Churn model

R Markdown

I will start by loading up my packages

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:randomForest':
##
##     margin
```

```
library(ResourceSelection)
```

```
## ResourceSelection 0.3-5   2019-07-22
```

```
user_data <- read.csv("~/Downloads/churn-data-full (1).csv")
```

Including Plots

You can also embed plots, for example:

```
##                            PASSPORT_ID         Region   device NET_REVENUE_USD
## 1 00ef1ab4-9a7b-4b0e-919c-5e4d1e512843             US   mobile           86.77
## 2 7c0240d8-8e93-40b4-946a-474a0f003d93             US   mobile           48.24
## 3 35eb405a-2b29-4e3e-a227-2bf3ffdb0897             US  desktop           89.80
## 4 ec27aa56-9322-42e1-adff-2b9526b1e36f United Kingdom  desktop           50.34
## 5 be1c4e8a-93be-44e6-ae8a-21e7d0124c0e             US   mobile          193.58
## 6 41ee14fe-6d64-4adb-9978-013c9e57d029             US  desktop           95.79
##              TITLE     JOB_CATEGORY CURRENCY total_visits time_on_site isChurn
## 1 General Manager          Manager      jpy            1           10       0
## 2 General Manager          Manager      usd            1           73       0
## 3 General Manager          Manager      jpy            4          372       0
## 4 General Manager          Manager      gbp            4         3147       0
## 5 Senior Engineer  Data / Software      usd          186        68913       0
## 6      Consultant         Strategy      usd            3          374       1
##    FIRST_RACE_TO_ZERO_EMISSIONS FIRST_MEMBERSHIP FIRST_PROMO FIRST_COVID
## 1                             0                0           0           0
## 2                             0                3           0           0
## 3                             0                0           0           0
## 4                             0                2           0           1
## 5                             0                1           0           0
## 6                             0                2           1           1
##    FIRST_SCIENCE_OF_TALENT FIRST_AFRICA FIRST_QZ_AT_WORK FIRST_QUARTZY
## 1                        0            0                0             0
## 2                        0            0                0             0
## 3                        0            0                0             0
## 4                        0            1                0             0
## 5                        0            0                0             0
## 6                        0            5                0             0
##    FIRST_SPACE_BIZ FIRST_FUTURE_FINANCE FIRST_DB FIRST_OBESSIONS
## 1                0                    0        0               0
## 2                0                    0        2              13
## 3                0                    0        0               0
## 4                1                    0        4               4
## 5                0                    0       12               0
## 6                4                    0       23               5
```

I do not need to do a full EDA on this data set since I created the curated list myself. I also did not perform a full ML project so I did not need to split into test/training.

```
#library(caret)
#set.seed(3456)
#trainIndex <- createDataPartition(user_data$isChurn, p = .8,
#                                  list = FALSE,
#                                  times = 1)
#head(trainIndex)
#churntrain <-user_data[ trainIndex]
#churntest <-user_data [-trainIndex]
```

```
model <- randomForest(as.factor(isChurn) ~ FIRST_DB + FIRST_QZ_AT_WORK + Region + device
+ time_on_site + JOB_CATEGORY + total_visits + FIRST_OBESSIONS , data=user_data, importa
nce=TRUE, omit.na=TRUE)
```

```
model$importance
```

```
##                              0            1 MeanDecreaseAccuracy
## FIRST_DB          0.0017965851  0.020622862          0.003818508
## FIRST_QZ_AT_WORK  0.0019344663  0.011992619          0.003013651
## Region            0.0010026555  0.009159631          0.001878575
## device           -0.0017241286  0.003380195         -0.001176185
## time_on_site      0.0130610291 -0.014141177          0.010143491
## JOB_CATEGORY     -0.0009917384  0.024695557          0.001767244
## total_visits      0.0142016463 -0.017270556          0.010820744
## FIRST_OBESSIONS   0.0020300617  0.010380968          0.002927477
##                  MeanDecreaseGini
## FIRST_DB                190.33072
## FIRST_QZ_AT_WORK         52.85356
## Region                  34.69698
## device                  47.87904
## time_on_site           371.21705
## JOB_CATEGORY           193.23030
## total_visits           185.59116
## FIRST_OBESSIONS         73.53839
```

```
modelglm <- glm(isChurn ~ FIRST_DB + FIRST_QZ_AT_WORK + Region + device + time_on_site +
JOB_CATEGORY + total_visits + FIRST_OBESSIONS , data=user_data)
```

```
summary(modelglm)
```

```
##
## Call:
## glm(formula = isChurn ~ FIRST_DB + FIRST_QZ_AT_WORK + Region +
##     device + time_on_site + JOB_CATEGORY + total_visits + FIRST_OBESSIONS,
##     data = user_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.34345  -0.12053  -0.10427  -0.08292   1.03744
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -7.318e-02  4.913e-02  -1.489 0.136382
## FIRST_DB                 -1.222e-03  3.385e-04  -3.611 0.000306 ***
## FIRST_QZ_AT_WORK         -4.470e-03  2.224e-03  -2.010 0.044415 *
## RegionUnited Kingdom      7.500e-02  1.828e-02   4.103 4.09e-05 ***
## RegionUS                  9.121e-02  1.545e-02   5.903 3.64e-09 ***
## devicemobile              1.354e-02  5.208e-03   2.601 0.009316 **
## devicetablet              2.507e-02  1.567e-02   1.600 0.109611
## time_on_site             -3.443e-07  3.131e-07  -1.100 0.271563
## JOB_CATEGORYArtist        1.281e-01  5.399e-02   2.372 0.017715 *
## JOB_CATEGORYC-Level       7.058e-02  4.736e-02   1.490 0.136229
## JOB_CATEGORYCEO           9.614e-02  4.724e-02   2.035 0.041857 *
## JOB_CATEGORYData / Software 6.365e-02 4.848e-02   1.313 0.189259
## JOB_CATEGORYDesign/Product 1.351e-01 4.980e-02   2.712 0.006688 **
## JOB_CATEGORYDirector      9.730e-02  4.713e-02   2.065 0.038977 *
## JOB_CATEGORYEducation     8.772e-02  5.041e-02   1.740 0.081863 .
## JOB_CATEGORYFinance       8.859e-02  5.129e-02   1.727 0.084134 .
## JOB_CATEGORYGovernment    7.713e-02  5.809e-02   1.328 0.184271
## JOB_CATEGORYHR            1.715e-01  5.846e-02   2.934 0.003350 **
## JOB_CATEGORYLegal         7.975e-02  5.541e-02   1.439 0.150128
## JOB_CATEGORYManager       6.640e-02  4.744e-02   1.400 0.161586
## JOB_CATEGORYMarketing     8.462e-02  4.831e-02   1.752 0.079848 .
## JOB_CATEGORYMedical       7.918e-02  5.318e-02   1.489 0.136565
## JOB_CATEGORYNews          8.839e-02  4.989e-02   1.772 0.076480 .
## JOB_CATEGORYOther         9.030e-02  4.693e-02   1.924 0.054354 .
## JOB_CATEGORYOwner/Founder 8.973e-02  4.718e-02   1.902 0.057215 .
## JOB_CATEGORYPscyology     4.154e-02  8.810e-02   0.472 0.637263
## JOB_CATEGORYResearch      3.807e-02  5.135e-02   0.741 0.458412
## JOB_CATEGORYRetired       9.401e-02  5.403e-02   1.740 0.081888 .
## JOB_CATEGORYSales         8.987e-02  5.049e-02   1.780 0.075089 .
## JOB_CATEGORYService       1.286e-01  7.632e-02   1.685 0.091938 .
## JOB_CATEGORYStrategy      1.098e-01  4.745e-02   2.314 0.020685 *
## JOB_CATEGORYStudent       2.047e-01  4.991e-02   4.100 4.15e-05 ***
## JOB_CATEGORYVP-Level      1.197e-01  4.849e-02   2.468 0.013595 *
## total_visits              3.184e-04  1.114e-04   2.859 0.004256 **
## FIRST_OBESSIONS          -1.149e-03  6.530e-04  -1.760 0.078502 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.09515234)
##
##     Null deviance: 1407.3  on 14656  degrees of freedom
```

```
## Residual deviance: 1391.3  on 14622  degrees of freedom
## AIC: 7154.4
##
## Number of Fisher Scoring iterations: 2
```

```
hoslem.test(modelglm$y, modelglm$fitted)
```

```
##
##   Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  modelglm$y, modelglm$fitted
## X-squared = 15.562, df = 8, p-value = 0.0491
```

```
modelglm_job <- glm(isChurn ~ FIRST_DB * JOB_CATEGORY + FIRST_QZ_AT_WORK * JOB_CATEGORY
+ FIRST_OBESSIONS , data=user_data)
```

```
summary(modelglm_job)
```

```
##
## Call:
## glm(formula = isChurn ~ FIRST_DB * JOB_CATEGORY + FIRST_QZ_AT_WORK *
##       JOB_CATEGORY + FIRST_OBESSIONS, data = user_data)
##
## Deviance Residuals:
##      Min       1Q    Median       3Q       Max
## -0.47693  -0.11949  -0.10329  -0.08106   1.00877
##
## Coefficients: (1 not defined because of singularities)
##                                          Estimate Std. Error t value
## (Intercept)                              0.0517930  0.0695967   0.744
## FIRST_DB                                -0.0054010  0.0103147  -0.524
## JOB_CATEGORYArtist                       0.1716780  0.0779723   2.202
## JOB_CATEGORYC-Level                      0.0405149  0.0705916   0.574
## JOB_CATEGORYCEO                          0.0503685  0.0703247   0.716
## JOB_CATEGORYData / Software              0.0641996  0.0720778   0.891
## JOB_CATEGORYDesign/Product               0.1220463  0.0731034   1.670
## JOB_CATEGORYDirector                     0.0788773  0.0702080   1.123
## JOB_CATEGORYEducation                    0.0131749  0.0737064   0.179
## JOB_CATEGORYFinance                     -0.0192904  0.0750483  -0.257
## JOB_CATEGORYGovernment                  -0.0174192  0.0856698  -0.203
## JOB_CATEGORYHR                           0.2059038  0.0812841   2.533
## JOB_CATEGORYLegal                        0.0733392  0.0824226   0.890
## JOB_CATEGORYManager                      0.0556274  0.0706748   0.787
## JOB_CATEGORYMarketing                    0.0733262  0.0714671   1.026
## JOB_CATEGORYMedical                      0.0910704  0.0784844   1.160
## JOB_CATEGORYNews                         0.0555964  0.0732338   0.759
## JOB_CATEGORYOther                        0.0699353  0.0699739   0.999
## JOB_CATEGORYOwner/Founder                0.0675602  0.0702436   0.962
## JOB_CATEGORYPscyology                    0.2372695  0.2300346   1.031
## JOB_CATEGORYResearch                     0.0202483  0.0749155   0.270
## JOB_CATEGORYRetired                      0.0010100  0.0795429   0.013
## JOB_CATEGORYSales                        0.0697439  0.0742741   0.939
## JOB_CATEGORYService                     -0.0463617  0.1235886  -0.375
## JOB_CATEGORYStrategy                     0.0895489  0.0706614   1.267
## JOB_CATEGORYStudent                      0.2377062  0.0735951   3.230
## JOB_CATEGORYVP-Level                     0.0340280  0.0715634   0.475
## FIRST_QZ_AT_WORK                        -0.0258965  0.0844245  -0.307
## FIRST_OBESSIONS                         -0.0009460  0.0006566  -1.441
## FIRST_DB:JOB_CATEGORYArtist             -0.0170356  0.0121956  -1.397
## FIRST_DB:JOB_CATEGORYC-Level             0.0045577  0.0104240   0.437
## FIRST_DB:JOB_CATEGORYCEO                 0.0064573  0.0103728   0.623
## FIRST_DB:JOB_CATEGORYData / Software    -0.0001501  0.0105549  -0.014
## FIRST_DB:JOB_CATEGORYDesign/Product      0.0009558  0.0108954   0.088
## FIRST_DB:JOB_CATEGORYDirector            0.0033292  0.0103497   0.322
## FIRST_DB:JOB_CATEGORYEducation           0.0110572  0.0105407   1.049
## FIRST_DB:JOB_CATEGORYFinance             0.0182609  0.0108980   1.676
## FIRST_DB:JOB_CATEGORYGovernment          0.0139777  0.0110461   1.265
## FIRST_DB:JOB_CATEGORYHR                 -0.0014701  0.0105742  -0.139
## FIRST_DB:JOB_CATEGORYLegal               0.0033146  0.0110272   0.301
## FIRST_DB:JOB_CATEGORYManager             0.0024726  0.0104252   0.237
## FIRST_DB:JOB_CATEGORYMarketing           0.0011781  0.0105495   0.112
```

```
## FIRST_DB:JOB_CATEGORYMedical                   -0.0032769  0.0113468  -0.289
## FIRST_DB:JOB_CATEGORYNews                        0.0044180  0.0106171   0.416
## FIRST_DB:JOB_CATEGORYOther                       0.0032361  0.0103403   0.313
## FIRST_DB:JOB_CATEGORYOwner/Founder               0.0027244  0.0103679   0.263
## FIRST_DB:JOB_CATEGORYPscyology                  -0.0180365  0.0290855  -0.620
## FIRST_DB:JOB_CATEGORYResearch                    0.0024423  0.0108101   0.226
## FIRST_DB:JOB_CATEGORYRetired                     0.0108204  0.0106291   1.018
## FIRST_DB:JOB_CATEGORYSales                       0.0033572  0.0106651   0.315
## FIRST_DB:JOB_CATEGORYService                     0.0336088  0.0187332   1.794
## FIRST_DB:JOB_CATEGORYStrategy                    0.0024502  0.0104189   0.235
## FIRST_DB:JOB_CATEGORYStudent                    -0.0138219  0.0111545  -1.239
## FIRST_DB:JOB_CATEGORYVP-Level                    0.0123876  0.0103882   1.192
## JOB_CATEGORYArtist:FIRST_QZ_AT_WORK              0.1350355  0.1639294   0.824
## JOB_CATEGORYC-Level:FIRST_QZ_AT_WORK             0.0296570  0.0850738   0.349
## JOB_CATEGORYCEO:FIRST_QZ_AT_WORK                 0.0293806  0.0844979   0.348
## JOB_CATEGORYData / Software:FIRST_QZ_AT_WORK     0.0225778  0.0849942   0.266
## JOB_CATEGORYDesign/Product:FIRST_QZ_AT_WORK     -0.0126691  0.0906738  -0.140
## JOB_CATEGORYDirector:FIRST_QZ_AT_WORK            0.0076306  0.0849703   0.090
## JOB_CATEGORYEducation:FIRST_QZ_AT_WORK           0.0363256  0.1030542   0.352
## JOB_CATEGORYFinance:FIRST_QZ_AT_WORK             0.0639979  0.0898462   0.712
## JOB_CATEGORYGovernment:FIRST_QZ_AT_WORK         -0.0017625  0.0868000  -0.020
## JOB_CATEGORYHR:FIRST_QZ_AT_WORK                 -0.0010787  0.0858969  -0.013
## JOB_CATEGORYLegal:FIRST_QZ_AT_WORK               0.0105754  0.0857217   0.123
## JOB_CATEGORYManager:FIRST_QZ_AT_WORK             0.0012321  0.0858079   0.014
## JOB_CATEGORYMarketing:FIRST_QZ_AT_WORK           0.0134992  0.0852529   0.158
## JOB_CATEGORYMedical:FIRST_QZ_AT_WORK            -0.0562216  0.1538124  -0.366
## JOB_CATEGORYNews:FIRST_QZ_AT_WORK                0.0928592  0.1005126   0.924
## JOB_CATEGORYOther:FIRST_QZ_AT_WORK               0.0097357  0.0847054   0.115
## JOB_CATEGORYOwner/Founder:FIRST_QZ_AT_WORK       0.0245532  0.0847574   0.290
## JOB_CATEGORYPscyology:FIRST_QZ_AT_WORK          -0.1694160  0.1900925  -0.891
## JOB_CATEGORYResearch:FIRST_QZ_AT_WORK            0.0251625  0.0849880   0.296
## JOB_CATEGORYRetired:FIRST_QZ_AT_WORK            -0.0019126  0.0947700  -0.020
## JOB_CATEGORYSales:FIRST_QZ_AT_WORK               0.0178282  0.0873067   0.204
## JOB_CATEGORYService:FIRST_QZ_AT_WORK                    NA         NA      NA
## JOB_CATEGORYStrategy:FIRST_QZ_AT_WORK            0.0238090  0.0852069   0.279
## JOB_CATEGORYStudent:FIRST_QZ_AT_WORK             0.1214173  0.0894194   1.358
## JOB_CATEGORYVP-Level:FIRST_QZ_AT_WORK           -0.0069320  0.0897516  -0.077
##                                                 Pr(>|t|)
## (Intercept)                                      0.45678
## FIRST_DB                                         0.60055
## JOB_CATEGORYArtist                               0.02770 *
## JOB_CATEGORYC-Level                              0.56602
## JOB_CATEGORYCEO                                  0.47386
## JOB_CATEGORYData / Software                      0.37311
## JOB_CATEGORYDesign/Product                       0.09504 .
## JOB_CATEGORYDirector                             0.26125
## JOB_CATEGORYEducation                            0.85814
## JOB_CATEGORYFinance                              0.79715
## JOB_CATEGORYGovernment                           0.83888
## JOB_CATEGORYHR                                   0.01132 *
## JOB_CATEGORYLegal                                0.37359
## JOB_CATEGORYManager                              0.43124
## JOB_CATEGORYMarketing                            0.30490
## JOB_CATEGORYMedical                              0.24592
```

```
## JOB_CATEGORYNews                              0.44777
## JOB_CATEGORYOther                             0.31759
## JOB_CATEGORYOwner/Founder                     0.33617
## JOB_CATEGORYPscyology                         0.30235
## JOB_CATEGORYResearch                          0.78695
## JOB_CATEGORYRetired                           0.98987
## JOB_CATEGORYSales                             0.34774
## JOB_CATEGORYService                           0.70757
## JOB_CATEGORYStrategy                          0.20507
## JOB_CATEGORYStudent                           0.00124 **
## JOB_CATEGORYVP-Level                          0.63444
## FIRST_QZ_AT_WORK                              0.75904
## FIRST_OBESSIONS                               0.14968
## FIRST_DB:JOB_CATEGORYArtist                   0.16247
## FIRST_DB:JOB_CATEGORYC-Level                  0.66195
## FIRST_DB:JOB_CATEGORYCEO                      0.53360
## FIRST_DB:JOB_CATEGORYData / Software          0.98865
## FIRST_DB:JOB_CATEGORYDesign/Product           0.93010
## FIRST_DB:JOB_CATEGORYDirector                 0.74771
## FIRST_DB:JOB_CATEGORYEducation                0.29419
## FIRST_DB:JOB_CATEGORYFinance                  0.09383 .
## FIRST_DB:JOB_CATEGORYGovernment               0.20575
## FIRST_DB:JOB_CATEGORYHR                       0.88943
## FIRST_DB:JOB_CATEGORYLegal                    0.76373
## FIRST_DB:JOB_CATEGORYManager                  0.81252
## FIRST_DB:JOB_CATEGORYMarketing                0.91108
## FIRST_DB:JOB_CATEGORYMedical                  0.77274
## FIRST_DB:JOB_CATEGORYNews                     0.67733
## FIRST_DB:JOB_CATEGORYOther                    0.75431
## FIRST_DB:JOB_CATEGORYOwner/Founder            0.79273
## FIRST_DB:JOB_CATEGORYPscyology                0.53519
## FIRST_DB:JOB_CATEGORYResearch                 0.82126
## FIRST_DB:JOB_CATEGORYRetired                  0.30870
## FIRST_DB:JOB_CATEGORYSales                    0.75293
## FIRST_DB:JOB_CATEGORYService                  0.07282 .
## FIRST_DB:JOB_CATEGORYStrategy                 0.81408
## FIRST_DB:JOB_CATEGORYStudent                  0.21532
## FIRST_DB:JOB_CATEGORYVP-Level                 0.23310
## JOB_CATEGORYArtist:FIRST_QZ_AT_WORK           0.41010
## JOB_CATEGORYC-Level:FIRST_QZ_AT_WORK          0.72739
## JOB_CATEGORYCEO:FIRST_QZ_AT_WORK             0.72806
## JOB_CATEGORYData / Software:FIRST_QZ_AT_WORK  0.79052
## JOB_CATEGORYDesign/Product:FIRST_QZ_AT_WORK   0.88888
## JOB_CATEGORYDirector:FIRST_QZ_AT_WORK         0.92844
## JOB_CATEGORYEducation:FIRST_QZ_AT_WORK        0.72448
## JOB_CATEGORYFinance:FIRST_QZ_AT_WORK          0.47629
## JOB_CATEGORYGovernment:FIRST_QZ_AT_WORK       0.98380
## JOB_CATEGORYHR:FIRST_QZ_AT_WORK               0.98998
## JOB_CATEGORYLegal:FIRST_QZ_AT_WORK            0.90182
## JOB_CATEGORYManager:FIRST_QZ_AT_WORK          0.98854
## JOB_CATEGORYMarketing:FIRST_QZ_AT_WORK        0.87419
## JOB_CATEGORYMedical:FIRST_QZ_AT_WORK          0.71473
## JOB_CATEGORYNews:FIRST_QZ_AT_WORK             0.35558
## JOB_CATEGORYOther:FIRST_QZ_AT_WORK            0.90850
```

```
## JOB_CATEGORYOwner/Founder:FIRST_QZ_AT_WORK      0.77206
## JOB_CATEGORYPscyology:FIRST_QZ_AT_WORK          0.37282
## JOB_CATEGORYResearch:FIRST_QZ_AT_WORK           0.76718
## JOB_CATEGORYRetired:FIRST_QZ_AT_WORK            0.98390
## JOB_CATEGORYSales:FIRST_QZ_AT_WORK              0.83820
## JOB_CATEGORYService:FIRST_QZ_AT_WORK                 NA
## JOB_CATEGORYStrategy:FIRST_QZ_AT_WORK           0.77992
## JOB_CATEGORYStudent:FIRST_QZ_AT_WORK            0.17454
## JOB_CATEGORYVP-Level:FIRST_QZ_AT_WORK           0.93844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.09466525)
##
##     Null deviance: 1407.3  on 14656  degrees of freedom
## Residual deviance: 1380.1  on 14579  degrees of freedom
## AIC: 7122
##
## Number of Fisher Scoring iterations: 2
```

```
hoslem.test(modelglm_job$y, modelglm_job$fitted)
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  modelglm_job$y, modelglm_job$fitted
## X-squared = 25.604, df = 8, p-value = 0.001227
```