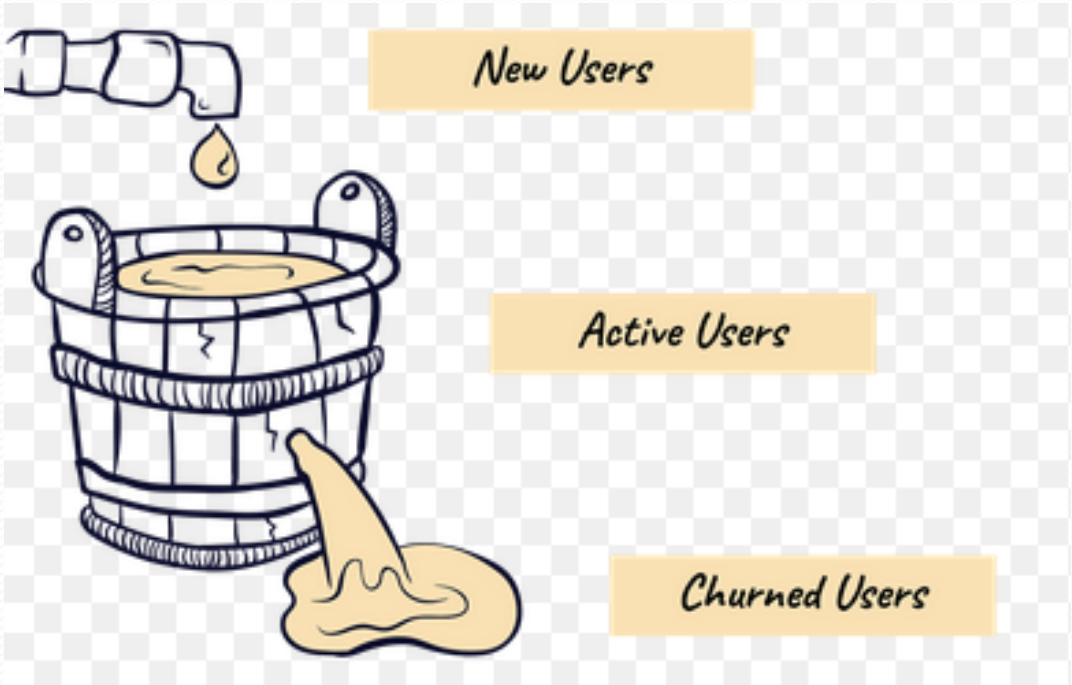
Telecom Churn Prediction Case Study

Business Problem

Telecommunications industry experiences an average of 15-25% annual churn rate.

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



Business Goal

Retaining high profitable customers.

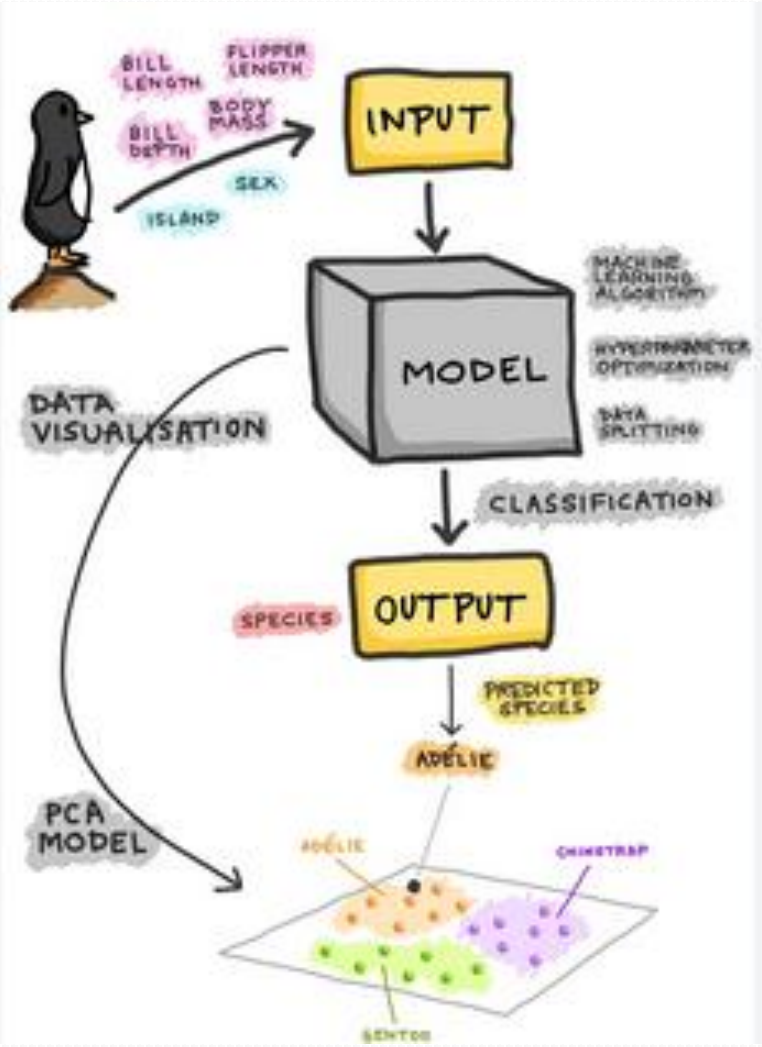
Predict which customers are at high risk of churn.

Steps for analysis

1. Importing data &Data preparation
2. Handling missing values in column
3. Delete unwanted columns which are not required for our analysis
4. Handling missing values in row
5. Checking records for MOU (Sept, Aug, July, June)
6. Tagging Churners
7. Delete all the attributes corresponding to churn phase
8. Checking churn %
9. Outlier treatment
10. Deriving new features
11. EDA (Exploratory data analysis) -
12. a) Univariate Analysis
13. b) Bi-variate Analysis
14. Data split into train & test
15. Data imbalance treatment
16. Feature scaling

Model Building

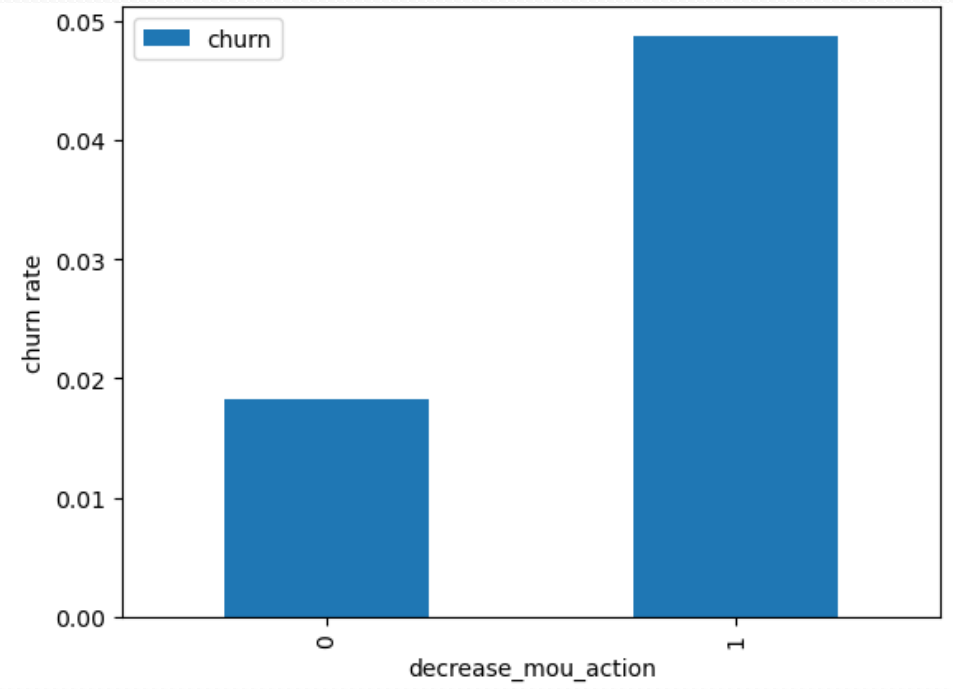
1) Model building
2) Model with PCA (Principle component Analysis)
3) Support Vector Machine (SVM) with PCA
4) Hyper parameter tuning
5) Build Logistic regression model with PCA
6) Build Decision tree model with PCA
7) Random forest model with PCA
8) Model without PCA
9) Feature selection using RFE & Checking VIFs
10) Model performance on train set
11) Metrics
12) Plotting ROC curve
13) Testing the models on test set



EDA - **Univariate analysis**

- **Decrease MOU in Action phase**

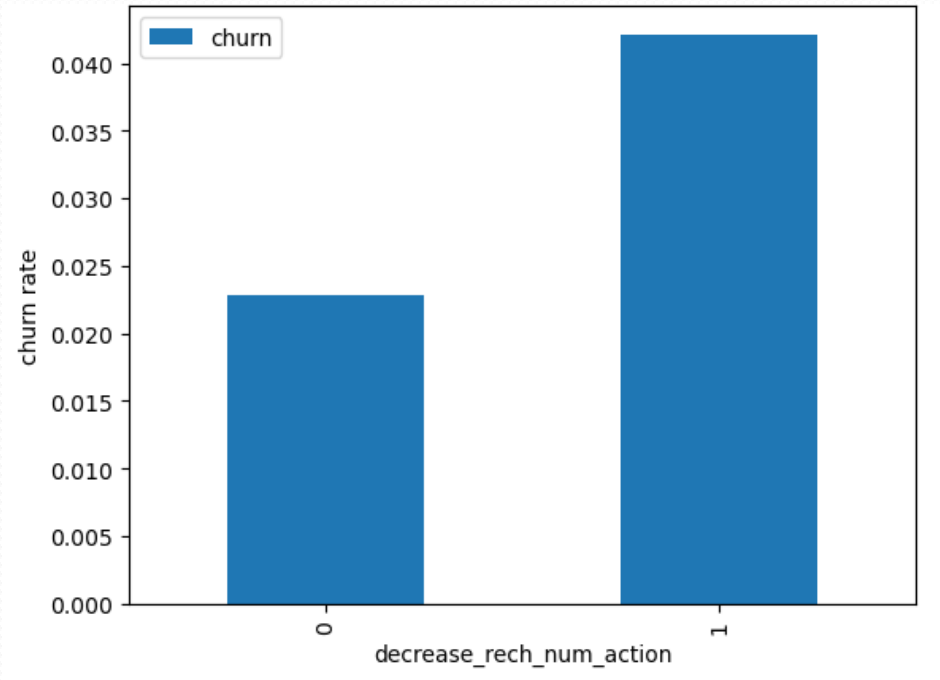
- **Observation –**
- It can be noticed that the churn ra is more for the customers, whose minutes of usage(mou) decreased the action phase than the good phase.



EDA

- **Decrease Recharge Number in Action Month.**

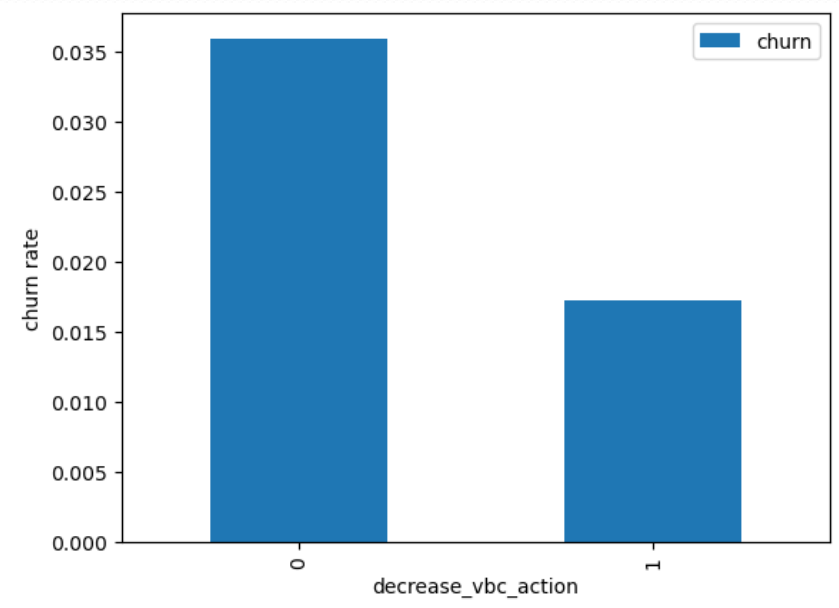
- **Observation –**
- Seems churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.



EDA

- **Decreased volume based cost in action phase**

- **Observation –**
- The churn rate is more for the customers, whose volume based cost in action month is increased. That means the customers do not do the monthly recharge more when they are in the action phase.



EDA

Action Phase ARPU

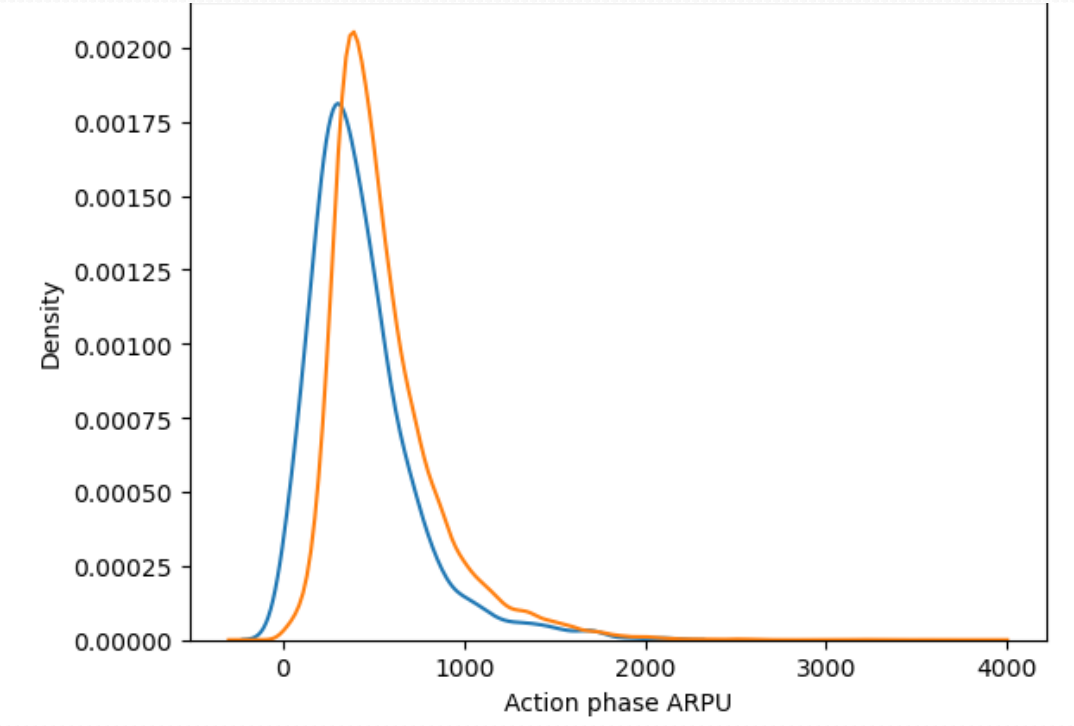
Observation –

Average revenue per user (ARPU) for the churned customers is mostly dense on the

0 to 900. The higher ARPU customers are less likely to be churned.

ARPU for the not churned customers is mostly dense on the

0 to 1000.

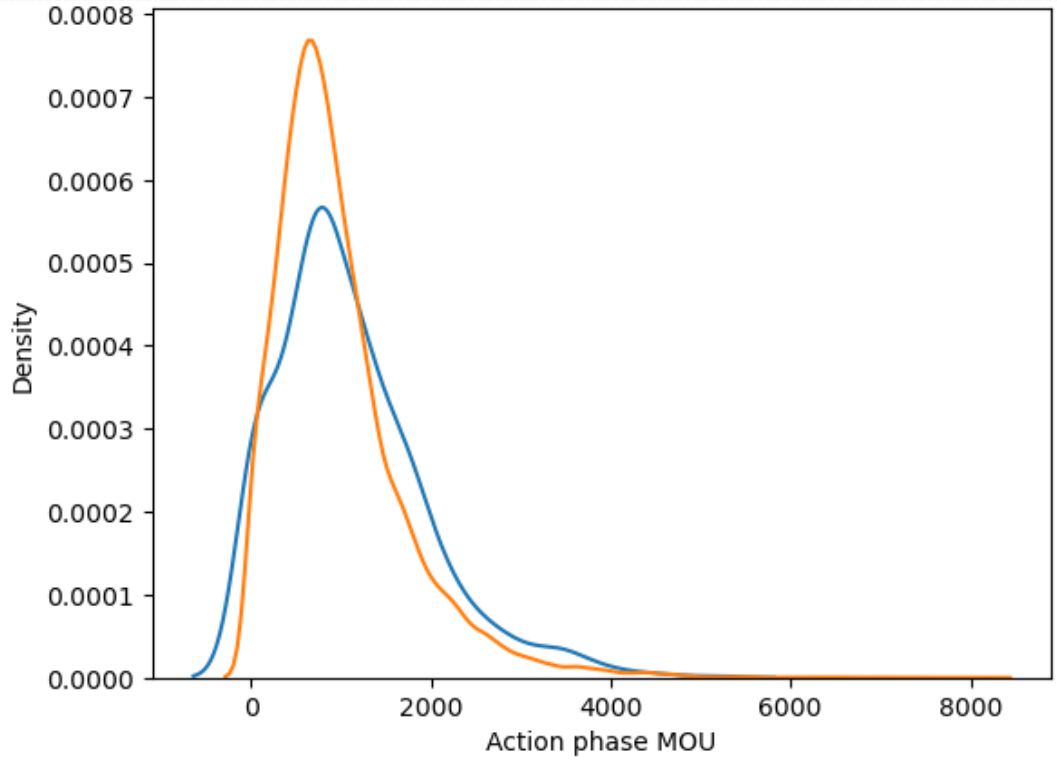


EDA

Action Phase MOU

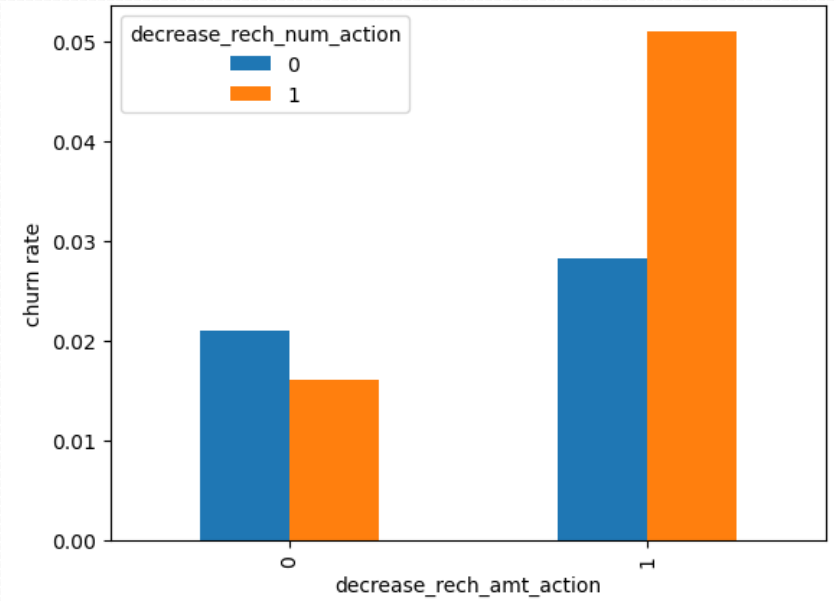
Observation –

Minutes of usage(MOU) of the churn customers is mostly populated on the 0 to 2500 range. Higher the MOU, lesser the churn probability.



EDA – **Bi-variate analysis**

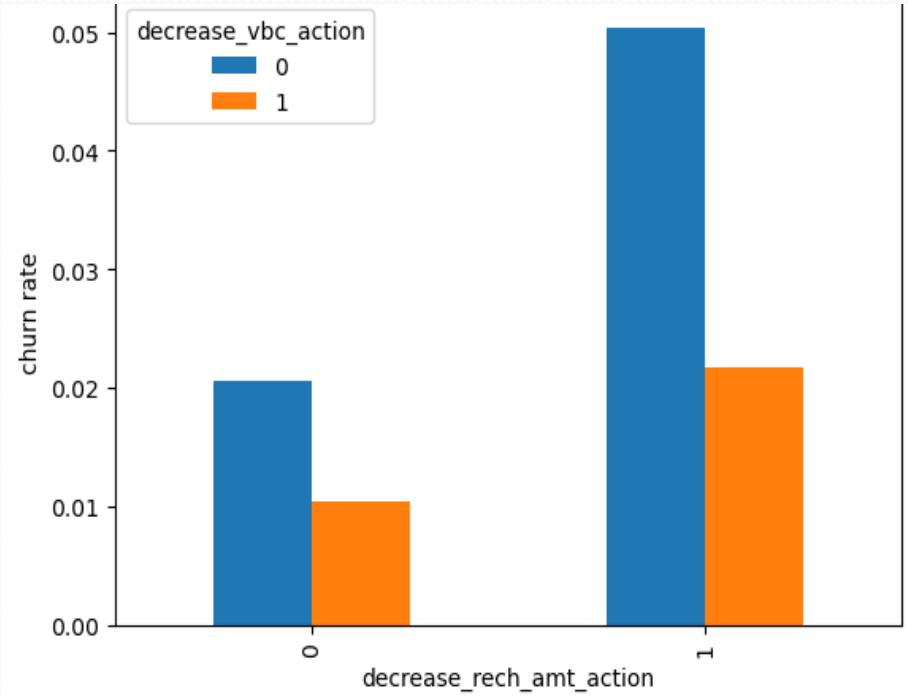
- **Decreasing recharge amount and number of recharge in the action phase.**
- **Observation –**
- The churn rate is more for the customers, whose recharge amount as well as number of recharge have decreased in the action phase than the good phase.



EDA

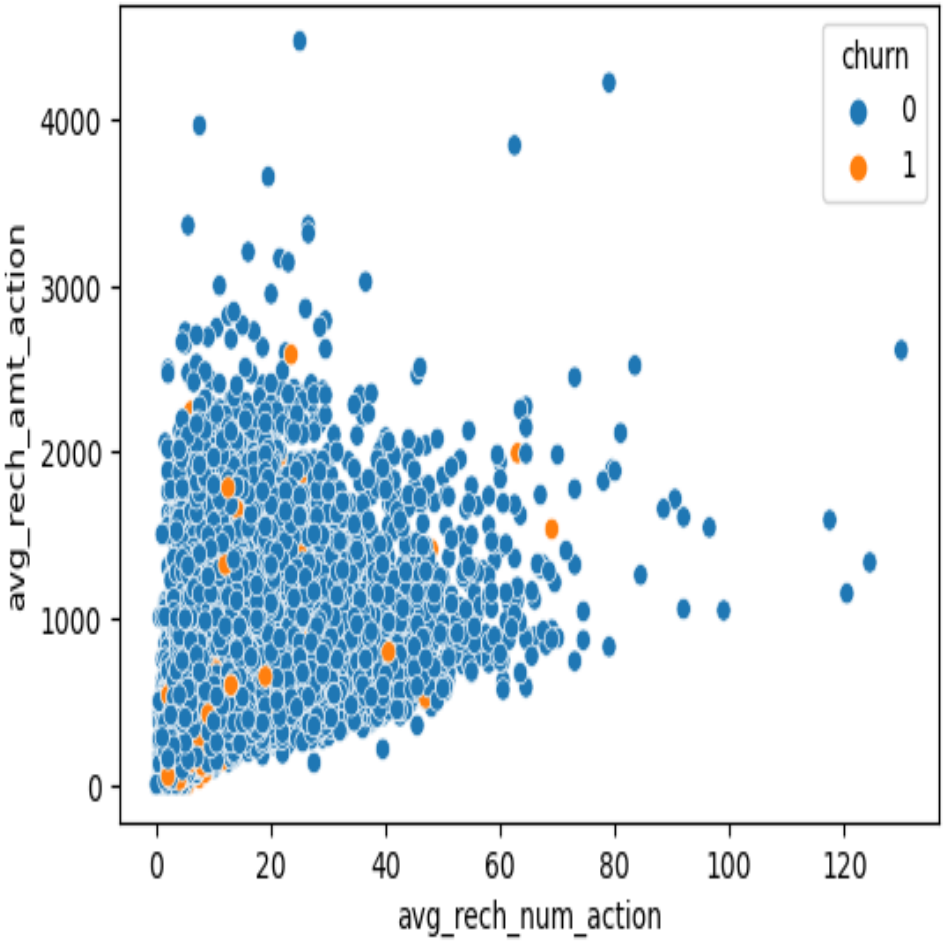
- **Decreasing recharge amount and volume based cost in the action phase**

- **Observation –**
- It can be seen that the churn rate is more for the customers, whose recharge amount is decreased along with the volume based cost is increased in the action month.



EDA

- **Recharge amount and number of recharge in action month.**
- **Observation –**
- The recharge number and the recharge amount are mostly proportional. More the number of recharge, more the amount of the recharge.



Outlier treatment & Train – Test split

Outlier treatment

- **Removing outliers below 10th and above 90th percentile**

Train – Test split

- **Feature variable into X**
- **Target variable into Y**
- **Splitted data into train and test set 80:20**

Derive new features

total_og_mou_6',
total_og_mou_7',
total_og_mou_8',
total_ic_mou_6',
total_ic_mou_7',
total_ic_mou_8',
total_rech_num_6',
total_rech_num_7',
total_rech_num_8',
total_rech_amt_6',
total_rech_amt_7',
total_rech_amt_8

Business recommendation - Top predictors

Top variables selected in the logistic regression model.

The top variables have negative coefficients. That means, the variables are inversely correlated with the churn probability

##### Variables	Coefficients
##### loc_ic_mou8	-3.3287
##### og_others_7	-2.4711
##### ic_others_8	-1.5131
##### isd_og_mou_8	-1.3811
##### decrease_vbc_action	-1.3293
##### monthly_3g_8	-1.0943
##### std_ic_t2f_mou_8	-0.9503
##### monthly_2g_8	-0.9279
##### loc_ic_t2f_mou_8	-0.7102
##### roam_og_mou_8	0.7135

Conclusion :-

- **During the action phase (as seen above in the month of August) keep are track of customers who's ISD, MOU, local calls are turned to be lesser.**

- **3G recharge if higher during any month those customers seems to churn.**
- **Customers decreasing monthly 2g usage for August are most probable to churn.**

- **Also, the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.**

THANKYOU

THE END