TELECOM CHURN CASE STUDY

Rohini Parthasarathy
Rohit Kumar

Business Problem overview

In the telecom industry, customers can 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

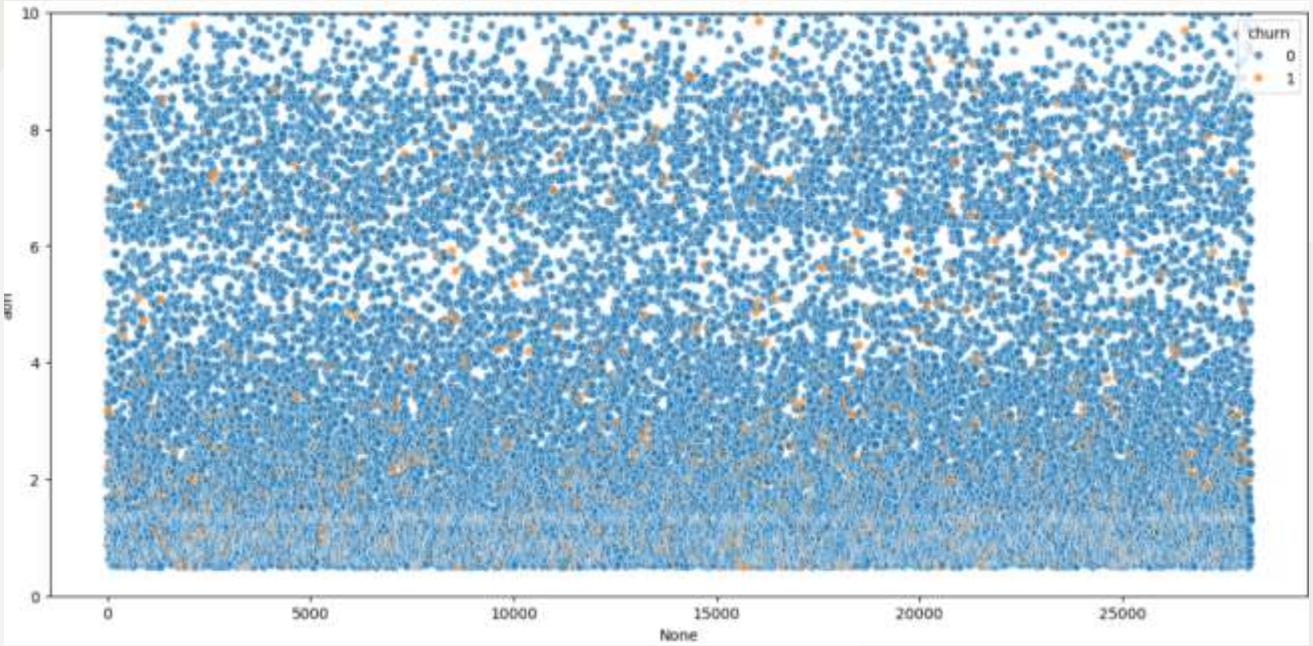
Business Objective

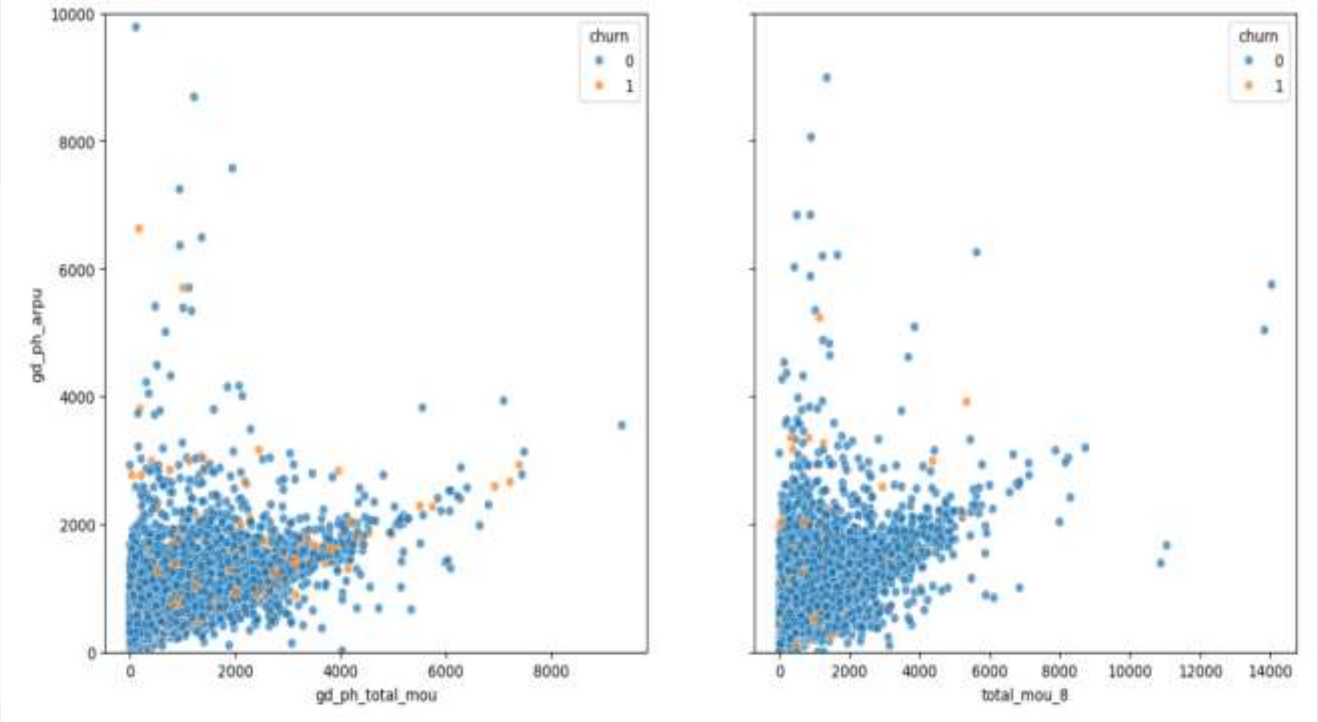
- The dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively.
- The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behavior during churn will be helpful

Steps involved in the Assignment

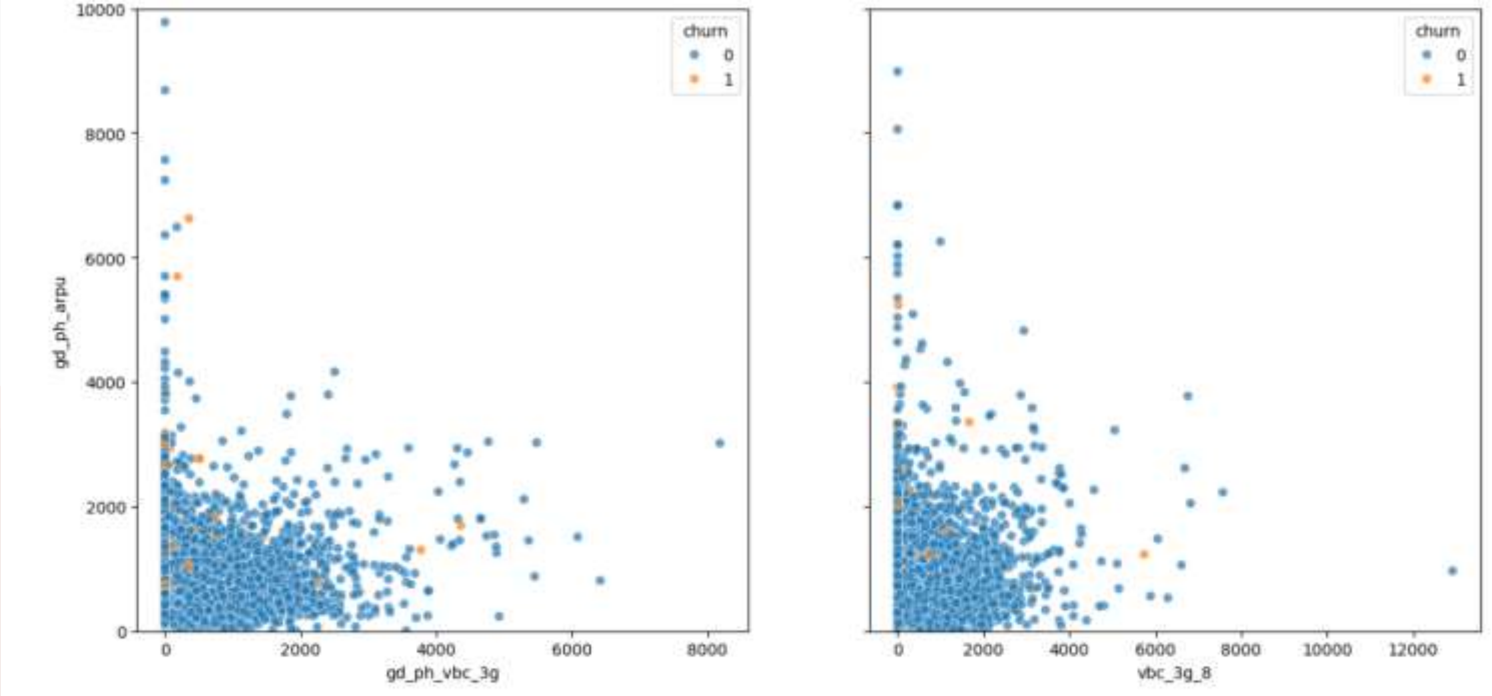
1. Reading the data
2. Finding the High value customer
3. Creating the target variable –'churn'
4. Cleaning the data and perform EDA
5. Deriving new features
6. Handling class Imbalance using SMOTE
7. Splitting data into train and test set
8. Model building using Logistic Regression, Decision Tree, Random Forest
9. Making Predictions and evaluating using all the three models.
10. ROC Curve
11. Prediction on test set
12. Finally validating the best model

EDA

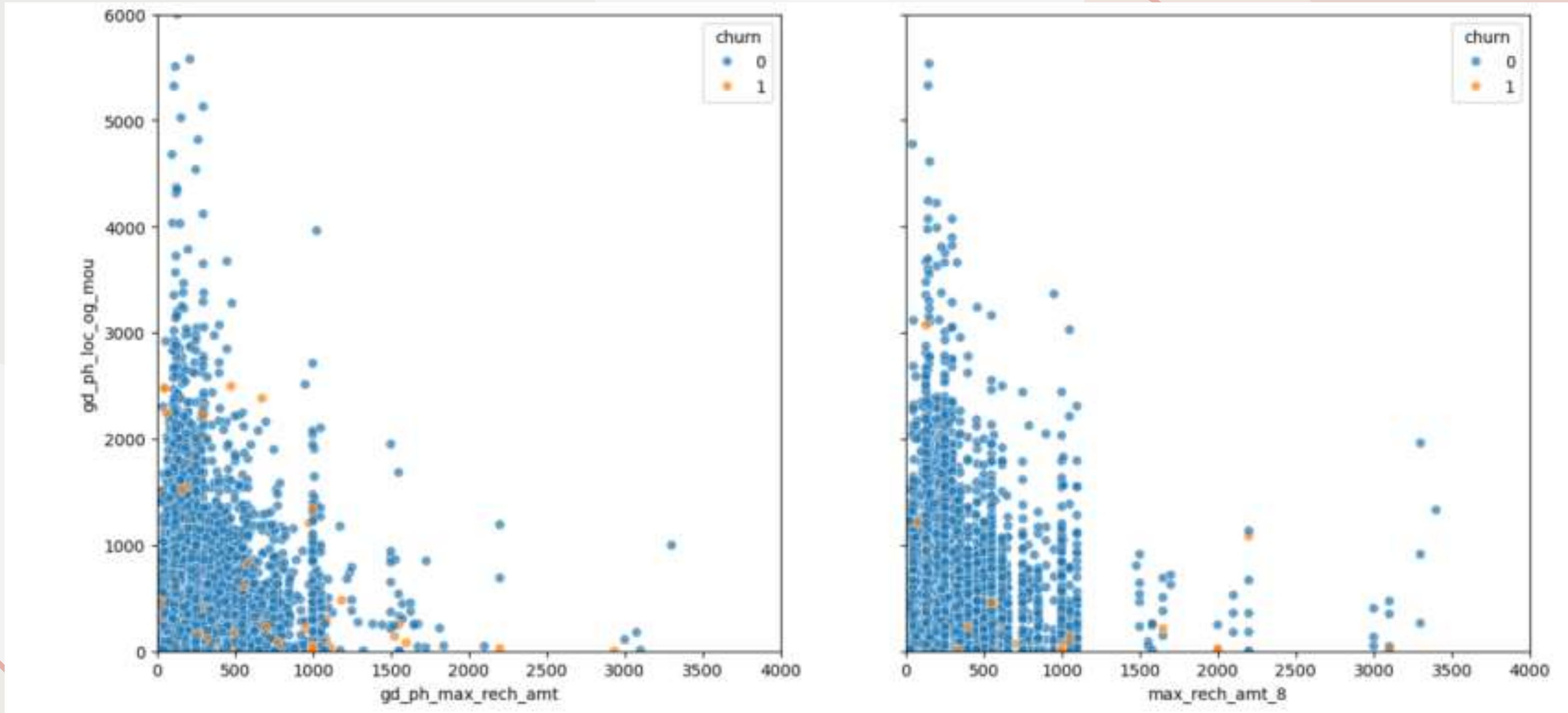




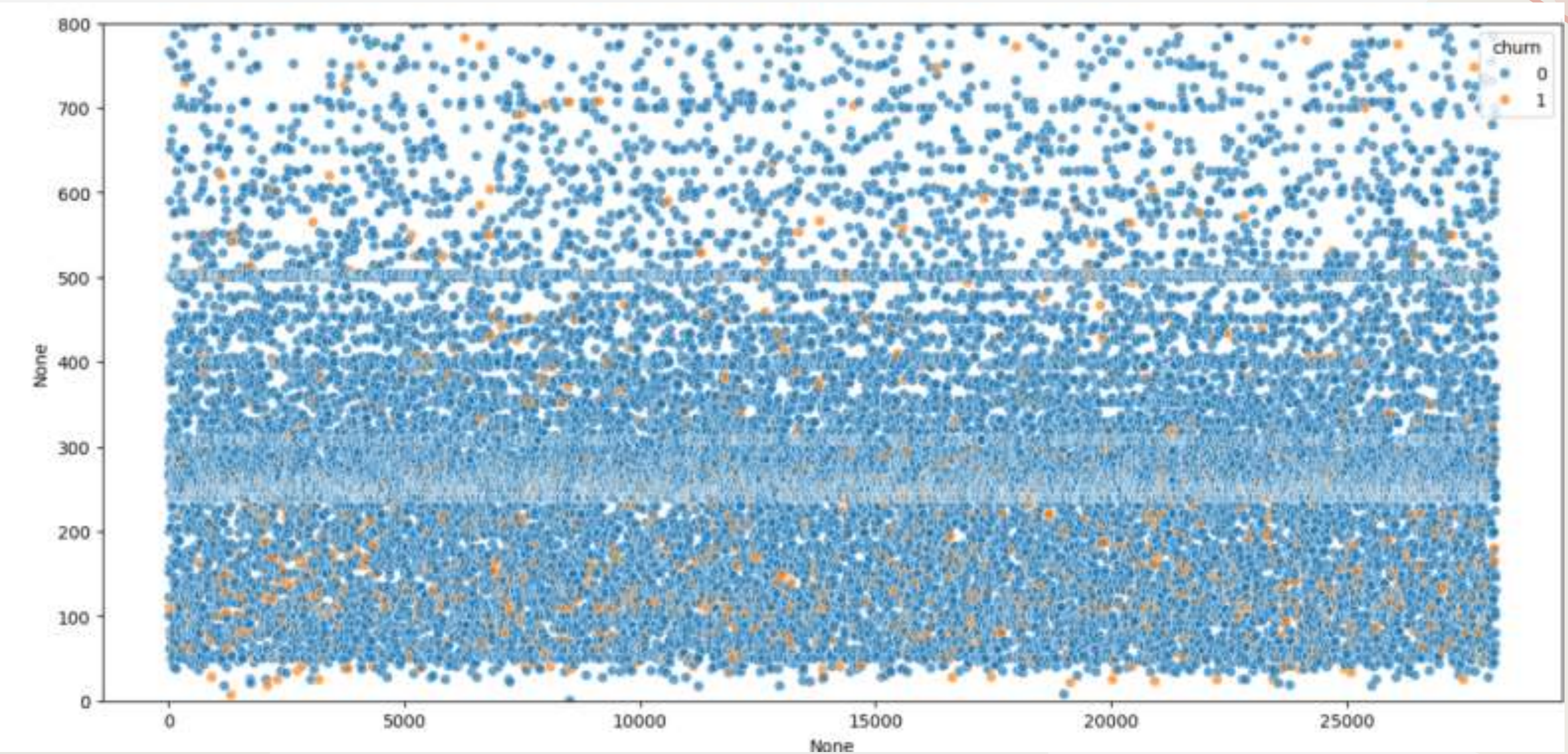
gd_ph_arpu
gd_ph_total_mou
total_mou_8
churn
0
1
10000
8000
6000
4000
2000
0
0
2000
4000
6000
8000
10000
12000
14000



churn
0
1
gd_ph_arpu
gd_ph_vbc_3g
10000
8000
6000
4000
2000
0
0
2000
4000
6000
8000
churn
0
1
vbc_3g_8
0
2000
4000
6000
8000
10000
12000



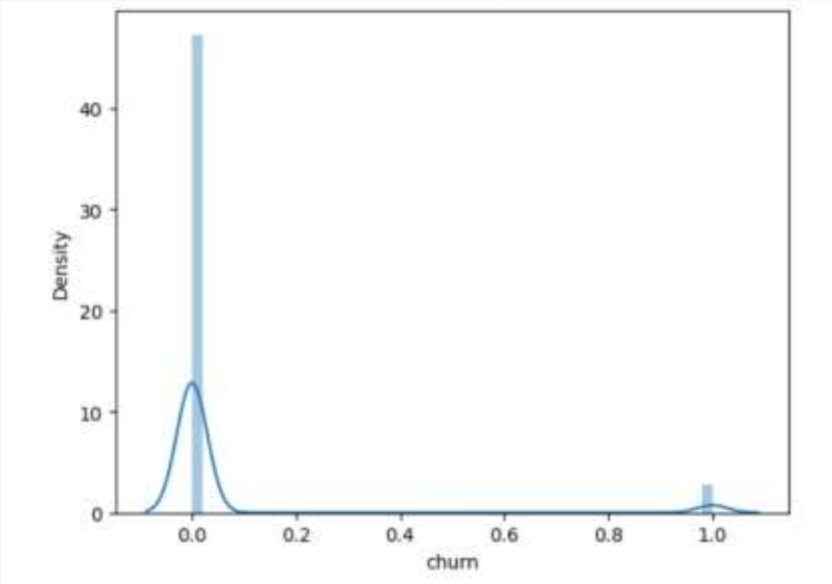
churn
0
1
gd_ph_loc_og_mou
gd_ph_max_rech_amt
0
1000
2000
3000
4000
5000
6000
0
500
1000
1500
2000
2500
3000
3500
4000
churn
0
1
max_rech_amt_8
0
500
1000
1500
2000
2500
3000
3500
4000

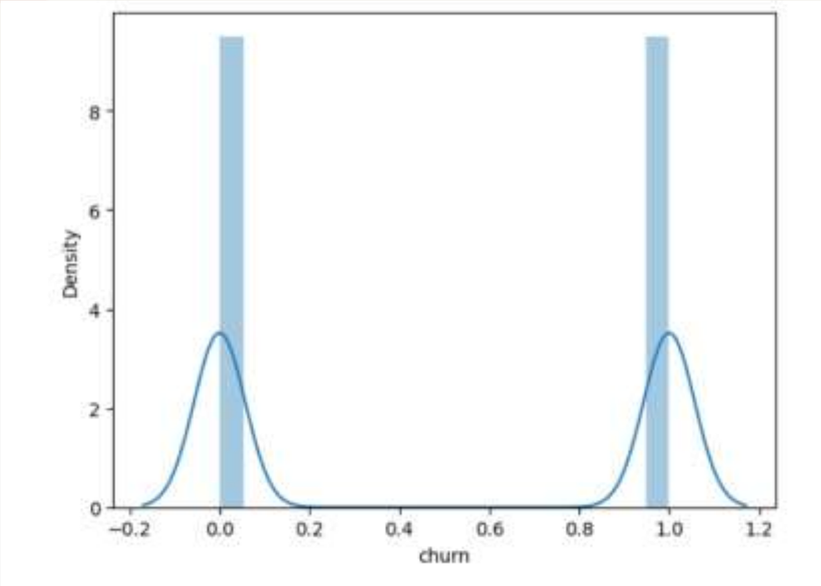

churn
0
1
800
700
600
500
400
300
200
100
0
None
0
5000
10000
15000
20000
25000
None

Observations

- We observe that those customers who have the tenure of less than 4 years turn to churn more.
- We find that MOU dropped significantly in the month_8 for the churners thus reducing the revenue.
- We find that the users whose recharge is high, used the local outgoing more.
- Also, we find that people whose max recharge amount as well as local out going were very less even in the good phase churned more.
- We find that users who had the max recharge amount less than 200 churned more
- We observe that those users who have max recharge amount on the higher end and still have low incoming call mou during the good phase, churned out more.

Handling class Imbalance





Principle Component Analysis

PCA

```
[258]: X.shape
```

```
[258]: (28163, 55)
```

```
[259]: from sklearn.decomposition import PCA

pca = PCA(n_components = 25)
X_pca = pca.fit_transform(X_res)
X_pca.shape
```

```
[259]: (53250, 25)
```

Model Building

- Since the target variable ‘Churn’ is a categorical variable, the first step is to use a simple model ,here we use ‘Logistic Regression’. We evaluated the model and found that the test accuracy was round 81.83%
- Then we used Decision tree classifier, and after hyper parameter tuning, The accuracy of the model was found to be 89%.
- Then we used Random Forest Classifier for our model building. We evaluated the model and after hyper tuning, the accuracy was found to be 94%.
- Hence Random Forest model came out to be the best.

Top 10 predictors using Random Forest classifier model

1. loc_ic_mou_8
2. monthly_3g_8
3. gd_ph_loc_ic_mou
4. total_mou_8
5. monthly_2g_8
6. sachet_2g_8
7. total_rech_num_8
8. last_day_rch_amt_8
9. loc_og_mou_8
10. gd_ph_std_og_mou

Conclusion and Recommendation

- We can see most of the predictors are from action phase.
- 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).
- Concentrate on those customers whose monthly 3g recharge and monthly 2G recharge in the month of August as they are more likely to get churned. Roll out some offers to those customers.
- Users that have been with the network less than 4 years, should be monitored time to time, as from data we can see that users who have been associated with the network for less than 4 years tend to churn more.
- MOU is one of the major factors, but data especially VBC if the user is not using a data pack if another factor to look out.