TELECOM CHURN CASE STUDY

INTRODUCTION

- 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.**
- 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.

PROJECT OBJECTIVE

- To predict Customer Churn.
- Highlighting the main variables/factors influencing Customer Churn.
- Use various ML algorithms to build prediction models, evaluate the accuracy and performance of these models.
- Finding out the best model for our business case & providing executive summary.

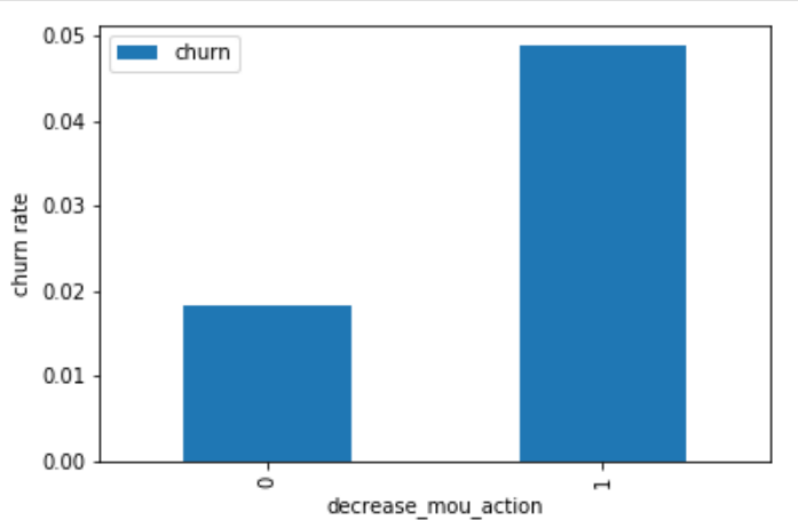
DATASET DESCRIPTION

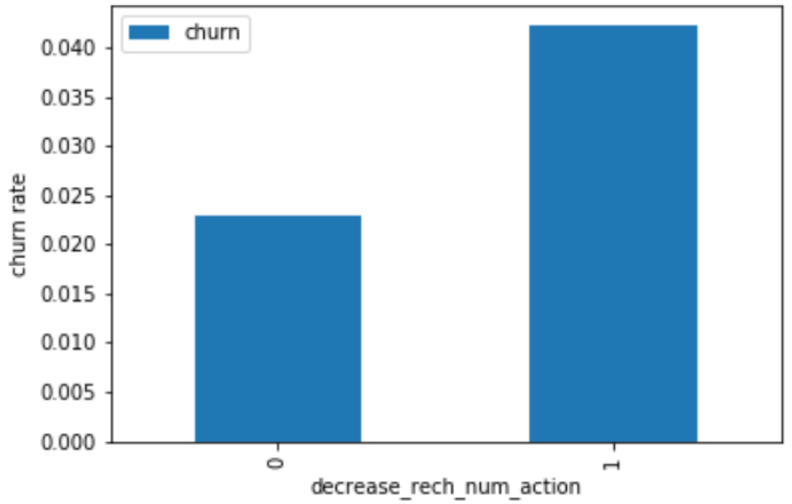
- Source dataset is in csv format.
- Dataset contains 99999 rows and 226 columns
- There are missing values for the provided input dataset which needs to be handled.
- After removing data with missing values, we have 27705 rows and 136 columns.
- Churn is the variable which notifies whether a particular customer is churned or not. And we will be developing our models to predict the churn.
- There are outliers present in the dataset.
- There is also class imbalance of the ‘churn’ variable.

METHODOLOGIES

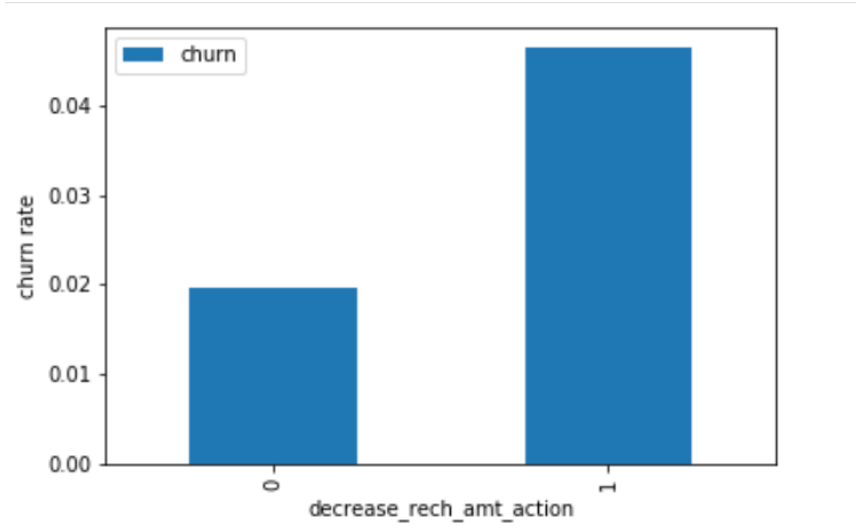
- Data Reading & Cleaning.
- Handling missing values & outliers with appropriate treatment method.
- EDA(Exploratory Data Analysis)
- Handling class imbalance
- Model building
- Evaluating of various models on required parameters.
- Selection of best model.

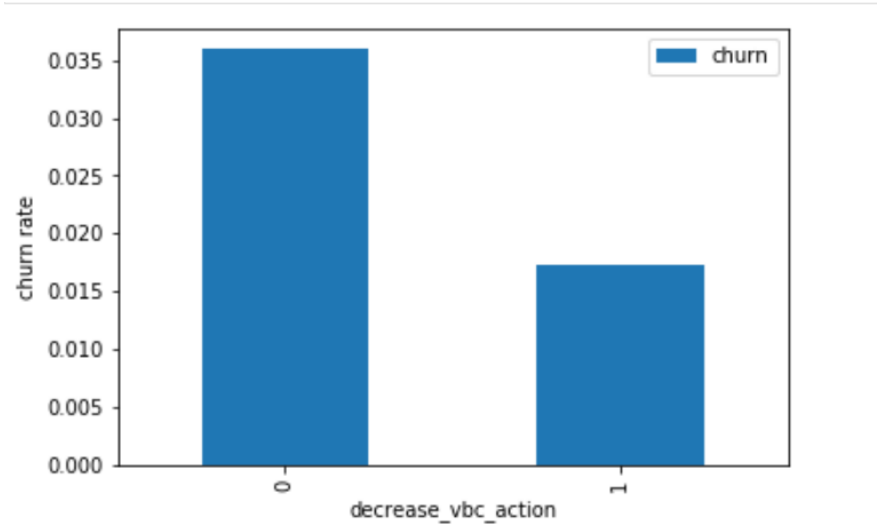
UNIVARIATE ANALYSIS





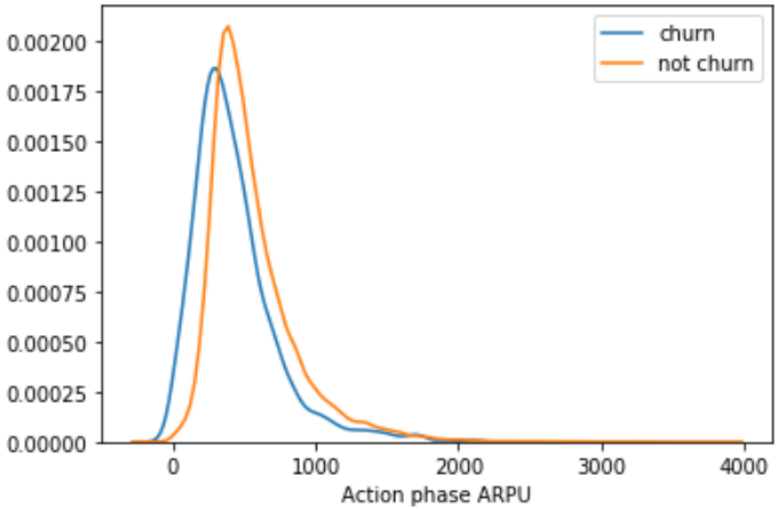
- We can see that the churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase than the good phase.
- Churn rate on the basis whether the customer decreased her/his number of recharge in action month
- As expected, the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.

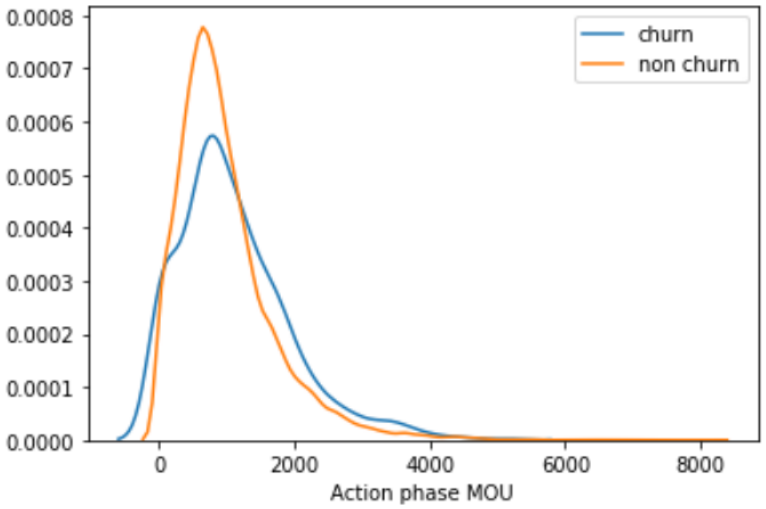




- Here also we see the same behaviour. The churn rate is more for the customers, whose amount of recharge in the action phase is lesser than the amount in good phase.

- Here we see the expected result. 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.

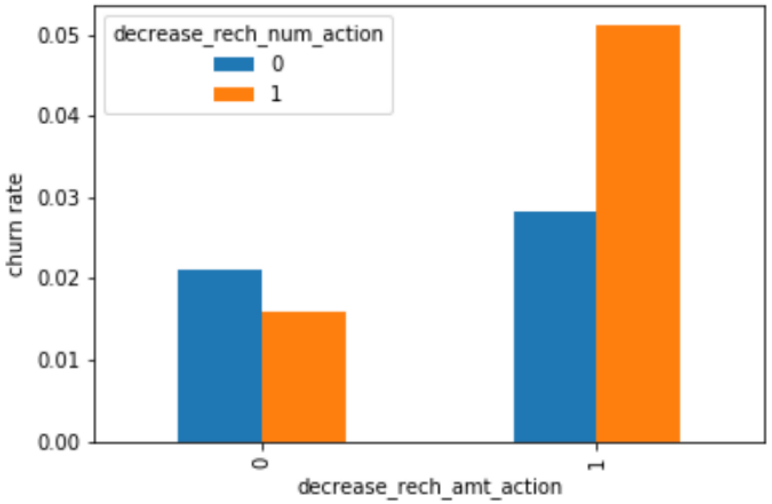


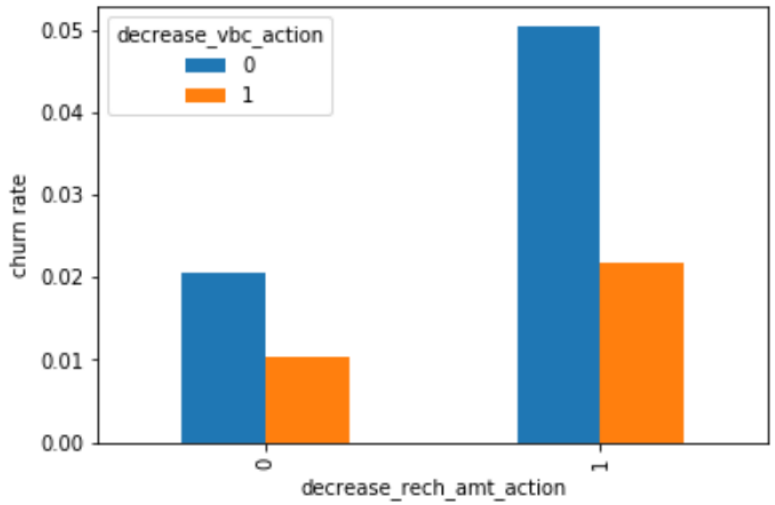


- Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned.
- ARPU for the not churned customers is mostly densed on the 0 to 1000.

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

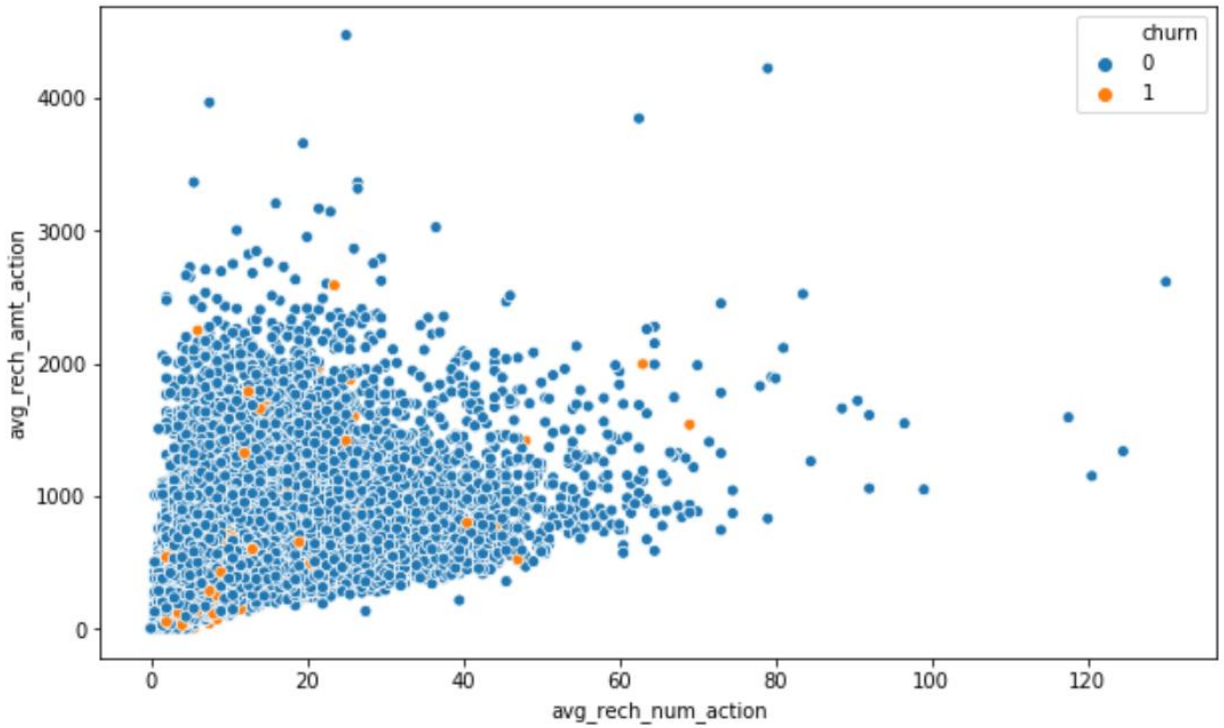
BIVARIATE ANALYSIS





- We can see from the above plot, that 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.

- Here, also we can see 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.



- We can see from the above pattern that the recharge number and the recharge amount are mostly proportional.
- More the number of recharge, more the amount of the recharge.

MODEL ANALYSIS

	Train Set			Test Set		
Model	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
Logistic Regression with PCA	0.86	0.89	0.83	0.83	0.81	0.83
Decision Tree with PCA	0.9	0.91	0.88	0.86	0.7	0.87
Random Forest with PCA	0.84	0.88	0.8	0.8	0.75	0.8

Final conclusion with PCA

- After trying several models we can see that for achieving the best sensitivity, which was our ultimate goal, the classic Logistic regression or the SVM models preforms well.
- For both the models the sensitivity was approx. 81%. Also we have good accuracy of approx. 85%.

	Train Set			Test Set		
Model	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
Logistic Regression with PCA	0.86	0.89	0.83	0.83	0.81	0.83
Logistic Regression with no PCA	0.84	0.81	0.83	0.78	0.82	0.78

Final conclusion with no PCA

- We can see that the logistic model with no PCA has good sensitivity and accuracy, which are comparable to the models with PCA.
- So, we can go for the more simplistic model such as logistic regression with PCA as it explains the important predictor variables as well as the significance of each variable.
- The model also helps us to identify the variables which should be act upon for making the decision of the to be churned customers.
- Hence, the model is more relevant in terms of explaining to the business.

FINDINGS AND SUGGESTIONS

- 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).
- Target the customers, whose outgoing others charge in July and incoming others on August are less.
- 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.
- Customers, whose monthly 3G recharge in August is more, are likely to be churned.
- 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.
- Customers decreasing monthly 2g usage for August are most probable to churn.
- Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
- 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.