# PHASE THREE PROJECT

**NAME**: AGGREY MUSONYE TIMBWA

**DATA BEING ANALYSED**: STRIATEL CUSTOMER CHUN

**STUDENT PACE**: PART TIME

**PROJECT PERIOD**: 22/08/2024 TO 30/08/2024

# BUSINESS UNDERSTANDING

- SyriaTel, like many telecommunications companies, faces the challenge of customers stop using their services because of one reason or another.
- This not only leads to a direct loss in revenue but also incurs additional costs related to acquiring new customers to replace those who have left through advertsing and marketing.
- The goal of this project is to predicts whether a customer is likely to stop doing business with SyriaTel in the near future.
- By identifying patterns and factors that contribute to customer stop using their services, the company can take proactive measures to retain at-risk customers and at the same same time gain new ones.

# OBJECTIVES

1. To develop a predictive model
2. understand the key features in the data set
3. Evaluate the model
4. Aid in decision making

# DATA UNDERSTANDING

The data used in the analysis include the following;

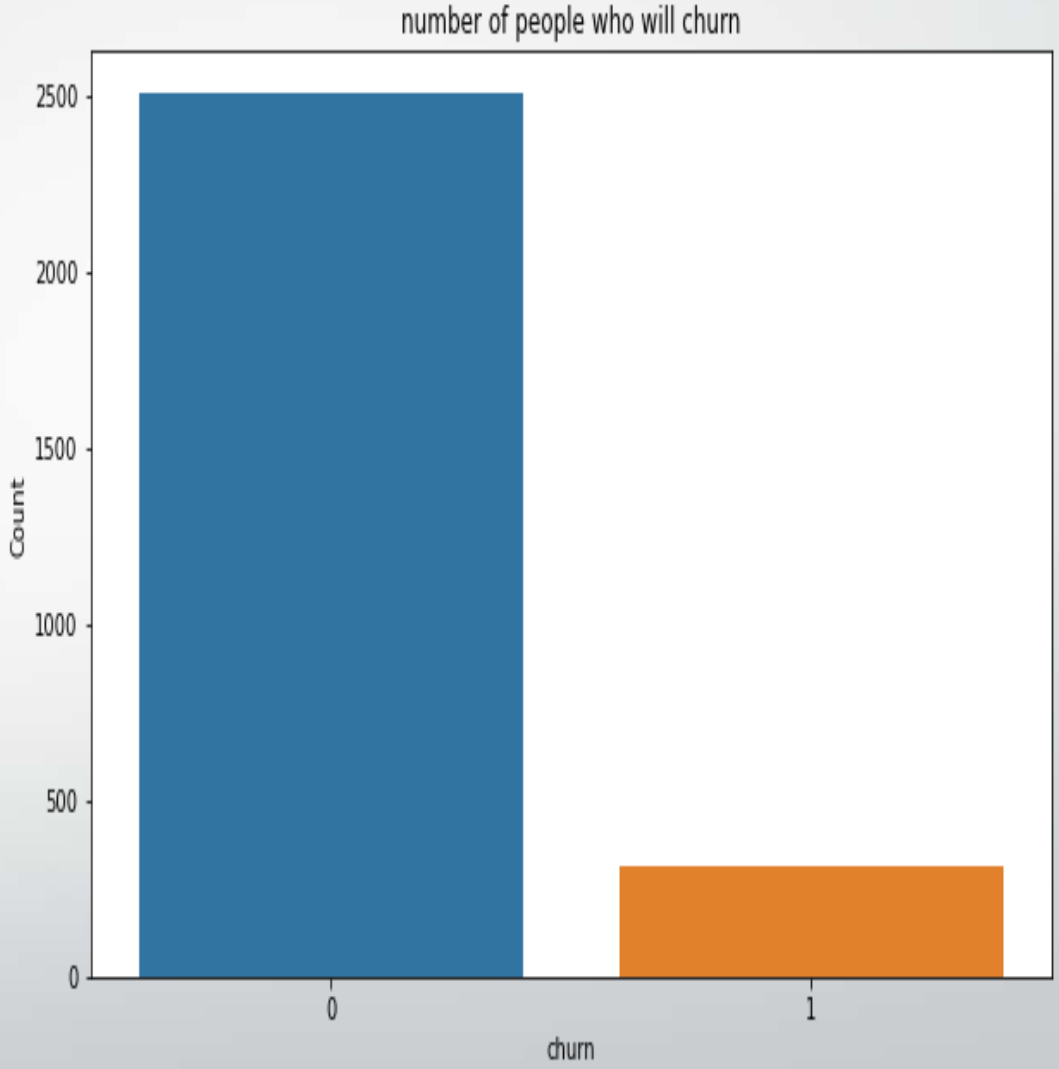
1. state
2. account length
3. area code
4. phone number
5. international plan
6. voice mail plan
7. number vmail messages
8. total day minutes
9. total day calls
10. total day charge
11. total eve minutes
12. total eve calls
13. total eve charge
14. total night minutes
15. total night calls
16. total night charge
17. total intl minutes
18. total intl calls
19. total intl charge
20. customer service calls
21. churn

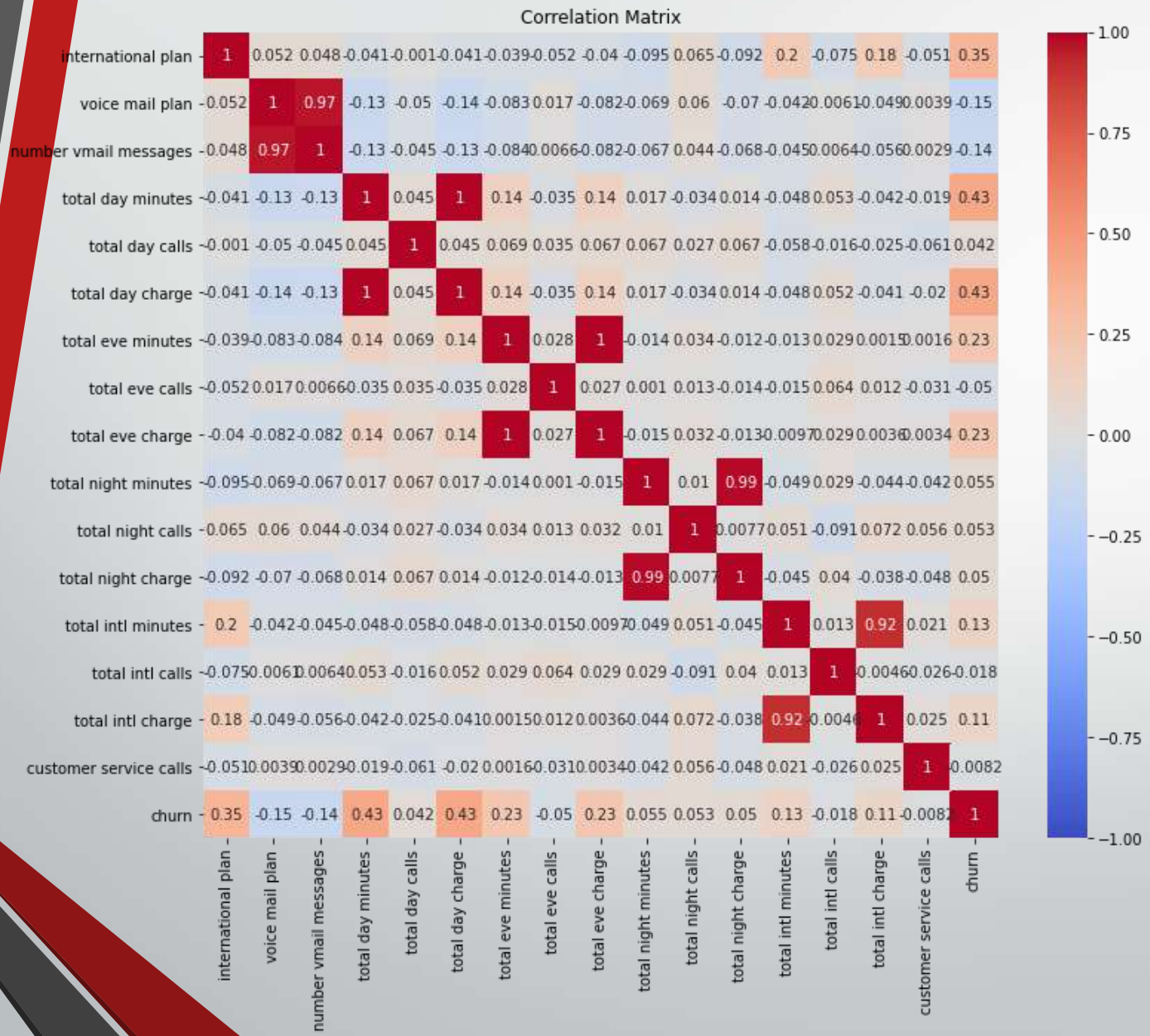
# MODELLING

The dependent variable that is to be observed is churn

However the data set is not well balanced. If the data is taken as is above there will be number of people who will churn is low therefore there will be over sampling of those who will not churn

The data should there fore be altered to prevent oversampling of "false" in the churning collumn before any training of the data set is done






# CHECKING CORRELATION

There are some attributes that have a high correlation to churn as compared to others.

The attributes that have a high correlation are likely to affect churn more.

The total day charge has the highest correlation to churn of 0.42822.

The lowest negative correlation is customer service call at -0.008192

# EVALUATION

The co-efficient matrix of the training data can be seen on the right

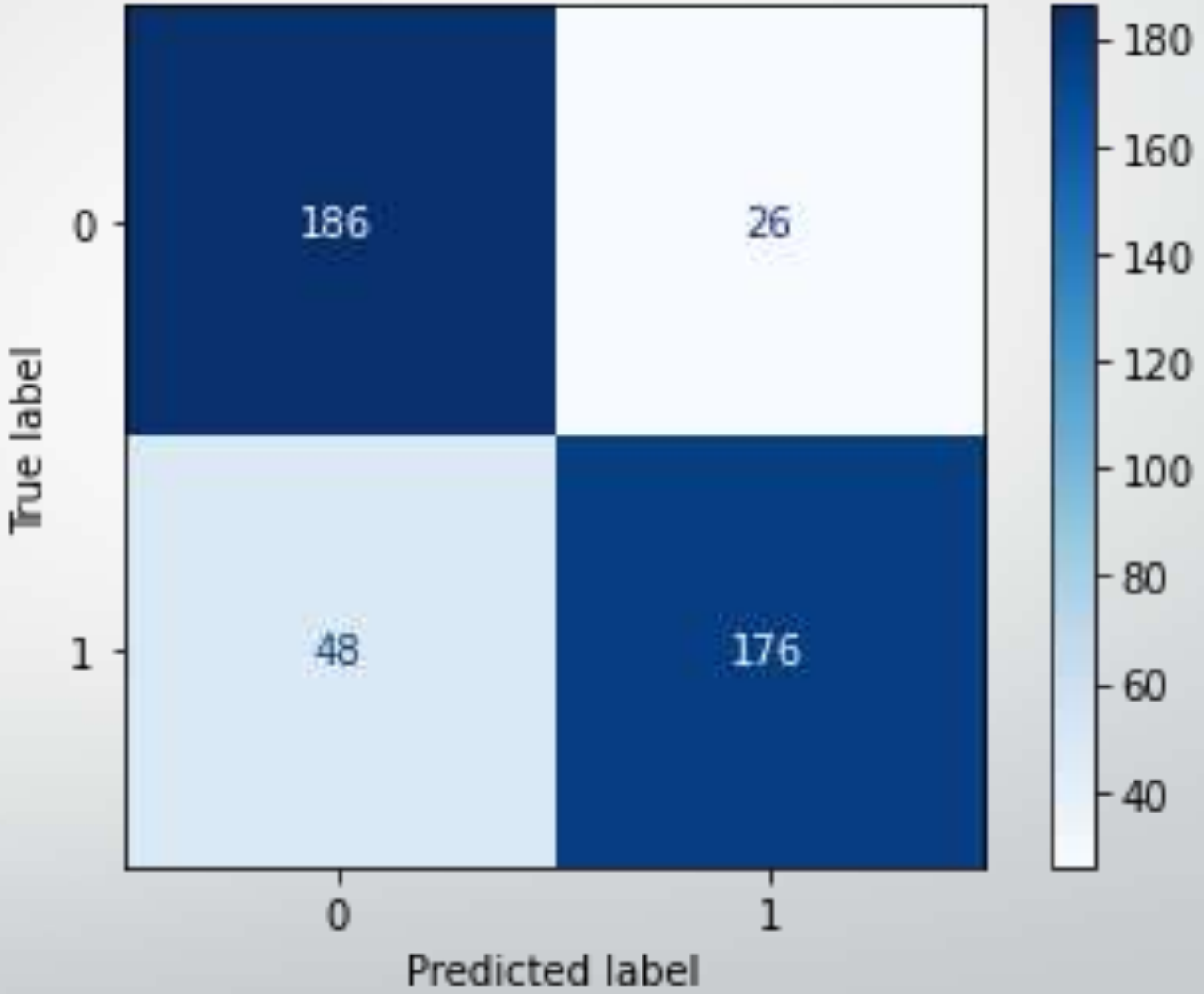
From the matrix the following calculated percentages can be obtained;

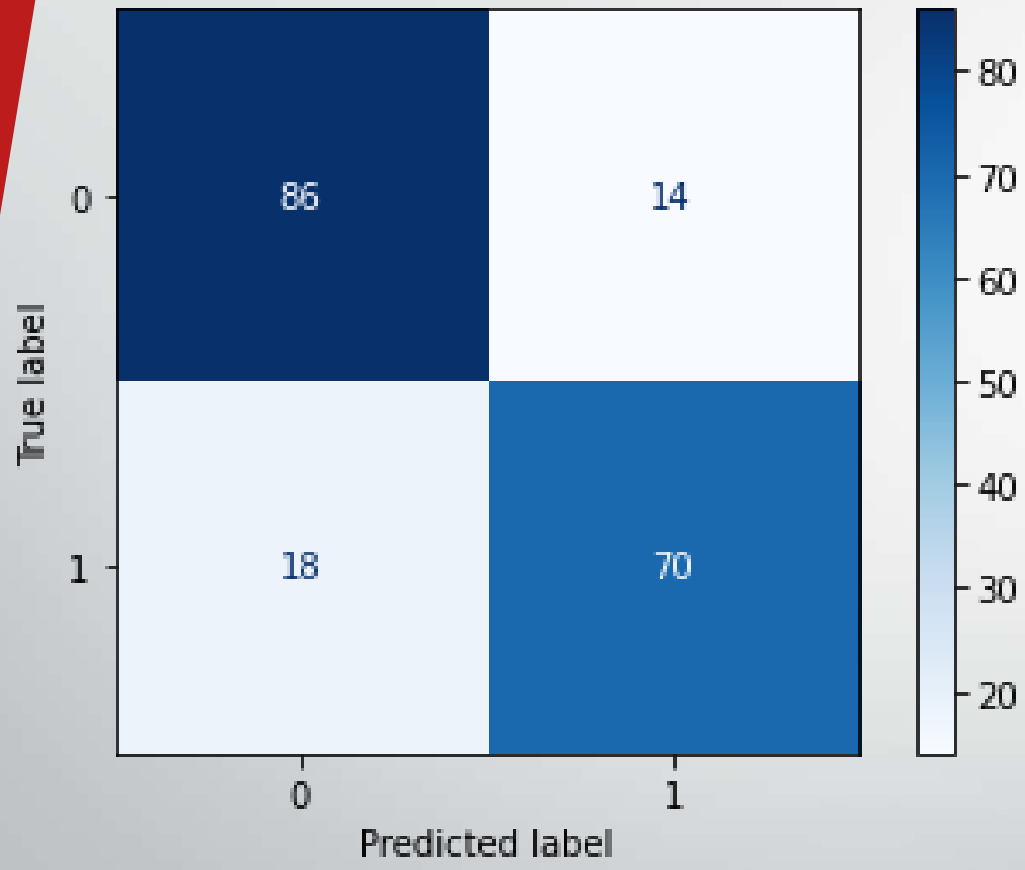
precision is 87.12871287128714

accuracy is 83.02752293577981

recall is 78.57142857142857

f1 score is 82.62910798122066






# TESTING DATA

The co-efficient matrix of the testing data can be seen on the left

From the matrix the following calculated percentages can be obtained;

precision is 83.33333333333334

accuracy is 82.97872340425532

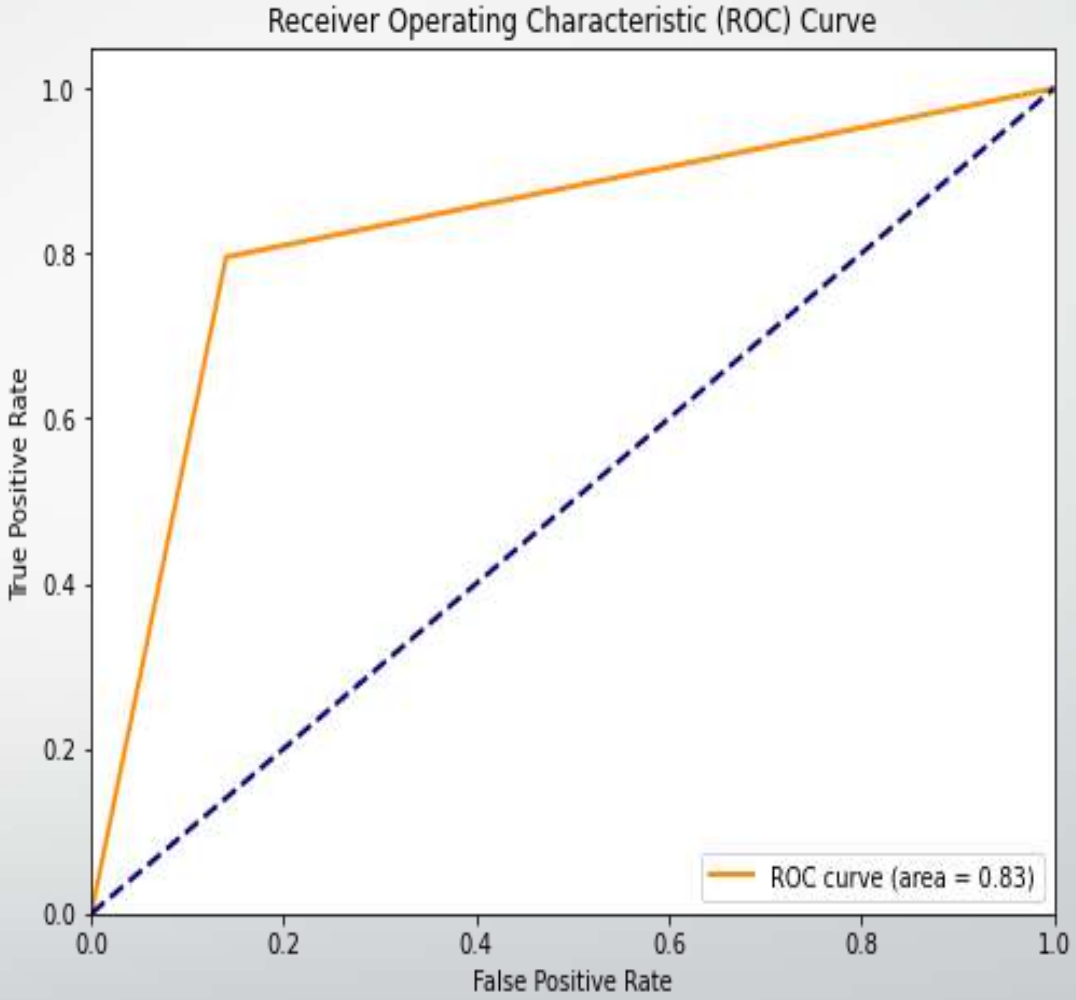
recall is 79.54545454545455

f1 score is 81.3953488372093

# AUC VALUE

An AUC of 0.83 suggests that the model is good at distinguishing between the positive and negative classes.

Specifically, there is an 83% chance that the model will rank a randomly chosen positive instance higher than a randomly chosen negative instance.




Thank you!