Machine Learning Project

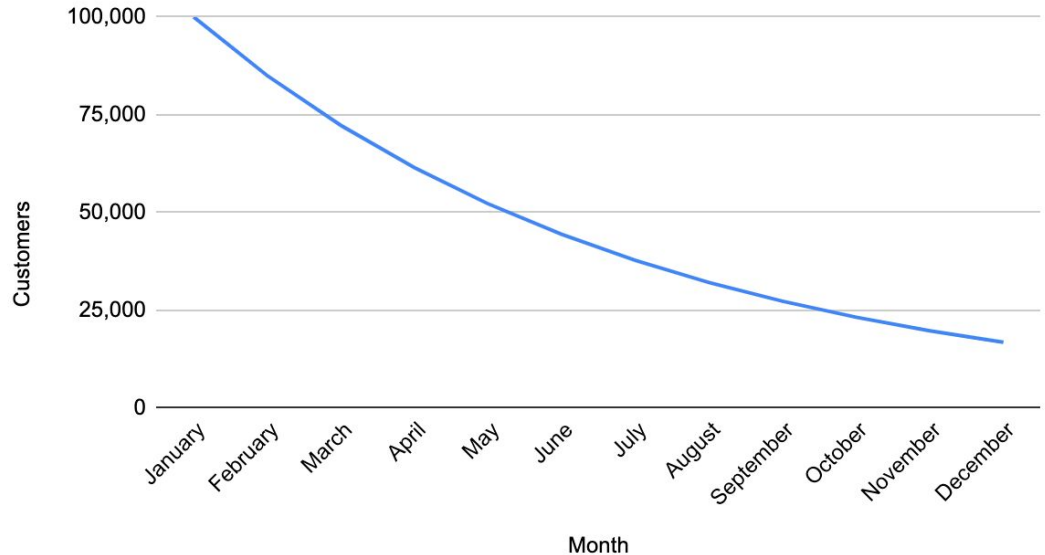Mac Lambert
Self-Paced Data Science
Flatiron School

Predicting Customer Churn for SyriaTel

Question: How do we predict whether a customer will churn (or stop doing business) with us?
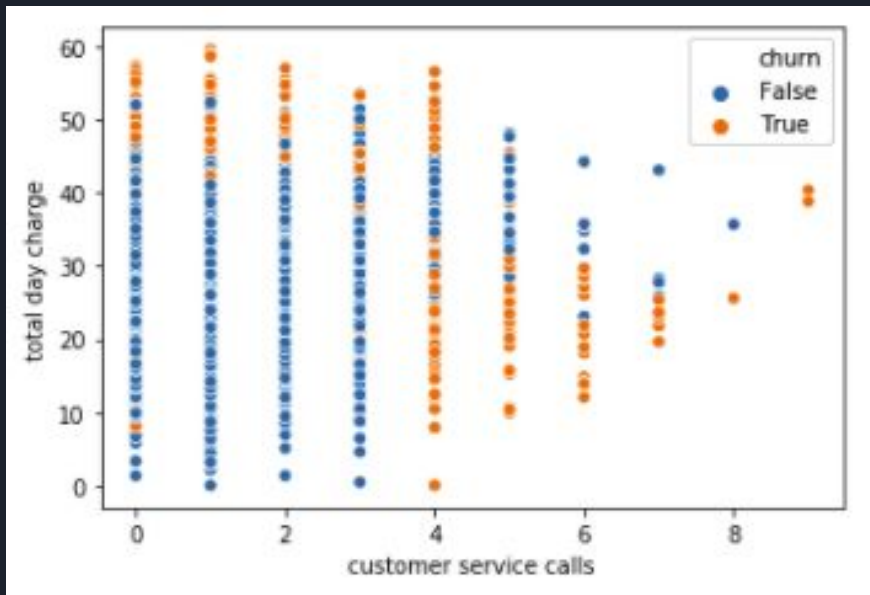
Roughly 15% of customers churn in any given month.

At this rate, over 80% of current customers will be gone within a year!
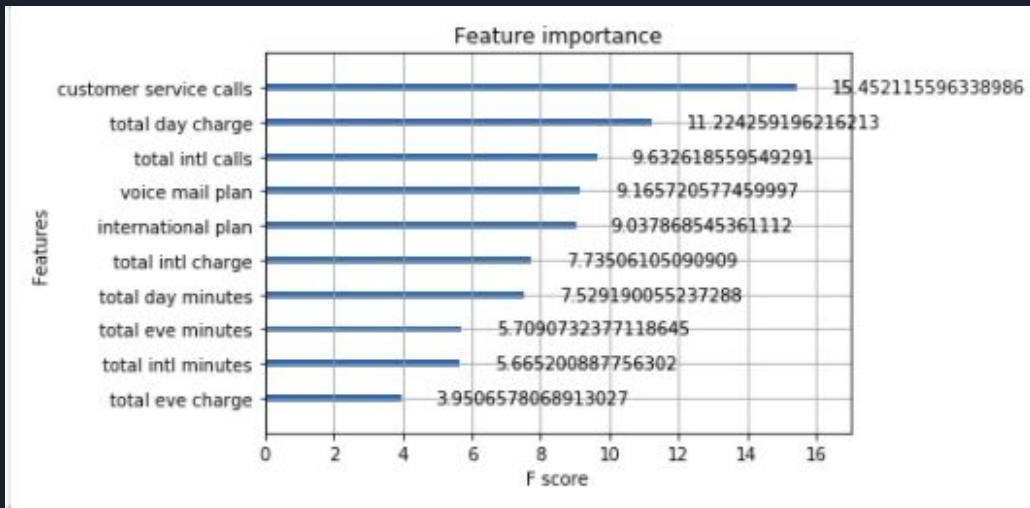

Customers vs. Month

What do we do about this?

Methodology:

- Use existing customer info to determine what makes a customer more likely to leave
- Our machine learning model built with customer info predicts churn to 97% accuracy

What should we focus on?

Feature importance

Features	F score
customer service calls	15.452115596338986
total day charge	11.224259196216213
total intl calls	9.632618559549291
voice mail plan	9.165720577459997
international plan	9.037868545361112
total intl charge	7.73506105090909
total day minutes	7.529190055237288
total eve minutes	5.7090732377118645
total intl minutes	5.665200887756302
total eve charge	3.9506578068913027

- The highest ranked features are shown on the left.
- Recommendation: focus on customer service calls, total day charge.

Further Work

- The next step in this analysis would be to examine charges for customers likely to churn, to determine cost savings impact. This is outside the scope of the predictive model discussed.
- Similarly, optimization was only run on the best performing model. Although 97% accuracy is fantastic, we could have spent the time to optimize all models.

Thank you & Appendix

- Helpful article on interpreting f score of XGBoost: https://towardsdatascience.com/be-careful-when-interpreting-your-features-importance-in-xgboost-6e16132588e7
- Link to Kaggle dataset: https://www.kaggle.com/becksddf/churn-in-telecoms-dataset
- Steps for one hot encoding state information: https://towardsdatascience.com/choosing-the-right-encoding-method-label-vs-onehot-encoder-a44344993149b