SyriaTel Churn Prediction

By Jinsol Cantrall

Contents

- Overview
- Business Problem
- Data Understanding
- Result
- Next Steps

Overview

SyriaTel is expanding its business and wants to increase its customers. However, before expanding, SyriaTel wants to find out if the company could stop the customers who are canceling the subscription. With the datas that are given, is it possible to know the reason for the canceling the service and predict the churn rate(a measurement of the percentage of accounts that cancel). If so, can the company improve the service and stop customers from canceling?

Business Problem

① What is current churn rate?

② What are the top 5 reasons that people are canceling?

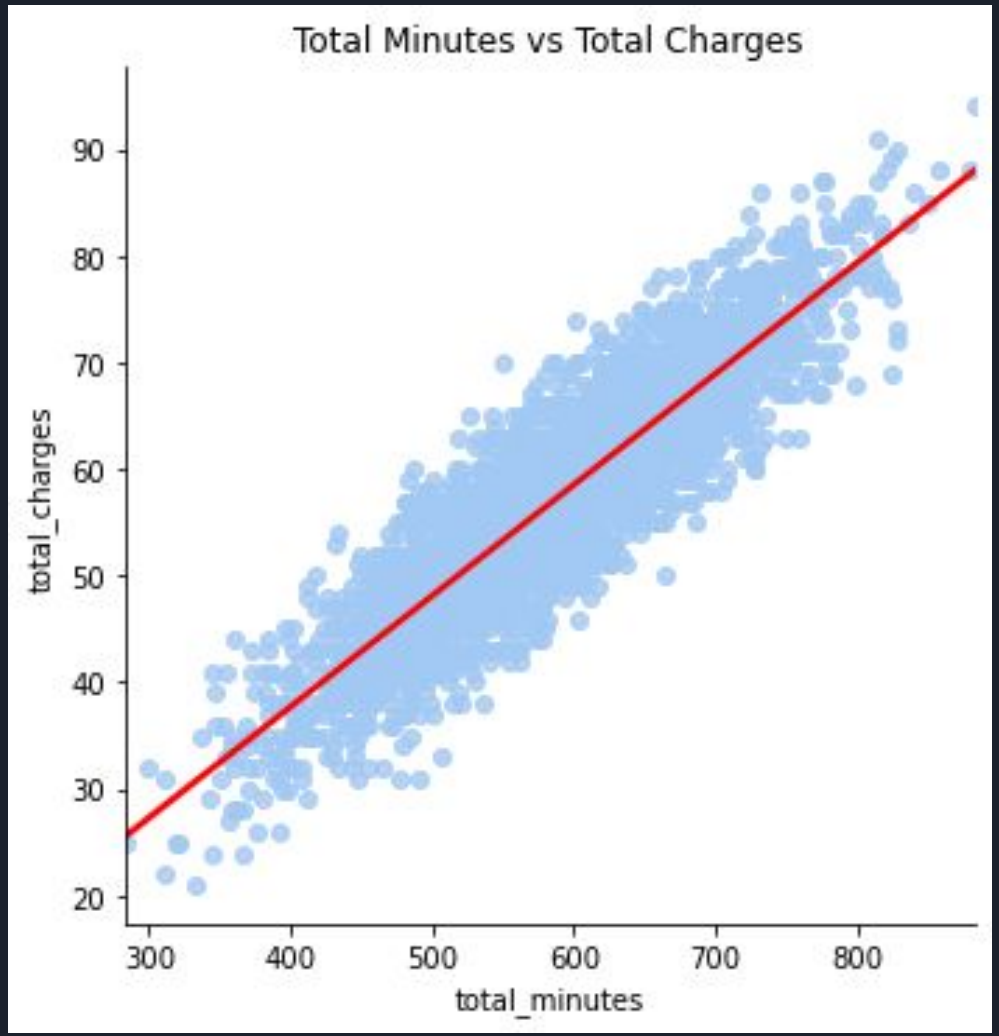
③ What are the recommendations to improve the satisfaction?

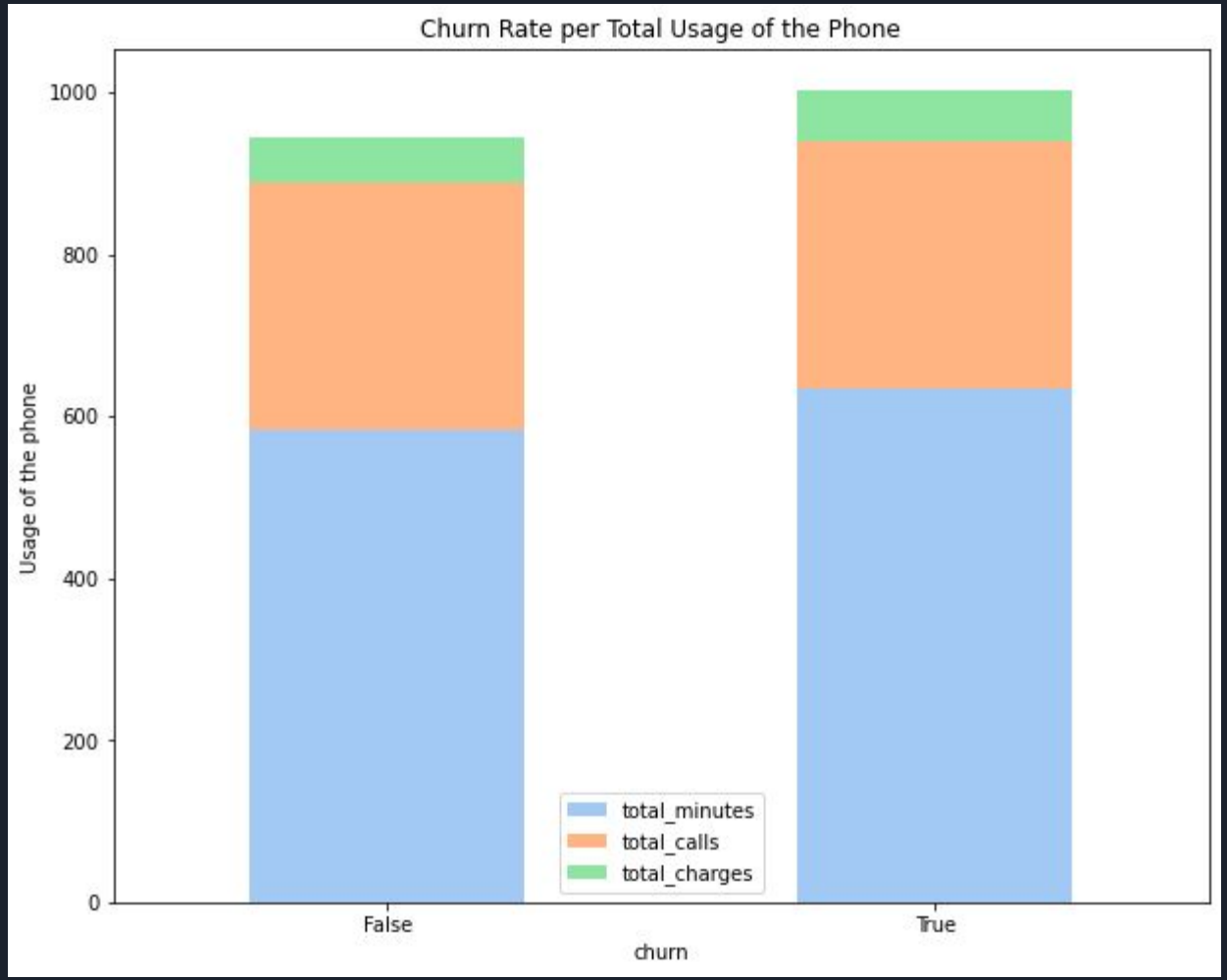
④ Should the business focus on getting new customers?

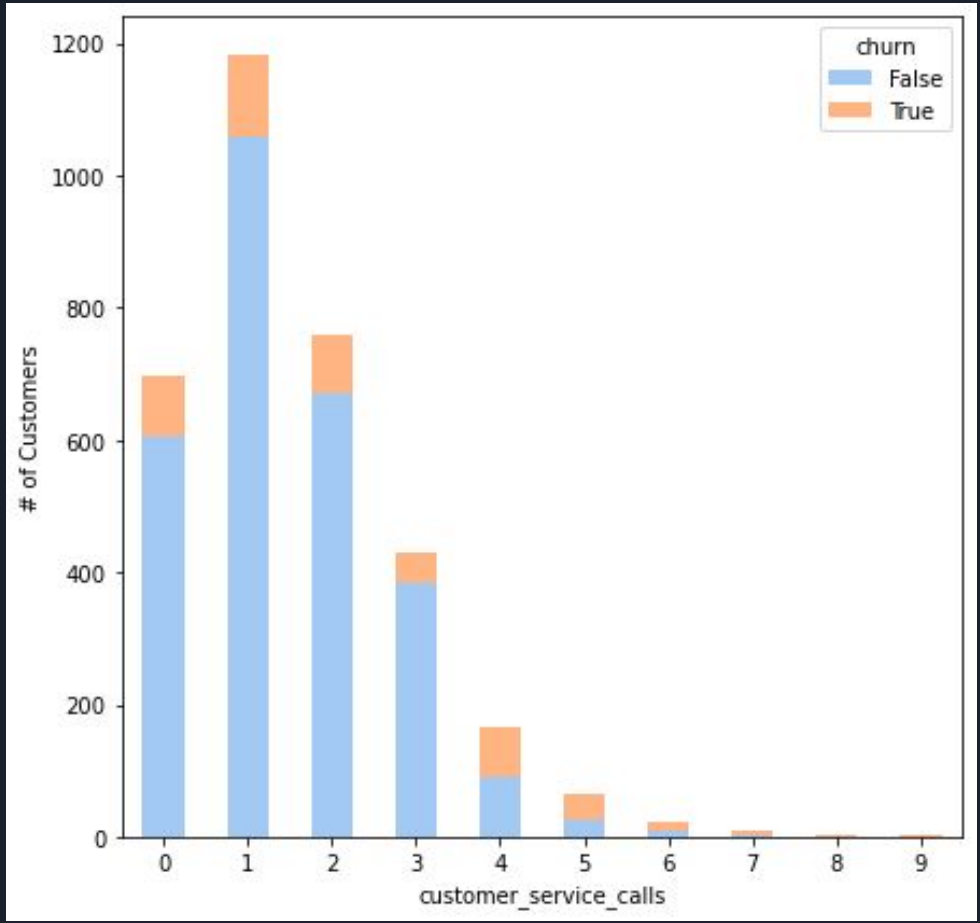
⑤ Can the business predict customers who are staying and leaving?

Dataset

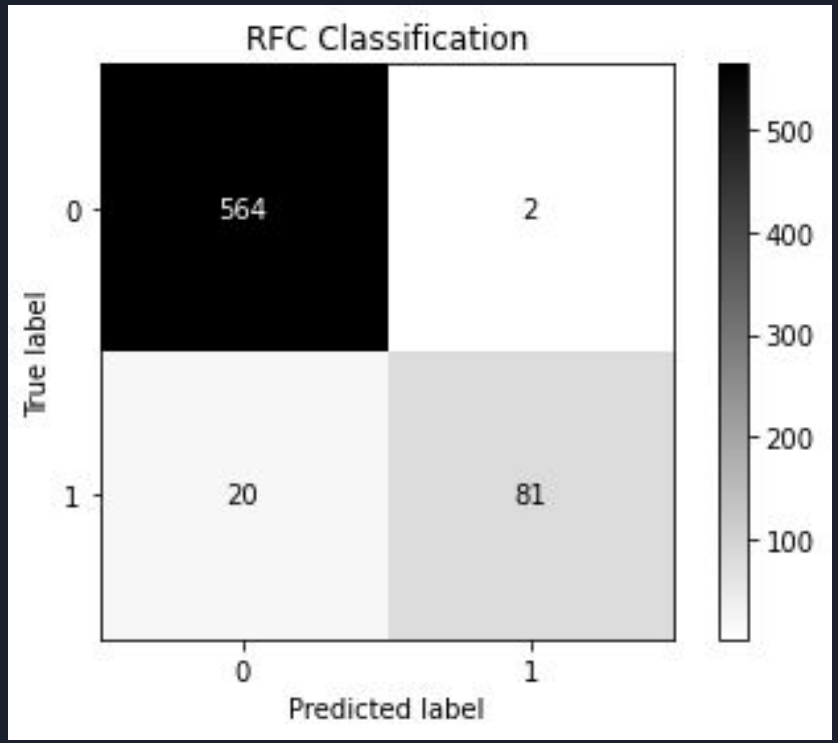
- SyriaTel dataset was obtained from https://www.kaggle.com/becksddf/churn-in-telecoms-dataset
- Dataset contained 21 columns and 3333 rows
- Target Variable will be ‘Churn’
- 5 Categorical Variables
- 1 Unique Identifier


Total Minutes vs Total Charges
total_charges
total_minutes
20
30
40
50
60
70
80
90
300
400
500
600
700
800


Churn Rate per Total Usage of the Phone
Usage of the phone
1000
800
600
400
200
0
False
True
churn
total_minutes
total_calls
total_charges


churn
False
True
# of Customers
customer_service_calls
0
200
400
600
800
1000
1200
0
1
2
3
4
5
6
7
8
9

Best Model
Random Forest Classification



- Predicted that 564 customers are staying and it is true.(True Positive)
- Predicted that 81 customers are going to churn and it is true. (True Negative)
- Predicted that 2 customers are staying but it is false.(False Negative)
- Predicted that 20 customers are leaving but it is false. (False Positive)

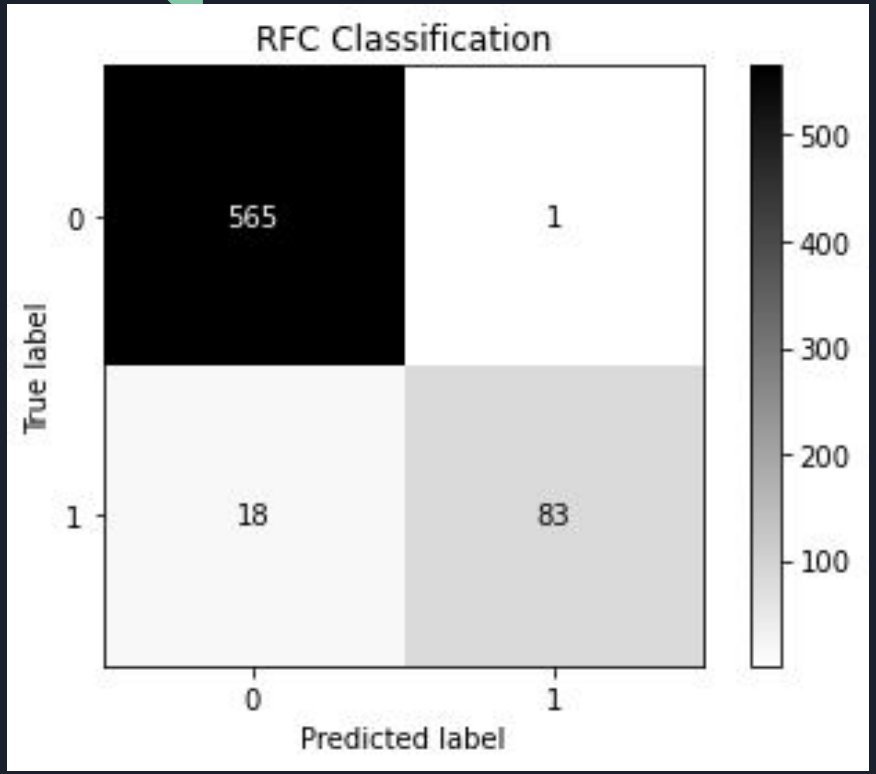
Top 5 Features

1. total_charges 0.178446
2. customer_service_calls 0.131498
3. total_minutes 0.077979
4. total_day_minutes 0.074535
5. total_day_charge 0.050467

Bottom 5 Features

21. 2total_night_charge 0.013311
22. total_intl_charge 0.011104
23. area_code_510 0.003812
24. area_code_415 0.003065
25. area_code_408 0.002974

Final Model



- Got rid of bottom 3 features.
- Predicted that 565(564 before) customers are staying and it is true.(True Positive)
- Predicted that 83 (81 before) customers are going to churn and it is true. (True Negative)
- Predicted that 1 (2 before) customers are staying but it is false.(False Negative)
- Predicted that 18 (20 before) customers are leaving but it is false. (False Positive)

Result

	precision	recall	f1-score	support
0	0.97	1.00	0.98	566
1	0.99	0.82	0.90	101
accuracy			0.97	667
macro avg	0.98	0.91	0.94	667
weighted avg	0.97	0.97	0.97	667

- The Random Forest Classifier model is 96.85% accurate in classifying churn of the customer.
- It predicted correctly that 564 customers are staying and 83 customers are going to churn and predicted incorrectly that 2 customers are staying and 18 customers are leaving.

Next Steps

This dataset is too small to generalize everything. Improving the size of dataset would give the business better idea of the problems. Collect more information about the customers and see if churn rate could be improved.