SYRIATEL CHURN PROJECT

Intro

Business Understanding

The problem statement present in SyriaTel customer churn dataset is a binary classification problem in which we are to determine if customer will soon stop doing business with syriaTel, telecommunications company in an effort of reducing money lost because of customers who don't stay long.

Data Understanding

the dataset encompasses a range of customer metrics, including service usage, billing information, and customer plans, all crucial for predicting churn. By analyzing features such as total charges, customer service interactions, and plan types, we can identify patterns and factors most strongly associated with customer attrition.

Objectives

main objective.

Determining customers who will soon stop using SyriaTel

Other objectives

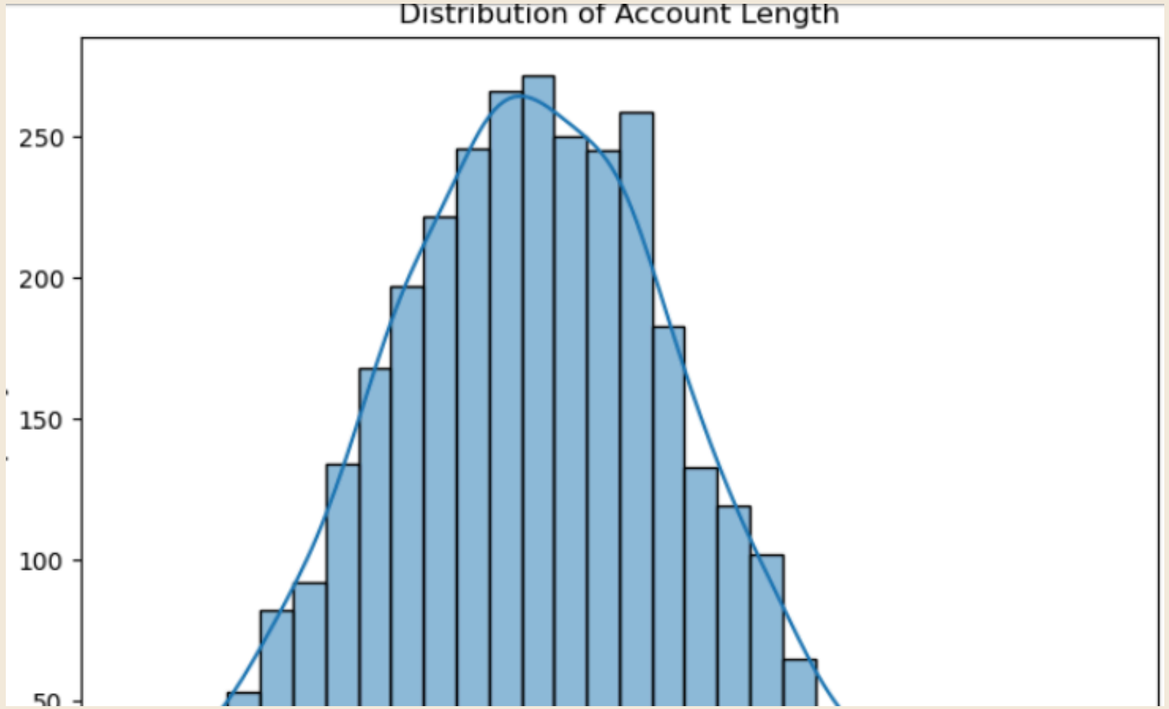
1. To determine most important features for model.
2. Evaluate and optimize model performance.

Contents

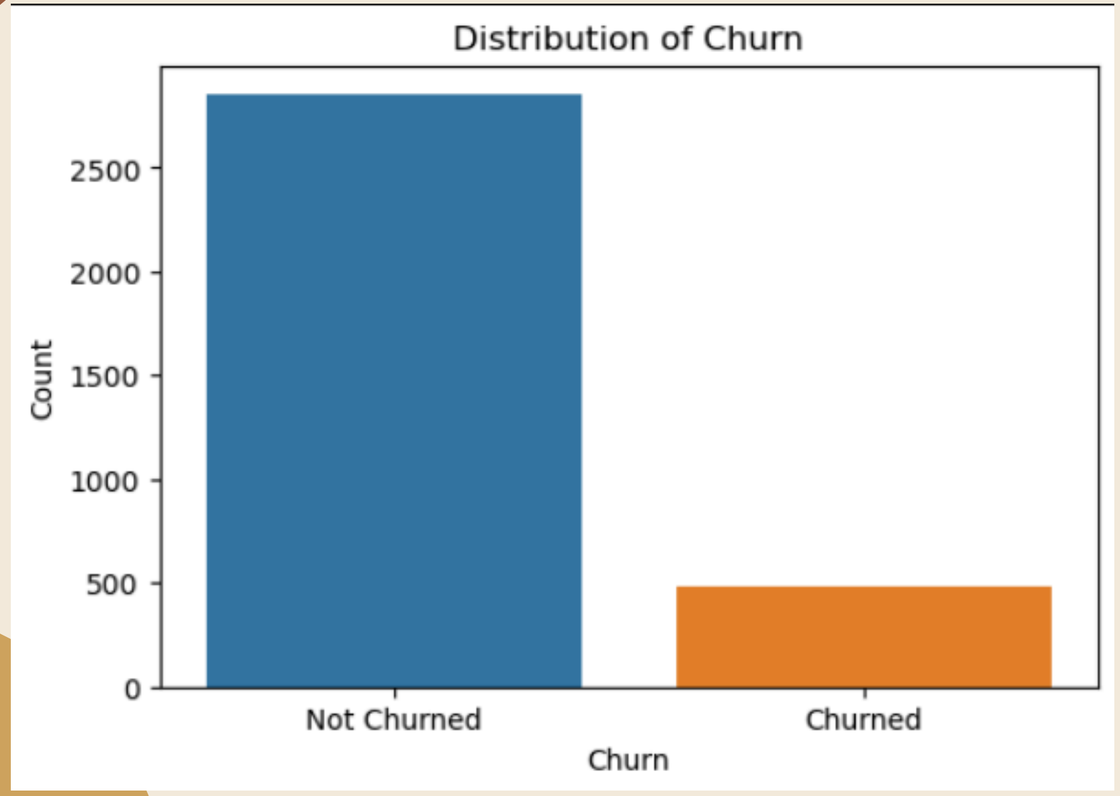
Exploratory Data Analysis

Chapter 1

Distribution of account length



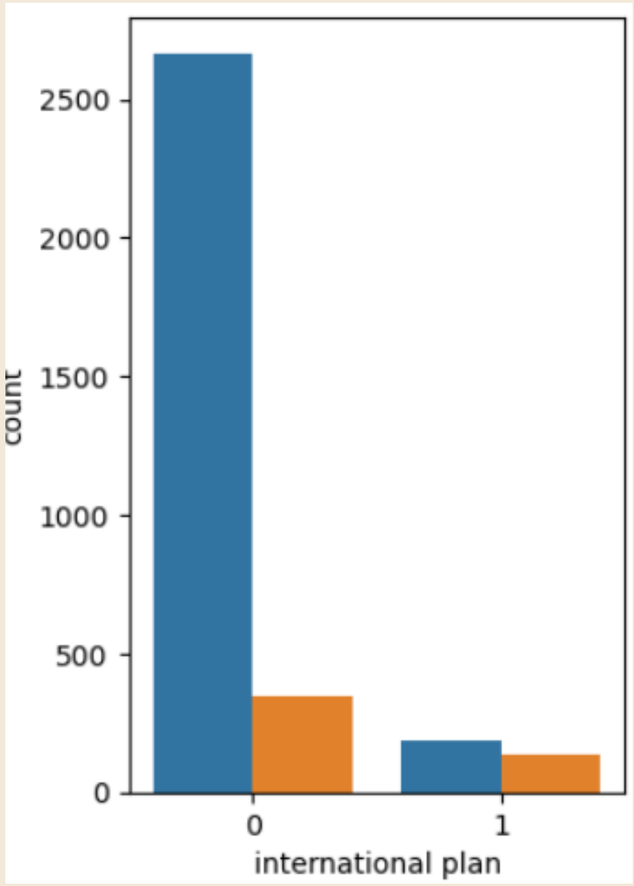
Write Your Subtopic



This helped to check for imbalance which was later on fixed before modelling.

Your paragraph text

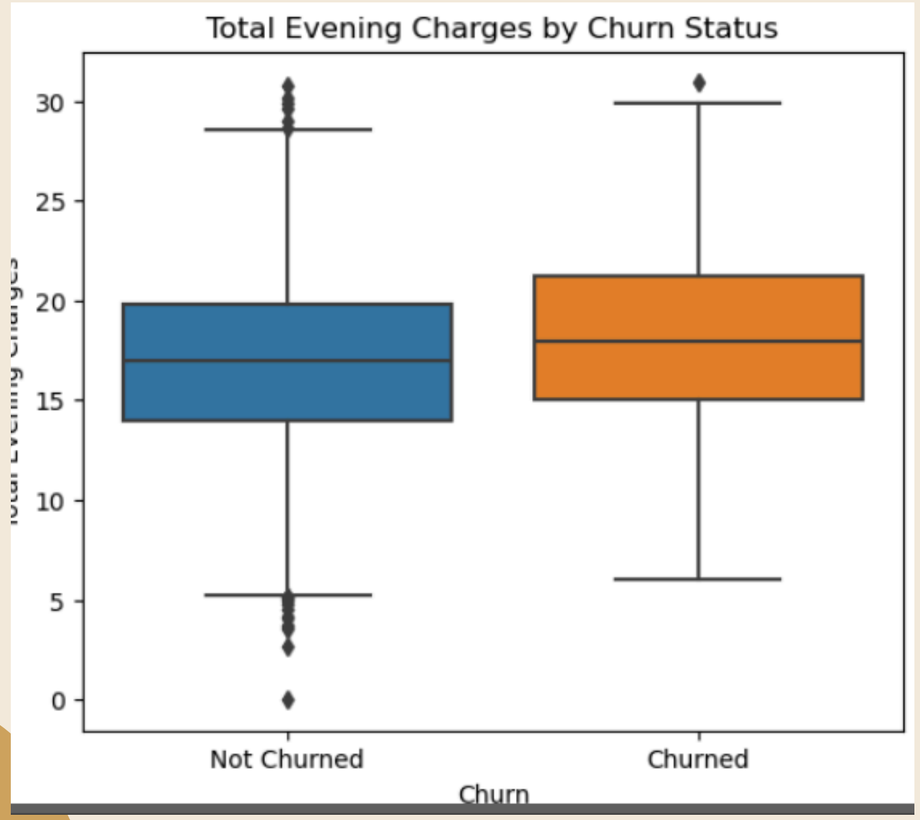
Write Your Subtopic



This was done to check for customers with international plan.

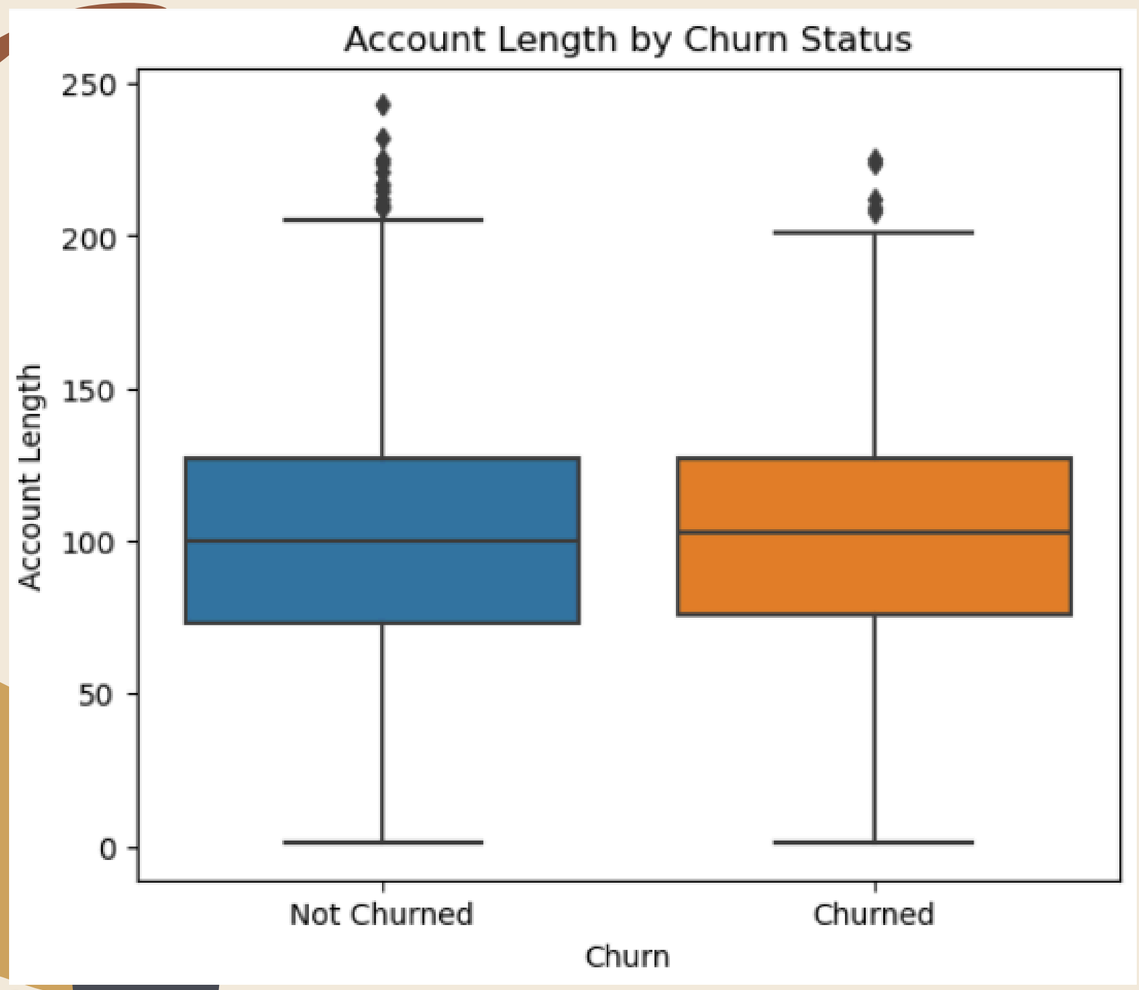
Chapter 2

Total Evening charges by churn



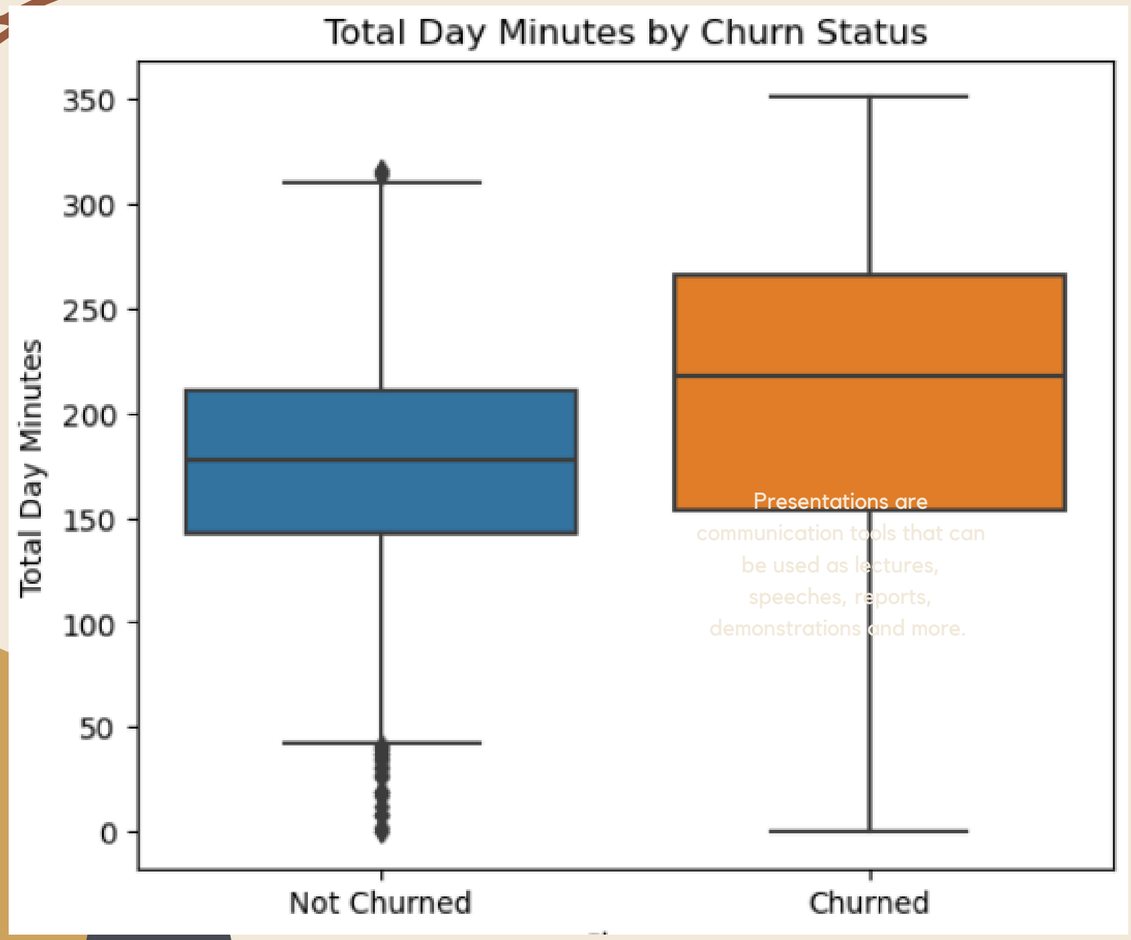
This is a box plot of total evening charges and churn status

Account length by churn status



This box plot shows how account length varies by churn status.

Total day minutes by churn



Distribution of total day minutes used by customer and their churn status.

Chapter 3

Modelling and Evaluation.

Models that were used were logistic regressional model as the baseline model and random forest model as the multiple feature model

Modelling and Evaluation

In evaluating the models confusion matrix was used and also precision, recall and accuracy were compared as shown.

Logistic Regression model

```
Logistic Regression Evaluation:
[[572  27]
 [ 71 527]]
              precision    recall  f1-score   support

       False       0.89      0.95      0.92       599
        True       0.95      0.88      0.91       598

    accuracy                           0.92      1197
   macro avg       0.92      0.92      0.92      1197
weighted avg       0.92      0.92      0.92      1197
```

From the results one can see that the model is performing well..

Chapter 4

Random forest model

```
Random Forest Evaluation:
[[575  24]
 [ 39 559]]
              precision    recall  f1-score   support

       False       0.94      0.96      0.95       599
        True       0.96      0.93      0.95       598

    accuracy                           0.95      1197
   macro avg       0.95      0.95      0.95      1197
weighted avg       0.95      0.95      0.95      1197
```

Same as for this one the model is performing well

Recommendations

1. Priorities features indicated above as most important for predicting churn because this will help in allocation of resources accordinglly to prevent high churn rate in areas.
2. Strategies on how to boost international services of the company.
3. Continuously monitor momdel performance and update it to ensure model remains accurate and relevant.
4.

Thank You.