SYRIATEL CUSTOMER CHURN PREDICTION MODEL:

ENHANCING RETENTION STRATEGIES WITH ADVANCED MODELING

TABLE OF CONTENTS

1. BUSINESS PROBLEM

- Syriatel Communications goal is to identify customers at risk of leaving, analyze the key factors contributing to churn, and offer actionable recommendations to boost customer retention.
  - **Goals**

1.Identify the primary factors leading to customer churn.

2.Build a prediction models to predict which customers are most likely to leave. A prediction model with a recall accuracy higher than **0.75** is preferred for the prediction.

3.Offer recommendations to proactively retain customers and reduce churn.

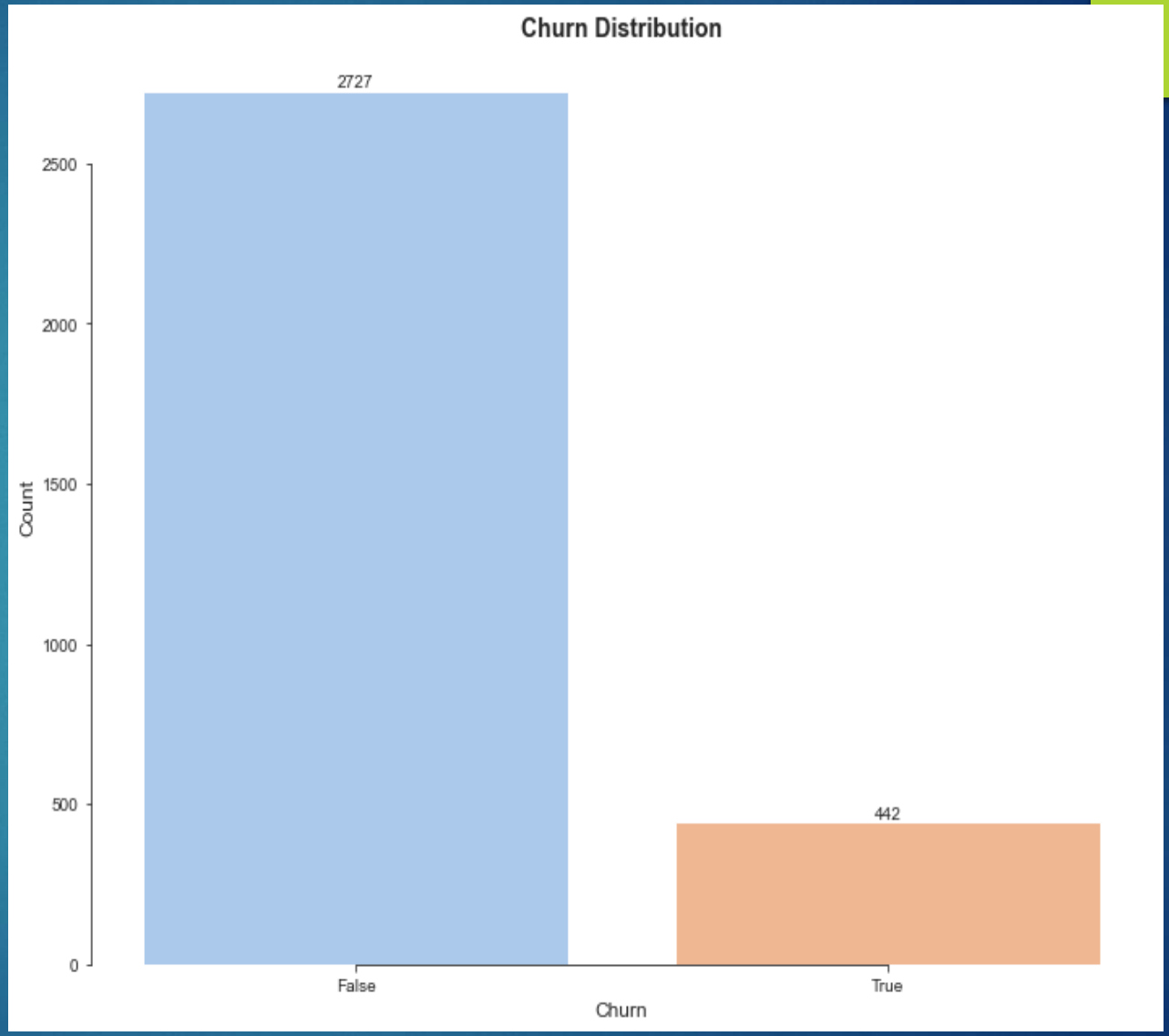
2. DATA ANALYSIS

- Data Exploration - where exploratory data analysis (EDA) to uncover patterns, correlations, and insights within the data, helping us to better understand the relationships between different variables.-:
  a) Churn Distribution
  b) Numerical Features Distribution
  c) Categorical Features Distribution
- Feature Engineering - new features were created and existing ones transformed, optimizing them to improve the performance of our predictive models.

DATA ANALYSIS: Data Exploration- Churn Distribution

Among the **3169** customers, **442** have churned, meaning they have cancelled their contracts with the company, while 2727 have remained.

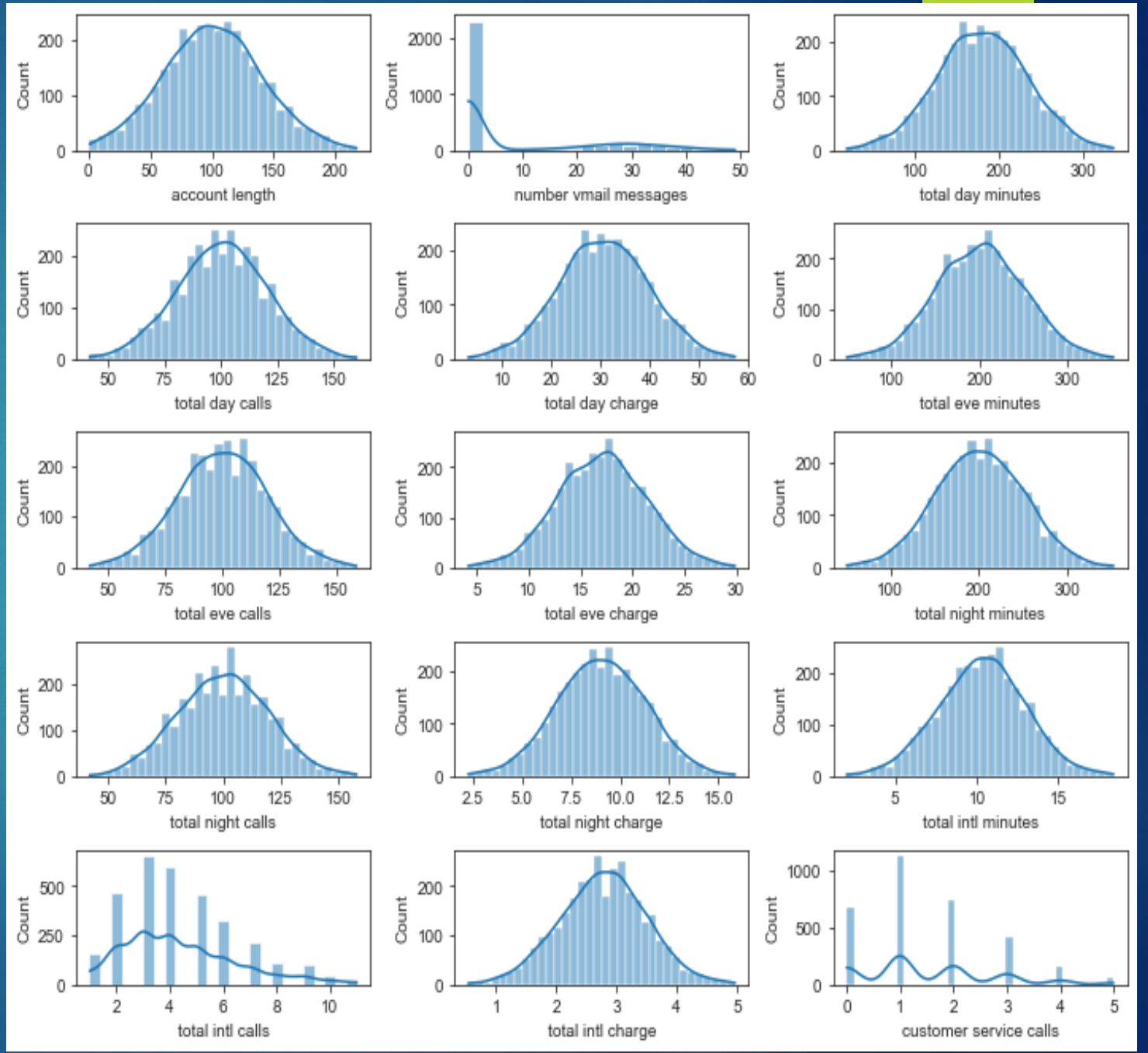
This distribution highlights a noticeable class imbalance, which needs to be addressed before modelling, as it could lead to skewed or inaccurate predictions.



DATA ANALYSIS: Data Exploration- Numerical Features distribution

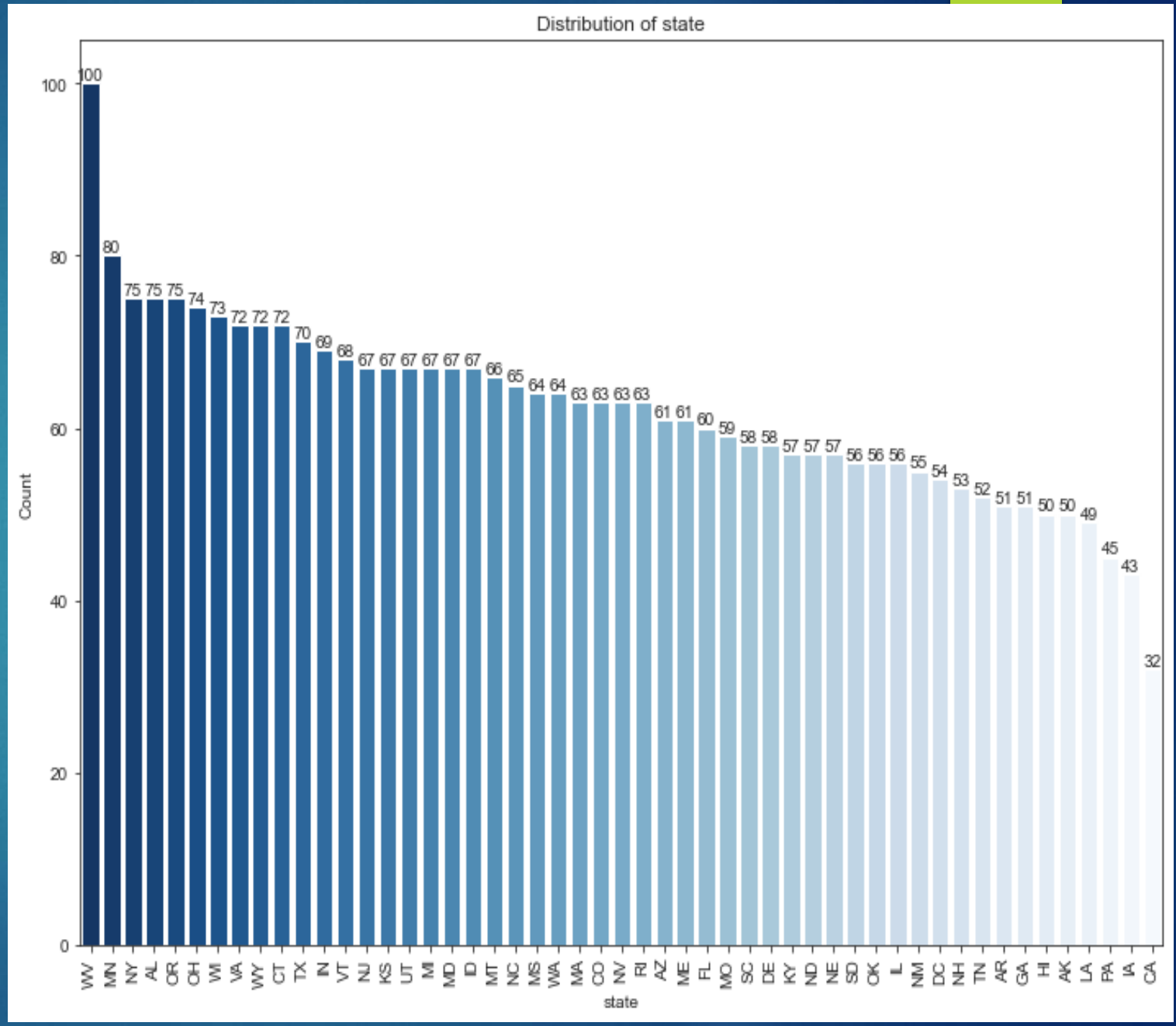
The majority of the features are normally distributed, especially usage-related ones (**minutes, calls, charges**), which suggests **uniform** customer behavior in these aspects.

Features like "**number of voicemail messages," "total intl calls," and "customer service calls"** show **skewed** distributions, indicating that only a minority of customers engage in these activities frequently.



DATA ANALYSIS: Data Exploration- Categorical Features distribution

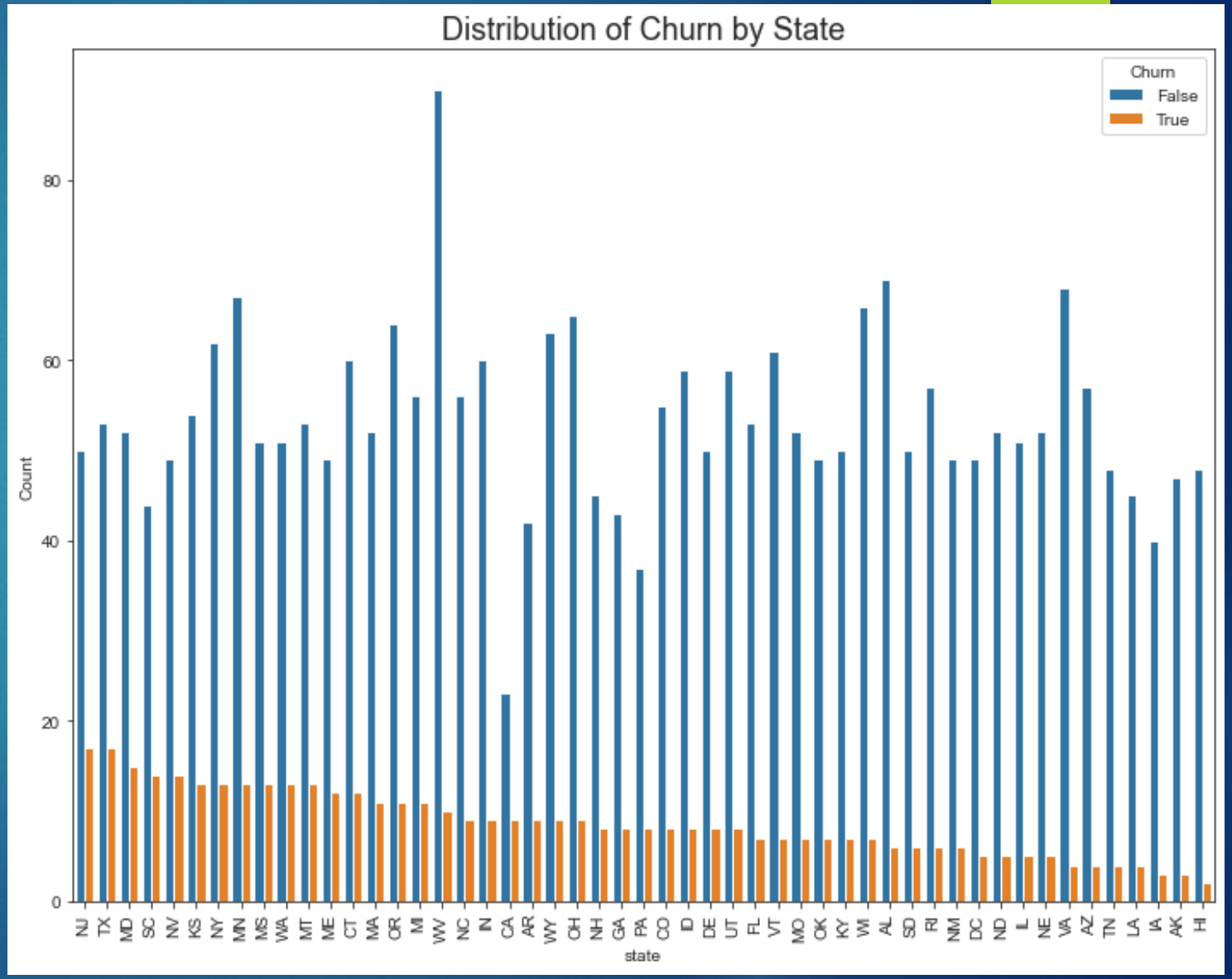
From the distibution the top 5 areas that have the most customers are : **West Virginia WV, Minnesota MN, New York NY, Ohio OH and AL Atlanta.**



DATA ANALYSIS: Data Exploration- Categorical Features distribution

The top 5 states with the highest churn (Orange bars) are:

WV (West Virginia), NJ (New Jersey), CA (California), TX (Texas) & OH (Ohio)



DATA ANALYSIS: Data Exploration- Categorical Features distribution

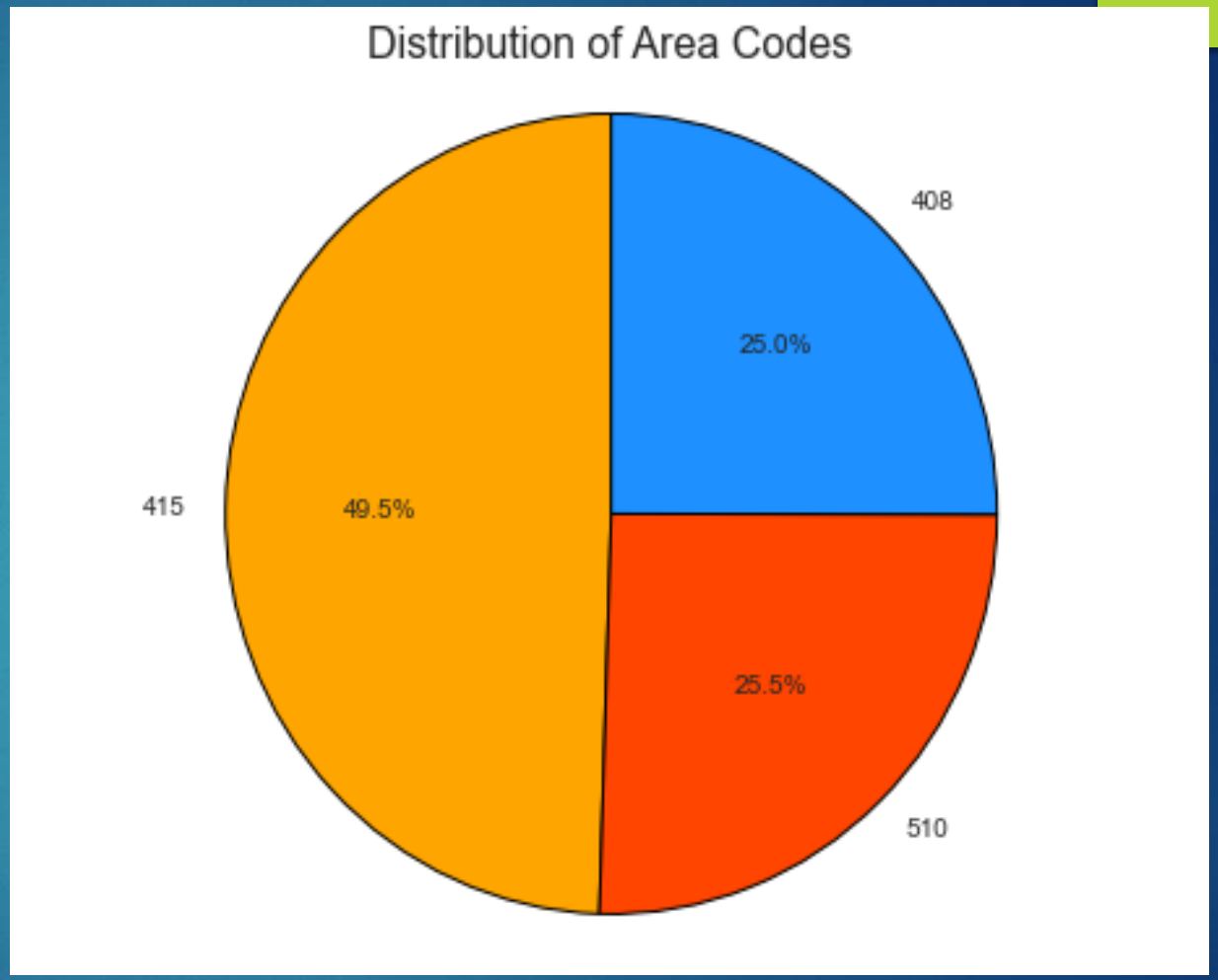
The pie chart shows the distribution of area codes among customers, with three distinct segments:

Area Code 415 represents the largest segment at **49.5%** of the total, indicating that this area has the highest concentration of customers. This area is likely to have the highest churn rate.

Area Code 510 constitutes **25.5%** of the total, which is the second-largest segment, suggesting a moderate customer base and potentially a significant churn rate.

Area Code 408 makes up **25.0%** of the total, representing the smallest segment, with a potentially lower churn rate compared to the other areas.

The largest segment, Area Code 415, is of particular concern as it likely has the highest churn rate, making it a critical focus for targeted retention strategies.



3. DATA MODELING

- Machine learning models are implemented to predict customer churn based on selected features.
- A **recall score** of 0.75 or higher will signify that the model is highly effective in identifying churners.
- Models that will be used are-:
  a) Logistic Regression
  b) Decision Tree
  c) Random Forest
  d) XGBoost.

The goal is to determine which model performs best in terms of recall. To further assess model performance, we will also utilize the **ROC-AUC** (Receiver Operating Characteristic - Area Under the Curve) metric. This comprehensive approach will allow us to select the most reliable model for churn prediction.

DATA MODELING: Recall Scores

Based on the recall scores for each model, **XGBoost** achieved the **highest** recall score of **0.7857**

This indicates that XGBoost is the best model for identifying customers at risk of churn in this case

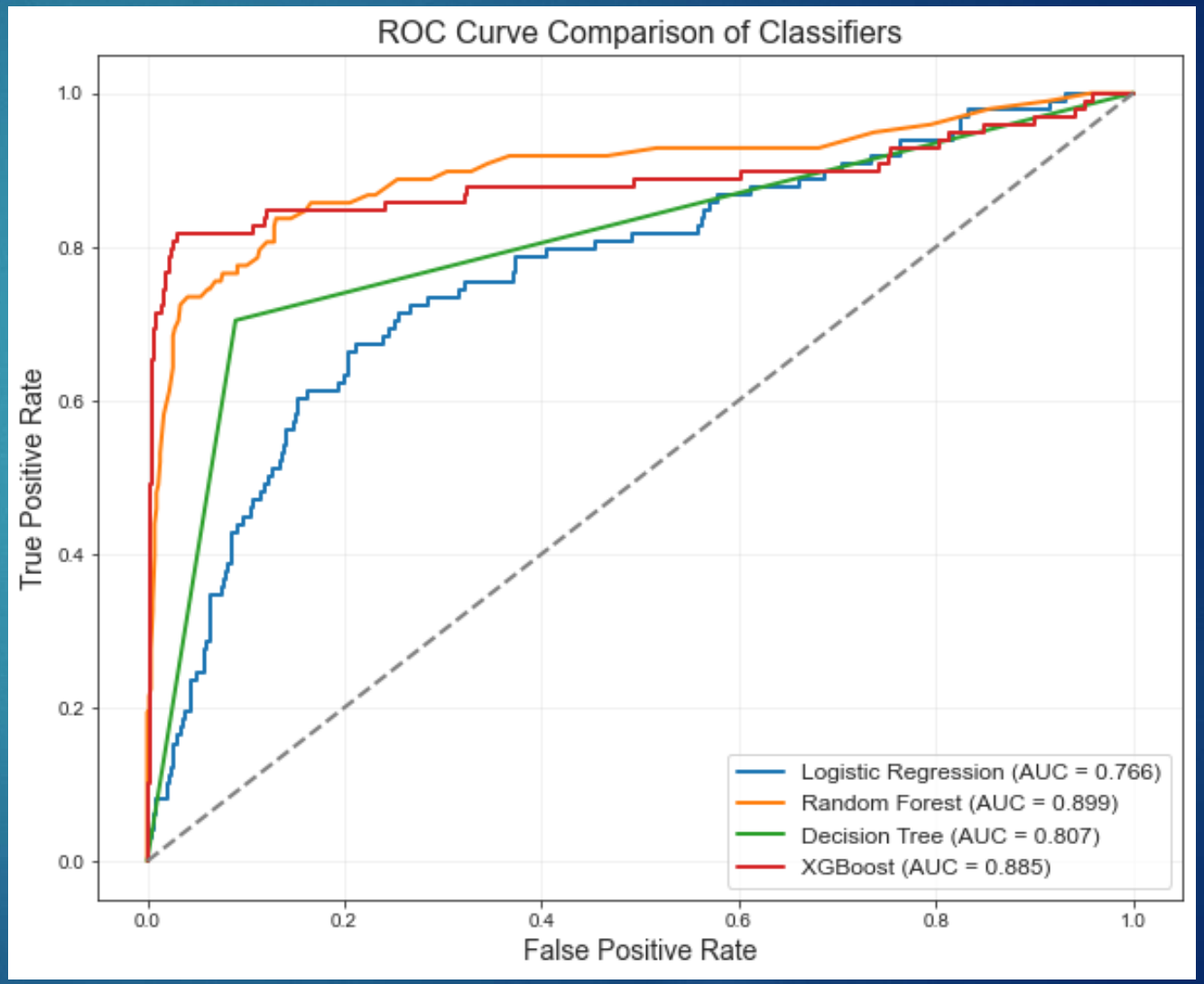
classifier	recall
Logistic Regression	0.673469
Random Forest	0.734694
Decision Tree	0.704082
XGBoost	0.785714

DATA MODELING: ROC Curve

Here **Random Forest** performs the best by having the highest **AUC = 0.899.**

Given the goal is to maximize predictive performance, Random Forest and XGBoost would be the preferred models for predicting churn.

They provide the best balance between true positive rate and false positive rate.



4. CONCLUSION

- **XGBoost** had the highest **recall score (0.786**), which indicates that it successfully identifies more churners compared to other models.
- However, **Random Forest** showed the highest **AUC (0.899)** on the ROC curve, meaning it generally performs better in balancing true positives and false positives, which makes it a strong overall model.
- Since the problem overview of SyriaTel seeks to create a predictive model to anticipate customer churn, **XGBoost would be the best choice.** It is better at catching potential churners even if it might come at the cost of a higher number of false positives.

5. RECOMMENDATION

1. Pricing Strategy Review:

Reassess pricing structures across various time periods (daytime, evening, and night) as well as international charges. Offering **discounted pricing plans** during peak times or for international calls may help retain customers who are currently considering switching due to cost concerns.

2. Area code Targeted Discounts and Promotions:

Focus on area code 415 , where the churn rate is notably high. Providing **personalized discounts or promotions** for customers in these regions could incentivize them to remain loyal to the company.

3. Improved marketing strategies

To address high churn rates in states like WV (West Virginia), NJ (New Jersey), CA (California), TX (Texas) & OH (Ohio) , it's essential to **implement targeted campaigns focused on customer retention.** Enhancing marketing strategies in these regions can include personalized offers, loyalty programs, and proactive customer support. By understanding the specific needs and pain points of customers in these states, you can create more effective engagement initiatives that strengthen customer relationships and reduce churn.

PRESENTATION BY:

ANGELA MAINA