FROM SATISFIED SUBSCRIBERS TO SILENT EXITS: UNVEILING CUSTOMER CHURN PATTERNS AT SYRIATEL WITH MACHINE LEARNING.

Presented by: Sonia Akinyi Ojay

Phase 3: Final Project Submission

Agenda

- Business Overview
- Business Problem
- Data Understanding
- Analysis
- Results
- Conclusions
- Next Steps

Overview

The telecommunications industry is a highly competitive landscape where customer retention plays a crucial role in a company's success. Customer churn, defined as the loss of subscribers to a competitor or service termination, poses a significant financial challenge for telecommunications companies. SyriaTel, a leading telecommunications provider, is no exception.

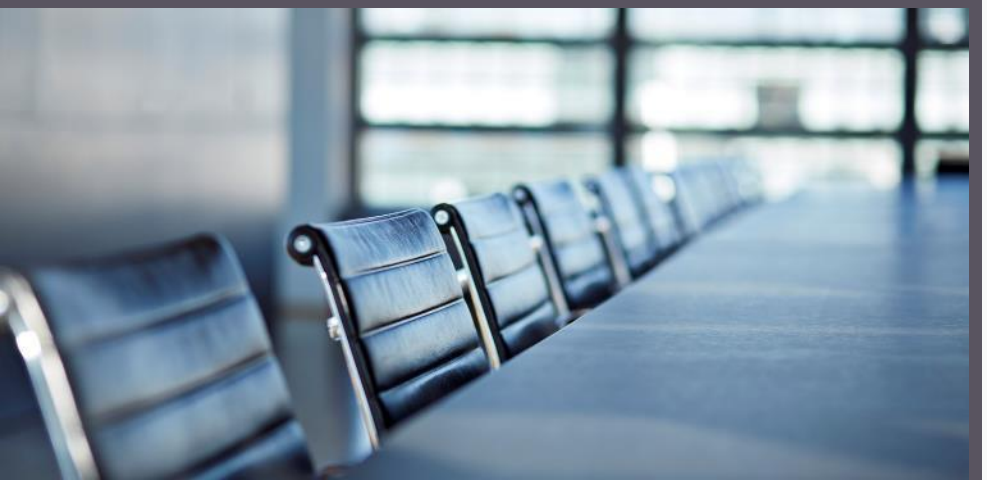
To maintain a competitive edge and minimize revenue loss, SyriaTel requires a robust system to predict customer churn and implement targeted retention strategies. The loss of subscribers to competitors or service termination, is a major financial concern for SyriaTel.

This research proposal outlines a plan to develop a machine learning classifier that can predict customer churn with high accuracy. By identifying customers at risk of churning, SyriaTel can proactively implement targeted retention strategies, minimizing revenue loss and maximizing customer lifetime value.

Business Problem

Customer churn is a significant financial burden for telecommunications companies like SyriaTel. The inability to accurately predict churn hinders the development of effective retention strategies.

Churn rate is very important because it is typically more expensive to obtain new customers than to retain existing customer. This research aims to address this problem by building a machine learning classifier to predict customer churn for SyriaTel inorder to improve their churn rate.



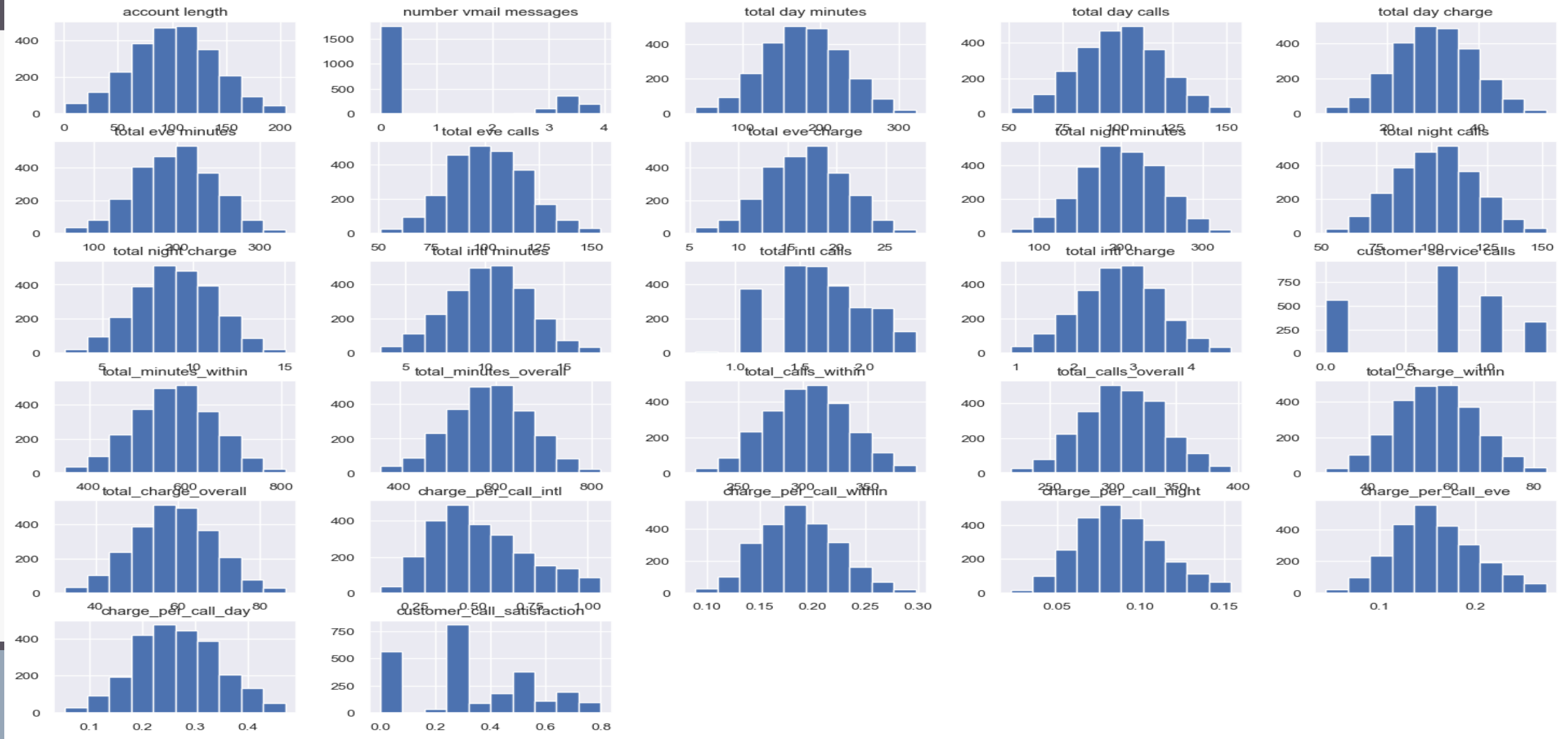
Data Understanding

The data frame used entails SyriaTel customer usage pattern and their churn status from Kaggle. The data included ~The dataset has 3,333 customers and 20 columns with no null values and a mix of data types. The data include:

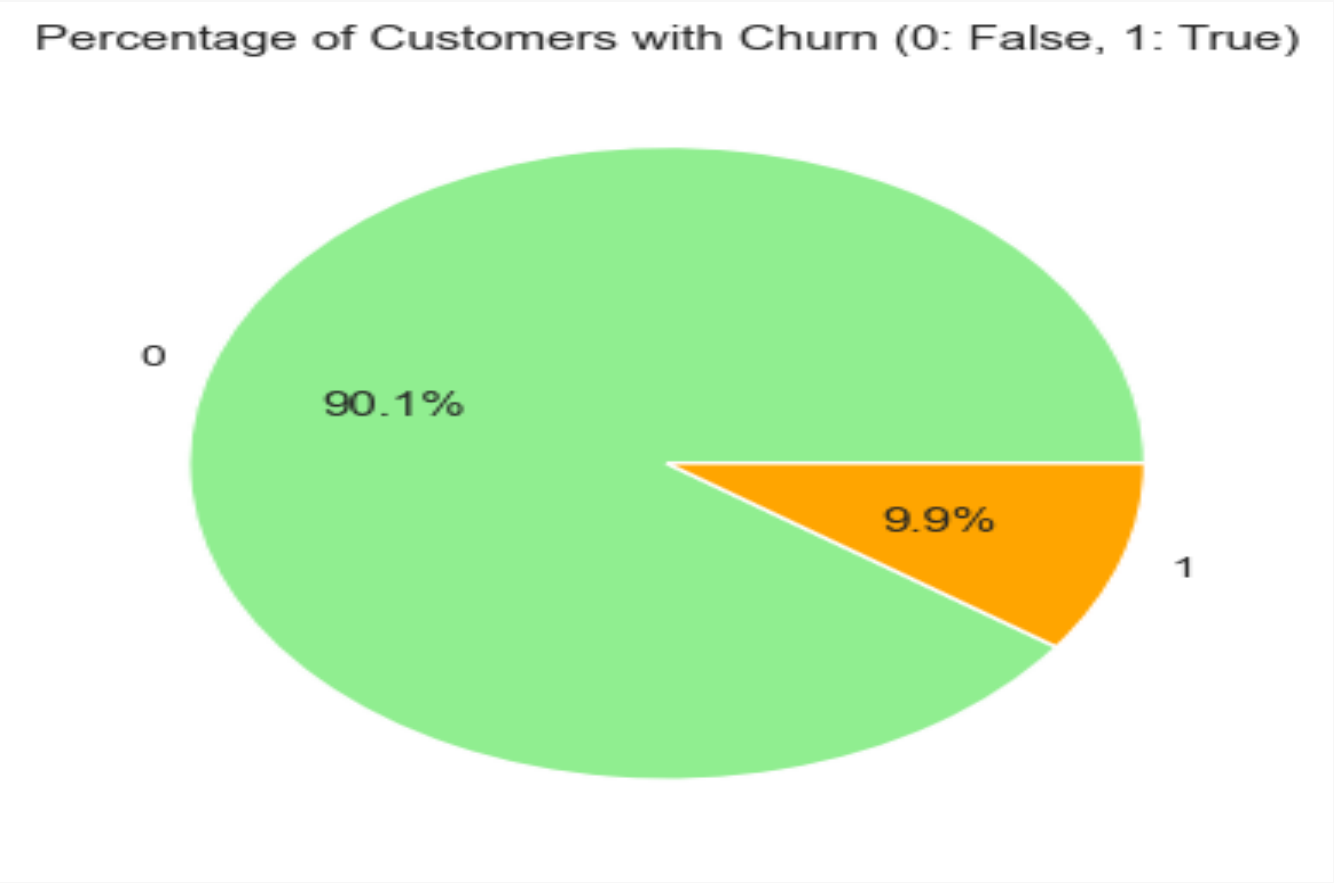
- The Target or Dependent Variable: **Churn**: the number of existing customers who may leave the service provider over a given period.
- Predictors or Independent Variables: such as total day charge, total evening calls etc

Analysis

Univariate Analysis – on Predictors



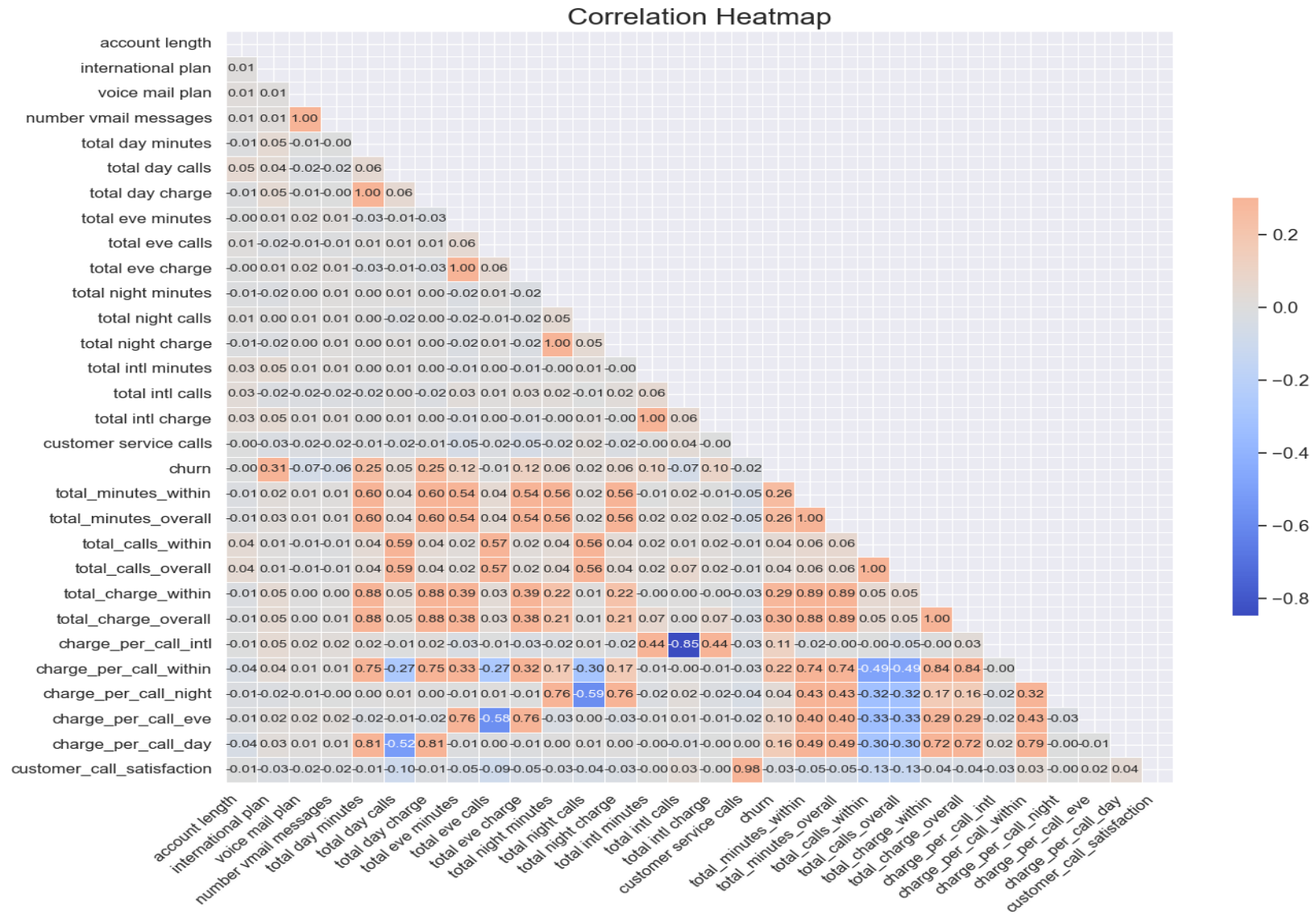
Univariate Analysis – on the Target “Churn”



There is an imbalance in the data where there are far more people who do not churn compared to those who churn.

Bivariate Analysis

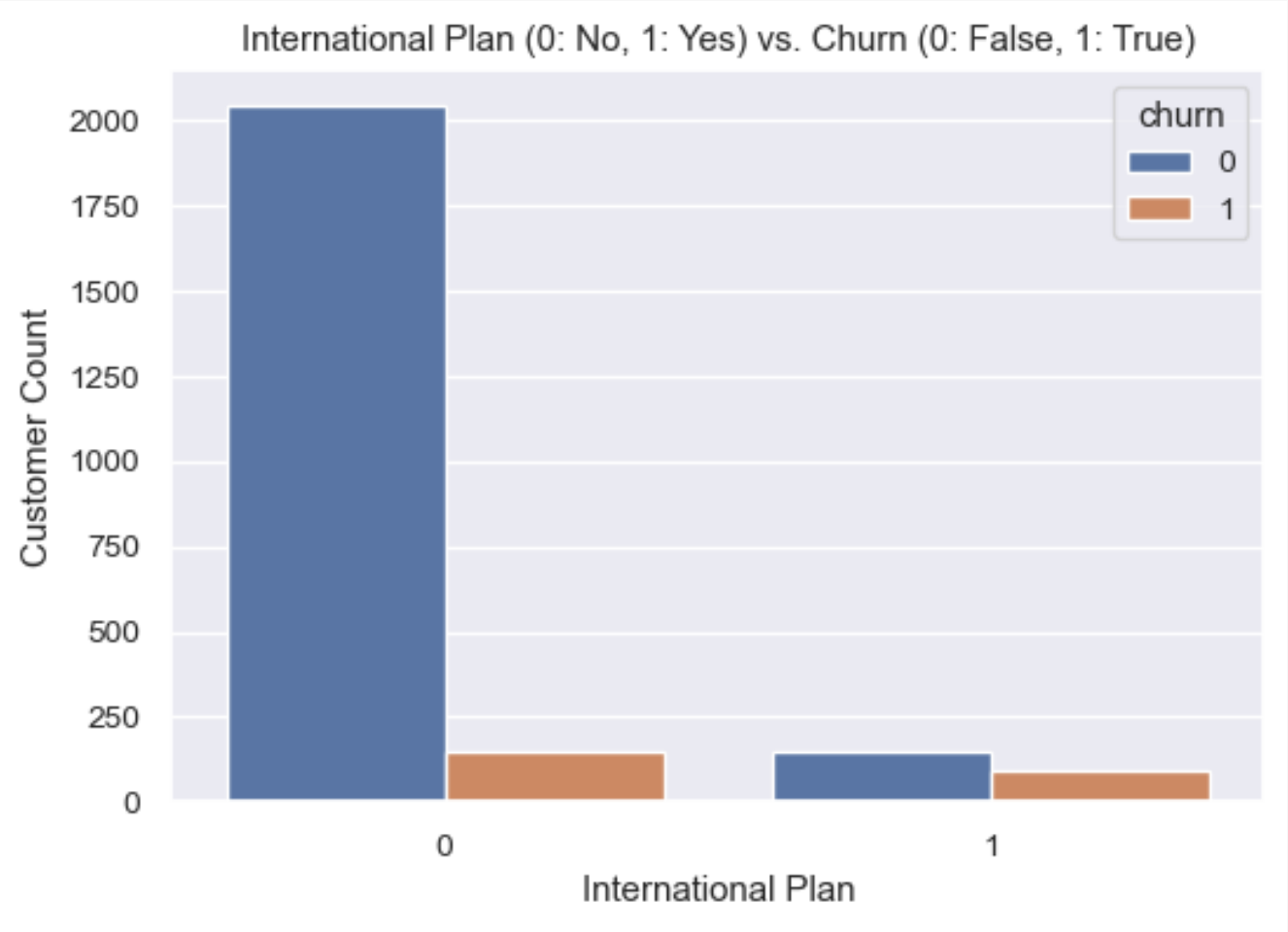
Correlation Matrix



Shows a Moderate Positive Correlations btn Churn and the below predictors.

- International Plan: 0.31,
- Total Day Charge: 0.25,
- Total Day Minutes: 0.25,
- Total charge overall:0.30,
- Total charge within:0.29,
- Total Minutes Overall:0.26,
- Total Minutes Within:0.26,
- Charge per Call Within:0.22

International Plan vs. Churn

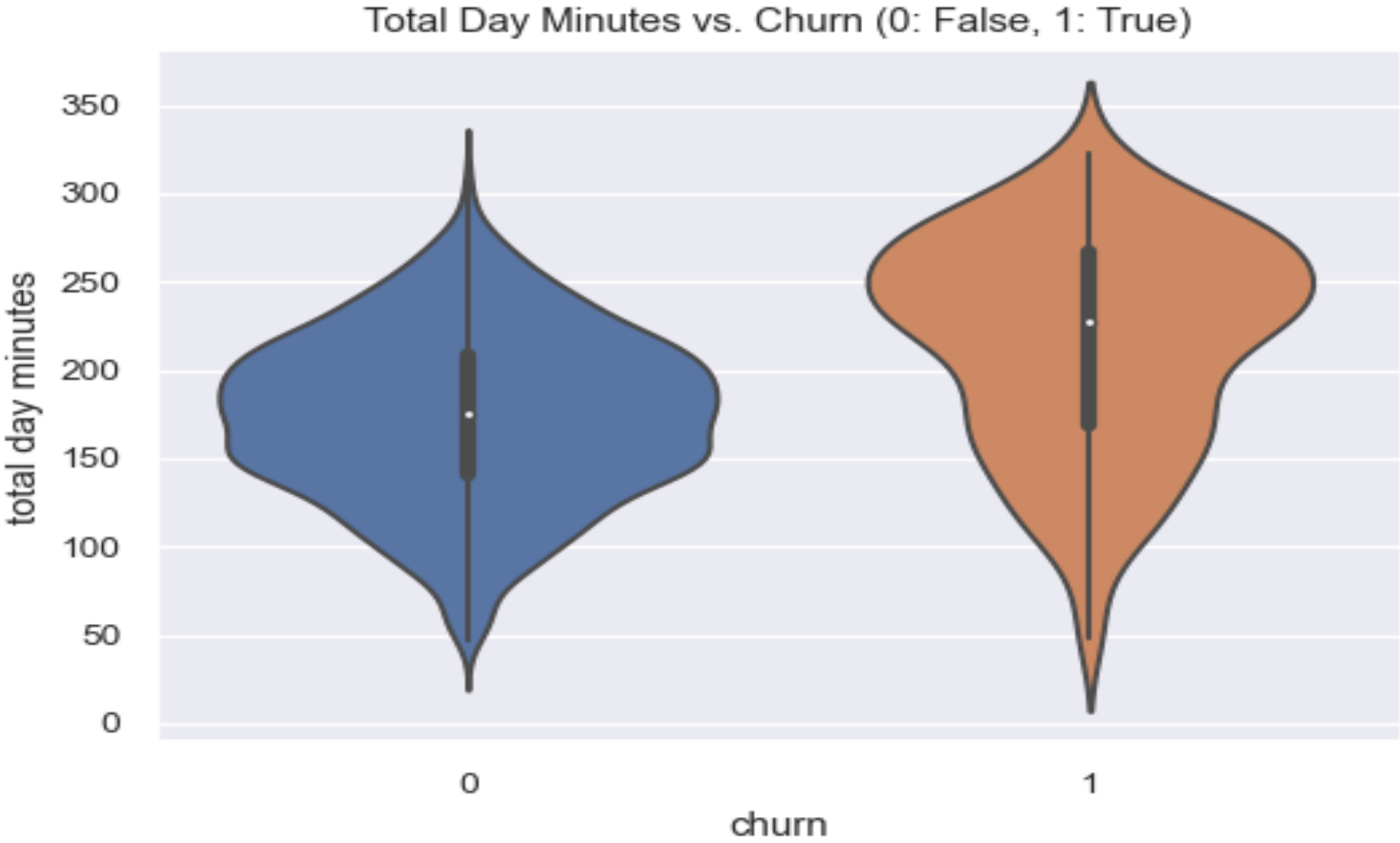


This plot shows the distribution of churned and non-churned customers among those with and without an international plan.

A higher proportion of non-churned customers among those without international plan visually confirm the correlation.

This suggests a very low churn rate (represented by the count) among customers without international plan compared to those who have it.

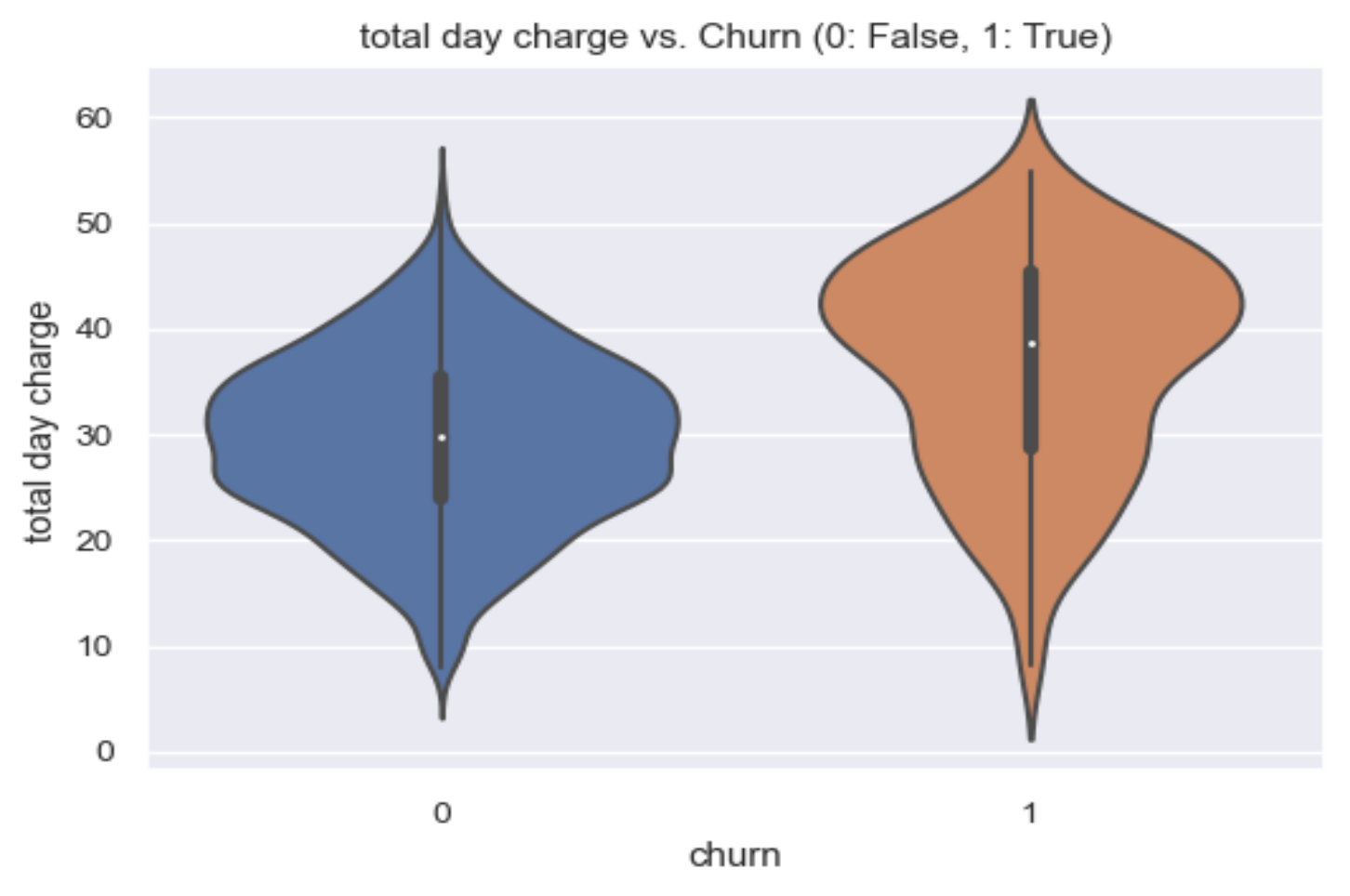
Total Day Minutes vs. Churn



The violin for non-churning users (0) is wider and shifted to the right compared to the violin for churning users (1). This indicates that users who spend more total day minutes tend to be less likely to churn (represented by 0).

Higher total day minutes usage appears to be associated with a lower churn rate. Non-churning users exhibit a wider range of usage patterns, with some using the service more frequently than others. Churning users tend to have lower total day minutes usage overall, with some not using the service at all.

Total Day Charge vs. Churn

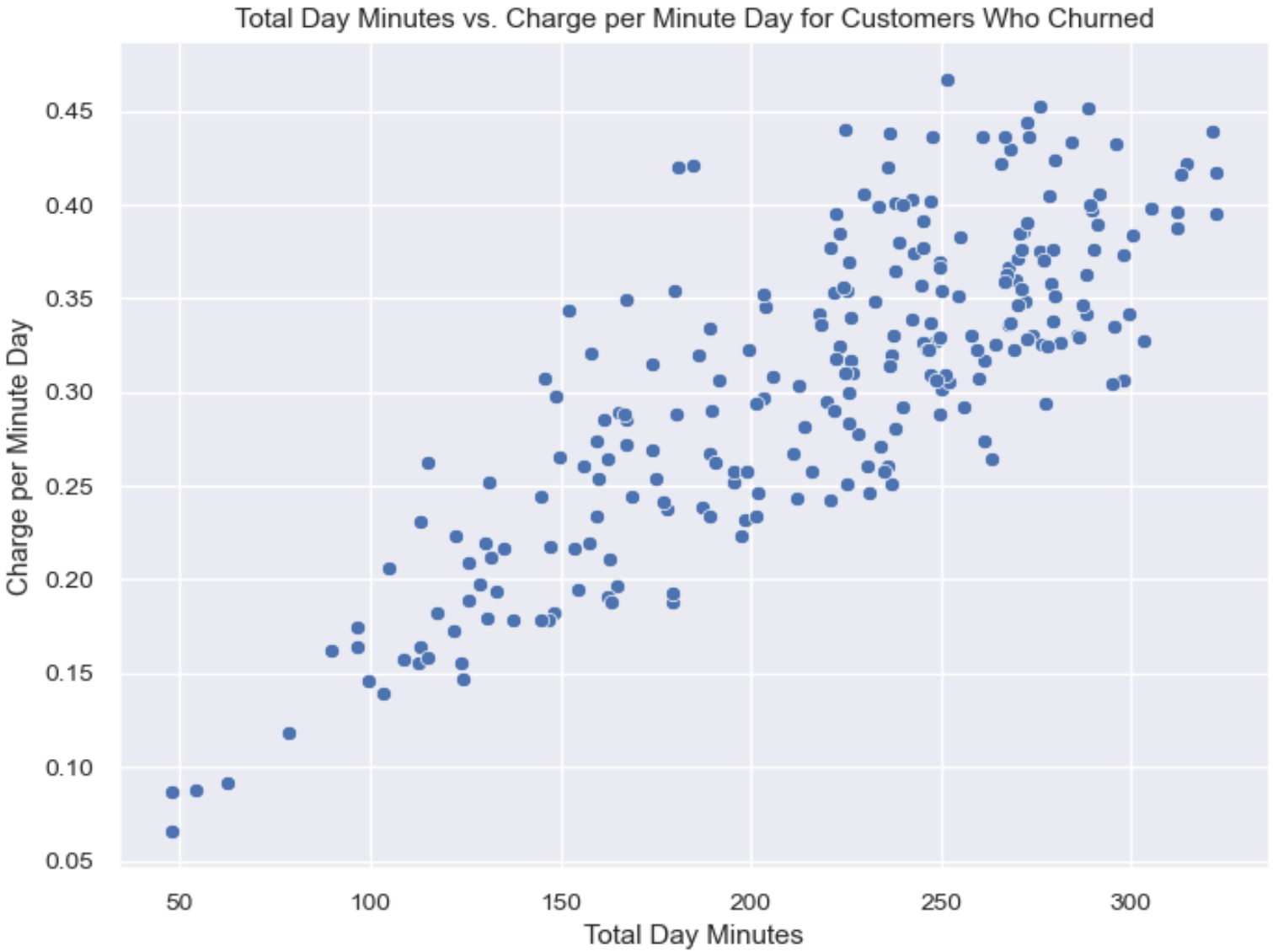


Users who pay a higher total day charge appears to be associated with a lower churn rate.

Non-churning users exhibit a wider range of usage patterns, with some paying for the service much more than others.

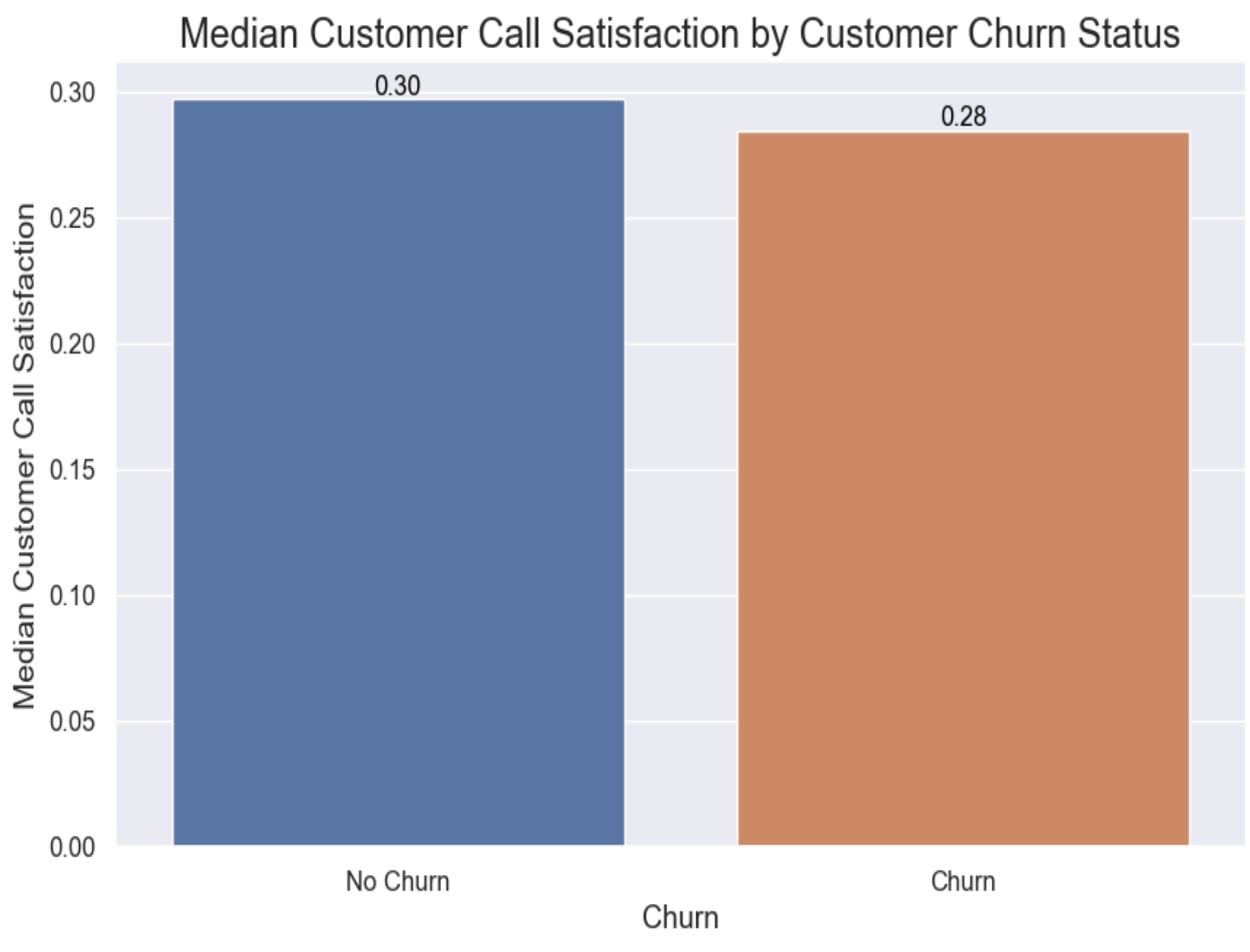
Churning users tend to have lower total day charge overall, with some not paying for the service at all.

Total Day Minutes vs. Charge per Minute Day for Customers Who Churned



From the graph, customer satisfaction or the proportion of customer service calls to total calls differed between the two classes.

Total Day Minutes vs. Charge per Minute Day for Customers Who Churned



The median customer call satisfaction is slightly higher for customers who did not churn (0.3) compared to those who did churn (0.28).

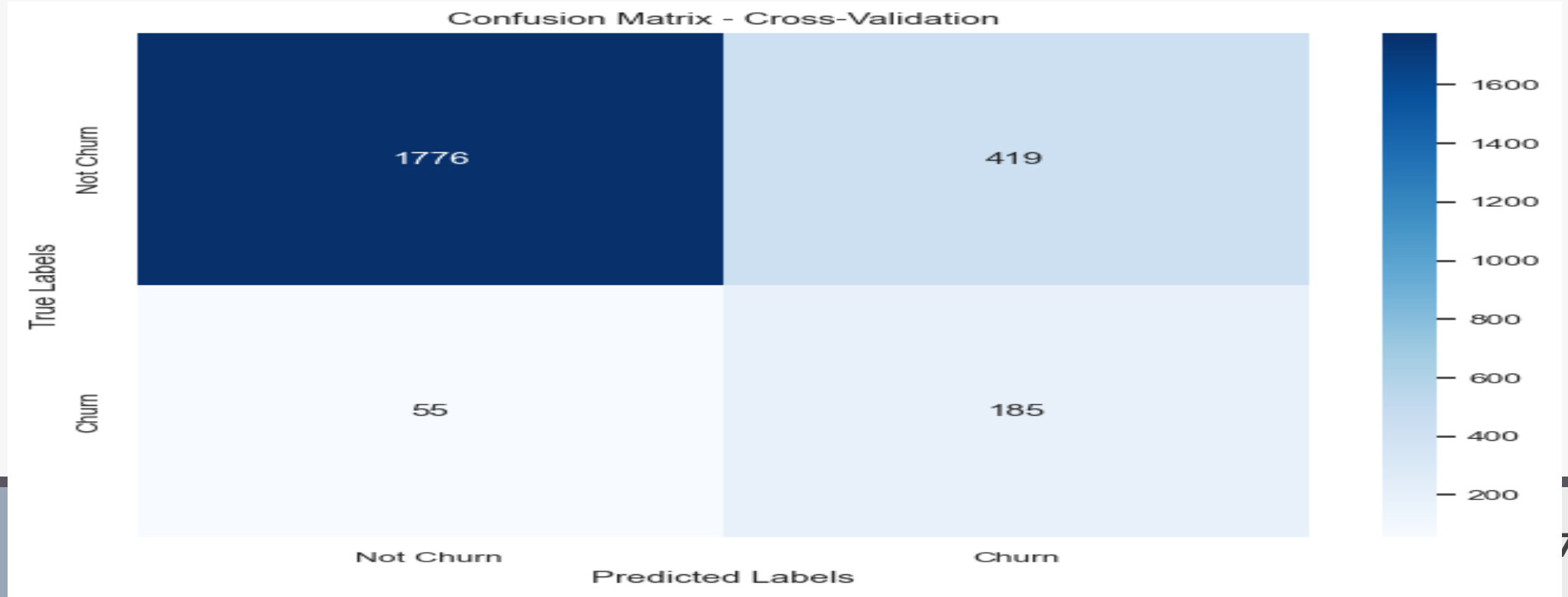
While the difference is small (0.02), it indicates a subtle pattern: customers who are less satisfied (having more customer service calls relative to their total calls) tend to churn slightly more than those who are more satisfied.

Modelling

- To build a classifier to predict whether a customer will churn from SyriaTel, three models were used to aid in the prediction:
- **Logistic Regression Model**
- **Decision Tree Model**
- **Random Forest Model**
- Assessment of the all models was based on the accuracy score, precision, recall and F1 Score. As a telecommunications company with the aim of reducing customer churn, the recall score of each model was a more applicable metric as the aim to minimize the amount of false negative predictions. Maximizing the recall score will ensure that the model does not miss out on any potential customers who are likely to churn and leave the phone company.

Result and Findings

- Across all of the models listed, Logistic Regression Model was the best model and the most appropriate model given its mix of high recall score and accuracy score on training data and testing data. The model achieved a recall of **81%** for the negative class (False) and **78%** for the positive class (True) with an accuracy of **80.4%.** This indicates that the model is effective at capturing a substantial portion of the positive cases.



Recommendation

Based on the Exploratory Data Analysis above such as correlation, the important features highlighted in the models led to a recommendations for SyriaTel to improve churn which is:

1. Adopt Low Fixed Charge

2. Enhanced Feature Engineering on handling Customer Satisfaction

The above recommendation have been run through the original dataset and the best model to predict the new churn status to see if churn decreases as churn rate was 14.49% from the original dataset calculated above.

Conclusion

Leveraging our logistic regression model, we've gained insights into predicting customer churn and informing retention strategies for our telecom service. Firstly, by capping the customer_call_satisfaction feature at the 75th percentile (0.72), we observed a predicted overall churn rate of approximately 25.50%. This suggests that addressing factors contributing to lower customer satisfaction could substantially impact churn reduction.

Secondly, when capping total_minutes_overall at the 50th percentile (around 700 minutes) and implementing a charge of $71.05 per minute for usage beyond this threshold, our model predicts a churn value of 0, indicating that customers are unlikely to churn under these conditions. Implementing targeted strategies, such as fixed charges for excess usage, based on these insights can significantly contribute to customer retention efforts. However, continuous monitoring and adaptation of strategies remain crucial for maintaining customer satisfaction and loyalty amidst evolving preferences and market dynamics.

Next Step

Conduct further market research on pricing plans can be conducted across SyriaTel's main competitors. To further improve on lowering the churn rate across all customers, we would need to collect more data from users. One way to implement a next step is to conduct a customer satisfaction survey to pinpoint more specifically other areas that can be improved upon in order to provide better customer experiences for all SyriaTel users.

THANK YOU

THE WAY TO GET STARTED IS TO QUIT TALKING AND BEGIN DOING.

Walt Disney