Analyze dataset

Upload dataset

Browse... Bank_Churn.csv

Upload complete

CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0
15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0
15574012	Chu	645	Spain	Male	44	8	113755.78	2	1	0	149756.71	1

Quantitative variables: CustomerId, CreditScore, Age, Tenure, Balance, NumOfProducts, EstimatedSalary
Qualitative variables: Geography, Gender, HasCrCard, IsActiveMember, Exited

Select outlier removal method

SD

Detect outliers

```
$CustomerId
integer(0)

$CreditScore
[1] 359 350 350 358 351 350 350 350

$Age
  [1] 75 73 72 79 80 75 72 82 74 71 72 74 76 71 73 77 74 74 74 74 74 72 77 74 88 71 72 71 75 73 76 85 74 76 72 71 74 72 72 84 71 74 84
 [44] 77 79 76 73 73 76 72 71 80 74 76 75 77 74 71 75 78 74 71 77 79 81 79 71 72 71 72 72 78 75 71 73 71 71 76 73 75 73 71 72 73 92 75
 [87] 71 77 92 72 71 76 72 77 74 72 73 77 71 72 81 76 74 71 76 72 81 73 71 75 71 71 71 73 72 71 81 73 74 83 71 78 72 74 80 72 76 71 71
[130] 78 78 77 77

$Tenure
integer(0)

$Balance
numeric(0)

$NumOfProducts
 [1] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4

$EstimatedSalary
numeric(0)
```

Visualize

Select Variable:

Exited

Run Model

```
[1] "Exited"
$summary

Call:
glm(formula = formula, family = binomial(link = "logit"), data = train_data)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)      -3.708e+00  6.600e+00  -0.562   0.5742
CustomerId        1.379e-08  4.205e-07   0.033   0.9738
CreditScore      -5.346e-04  3.130e-04  -1.708   0.0877 .
GeographyGermany  7.750e-01  7.543e-02  10.274  < 2e-16 ***
GeographySpain    5.420e-02  7.861e-02   0.690   0.4905
GenderMale       -5.022e-01  6.068e-02  -8.276  < 2e-16 ***
Age               7.161e-02  2.872e-03  24.936  < 2e-16 ***
Tenure           -1.303e-02  1.040e-02  -1.253   0.2101
Balance           2.596e-06  5.738e-07   4.525 6.04e-06 ***
NumOfProducts    -7.693e-02  5.204e-02  -1.478   0.1394
HasCrCard1       -3.021e-02  6.620e-02  -0.456   0.6482
IsActiveMember1  -1.034e+00  6.391e-02 -16.172  < 2e-16 ***
EstimatedSalary   2.020e-07  5.285e-07   0.382   0.7023
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8088.9  on 7999  degrees of freedom
Residual deviance: 6898.6  on 7987  degrees of freedom
AIC: 6924.6

Number of Fisher Scoring iterations: 5


$conf_matrix
      Predicted
Actual    0    1
     0 1547   46
     1  318   89

$accuracy
[1] 0.818

$precision
[1] 0.6592593

$sensitivity
[1] 0.2186732

$specificity
[1] 0.9711237

$f1_score
[1] 0.3284133

$plot

Call:
roc.default(response = actual, predictor = test_probabilities)

Data: test_probabilities in 1593 controls (actual 0) < 407 cases (actual 1).
Area under the curve: 0.7868
```