

# SYRIATEL

## CUSTOMER CHURN PREDICTION

Forecasting Client Turnover

# Overview

- The telecommunication industry is a key player in our daily lives, particularly in media transmission, experiencing continuous growth.
- Operators in this sector face the ongoing challenge of staying competitive amidst rapid advancements.
- Retaining existing customers is paramount, as customer churn, the rate at which subscribers switch to competitors, poses a significant threat.
- Customer churn prediction involves employing data mining techniques to analyze customer data and anticipate churn.



# Project Workflow

```
graph LR; 1((1 Business Understanding)) --- 2((2 Data Understanding)); 2 --- 3((3 Data Preparation)); 3 --- 4((4 Exploratory Data Analysis)); 4 --- 5((5 Modelling)); 5 --- 6((6 Recommendations and Conclusions));
```

The diagram illustrates a six-step project workflow. The steps are arranged in two columns and connected by horizontal lines. Each step is represented by a rounded rectangle with a colored border and a numbered circle. The steps are: 1. Business Understanding (yellow border), 2. Data Understanding (orange border), 3. Data Preparation (blue border), 4. Exploratory Data Analysis (red border), 5. Modelling (teal border), and 6. Recommendations and Conclusions (yellow border). The workflow proceeds from left to right, starting with step 1 and ending with step 6.

1

**Business  
Understanding**

**Exploratory Data  
Analysis**

4

2

**Data  
Understanding**

**Modelling**

5

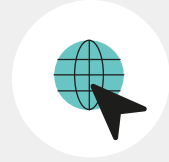
3

**Data Preparation**

**Recommendations  
and Conclusions**

6

# Business Understanding



## Business Problem

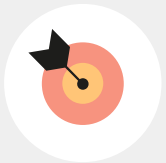
- Customer churn, the rate at which customers terminate services with a company, presents a significant challenge for telecom companies like SyriaTel.
- It impacts revenue, marketing expenses, and brand perception, directly affecting financial stability and market standing.

## Objectives

- Develop a precise machine learning model for forecasting customer churn, leveraging pertinent dataset features and classification training to identify customers prone to churn.

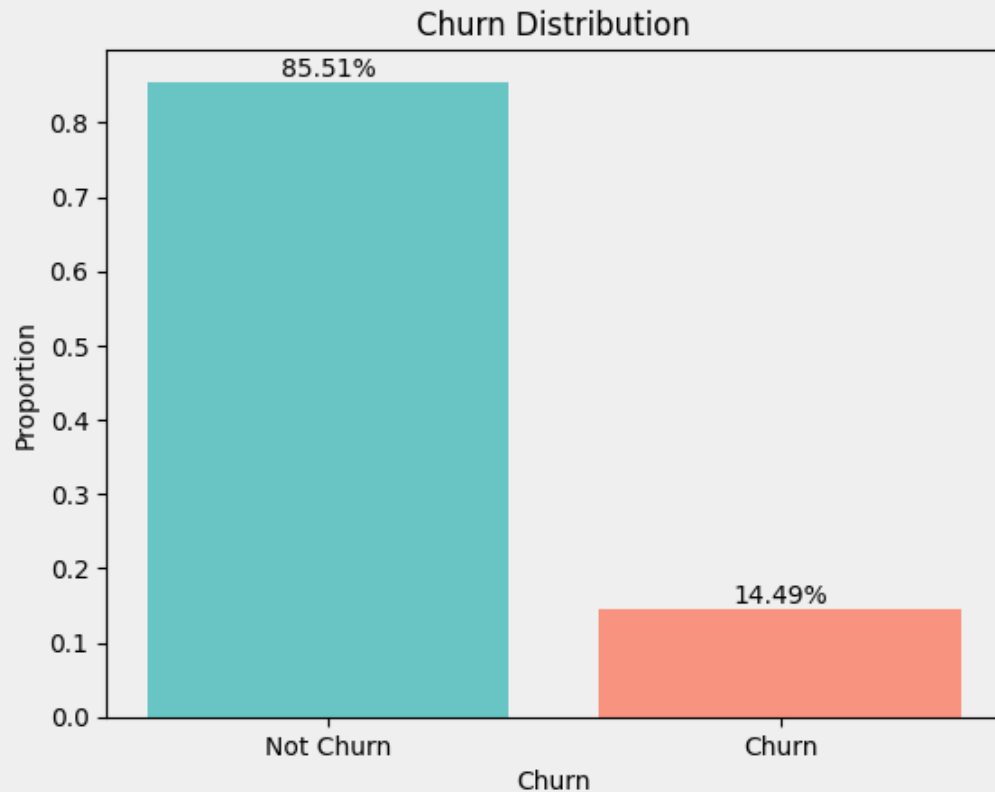
# Data Understanding

- The data for this project comes from the "Churn in Telecom" Kaggle dataset and focuses on customer activity and churn, which is when a customer cancels their service.
- This data will be used to build models that help SyriaTel reduce financial losses from customer churn.
- The dataset includes 3,333 customer records with 21 features about each customer, including demographics, service plans, call metrics, and a key variable indicating whether they churned (cancelled).
- The target column, "churn," is a boolean column. In this context, "True" indicates that the customer churned, while "False" signifies that the customer did not churn. Thus, this situation presents a binary classification problem.



# Data Analysis

## Univariate EDA



Total customer count:  
3,333

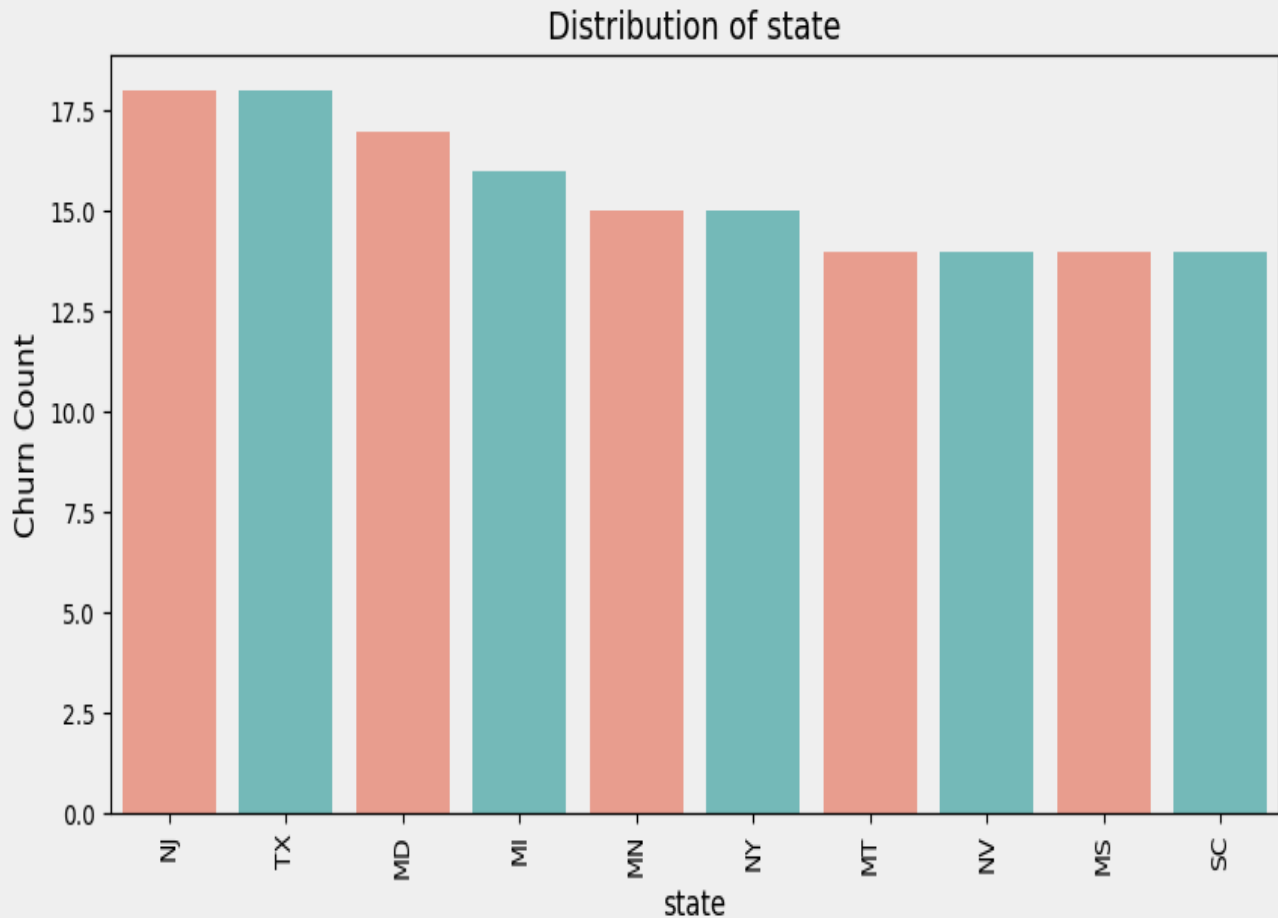
Count of churned customers:  
483

Churn rate:  
14.5%

# Data Analysis

## Bivariate Analysis

The majority of churned customers are from New Jersey, Texas, Maryland, Miami, and New York.

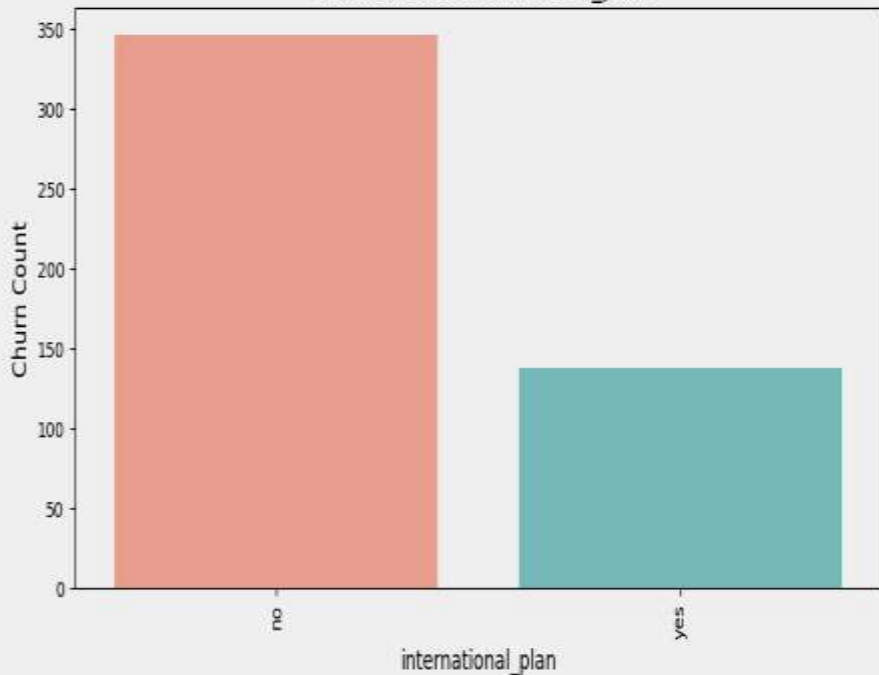


# Data Analysis

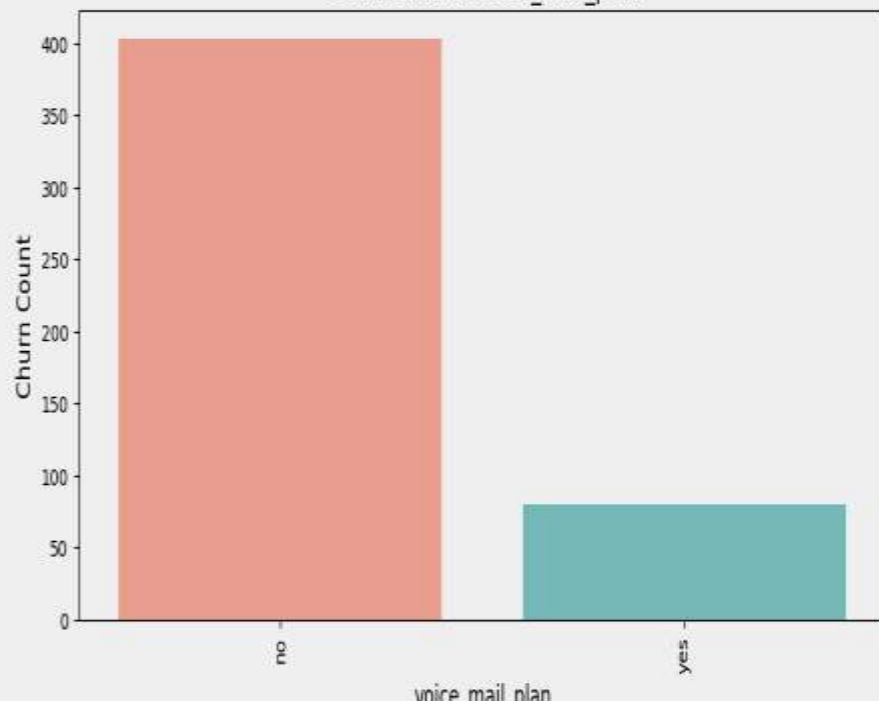
## Bivariate Analysis

- Subscribers to the international plan exhibit a notably higher likelihood of churning.
- For the voice mail plan, there is a reduced likelihood of churning compared to the general churn rate.

Distribution of international\_plan



Distribution of voice\_mail\_plan





# Modeling



The data posed a binary classification challenge, where various classification models were applied, and the one demonstrating the highest accuracy was chosen.

The project employed classification models such as:

- Logistic Regression.
- Decision Tree.
- Random Forest,.
- XG Boost.

XGBoost was chosen as it proved a better model due to its higher accuracy.

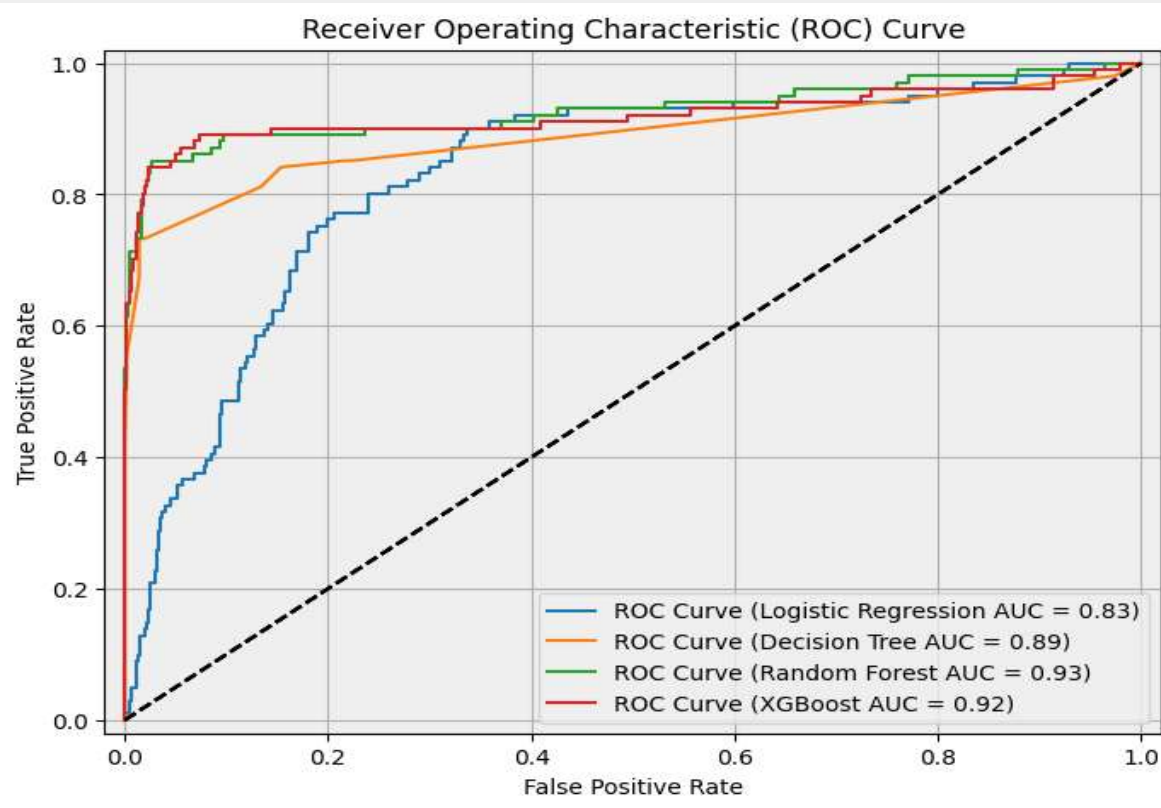
# Model Evaluation

	Model	Recall	Precision	F1-Score	Accuracy
0	Random Forest	0.643600	0.850000	0.730000	0.940000
1	Decision Tree	0.722800	0.960000	0.820000	nan
2	XGBoost	0.762400	0.930000	0.820000	0.950000

- XGBoost achieved the highest overall performance with a recall of 0.76, precision of 0.93, F1-score of 0.82, and accuracy of 0.95.
- Random Forest follows closely with a recall of 0.64, precision of 0.85, F1-score of 0.73, and accuracy of 0.94.
- Decision Tree shows improvement over Logistic Regression but falls short of Random Forest and XGBoost.
- Logistic Regression performance details are missing from this section but were presumably reported earlier.

# Model Evaluation

## ROC curve interpretation



- Logistic Regression: AUC = 0.83
- Decision Tree: AUC = 0.89
- Random Forest: AUC = 0.93
- XGBoost: AUC = 0.92

The Random Forest model has the highest AUC value, indicating that it has the best performance among the four models according to this metric.

# Conclusion



This analysis successfully achieved the objectives of building a machine learning model for customer churn prediction and utilizing it to estimate churn probability.

- Two models, Random Forest and XGBoost, were compared for their effectiveness in predicting churn.
- Both models demonstrated strong performance, with Random Forest showcasing a robust overall performance and XGBoost achieving a high recall score.

Discussion: SyriaTel should prioritize the Random Forest Classifier as the primary model for predicting customer churn due to its superior overall performance and ability to accurately identify potential churners.

# Recommendation



The analysis identified key factors significantly impacting customer churn prediction (Feature Importance):

- Call minutes and charges (daytime, evening, international)
- Customer service calls
- Usage of value-added services (voicemail plan)

Based on these insights, SyriaTel should implement strategic customer retention efforts that address these factors:

- Personalized offers and discounts: Target customers based on their usage patterns (e.g., high day charges) with relevant incentives.
- Reduce customer service calls: Invest in strategies like interactive voice response (IVR) systems to improve customer experience and decrease churn associated with frequent calls.
- Promote value-added services: Highlight the benefits of services like voicemail plans and incentivize customer adoption.