Initial Findings

Model Performance

No Oversample vs Oversample

Class Imbalance

27% **Average Rate of Churn**

73% **Negative class (Typically problematic at 80%)**

- **Addressed class imbalance with oversampling**
- **Optimizing for Recall metric**

Model Scores

Oversample vs No Oversample

Default Hyperparameters

Model	AUC	Recall
Logistic Regression Oversample	.75	80%
Logistic Regression No Oversample	.83	51%
XGBoost Oversample	.73	68%
XGBoost No Oversample	.68	51%
Random Forest Oversample	.71	59%
Random Forest No Oversample	.69	49%
Decision Tree Oversample	.64	47%
Decision Tree No Oversample	.62	46%

Oversampling improved each model's prediction performance:

- Improved AUC
- In most cases greatly improved recall

Score Definitions

- **AUC**: Aggregate measure of performance across all possible classification thresholds
  - AUC ranges in value from 0 to 1.
  - Model whose predictions are 100% wrong has an AUC of 0.0;
  - model whose predictions are 100% correct has an AUC of 1.0.
- **Recall:** Proportion of actual positives that were correctly identified

Best Performing Models

Based on having best recall score and acceptable AUC score

Model	AUC	Recall
Logistic Regression Oversample	.75	80%
XGBoost Oversample	.73	68%
Random Forest Oversample	.71	59%

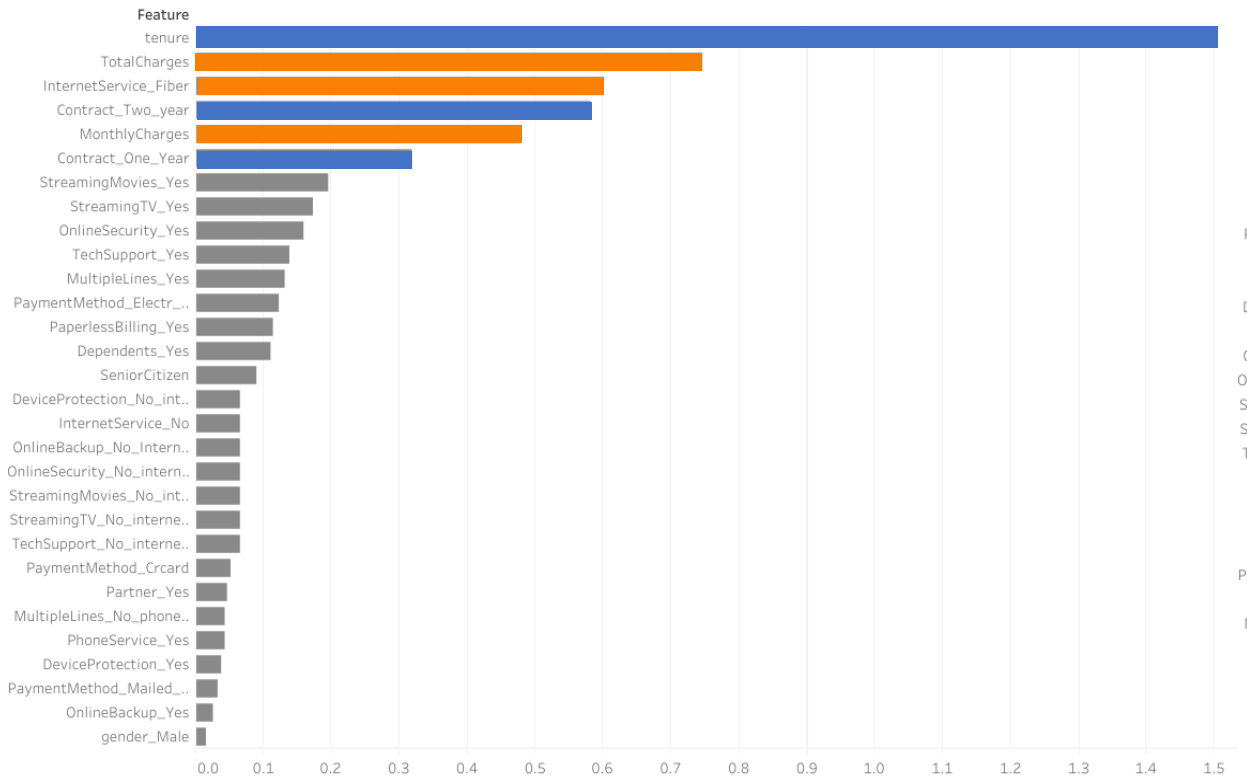
AUC Conventional Scoring
0.5 = No discrimination
0.5-0.7 = Poor discrimination
0.7-0.8 = Acceptable discrimination
0.8-0.9= Excellent discrimination
>0.9 = Outstanding discrimination

Logistic Regression Important Features

Important Features

Oversampling impacts the magnitude of importance

No Oversampling(LogReg)



Oversampling(LogReg)