PHASE 3 PROJECT

Syriatel Customer Churn Classification Project

BUSINESS UNDERSTANDING

- Syriatel's major challenge is losing customers leading to revenue loss, negative brand perception and increased acquisition costs. It aims at reducing customer churn .
- The management would like to understand the factors that drive churn and create a model that predict it accurately hence helping Syriatel take targeted actions to retain valuable customers.

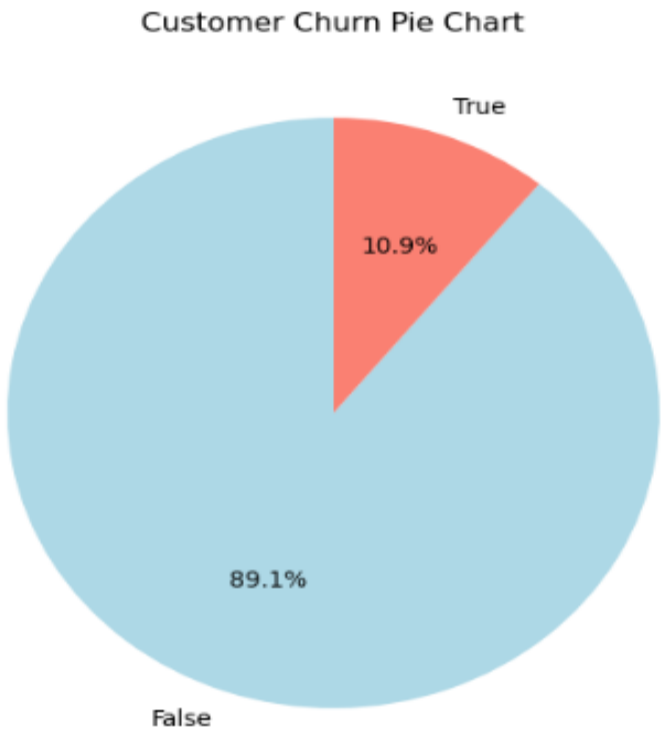
BUSINESS OBJECTIVES

- Identify the key factors driving customer churn.
- Develop a predictive model to classify customers at risk of leaving.
- Enable Syriatel to take proactive actions to improve customer retention.
- Enhance customer satisfaction and business sustainability.

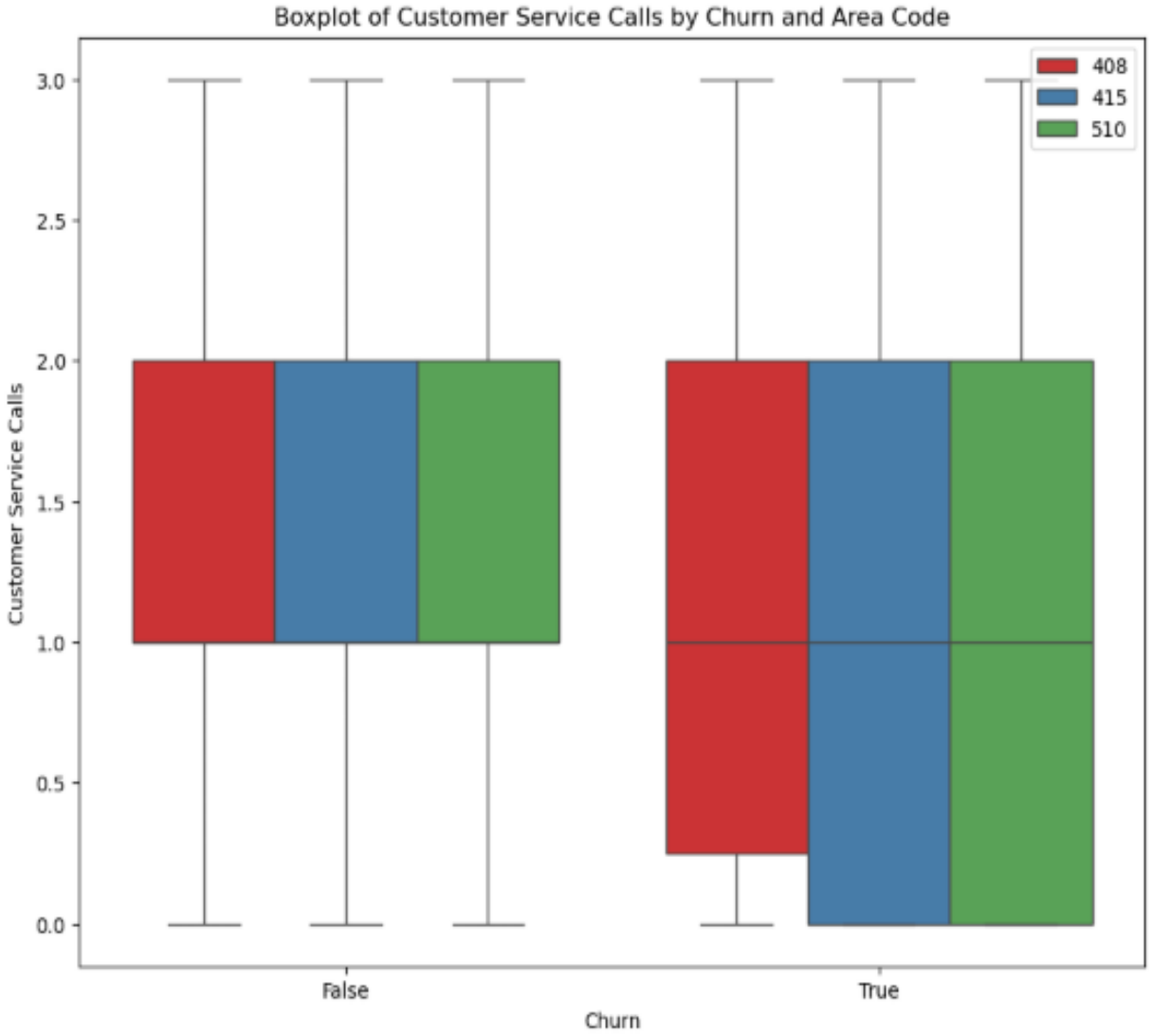
DATA PREPARATION

- Data cleaning
- Data Analysis (Univariate Analysis , Bivariate Analysis and Multivariate Analysis)Handled missing values and removed duplicate records.
- Standardized numerical features using standard scaler.
- Encoding categorical variables using Label Encoding.
- Splitting the dataset into training (80%) and testing (20%) sets

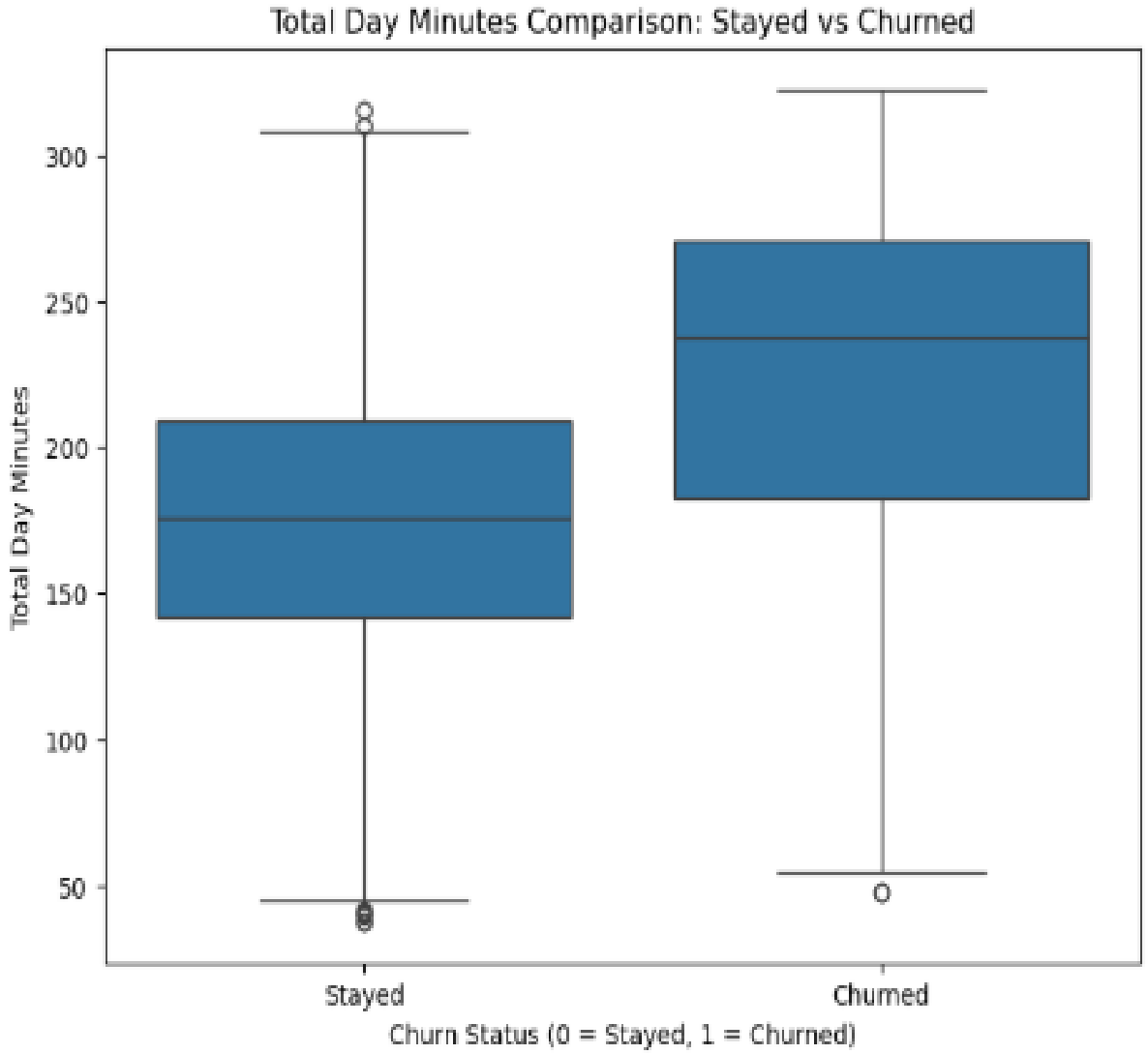
EXPLORATORY DATA ANALYSIS



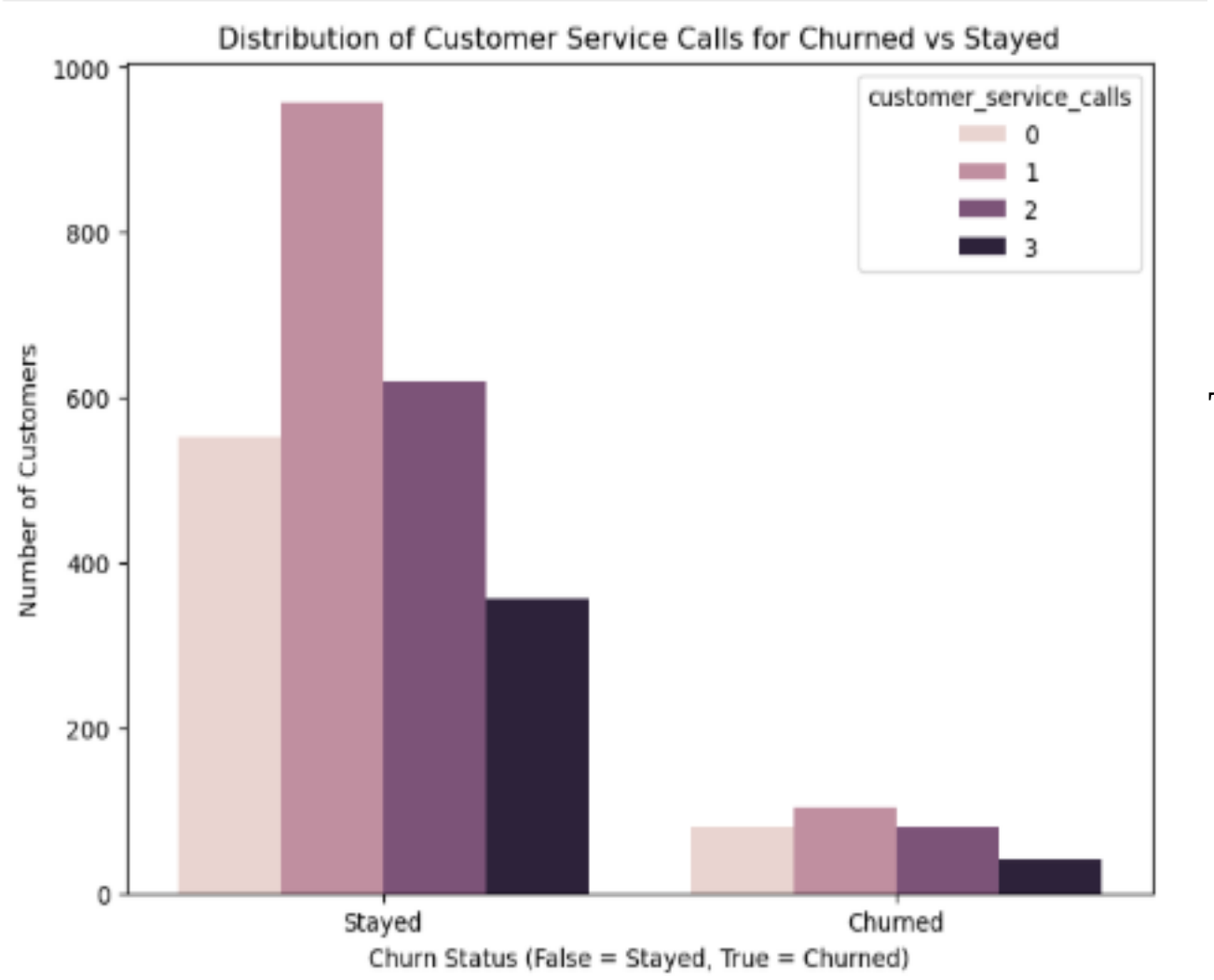
This means that 90% of the customers in the dataset did not churn (False), while 10% did churn (True).



The boxplot suggests that a significant portion of customers who terminated their contracts belong to area codes 415 and 510. Additionally, there are numerous outliers present in the dataset.



The box plot above shows that customers who spent more time on the phone during the day are more likely to churn.

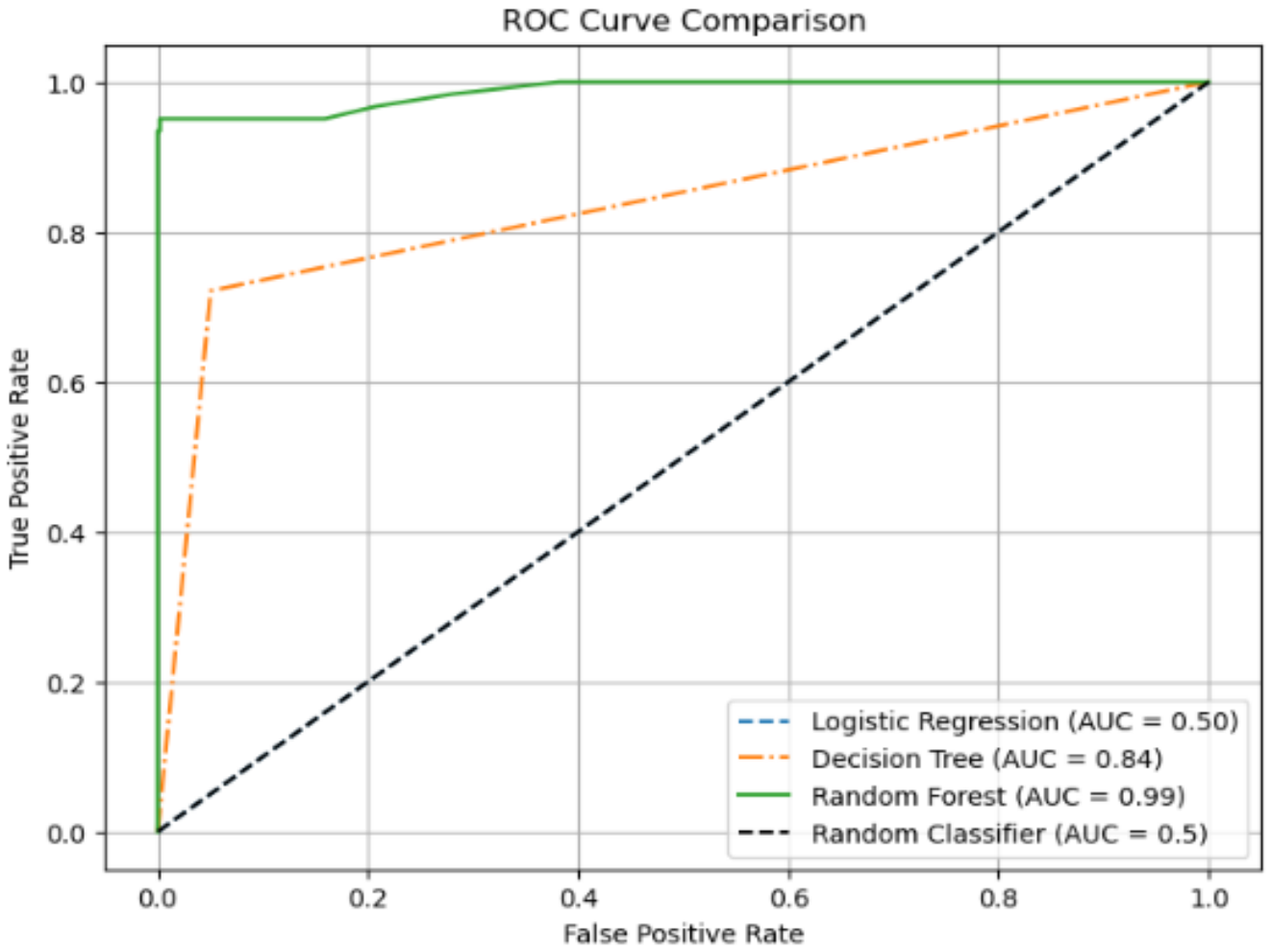


The bar graph shows that most churners called least times unlike those who stayed.

MODELING AND EVALUATION

- The dataset was analyzing presents a binary classification challenge aimed at predicting churn, i.e., whether a customer will leave or not.
- My approach involves utilizing a basic logistic regression model, with evaluation based on the recall score. Success criteria are set at achieving a recall score of 75% or higher. I employed the algorithms below;

✓ Logistic Regression
✓ Decision Tree
✓ Random Forest



- Logistic Regression (0.50) performs poorly, suggesting poor feature selection.
- The Random Forest model (AUC = 0.99) performs strongly and is likely more stable but maybe overfitting.
- Logistic Regression (AUC = 0.50) has the lowest AUC, suggesting it may not be capturing complex relationships in the data effectively.

CONCLUSION

Based on the findings, the business conclusion is as follows:

- Recall:This strategy aimed to minimize the misclassification of churners as non-churners, ensuring that the model effectively identifies customers at risk of leaving.
- Optimal Model: Among the explored models, the Random Forest classifier stood out as the top performer,
- Continuous Improvement: Predicting customer churn is an ongoing endeavor, necessitating constant refinement of the model.

RECOMMENDATIONS

- Offer discounts or loyalty programs for high-risk customers.
- Improve customer support to reduce dissatisfaction.
- Encourage long-term contracts by offering incentives.
- Monitor high-churn segments and take proactive retention measures.

THANK YOU