TELECOM CHURN CASE STUDY

BY PAIDA HARINI

PROBLEM STATEMENT

- In the telecom industry, customers are able to 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. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, **customer retention** has now become even more important than customer acquisition.
-
- 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 analyse 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.

BUSINESS OBJECTIVE

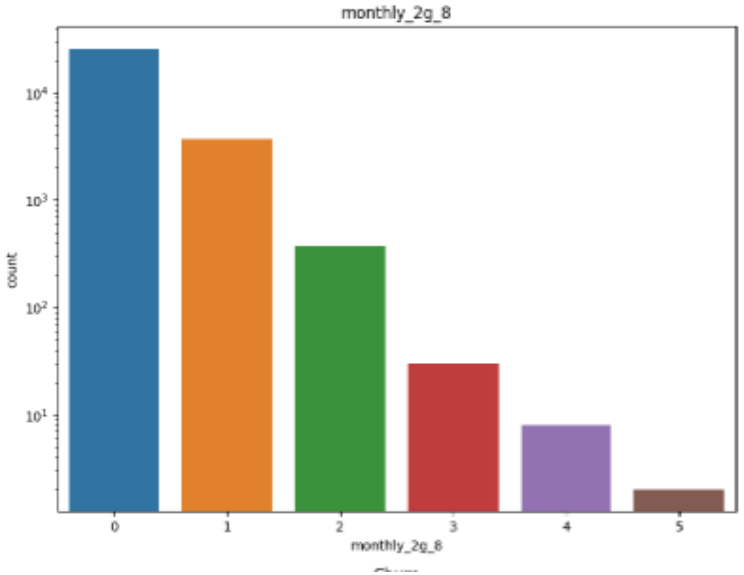
- The **business objective** is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behaviour during churn will be helpful.

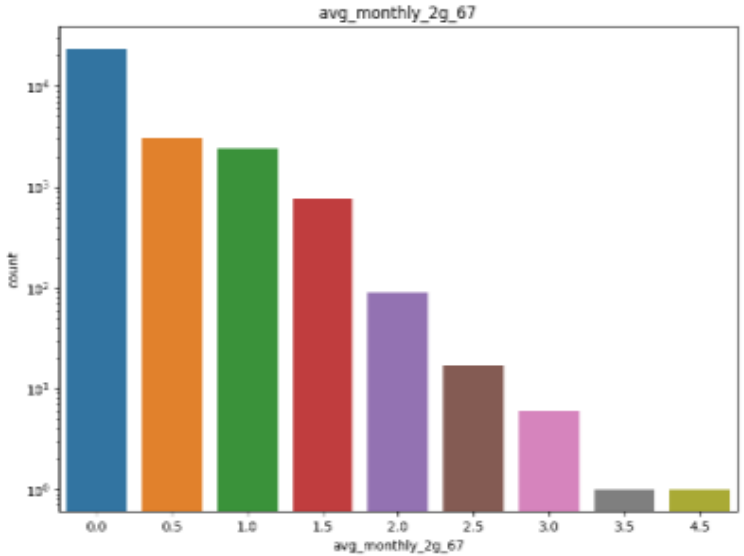
SOLUTION METHODOLOGY

- Data cleaning and data manipulation. 1. Check and handle duplicate data. 2. Check and handle NA values and missing values. 3. Drop columns, if it contains large amount of missing values and not useful for the analysis. 4. Imputation of the values, if necessary. 5. Check and handle outliers in data.
- EDA
- Model building
- Model presentation
- Conclusion & Recommended Strategies

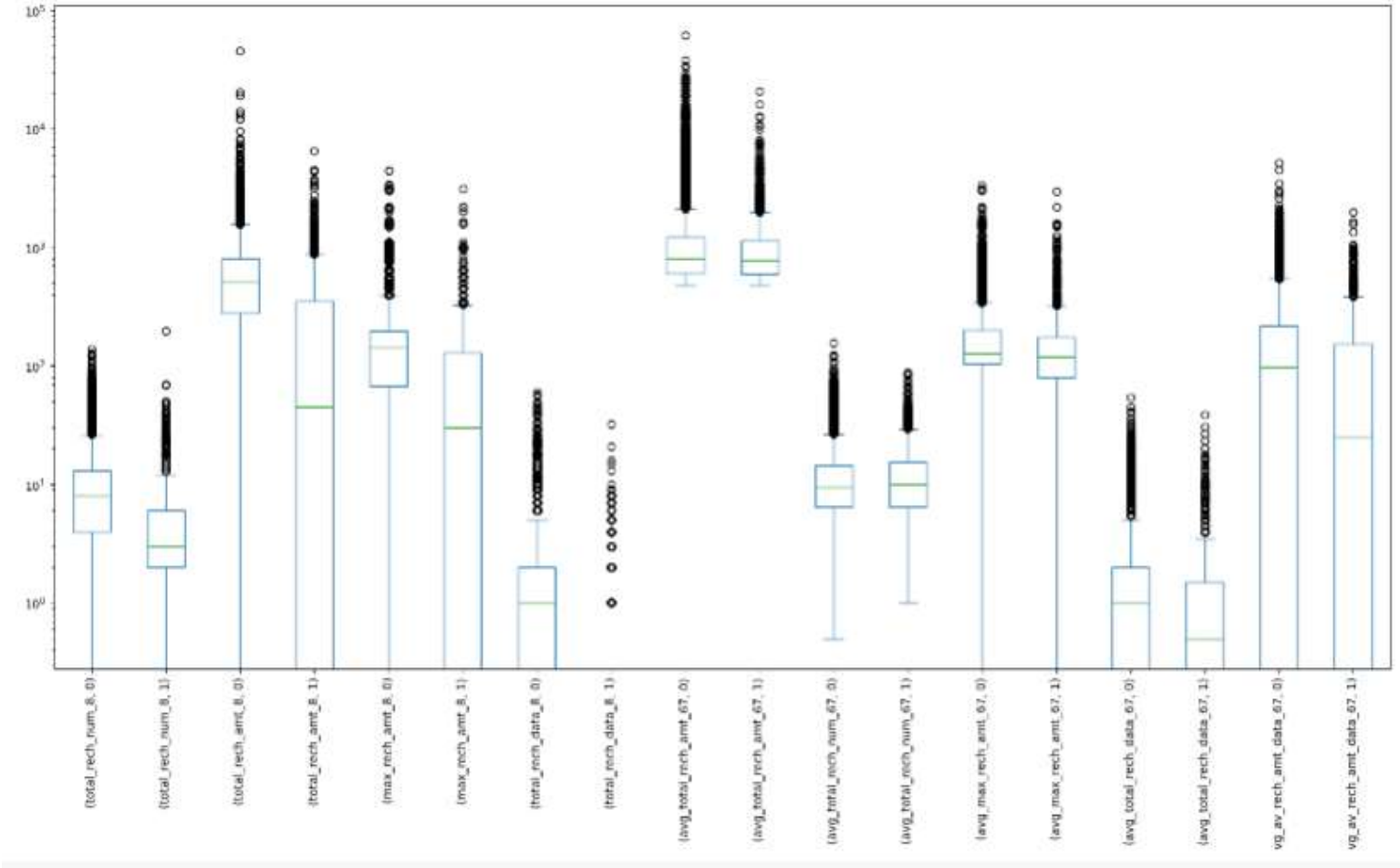
EDA

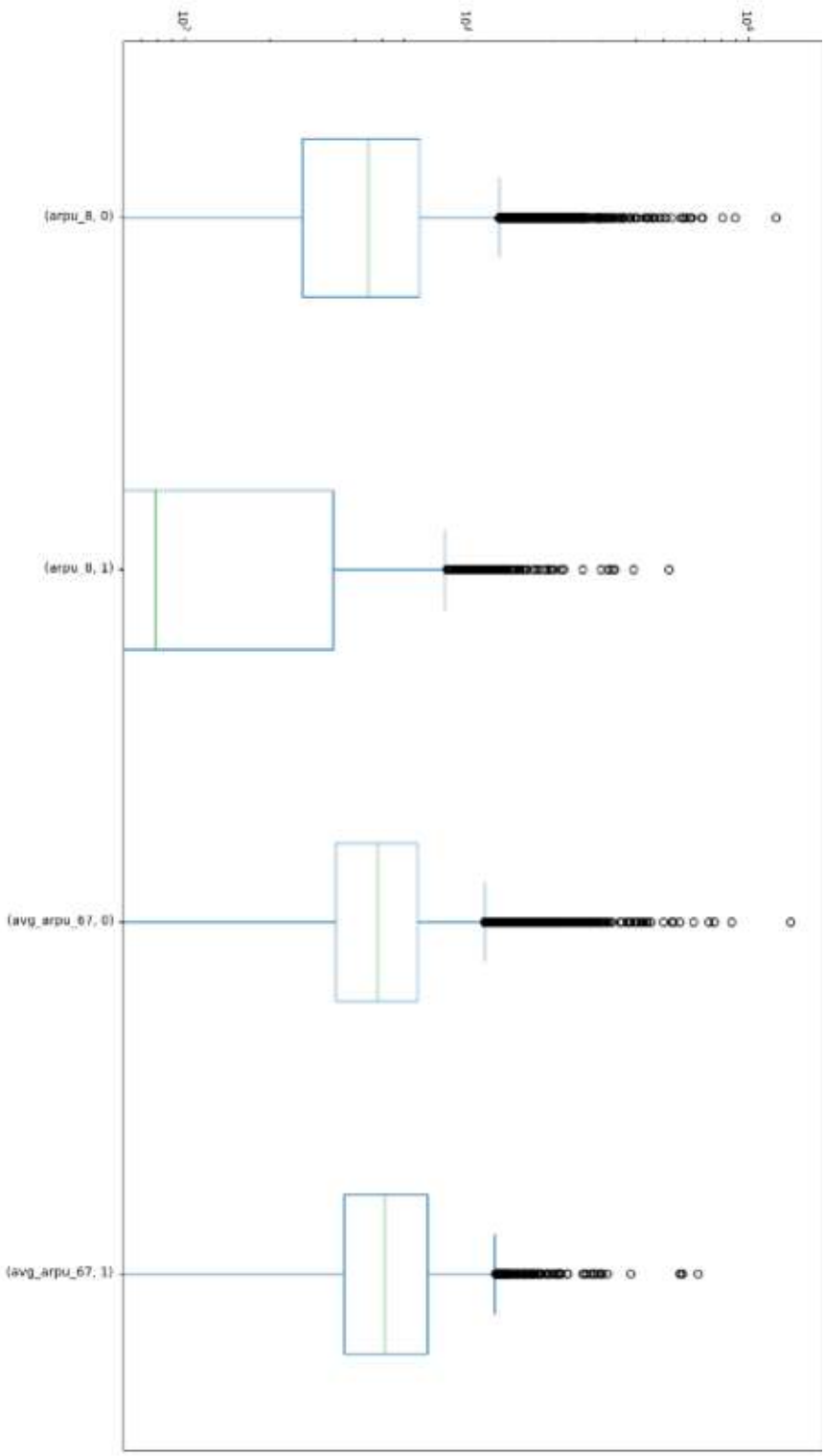
BAR CHART

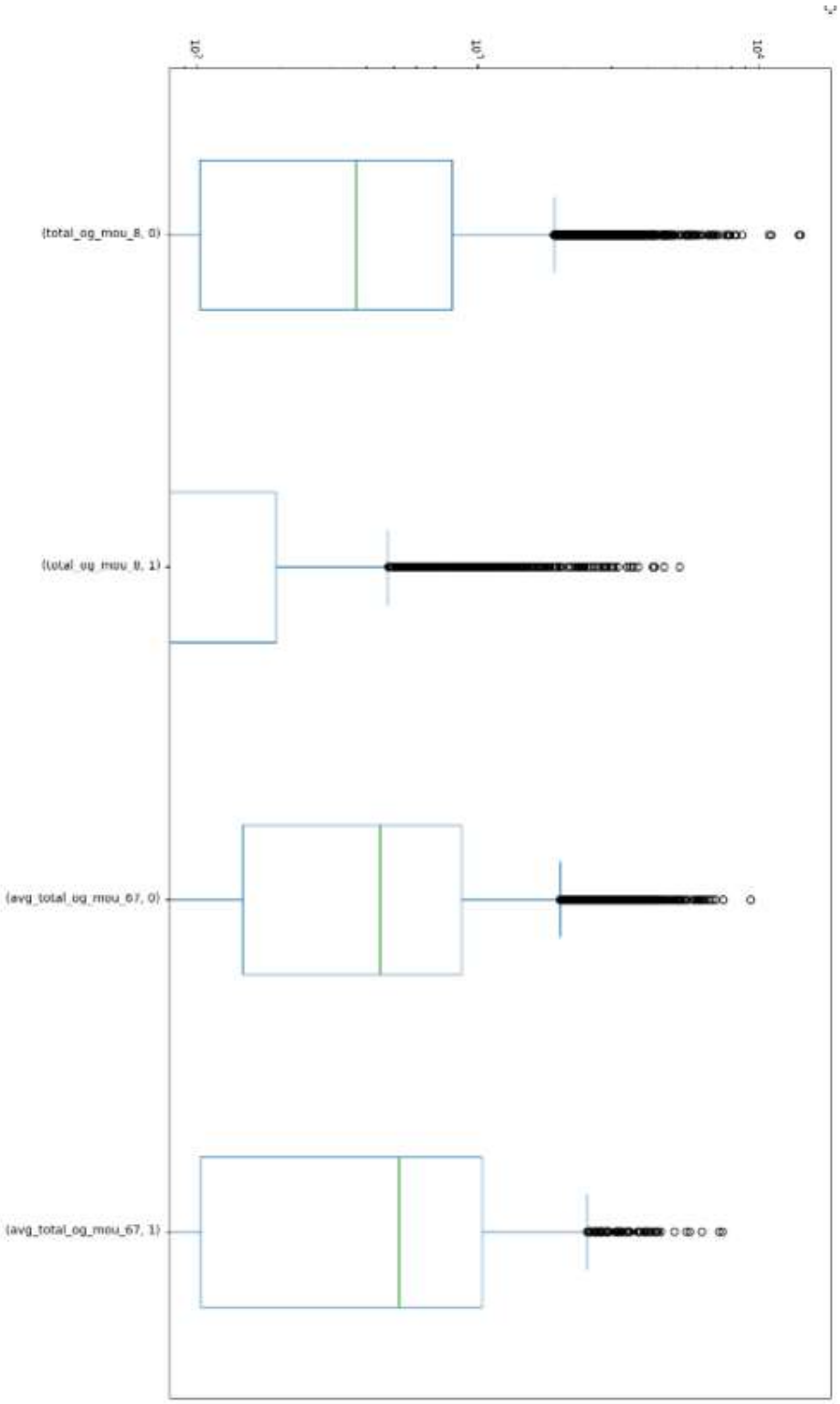


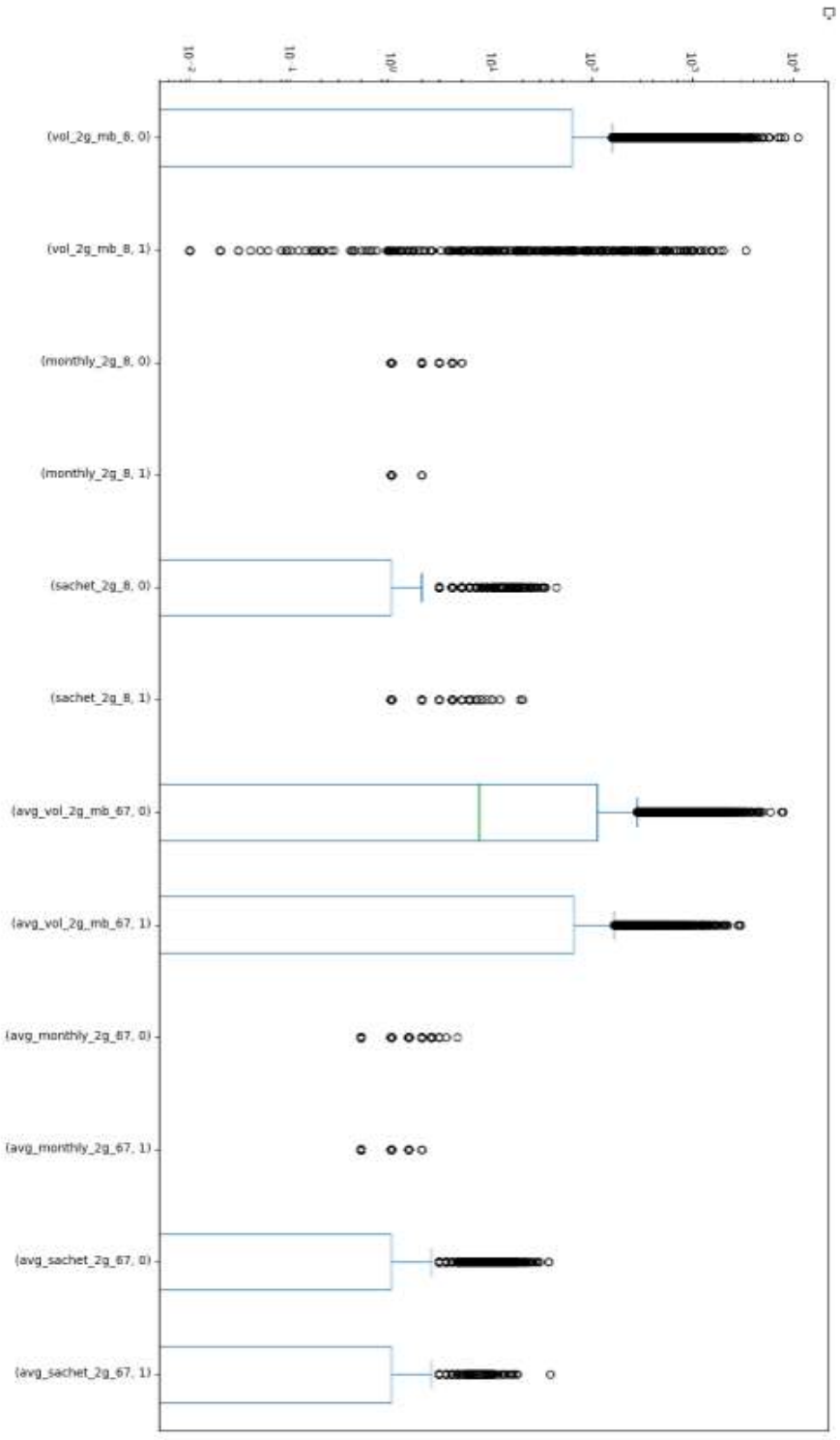


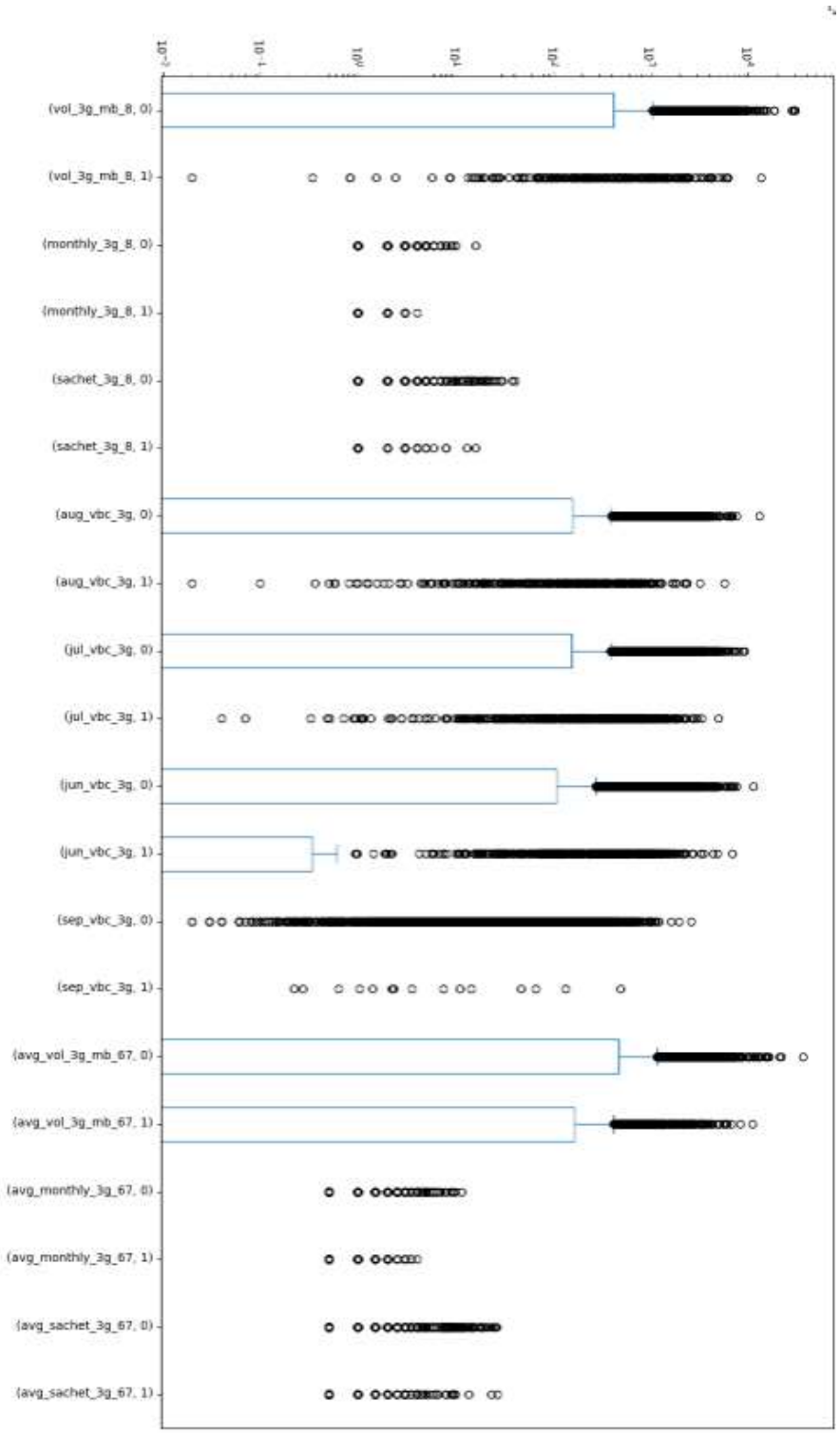
BOX PLOT

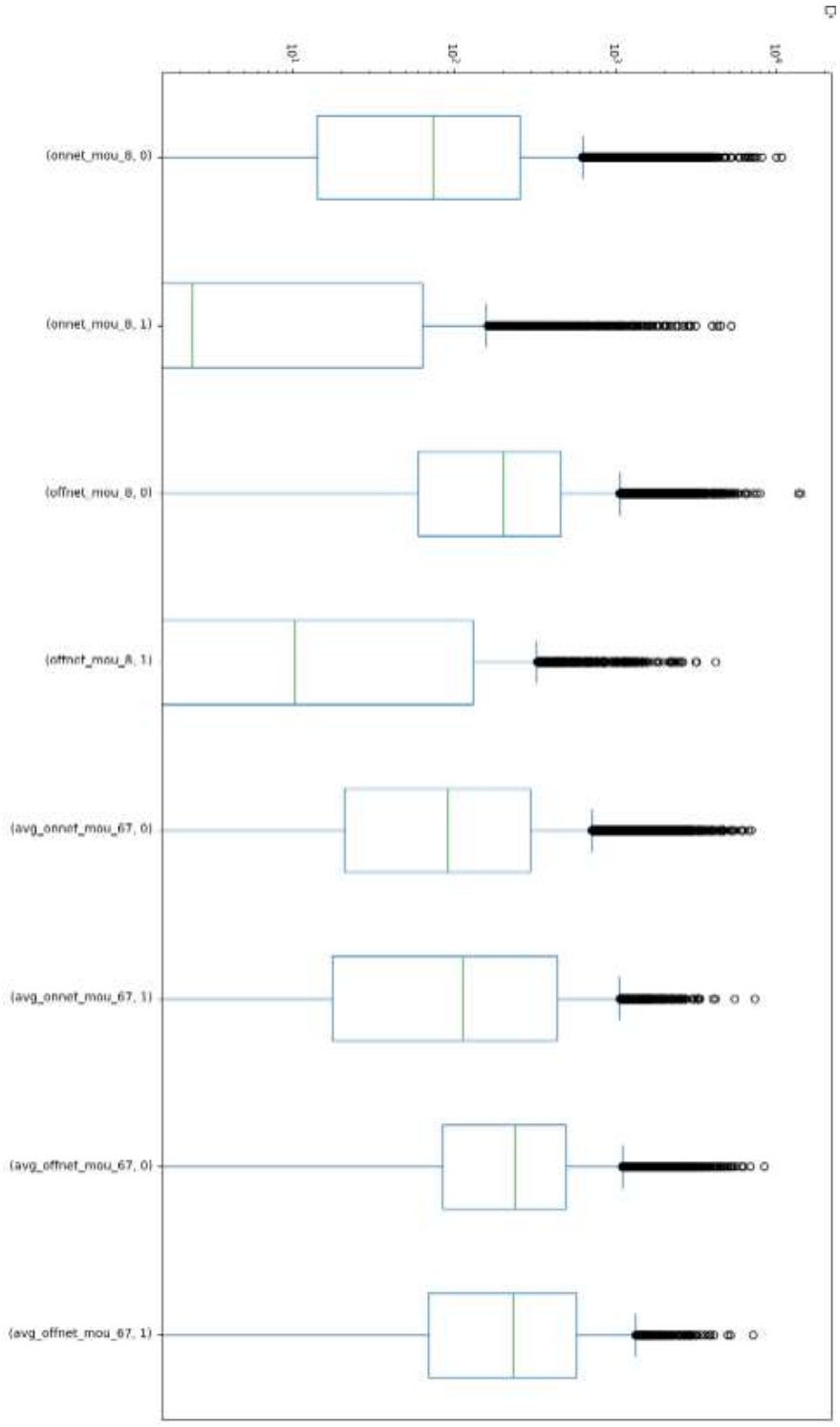

10^5
10^4
10^3
10^2
10^1
10^0
(total_rech_num_8, 0)
(total_rech_num_8, 1)
(total_rech_amt_8, 0)
(total_rech_amt_8, 1)
(max_rech_amt_8, 0)
(max_rech_amt_8, 1)
(total_rech_data_8, 0)
(total_rech_data_8, 1)
(avg_total_rech_amt_67, 0)
(avg_total_rech_amt_67, 1)
(avg_total_rech_num_67, 0)
(avg_total_rech_num_67, 1)
(avg_max_rech_amt_67, 0)
(avg_max_rech_amt_67, 1)
(avg_total_rech_data_67, 0)
(avg_total_rech_data_67, 1)
vg_av_rech_amt_data_67, 0)
vg_av_rech_amt_data_67, 1)


10^3
10^4
10^5
(arpu_8, 0)
(arpu_8, 1)
(avg_arpu_67, 0)
(avg_arpu_67, 1)


10^0
10^1
10^2
(total_og_mou_8, 0)
(total_og_mou_8, 1)
(avg_total_og_mou_67, 0)
(avg_total_og_mou_67, 1)

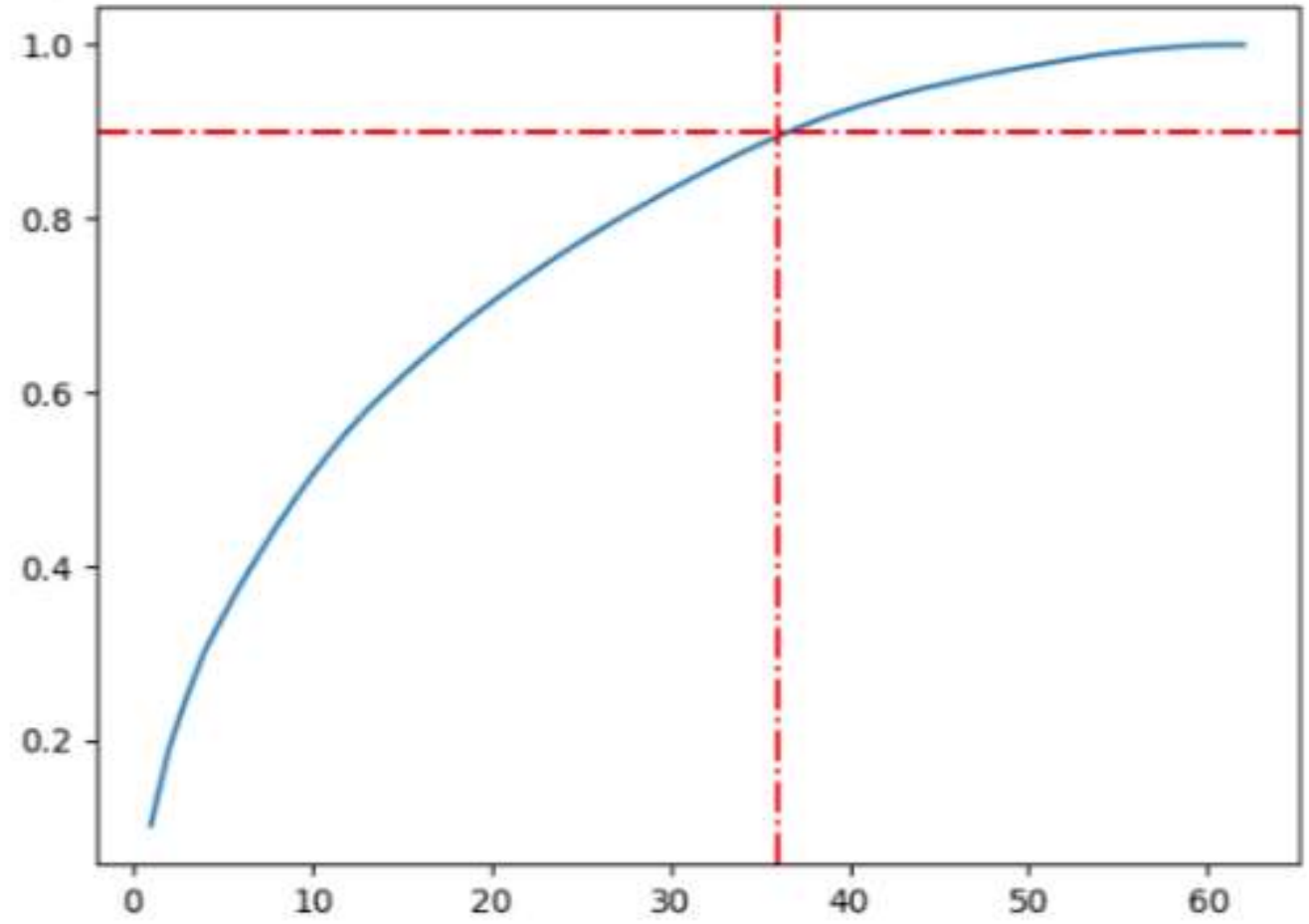

(vol_2g_mb_8, 0)
(vol_2g_mb_8, 1)
(monthly_2g_8, 0)
(monthly_2g_8, 1)
(sachet_2g_8, 0)
(sachet_2g_8, 1)
(avg_vol_2g_mb_67, 0)
(avg_vol_2g_mb_67, 1)
(avg_monthly_2g_67, 0)
(avg_monthly_2g_67, 1)
(avg_sachet_2g_67, 0)
(avg_sachet_2g_67, 1)


(vol_3g_mb_8, 0)
(vol_3g_mb_8, 1)
(monthly_3g_8, 0)
(monthly_3g_8, 1)
(sachet_3g_8, 0)
(sachet_3g_8, 1)
(aug_vbc_3g, 0)
(aug_vbc_3g, 1)
(jul_vbc_3g, 0)
(jul_vbc_3g, 1)
(jun_vbc_3g, 0)
(jun_vbc_3g, 1)
(sep_vbc_3g, 0)
(sep_vbc_3g, 1)
(avg_vol_3g_mb_67, 0)
(avg_vol_3g_mb_67, 1)
(avg_monthly_3g_67, 0)
(avg_monthly_3g_67, 1)
(avg_sachet_3g_67, 0)
(avg_sachet_3g_67, 1)

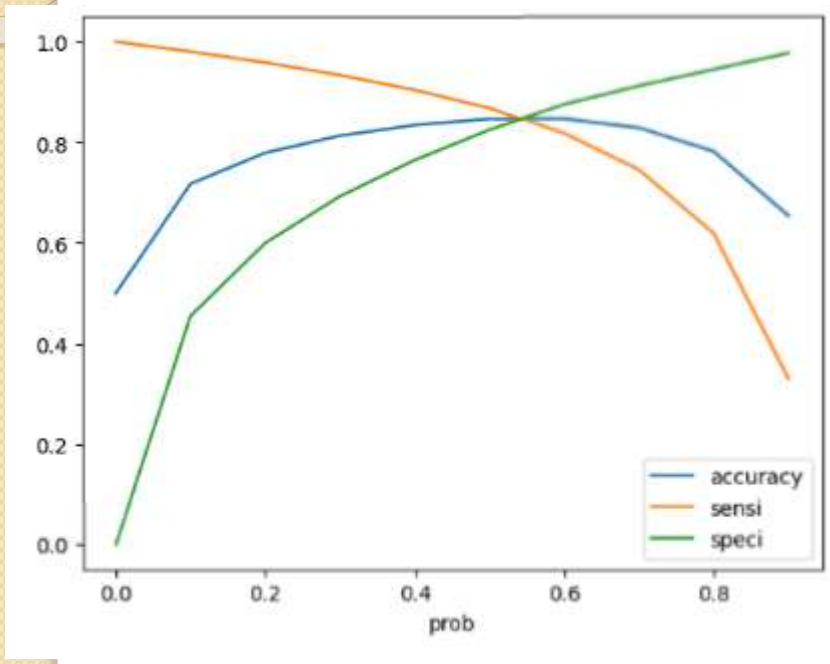

10¹
10²
10³
10⁴
(onnet_mou_8, 0)
(onnet_mou_8, 1)
(offnet_mou_8, 0)
(offnet_mou_8, 1)
(avg_onnet_mou_67, 0)
(avg_onnet_mou_67, 1)
(avg_offnet_mou_67, 0)
(avg_offnet_mou_67, 1)

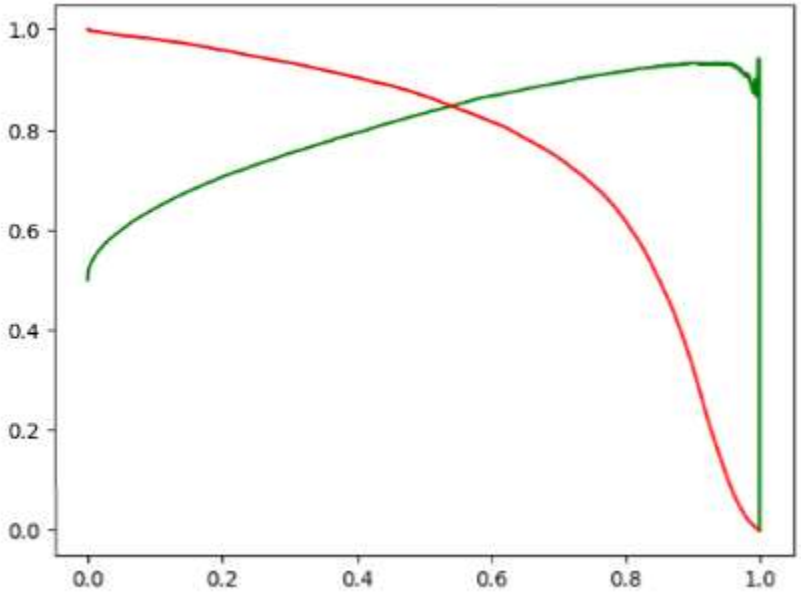
PCA

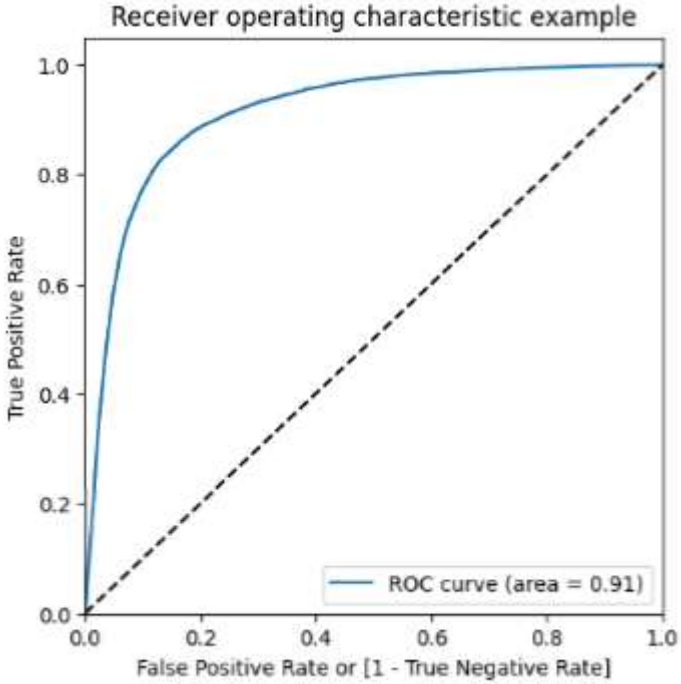
<matplotlib.lines.Line2D at 0x7fc68a59e2b0>

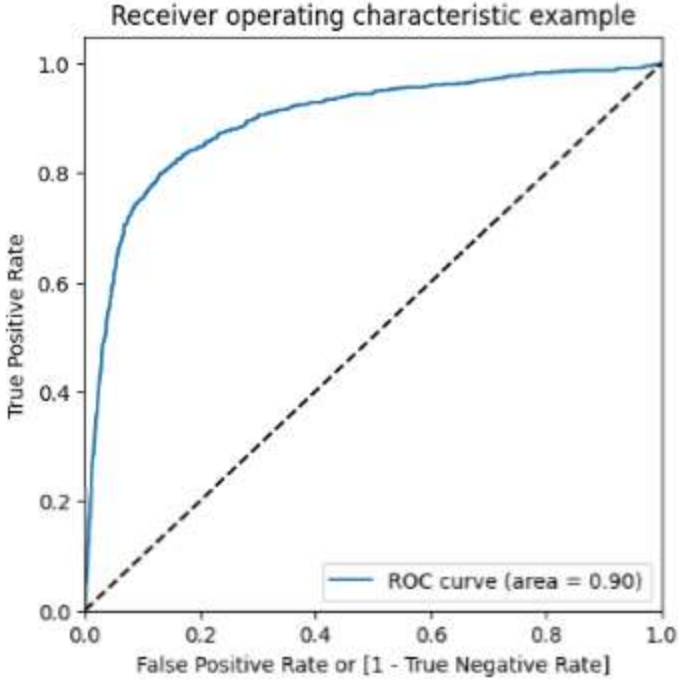


Logistic Regression with PCA

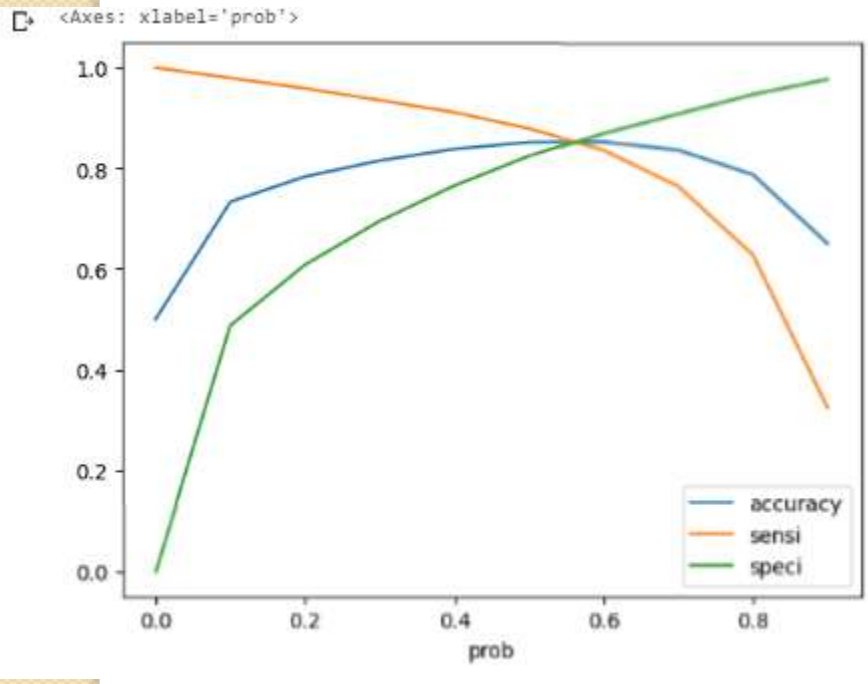


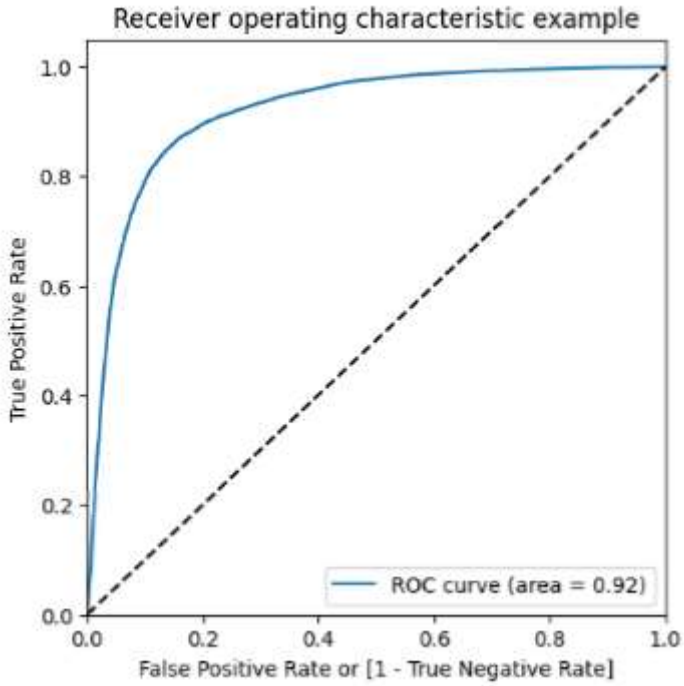


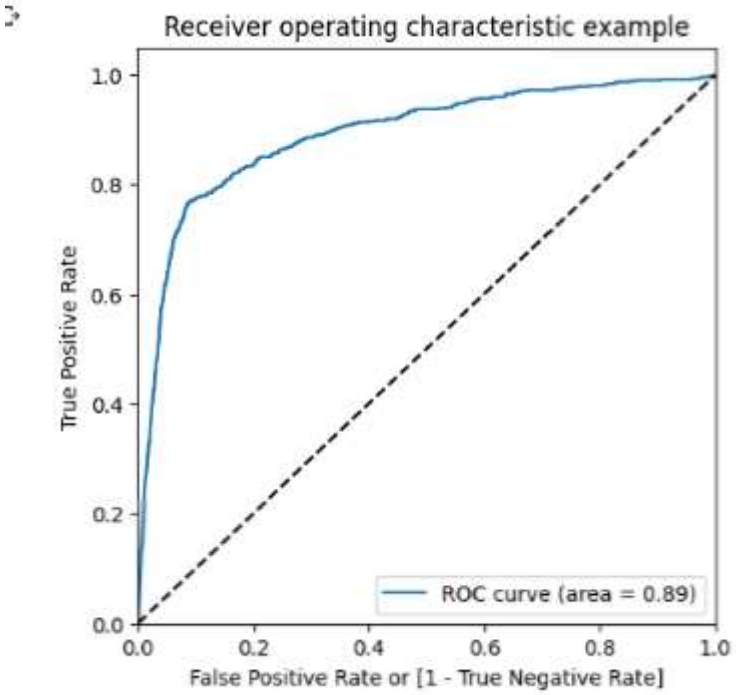

Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
ROC curve (area = 0.91)
0.0
0.2
0.4
0.6
0.8
1.0


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
ROC curve (area = 0.90)
0.0
0.2
0.4
0.6
0.8
1.0

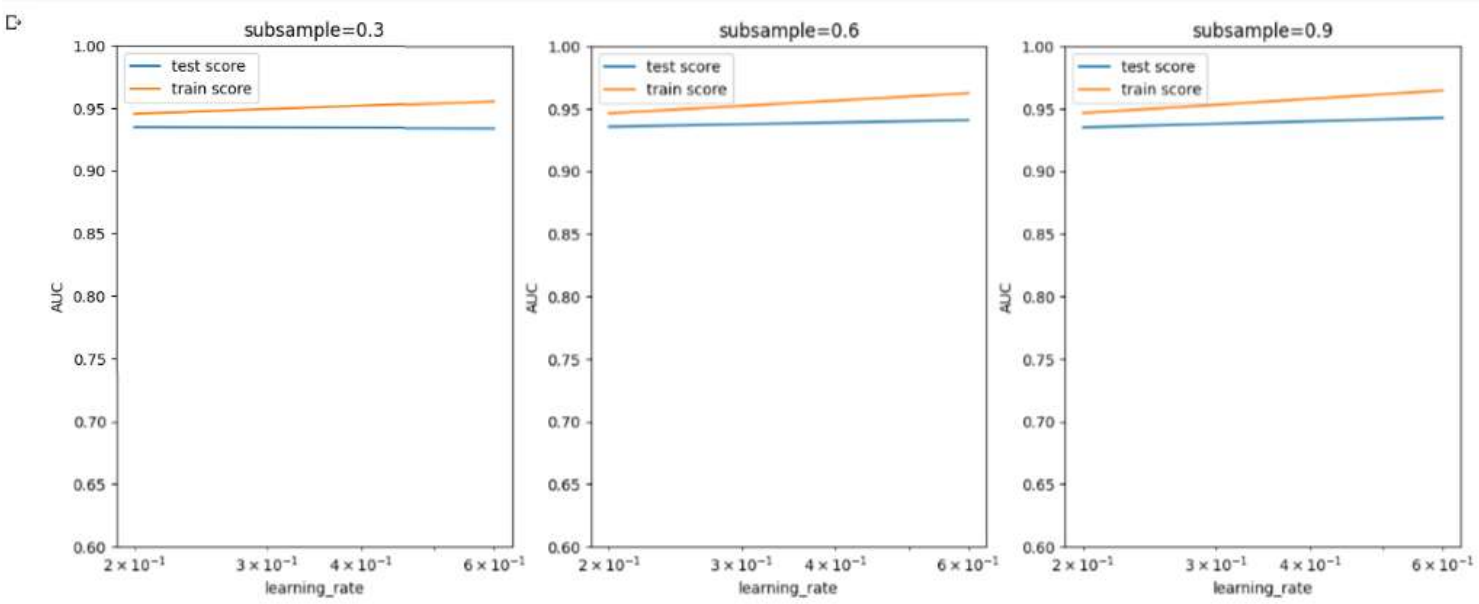
Logistic Regression with RFE

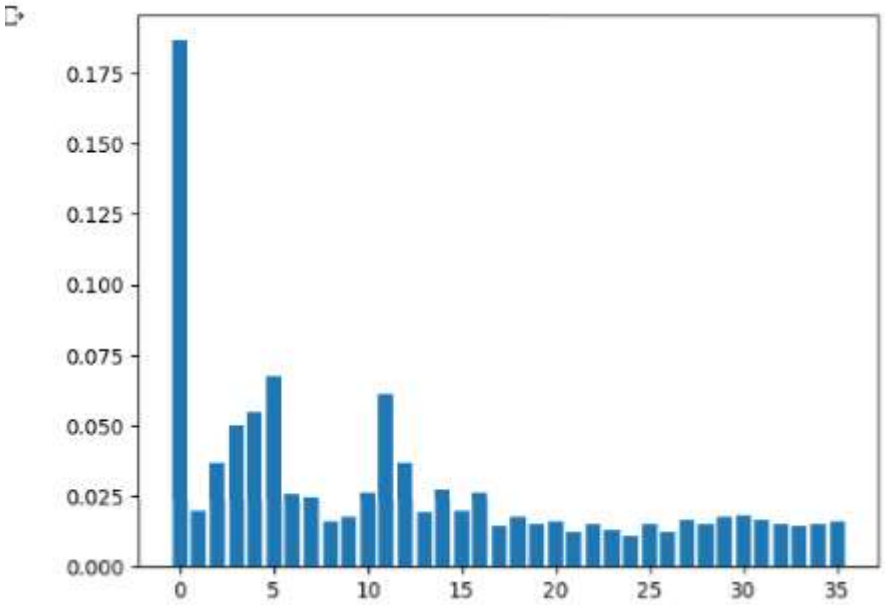





Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
ROC curve (area = 0.89)
0.0
0.2
0.4
0.6
0.8
1.0

XGBoost with PCA




0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
5
10
15
20
25
30
35

Model Results

- Model Accuracy is better along with Sensitivity and Specificity while using Logic Regression along with PCA. Model which is built with probablity of 0.6 or greater yielded Better results.
- Area under the curve also resulted in 0.92

Important Predictive variables which decide Churn are as follows:

- 1.total_og_mou_8
- 2.sep_vbc_3g
- 3.onnet_mou_8
- 4.loc_ic_t2m_mou_8
- 5.std_og_t2m_mou_8
- 6.avg_total_ic_mou_67
- 7.loc_ic_t2t_mou_8
- 8.og_others_8

Recommended Strategy

- Based on the above variables, reduced usage of the mobile either for call or data is a clear indication of churn.
- Hence on any month, if the usage of customer reduces or go below the average usage in previous month, then we do the following:
- 1.Roll out offers to retain them
- 2.Assign a dedicated customer service specialist to understand the reason for leaving and bring the issue to closure in case of any pending problems reported by customer.

THANK YOU