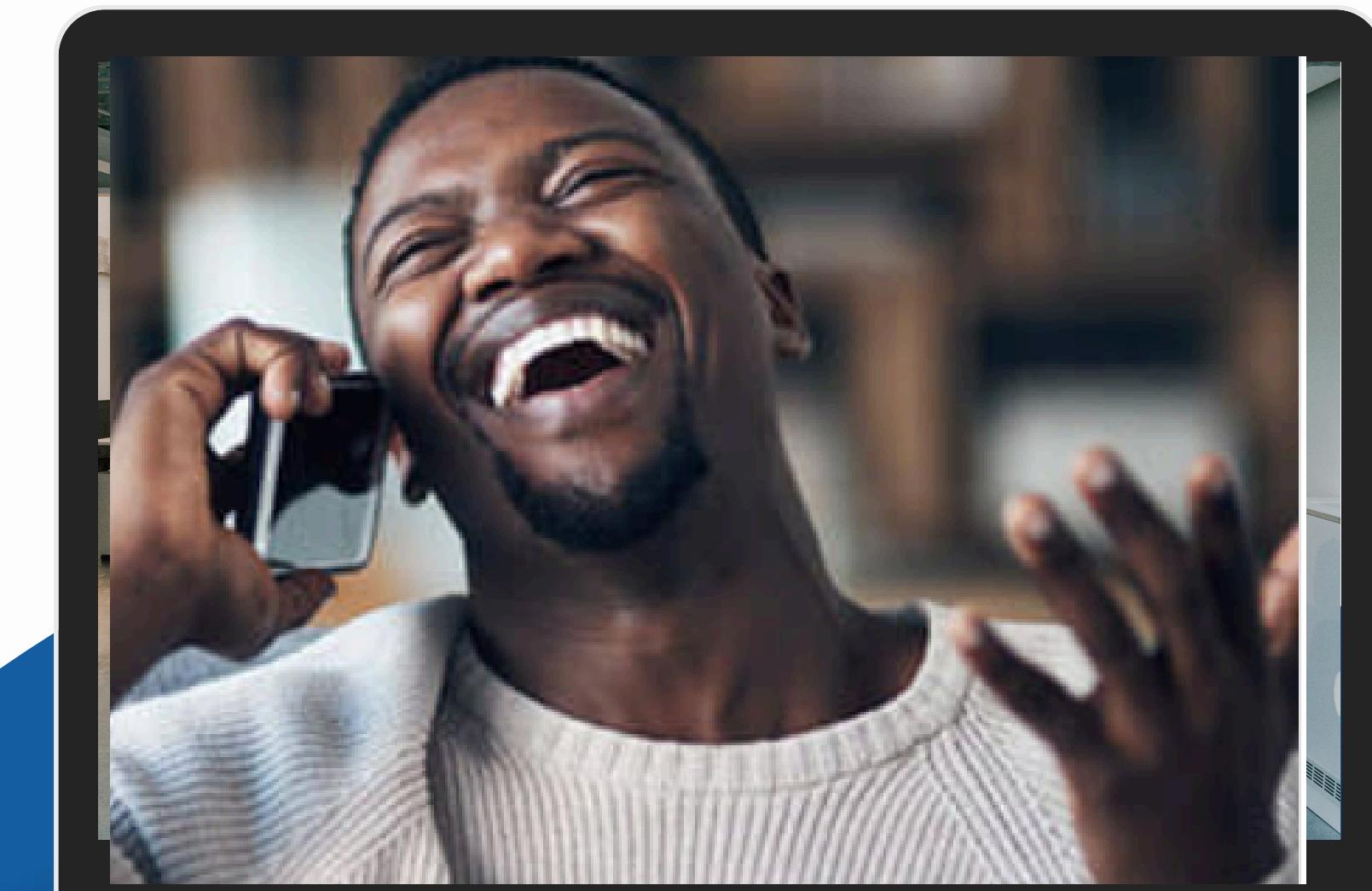


SYRIA TEL COMPANY

PREDICTING CUSTOMER CHURN USING MACHINE LEARNING ALGORITHMS

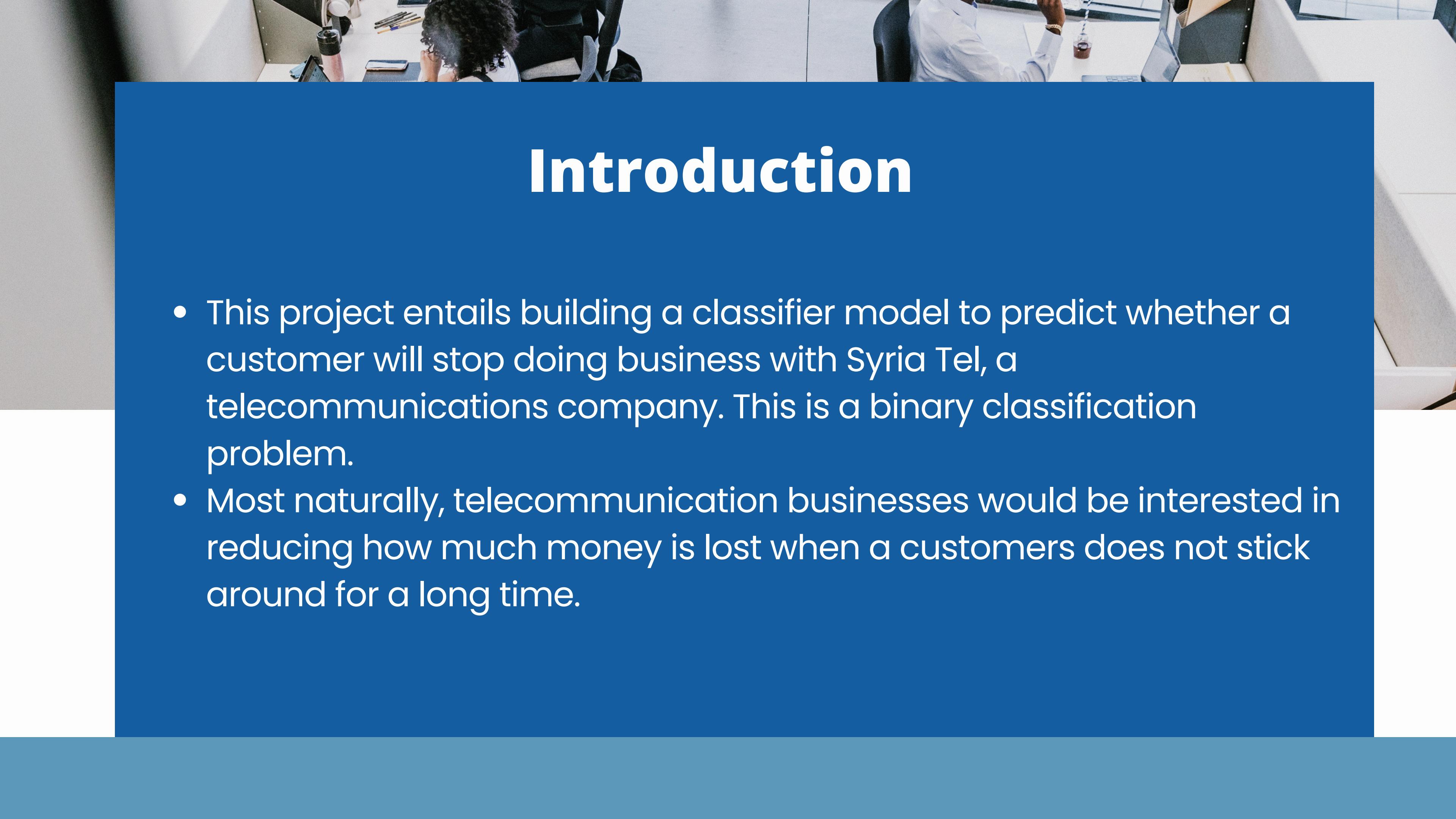
By: Cynthia Dalmas



Overview

- ▶ BUSINESS PROBLEM 01
- ▶ DATA UNDERSTANDING 02
- ▶ DATA PREPARATION 03
- ▶ MODELING 04
- ▶ EVALUATION OF METRICS 05
- ▶ DEPLOYMENT OF MODEL 06





Introduction

- This project entails building a classifier model to predict whether a customer will stop doing business with Syria Tel, a telecommunications company. This is a binary classification problem.
- Most naturally, telecommunication businesses would be interested in reducing how much money is lost when a customers does not stick around for a long time.

DATA UNDERSTANDING

The dataset contains call details of 3333 customers with 21 features and a dependent churn parameter with two values: Yes/No.

Some features include information about the number of incoming and outgoing messages and voicemail for each customer.

Most independent variables have a normal distribution

METHODOLOGY

The following methods were followed in order to meet the objectives of the project.

DATA CLEANING

Irrelevant columns were dropped from the dataset, Checking for missing values, checking and dropping duplicates & label encoding was done.

EXPLORATORY DATA ANALYSIS

Performing a descriptive statistics, Checking the distribution of the columns of the data set, using chart and diagrams

MODELING

Six Supervised Machine learning models were fit on the cleaned dataset and classification metrrics recorded

EVALUATION OF METRICS

All model performances was evaluated against each other.

MODEL DEPLOYMENT

Checking the dictrition of data set, using chrt and diagrams

EXPLORATORY DATA ANALYSIS

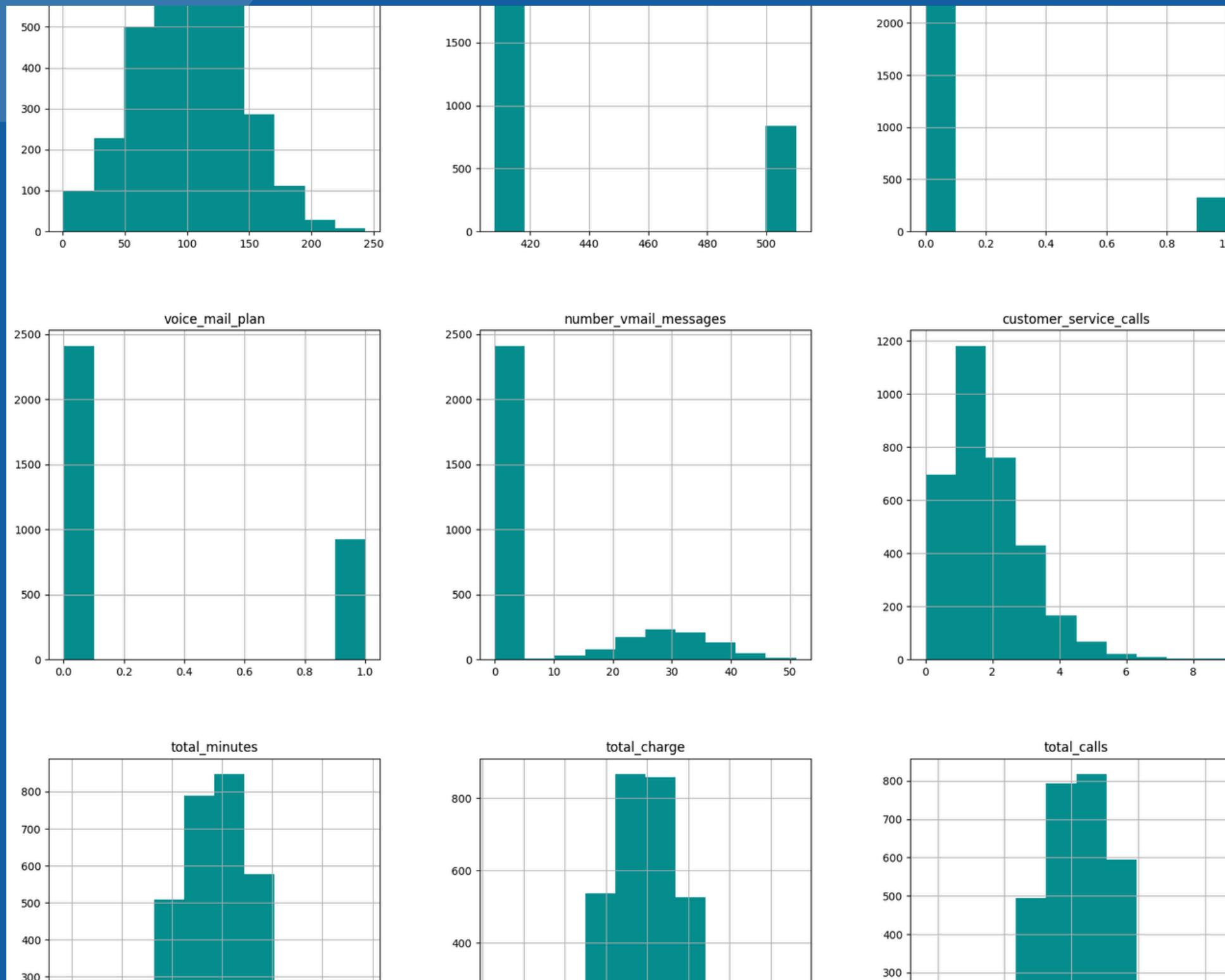


Fig. 1: Histogram, indicates that most variable have a normal type of distribution

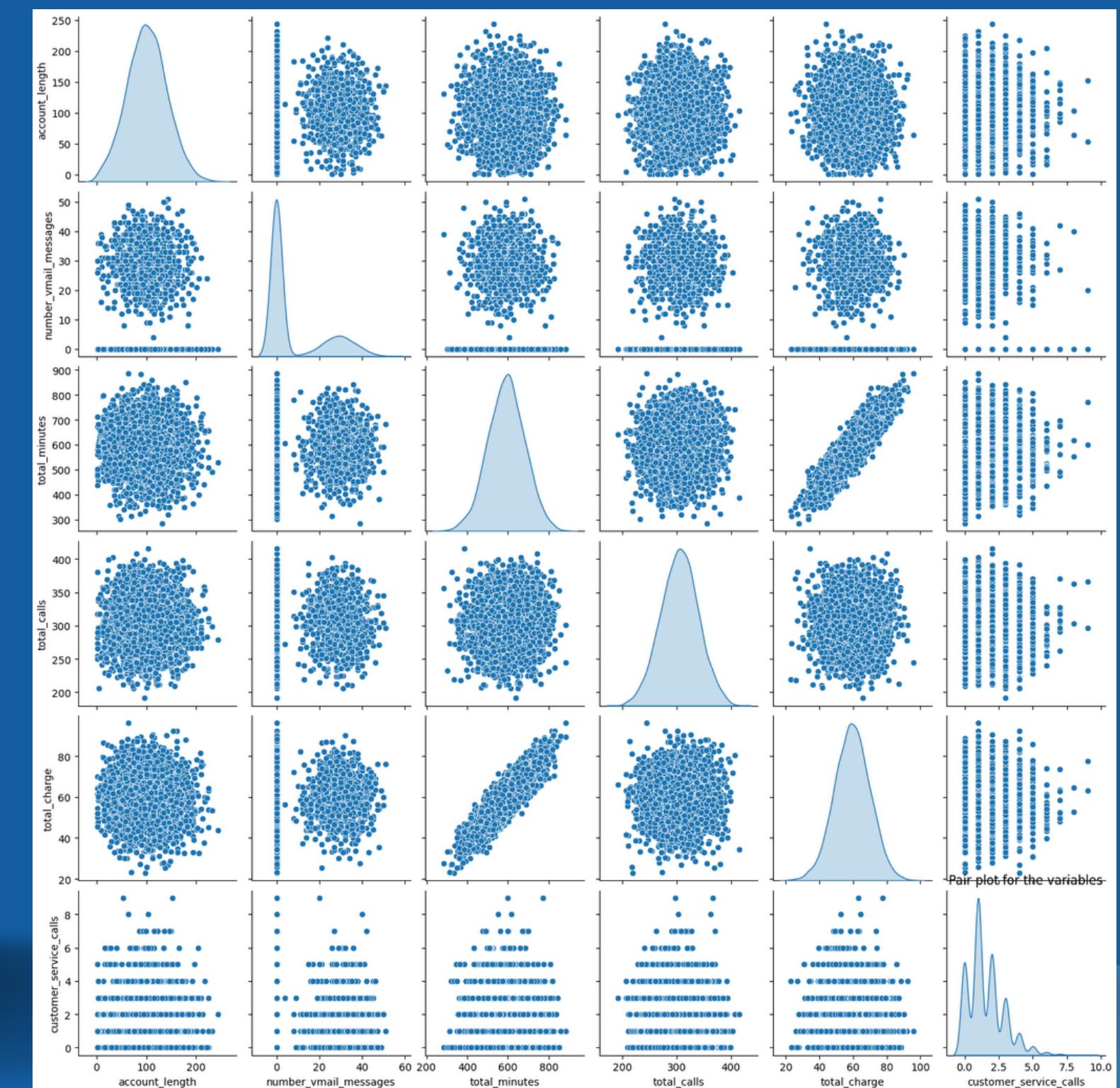
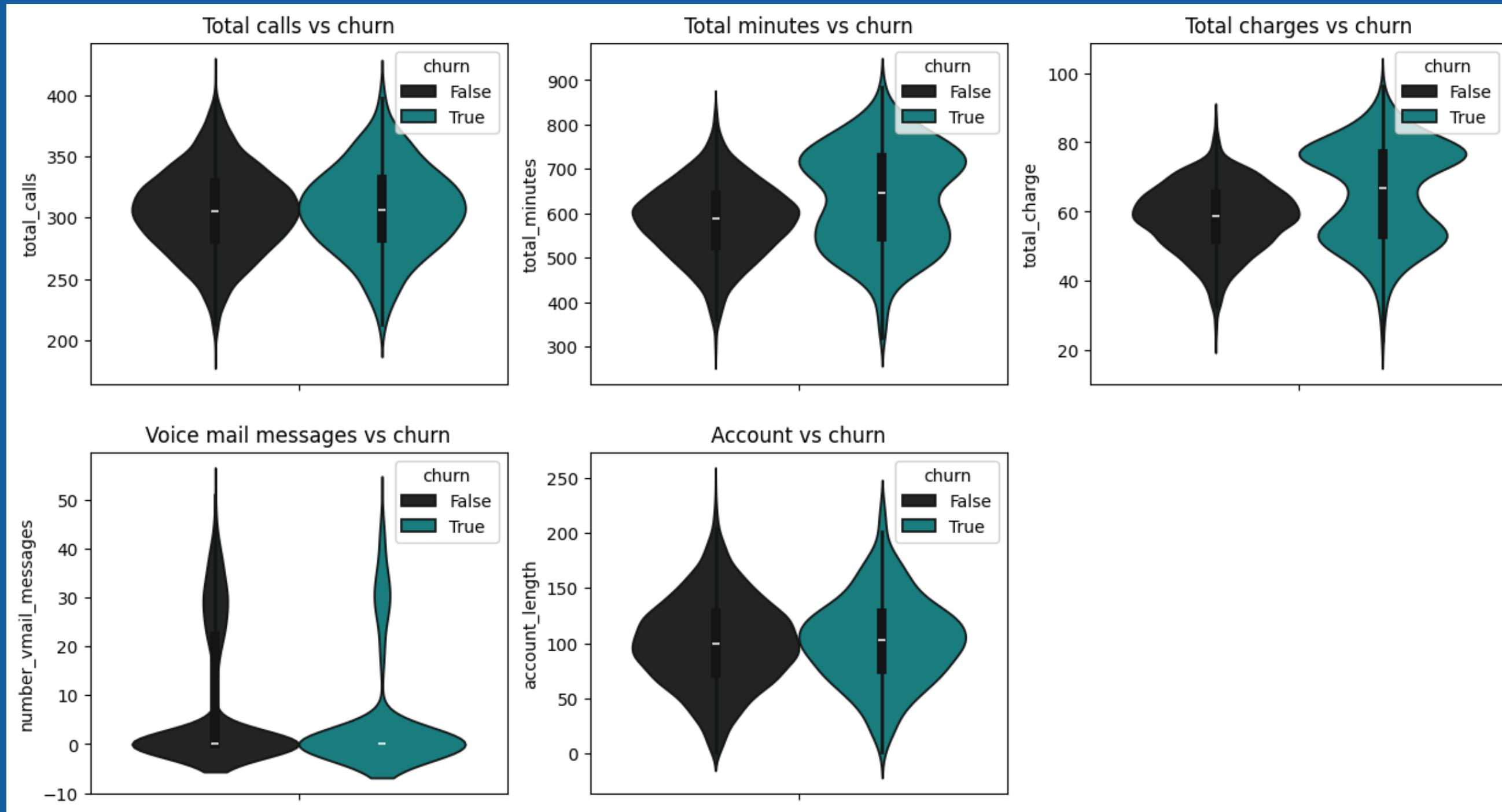


Fig. 2: Pairplot for checking for relationships between the variables

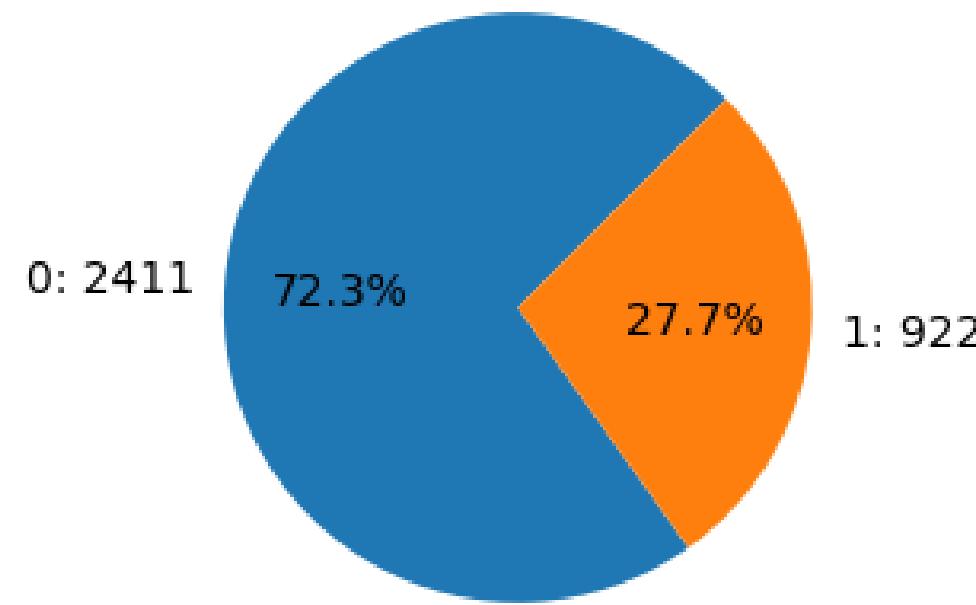
EXPLORING CHURNING FACTOR WRT TOTAL CALLS,TOTAL MINUTES,TOTAL CHARGE,ACCOUNT LENGTH,No of VOICE MAIL MESSAGES



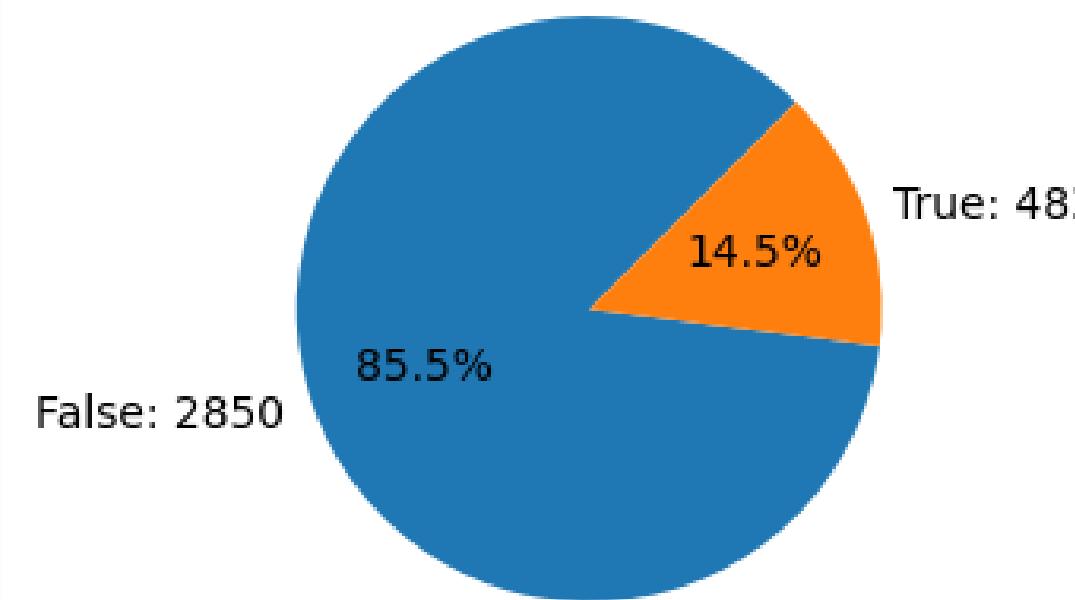
The violin plots, Churning factor is quite visible among customers who incurred high charges.

EXPLORING INTERNATIONAL PLAN, VOICE MAIL PLAN & CHURNING

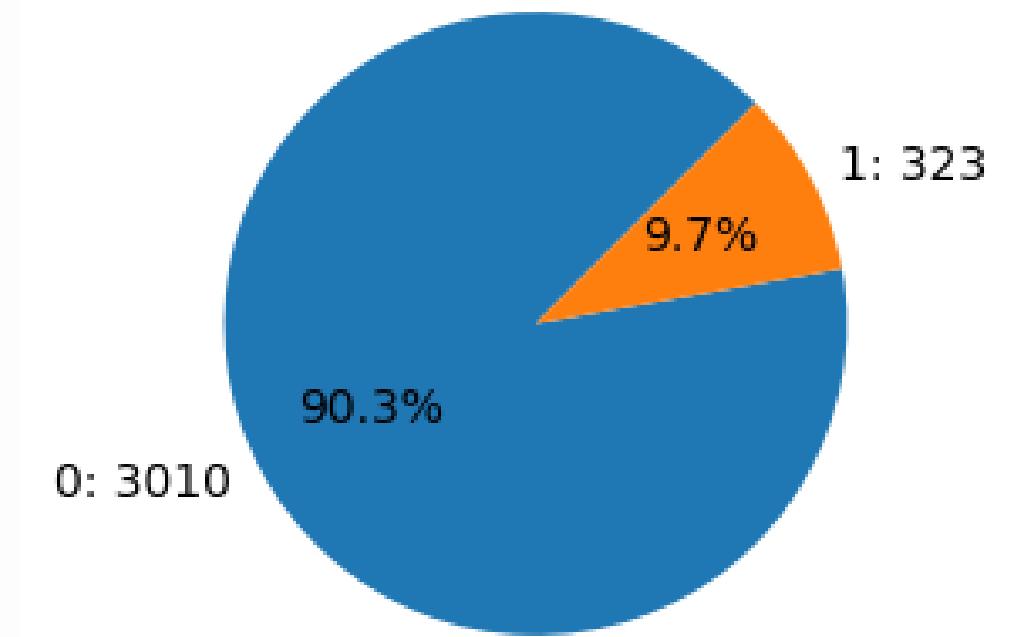
Customers with Voice mail plan



customer churning ratio



Customers with international plan



A bi-variate analysis of these variables indicates that 27.7% had voice mail plan, 14.5% churned from the company, 9.7% had international plan as shown by the pie plots.

EXPLORING CALLS, TALK TIME MINUTES VS CHARGES

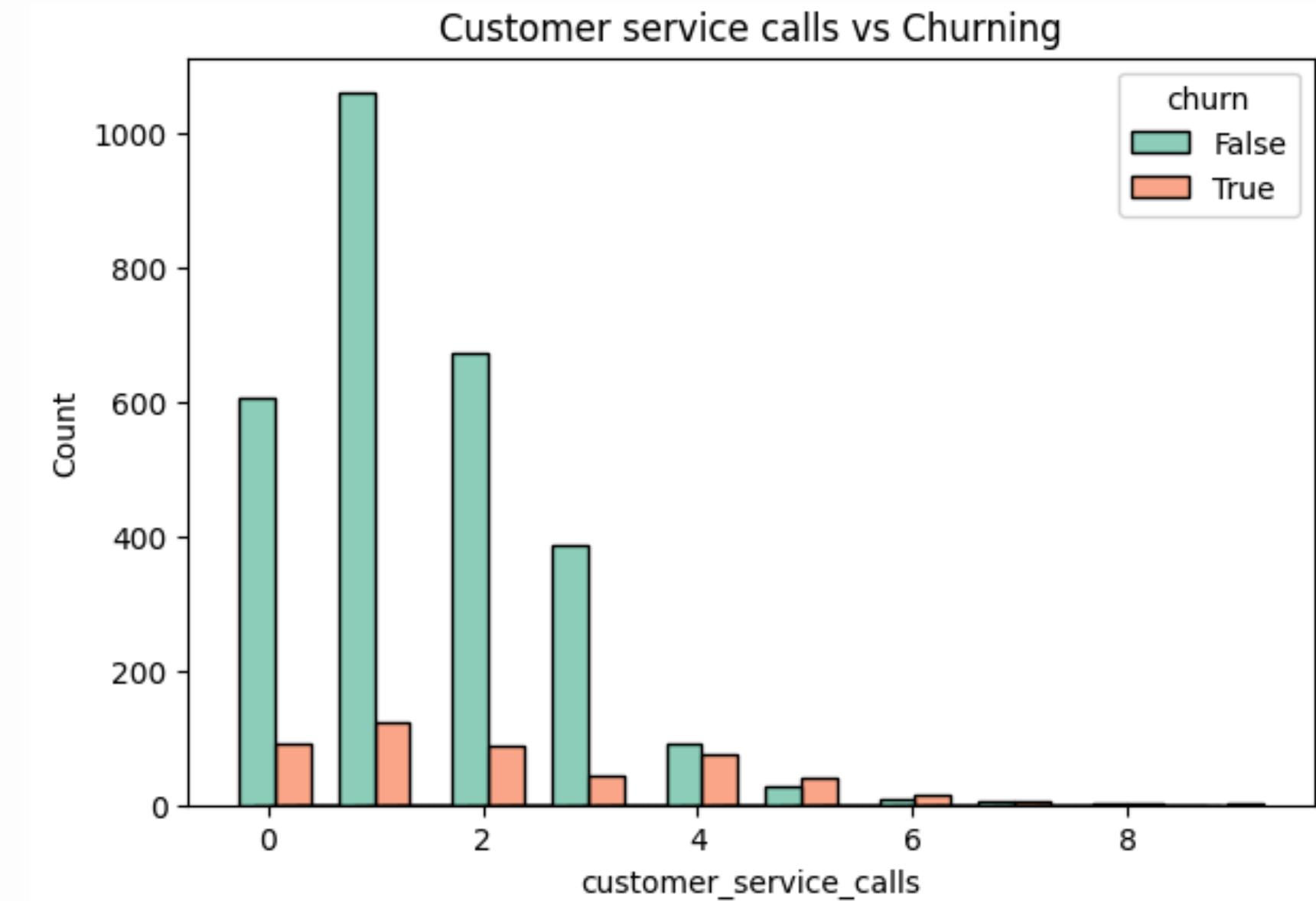
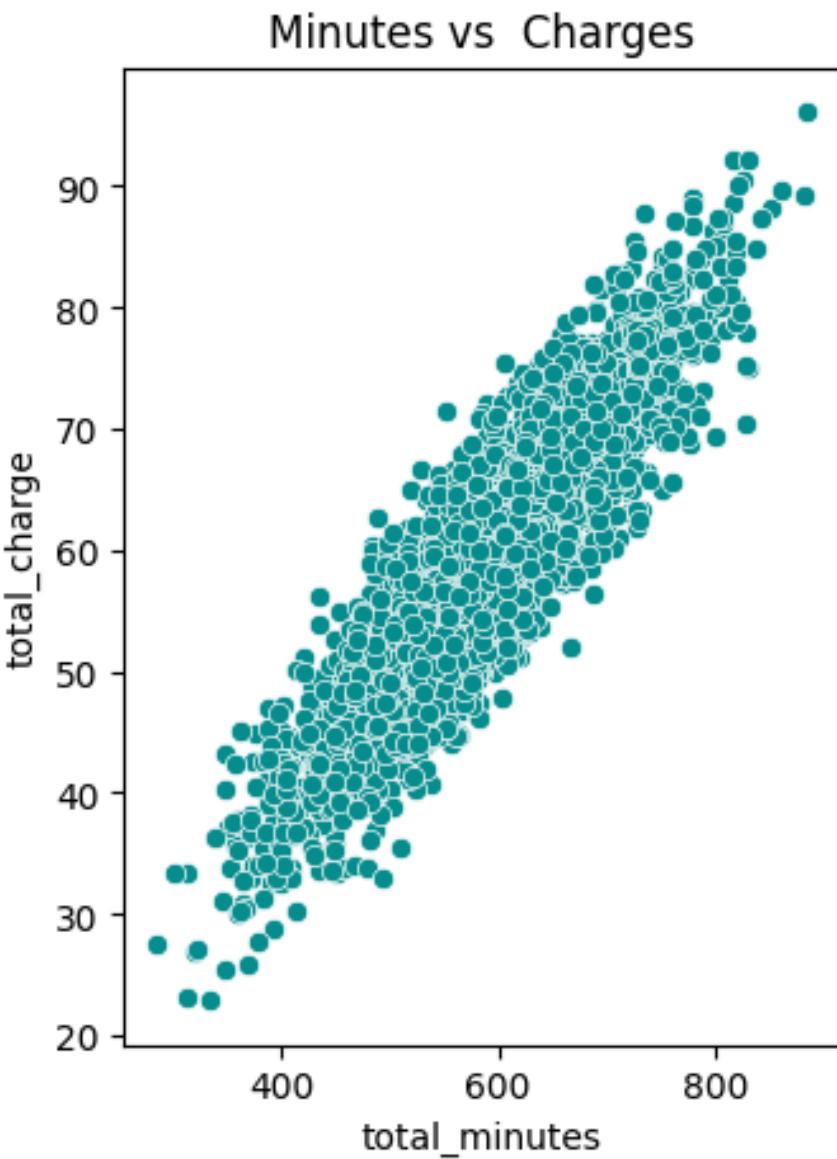
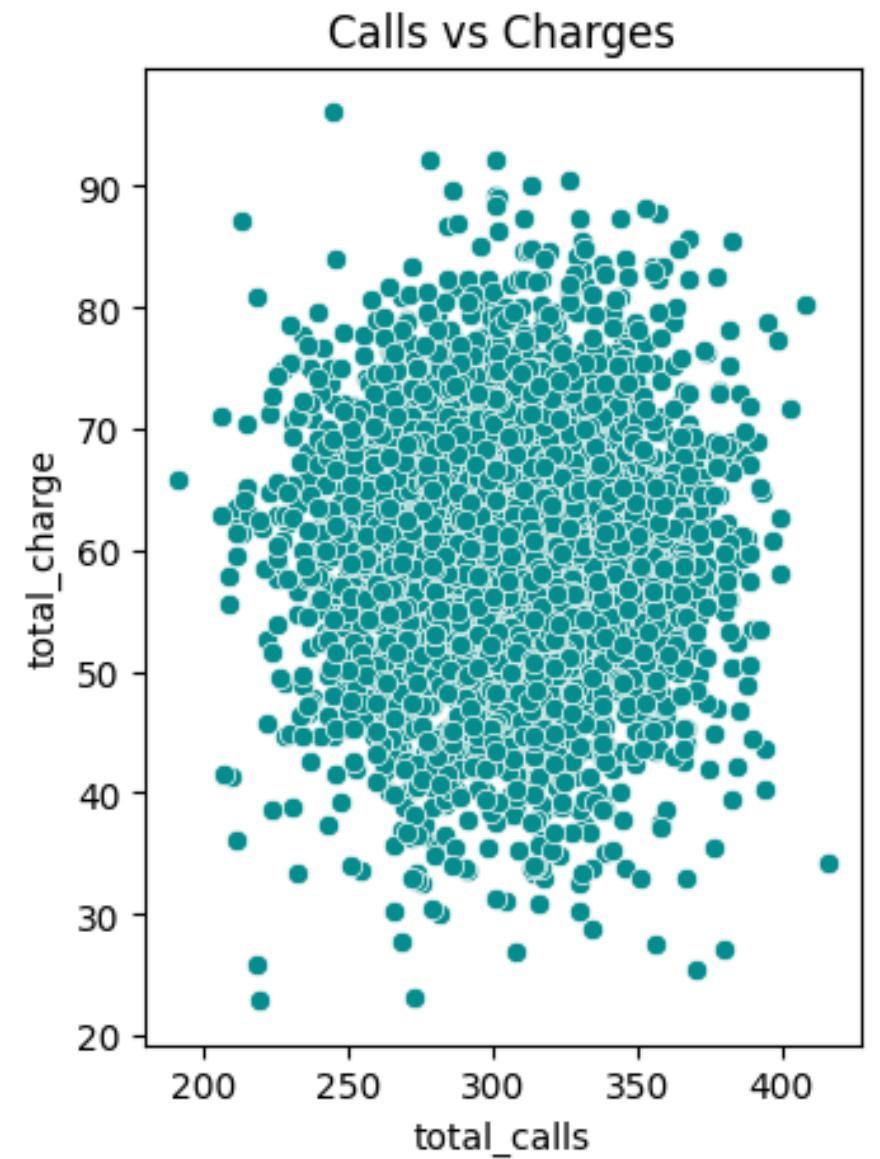


Fig 1:The Scatter plot Indicates low or no relationship between calls and charges

Fig 2:The Scatter plot,Indicates a perfect relationship between Minutes and Charges

Fig 3:The Histogram Indicates that most customer who left the company had contacted customer service more than once.

EXPLORING STATES WRT CHURNING

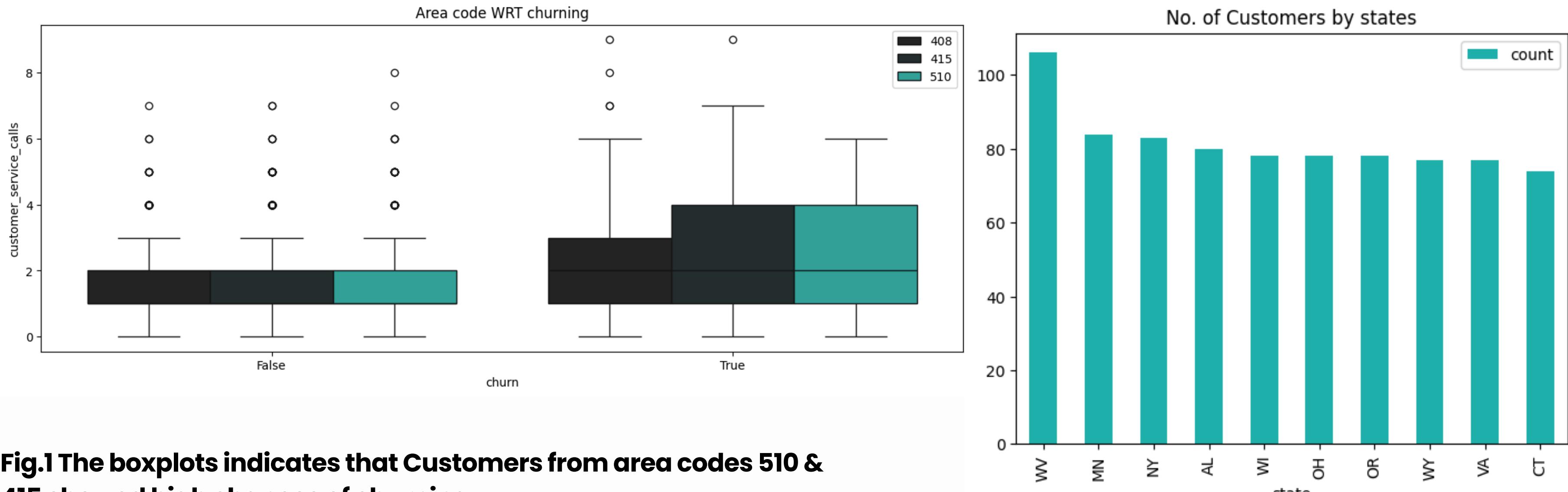
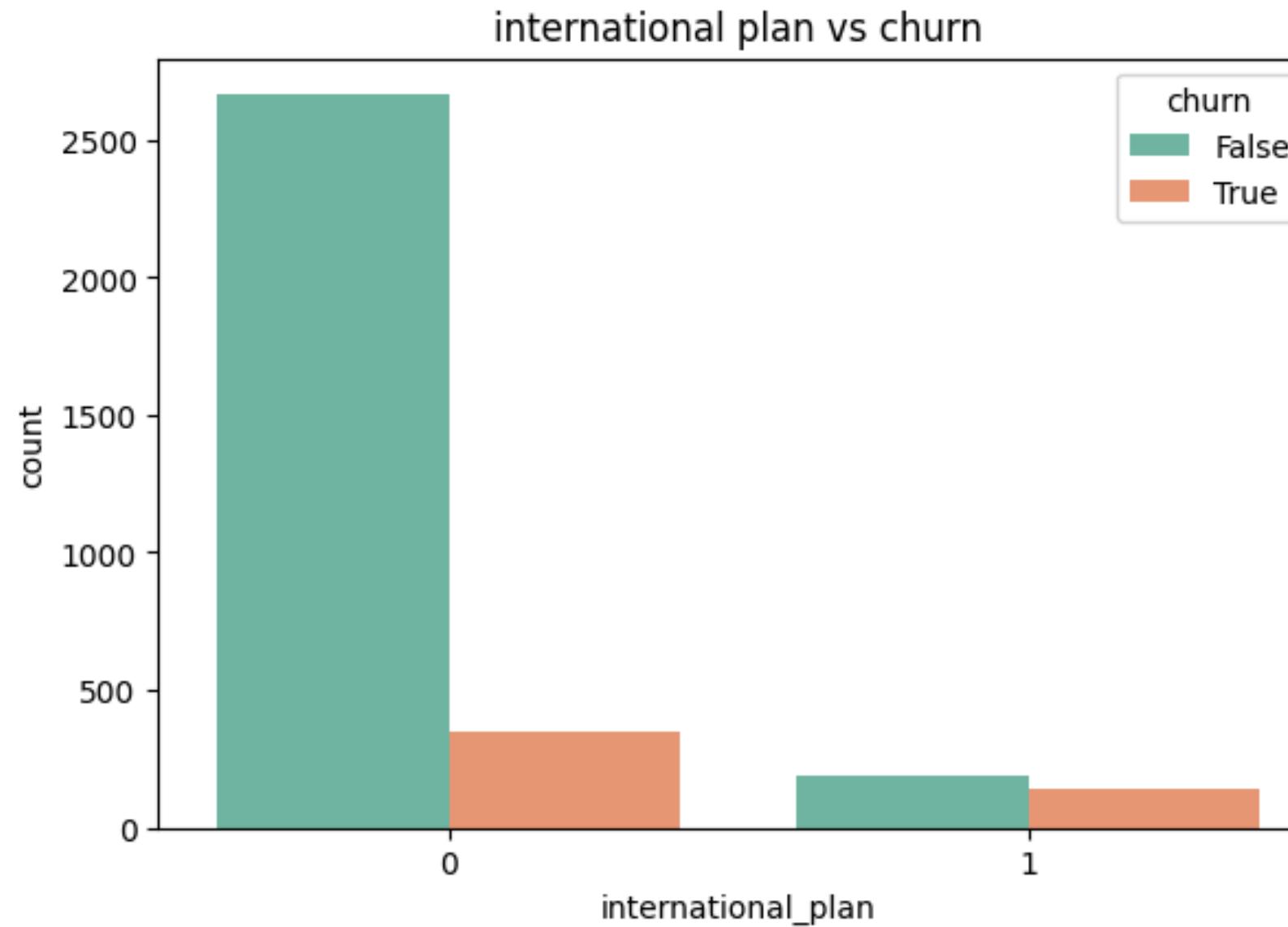


Fig.1 The boxplots indicates that Customers from area codes 510 & 415 showed high chances of churning.

Fig.2. The Histogram indicates that WV, MN, NY, AL, WI, OH, OR, WY, VA and CT are amongst the top ten states that telecom services were commonly used.

HYPOTHESIS TESTING



Hypotheses:

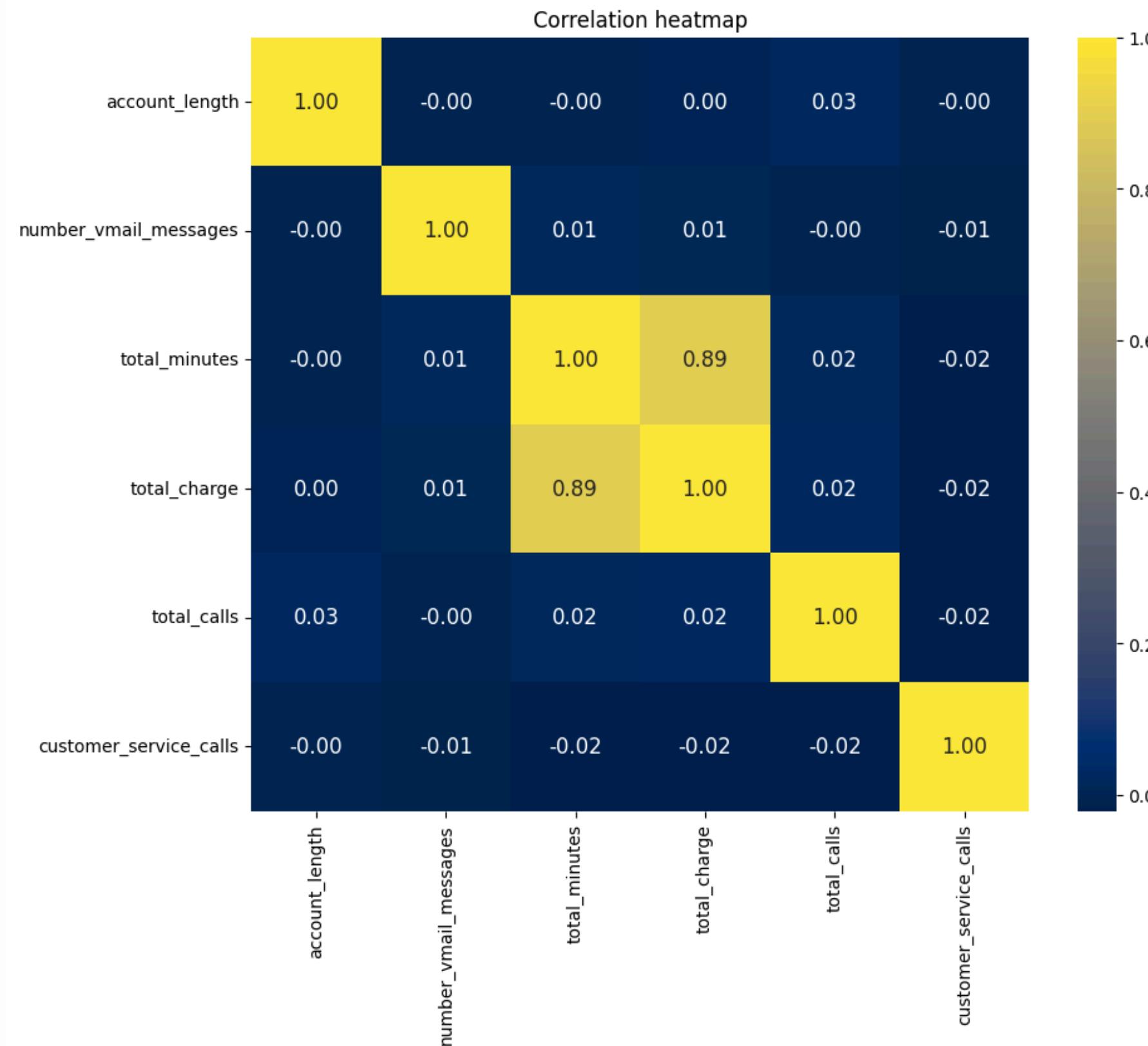
H_0 : Having an international plan has no affect on whether a customer would churn or not

H_A : Having an international plan has an affect on whether a customer would churn or not

TtestResult(statistic=inf, pvalue=0.0, df=1.0)

The p-value is less than the significance level 5%, we can conclude that it is statistically significant that having an international plan has an affect on whether a customer would churn or not.

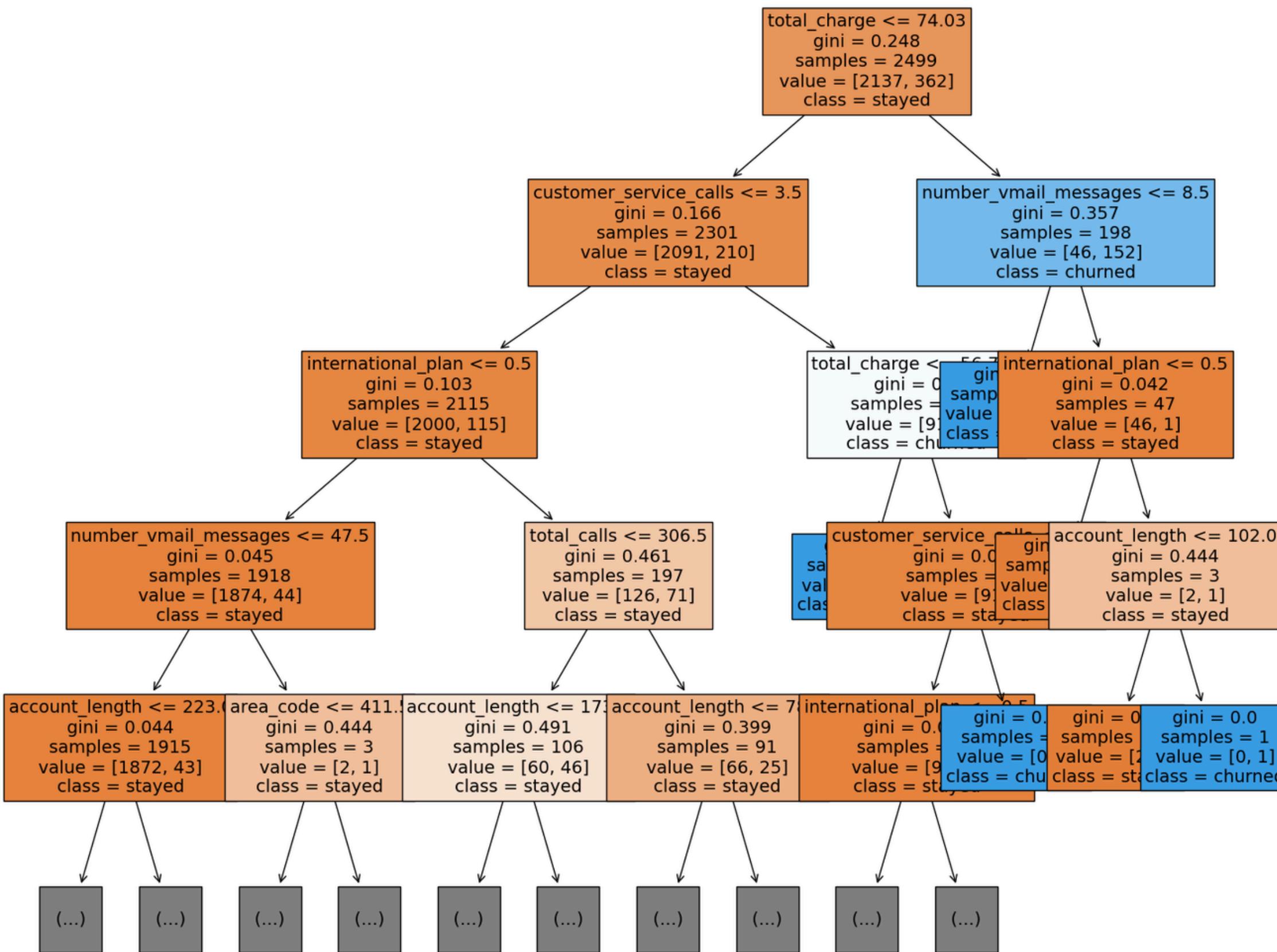
CORRELATION



The correlation heatmap plot shows a negative or no relationship among the variables, which is a good thing as we do not want multicollinearity

There is a perfect correlation between Total charges and total minutes

DECISION TREE MODEL

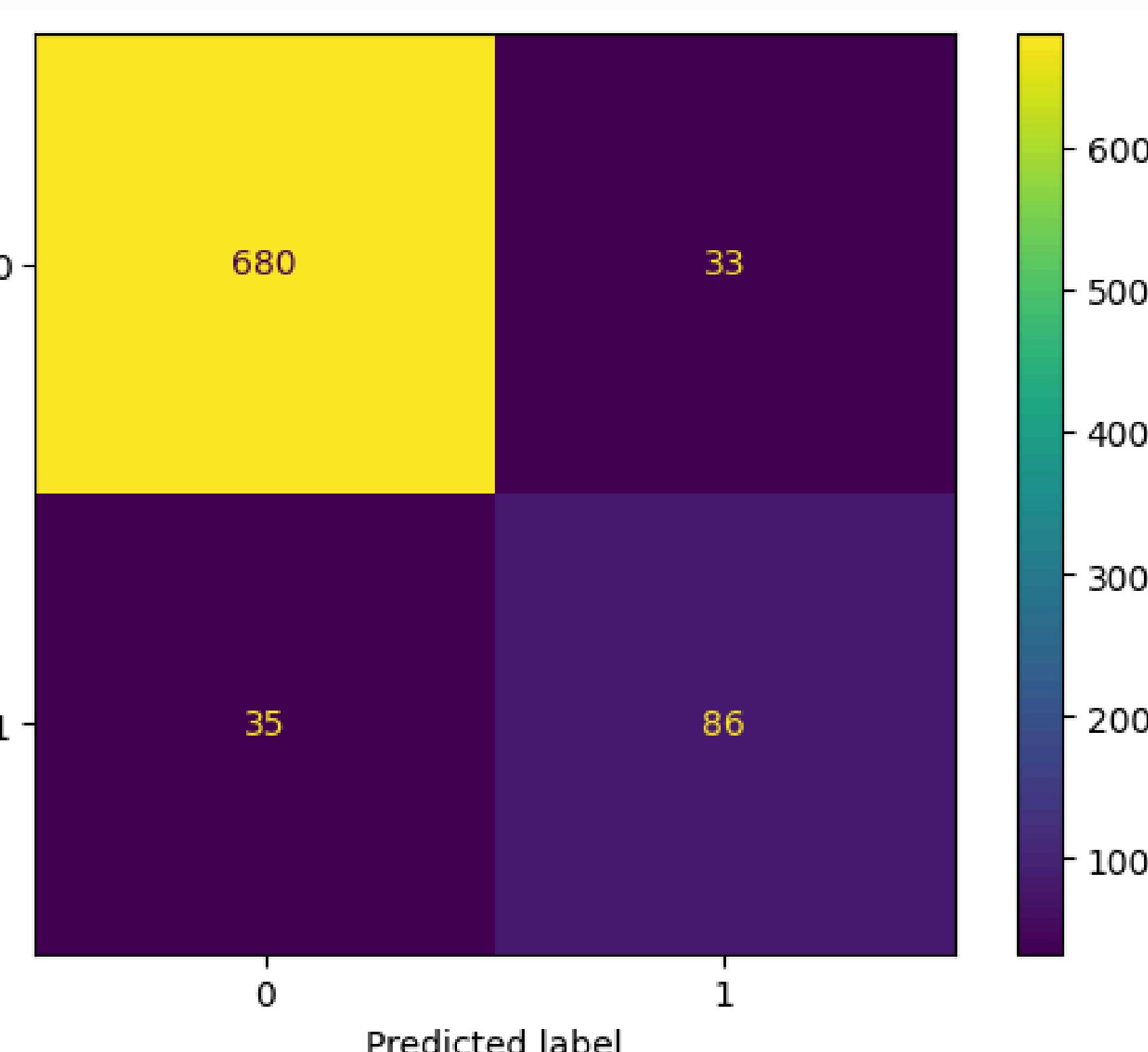


Gini value ranges from 0 to 0.5.

A Gini score of 0 means there is no impurity—the node is a leaf, and all of its samples are of a single class

. A score of 0.5 means the classes are all equally represented in that node.

ANALYZING THE CONFUSION MATRIX



The upper-left quadrant displays the number of true negatives.-680

The bottom-left quadrant displays the number of false negatives.-35

The upper-right quadrant displays the number of false positives-33

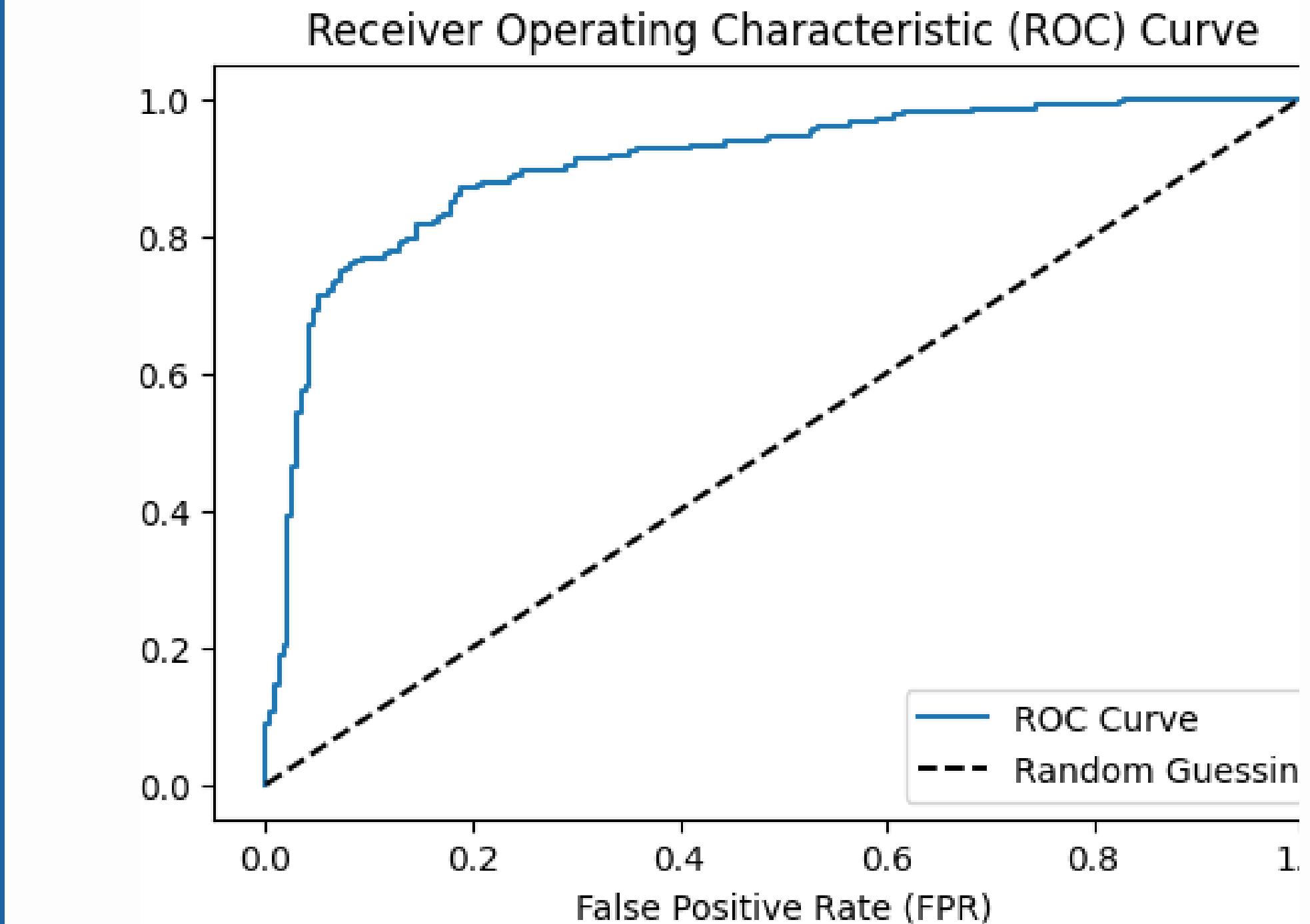
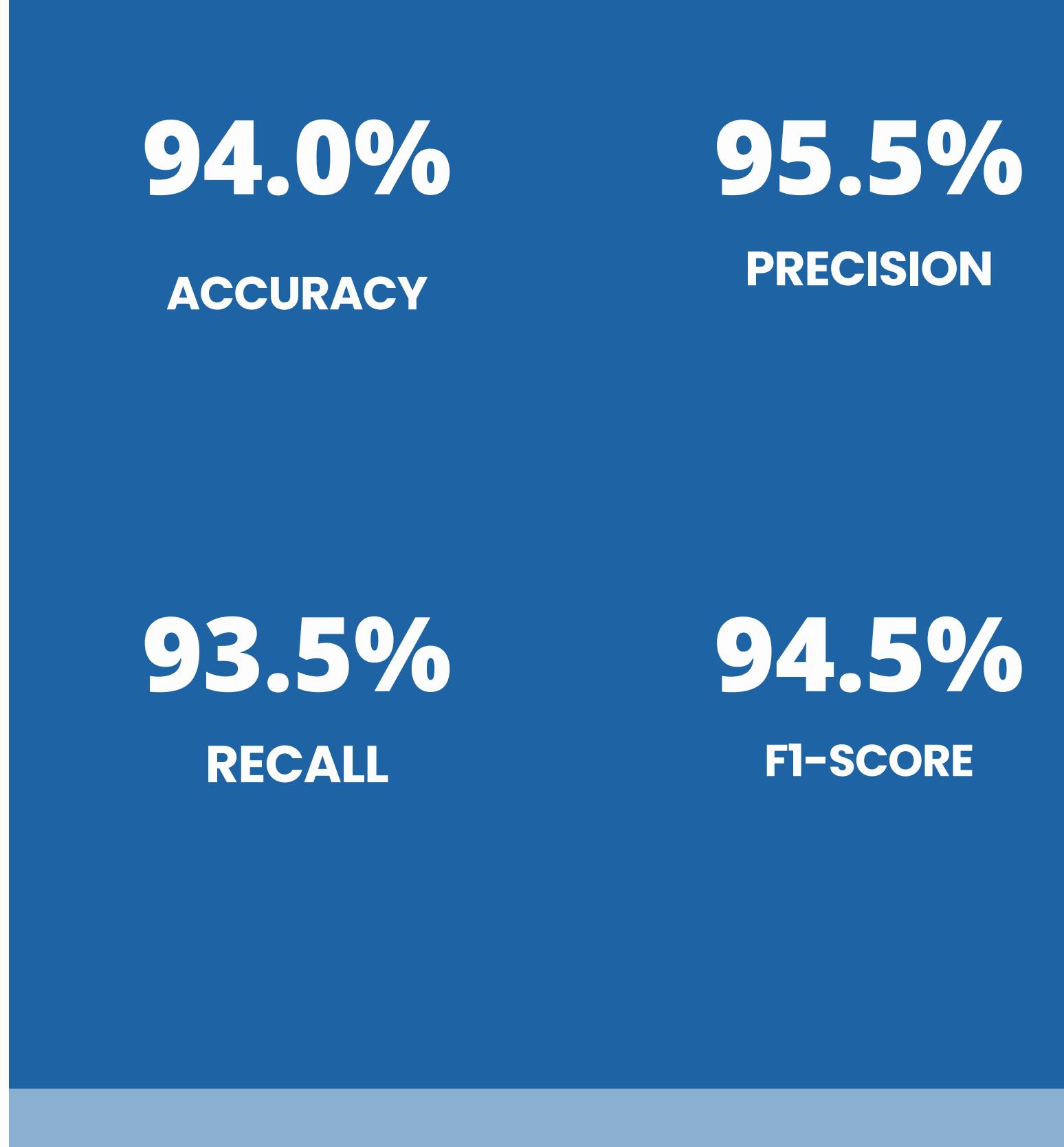
The bottom-right quadrant displays the number of true positives.-86

From this confusion matrix the model correctly predicts many true negatives.-680

It appears slightly more likely to predict a false positive than a false negative.

The decision tree model also predicts lesser false positive -33

ANALYZING THE CLASSIFICATION METRICS OF THE DT



With an accuracy of 94% after tuning the Decision tree with a max depth =5. We therefore conclude that Decision Trees is the perfect model for this classification model.

FINDINGS

- Decision tree model was the best to predict churning among customers in this analysis.
- The classification metrics of the model were: Accuracy-94%, Precision-95%, Recall-94% and F1-Score-95%.

RECOMMENDATIONS

- **I would recommend that Syria Tel Company should use large datasets to test the performance model.**
- **Syria Tel company can do product development more frequently to keep up with the consistent change in customer taste.**
- **Add more enticing features to products according to daily trends in fashion, music, entertainment, sports, holidays.**
- **Invest more on advertising to maintain awareness amongst it's customers ang even gain new ones.**
- **Take customer service seriously, analyze daily calls and ensure all reported issues are resolve in a timely manner. This is because the analysis showed, most customers who churn, had contacted the Customer service more than once.**
- **Reduce charges on minute rate to prevent the 19% of customers who churned due to charges incurred from churning.**



THANK YOU!

CYNTHIA DALMAS



Cakinyi97@gmail.com



NAIROBI, KENYA