SyriaTel Customer Chun

A customer churn predictive modell

By: JOHN ELVIS

BUSINESS PROBLEM

Business Problem: "Reducing Customer Churn for SyriaTel Telecom"

Background: SyriaTel, a leading telecommunications company, is facing a significant challenge with customer churn. Churn is when customers decide to terminate their subscriptions with SyriaTel, resulting in revenue loss. To address this issue, SyriaTel aims to build a predictive model to identify customers who are likely to churn. By proactively targeting these at-risk customers with retention strategies, SyriaTel hopes to reduce churn rates and retain valuable customers.

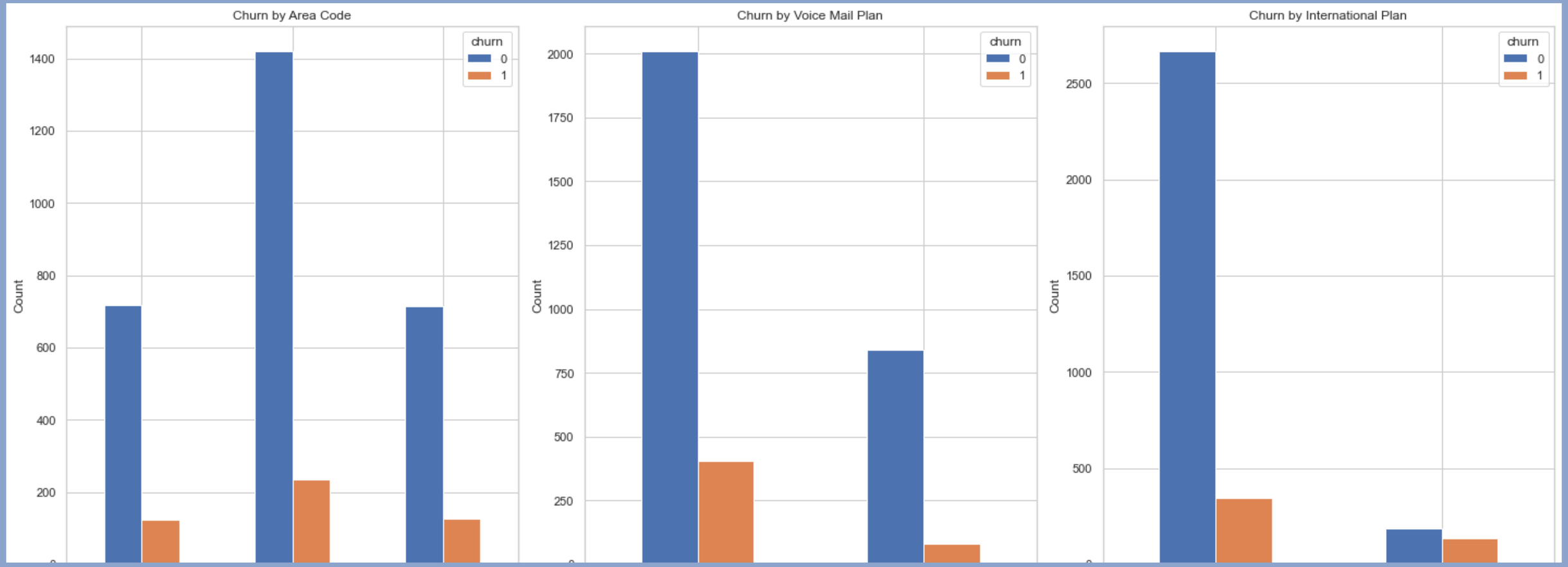
Problem Statement: "Can we predict customer churn for SyriaTel and identify the key factors driving churn, enabling the company to implement effective retention strategies?"

Key Objectives

- To build a Churn Prediction model that can accurately predict customers who will churn based on the information available in the dataset.
- To identify the predictive patterns/features that are important for predicting customer churn.

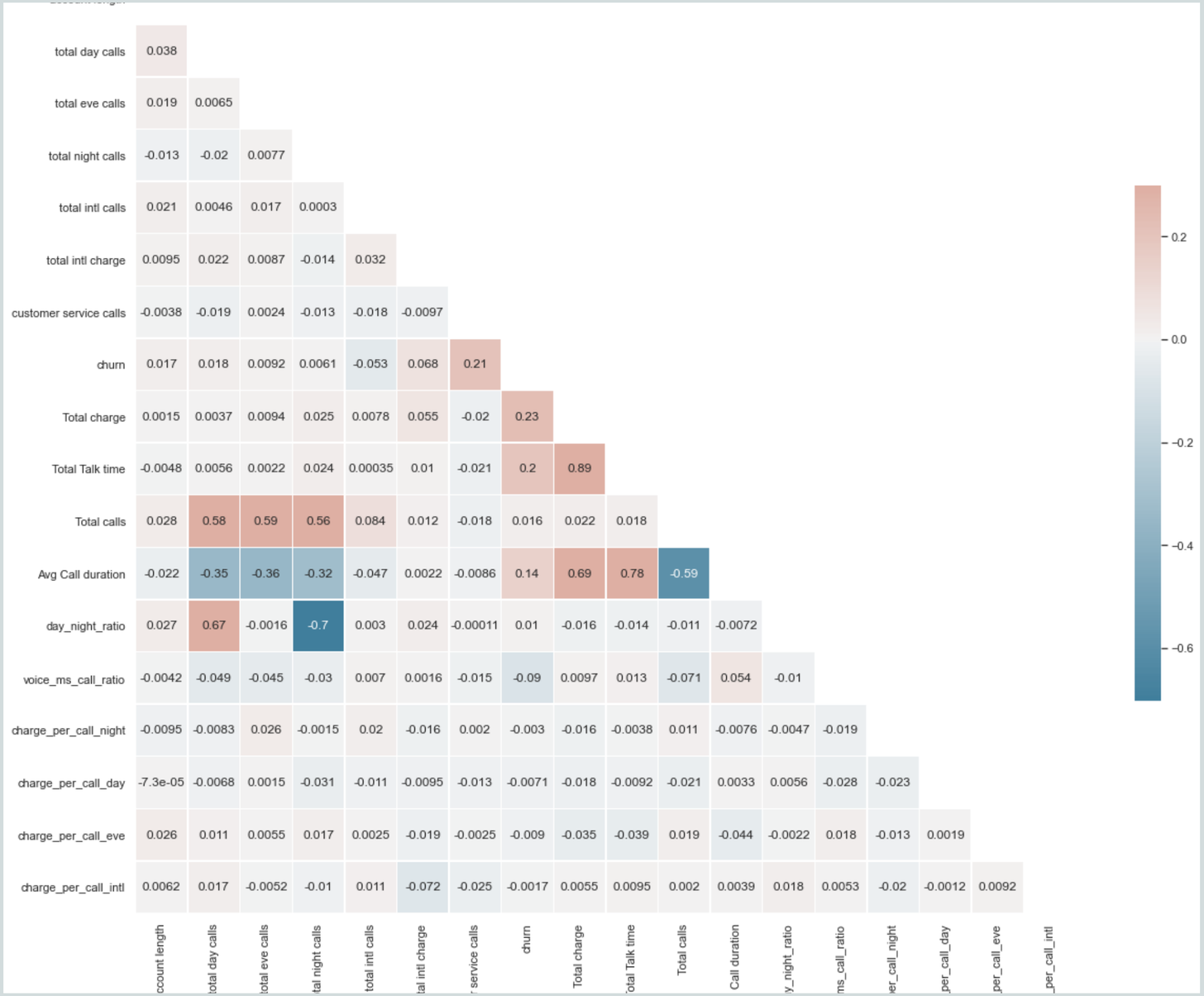
The business problem aligns with SyriaTel's goal of retaining customers and reducing revenue loss due to churn, emphasizing the importance of data-driven decisions in the telecommunications industry.

EXPLORATORY DATA ANALYSIS

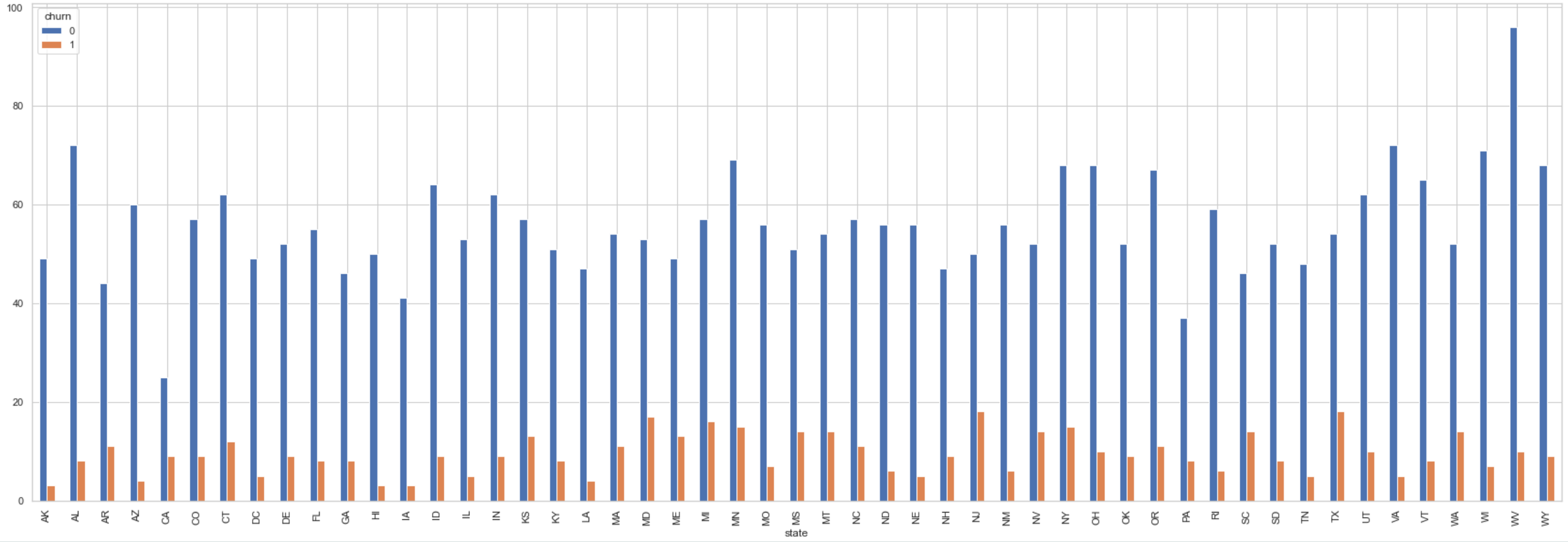


Here, we explore Churn rate by area code, voice mail plan and international plan.

Co-relation of the various features.

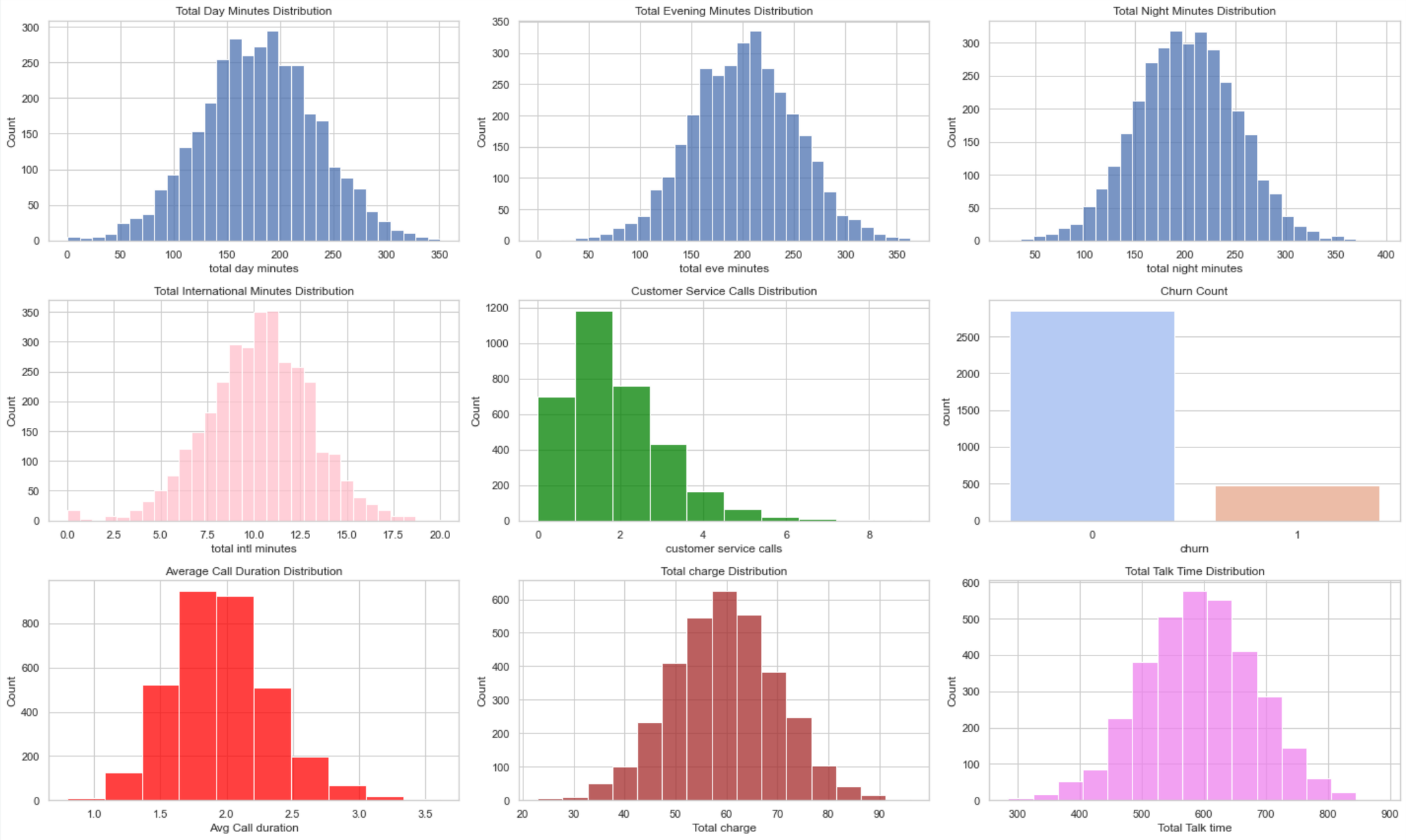


Churn By State



Some states have relatively higher churn rates like WV, VT, NY, OH with a significant number of churned customers (churn 1) while other states have lower churn rates like AR, AZ, CA, CO with a higher count of customers who did not churn (churn 0)

Distribution of Numerical features



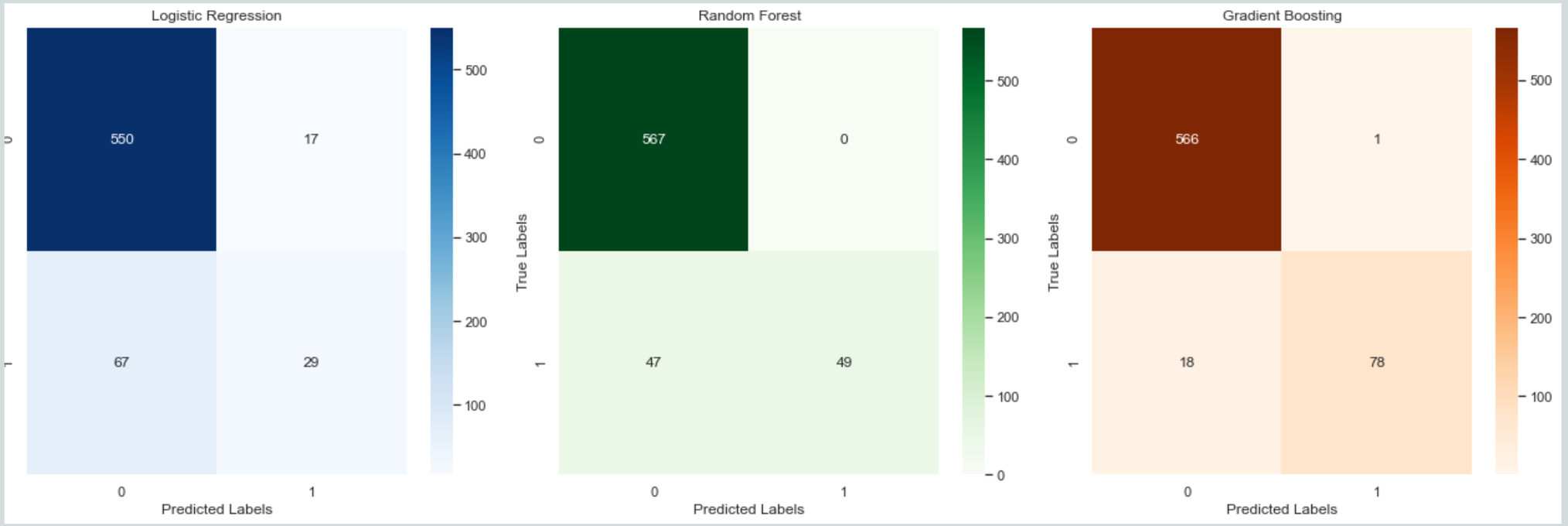
Modelling

For prediction, we used Logistic regression, random forest, and gradient boosting. They were trained using a resampled dataset and optimized with hyperparameter tuning. Below is the results after tuning:

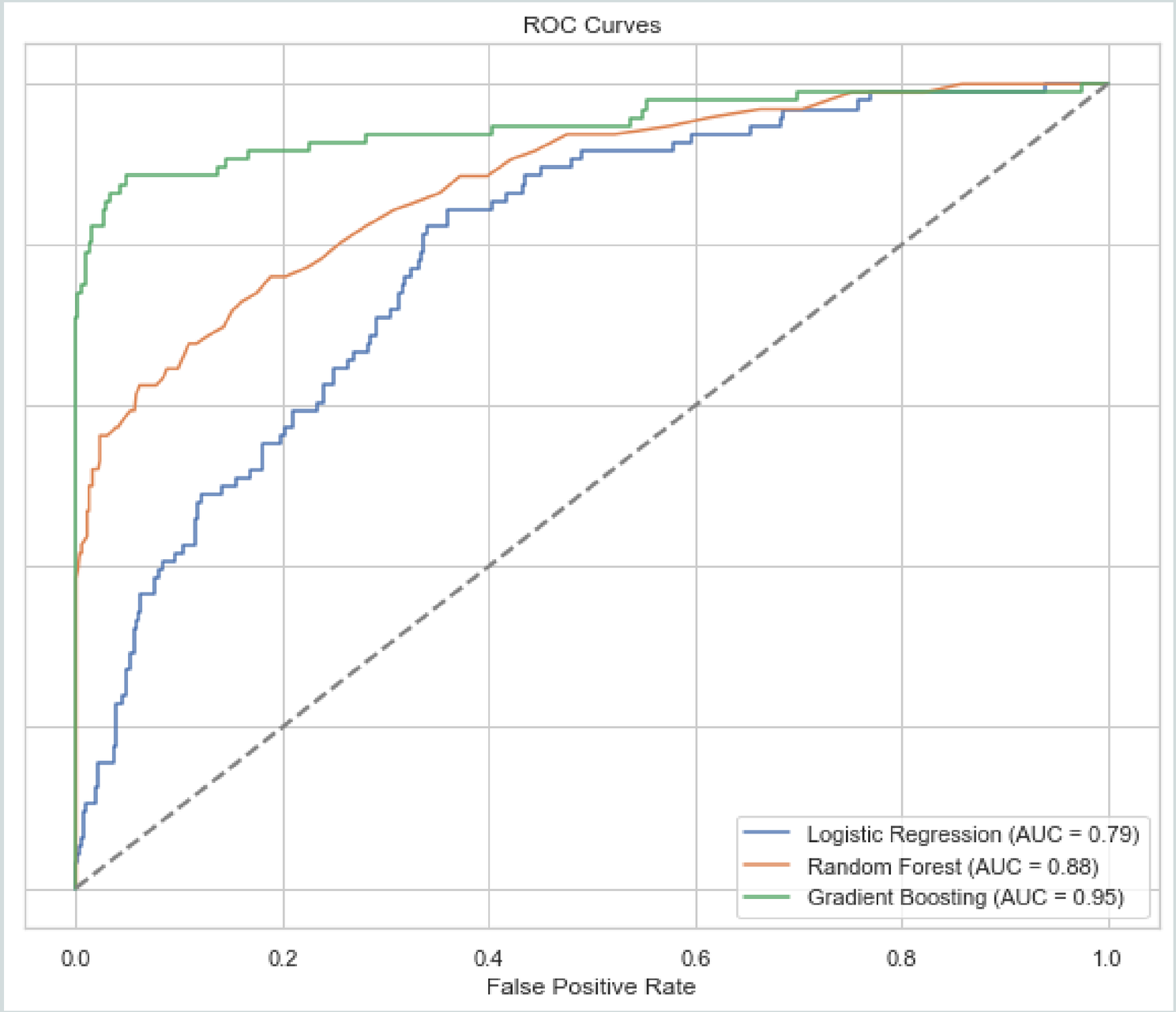
		Accuracy	Precision	Recall	F1-score
Logistic Regression	Test	0.767722	0.325301	0.562500	0.412214
	Train	0.770857	0.327613	0.544041	0.408958
Random Forest	Test	0.904977	0.714286	0.572917	0.635838
	Train	1.000000	1.000000	1.000000	1.000000
Gradient Boosting	Test	0.960784	0.897727	0.822917	0.858696
	Train	0.975840	0.976331	0.854922	0.911602

Model Evaluation

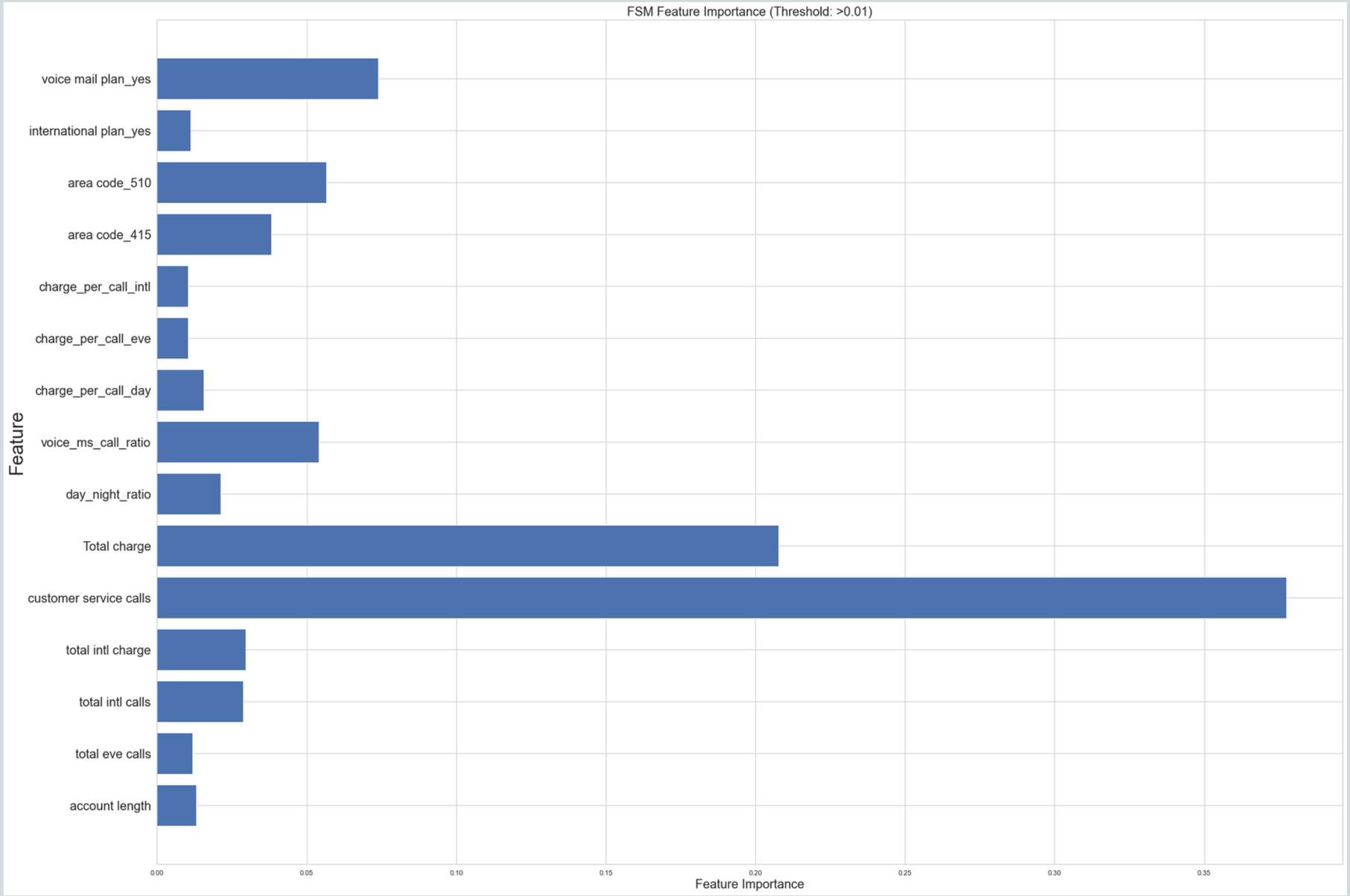
Gradient Boosting model: performs excellently with an accuracy of approximately 96.08%. It has high precision (89.77%) and recall score(82.29%), which indicates it can correctly identify a significant portion of the positive cases while minimizing false positives. The F1-score is 85.87%, indicating a well-balanced performance.



ROC curves for the tuned models



Feature importance selection graph



Observations , Findings & Insights

Model Performance

- The churn prediction models seem to have reasonably good performance based on the metrics used (accuracy, precision, recall, and F1-score), with Gradient Boosting performing the best with an Accuracy of 96.07% and Recall of 82.2% after tuning.

- The model tuning for the Decision Tree model led to an improvement in accuracy and precision but a decrease in recall. This indicates a trade-off between correctly predicting positive instances and capturing all actual positive instances.

Recommendations

Based on our findings, the following is recommended:

1. **Review International Plan**: Given its importance in predicting churn, it would be beneficial to review the structure and pricing of the international plan to ensure it meets customer needs.
2. **Improve Customer Service**: The number of customer service calls is a strong predictor of churn. Efforts should be made to improve the customer service experience to reduce the likelihood of churn.
3. **Analyze Pricing Structure**: Customers with a higher total charge are more likely to churn. A review of pricing strategies and structures could help to ensure they are competitive and provide value to customers.

~THE END~