Phase 3 Project

Heath Rittler

Business Problem

Skyvia is trying to **optimize their retention strategies**, and to do so they are trying to understand what features of different customers indicate churn. The presentation will be made available for the VP of Customer Success, and the Chief Revenue Officer.

*Source : Customer Acquisition vs. Customer Retention: What Data Says?

Data

3,333 records from Skyvia customers

US based customers

21 data attributes

Approach & Goals

Classification model to determine who is likely to churn

Minimize false negatives

Evaluate multiple methodologies, and pick the best one (Logistic Regression, Random Forest, Decision Tree, XGBoost, Ridge)

Final Model

Tuned XGBoost
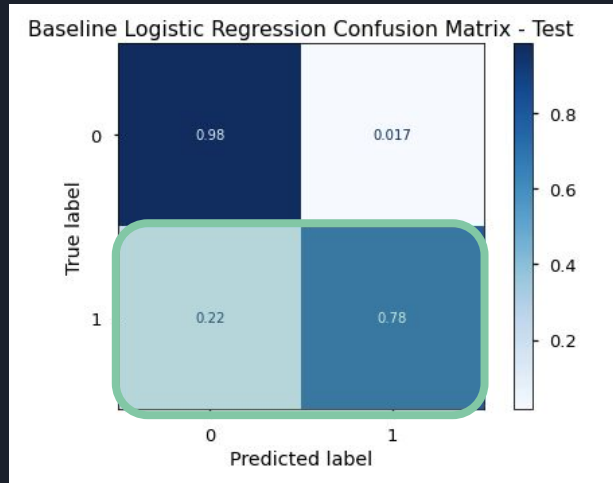
78% Churn Prediction (Recall)

89% Precision


```
Performance Comparison
-------------------------------------------------
Train Roc_Auc Score: 99.99%
Test Roc_Auc Score: 93.80%
-------------------------------------------------
Test Recall score: 78.40%
Mean Cross Validated Recall Score: 93.89%
-------------------------------------------------
Train F1 score: 99.22%
Test F1 score: 83.40%
-------------------------------------------------
Train Precision score: 99.91%
Test Precision score: 89.09%
```


Baseline Logistic Regression Confusion Matrix - Test

To put it simply, precision is the ability to get the highest ratio of at risk customers, recall is the ability to not miss out on high risk customers.
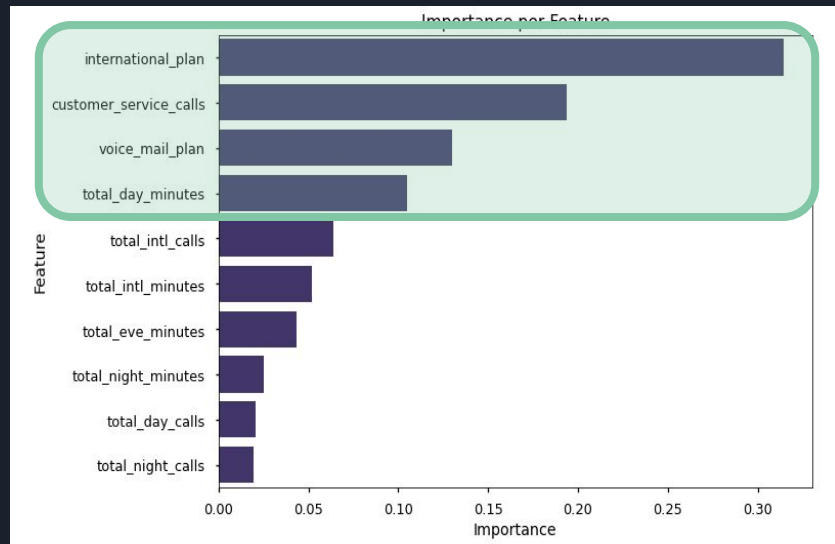
Feature Importance

Product Features

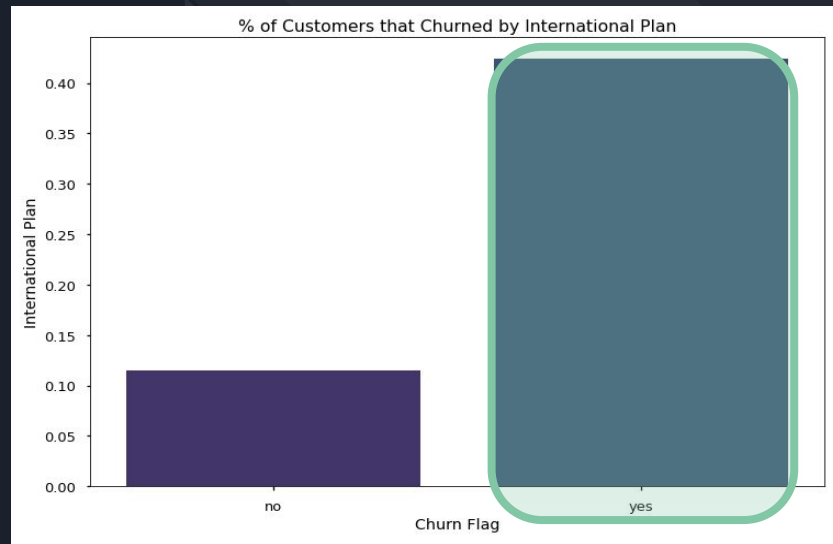- International
- Voice Mail

Usage
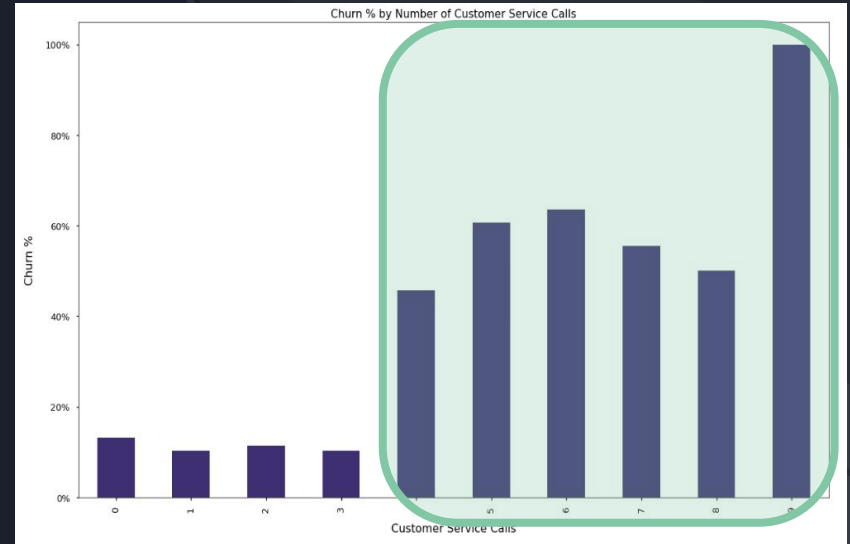
- Total Day Minutes

Account

- Customer Service Calls

4x

More likely to churn

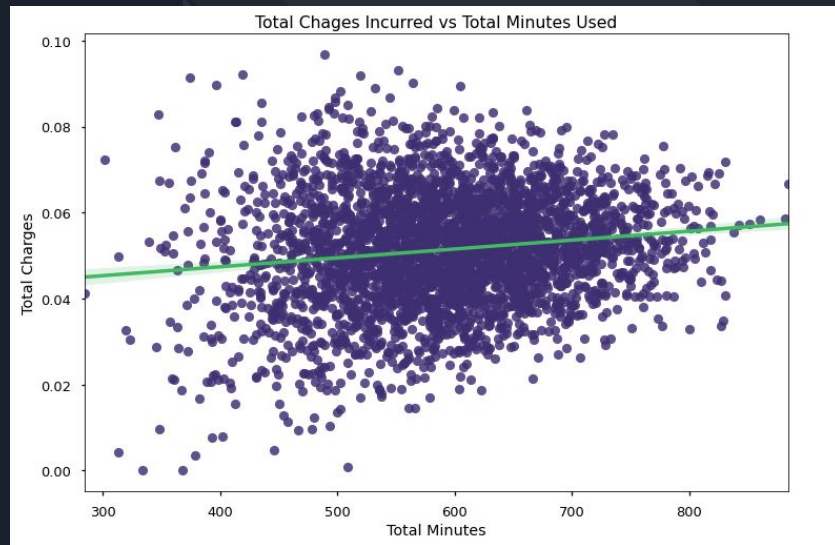% of Customers that Churned by International Plan

50%

Likely to churn once reaching >= 4 customer service calls

$0.05

Per minute average for customers, with churn customer paying close to 10% more

Total Chages Incurred vs Total Minutes Used

Recommendations

Survey customers that have the international plan and the voicemail plan

Once a customer reaches 3 customer service calls, **flag** the account as at risk.

Evaluate pricing strategies