
SYRIATEL
TELECOMMUNICATION
COMPANY

introduction

SyriaTel is a telecommunication company in Syria.it's one of the companies that provide communication services, it cuurently experience decrease in revenue returns. Customer churn is one of the problem the company is facing. Due to the direct effect on the company revenue, the company is looking to identify means to predict potential costumers to churn. therefore finding factors that increases customer churn is important in order to mitigate the risk by taking the neccesary actiont

Problem statement

Predicting Customer Churn for SyriaTel Telecommunications Company

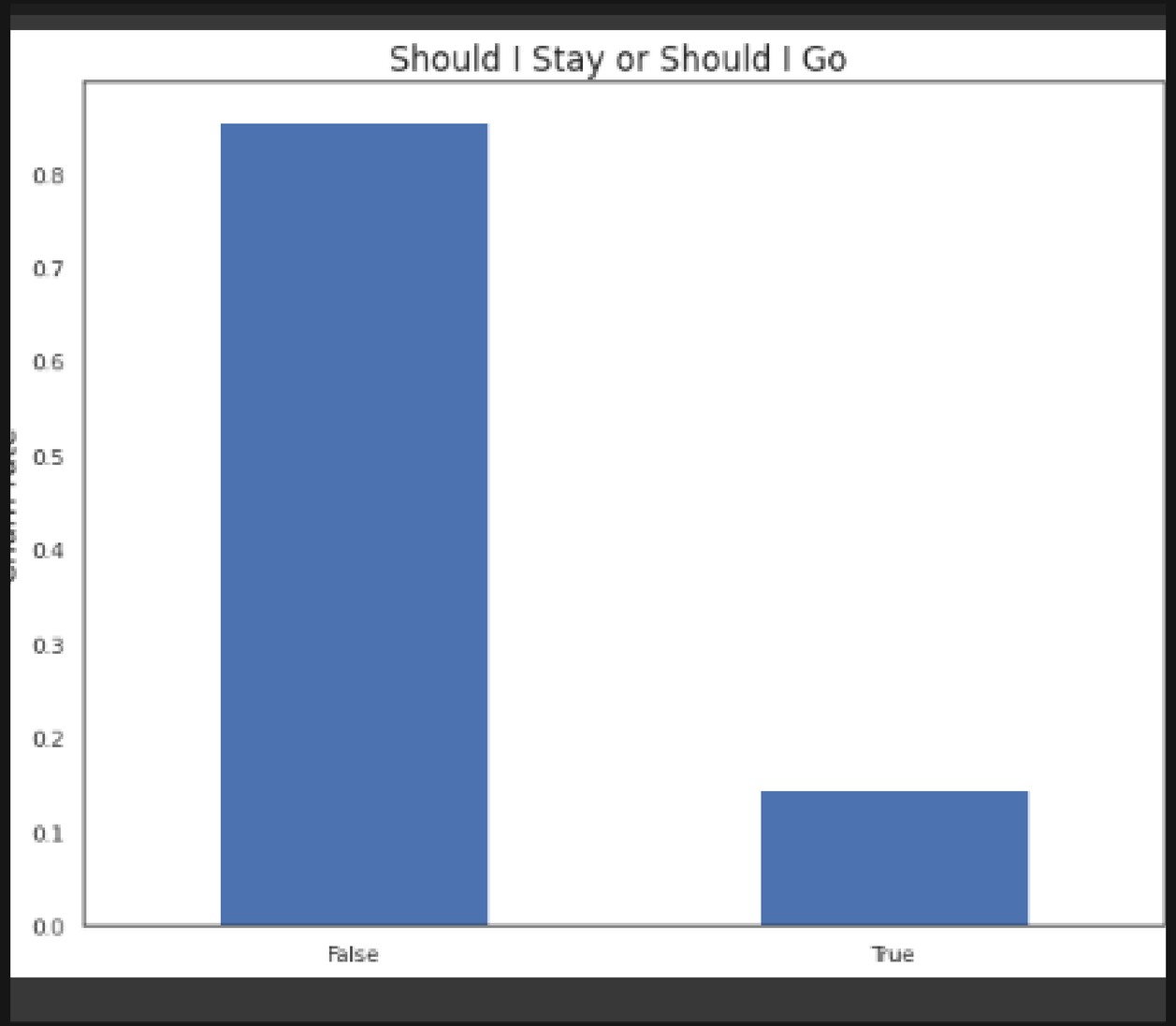
In todays world every company is looking to increase its revenue same to SyriaTel company. But recently it has experience some decline in its revenue and thus requested to look into the problem causing them

Objectives

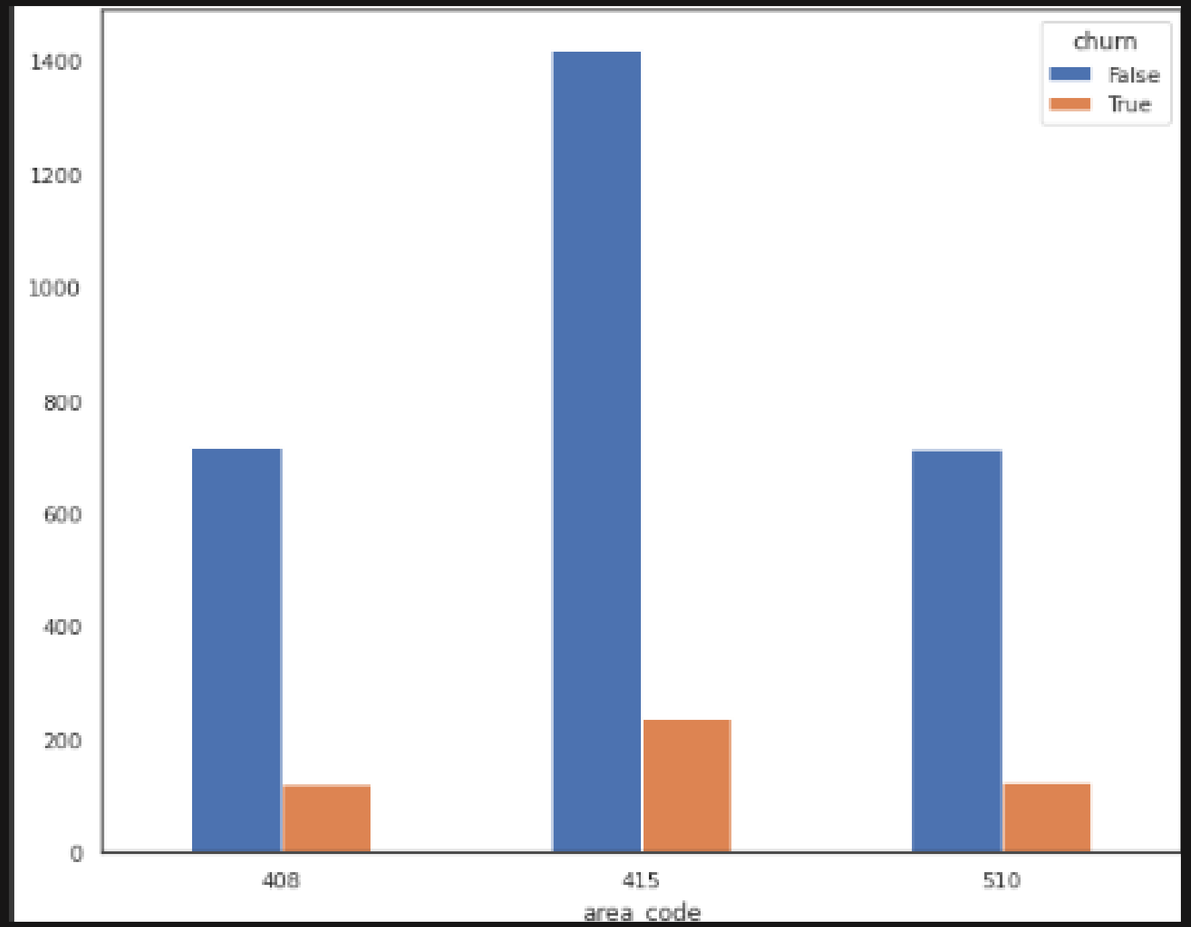
The objective of the project is to identify the best machine learning model that can predict causes of customer churn and how to metigate it in order to avoid lossing the customers and what to improve on the services provided

explonatory Data Analysis (EDA)

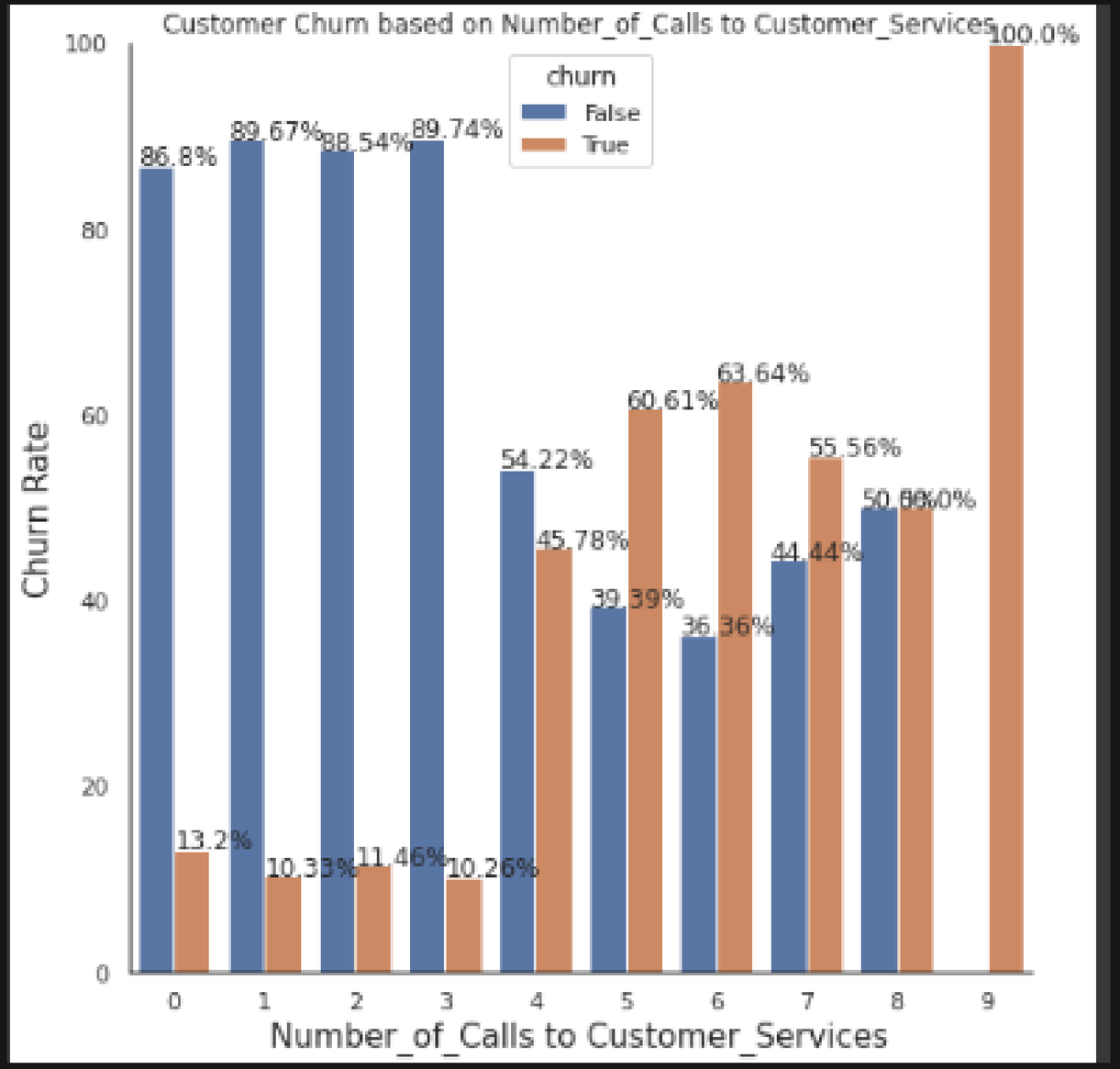
graph showing customers who left versus those that remained



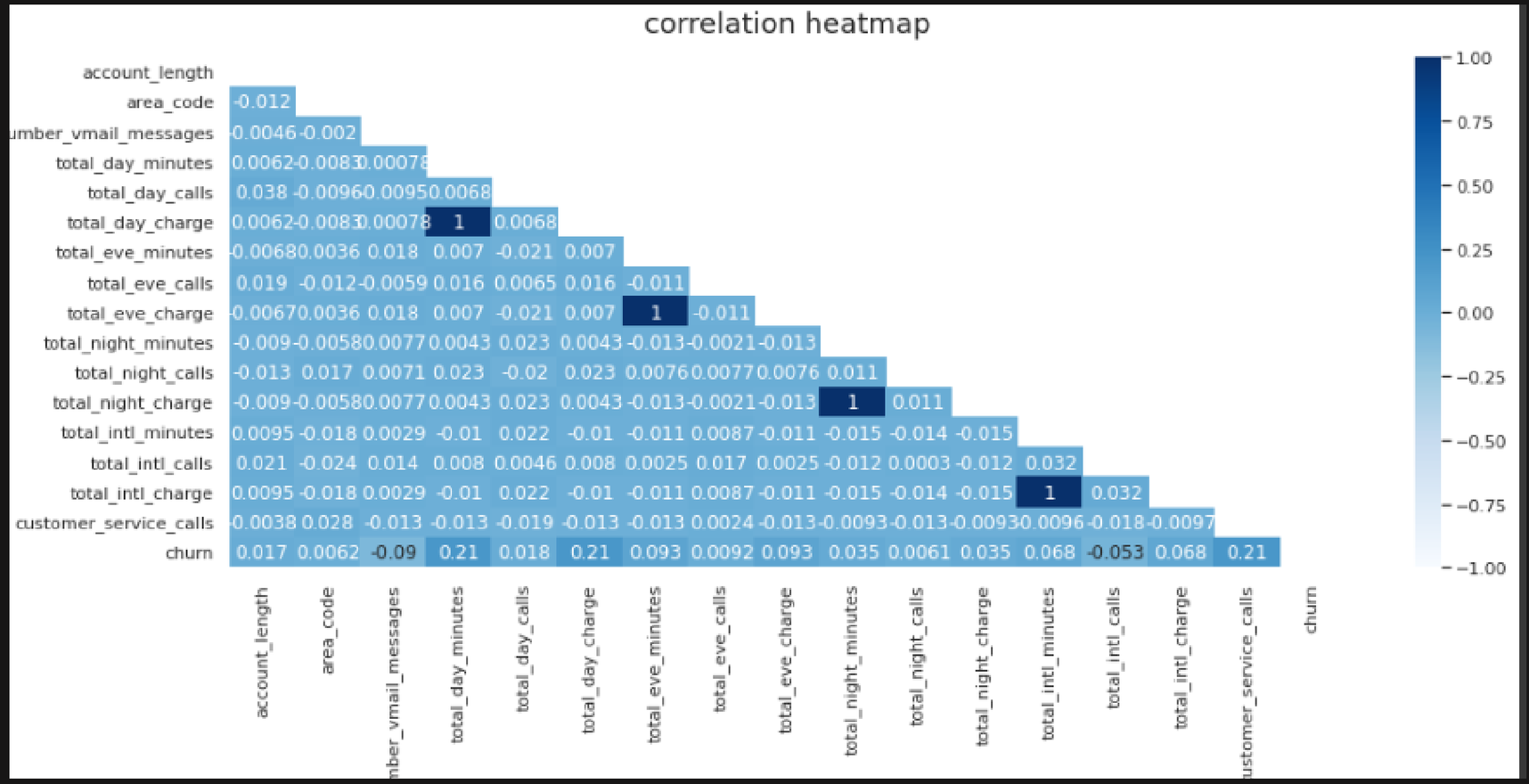
graph showing area code for customer who left



Graph showing number of customers that called more than four times left



graph showing relationship between different varaibles



```
churn                          1.000000
customer_service_calls         0.208750
total_day_minutes              0.205151
total_day_charge               0.205151
total_eve_minutes              0.092796
total_eve_charge               0.092786
total_intl_charge              0.068259
total_intl_minutes             0.068239
total_night_charge             0.035496
total_night_minutes            0.035493
total_day_calls                0.018459
account_length                 0.016541
total_eve_calls                0.009233
area_code                      0.006174
total_night_calls              0.006141
total_intl_calls              -0.052844
number_vmail_messages         -0.089728
Name: churn, dtype: float64
```

Correlation between variables and churn

modeling and evaluation

Accuracy and evalution of logistic regression model

```
----------------------------------------
Accuracy score for Training Dataset =  0.6634653861544618
Accuracy score for Testing Dataset =  0.6462829736211031
----------------------------------------
------------Confusion Matrix------------
Train set:
[[1336  801]
 [  40  322]]

Test set:
[[441 272]
 [ 23  98]]

----------------------------------------
---------Classification Matrix:---------
Train set:
              precision    recall  f1-score   support

           0       0.97      0.63      0.76      2137
           1       0.29      0.89      0.43       362

    accuracy                           0.66      2499
   macro avg       0.63      0.76      0.60      2499
weighted avg       0.87      0.66      0.71      2499


Test set:
              precision    recall  f1-score   support

           0       0.95      0.62      0.75       713
           1       0.26      0.81      0.40       121

    accuracy                           0.65       834
   macro avg       0.61      0.71      0.57       834
weighted avg       0.85      0.65      0.70       834
```

Accuracy and Evalution of GridSearch model

```
[ ] ----------------------------------------
    Accuracy score for Training Dataset =  0.8727490996398559
    Accuracy score for Testing Dataset =  0.8621103117505995
    ----------------------------------------
    ------------Confusion Matrix------------
    Train set:
    [[2072   65]
     [ 253  109]]

    Test set:
    [[686  27]
     [ 88  33]]

    ----------------------------------------
    ---------Classification Matrix:---------
    Train set:
                  precision    recall  f1-score   support

               0       0.89      0.97      0.93      2137
               1       0.63      0.30      0.41       362

        accuracy                           0.87      2499
       macro avg       0.76      0.64      0.67      2499
    weighted avg       0.85      0.87      0.85      2499


    Test set:
                  precision    recall  f1-score   support

               0       0.89      0.96      0.92       713
               1       0.55      0.27      0.36       121

        accuracy                           0.86       834
       macro avg       0.72      0.62      0.64       834
    weighted avg       0.84      0.86      0.84       834
```

Accuracy and Evaluation of SMOTE

```
----------------------------------------
Accuracy score for Training Dataset =  0.7690687880205896
Accuracy score for Testing Dataset =  0.6247002398081535
----------------------------------------
------------Confusion Matrix------------
Train set:
[[1285  852]
 [ 135 2002]]

Test set:
[[423 290]
 [ 23  98]]

----------------------------------------
---------Classification Matrix:---------
Train set:
              precision    recall  f1-score   support

           0       0.90      0.60      0.72      2137
           1       0.70      0.94      0.80      2137

    accuracy                           0.77      4274
   macro avg       0.80      0.77      0.76      4274
weighted avg       0.80      0.77      0.76      4274


Test set:
              precision    recall  f1-score   support

           0       0.95      0.59      0.73       713
           1       0.25      0.81      0.39       121

    accuracy                           0.62       834
   macro avg       0.60      0.70      0.56       834
weighted avg       0.85      0.62      0.68       834
```

Accuracy and Evalution of DecisionTree model

```
----------------------------------------
Accuracy score for Training Dataset =  1.0
Accuracy score for Testing Dataset =  0.9088729016786571
----------------------------------------
------------Confusion Matrix------------
Train set:
[[2137    0]
 [   0  362]]

Test set:
[[677  36]
 [ 40  81]]

----------------------------------------
---------Classification Matrix:---------
Train set:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      2137
           1       1.00      1.00      1.00       362

    accuracy                           1.00      2499
   macro avg       1.00      1.00      1.00      2499
weighted avg       1.00      1.00      1.00      2499


Test set:
              precision    recall  f1-score   support

           0       0.94      0.95      0.95       713
           1       0.69      0.67      0.68       121

    accuracy                           0.91       834
   macro avg       0.82      0.81      0.81       834
weighted avg       0.91      0.91      0.91       834
```

Accuracy and Evaluation of Random Forest model

```
----------------------------------------
Accuracy score for Training Dataset =  0.8899559823929571
Accuracy score for Testing Dataset =  0.8776978417266187
----------------------------------------
------------Confusion Matrix------------
Train set:
[[2137    0]
 [ 275   87]]

Test set:
[[711   2]
 [100  21]]

----------------------------------------
---------Classification Matrix:---------
Train set:
              precision    recall  f1-score   support

           0       0.89      1.00      0.94      2137
           1       1.00      0.24      0.39       362

    accuracy                           0.89      2499
   macro avg       0.94      0.62      0.66      2499
weighted avg       0.90      0.89      0.86      2499


Test set:
              precision    recall  f1-score   support

           0       0.88      1.00      0.93       713
           1       0.91      0.17      0.29       121

    accuracy                           0.88       834
   macro avg       0.89      0.59      0.61       834
weighted avg       0.88      0.88      0.84       834
```

Accuracy and Evaluation of XGBoost modelt

```
[ ]
----------------------------------------
Accuracy score for Training Dataset =  0.9683873549419768
Accuracy score for Testing Dataset =  0.9436450839328537
----------------------------------------
------------Confusion Matrix------------
Train set:
[[2133    4]
 [  75  287]]

Test set:
[[703  10]
 [ 37  84]]

----------------------------------------
---------Classification Matrix:---------
Train set:
              precision    recall  f1-score   support

           0       0.97      1.00      0.98      2137
           1       0.99      0.79      0.88       362

    accuracy                           0.97      2499
   macro avg       0.98      0.90      0.93      2499
weighted avg       0.97      0.97      0.97      2499


Test set:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97       713
           1       0.89      0.69      0.78       121

    accuracy                           0.94       834
   macro avg       0.92      0.84      0.87       834
weighted avg       0.94      0.94      0.94       834
```

XGBoost after tuning

```
[ ] ----------------------------------------
    Accuracy score for Training Dataset =  0.976390556222489
    Accuracy score for Testing Dataset =  0.9448441247002398
    ----------------------------------------
    ------------Confusion Matrix------------
    Train set:
    [[2135    2]
     [  57  305]]

    Test set:
    [[701  12]
     [ 34  87]]

    ----------------------------------------
    ---------Classification Matrix:---------
    Train set:
                  precision    recall  f1-score   support

               0       0.97      1.00      0.99      2137
               1       0.99      0.84      0.91       362

        accuracy                           0.98      2499
       macro avg       0.98      0.92      0.95      2499
    weighted avg       0.98      0.98      0.98      2499


    Test set:
                  precision    recall  f1-score   support

               0       0.95      0.98      0.97       713
               1       0.88      0.72      0.79       121

        accuracy                           0.94       834
       macro avg       0.92      0.85      0.88       834
    weighted avg       0.94      0.94      0.94       834
```

Conclusion
&
Recommendation

Conclusion

Customers who called more than four times left, this might be due to their concern not addressed or the feedback recieved was not helpful and didn't sort their problem and they opted to try other providers.
There were also some states that had highest number of customers who churned. this can be due to poor network coverage or were not being given good services compared to other telecommunication networks in the area.
Also the charges and minutes of day calls were contributing highly to the costomers churning out. this might be due to high charges imposed on the services provided.

Recommendation

1. XGBoost model can be used for deployment as its the one that produced the highest performance compared to the other model
2. There is need for the company to check on the customers who called more and identify there problem first to avoid losing them
3. Company to develop a department that can deal with customer concerns quickly and to prioristise the areas that are more affected.
4. The model can also be further evaluated to improve its performance even better.
5. Need to train their employees to identify and provide feedback that are more helpful to the company customers to avoid losing them
6. Check on the charges and the day calls, provide affordable services that are standard in regards to the services they offer and compared to other companies
7.

Limitations

- No data regarding competitors in states with higher churn
-
-

THANK YOU