

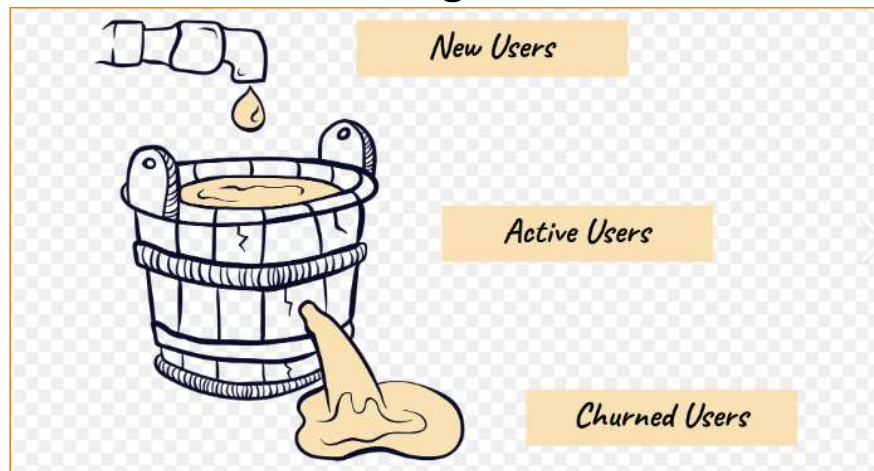
Telecom churn Case Study

- IIITB & UPGRAD (VIVEK, AFZAL AND ANGAD)

CONTENTS:-

1. Problem statement.
2. Objective
3. Source Data provided
4. Exploratory Data Analysis (EDA)
5. Modelling
6. Recommendation / findings

- We found a good analogy online.
 - Here if we observe the input of New Users is very low. Represented by drop of water.
 - There is a pool of Active users which needs to be safeguarded in the vessel or barrel.
 - There is a heavy flow of users outside from the leakage. This is nothing but churned users which is high than the input of New Users. Here leakage is nothing but the significant parameters which if not handled effectively will result in leakage of users and moving to better network.



Problem Statement :

- As we observed in previous slide, customer churn is a big challenge in telecom industry. Retaining an existing customer is more preferable and profitable compared to adding a new customer to your network.
- To predict the churn in the ninth month with help of the data provided for first three months.
- By detecting the customers who are likely to move away from current network (based on features significance) and align with new network is primary goal of this problem.
- Here we will need to identify the driving features which will help address the problem statement and ensure that the attrition from current network is least. For this we will need the past and current data related to plans, churned and not churned customers, their demography and usage information.

Objective :

Our objective here would be help address the Problem statement which is to reduce the churn.

Propose a solution where the current provided attributes will help to maximize the customer satisfaction and minimize the churn. Utilizing the given data and applying the data analytics, prediction formulae's and forecasting the driving features we will build a model which will help control the churn and help this networks business grow.

Here our predictor variable is Churn Yes /No. Hence this is classification model and as a Business Analyst role we will design a model accordingly which will help predict the features which will play vital role is reducing the customer churn and subsequently enhance the profits through customer retentions.

Data details :

- CSV file of 99999 rows and 226 columns / features is available.
- Data dictionary to help understand the relevance of features is available.
- The data contains details for 4 months June, July, August and September precisely. Here first two months are good phase when customer is happy with network and its services, third month is action phase where the unhappiness with services start basically due to offers from competitors. This is the month when customer starts showing signs that he will probably think of alternative network. Here corrective actions can be applied based on our findings and provide offers / improve services to retain.
- Fourth month is the churn month, when the customer has already churned. So for first three months we did not had the churn data, but as a residual of first 3 months the customer churned which we only identify in fourth month.

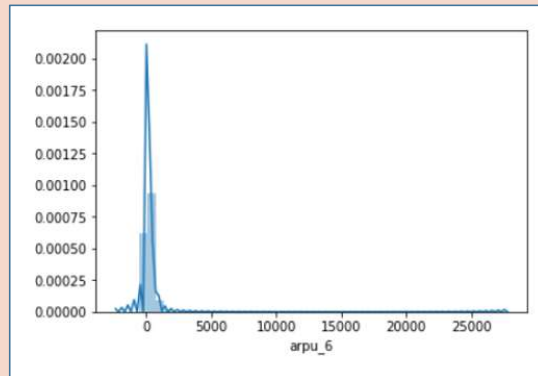
EDA :

- In EDA routine NULL Value imputation and ignoring the not required columns is performed.
- Apart from this another important step is to restrict the analysis to subset of customers who impact the business. In other words keep the High valued customers in scope and filter out the rest data. Here we keep the data where the customer recharge is more or equal to 70th percentile of average recharges. Post applying this we received 29953 rows inscope. This will be our data for further analysis.
- Replace the churned customers with 1 and others with 0 based on fourth month details.
 - Customers who have not made any outgoing calls or received any incoming calls. (total_ic_mou_9, total_og_mou_9)
 - Customers who have not used internet at all. (vol_2g_mb_9, vol_3g_mb_9)
 - Also remove other attributes for 9 month.
- Handle class imbalance as dependent variable (churn data) is 6% which will not help to make unbiased decisions. Hence handling class imbalance will be one step.

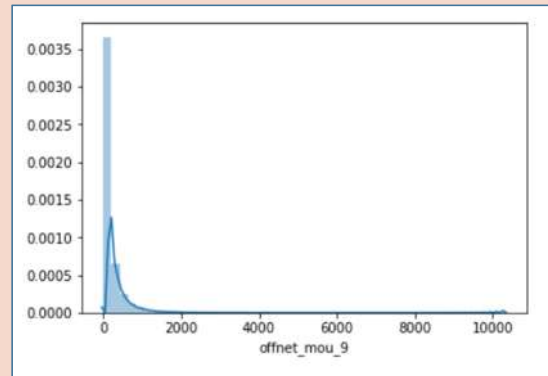
Modeling : Univariate analysis

Plot distribution plot to represent the univariate distribution of data. (Some examples are below)

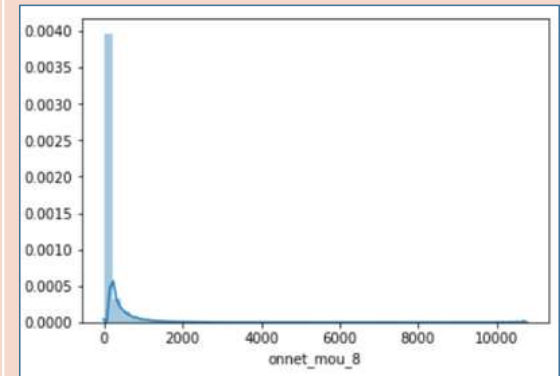
Average Revenue per user for 6th Month



Calls outside the network for 9th Month



Same network calls for 8th Month.

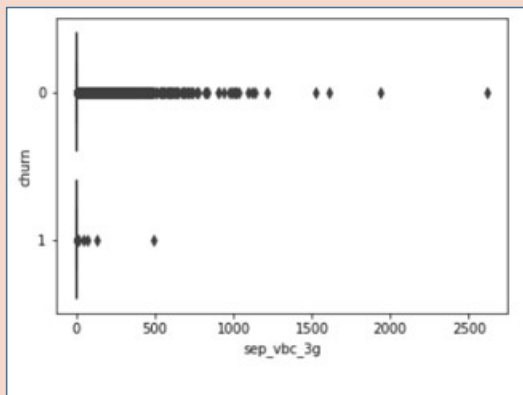


Bivariate analysis

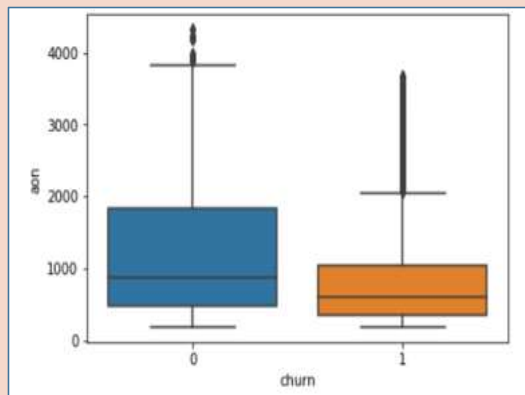
Plot Box plot to represent the bivariate distribution of data aligned with the dependent variable (Churned data). (Some examples are below)

Shows that customer who have spent more time with network are less churning out.

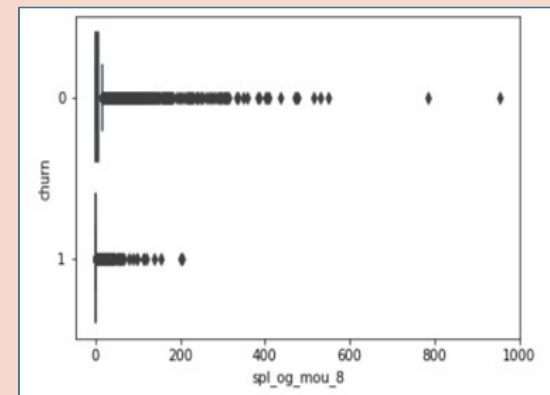
Volume based cost (Sep)



Age on current network

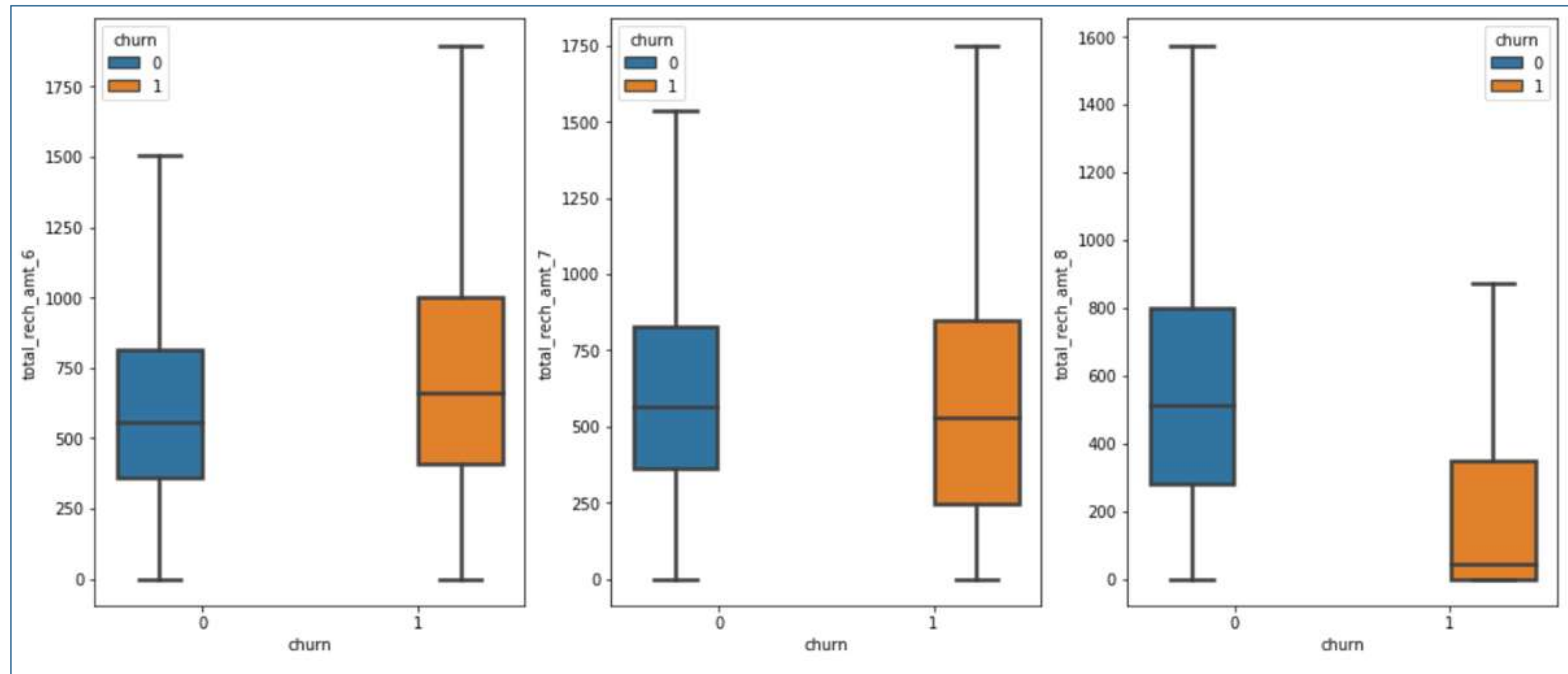


Special Outgoing calls for 8th Month

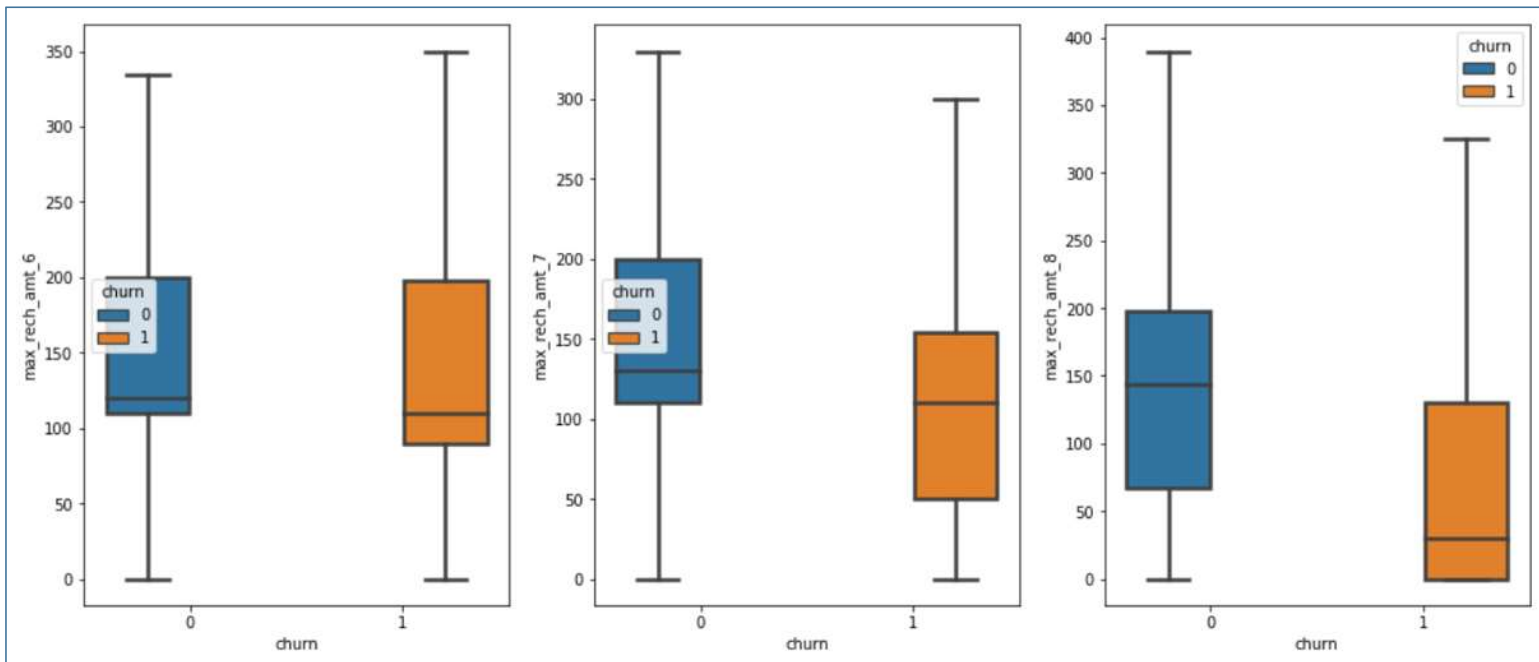


Shows trend of Total Recharge amount and how it reduced month over month.

The downfall of recharge amount is evident that these are churning customers.



Similar trend will be observed for maximum recharge for Jun, Jul and August



Feature details :

Random Forest model showed over fitting hence going ahead with Logistic regression

Sensitivity at 80%

Specificity at 84%

ROC at 88% looks to be a better model and fetching the driving features based on these parameters.

Findings :

1. Based on our analysis it is clear that total incoming calls, considerable drop in recharge amounts, data usage, roaming incoming and outgoing calls in the 8th month are the factors affecting the churn.

Those specific features are :

total_ic_mou_8, total_rech_amt_diff, total_og_mou_8, arpu, roam_ic_mou_8, roam_og_mou_8, std_ic_mou_8, std_og_mou_8, av_rech_amt_data_8.

2. Recharge drop in 7th Month could have been the trigger point as the recharge fell down where the corrective actions could be taken predicting these are probably customers planning to go away from network.

3. Usage considerations are very crucial as Local and Roaming Minutes of usage are factors that can be seen as important feature as the trend falls and customer churn in following month.

Recommendations :

- ✓ Give special promotional offers, discounts to customers based on their usage.
 - ✓ If roaming or out going calls are more, provide calling discounts.
 - ✓ If customer uses internet services provide better data plans. (2G where 3G is not available or expand 3G network.
- ✓ Offer lowering of existing tariff plans to make the efforts to retain customers.
- ✓ Even for small recharges, calls can be made to customers to understand their experiences, grievances and provide attractive offers.
- ✓ If the grievance is related to poor network coverage, provide resolution timelines and raise a technical issue with the team internally, once resolved perform a follow-up to confirm customer satisfaction.

From our group perspective 7th Month will be crucial where drop in call time, drop in recharge spend can be analyzed in more depth and actions can be taken to provide a better plan, pricing and experience to customers in 8th Month with this we can see low churn in 9th Month.

NEVER STOP LEARNING

Similarly the incremental data should be regressively utilized to learn ahead as rightly said learning never stops and hence Machine learning.

