

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light greenish-blue. They are positioned diagonally, with the blue one partially covering the green one.

# SyriaTel Customer Churn

Eric Cusick

# Business Case

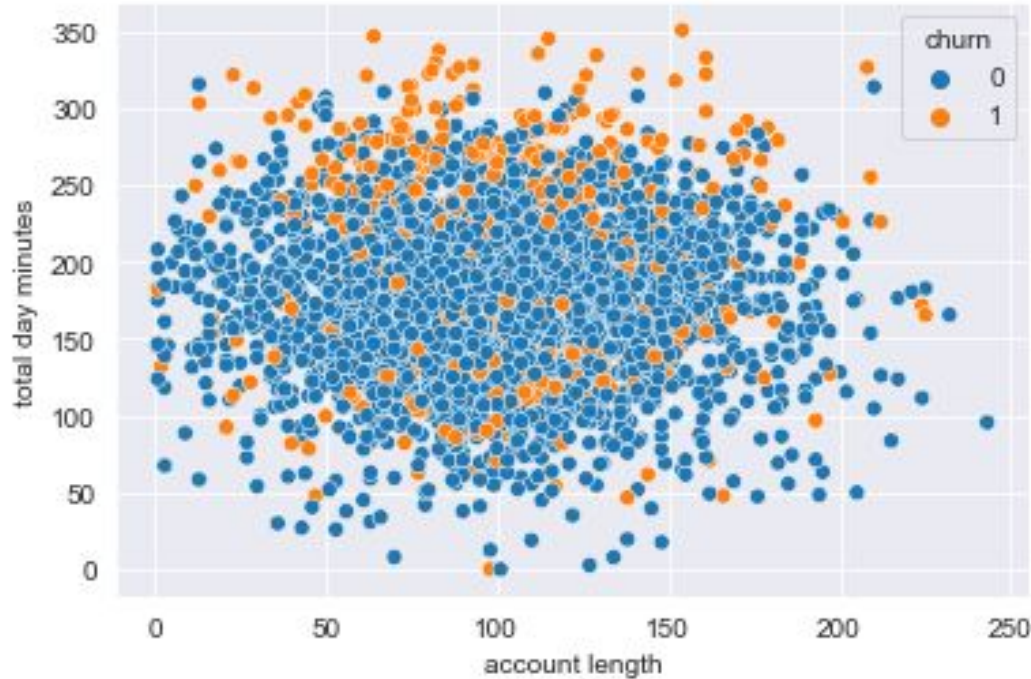
A telecommunication company call SyriaTel has enlisted the help of a data scientist to create a model that predict when the customers are leaving their business.

Given the 'customer\_churn.csv' file the data scientist is expected to:

- Build an accurate model
- Report features that may be an indicator that the customers will be more likely to leave



# Exploratory Data Analysis



The Scatterplot shows the relationship between the account length and total day minutes used by each customers.

There is a higher concentration of customer churning towards the top of the graphs which could indicate the cost may not be as agreeable for some people

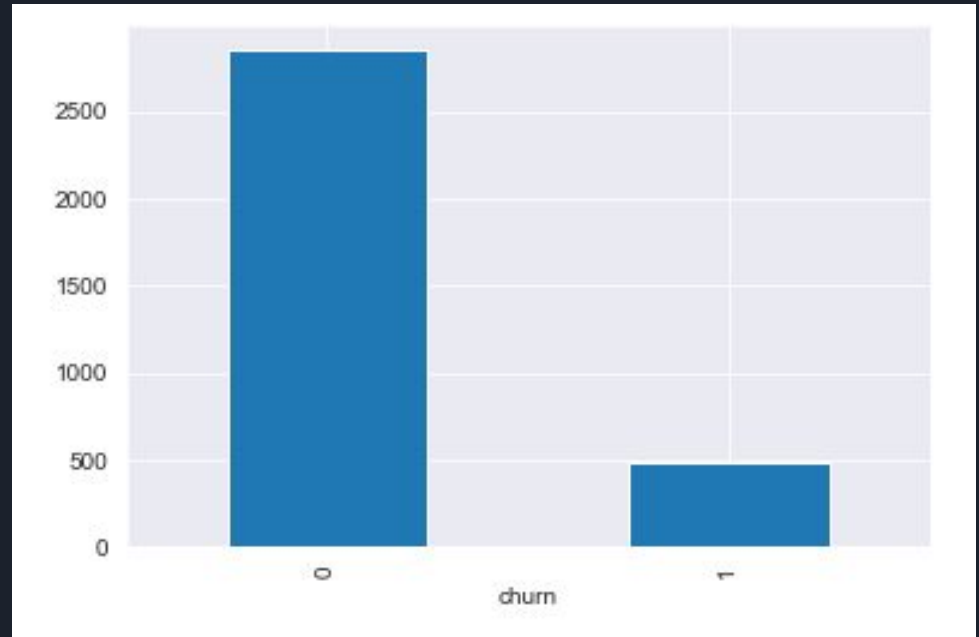
# Exploratory Data Analysis

There is a trend in the graph that shows that the more times the customers has to call for services the more likely that they will leave.



# Class Imbalance

There is a problem with this dataset, out of over 3000 customers, almost 500 has left the company. This shows a significant class imbalance the first model done had a high accuracy score however the recall score was low.



# Model Process Random Forest

During the model process for Random Forest:

- I fitted SMOTE to the training data from the train test split of the original data to improve the class imbalance
- I took advantage of the Grid Search to find the best parameters for the Random Forest Classifier

```
Before SMOTE:  
0    2284  
1     382  
Name: churn, dtype: int64
```

```
After SMOTE:  
1    2284  
0    2284  
Name: churn, dtype: int64
```

```
Optimal Parameters:  
  
criterion : entropy  
  
max_depth : None  
  
min_samples_leaf : 2  
  
min_samples_split : 5  
  
n_estimators : 100
```

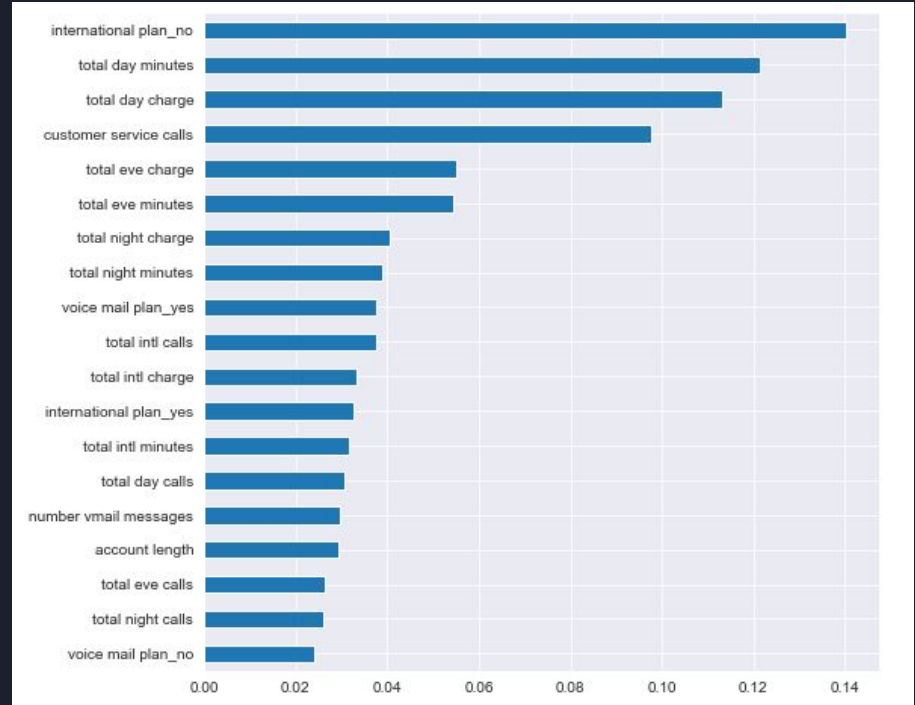
# Final Random Forest

The feature importances graph is sorted by how strongly the factors influence customer churn.

Below is the final metric score of the Random Forest Model

```
Precision Score: 0.8282828282828283  
Recall Score: 0.8118811881188119  
Accuracy Score: 0.9460269865067467  
F1 Score: 0.8200000000000001
```

Feature Importances



# Recommendations

Looking at the model, I would recommend:

- Increase the quality of customer services to decrease the amount of time called
- Provide decreasing charge rate after hitting a cap of 240 call minutes instead of a flat rate per minutes
- Give out some unique benefits or rewards for customers who does not have an international plan. This can be done by account length of stay

The SyriaTel communication company will be able to utilize this information to make decision that best serve their interests in customer retention.





# Future Work



- Gather more data; Monthly Costs, Family Plans, Data Services, etc. to see if there is any other unique factors that affects churn.
- Make 24 hour columns that includes day, evening, and night to see if total accumulation of charge, calls, and minutes could show a more impactful outcome of customer churning.
- Take a look at locations, see if one state has more churn than other which could be a sign of poor cellular coverage.



# Thank You

Thank you for giving me your time  
during my presentation today.