SyriaTel Customer Churn Prediction.

Dorine Langat

24th May 2023

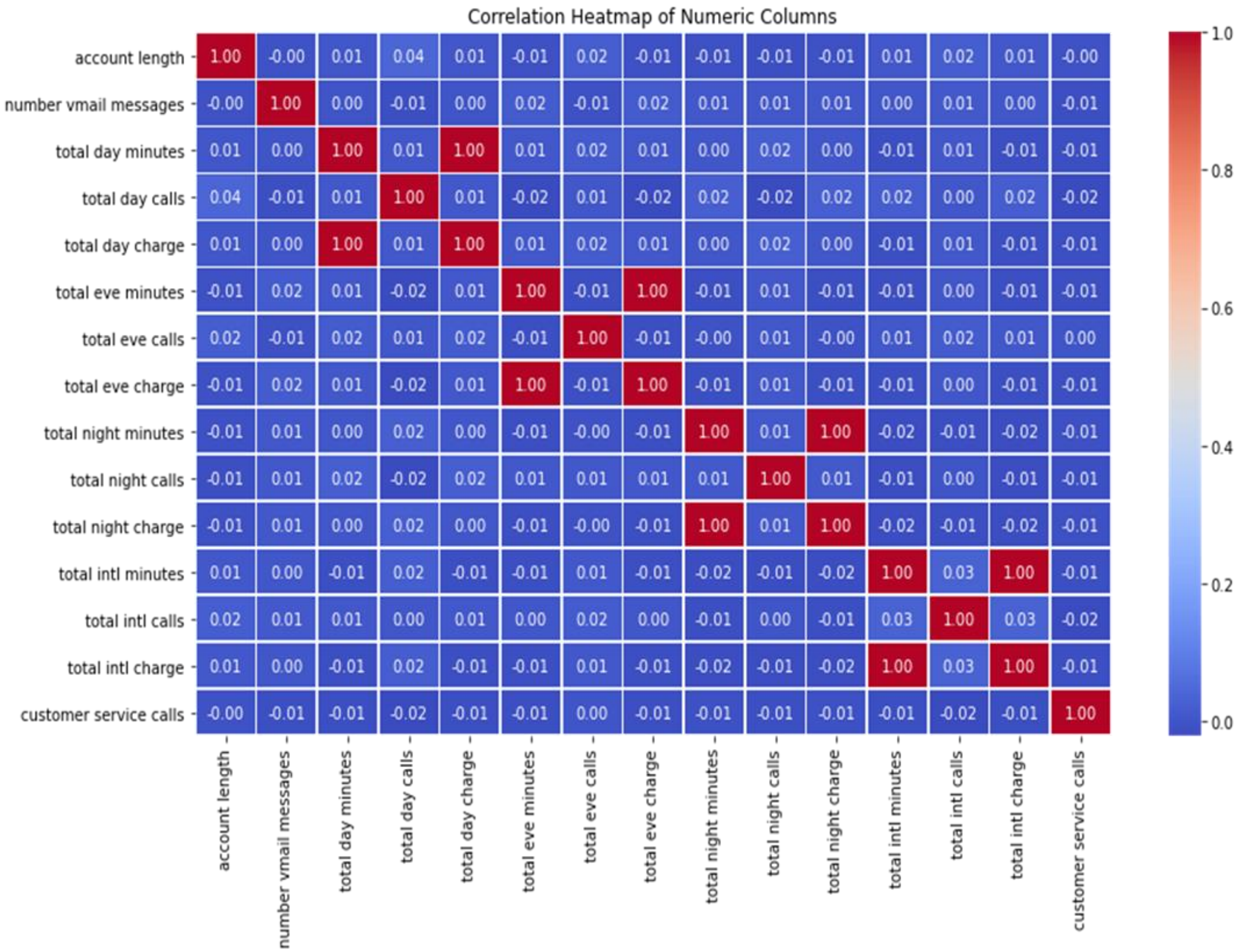
Overview

- SyriaTel, a telecommunication company is interested in knowing whether a customer will stop doing business with them.
- Understanding and predicting customer churn is of paramount importance to telecommunication companies, as it directly impacts their revenue and market share.
- This understanding enables such companies to develop proactive retention strategies, improve customer satisfaction, and ultimately reduce churn rates.
- Furthermore, it can help allocate resources efficiently, and optimize marketing efforts to attract new customers.

Data Exploration

- For this analysis, we used SyriaTel’s customer information.
- The customer information contained in the dataset include customer’s personal information, call details, subscription details, and churn indicators.
- The dataset consists of 3333 entries in each of the 21 columns: (20 feature columns and 1 target column).

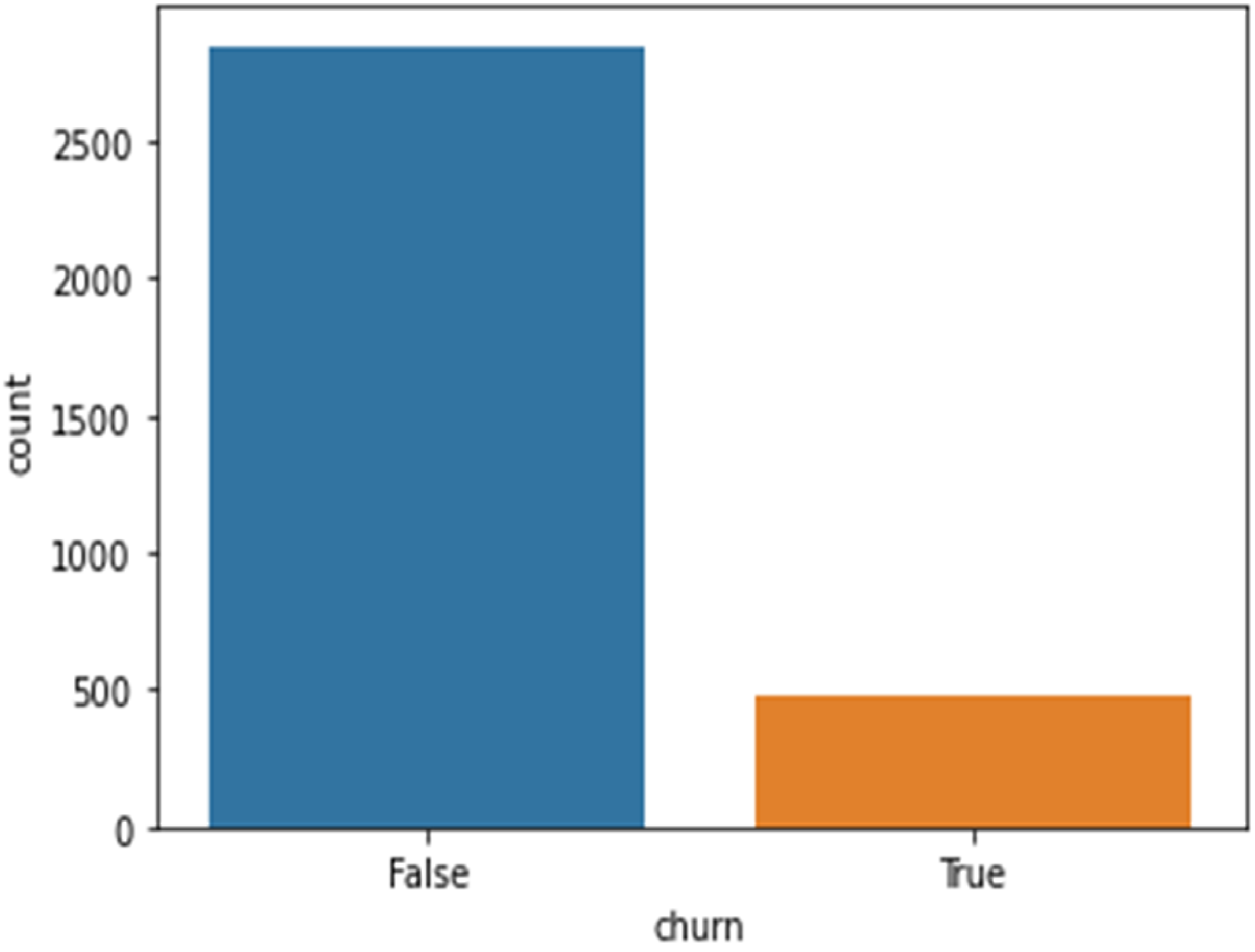
Data Understanding



While most of the features in the dataset do not show significant correlation, there are some pairs of features that exhibit perfect positive correlation.

This perfect positive correlation is expected and makes sense since the charge is directly influenced by the minutes used.

Data Understanding



Out of the 3,333 customers included in the dataset, 483 customers have ended their contract with SyriaTel.

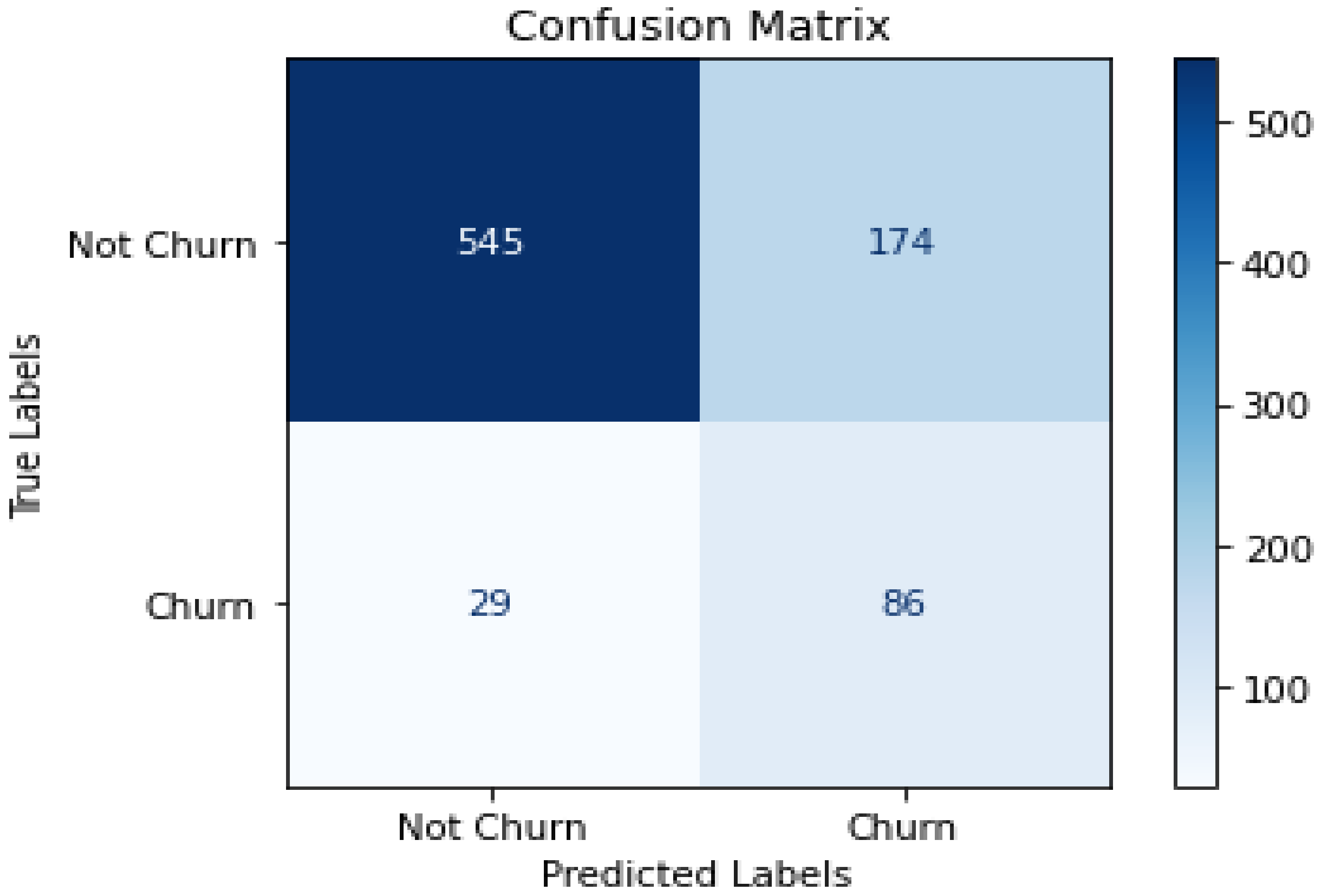
This accounts for approximately **14.5%** of the total customers, indicating a loss in customer base.

Objectives

The objectives of this analysis include:

- To predict customer churn.
- To identify customer churn drivers.

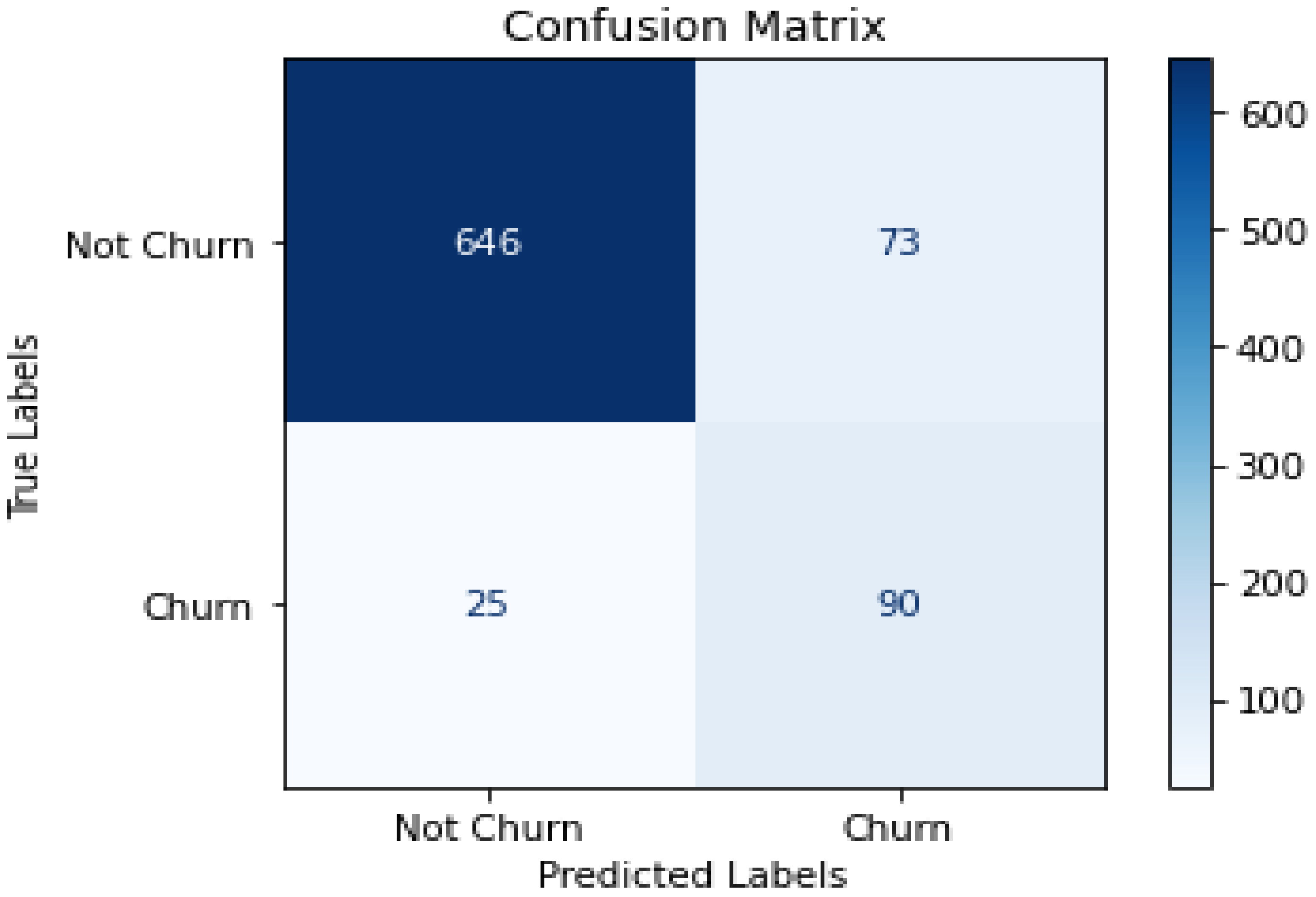
Modeling (Logistic Regression)



The model correctly predicted 86 instances of customer churn (TP), but it incorrectly classified 174 instances of non-churn as churn (FP).

This suggests that the model may not be performing optimally and may require further evaluation and refinement.

Modeling (Decision Tree)

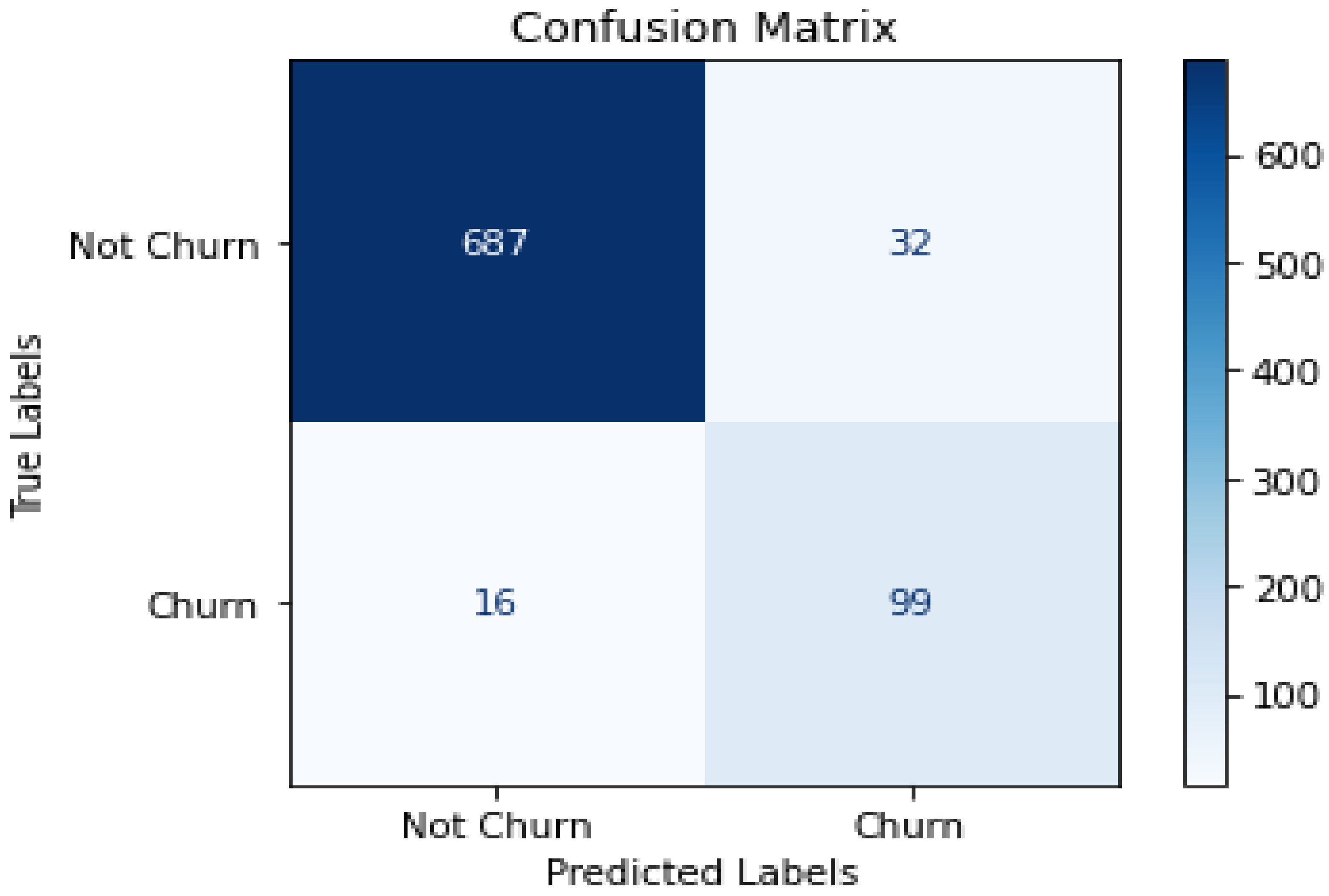


The model correctly predicted 90 instances of customer churn (TP), but it incorrectly classified 73 instances of non-churn as churn (FP).

Modeling

- The decision tree classifier appears to be a more promising model for the given task.
- It shows better performance across various metrics, indicating that it has learned the underlying patterns in the data more effectively.
- However, while the decision tree classifier performs well on the testing dataset, it achieves perfect precision and recall on the training dataset, suggesting potential overfitting.
- Therefore, we fitted a random forest model, which combines multiple decision trees to enhance predictive performance and mitigate the risk of overfitting.

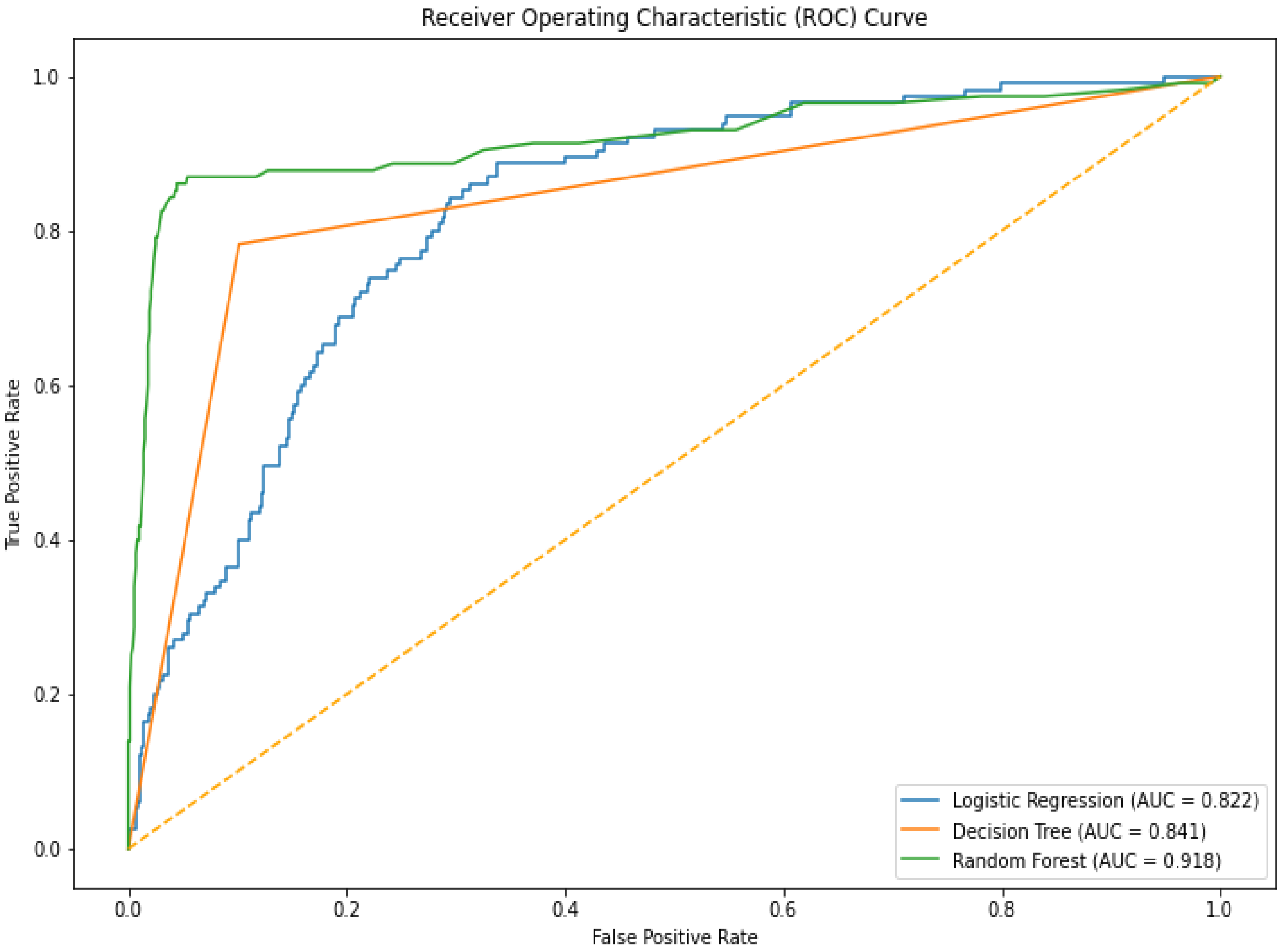
Modeling (Random Forest)



The model correctly predicted 99 instances of customer churn (TP), but it incorrectly classified 32 instances of non-churn as churn (FP).

The false negatives (16) indicate potential missed opportunities to identify customers at risk of churn, which could impact the effectiveness of retention strategies.

Evaluation (ROC)

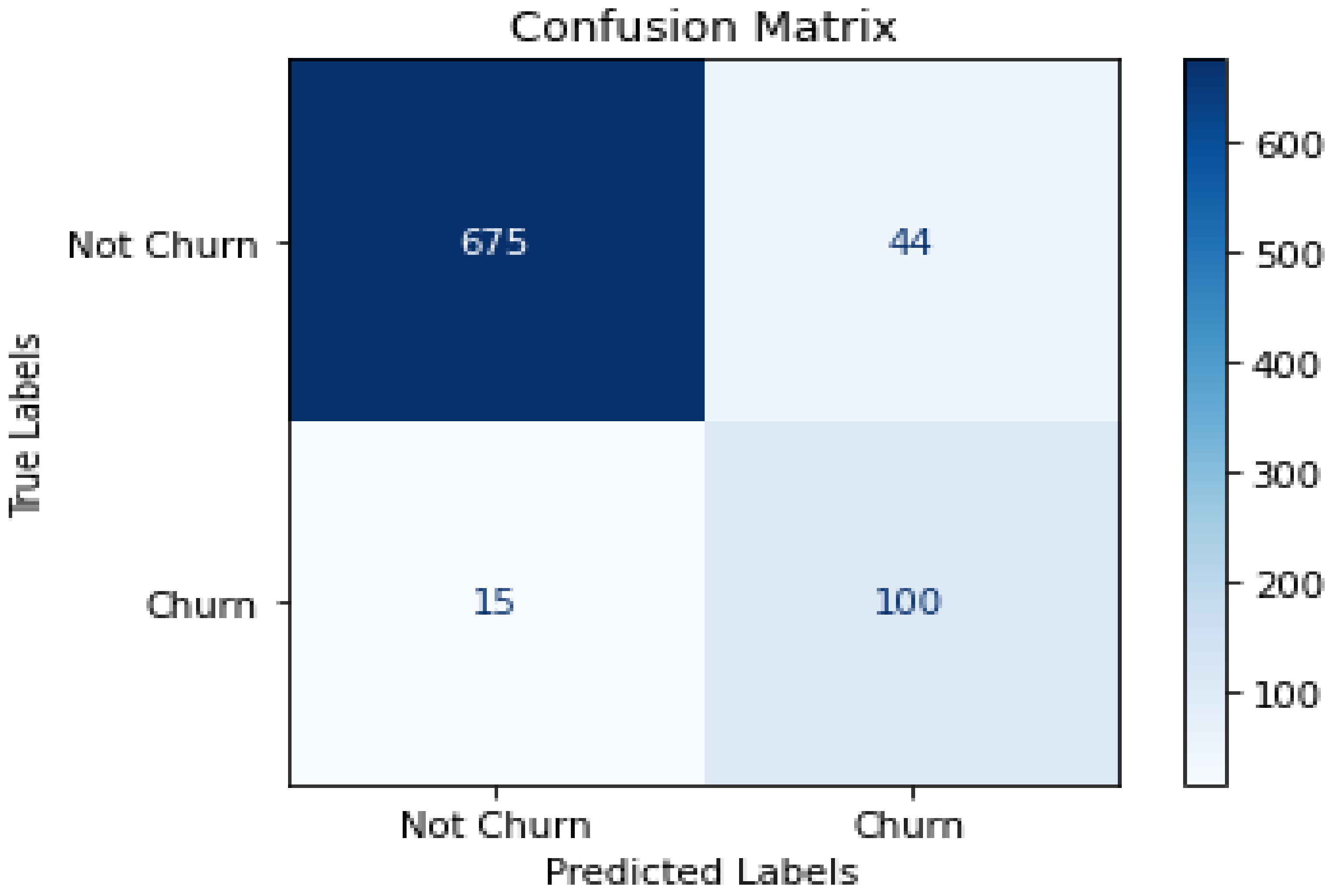


1. **Logistic Regression**: has a reasonably good performance a trade-off between correctly identifying churned customers and minimizing false positives.

2. **Decision Tree**: has a slightly better predictive performance, and better discrimination between churn and non-churn customers.

3. **Random forest**: has an excellent overall predictive performance, and effectively distinguishes between churn and non-churn customers with a high degree of accuracy.

False Positive - False Negative Trade-off



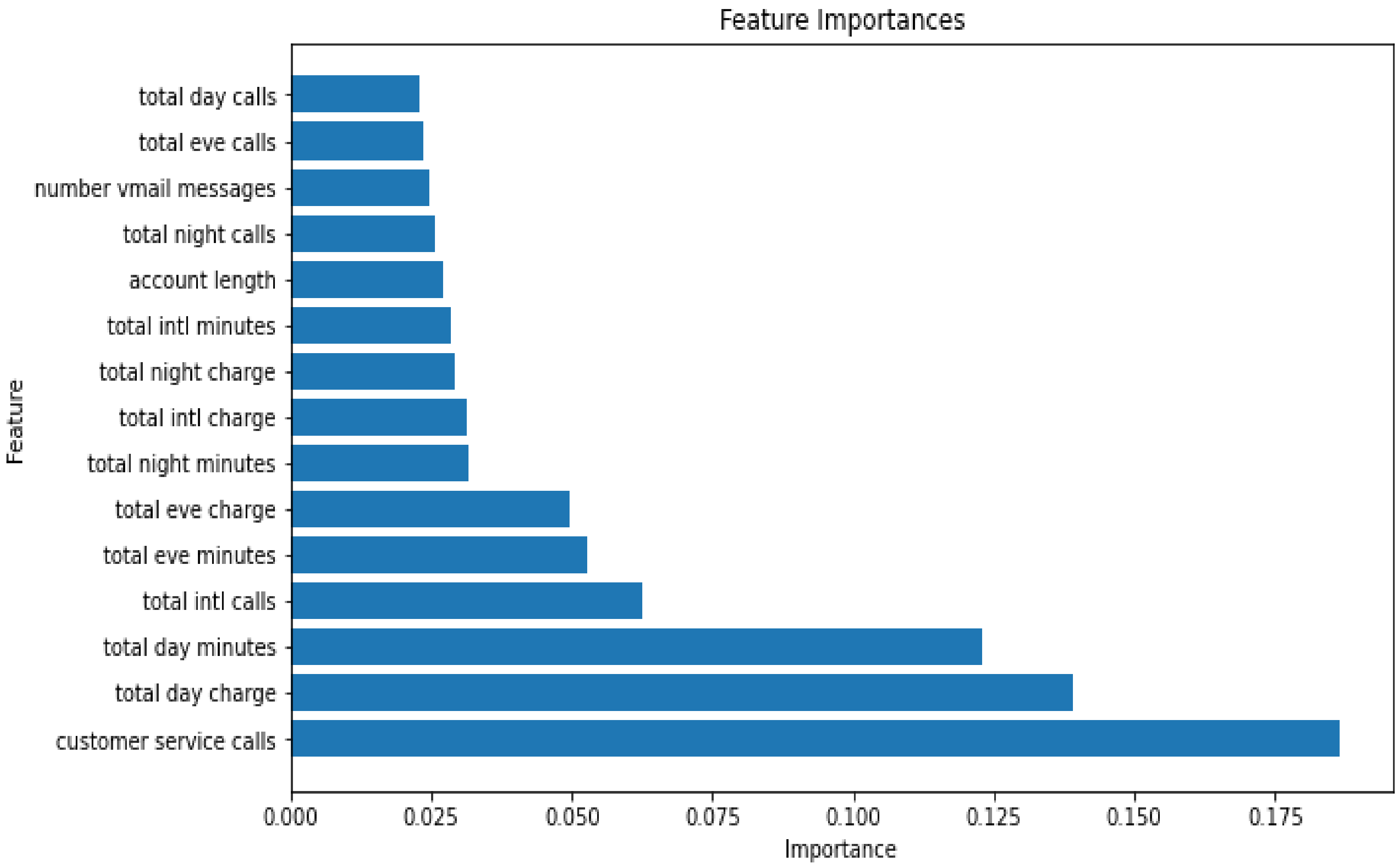
Here, we prioritized reducing false positives at the expense of slightly increasing false negatives.

This adjustment aligns with our business objective, as it reduces the risk of losing customers who were mistakenly classified as non-churned.

Evaluation

- As we progress from the logistic regression model to the decision tree, random forest, and the false positive and false negative trade-off matrix, there is an overall improvement in the model's performance.
- The accuracy in correctly identifying churned and not churned customers increases, and the false positive rate generally decreases.
- The trade-off between false positives and false negatives was intentional for our objective as it reduces the risk of losing customers who were mistakenly classified as non-churned.

Feature Importance



The features that contribute the most to whether a customer churns or not include **customer service calls**, **total day charge**, **total day minutes**, **total international calls**, and **total eve minutes**.

Conclusion

- Our model can correctly make predictions for approximately 93.65% of the customers, indicating that the model's predictions were accurate for the majority of the customers.
- Out of all the customers predicted as churned, approximately 72.46% of them actually churned, indicating that when the model identified a customer as churned, it was correct around 72.46% of the time.
- Our model successfully captured about 86.96% of the customers who truly churned.
- The F1 score of 79.05% indicates that our model achieved a balanced trade-off between correctly identifying churned customers and minimizing false predictions.

Recommendations

- **Targeted Retention Strategies**: The company should focus on customers who have a high likelihood of churning based on the model's predictions. By identifying these customers in advance, the company can proactively engage with them, offer personalized incentives, discounts, or improved services to encourage their continued loyalty.
- **Customer Segmentation**: SyriaTel should categorize customers into different segments based on their churn propensity hence tailoring its marketing and retention efforts more effectively.
- **Review Pricing Structure**: SyriaTel should consider analyzing the charges and exploring options to optimize pricing plans or introduce flexible pricing options that meet the diverse needs of customers.
- **Improve customer service**: since customers who contact customer service frequently are more likely to churn, the company should empower customer service representatives to provide personalized and empathetic support, as well as proactive follow-ups to address any lingering concerns and increase customer satisfaction.

Next Steps

- **Ongoing monitoring and evaluation**: there is need for regular model evaluation and iteration to boost model performance.
- **Enhancing False Negative Predictions**: Although the recall score indicates that our model captured a significant proportion of churned customers, there is room for improvement to reduce false negative predictions. It is crucial to identify these customers to implement proactive retention strategies and prevent churn.

Thank You.