Customer Churn Analysis

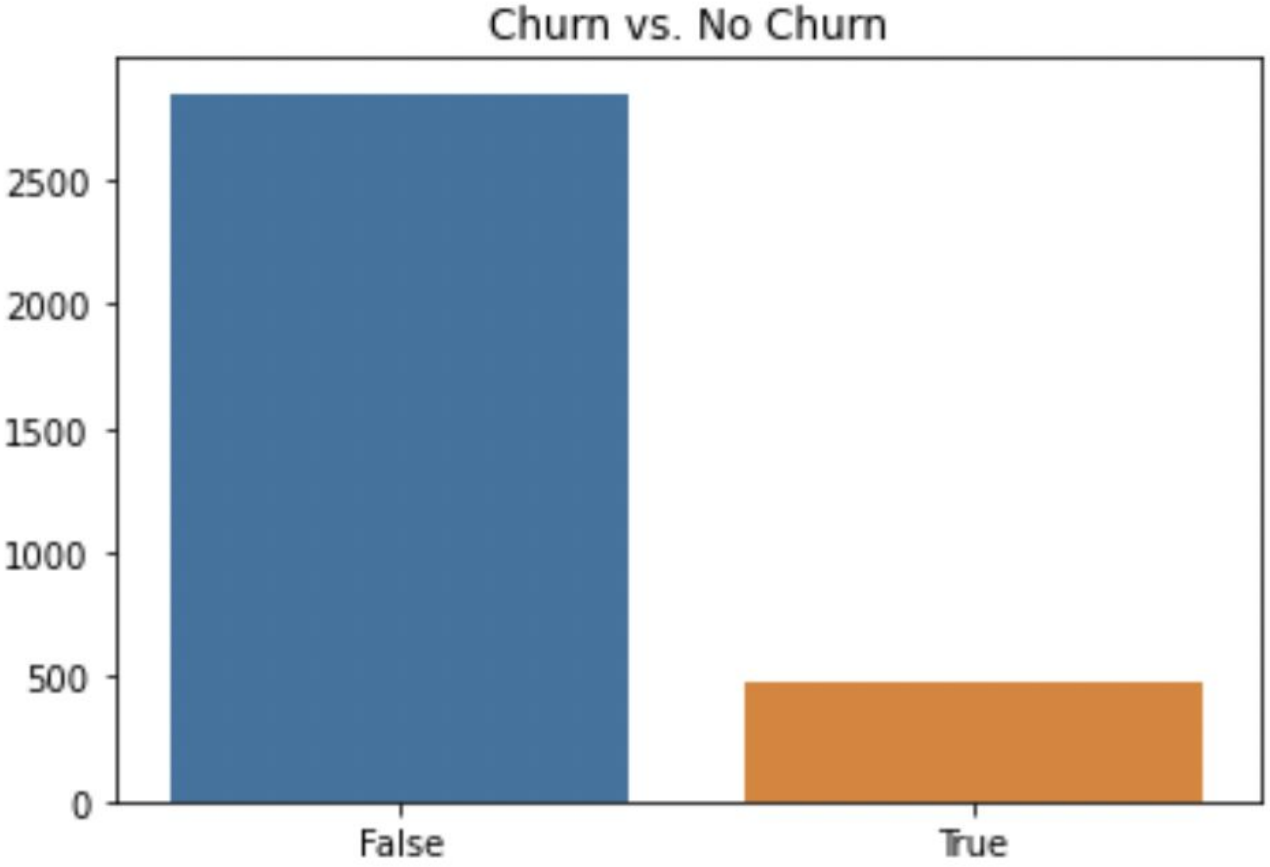
Kregg Jackson

Business Context

- The goal of this project is to show stakeholders what factors have the greatest impact on if a customer will churn
- With analysis of the customer data I will be able to identify which customers are most likely to churn

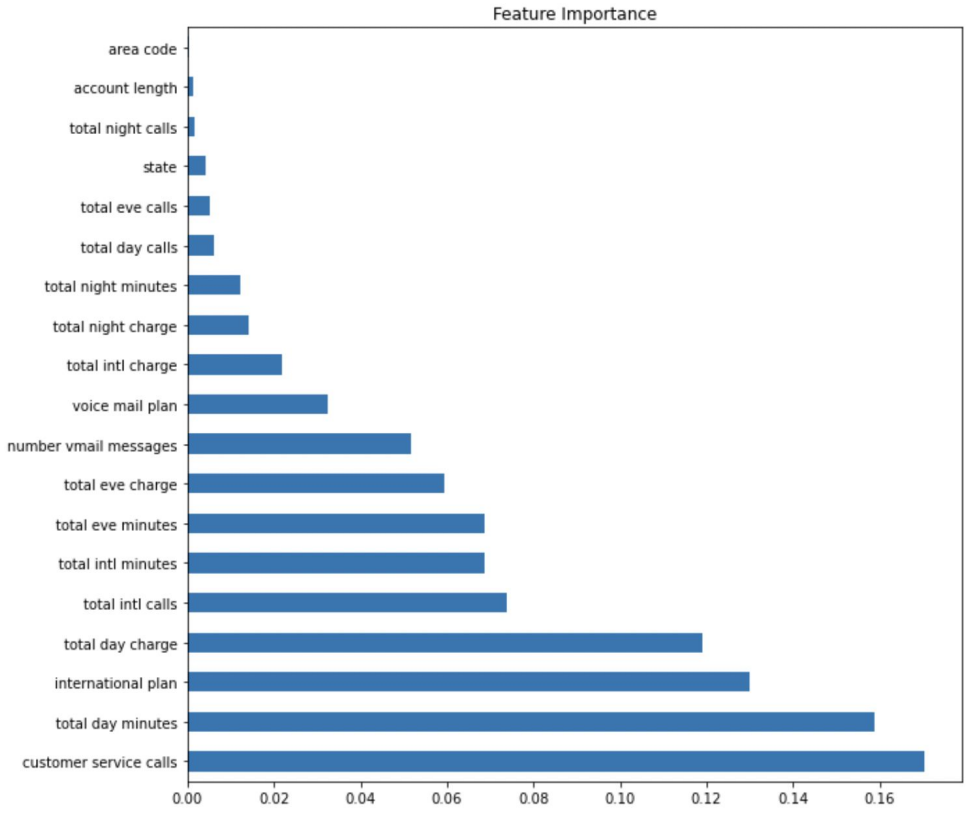
Phone Company Data

- The phone company's data frame originally consisted of 21 columns with about 3,300 rows
  - Some of the key columns were :
    - 'International plan'
    - 'Customer service calls'
    - 'Total day minutes'
- My first step was to explore and clean the data and during my EDA I learned the independent variable has class imbalance


Churn vs. No Churn
2500
2000
1500
1000
500
0
False
True

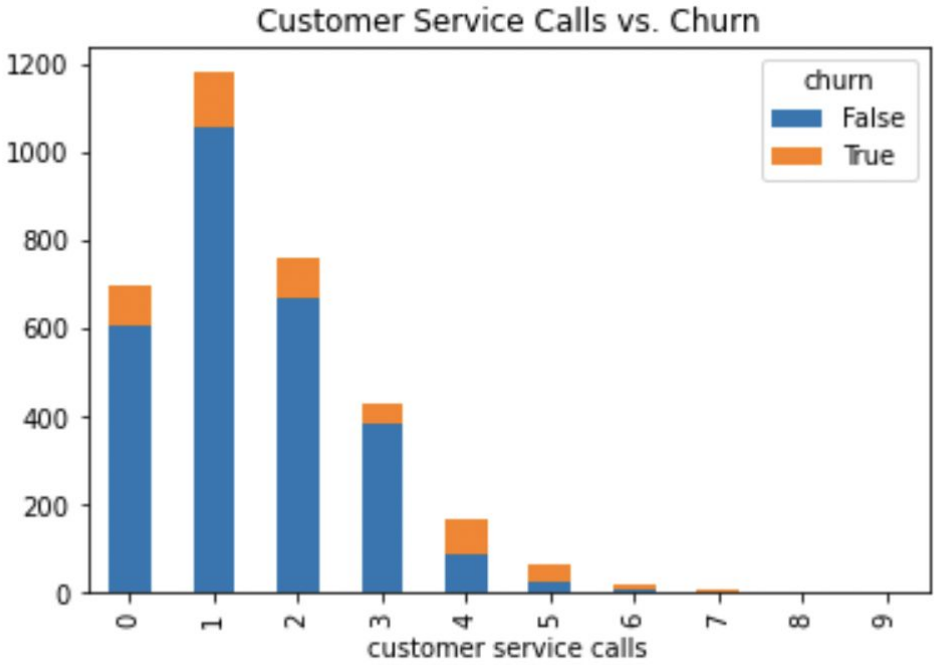
Process Steps & Results

- Since the churn data was imbalanced my next step was to balance the dependent variable
- Then I built my models starting with a dummy classifier, I analyzed the classification reports until I got my best performing model
- The most important features are ‘customer service calls’, ‘total day minutes’, and ‘international plan’



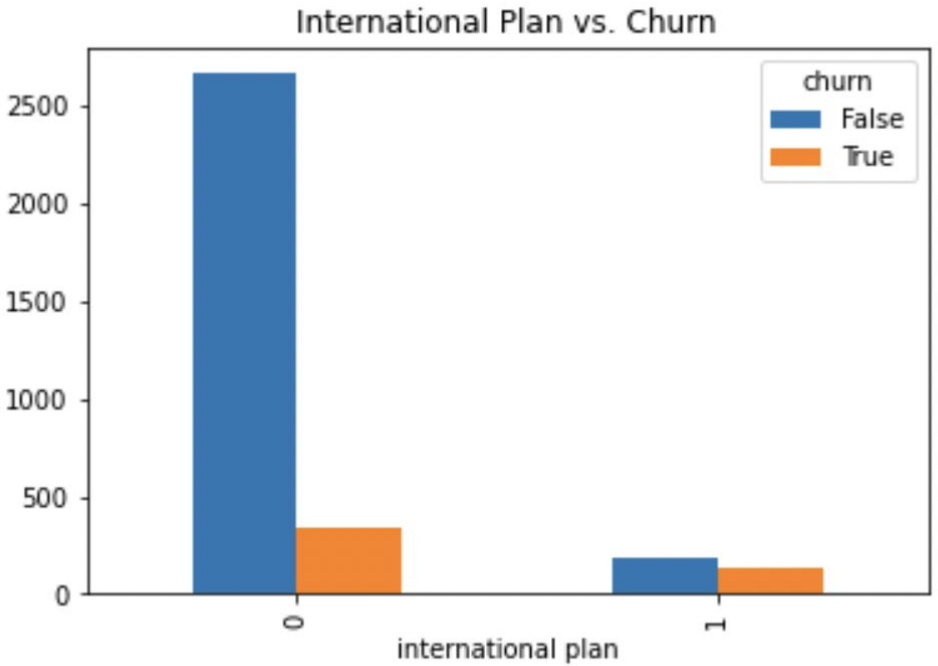
Customer Service Calls

- The gradient boosting model reported customer service calls was the most important feature
- Once customers make four customer service calls it is much more likely that they cancel their subscription
- This could be due to a problem with the phone company's customer service department



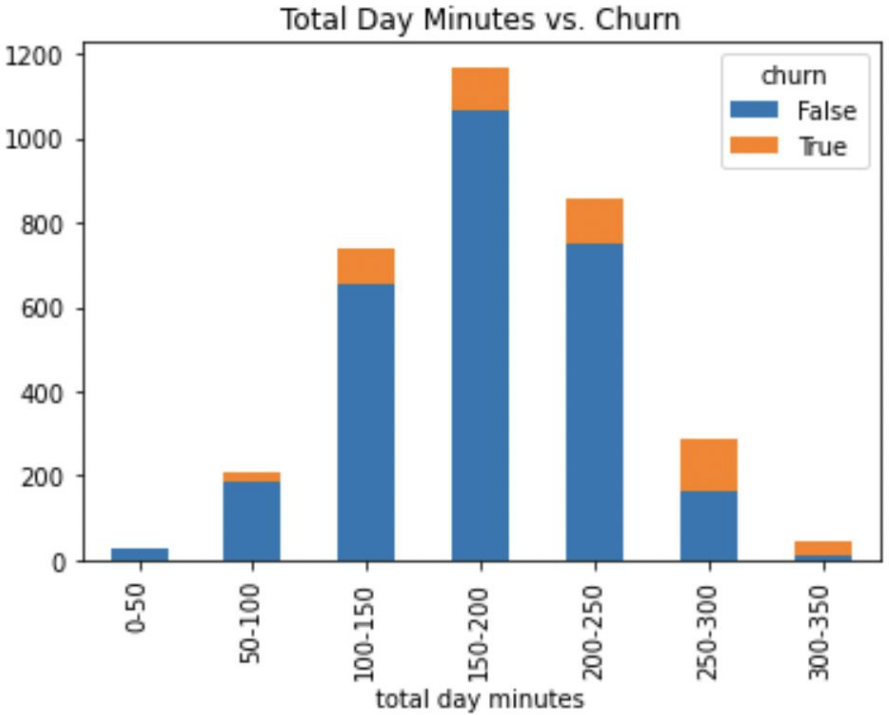
International Plan

- The attribute with the third highest feature importance score is international plan
- Almost half of the customers who opt in for the international plan will cancel their subscription
- This could be due to the international services being overpriced



Total Day Minutes

- The feature with the second highest score is total day minutes
- When customer pass 250 minutes per day it is much more likely they end their subscription
- Customers who use this many minutes may feel the plans available don't suit them, I'd suggest adding an unlimited minutes phone plan option



Conclusive Evaluation

- Based on the feature importance score of my model the most important feature is customer service calls
- I advise the phone company to invest money in improving the customer service department
- I also recommend lowering the price of the international plan and offering an unlimited minutes option

Future Improvements

- Obviously more time for analysis and data to analyze would make the models more accurate
- I could run cross validation on the churn data to correct the class imbalance differently and see if that is more accurate than class weight and SMOTE
- Remove more redundant columns the ‘charge’ and ‘minutes’ columns are too correlated

Thank You

Email: kreggthegoat@gmail.com

Github: @kreggthegoat