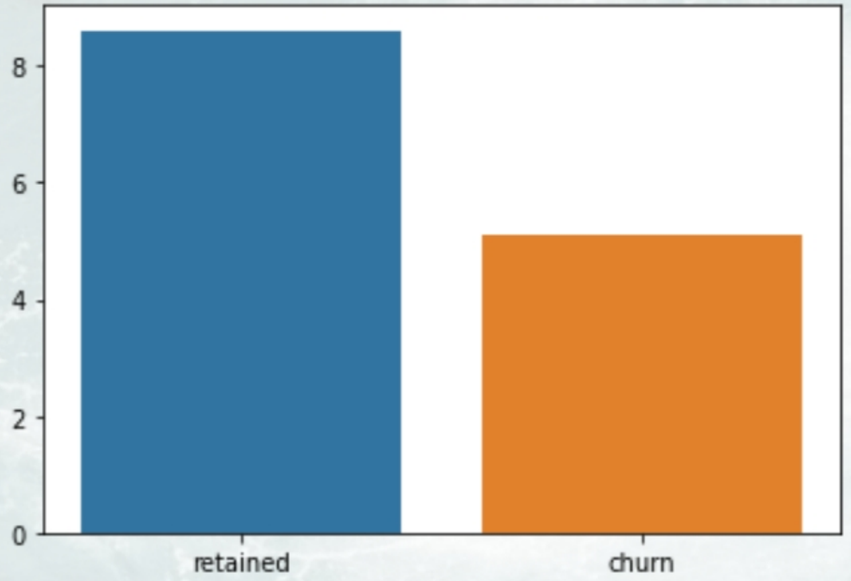
# Predicting Customer Churn for SyriaTel

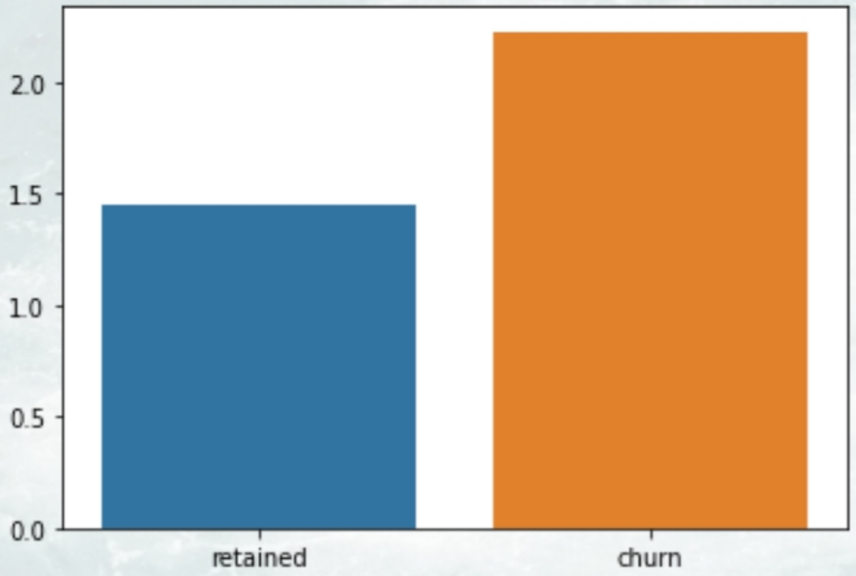
# Voicemails

The average churn customer makes 5 voicemails and the average retained customer makes 9.



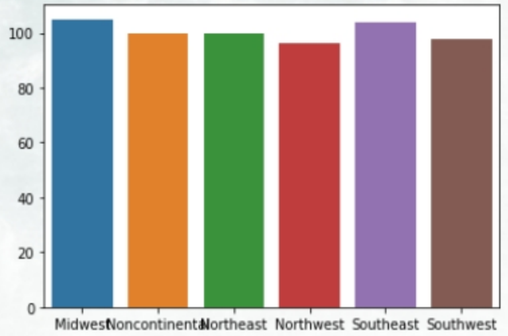
# Biggest correlation

Churn customers make more calls to customer service. 56% of customers who leave make 2 or more customer service calls, compared to 42% for those who stay.



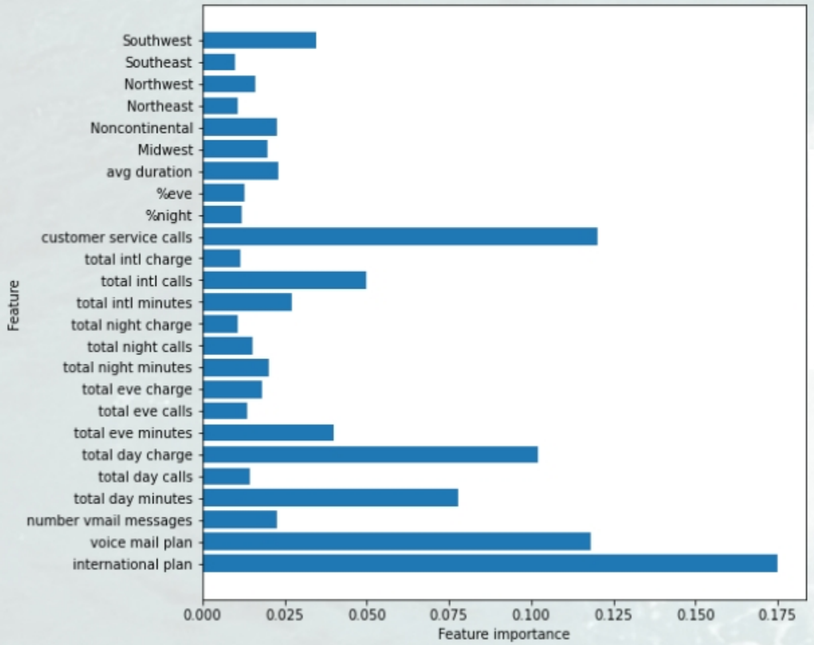
# EDA Conclusions

The data shows that customers who leave, on average, make more customer service calls, make fewer but longer calls, and fewer voicemails than the customers who stay. No strong correlations were found between region, however customers who leave and live on the west coast make fewer calls than any other group (combination of region and churn or not). From this we can conclude that region does not play a significant role in churn, and providing quality customer support should be a priority because most customers who leave make 2 or more calls to customer service.




# What features are important?

This graph shows the feature importance of all of the features fed into the original model, excludes phone number and area code




$$normfi_i = \frac{fi_i}{\sum_{j \in all\ features} fi_j}$$

# The Model

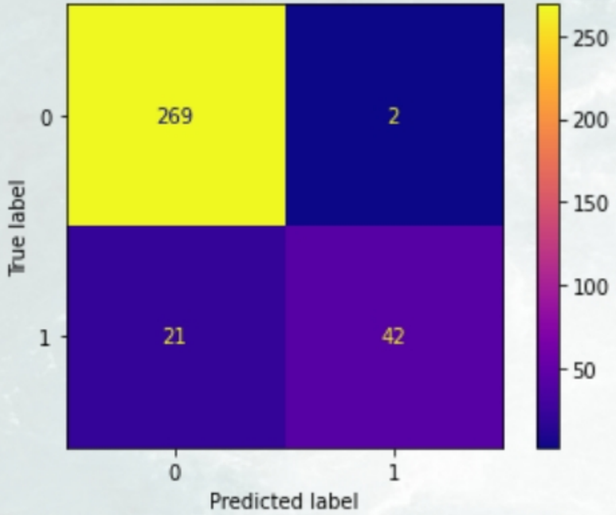
I chose the Machine Learning Model by training several models on the raw data and picking two that had high test accuracy, but more importantly test accuracy and train accuracy being close together. I ended up going with the Gradient Boosted Random Forests (GB and XGB).

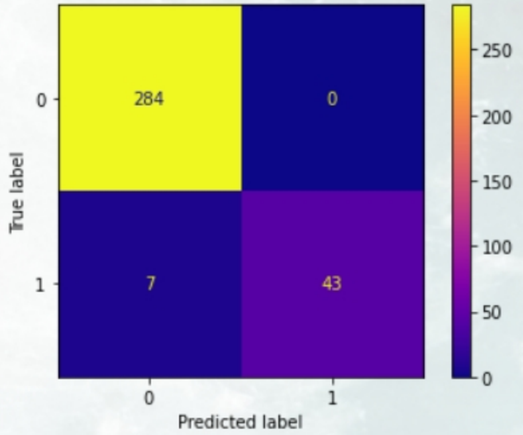
Out[11]:

| | Model | train_acc | test_acc |
|---|---|---|---|
| **0** | KNN | 0.915073 | 0.878667 |
| **1** | NB | 0.878613 | 0.874667 |
| **2** | DT | 1.000000 | 0.909333 |
| **3** | RF | 1.000000 | 0.960000 |
| **4** | AB | 0.897288 | 0.881333 |
| **5** | GB | 0.976434 | 0.954667 |
| **6** | XGB | 0.969320 | 0.960000 |
| **7** | Log | 0.876834 | 0.882667 |
| **8** | SVM | 0.943530 | 0.916000 |

# The Final product

After tuning the model and making modifications to the data. I had a model that had perfect accuracy and recall, correctly labeling positives. No tuning, trained on raw data... tuned model, data has been normalized classes are balanced...



# Future Work

*Following this project it would be great to make a classifier that could help customers figure out what plan is best for them. This could offer a personal touch and likely prevent churn.*

# The End