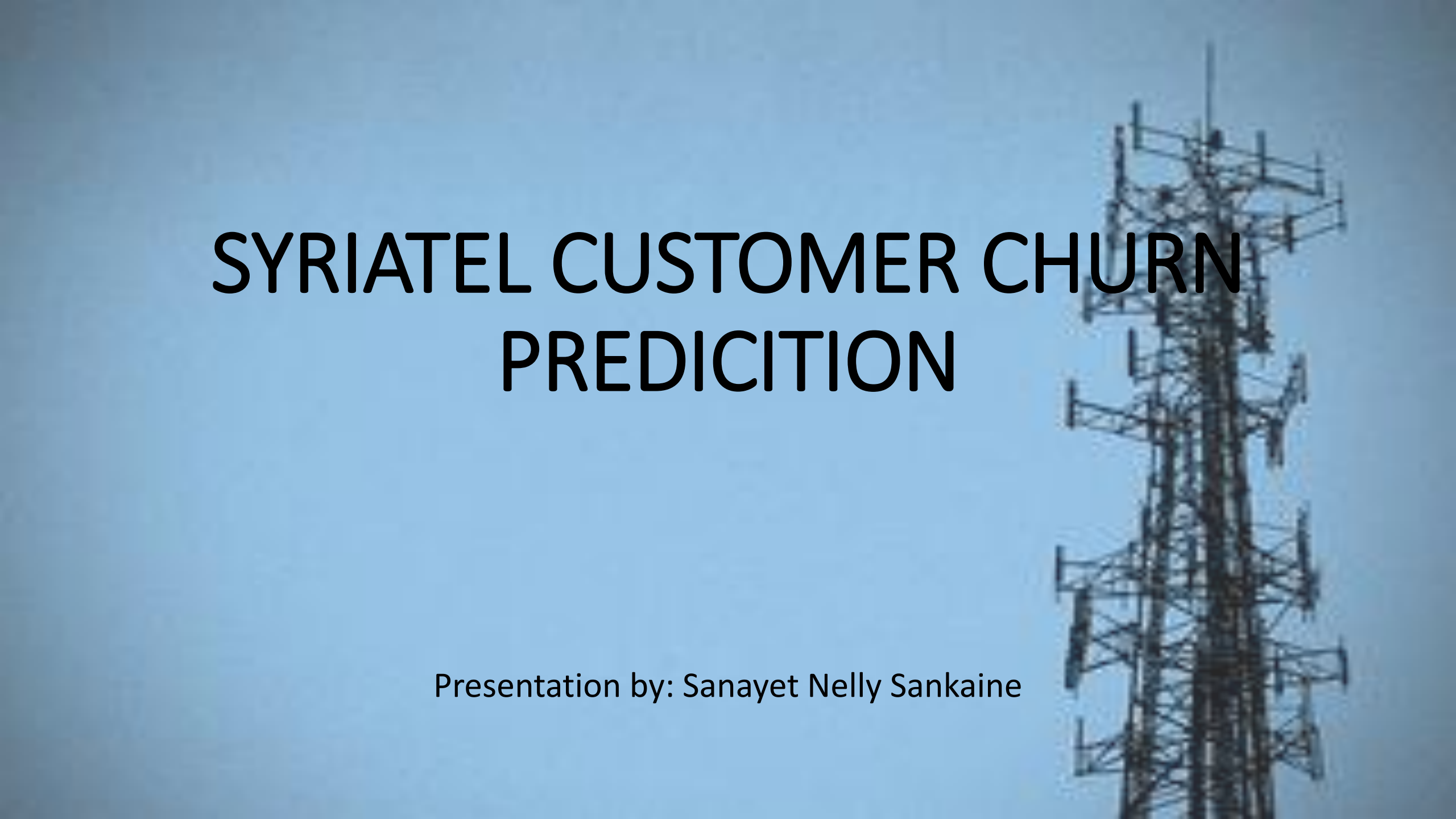
# SYRIATEL CUSTOMER CHURN PREDICITION



Presentation by: Sanayet Nelly Sankaine

# PROJECT OVERVIEW

- Customer churn is a critical challenge for businesses, especially in the telecommunications industry. This project focuses on predicting customer churn for SyriaTel, a telecom company, using machine learning techniques. The goal is to build a model that can accurately classify whether a customer is likely to churn, allowing the company to take proactive measures to retain them
- By identifying the most influential factors contributing to customer churn, this project provides valuable insights that can help SyriaTel develop targeted retention strategies.

# PROJECT WORKFLOW

Business Understanding

Data Understanding

Modeling

Evaluation

Conclusion

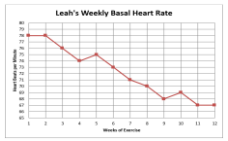
Recommendation

# BUSINESS UNDERSTANDING

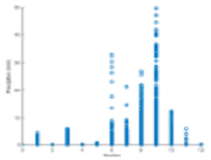
## PROBLEM STATEMENT

The objective of this project is to predict customer churn for SyriaTel, a telecommunications company. The goal of this analysis is to be able to build a model that accurately predicts the customers that are likely to churn so that SyriaTel can take proactive measures to retain them leading to long term revenue generation and company growth.

# SPECIFIC OBJECTIVES

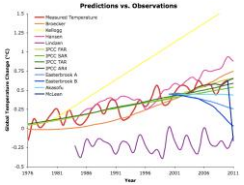



Uncover the trends and patterns that lead to customer churn



Build a model that accurately predicts customers that are likely to churn


Develop practical insights that SyriaTel can use to reduce churn and retain valuable customer

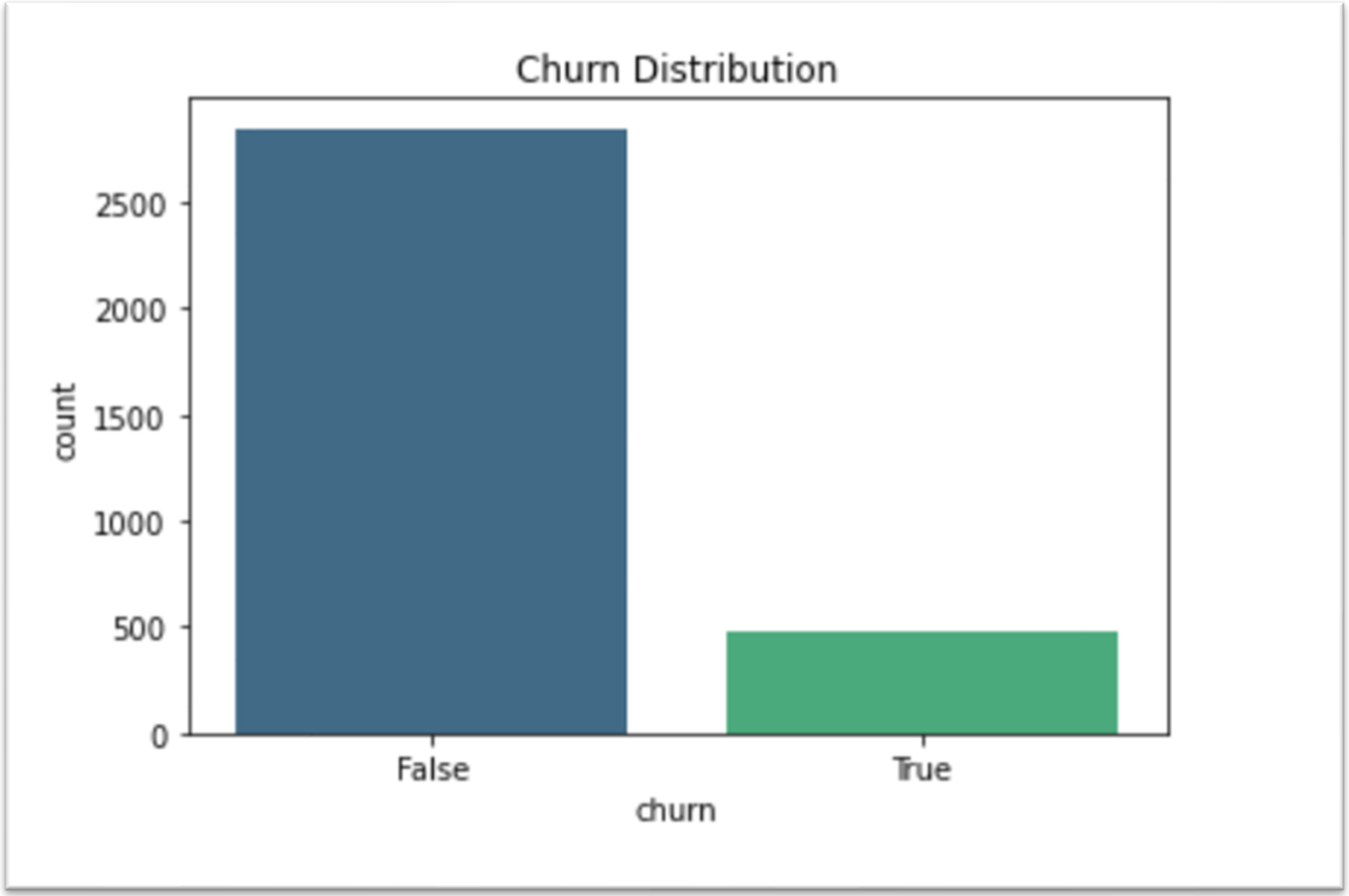



Determine features with most impact on predicting customer churn to inform targeted retention strategies

# DATA UNDERSTANDING

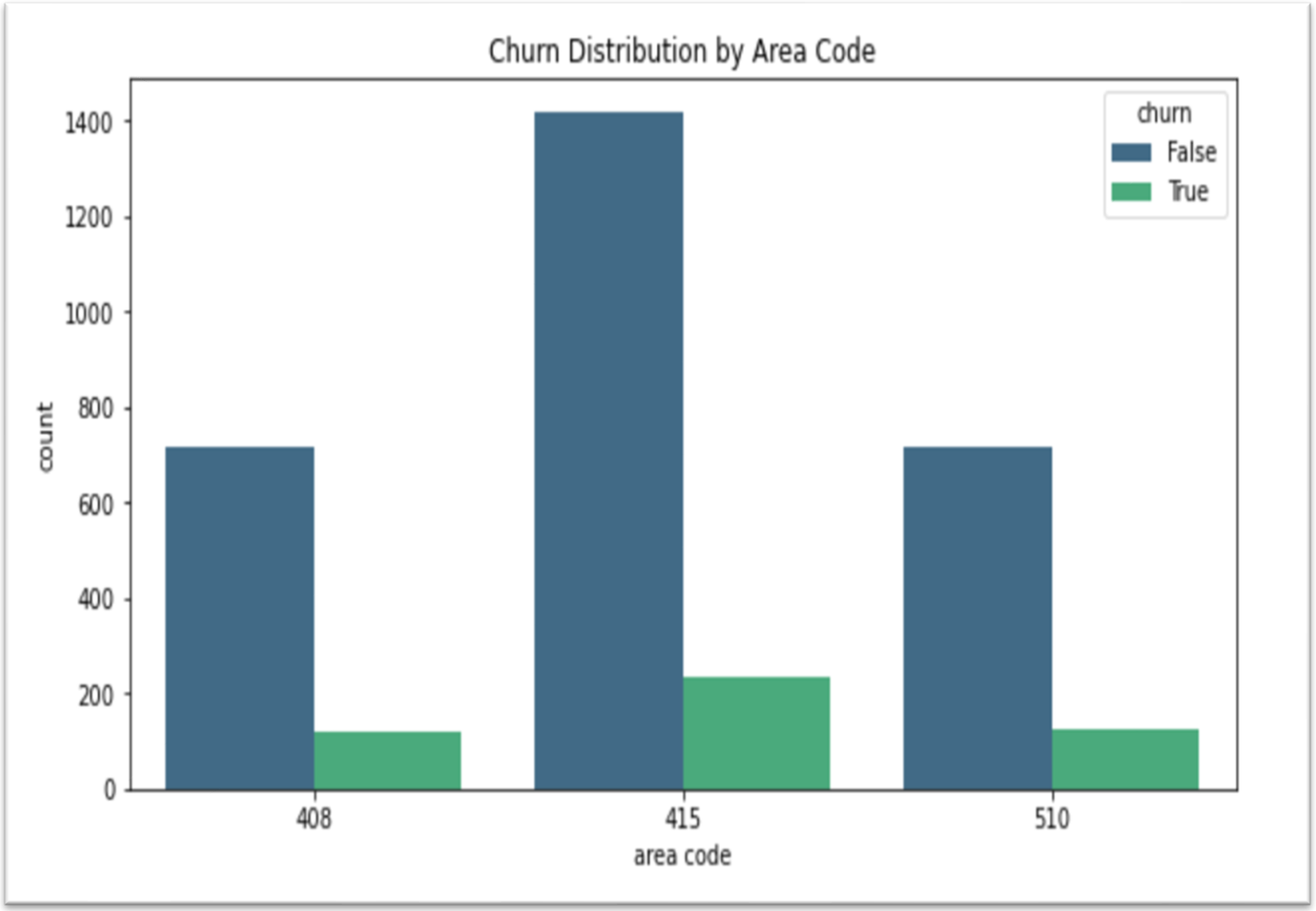
- The dataset we are working with has been obtained from Kaggle website.
- It contains 3333 records and 21 columns
- It contain 4 categorical features and 17 numerical features

# FINDINGS



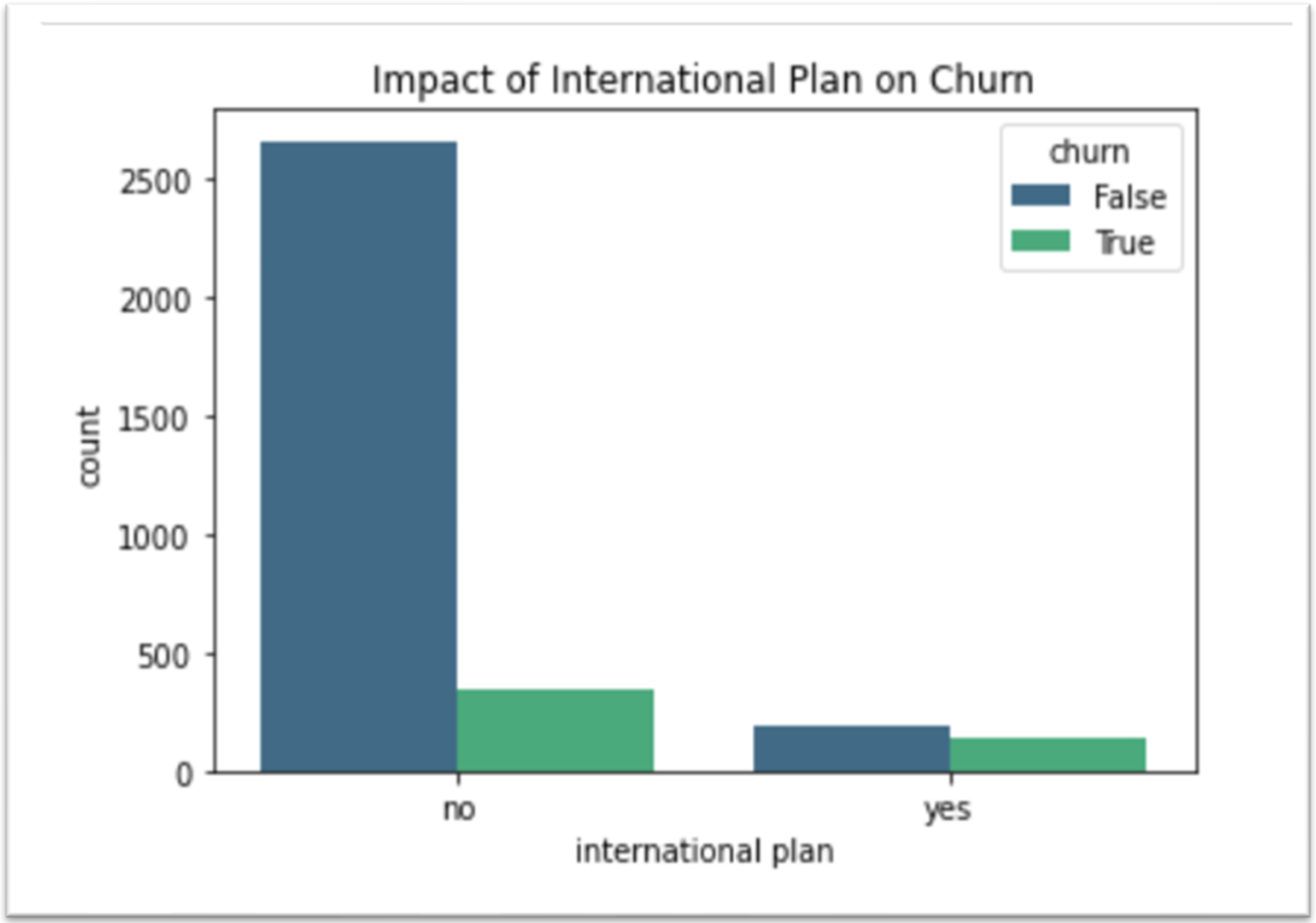

Churn Distribution?

- ❑ No of Customers that Churned : 483
- ❑ Churn Percentage : 14.5%
- ❑ Non-Churn Percentage: 85.5%



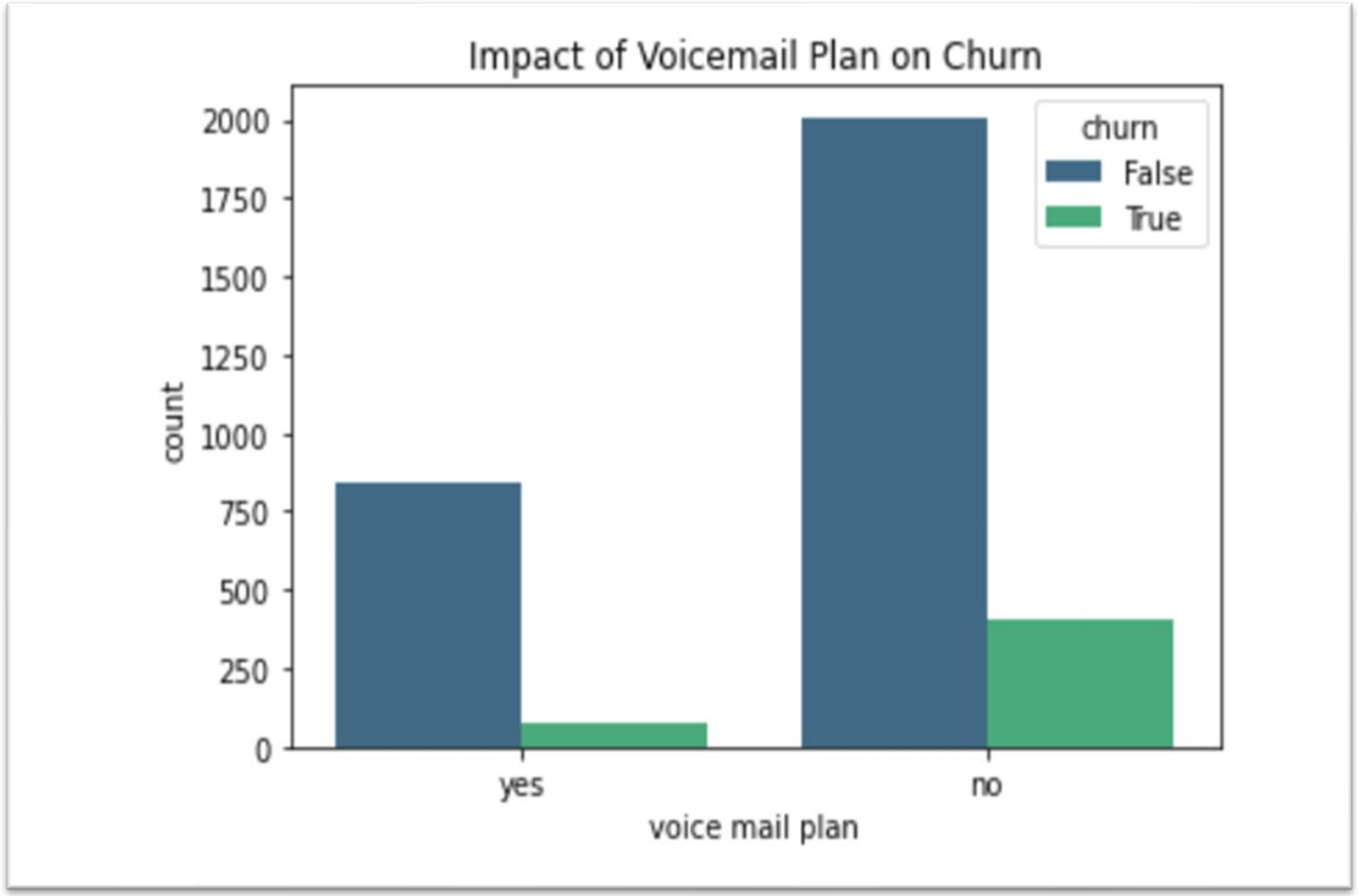

Churn Distribution by Area Code:

- Area Code 415 had the highest number of customers who churned



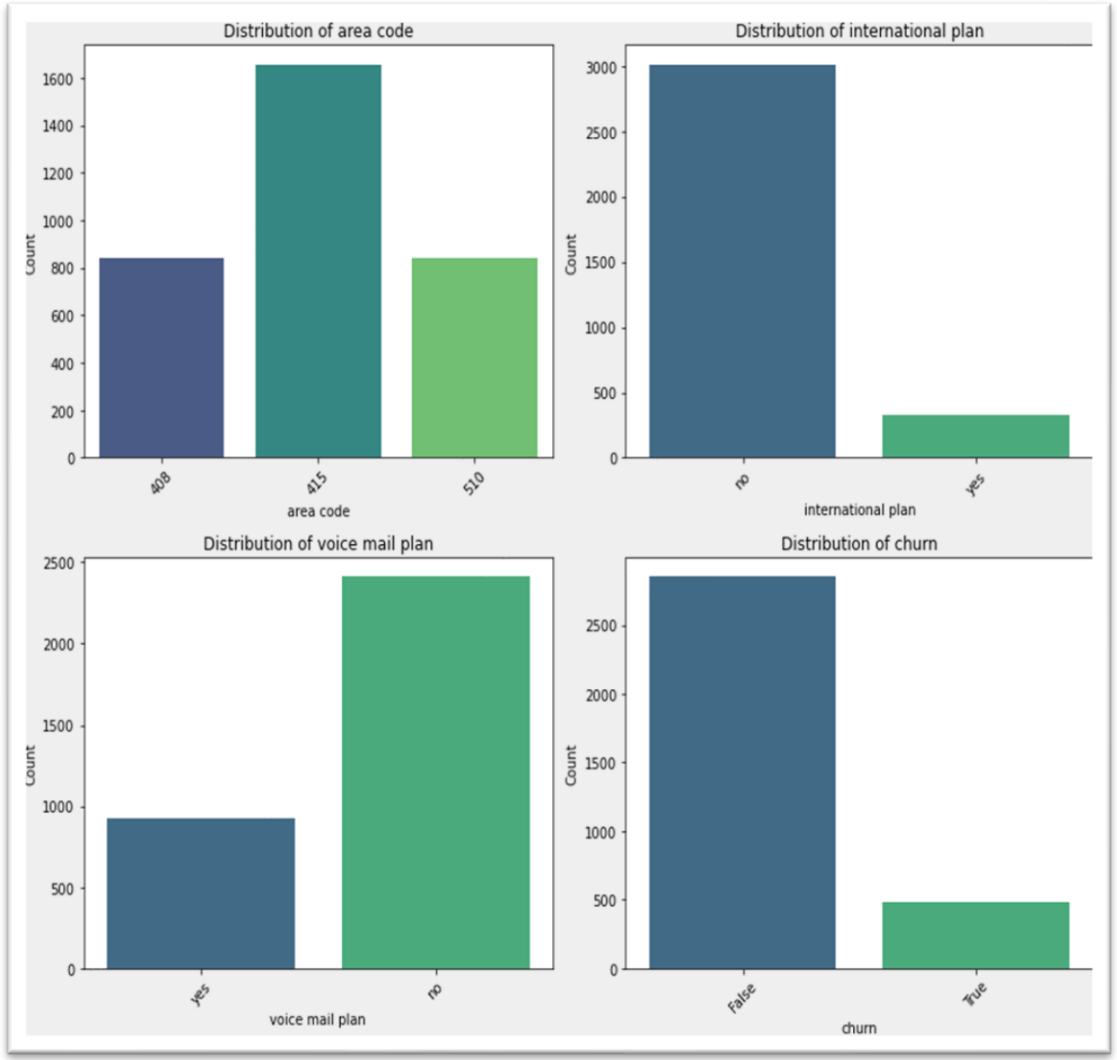

Churn Distribution by International Plan:

- ❑ The majority of the customers who churned had not subscribed to an international plan



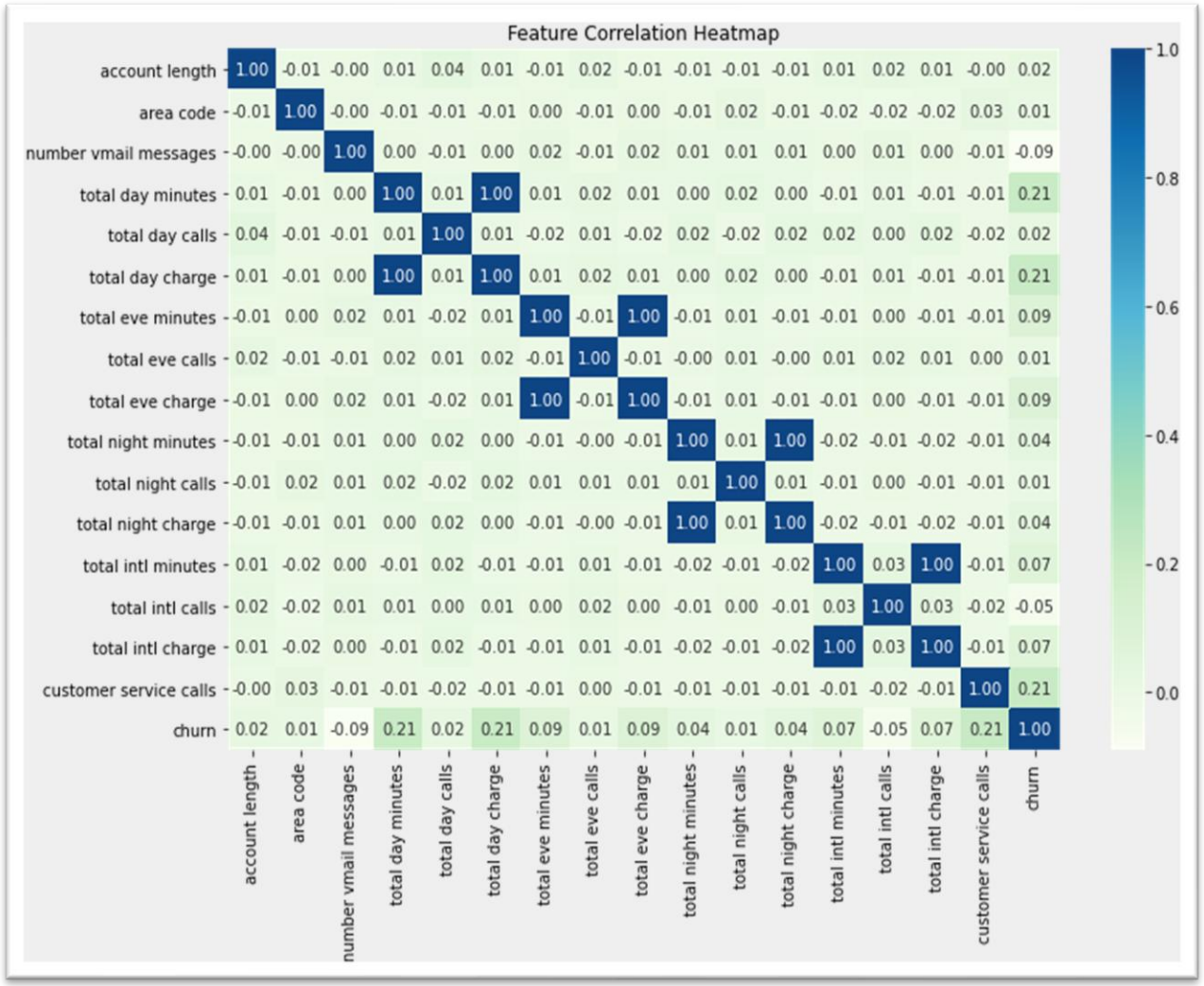

Churn Distribution by Voicemail Plan:

- ❑ The majority of the customers who churned had not subscribed to a voicemail plan




Distribution of Categorical Features:

- ❑ Out of 3333, only 323 of customers have an international plan which is about 0.1%
- ❑ About 38% of SyriaTel customers are subscribed to a voicemail plan




Features Correlation:

- ❑ Total int charge and total int minutes, highly positively correlated
- ❑ Total eve charge and total eve minutes, highly positively correlated
- ❑ Total day charge and total day minutes, highly positively correlated
- ❑ Total night charge and total night minutes, highly positively correlated.

# MODELING

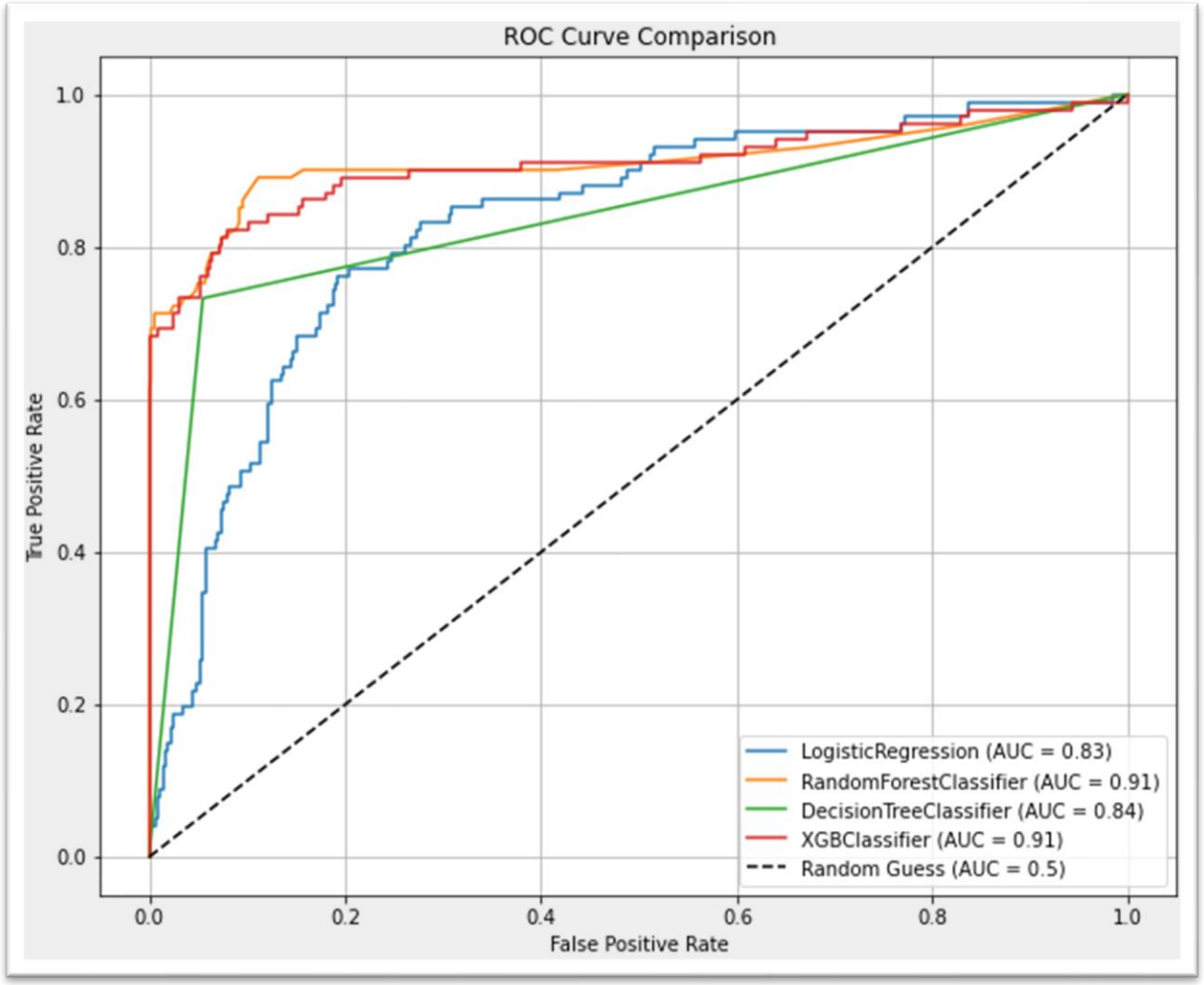
Algorithms Used:

- ❑ Logistic Regression
- ❑ Decision Tree
- ❑ Random Forest
- ❑ XG Boost

Evaluation Metrics Used:

- ❑ Recall focus to identify correctly predicted positive classes
- ❑ ROC_AUC Curve plots

# MODEL RESULTS



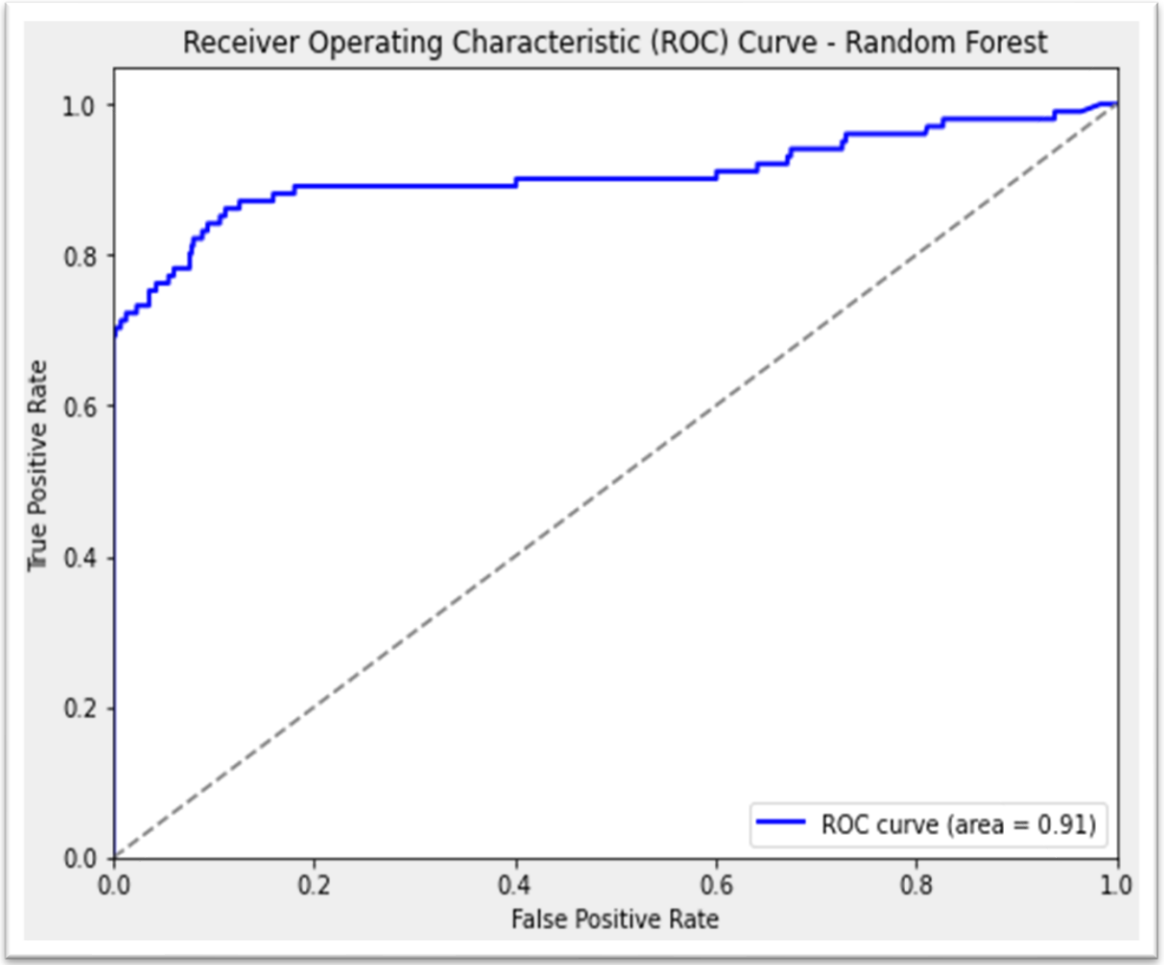

Recall Scores:

- ❑ LogisticRegression 0.772277
- ❑ RandomForestClassifier 0.693069
- ❑ DecisionTreeClassifier 0.732673
- ❑ XGBClassifier 0.693069

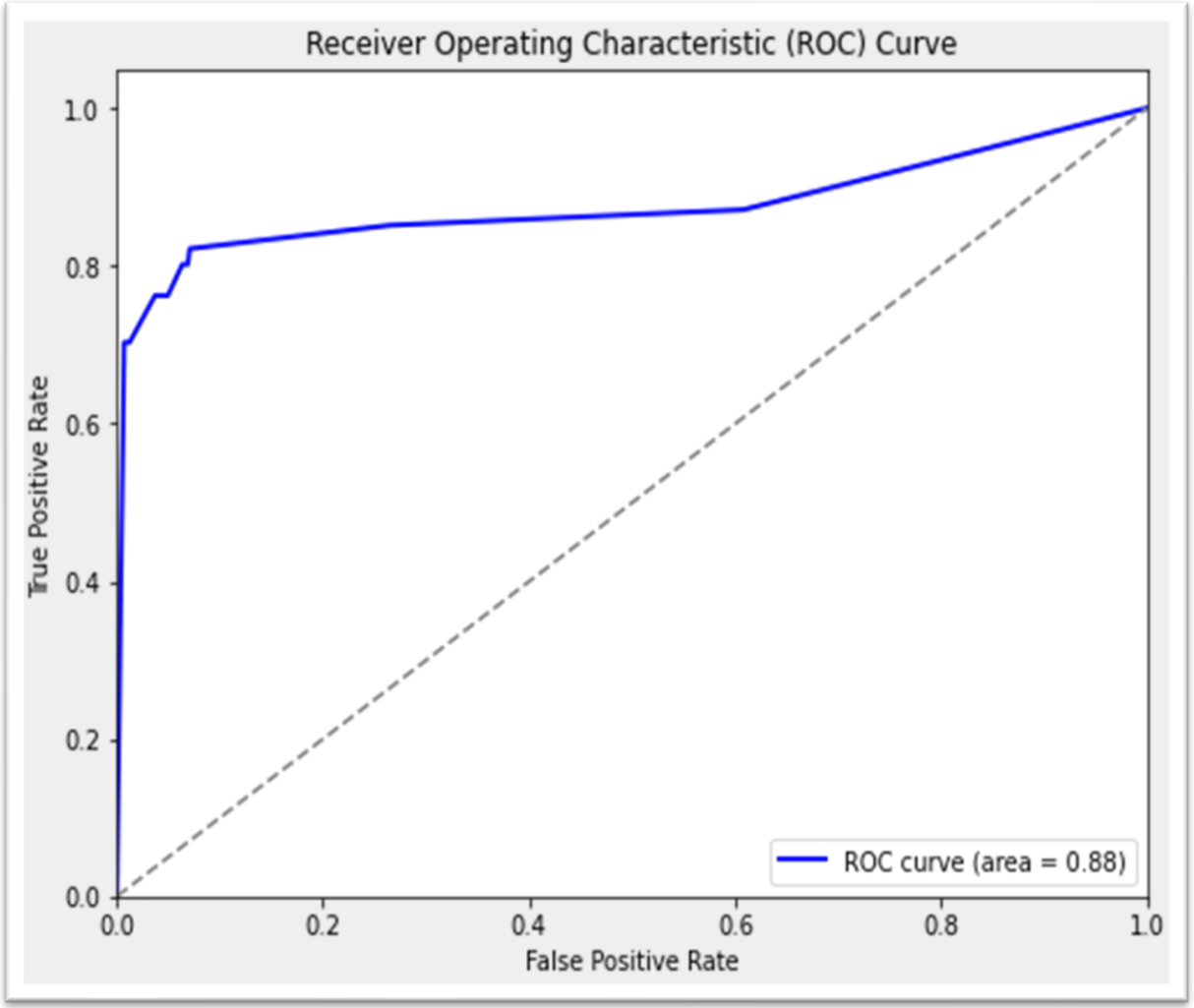
ROC Scores:

- ❑ LogisticRegression 0.83
- ❑ RandomForestClassifier 0.91
- ❑ DecisionTreeClassifier 0.84
- ❑ XGBClassifier 0.91

# MODEL TUNING RESULTS

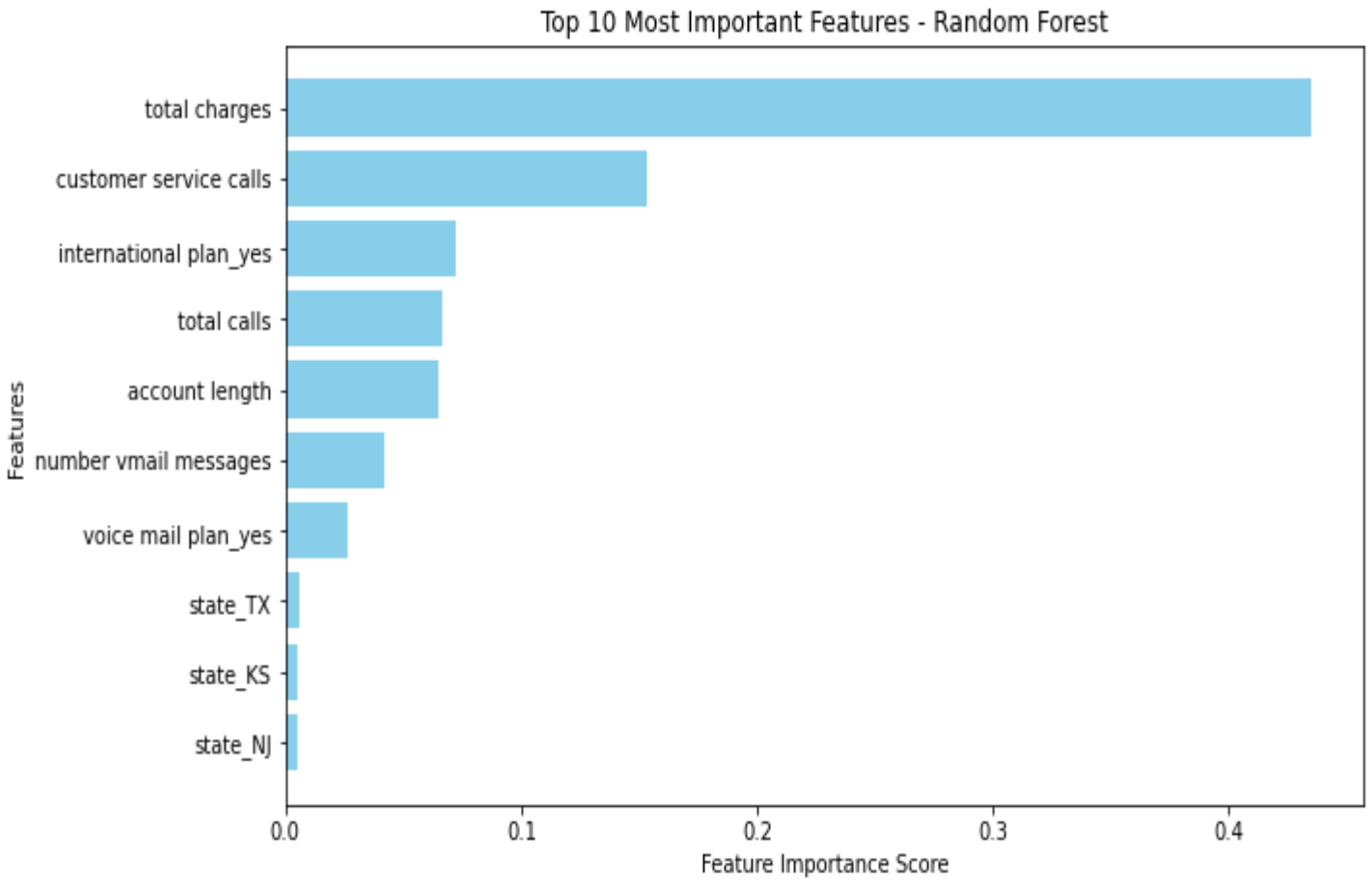



- Recall score for the tuned Random Forest model improves to 0.71 with an ROC score of 0.91, an improvement from the untuned model
- R.F model is better at correctly identifying positive instances (true positives) while maintaining a good balance between correctly classifying both positive and negative instances.
- The ROC AUC score indicates the model's strong performance in distinguishing between the two classes across different decision threshold




- The recall score for the tuned Decision Tree model slightly drops to 0.69 with an ROC score of 0.88. which is an improvement from the untuned model.

- Model misses a few more positive instances compared to the untuned model. However, the improvement in the ROC AUC score (0.88) shows that the model has improved its ability to distinguish between positive and negative instances overall, indicating a better overall classification performance




Significantly Important Features:

- ❑ Total Charges
- ❑ Customer Service Calls
- ❑ International Plan

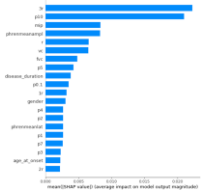
# CONCLUSIONS

- Random Forest is the superior model in this case, with a better recall and ROC AUC score, making it more reliable for identifying customers who are likely to churn
- Both models benefit from tuning, with improvements in their ability to distinguish between positive and negative instances
- Total Charges, Customer Service Calls, and International Plan are important features influencing customer churn predictions

# RECOMMENDATIONS


Targeted Promotions and Discounts


Improved Plan Pricing Strategies



Significant Features Enhancement


Customer Retention Strategies

# Q & A SESSION

THANK YOU 3>