Predicting Customer Churn To Defend Subscription Revenue Streams

March 2020

Agenda

- Problem Statement & Key Objectives
- Classification Modeling & Evaluation
- Recommendations & Next Steps

Problem Statement & Key Objectives

Background:

- Businesses that rely on subscription-based revenue models can be greatly impacted by customer churn
- In order to defend topline revenue, businesses should seek to identify customers with high churn risk and escalate them to Customer Success/Outreach teams to mitigate churn and emphasize value prop realization
- Churn data typically highly confidential therefore telco churn data obtained from Kaggle and includes:
    - *Demographics (age, gender)*
    - *Services (e.g. Internet, TV, Phone)*
    - *Total and Monthly Payments*

Key Objectives:

- ***Improve Churn and defend key revenue streams by establishing a Customer Churn Prediction program by applying Machine Learning to customer data***
- ***Prototype Classification System for ongoing identification of customer risk of churn***
- ***Calculate potential salvageable revenue***

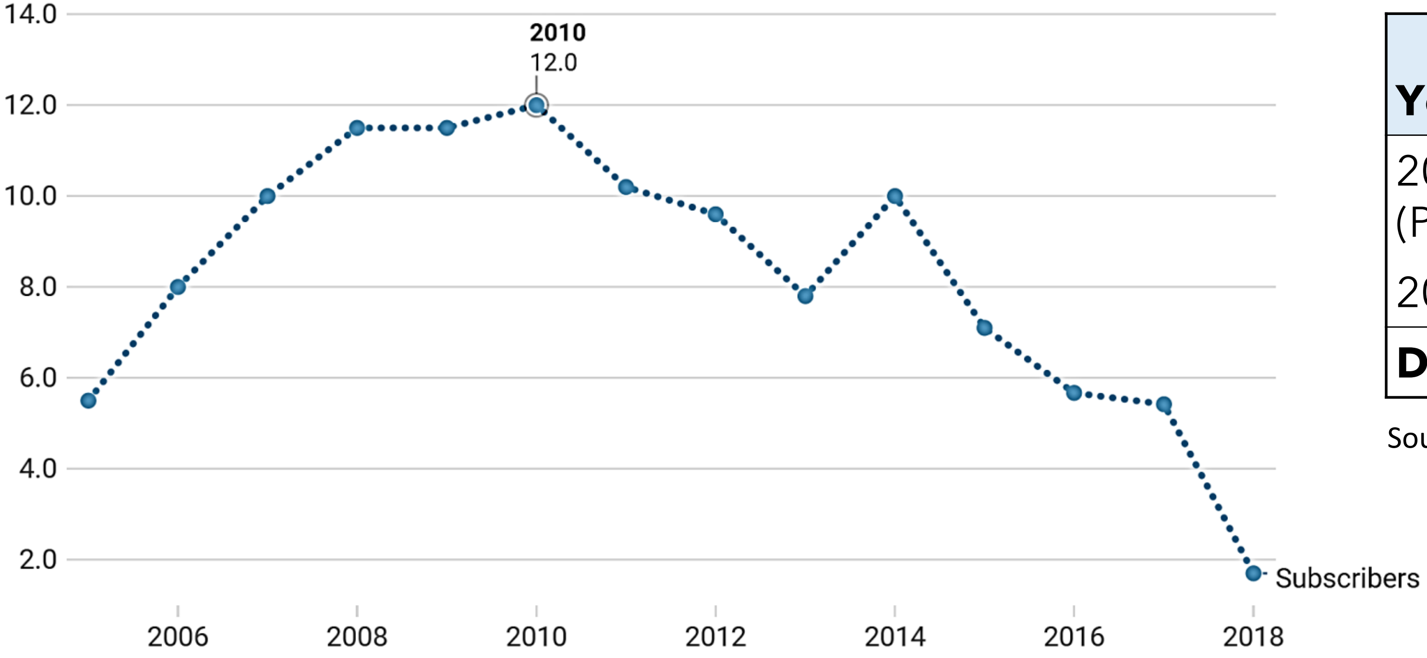
Executive Summary

- **Key objectives: Improve Churn and defend key revenue streams by establishing a Customer Churn Prediction program by applying Machine Learning to Customer Data. Prototype Classification System for ongoing identification of customer risk of churn and estimate potential salvageable revenue**
- Most churned customers were on monthly contracts and canceled within first month likely due to trials. Poor product market fit for Fiber Optic Internet Service likely influenced churn, supported by high correlation, high concentration, and coefficient results
- Myriad Machine Learning Algorithms were applied. Ultimately Logistic Regression and Neural Network exhibited the highest performance and were then optimized for recall/sensitivity (testing positive for Churn)
- Fiber Optic Internet, Paperless Billing, Elec Check Payments, Senior Citizen status suggest higher likelihood to churn
- In contrast, having a contract, age of customer, higher spend, being subbed to more features (OnlineSecurity, TechSupport Prioritization) suggest higher likelihood to remain subscribed
- Approximately **$803K** in ARR are conservatively estimated to be recoverable with the Customer Churn Prediction program. Results will vary depending on product market fit, industry, customer outreach team ability to execute
- **Recommendations:** trial and refine program. Prioritize customer outreach to predicted churn customers with 0.35 to 0.75 probability, emphasize value proposition and warm transition to customer support if issues exist. At renewal time consider promotional incentives to boost customer retention

Example: WoW At Its Peak Had 12M Subscribers Driving $2.1B ARR, After Churn It Decreased To $306M In 2018 (1.7M Subs)

World of Warcraft Subscribers Since 2005

Number of WoW Subscribers since it's release in 2005.*



Number of subscribers in millions

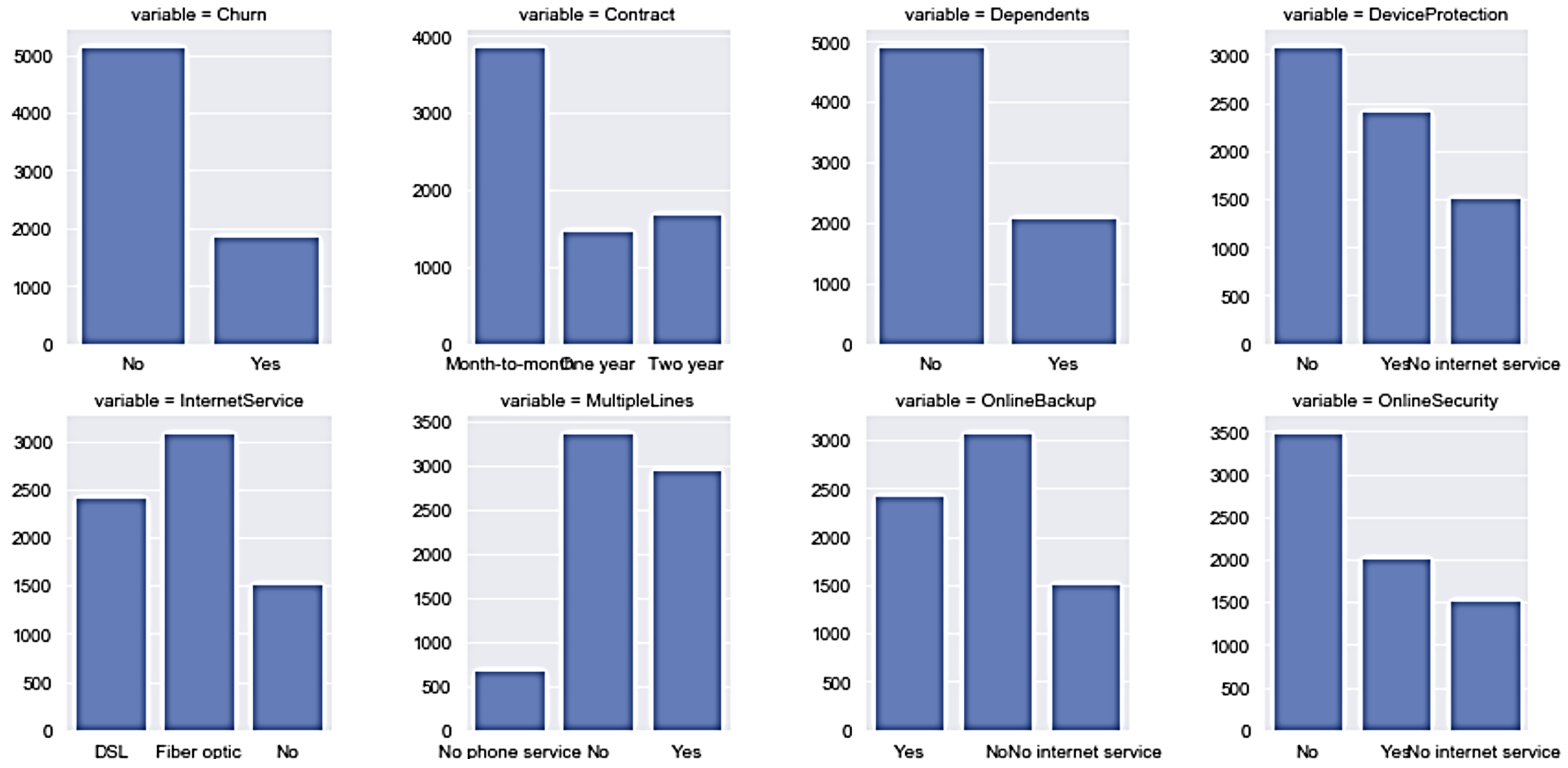
Source: Expanded Ramblings

Year	Subscriptions (M)	MRR ($M)	ARR ($M)
2010 (Peak)	12	$180	$2,159
2018	1.7	$25	$306
Delta	**10.3**	**$154**	**$1,853**

Source: Power Word Gold

Assumes $14.99 monthly fee across all customers.
Note: can't calculate churn directly since figures include new subscribers as well

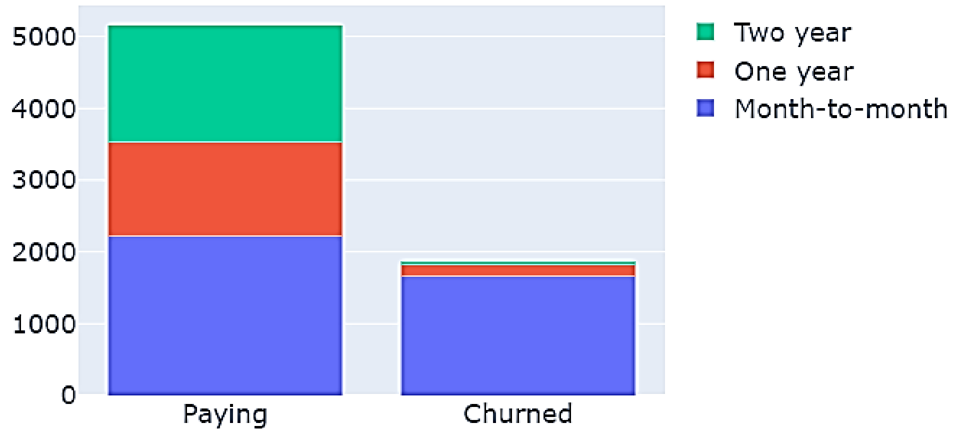
Telco Dataset Composition



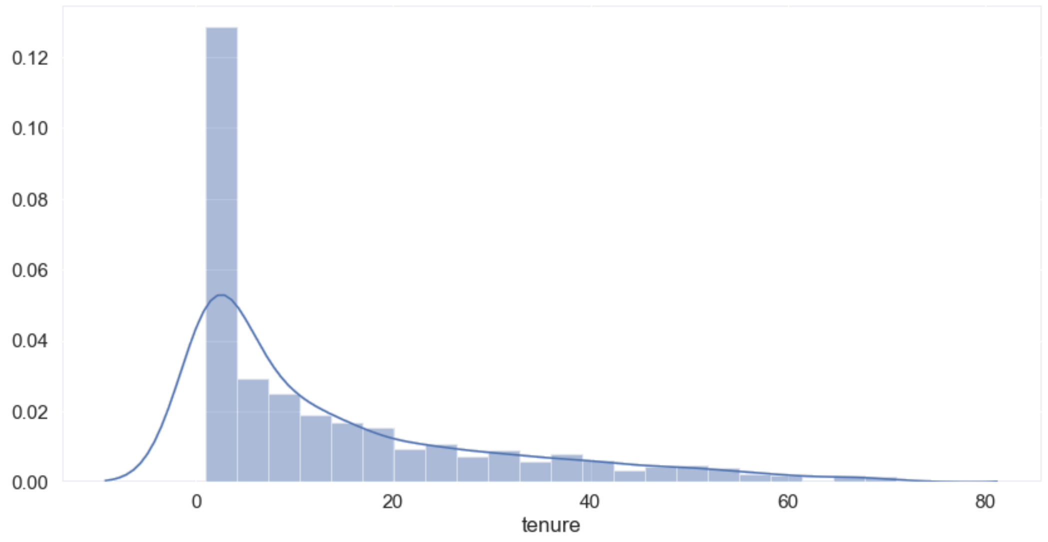
- **7042** unique customers with **21** columns
- **26.5%** of customers in Dataset churned
- Most customers in dataset have:
  - ***month-to-month contracts***
  - ***no dependents***
  - ***no DeviceProtection***
  - ***FiberOptic Internet***
  - ***One Phoneline***
  - ***No online backup***
  - ***No OnlineSecurity***

Most Churned Customers Were On Monthly Contracts And Canceled Within First Month Likely Due To Trials And Poor Product Market Fit

Customer Contract Types by Churned Status



Churned Customers on Month-to-Month plans by Churn Month

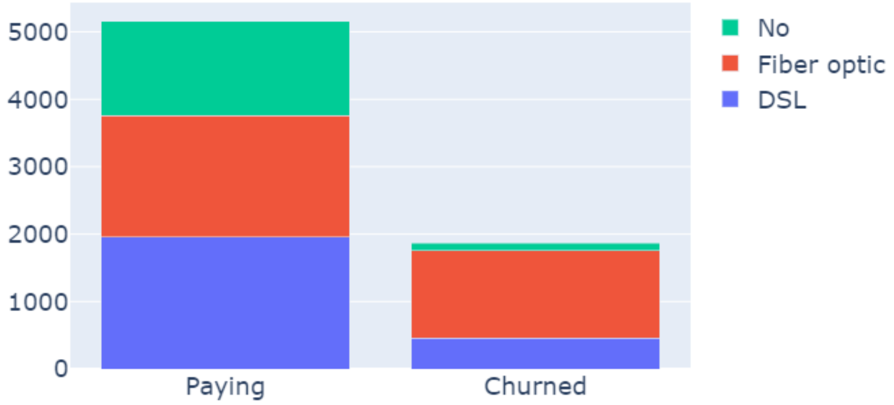


- Churned customers have a higher concentration of Month-to-Month contracts, which are typically more expensive than monthly rates from annual or multi-annual contracts. These customers may have opted for cheaper rates from competitor annual plans

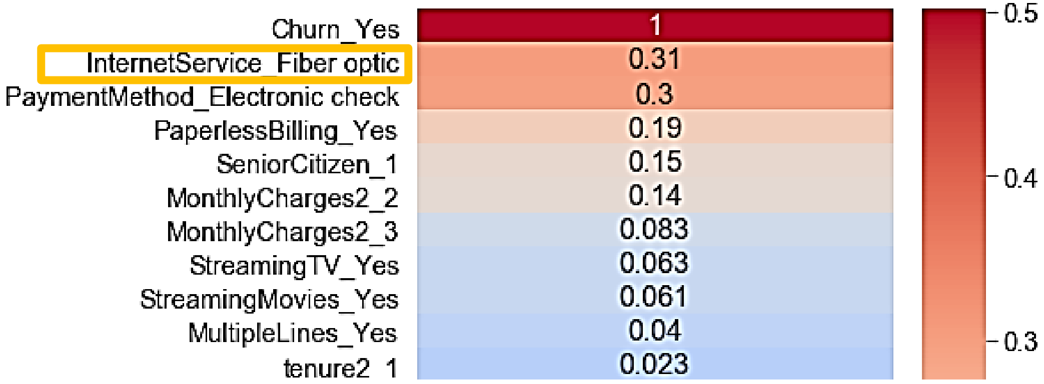
- 23% of Churned Customers cancelled within the first month. This is likely due to a trial period
- 45% of Churned Customers cancelled within first 6 months

Churned Customers Also Had High Concentration Of Fiber Optic Internet Service, Poor Product Market Fit Likely Influenced Churn

Customer Internet Types by Churned Status



Top 10 Positively Correlated Churn Features



- 70% of churned customers had Fiber Optic Internet
- Of the churned customers with month-to-month contracts 70% also had Fiber optic internet

- Fiber Optic Internet was the leading feature correlated with Churn
- The data suggests that the Fiber Optic internet offering may not have a strong Product Market fit
- High price, slow speeds, poor reliability may be driving churn. Additional customer surveys and reviews with Product Managers are suggested to corroborate this

Modeling
& Evaluation

Model Overview

Data Import & Wrangling

- Read-in myriad data sources and consolidate
- Check for nulls, missing values appropriate data types, and outliers
- Assess whether any numerical values might be better suited for categorical type (non-continuous)
- Assess whether any categorical data may be suited for ordinal numerical type

EDA & Data Pre-Processing

- Correlations
- Dataset composition – characteristics of churned vs paying customers
- Feature Engineering, Interaction terms
- Feature selection
- Dummy encoding
- Train test split
- Standard Scaler

Modeling & Evaluation

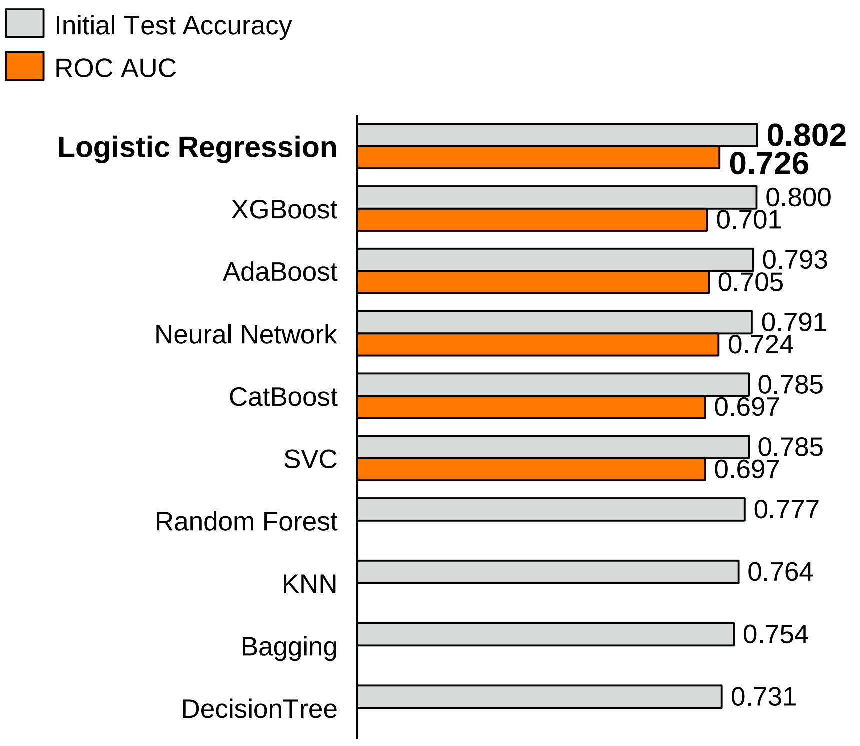
- Logistic Regression
- Neural Network
- XGBoost
- Random Forest
- DecisionTree
- AdaBoost
- CatBoost
- SVC
- Bagging
- KNN

Recall-Optimization and EDA on Results

- Optimized Logistic Regression and Neural Network for recall
- Analyzed results to estimate salvageable revenue

Model Evaluation: Logistic Regression Had The Highest Initial Performance

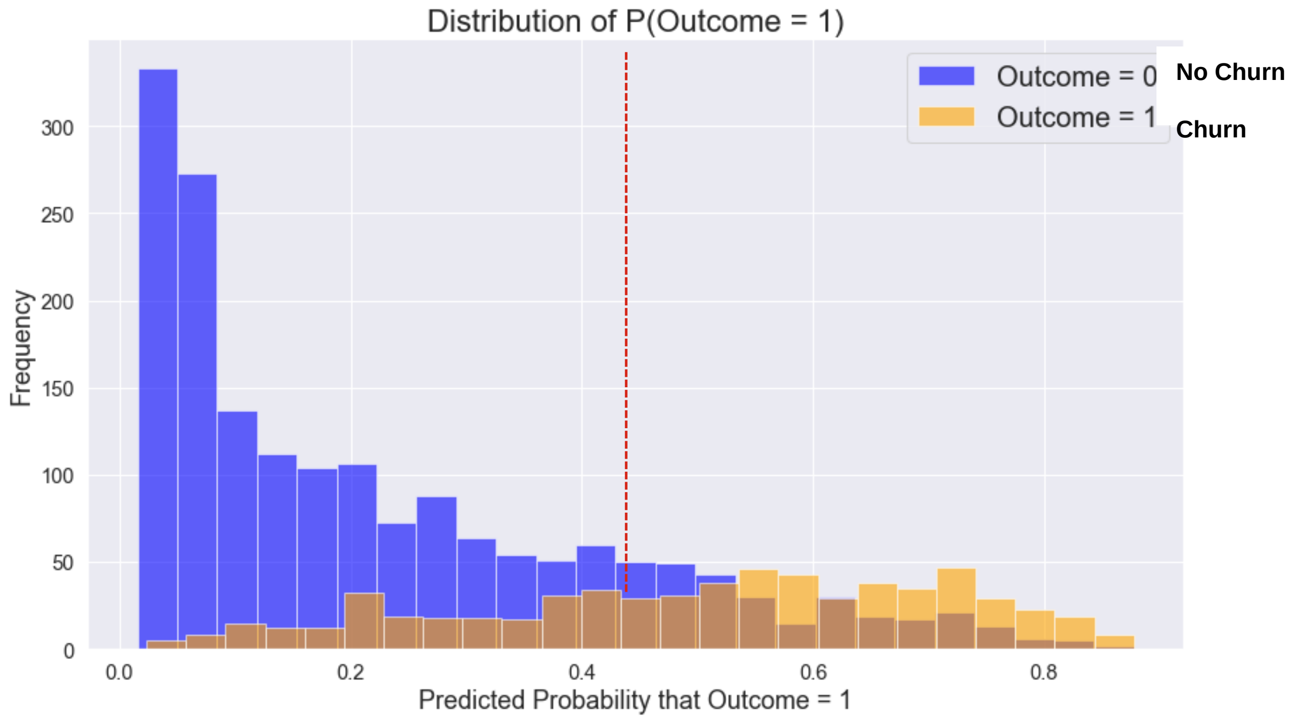
Model Accuracy Scores



Baseline Accuracy Score is 0.7341

Note: I did not proceed with Random Forest, KNN, Bagging, or Decision Trees

Logistic Regression



Initial Logistic Regression Recall at 0.56 shows room for optimization

Recall-Optimization Results

Model	Initial Accuracy	Recall-optimized Acc	Initial Recall	Recall-optimized	Initial ROC AUC	Recall-optimized ROC AUC
Logistic Regression	0.801757	0.761606	0.564465	0.761006	0.726107	0.761415
Neural Network	0.790882	0.774571	0.581761	0.716981	0.724214	0.756211

- Optimized for recognizing positive Churn, Customers with 38%+ probability of churn
- Logistic regression optimized +20 pts to 0.76 recall
- Accuracy impacted slightly but ROC AUC (ability to distinguish between classes) improved +4 pts
- Neural Network also optimized recall below Logistic Regression performance

Fiber Optic Internet, Paperless Billing, Elec Check Payments , Senior Citizen Status Suggest Higher Likelihood To Churn

features	coefficients
fiber_internet+elec_check	0.362354
InternetService_Fiber optic	0.255167
PaperlessBilling_Yes	0.190131
SeniorCitizen_1	0.126046
PaymentMethod_Electronic check	0.107188
MonthlyCharges2_3	0.096713
StreamingMovies_Yes	0.076301
StreamingTV_Yes	0.073607
MultipleLines_Yes	0.058025
PaymentMethod_Mailed check	0.016887
MonthlyCharges2_2	-0.007192
MultipleLines_No phone service	-0.010570
gender_Male	-0.048143
PaymentMethod_Credit card (automatic)	-0.112110
MonthlyCharges2_1	-0.119963
DeviceProtection_Yes	-0.123661
Partner_Yes	-0.158171
TotalCharges2_1	-0.161258

- *Fiber Optic Internet*
- *Paperless Billing*
- *Senior Citizen*
- *Electronic Check payments*
- *High Monthly Charges*

- **These factors suggest higher likelihood to churn**

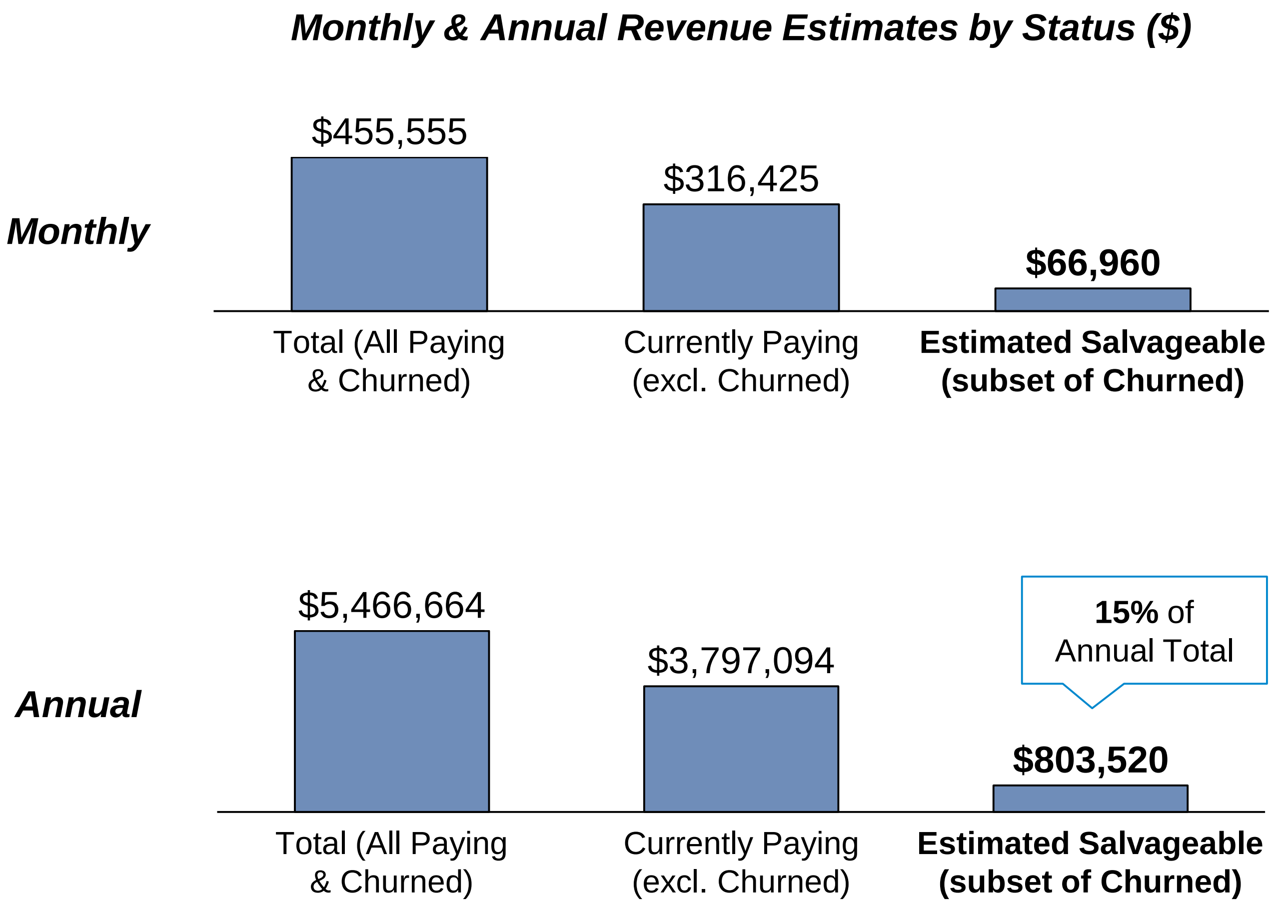
	features	coefficients
22	StreamingTV_No internet service	-0.178749
24	StreamingMovies_No internet service	-0.178749
18	DeviceProtection_No internet service	-0.178749
20	TechSupport_No internet service	-0.178749
16	OnlineBackup_No internet service	-0.178749
14	OnlineSecurity_No internet service	-0.178749
13	InternetService_No	-0.178749
0	PhoneService_Yes	-0.198795
9	Dependents_Yes	-0.206553
26	tenure2_1	-0.241895
17	OnlineBackup_Yes	-0.254846
33	TotalCharges2_2	-0.329548
27	tenure2_2	-0.358495
21	TechSupport_Yes	-0.360700
15	OnlineSecurity_Yes	-0.363053
34	TotalCharges2_3	-0.419491
1	Contract_One year	-0.425887
28	tenure2_3	-0.515219
2	Contract_Two year	-0.604761

- *One or Two year contracts*
- *Age of customer*
- *High spend*
- *Subbed to more features (OnlineSecurity, TechSupport prioritization)*

- **These factors suggest higher likelihood to remain subscribed**

Coefficients derived from unscaled Logistic Regression model

Results: Defend $800K (15%) Of Annual Recurring Revenue By Predicting Customers At High Risk Of Churn



Results will vary depending on product market fit, industry, and customer outreach team's ability to execute

Assumptions

- Customers with churn probability of **75%** or higher are unsalvageable (too unlikely to recover) (approx. 95 unique customers in dataset)
- Assume we can recover a conservative **50%** of customers with churn probability between **0.4** and **0.75** (approx. 705 customers)
- Customers with churn probability between **0.2** and **0.4** are salvageable since they have less than half chance of churning
- Customers with less than 20% probability are likely to remain paying customers
- Average monthly churn price: $74.44
- Average monthly paying price: $61.30
- Average monthly total price: $64.79

Recommendations & Next Steps

- **Prioritize customer outreach to predicted churn customers** with 0.35 to 0.75 probability, emphasize value proposition and warm transition to customer support if issues exist. At renewal time consider promotional incentives to boost customer retention
- **Evaluate fail fast mentality:** customers with 75% or higher probability of churn are unlikely to remain, are there any profitability gains by letting them out early and saving on forbiddingly high cloud (e.g. AWS, Azure, Google Cloud) costs, in addition to defending NPS scores
- **Additional data: NPS, Product Usage,** and **Support Case** volume (well-known pre-cursors to churn) will drastically increase accuracy of model
- **Date/time data** and corresponding product releases are critical for SaaS businesses since continuous development and continuous integration will incrementally improve product. Evaluating churn with a substantially improved product will have different results
- Field **customer exit surveys** to evaluate leading causes of churn and feed back into product roadmap. Based on this analysis the data suggests poor product market fit for fiber optic internet service (potentially high price, slow speeds, poor reliability)
- Expand to include **Customer Lifetime Value** analysis leveraging frequency and recency transactional data