TELECOM CHURN CASE STUDY

VIRESH MOHAN

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

UNDERSTANDING CHURN

- There are two main models of payment in the telecom industry - postpaid (customers pay a monthly/annual bill after using the services) and prepaid (customers pay/recharge with a certain amount in advance and then use the services).
- In the postpaid model, when customers want to switch to another operator, they usually inform the existing operator to terminate the services, and you directly know that this is an instance of churn.
- However, in the prepaid model, customers who want to switch to another network can simply stop using the services without any notice, and it is hard to know whether someone has actually churned or is simply not using the services temporarily (e.g. someone may be on a trip abroad for a month or two and then intend to resume using the services again).
- Thus, churn prediction is usually more critical (and non-trivial) for prepaid customers, and the term ‘churn’ should be defined carefully. Also, prepaid is the most common model in India and Southeast Asia, while postpaid is more common in Europe in North America.
- Revenue-based churn: Customers who have not utilised any revenue-generating facilities such as mobile internet, outgoing calls, SMS etc. over a given period of time. One could also use aggregate metrics such as ‘customers who have generated less than INR 4 per month in total/average/median revenue.
- Usage-based churn: Customers who have not done any usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.

PROBLEM SOLUTION

Proposed Solution

Usage-based definition used to define 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 (mentioned later below) and predict churn only on high-value customers.

IMPLEMENTATION METHODOLOGY

Data Loading & Understanding

Loading & Understanding the past data provided by the Company

Data Cleaning

Handling of missing values, null values, unnecessary column elimination

EDA

Bivariate and Univariate Analysis

Data Preparation

Train-test split, scaling and standardisation, SMOTE for addressing class imbalance

IMPLEMENTATION METHODOLOGY

Model Building and Evaluation-I
- Logistic Regression with RFE & Manual Elimination (Interpretable Model)

Model Building and Evaluation-II
- PCA + Logistic Regression

Model Building and Evaluation-III
- PCA + Random Forest Classifier

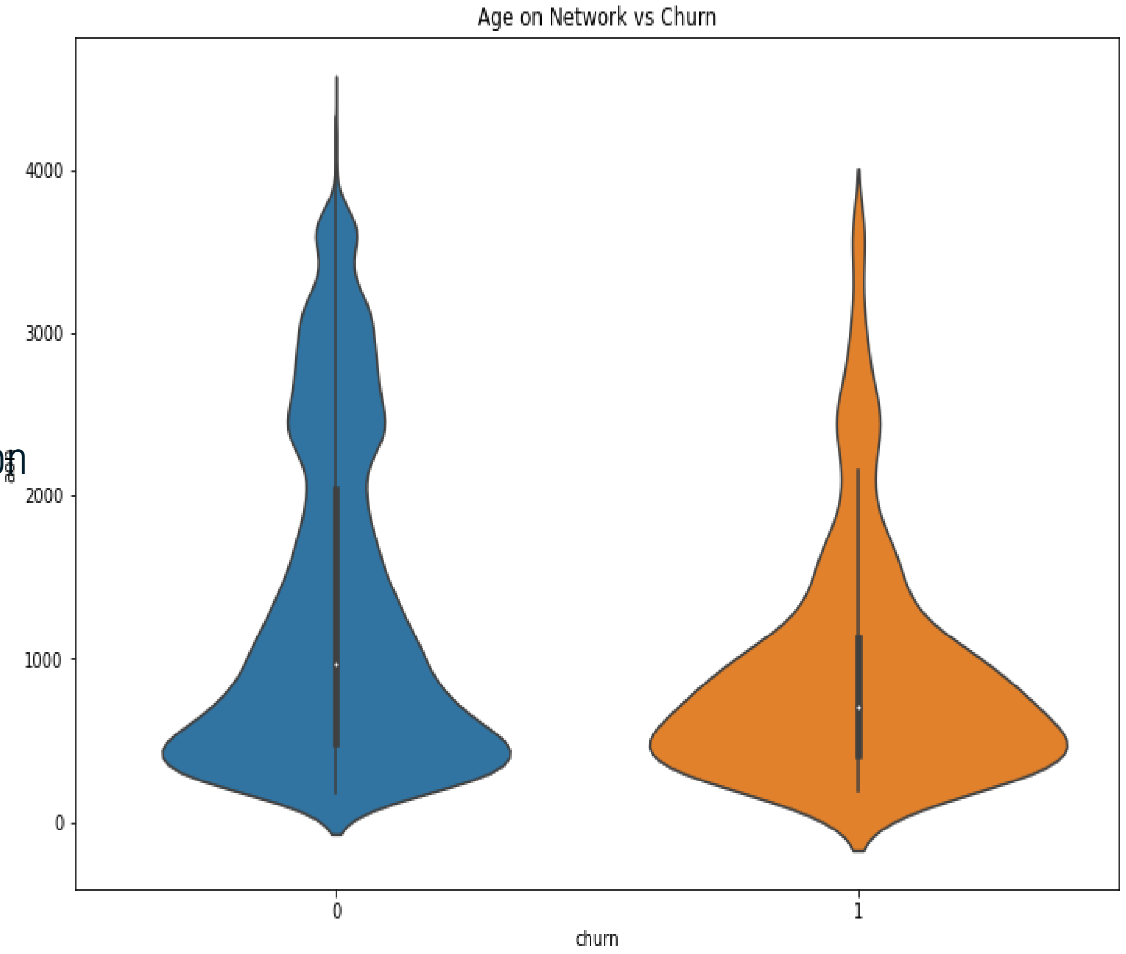
Model Building and Evaluation-IV
- PCA + XGBoost

Business Solutions
- Selection of best model and business recommendations

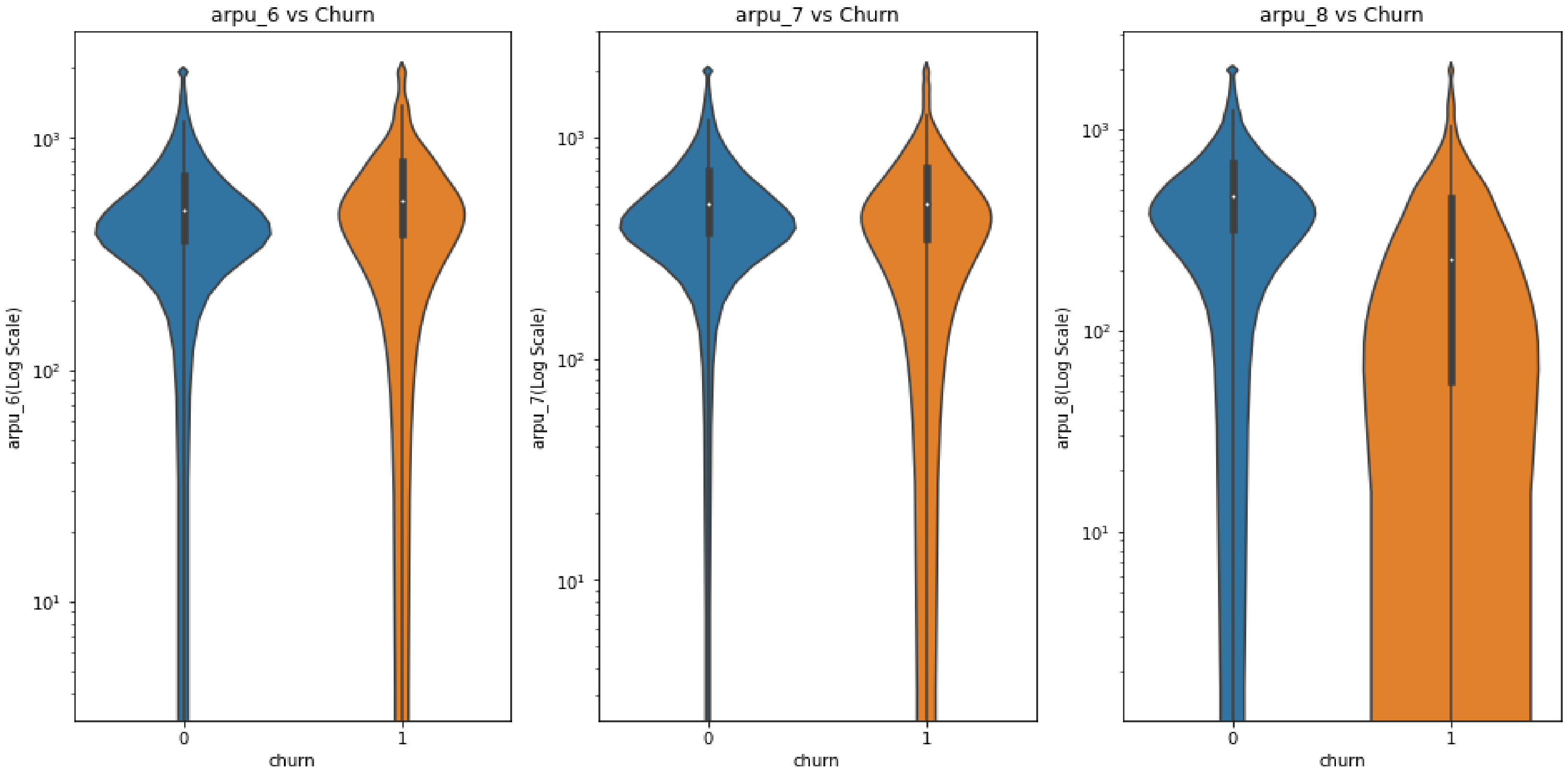
INSIGHTS FROM EDA

EFFECT OF AGE ON NETWORK ON CHURN

- It is clear to deduce that the clients with lesser AON(Age on Network) are more likely to churn

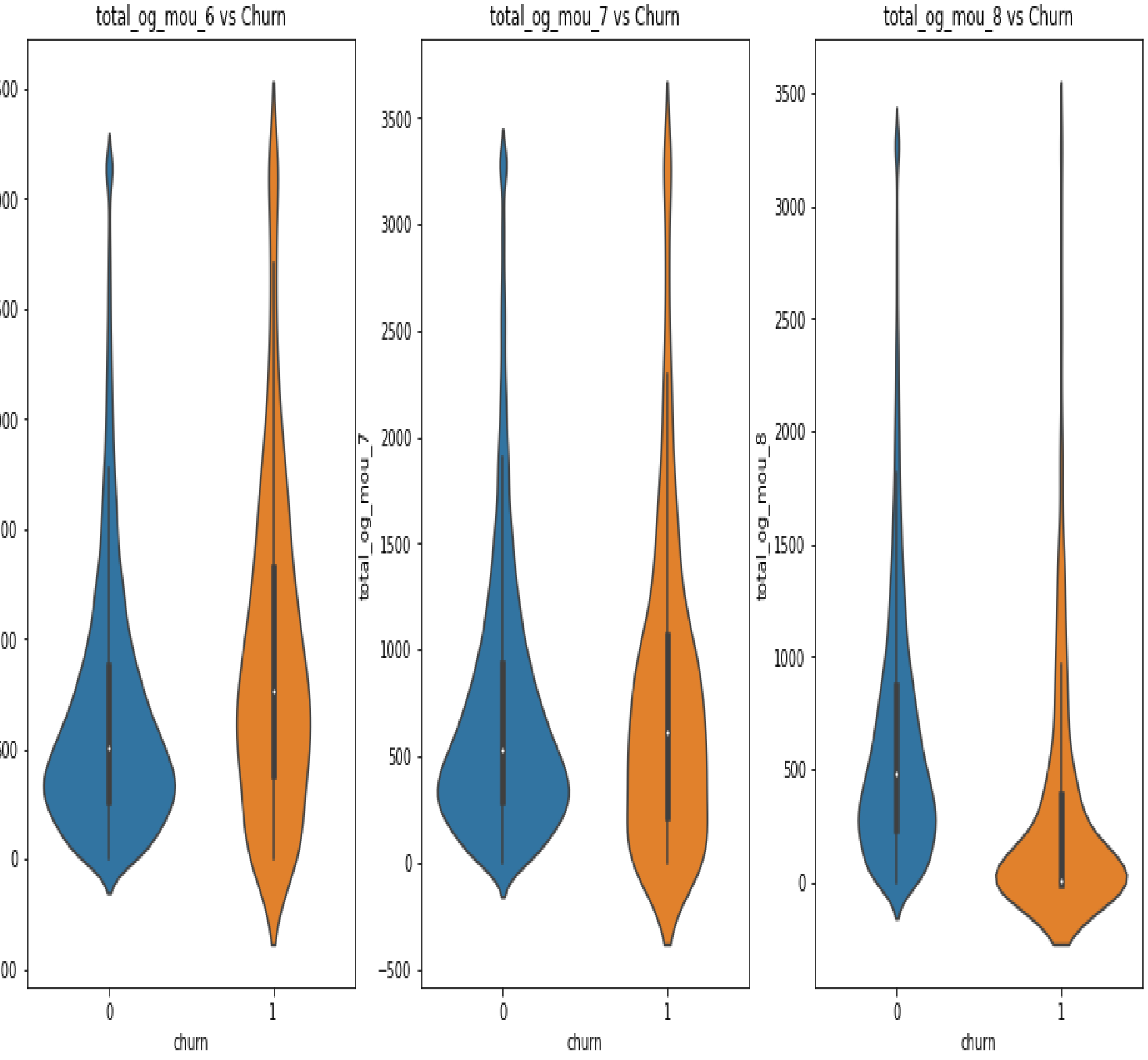


EFFECT OF AVERAGE REVENUE ON CHURN



- ARPU for churn customers start reducing from 8th month - the action phase. This is the phase where they start getting upset with the services available

- The customers more likely to churn have a decreasing trend in ARPU right from good phase , initially very low increase and then a drastic decrease



EFFECT OF OUTGOING CALLS ON CHURN

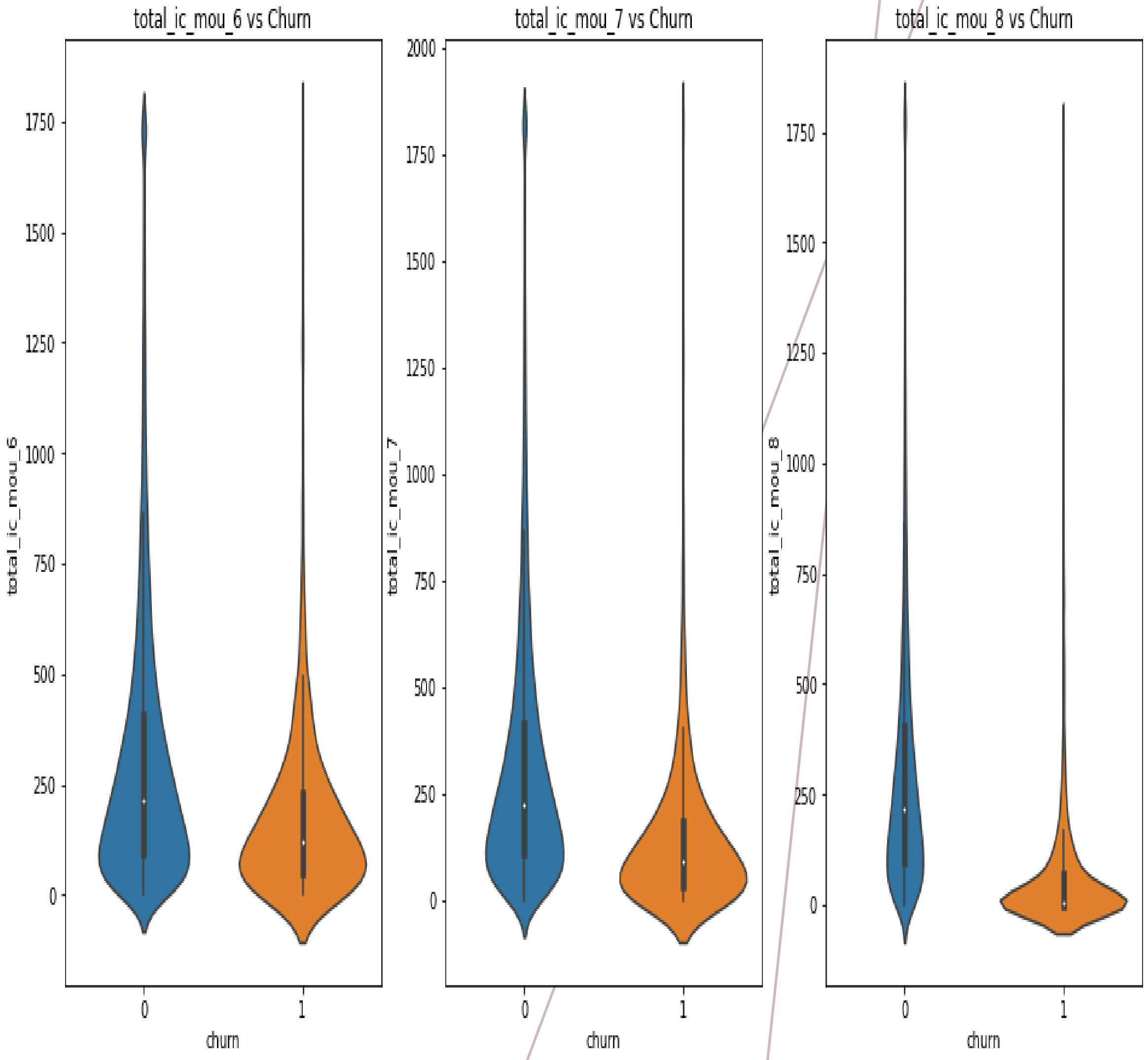
- For customers likely to churn , it is very clear that the outgoing calls have reduced to almost 1/3rd over 3 months. This is a very clear indication of churning

- Hence customers with decrease in mou from 6th to 7th month is more likely to churn

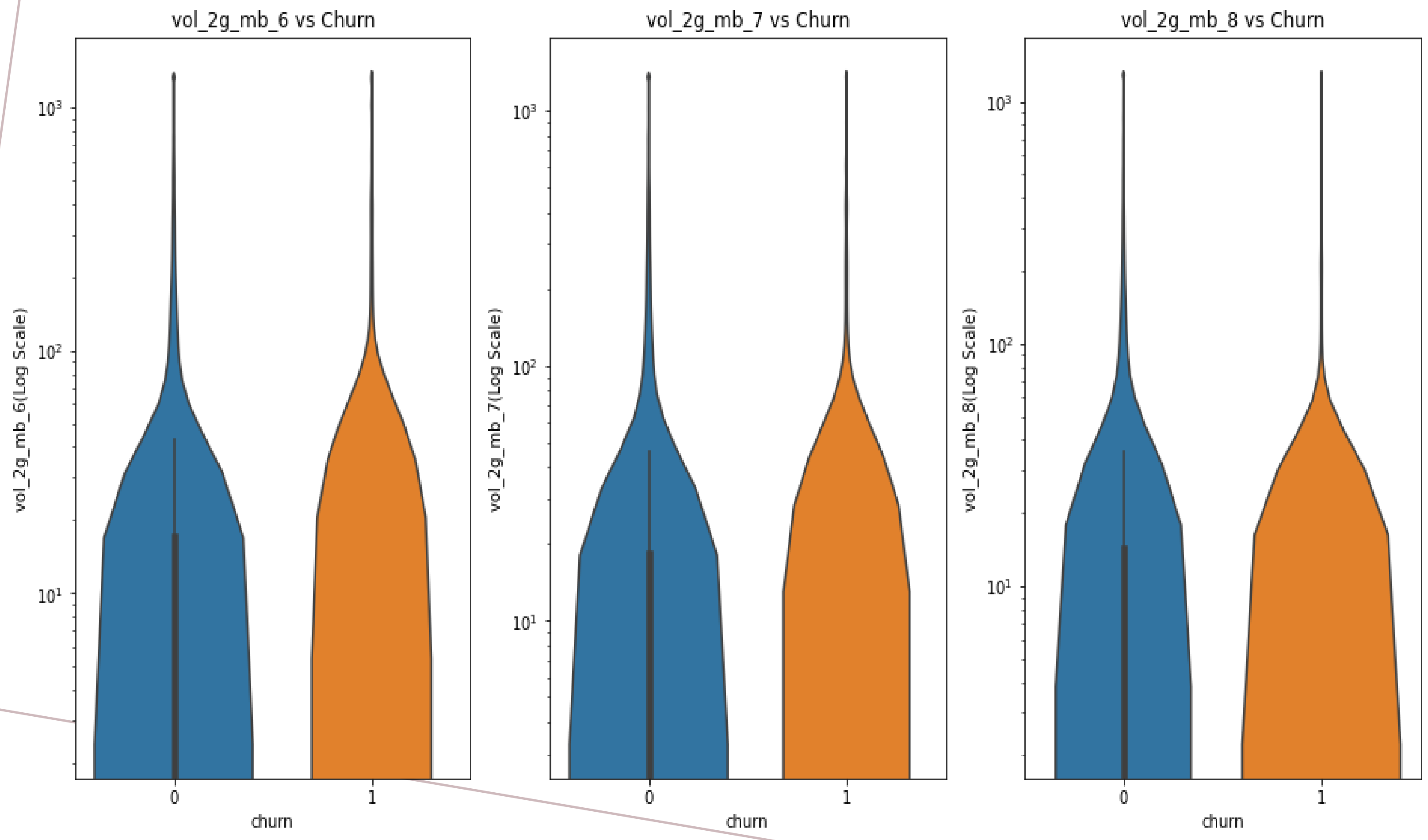
EFFECT OF INCOMING CALLS ON CHURN

- For customers likely to churn , it is very clear that the incoming calls have reduced to almost 1/3rd over 3 months. This is a very clear indication of churning. It has same pattern as outgoing calls

- Hence customers with decrease in incoming from 6th to 7th month is more likely to churn



EFFECT OF 2G USAGE

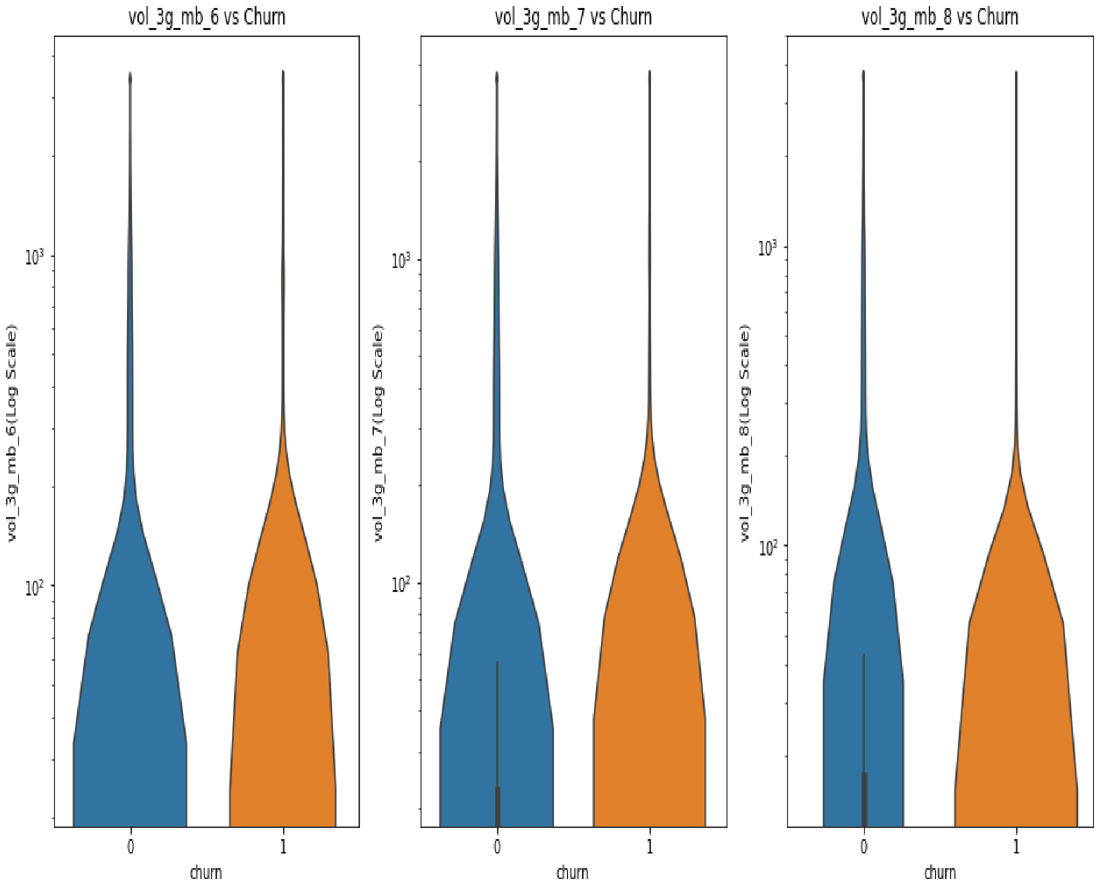


- No major changes detected . There is no clear pattern emerging from 2g usage

EFFECT OF 3G USAGE

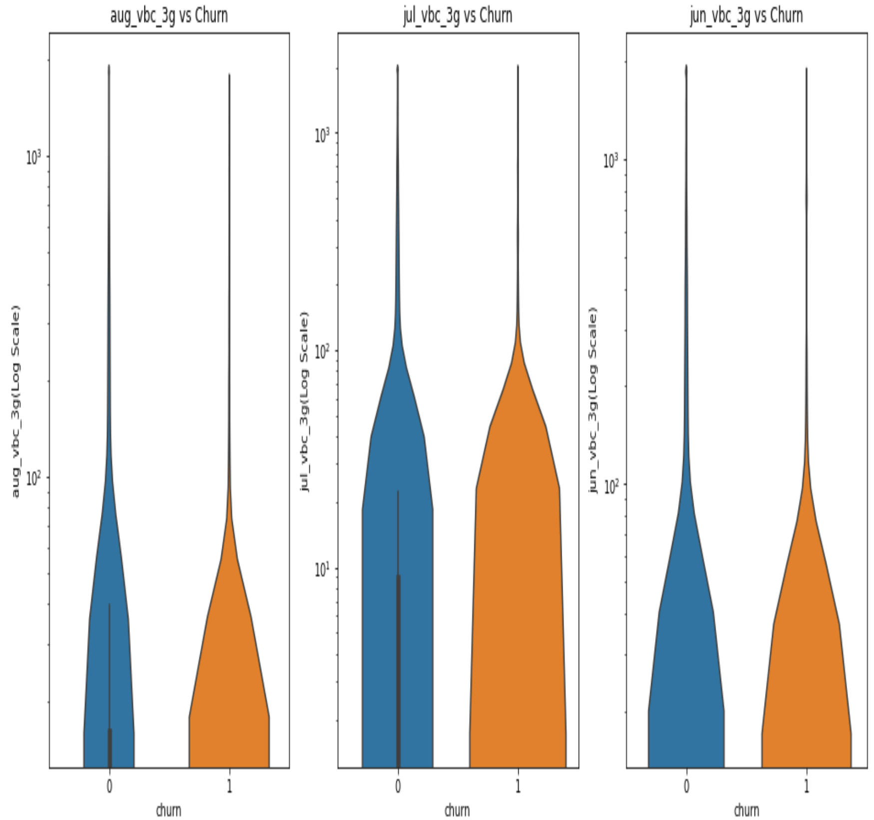
- For customers likely to churn , it is very clear that the 3g usage have reduced to almost 1/3rd over 3 months. This is a very clear indication of churning. It has same pattern as outgoing and incoming calls

- Hence customers with decrease in 3g usage from 6th to 7th month is more likely to churn

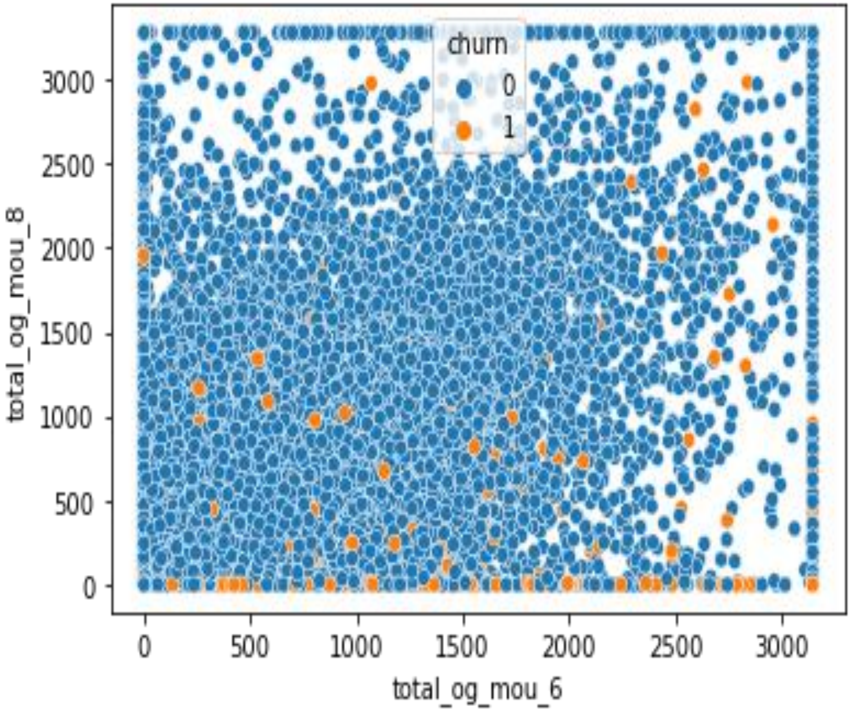


EFFECT OF 3G USAGE

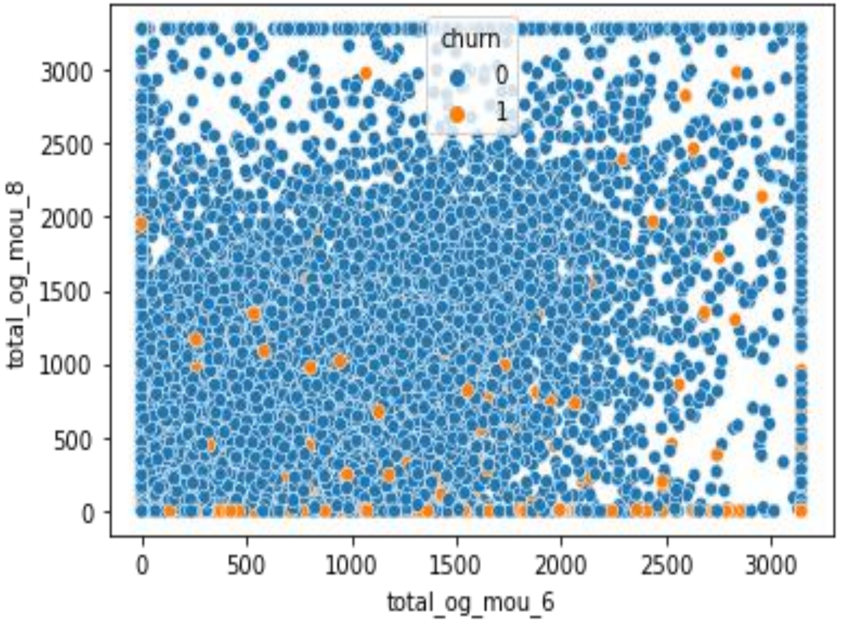
- As per univariate analysis , we can conclude incoming, outgoing and 3g usage decreasing trend contributes to more chances of churn
- We have also checked the monthly sachet usage , 3g usage , monthly recharges and we can see similar trends on the same



BIVARIATE ANALYSIS- CHECKING FOR THE OUTGOING CALLS IMPACT ON CHURN TAG



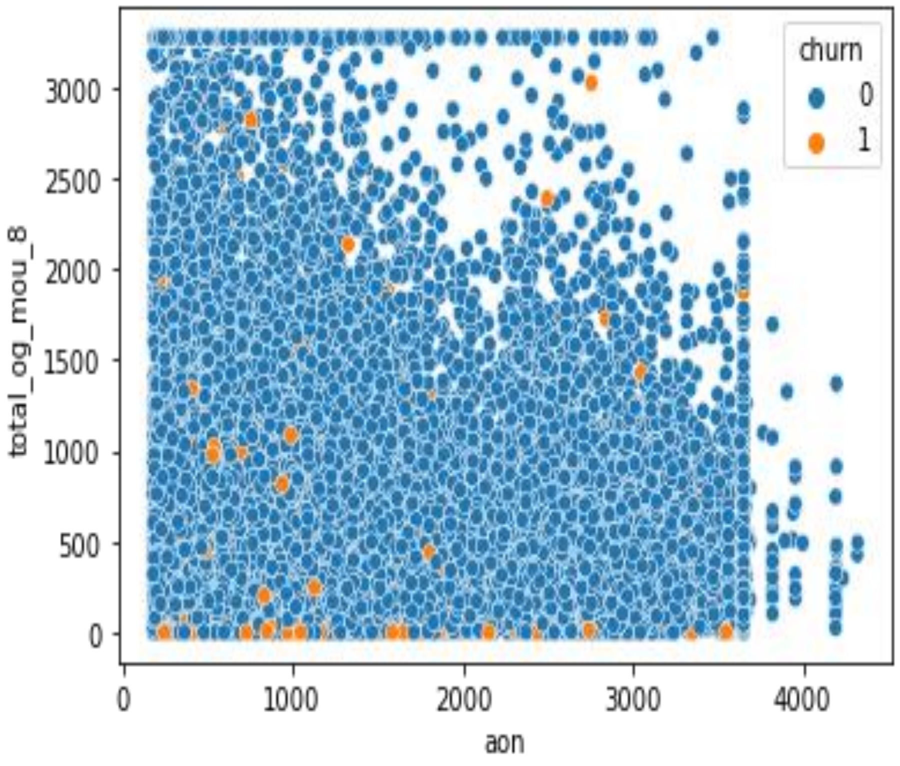
6th and 8th month(good phase to action phase)



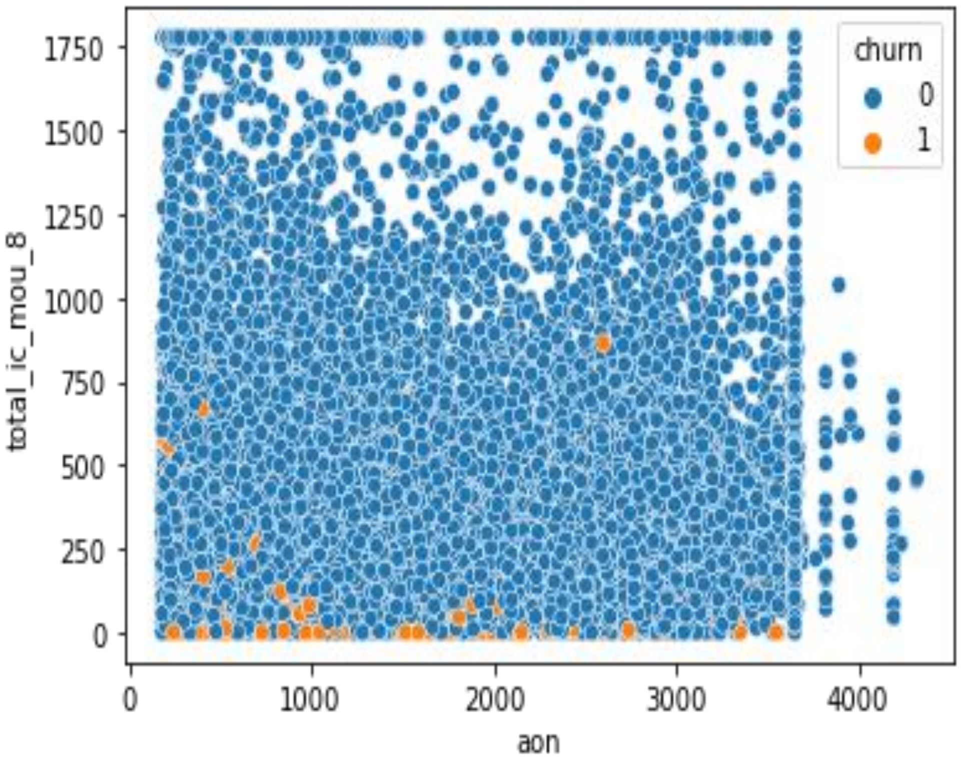
6th and 7th month(good phase to action phase)

- Clear indication that there is slight reduction in outgoing calls from 6th to 7th month are more likely to churn
- Clear indication that sharp reduction in outgoing calls from 6th to 8th month are more likely to churn

IMPACT OF AON ON INCOMING AND OUTGOING CALLS IN 8TH MONTH



Outgoing Calls



Incoming Calls

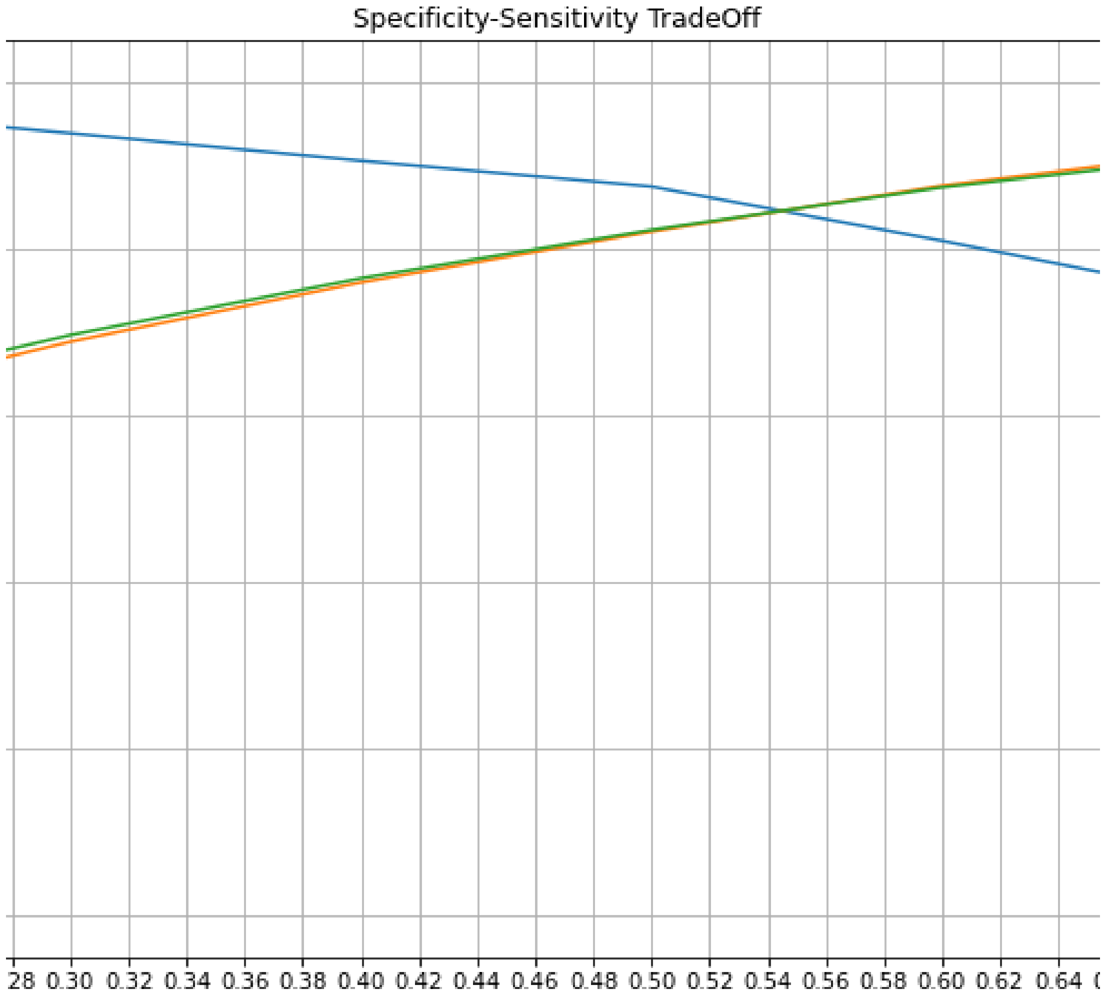
- The customers with less total_ic_mou_8 are more likely to churn irrespective of aon.
- The customers with total_ic_mou_8 > 2000 are very less likely to churn.

MODEL BUILDING AND EVALUATION

MODEL I :LOGISTICS REGRESSION

- Base Model : Base Model developed and cut-off of specificity – sensitivity trade-off identified at 0.55

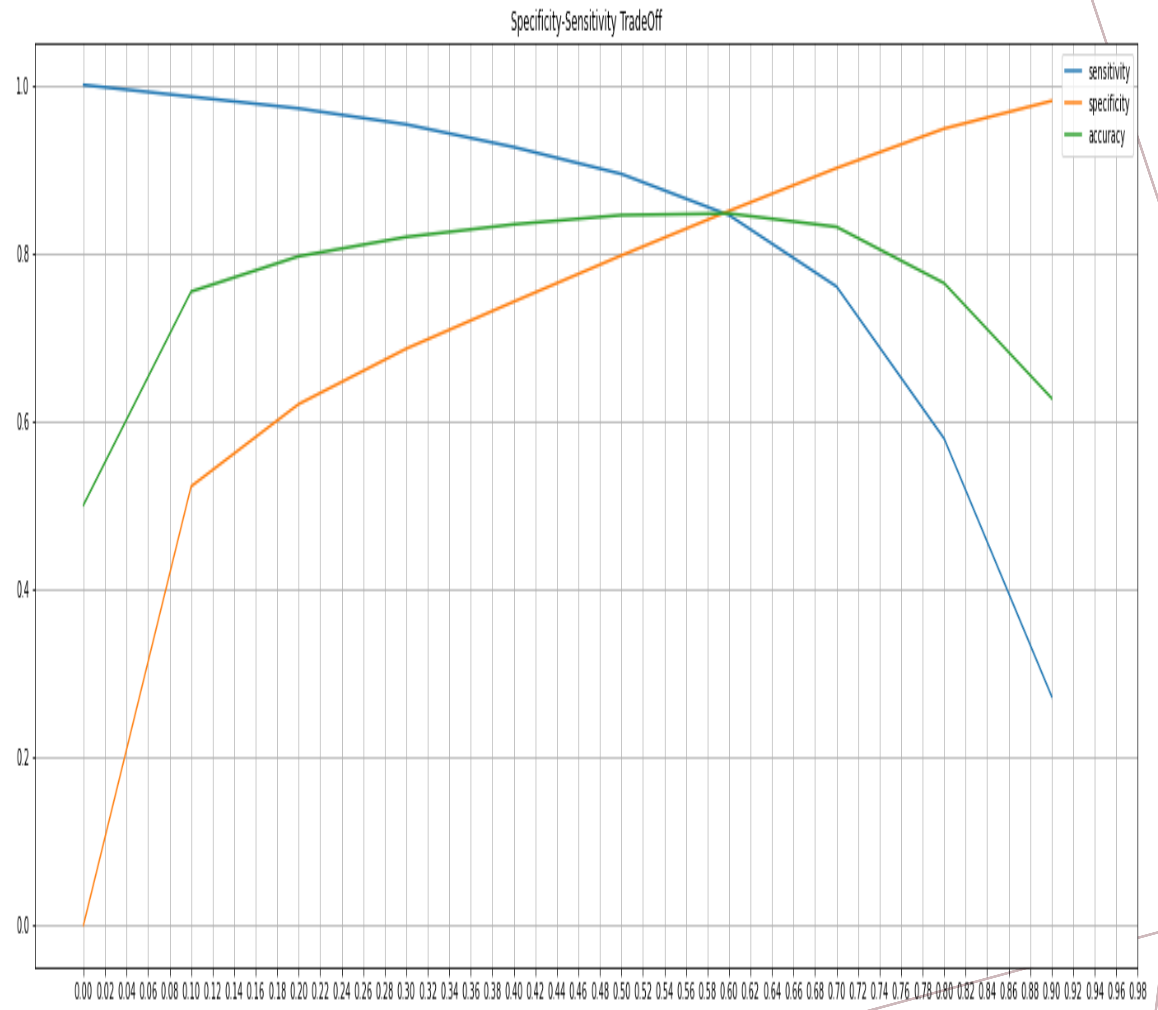
- Further, feature selection / engineering is done using RFE as well as manual feature elimination techniques



MODEL I :LOGISTICS REGRESSION- FINAL MODEL

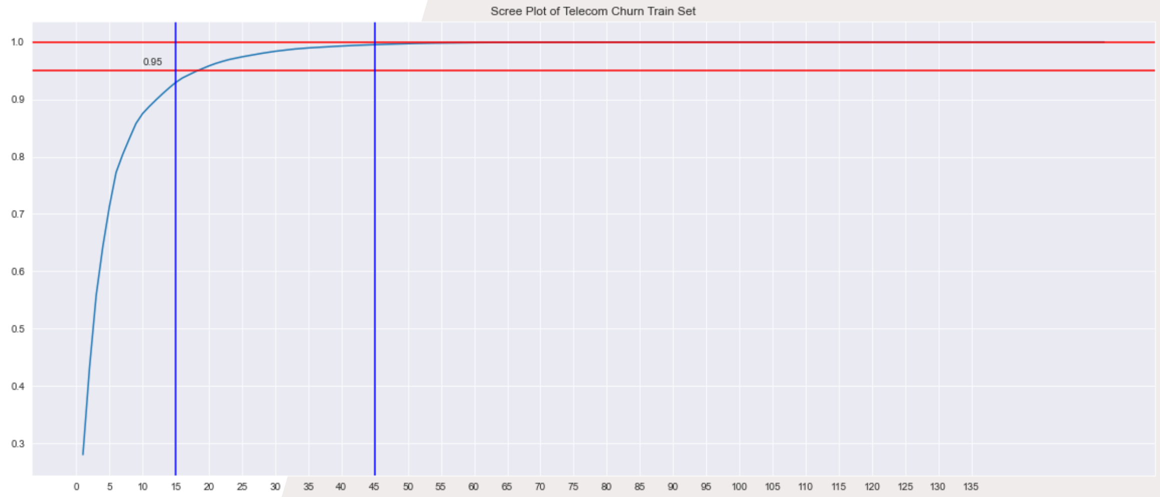
- Final Model : Final Model developed after 7 iterations and optimal cutoff identified as 0.59
- Metrics of test – train as below:

Metrics	Train Data	Test Data
Accuracy	0.848	0.843
Sensitivity / TPR / Recall	0.852	0.806
Specificity/TNR	0.845	0.845
Precision	0.846	0.156
F1-Score	0.849	0.261
ROC AUC score	0.919	0.896



MODEL II : PCA + LOGISTICS REGRESSION

- Scree Plot indicates that 95% of variance in the train set can be explained by first 18 principal components and 100% of variance is explained by the first 48 principal components



- PCA done and scatter plot of PCA wrto churn plotted

MODEL II : PCA + LOGISTICS REGRESSION

- Baseline Model developed
- Hyperparameter tuning done on model and the metrics are identified for test and train sets as below:

Metrics	Baseline		Final	
	Train Data	Test Data	Train Data	Test Data
Accuracy	0.747	0.361	0.728	0.334
Sensitivity / TPR / Recall	0.893	0.844	0.899	0.847
Specificity/TNR	0.742	0.344	0.723	0.316
Precision	0.108	0.044	0.102	0.042
F1-Score	0.193	0.084	0.183	0.08

MODEL III :PCA + RANDOM FOREST

- Hyperparameter tuning done and random forest model developed
- Metrics as shown:
- OOB score seen as 90.4%
- This indicates the model is overfitted with very high accuracy

Metrics	Train Data	Test Data
Accuracy	0.918	0.966
Sensitivity / TPR / Recall	0.865	0
Specificity/TNR	0.92	1
Precision	0.274	-
F1-Score	0.416	-

MODEL IV: PCA+XGBOOST

- Hyperparameter tuning done and model developed
- Metrics as shown:

Metrics	Train Data	Test Data
Accuracy	0.932	0.887
Sensitivity / TPR / Recall	0.995	0.233
Specificity/TNR	0.93	0.910
Precision	0.331	0.084
F1-Score	0.497	0.123
ROC-AUC Score	0.987	0.676

BUSINESS RECOMMENDATIONS

- Customers who churn show higher average monthly local outgoing calls in the action period by 1.31 standard deviations , compared to users who don't churn , when all other factors are held constant. This is the strongest indicator of churn.
- Customers who churn show higher roaming incoming calls in action period by 0.96 standard deviations, when all other factors are held constant. This is the second strongest indicator of churn.
- Further customers who churn have done 0.83 standard deviations higher roaming incoming calls than non-churn customers. This factor when coupled with above factors is a good indicator of churn.
- Customers who churn are more likely to have average revenue lesser by 0.66 points in action period , when all other factors are held constant.
- Customer who churn show lesser outgoing calls to mobile by 0.66 standard deviations compared to action phase, when all other factors are kept constant.
- There are multiple other factors which are affecting the churn factor , but above listed ones are the top 5

BUSINESS RECOMMENDATIONS

- Concentrate on users with 1.31 std devations lower than average local outgoing calls. They are most likely to churn.
- Concentrate on users who have lesser roaming incoming calls in the 8th month by 0.96 standard deviations compared to action phase. They are second most likely to churn.This may show the rual customers as indicated in the earlier business problem explanation
- Models with high sensitivity are the best for predicting churn. Use the PCA + Logistic Regression model to predict churn. It has a test sensitivity of 84.7% highest among all 4 models with logistics regression also having test sensitivity of 80.6%

THANK YOU