

SYRIATEL CHURN STUDY

REPORT PREPARED FOR SYRIATEL

24.05.2023





PRESENTATION AGENDA

Business Understanding

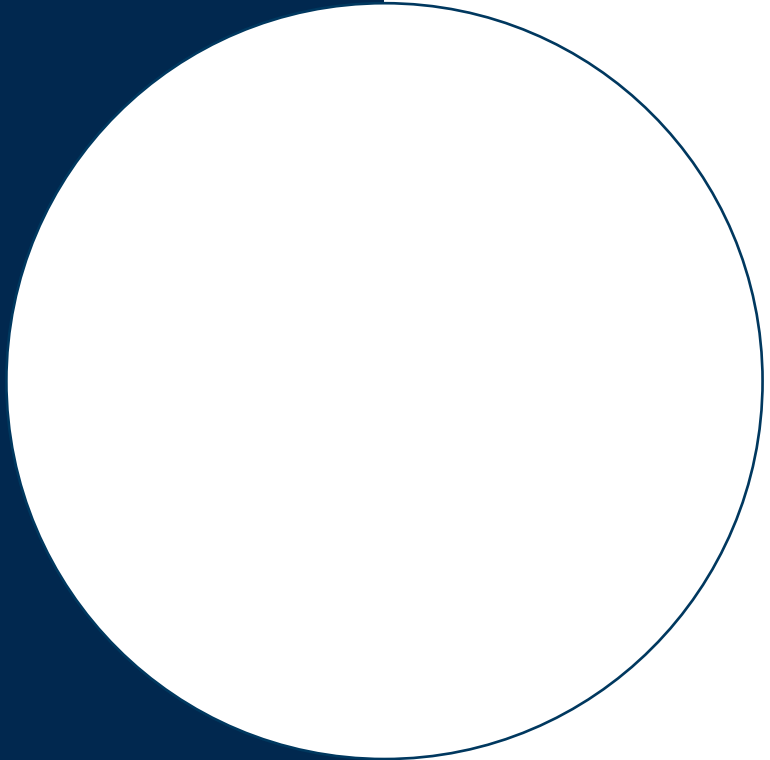
Exploratory Analysis

Modeling

CONCLUSIONS

Business Understanding

- Problem Statement
 - The marketing team in syriatel would like to understand churn trends help them become more competitive against competition. This will help to improve their customer acquisition and retention strategy
- Objectives
 - Understanding the reasons behind customer churn
 - Build a prediction model to help proof the business against churn
 - Reduce churn to improve business performance

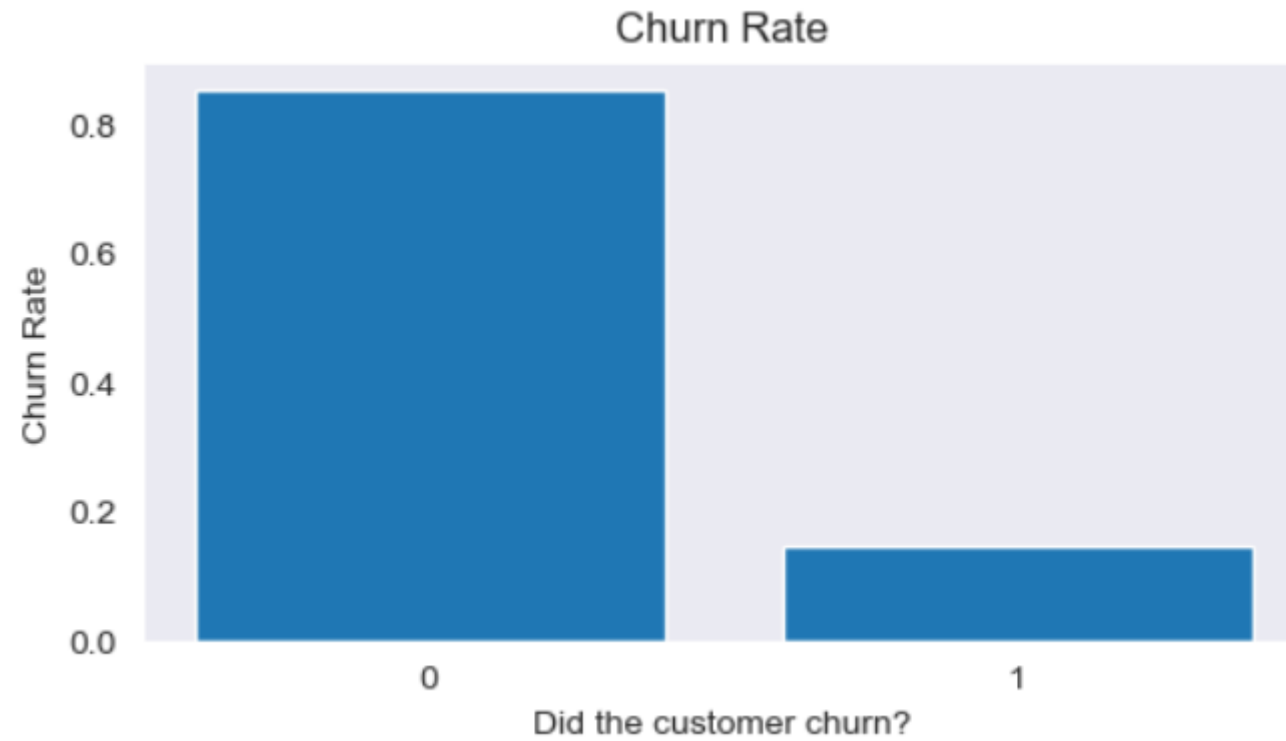


EXPLORATORY DATA ANALYSIS



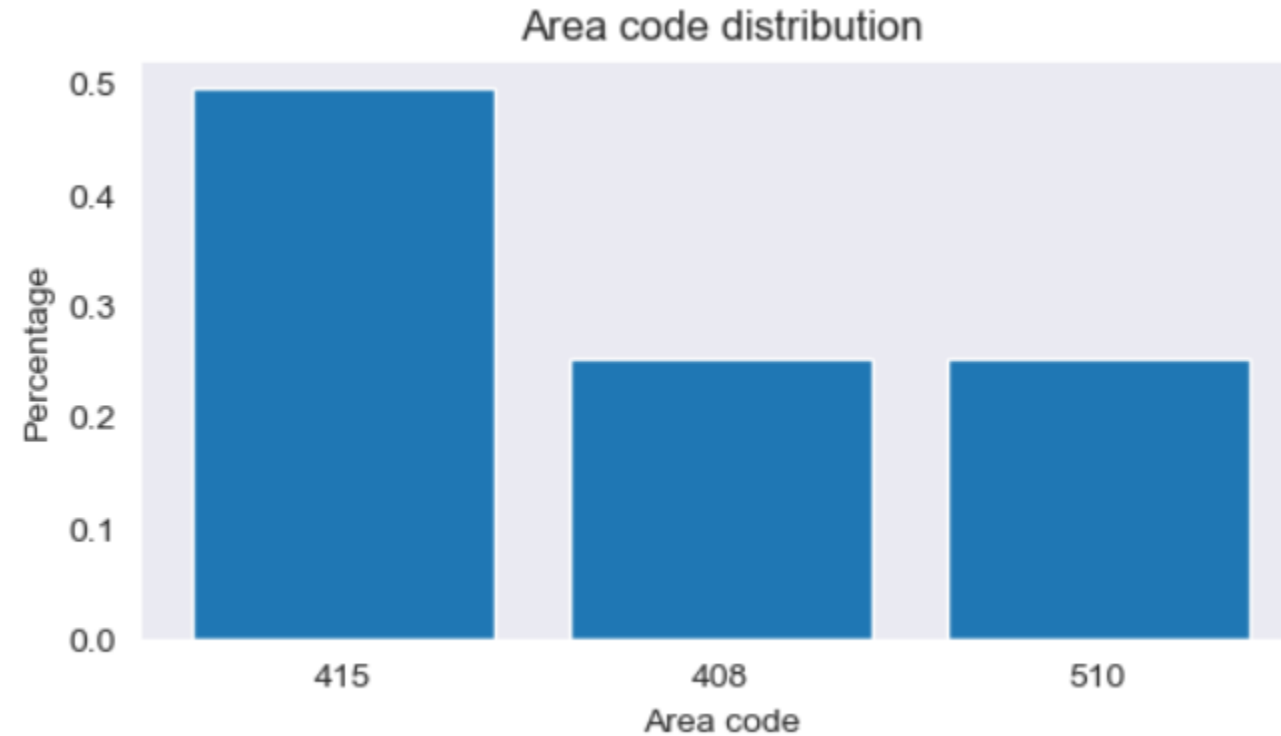
Churn Rate

- The data shows a churn rate of 14%, meaning that our target variable is imbalanced. We will therefore have to correct for the imbalances when modeling



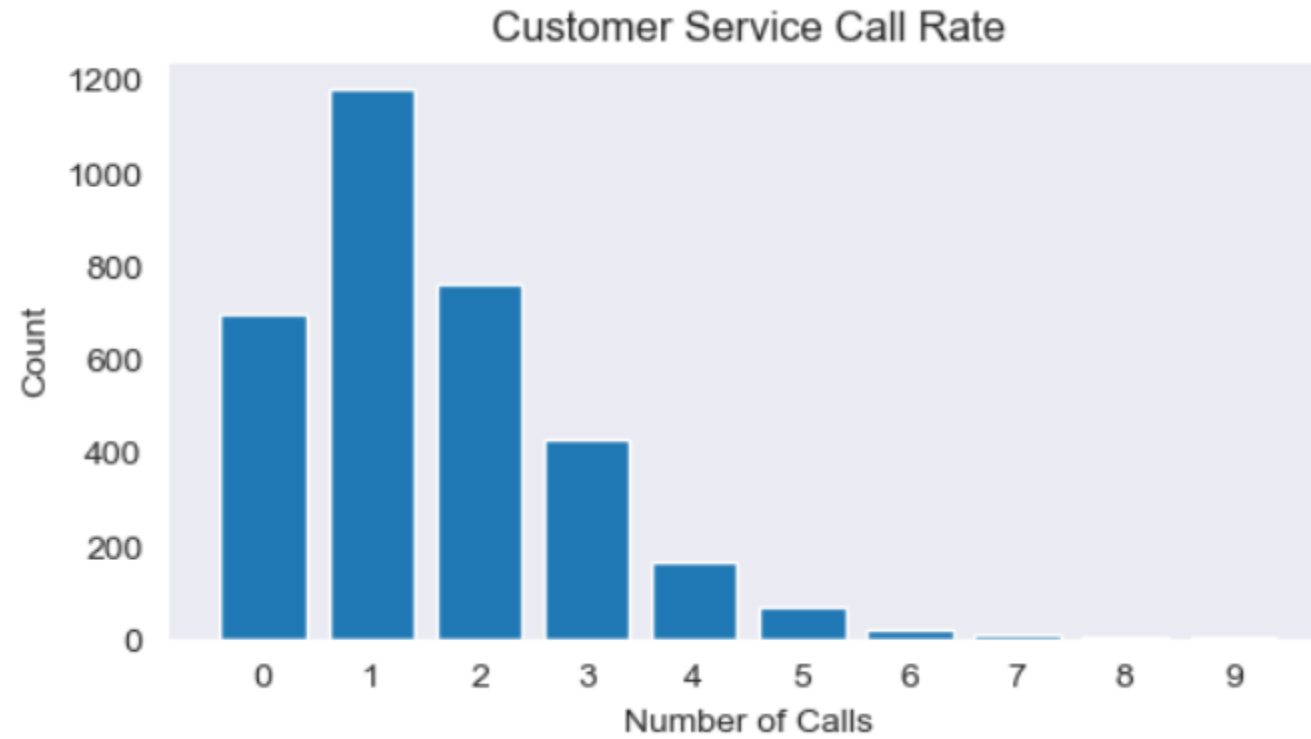
Geographic Distribution

- Roughly half of the subscribers are located in area code 415. The remainder are evenly distributed between 408 and 510



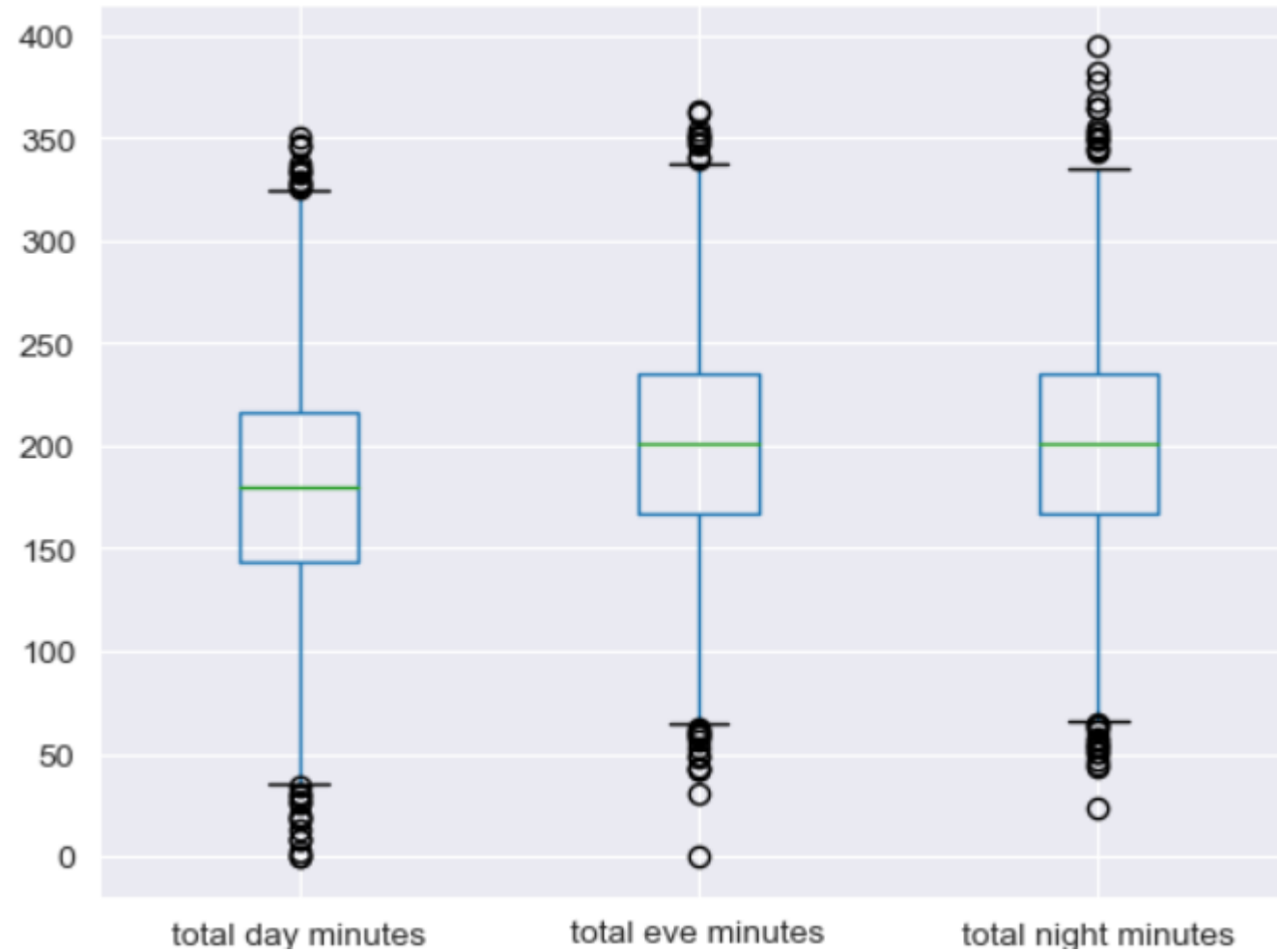
Customer Service Calls

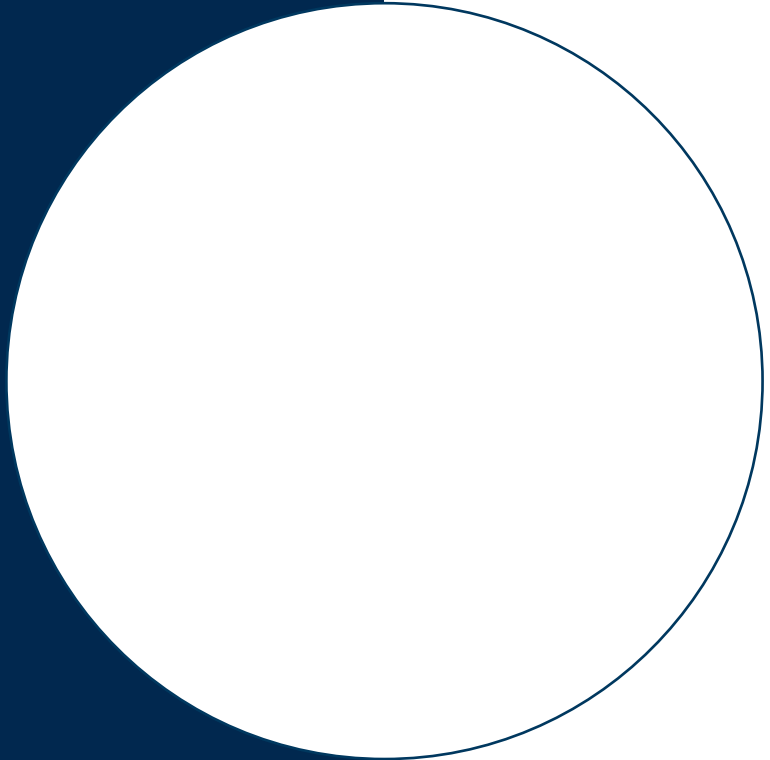
- Calls to customer service is binomially distributed with most people making 1 to 3 calls



How Does Call Duration Vary By Day Part?

- Call duration increases by day part, thus evening and night calls last longer than day calls





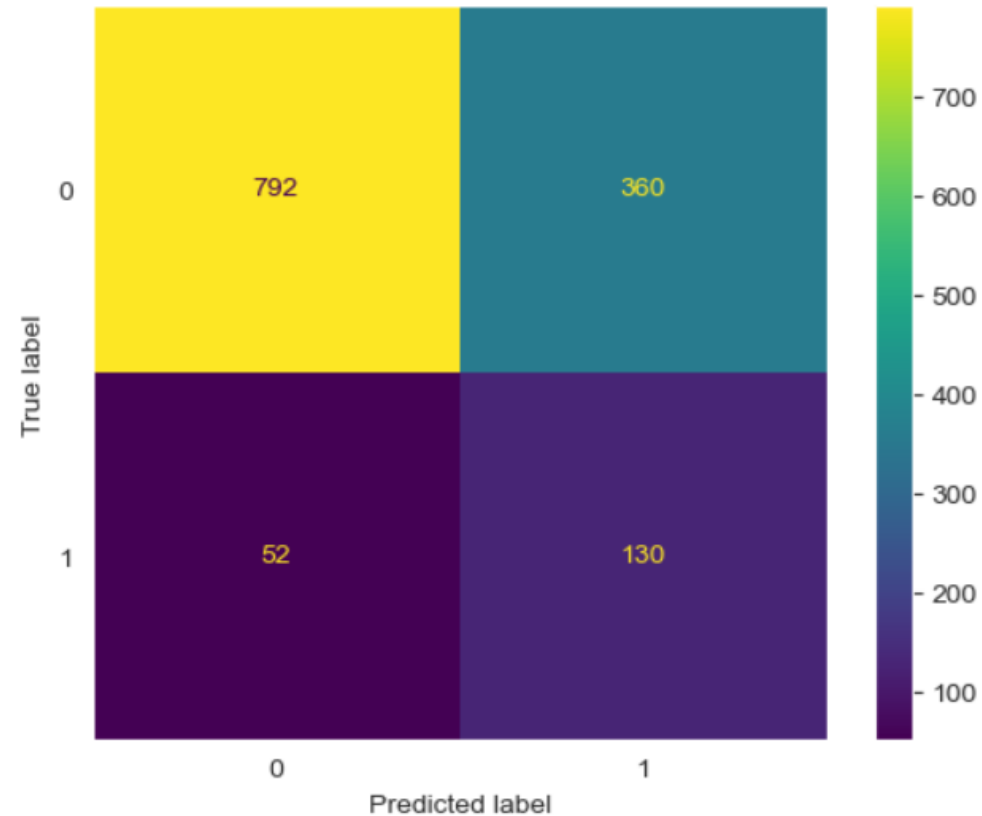
MODELING



- We start with logistic regression as our baseline model. We will use a pipeline to streamline our work and balanced class weight to account for class imbalances. From the metrics below, we see that our baseline model can be improved. We have a high number of false positives affecting the precision score. However, our key metrics to consider are the accuracy and auc scores

Baseline Model

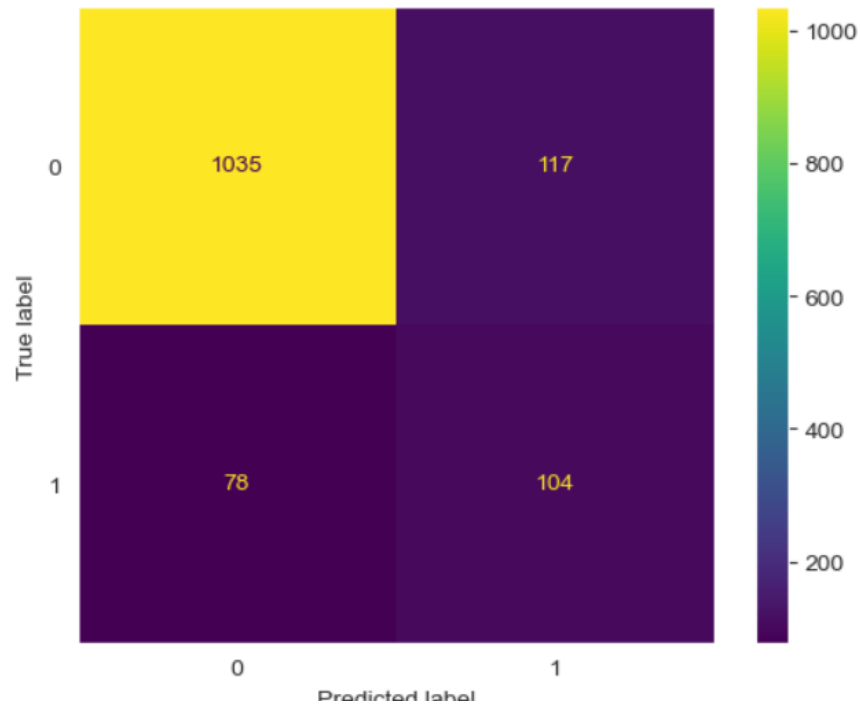
```
precision: 0.265  
recall: 0.714  
accuracy: 0.691  
AUC: 0.701
```



Improved Model: Decision Tree

- Next we will use DecisionTree with grid search to look for optimal solutions.

precision: 0.471
recall: 0.571
accuracy: 0.854
AUC: 0.735



- With Decision Tree classifier, we have been able to improve the metrics as shown below
 - accuracy improves from 69% to 85%
 - auc score improves from 70% to 73%

Conclusion and Recommendations

We therefore conclude that a decision tree model of `max_depth` None and `min_samples_split` 5 is the better model at predicting churn

Areas of further investigation include:

- trying other models like ensemble methods
- further tuning of the model
- applying dimensionality reduction to engineer correlated features

