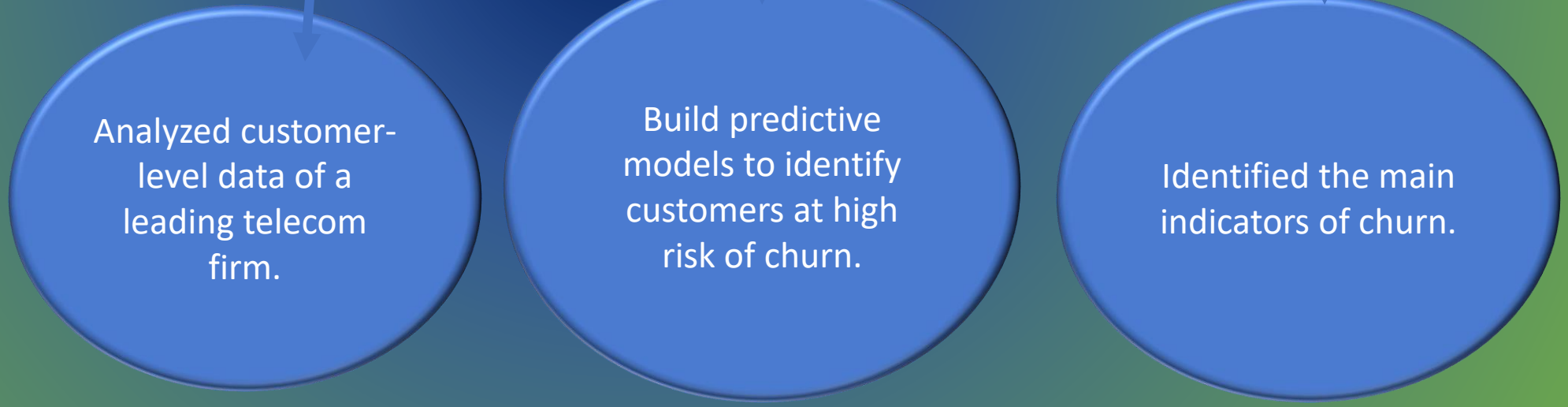
Telecom Churn Analysis Case Study

TEAM MEMBERS

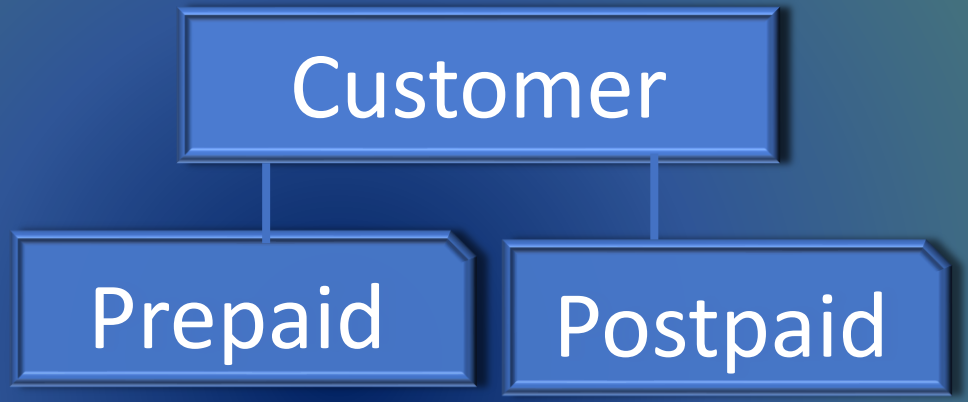
- Ojas Joshi
- Nibedita Sahoo
- Pooja Kumari

Business problem overview

To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn .**



Understanding and defining churn



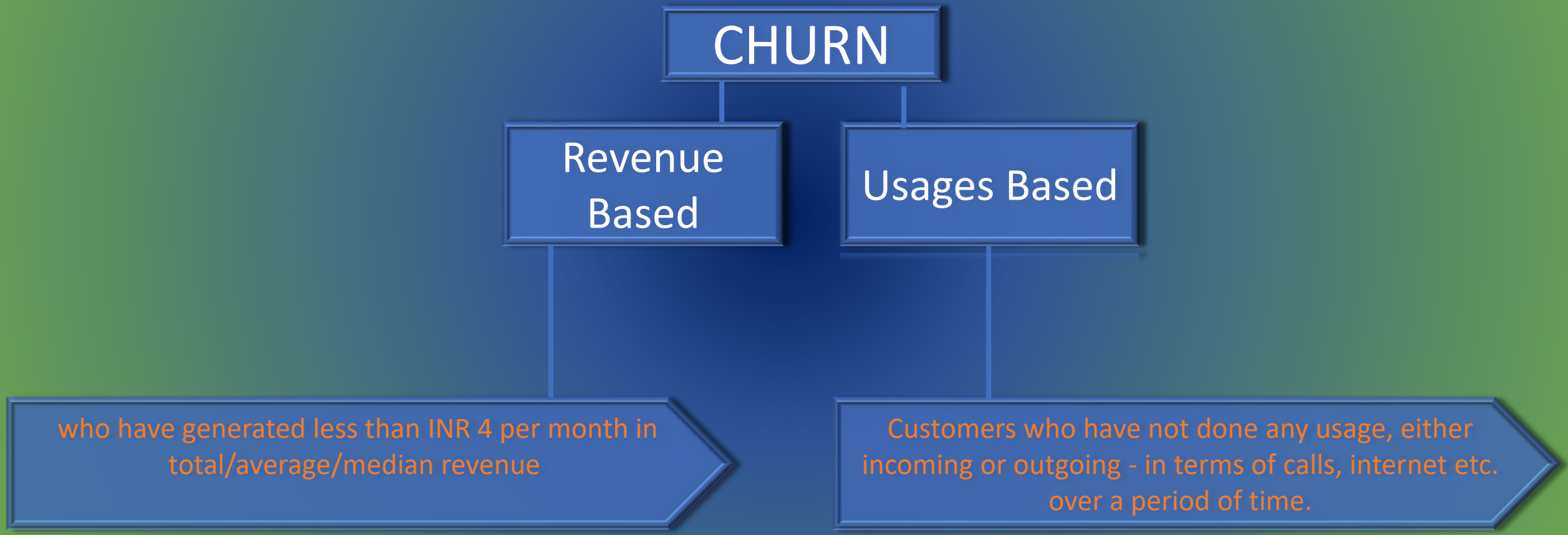
Prepaid: customers pay/recharge with a certain amount in advance and then use the services.

- Customers can switch to another network easily.
- Can simply stop using the services without any notice.
- Is hard to know whether someone has actually churned or is simply not using the services temporarily.

Postpaid: customers pay a monthly/annual bill after using the services.

- Customers have to inform the existing operator before switching to another operator to terminate the services.

Understanding and defining churn



High Value Churn

In the Indian and Southeast Asian markets, approximately 80% of revenue comes from the top 20% of customers (called high-value customers). Thus, if we can reduce the churn of high-value customers, we will be able to reduce significant revenue leakage.

In this project, we will define high-value customers based on a certain metric and predict churn only on high-value customers.

Understanding the business objective and the data

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.

The **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

Customer behavior during churn

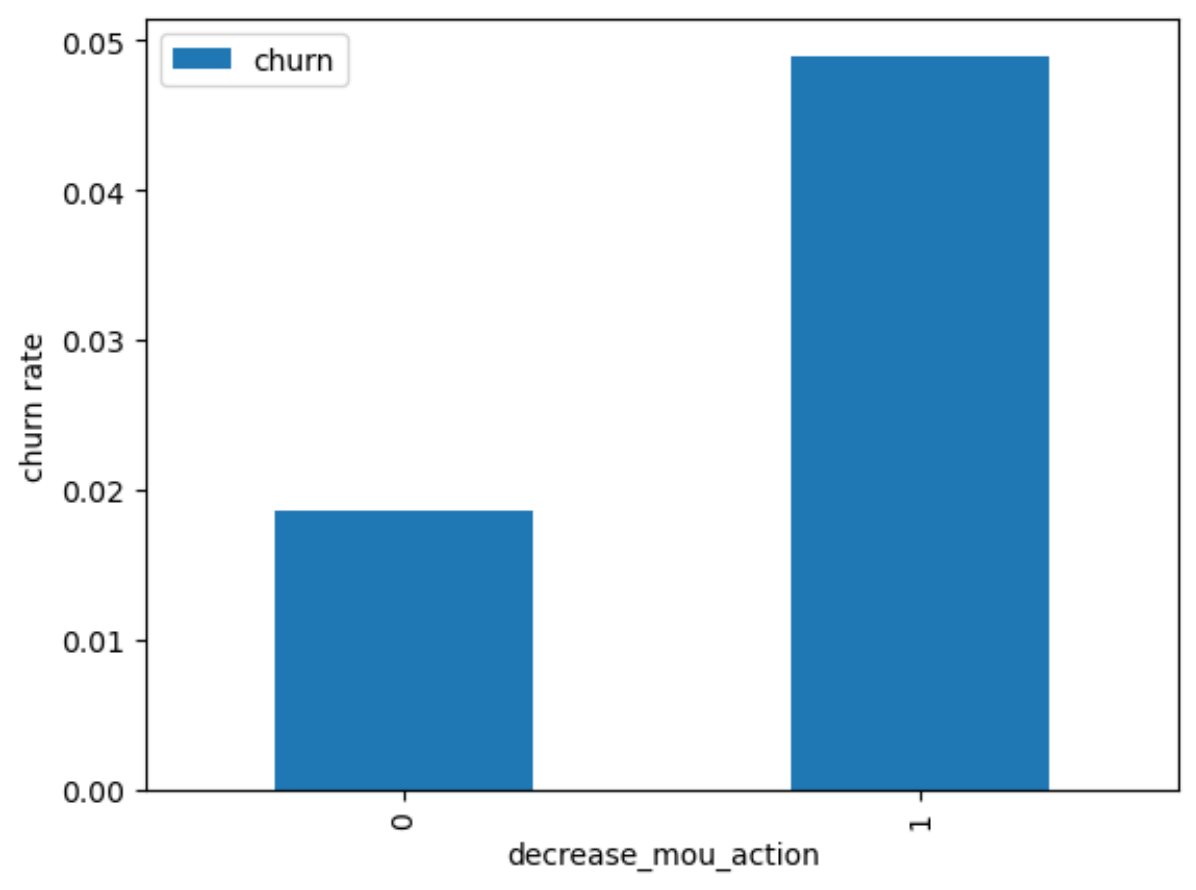
Good Phase → Action Phase → Churn Phase

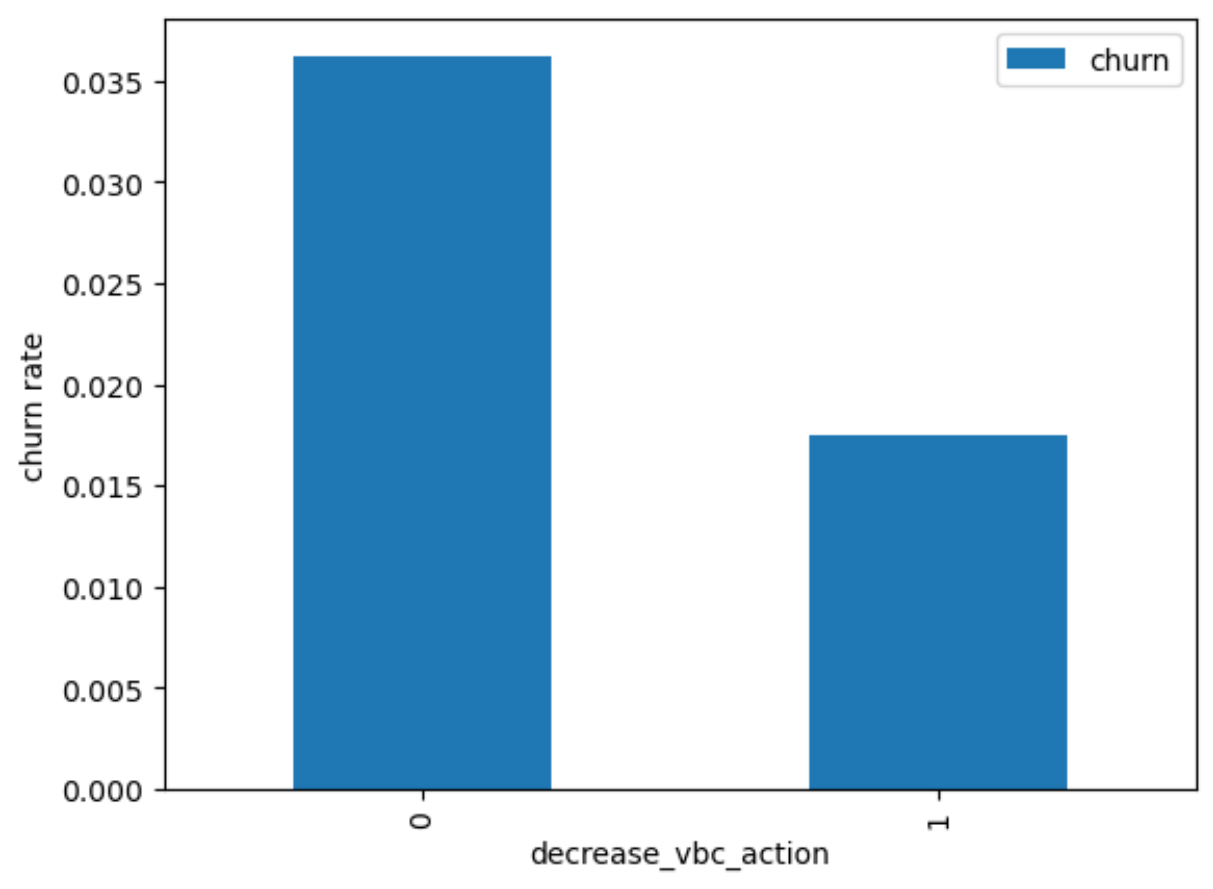
In this case, since we are working over a four-month window, the first two months are the ‘good’ phase, the third month is the ‘action’ phase, and the fourth month is the ‘churn’ phase.

PROCEDURE

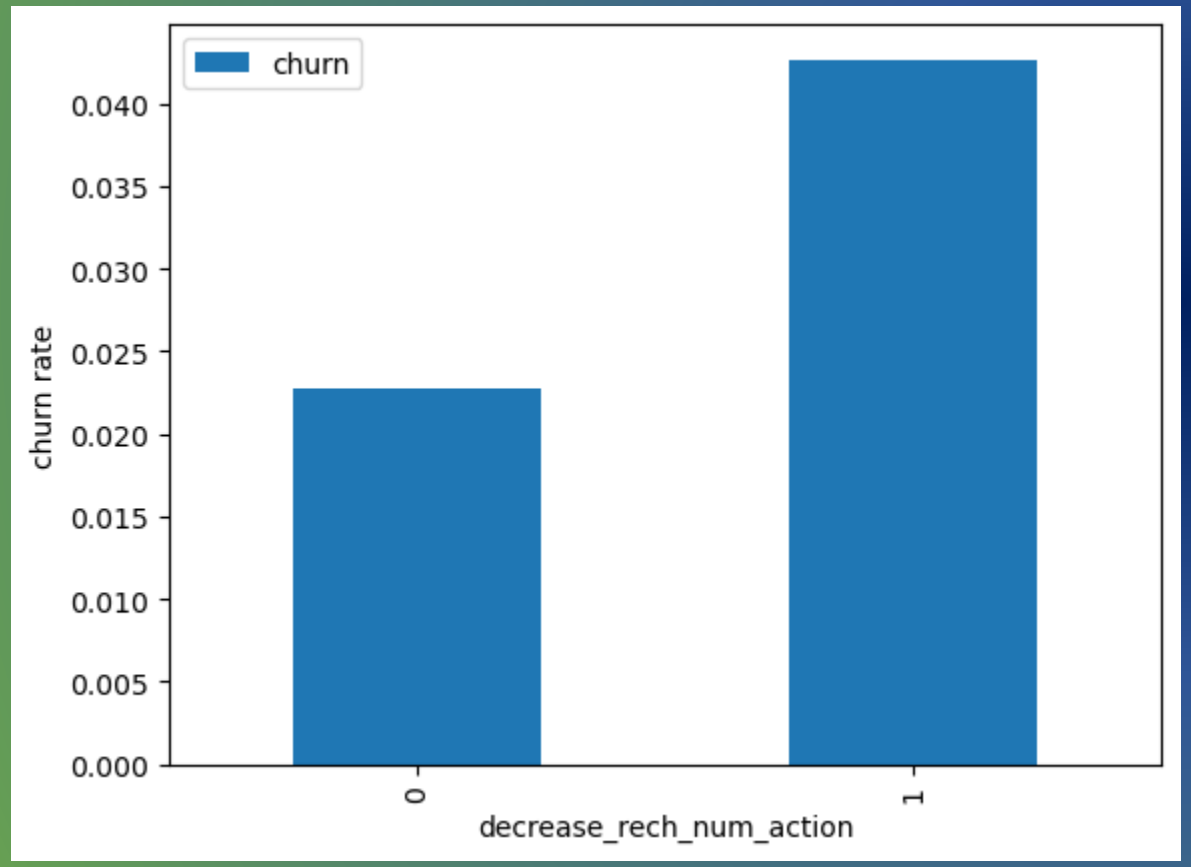
- Test Train Split
- Class Imbalance
- Standardization
- Modelling
- Model 1 : Logistic Regression with RFE & Manual Elimination (Interpretable Model)
- Model 2 : PCA + Logistic Regression
- Model 3 : PCA + Random Forest Classifier

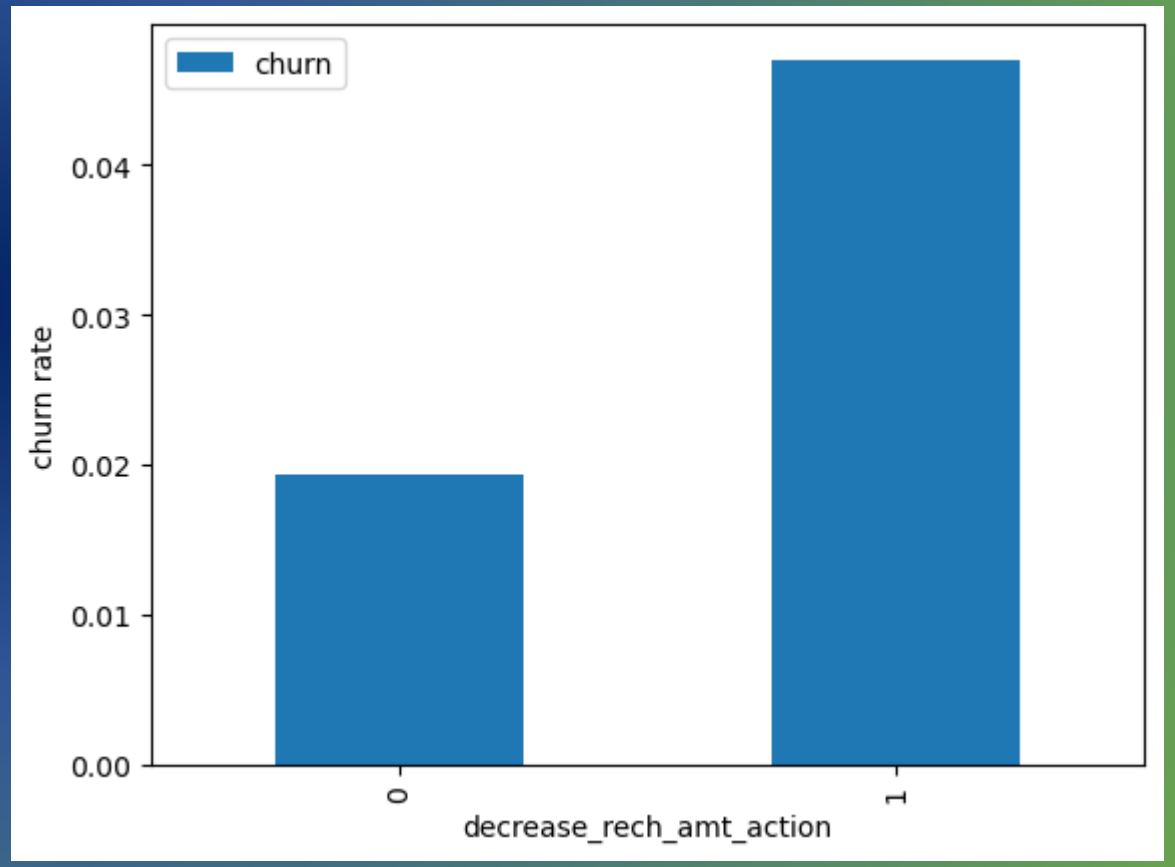
Univariate analysis



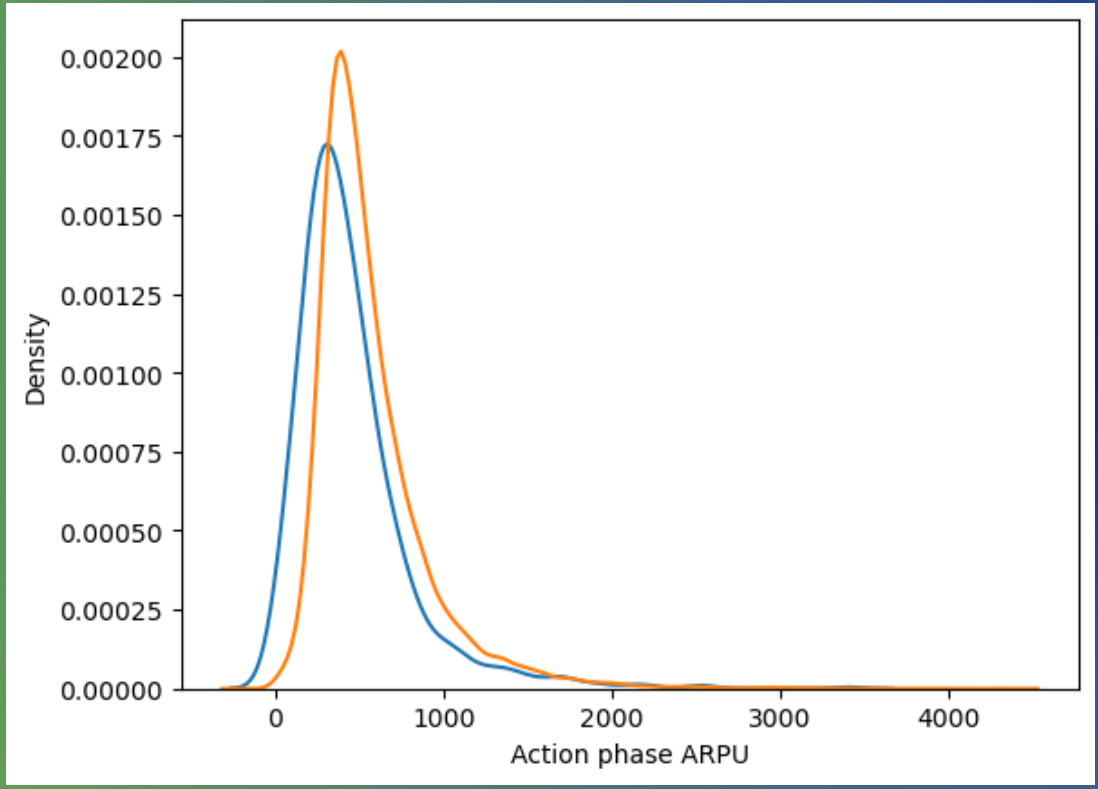


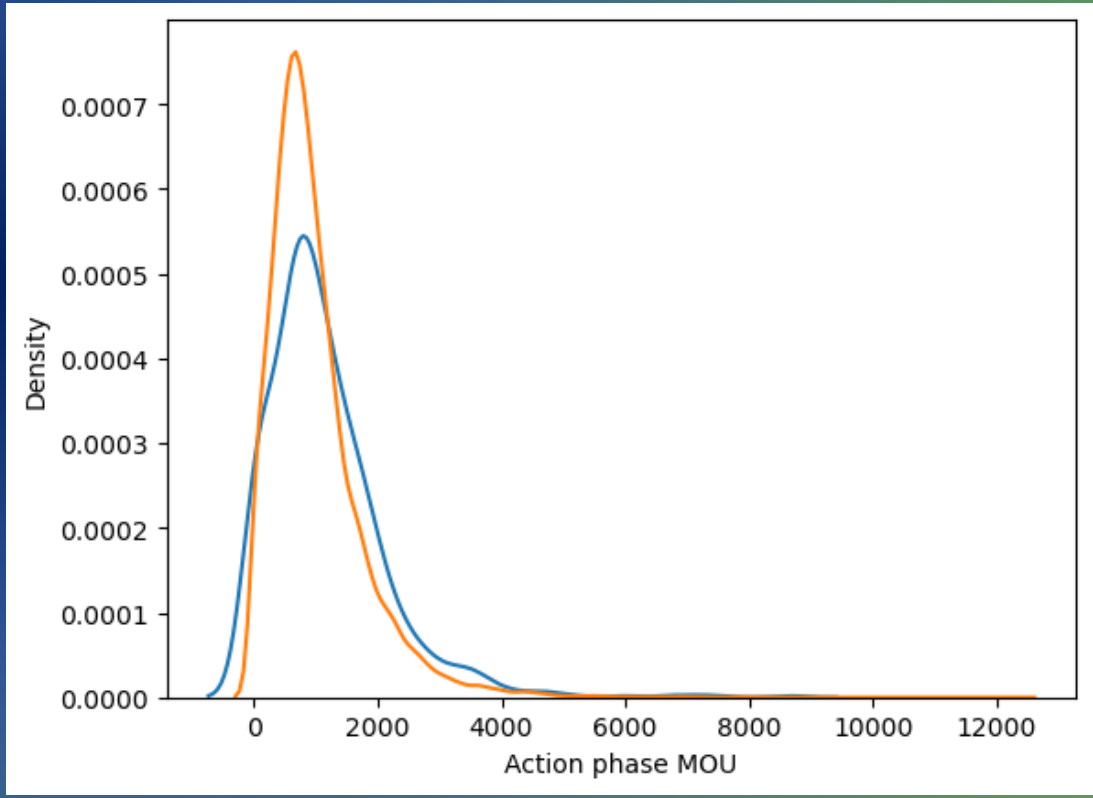
Univariate analysis



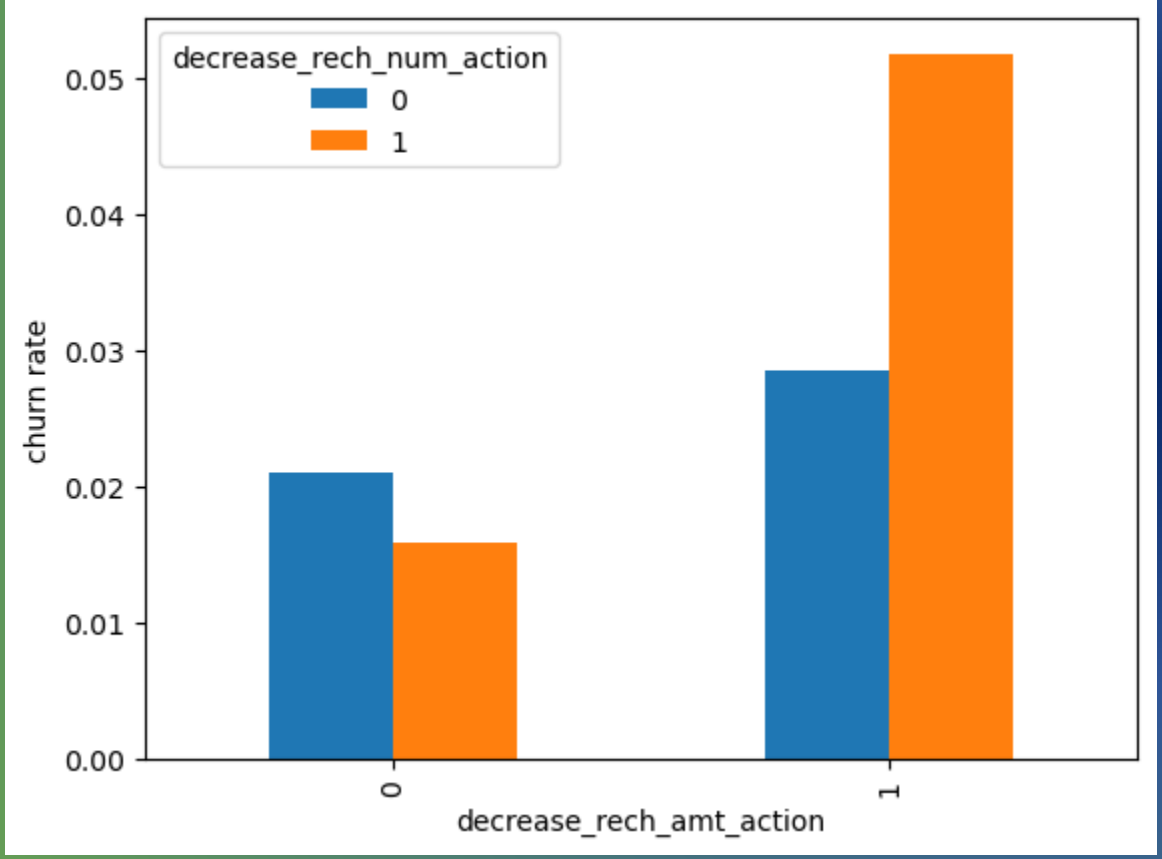


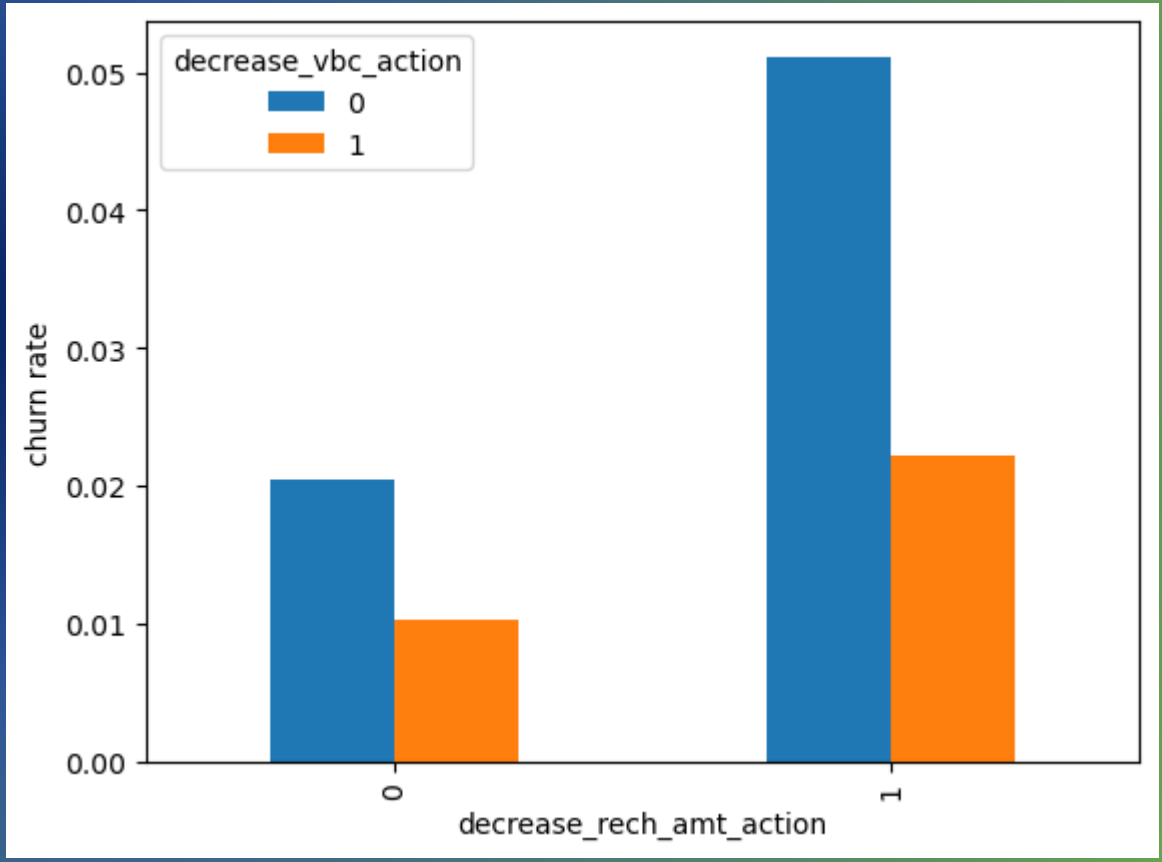
Distribution Plot



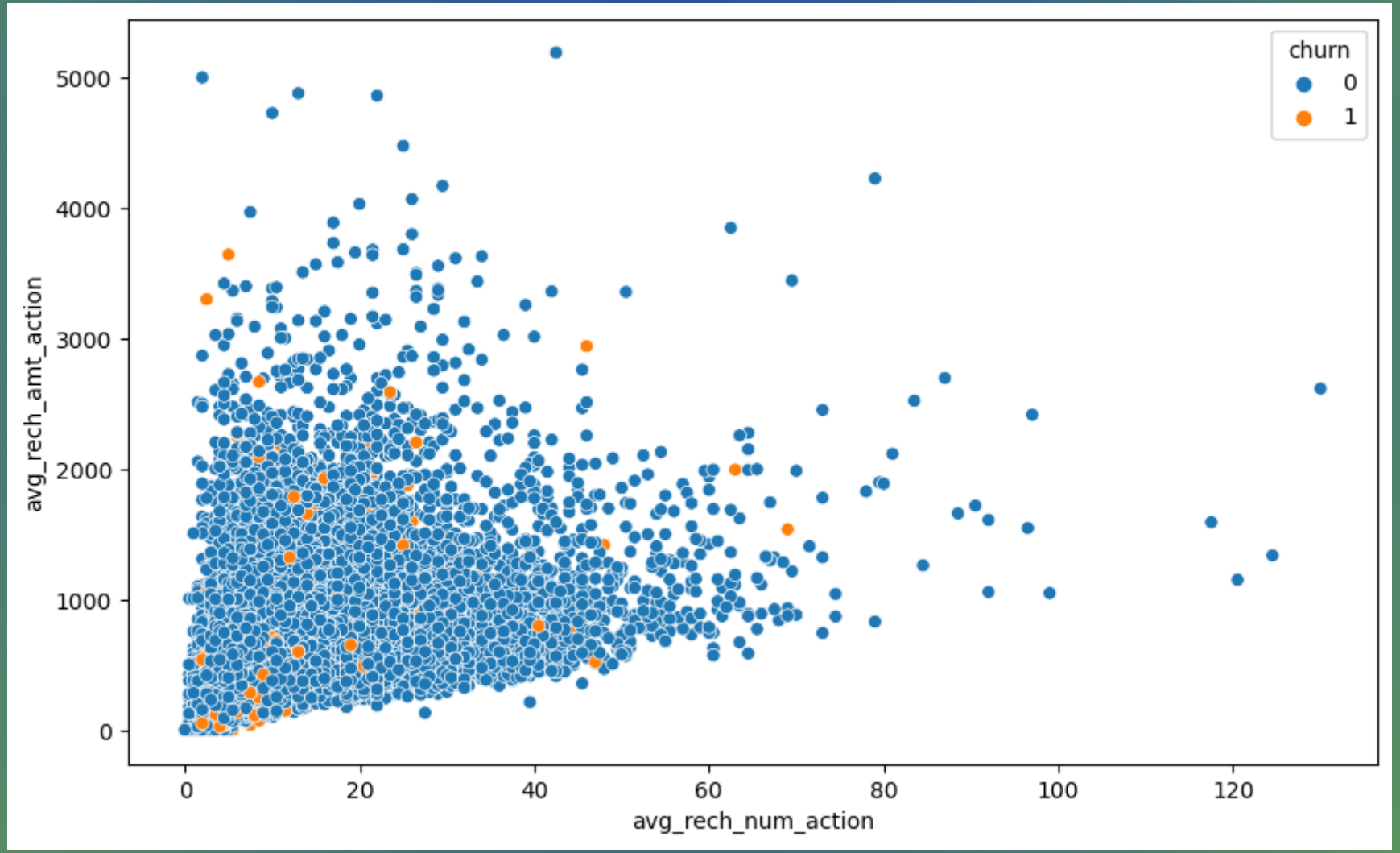


Bivariate analysis

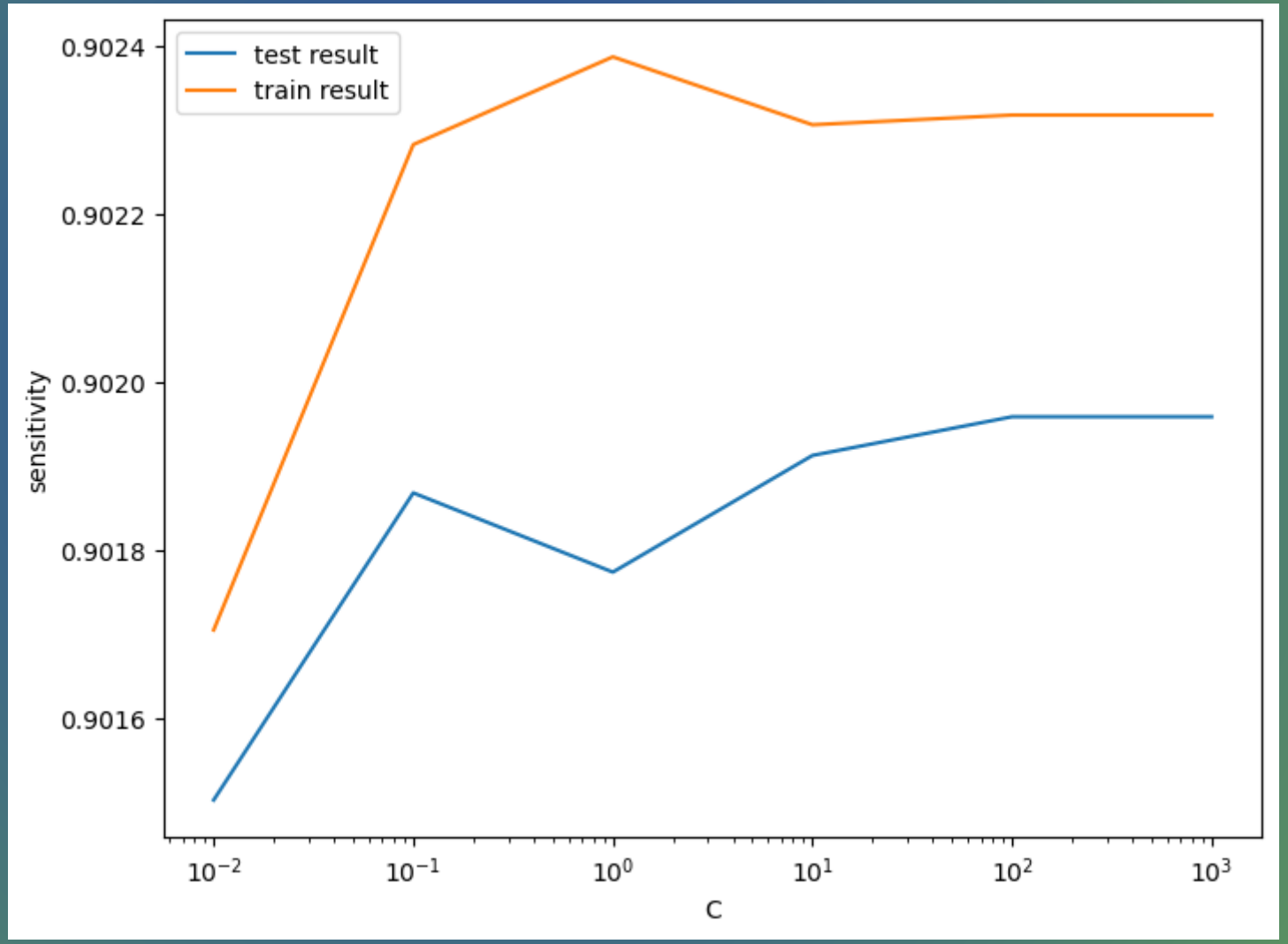




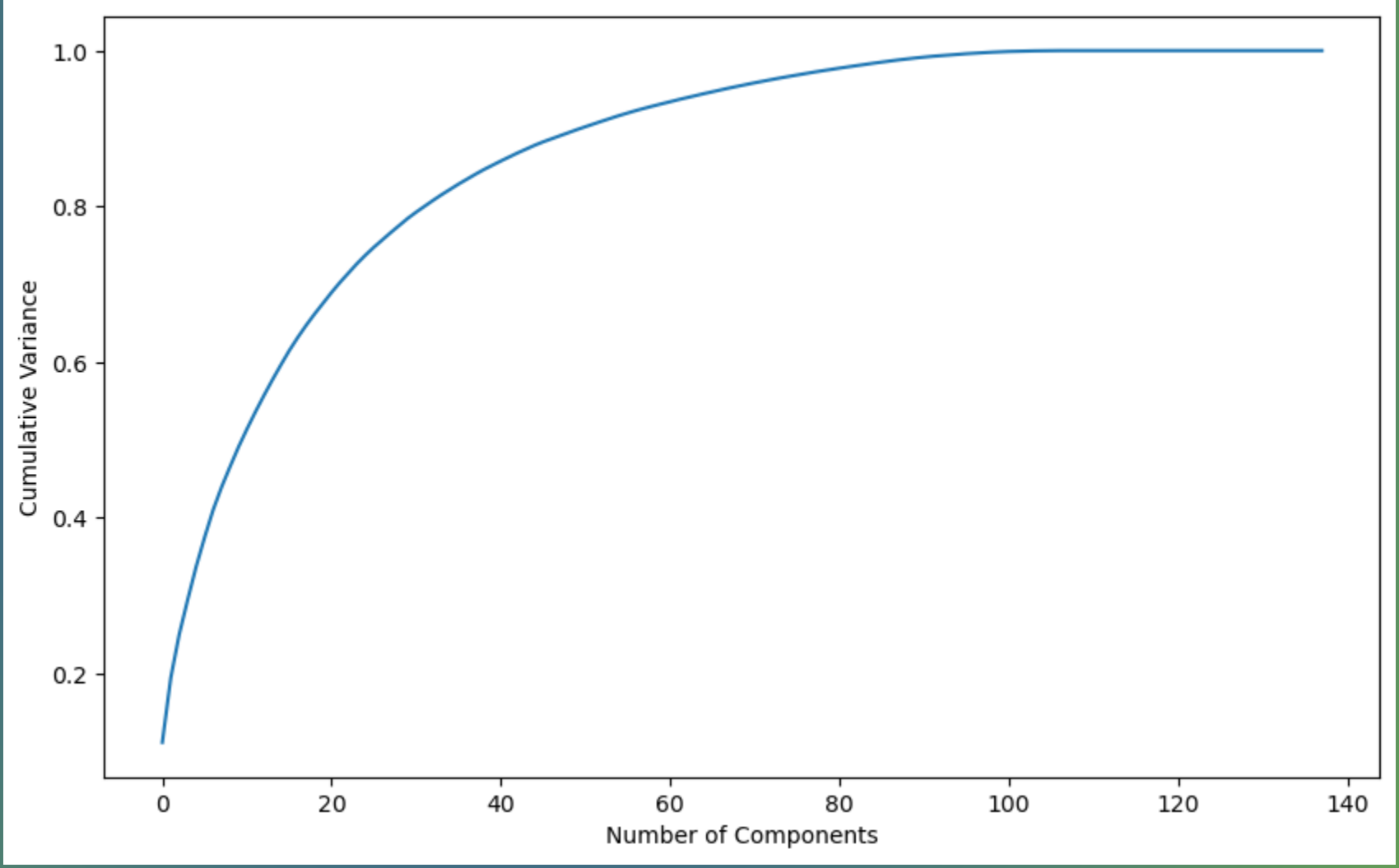
Scatter Plot



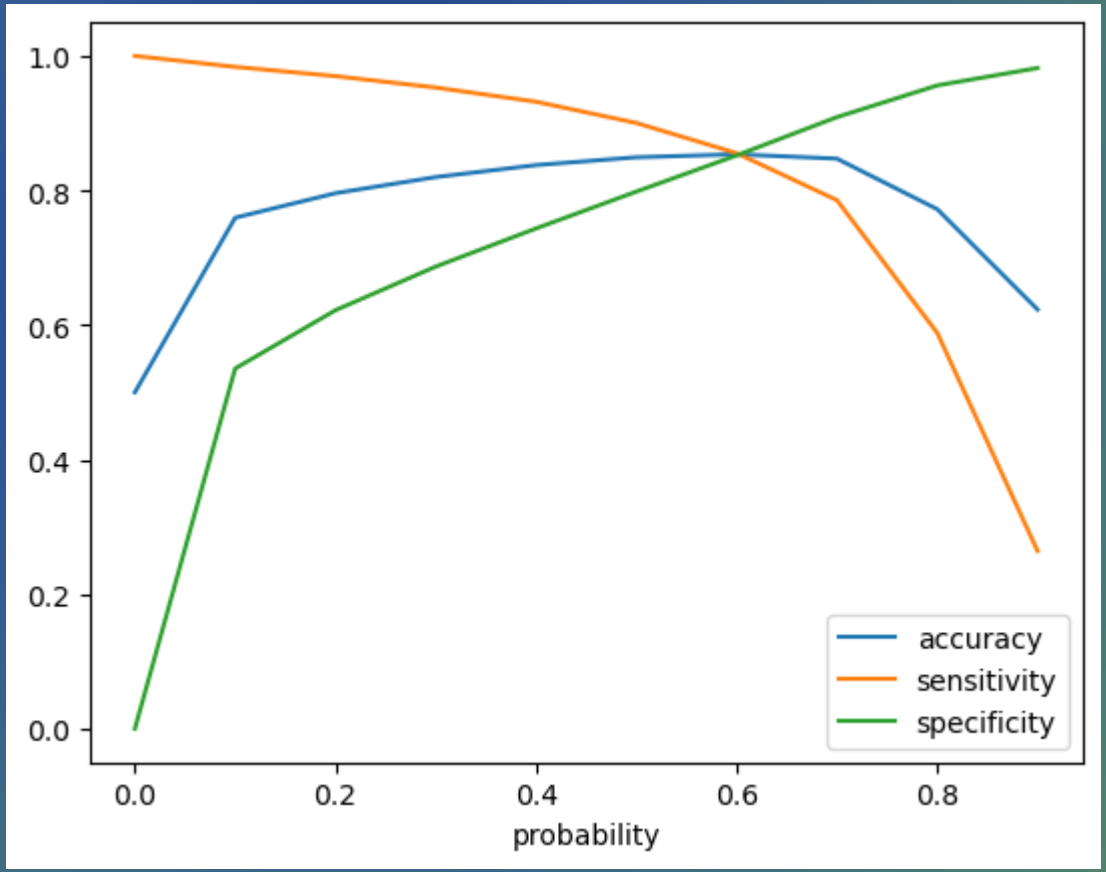
Plot of C versus train and validation scores



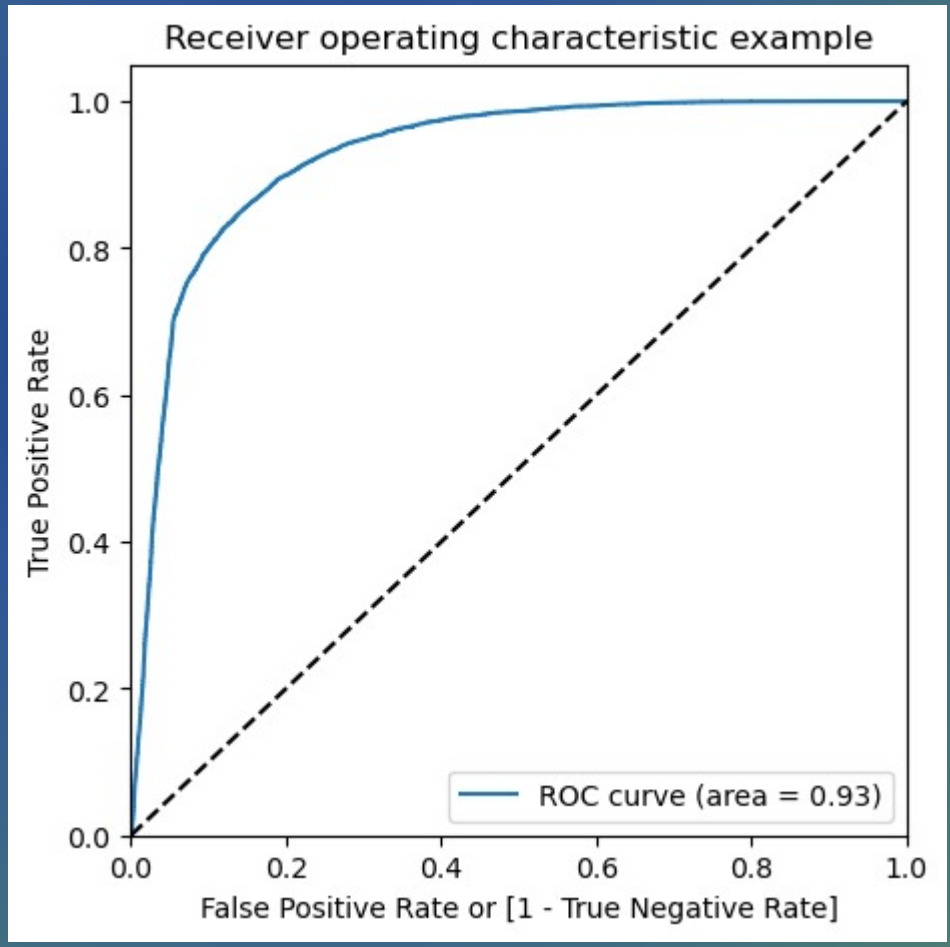
Model with PCA



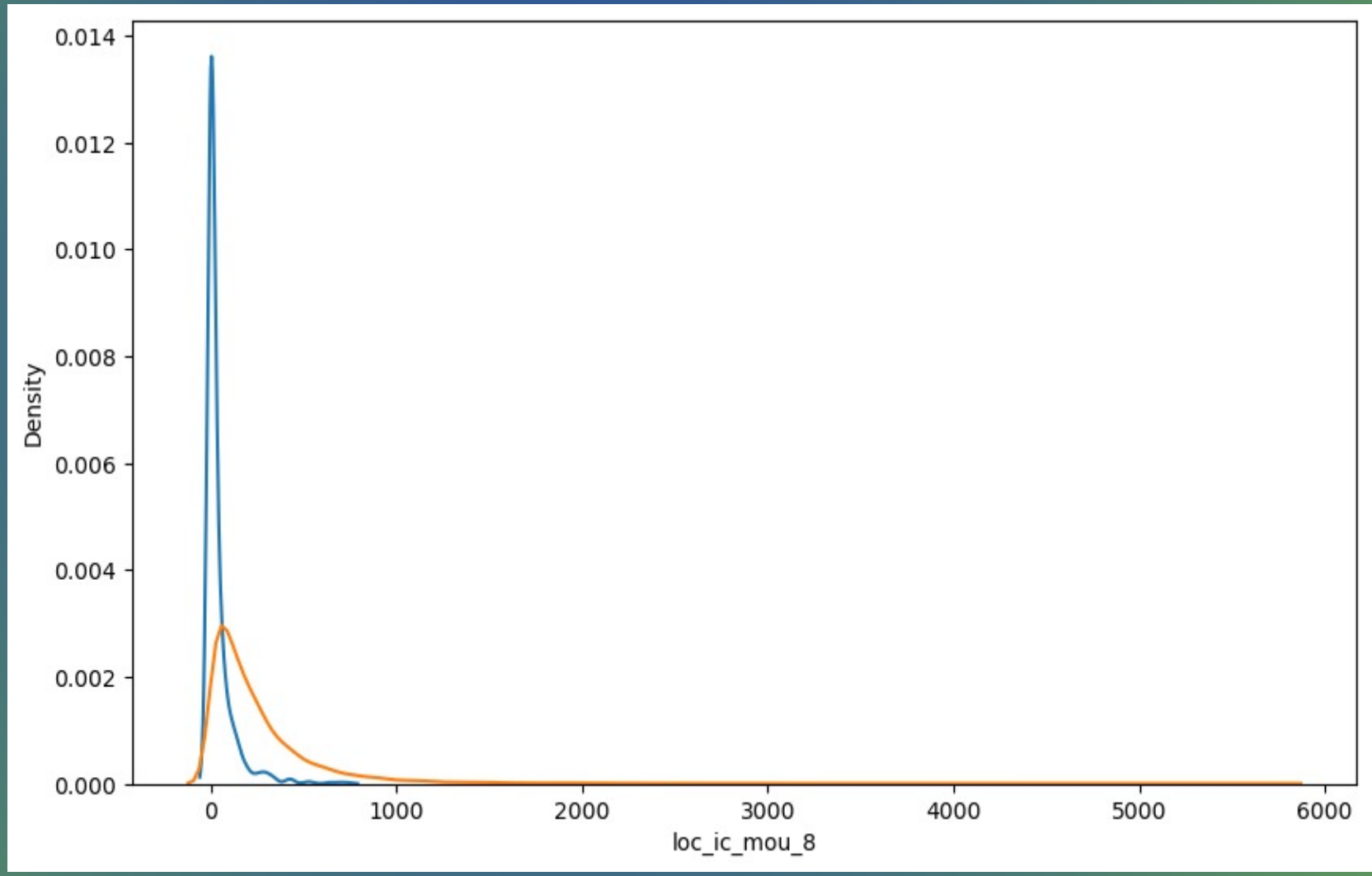
Feature Selection Using RFE



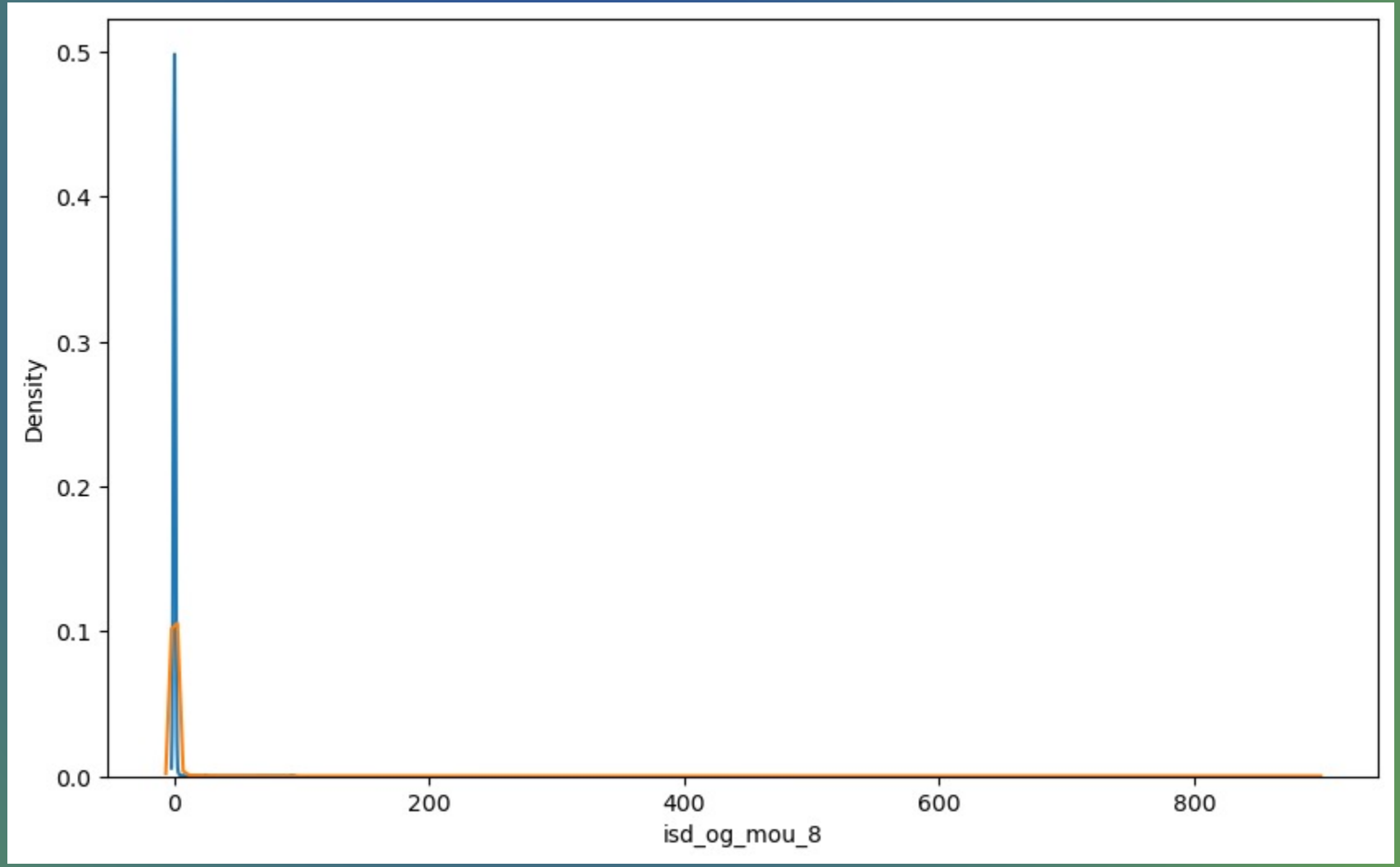
Plotting the ROC Curve (Trade off between sensitivity & specificity)



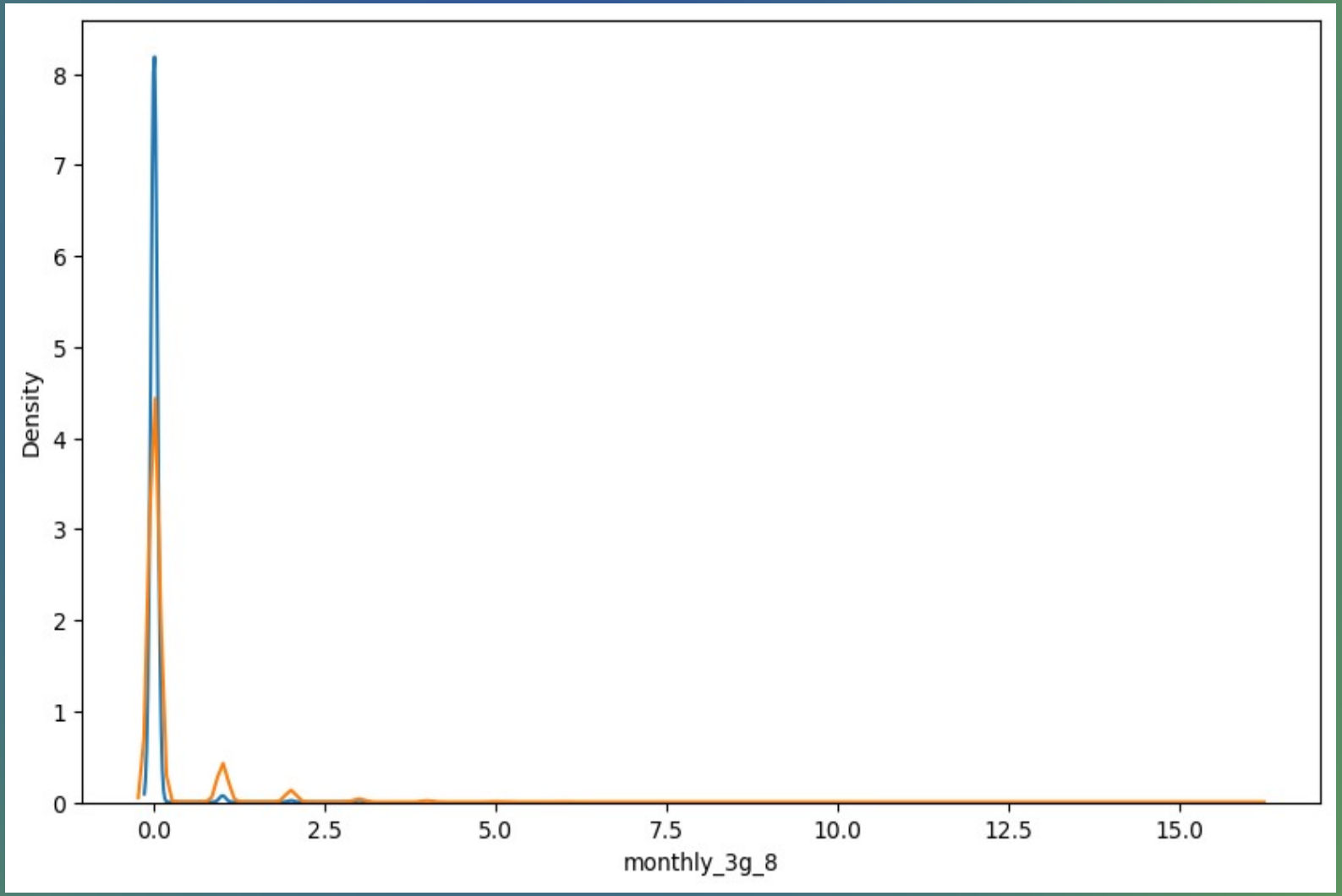
Plots of important predictors for churn and non churn customers



Plots of important predictors for churn and non churn customers



Plots of important predictors for churn and non churn customers



CONCLUSION

- Std Outgoing Calls and Revenue Per Customer are strong indicators of Churn.
- Local Incoming and Outgoing Calls for 8th Month and avg revenue in 8th Month are the most important columns to predict churn.
- Cutomer’s with tenure less than 4 years are more likely to churn.
- Max Recharge Amount is a strong feature to predict churn.
- Random Forest produced the best prediction results followed by SVM.

Recommendations

- Target the customer’s, whose monthly 3G recharge in August is more, as they are likely to be churned.
- Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
- Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
- Customer’s decreasing monthly 2g usage for August are most probable to churn target them.
- Target the customers, whose outgoing others charge in July and incoming others on August are less.

THANK YOU