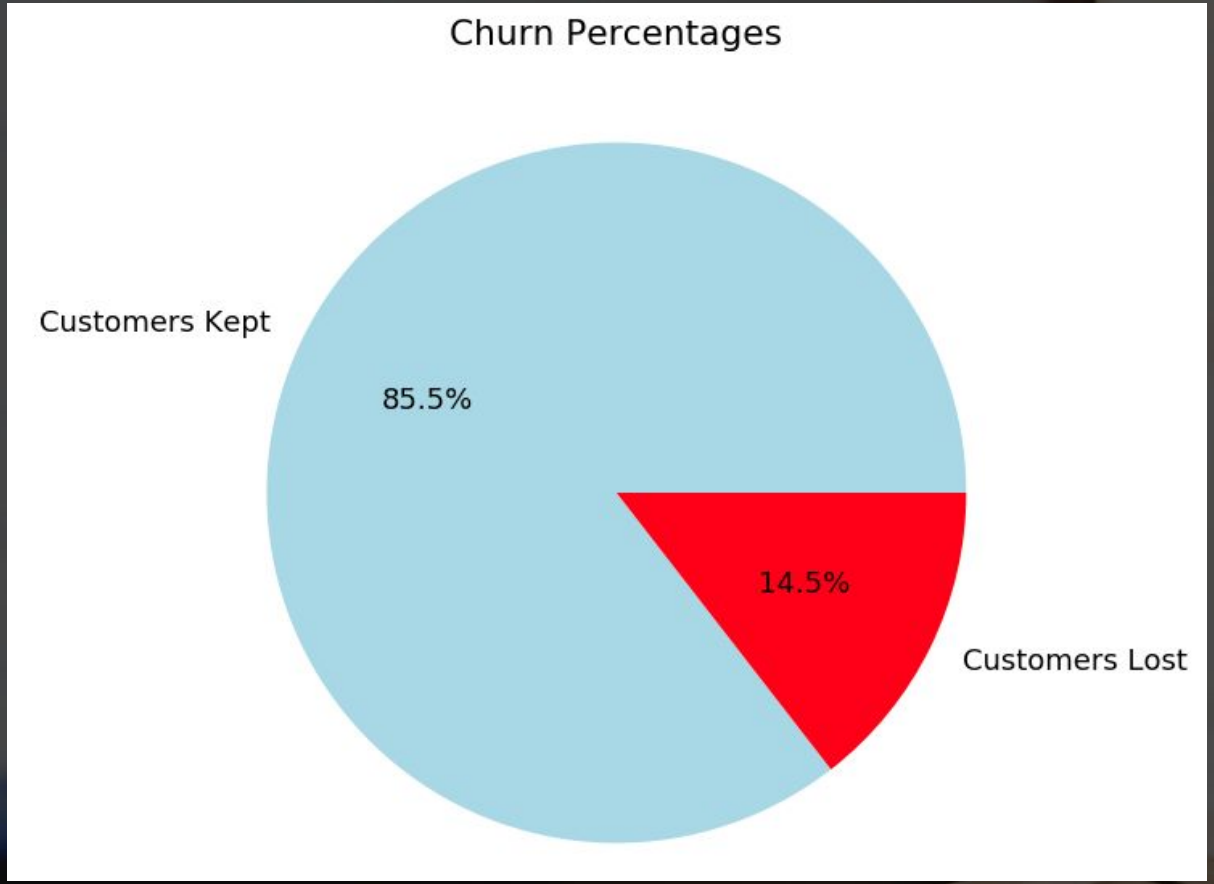
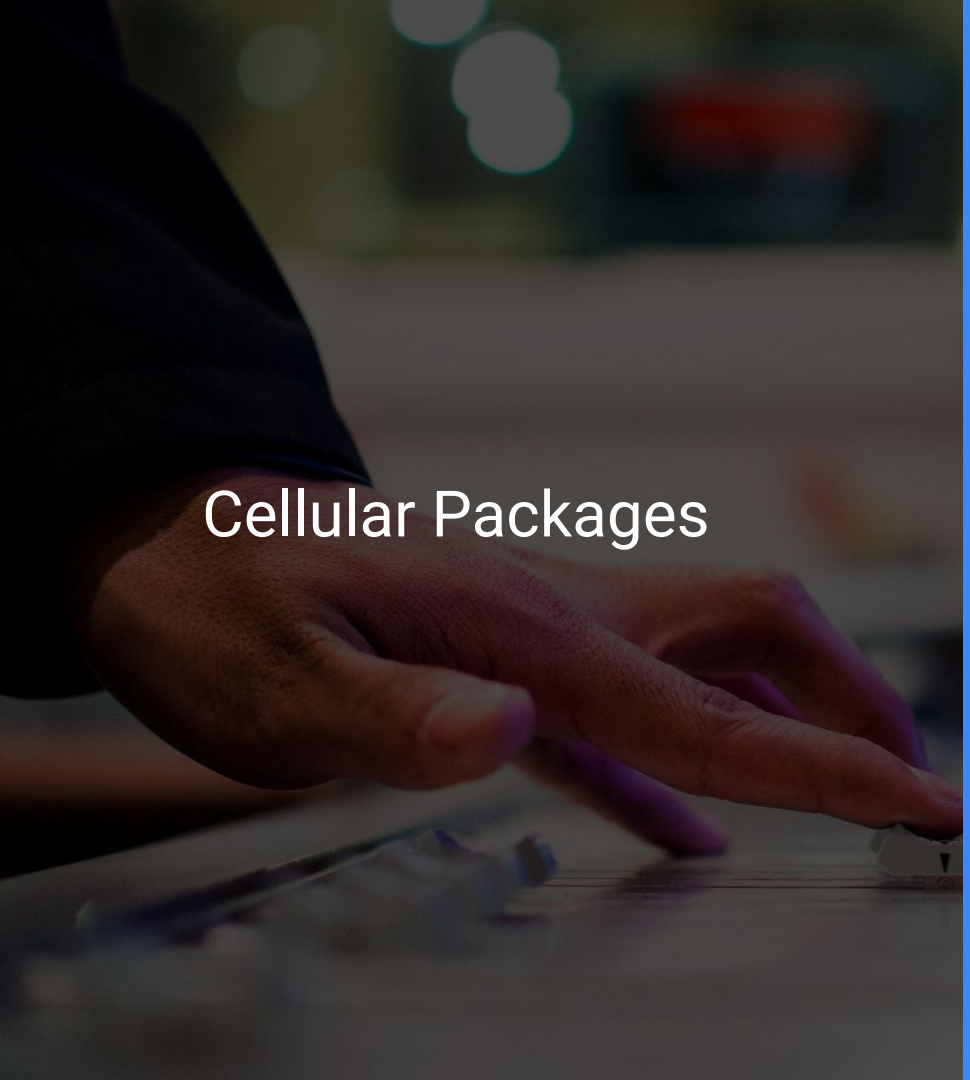
Churn at SyriaTel

Mod 3 Project

Alex H. Macy

SyriaTel: A telecommunications company concerned with their rate of churn.


Churn Percentages
Customers Kept
85.5%
14.5%
Customers Lost



Cellular Packages

Customers across the US

- International Plans
- Voicemail Plans
    - Minutes
    - Charges

Logistic Regression

Feature Selection

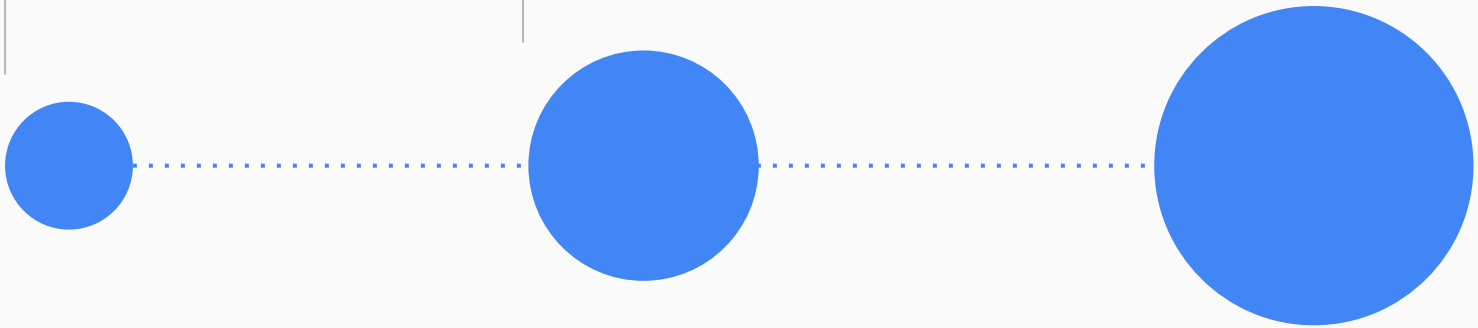
Statistical significance at the P > 0.05 level

Normalizing Data

Standardized scales; class corrections with SMOTE.

Cost-Benefit Analysis

Confusion Matrix, ROC curves, AUC scores, Ensemble Methods

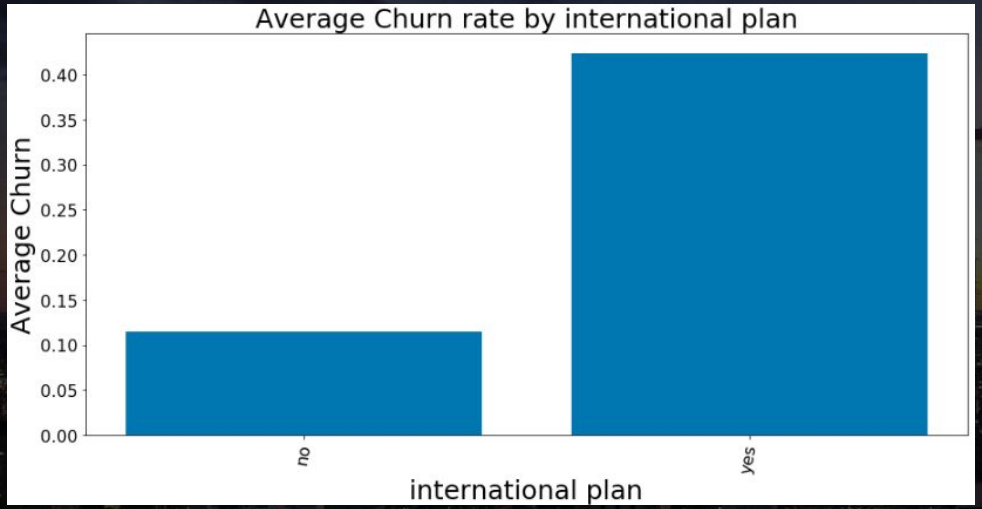


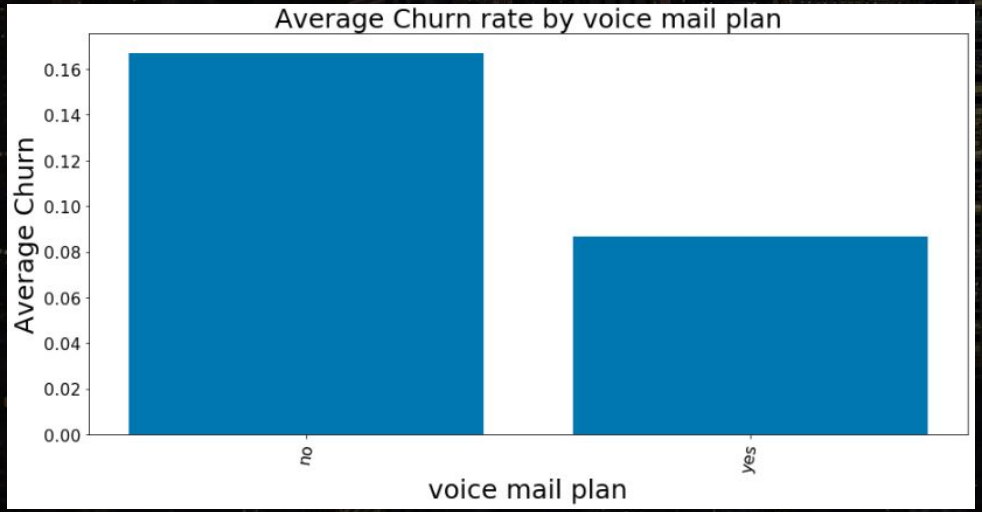
Feature Selection:

```
Optimization terminated successfully.
         Current function value: inf
         Iterations 7
```

Dep. Variable:	churn	**No. Observations:**	3333
Model:	Logit	**Df Residuals:**	3327
Method:	MLE	**Df Model:**	5
Date:	Thu, 02 Apr 2020	**Pseudo R-squ.:**	inf
Time:	08:57:45	**Log-Likelihood:**	-inf
converged:	True	**LL-Null:**	0.0000
Covariance Type:	nonrobust	**LLR p-value:**	1.000

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	-2.3652	0.139	-16.967	0.000	-2.638	-2.092
international plan	1.9657	0.136	14.460	0.000	1.699	2.232
voice mail plan	-1.8059	0.534	-3.381	0.001	-2.853	-0.759
number vmail messages	0.0323	0.017	1.924	0.054	-0.001	0.065
total intl calls	-0.0754	0.023	-3.214	0.001	-0.121	-0.029
customer service calls	0.4418	0.037	12.090	0.000	0.370	0.513

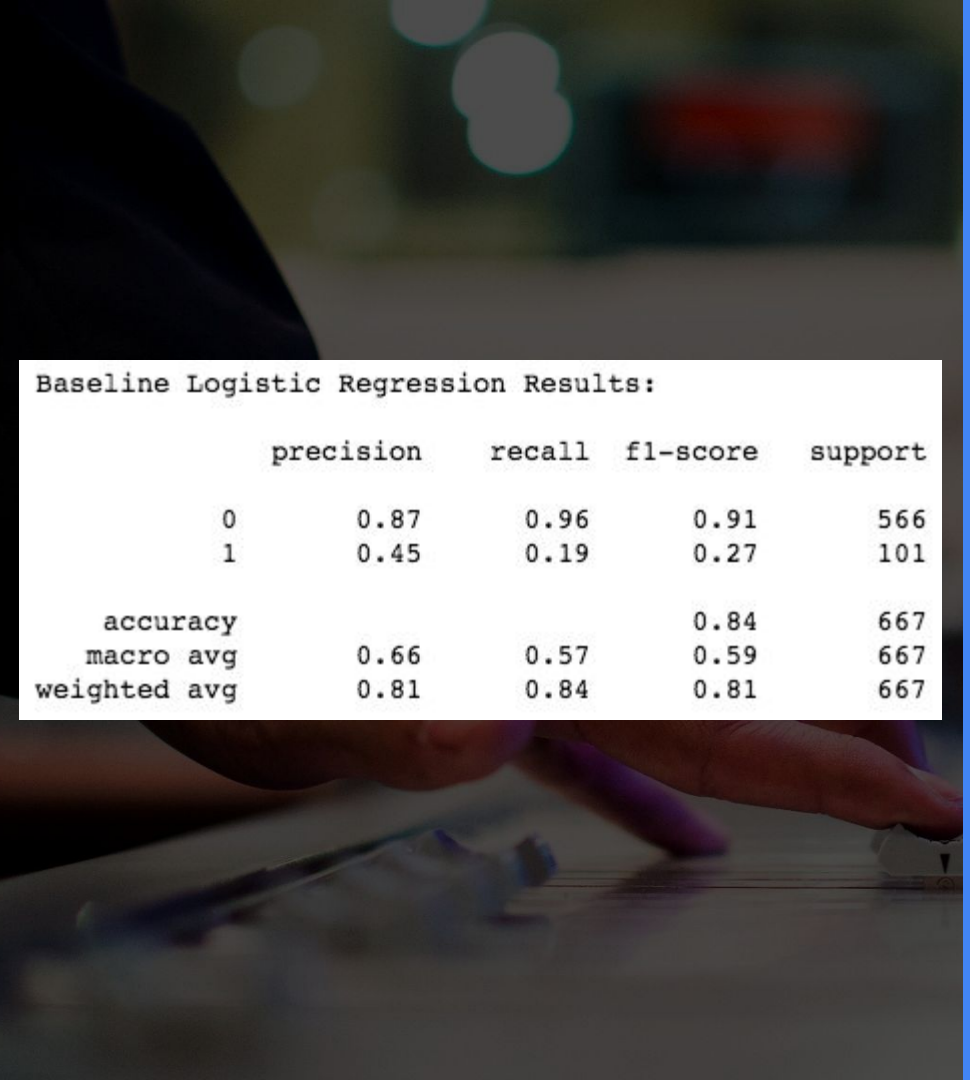




At first glance,

Customers with an international plan seem more likely to churn.

Let’s investigate further.



Baseline Logistic Regression Results:

	precision	recall	f1-score	support
0	0.87	0.96	0.91	566
1	0.45	0.19	0.27	101
accuracy			0.84	667
macro avg	0.66	0.57	0.59	667
weighted avg	0.81	0.84	0.81	667

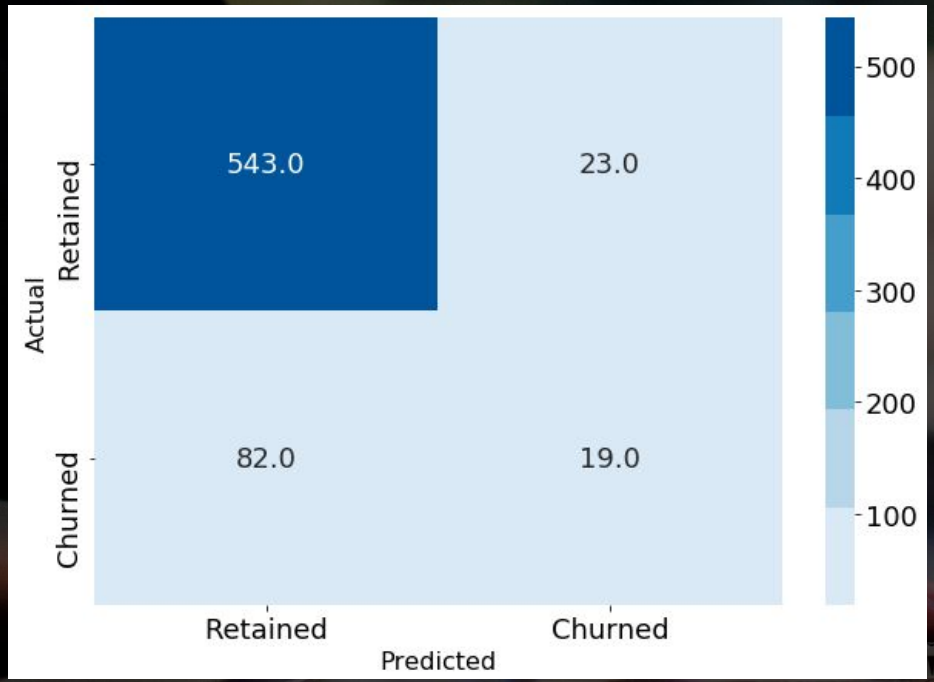
‘Vanilla’

Logistic Regression

Results

84% Accuracy

Confusion Matrix

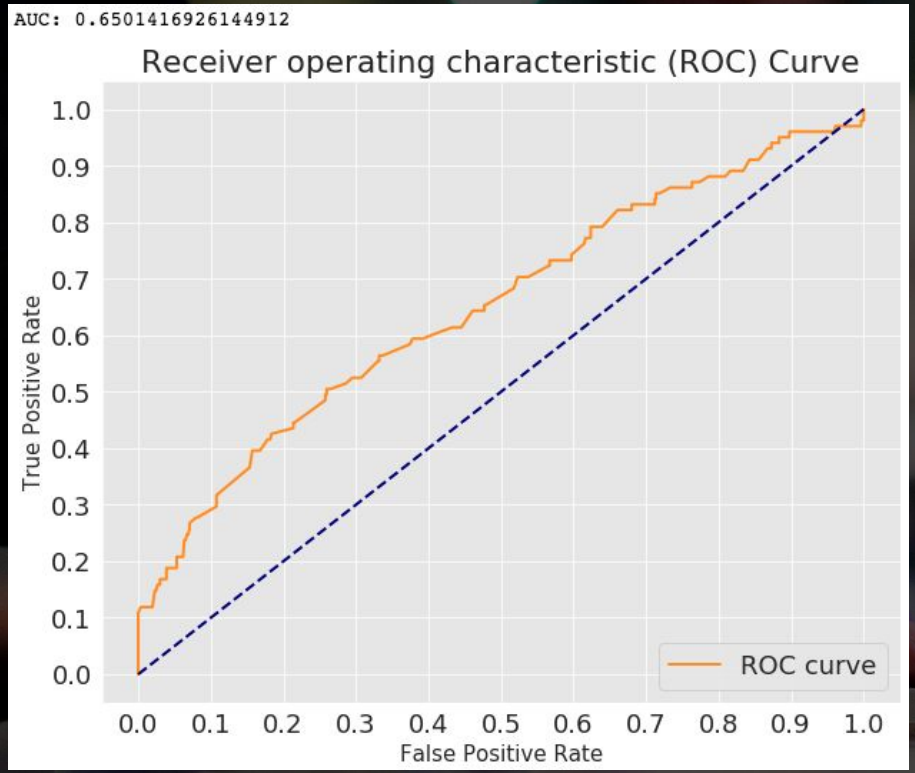


Cost-Benefit Analysis:

SyriaTel should focus on False Negatives.

Predicted ‘Retained’ but actually ‘Churned’

Baseline



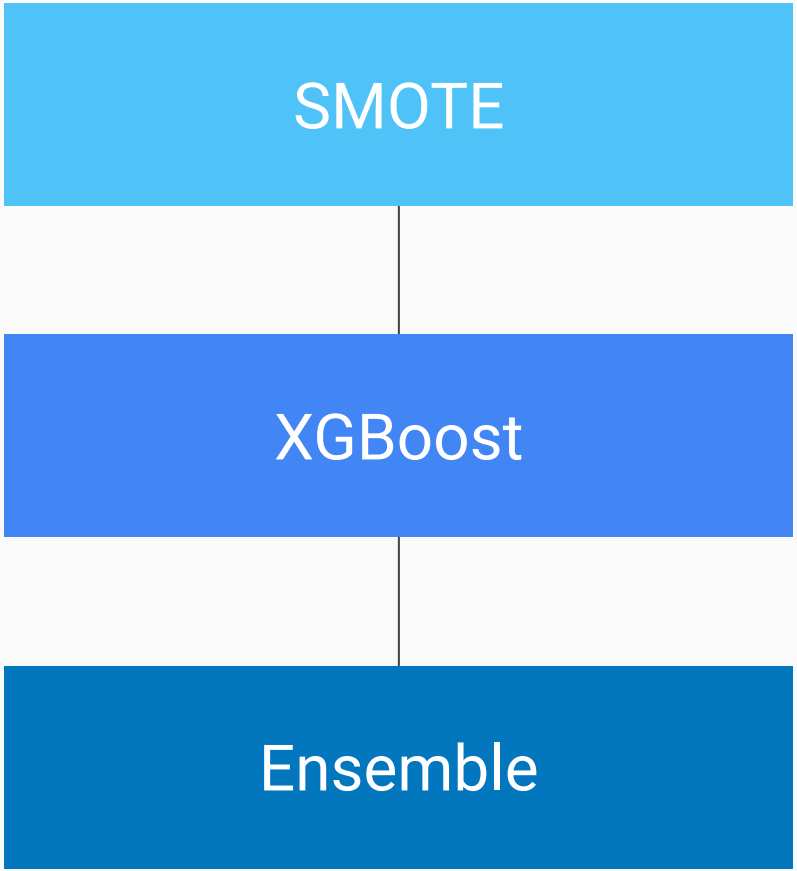
Cost-Benefit Analysis:

To gain an 80% True Positive Rate, SyriaTel would have to risk a 60% False Positive Rate.

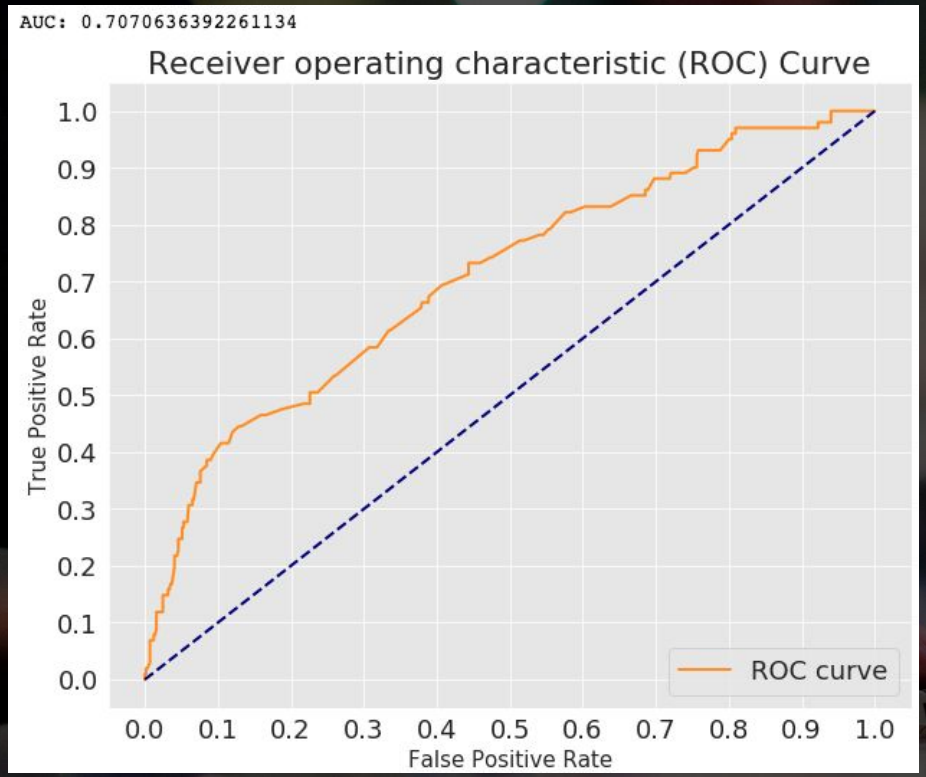
Too much risk.

Improving Performance

Shift ROC curve, and adjust AUC score



SMOTE

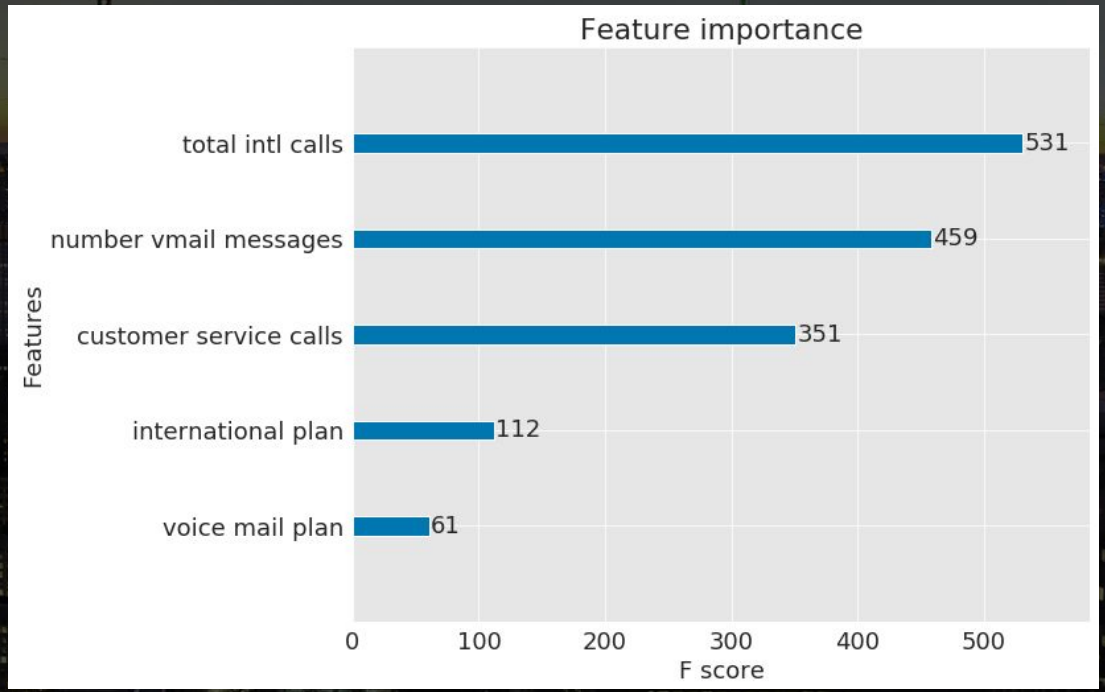


Cost-Benefit Analysis:

Synthetic Minority Oversampling Technique improved performance by 5%!

80% TPR now only risks 55% FPR

XGBoost Feature Selection:



XGBoost Performance Evaluation Before SMOTE:

XGBoost Baseline Classification Report:

	precision	recall	f1-score	support
0	0.88	0.98	0.93	566
1	0.73	0.24	0.36	101
accuracy			0.87	667
macro avg	0.80	0.61	0.64	667
weighted avg	0.86	0.87	0.84	667

Baseline LogReg Classification Report:

	precision	recall	f1-score	support
0	0.87	0.96	0.91	566
1	0.45	0.19	0.27	101
accuracy			0.84	667
macro avg	0.66	0.57	0.59	667
weighted avg	0.81	0.84	0.81	667

XGBoost Performance Evaluation After SMOTE:

XGBoost SMOTE Classification Report:

	precision	recall	f1-score	support
0	0.90	0.92	0.91	566
1	0.47	0.42	0.44	101
accuracy			0.84	667
macro avg	0.68	0.67	0.67	667
weighted avg	0.83	0.84	0.84	667

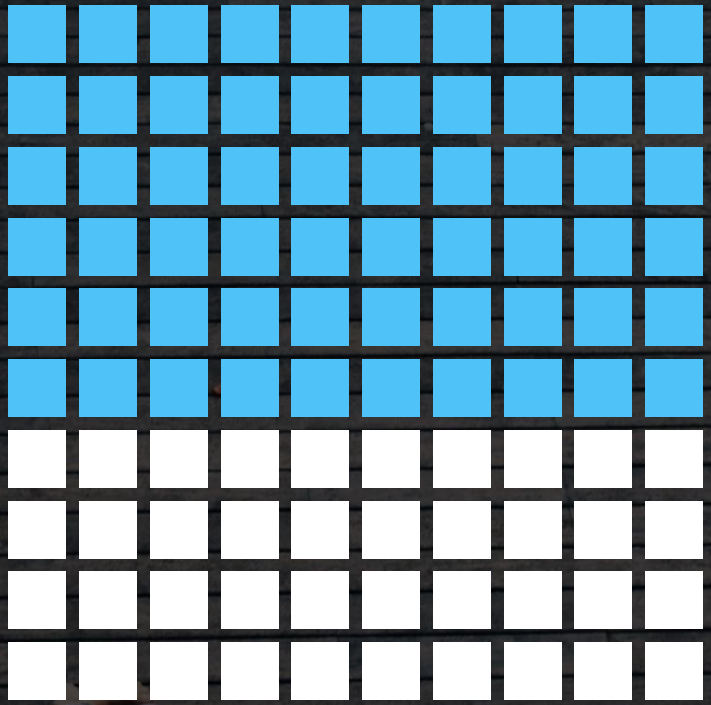
XGBoost Baseline Classification Report:

	precision	recall	f1-score	support
0	0.88	0.98	0.93	566
1	0.73	0.24	0.36	101
accuracy			0.87	667
macro avg	0.80	0.61	0.64	667
weighted avg	0.86	0.87	0.84	667

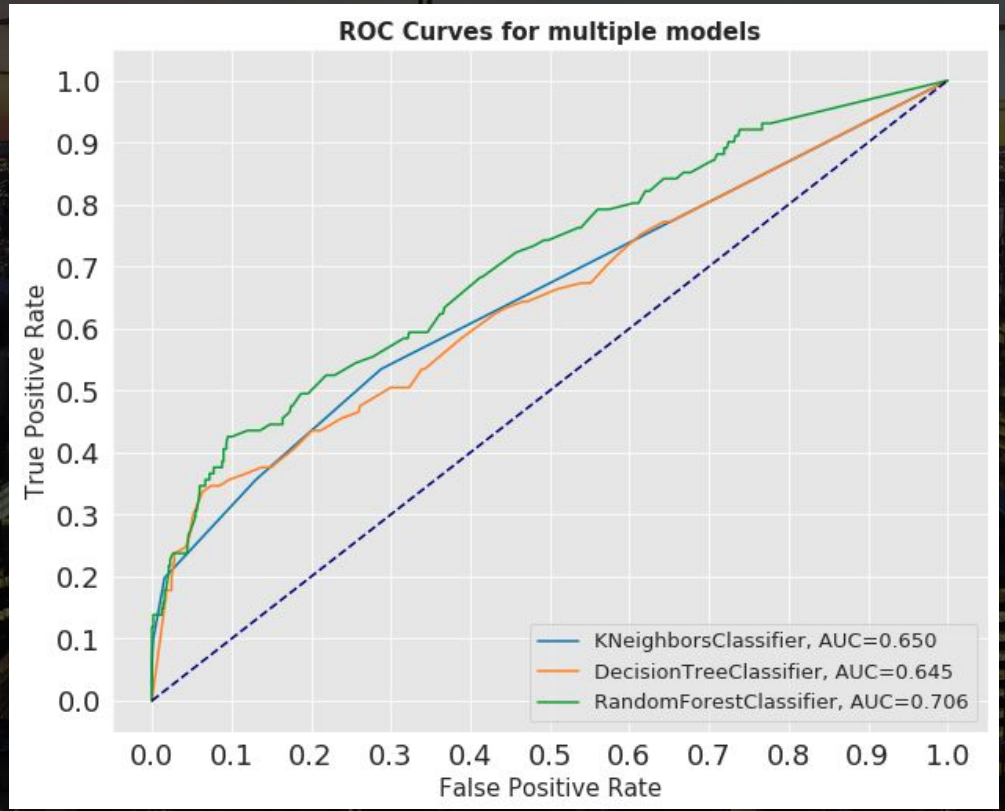
XGBoost

With Synthetic Minority Oversampling was able to reach 84% accuracy on our data

Ensemble methods may do better



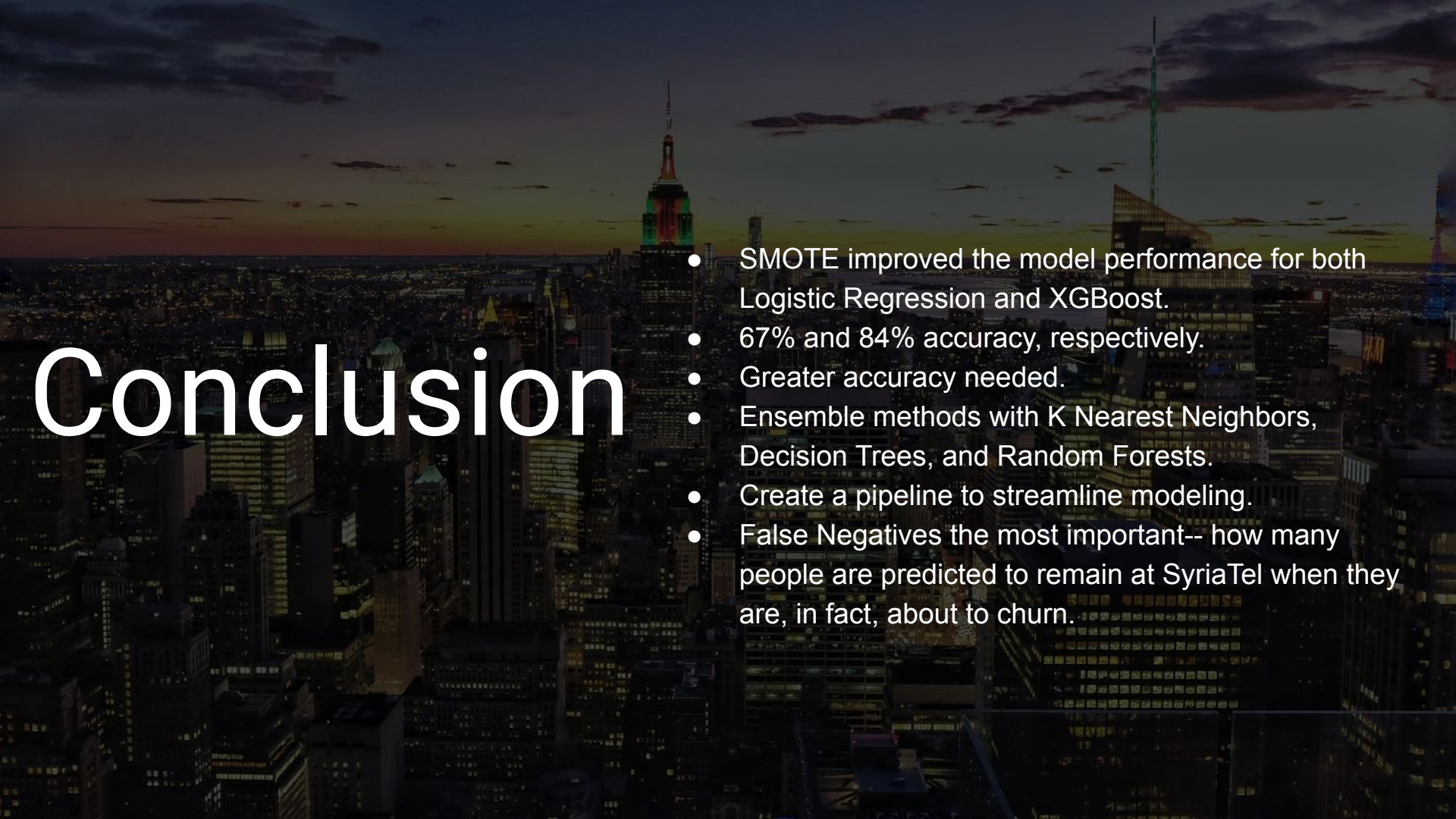
Baseline Ensemble Method Performance Example



Recommendations:

Model churn data across multiple models using

- ***Different forms of Feature Selection***
  - ***KBest***
  - ***Recursive Feature Elimination***
  - ***Principal Component Analysis***
- ***Cross Validation techniques***



Conclusion

- SMOTE improved the model performance for both Logistic Regression and XGBoost.
- 67% and 84% accuracy, respectively.
- Greater accuracy needed.
- Ensemble methods with K Nearest Neighbors, Decision Trees, and Random Forests.
- Create a pipeline to streamline modeling.
- False Negatives the most important-- how many people are predicted to remain at SyriaTel when they are, in fact, about to churn.