Team Bravo

How to improve customer retention of bank

Junchi Tian
Bixuan Huang

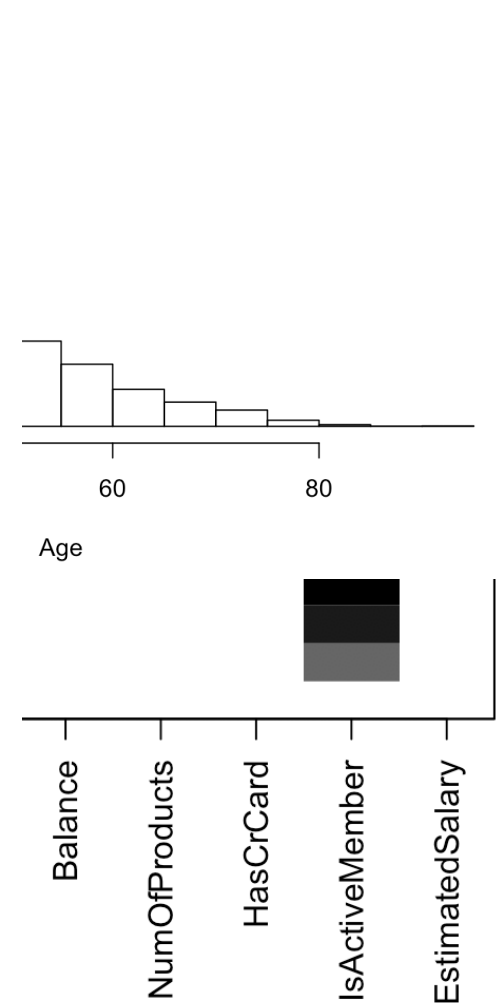# EXPLORATORY DATA ANALYSIS

```
'data.frame':    10000 obs. of  11 variables:
$ CreditScore    : int  619 608 502 699 850 645 822 376 50
$ Geography      : Factor w/ 3 levels "France","Germany"
```

Histogram of Age

Histogram of IsActiveMember
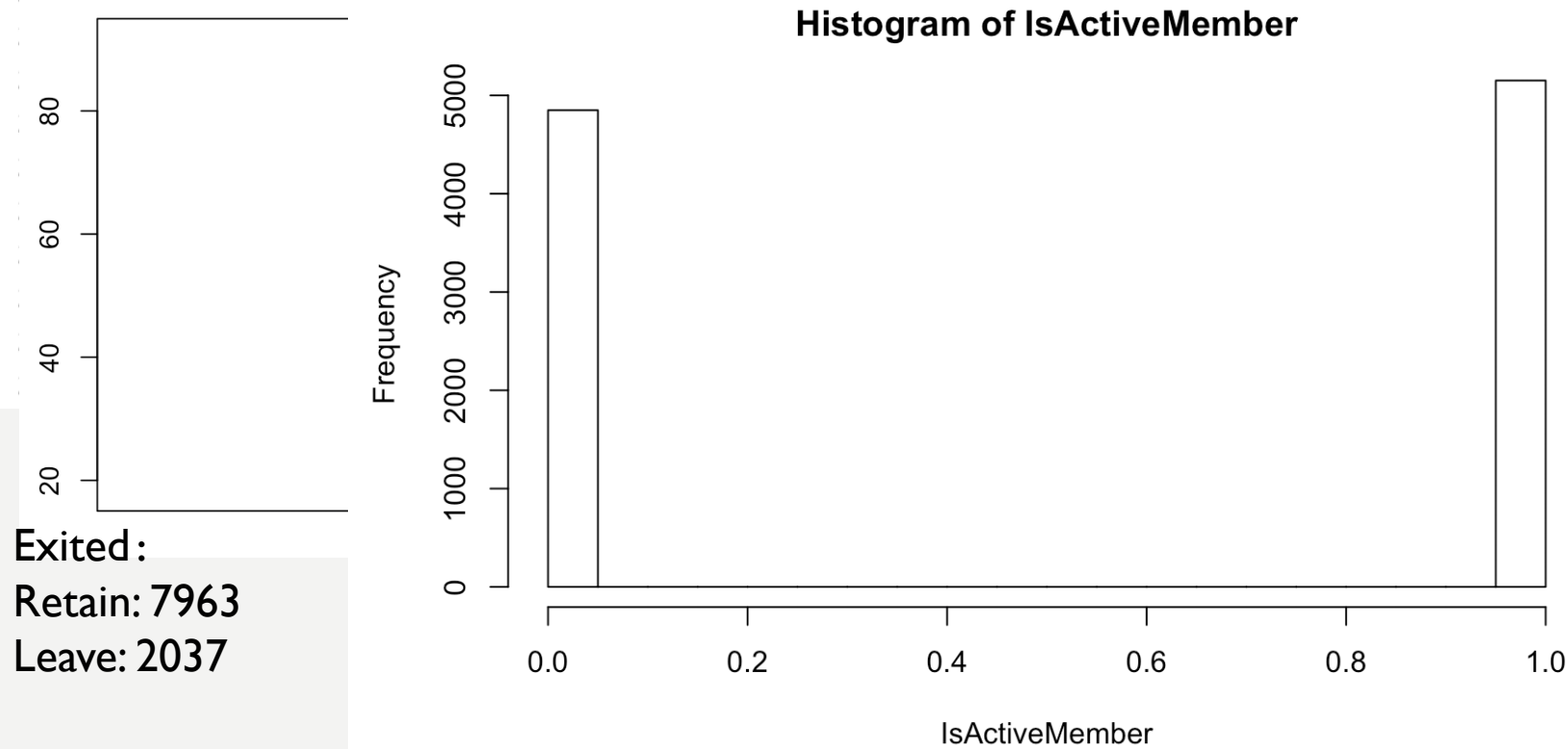
Exited:
Retain: 7963
Leave: 2037

Age

Balance NumOfProducts HasCrCard IsActiveMember EstimatedSalary

(Inte Credit graphyGer Geography Gende T

MODEL DIAGNOSTICS

Dependent: Exited
Independent: CreditScore, Geography, Gender, Age, Balance, IsActiveMember

Multicollinearity:

```
Cal                         GVIF Df GVIF^(1/(2*Df))
glr CreditScore        1.001354  1           1.000677   Age +
    Geography          1.196809  2           1.045939   git"),
    Gender             1.003773  1           1.001885
Dev Age                1.091707  1           1.044848
    Balance            1.195835  1           1.093543
-2. IsActiveMember     1.086432  1           1.042321
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)       -3.717e+00  2.417e-01 -15.379  < 2e-16 ***
CreditScore       -5.973e-04  3.134e-04  -1.906   0.0567 .
GeographyGermany   7.333e-01  7.548e-02   9.716  < 2e-16 ***
GeographySpain     8.375e-03  7.861e-02   0.107   0.9152
GenderMale        -5.780e-01  6.100e-02  -9.476  < 2e-16 ***
Age                7.545e-02  2.899e-03  26.022  < 2e-16 ***
Balance            2.999e-06  5.521e-07   5.431 5.61e-08 ***
IsActiveMember    -1.147e+00  6.502e-02 -17.637  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8135.0  on 7999  degrees of freedom
Residual deviance: 6832.4  on 7992  degrees of freedom
AIC: 6848.4

Number of Fisher Scoring iterations: 5
```
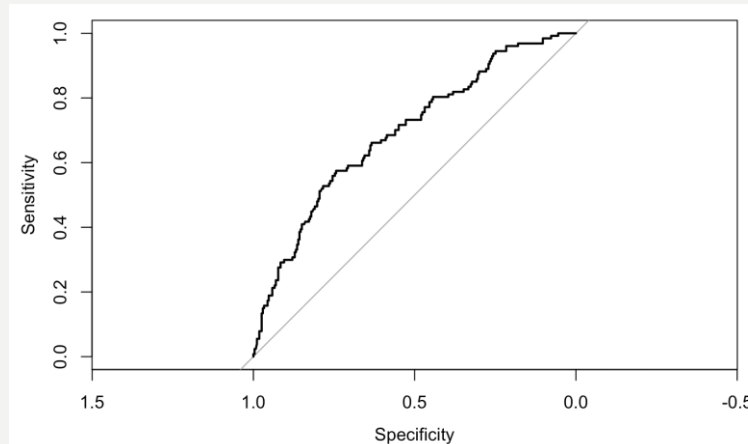
```
  (Intercept)    CreditScore GeographyGermany   GeographySpain
   0.02430666     0.99940291       2.08199990       1.00841011
IsActiveMember
   0.31765101
    GenderMale            Age          Balance
   0.56099703     1.07836574       1.00000300
```

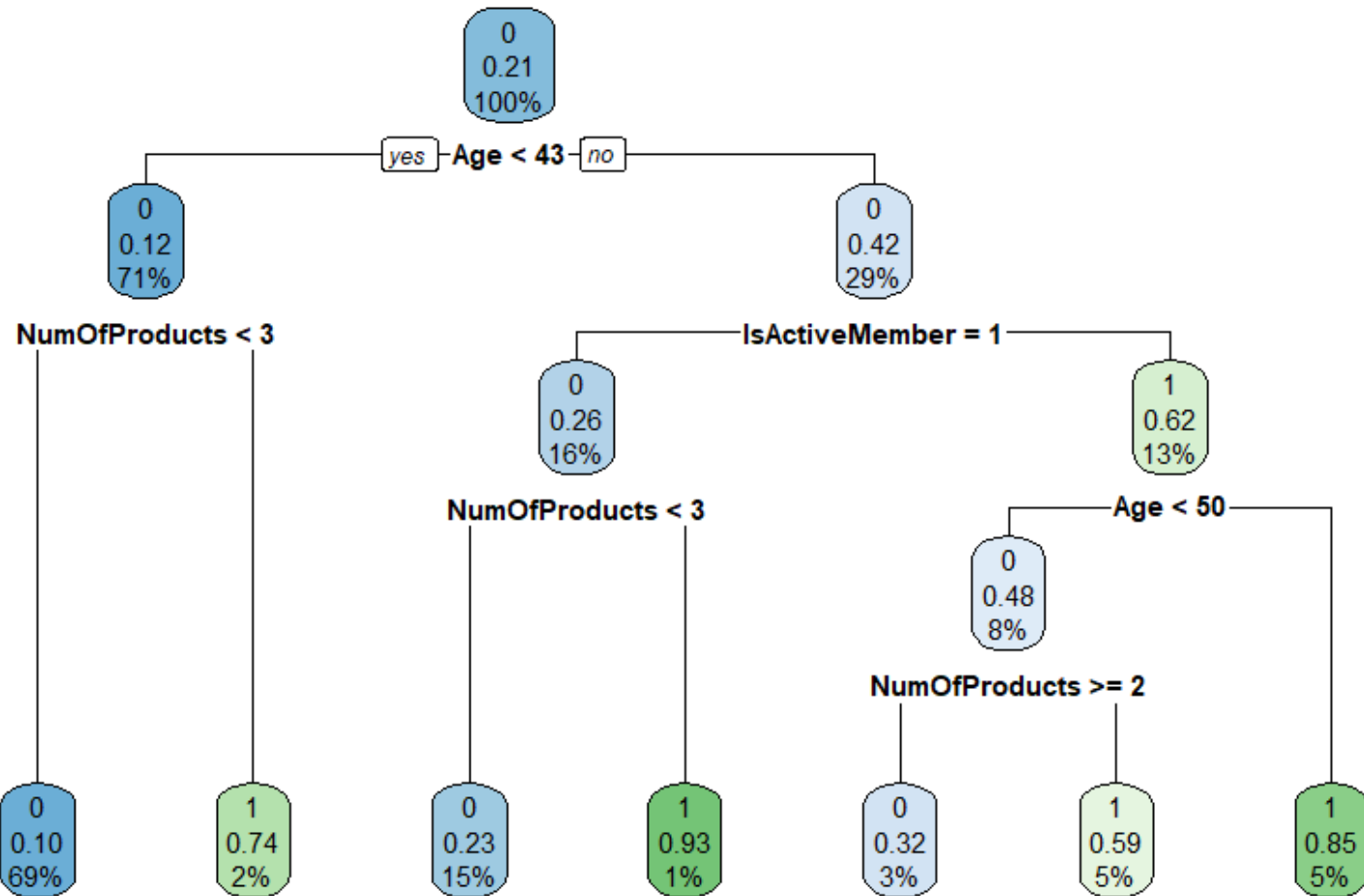Hosmer and Lemeshow goodness of fit (GOF) test

```
data:  train$Exited, fitted(bank_log)
X-squared = 13.407, df = 8, p-value = 0.0986
```

Hit rate : 0.809

AUC : 0.7418554

DECISION TREE


```
bank_model    0     1
         0  6192   957
         1   161   690
```

"Real error rate is: 13.975%"

Hit rate : 0.805

CONCLUSION/SOLUTION

- 1, The bank should held promotional activities to the old, particularly age is older than 43.

- 2, Create more useful products and enable customers to have more accounts, better than 3.

- 3, Focus on German and Female, compared to France, Spain and Male.

- 4, Establish a good membership system to make more active members.

Team Bravo

Thank you

Junchi Tian
Bixuan Huang