# Domain Oriented Assignment Telecom Churn Case Study

Submitted By: Rajeev Punjabi
D57 June Batch

# Problem Statement (Overview)

- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another
- 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.
- For many incumbent operators, *retaining high profitable customers is the number one business goal*.
- To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn.**

# Definitions of churn

There are various ways to define churn, such as:

**Revenue-based churn**: Customers who have not utilised any revenue-generating facilities such as mobile internet, outgoing calls, SMS etc. over a given period of time. One could also use aggregate metrics such as 'customers who have generated less than INR 4 per month in total/average/median revenue.

The main shortcoming of this definition is that there are customers who only receive calls/SMSes from their wage-earning counterparts, i.e. they don't generate revenue but use the services. For example, many users in rural areas only receive calls from their wage-earning siblings in urban areas.

**Usage-based churn**: Customers who have not done any usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.

A potential shortcoming of this definition is that when the customer has stopped using the services for a while, it may be too late to take any corrective actions to retain them. For e.g., if you define churn based on a 'two-months zero usage' period, predicting churn could be useless since by that time the customer would have already switched to another operator.

**In this project, I have used the usage-based definition to define churn.**

# Customer behaviour during churn

In churn prediction, I have assumed that there are three phases of the customer lifecycle as per the defined guideline :

The ‘good’ phase: In this phase, the customer is happy with the service and behaves as usual.

The ‘action’ phase: The customer experience starts to sore in this phase, for e.g. he/she gets a compelling offer from a competitor, faces unjust charges, becomes unhappy with service quality etc. In this phase, the customer usually shows different behaviour than in the ‘good’ months. Also, it is crucial to identify high-churn-risk customers in this phase, since some corrective actions can be taken at this point (such as matching the competitor’s offer/improving the service quality etc.)

The ‘churn’ phase: In this phase, the customer is said to have churned. Churn is defined based on this phase. Also, at the time of prediction (i.e. the action months), this data was not available to me for prediction. Thus, after tagging churn as 1/0 based on this phase, I have discarded all data corresponding to this phase.

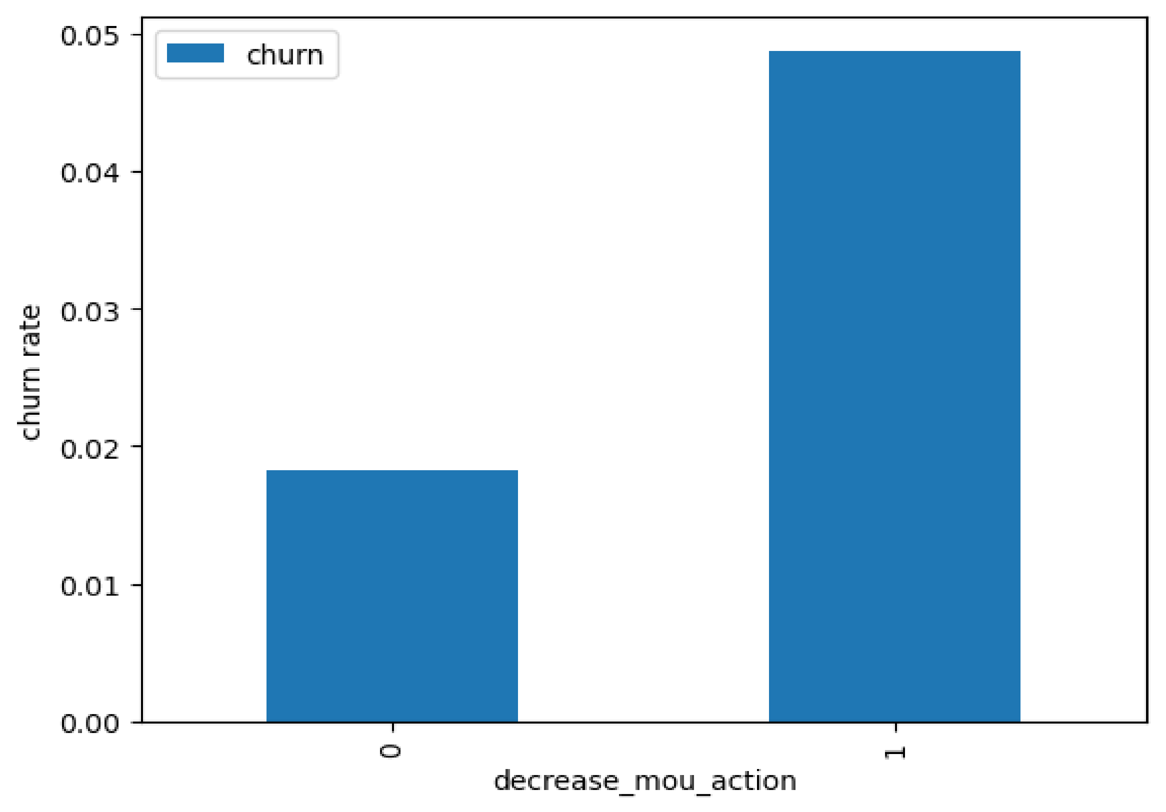
# Steps taken for Data preparation

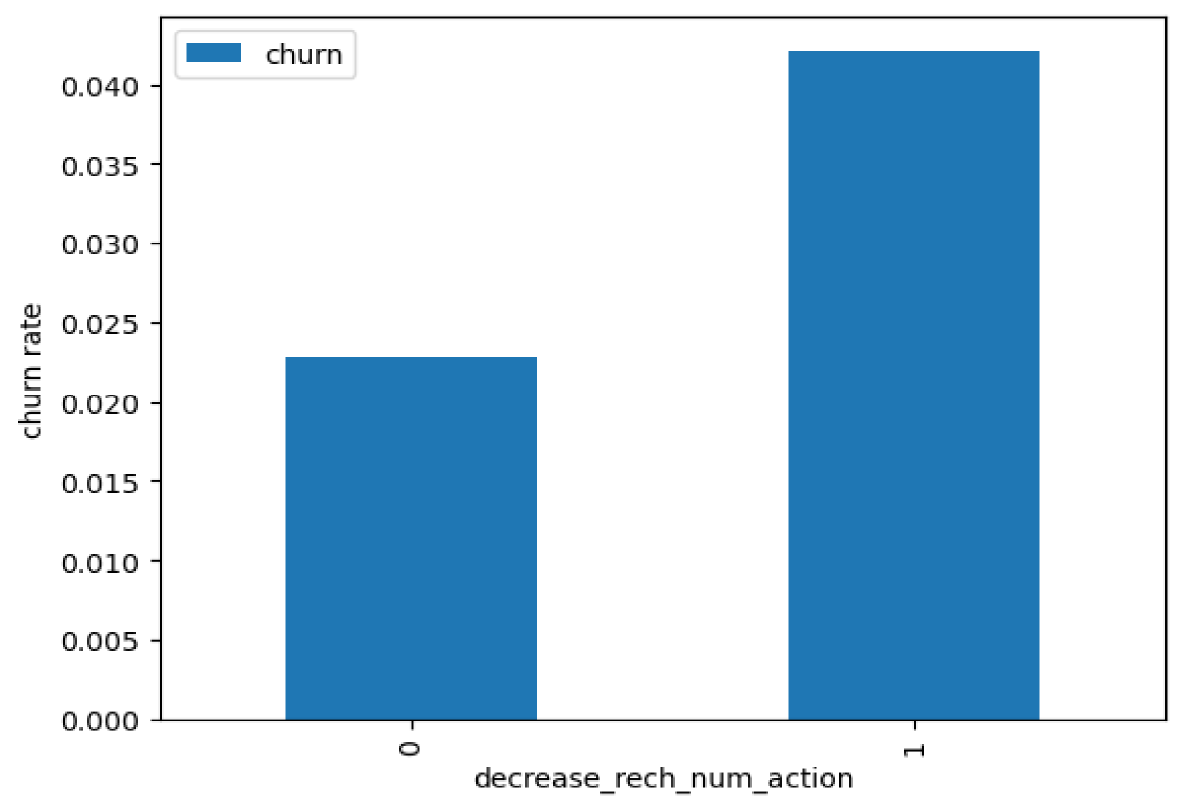
1. Handling missing values and Outlier Treatment
2. Filter high-value customers
3. Derived New Features (Columns) wherever necessary
4. Tag churners and remove attributes of the churn phase
5. Performed Univariate and Bivariate EDA
6. Built predictive Models to predict churn – Train-Test Split
7. Applied Logistic Regression concept in Modelling
8. Created Decision Trees and Random Forests

# Observations and Analysis

It can thus be seen that the churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase than the good phase.

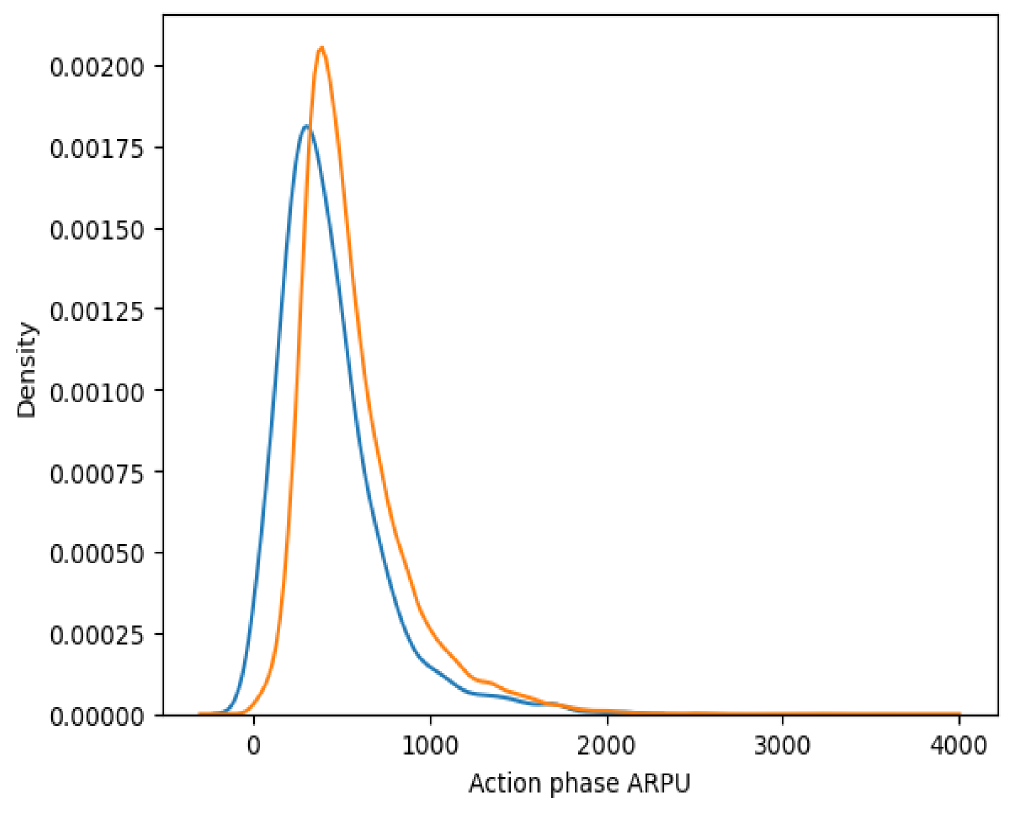
As expected, the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.

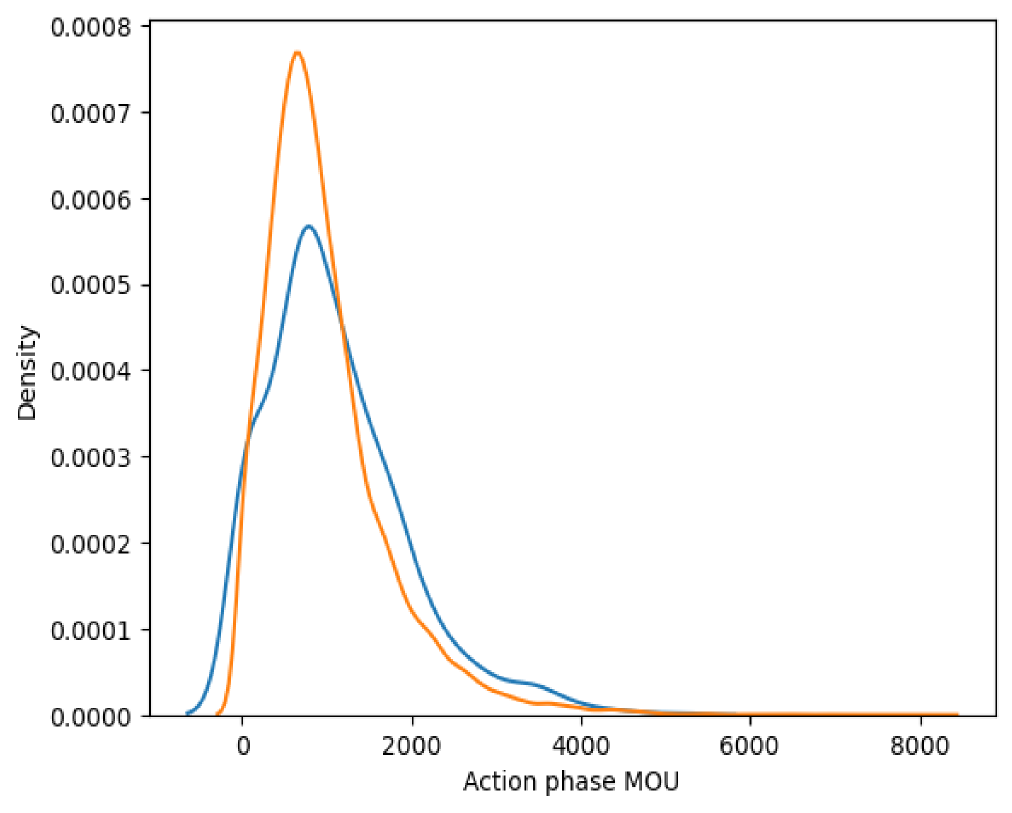


# Action Phase - Analysis

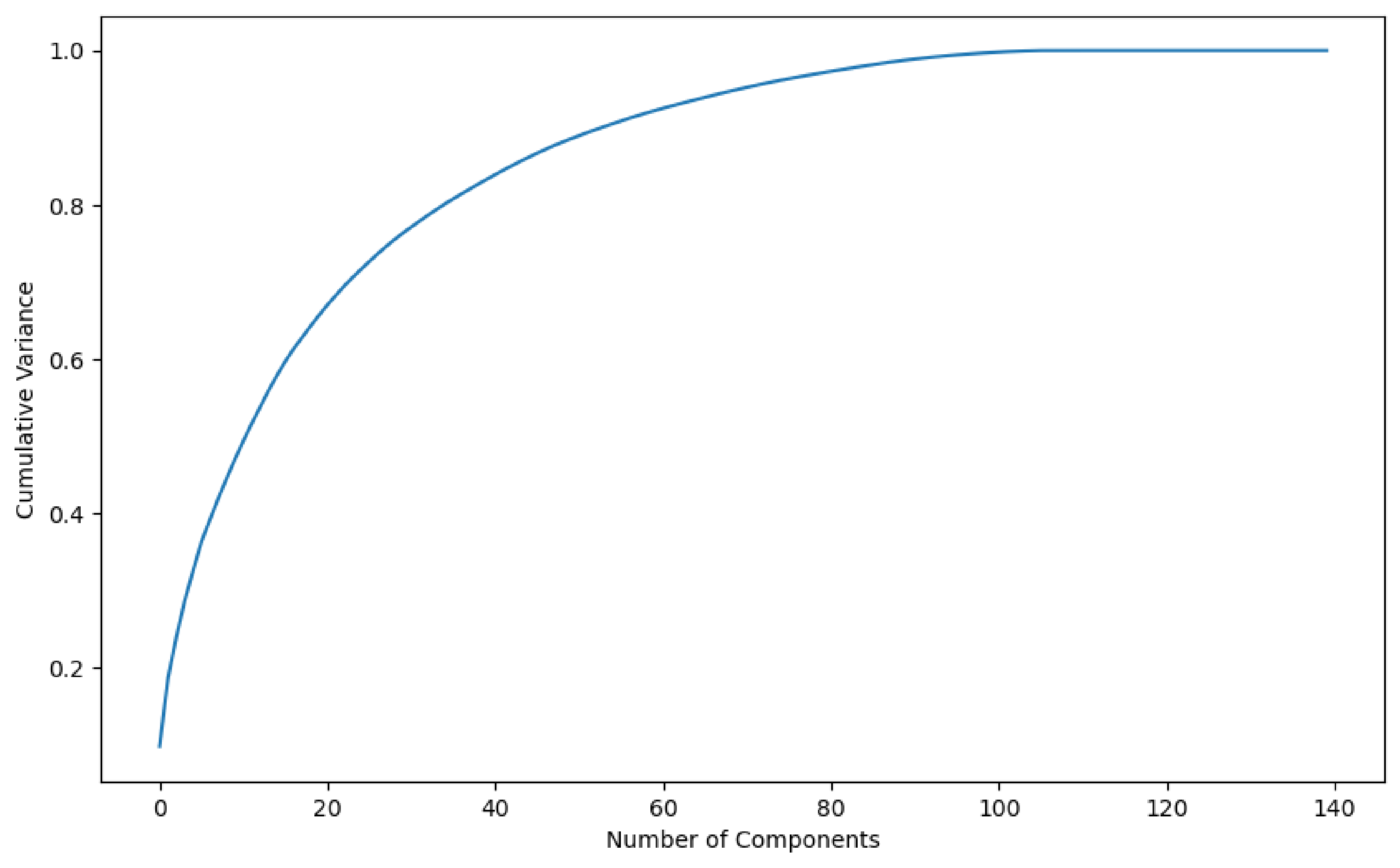
- Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned.
- ARPU for the not churned customers is mostly densed on the 0 to 1000




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

# PCA Model Analysis

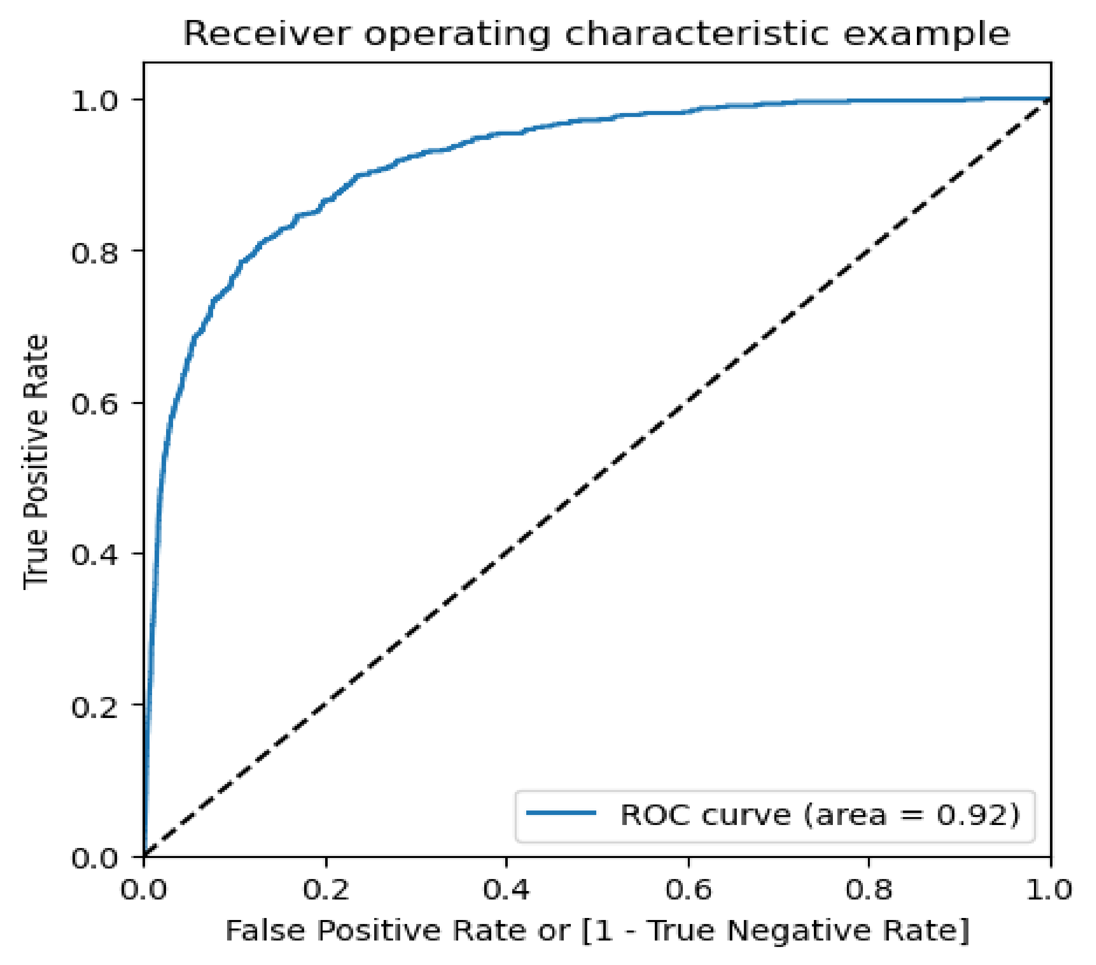



- **Principal Component Analysis (PCA)** is used to reduce the dimensionality of a data set by finding a new set of variables, smaller than the original set of variables, retaining most of the sample's information, and useful for the regression and classification of data
- We can see that 60 components explain amost more than 90% variance of the data. So, we will perform PCA with 60 components

# ROC Analysis

ROC curve is showing the performance of a classification model at all classification thresholds.
This curve plots two parameters:
- True Positive Rate
- False Positive Rate

**We can see the area of the ROC curve is closer to 1, which is the Gini of the model.**



# Recommendations

- ✓ Goal should be to 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)
- ✓ Customers whose outgoing others charge in July and incoming others on August are less can also be the target group
- ✓ 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 with 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

# THANK YOU