

# Telecom Churn Case Study

- Group members:**
- Aditya Kumar Roy
  - Dona Maria Joseph

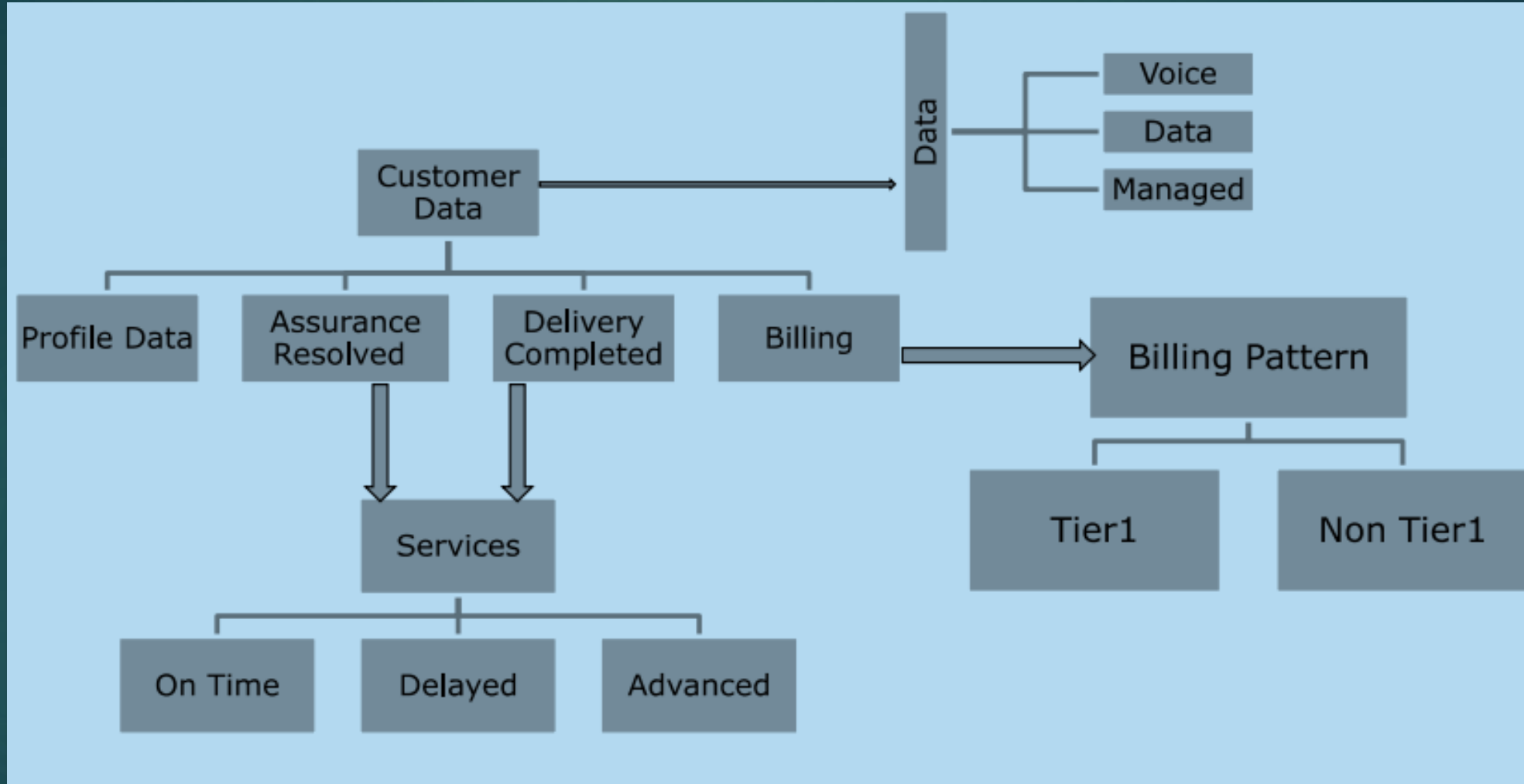
# Introduction

- Global Top Trends in technology and computing includes mobile technology.
- Landscape of Telecom Industry has changed.
  - ✓ Large Number of Private Service Providers have evolved.
  - ✓ To Survive in current Scenario new innovative business models are a must.
- Churn is huge factor in Telecom Industry.
- Major initiators of churn include.
  - ✓ Quality of service.
  - ✓ Tariffs.
  - ✓ Dissatisfaction in post sales service etc.
- Interesting facts surrounding churn.
  - ✓ Annual churn rate is estimated to be 25-30%.
  - ✓ Acquiring new customers is costlier than retaining them.

# OBJECTIVE

- ▶ To reduce customer churn
- ▶ Retaining high profitable customers by predicting customers who are at high risk of churn.

# Telecom Data - Classification



# Steps undertaken

- ▶ Data preparation
- ▶ Exploratory Data Analysis
- ▶ Treatment of data imbalance
- ▶ Modelling
- ▶ Model Evaluation
- ▶ Conclusion based on the right model
- ▶ Recommendations to reduce churn

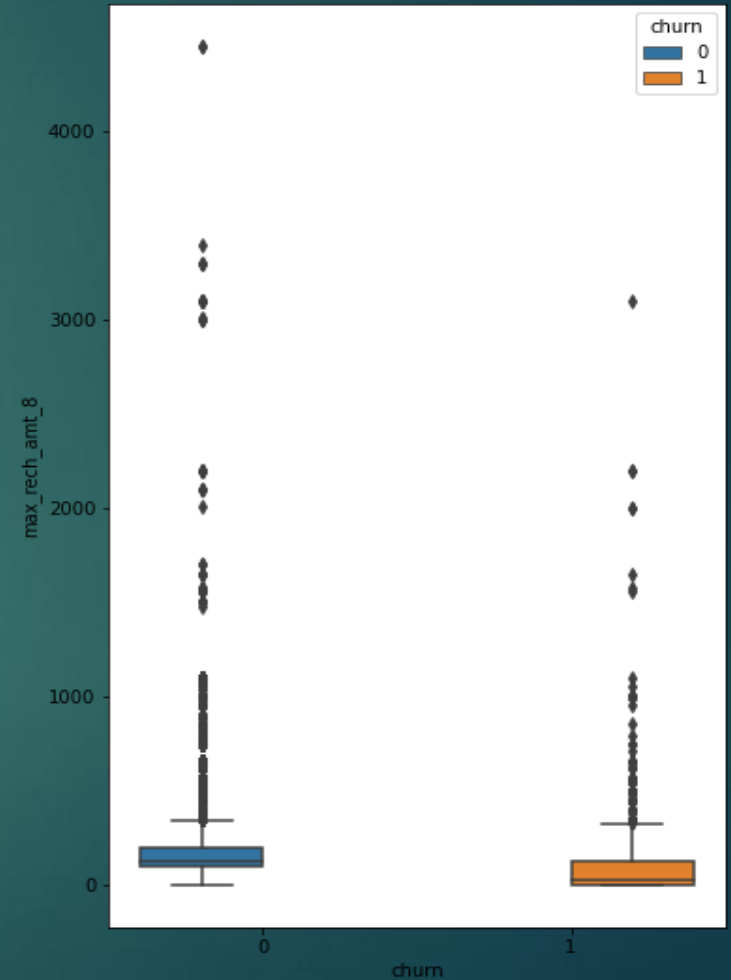
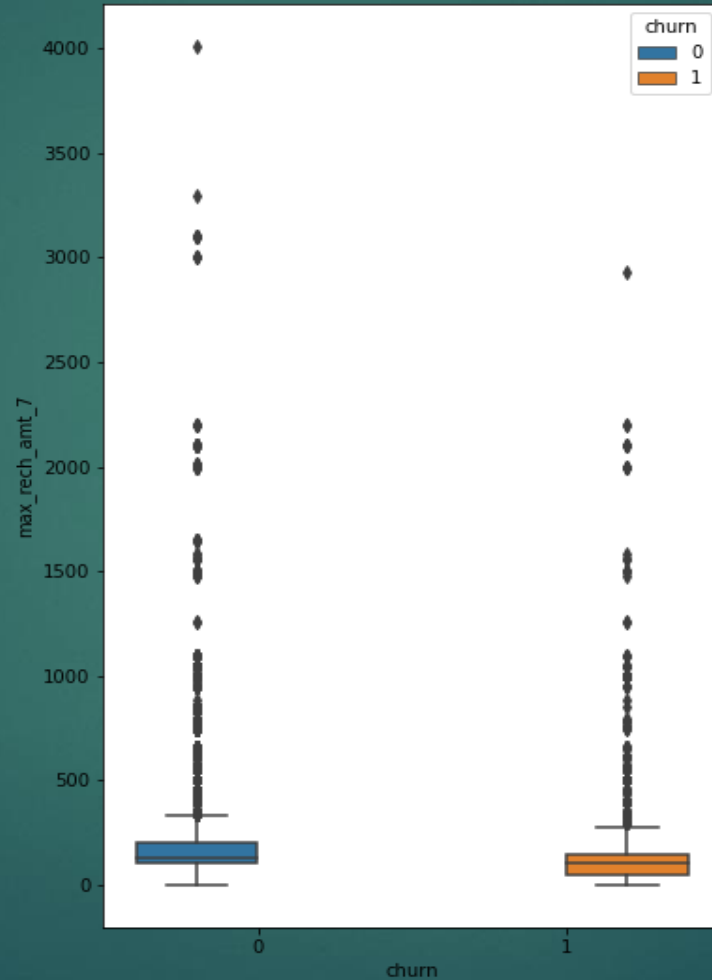
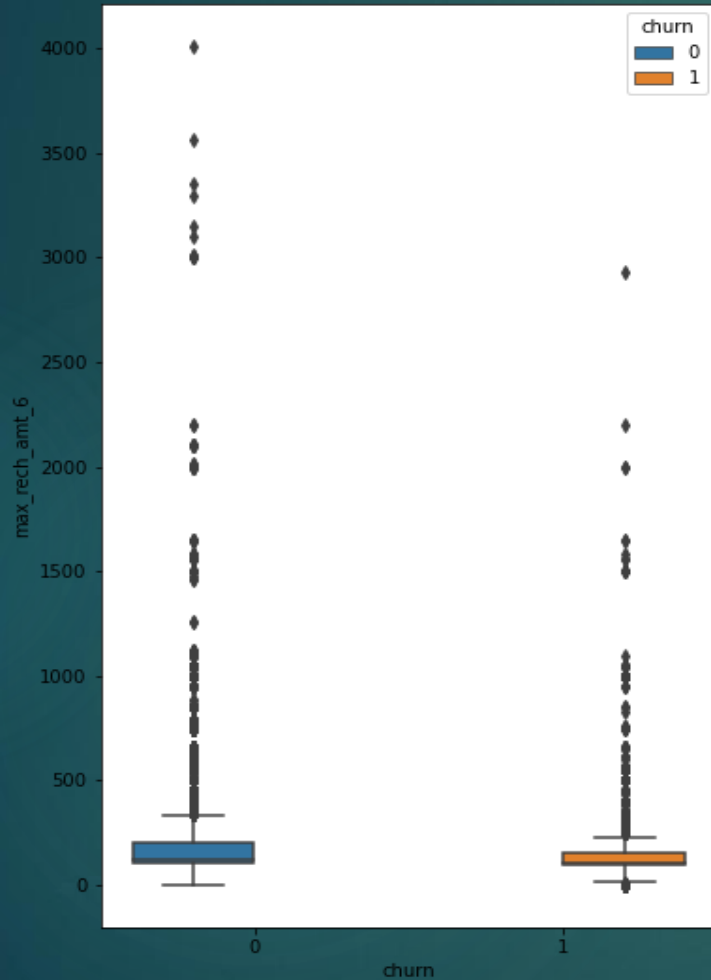
# Data preparation

- ▶ High value customers were filtered from the given data by looking at the 70th percentile of the average recharge amount in the first two months (the good phase)
- ▶ Missing values were imputed for better analysis
- ▶ Churners were tagged and attributes of the churn phase were removed

# EDA

- ▶ Univariate, bivariate and multivariate analysis were done.
- ▶ Heat map was plotted to check for multicollinearity.

# Bivariate analysis for maximum recharge amount for the months 6,7,8





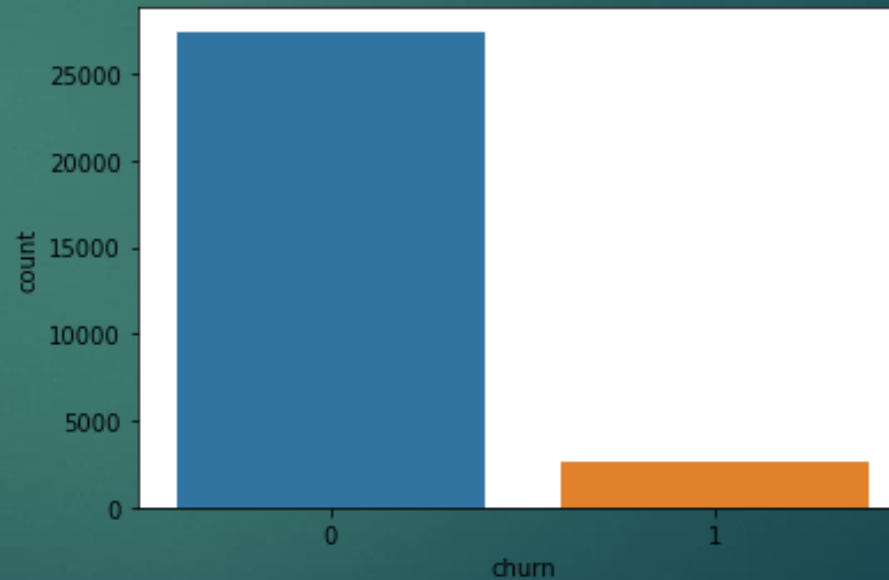
# Data Imbalance

Handling imbalance:

- The imbalance in data was treated using SMOTE.

1 = Churn

0 = Non Churn



# Modelling

Models used are

1. Logistic Regression
2. Decision Tree
3. Random forest

# Logistic regression With PCA

► The scores obtained are:

Train set

- Accuracy = 0.89
- Sensitivity = 0.92
- Specificity = 0.85

Test set

- Accuracy = 0.85
- Sensitivity = 0.81
- Specificity = 0.85

# Logistic Regression Without PCA

- Train set
  - Accuracy = 0.84
  - Sensitivity = 0.81
  - Specificity = 0.83
- Test set
  - Accuracy = 0.78
  - Sensitivity = 0.82
  - Specificity = 0.78

# Decision Tree

## Train set

- Accuracy = 0.90
- Sensitivity = 0.91
- Specificity = 0.88

## Test set

- Accuracy = 0.86
- Sensitivity = 0.70
- Specificity = 0.87

# Random Forest



## Train set

- Accuracy = 0.84
- Sensitivity = 0.88
- Specificity = 0.80

## Test set

- Accuracy = 0.80
- Sensitivity = 0.75
- Specificity = 0.80

# Conclusions

- Logistic regression with PCA explains the important predictor variables as well as the significance of each variable.
- Decrease in Total data recharge amount and Maximum recharge amount in month 8 indicates high chances of churn.
- A drop in 2G usage for Month 8 can be another indication of churn.
- Decrease in incoming and outgoing calls in month 8 shows high churn probability.

# Recommendations

- ▶ Providing more attractive tariffs than the competitors to the customers who are most likely to churn.
- ▶ Providing discount rates for outgoing calls during non peak hours.
- ▶ Offering better roaming charges targeting roaming customers.
- ▶ Improving the customer service quality.
- ▶ Providing better data plans than the competitors for those who highly use data.