



# Telecom Churn Prediction Case Study

Submitted By:

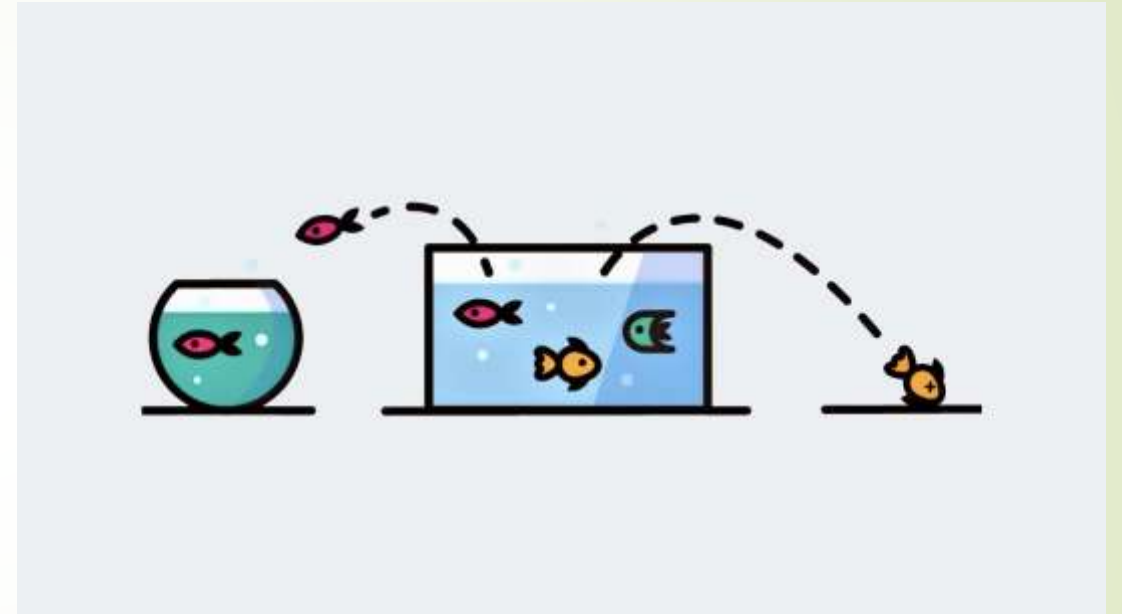
Vikas Ware

Vinayak Jadhav

Vedant Patel

## 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 of 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 percentage
9. Outlier treatment
10. Deriving new features
11. EDA (Exploratory data analysis): -
  1. Univariate Analysis
  2. Bi-variate Analysis
12. Data split into train & test
13. Data imbalance treatment
14. Feature scaling

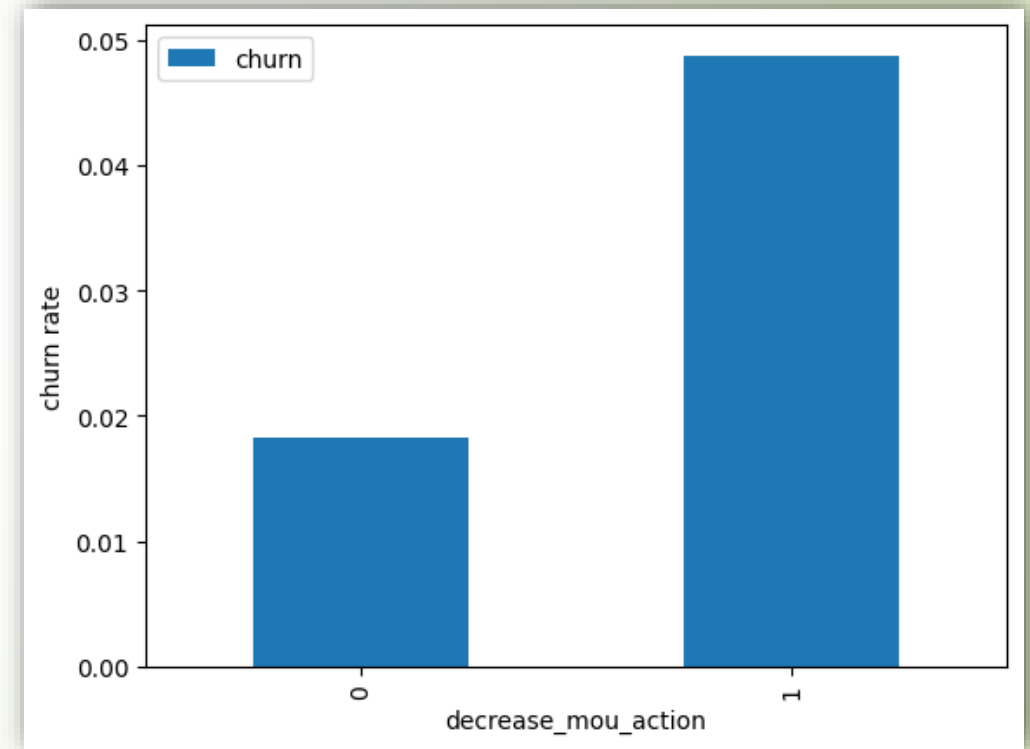


# Model Building

1. Model building
2. Model with PCA (Principle Component Analysis)
3. Support Vector Machine (SVM) with PCA
4. Hyperparameter Tuning
5. Logistic regression model with PCA
6. 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

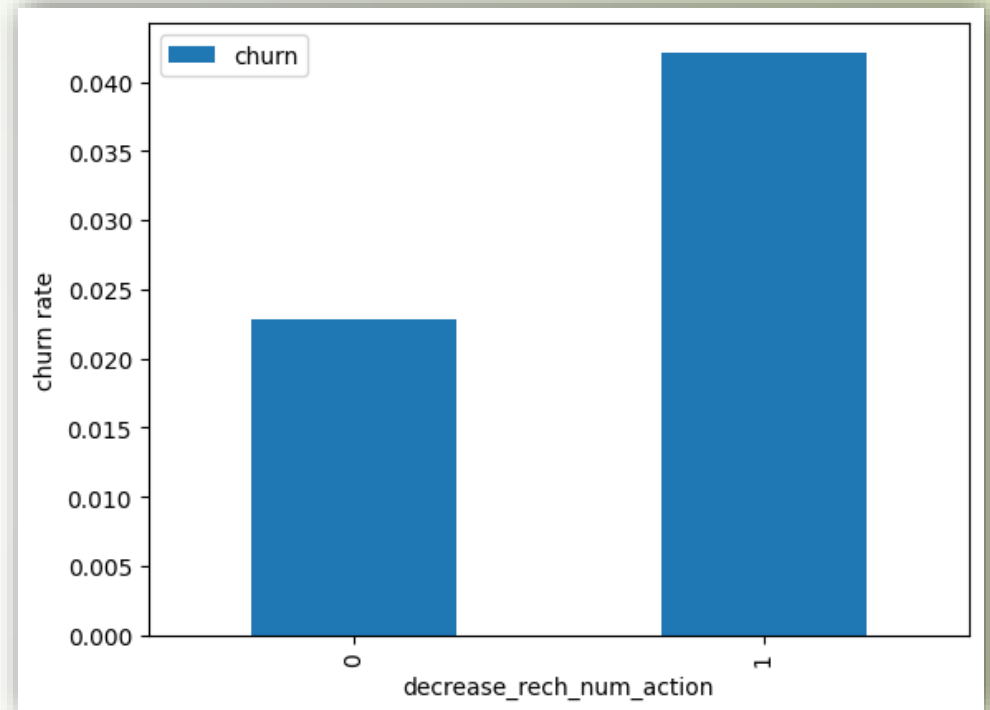
- **Decrease MOU in Action phase**
- **Observation:-**
  - It can be noticed that the churn rate is more for the customers, whose minutes of usage(MOU) decreased in the action phase than the good phase.





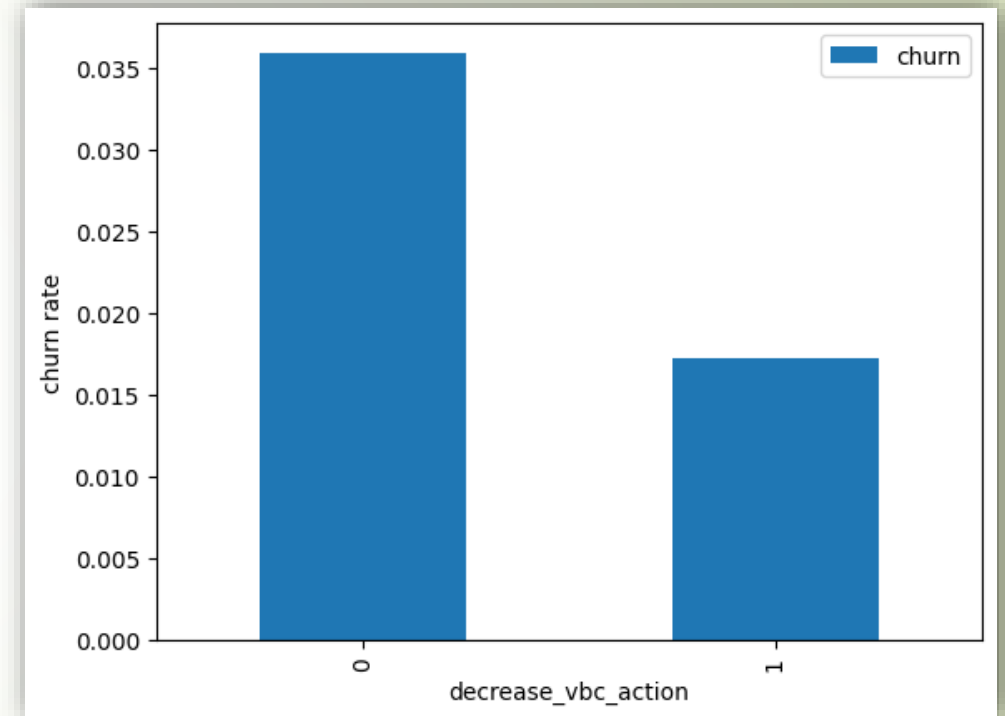
# EDA

- **Decrease Recharge Number in Action Month.**
- **Observation:**
  - Seems the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.



# EDA

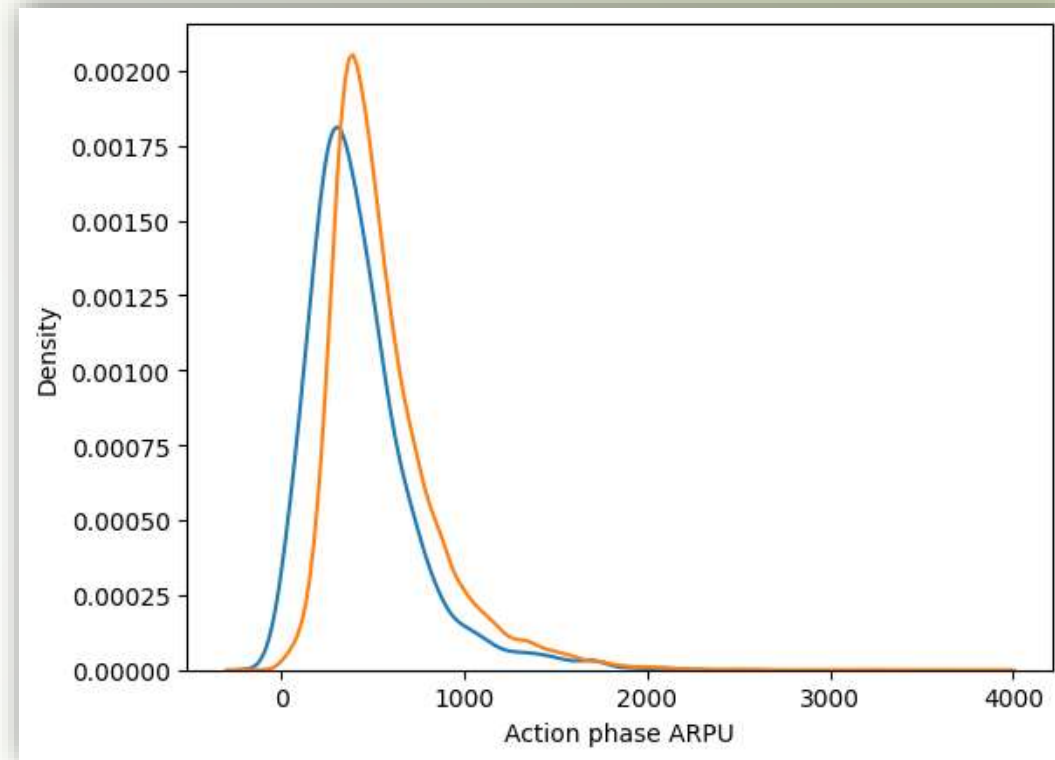
- **Decreased volume based cost in action phase**
- **Observation:**
  - The churn rate is more for the customers, whose volume based cost in action month is increased. That means the customers do not do the monthly recharge more when they are in the action phase.





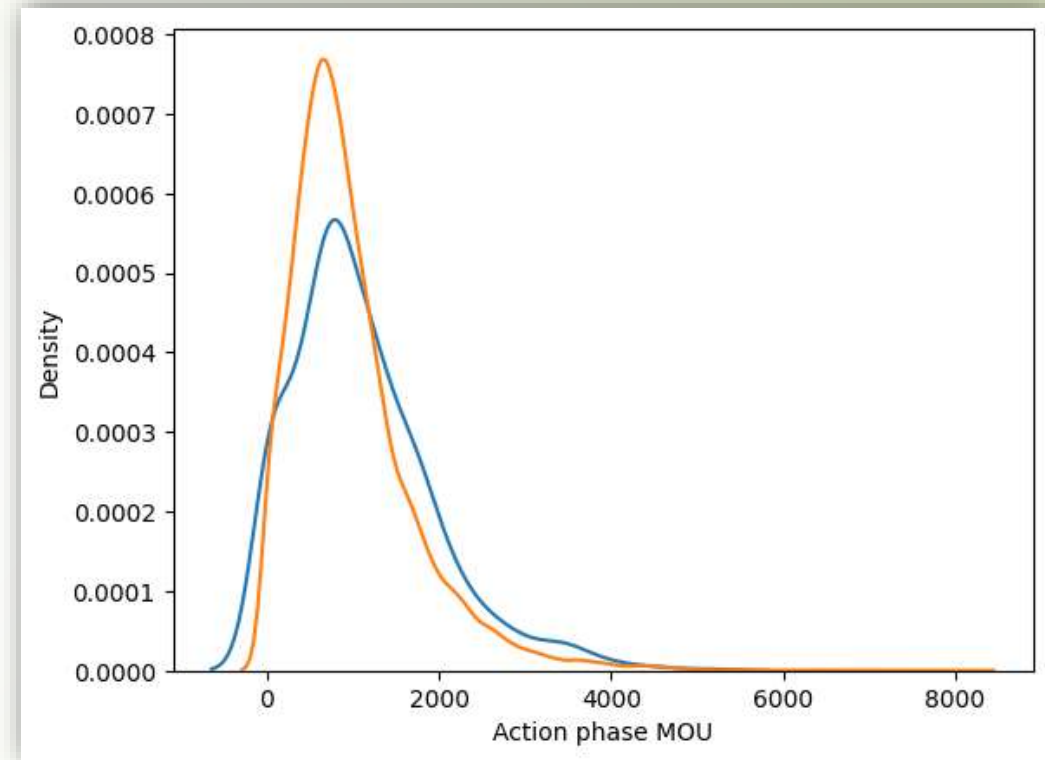
# EDA

- **Action phase ARPU**
- **Observation:**
  - Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned.
  - ARPU for the not churned customers is mostly densed on the 0 to 1000.



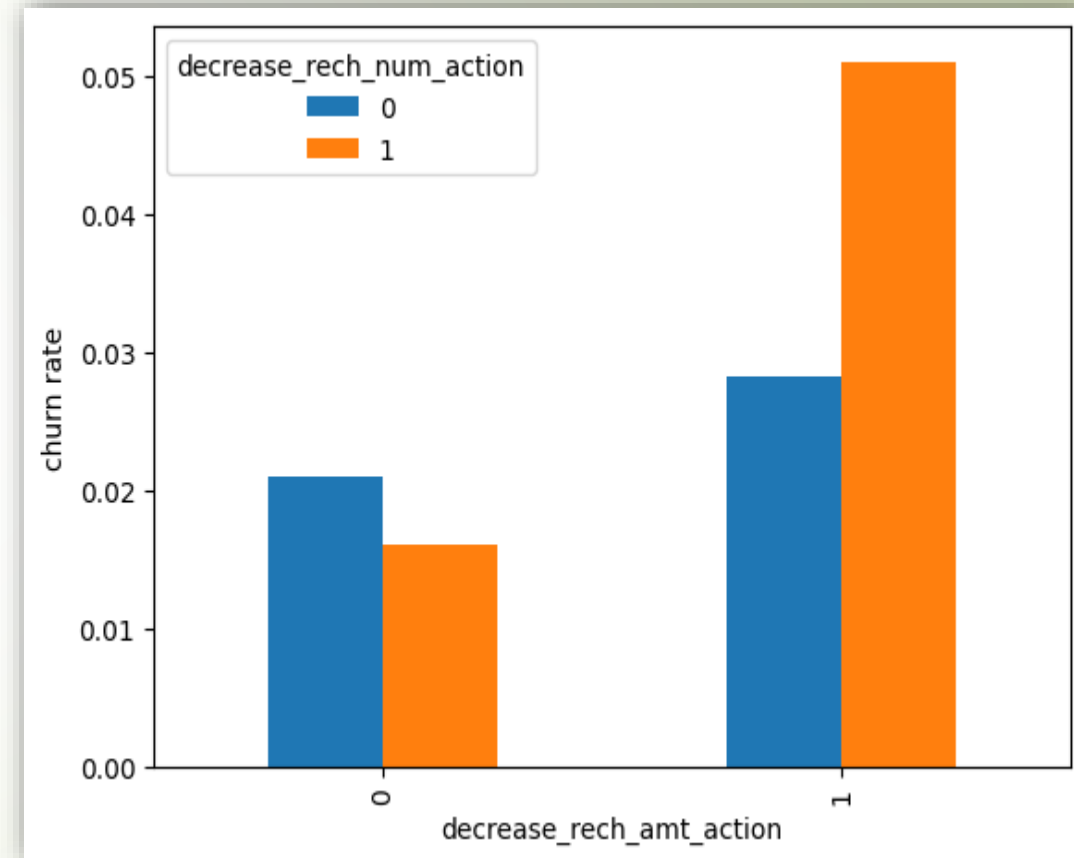
# EDA

- **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.



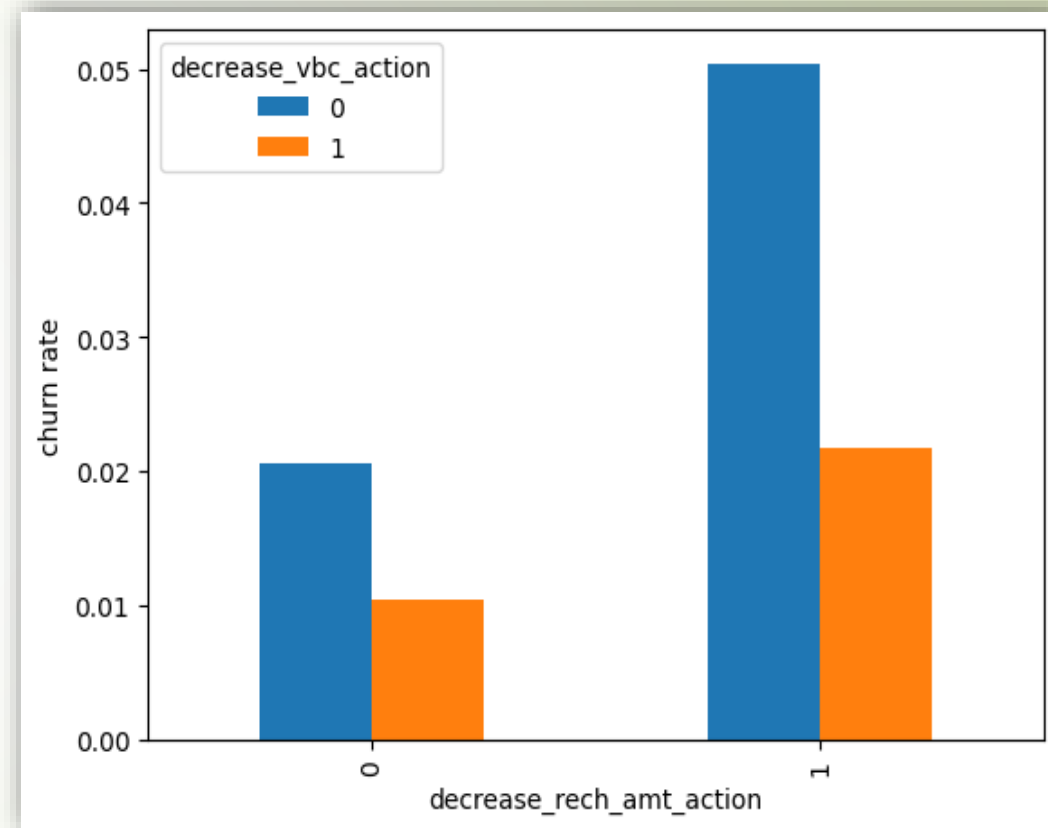
# EDA- Bi-variate Analysis

- 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.



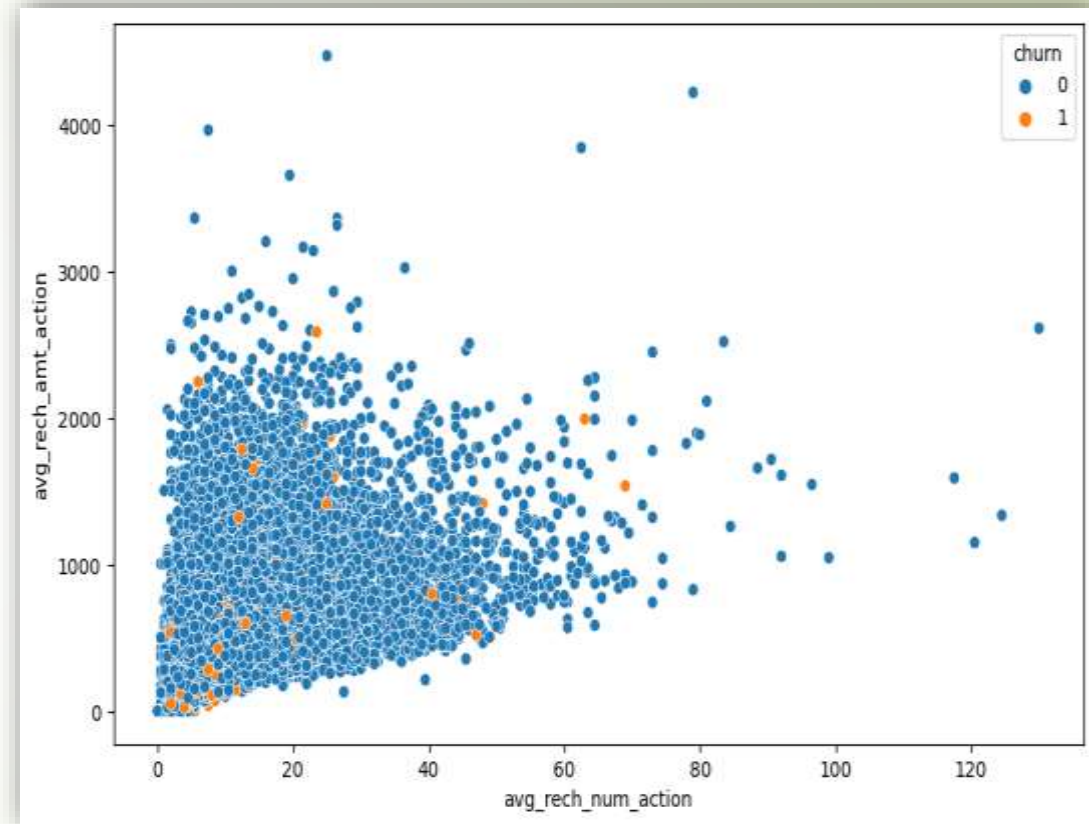
# EDA

- **Decreasing recharge amount and volume based cost in the action phase**
- **Observation:**
  - 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
- Spitted data into train and test set 80:20



# Derive new feature

- 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_mou_8	-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 (*Recommendations*):-

1. Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
2. Target the customers, whose outgoing others charge in July and incoming others on August are less.
3. Also, the customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be a good target to provide offer.
4. Customers, whose monthly 3G recharge in August is more, are likely to be churned.
5. Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
6. Customers decreasing monthly 2g usage for August are most probable to churn.
7. Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
8. roam\_og\_mou\_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.



# THANK YOU