Telecom Churn Case Study

Aazeenelsa Shaikh

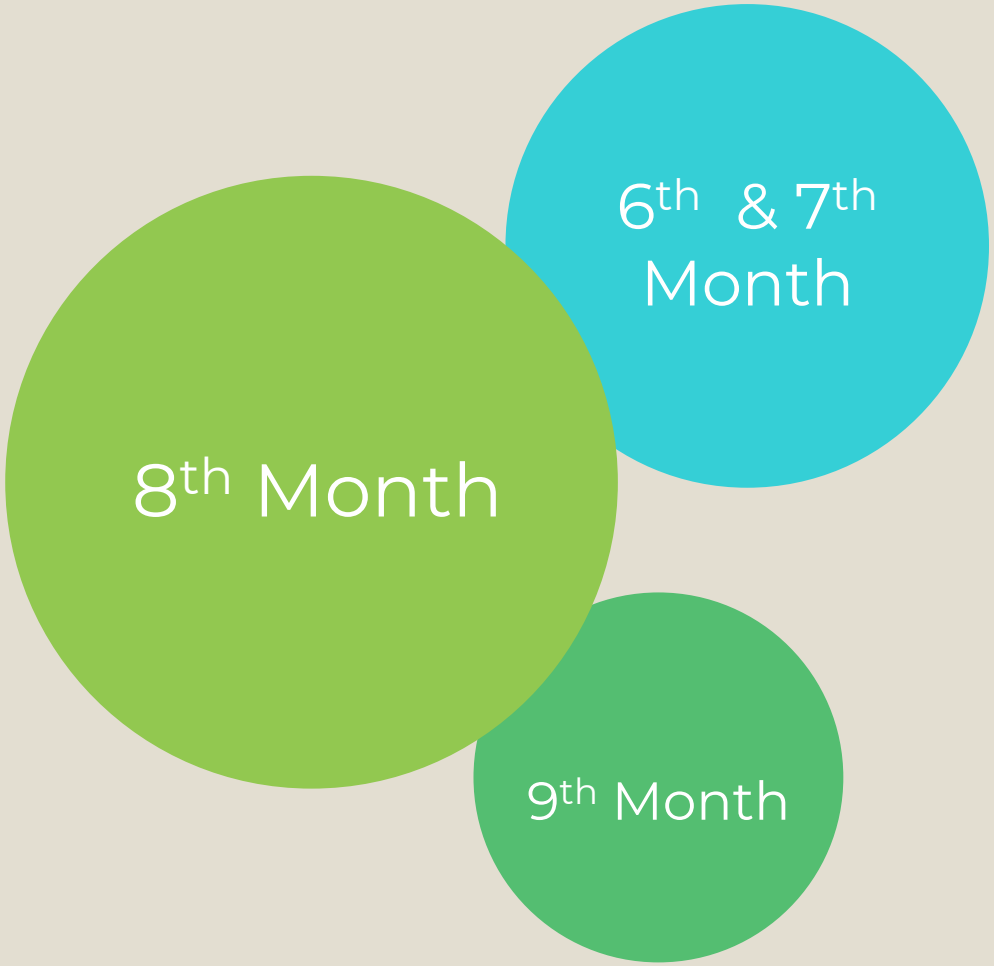
Data Analysis

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, while the fourth month is the 'churn' phase.

Good Phase.

Action Phase.

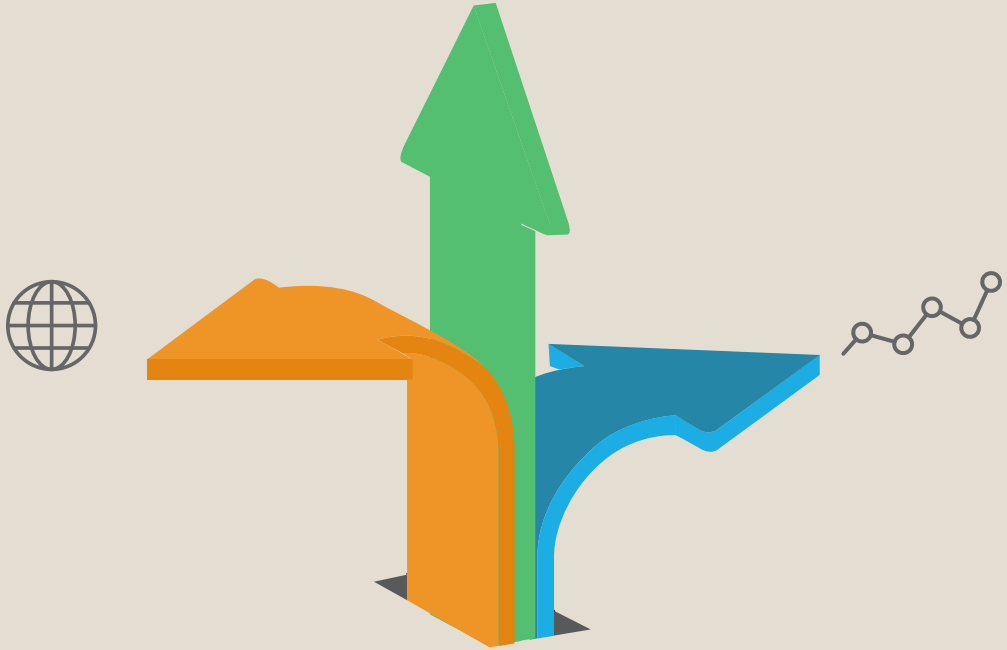
Churn Phase.



Data Preparation

Filter High Value Customers

Those who have recharged with an amount more than or equal to X is greater than 70th percentile of the average recharge mount in the first two months (the good phase)
70 percentile of 6th and 7th months avg recharge amount: 478.0
Data frame Shape after filtering High Value Customers: (29953, 230)



Derive New Features

We have derived more meaningful information

- Total recharge amount
- Total recharge for data
- Last date of recharging the data
- Average recharge amount for data
- Maximum recharge for data

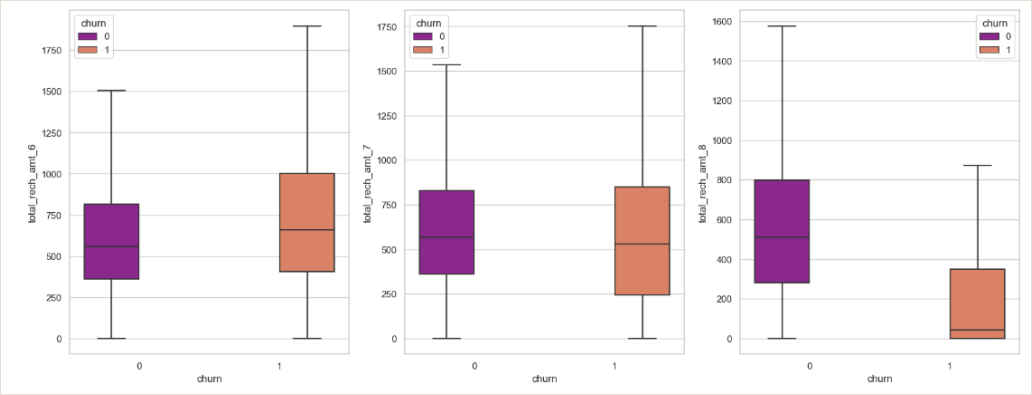
Tag Churners and remove attributes of the churn phase

Now tag the churned customers [churn = 1, else 0] based on the fourth month as follows:

- Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase.
- The attributes you need to use to tag churners are **total_ic_mou_9, total_og_mou_9, vol_2g_mb_9, vol_3g_mb_9
- After tagging churners, remove all the attributes corresponding to the churn phase(all attributes having '_9', etc. in their names)

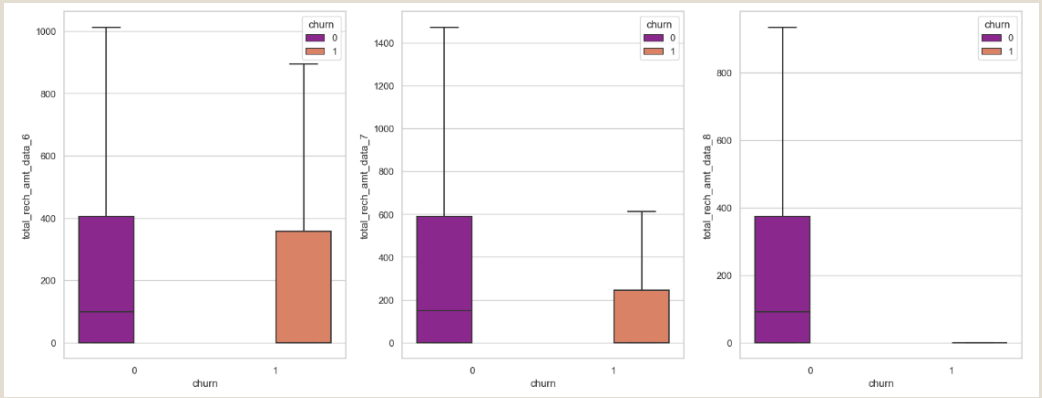
Exploratory Data Analysis

a. Recharge amount related variables



Total Recharge Amount

We can observe a drop in the total recharge amount for churned customers in the 8th Month(Action Phase)

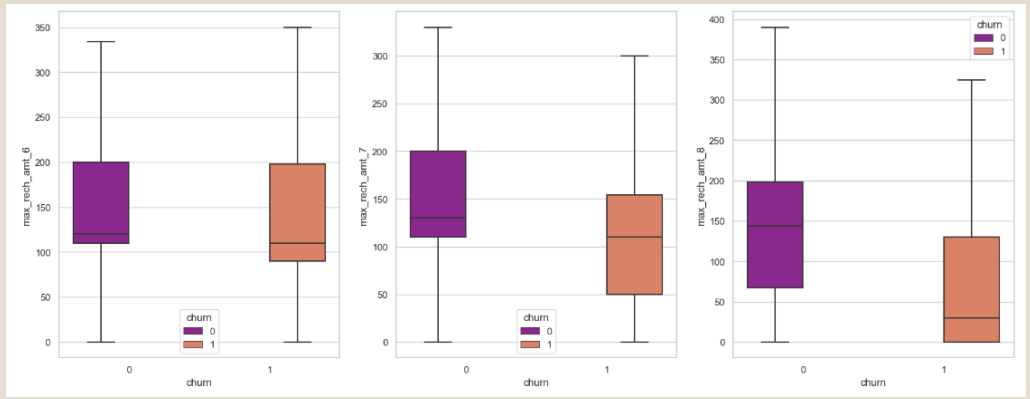


Total Recharge Amount for Data

We can observe that there is a huge drop in total recharge amount for data in the 8th Month(Action Phase) for churned customers

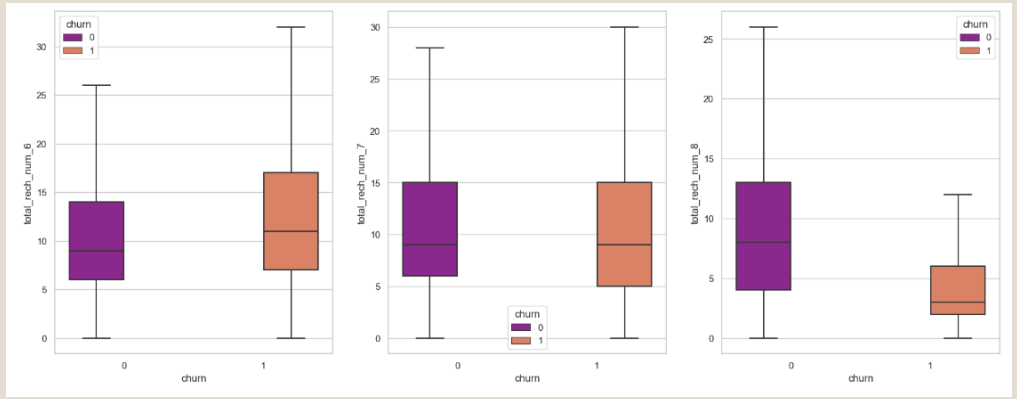
Exploratory Data Analysis

a. Recharge amount related variables



Maximum Recharge Amount for Data

We can observe that there is a huge drop in maximum recharge amount for data in the 9th Month(Action Phase) for churned customers.

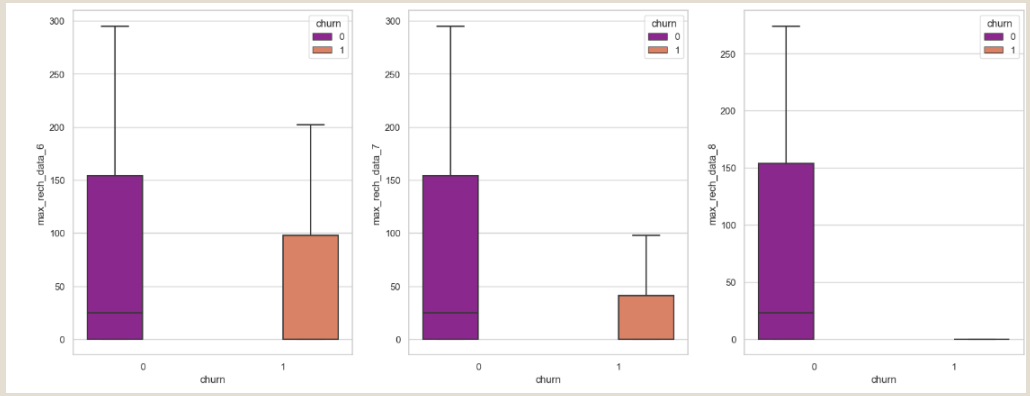


Total Recharge for number

We can see that there is a huge drop in total recharge number also in the 8th Month(Action Phase) for churned customers.

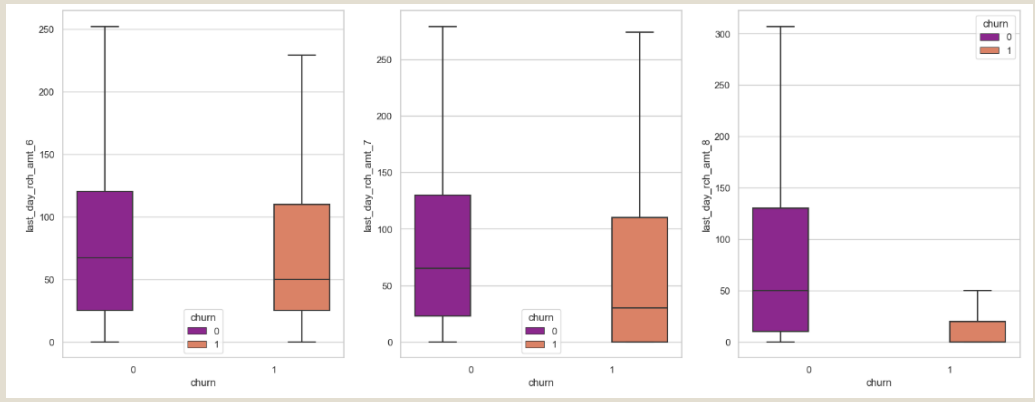
Exploratory Data Analysis

a. Recharge amount related variables



Maximum Recharge for Data

We can observe that there is a huge drop in maximum recharge for data also in the 8th Month(Action Phase) for churned customers.



Last Day recharge Amount

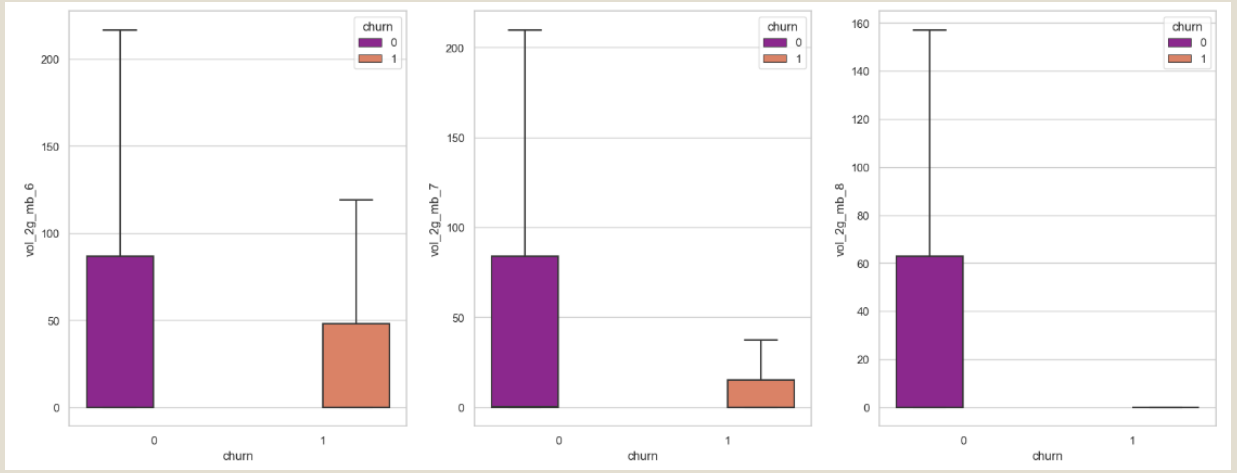
We are getting a huge drop in 8th Month(Action Phase) recharge amount for churned customers..

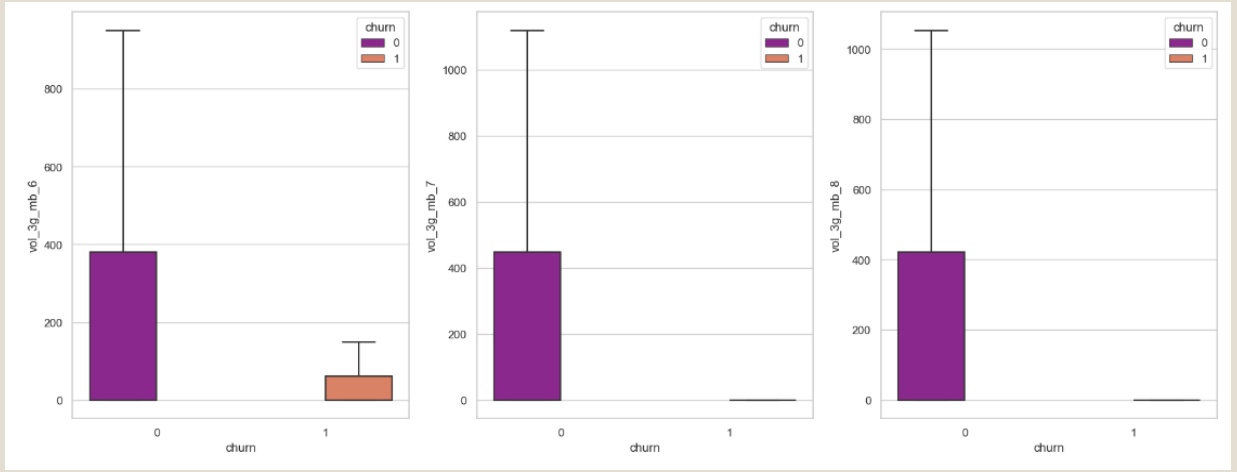
Exploratory Data Analysis

b. 2G and 3G Usage Related Attributes

We have two observations from above:

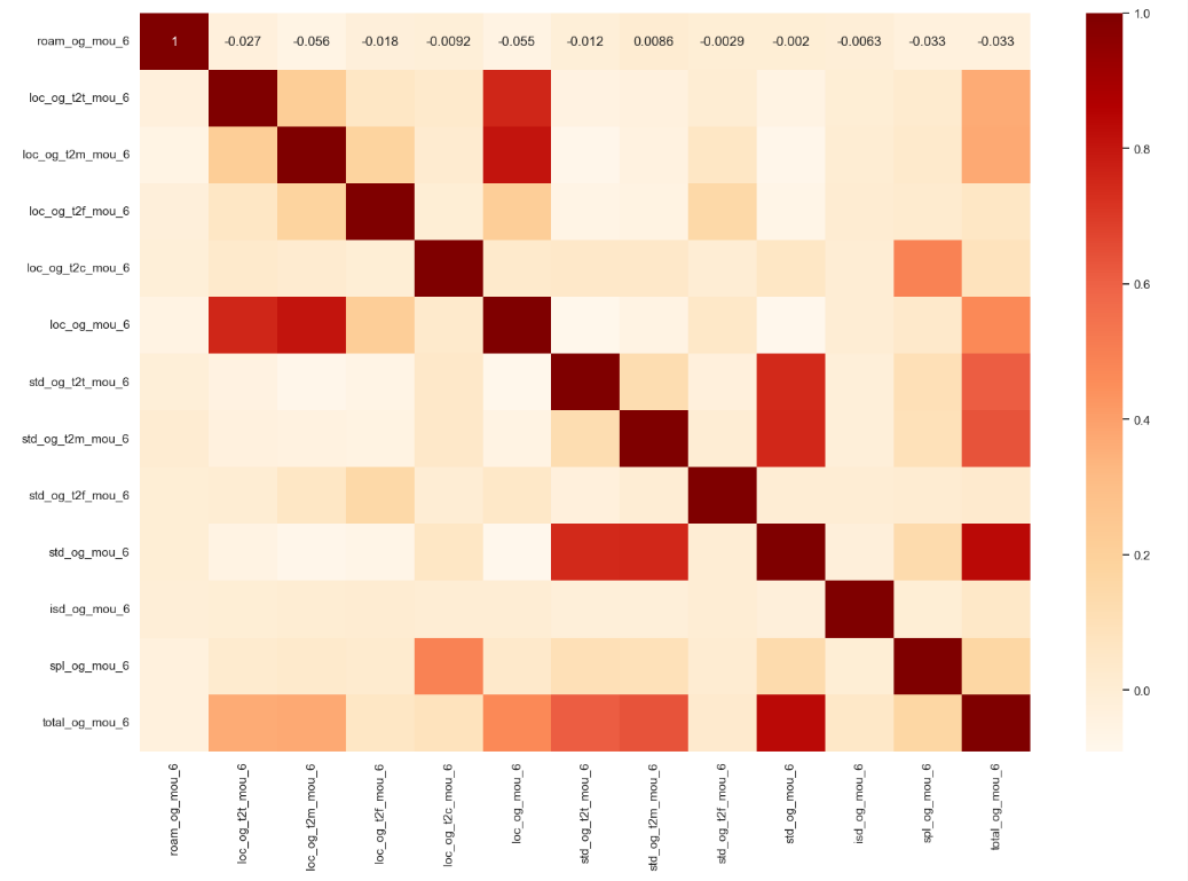
1. 2G and 3G usage for churned customers drops in 8th Month(Action Phase)
2. We can also observe that 2G/3G usage is higher for non-churned customers indicating that churned customers might be from areas where 2G/3G service is not properly available





Exploratory Data Analysis

c. Minutes of usage - voice calls



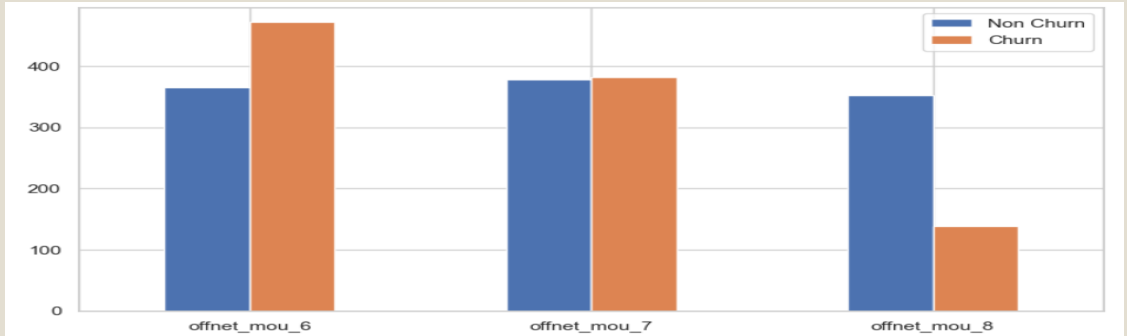
We can see that total_og_mou_6, std_og_mou_6 and loc_og_mou_6 seems to have strong correlation with other fields and they needs to be inspected to avoid any multi collinearity issues.

Exploratory Data Analysis

d. Offnet Usage

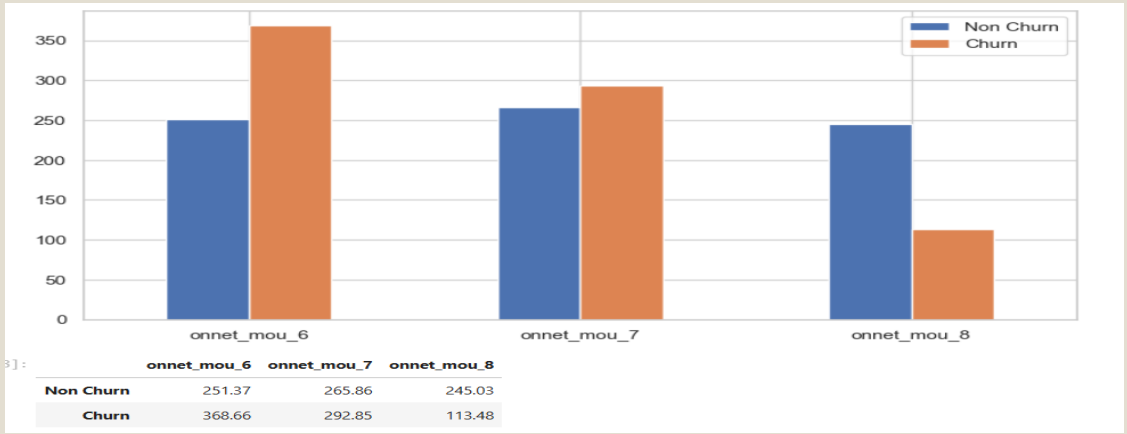
e. ONNET: All kind of calls within the same operator network

We can observe the drop for offnet mou services in the 8th month.



	offnet_mou_6	offnet_mou_7	offnet_mou_8
Non Churn	365.12	377.88	352.50
Churn	471.95	382.28	138.52

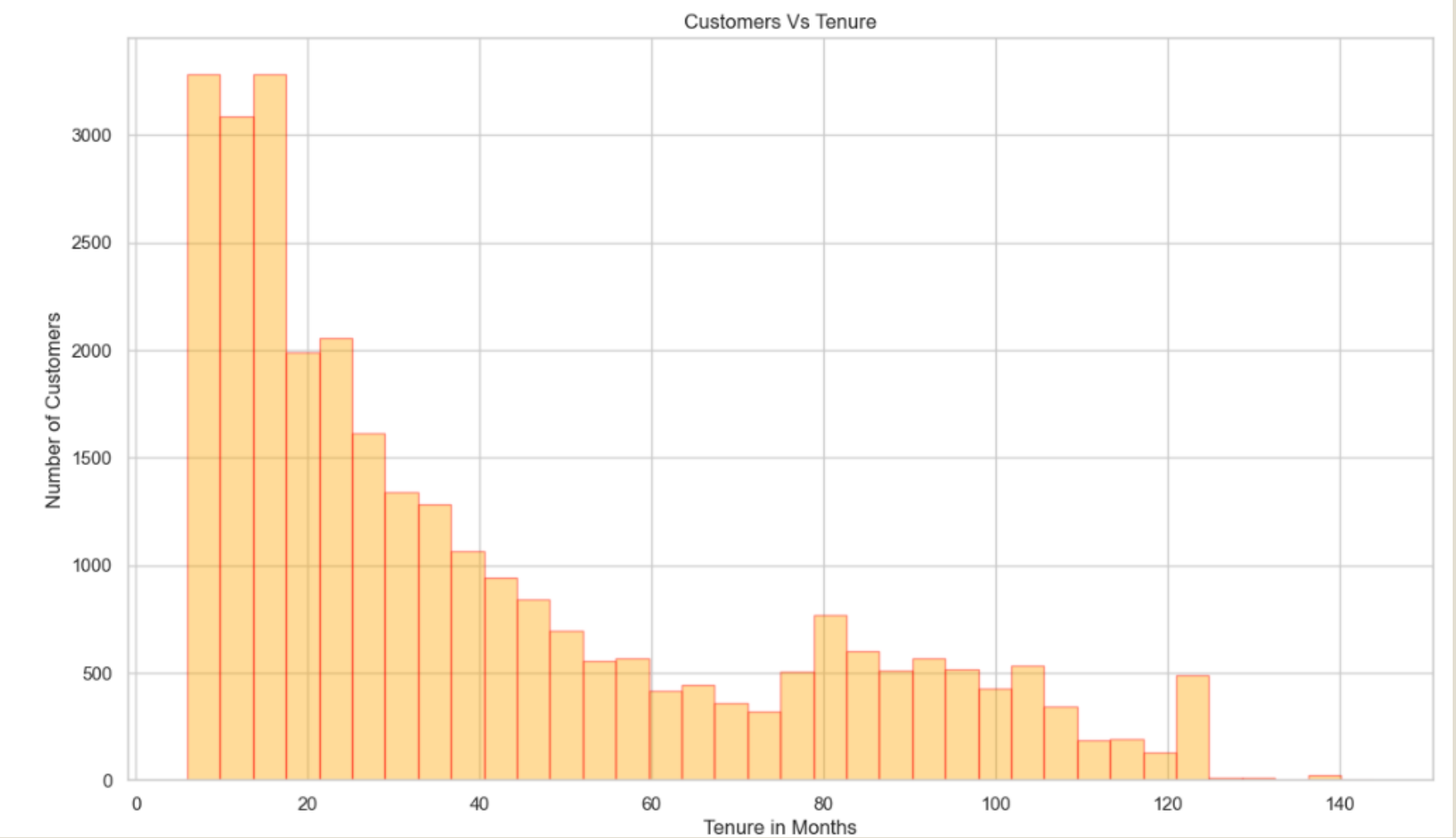
We can also observe here that there is a drop in On-net usage in 8th month for churned customers.



	onnet_mou_6	onnet_mou_7	onnet_mou_8
Non Churn	251.37	265.86	245.03
Churn	368.66	292.85	113.48

Exploratory Data Analysis

f. Tenure Analysis for Customers

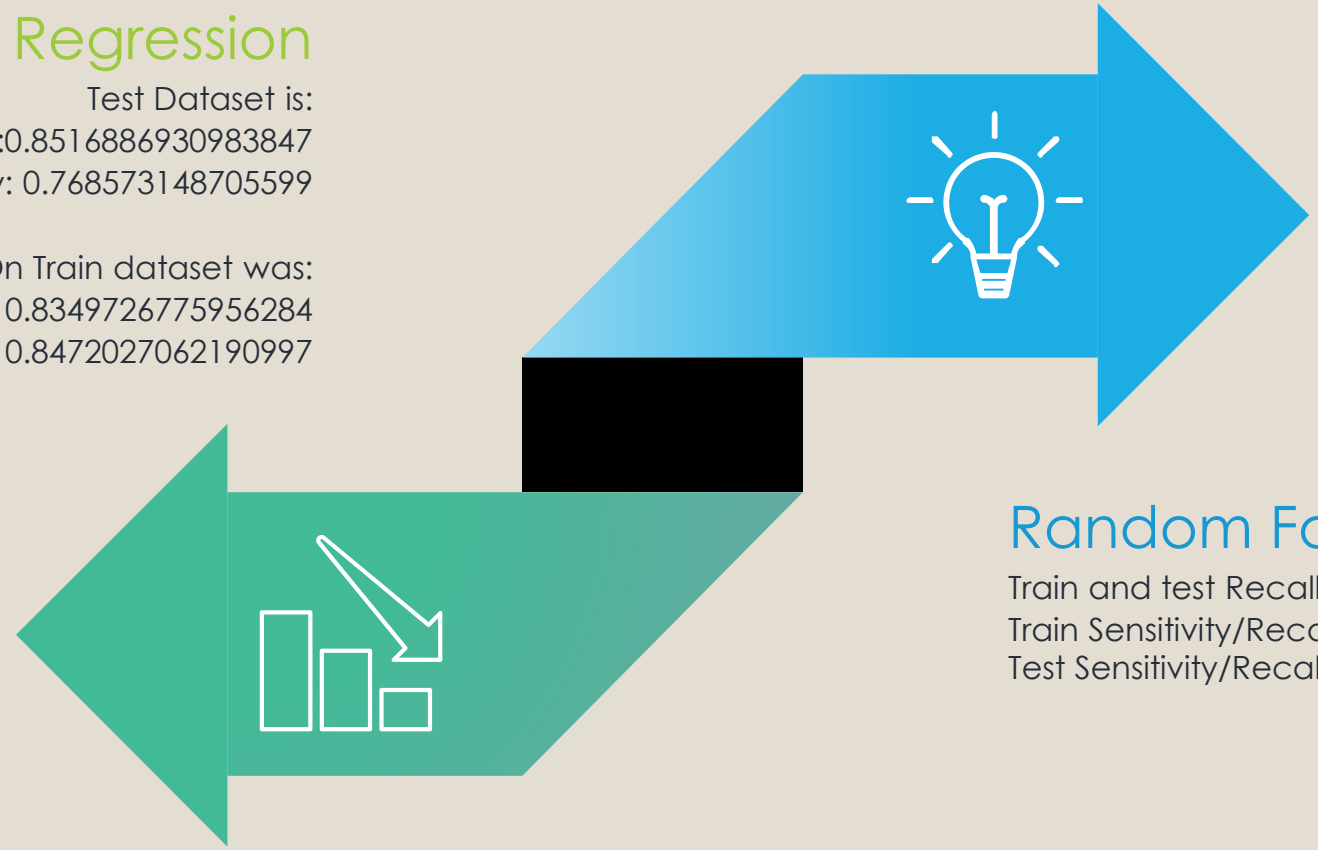


MODELLING

Logistic Regression

Test Dataset is:
Sensitivity/Recall :0.8516886930983847
Specificity: 0.768573148705599

Stats On Train dataset was:
Sensitivity/Recall : 0.8349726775956284
Specificity: 0.8472027062190997



Random Forest

Train and test Recall score of Random forest:
Train Sensitivity/Recall: 0.8918032786885246
Test Sensitivity/Recall: 0.7870778267254038

OBSERVATIONS

Recall/Sensitivity score need to be considered in this case study as Telecom company will not like any high value customer to churn so will try to find out all high value customers who may leave in future. Company may willing to even bear burden of few customers who may not churn but will be classified as churn.so overall recall score will be considered for good model. Logistics Regression will be chosen in this case.

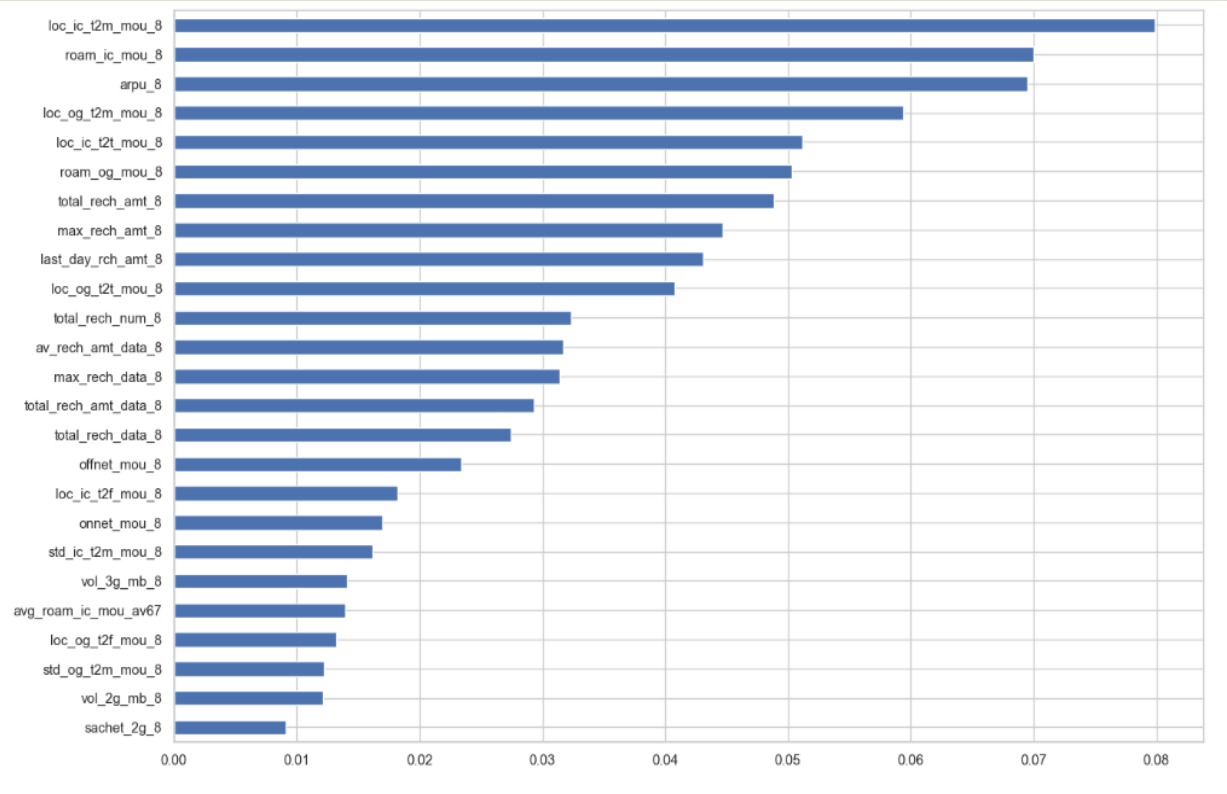
As we can see that Train Recall score of Random forest is better than Logistics Regression but Test Recall score of Logistics Regression is better as compared to Random forest. Difference in between Train and test score is less for Logistics Regression which seems model is stable.

Build another model with the main objective of identifying important predictor attributes which help the business understand indicators of churn.

A good choice to identify important variables is a logistic regression model or a model from the tree family.

In case of logistic regression, make sure to handle multi-collinearity.

Recommend strategies to manage customer churn - Random Forest



We could see from plot that almost all the features of top 25 most important features are from action phase that is August month. Hence, we need to focus on these features to identify high value customers who may churn in future.

Top 10 most important features are as follows:
1.loc_ic_t2m_mou_8: Local incoming call operator T to other operator mobile minutes of usage - voice calls in 8th month.
2.total_rech_amt_8: Total recharge amount in 8th month.
3.max_rech_amt_8: Max recharge amount in 8th month.
4.roam_ic_mou_8: Roaming incoming calls minutes of usage - voice calls in 8th month.
5.loc_ic_t2t_mou_8: Local incoming calls within same operator mobile minutes of usage - voice calls in 8th month.
6.arpu_8: Average revenue per usage in month 8.
7.total_rech_data_8: Total data recharge in month 8.
8.roam_og_mou_8: Roaming outgoing calls minutes of usage - voice calls in 8th month.
9.loc_og_t2t_mou_8: Local outgoing calls within same operator mobile minutes of usage - voice calls in 8th month.
10.total_rech_num_8: Total number of recharges done in the month 8.

Recommended Strategies and Suggestions

- Number of Incoming and outgoing calls from a mobile number in particular month by customer. If number of calls starts reducing then it may be sign of customer trying to switch from one netwrok to another network or he is has already switched and using current netwrok for few days.
- Recharge amount is very important factor to notice if it starts reducing month by month then it need to be looked as customer may not be happy with the services he is getting that is why he started recharging with less amount.
- If data usage starts decreasing and in august month it is minimal then it shows customer is not getting good speed of internet.
- If internet speed that customer is getting is good, customer will finish data soon and recharge it again but if network is poor and speed is not good then customer will not be able to finish it and will not recharge it multiple times. So need to look into areas where network is poor and customer care is receiving complaints multiple times.
- If all kinds of call and data usage reduces then it is serious concern as customer may be planning to churn and just timepassing for few more days. So, company need to look into these areas.
- If customer is using the services for incoming calls only and has stopped using outgoing calls then he is finding the services very costly and may switch to network where incoming and outgoing services are in reasonable rate.

THANKS!