

BY : LEE NDUNG'U

Customer Churn Prediction in Telecommunications

SYRIATEL

Introduction

Customer churn is a critical concern for businesses in the telecommunications sector.

Predicting churn and implementing effective retention strategies are vital for maintaining market share and profitability.

This project aims to leverage predictive analytics and machine learning to identify churn risks and develop targeted retention initiatives.

Problem Statement

The problem to be addressed in this project is customer churn within Syriatel, a major telecommunications company. Customer churn poses a significant challenge to Syriatel's revenue and long-term viability.

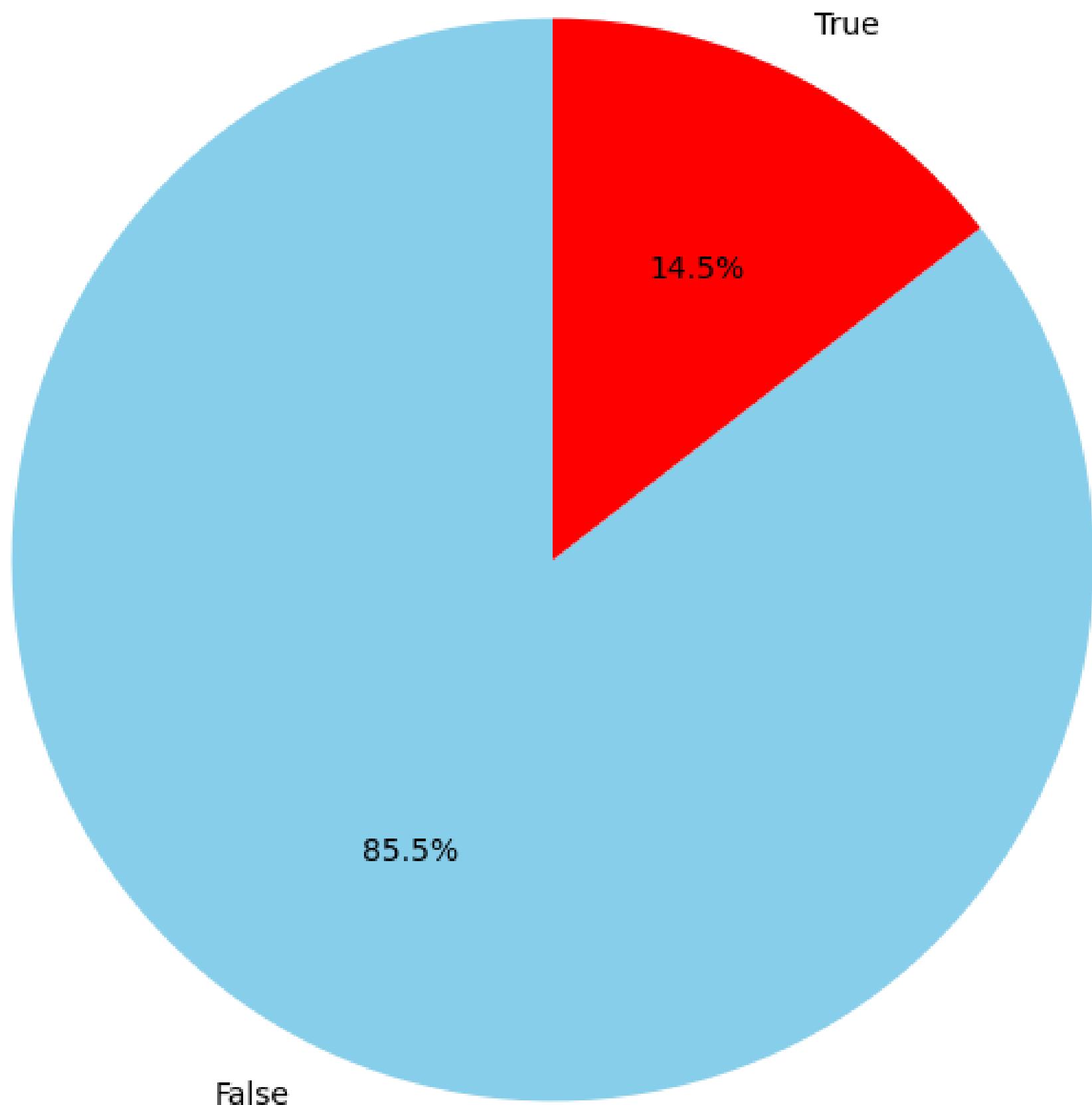
Syriatel's income and sustainability are directly impacted by customer churn which may impact revenue, market share, and the long-term sustainability of the business.

This impact is what has The stakeholders involved in this project to each have a vested interest in reducing churn rates and enhancing customer retention strategies.

Objectives

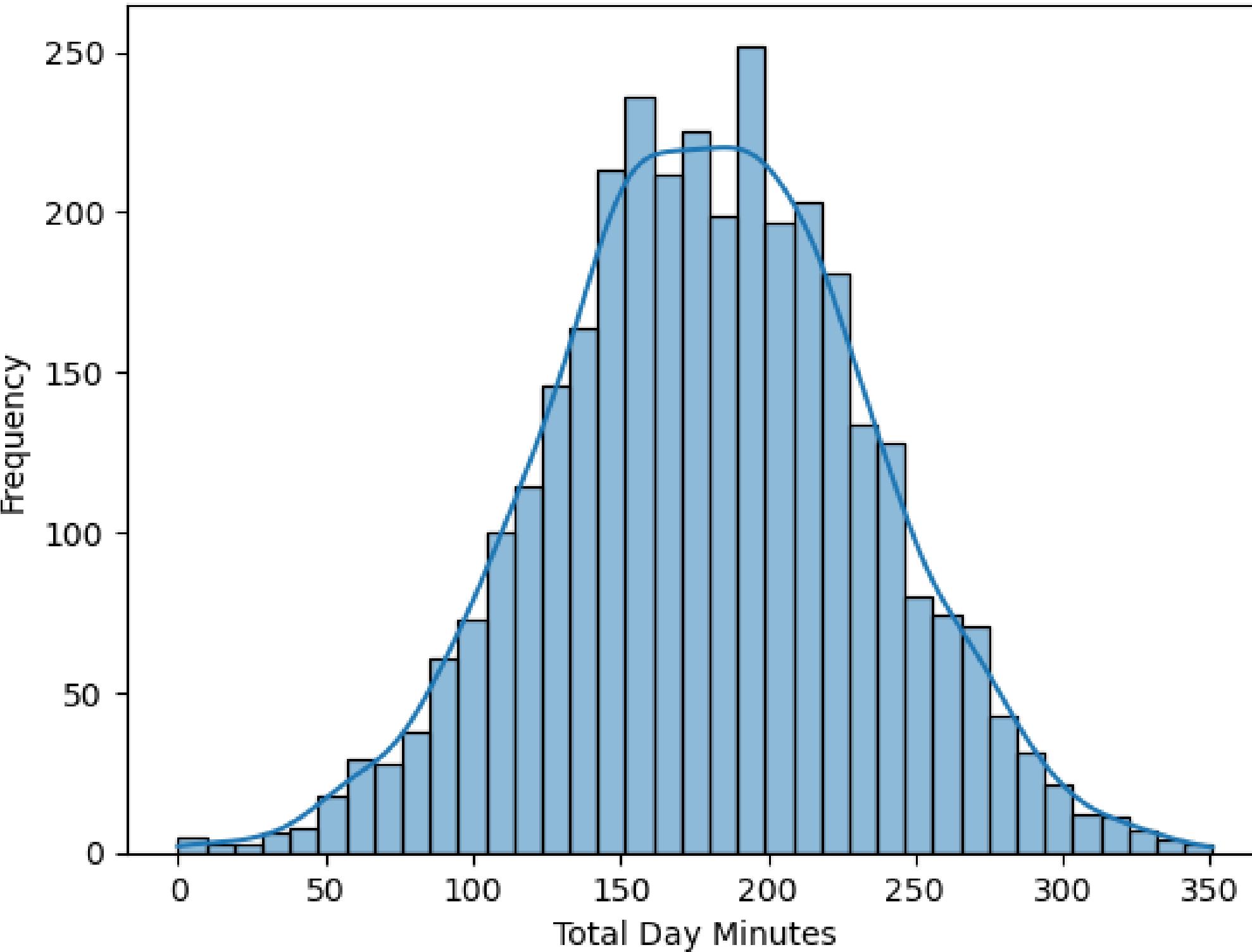
- Churn Prediction
- Retention Strategies
- Enhanced Customer Experience
- Business Growth

Churn Distribution Representation



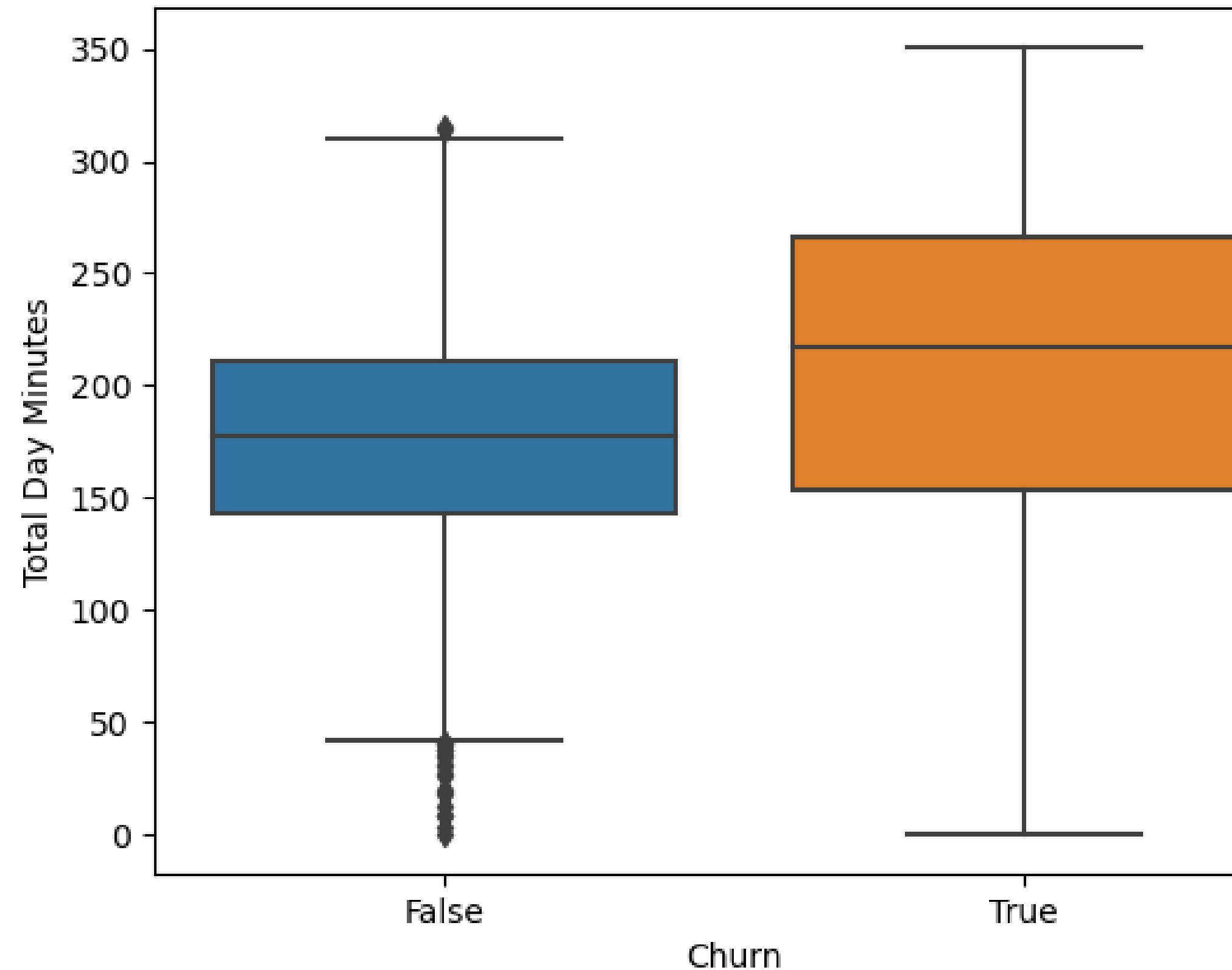
THIS PLOT VISUALIZES THE DISTRIBUTION OF CHURNED VERSUS NON-CHURNED CUSTOMERS IN THE DATASET; PROVIDING INSIGHTS INTO THE PROPORTION OF CUSTOMERS WHO HAVE CHURNED COMPARED TO THOSE WHO HAVE NOT.

Distribution of Total Day Minutes



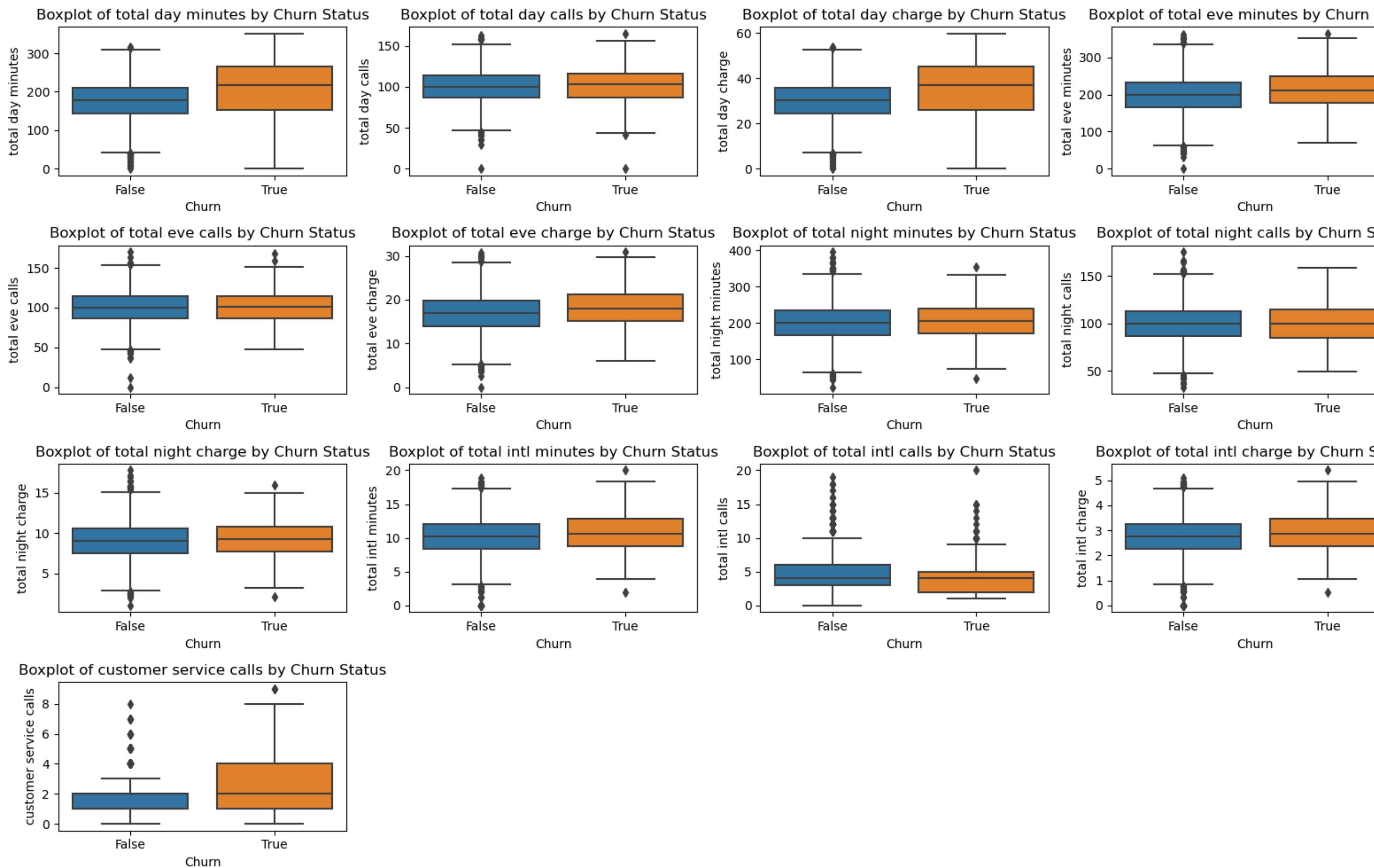
THIS HISTOGRAM SHOWS THE DISTRIBUTION OF TOTAL DAY MINUTES FOR ALL CUSTOMERS IN THE DATASET. IT HELPS VISUALIZE THE SPREAD AND CONCENTRATION OF TOTAL DAY MINUTES AMONG CUSTOMERS. THE X-AXIS REPRESENTS THE RANGE OF TOTAL DAY MINUTES, AND THE Y-AXIS REPRESENTS THE FREQUENCY OF OCCURRENCE.

Boxplot of Total Day Minutes by Churn Status



THIS HISTOGRAM SHOWS THE DISTRIBUTION OF TOTAL DAY MINUTES FOR ALL CUSTOMERS IN THE DATASET. IT HELPS VISUALIZE THE SPREAD AND CONCENTRATION OF TOTAL DAY MINUTES AMONG CUSTOMERS. THE X-AXIS REPRESENTS THE RANGE OF TOTAL DAY MINUTES, AND THE Y-AXIS REPRESENTS THE FREQUENCY OF OCCURRENCE.

THIS SERIES OF BOXPLOTS VISUALIZES THE DISTRIBUTION OF VARIOUS NUMERICAL FEATURES IN THE DATASET.
EACH BOXPLOT COMPARES THE DISTRIBUTION OF A SPECIFIC FEATURE BETWEEN CHURNED AND NON-CHURNED CUSTOMERS.
IT HELPS IDENTIFY POTENTIAL DIFFERENCES OR RELATIONSHIPS BETWEEN NUMERICAL FEATURES AND CHURN STATUS.

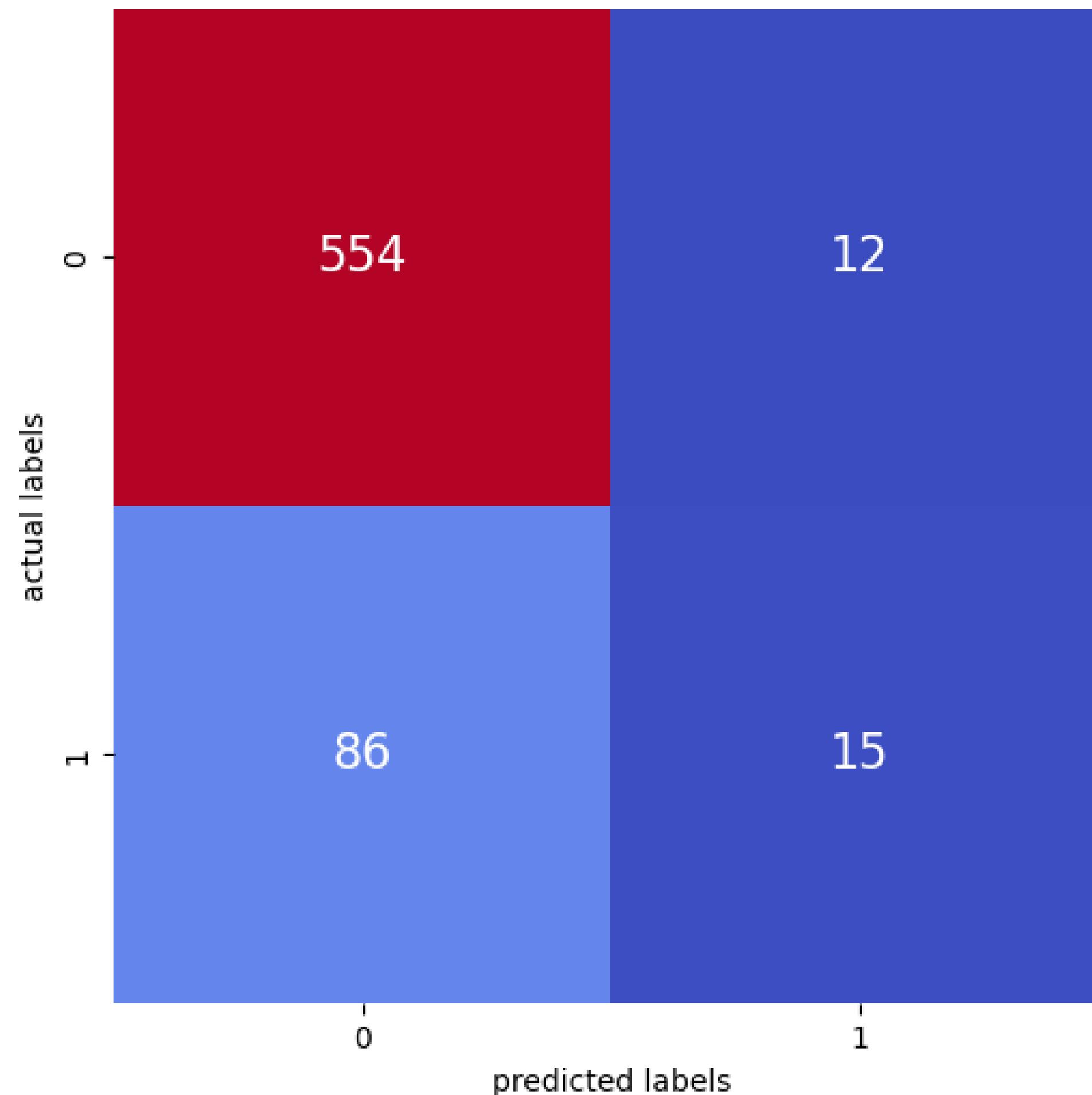


Model Development

THE MODELS USED INCLUDE:

- Random Forest Classifier: Random Forest stems from its prowess in handling a diverse range of features, both categorical and numerical.
- Logistic Regression Classifier: Logistic Regression is a foundational binary classification approach, optimal for scenarios where the relationship between features and the target is linear.
- Decision Tree: Decision Trees are intuitive models that excel in identifying patterns in customer behavior.

Confusion matrix for logistic regression

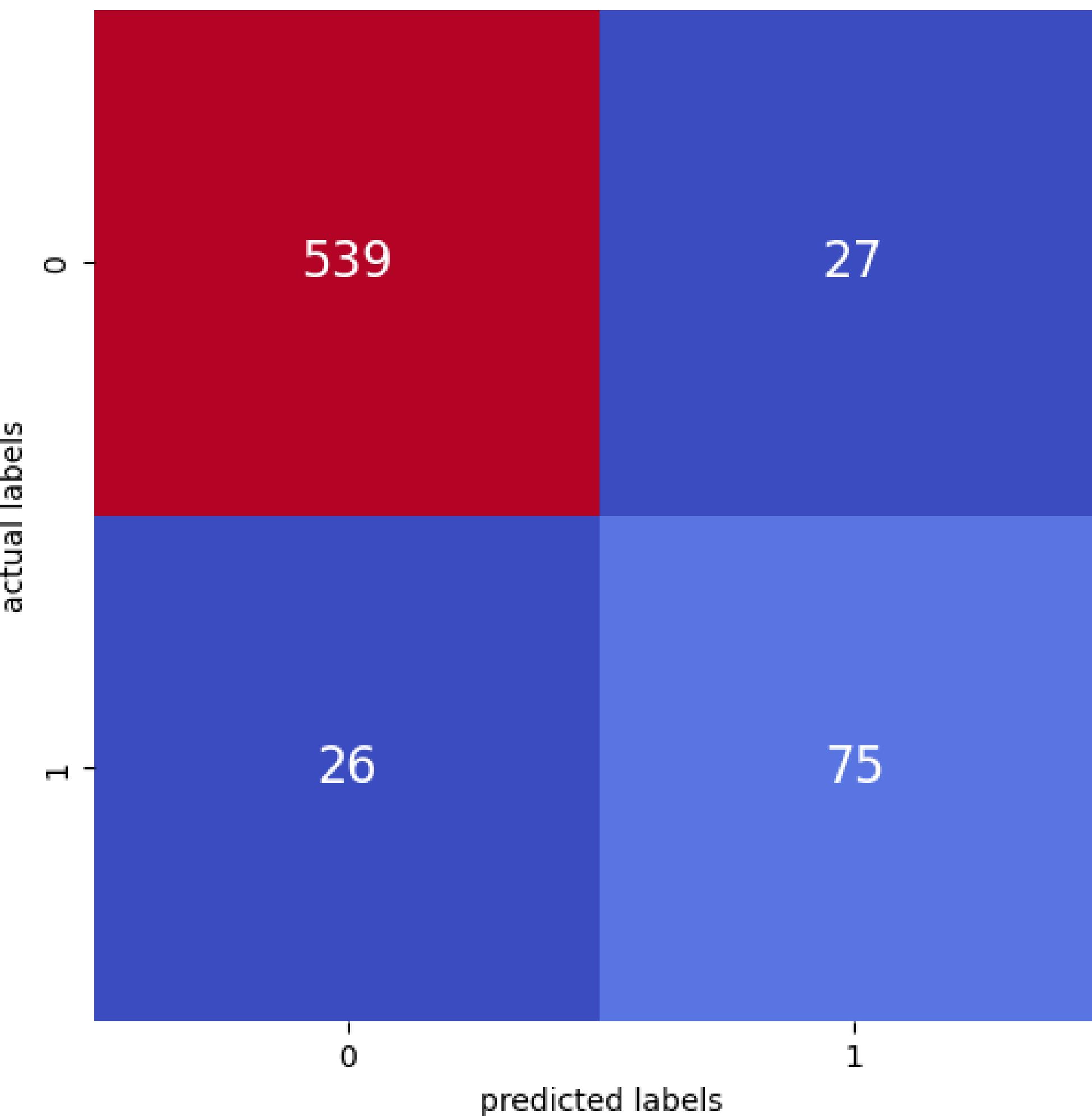


LOGISTIC REGRESSION MODEL

This model is exhibiting signs of overfitting; demonstrating strong performance on the training data but failing to generalize well to the test data

Accuracy of 0.85

Confusion matrix for Decision Trees

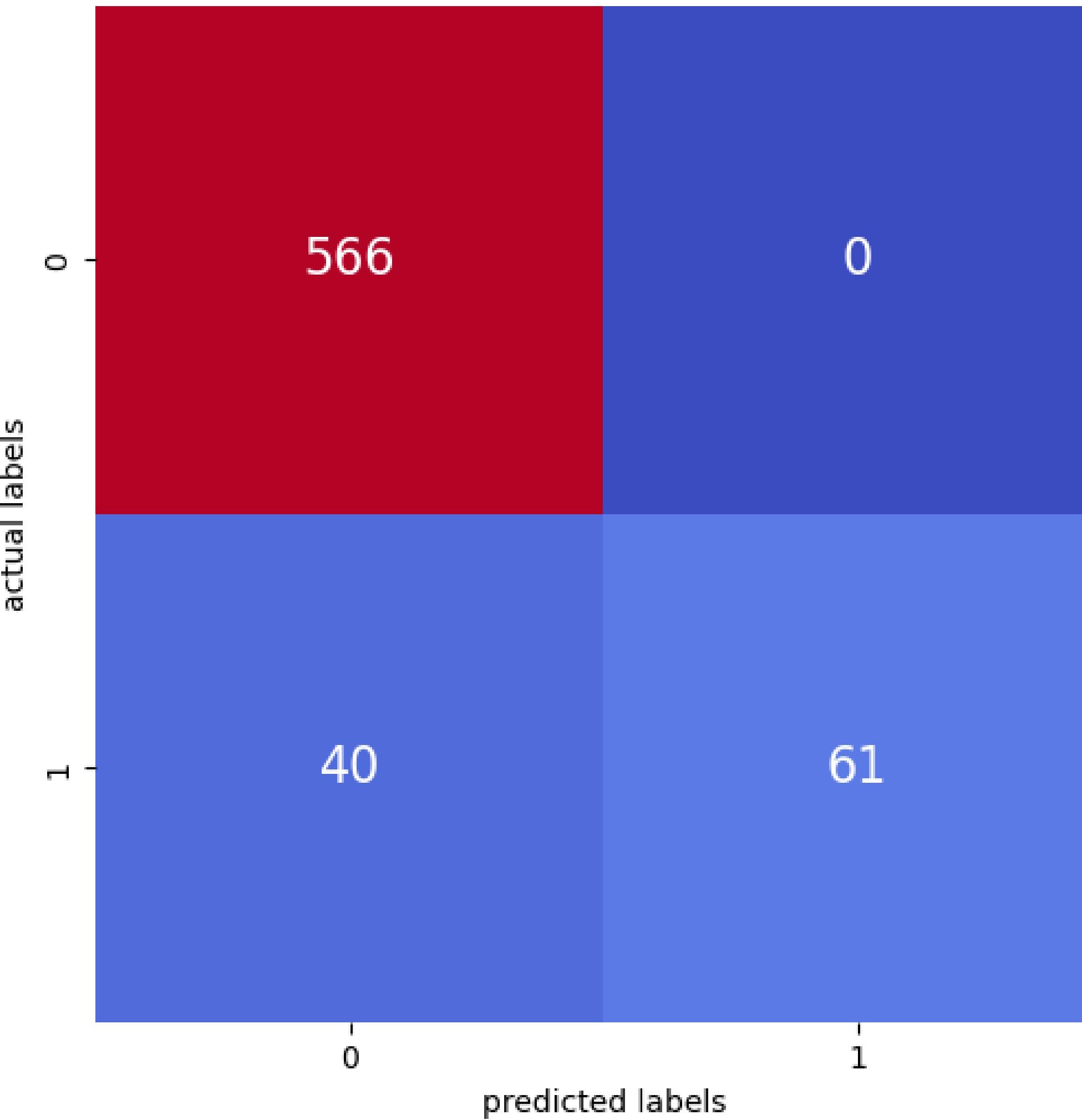


DECISION TREES MODEL

The decision tree model outperforms the logistic regression model, though still necessary to note it exhibits signs of overfitting.

Accuracy score of 0.92

Confusion matrix for Random Forest



RANDOM FOREST MODEL

The random forest model demonstrates a higher performance than the logistic regression model and the decision tree model, but still exhibits overfitting.

Accuracy score of 0.94

Conclusion



In this customer churn prediction project for Syriatel, we aimed to develop predictive models to identify customers at risk of churning and provide insights for targeted retention strategies. Three machine learning models, logistic regression, decision trees, and random forest, were trained and evaluated on the dataset consisting of various customer attributes and usage metrics.

The decision tree and random forest models outperformed logistic regression in terms of accuracy, precision, recall, and F1-score. However, random forest achieved the highest accuracy of approximately 94%, making it the most suitable model for predicting customer churn in this scenario. Furthermore, hyperparameter tuning using grid search cross-validation helped optimize the random forest model's performance, resulting in improved accuracy and model generalization.

Recommendations

Based on the findings of this project, the following recommendations are proposed for Syriatel:

1. Implement Predictive Model: Deploy the tuned random forest model into production to make real-time predictions on new customer data. This will enable Syriatel to identify customers at high risk of churn and take proactive retention measures.
2. Targeted Retention Strategies: Utilize insights from the predictive model to design targeted retention initiatives aimed at reducing churn rates. Tailor marketing campaigns and offers to address specific customer needs and preferences identified by the model.
3. Enhance Customer Experience: Leverage predictive analytics to provide personalized services and proactive support to customers identified as at-risk of churn. Address customer concerns promptly and offer solutions to improve overall satisfaction and loyalty.
4. Continuous Monitoring and Improvement: Regularly monitor model performance and customer churn trends to adapt retention strategies accordingly. Continuously update the model with new data to enhance its accuracy and effectiveness over time.

**The network of the future
is already being built as
we speak. 5G is an
important part of that.**

— Eben Albertyn

THANKYOU