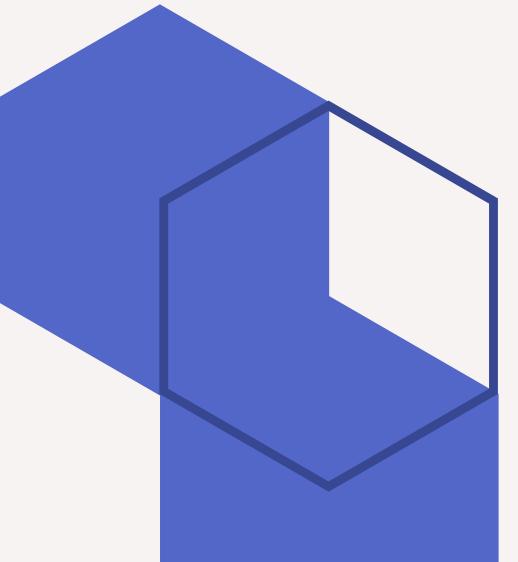
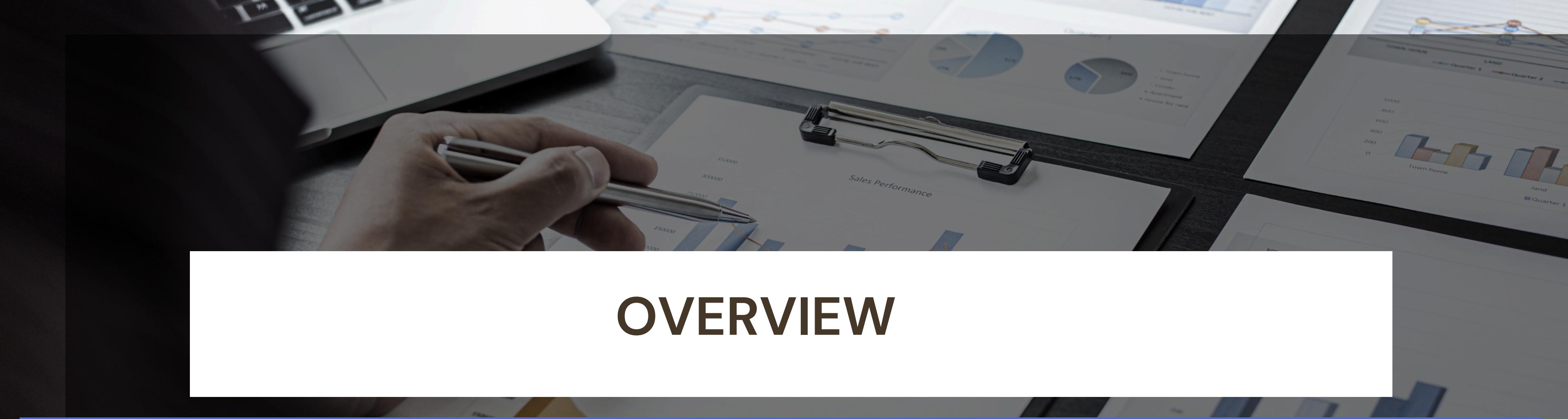


# Customer behaviors Analysis

Machine Learning Project





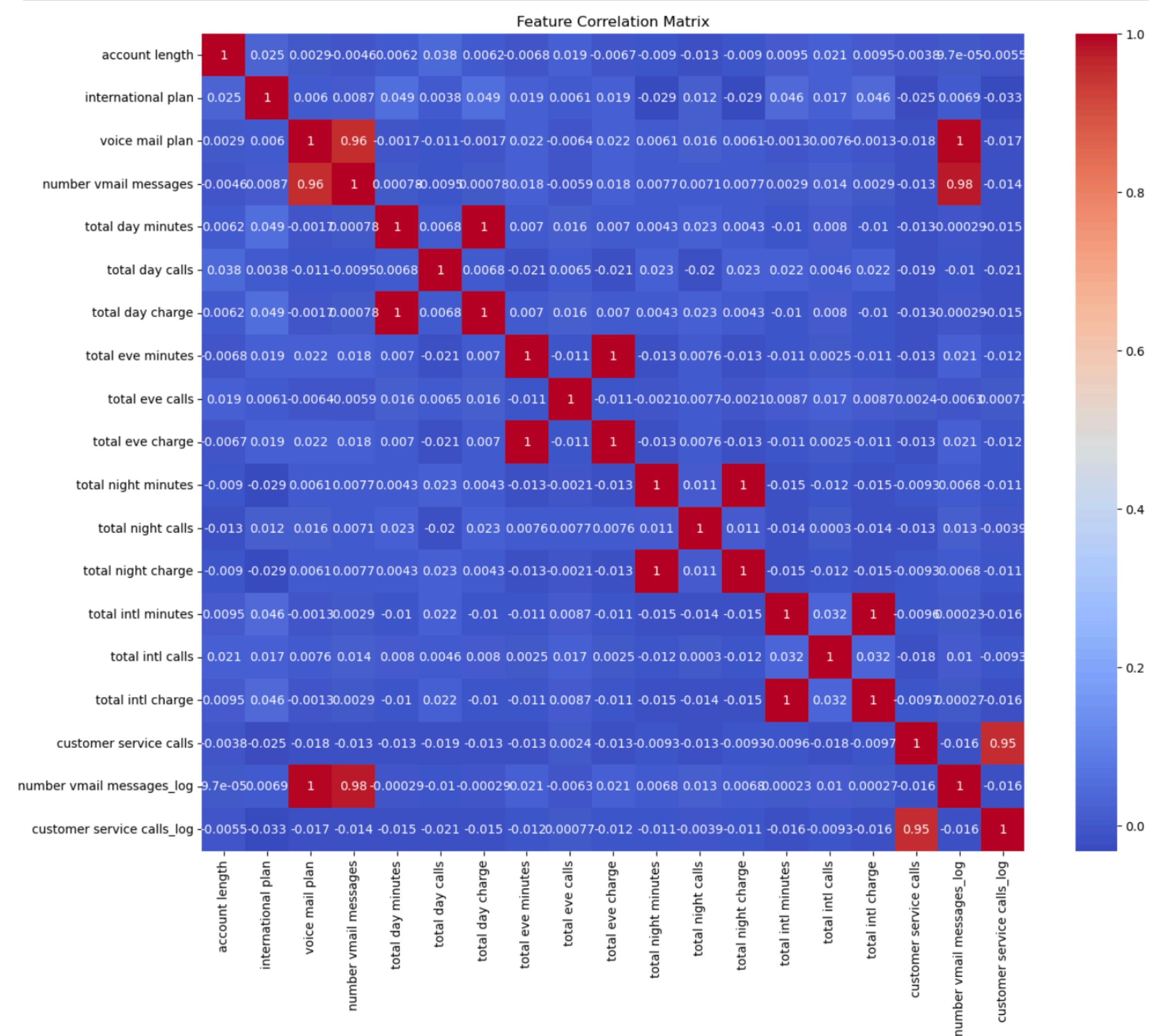
# OVERVIEW

This project focuses on predicting which customers are likely to stop using a telecom company's services (customer churn). The dataset consists of 3,333 customer records, containing details such as call usage, subscription plans, and customer service interactions. The main goal is to build a machine learning classification model that can proactively identify customers at risk of churn. Early identification enables the company to take targeted actions like offering personalized services or special deals to retain valuable customers. By analyzing usage behavior, plan details, and customer service interactions, we aim to discover key patterns and insights that contribute to customer churn, thereby improving customer satisfaction and reducing churn rates.

# OBJECTIVE

- Develop a binary classification model to predict customer churn.
  - Identify and rank the most significant features influencing churn.
  - Evaluate model performance using appropriate classification metrics.
  - Provide actionable insights into customer behaviors that lead to churn.
  - Recommend targeted retention strategies to reduce churn and increase customer lifetime value.
  - Enable the telecom company to proactively intervene with at-risk customers, improving overall customer satisfaction.

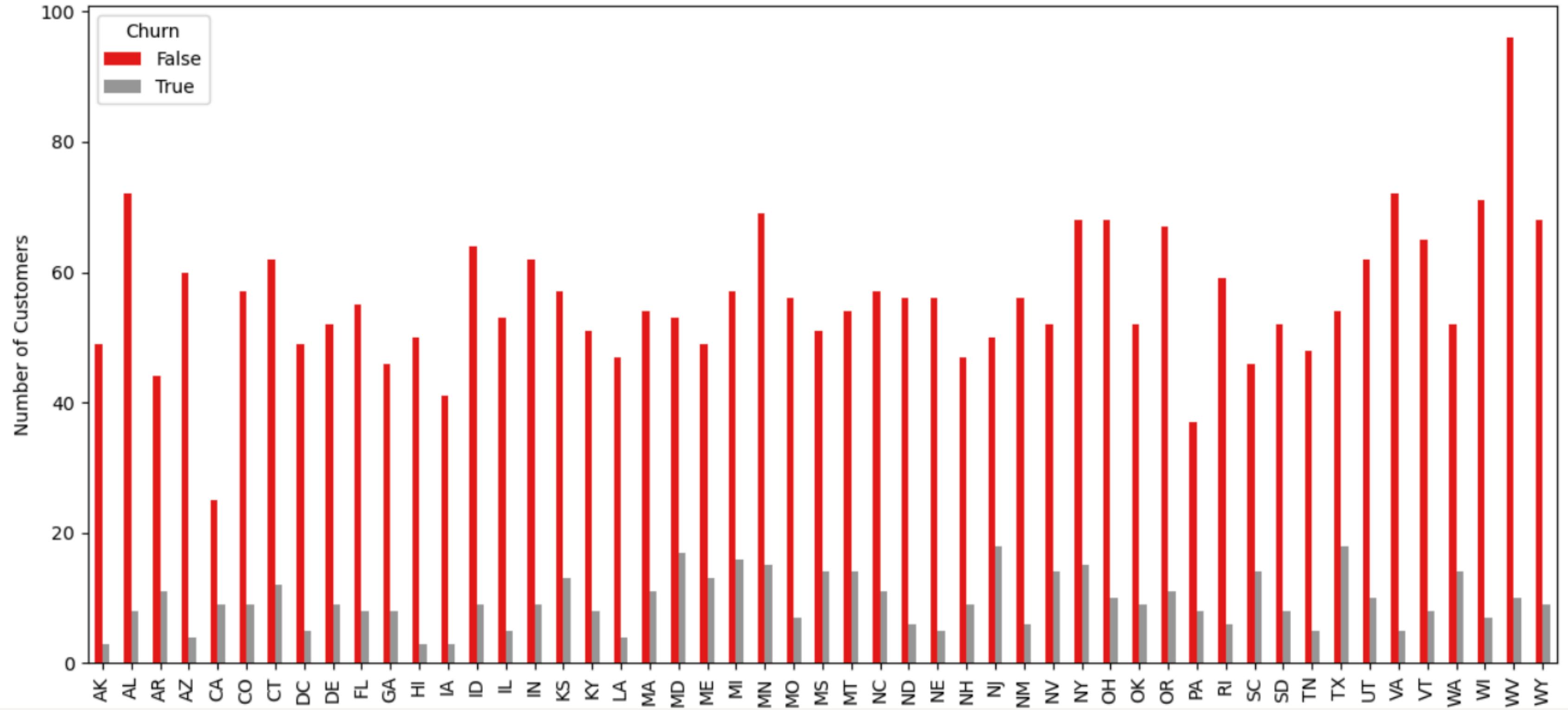




## Feature Correlation Matrix

- From correlation matrix some features have very strong correlations meaning they carry the same information.
- This heatmap shows the correlation between the numerical features in our dataset
- Most features have very low correlations with each other, suggesting minimal multicollinearity
- Low correlation is good because it means features are mostly independent, reducing redundancy.
- Highly correlated features could cause multicollinearity issues in some models (like linear regression).
- I will consider dropping one of the two highly correlated features or using regularization techniques.

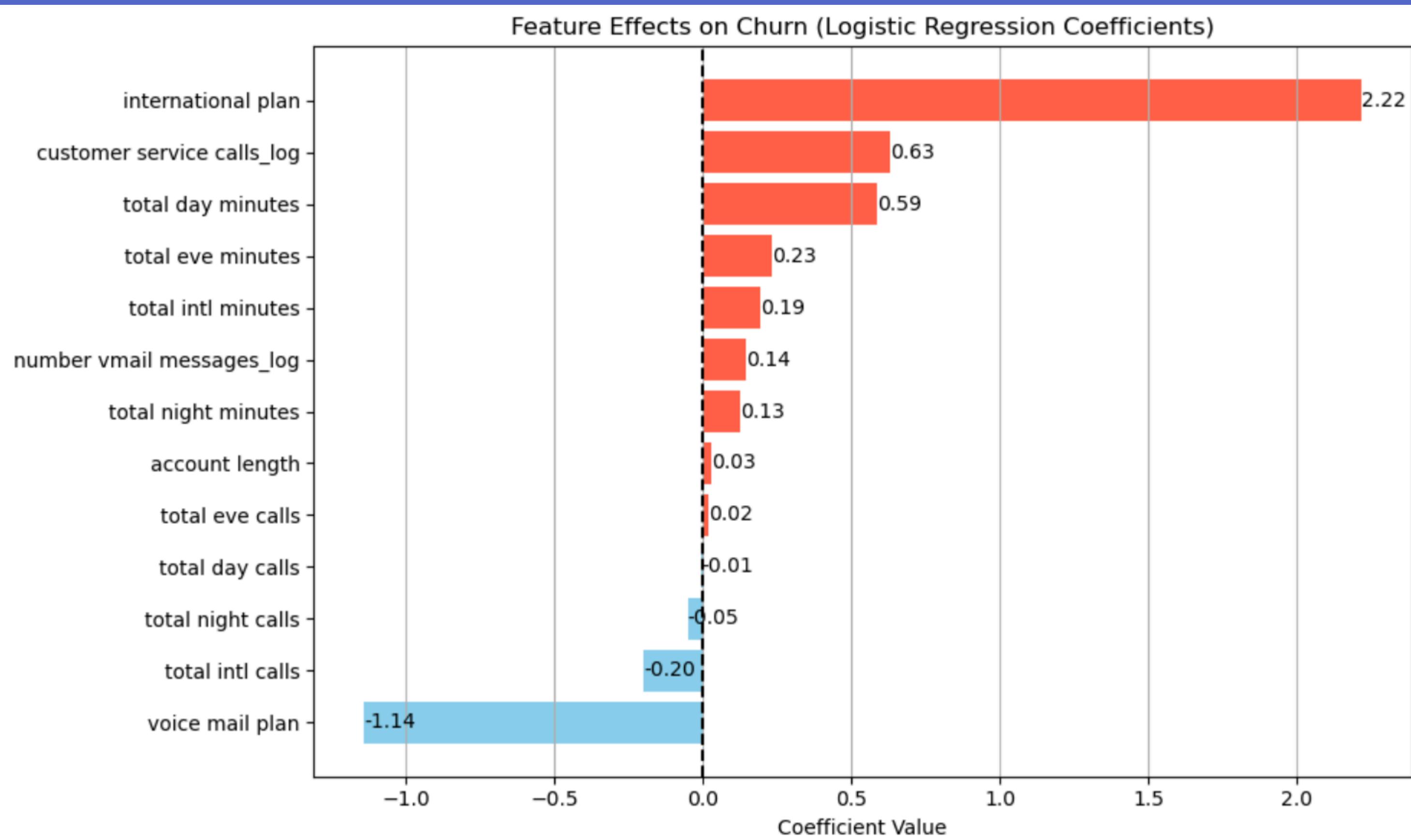
### Churn by State



- In every state, the number of customers who did not churn (False, in red) is much higher than those who did churn (True, in gray)
- This suggests that churn is relatively low across the board, which could be a good sign for customer retention.

- The top states like NJ, CA, TX, MS, and WA show churn rates of 20% and above.
- High and low churn states appear scattered there's no clear regional pattern (e.g., East vs. West), implying that churn is likely influenced more by other features.

## Logistic Regression Coefficients

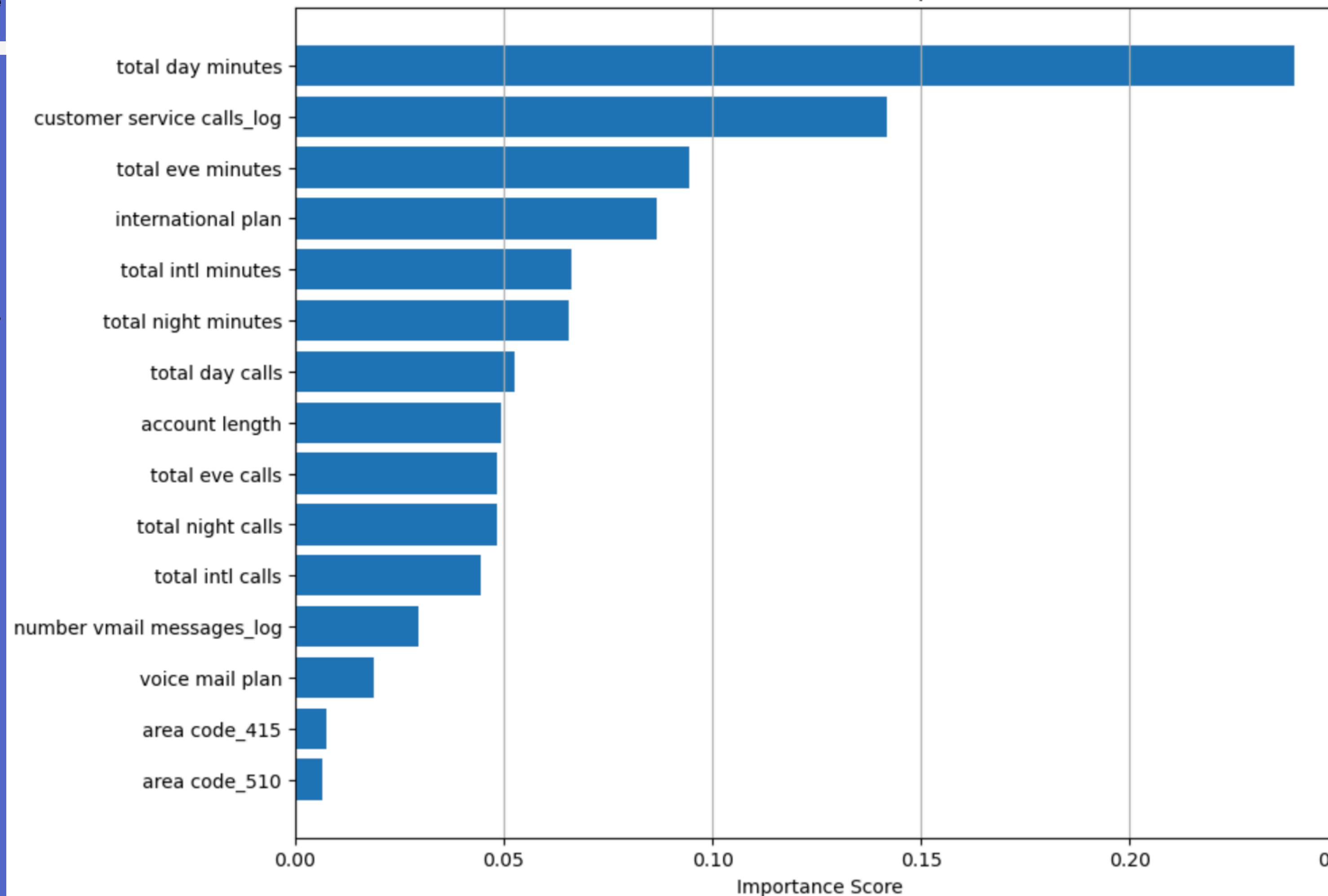


- Customers with international plans are more likely to churn maybe high cost or dissatisfaction with international services.
- Customers with voicemail plans are less likely to churn , voicemail users are more engaged or satisfied.
- Customers who contact customer service more frequently are more likely to churn possible dissatisfaction.
- High day-time phone usage correlates with higher churn heavy users might expect better service/discounts.
- Evening minutes also show a positive relation with churn ,active users might churn if dissatisfied.
- Longer accounts slightly increase churn maybe loyalty fatigue.

## Random Forest Feature Importance

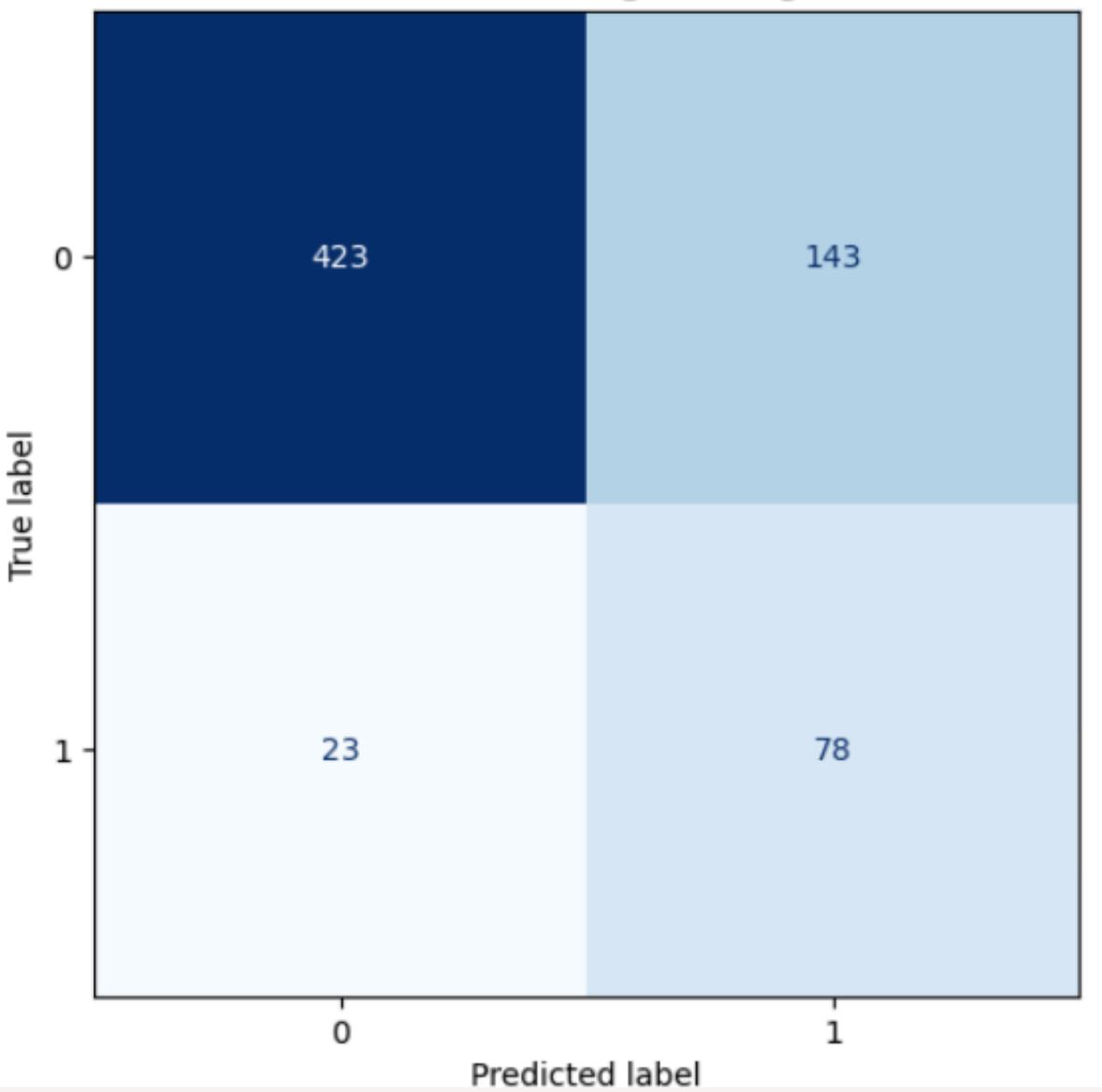
- total day minutes is very important with 24% contribution meaning strong predictor.
- customer service calls\_log is second churners called customer service more.
- international plan is also significant its a binary feature churners tend to have it more.
- number vmail messages\_log and voice mail plan have lower importance, but they are not zero.
- area code\_415 and area code\_510 have very low importance they are almost noise.

## Random Forest Feature Importance

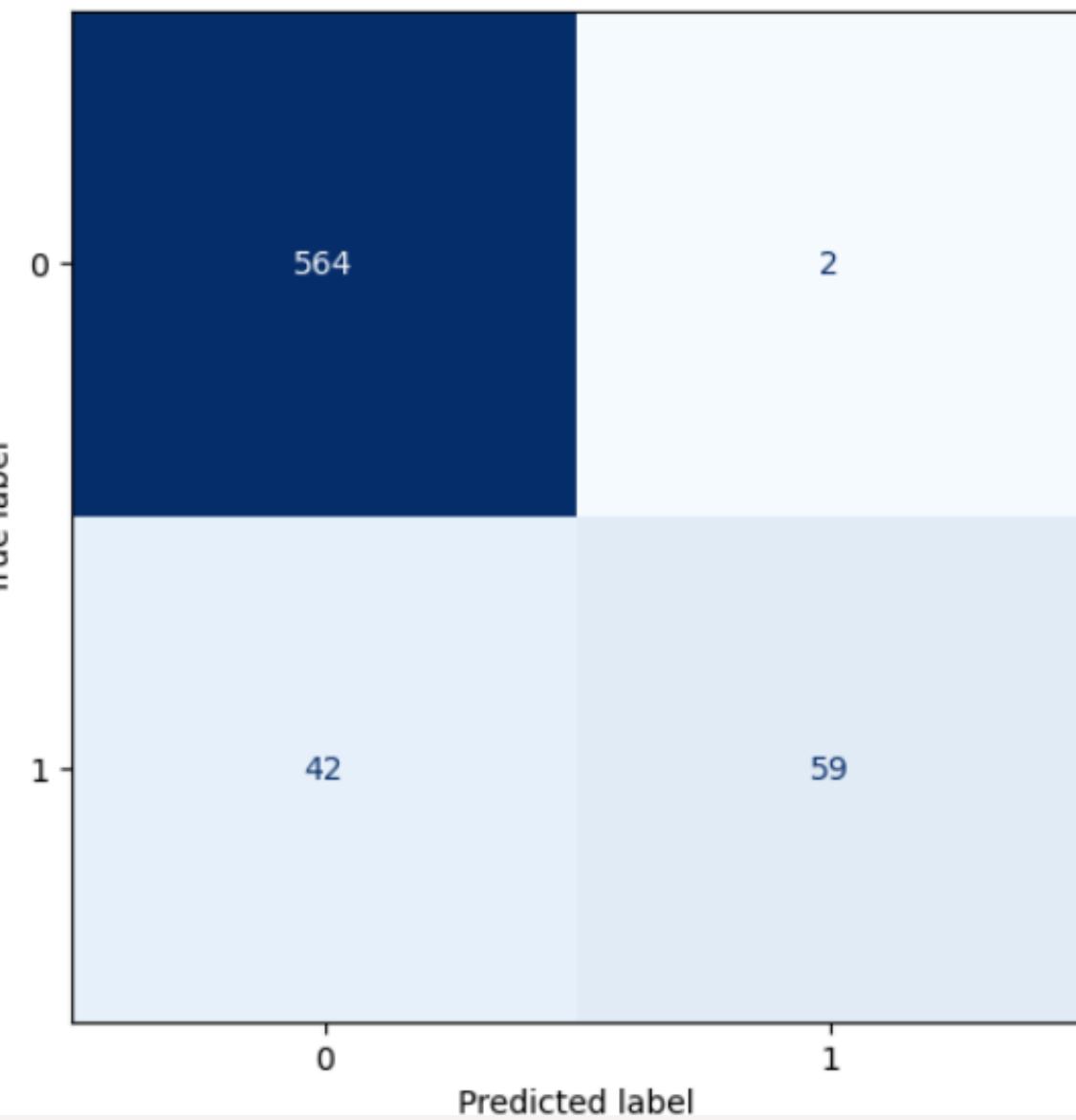


## Confusion Metrics

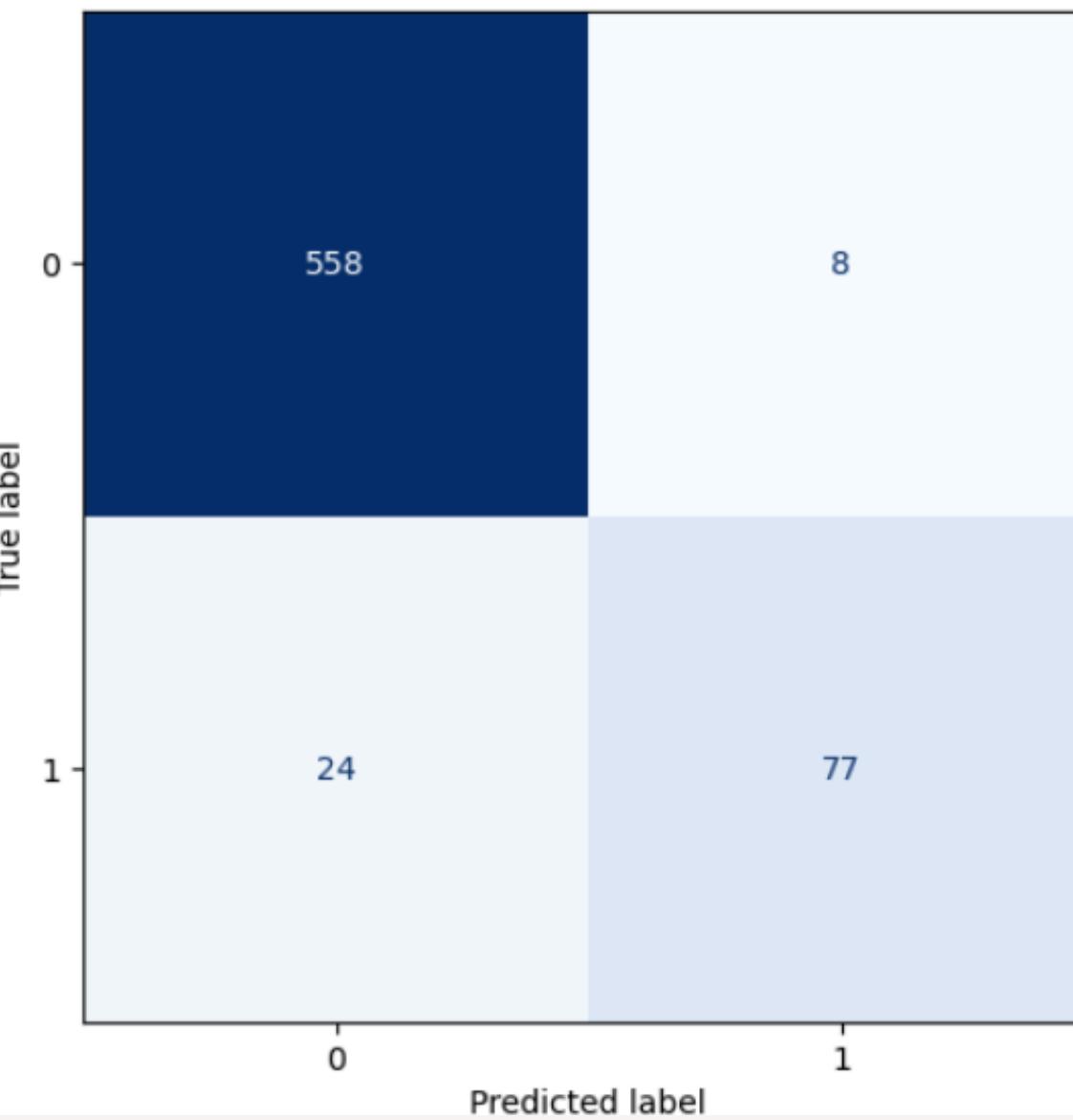
Confusion Matrix: Logistic Regression



Confusion Matrix: Random Forest



Confusion Matrix: XGBoost



### Logistic Regression

It has high Recall (77%) meaning it captures many churners it also has very low Precision (35%) ,many false positives and its inefficient

### Random Forest

It shows excellent Precision (97%) with almost no false positives it also shows Low Recall (58%) probaly misses many churners

### XGBoost

It shows a balance between precision and Recall

# Threshold Tuning

Best Threshold

0.4430

Best F1-Score

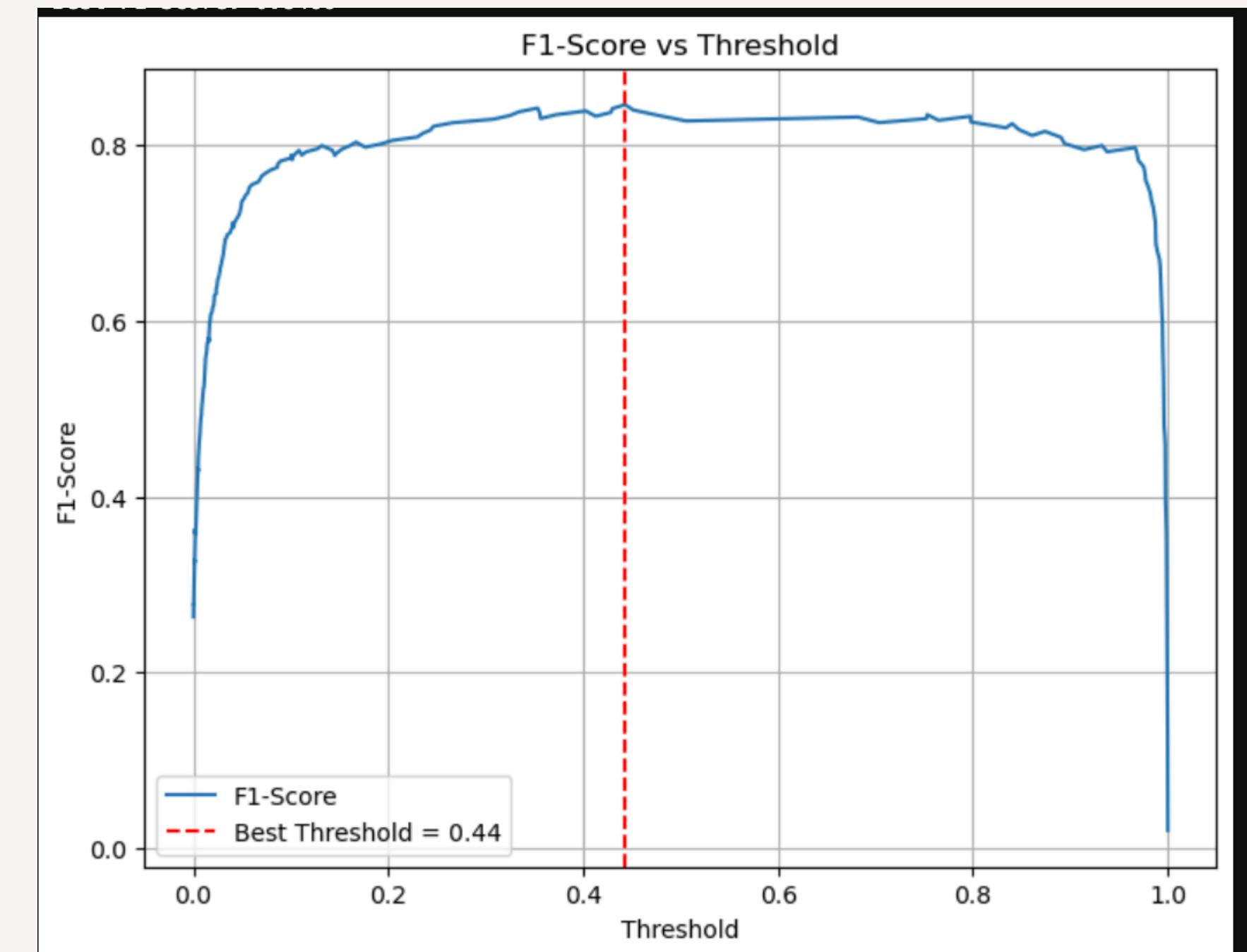
0.8466

NOTE

Precision improves.

Recall decreases.

- False Negatives reduce.



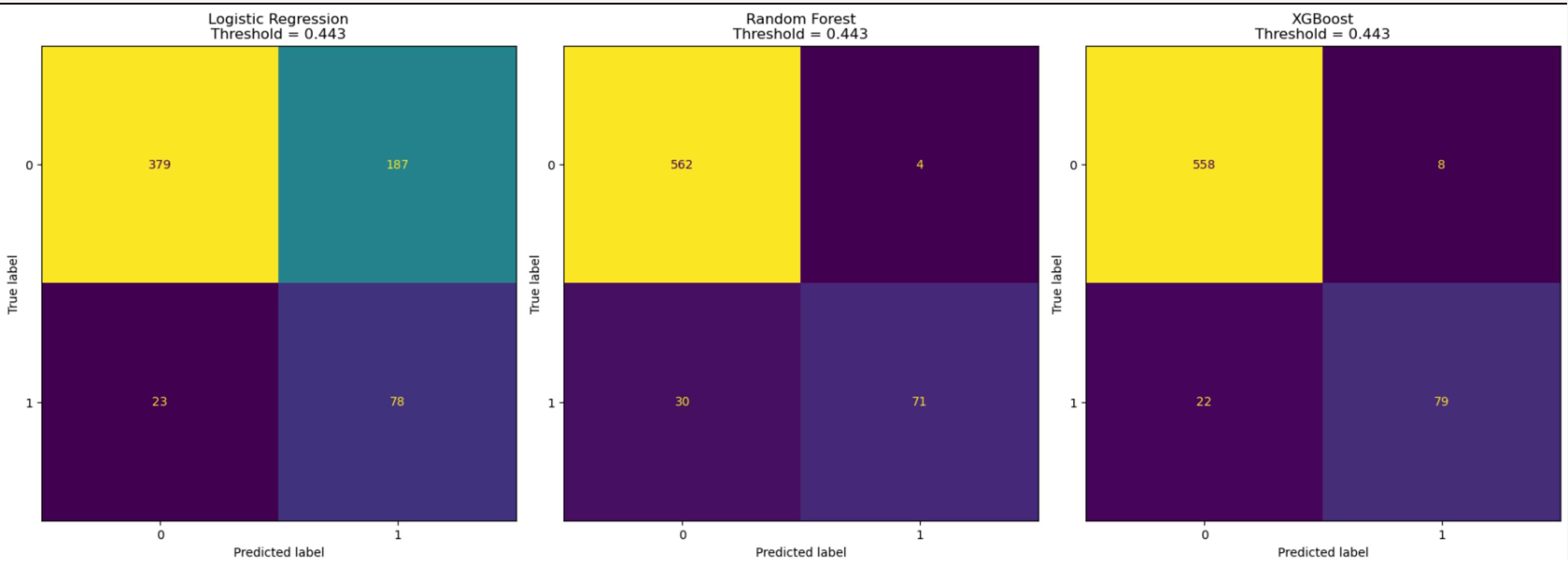
# Adjusting the threshold for the models

BEST MODEL

XGBoost

XGBoost Details

Accuracy = 0.9550  
Precision = 0.9080  
Recall = 0.7822  
F1-Score = 0.8404  
AUC = 0.8116



## OBSERVATIONS AND RECOMMENDATIONS

Observations	Recommendations
<b>XGBoost Model Performance:</b> XGBoost outperformed others; threshold tuning improved recall without big precision loss.	Deploy the XGBoost model for Churn Prediction: Deploy the model in production, use adjusted threshold (0.443), integrate with CRM for proactive churn management, and retrain periodically.
<b>International Plan:</b> Customers with an international plan are significantly more likely to churn.	Investigate and Optimize International Plans: Review pricing and service quality for international plans to improve customer satisfaction and reduce churn.
<b>Voice Mail Plan:</b> Having a voice mail plan reduces churn risk.	Promote Voice Mail Plan Adoption: Encourage customers to adopt voicemail plans through promotions or bundled services.
<b>Customer Service Calls:</b> Frequent customer service calls are associated with a higher churn risk.	Enhance Customer Service: Improve service quality and responsiveness; prioritize customers with frequent service interactions for retention.
<b>Total Day Minutes:</b> Heavy daytime usage correlates with increased churn risk.	Loyalty Programs for Heavy Users: Offer loyalty rewards, discounts, or premium service tiers targeting high daytime and evening users.
<b>Total Evening Minutes:</b> Higher evening call minutes are associated with a moderate increase in churn risk.	Loyalty Programs for Heavy Users: Same — address high service expectations to prevent churn.
<b>International Calls:</b> Customers making more international calls are less likely to churn.	Monitor International Usage Patterns: Focus more on customers with long call durations to manage dissatisfaction.
<b>Total International Minutes:</b> Higher total international call duration slightly increases churn risk.	Monitor International Usage Patterns: Identify high international usage customers and offer special packages or discounts to reduce churn risk.