

SYRIATEL CUSTOMER CHURN PREDICTION.

Presented by: Esther Francis



OVERVIEW

01

Introduction

02

Problem Statement

03

Objectives

04

Data Understanding

05

Data Preparation

06

EDA

07

Modeling

08

Recommendations



INTRODUCTION

The telecommunications industry is marked by fierce competition and rapid technological advancements, where customer retention stands as a pivotal factor for sustainable growth. Customer churn, the phenomenon of customers discontinuing services, poses a significant challenge for telecom companies like SyriaTel, leading to revenue loss, decreased market share, and escalated acquisition costs.

This project seeks to address the critical issue of customer churn prediction for SyriaTel by leveraging machine learning techniques on available data. By developing a predictive model tailored to SyriaTel's customer base, the aim is to provide actionable insights for strategic decision-making and customer retention efforts. Through an understanding of industry dynamics, SyriaTel's market position, and the significance of churn reduction.



PROBLEM STATEMENT

In the competitive landscape of the telecommunications industry, SyriaTel faces the challenge of retaining its customer base amidst increasing churn rates. Hence, will aim to develop a robust predictive model capable of identifying customers at risk of churn, thereby enabling proactive retention strategies. By analyzing historical customer data, the aim is to uncover patterns and indicators that precede churn behavior.



OBJECTIVES

General Objectives

To build a classifier for SyriaTel for predicting customer churn.

Specific Objectives

- 1 Identify relevant features correlated with churn.
- 2 Develop and train predictive models of an accuracy of at least 90% using machine learning algorithms that will identify customers at risk of churning.
- 3 Interpret model results and derive actionable insights for SyriaTel's business strategy.

DATA PREPARATION



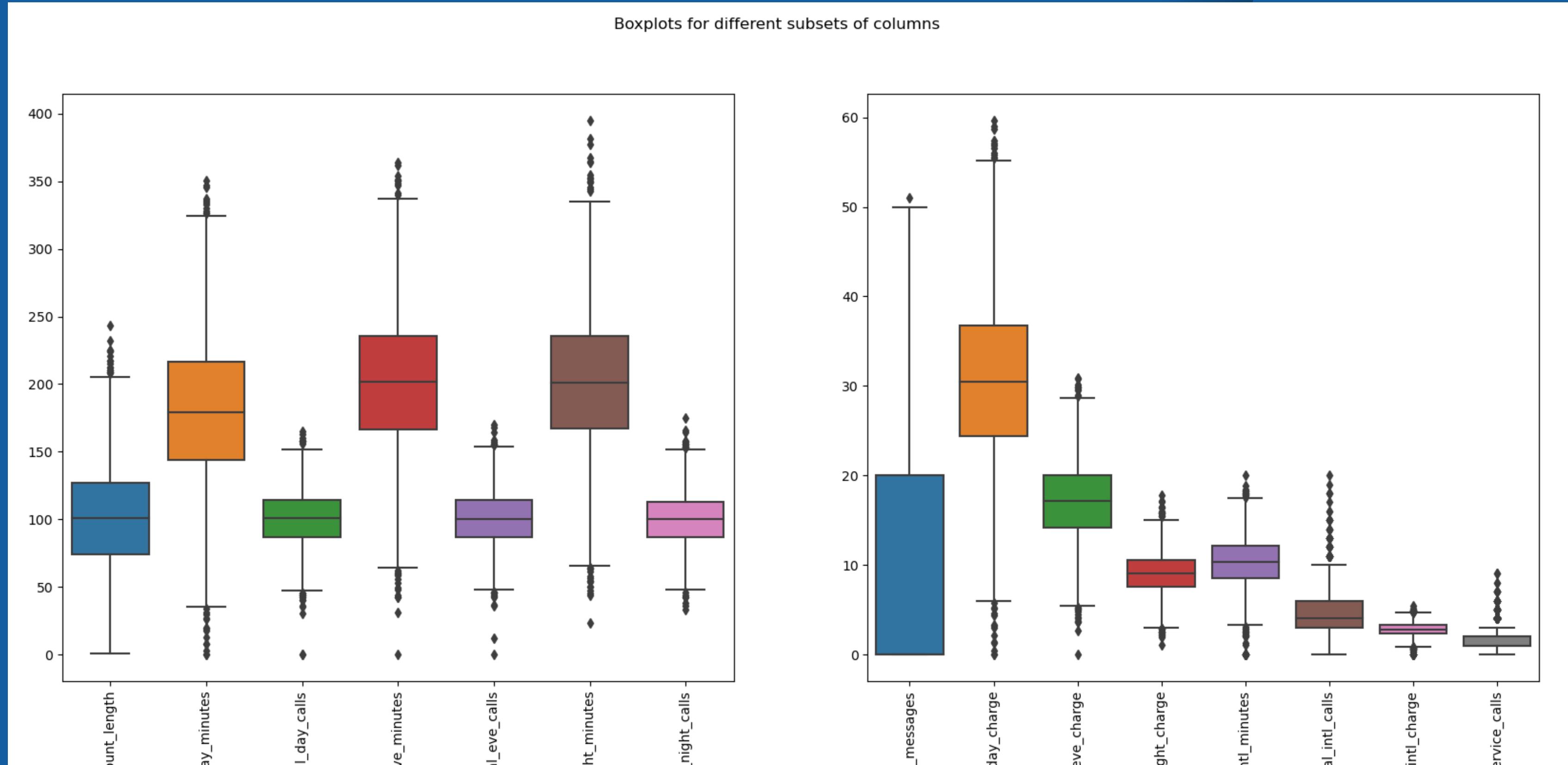
The SyriaTel dataset is from Kaggle, and it contains information about customer activity and whether or not they canceled their subscription with the Telecom firm. The dataset contains 3333 entries and 21 columns.

Summary of Features in the Dataset

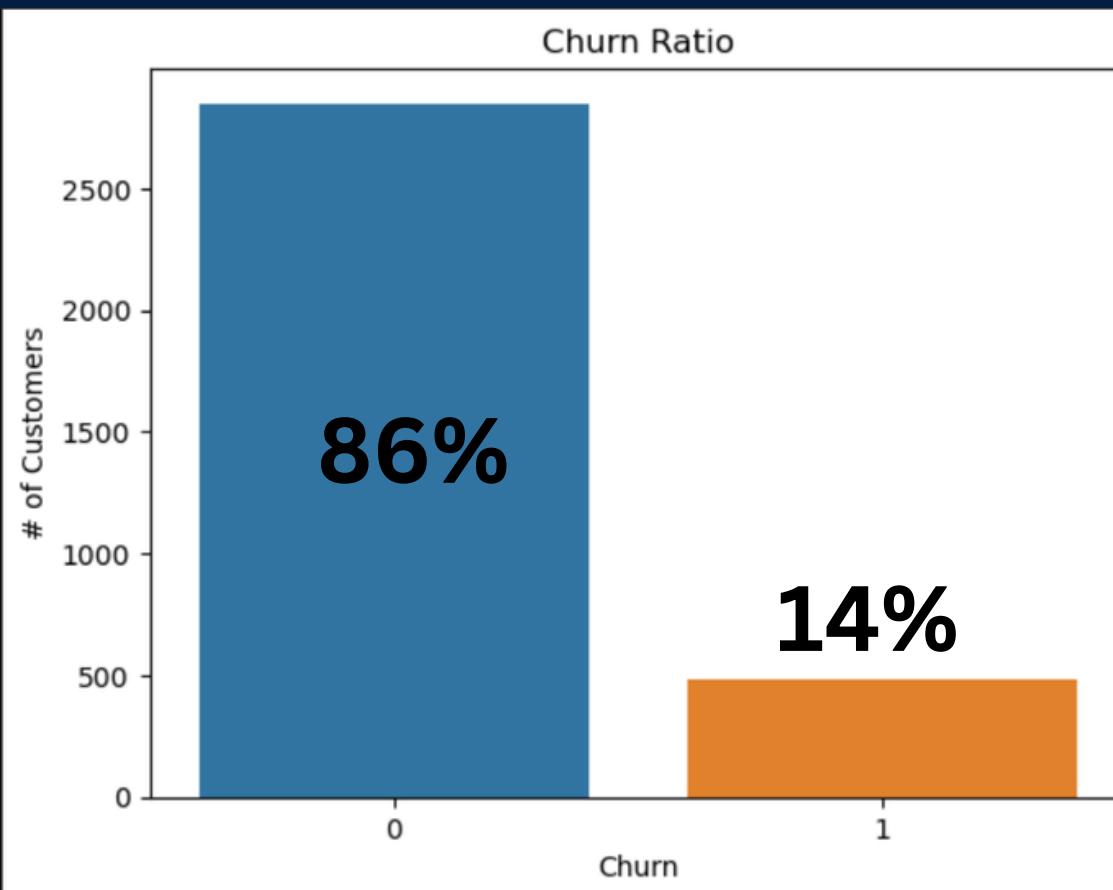
state : Customer's different states.
account_length: active number of days for a customer's account.
area_code : 'customer's location
phone_number : customer's phone number
international_plan : "yes" if the user has the international plan, otherwise "no"
voice_mail_plan : "yes" if the user has the voice mail plan, otherwise "no"
number_vmail_messages : customer has a vmail plan and how many vmail messages do they get
total_day_minutes : total number of call minutes used during the day
* total_day_calls : total number of calls made during the day
* totalday_charge : day calls' total charge
* total_eve_minutes : total number of call minutes used in the evening
* total_eve_calls : total calls made in the evening
* total_eve_charge : total charge on evening calls
* total_night_minutes: Total number of call minutes used at night
* total_night_calls : Total number of night calls
* total_night_charge : Total charge on night calls
* total_intl_minutes : total international minutes used
* total_intl_calls : total number of international calls made
* total_intl_charge : total charge on international calls
* customer_ervice_calls : number of calls made to customer service
* churn : boolean on whether the customer left or not

DATA PREPARATION

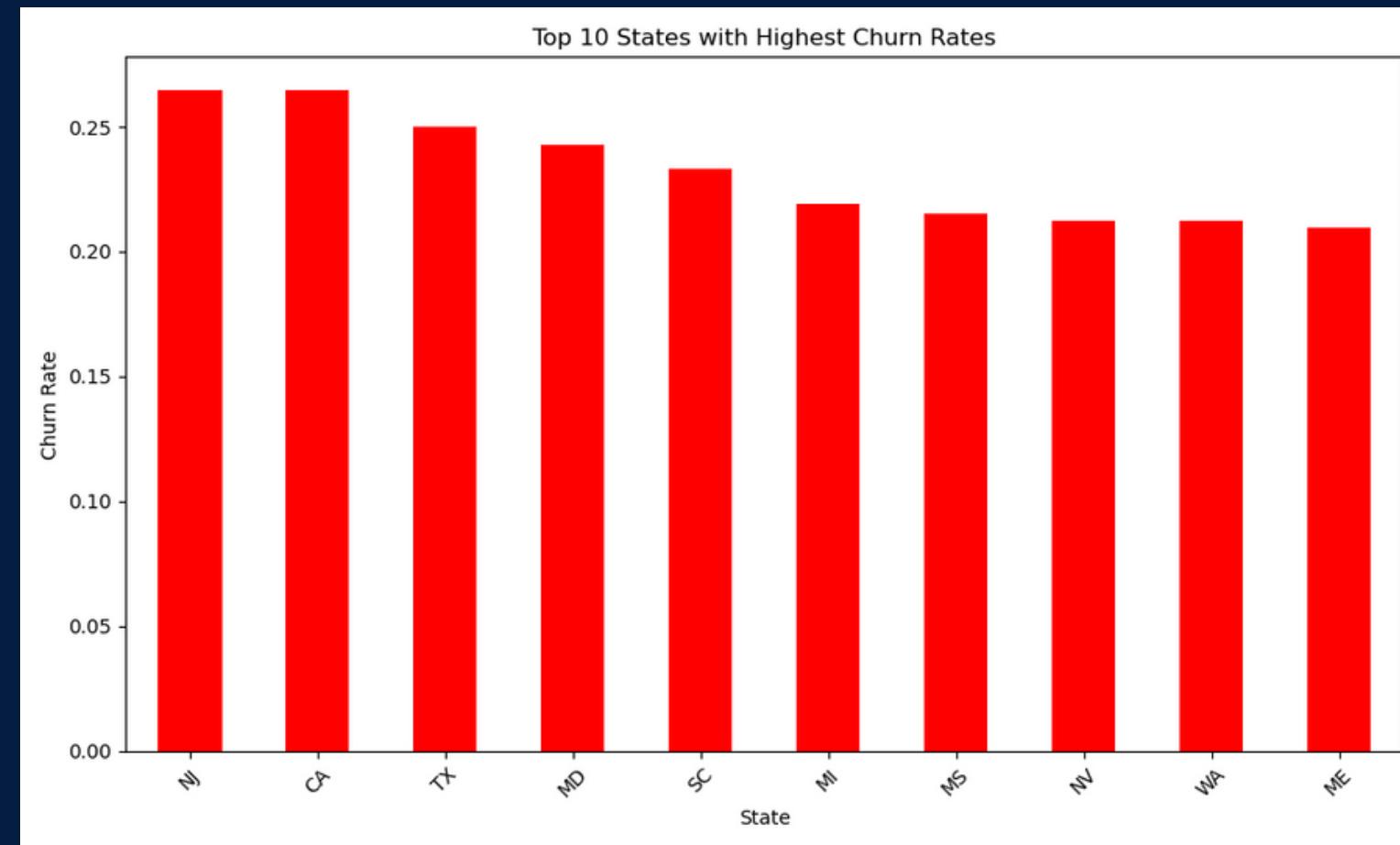
Dropped unnecessary rows, like area code since we had states, Made the phone number the unique identifier, and checked for outliers to help analyze our data.



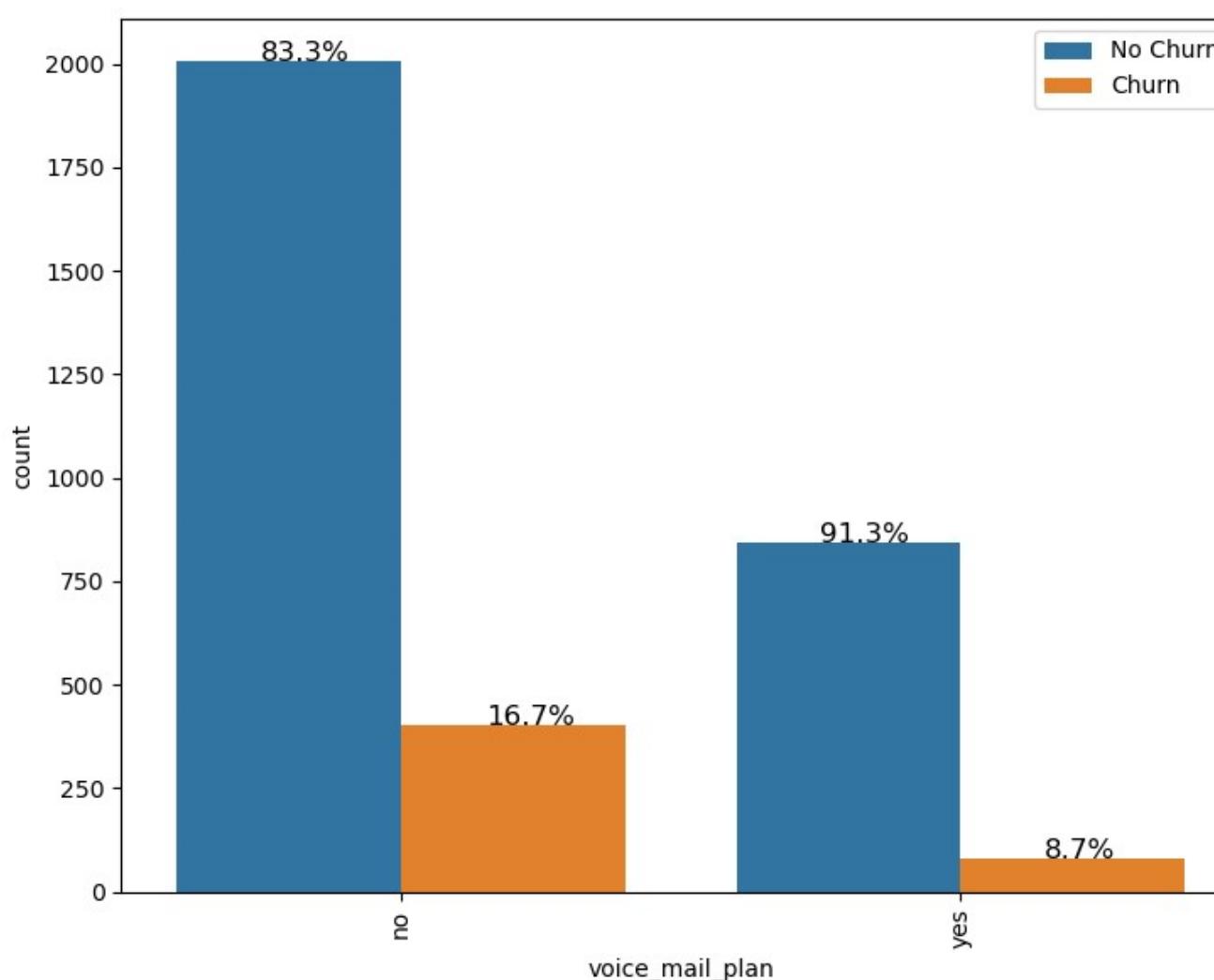
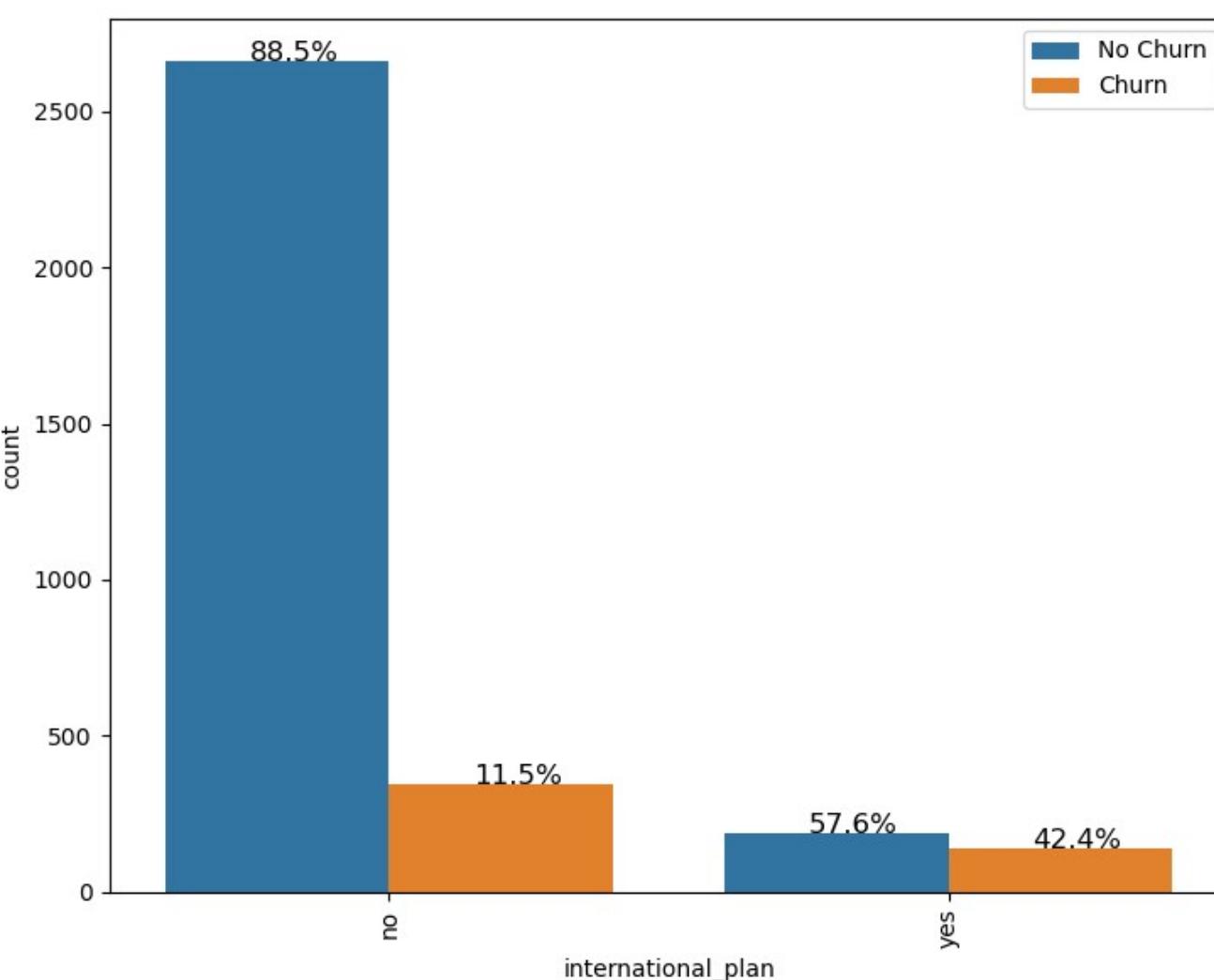
EXPLORATORY DATA ANALYSIS



Percentage of Customers Churning
It may indicate that just a small percentage of 14% of customers are churning but it posses a great challenge to the stakeholders and the new recommendations from the data scientist should be taken into considerations.

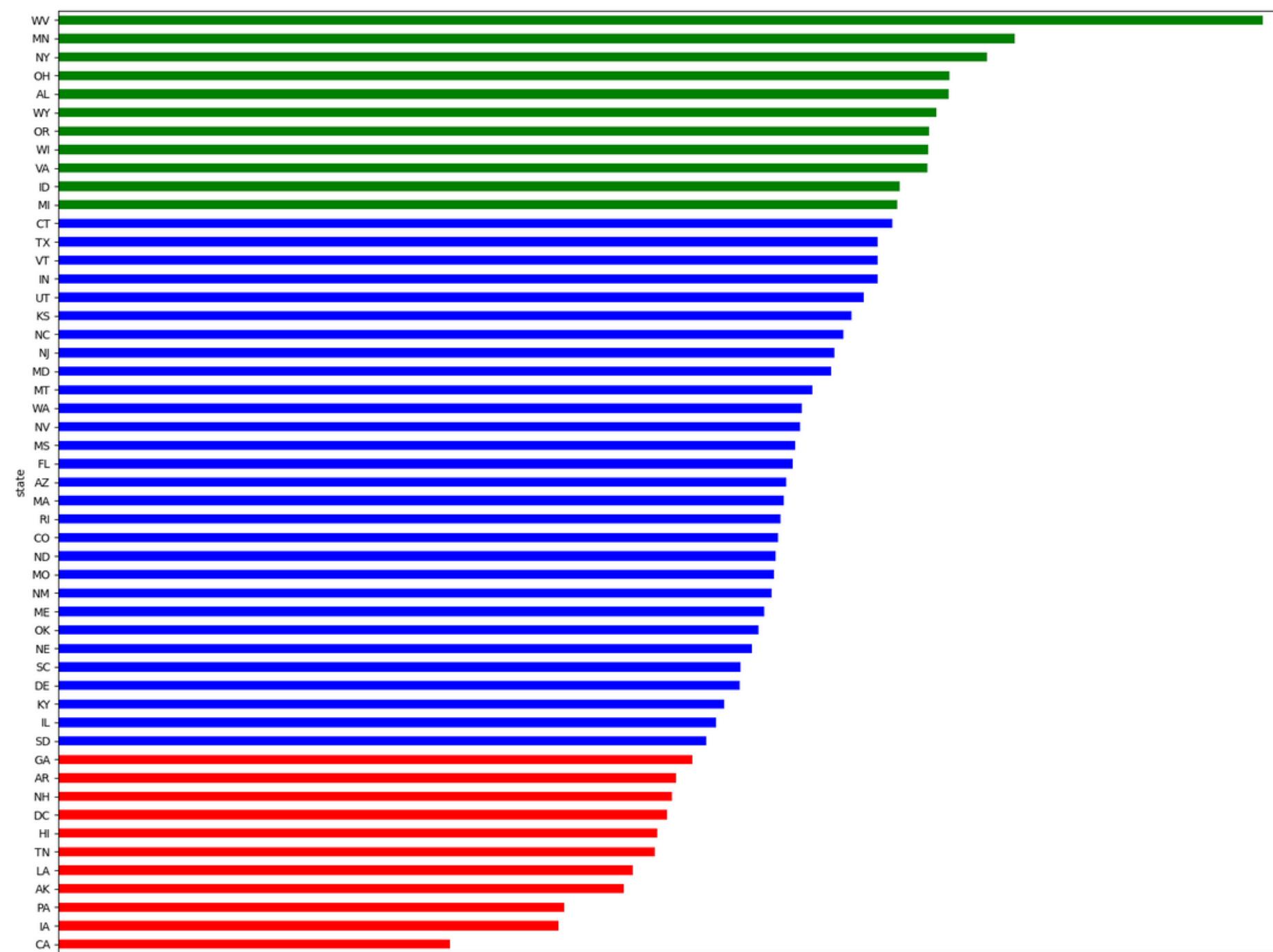


Top 10 states with the highest churning rate that stakeholders should note



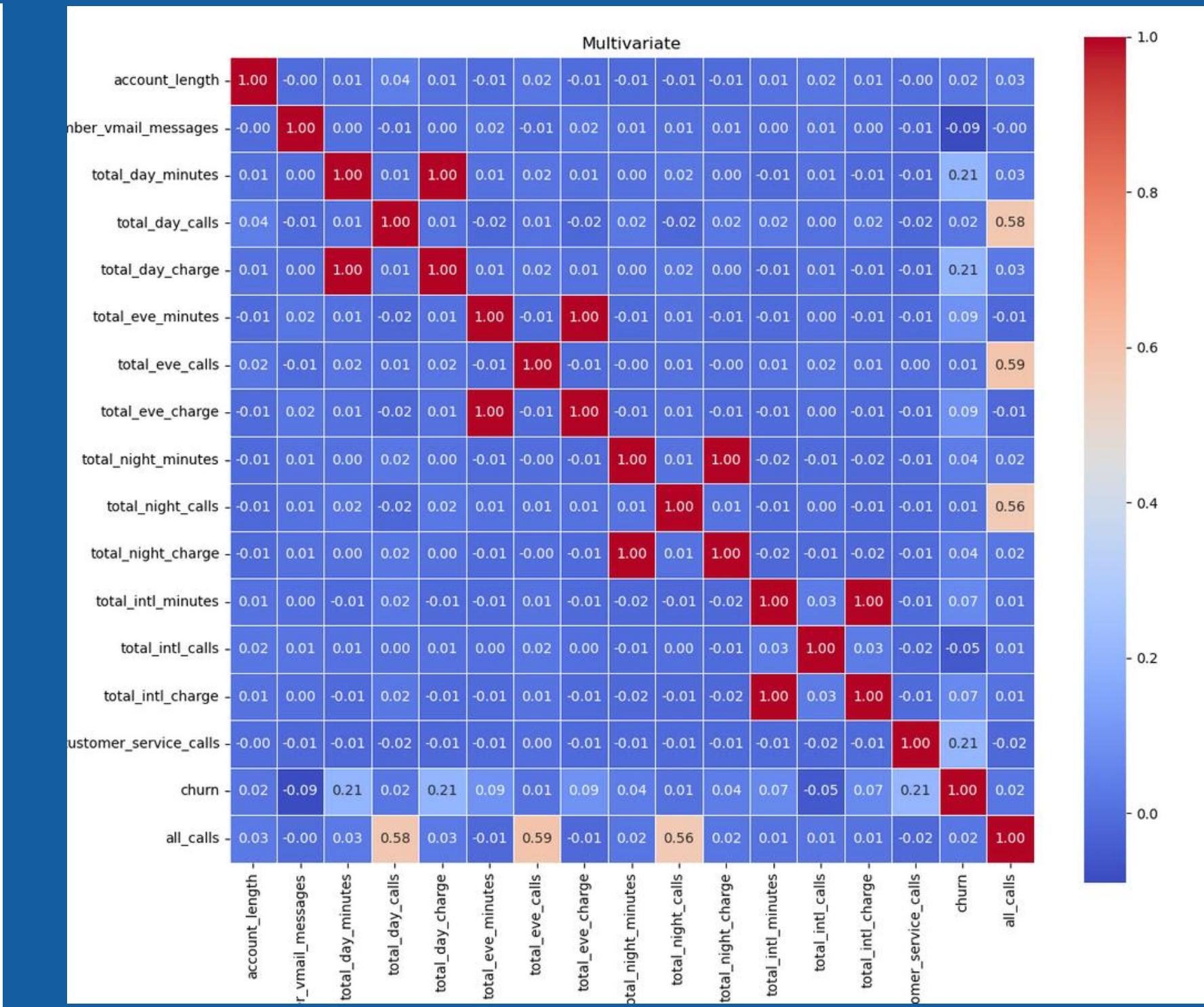
1. A higher proportion of customers who subscribed to the `international_plan` churned. The observation gives an impression that subscribing to the `international_plan` may be increasing the chances of customer churn. Hence stakeholders should investigate if there are any issues associated with the `international_plan` that could be making the plan unsatisfactory to the client.
2. A lower proportion of customers who subscribed to the `voice_mail_plan` churned. This suggests that subscribing to the `voice_mail_plan` may be associated with a lower likelihood of churning.

EXPLORATORY DATA ANALYSIS CONT'...



Total calls made per State

The states highlighted in red made the least number of calls, and states like CA that is at the bottom is topping in the list of states with the highest rate of churning, which indicates that stakeholders should consider to review.



Correlation between variables

There is no much correlation between the variables.



MODELING

Logistic Regression Model

Train Accuracy: 0.85

Test Accuracy: 0.85

Classification Report (Test Data):

	precision	recall	f1-score	support
0	0.86	0.99	0.92	857
1	0.39	0.05	0.09	143
accuracy			0.85	1000
macro avg	0.63	0.52	0.50	1000
weighted avg	0.79	0.85	0.80	1000

Confusion Matrix (Test Data):

```
[[846  11]
 [136   7]]
```

K-Nearest Neighbors

Accuracy: 0.885

Precision: 0.7333333333333333

Recall: 0.3076923076923077

F1-score: 0.43349753694581283

0.9151307329618517

0.885

Decision Tree Classifier

Accuracy: 0.906

Classification Report:

	precision	recall	f1-score	support
0	0.95	0.94	0.94	857
1	0.66	0.70	0.68	143
accuracy			0.91	1000
macro avg	0.81	0.82	0.81	1000
weighted avg	0.91	0.91	0.91	1000

Confusion Matrix:

```
[[806  51]
 [ 43 100]]
```

XGBoost

Accuracy: 0.956

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.99	0.97	857
1	0.93	0.75	0.83	143
accuracy			0.96	1000
macro avg	0.94	0.87	0.90	1000
weighted avg	0.96	0.96	0.95	1000

Confusion Matrix:

```
[[849  8]
 [ 36 107]]
```

RECOMMENDATIONS



We recommend to the stakeholders to adopt the XGBoost model for predicting customer churn, since it achieved an accuracy of 95.1% when classifying customer churn. This means that it correctly predicted 95.1% of the cases.



Check on states like NJ, CA, TX, MD, SC, MI, MS, NV, WA, ME which have the highest churn rate. There could be different reasons that lead to customer churning.



Feedback Mechanisms, by creating feedback mechanisms that encourage customers to provide input on their customer service experience will improve service quality. This can be due to the fact that there was a higher proportion of customers who subscribed to the international_plan churning with a percentage of 42.4% which is more than those customers that did not subscribe at 11.5%.

Hence stakeholders should investigate if there are any issues associated with the international plans that could be making the plan unsatisfactory to the client. And this can be achieved by developing a feedback mechanism.



Regularly Review and Update Plans, since the telecommunications industry is highly competitive and dynamic. Regularly assess and update your service plans to ensure they remain competitive and meet evolving customer needs. Offering flexibility and customization options can be an attractive feature for customers.



There is an association with total calls made (total day calls, total evening calls, total night call) from a state and its probability of having a higher number of customers churning. Like CA had the lowest total calls made and it was leading in the percentage of having numbers of customers churning. As stakeholders increase and work on their marketing, the total call should be considered.

Thank you!

