CUSTOMER CHURN PREDICTION FOR SYRIATEL

PREPARED BY:
Emmanuel Chol for Data Science Project

Moringa School and Flatiron student

BUSINESS UNDERSTANDING

SyriaTel is a mobile telecommunications company that has built an unquestionable reputation in the industry over the years.

The company wish to mitigate the effects that would come with churning customers and maximise on profits by coming up with a machine learning model that predict whether a customer stays or churn. In addition, the company also needs to identify the factors making customers churn SyriaTel.

BUSINESS OBJECTIVES

Builds a Machine learning model to predict if a customer stays or churns

Understand the factors contributing to churn

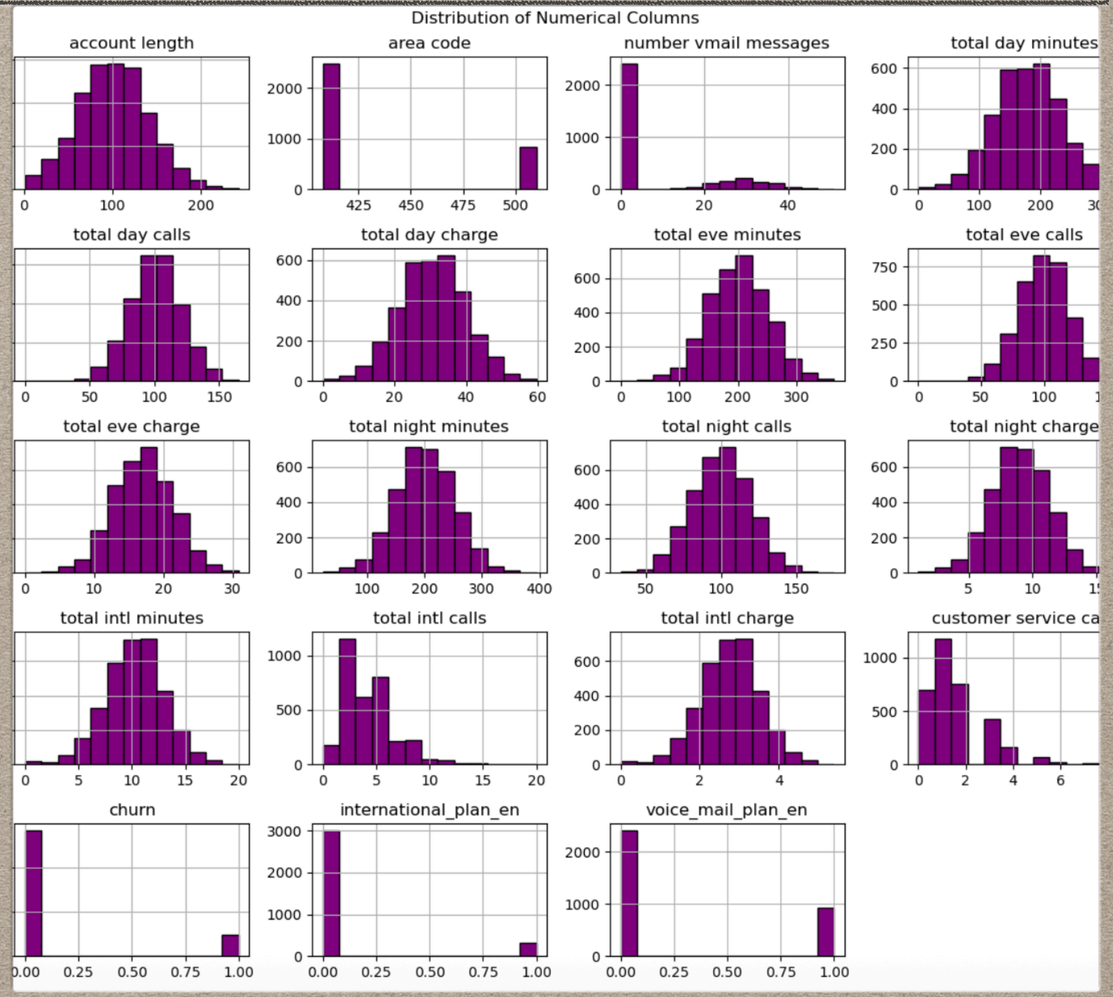
DATA UNDERSTANDING

The data for this project was downloaded from canvas as a zipped file.
It contains 3,333 records and 21 rows.

DESCRIPTIVE ANALYTICS

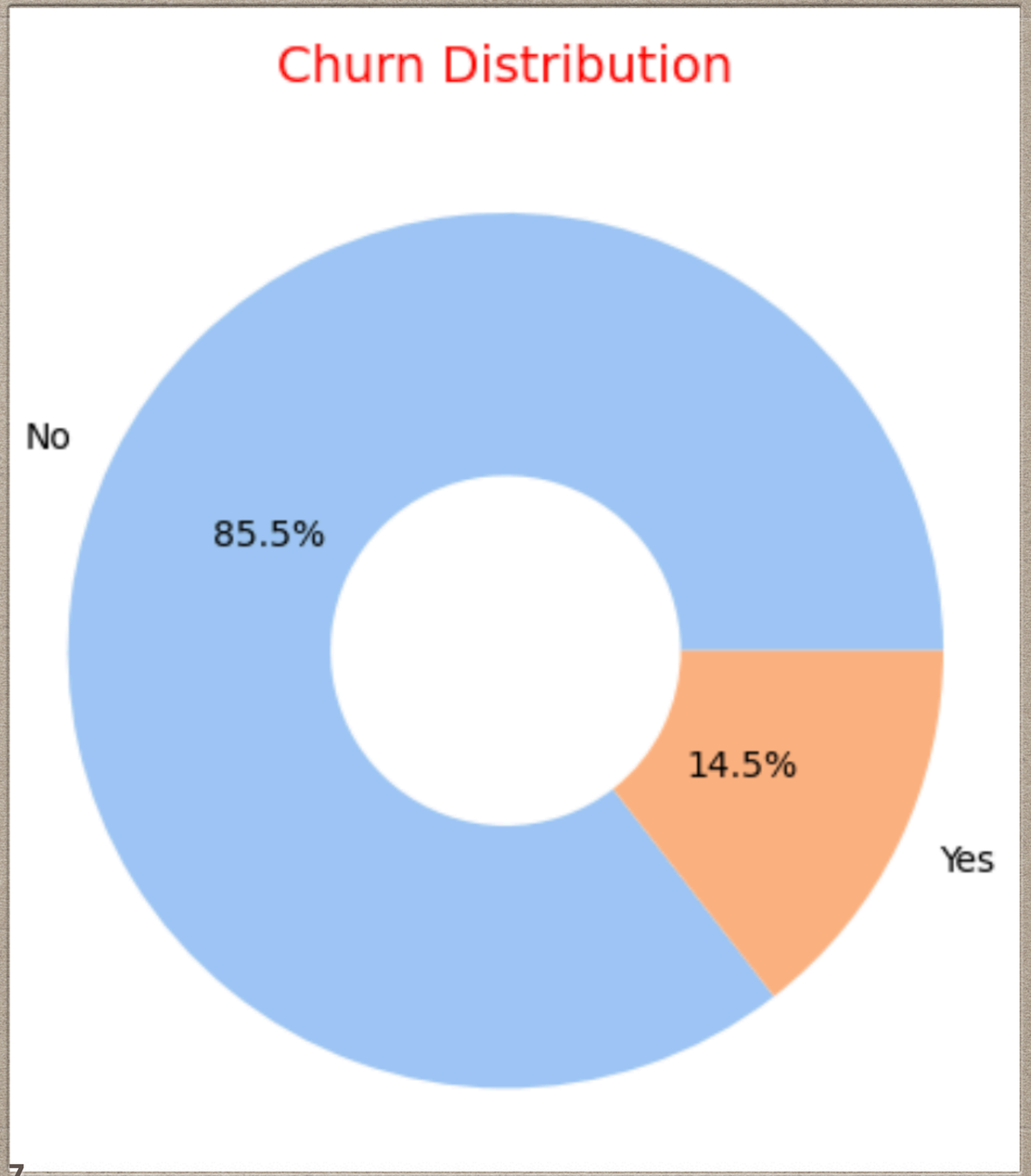
Univariate Analysis of the DISTRIBUTION OF VARIOUS FEATURES

- Most of the features have Gaussian-like Distributions, for example total day minutes, total eve minutes, total night minutes, total intl minutes, and account length.
- Others are Skewed e.g, number vmail messages

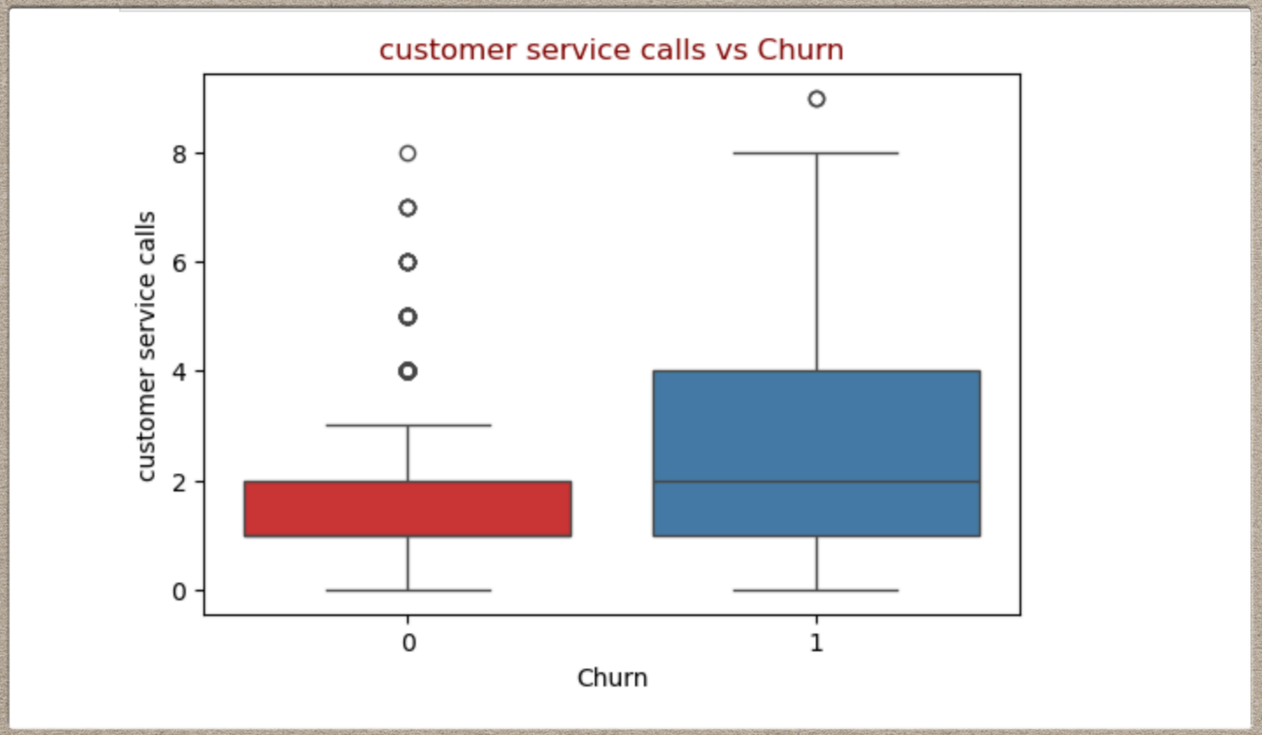


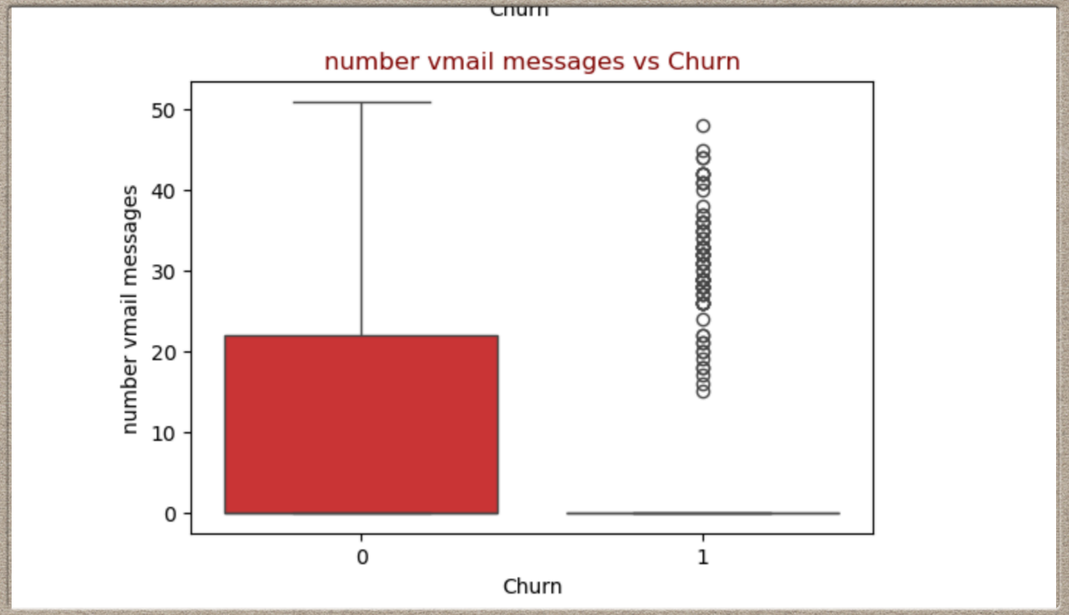
UNIVARIATE ANALYSIS OF THE TARGET

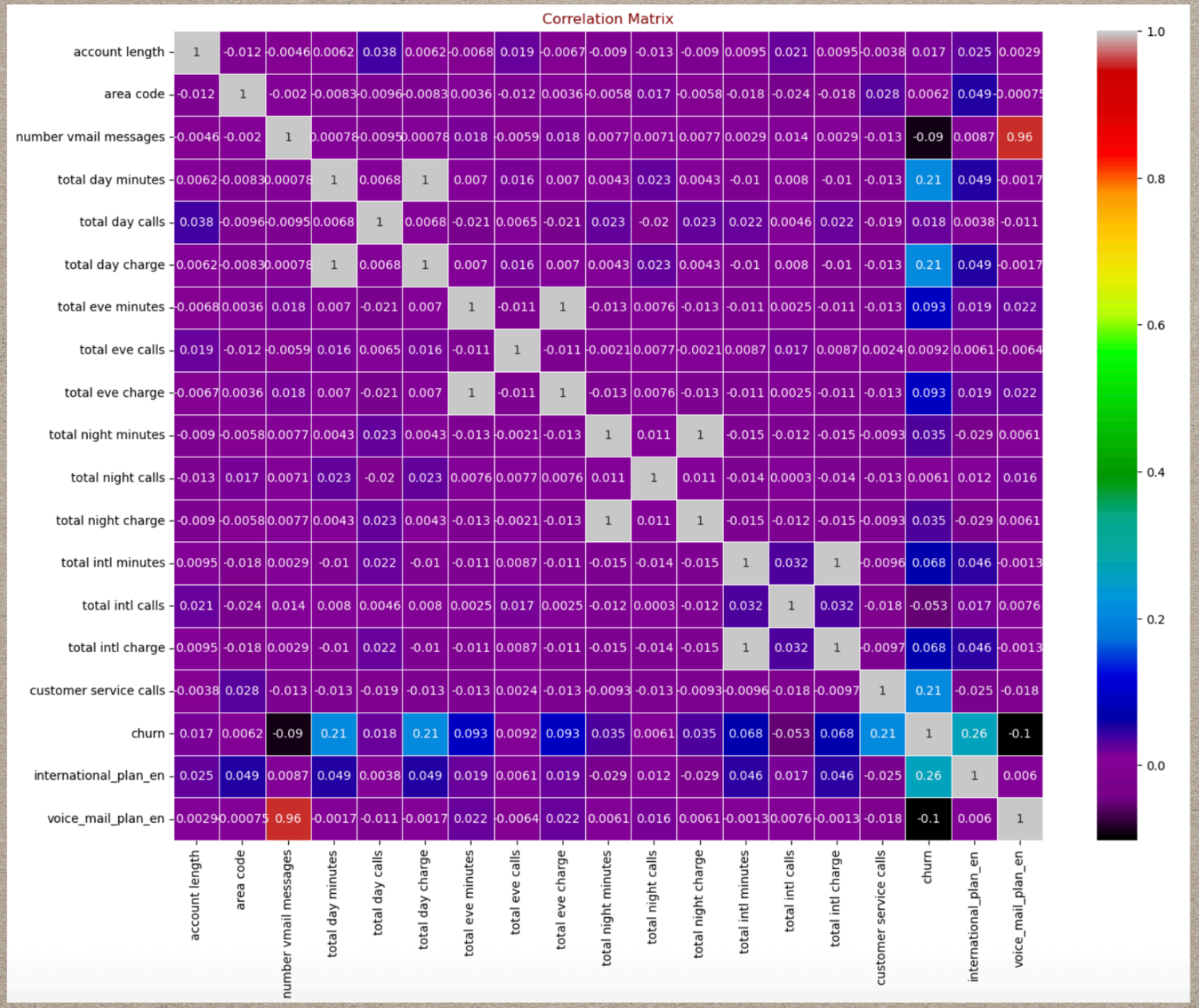
The data is imbalance with "No" representing the larger chunk(85.5%), while "Yes"represents 14.5%



Bivariate Analysis of Churn vs predictors







Correlation Metric of all features

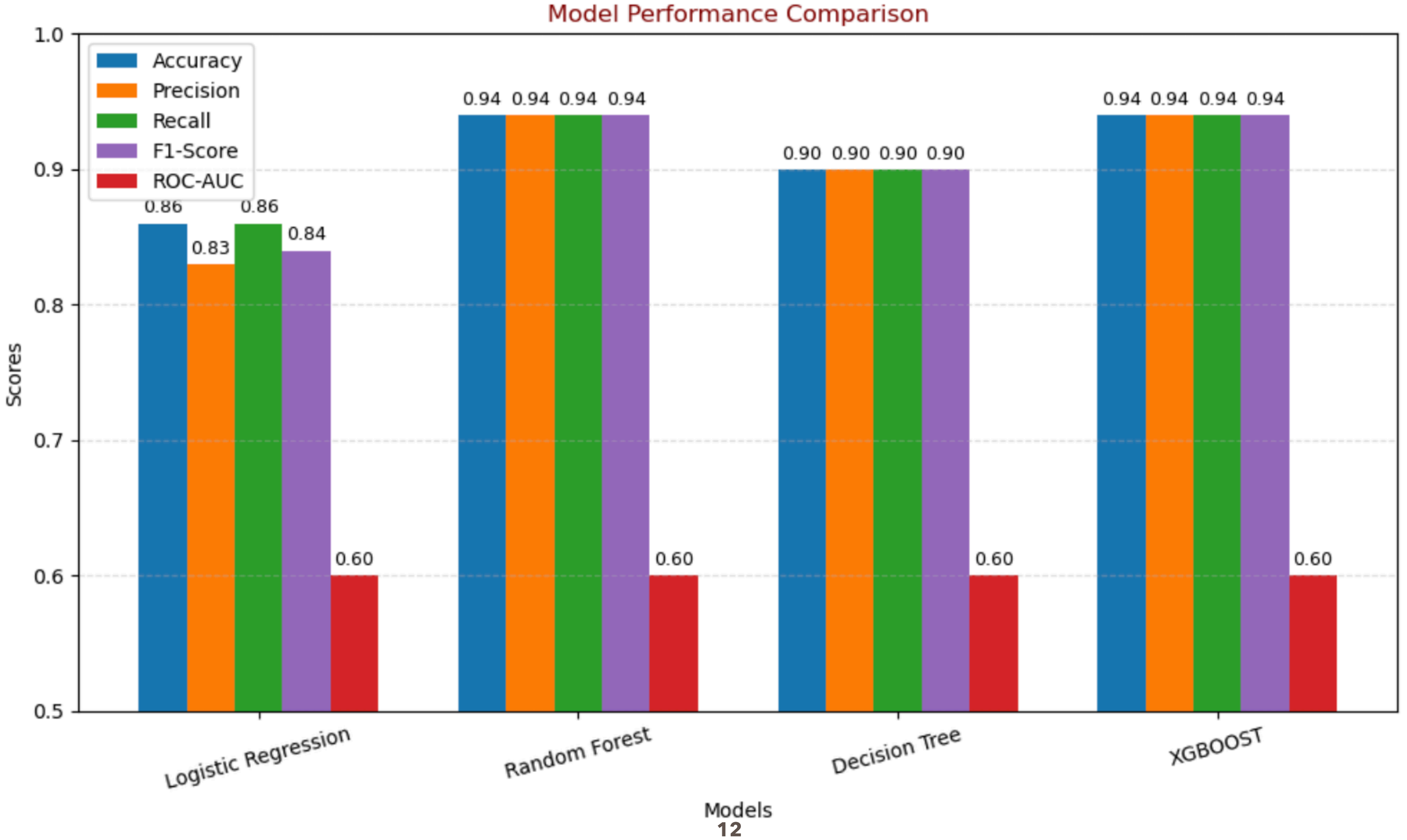
Some features such as international plan, number of voicemail messages, and customer service call have strong correlation with churn.

MODELING APPROACHES

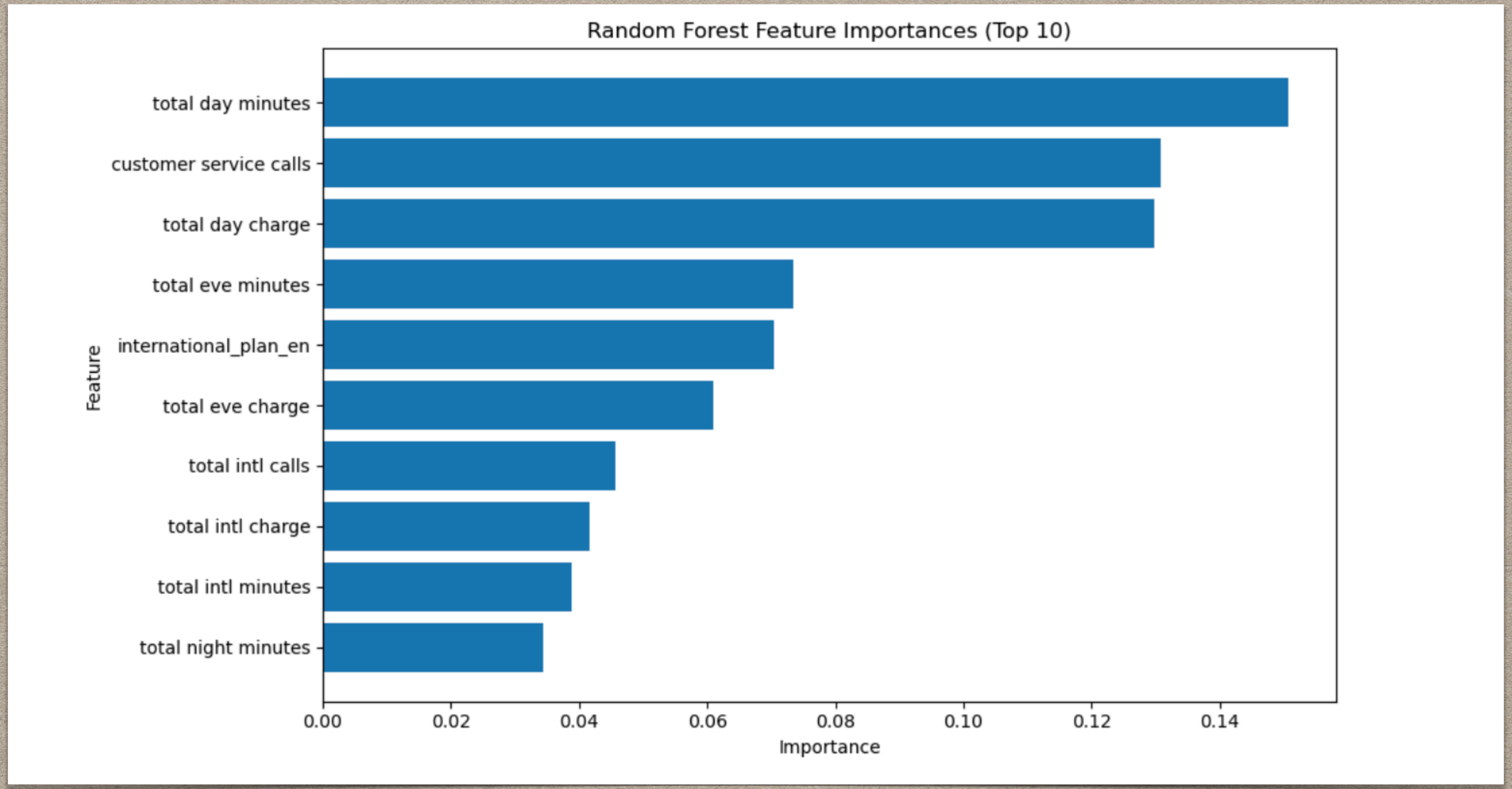
The following machine learning models were used:

- Logistic Regression
- Random Forest
- XGBoost
- Decision Tree

Of the four models, Random Forest out perform the others in most of the metric. Random Forests only out perform XGBOOST in only one metric i.e the precision.


Model Performance Comparison
Accuracy
Precision
Recall
F1-Score
ROC-AUC
Scores
1.0
0.9
0.8
0.7
0.6
0.5
0.86
0.83
0.86
0.84
0.60
0.94 0.94 0.94 0.94
0.60
0.90 0.90 0.90 0.90
0.60
0.94 0.94 0.94 0.94
0.60
Logistic Regression
Random Forest
Decision Tree
XGBOOST
Models

Feature Importance of The Best Performing Model - Random Forest



Conclusions

- High churns are associated with day time usages.
- Customers with more customer serive calls high chances of leaving/ churning.
- Random Forest is the best model for handling class imbalances

Recommendations

- Improve customer service
- Construe policies to hold customers with high chances of churning
- Use Random Forest for SyriaTel data modeling.

Thanks