CASE STUDY TELECOM CHURN

GROUP MEMBERS:

Sandhya Sahu
Vivek Saxena
Sangeethapriya Jayakumar

AGENDA

INTRODUCTION:

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.**

BUSINESS OBJECTIVE

In this project, we will analyze customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months.

We assume that there are three phases of the customer lifecycle :

- 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.
- The 'churn' phase: In this phase, the customer is said to have churned. We define churn based on this phase.

DATA PREPARATION

The following crucial steps were performed as part of preparing the data for our analysis:

1) Filtering High Value Customers:

The objective is to predict churn for High Value customers only, i.e. Those who have recharged with an amount more than or equal to X, where X is the 70th percentile of the average recharge amount in the first two months

2) Tag Churners and create the target variable:

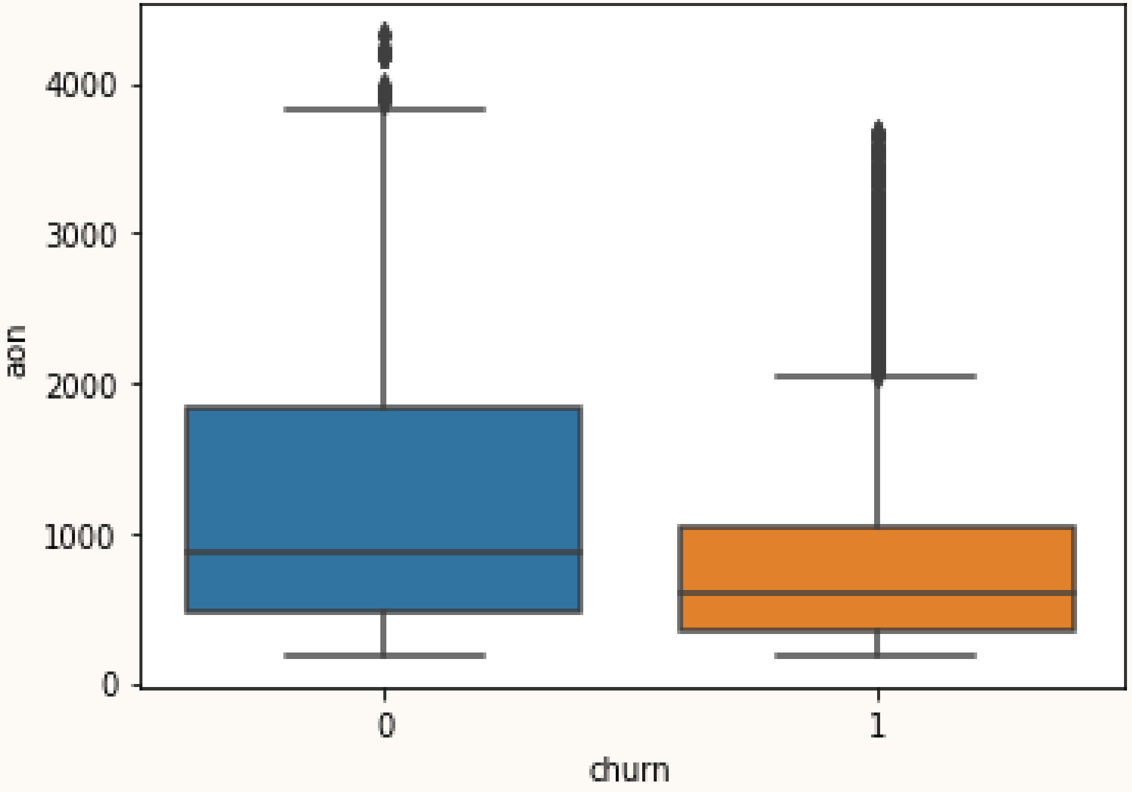
The churned customers (churn=1, else 0) are tagged based on their usage in the fourth month. Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase are tagged as churners .

The attributes used to tag churners are:

- total_ic_mou_9
- total_og_mou_9
- vol_2g_mb_9
- vol_3g_mb_9

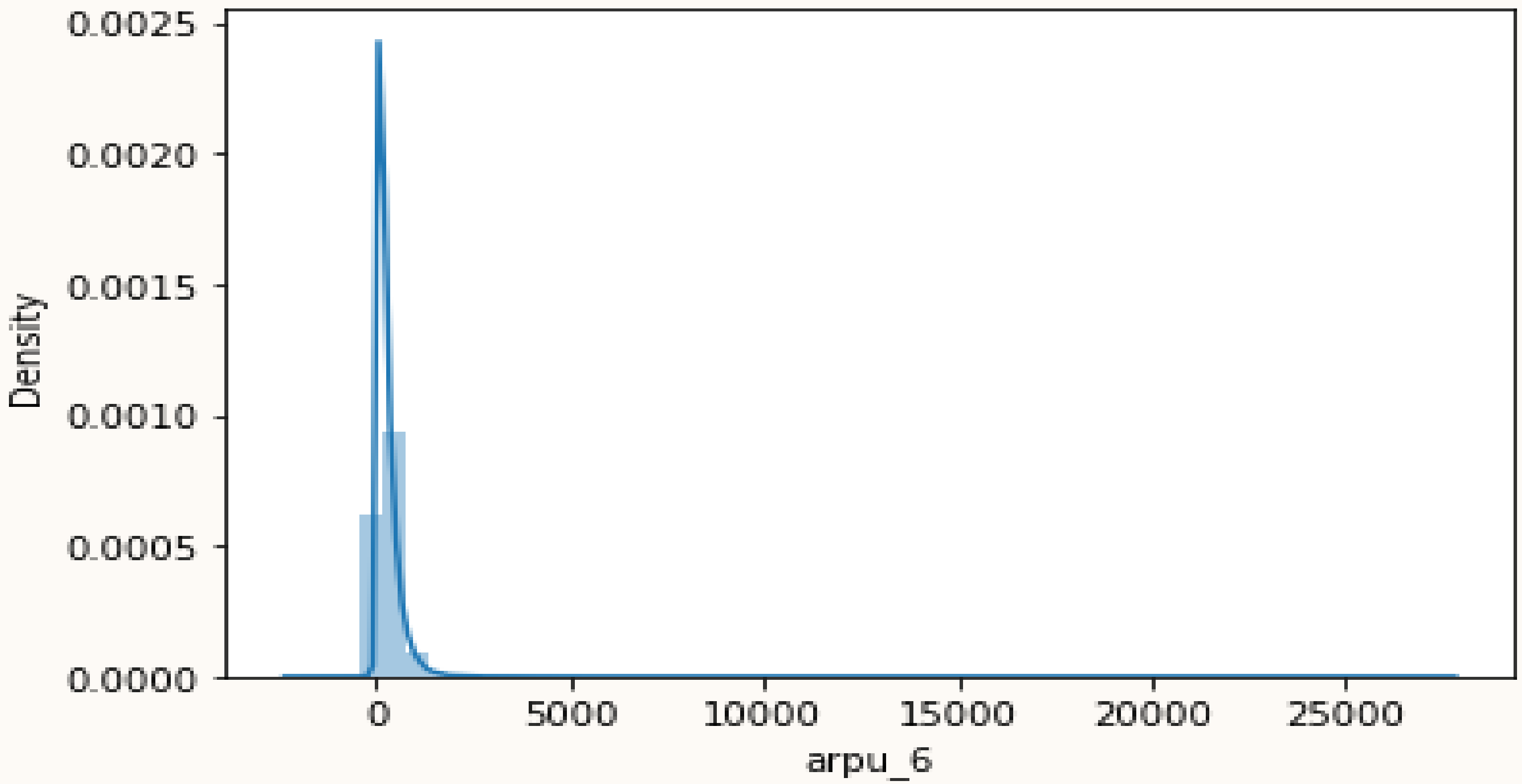
EXPLORATORY DATA ANALYSIS (EDA)

VISUALIZING THE DISTRIBUTION OF THE AGE ON NETWORK VALUES SHOWED US SKEWED DATA, WHICH SHOWS THAT PEOPLE TEND TO CHURN AFTER USING THE TELECOM SERVICES FOR A FEW YEARS.



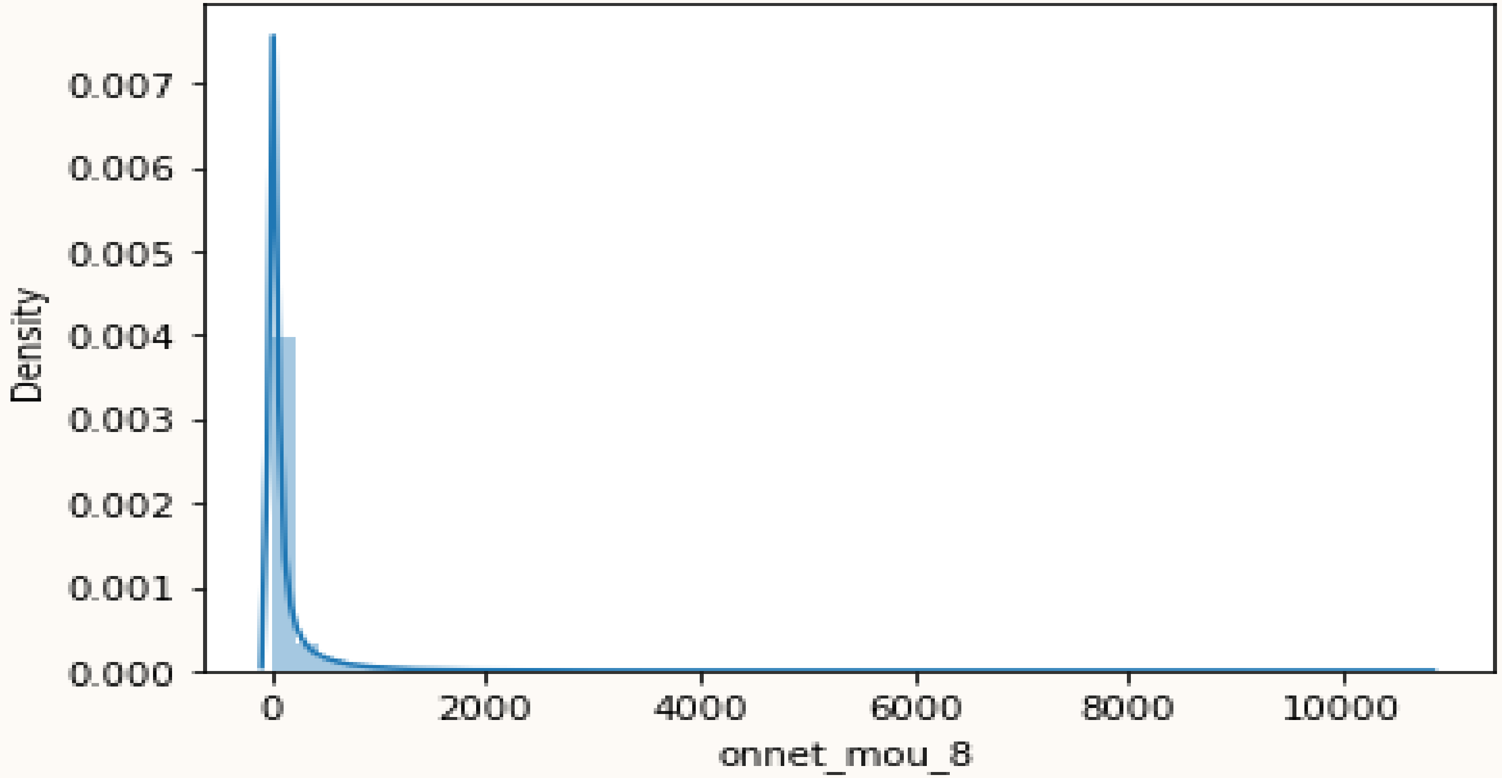
EDA

A decrease in Average revenue per user results in a higher churn rate



EDA

A decrease in all kinds of calls within the same operator network results in a higher churn rate



MODEL BUILDING

To predict the churn, we have used Logistic regression and Principal component analysis (PCA) as it can reduce the dimensionality of data sets so they can become a smaller set of variables and give better results.

After hyper parameter tuning, we concluded that an Accuracy of 87.37%.

Making predictions on the train set and test set, we calculated the following metrics:

```
y_pred = model.predict(X_test)

# creating onfusion matrix
print(confusion_matrix(y_test, y_pred))

# checking sensitivity and specificity
sensitivity, specificity, _ = sensitivity_specificity_support(y_test, y_pred, average='binary')
print("Sensitivity: \t", round(sensitivity, 2), "\n", "Specificity: \t", round(specificity, 2), sep='')

# checking area under curve
y_pred_prob = model.predict_proba(X_test)[:, 1]
print("ROC:    \t", round(roc_auc_score(y_test, y_pred_prob),2))
```

```
[[5795 1086]
 [ 120  488]]
Sensitivity:    0.8
Specificity:    0.84
ROC:             0.88
```

FEATURE IMPORTANCE

Fitting the model in the Random Forest classifier helped us get the importance of the features in the data set.

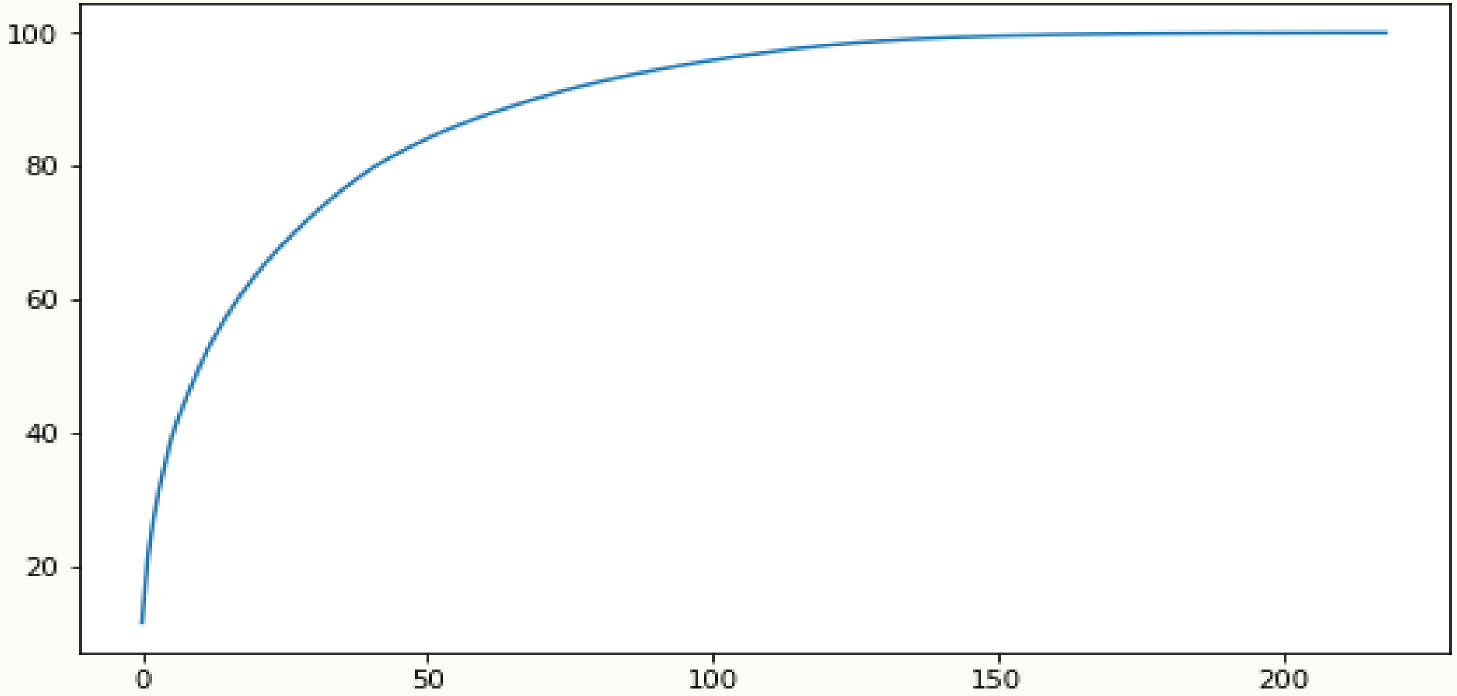
The top 10 features according to our model are as follows:

	variables	importance_percentage
0	loc_ic_t2m_mou_9	10.812733
1	total_ic_mou_9	8.619926
2	loc_ic_mou_9	8.054322
3	loc_og_mou_9	6.561208
4	loc_og_t2t_mou_9	6.318156
5	arpu_9	4.849755
6	loc_ic_t2t_mou_9	4.704393
7	max_rech_amt_9	4.535522
8	total_og_mou_9	4.523905
9	loc_og_t2m_mou_9	4.078045
10	std_ic_mou_9	3.003315

ROC CURVE

The ROC curve provides valuable insights into the trade-off between true positive rate and false positive rate for different threshold settings.

The Area Under Curve (AUC) shows a healthy 88% for our model.



IMPORTANT CONSIDERATIONS

Action Phase: The focus on the "action phase" likely refers to a specific period within the customer's lifecycle (e.g., post-activation, renewal period, etc.). Understanding this phase is essential for contextualizing the data.

Correlation, Not Causation: These factors show a correlation with churn; it's essential to investigate further to determine the actual causes of customer churn.

Steps to reduce the churn:

1. **Personalized Incentives:** Offer targeted discounts or promotions based on a customer's individual usage patterns and spending habits.
2. **Value-Added Services:** Bundle internet-related services with recharges, providing additional value and increasing customer engagement.
3. **Proactive Customer Engagement:** Reach out to customers directly through surveys or calls to understand their needs and address potential areas of dissatisfaction.
4. **Competitive Pricing:** Review data usage tariffs to stay competitive in the market, reducing the reason for customers to switch to alternative providers.
5. **Network Improvement:** Prioritize the expansion of 3G networks for wider coverage and improve 2G signal strength in areas with limited 3G accessibility.

THANK YOU