# SYRIATEL CUSTOMER CHURN PREDICTION

By **RODGERS NDEMO**

# **BUSINESS PROBLEM:** REDUCING CUSTOMER CHURN FOR SYRIATEL TELECOM

**BACKGROUND:** SyriaTel, a leading telecommunications company, is facing a significant challenge with customer churn. Churn is when customers decide to terminate their subscriptions with SyriaTel, resulting in revenue loss. To address this issue, SyriaTel aims to build a predictive model to identify customers who are likely to churn. By proactively targeting these at-risk customers with retention strategies, SyriaTel hopes to reduce churn rates and retain valuable customers.

**PROBLEM STATEMENT:** "Can we predict customer churn for SyriaTel and identify the key factors driving churn, enabling the company to implement effective retention strategies?

The business problem aligns with SyriaTel's goal of retaining customers and reducing revenue loss due to churn, emphasizing the importance of data-driven decisions in the telecommunications industry

# KEY OBJECTIVES

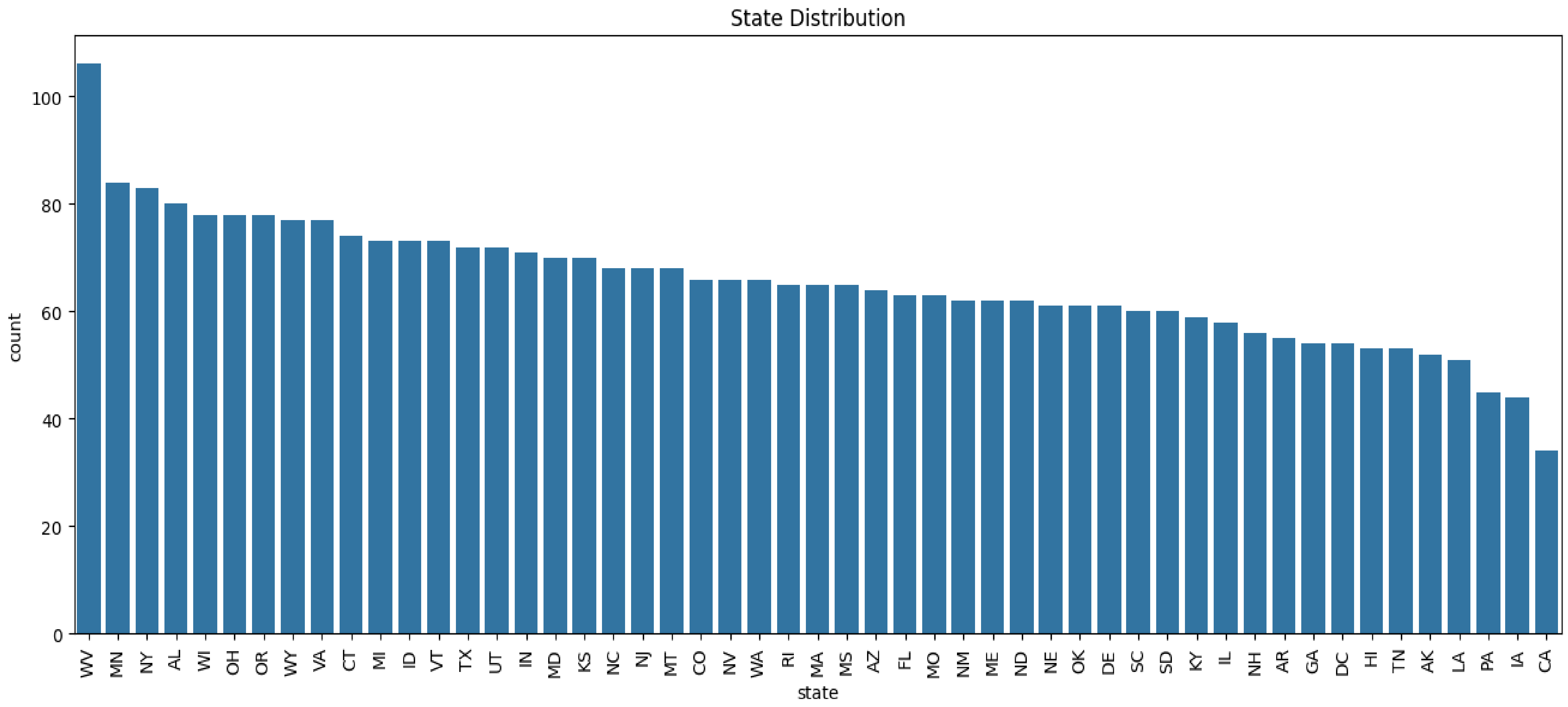
- To build a Churn Prediction model that can accurately predict customers who will churn based on the information available in the dataset.
- To identify the predictive patterns/features that are important for predicting customer churn.

The dataset includes customer information such as:

- Service plans
- Call and charge details
- Customer interactions with support

# EXPLORATORY DATA ANALYSIS

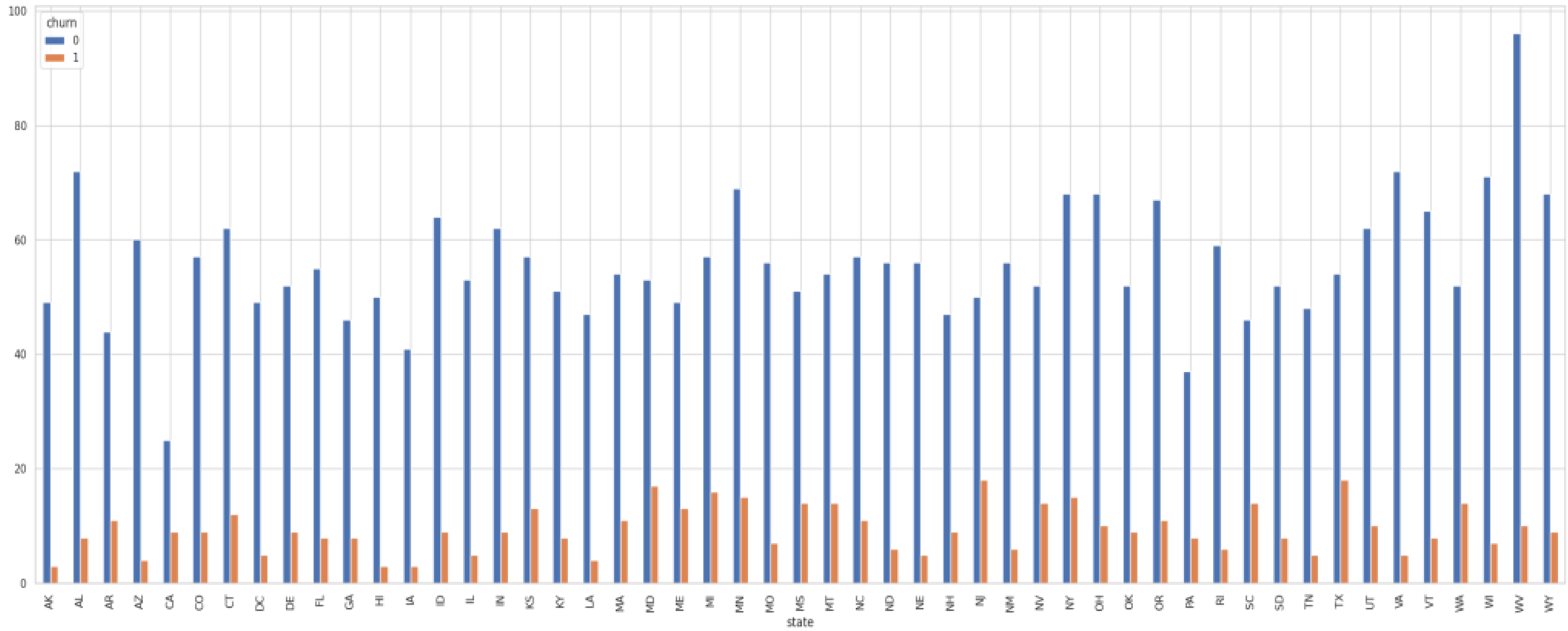
Here we explore some visualizations which help us in understanding our data more.


State Distribution
count
state
0
20
40
60
80
100
WV
MN
NY
AL
WI
OH
OR
WY
VA
CT
MI
ID
VT
TX
UT
IN
MD
KS
NC
NJ
MT
CO
NV
WA
RI
MA
MS
AZ
FL
MO
NM
ME
ND
NE
OK
DE
SC
SD
KY
IL
NH
AR
GA
DC
HI
TN
AK
LA
PA
IA
CA

# FROM THE PREVIOUS SLIDE WE CAN SEE THAT:

- The state with the highest count is WV (West Virginia) with 106 occurrences, indicating it is the most frequent state in the dataset.
- MN (Minnesota) follows closely with 84 occurrences, making it the second most common state.
- NY (New York) comes next with 83 occurrences, showing a similar frequency to MN.
- AL, WI, OR and OH all have 78 occurrences, placing them among the top states in terms of frequency.
- The state with the lowest count is CA with only 34 occurrences, suggesting it is the least frequent state in the dataset.

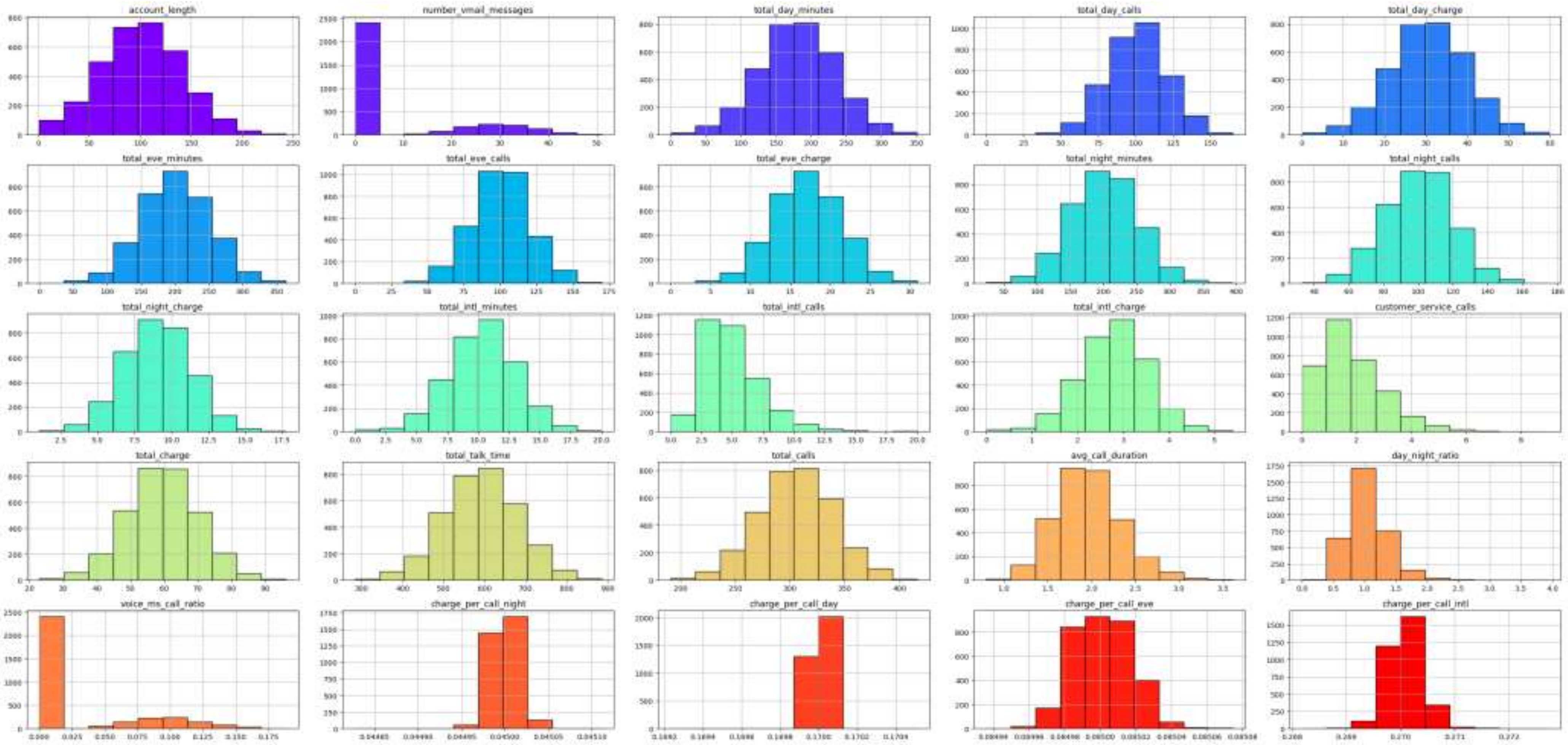
# CHURN BY STATE



# FROM THAT VISUAL WE CAN SEE THAT:

Some states have relatively higher churn rates like WV, VT, NY, OH with a significant number of churned customers (churn 1) while other states have lower churn rates like AR, AZ, CA, CO with a higher count of customers who did not churn (churn 0).

# DISTRIBUTION OF NUMERICAL FEATURES:



# MODELING

Classification models were used to predict churn. They were trained using a resampled dataset and optimized with hyperparameter tuning.

Key models tested:

- Logistic Regression
- Random Forest
- Gradient Boosting

# MODEL EVALUATION

The models were evaluated both before tuning and after tuning, using various metrics e.g. recall, precision and F1-scores.

The main metric for evaluating the classification model's performance was 'recall,' which measures how well the model identifies customers likely to churn. The goal was to minimize false negatives since missing a potential churner is more costly for the business than incorrectly predicting a non-churner. The target was to achieve at least 80% recall.

However, a balance was necessary. Predicting that all customers will churn would result in perfect recall but offer no real business value, as not all customers are at risk. Therefore, in addition to recall, 'precision' and 'accuracy' were used as secondary metrics to ensure a comprehensive assessment of the model's performance.

# OVERALL INSIGHTS AND BEST MODEL CHOICE

**Gradient Boosting (After Tuning) Performs Best**

Highest test accuracy (96.68%) with good recall (81.25%) and excellent precision (95.12%). Outperforms other models in balancing recall and precision. Slight risk of overfitting, but generalizes well to test data.

**Decision Tree (After Tuning) is a Good Alternative**

Good recall (86.46%) but lower precision (74.11%) compared to gradient boosting. If high recall is the priority (capturing all churn cases), this model is preferable.

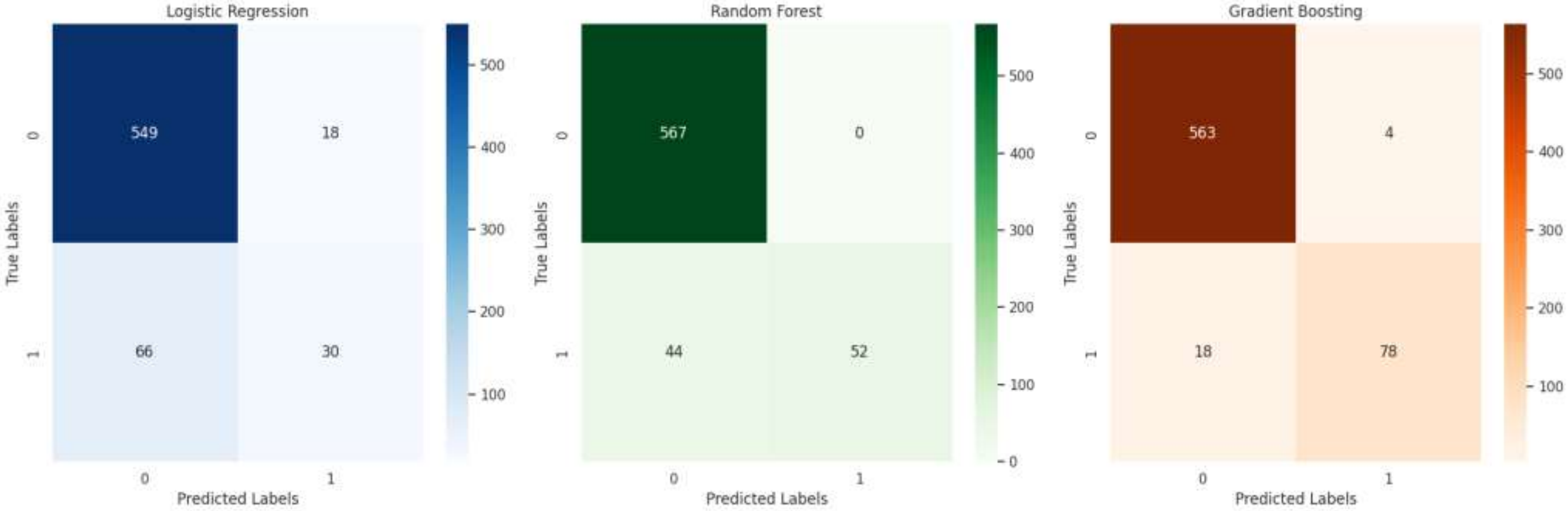
**Random Forest (After Tuning) is Too Conservative**

100% precision but only 54.17% recall—it captures very few churn cases. Would only be useful if the cost of false positives is extremely high. If computational efficiency or interpretability is a concern, this model is a great alternative.

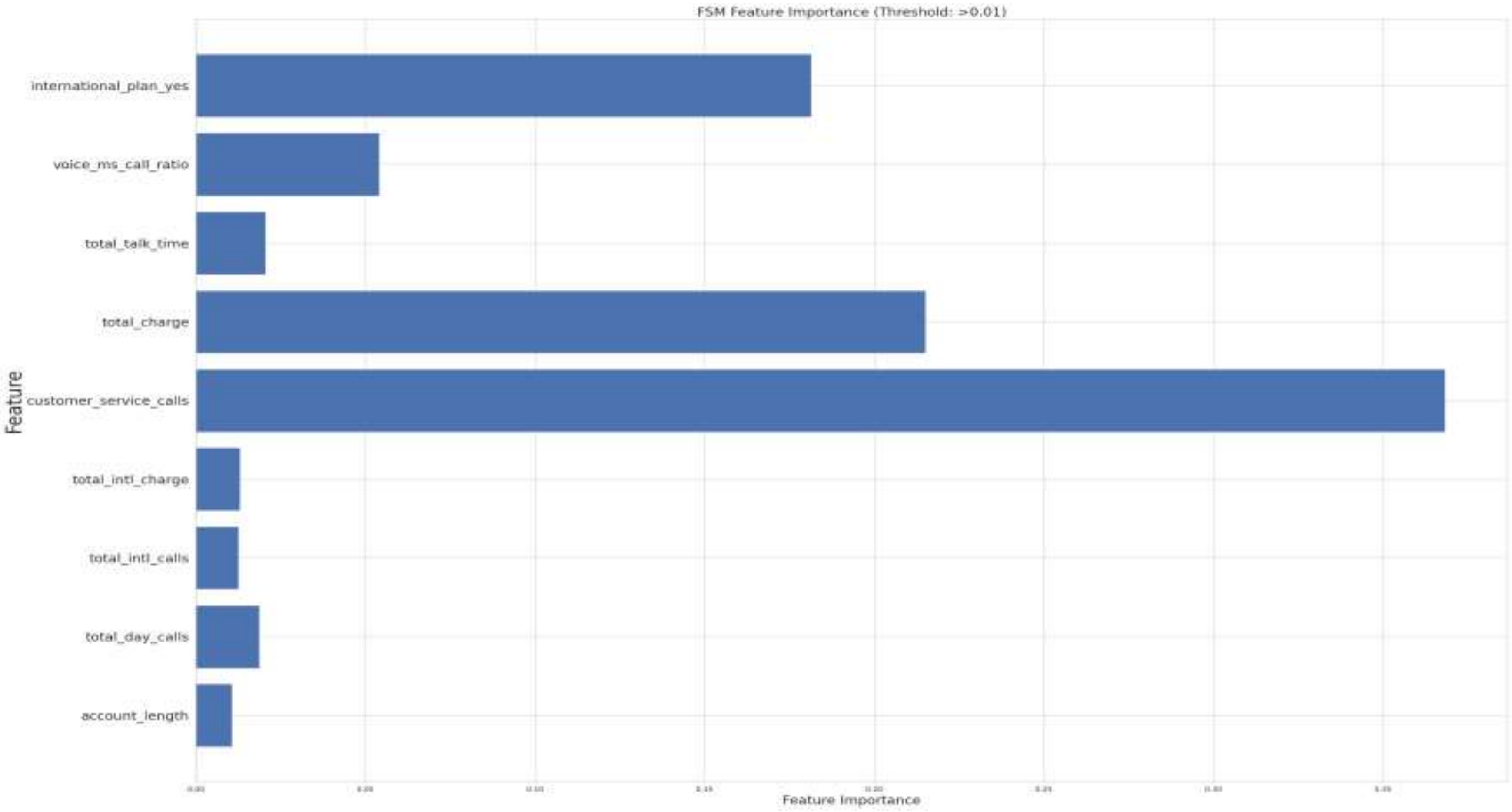
**Logistic Regression is the Weakest Model**

Even after tuning, recall drops significantly (31.25%), making it unreliable for churn prediction.

# CONFUSION MATRICES OF THE MODELS



# FEATURE IMPORTANCE SELECTION GRAPH



# RECOMMENDATIONS

1. Improve Customer Support:

Since customer service calls are the biggest churn factor, identify common complaints and resolve them proactively. Use AI-driven support or faster resolution times to improve customer satisfaction.

2. Target High-Charge Customers:

Customers with high total charges should receive loyalty rewards, exclusive discounts, or personalized offers to reduce price sensitivity.

3. Re-Evaluate International Plans:

Since international plan users are more likely to churn, conduct a survey to understand why. Offer flexible international calling packages or discounts.

4. Enhance Retention Efforts Based on Usage:

Customers with high talk time and high call volumes should be offered personalized plans to match their needs.

5. Monitor Voicemail Users:

Users with high voicemail usage may prefer messaging or digital alternatives. Consider promoting chat-based or unlimited text/calling plans.

# NEXT STEPS

To further improve our model and retention strategies, the following would help:

- Refining model features with additional customer data
- Implementing targeted retention campaigns
- Conducting A/B testing on new engagement strategies

# THANK YOU!

~THE END~