Telecom Churn Prediction

BY:
Vinod Kumar
Rakhi Chandrasekhar
Abhinesh Durai

PROBLEM STATEMENT

- In the telecom industry, customers can choose from multiple service providers and actively switch from one operator to another.

- In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate.

- For many incumbent operators, retaining high profitable customers is the number one business goal.

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

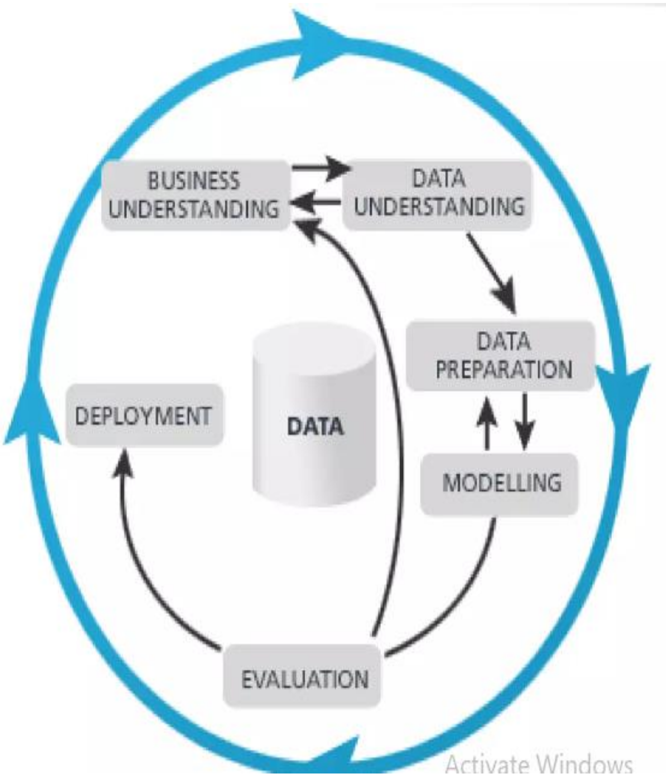
- In this project, you will analyze customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

MAIN OBJECTIVES

- Understand the data variables responsible for churn.
- Use ML algorithms to build the model and evaluate the accuracy of the model.
- Finding out the best model for business case and provide executive summary.
- Highlight the main variables/factors influencing the customer churns.

Method of Problem Solving

- Identify problem statement
- Data collection
- EDA(Exploratory data analysis
- Feature engineering
- Feature Selection
- Handling imbalance data
- Model selection
- Business Recommendations



Data Understanding and EDA

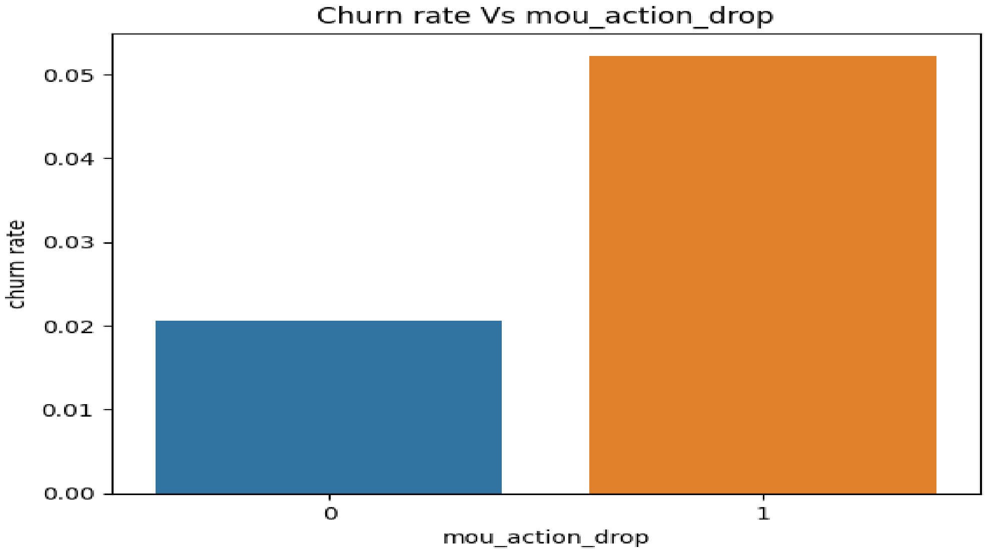
- The size of the Data Set is (99999, 226)
- Number of Rows = 99999, Number of Columns = 226.
- Dropping the columns in the table which are not

helpful for the analysis.

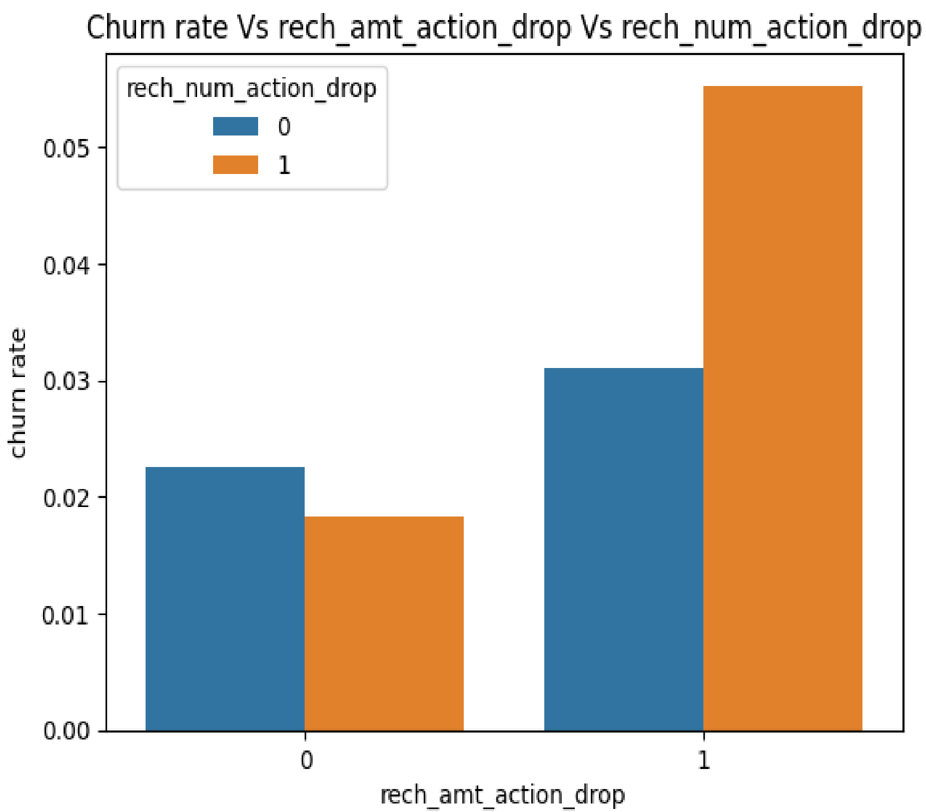
- Dropping Duplicates also, if any.
- Treating the Columns that has null values.

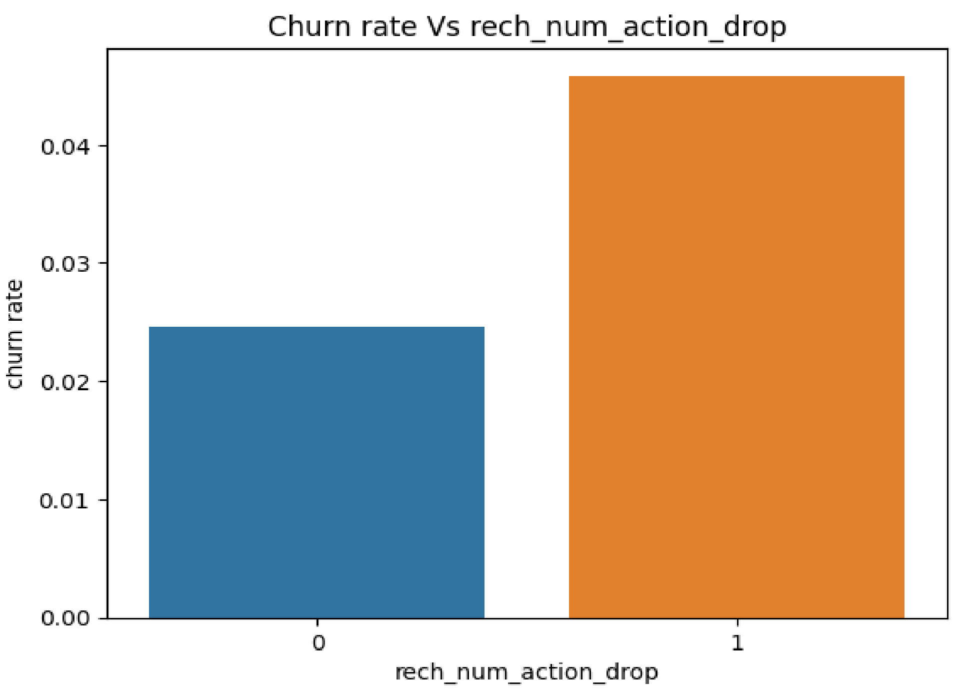
Columns dropped
date_of_last_rech_8
last_date_of_month_6
last_date_of_month_7
'last_date_of_month_8
date_of_last_rech_9
last_date_of_month_9
date_of_last_rech_6
date_of_last_rech_7

EDA and Categorical Variable Relation



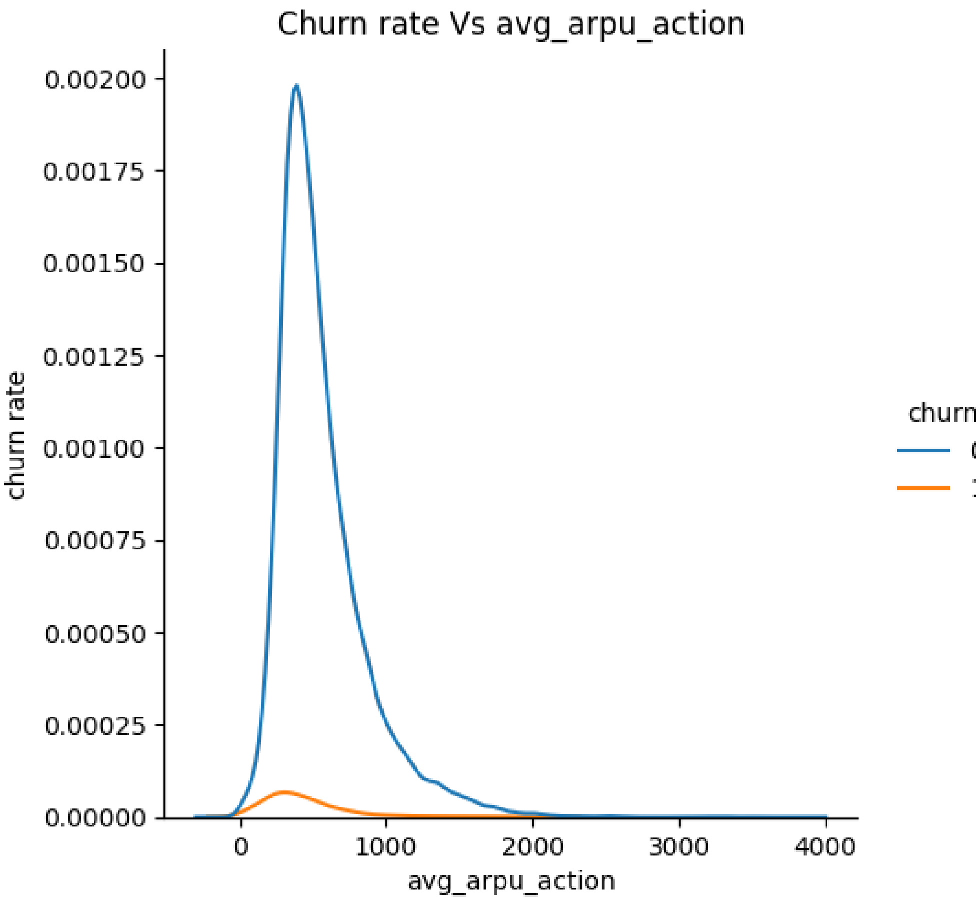
Churn rate for the customers whose minutes of usage (mou) decreased in the action phase than the good phase is more.

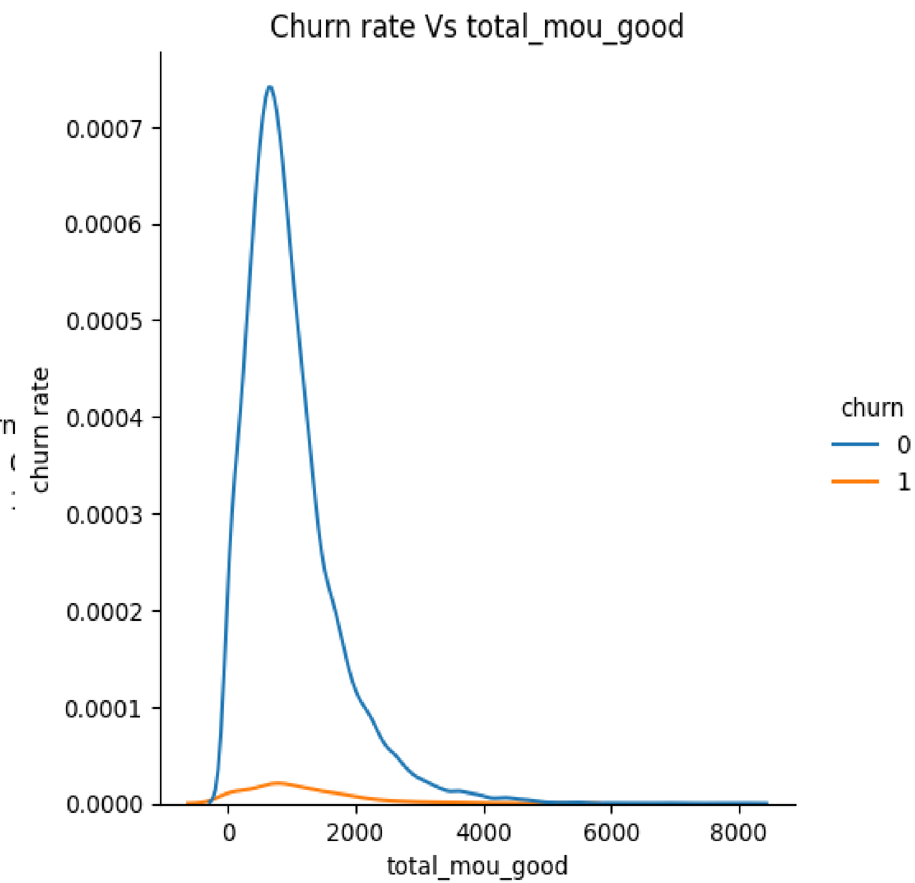




The accompanying graph demonstrates that consumers with lower recharge amounts and/or fewer recharges in the action phase compared to the good phase have higher churn rates.

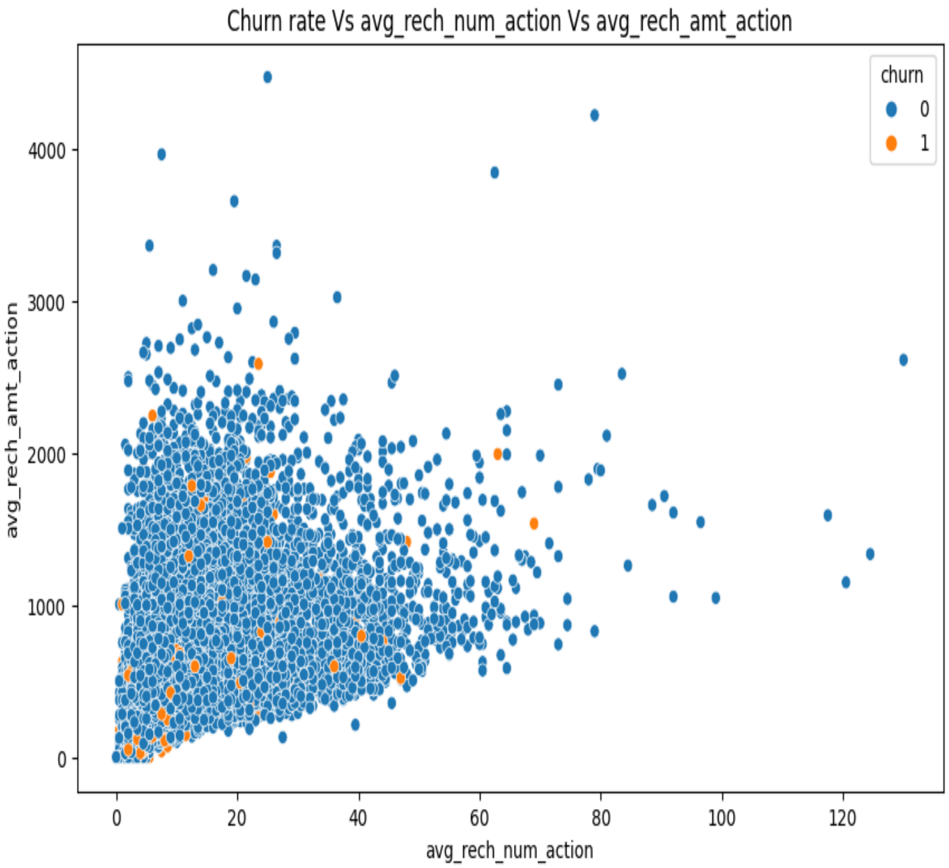
The churn rate is higher for customers whose numbers of recharges during the action phase is lower than their number during the good phase as is to be expected.

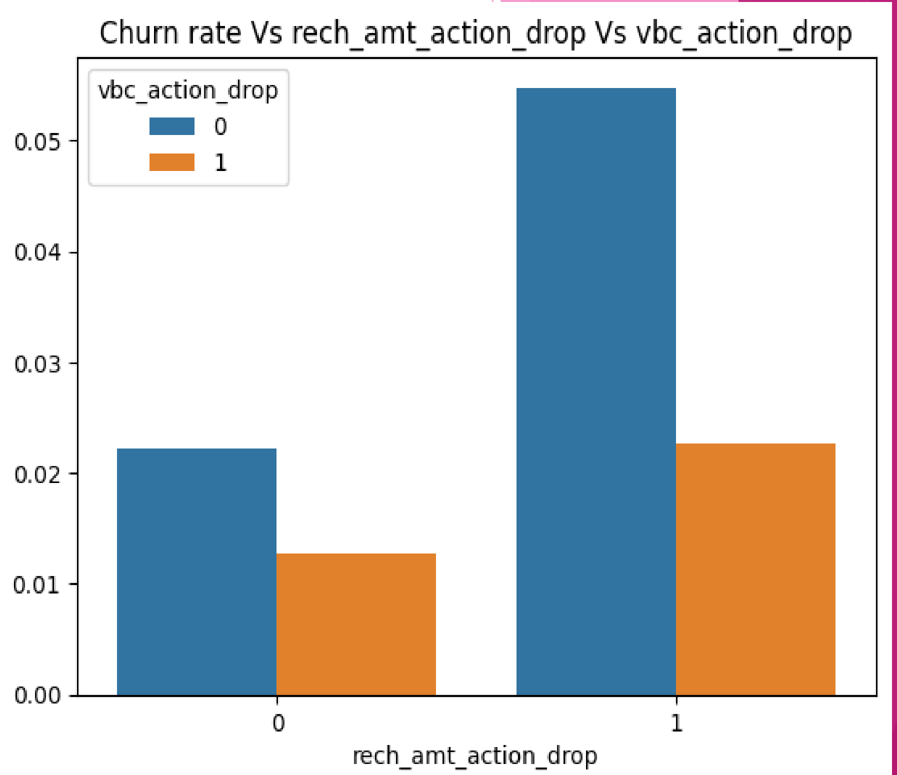




- The range of every revenue per user (ARPU) for churned cutomers is primarily 0 to 900. Customers with higher ARPU are less likely to leave The company
- The range ARPU for non churned clients is primarily 0 to 1000

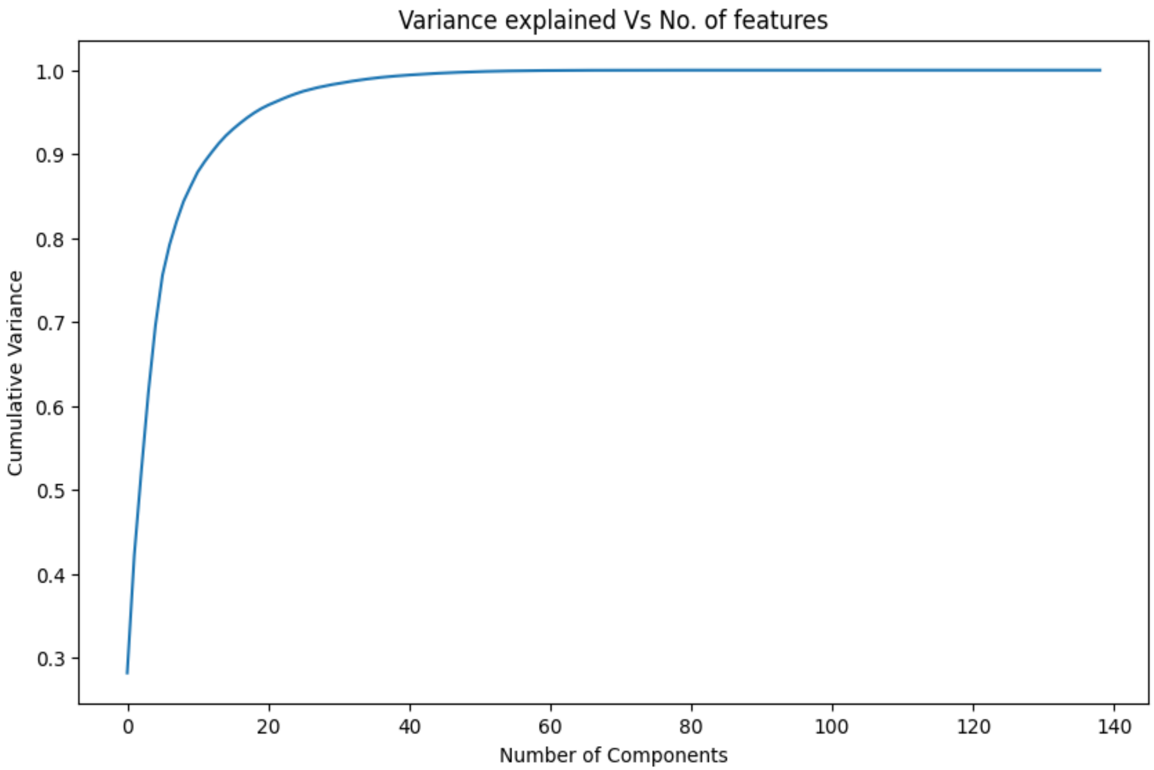
The churn customer's minutes of usage (MOU) are primarily in the 0 to 2500 range.Less turnover is likely the stroner than MOU



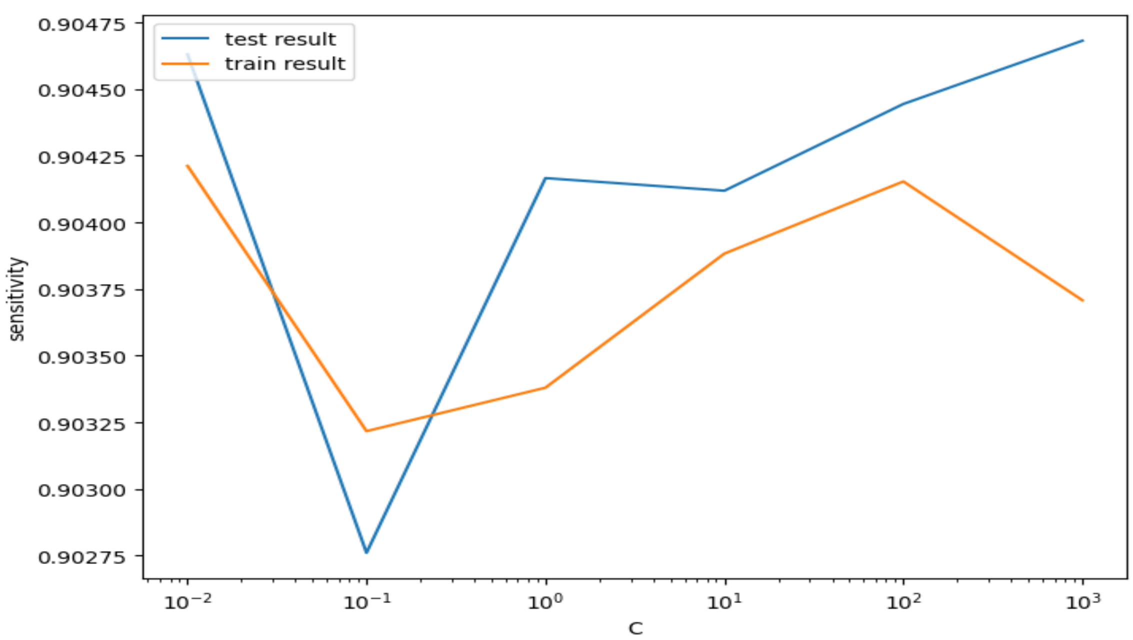


The pattern shown above demonstrates that the recharge quantity and amount are mostly proportional. The amount of the recharge increases with the number of recharges.

We can see that the churn rate is higher in This instance as well for consumers whose Recharge amounts are reduced as the volume based costs rise during the action month



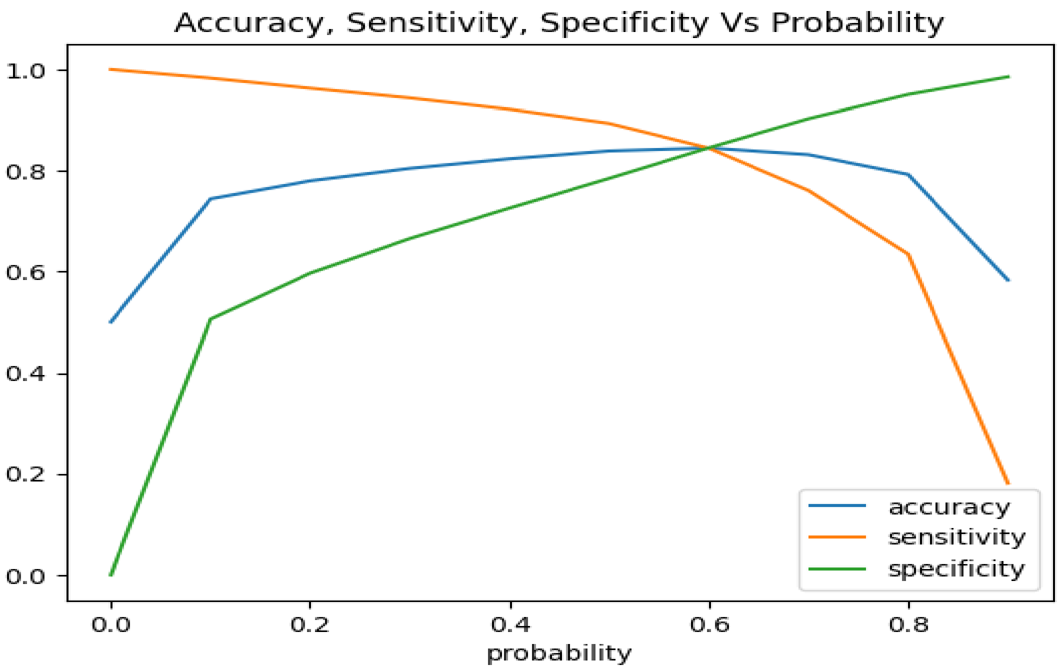
We can observe that 40 components account for almost 90%of the data’s variation. Thus, we will run PCA on 40 components



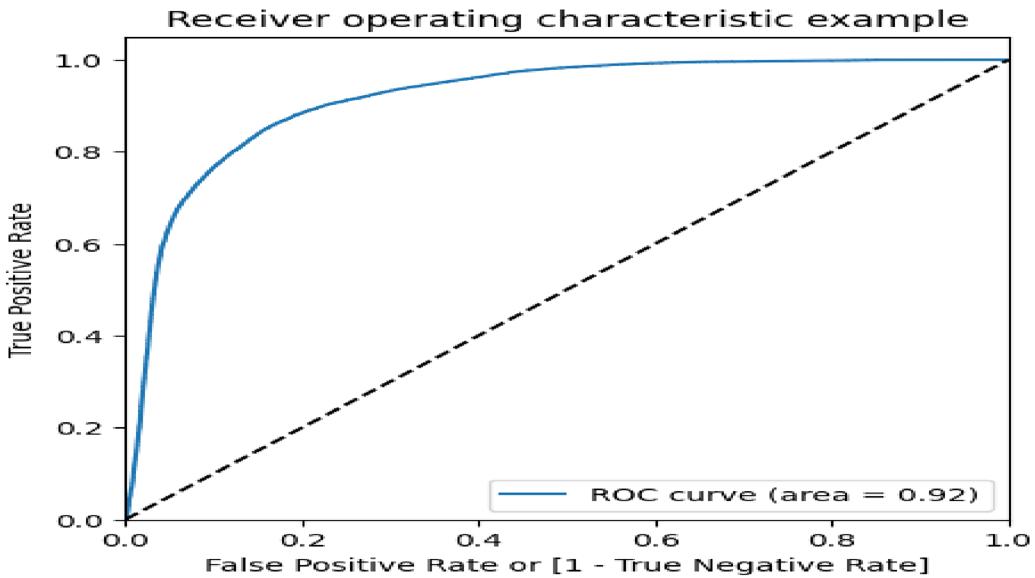
Metric	Train set	Test set
Accuracy	0.84	0.77
Sensitivity	0.9	0.9
Specificity	0.76	0.78

Overall the model is applying what it had learned from the train set well in the test set and is giving goo d performance

MODEL INTERPRETATION

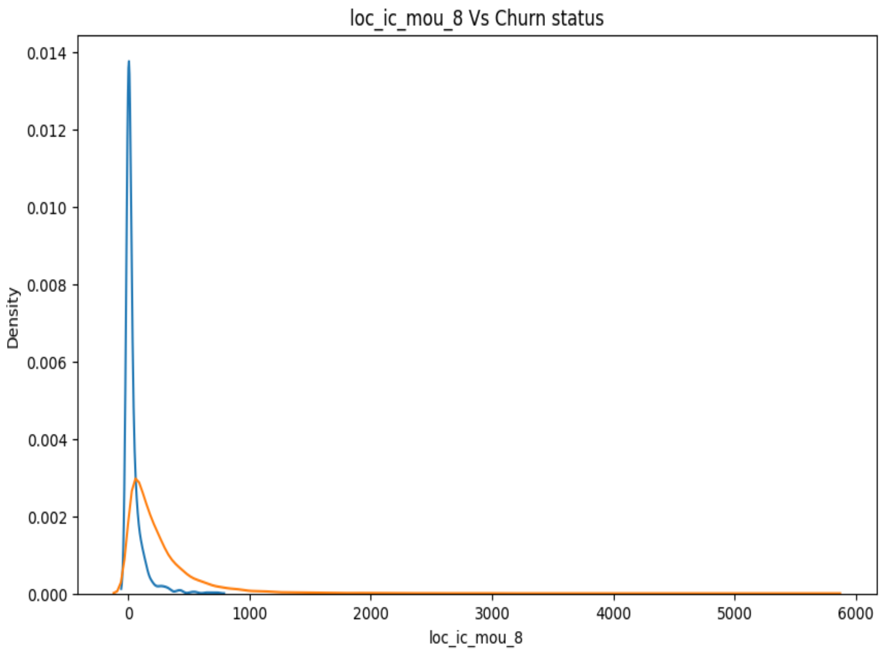


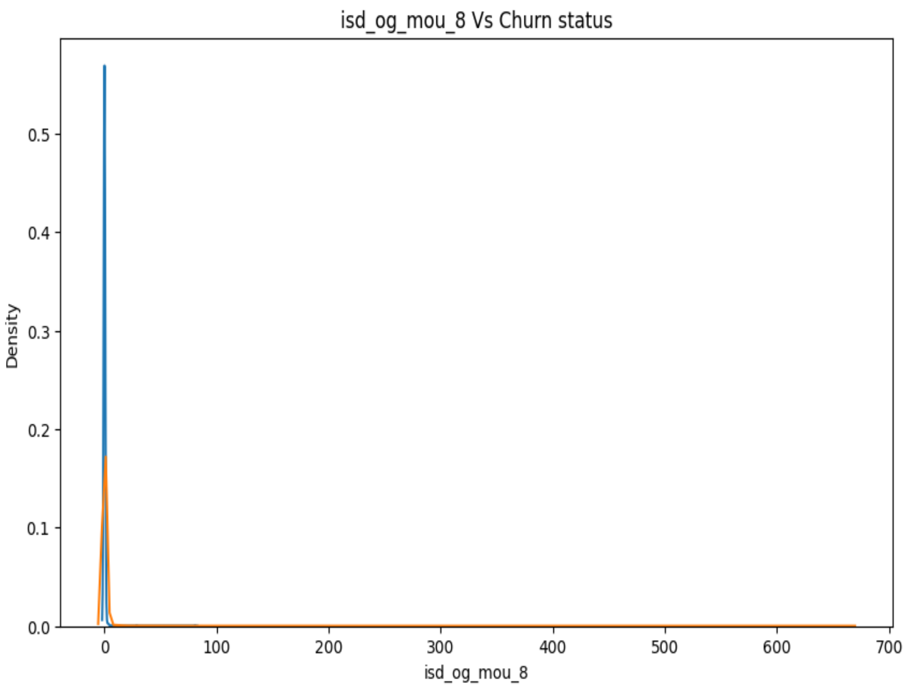
- We can see that there is accurate balance between sensitivity and specificity at 0.6, where the three parameters intersect.
- Here sensitivity is more important than accuracy and specificity to us. Although 0.6 should be the ideal probability cutoff according to the above curve ,we are using 0.5 in order to get increased sensitivity, which is our primary objective.



We can observe that the ROC curve's area is closer to 1 than it is to the model's Gini coefficient.

MODEL INTERPRETATION





We can see that the minutes of consumption for the churn customers throughout the month of August are typically lower than those for non-churn customers.

- For August the number of monthly 3G data for turnover clients is quiet densely populated at around 1,yet it dispersed throughout various values for non-churn customers.
- We may also plot the churn distribution for each variable that has greater coefficients.

BUSINESS RECOMMENDATION

- Focus on clients who use fewer minutes for inbound local calls and outgoing ISD calls during the action phase (mostly in the month of August).
- Pay special attention to the clients who pay more in July than they do in August.
- In addition, clients that experience increasing value-based costs throughout the action phase are more likely to leave than other customers. So, making an offer to these customers may be a good idea.
- Customers have a greater chance to be churned if their monthly 3G recharge is higher in August.
- Customers who used fewer STD incoming minutes on fixed T lines from operators T in August are more likely to churn.
- Customers who use less 2G data each month in August are more likely to churn.
- Consumers who used less incoming minutes on fixed T lines from operators in August are more likely to leave.
- Variables in roam_og_mou_8 have positive coefficients (0.7135). That means for the clients, whose roaming outbound minutes of consumption is increasing are more likely to churn.

THANK YOU