SyriaTel Customer Churn Prediction

CODE TRIBE_3.0

19TH JULY 2023

Outline

- Project Overview
- Business Problem
- Data Understanding
- Data Cleaning
- Data Analysis
- Data Modelling
- Conclusion
- Recommendations

Project Overview

- This project aims at identifying customer churn patterns and build a customer churn prediction model to help **SyriaTel Telecommunication** company take proactive measures to retain at-risk customers.

- To address this issue, SyriaTel have requested **CodeTribe3** researchers to build a churn prediction system that can identify customers likely to churn in the near future.

Business Problem

The Project seeks to investigate :

- Any predictable or discernible patterns in customer behaviors that can aid in identification of customers who are likely to churn from SyriaTel company, enabling SyriaTel to implement proactive retention strategies and reduce churn rate.

Data Understanding

The SyriaTel data contains information about customer attributes, call usage, charges and customer service interactions with the churn column acting as our target variable

The dataset contains;

- 3333 rows
- 20 columns

The data has 4 data types namely;

- Int64
- Object
- Bool
- Float64

Columns included;

- total day min
- total day charge
- total eve min
- total intl charge
- phone number
- voice mail plan etc.

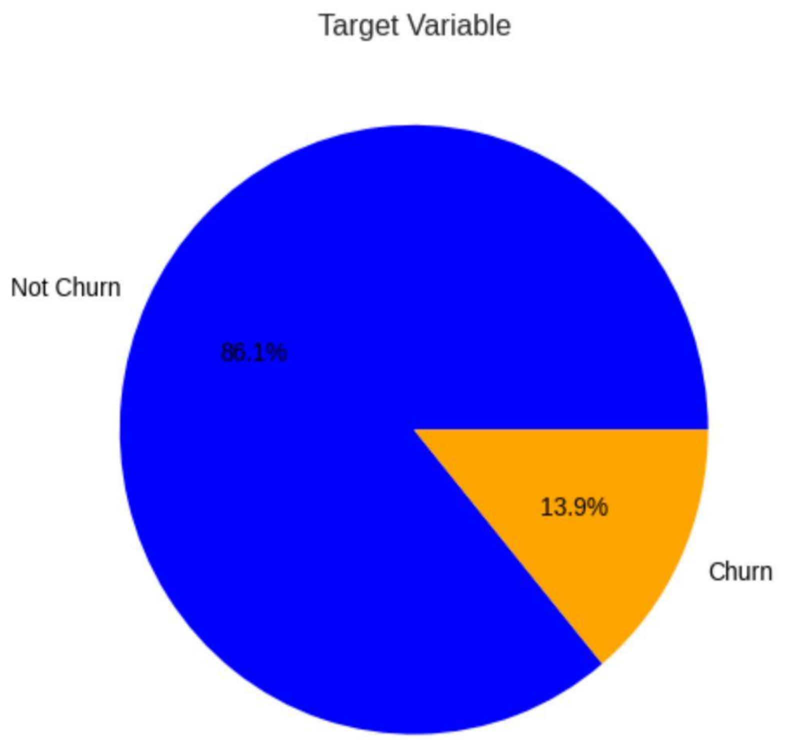
Data Cleaning

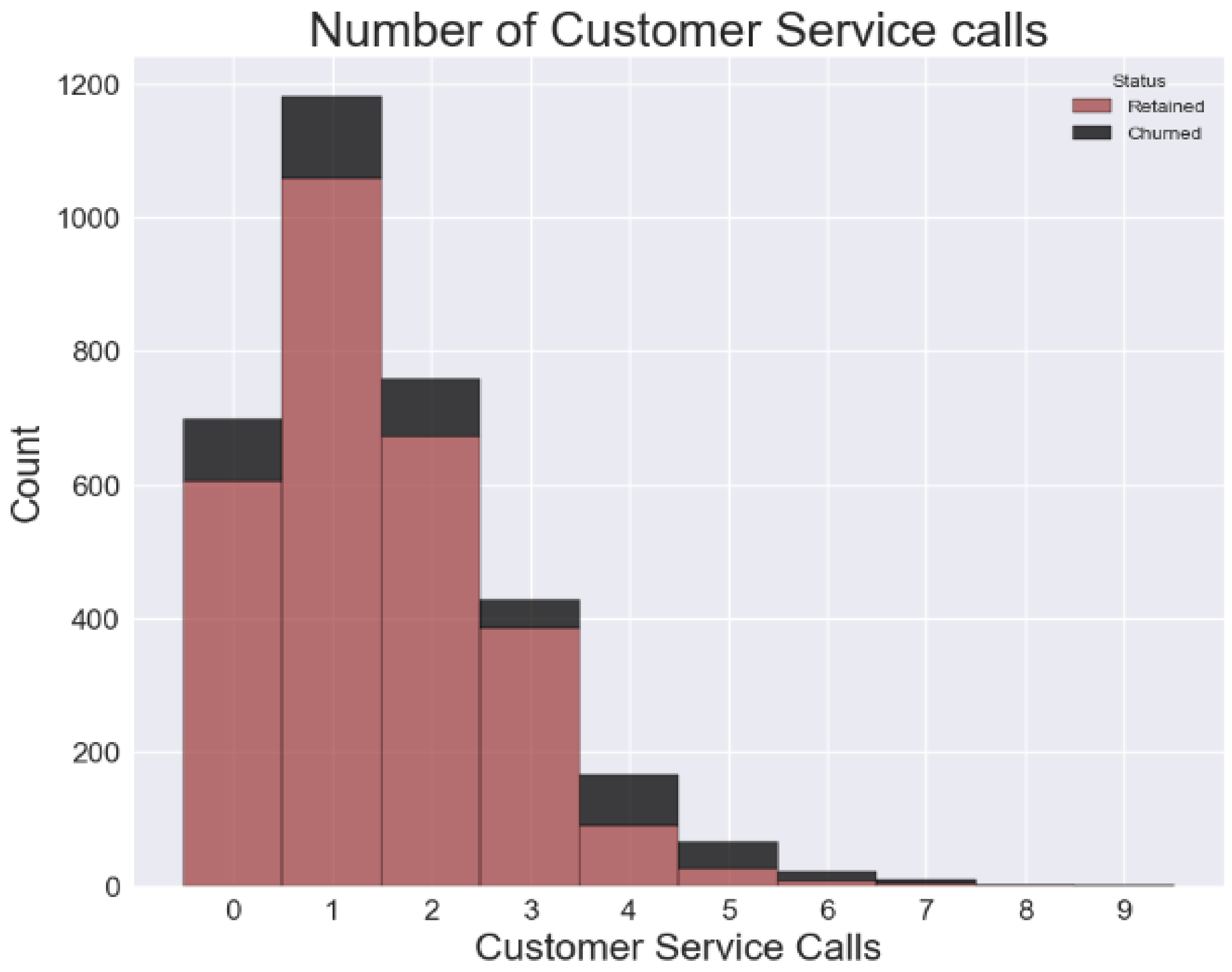
The data preparation included the following steps:

- Dropping unnecessary columns such as "phone number"
- Identifying and handling missing values if any
- Check for duplicate entries
- Deal with outliers

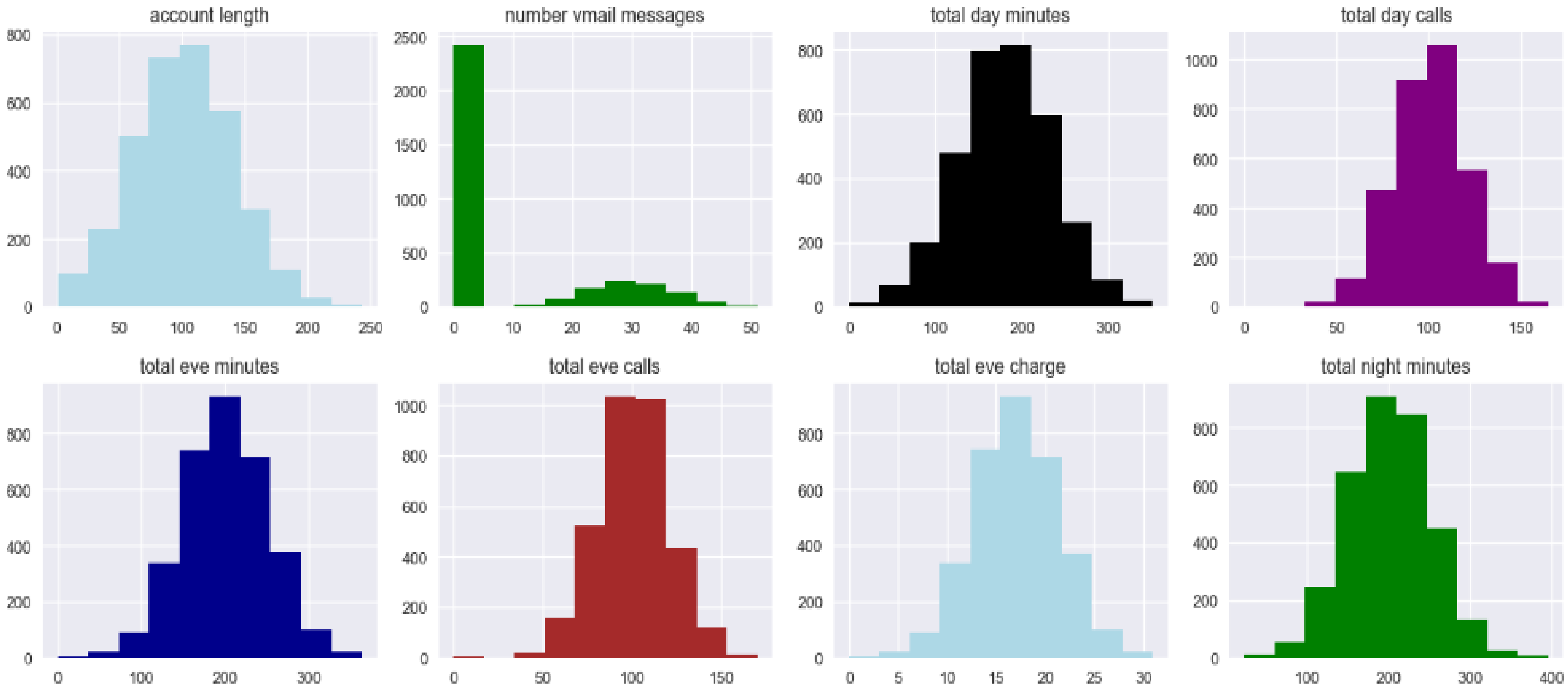
Exploratory Data Analysis

❖ **Univariate Analysis**

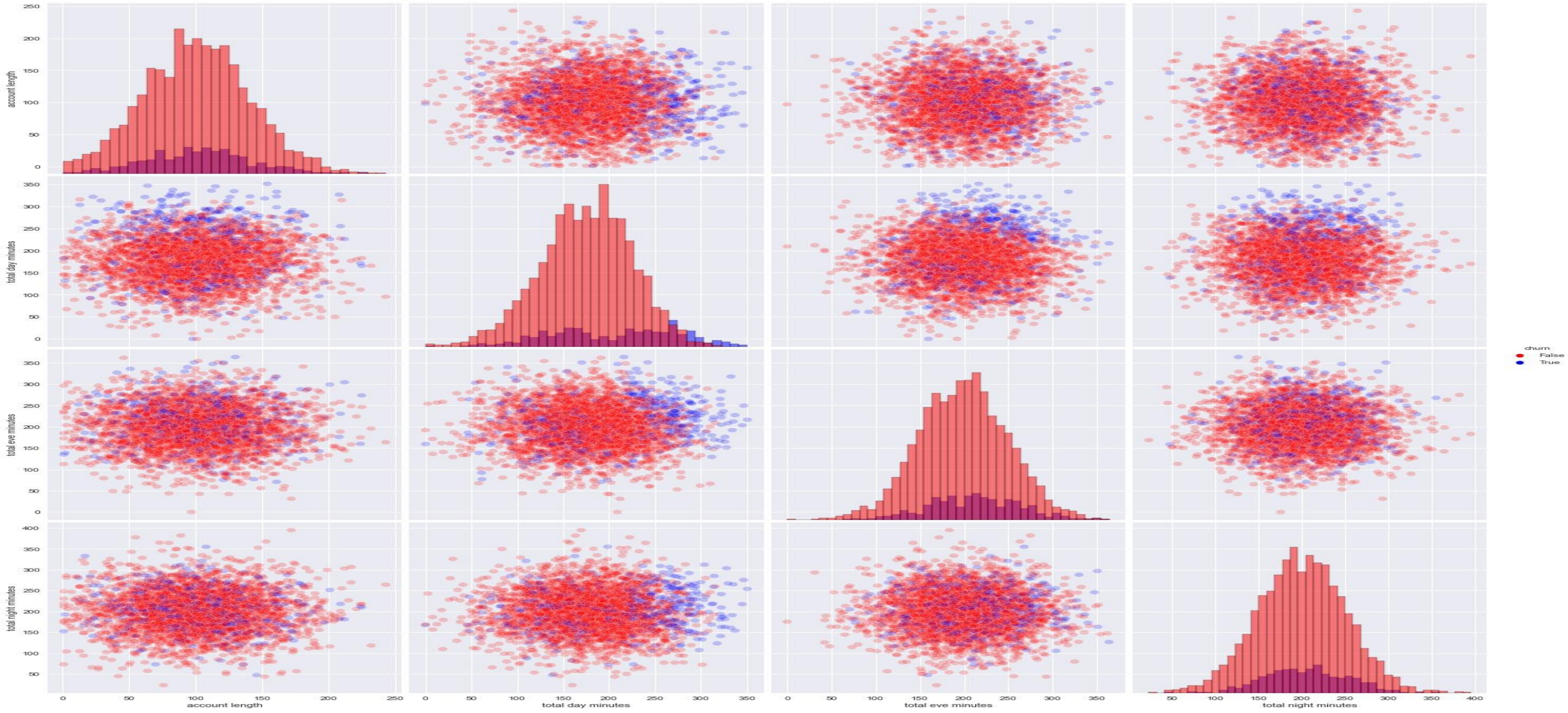




❖ Univariate Analysis



❖ Bivariate & Multivariate Analysis



Data Modelling

Data Pre-preprocessing

- Label Encoding
- One Hot encoding
- Scaling of the dataset using StandardScaler.
- Handling class imbalance using SMOTE.

Vanilla Model : Decision Tree Classifier

```
[[532  34]
 [ 21  80]]
****************************************
              precision    recall  f1-score   support

           0       0.96      0.94      0.95       566
           1       0.70      0.79      0.74       101

    accuracy                           0.92       667
   macro avg       0.83      0.87      0.85       667
weighted avg       0.92      0.92      0.92       667
```

Model 2 : Random Forest Classifier

```
confusion_matrix for  Random Forest
 [[546  20]
 [ 28  73]]
************************************************
DecisionTreeClassifier Random Forest
              precision    recall  f1-score   support

           0       0.95      0.96      0.96       566
           1       0.78      0.72      0.75       101

    accuracy                           0.93       667
   macro avg       0.87      0.84      0.86       667
weighted avg       0.93      0.93      0.93       667
```

Model 3 : K - Nearest Neighbors Classifier

```
confusion_matrix for KNN
 [[424 142]
 [ 38  63]]
KNN classification_report
              precision    recall  f1-score   support

           0       0.92      0.75      0.82       566
           1       0.31      0.62      0.41       101

    accuracy                           0.73       667
   macro avg       0.61      0.69      0.62       667
weighted avg       0.83      0.73      0.76       667
```

Model 4 : XGBOOST Classifier

```
confusion_matrix for XGBoost
 [[546  20]
 [ 28  73]]
***************************************************
              precision    recall  f1-score   support

           0       0.97      0.99      0.98       566
           1       0.91      0.81      0.86       101

    accuracy                           0.96       667
   macro avg       0.94      0.90      0.92       667
weighted avg       0.96      0.96      0.96       667

Test ROC AUC Score: 0.8988734562502186
```

Model 5 :Tuned XGBOOST

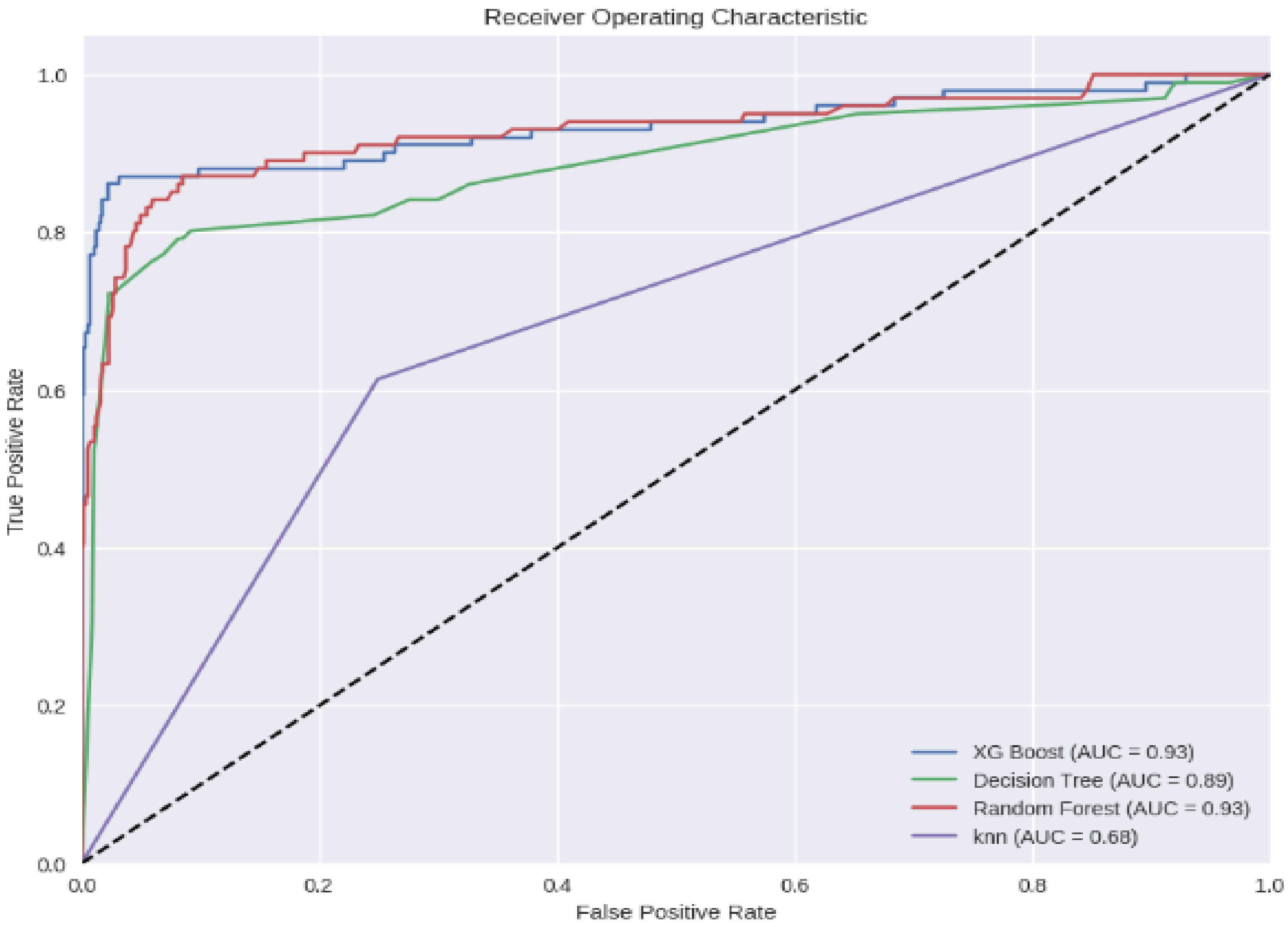
```
confusion_matrix for XGBoost
 [[556  10]
 [ 21  80]]
*************************************************
              precision    recall  f1-score   support

           0       0.96      0.98      0.97       566
           1       0.89      0.79      0.84       101

    accuracy                           0.95       667
   macro avg       0.93      0.89      0.91       667
weighted avg       0.95      0.95      0.95       667

Test ROC AUC Score: 0.8872056816989119
```

Model Evaluation



We assessed their performance using two metrics: F1-score and Test ROC AUC Score.

Among the four models, XG Boost had the highest F1-score of 0.96 and ROC AUC Score of 0.89 which indicates that it can make more accurate predictions compared to the other models.

Conclusion

Gradient Boosting (XGBoost) was our best model to predict churn patterns.

Customer churn is existent in each state however we can’t fully attribute the relationship to a specific state or certain reason. Other attributable factors include:

- Regional Preferences
- Competition
- Service Quality
- Demographics
- Regulations

There is an increasing relationship between the number of customer service calls and customer churn

The different times of day a call was made influences the likelihood of churn, however not directly.

The area code does not highly affect churning but could be attributable due to insufficient network masts or lack of product knowledge in specific area code.

Recommendations

- Enhance Network Coverage
- Putting the Customer First with Personalized Experience
- Proactive Customer Support
- Introduce Value-Added Services and Offers
- Regular Communication
- Customer Feedback and Surveys
- Community Engagement

Thank You

?

Any Questions..

That’s All Folks!

CodeTribe_3.0

TEAM

Stephen Ndirangu

Kithinji Murungi

Stephen Gathai

Jael Akech

MariaCharlote Mbiyu

Muthoni Kahura

Maureen Wangonyo