



# Customer Churn Prediction for SyriaTel

Using Machine Learning to Improve Customer Retention

**LEWIS KARANJA**

23<sup>rd</sup> July 2024



01

02

03

# Introduction



## Objective:

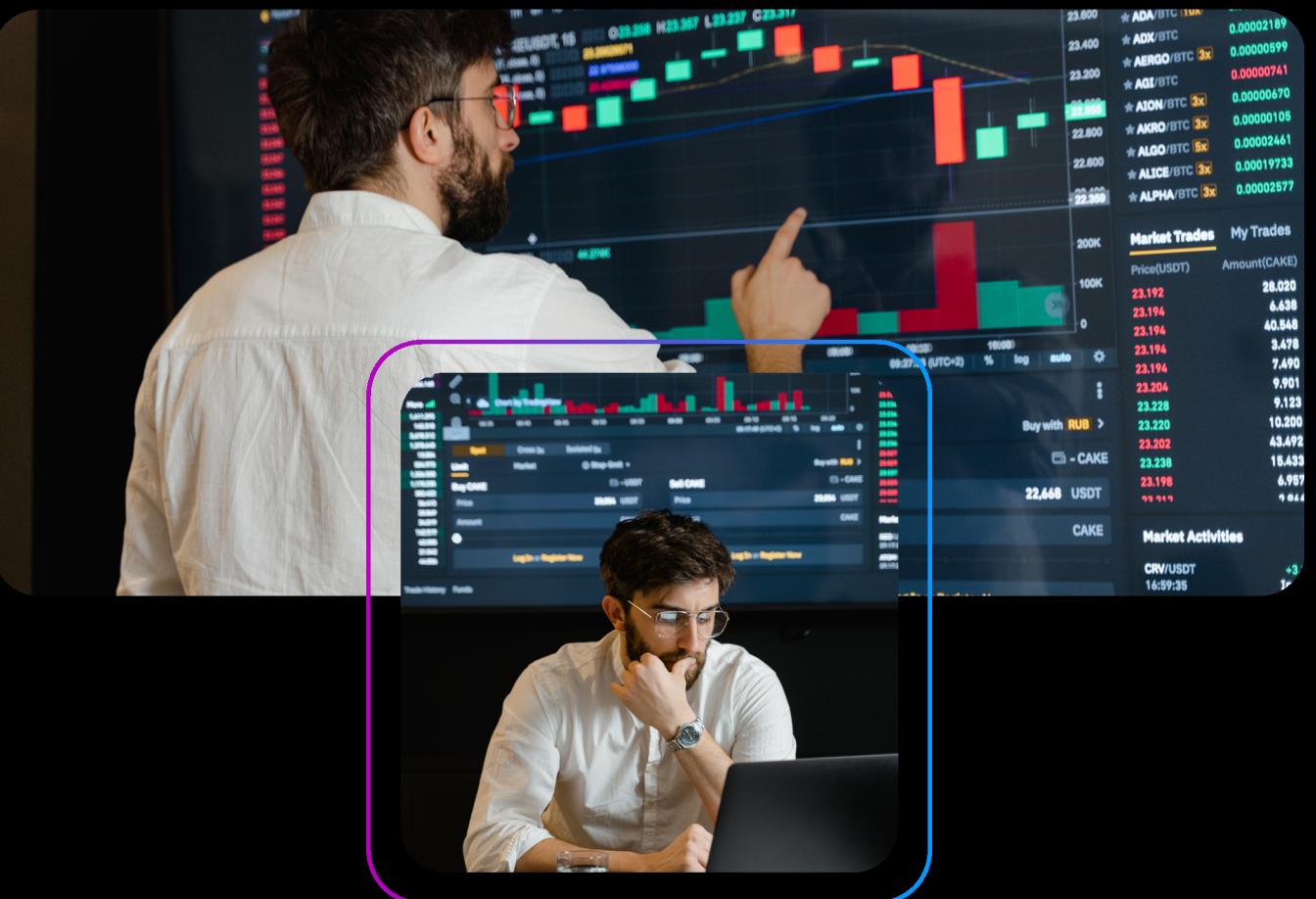
- Predict which customers are likely to churn (leave the service) using machine learning models.

## Business Impact:

- Churn is a major problem for telecommunications companies as it leads to revenue loss.
- Reducing churn can significantly improve the company's long-term profitability by retaining customers.
- 

## Goal:

- To create a model that predicts churn and allows SyriaTel to take proactive actions (like targeted retention efforts).



01

02

03



# Problem Statement

## Churn Definition:

- A customer is considered to have churned when they stop using SyriaTel's services.

## Stakeholder Focus:

- SyriaTel's leadership needs a way to identify and prevent churn before it happens, thus protecting revenue and improving customer lifetime value.

## Key Question

- Can we predict which customers are most likely to churn so that SyriaTel can take timely action?

## Goal:

- Build a predictive model that can flag high-risk customers for retention efforts.



01

02

03

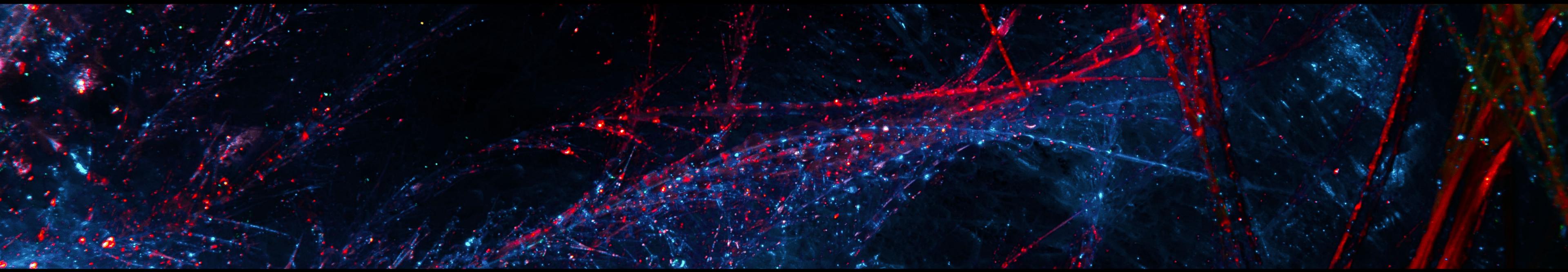


# Data Overview

- Dataset:
- We used a dataset from Kaggle containing customer data for telecommunications churn.
- Key Features:
- Demographics: Account length, area code, phone number.
- Usage Metrics: Total minutes (day, evening, night), total calls, service plan details.
- Target: Churn (1 for churned, 0 for non-churned).
- Data Size:
- 667 instances, 21 features (including service usage and customer data).



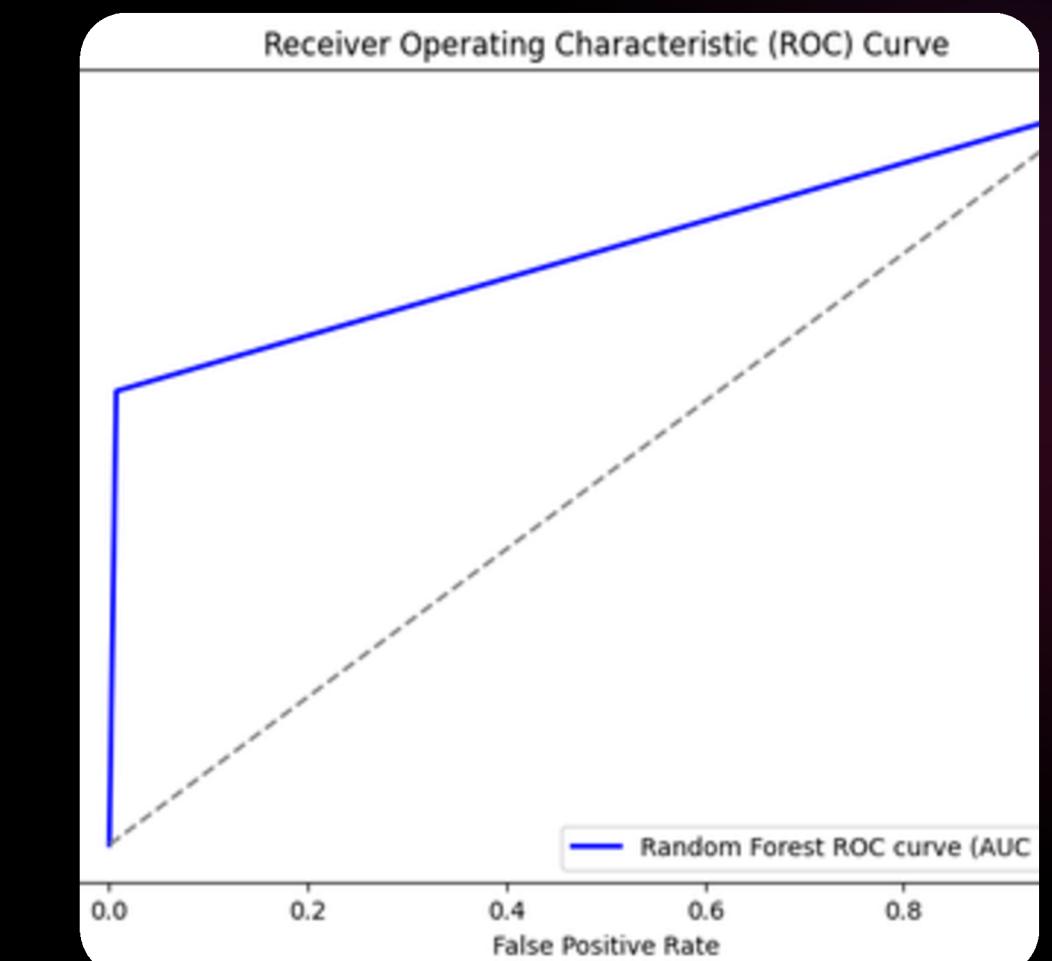
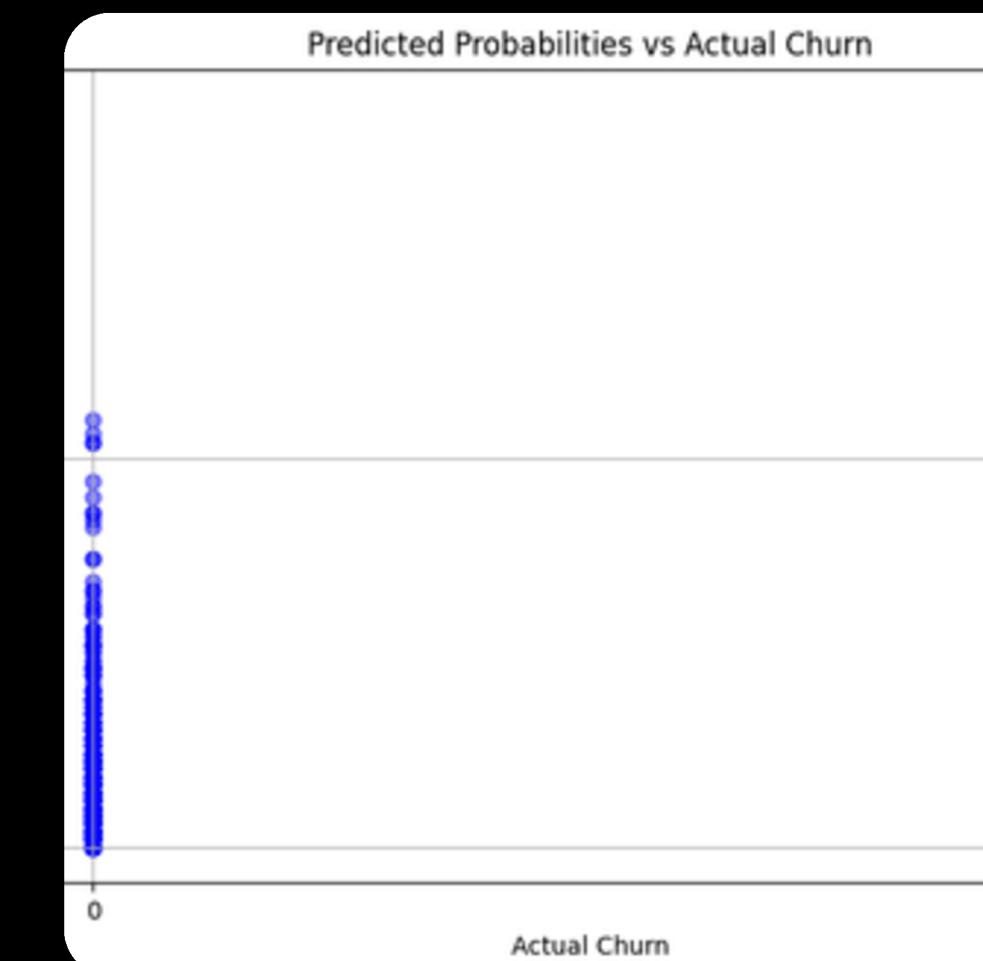
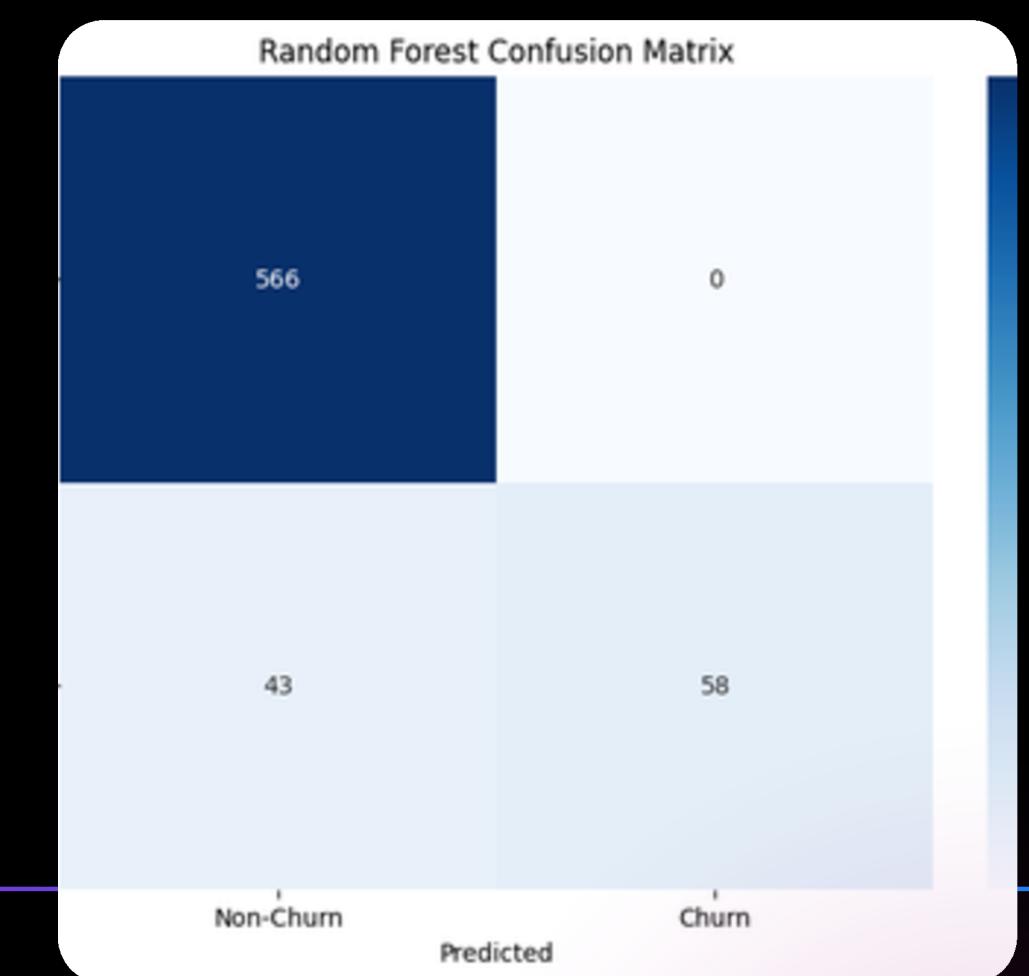
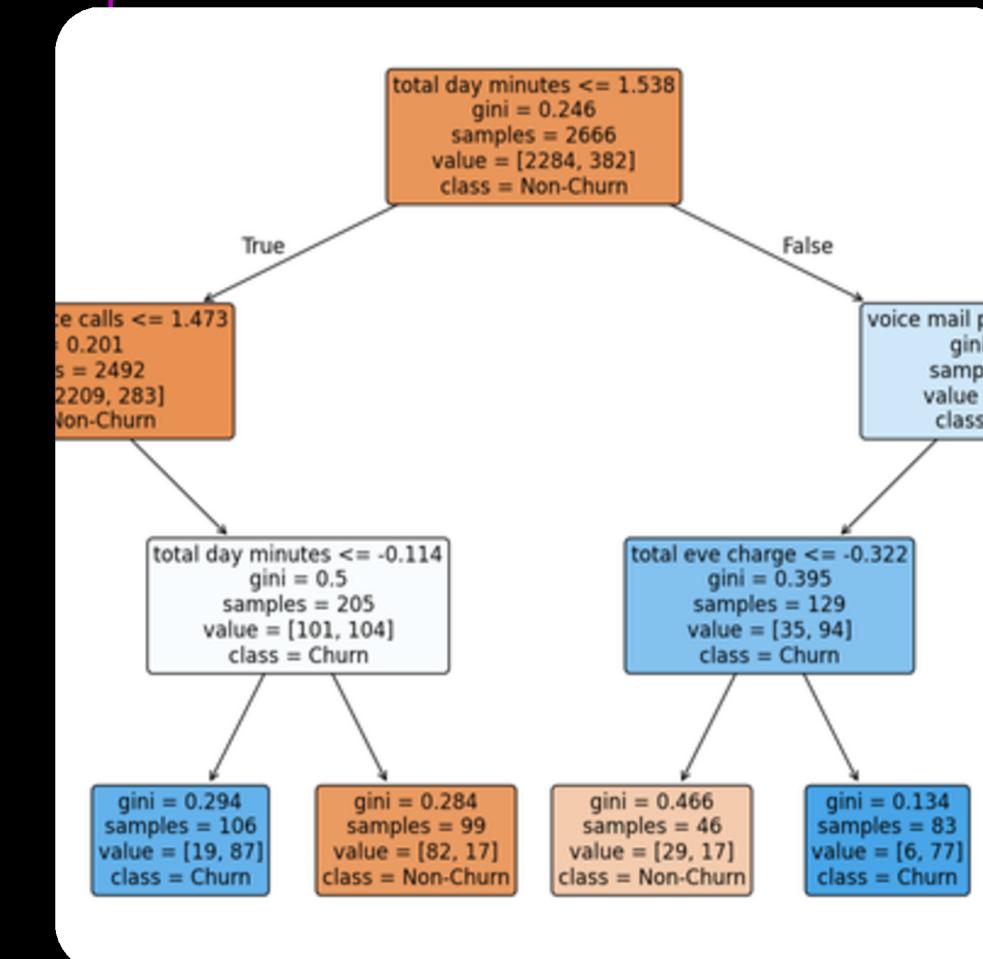
# Exploratory Data Analysis (EDA)



- Key Insights:
  - Customers with more customer service calls are more likely to churn.
  - Lower total day minutes correlated with higher retention.
  - Customers with international and voice mail plans tend to churn more.
- Data Imbalance:
  - Most customers do not churn, resulting in an imbalanced dataset (non-churners are much more frequent than churners).
- Visualizations:
  - Churn vs non-churn distribution.
  - Correlation between usage patterns (e.g., total day minutes) and churn.

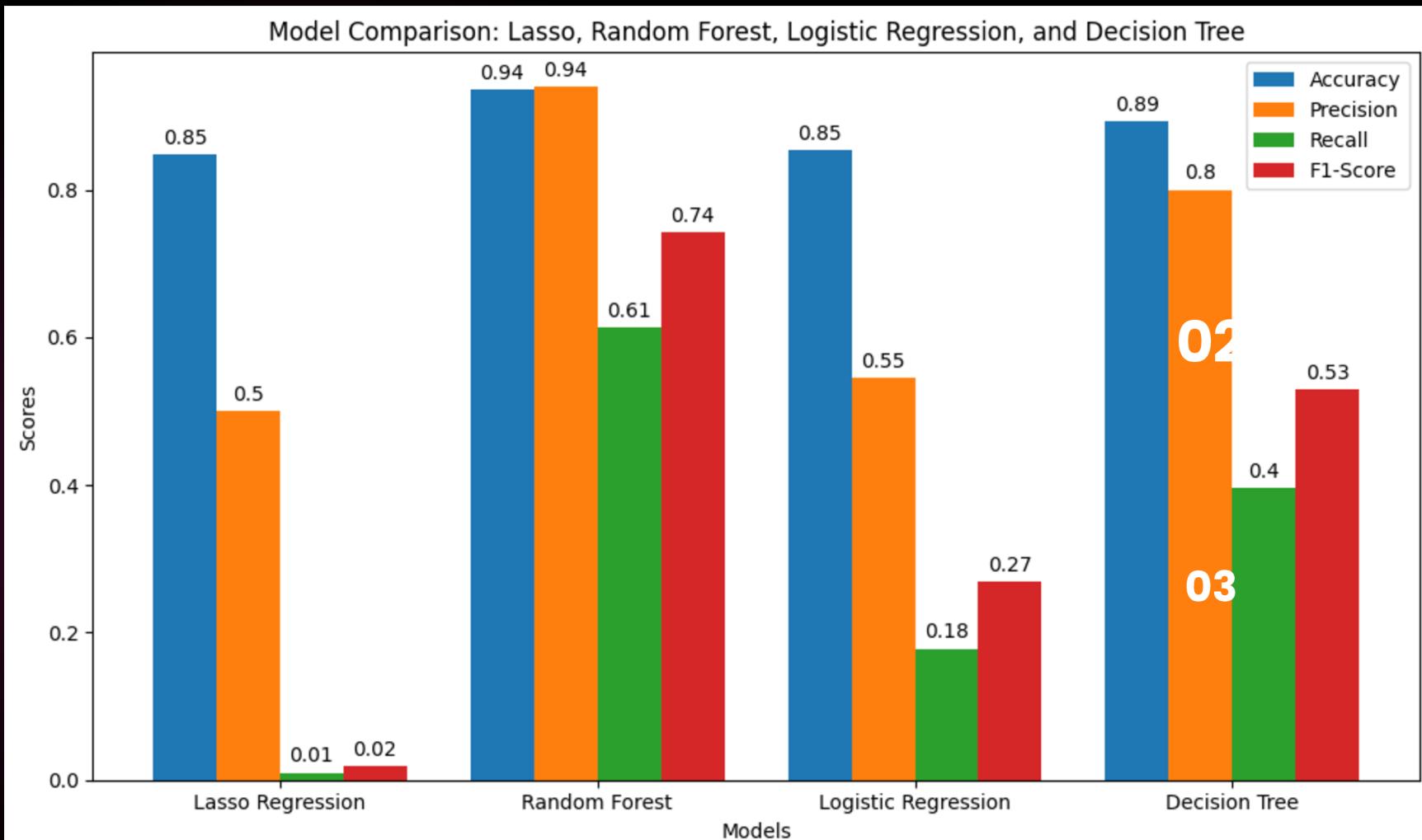
# Models Tested:

- Logistic Regression: A basic linear model for binary classification.
- Random Forest: Ensemble model to improve prediction by using multiple decision trees.
- Lasso Regression: Linear regression with regularization to reduce overfitting.
- Decision Tree: Tree-based model for decision-making.
- Evaluation Metrics:
  - Accuracy, Precision, Recall, F1-Score, AUC (ROC curve).
  - We want the model to predict accurately while not missing too many churners (minimize false negatives).



# Model Performance Comparison

- Lasso Regression: A linear model with regularization that reduces overfitting. However, it showed poor results with low recall (0.01) and low F1-Score (0.02).
- Random Forest: An ensemble model with high accuracy (93.55%), perfect precision (1.00), and a balanced recall (0.57). It performed the best overall.
- Decision Tree: High precision (0.80) and moderate recall (0.40). It was more interpretable than Random Forest but less effective.
- Logistic Regression: A basic model with moderate accuracy (85.31%), but low recall (0.18), meaning it misses many churners.
- Random Forest outperforms all other models, providing the best balance between accuracy, precision, and recall.
- Logistic Regression misses many churners (low recall), while Random Forest captures more churners without sacrificing precision.



# Decision Tree Insights



## Key Features in Decision Tree:

- Total Day Minutes: Customers with lower day minutes are more likely to stay (non-churn).
  - Customer Service Calls: More calls result in a higher risk of churn.
  - Service Plans: Customers with an international plan or voice mail plan are more likely to churn.
- 
- Visualizing the Tree:
  - Show the decision splits based on Total Day Minutes and Customer Service Calls, emphasizing how the model identifies churn.
  - Value to Stakeholders:
  - Helps decision-makers identify key customer segments that are at higher risk of churn, allowing for targeted actions.



01

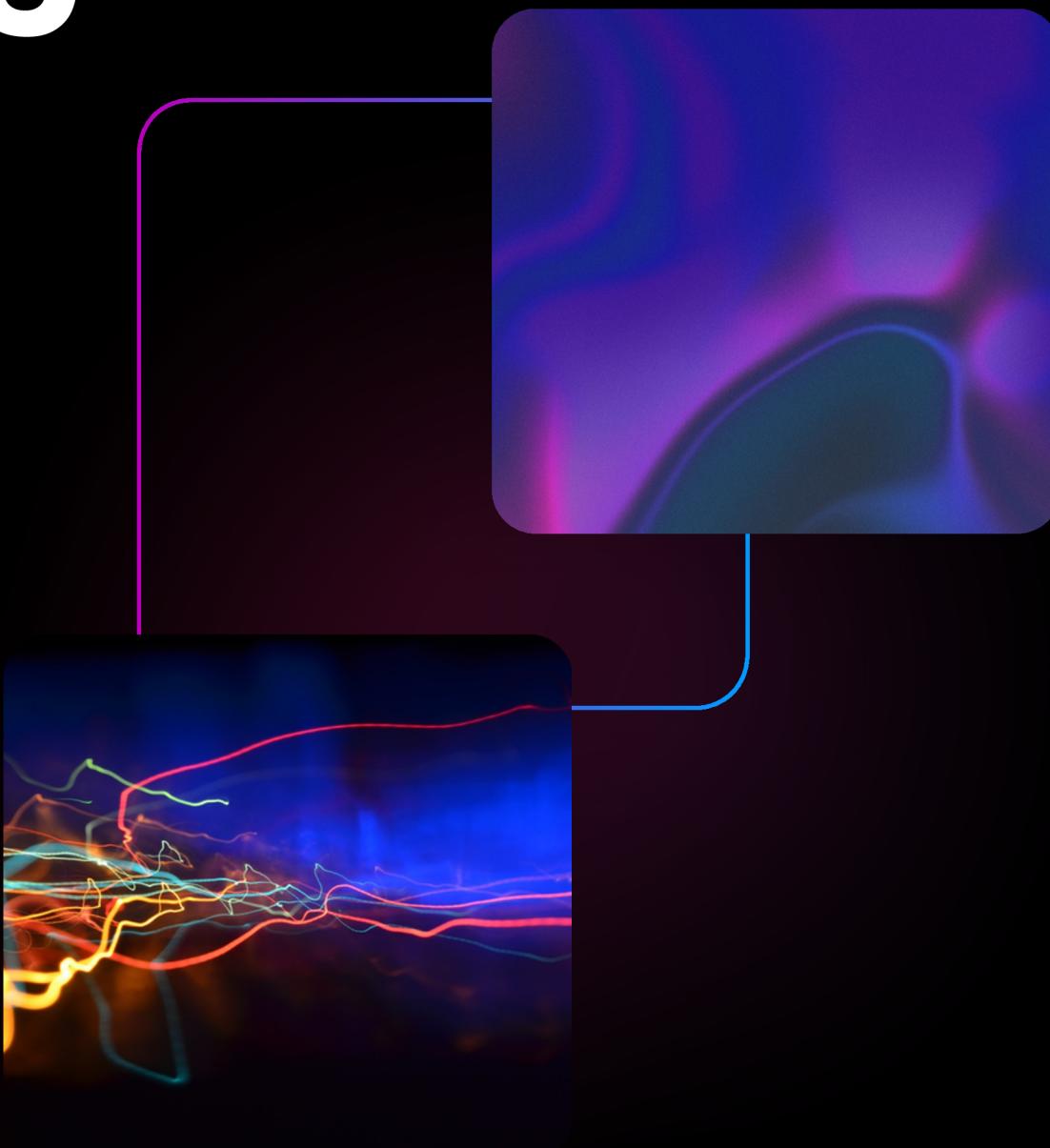
02

03



# Predictive Insights and Actionable Strategies

- Target High-Risk Customers:
  - Random Forest predictions can flag customers with a high churn probability. Focus retention efforts on this group.
- Improve Customer Service:
  - If customers make frequent customer service calls, improve the service quality to prevent churn.
- Reevaluate Service Plans:
  - Consider adjusting the international and voice mail plans as they appear to be associated with higher churn.
- Personalized Retention:
  - Offer tailored plans or discounts to high-risk customers to reduce churn.



01

02

03

# summary

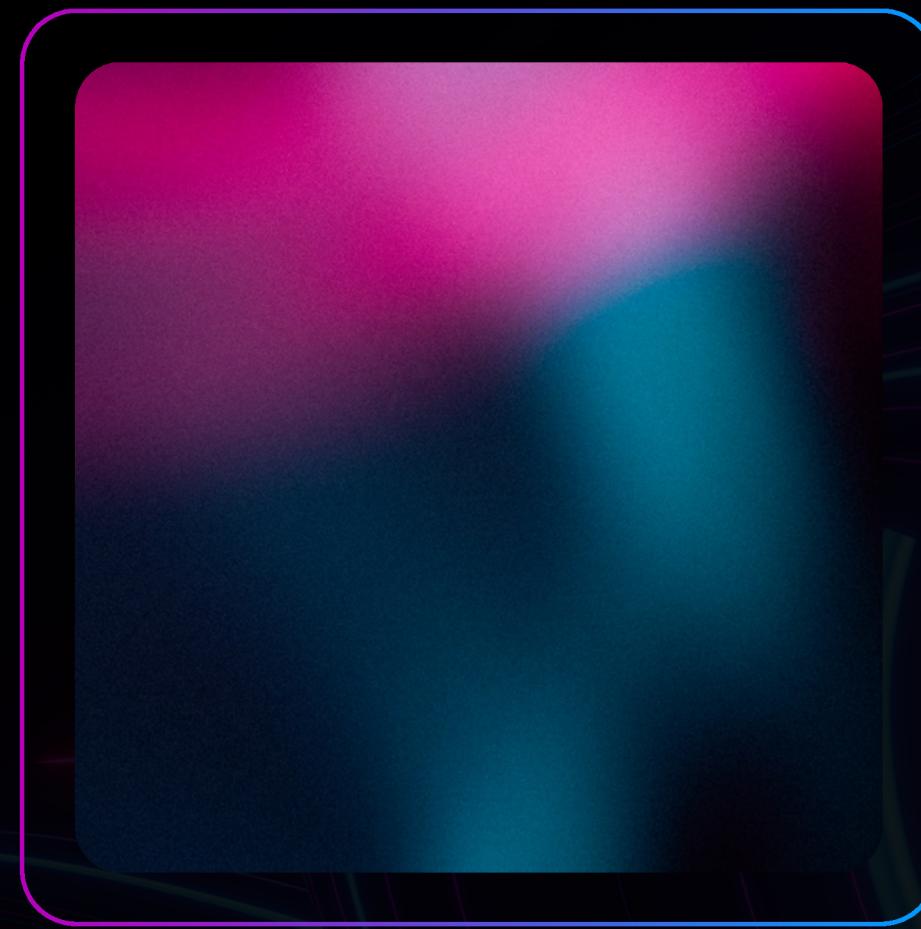


- Random Forest is the best model for predicting churn. The model offers accurate predictions and identifies the right customers for retention.
- Business Impact:
  - By focusing on high-risk customers, SyriaTel can proactively reduce churn and increase customer lifetime value.
- Next Steps:
  - Implement the churn prediction model into SyriaTel's CRM system.
  - Begin targeted retention campaigns for high-risk customers.
- Closing Remarks:
  - "By leveraging machine learning to predict churn, SyriaTel can save money and foster long-term customer relationships."



# Thank You

FOR YOUR ATTENTION



01

02

03