SYRIA TEL CUSTOMER CHURN ANALYSIS.

MORINGA SCHOOL

PHASE 3 PROJECT

THORNE MAKAU MUSEMBI

INTRODUCTION

In the fiercely competitive telecommunications industry, customer retention holds equal significance to customer acquisition. SyriaTel, currently grappling with escalating customer churn, aims to employ data-driven insights to comprehensively grasp and alleviate the factors contributing to customer attrition. This analysis navigates the intricacies of customer behaviour,

service utilization, and interaction patterns to unveil the underlying causes of churn. The goal is to pinpoint vulnerable customers and the elements influencing their departure, proposing precise strategies to boost satisfaction, nurture loyalty, and ultimately curtail churn. SyriaTel envisions that this initiative will fortify its market standing, redefining its customer relationship management approach and setting a new benchmark for telecommunications service excellence.

Problem Statement

SyriaTel faces a critical challenge in retaining its customer base, with an upward trend in customer churn. This churn not only impacts the company's revenue but also escalates the expenses associated with acquiring new customers to maintain market share. The identification of factors contributing to churn and the prediction of at-risk customers empower SyriaTel to implement focused retention strategies, fostering customer loyalty and mitigating turnover.

This analysis scrutinizes the patterns and predictors of customer churn at SyriaTel, a prominent telecommunications provider. Leveraging customer usage data, service plans, and interaction history, we employ statistical analysis and machine learning models to discern primary drivers of churn. The study aims to segment the customer base, assess the impact of diverse service features on customer retention, and construct a predictive model to pinpoint at-risk customers. The analysis outcomes guide targeted interventions geared towards enhancing customer satisfaction and loyalty, reducing churn rates, and nurturing sustained business growth.

Objectives

1. Identifying Churn Factors: Understand the factors contributing to customer churn by analyzing data on customer behavior, service usage, and interaction patterns.

2. Predictive Modelling: Develop machine learning models to predict which customers are at risk of churning. This involves using historical data to create models that can anticipate future customer behavior.

3. Improve the model using hyper parameter tuning to increase its accuracy.

Data Understanding

- Of the 3,333 customers in the dataset, 483 have terminated their contract with SyriaTel. That is 14.5% of customers lost.

- The distribution of the binary classes shows a data imbalance. This needs to be addressed before modelling as an unbalanced feature can cause the model to make false predictions.
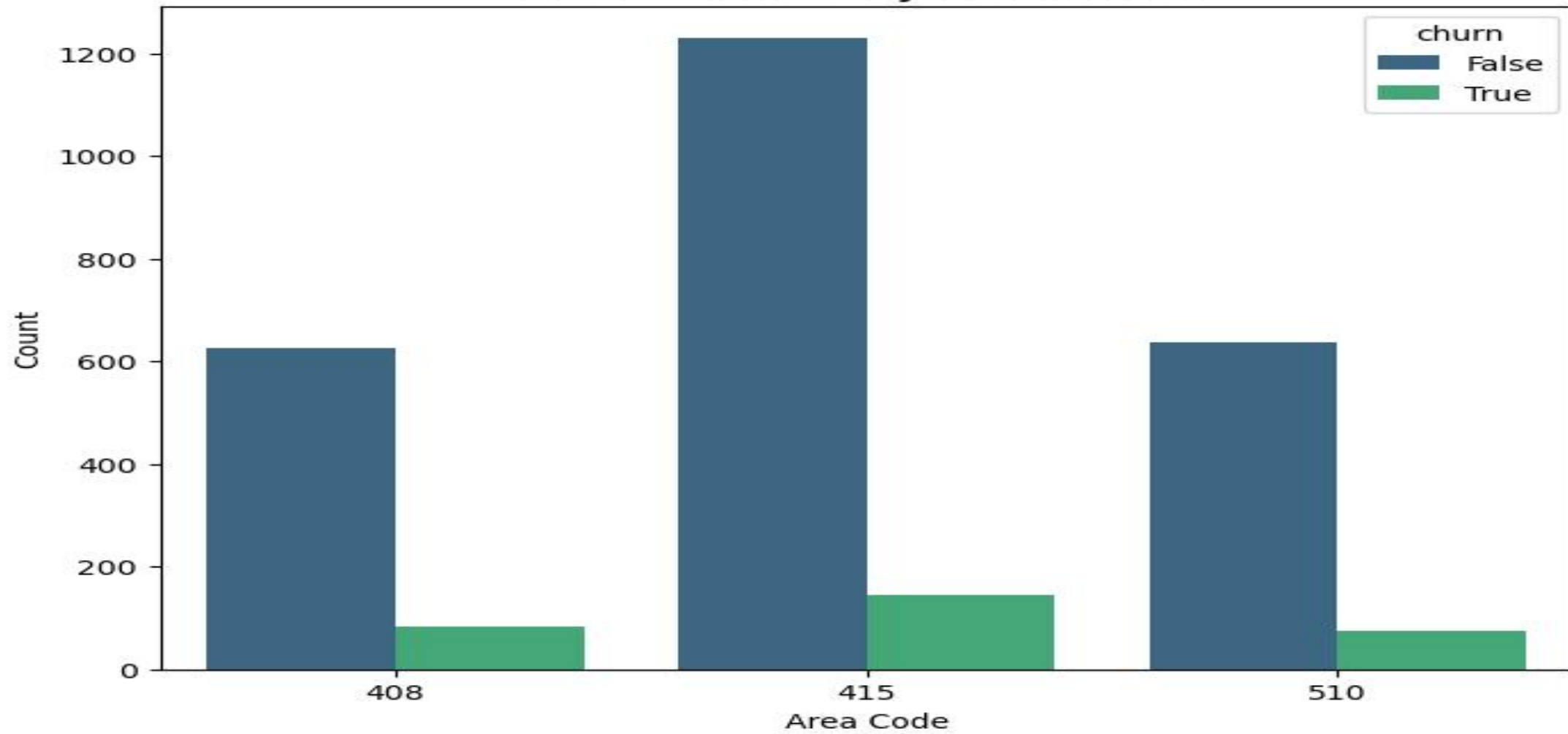

Churned Counts
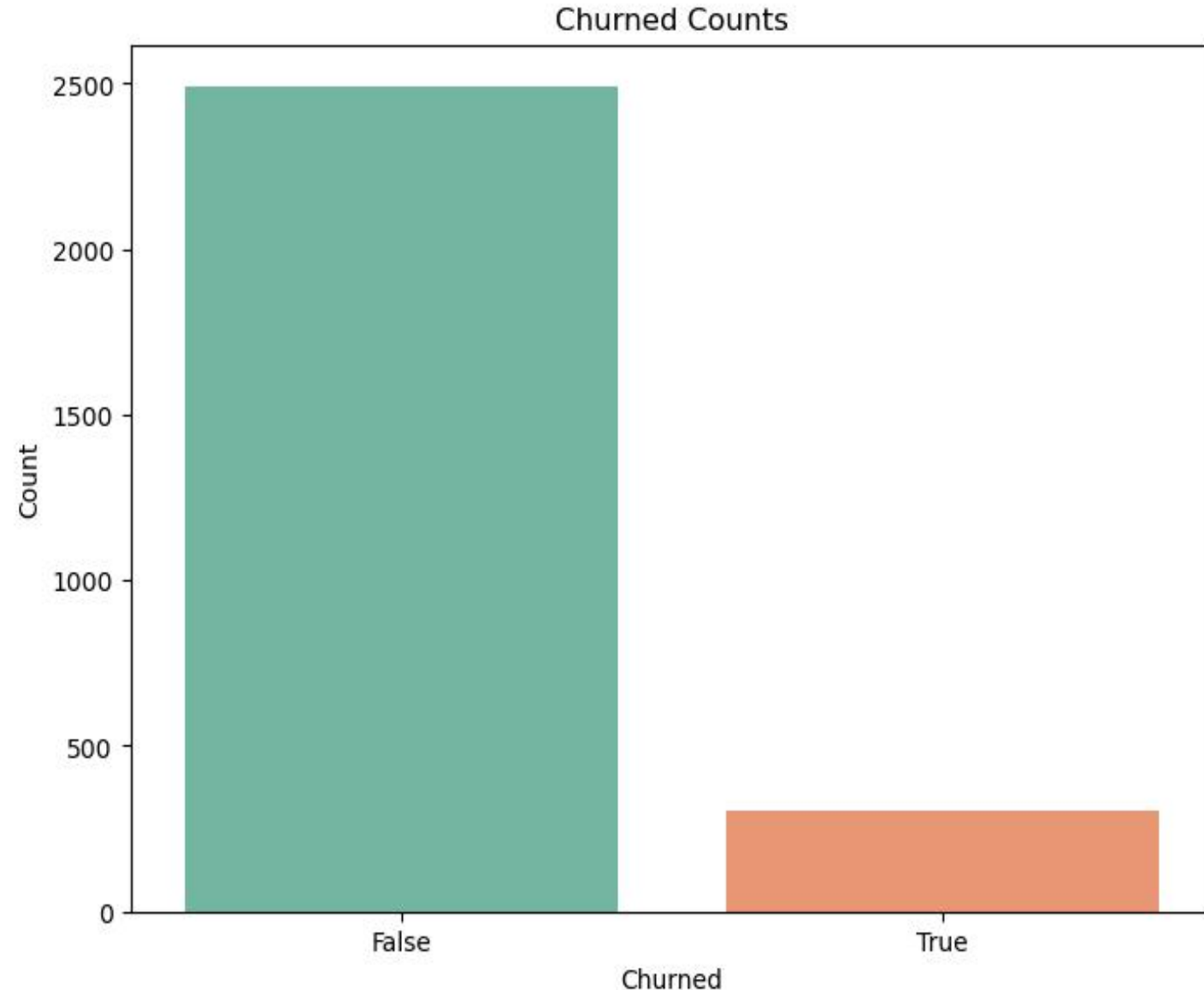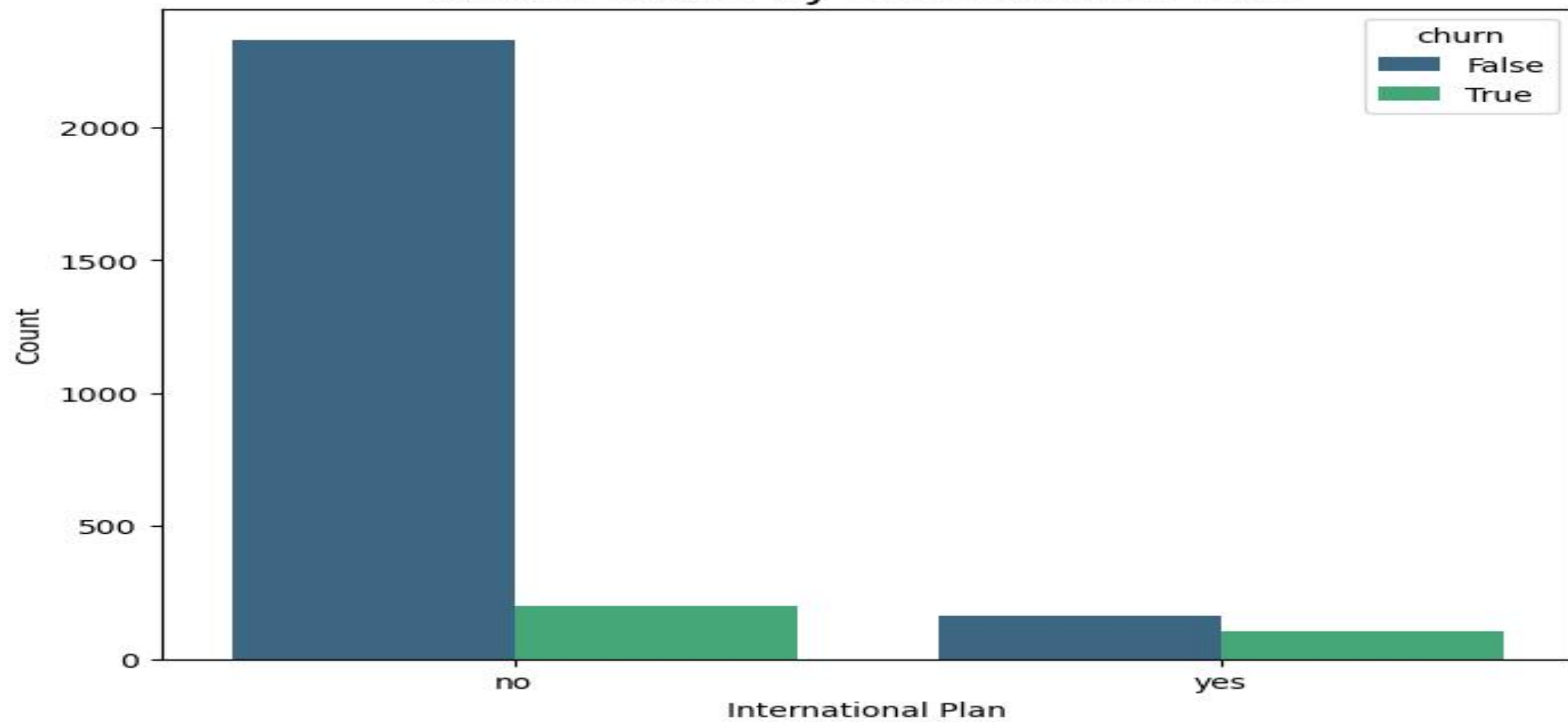
Churn plots for categorical Features.

Churn Count by State
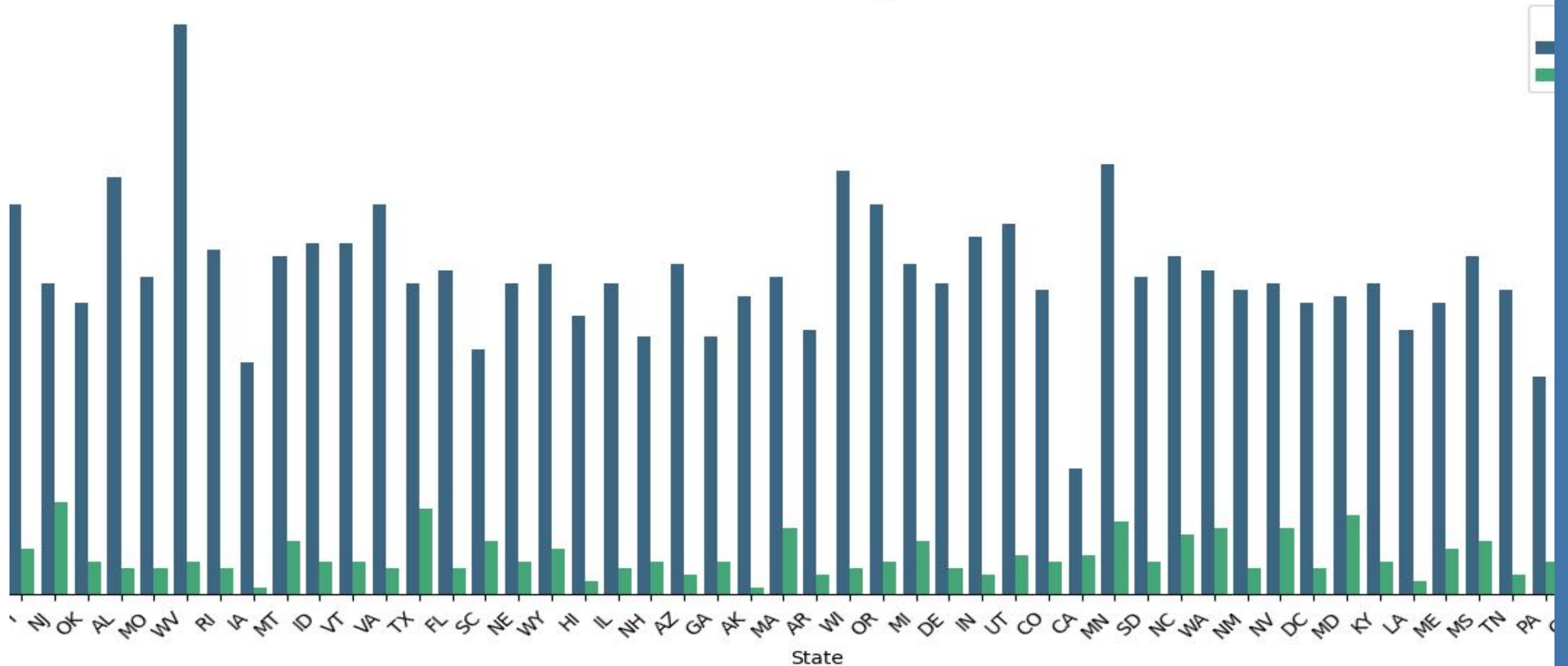
Churn Count by Area Code

Churn Count by International Plan

Modelling

Logistic regression

 Accuracy: 0.82

Precision (Class 0.0): 0.96

Precision (Class 1.0): 0.32

Recall (Class 0.0): 0.84

Recall (Class 1.0): 0.67

F1 Score (Class 0.0): 0.89

F1 Score (Class 1.0): 0.44

Decision Tree Classifier

Accuracy: 0.90571Precision: 0.53125Recall: 0.70833F1 Score: 0.60714

Random Forest Classifier

Accuracy: 0.94429Precision: 0.72603Recall: 0.73611F1 Score: 0.73103

Gradient Boosting Classifier

Accuracy: 0.94143

Precision: 0.69620

Recall: 0.76389

F1 Score: 0.72848

Models Comparison

The ROC curve illustrates the true positive rate against the false positive rate of our classifier. The best-performing models will have a curve that hugs the upper left of the graph, which is the random forest classifier in this case.

Receiver Operating Characteristic (ROC) Curve
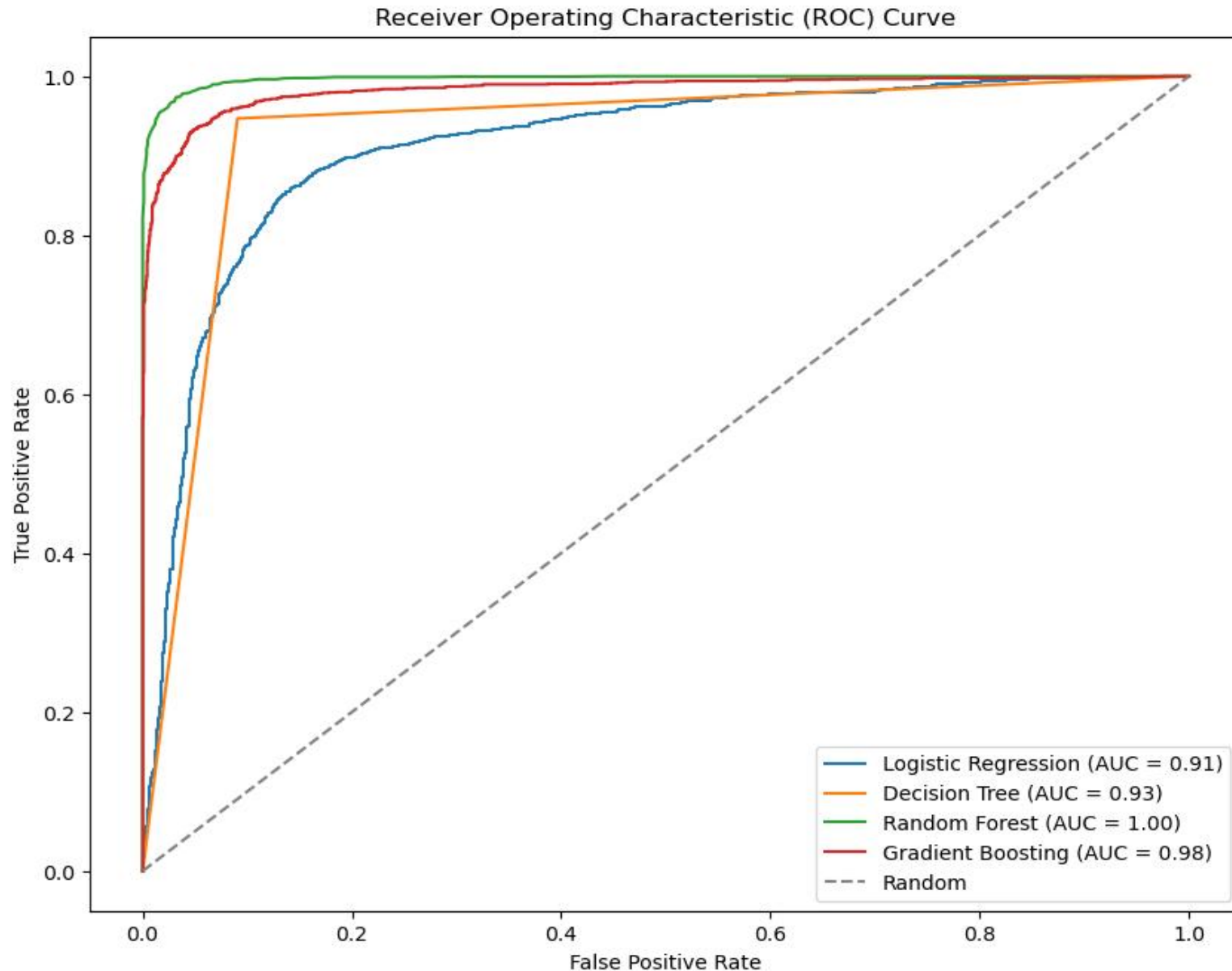
Hyperparameter Tuning of Random Forest Classifier

Accuracy with the best hyperparameters: 0.9328571428571428

Hyperparameter Tuning of Gradient Boosting Classifier

Accuracy with the best hyperparameters: 0.96

Modeling Insights

Both models show improved performance through hyperparameter tuning, with

Gradient Boosting slightly outperforming Random Forest.

For the predictor and based on the results of the Hyperparameter tuning, we choose

The Gradient Boosting Classifier Model for prediction.

Conclusion

In conclusion, the Gradient Boosting Classifier, with tuned hyperparameters, is recommended for predicting customer churn in the telecom dataset. By acting on the insights gained from feature importance and model evaluations, the telecom company can implement targeted strategies to minimize churn and enhance customer satisfaction.

Recommendations

- **Customer Segmentation for Revenue Maximization:** Identify high-value customers by analyzing their usage patterns, plan subscriptions, and service calls. Tailor strategies to retain these customers who significantly contribute to revenue.

- **Risk-Based Prioritization:** Segment customers based on churn risk to prioritize retention efforts. Target high-risk customers with personalized campaigns to reduce the likelihood of churn.

- **Optimizing Service Offerings:** Analyze peak vs. off-peak usage patterns to guide promotional offers or adjustments in service plans, enhancing overall customer satisfaction.

- **Customized Service Utilization:** Understand how different services (voice, international, voicemail) influence churn to tailor service offerings based on customer preferences and needs.

- **Feedback-Driven Improvement:** High numbers of customer service calls may indicate areas for service improvement. Analyze the reasons behind these calls to address specific pain points.

- **Holistic Customer Feedback:** Incorporate customer feedback through surveys or feedback mechanisms for qualitative insights that complement quantitative findings from the dataset.