

SYRIATEL CUSTOMER CHURN RATE ANALYSIS

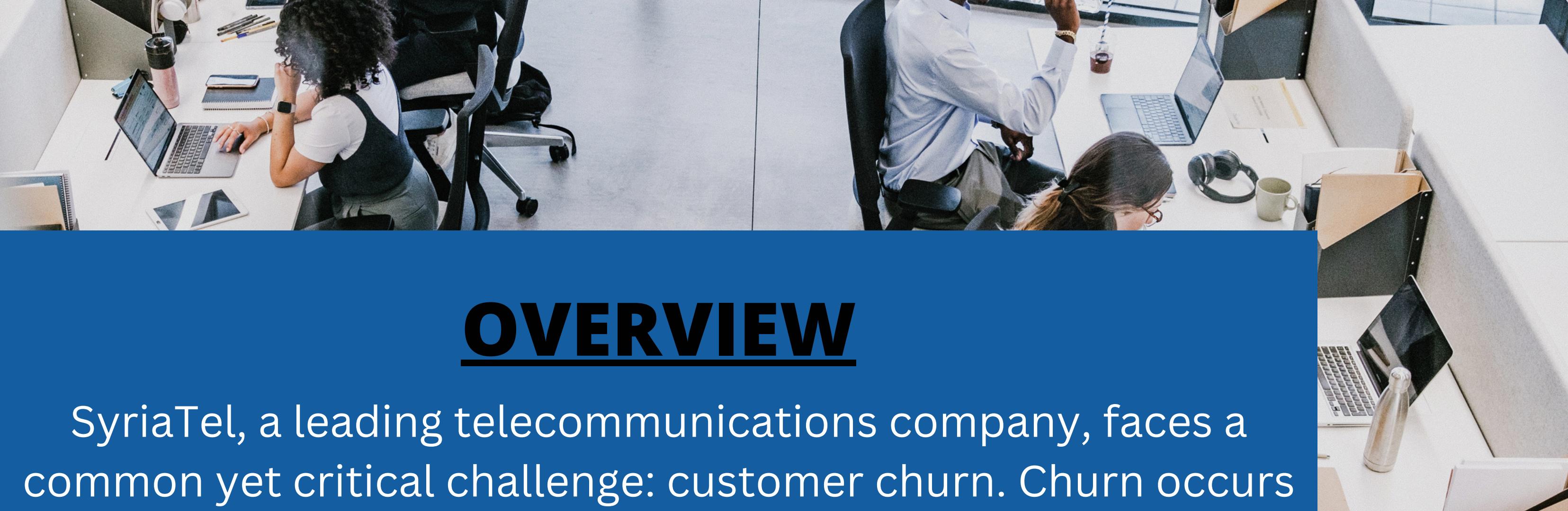
By: Victor Maina



CONTENTS

➤ Overview	03
➤ Business and Data Understanding	04
➤ Modelling	14
➤ Evaluation	20
➤ Recommendations	22
➤ Next Steps	22
➤ Thank you	24





OVERVIEW

SyriaTel, a leading telecommunications company, faces a common yet critical challenge: customer churn. Churn occurs when customers stop using the company's services, leading to a loss of revenue. Retaining customers is often more cost effective than acquiring new ones. This project aims to uncover patterns and predict which customers are likely to churn.

Business Understanding

Objectives

- Who are the customers most likely to churn?
- What are the main factors influencing customer churn?
- How can SyriaTel reduce the churn rate effectively?



Potential Impact and Benefits

- Revenue Retention
- Customer Lifetime Value
- Improved Customer Satisfaction
- Competitive Advantage



Data Understanding

1.Data collection

Data was sourced from SyriaTel's customer databases, including CRM systems, billing systems, customer support logs and usage analysis.

2.Key Data Features

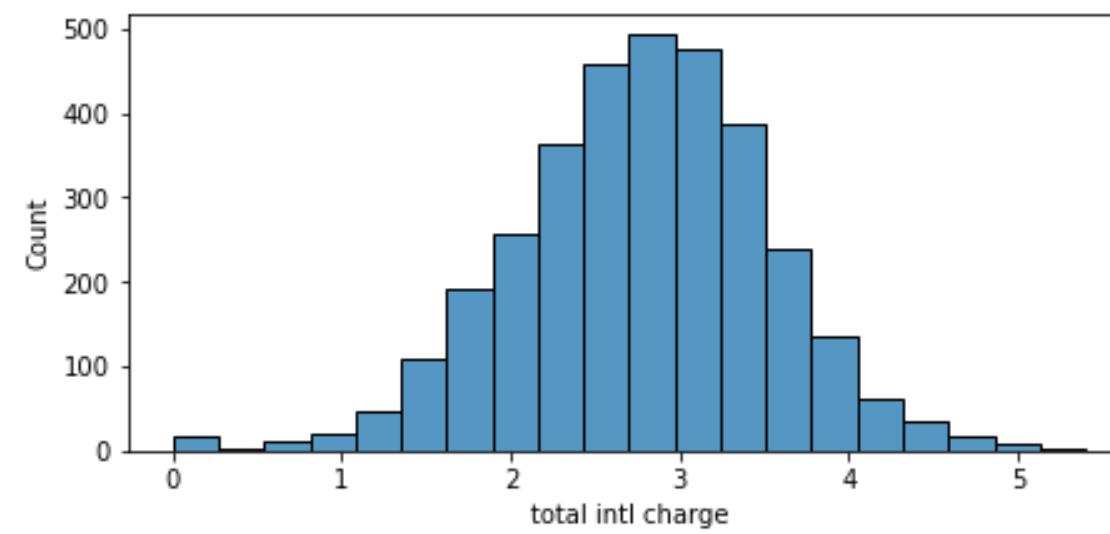
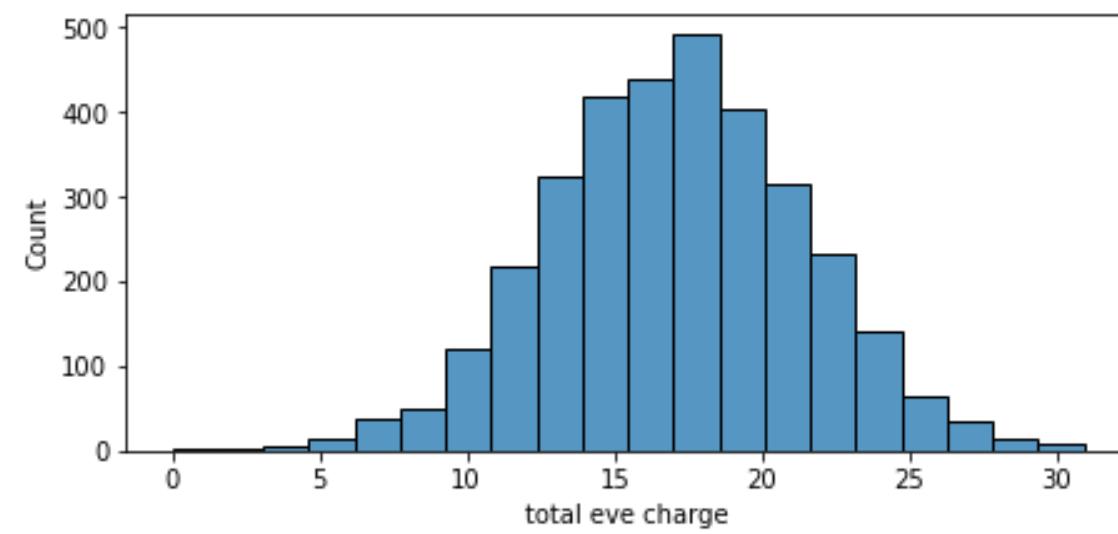
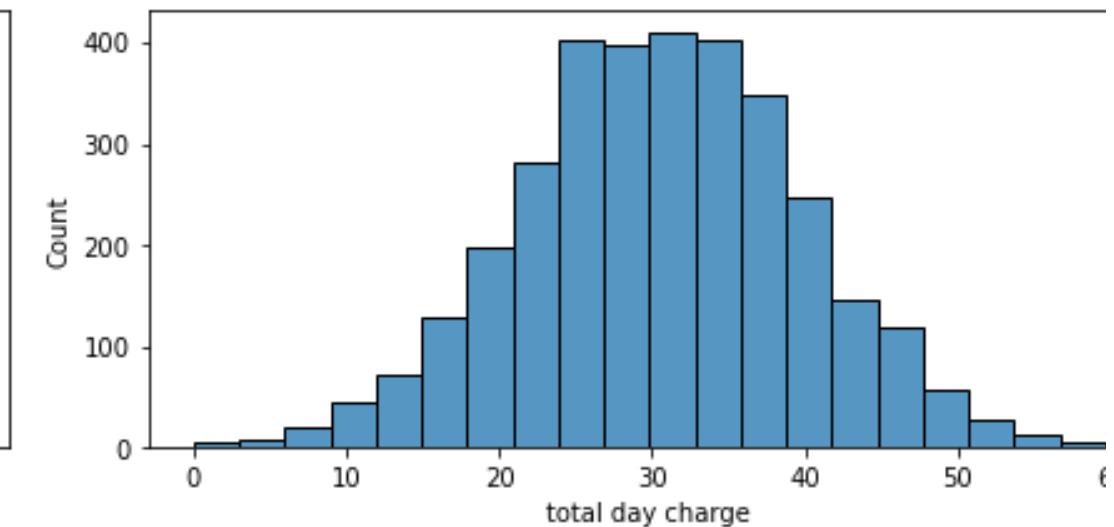
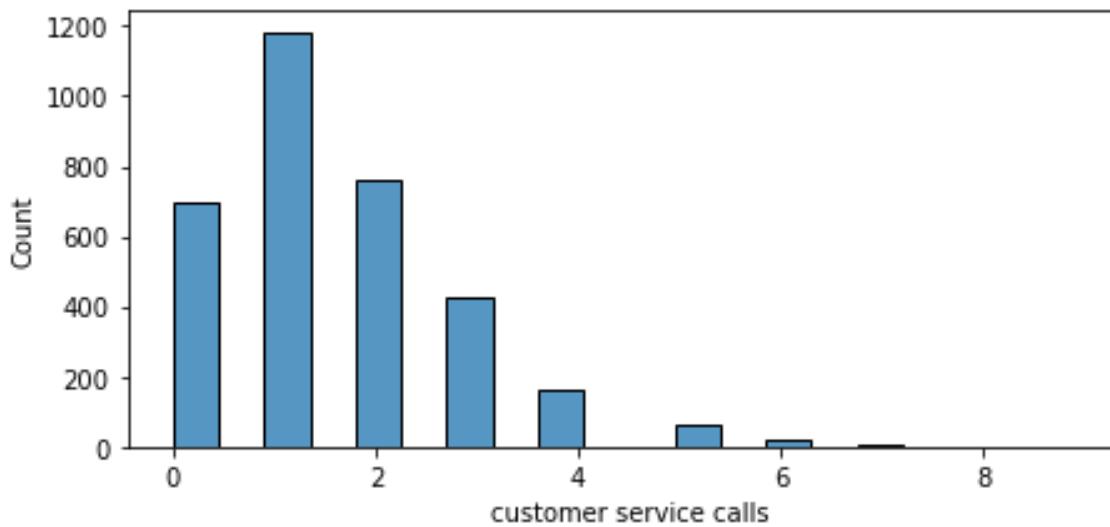
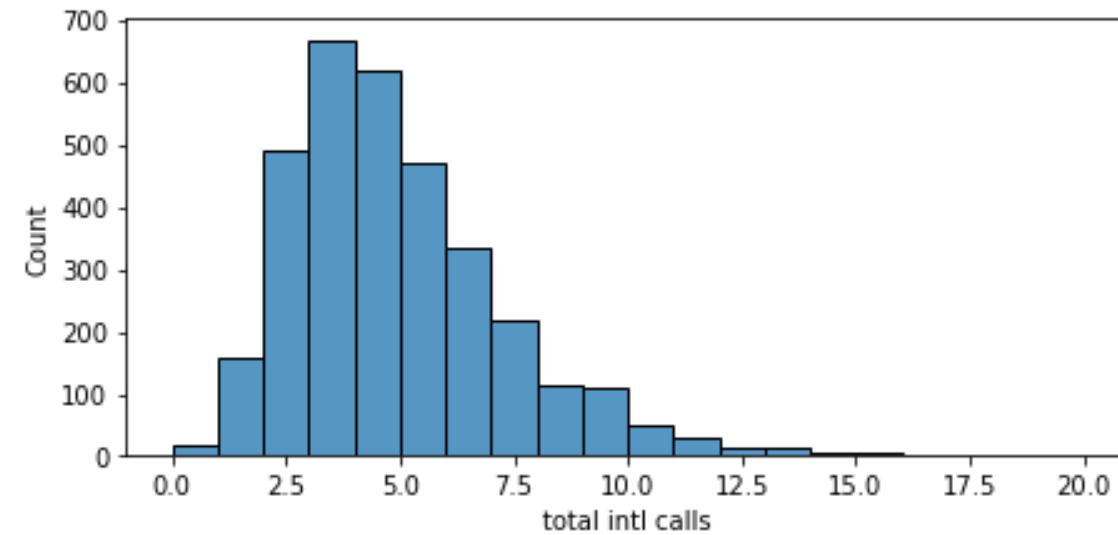
- total day charge: total cost of calls during the day
- total eve charge : total cost of calls during evening hours
- customer service calls : the number of calls a customer has made to customer service
- total intl charge : total cost of making international calls
- total intl calls : total number of international calls made

Data Analysis

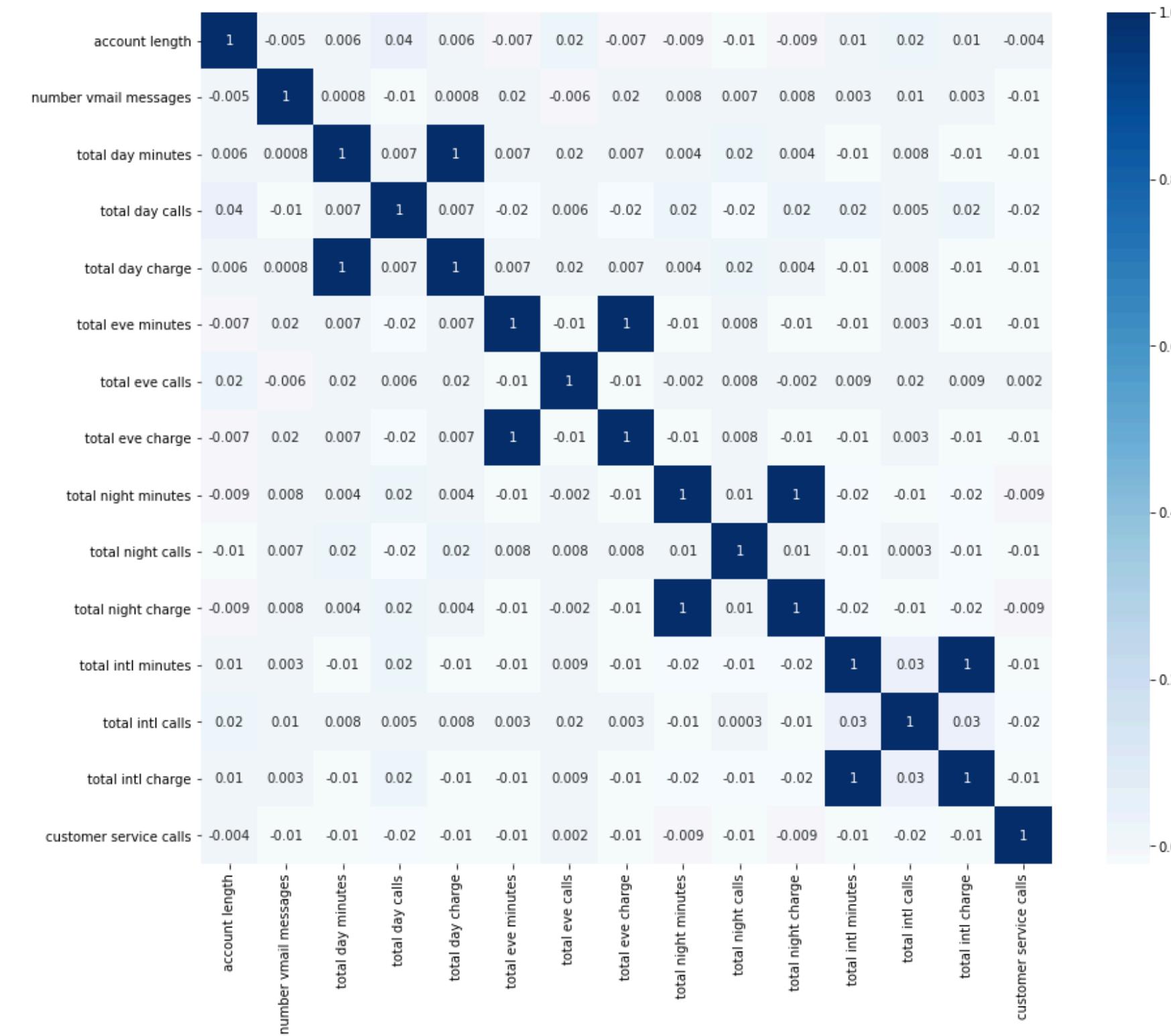
- Conducted exploratory data analysis to understand data distributions, correlations, and potential anomalies.
- No missing values were found in the dataset .



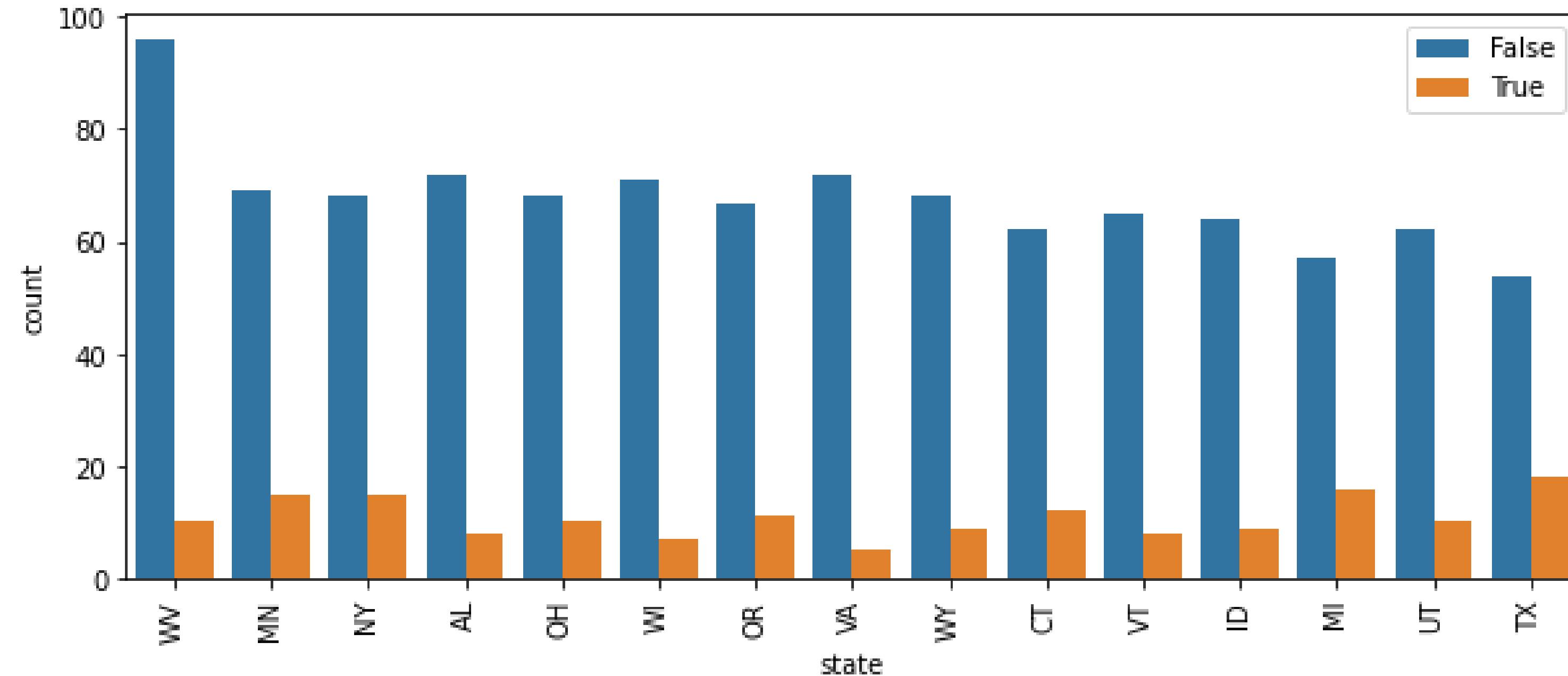
Plot Histograms to visualize data distribution



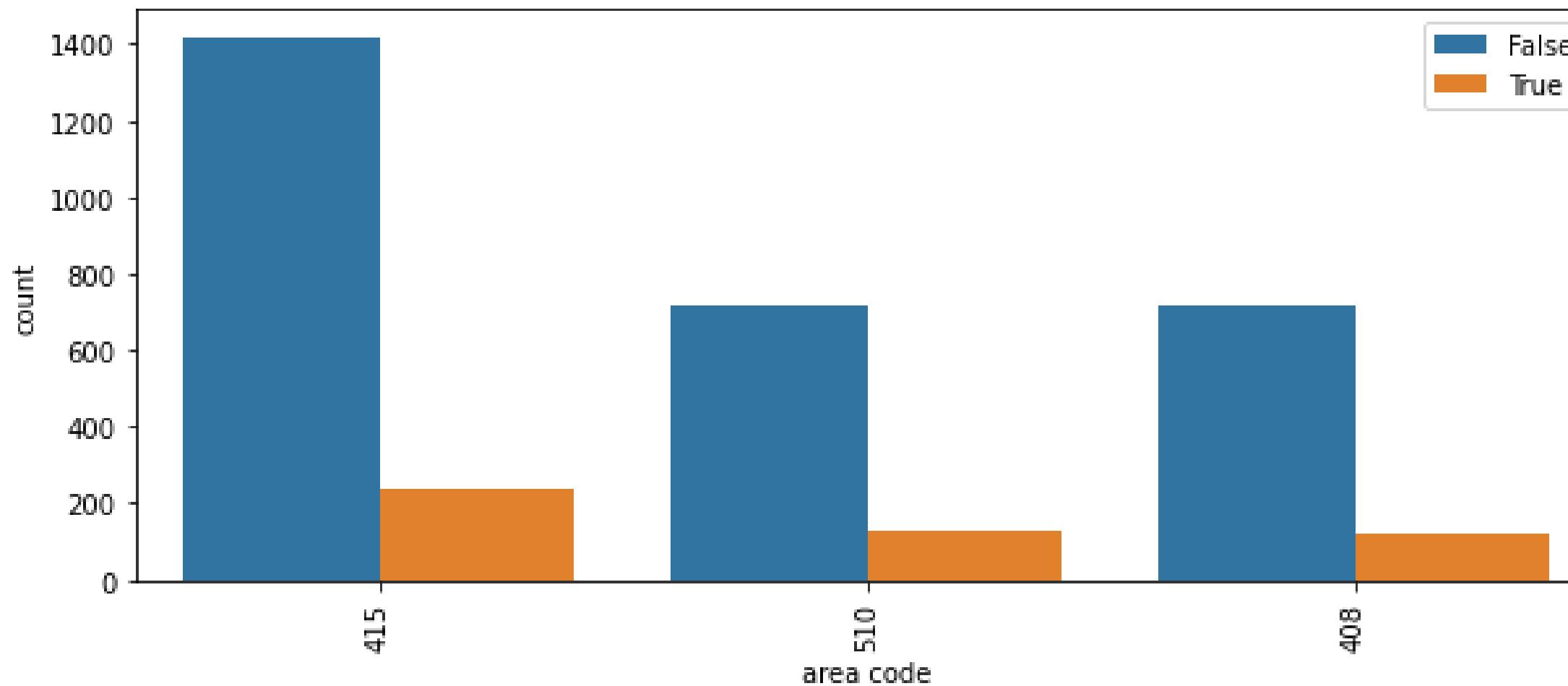
Correlation heatmap for numerical features



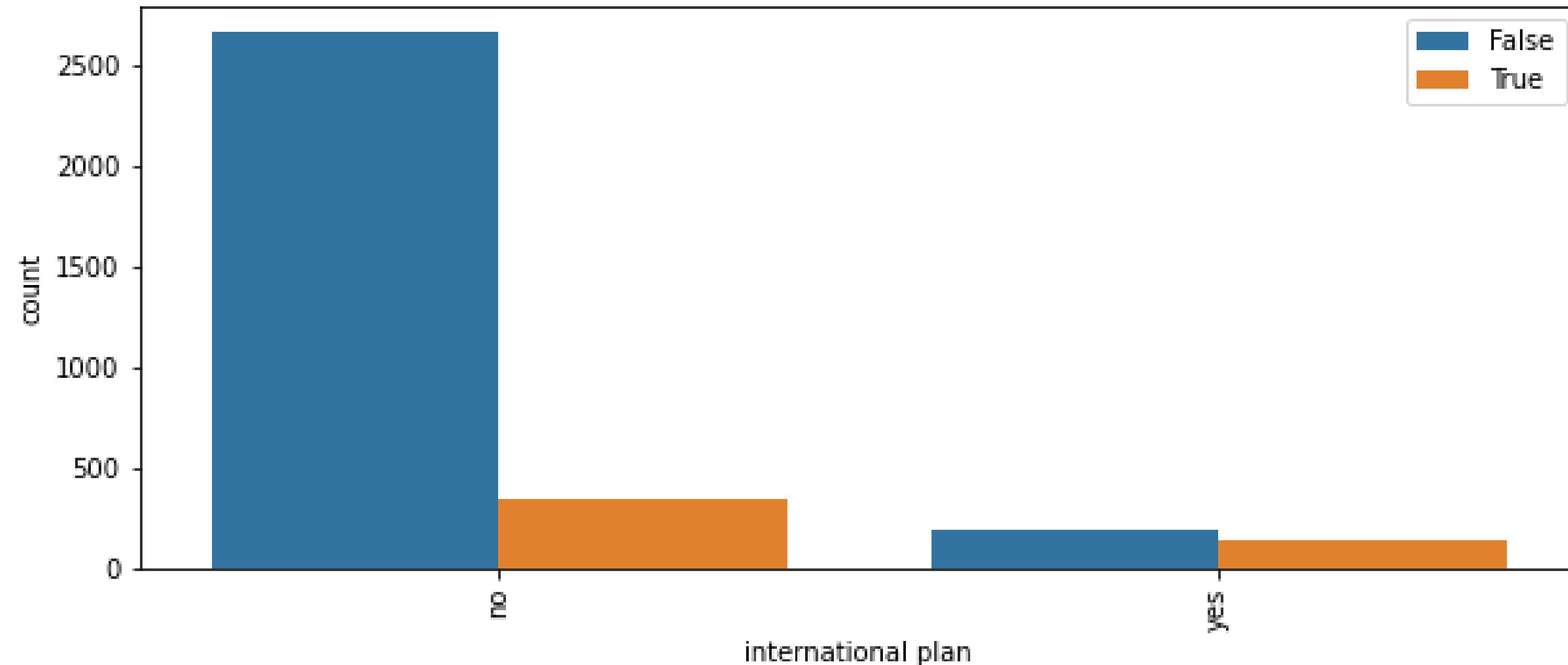
Analysis of state variable



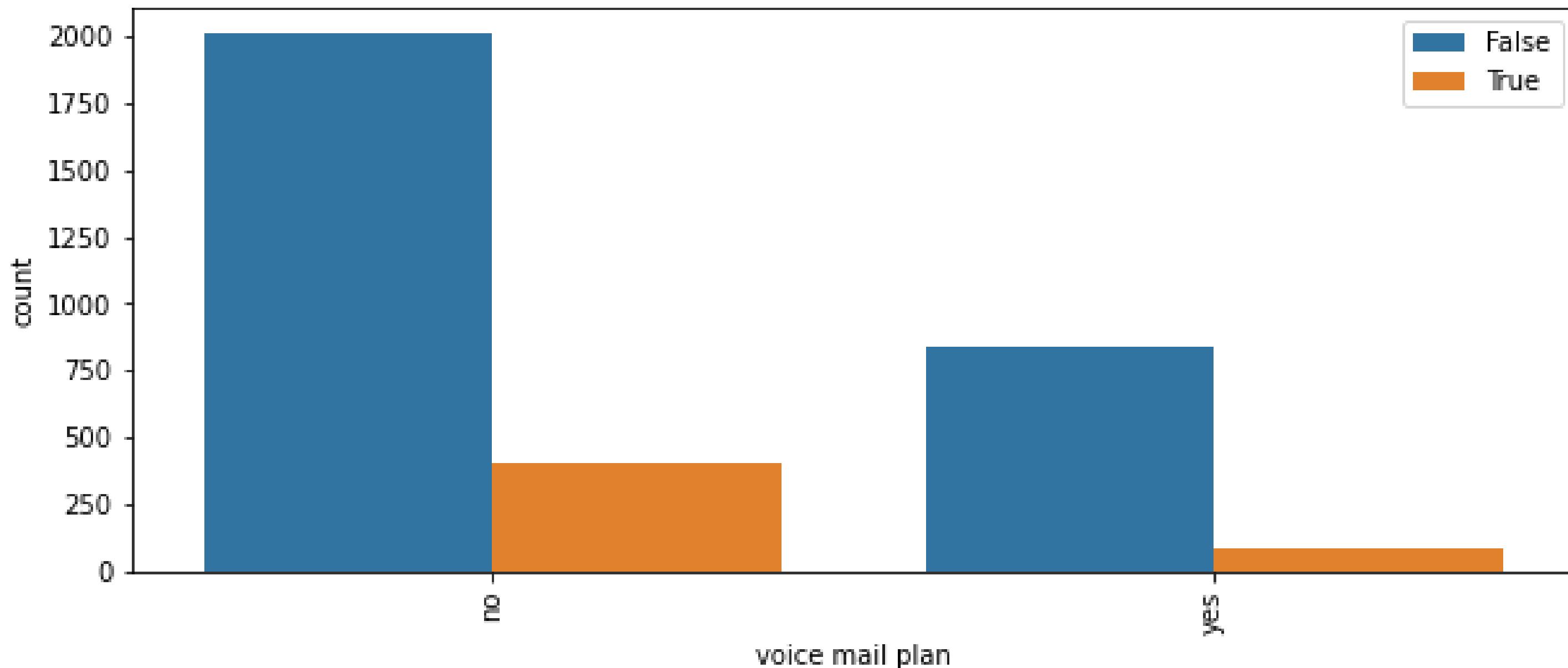
Analysis of area code variable



Analysis of international plan variable



Analysis of voice mail plan variable

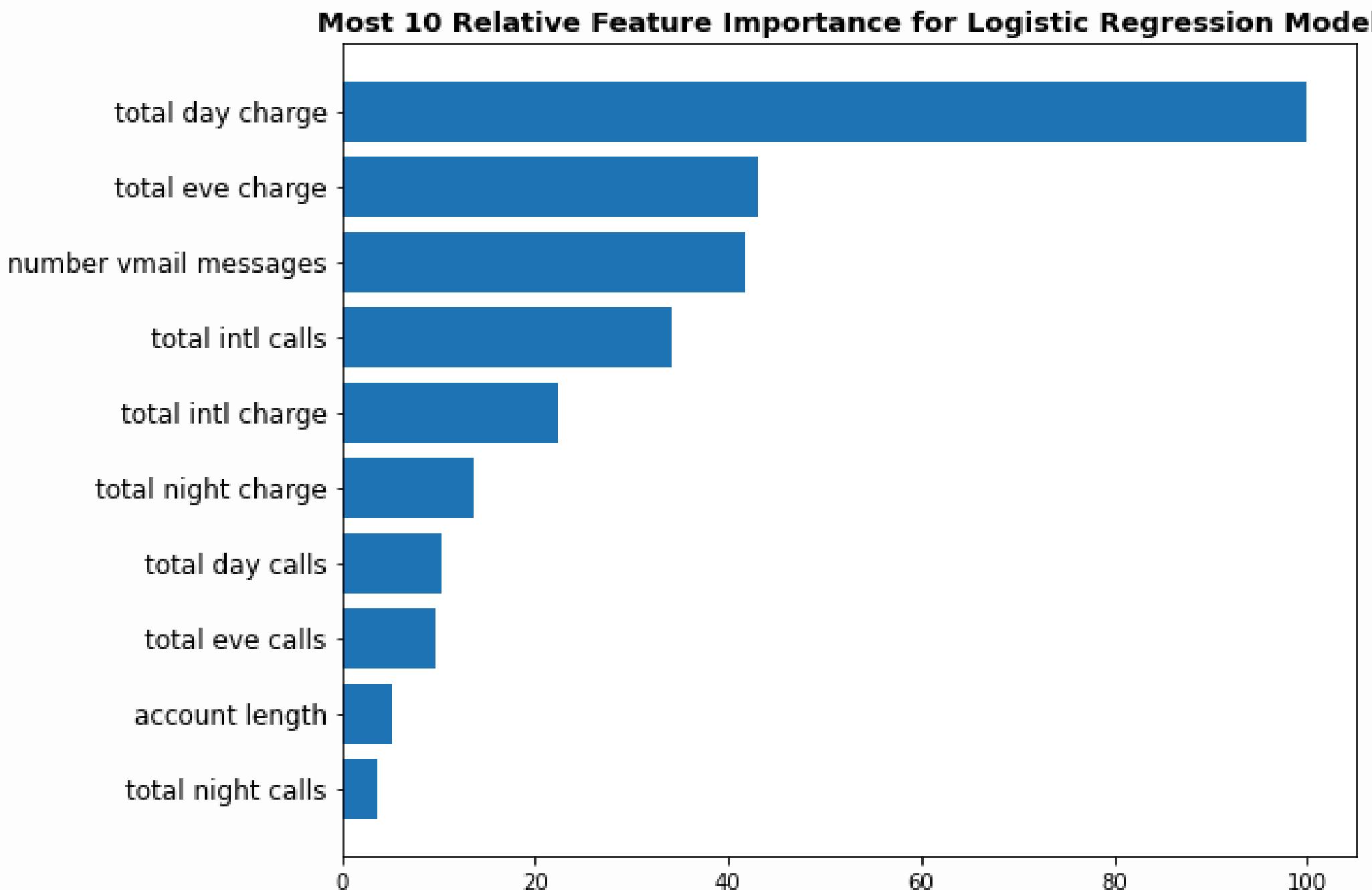


Modelling

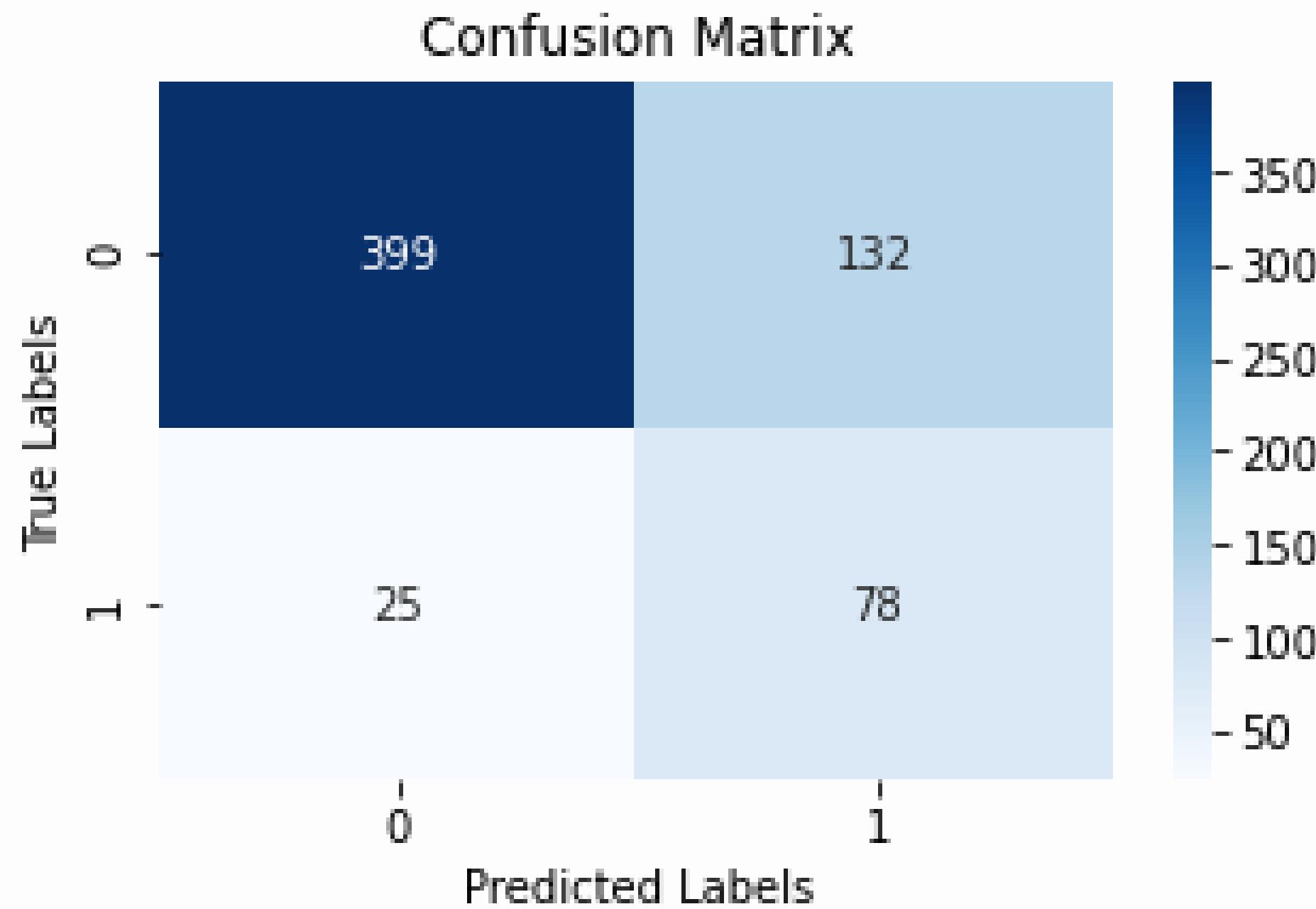
1. Logistic Regression Model

- The Logistic Regression Model has an accuracy score of 75.2% which means that the model correctly classified 75.2% of the total instances in the training set.
- The logistic regression model has a precision score of 37.1% which suggests that the model has a high rate of false positives
- Logistic Regression model has a recall score of 75.7% which is relatively high indicating that the model is good at capturing most of the positive cases
- To improve the model's precision, feature engineering through adding or refining features can help the model make more accurate predictions

Most Relative feature importance for Logistic Regression model



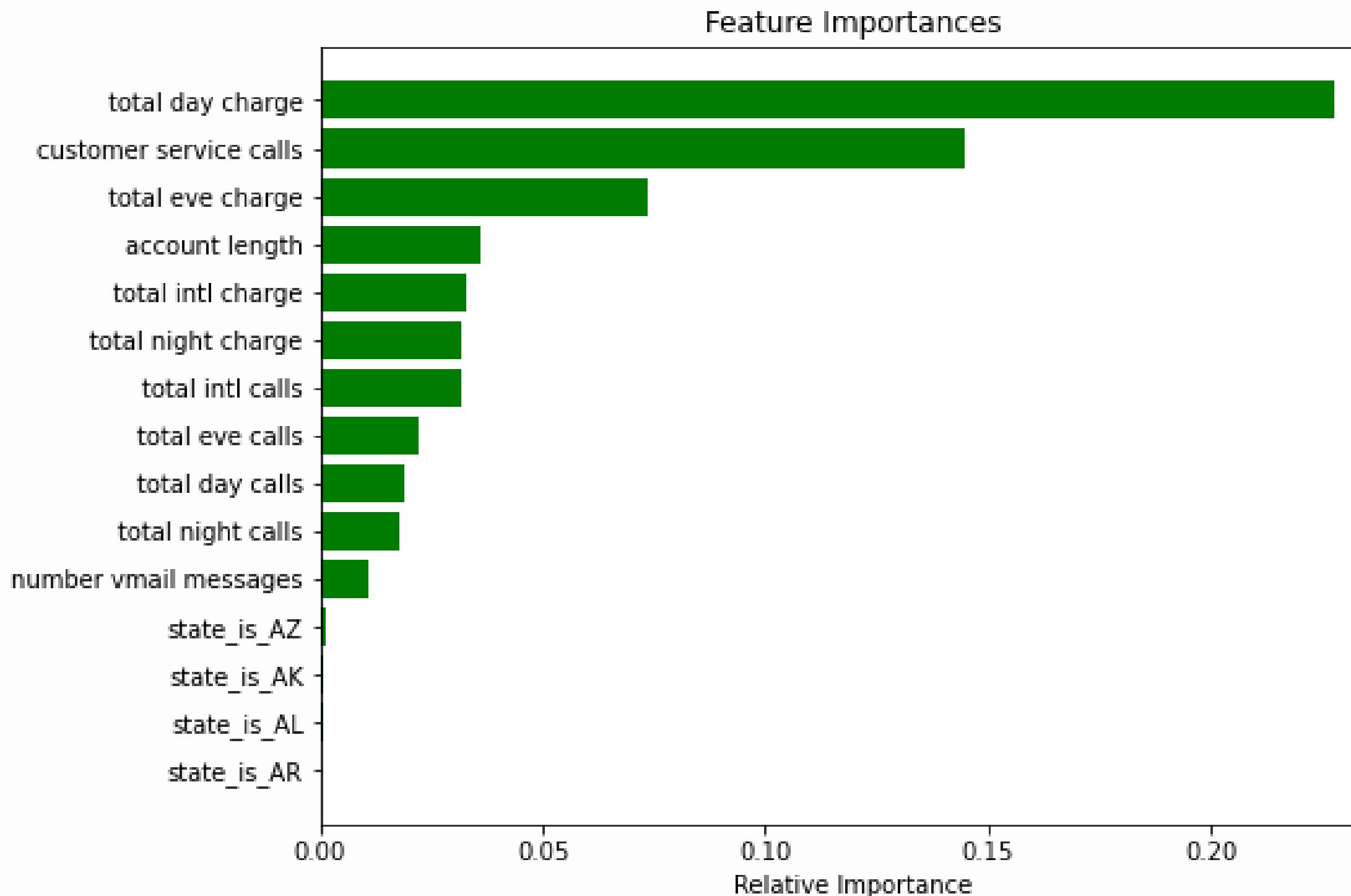
Confusion Matrix for Logistic Regression model



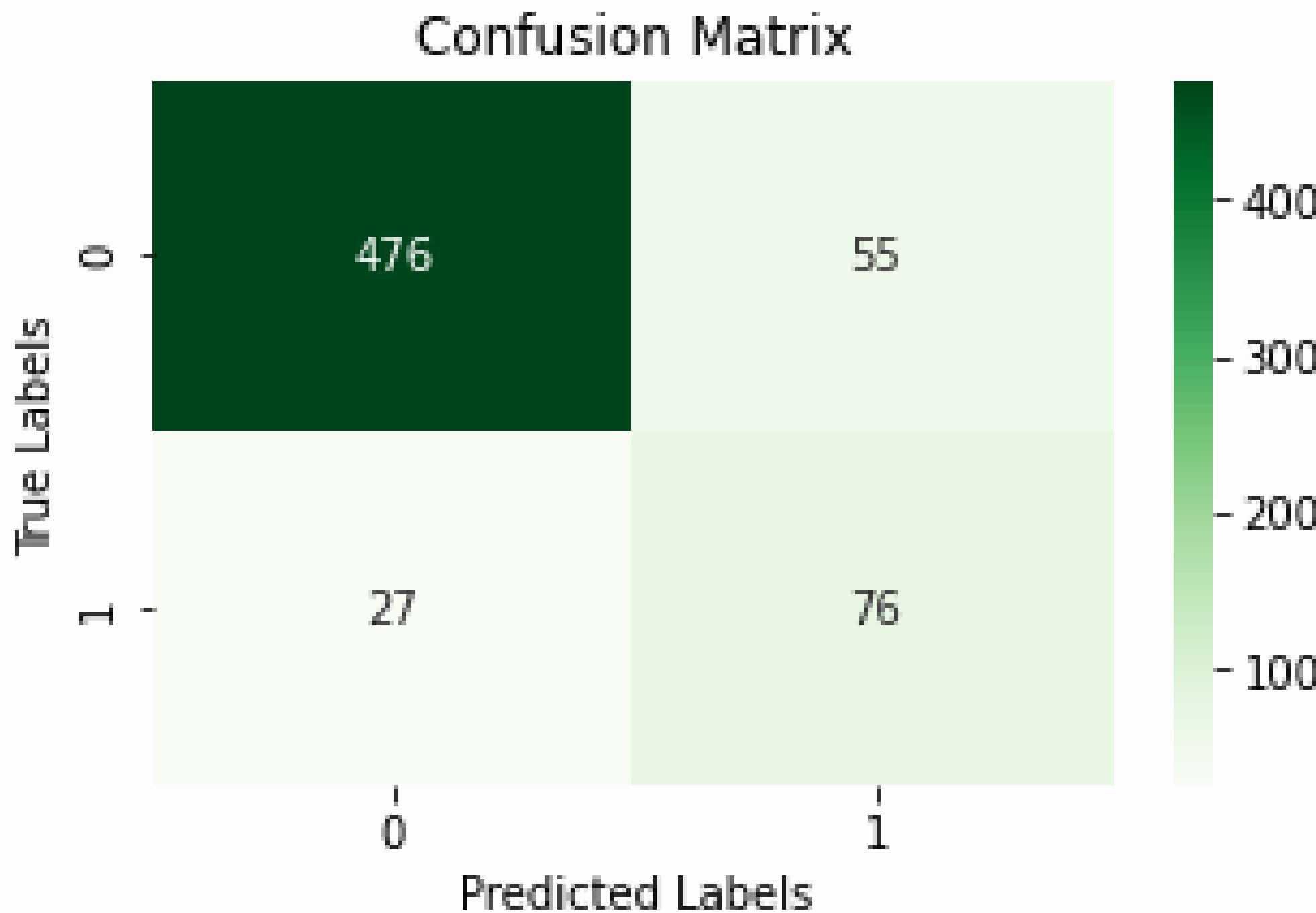
2. Decision Tree Model

- The Decision Tree model has an accuracy score of 87.4% indicating it performs well overall in correctly classifying instances in the training set.
- The Decision Tree model has a precision of 59% which means that about 59% of the predicted positives are actually positive.
- The model has a recall score of 74% which means the model is good at capturing most of the positive cases.
- Hyperparameter tuning can be used to improve the model.

Most relative feature importance for Decision Tree Model



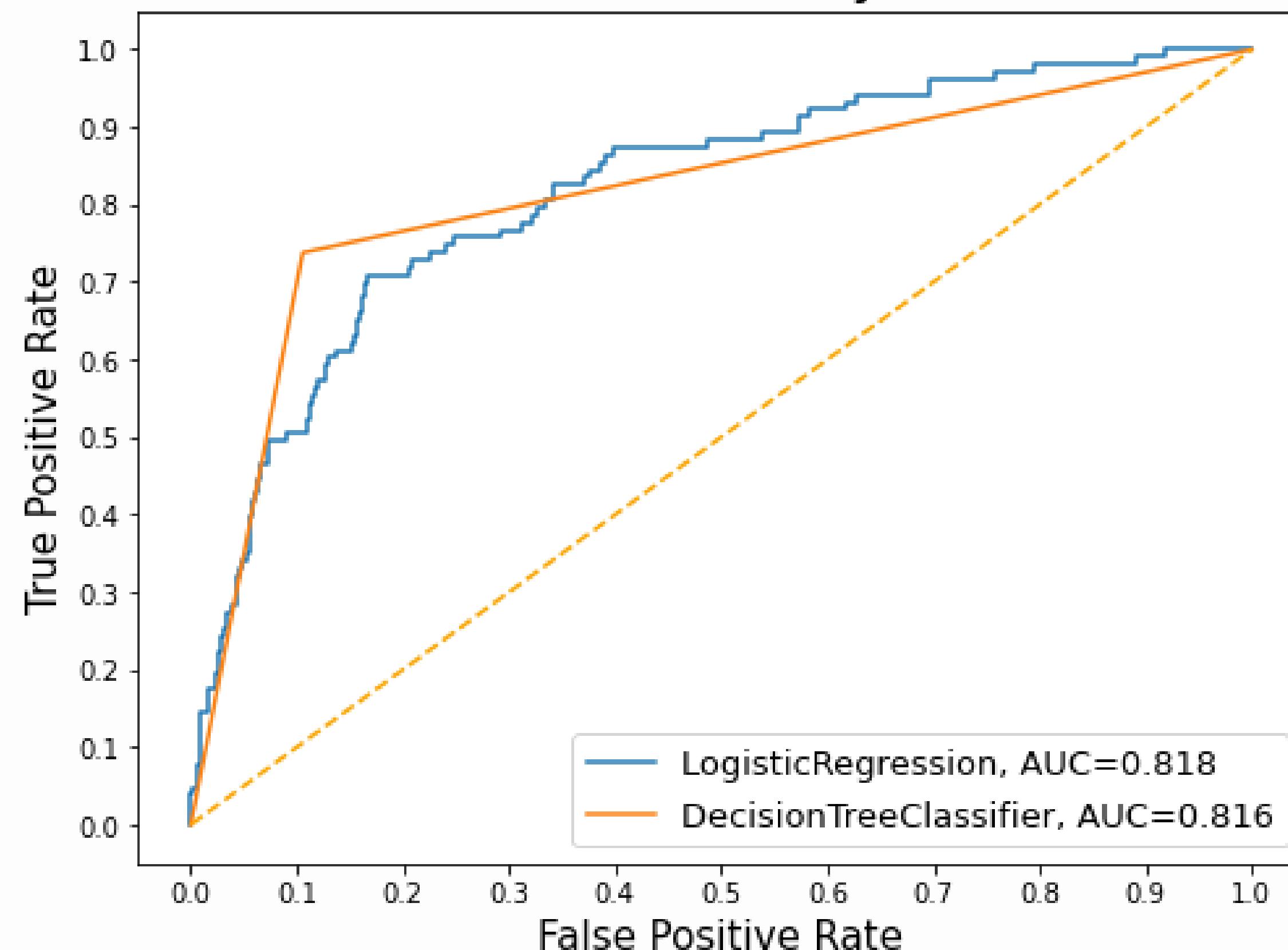
Confusion Matrix for Decision Tree model



MODEL COMPARISON

- The AUC for logistic regression is 0.818 which indicates that the model is performing reasonably well, but theres still room for improvement. The curve`s shapes suggests that the model is generally good at distinguishing between positive and negative churn cases, but there might be some trade-off between sensitivity and specificity
- The AUC for Decision Tree Classifier is slightly higher than Logistic Regression which suggests that the Decision Tree Model might be performing better marginally in terms of overall classification accuracy

ROC Curve Analysis



Findings and Recommendations

- The Decision Tree model outperforms the Logistic Regression model in terms of; accuracy where it has a higher accuracy of 87% indicating it performs better overall in correctly classifying instances in the training set.
- Precision score of 59% meaning it is more reliable in terms of the proportion of predicted positives that are truly positive compared to the logistic regression model.
- The logistic regression model though does have a slightly higher recall, indicating it identifies a slightly higher proportion of actual positives. However, the difference is not significant.

Findings and Recommendations

- Based on these results, the Decision Tree model seems to better predict churn rate among SyriaTel's customers as it does a better job at predicting churn rate based on accuracy score and precision
- Customers who incur high total day charges, make many customer service calls and have high total evening charges are most likely to stop using SyriaTel's telecommunications services
- The company should prioritise lowering their day charges through probably offering discounts during day time hours, ensuring customer queries are sorted in a first and timely manner and also lowering evening charges to retain customers and reduce the churn rate among their customers

THANK YOU!

VICTOR MAINA
0743235926
victorkamau420@gmail.com

