SYRIATEL CHURN PREDICTION

OVERVIEW

- Problem Statement
- Business Understanding
- General Objectives
- Data Understanding
- Data Preparation
- Data Analysis and Visualization
- Conclusion
- Contact

PROBLEM STATEMENT

SyriaTel is facing the challenge of customer churn, where customers discontinue their services. This churn results in lost revenue and increased customer acquisition costs.

My project aims to address this issue by using machine learning to build a customer churn prediction model.

BUSINESS UNDERSTANDING

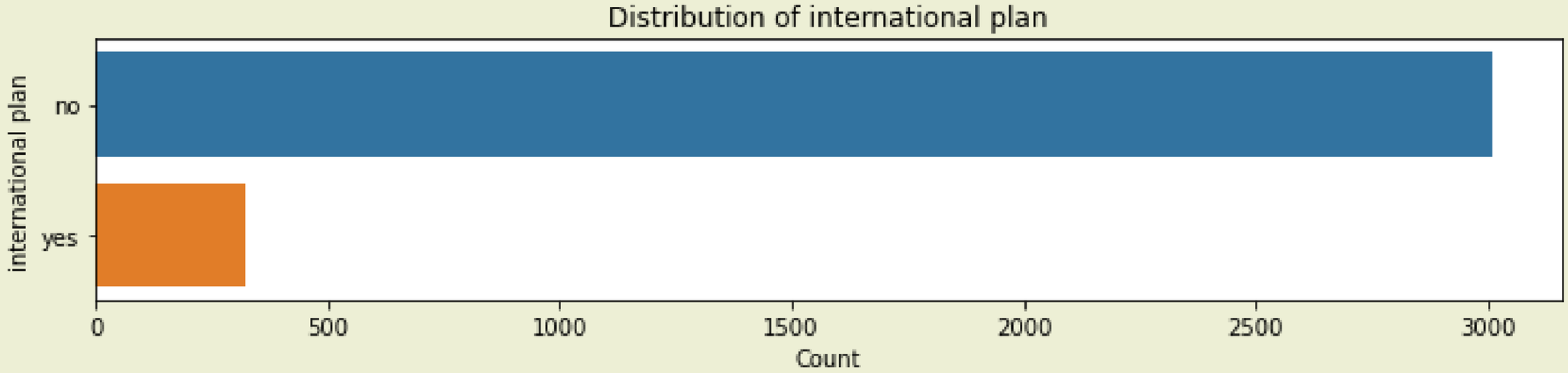
Syriatel is a telecommunications company that operates in Syria. It is one of the two major mobile network operators in the country.

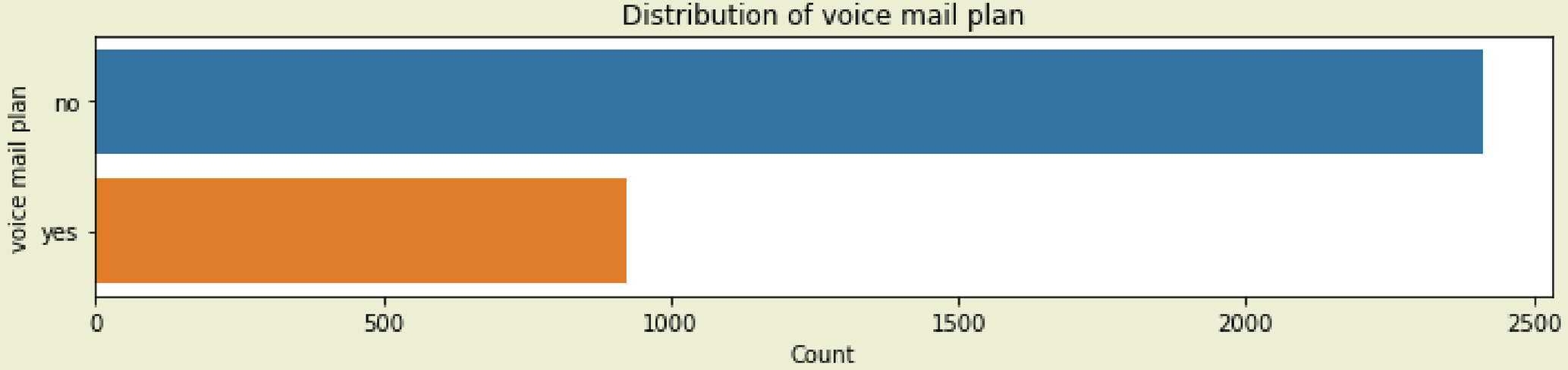
SyriaTel targets individual consumers in Syria, offering mobile subscriptions with voice, text, and data options to suit various usage patterns (personal, professional). Subscription-Based Revenue: SyriaTel's primary revenue stream comes from customer subscriptions to their mobile network plans. These plans offer a set amount of voice minutes, text messages, and data for a fixed monthly fee.

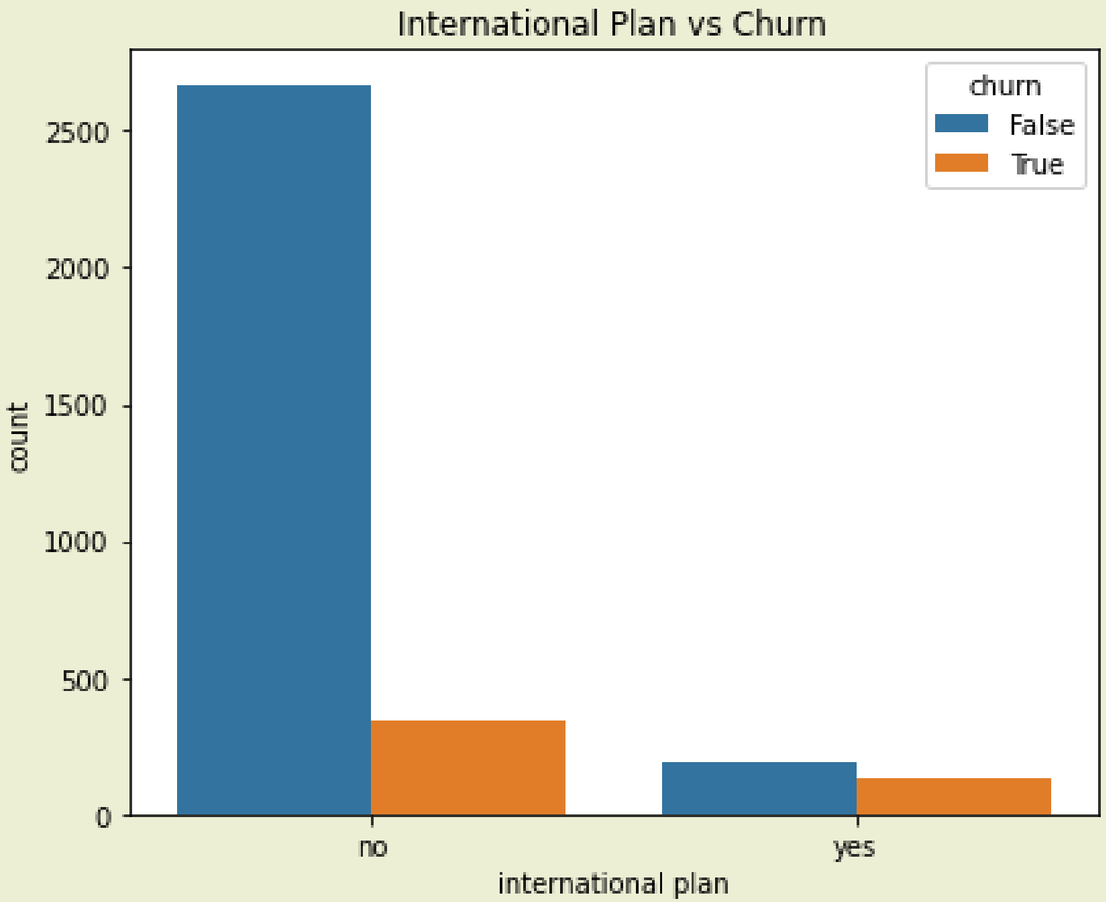
OBJECTIVES

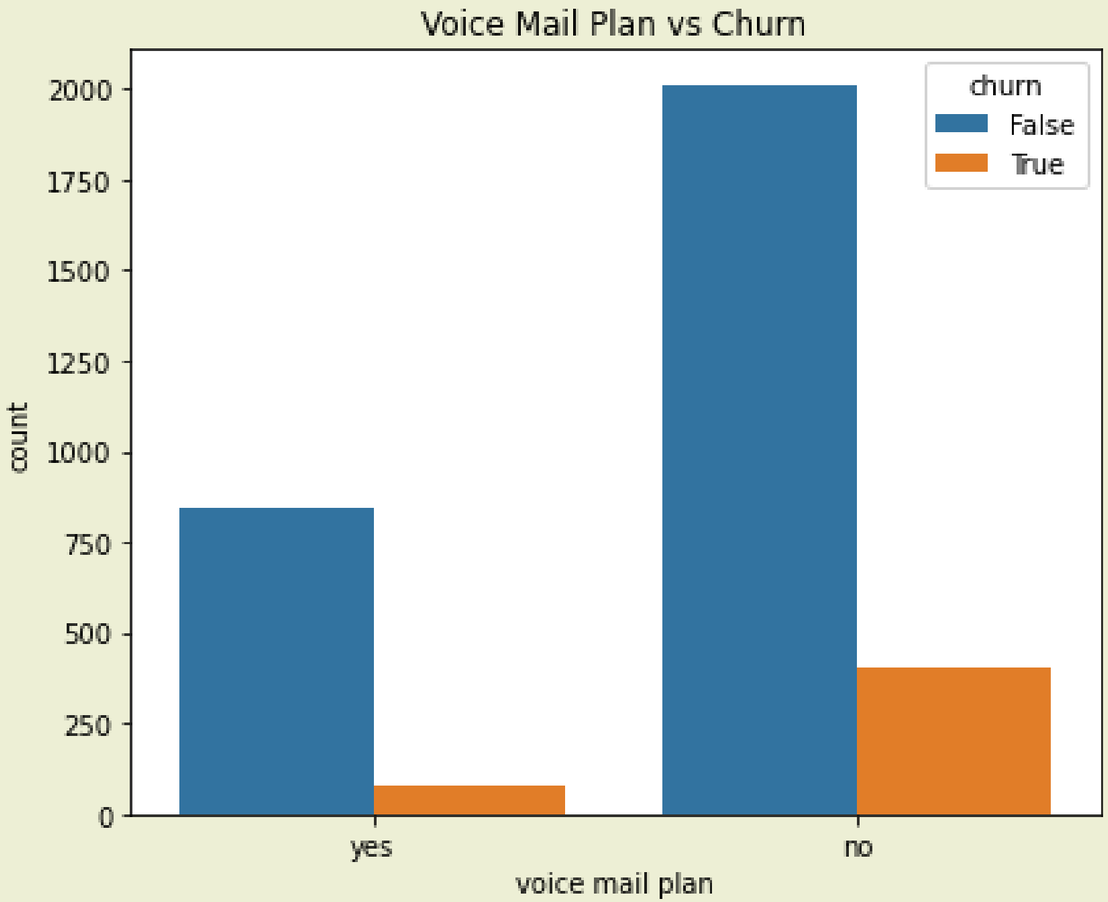
- To understand the dataset's characteristics.
- To create a training and testing set for model evaluation.
- To develop a simple baseline model to establish a performance benchmark.
- To develop more complex models to improve performance.
- To tune on the complex models using GridSearchCV.
- To assess the final model's performance and interpret the results.
- To offer actionable insights and recommendations.

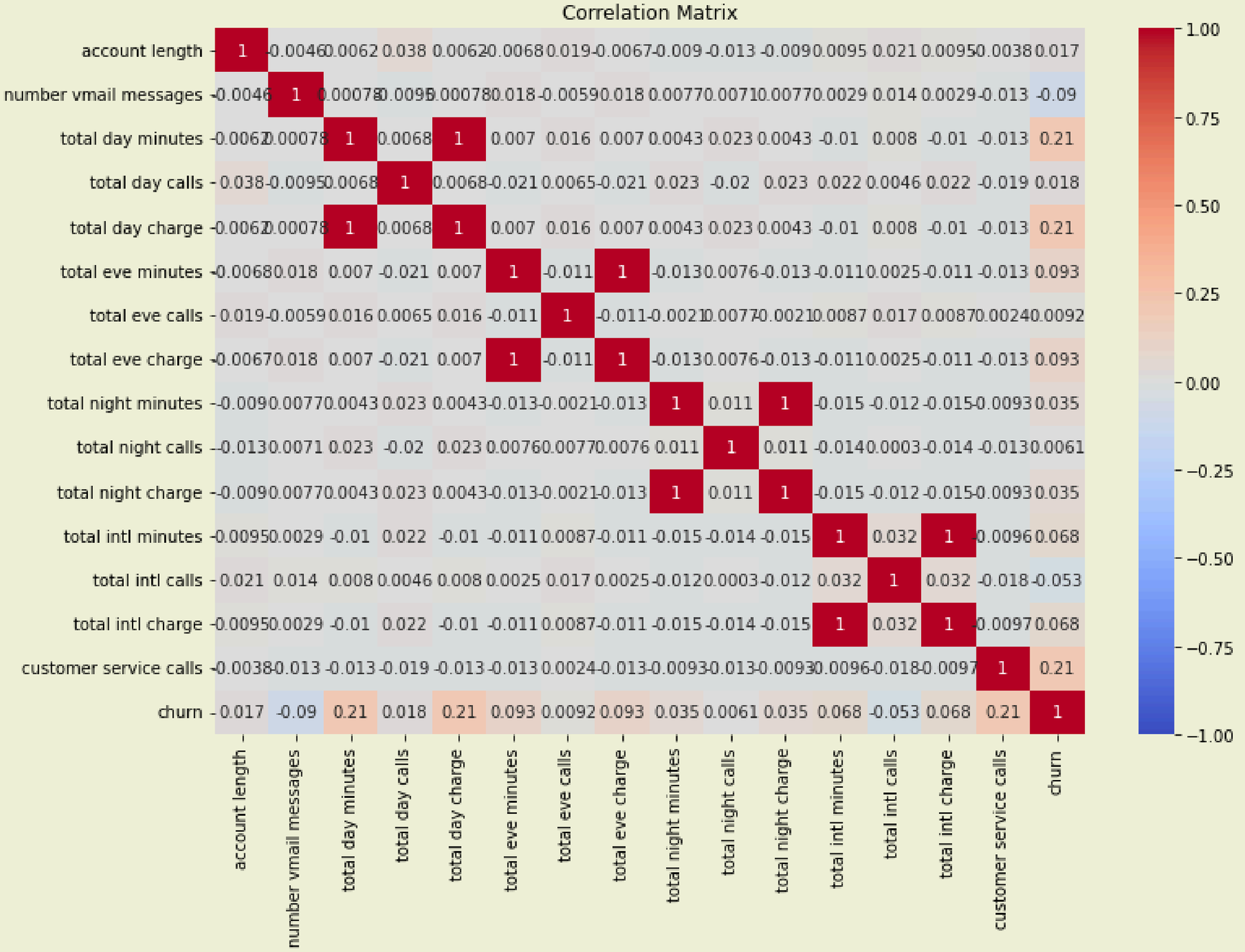
EXPLORATORY DATA ANALYSIS






International Plan vs Churn
churn
False
True
count
2500
2000
1500
1000
500
0
no
yes
international plan

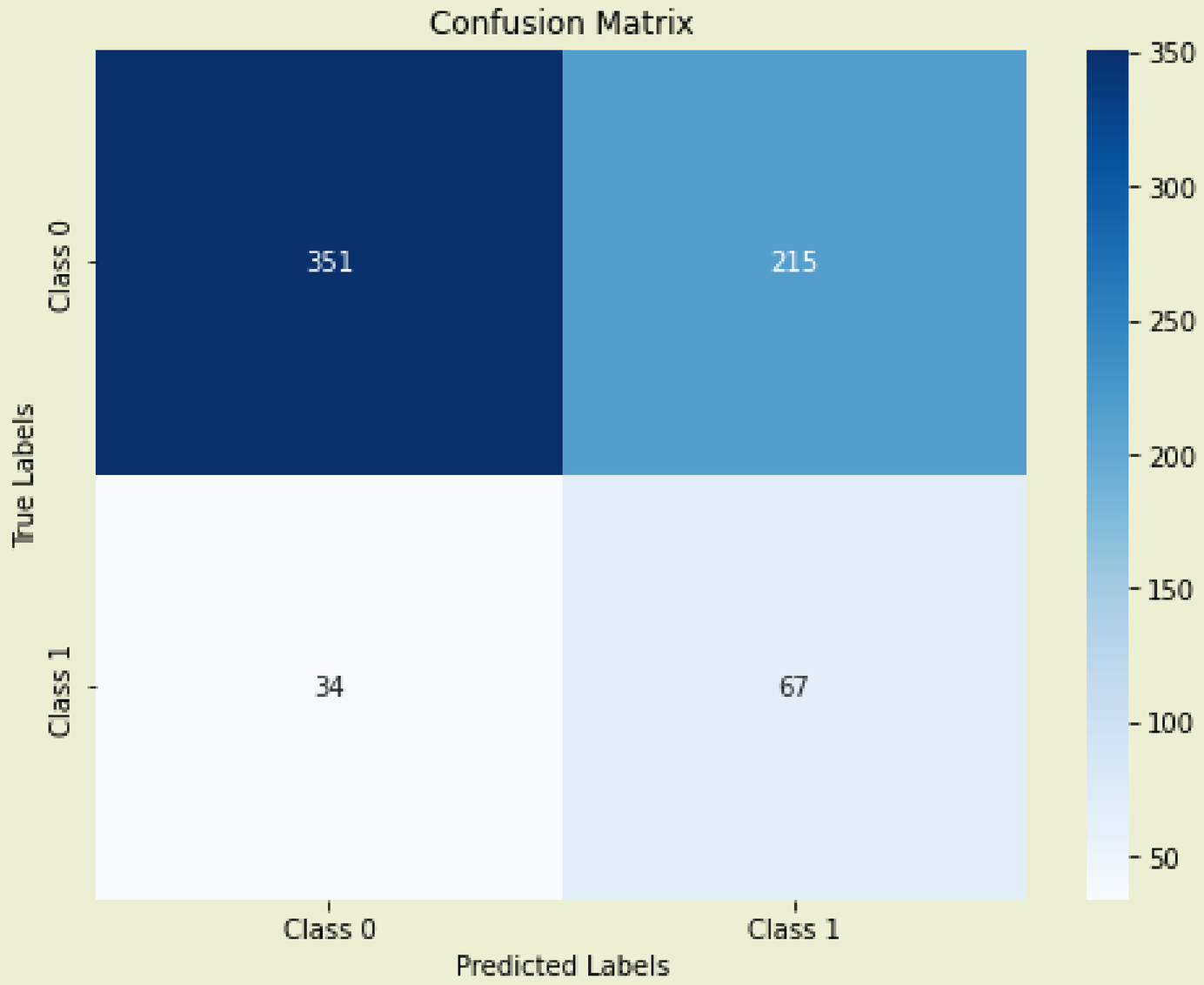

Voice Mail Plan vs Churn
churn
False
True
count
2000
1750
1500
1250
1000
750
500
250
0
yes
no
voice mail plan


Correlation Matrix
account length
number vmail messages
total day minutes
total day calls
total day charge
total eve minutes
total eve calls
total eve charge
total night minutes
total night calls
total night charge
total intl minutes
total intl calls
total intl charge
customer service calls
churn
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

PREDICTION MODELS

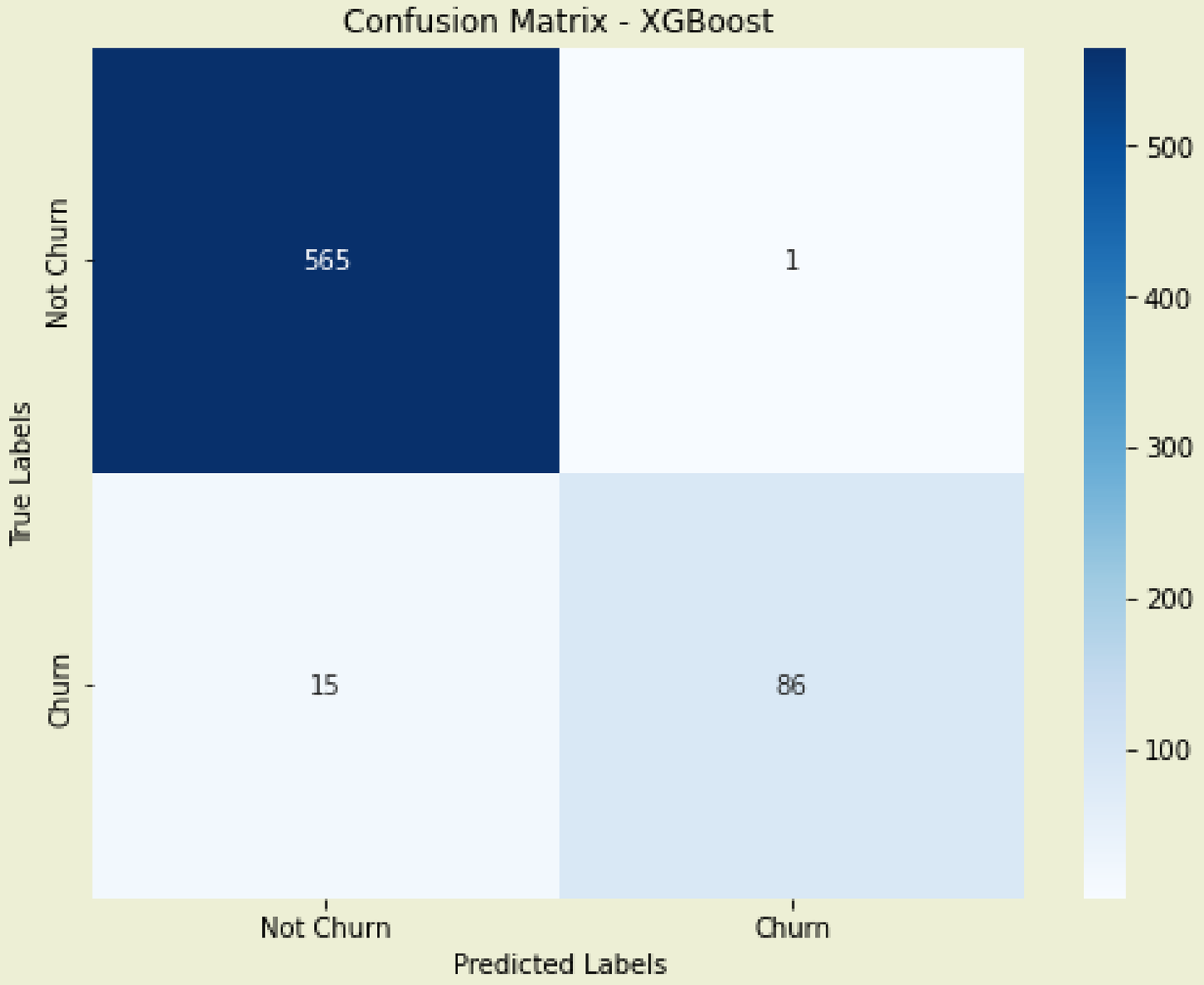
Base Model - Linear Regession

Accuracy: 62.6%



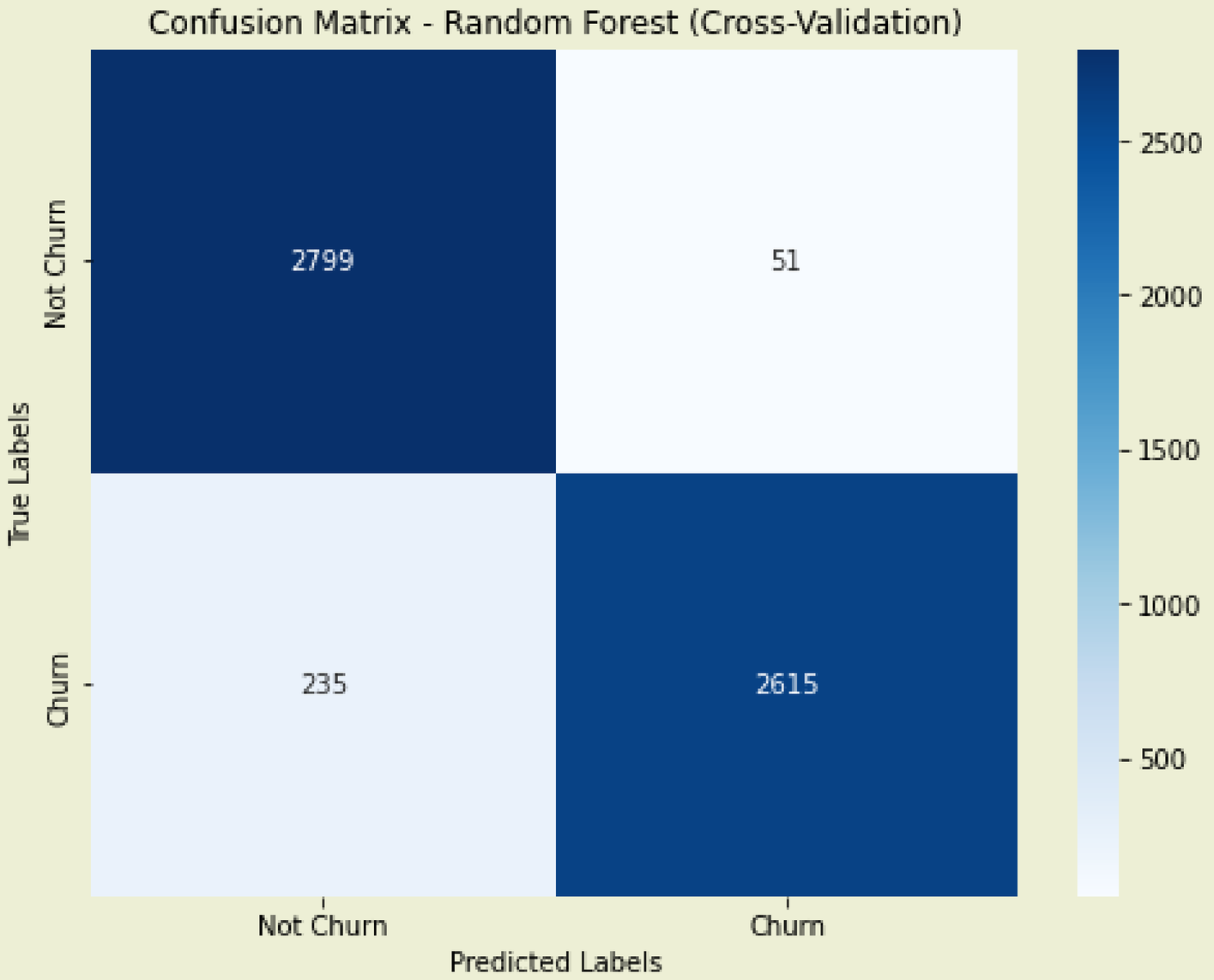
XGBoost Model

Accuracy: 97.6%

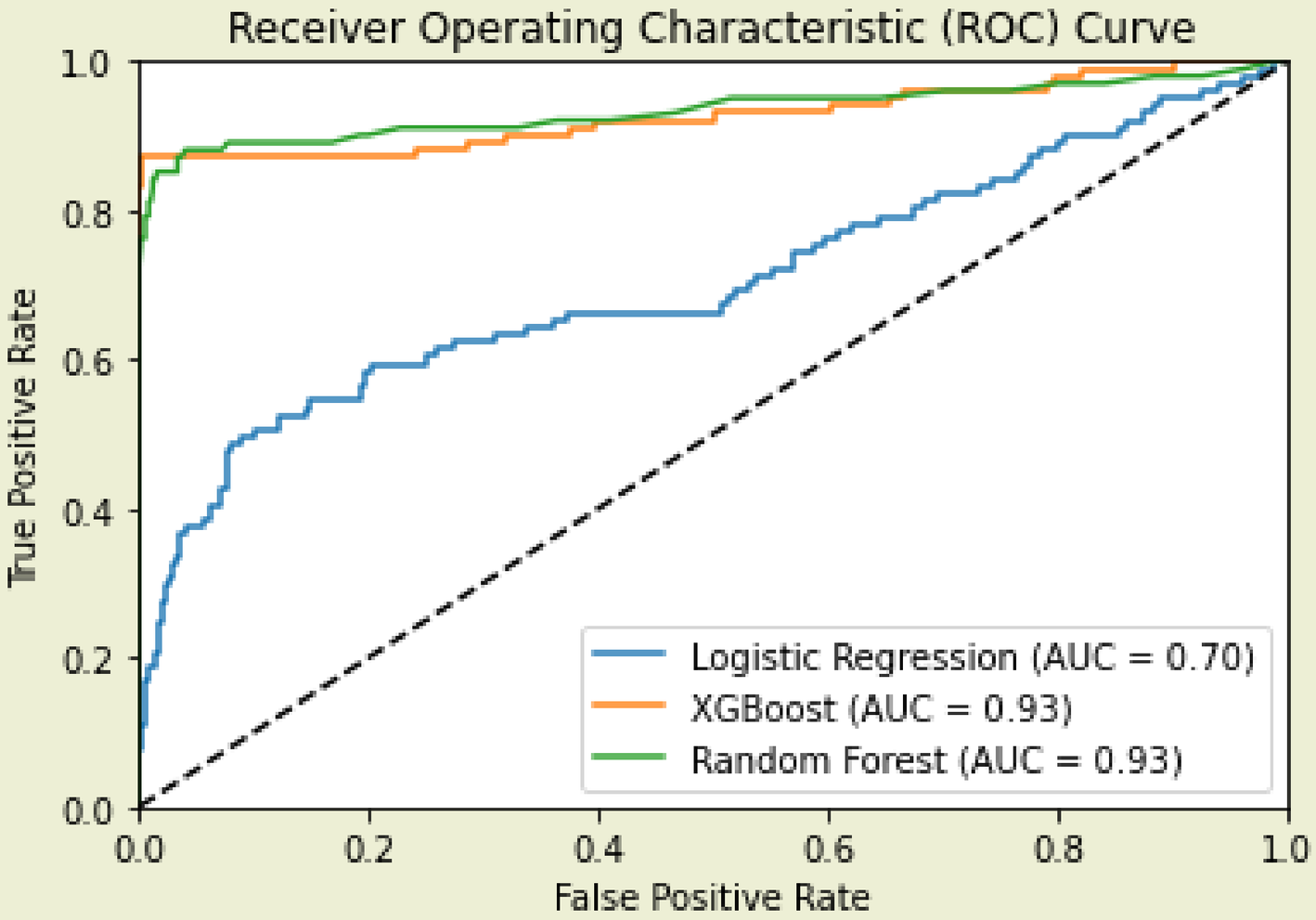


Random Forest Model

Accuracy: 95.0%



ROC CURVE


Receiver Operating Characteristic (ROC) Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (AUC = 0.70)
XGBoost (AUC = 0.93)
Random Forest (AUC = 0.93)

CONCLUSION

XGBoost demonstrates the best overall performance across multiple metrics. It consistently outperforms both Logistic Regression and Random Forest, indicating its effectiveness in predicting churn.

RECOMMENDATIONS

- Focus on improving customer service quality.
- Implement proactive measures such as follow-up calls.
- Review the pricing strategy and provide transparent billing to mitigate churn caused by the high total charges.
- Offer attractive international calling plans and promotions so as to retain customers who make frequent international calls.

Questions?

Thank You

email: wanjirudoreen13@gmail.com
github link: DoreenMolly

Presented By : Doreen Wanjiru