PHASE 3 PROJECT

THIS PROJECT WILL FOCUS ON BUILDING A MACHINE LEARNING MODEL TO PREDICT CUSTOMER CHURN(THIS IS THE RATE AT WHICH CUSTOMERS STOP DOING BUSINESS WITH AN ENTITY) WITHIN SYRIATEL, A TELECOMMUNICATIONS COMPANY.

KEY PHASES

1.BUSINESS UNDERSTANDING

2.DATA UNDERSTANDING AND PREPARATION

3.PERFORMING EXPLORATORY DATA ANALYSIS

4.MODELLING

5.EVALUATING THE MODEL'S EFFECTIVENESS

6.RECOMMENDATIONS AND CONCLUTIONS.

1. BUSINESS UNDESTANDING

- RESEARCH OBJECTIVES:

1. TO IDENTIFY THE PRIMARY FACTORS CONTRIBUTING TO CUSTOMER CHURN.
2. TO EVALUATE AND DETERMINE THE MOST EFFECTIVE PREDICTIVE MODEL FOR CUSTOMER CHURN.
3. TO DEVELOP STRATEGIES AIMED AT IMPROVING CUSTOMER RETENTION AND DECREASING CHURN RATES.

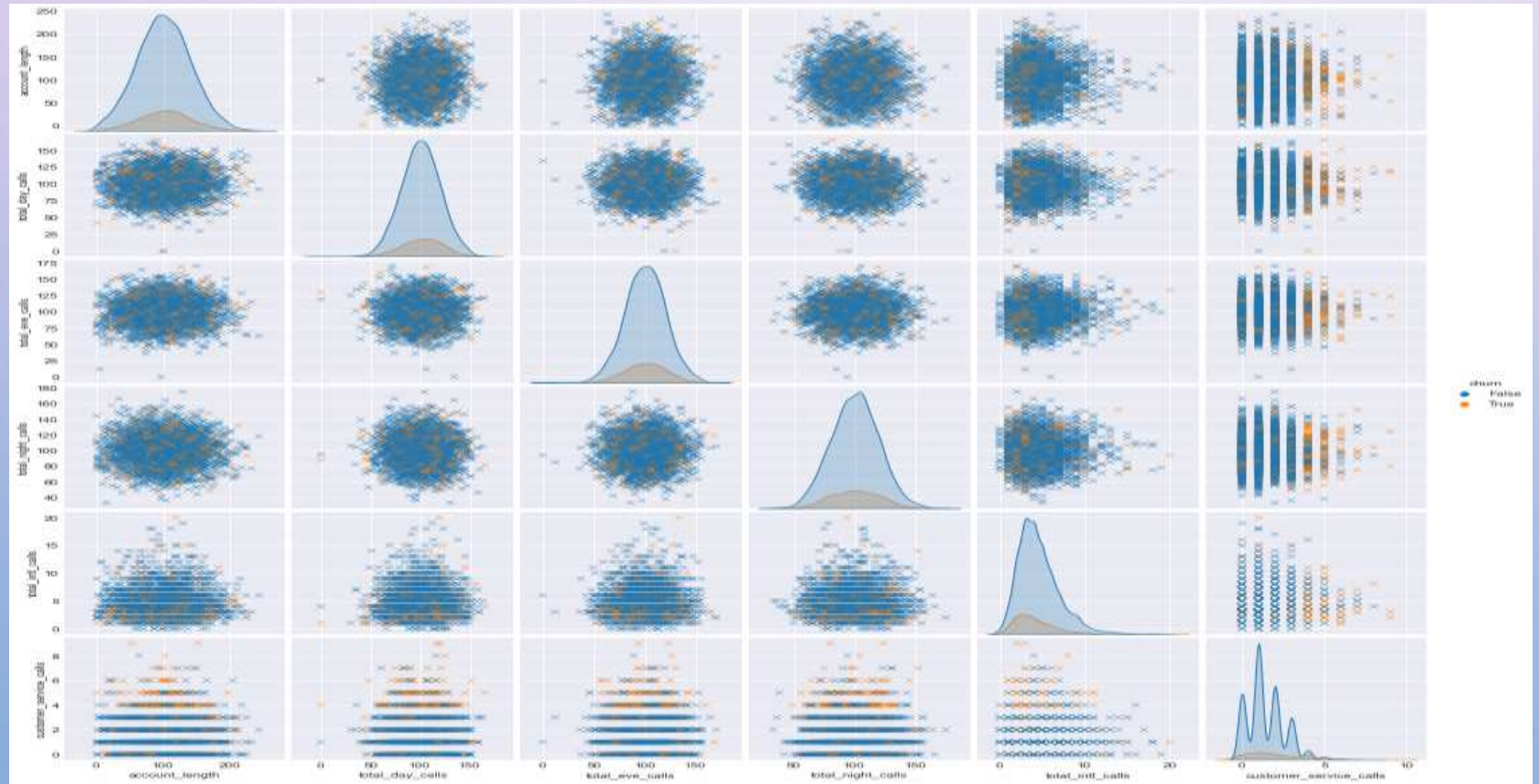
2. DATA UNDERSTANDING AND PREPARATION

THE DATASET CONSISTS OF 3,333 ENTRIES AND 21 COLUMNS. IT INCLUDES VARIOUS DETAILS SUCH AS THE STATE, ACCOUNT LENGTH, AREA CODE, PHONE NUMBER, INTERNATIONAL PLAN, VOICEMAIL PLAN, THE NUMBER OF VOICEMAIL MESSAGES, AND SEVERAL METRICS RELATED TO CALL USAGE, INCLUDING:

1. TOTAL DAY MINUTES, CALLS, AND CHARGES
2. TOTAL EVENING MINUTES, CALLS, AND CHARGES
3. TOTAL NIGHT MINUTES, CALLS, AND CHARGES
4. TOTAL INTERNATIONAL MINUTES, CALLS, AND CHARGES
5. ADDITIONALLY, IT CONTAINS INFORMATION ABOUT CUSTOMER SERVICE CALLS AND A COLUMN WITH A LABEL INDICATING WHETHER THE CUSTOMER HAS CHURNED.

3. PERFORMING EXPLORATORY DATA ANALYSIS

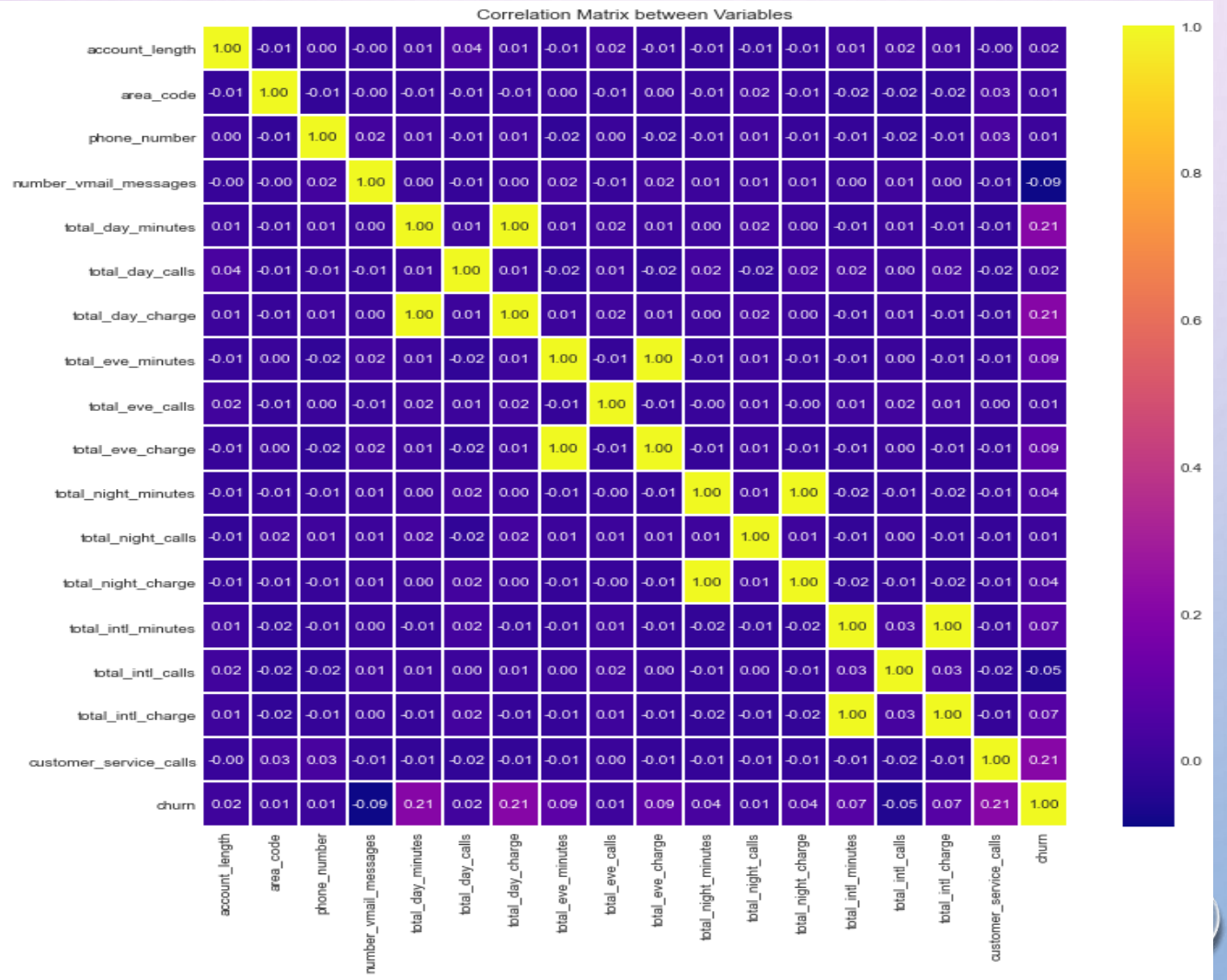
- PAIR PLOTS FOR NUMERIC VARIABLES



FINDINGS

- THERE IS A STRONG RELATIONSHIP BETWEEN CUSTOMER SERVICE CALLS AND TRUE CHURN VALUES.
- AFTER 4 CALLS, CUSTOMERS ARE A LOT MORE LIKELY TO DISCONTINUE THEIR SERVICE.
- BESIDES, MOST CUSTOMER CALLS ARE ARE ASSOCIATED WITH DISATISFACTION WITH CUSTOMER SERVICE.
- AT THIS POINT MORE THAN 4 CUSTOMER CALLS INDICATE THAT IT TAKES LONG FOR THEIR ISSUES TO BE ADDRESSED, AND THUS A POSSIBILITY OF THEM LEAVING INCREASES.

CORRELATION HEATMAP



FINDINGS

- THE CORRELATION MATRIX SHOWS THAT TOTAL INTERNATIONAL CHARGE AND TOTAL INTERNATIONAL MINUTES ,TOTAL NIGHT MINUTES AND TOTAL NIGHT CHARGE, TOTAL EVENING MINUTES AND TOTAL EVENING CHARGE AND TOTAL DAY MINUTES AND TOTAL DAY CHARGE HAVE PERFECT CORRELATION WHICH INDICATES MULTICOLLINEARITY.

4.MODELLING

- FITTING THE MODEL

```
LogisticRegression
LogisticRegression(C=100000000000.0, fit_intercept=False, solver='liblinear')
```

5. MODELLING EVALUATION AND

- 5.1 LOGISTIC REGRESSION
- 5.2 DECISION TREE
- 5.3 RANDOM FOREST

5.1 LOGISTIC REGRESSION RESULTS

NAME: RESIDUALS (COUNTS), DTYPE: INT64

0 2284

1 382

NAME: RESIDUALS (PROPORTIONS), DTYPE: FLOAT64

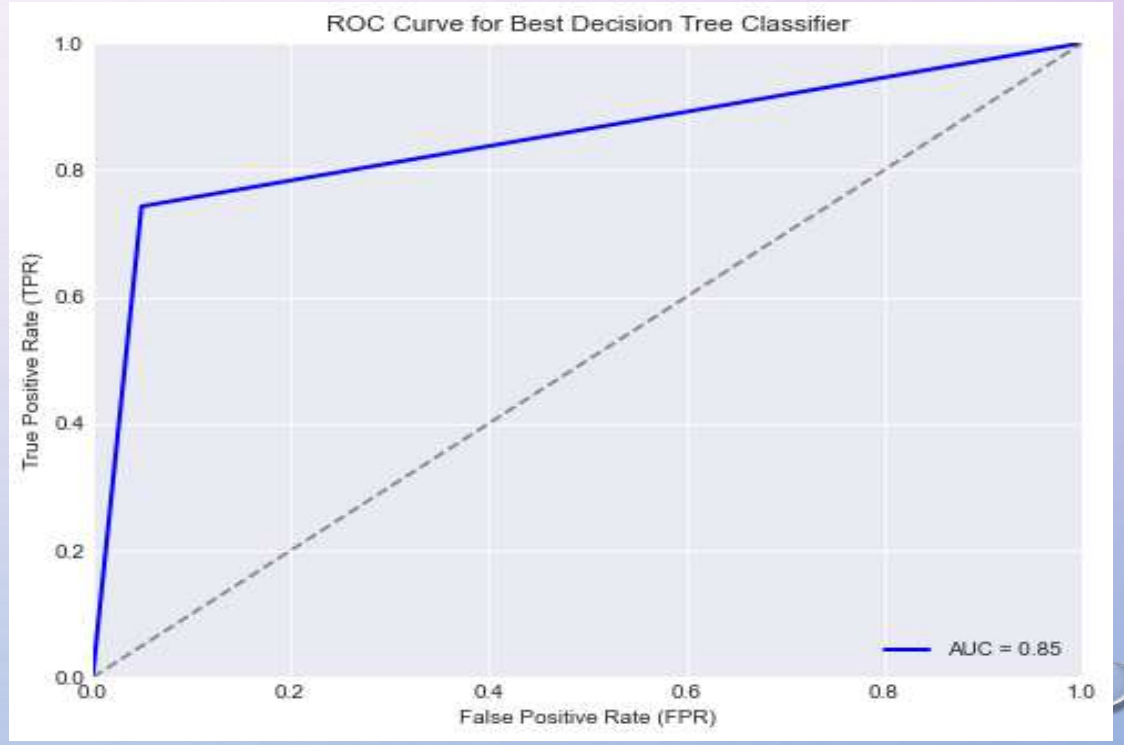
0 0.856714

1 0.143286

- THE CLASSIFIER WAS ABOUT 85% CORRECT ON THE TRAINING DATA WHICH IS CONSIDERED A FAIRLY GOOD PERFORMANCE.

5.2 DECISION TREE FINDINGS

- ACCURACY: 0.8740629685157422
- PRECISION: 0.5620437956204379
- RECALL: 0.7623762376237624
- F1-SCORE: 0.6470588235294118
- TRAIN SCORE 1.0
- TEST SCORE 0.8740629685157422

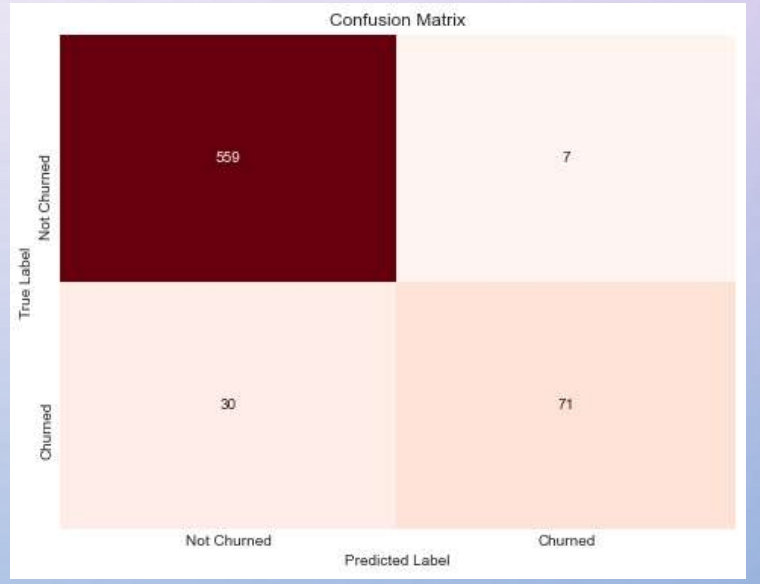


5.3 RANDOM FOREST FINDINGS

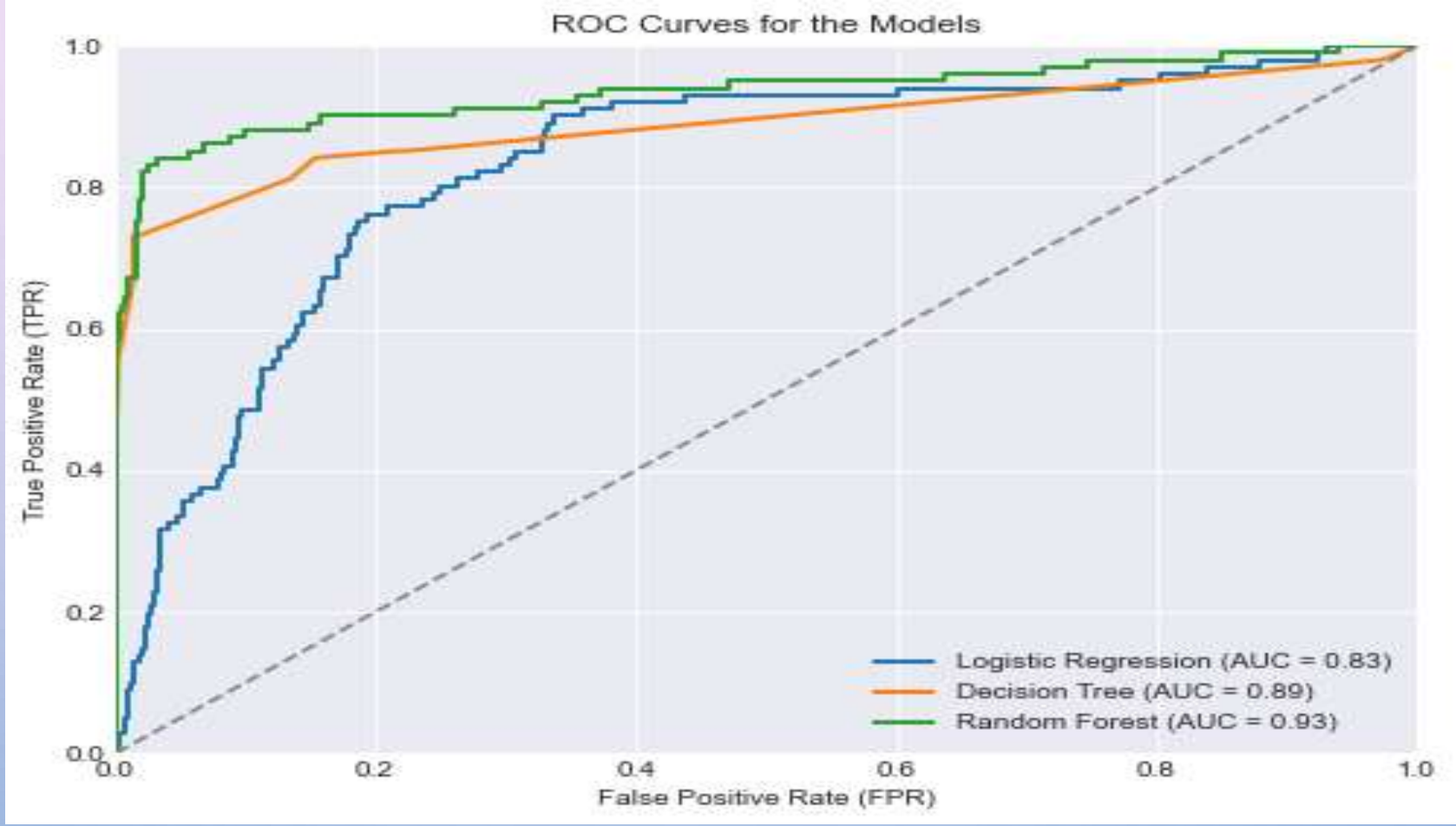
MODEL'S PERFORMANCE

- CROSS-VALIDATION ACCURACY: 0.947861374032927 8
- PRECISION: 0.910256410256410 2
- RECALL: 0.702970297029702 9
- F1-SCORE: 0.793296089385474 8
- TRAIN SCORE: 0.999624906226556 6
- TEST SCORE: 0.944527736131934 1

CONFUSION MATRIX



ROC CURVES FOR THE MODELS



COMPARING THE THREE MODELS

- THE FIGURE ABOVE SHOW THE ROC CURVES FOR LOGISTIC REGRESSION, DECISION TREE, AND RANDOM FOREST MODELS. THE RANDOM FOREST MODEL OUTPERFORMED THE OTHERS, SHOWING A HIGHER AREA UNDER THE CURVE (AUC) AND BETTER CLASSIFICATION PERFORMANCE, MAKING IT THE MOST EFFECTIVE MODEL FOR THE GIVEN TASK.

6. RECOMMENDATION AND CONCLUSION

- IN CONCLUSION, THE ANALYSIS SUGGESTS THAT WE CAN ACCURATELY PREDICT CUSTOMER CHURN USING A MACHINE LEARNING MODEL, WITH THE RANDOM FOREST CLASSIFIER BEING OUR RECOMMENDED MODEL DUE TO ITS STRONG OVERALL PERFORMANCE. AS THIS IS THE BEST PERFORMING MODEL WITH AN ROC CURVE THAT HUGS THE UPPER LEFT CORNER OF THE GRAPH, HENCE GIVING US THE LARGEST AUC (AREA UNDER THE CURVE).
- I WOULD RECOMMEND THAT SYRIATEL MAKE USE OF THE RANDOM FOREST CLASSIFIER AS THE PRIMARY MODEL FOR PREDICTING CUSTOMER CHURN. THIS MODEL HAS A HIGHER ROC CURVE AND STRONG OVERALL PERFORMANCE IN TERMS OF ACCURACY, F1-SCORE, RECALL, AND PRECISION ON THE TEST SET, MAKING IT WELL-SUITED FOR ACCURATELY CLASSIFYING CUSTOMERS AS LIKELY OR UNLIKELY TO CHURN.
- IN TERMS OF BUSINESS STRATEGIC RECOMMENDATIONS FOR SYRIATEL, I WOULD RECOMMEND A CUSTOMER RETENTION STRATEGY THAT ADDRESSES KEY FEATURES IN RELATION TO CALL MINUTES AND CHARGES. THESE EFFORTS COULD INCLUDE PERSONALIZED OFFERS OR DISCOUNTS ON DAY CHARGES. BY IMPLEMENTING COST-EFFECTIVE STRATEGIES THAT ADDRESS THE KEY FACTORS DRIVING CUSTOMER CHURN, SYRIATEL CAN RETAIN CUSTOMERS AND MINIMIZE REVENUE LOSS.
- WE WOULD RECOMMEND, THAT SYRIATEL COMES UP WITH STRATEGIES TO REDUCE ON CUSTOMER SERVICE CALLS, AS THIS IS AMONG THE TOP FEATURES THAT WOULD LIKELY LEAD TO CUSTOMER CHURN. EXAMPLE: COME UP IV