Determining Telco Churning.

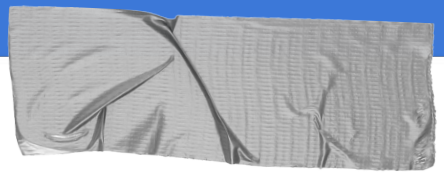
By RMM - June 2018

Mid-term Capstone.

This project corresponds to the final part of Unit 3 at the mid time of the Thinkful Data Science Bootcamp. You can visit this notebook at my GitHub repository.

Outline

- Introduction
- Dataset information & modelling aspects
- Feature & variable analysis
- Most influential features
- Model performance for predicting customer churn.
- Most important features and variables in customer churn
- Probability of customers leaving
- How can this information help us?
- How could we optimize our conclusions?



Intro

Customer churn (cc) occurs when customers or subscribers stop doing business with a company or service.

➔ **Which features influence cc?**
Valuable for building a retention campaign

➔ **Which features are most important?**
Knowing this will allow us to reduce computational costs and focus our resources when building our retention campaign

➔ **Which clients are most likely to leave?**
Applying our best model we'll determine which clients are more likely to leave.

Telco customer Churn - IBM Watson Analytics community

7042
Telco Customers

8
Models tested

21
Features present

3
Techniques to enhance hyper parameters within each model

Month-to-month contracts = 10 month median tenure

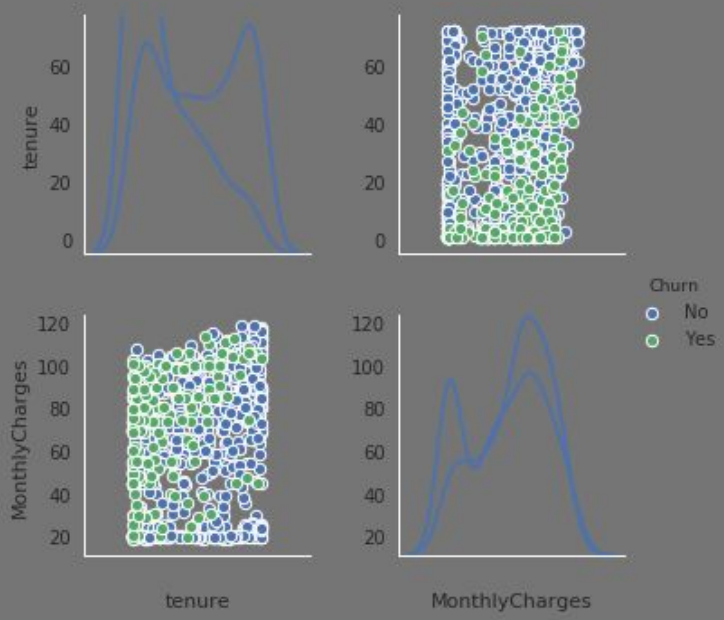
Two year contracts = 70 month median tenure

Tip

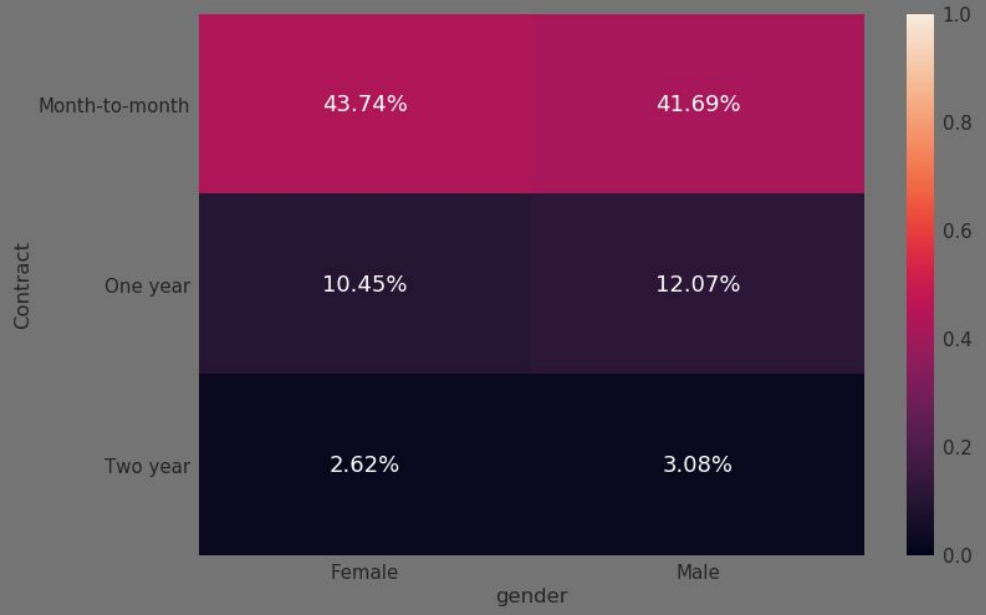
Tenure is the measurement of the amount of time a person is your customer, or in other words it's the age of a customer in your system.

Customers with online backup have a median age in the company of 32 months more in contrast to those without this service

Monthly charges and tenure share a close relationship



Month-to-month contracts have a higher churn probability



Electronic check payers churn most often

Most influential features

>>

- Tenure
- Monthly charges
- Internet service
- Online security
- Online backup
- Tech support
- Contract
- Payment method

Model performance for predicting customer churn.

>>

Logistic Regression	79.8 %
Ridge Classification	79.7 %
Gradient Boost Classification	79.6 %
Decision Tree	79.3 %
Random Forest C.	79.1 %
Support Vector Classification	78.7 %
KNN	77.1 %

SelectKBest outperformed PCA - GradientSearchCV applied on all models.

Most important features and variables in customer churn

	Coefficient
Contract - Month-to-month	0.350954
Internet Service - Fiber optics	0.292869
Payment Method - Electronic check	0.106498

Probability of customers leaving.

Probability	Customer count	Predicted probability (mean)	True probability (mean)
0 - 10 %	2117	0.038463	0.042192
10 - 20 %	833	0.147585	0.137267
20 - 30 %	512	0.248014	0.267693
30 - 40 %	539	0.349264	0.304379
40 - 50 %	463	0.448919	0.443929
50 - 60 %	408	0.552887	0.551948
60 - 70 %	390	0.64793	0.666568
70 - 80 %	304	0.744805	0.721296
80 - 90 %	59	0.824136	0.878205
90 - 100 %	0	NaN	NaN

How can this information help us?

Decide into which clients we should focus our resources

363 clients have over 70% probability of leaving

Address the most important issues that influence churn

Type of contract, payment method, internet service

Reduce significantly campaign costs

Budgets are tight and we need to maximize our resources

Aspects we could improve to increase our predictability.

>>

- Dataset size
- Hyper parameter tuning
- Multiple algo testing
- Ensemble models
- Surveying for more features


Thank you.

Questions?