TELECOM

PREDICT CUSTOMER CHURN

What drives higher churn rate of customer & how can Telecoms tackle this problem?

Agenda

Goal

Data Overview

Model

Implication

Summary

Goal

Develop a model to predict customers likely to churn

Challenges:

- 400 millions subscribers in the US telecommunication industry
- There are 54 telecommunication companies, according to Forbes 2018
- Monthly loss from Churn is $65M

Solution:

- Predict customers' churn decisions
- Develop retention plans

Data Overview

Customer Demographics

Services

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No

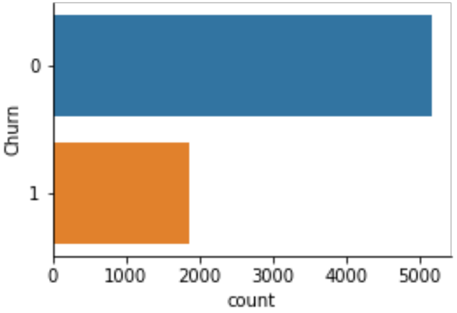
Customer who left within last month

Services

Account Info

DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
Yes	Yes	No	No	One year	No	Bank transfer (automatic)	42.30	1840.75	No
No	No	No	No	Month-to-month	Yes	Electronic check	70.70	151.65	Yes

Data Source: IBM

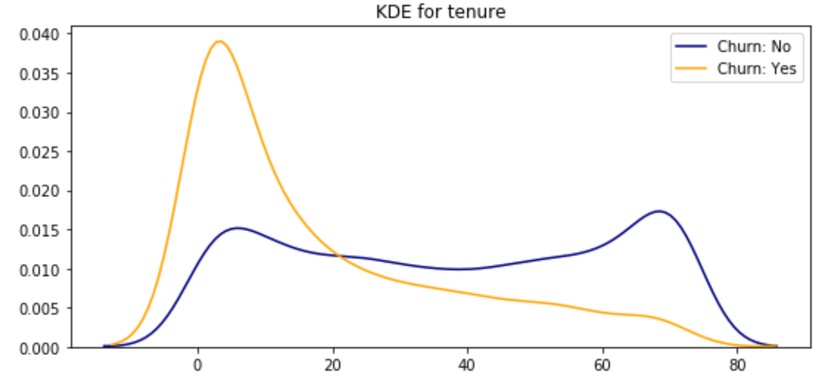


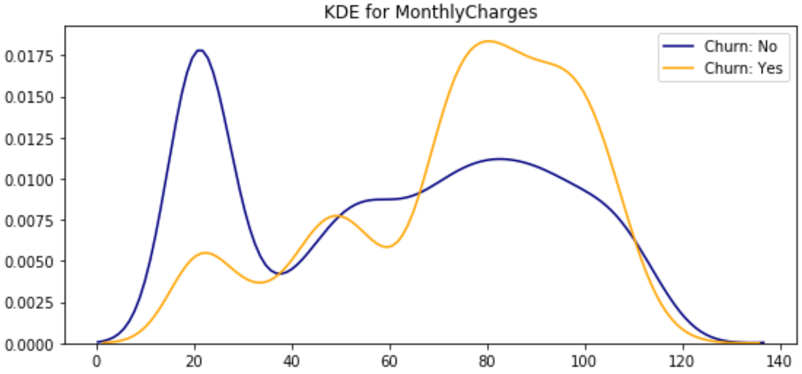
Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
gender	7,043	0.5	0.5	0	0	1	1	1
SeniorCitizen	7,043	0.2	0.4	0	0	0	0	1
Partner	7,043	0.5	0.5	0	0	0	1	1
Dependents	7,043	0.3	0.5	0	0	0	1	1
tenure	7,043	32.4	24.6	0	9	29	55	72
PhoneService	7,043	0.9	0.3	0	1	1	1	1
PaperlessBilling	7,043	0.6	0.5	0	0	1	1	1
MonthlyCharges	7,043	64.8	30.1	18.2	35.5	70.3	89.8	118.8
TotalCharges	7,032	2,283.3	2,266.8	18.8	401.4	1,397.5	3,794.7	8,684.8
Churn	7,043	0.3	0.4	0	0	0	1	1
Contract_MonthtoMonth	7,043	0.6	0.5	0	0	1	1	1
Contract_OneYear	7,043	0.2	0.4	0	0	0	0	1
Contract_TwoYear	7,043	0.2	0.4	0	0	0	0	1
DSL	7,043	0.3	0.5	0	0	0	1	1
FiberOptic	7,043	0.4	0.5	0	0	0	1	1
NoInternetService	7,043	0.2	0.4	0	0	0	0	1
OnlineSecurity_Yes	7,043	0.3	0.5	0	0	0	1	1
OnlineBackup_Yes	7,043	0.3	0.5	0	0	0	1	1
DeviceProtection_Yes	7,043	0.3	0.5	0	0	0	1	1
TechSupport_Yes	7,043	0.3	0.5	0	0	0	1	1
StreamingTV_Yes	7,043	0.4	0.5	0	0	0	1	1
StreamingMovies_Yes	7,043	0.4	0.5	0	0	0	1	1
PaymentMethod_ElectronicCheck	7,043	0.3	0.5	0	0	0	1	1
PaymentMethod_MailedCheck	7,043	0.2	0.4	0	0	0	0	1
PaymentMethod_BankTransfer	7,043	0.2	0.4	0	0	0	0	1
PaymentMethod_CreditCard	7,043	0.2	0.4	0	0	0	0	1
MultipleLines_No	7,043	0.5	0.5	0	0	0	1	1
MultipleLines_Yes	7,043	0.4	0.5	0	0	0	1	1

Numerical Features

Tenure & Monthly Charges

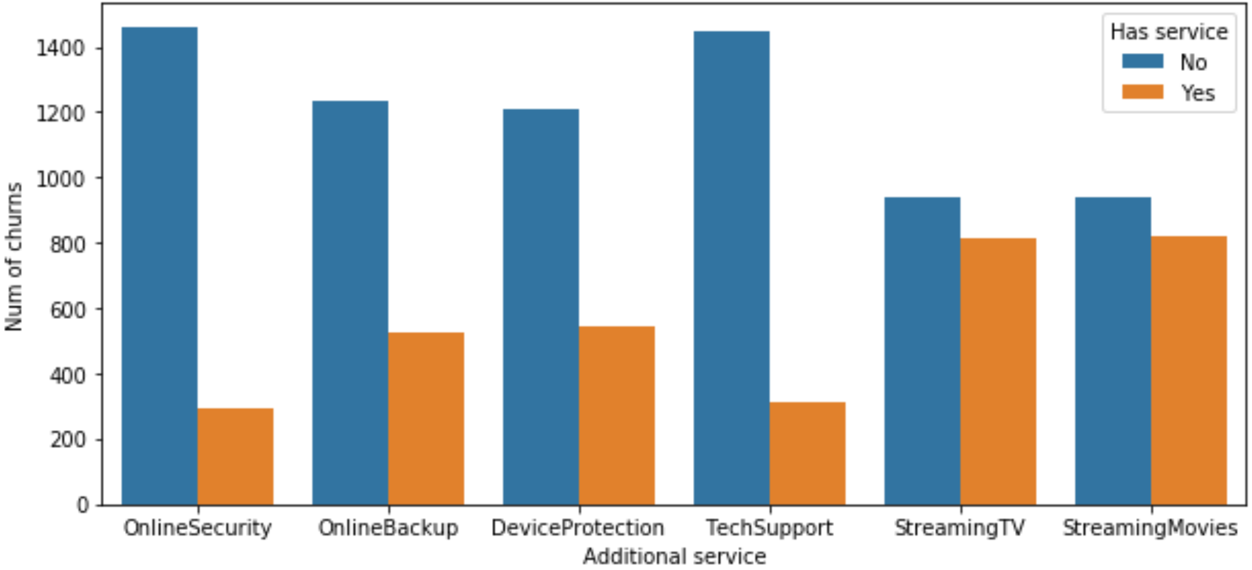




1. New clients are more likely to churn
2. Clients with higher *Monthly Charges* are also more likely to churn
3. *Tenure* and *Monthly Charges* are probably important features

Categorical Features
Six Additional Services

- Customers with the first 4 additional services (***Security, Backup, Protection, Tech support***) are less likely to churn
- ***Streaming services*** are not likely to associate with churn



Model

Logistic

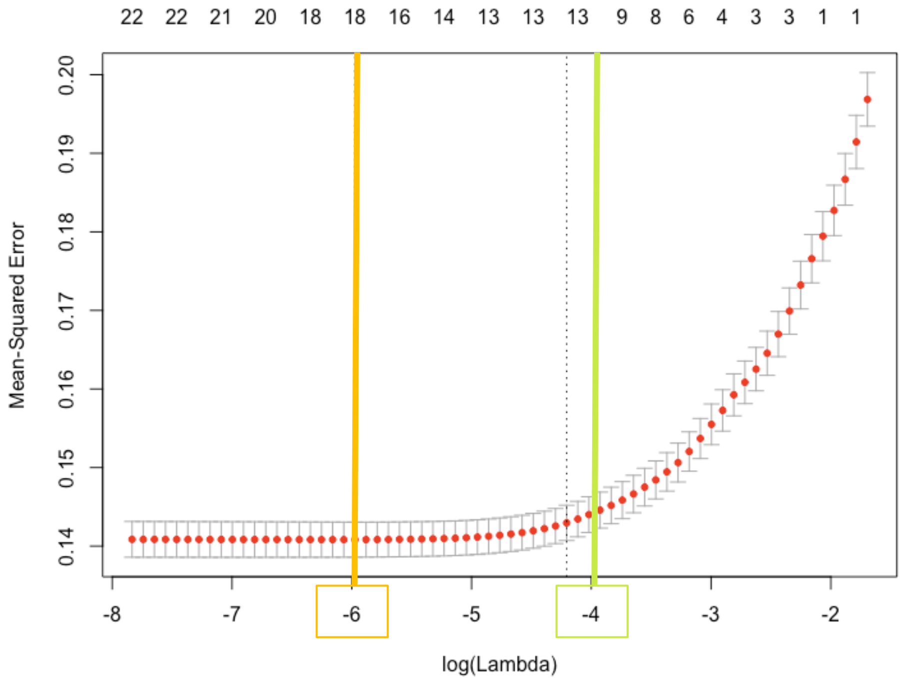
```
                               Estimate Std. Error z value Pr(>|z|)    
(Intercept)                  -2.865e+00  2.819e-01 -10.165  < 2e-16 ***
gender                        7.109e-02  7.749e-02   0.917 0.358939    
SeniorCitizen                 1.058e-01  1.015e-01   1.042 0.297202    
Partner                       3.189e-02  9.161e-02   0.348 0.727773    
Dependents                   -2.984e-01  1.073e-01  -2.781 0.005421 ** 
PhoneService                  3.911e-01  7.822e-01   0.500 0.617065    
PaperlessBilling             -3.169e-01  8.897e-02  -3.562 0.000368 ***
MonthlyCharges               -3.285e-02  3.834e-02  -0.857 0.391535    
TotalCharges                 -3.704e-04  3.375e-05 -10.976  < 2e-16 ***
MultipleLines_Yes             5.405e-01  2.125e-01   2.543 0.010989 *  
InternetService_DSL           1.681e+00  9.734e-01   1.727 0.084201 .  
InternetService_Fiber_optic   3.485e+00  1.928e+00   1.808 0.070631 .  
InternetService_No                   NA         NA      NA       NA    
OnlineSecurity_Yes           -2.136e-01  2.173e-01  -0.983 0.325636    
OnlineBackup_Yes              8.951e-02  2.115e-01   0.423 0.672213    
DeviceProtection_Yes          2.515e-01  2.135e-01   1.178 0.238883    
TechSupport_Yes              -1.041e-01  2.198e-01  -0.473 0.635861    
StreamingTV_Yes               5.955e-01  3.930e-01   1.515 0.129710    
StreamingMovies_Yes           7.518e-01  3.933e-01   1.911 0.055960 .  
Contract_Month_to_month       1.724e+00  2.061e-01   8.365  < 2e-16 ***
Contract_One_year             9.306e-01  2.097e-01   4.437  9.1e-06 ***
Contract_Two_year                    NA         NA      NA       NA    
PaymentMethod_Bank_transfer  -4.094e-02  1.339e-01  -0.306 0.759779    
PaymentMethod_Credit_card    -1.301e-01  1.362e-01  -0.955 0.339446    
PaymentMethod_Electronic_check  2.824e-01  1.140e-01   2.477 0.013255 *  
PaymentMethod_Mailed_check           NA         NA      NA       NA    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5705.6  on 4892  degrees of freedom
Residual deviance: 4115.0  on 4870  degrees of freedom
AIC: 4161

Number of Fisher Scoring iterations: 6
```

Logistic + LASSO



Minimum MSE at -6

More recommended model at -4

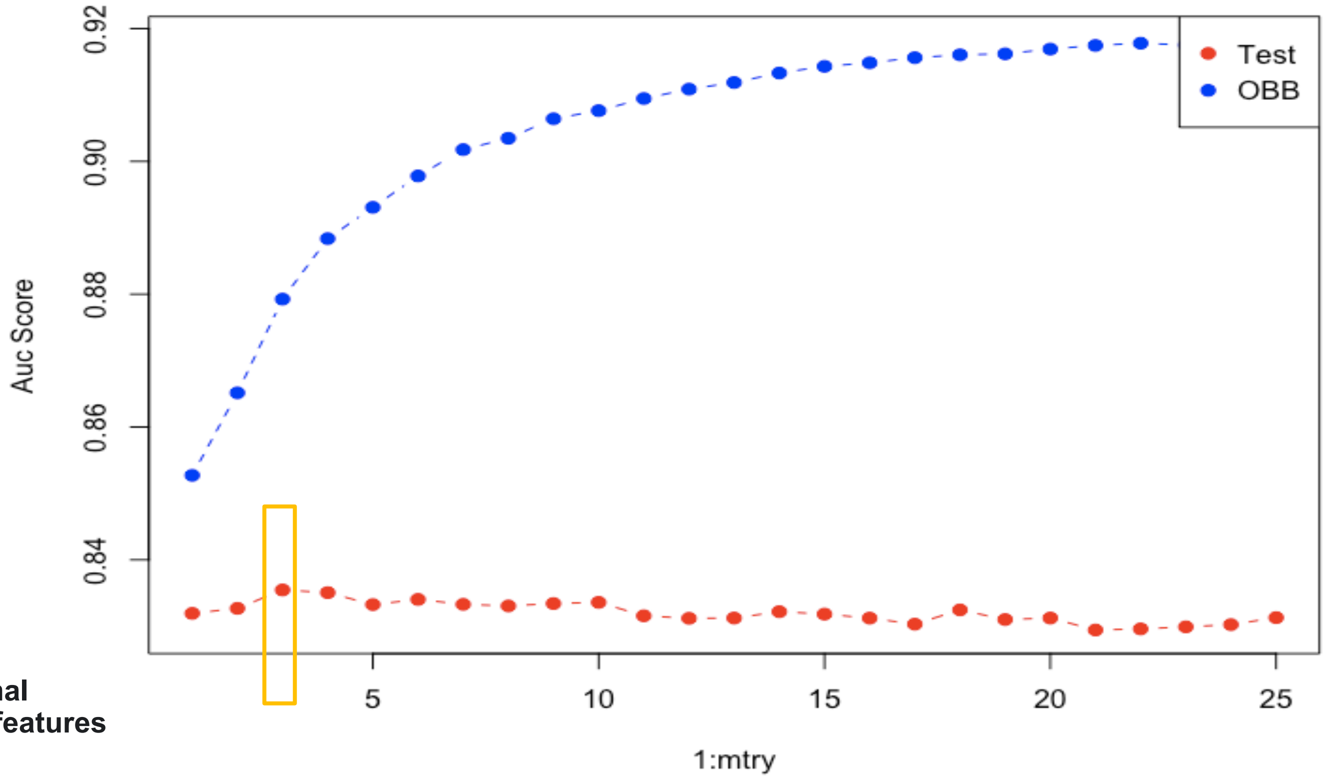
```
26 x 1 sparse Matrix of class "dgCMatrix"
                                             1
(Intercept)                       2.240950e-01
gender                            .
SeniorCitizen                     8.873877e-03
Partner                           .
Dependents                       -1.984315e-02
PhoneService                      .
PaperlessBilling                 -3.398664e-02
MonthlyCharges                    .
TotalCharges                     -4.728939e-05
MultipleLines_Yes                 1.756528e-02
InternetService_DSL               .
InternetService_Fiber_optic       1.734897e-01
InternetService_No               -1.095859e-01
OnlineSecurity_Yes               -2.877791e-02
OnlineBackup_Yes                  .
DeviceProtection_Yes              .
TechSupport_Yes                  -1.047338e-02
StreamingTV_Yes                   1.165585e-02
StreamingMovies_Yes               3.939471e-02
Contract_Month_to_month           1.458370e-01
Contract_One_year                 .
Contract_Two_year                 .
PaymentMethod_Bank_transfer       .
PaymentMethod_Credit_card         .
PaymentMethod_Electronic_check    7.056601e-02
PaymentMethod_Mailed_check        .
```

Random Forest Model Training:
Change some features of Random Forest

N_estimators:
The number of trees in the forest.
Change default value from 10 to 1000

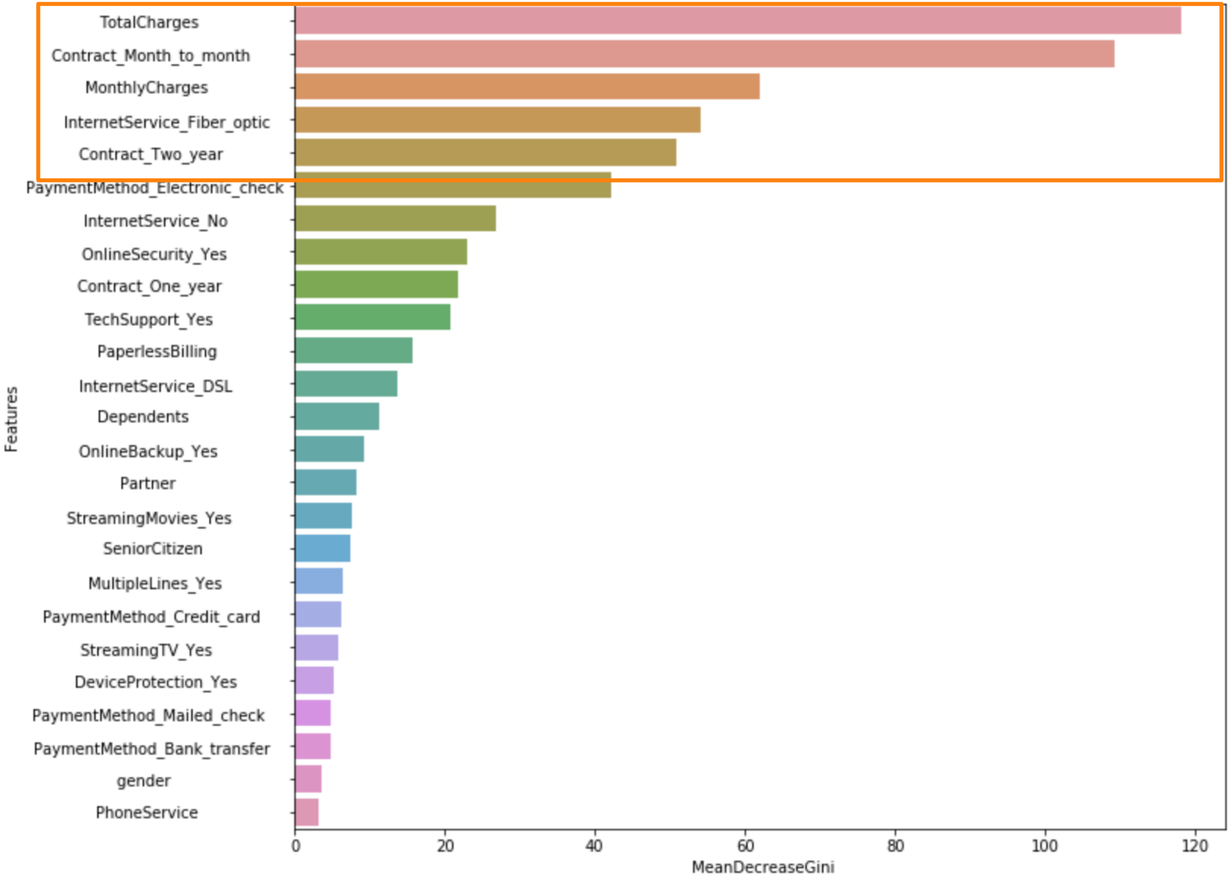
Min_samples_leaf:
The minimum number of samples required to be at a leaf node.
Change default value from 2 to 35

Max_features
The number of features to consider when looking for the best split
Change default value to 3

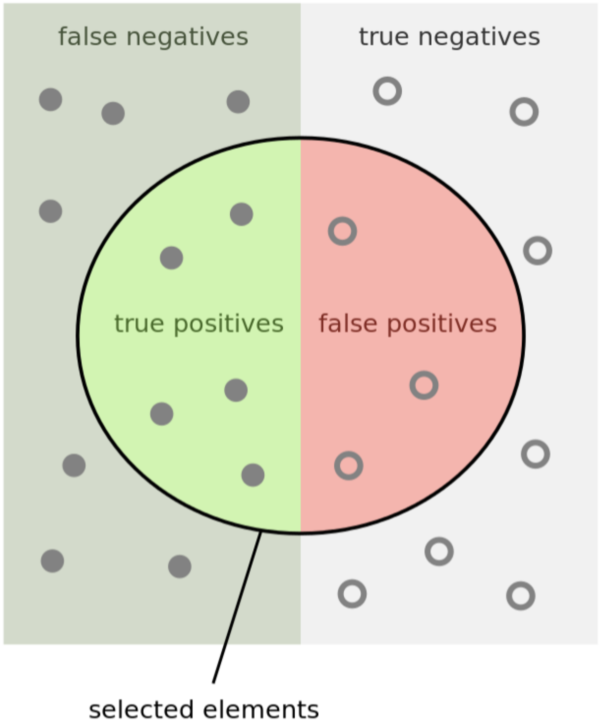


Optimal Max_features = 3

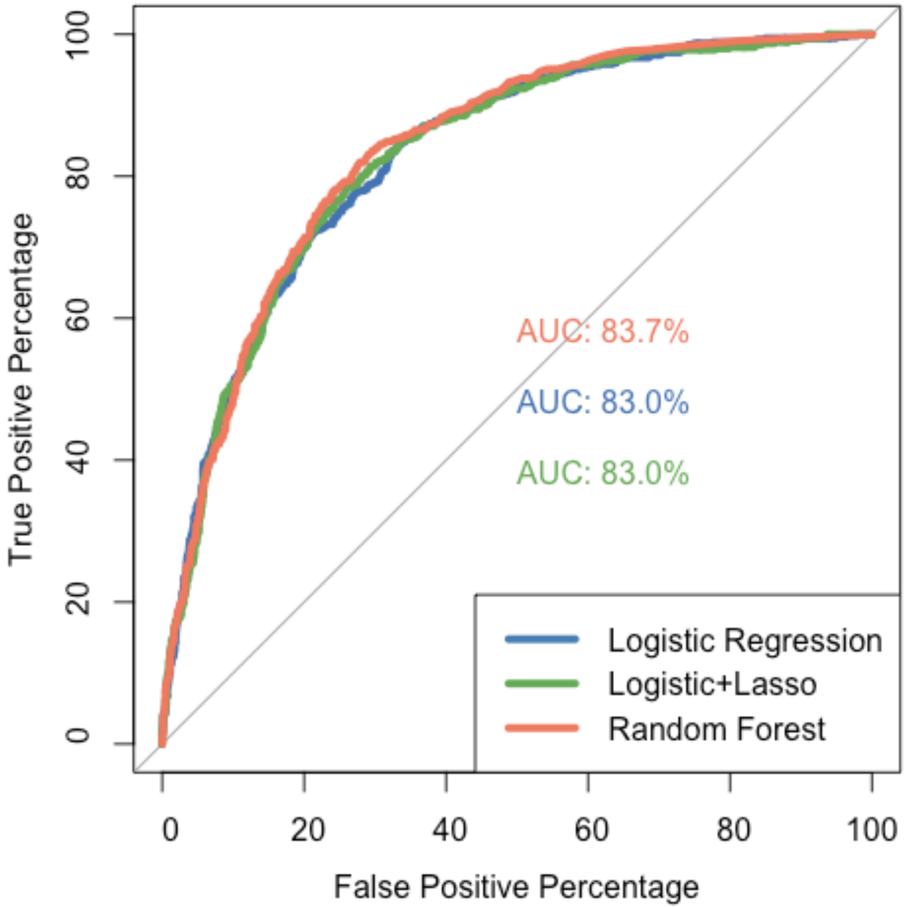
Random Forest



ROC _ AUC



false negatives
true negatives
true positives
false positives
selected elements

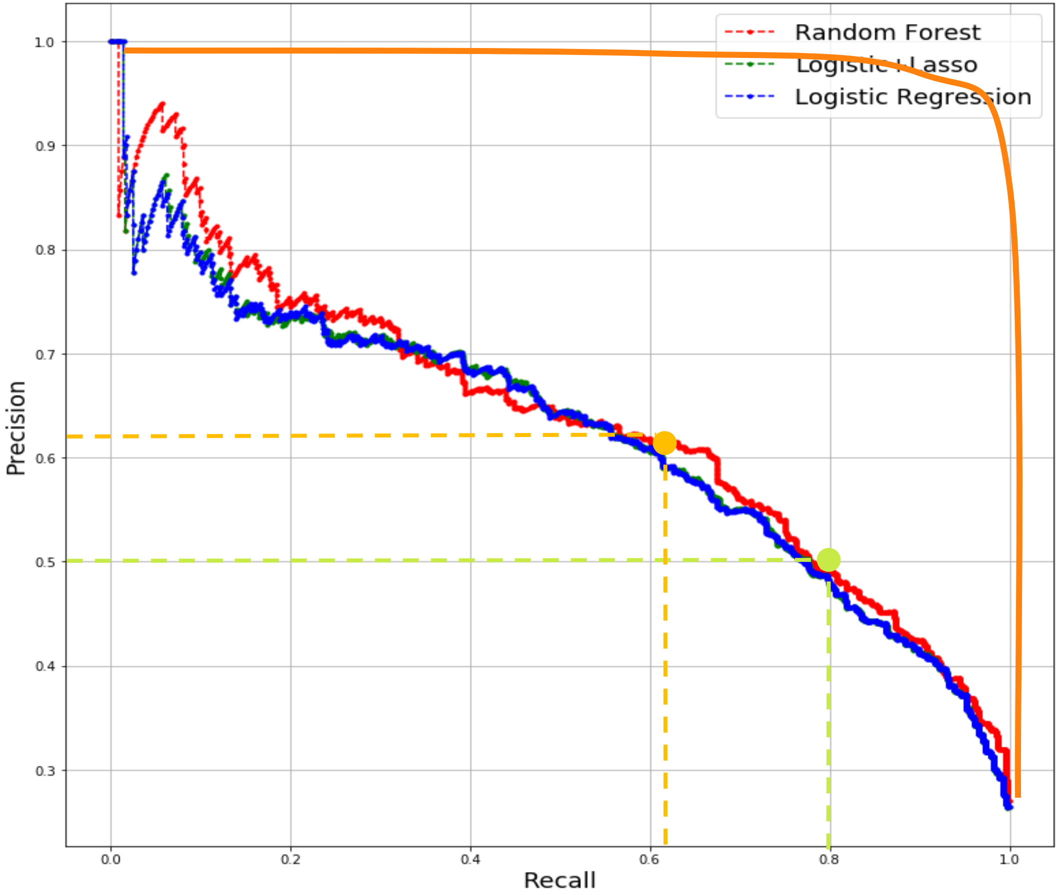


AUC: 83.7%
AUC: 83.0%
AUC: 83.0%
Logistic Regression
Logistic+Lasso
Random Forest
True Positive Percentage
False Positive Percentage
0
20
40
60
80
100

Precision

$$\textbf{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\textbf{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$



Implications

Offer Monthly Discount of 2.2% to Break Even
Do Nothing to Maximize Net Profit of $561,324

Avg Churn Month	18
Avg Lifetime	32
Discount	2.24%
Revenue/month	$ 65.00
Gross Margin	0.55
Gross Profit	$ 35.75

Recall	0.6
Precision	0.62
Sample Size	7043
Chun Rate	26.54%

TP	1122
FP	687
FN	748

	Cost	Num	Total Cost	Margin	Num	Total Margin	Net Profit
True Positive	$ (11.23)	1122	$ (12,591.50)	$ 500.50	1122	$ 561,324.42	$ 548,732.92
True Negative							$ -
False Positive	$ (11.23)	687	$ (7,717.37)				$ (7,717.37)
False Negative				$ (723.59)	748	$ (541,014.73)	$ (541,014.73)
							$ 0.82

TP: We predict the customer will leave, and they would have, but we stopped them.
TN: We predict the customer will leave, and they do.
FP: We predict the customer will leave. they weren't going to, but we gave them discount.
FN: We predict the customer would stay, but they left.

Offer Monthly Discount of 5.5% to Break Even
Do Nothing to Maximize Net Profit of $748,433

Avg Churn Month	18
Avg Lifetime	32
Discount	5.52%
Revenue/month	$ 65.00
Gross Margin	0.55
Gross Profit	$ 35.75

Recall	0.8
Precision	0.5
Sample Size	7043
Chun Rate	26.54%

TP	1495
FP	1495
FN	374

	Cost	Num	Total Cost	Margin	Num	Total Margin	Net Profit
True Positive	$ (27.63)	1495	$ (41,321.33)	$ 500.50	1495	$ 748,432.56	$ 707,111.24
True Negative							$ -
False Positive	$ (27.63)	1495	$ (41,321.33)				$ (41,321.33)
False Negative				$ (1,780.94)	374	$ (665,789.91)	$ (665,789.91)
							$ (0.00)

TP: We predict the customer will leave, and they would have, but we stopped them.
TN: We predict the customer will leave, and they do.
FP: We predict the customer will leave. they weren't going to, but we gave them discount.
FN: We predict the customer would stay, but they left.

- Marketers should be careful with the tradeoff between precision and recall.
- We recommend future tuning out prediction model before we offer discount to retain customers.

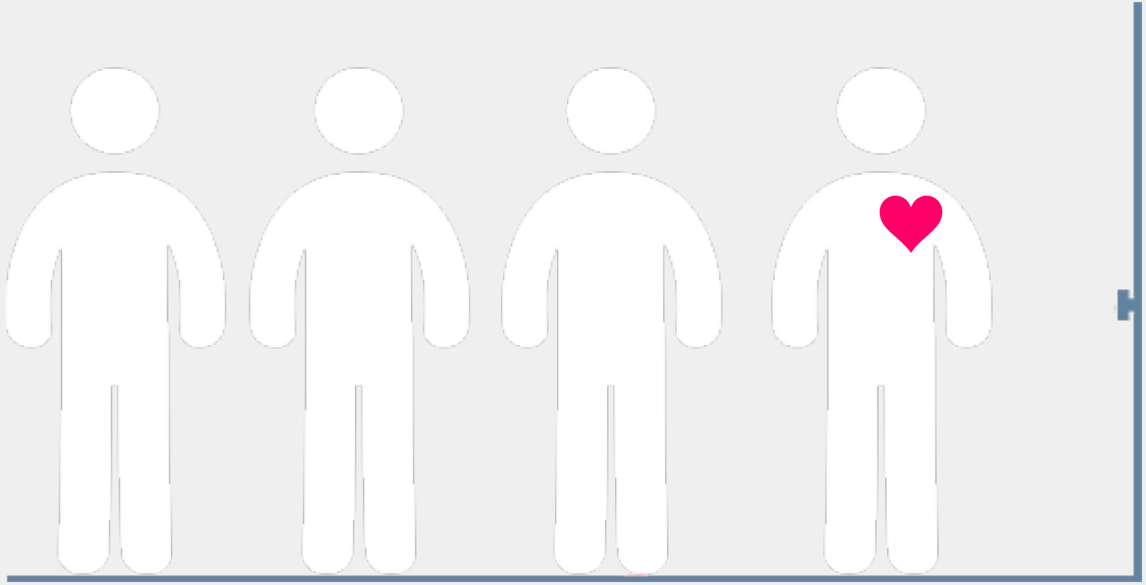
Summary

Limitations

- Limited Data (7,043 observations with 26 variables)
- Imbalanced Data (26.54% of churned customers)
- Bias: a point in time
- More features and more data to train model
- Not possible to retain high precision when aiming high recall

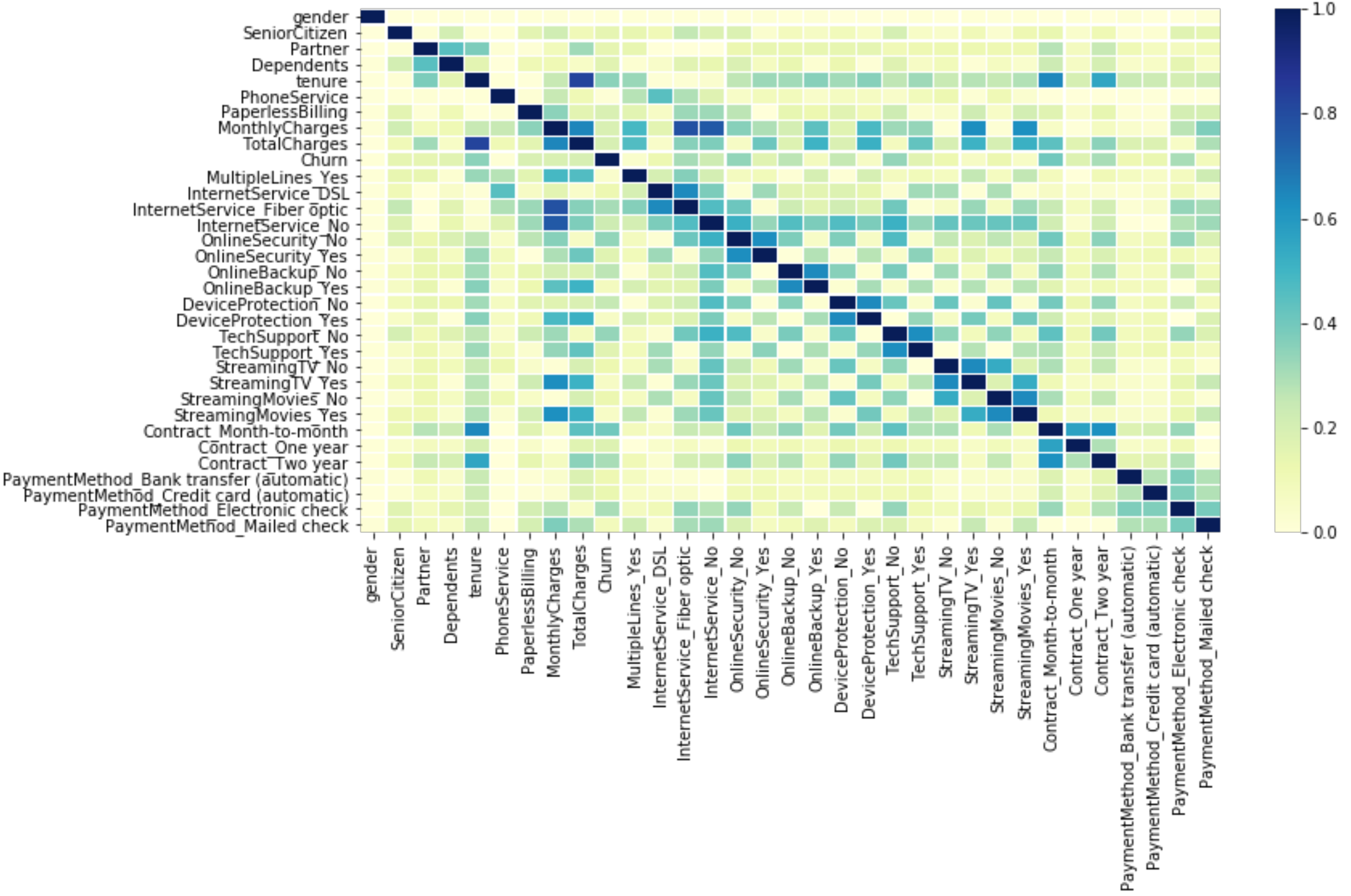
Conclusions

- Careful revision before implementing any actions on churned customers



Thank You

Appendix


gender
SeniorCitizen
Partner
Dependents
tenure
PhoneService
PaperlessBilling
MonthlyCharges
TotalCharges
Churn
MultipleLines_Yes
InternetService_DSL
InternetService_Fiber optic
InternetService_No
OnlineSecurity_No
OnlineSecurity_Yes
OnlineBackup_No
OnlineBackup_Yes
DeviceProtection_No
DeviceProtection_Yes
TechSupport_No
TechSupport_Yes
StreamingTV_No
StreamingTV_Yes
StreamingMovies_No
StreamingMovies_Yes
Contract_Month-to-month
Contract_One year
Contract_Two year
PaymentMethod_Bank transfer (automatic)
PaymentMethod_Credit card (automatic)
PaymentMethod_Electronic check
PaymentMethod_Mailed check
1.0
0.8
0.6
0.4
0.2
0.0

Highly Correlated

Total Charges & Tenure

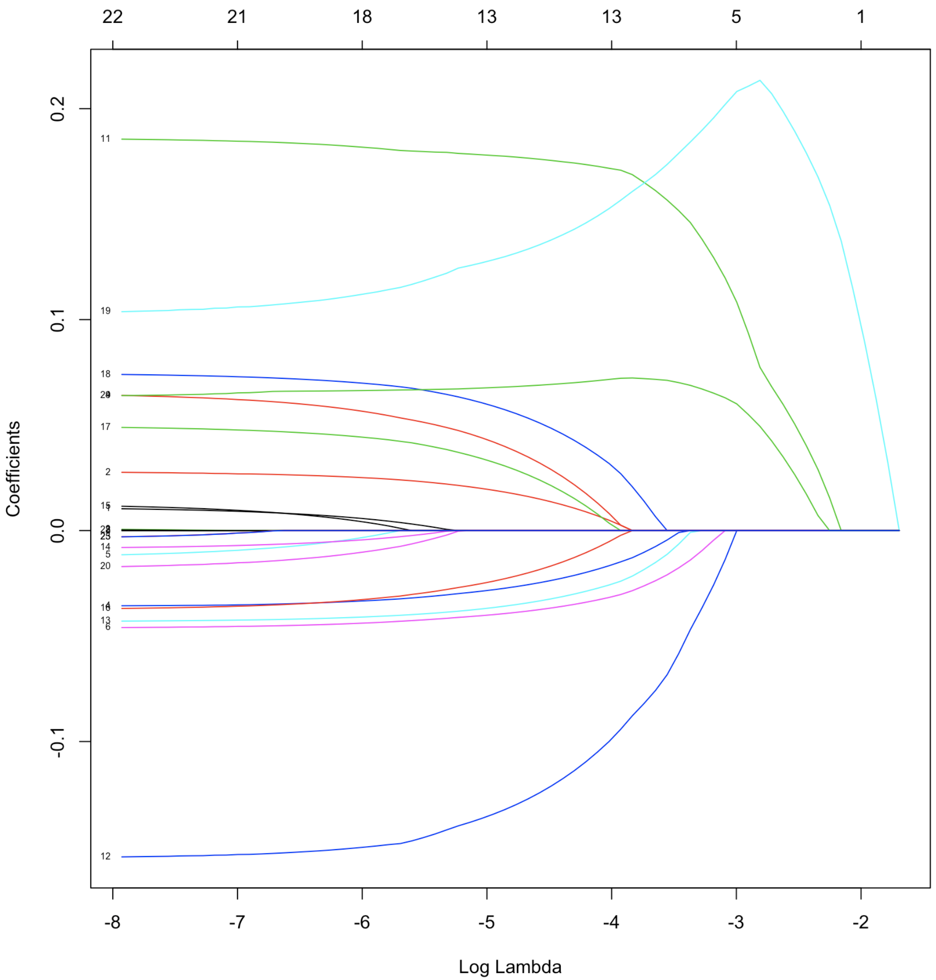
In [722]:
```
correlated_features = set()
correlation_matrix = df.corr()
```

In [723]:
```
for i in range(len(correlation_matrix.columns)):
    for j in range(i):
        if abs(correlation_matrix.iloc[i, j]) > 0.8:
            colname = correlation_matrix.columns[i]
            colname2 = correlation_matrix.columns[j]
            correlated_features.add(colname)
            correlated_features.add(colname2)
```

In [724]:
```
print(correlated_features)
```

```
{'TotalCharges', 'tenure'}
```

Logistic LASSO



Avg Churn Month	18
Avg Lifetime	32
Discount	0.00%
Revenue/month	$ 65.00
Gross Margin	0.55
Gross Profit	$ 35.75

Recall	0.6
Precision	0.62
Sample Size	7043
Chun Rate	26.54%

TP	1122
FP	687
FN	748

	Cost	Num	Total Cost	Margin	Num	Total Margin	Net Profit
True Positive	$ -	1122	$ -	$ 500.50	1122	$ 561,324.42	$ 561,324.42
True Negative							$ -
False Positive	$ -	687	$ -				$ -
False Negative				$ -	748	$ -	$ -
							$ 561,324.42

TP: We predict the customer will leave, and they would have, but we stopped them.

TN: We predict the customer will leave, and they do.

FP: We predict the customer will leave. they weren't going to, but we gave them discount.

FN: We predict the customer would stay, but they left.

Avg Churn Month	18
Avg Lifetime	32
Discount	0.00%
Revenue/month	$ 65.00
Gross Margin	0.55
Gross Profit	$ 35.75

Recall	0.8
Precision	0.5
Sample Size	7043
Chun Rate	26.54%

TP	1495
FP	1495
FN	374

	Cost	Num	Total Cost	Margin	Num	Total Margin	Net Profit
True Positive	$ -	1495	$ -	$ 500.50	1495	$ 748,432.56	$ 748,432.56
True Negative							$ -
False Positive	$ -	1495	$ -				$ -
False Negative				$ -	374	$ -	$ -
							$ 748,432.56

TP: We predict the customer will leave, and they would have, but we stopped them.

TN: We predict the customer will leave, and they do.

FP: We predict the customer will leave. they weren't going to, but we gave them discount.

FN: We predict the customer would stay, but they left.