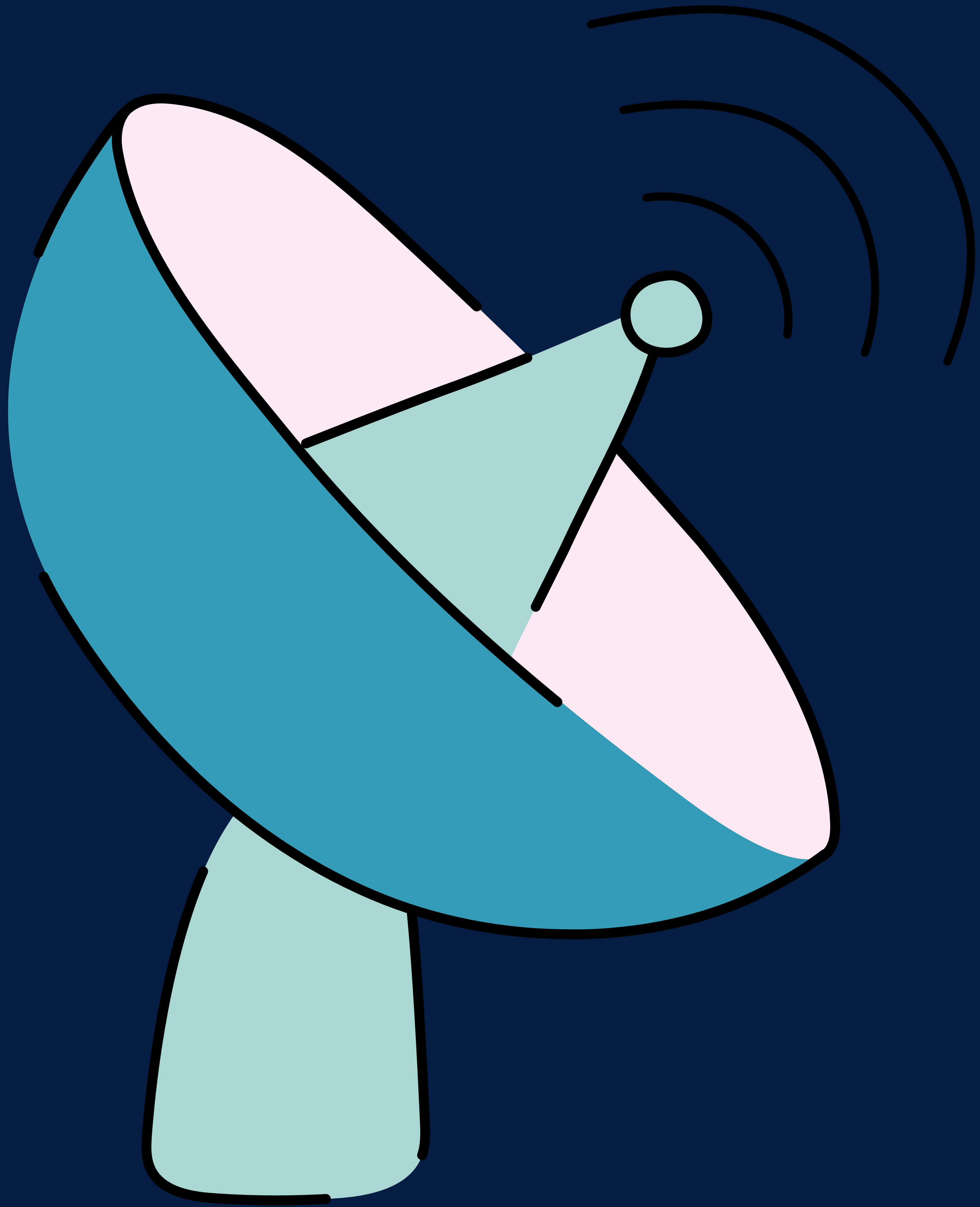
SyriaTel

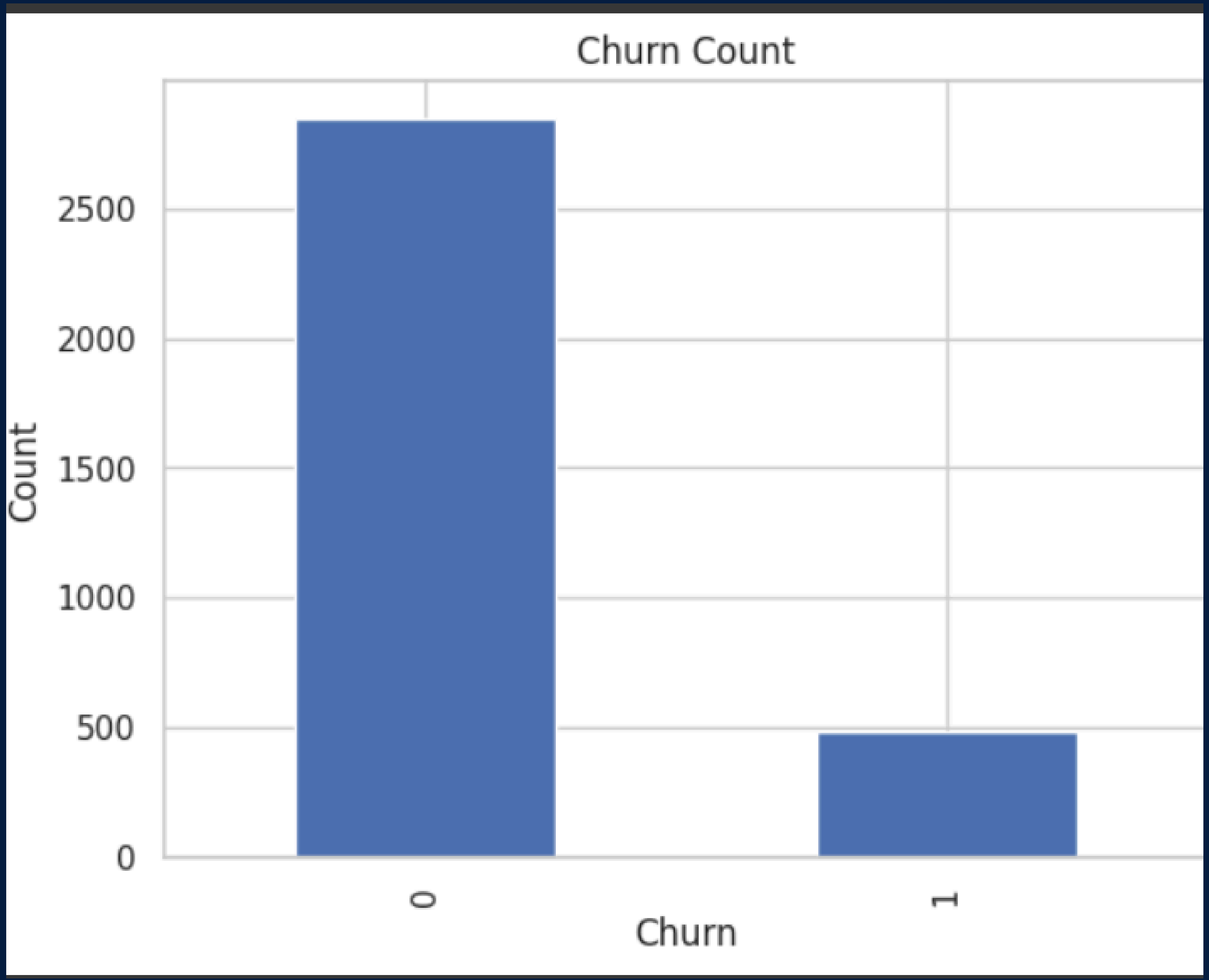
BUSINESS PROJECT

Presented by: Brandon Kanyi Mwangi

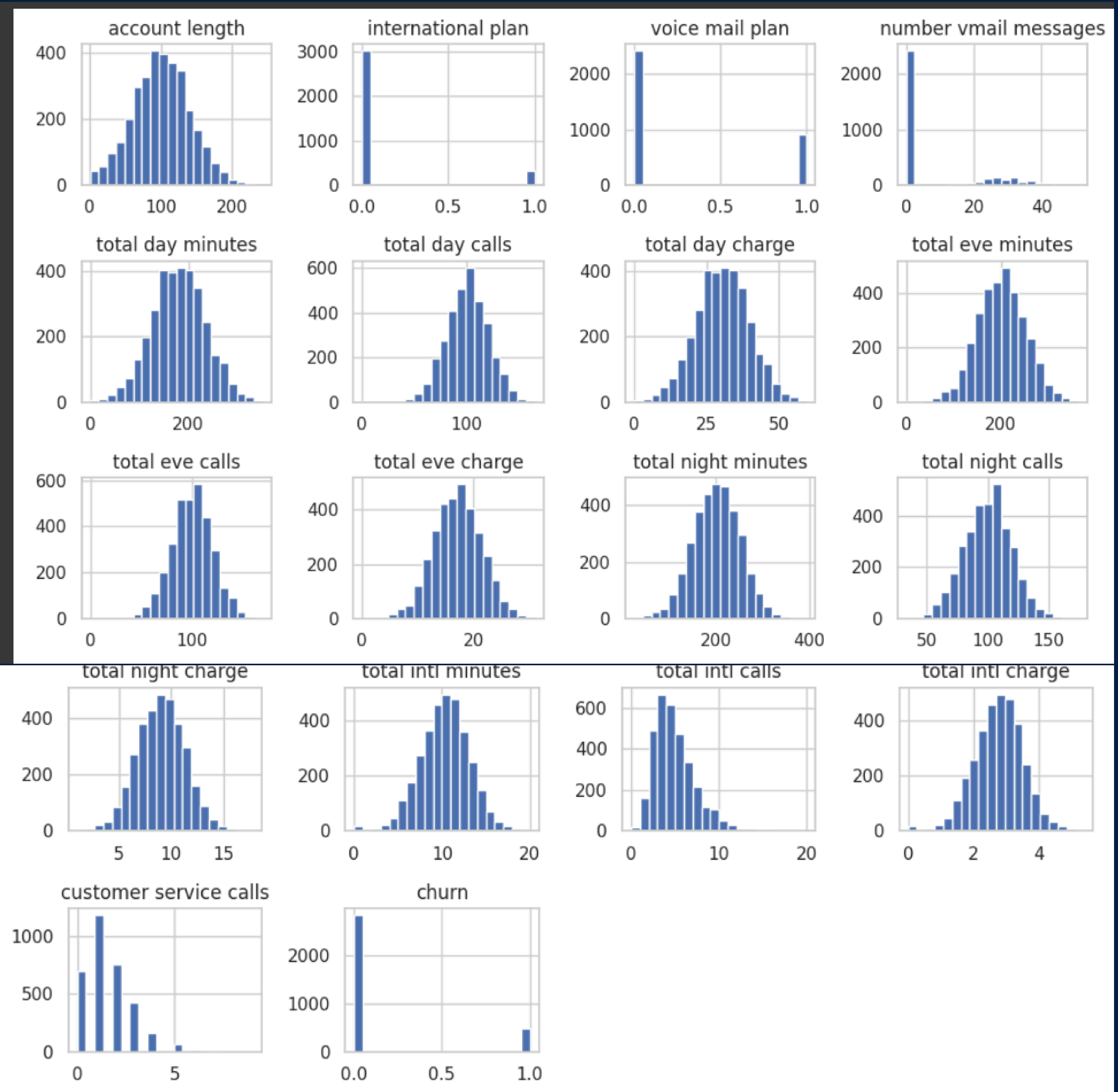
BUILD A CLASSIFIER TO PREDICT WHETHER A CUSTOMER WILL ("SOON") STOP DOING BUSINESS WITH SYRIATEL, A TELECOMMUNICATIONS COMPANY. THIS IS A BINARY CLASSIFICATION PROBLEM. MOST NATURALLY, YOUR AUDIENCE HERE WOULD BE THE TELECOM BUSINESS ITSELF, INTERESTED IN REDUCING HOW MUCH MONEY IS LOST BECAUSE OF CUSTOMERS WHO DON'T STICK AROUND VERY LONG. THE QUESTION YOU CAN ASK IS: ARE THERE ANY PREDICTABLE PATTERNS HERE?



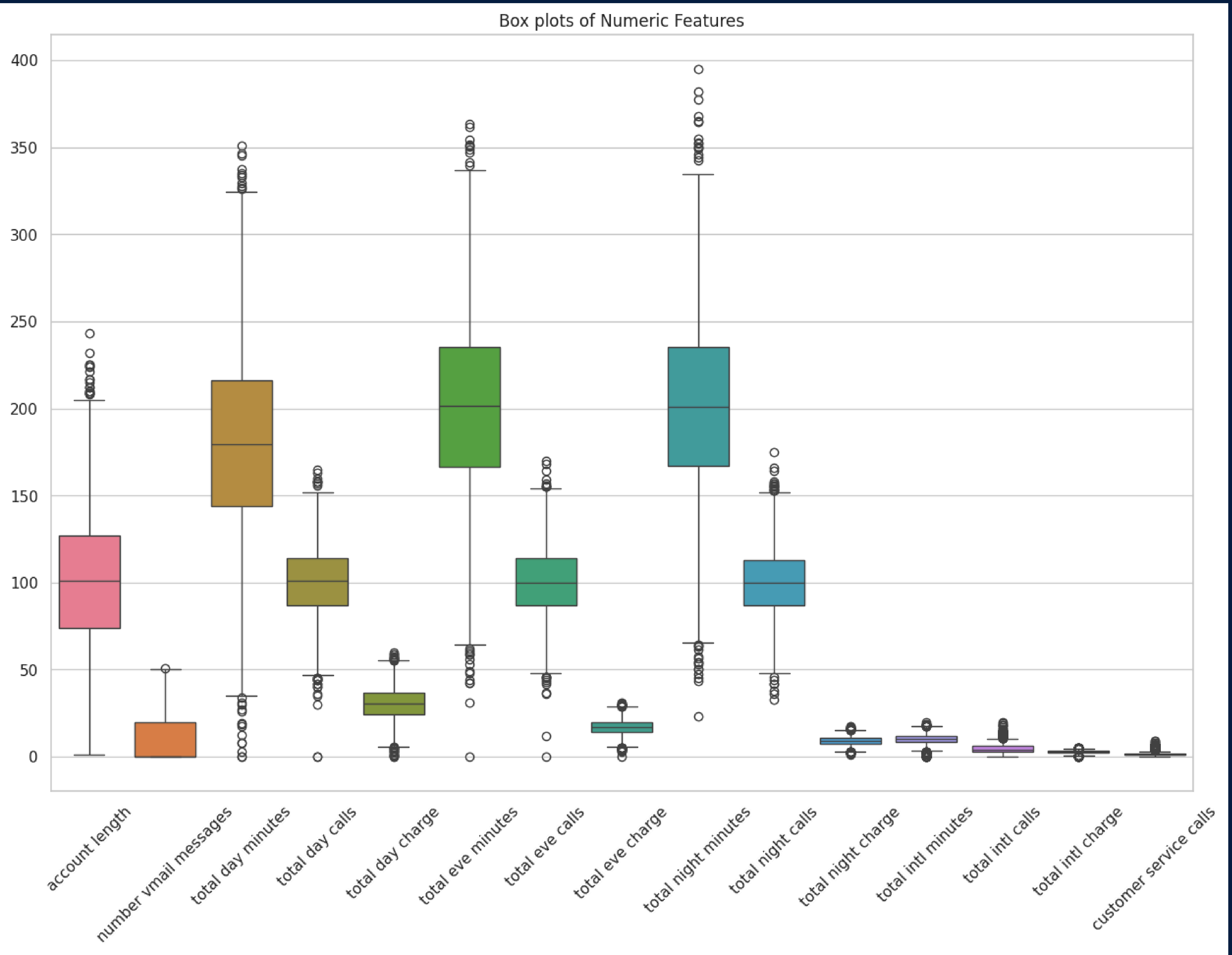
AFTER ANALYSIS WE ALSO BUILD A PREDICTOR MODEL TO ASSIST IN FUTURE PREDICTIONS



Churn is our target variable. From our graph we have an imbalance with our data.

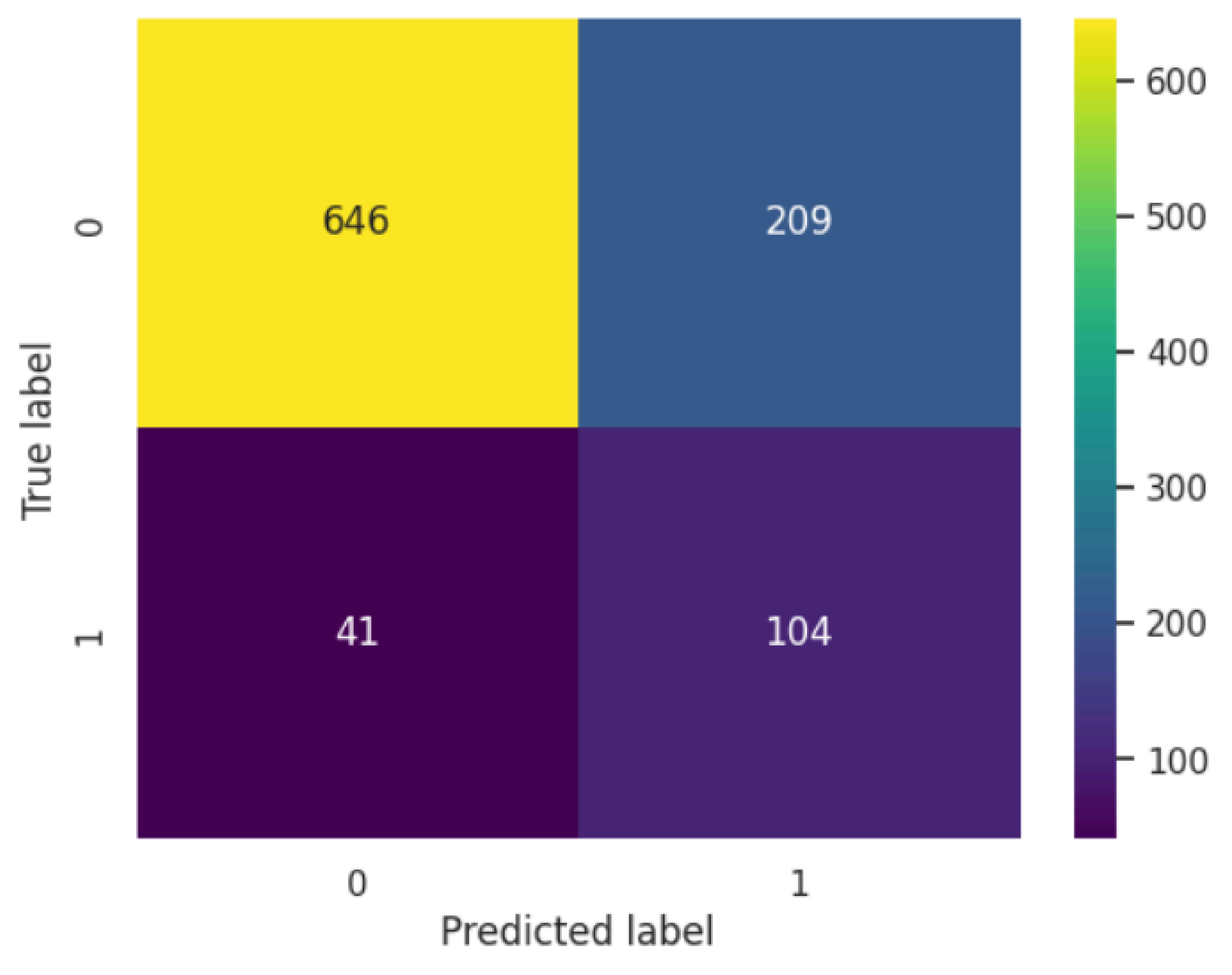


As seen, most of our variables are normally distributed.



- The box plots reveal outliers within the numeric features, indicating the presence of potentially significant data points. While these outliers deviate from the typical distribution, retaining them acknowledges the possibility of important and impactful observations that could enrich the analysis and yield valuable insights.

Baseline logistic model



Imbalance Issue: The model struggles with predicting the minority class (churn). Precision for churn (class 1) is quite low at 0.33, indicating a high number of false positives.

Recall for Churn: The recall for churn (0.72) is relatively good, meaning the model is doing a decent job of identifying actual churn cases, though it comes at the cost of many false positives

Precision:

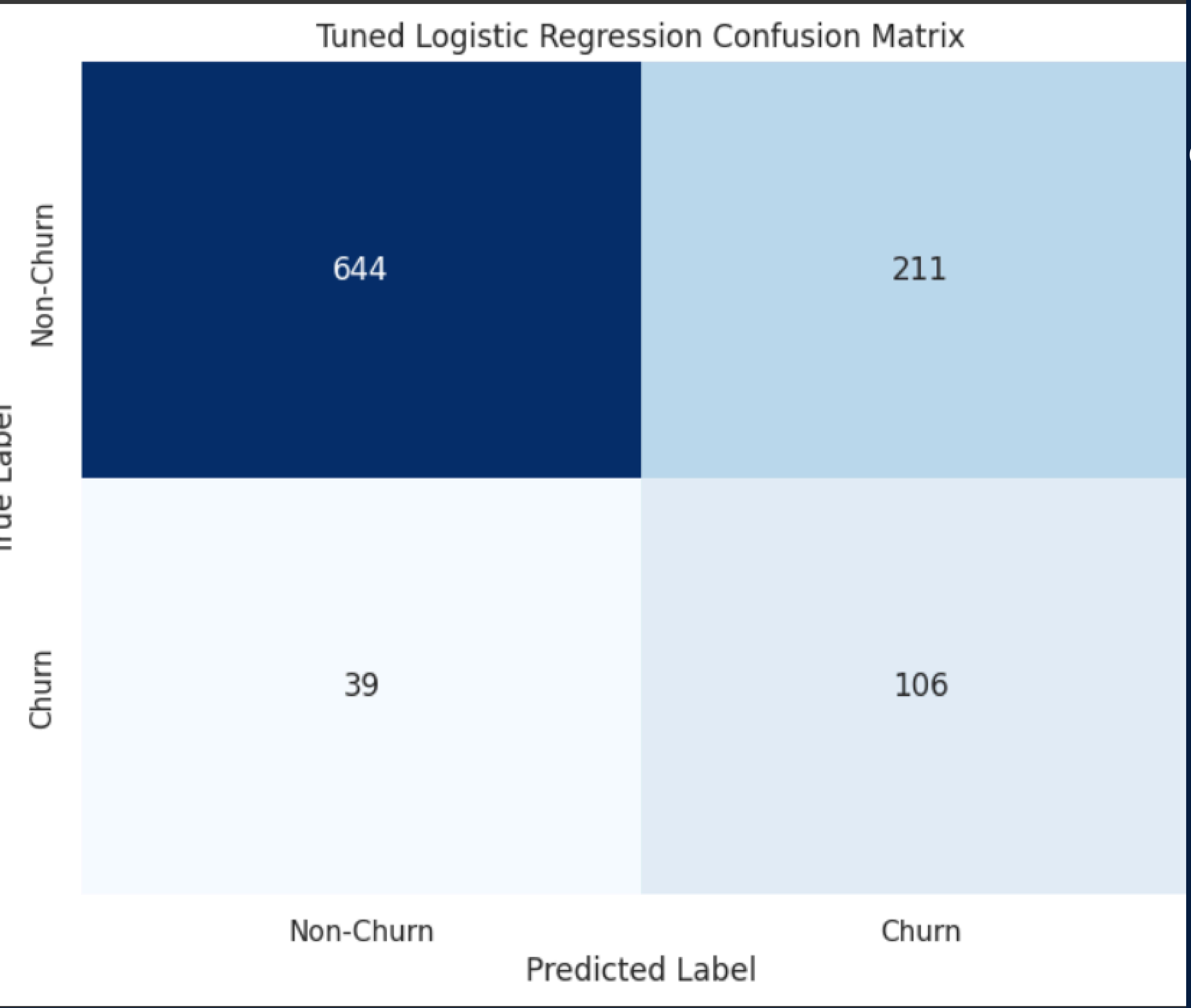
For class 0 (non-churn): 0.94, indicating that when the model predicts non-churn, it's correct 94% of the time.

For class 1 (churn): 0.33, indicating lower precision, so when the model predicts churn, it's correct 33% of the time.

Accuracy Score:

0.75: The model correctly classified 75% of the instances. While this may seem reasonable, the imbalanced nature of the dataset means that accuracy alone might not be the best measure of model performance.

MODEL TUNED LOGISTIC

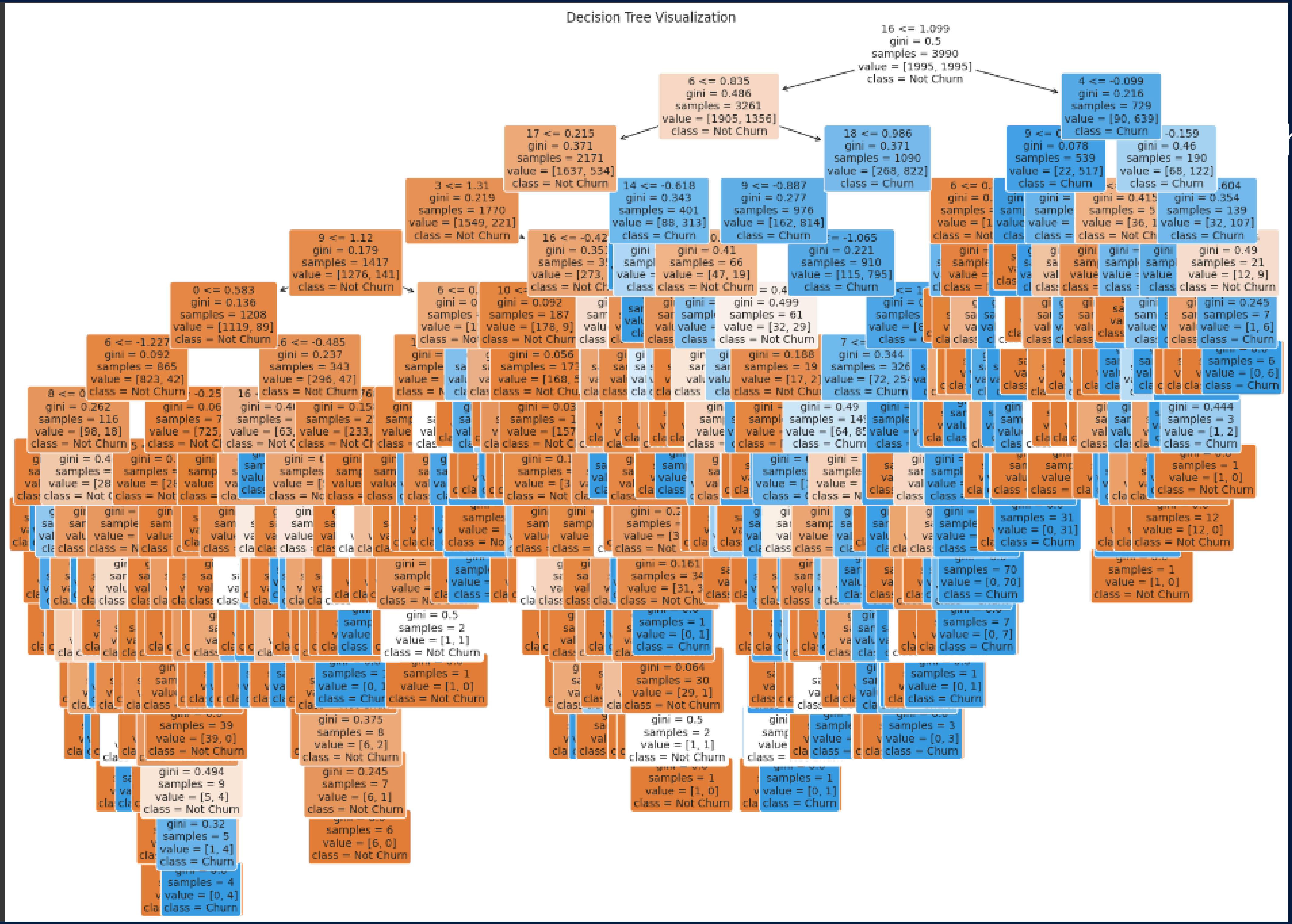


Improvement in Recall:- The recall for the churn class (1) has improved slightly to 0.73 from 0.72, meaning the model is slightly better at identifying actual churn cases.

Precision for Churn:- The precision for the churn class remains low at 0.33, indicating that while more churn cases are correctly identified, there are still many false positives.

Overall Accuracy:- The accuracy of 0.75 hasn't changed, which suggests that while some metrics (like recall) improved, the overall balance of the model's performance is similar.

Basic Decision Tree

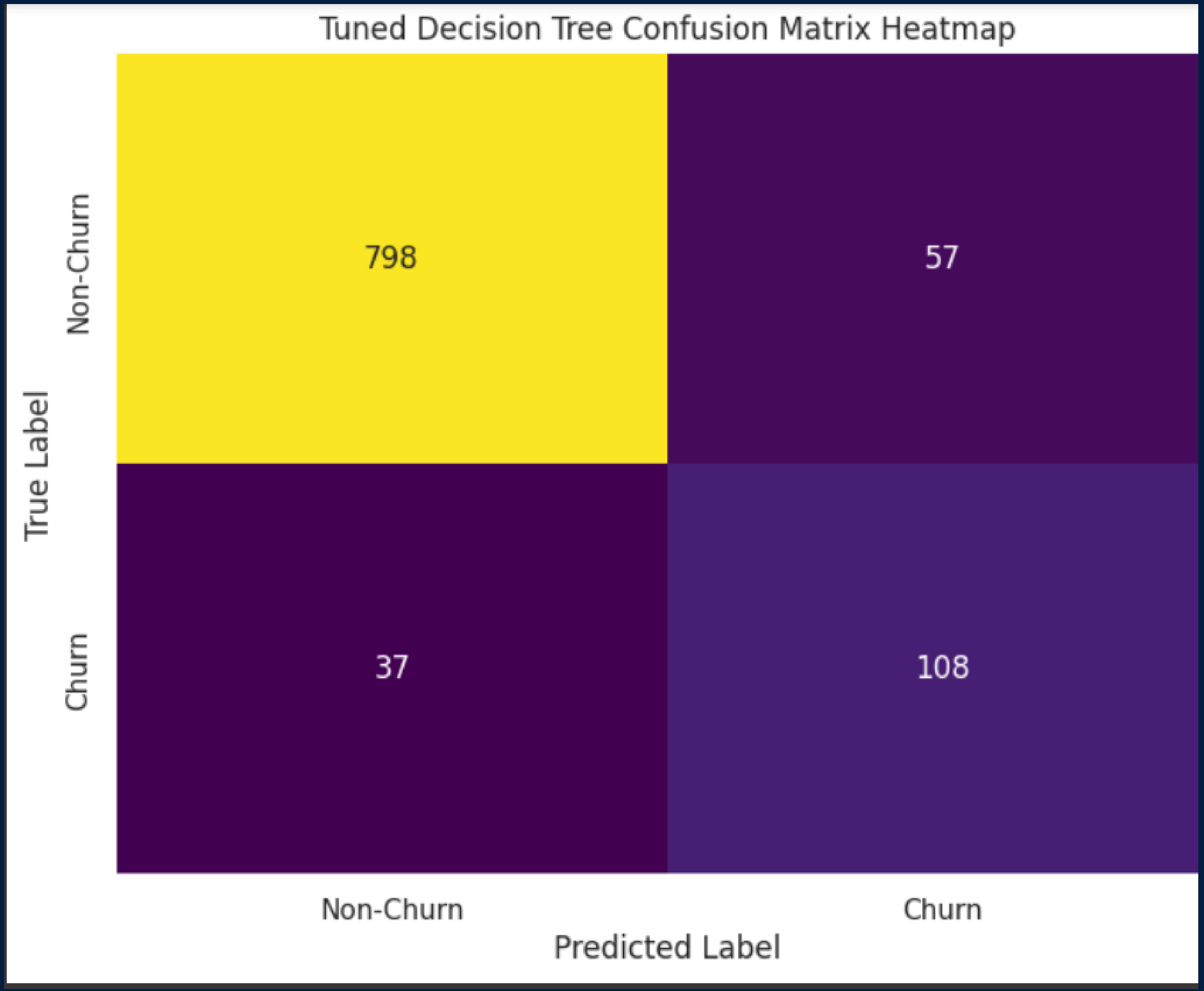


Significant Improvement: The Decision Tree model has significantly improved both precision and recall for the churn class, leading to a much better F1-score (0.61) for this class.

Higher Accuracy: The model achieves a high accuracy of 86.5%, a noticeable improvement over the previous models.

Better Handling of Imbalance: The Decision Tree seems to handle the class imbalance better, particularly with the improved recall for the churn class.

Tuned Decision tree

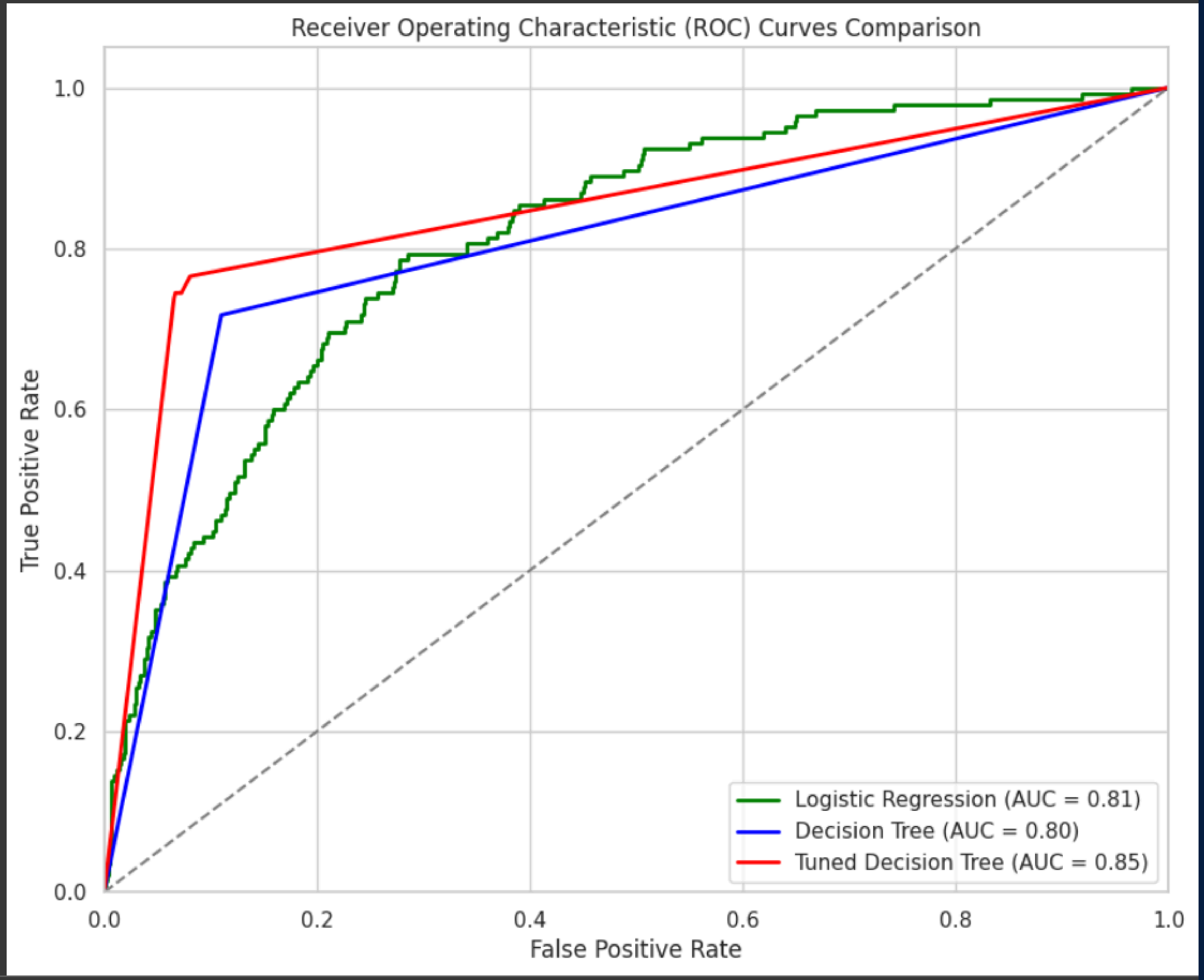


Improved Accuracy: The accuracy of 90.6% indicates that the tuned Decision Tree model is effective in predicting customer churn.

Balanced Performance: The precision and recall for the churn class (class 1) have improved, resulting in a more balanced model with an F1-score of 0.70 for churn prediction.

Effective Tuning: The tuning process has led to a significant improvement in the model's performance, particularly in handling the imbalanced classes.

model comparisons



Tuned Decision Tree: This model has the highest AUC (0.85), indicating the best performance among the three models. Its curve is closest to the top left corner, which signifies better classification.

Logistic Regression: With an AUC of 0.81, it performs slightly better than the Decision Tree but not as well as the Tuned Decision Tree.

Decision Tree: This model has the lowest AUC (0.80) among the three, indicating it has the least effective classification performance.

RECOMMENDATIONS

The predictive model we developed can accurately forecast customer churn based on various customer attributes and behaviors. By leveraging this model, SyriaTel can:

- Target At-Risk Customers: Identify customers at high risk of churn and take proactive steps to retain them, such as offering personalized incentives or improving customer service.
- Optimize Marketing Strategies: Focus marketing efforts on features and services that reduce churn, based on the insights gained from feature importance.
- Enhance Customer Experience: Understand common factors leading to churn and address underlying issues, such as billing disputes, service dissatisfaction, or lack of engagement.

The predictive model not only answers the business question but also provides a practical tool for reducing customer churn, thereby minimizing revenue loss and improving customer loyalty.