Should I Stay or Should I Go?

Telecom Customer Retention Analysis

Sawyer Shader-Seave

01.04.21

Scenario

As a consultant for a telecom company how can I get more existing customers to stay and potentially attract new customer?

Data

- Data: telecom churn data (provided)
- Source: https://www.kaggle.com/becksddf/churn-in-telecoms-dataset
- 3300 observations with 20 atributes (including churn)
- Target Variable: Churn
- 2,850 out of 3300 observations are False

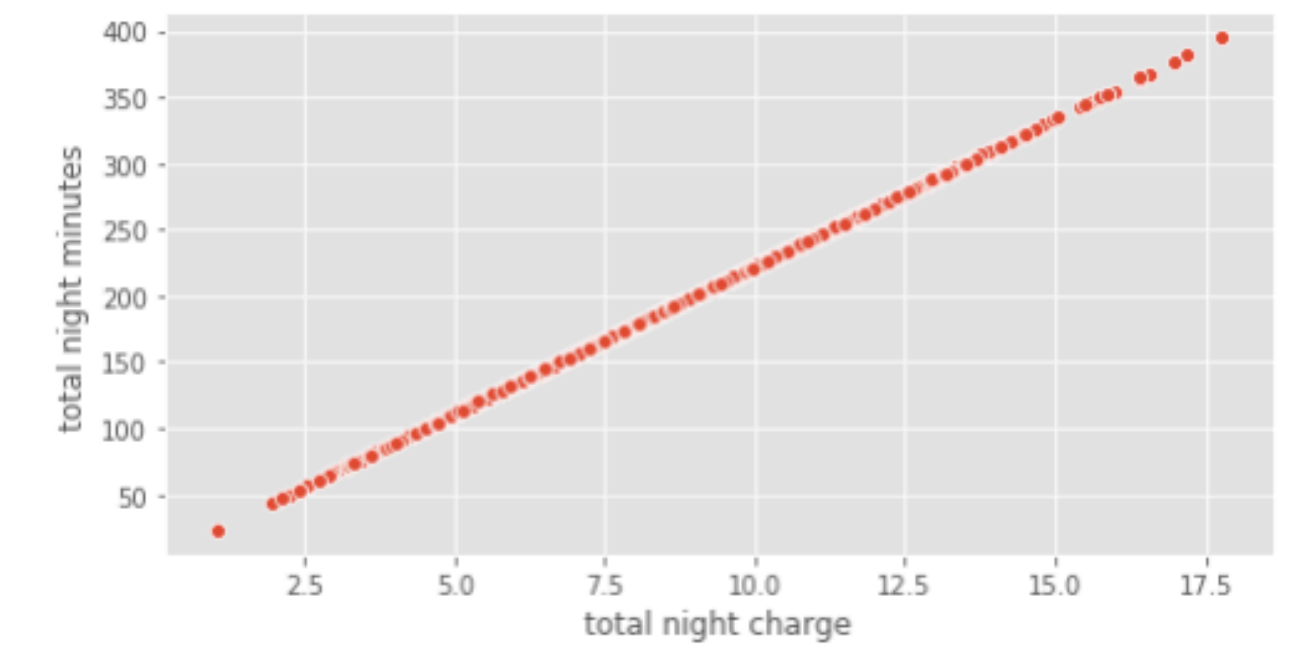
```
Column
------
state
account length
area code
phone number
international plan
voice mail plan
number vmail messages
total day minutes
total day calls
total day charge
total eve minutes
total eve calls
total eve charge
total night minutes
total night calls
total night charge
total intl minutes
total intl calls
total intl charge
customer service calls
churn
```

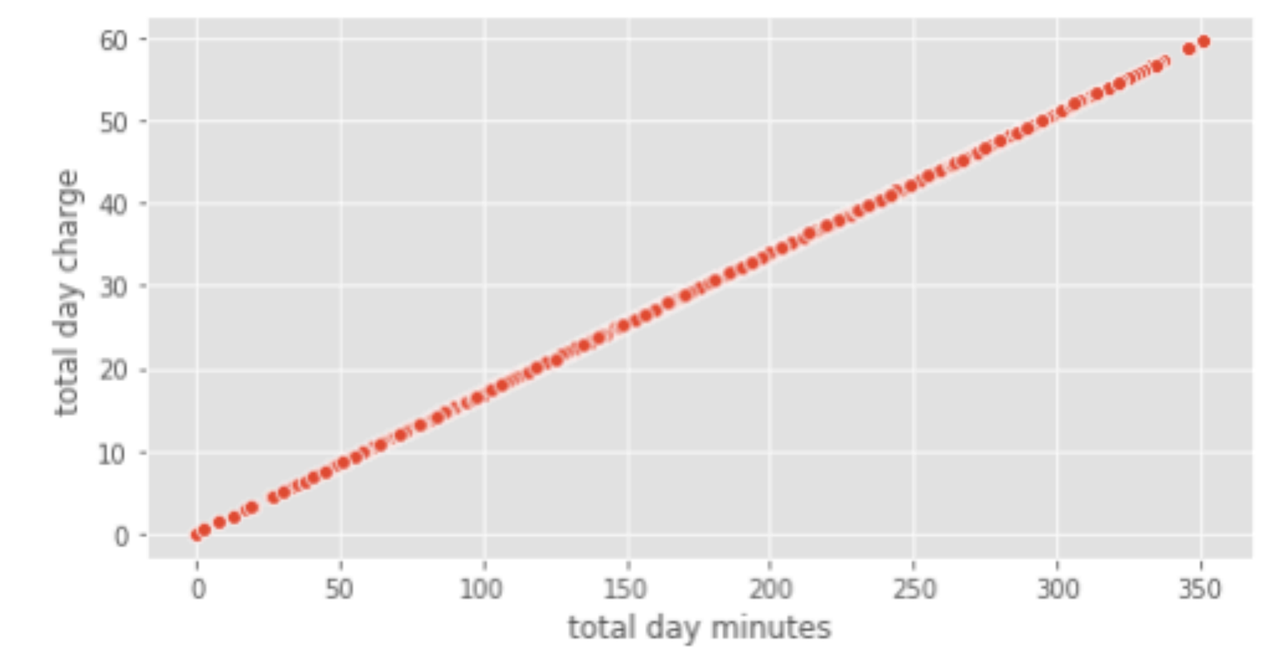
EDA

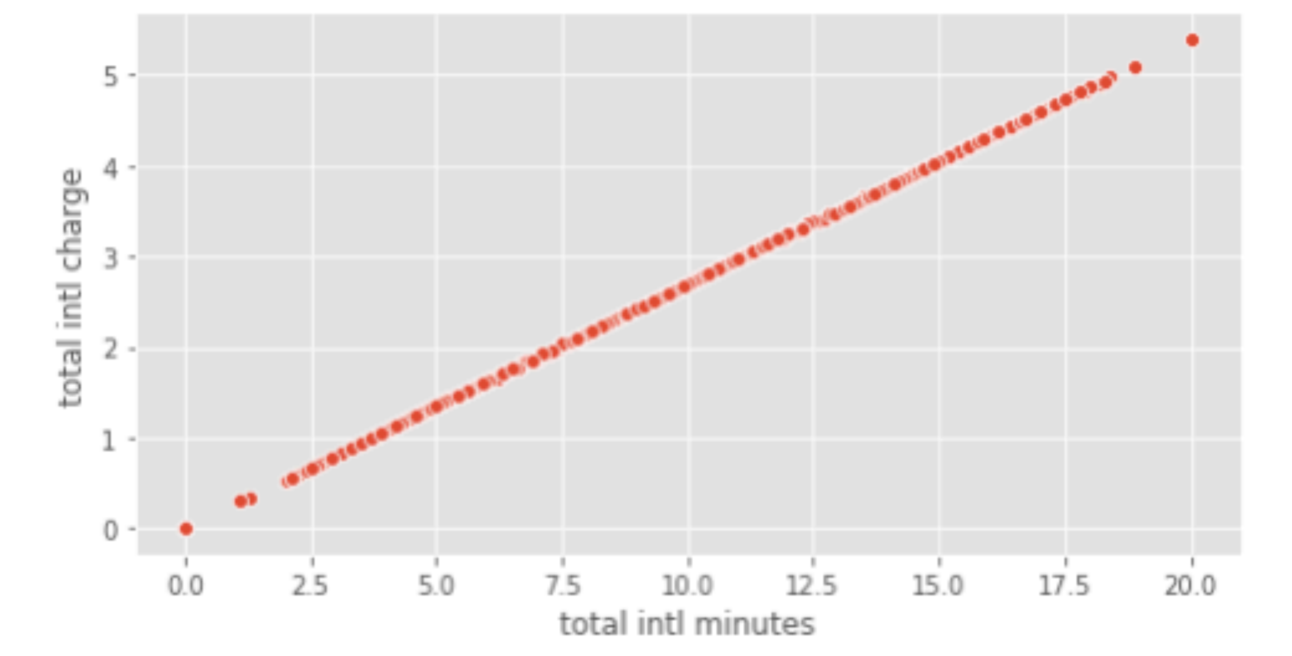
- None of the date is very correlated with churn
- Much of the data is highly correlated
- Converted states and area code to dummies
- Dropped phone number

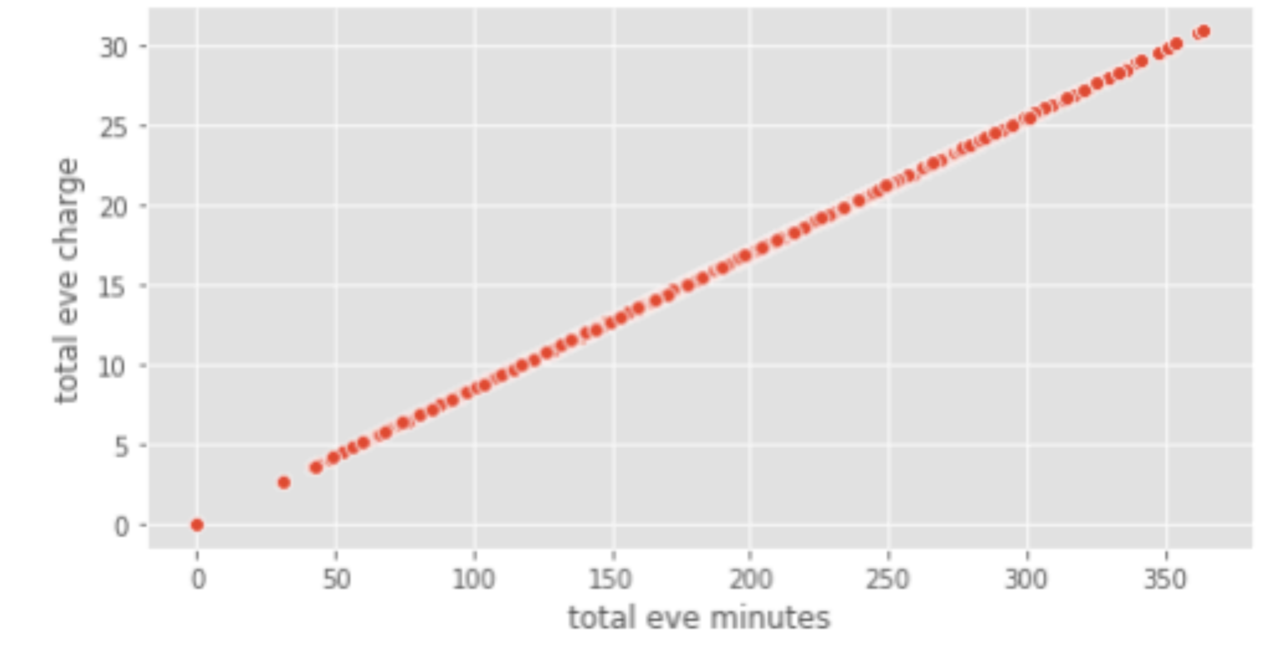
```
In [99]: df2.corr()['churn'].sort_values(ascending=False)

Out[99]: churn                       1.000000
         international plan          0.259852
         customer service calls      0.208750
         total day charge            0.205151
         total eve minutes           0.092796
                                       ...
         state_cat_HI               -0.031890
         state_cat_AZ               -0.032759
         state_cat_VA               -0.034940
         total intl calls           -0.052844
         voice mail plan            -0.102148
```

Pairs	
(total day minutes, total day charge)	1.000000
(total eve minutes, total eve charge)	1.000000
(total night charge, total night minutes)	0.999999
(total intl minutes, total intl charge)	0.999993
(voice mail plan, number vmail messages)	0.956927







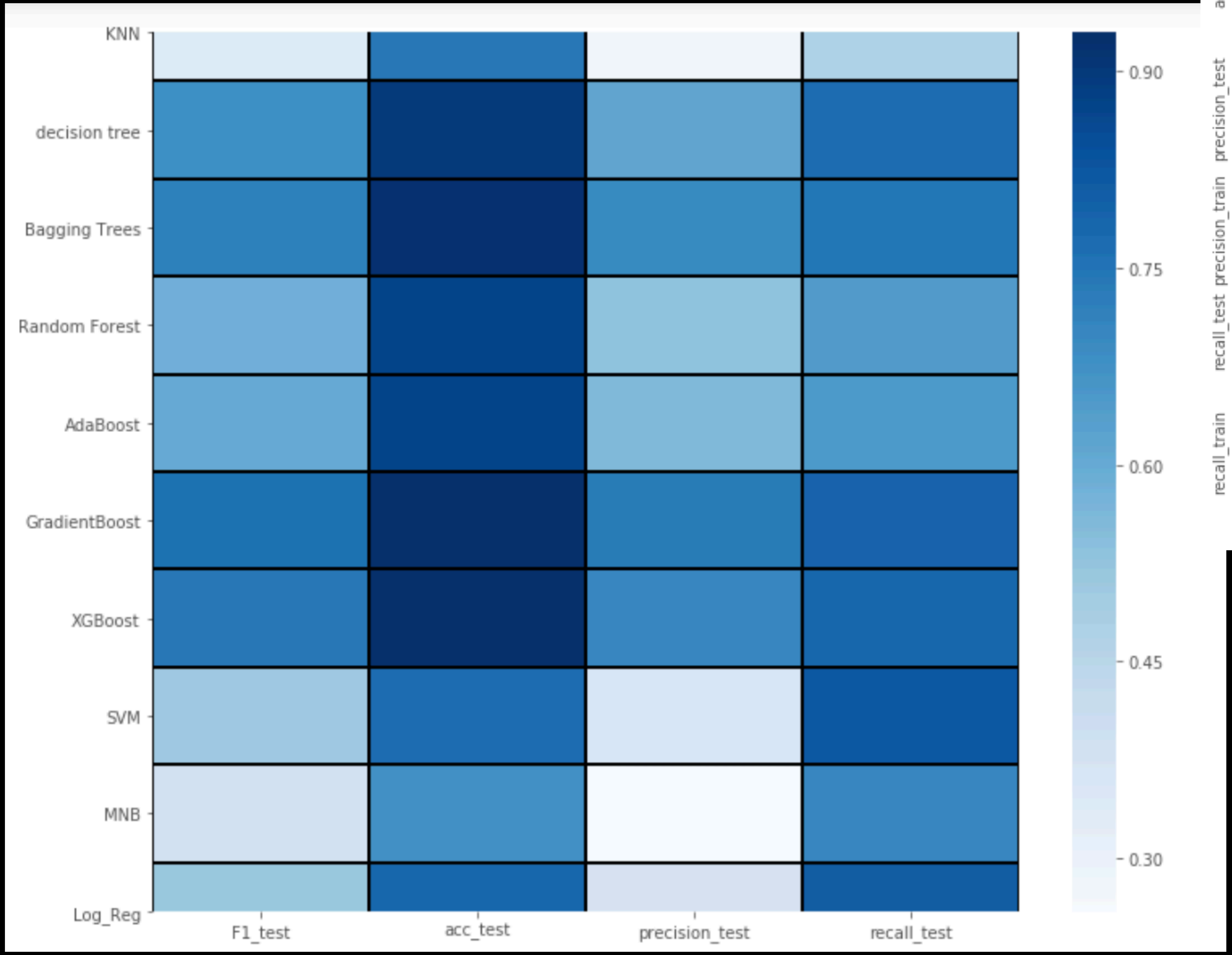


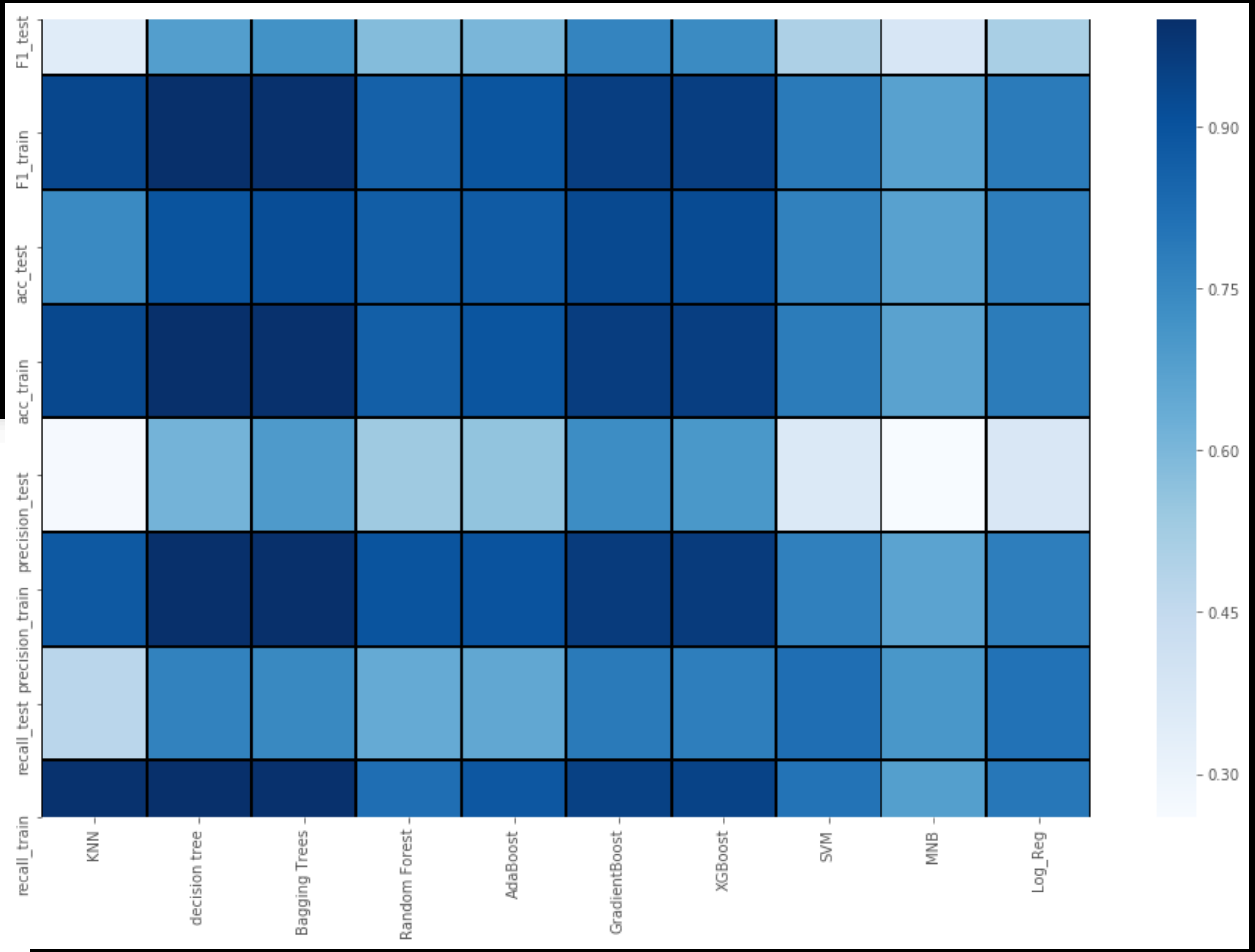
Initial Models

- KNN and Decision Tree over trained
- Gradient Boost and XGBoost tested the best
- KNN = baseline model

	F1_test	F1_train	acc_test	acc_train	precision_test	precision_train	recall_test	recall_train
KNN	0.344828	0.935344	0.743628	0.931299	0.271084	0.883340	0.473684	0.993854
decision tree	0.682243	1.000000	0.898051	1.000000	0.613445	1.000000	0.768421	1.000000
Bagging [illegible]	0.717172	0.996037	0.916042	0.996049	0.689320	0.999117	[illegible]	0.992976
Random Forest	0.583732	0.858519	[illegible]	[illegible]	[illegible]	0.898704	0.642105	0.821773
[illegible]	0.601942	0.894040	0.877061	0.894644	0.558559	0.899201	0.652632	[illegible]
GradientBoost	0.761421	0.959328	0.929535	0.959614	0.735294	0.966162	0.789474	0.952590
XGBoost	0.740000	0.957216	0.922039	0.957638	0.704762	0.966861	0.778947	0.947761
SVM	[illegible]	0.790638	0.770615	0.785996	0.364486	[illegible]	[illegible]	0.808165
MNB	0.379603	0.672605	0.671664	0.669227	0.259690	0.665806	0.705263	0.679543
Log_Reg	0.511628	0.787511	0.779610	0.784899	0.373786	0.778063	0.810526	0.797191

Initial Models
Visuals

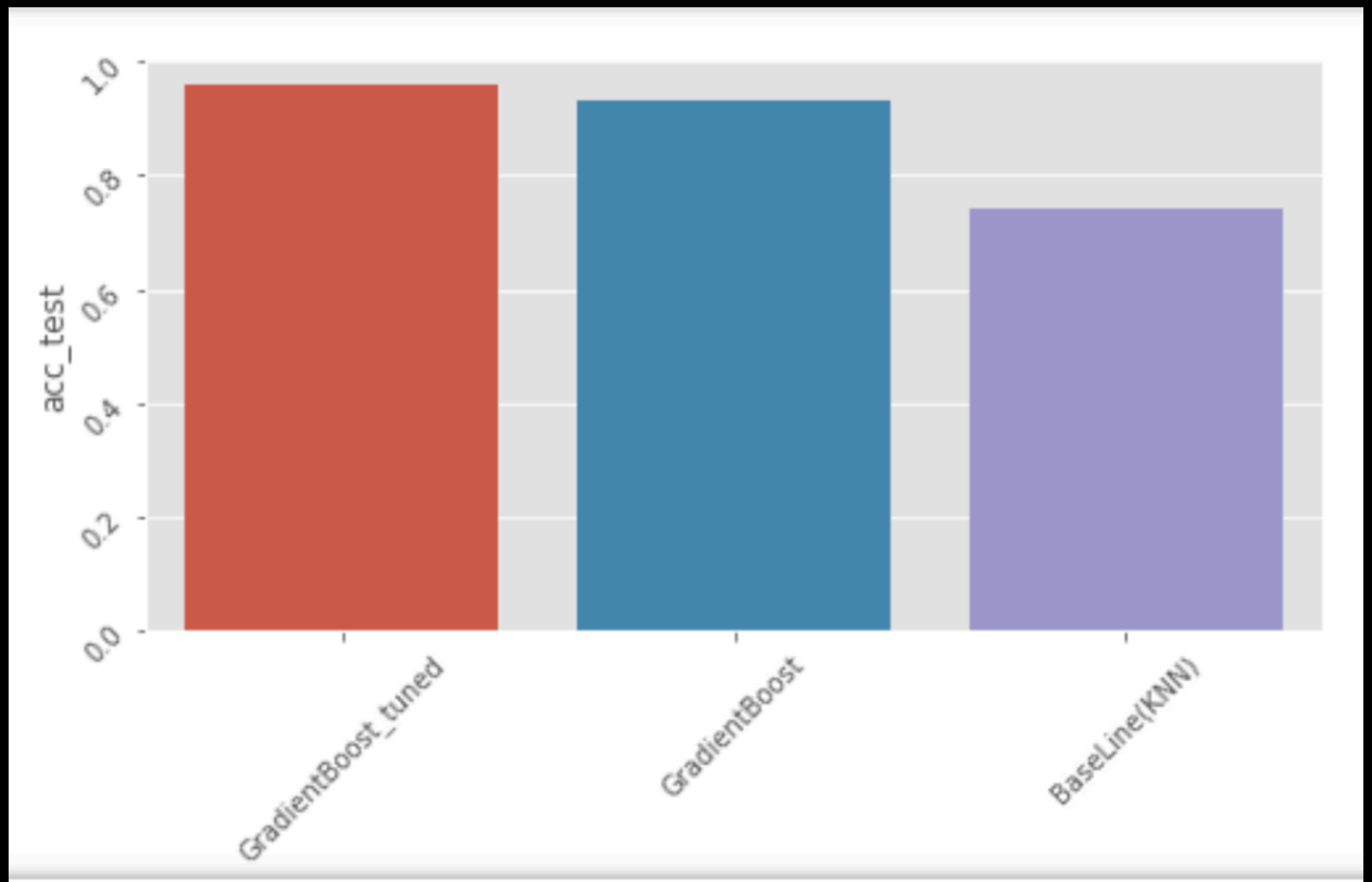


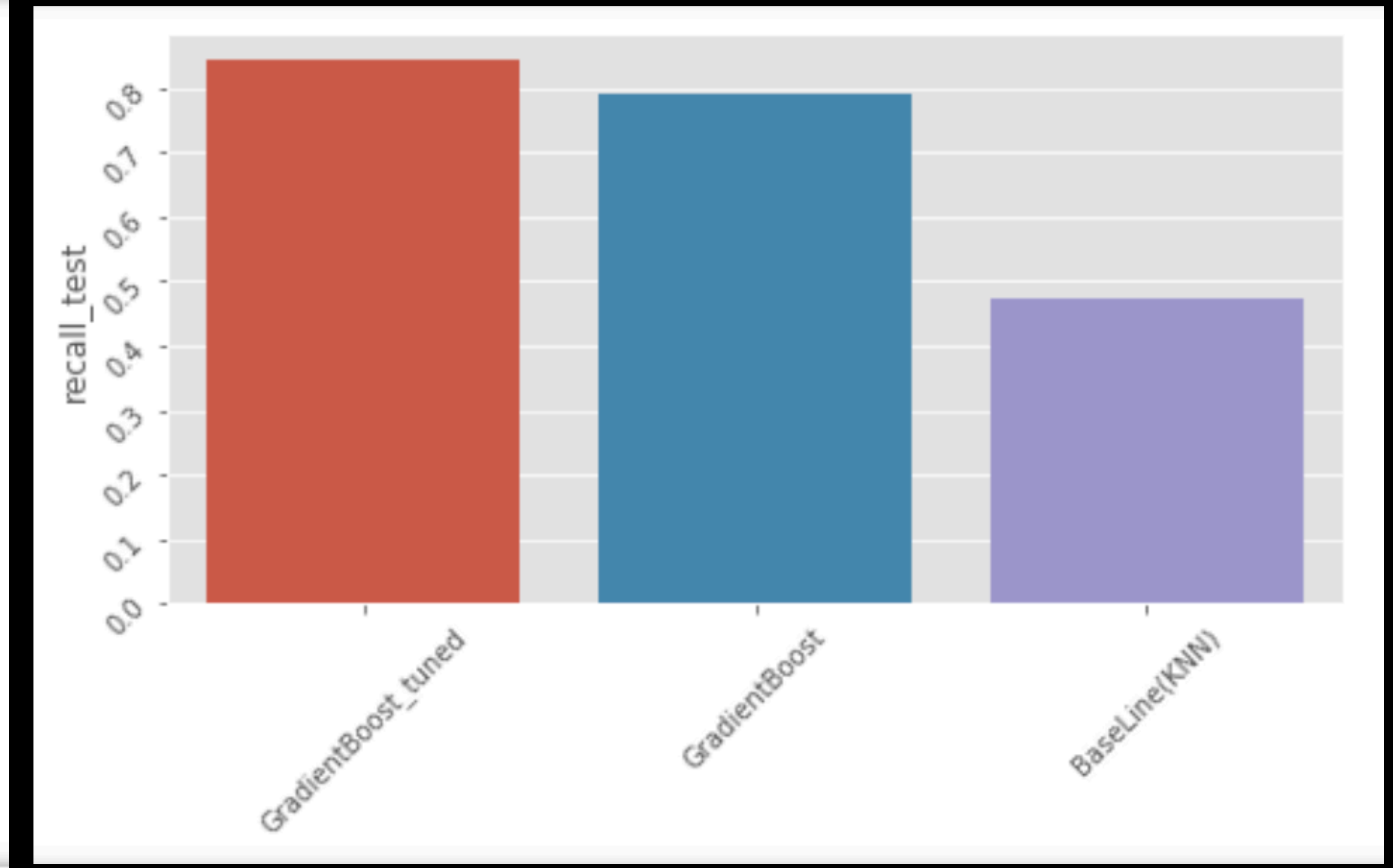


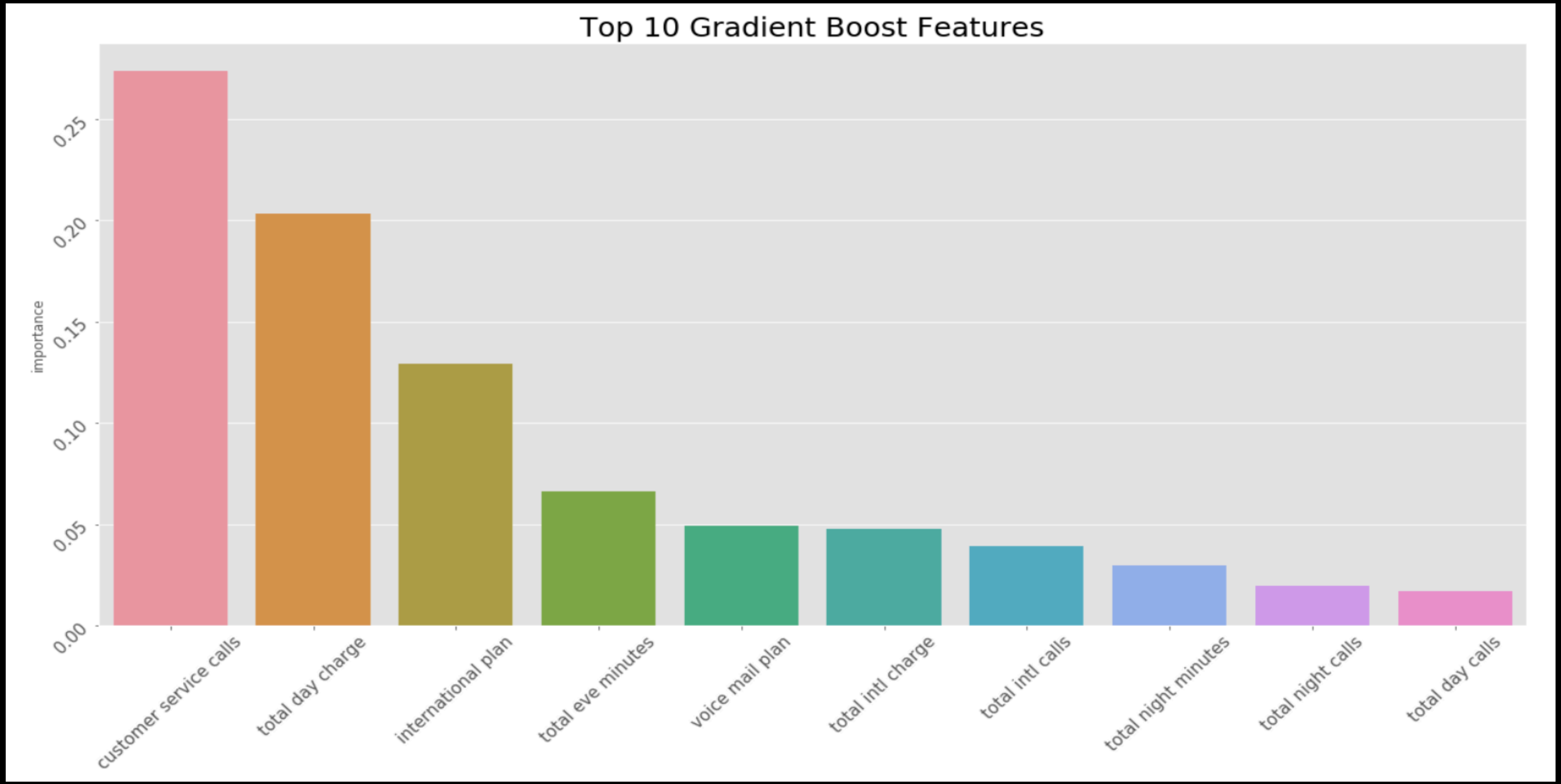
Final Models

Gradient Boost

- Accuracy: up 0.03 and 0.21 from original and baseline
- Recall: up 0.05 and 0.4 from original and baseline
- Top features: customer service calls, day charge, international plan

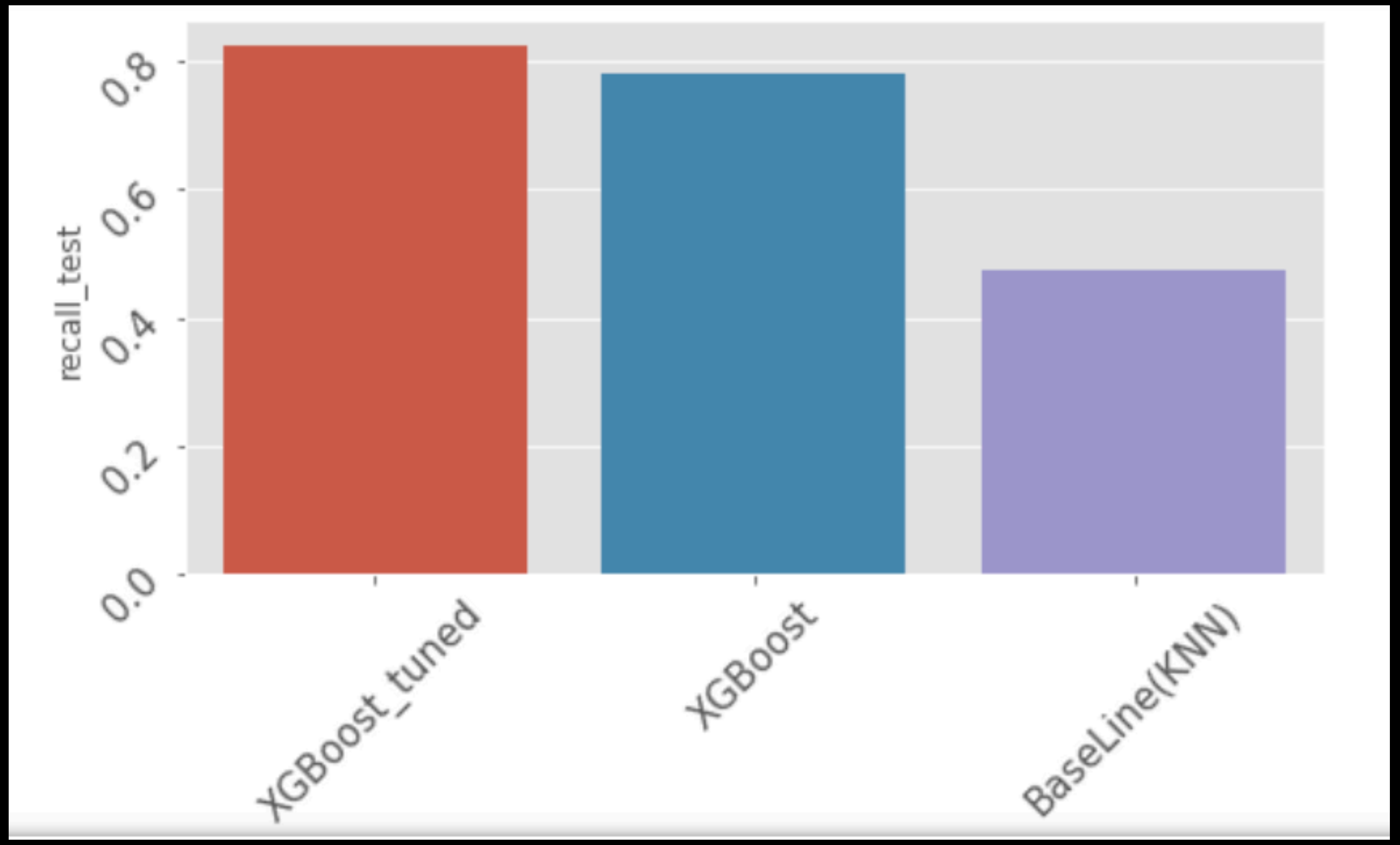


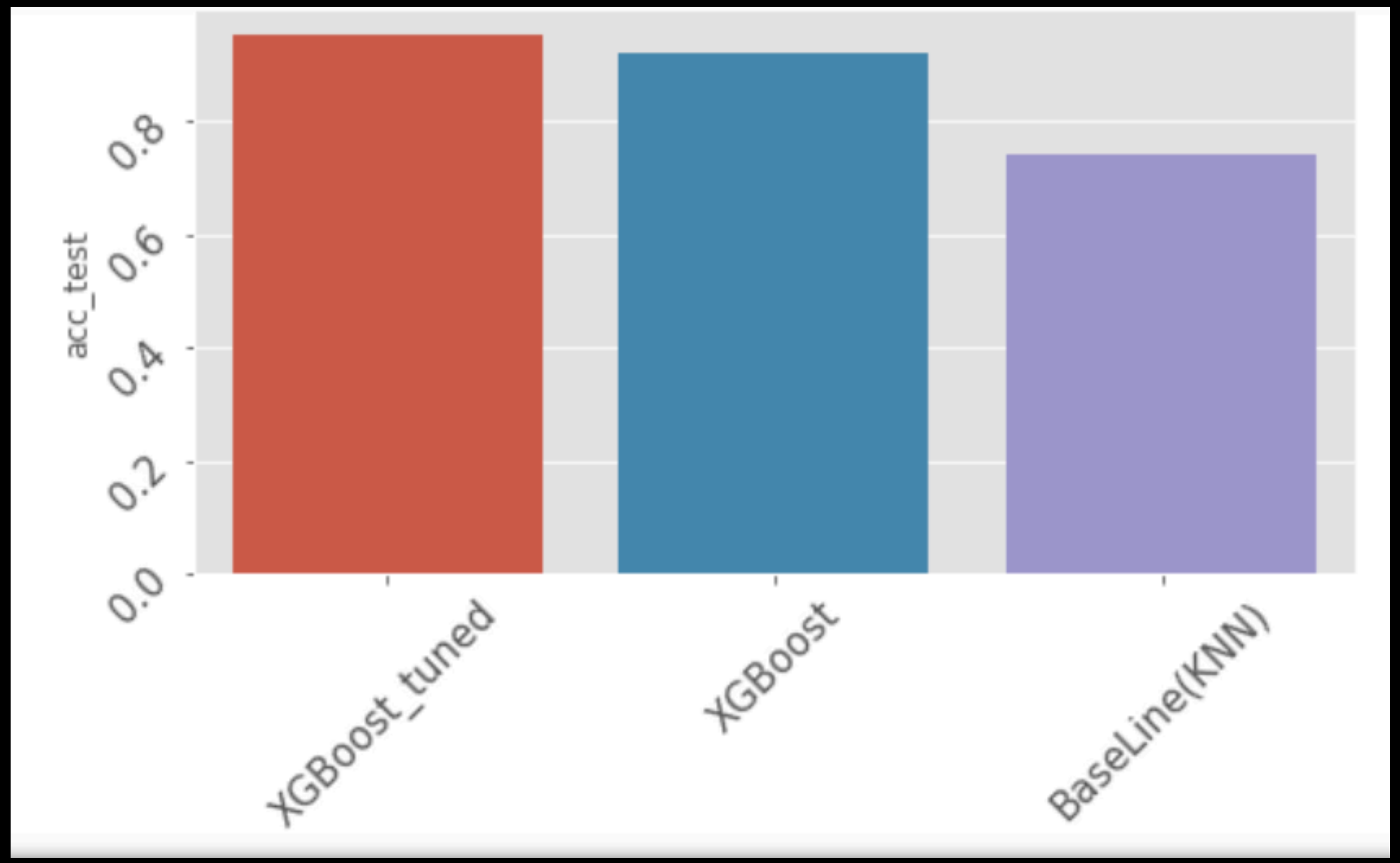


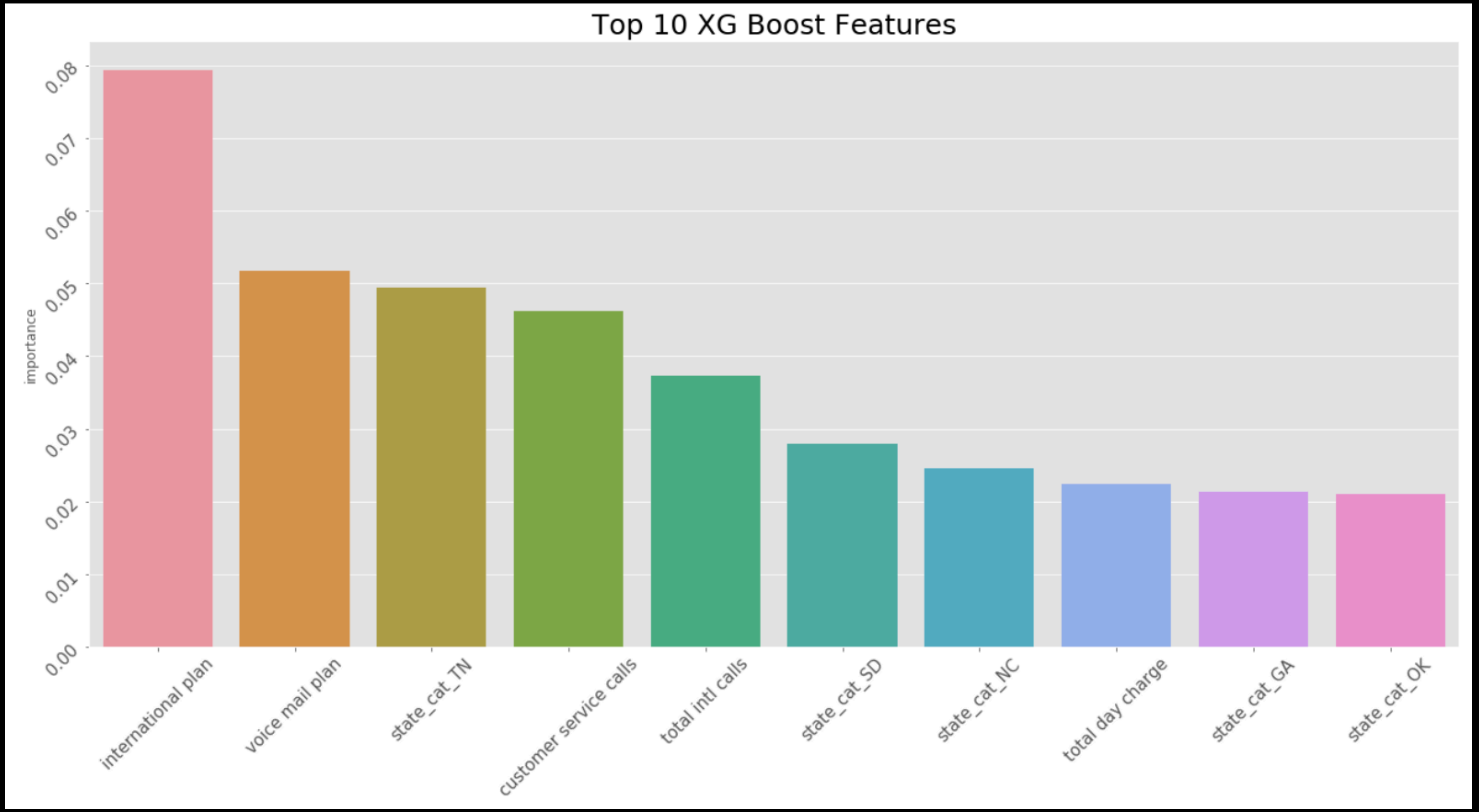


Final Models
XG Boost

- Accuracy: up 0.03 and 0.22 from original and baseline
- Recall: up 0.05 and 0.35 from original and baseline
- Top features: international plan, voicemail plan, Tennessee
- No feature over 0.10



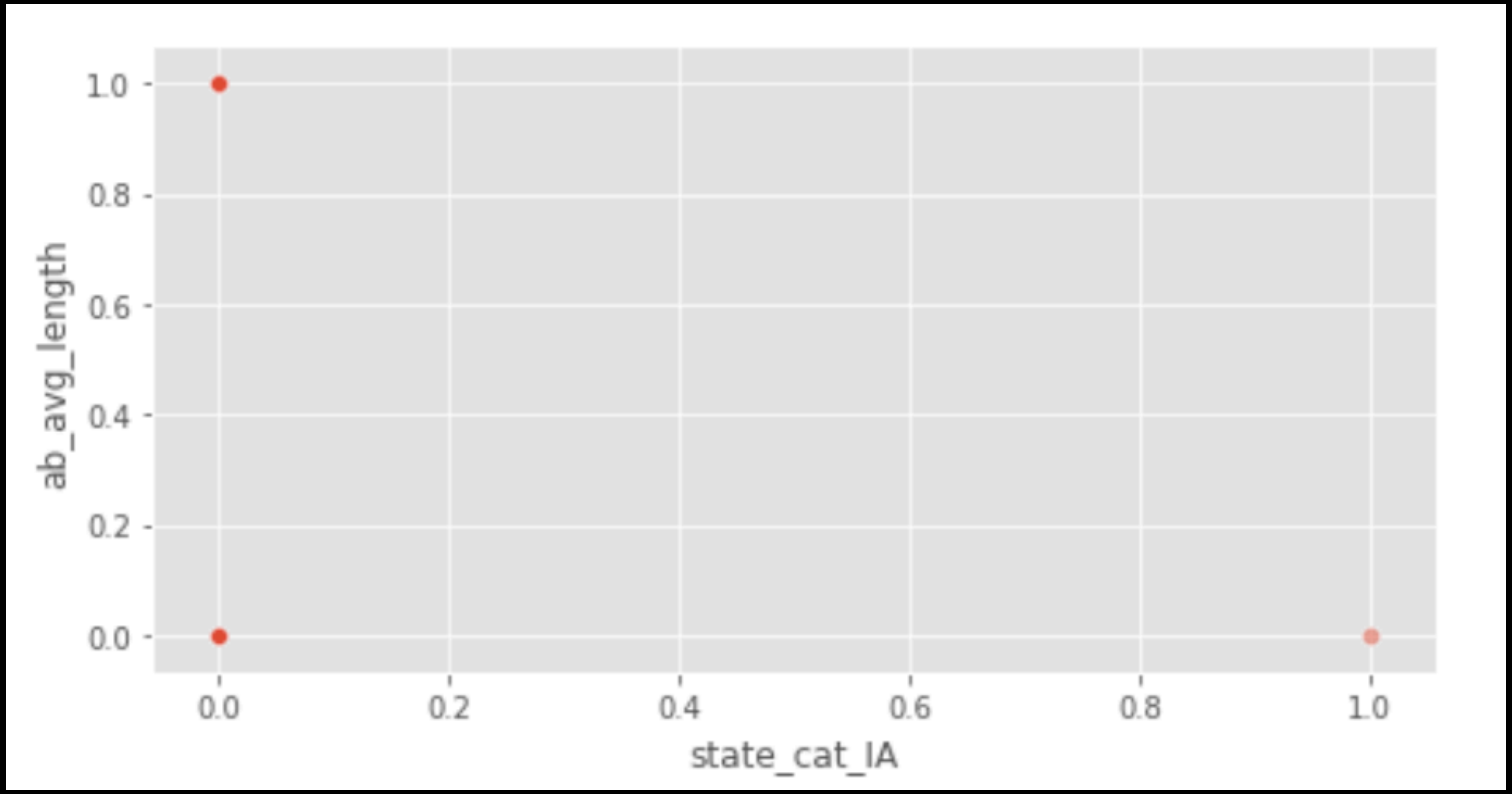




Secondary Analysis
Account Length

- EDA:
  - looked only at accounts that left
  - Created binary variable ‘Above_AVG_Length’
- Almost no one in RI, NM, LA stayed less than average
- No one in IA stayed longer than average

```
df3.corr()['ab_avg_length'].sort_values(ascending=False)

ab_avg_length          1.000000
state_cat_LA           0.093488
state_cat_NY           0.087647
state_cat_CO           0.079711
state_cat_RI           0.077344
                          ...
areaCode_510          -0.066922
state_cat_IA          -0.077277
state_cat_KY          -0.094402
total night minutes   -0.100395
state_cat_ME          -0.111380
Name: ab_avg_length, Length: 64, dtype: float64
```



Conclusion

- Important features:
  - Number of customer service calls
  - International plan
  - Voicemail plan
- Invest in customer service
- Target domestic customers
  - potentially in RI, NM, LA
- Promote voicemail plan

Next steps

- Get more data on accounts/ customers that left (or churned).
- Find other variable about customers to analyze
- Get data on broader industry/competitors

Thank you

Questions?

Link: https://github.com/s-shader/Telecom-Customer-Retention-Analysis

Appendix

Gradient boost Tuned

Train/test results: Gradient Boost Tuned Model

	GradientBoost_tuned	GradientBoost	BaseLine(KNN)
F1_test	0.851064	0.761421	0.344828
F1_train	1.000000	0.959328	0.935344
acc_test	0.958021	0.929535	0.743628
acc_train	1.000000	0.959614	0.931299
precision_test	0.860215	0.735294	0.271084
precision_train	1.000000	0.966162	0.883340
recall_test	0.842105	0.789474	0.473684
recall_train	1.000000	0.952590	0.993854

Top 10 Features

	importance
customer service calls	0.274019
total day charge	0.203235
international plan	0.129018
total eve minutes	0.066390
voice mail plan	0.049568
total intl charge	0.047704
total intl calls	0.039292
total night minutes	0.029662
total night calls	0.019661
total day calls	0.016852
total eve calls	0.016702
account length	0.015789

Appendix
XG Boost

Train/test results: XG Boost Tuned Model

	XGBoost_tuned	XGBoost	BaseLine(KNN)
F1_test	0.829787	0.740000	0.344828
F1_train	1.000000	0.957216	0.935344
acc_test	0.952024	0.922039	0.743628
acc_train	1.000000	0.957638	0.931299
precision_test	0.838710	0.704762	0.271084
precision_train	1.000000	0.966861	0.883340
recall_test	0.821053	0.778947	0.473684
recall_train	1.000000	0.947761	0.993854

Top 10 Features

	importance
international plan	0.079393
voice mail plan	0.051786
state_cat_TN	0.049446
customer service calls	0.046269
total intl calls	0.037305
state_cat_SD	0.027997
state_cat_NC	0.024579
total day charge	0.022493
state_cat_GA	0.021376
state_cat_OK	0.021074