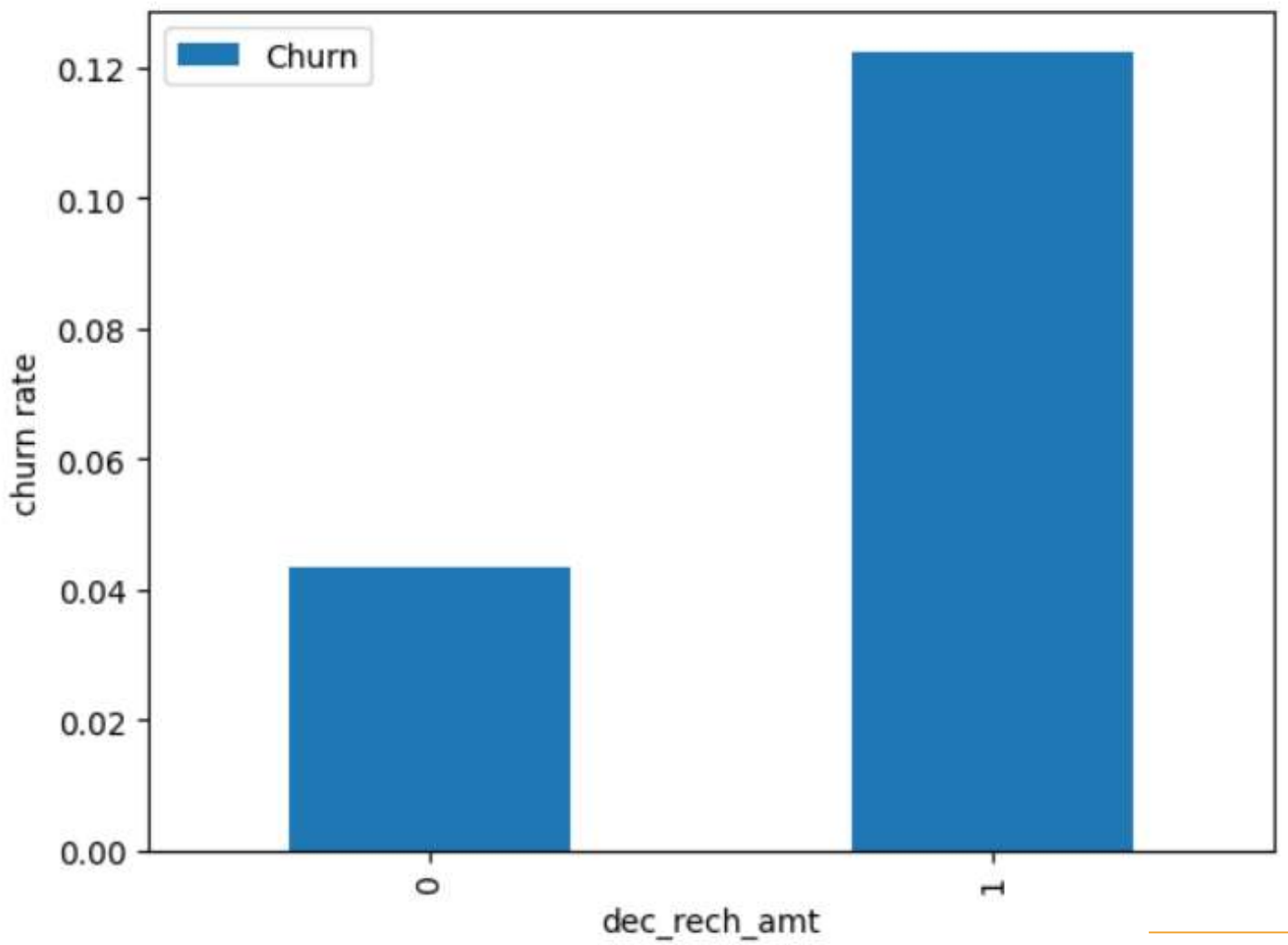
Churn rate based on whether the customer decreased his/her MOU in action month

-customer whose MoU decreased in action phase have higher churn rate than those whose minutes increased in good phase.



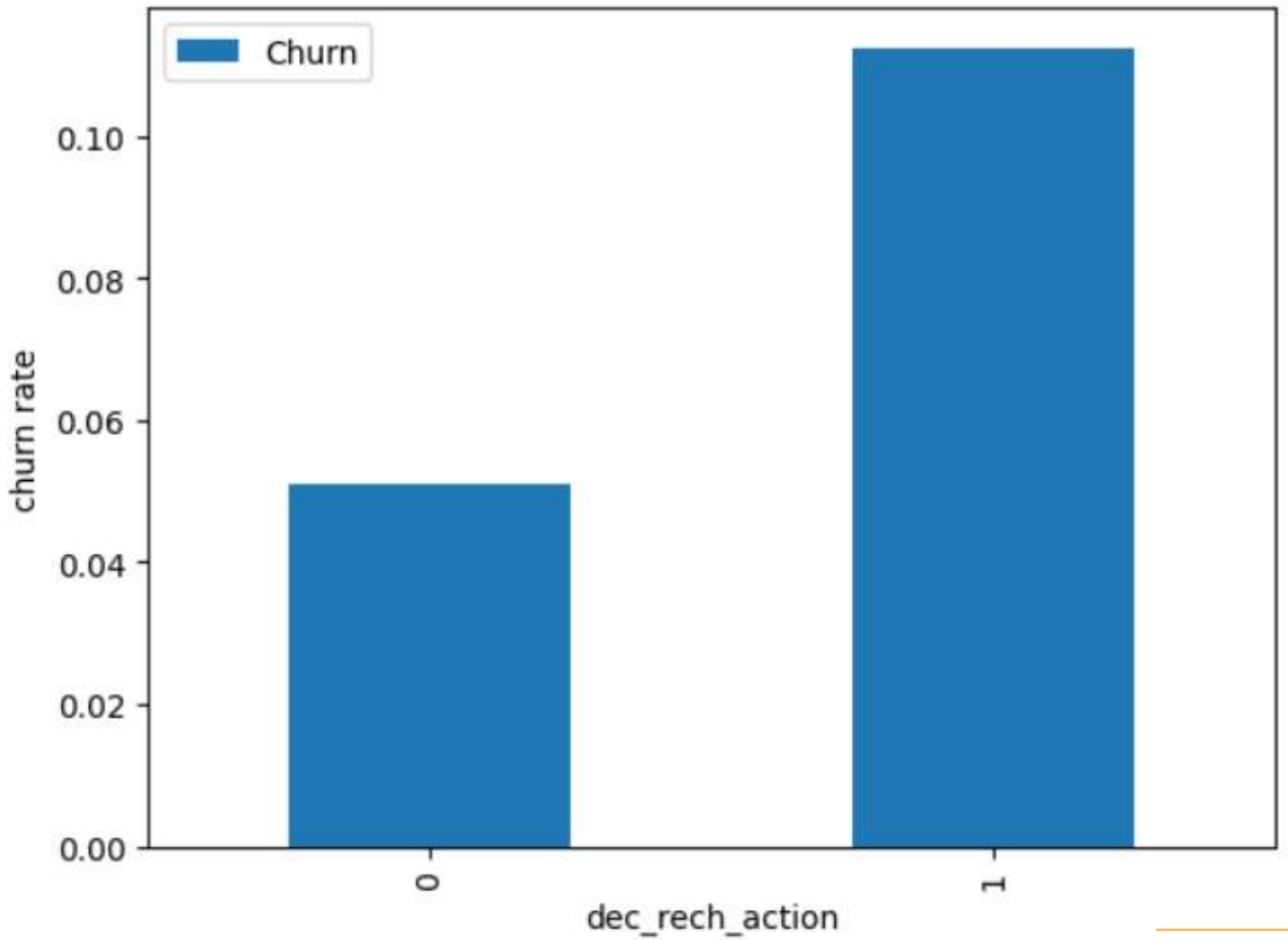
Churn rate based on whether customer decreased amount of recharge in action month

- **churn rate is greater for customers whose amount of recharge is less in action phase than the good phase**



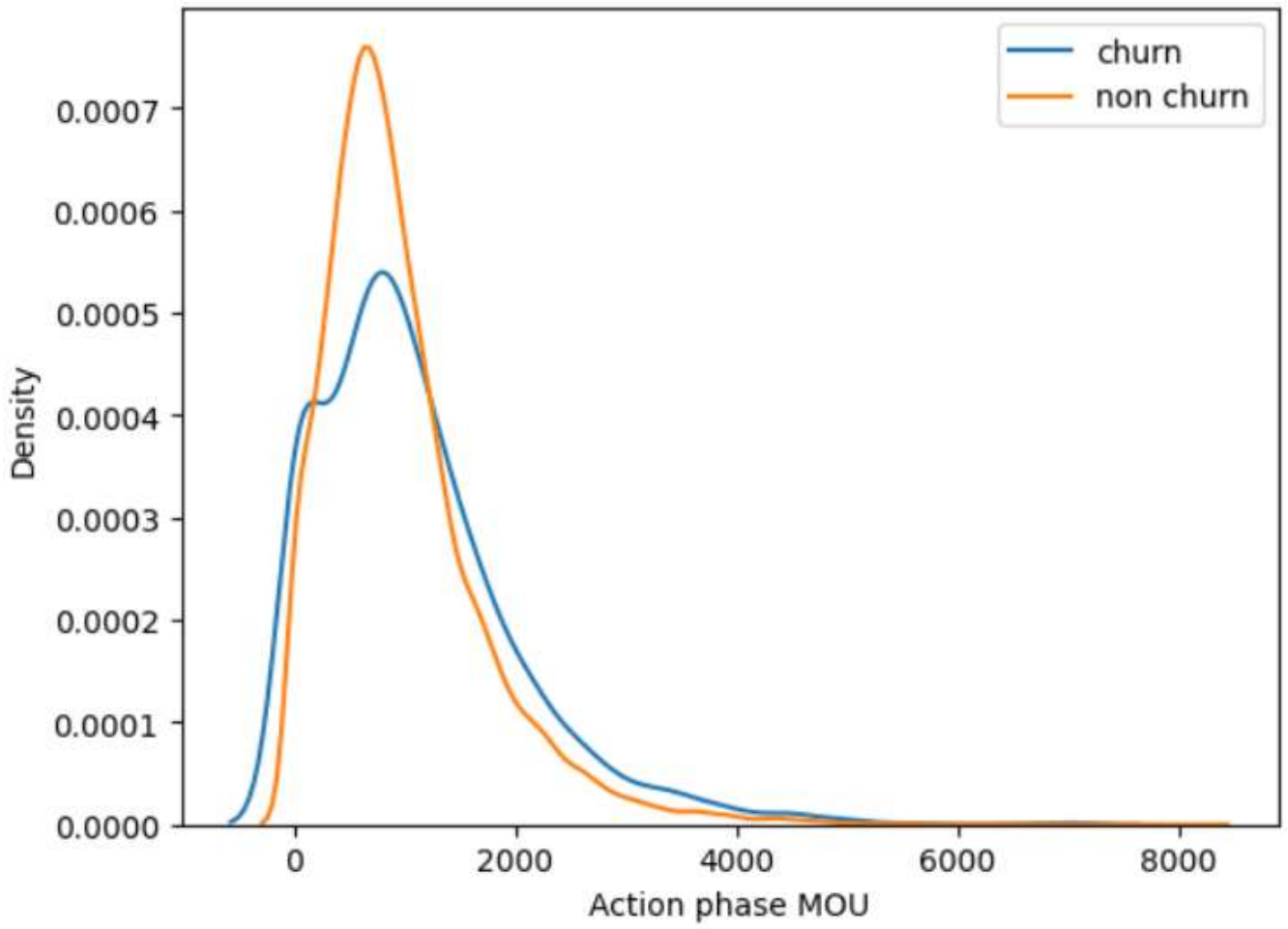
Churn rate based on number of recharge in action month

-the churn rate is more for the customers whose no. of recharge in action phase is lesser



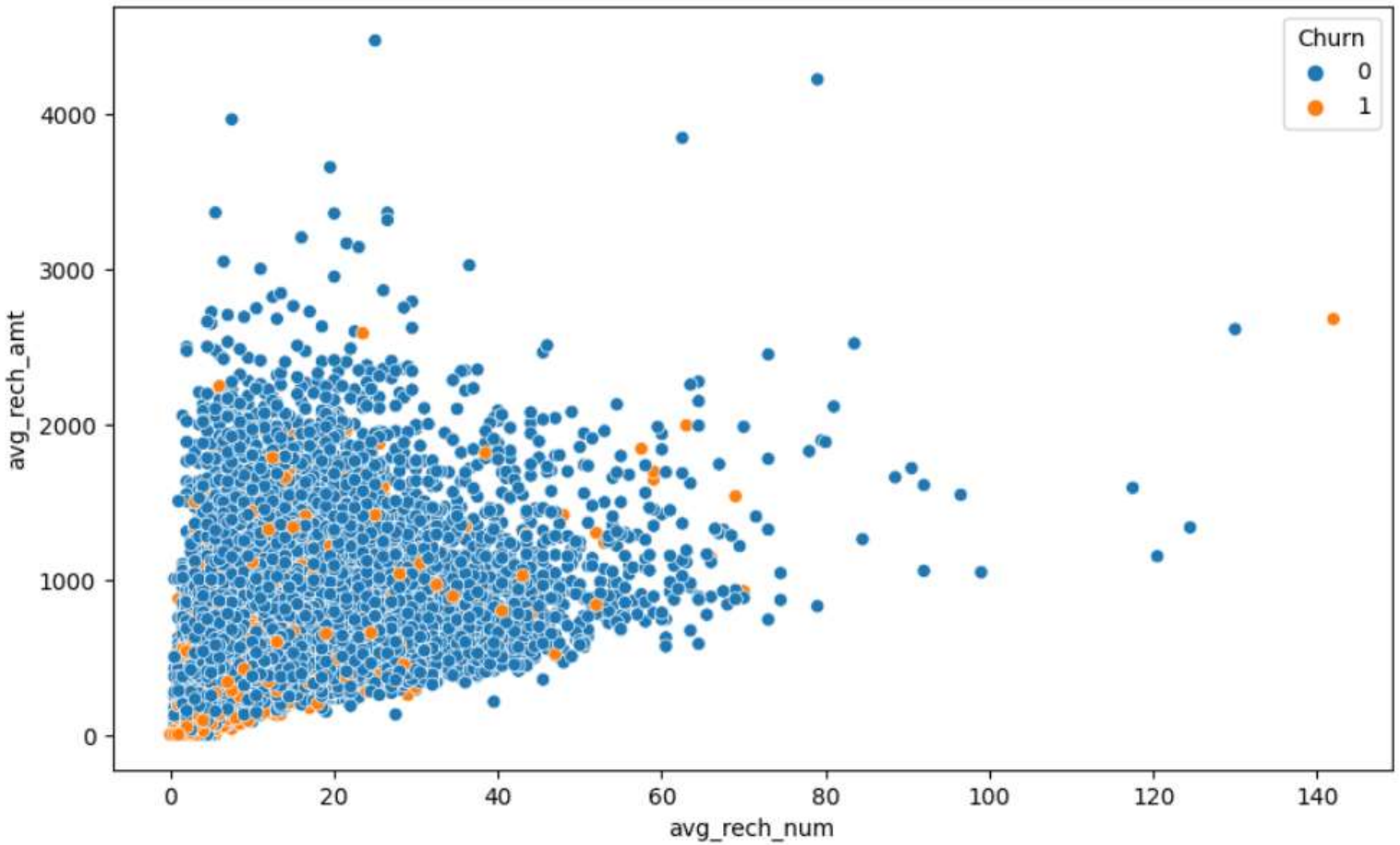
Analysis of the minutes of usage MOU (churn and not churn) in action phase

-MOU of churn customers is populated on the 0 to 2500 range. higher the MOU, lesser the churn probability



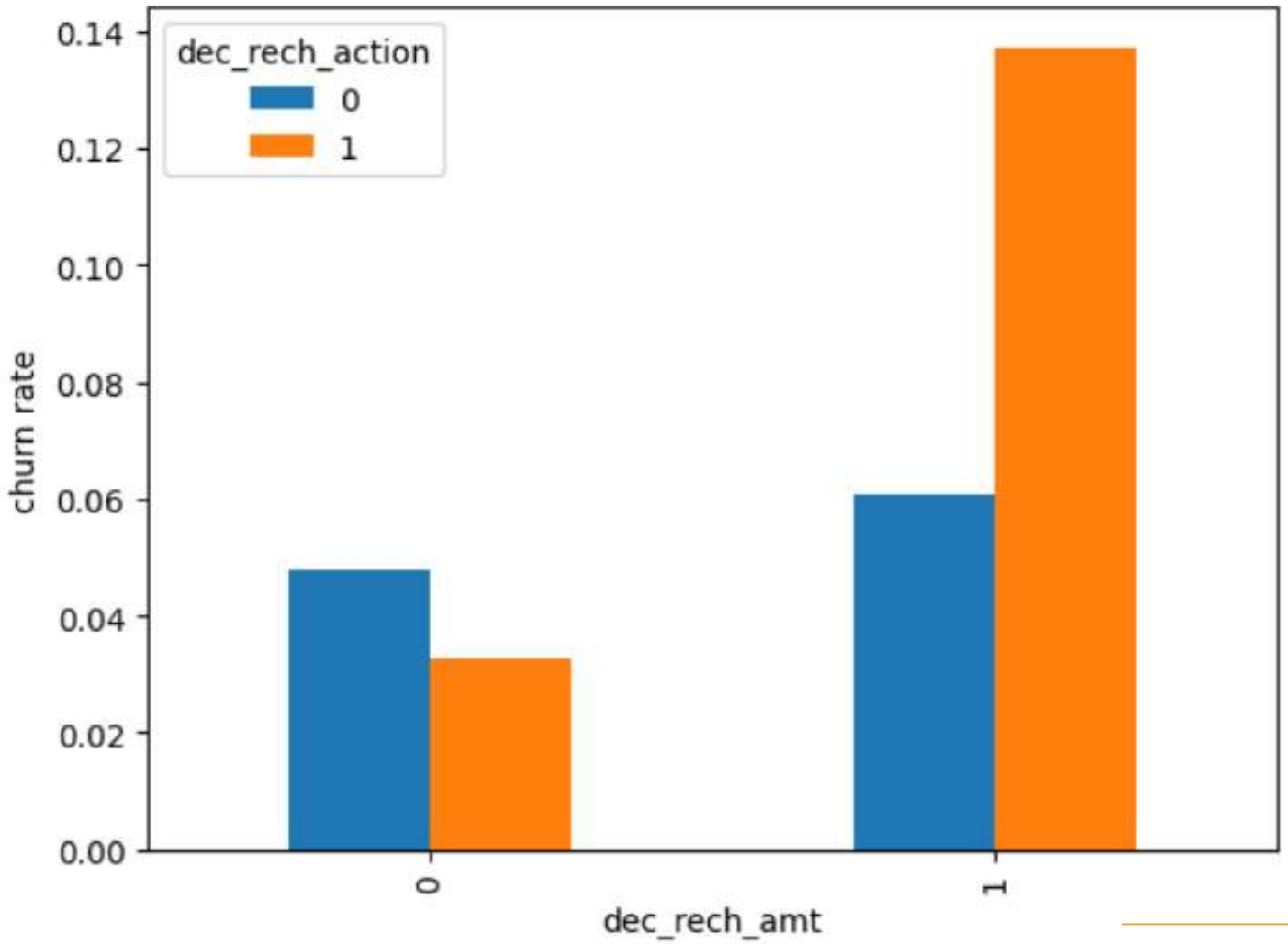
Analysing recharge amount and number of recharge in action month

-as can be seen from the pattern recharge number and recharge amount are proportional.



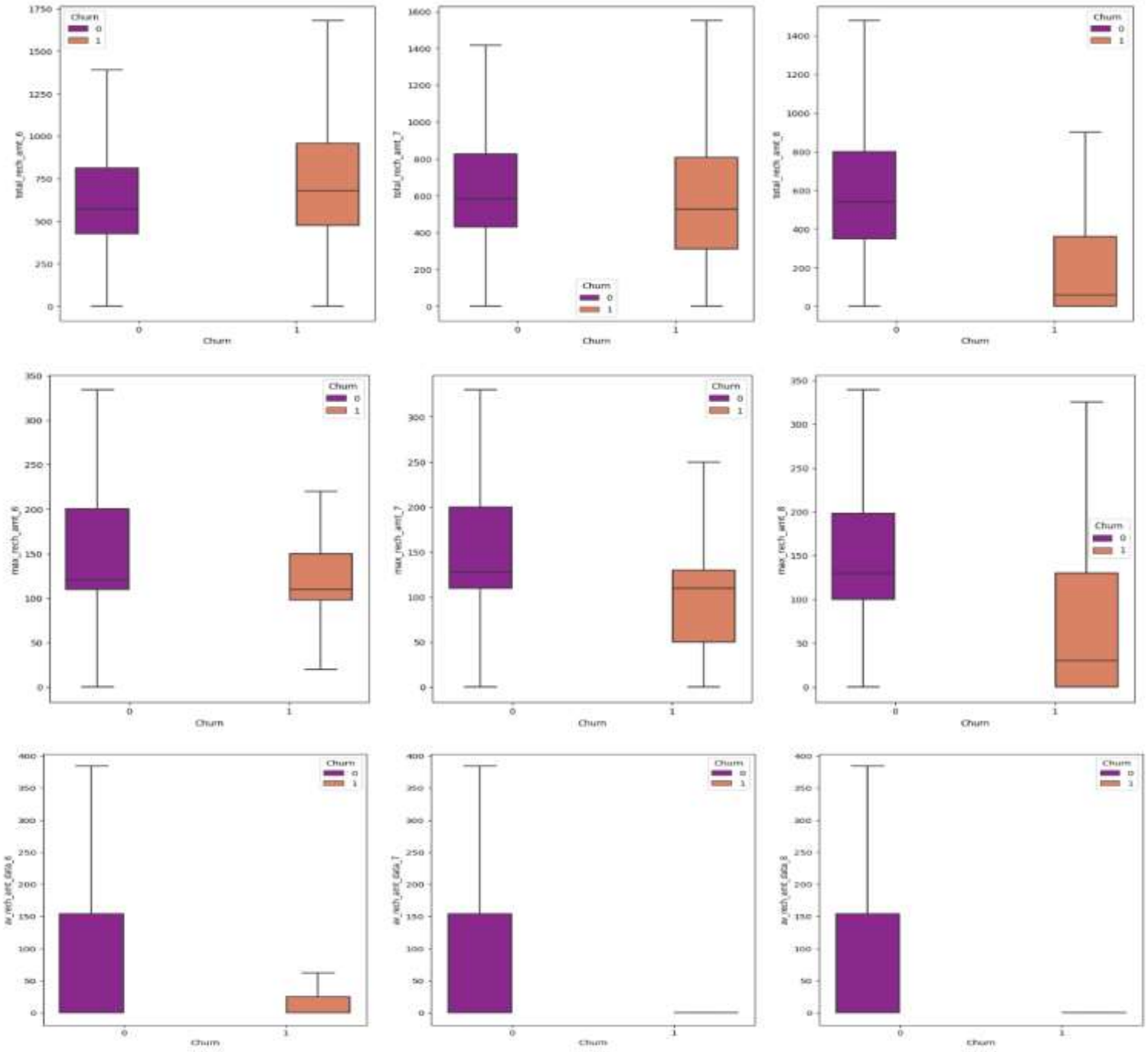
Analysing churn rate WRT the decreasing recharge amount and number of recharge during action phase.

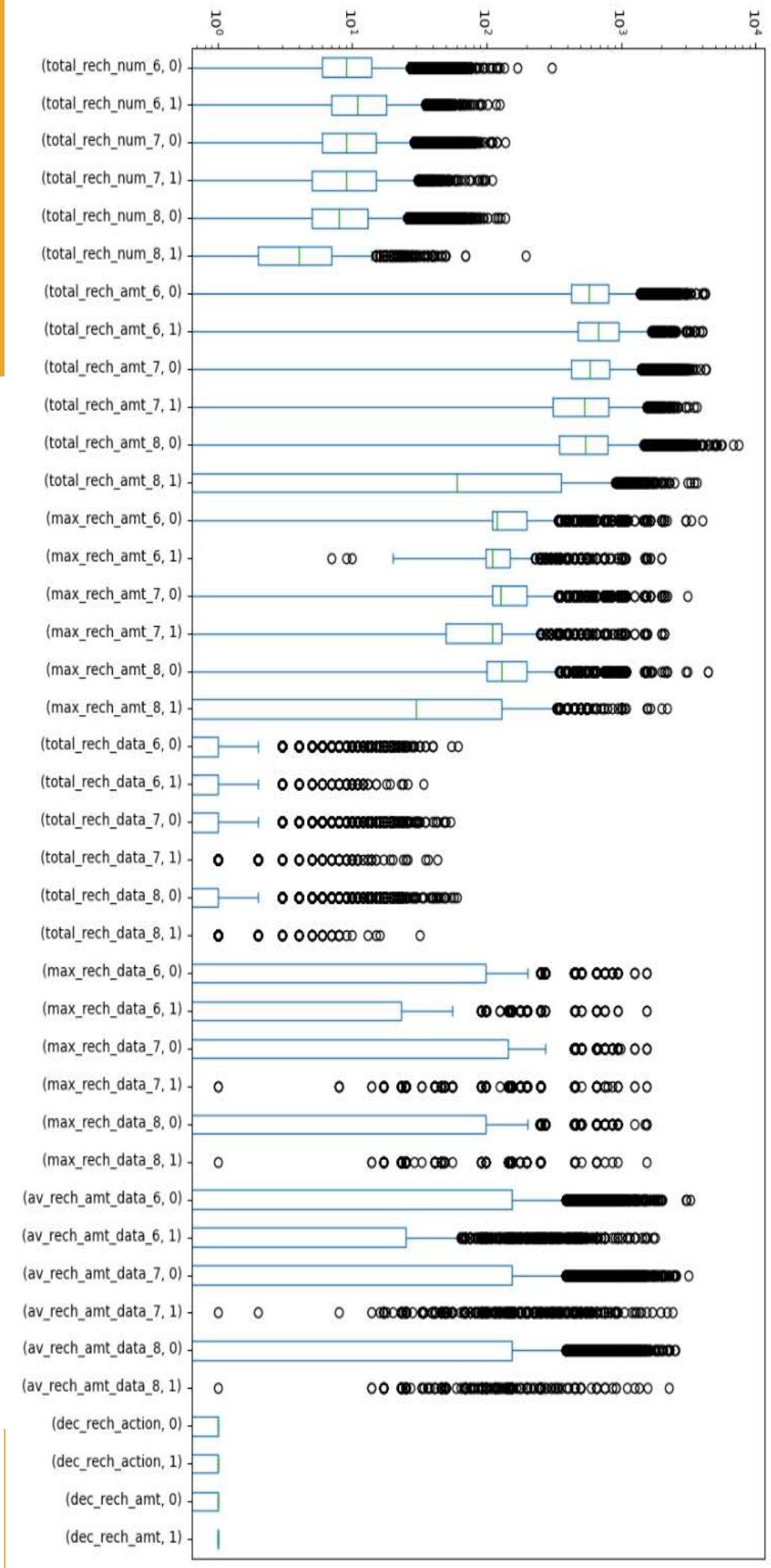
-churn rate is higher for the customers whose recharge amt as well as no. of recharge has decreased in action phase



Recharge amounts total and maximum for 6th, 7th and 8th month

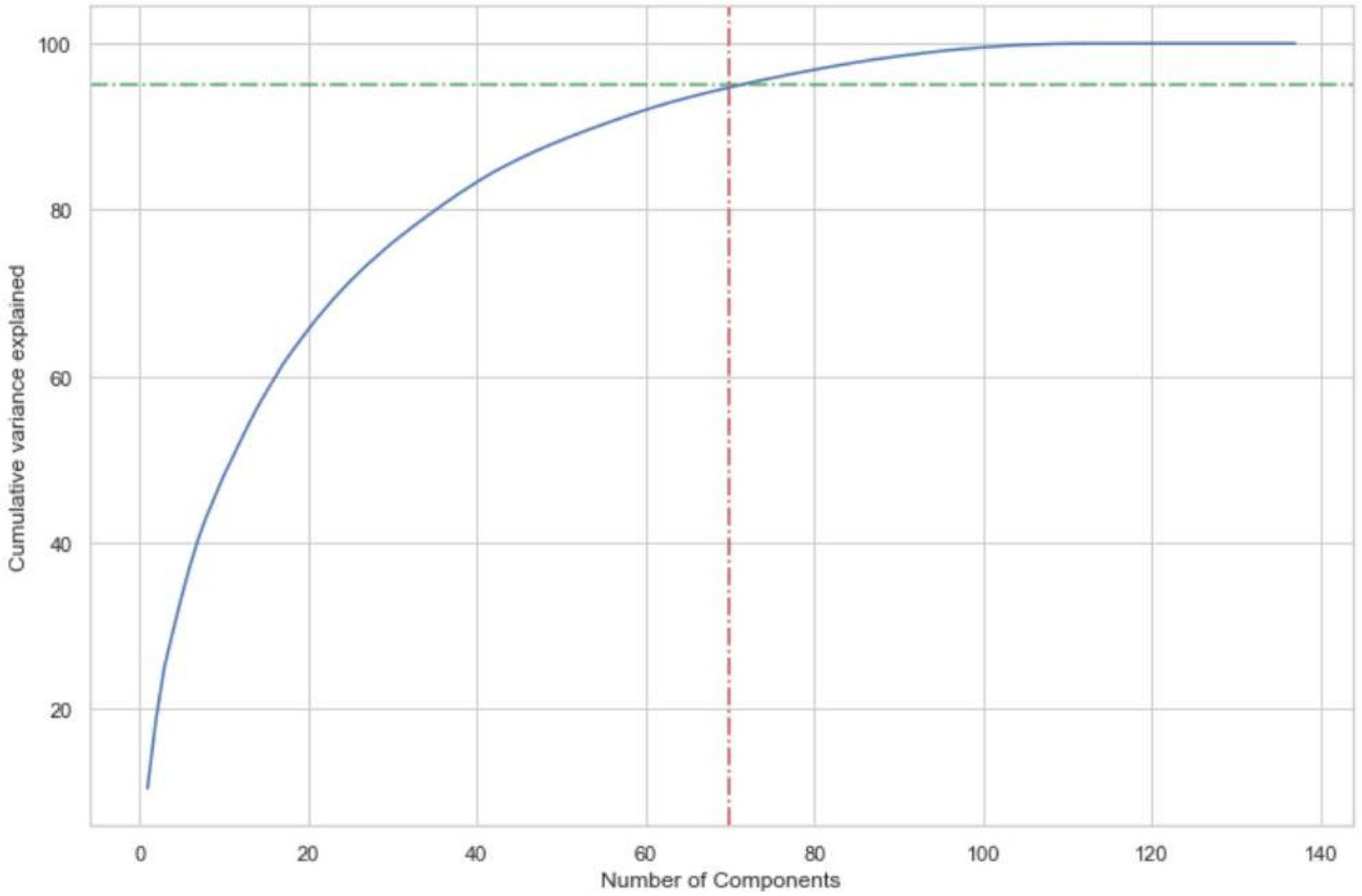
-from above plots its crystal clear that recharge amounts(total & maximum) started to fall in the 8th month meaning near to the churn phase



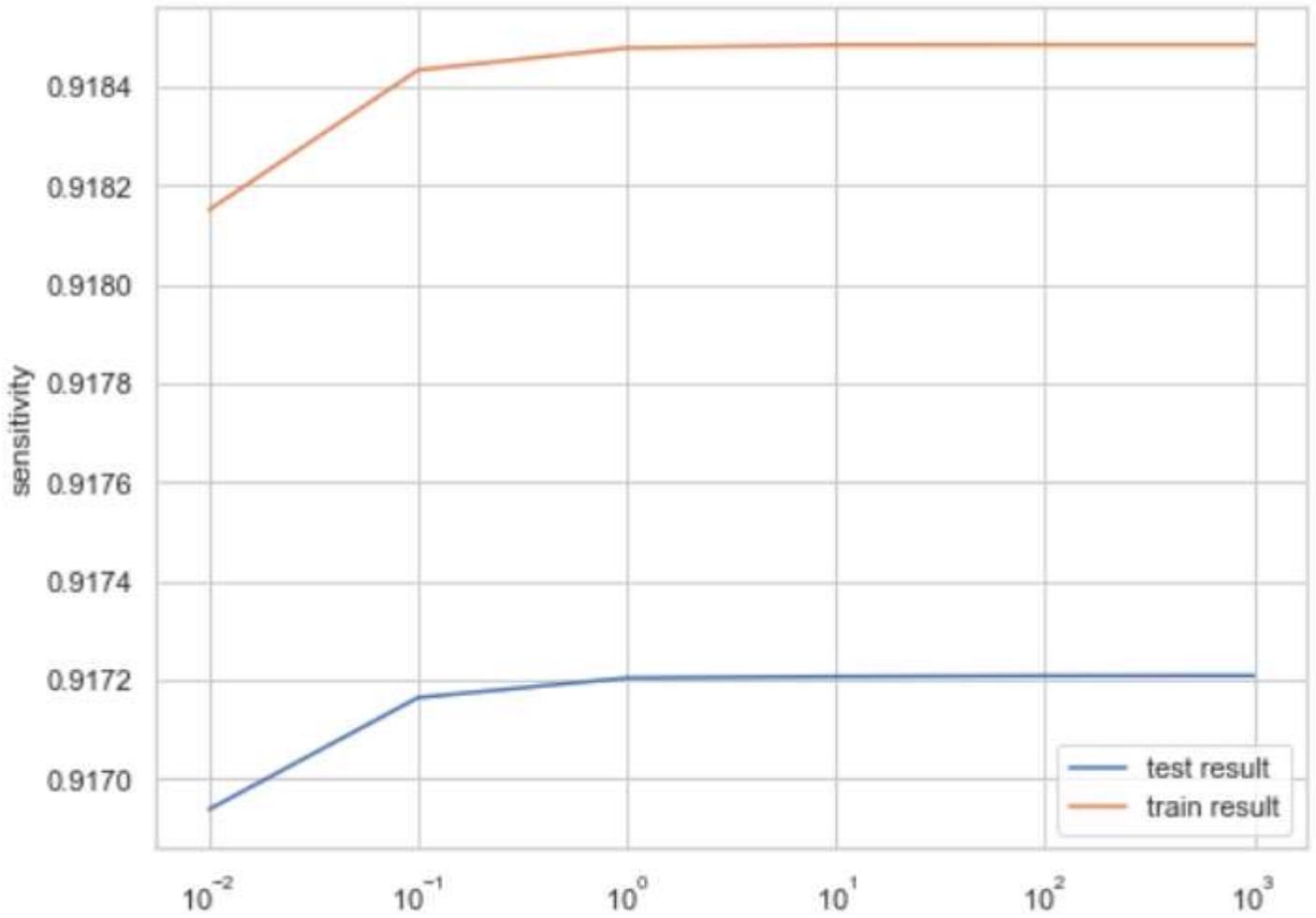

10^0
10^1
10^2
10^3
10^4
(total_rech_num_6, 0)
(total_rech_num_6, 1)
(total_rech_num_7, 0)
(total_rech_num_7, 1)
(total_rech_num_8, 0)
(total_rech_num_8, 1)
(total_rech_amt_6, 0)
(total_rech_amt_6, 1)
(total_rech_amt_7, 0)
(total_rech_amt_7, 1)
(total_rech_amt_8, 0)
(total_rech_amt_8, 1)
(max_rech_amt_6, 0)
(max_rech_amt_6, 1)
(max_rech_amt_7, 0)
(max_rech_amt_7, 1)
(max_rech_amt_8, 0)
(max_rech_amt_8, 1)
(total_rech_data_6, 0)
(total_rech_data_6, 1)
(total_rech_data_7, 0)
(total_rech_data_7, 1)
(total_rech_data_8, 0)
(total_rech_data_8, 1)
(max_rech_data_6, 0)
(max_rech_data_6, 1)
(max_rech_data_7, 0)
(max_rech_data_7, 1)
(max_rech_data_8, 0)
(max_rech_data_8, 1)
(av_rech_amt_data_6, 0)
(av_rech_amt_data_6, 1)
(av_rech_amt_data_7, 0)
(av_rech_amt_data_7, 1)
(av_rech_amt_data_8, 0)
(av_rech_amt_data_8, 1)
(dec_rech_action, 0)
(dec_rech_action, 1)
(dec_rech_amt, 0)
(dec_rech_amt, 1)

Model building with PCA

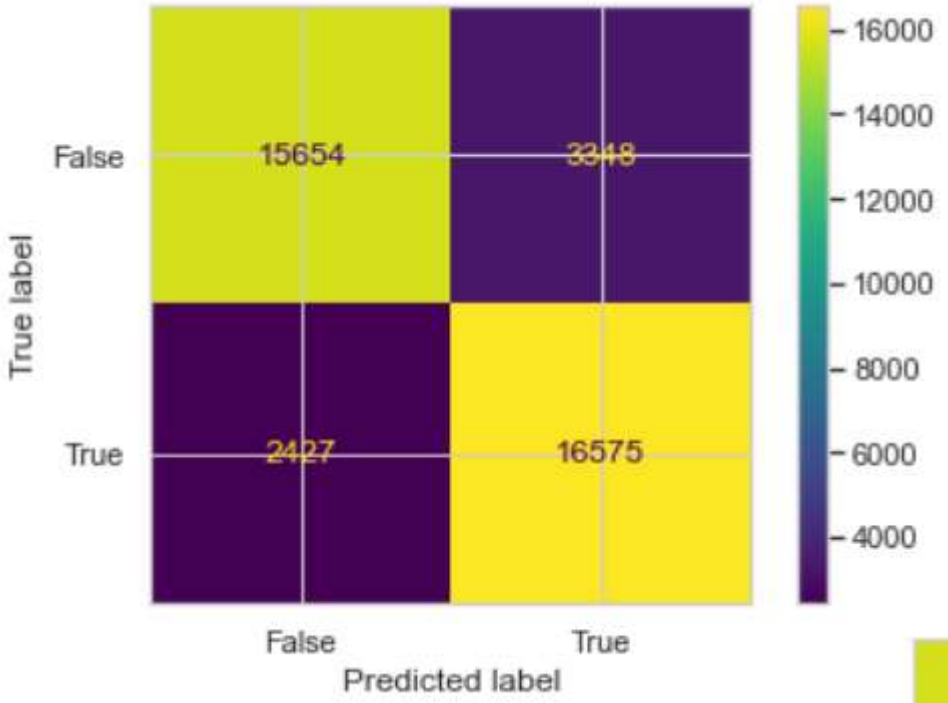
As per above looks like 70 components are enough to describe 95% of the variance in the dataset. We'll choose 70 components for our modeling

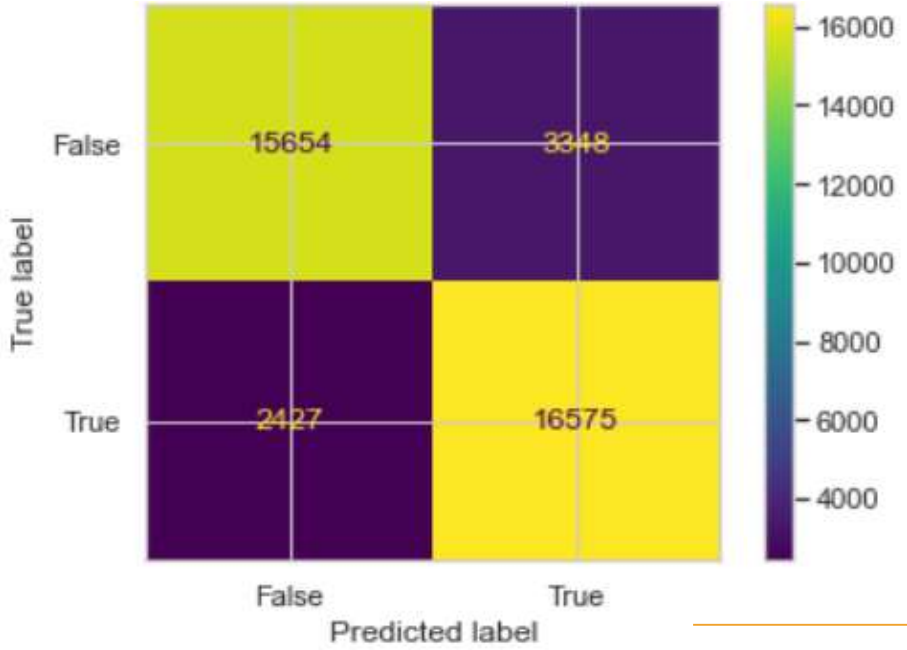


Logistic regression with PCA



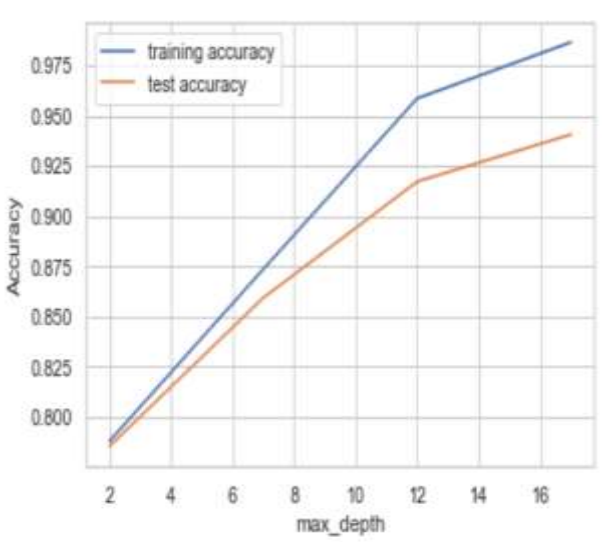
Confusion Matrix- predicted label-train and test set

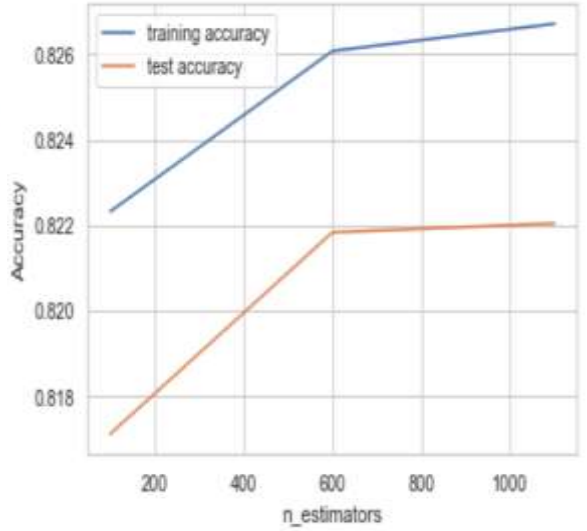


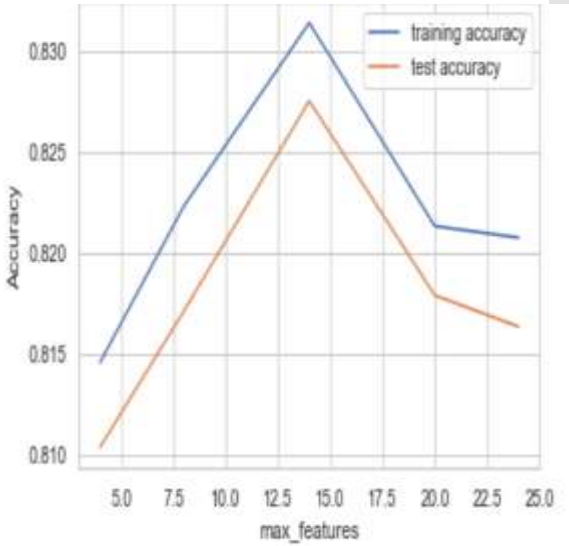


Training and Test Accuracy(max_depth, n_estimators and max _feature)

We see that as we increase the value of max_depth, both train and test scores increase till a point. 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.







Complete MODEL STATS

	Model	Recall	Test Accuracy	Roc_auc_score
1	Decision Tree with PCA	0.89	0.83	0.77
0	Logistic Regression with PCA	0.87	0.83	0.88
3	Logistic without PCA	0.82	0.79	0.76
2	Random Forest with PCA	0.70	0.87	0.88

CONCLUSION

- EDA observations: there is a considerable drop in recharge, call usage and data usage in the 8th month.
- Factors having significant impact on churn are:

1. Tenure: customers with longer durable of time within company are less likely to churn.
2. Contract type: month-to-month customers are more likely to churn than longer term contract clients.
3. Payment method: people with electronic cheque are more likely to churn in comparison with other payment modes
4. Monthly charges: highly monthly charges customers are more likely to churn
5. Minutes of Usage: important factor affecting the churn

RECOMMENDATION

- Analysis provide valuable insights on customer churn factors. Analysis predicted random forest classifier as the best model in predicting churn with accuracy of 80.52%.

1. Provide incentives to loyal customers or high-value customers i.e. stayed with the company for a longer period of time.
2. A sudden drop in local minutes might be due to the poor network or unsuitable customer plans, efforts shall be made on both the parts.
3. Encouraging customers for other payment modes than electronic cheques.
4. Routine feedback calls for customer retention
5. Various attractive offers to customers showing sudden drop in total amount spent on calls & data recharge in action phase
6. Provide better value for money by offering packages.