#Output variable -> y

#y-> Whether the client has subscribed a term deposit or not

#Binomial ("yes" or "no")

#Problem Statement for Bank Data

Bank Output:-

> sum(is.na(bank)) # To get the count of NA Values

[1] 0

> str(bank)

'data.frame': 45211 obs. of 17 variables:

$ age : int 58 44 33 47 33 35 28 42 58 43 ...

$ job : chr "management" "technician" "entrepreneur" "blue-collar" ...

$ marital : chr "married" "single" "married" "married" ...

$ education: chr "tertiary" "secondary" "secondary" "unknown" ...

$ default : chr "no" "no" "no" "no" ...

$ balance : int 2143 29 2 1506 1 231 447 2 121 593 ...

$ housing : chr "yes" "yes" "yes" "yes" ...

$ loan : chr "no" "no" "yes" "no" ...

$ contact : chr "unknown" "unknown" "unknown" "unknown" ...

$ day : int 5 5 5 5 5 5 5 5 5 5 ...

$ month : chr "may" "may" "may" "may" ...

$ duration : int 261 151 76 92 198 139 217 380 50 55 ...

$ campaign : int 1 1 1 1 1 1 1 1 1 1 ...

$ pdays : int -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...

$ previous : int 0 0 0 0 0 0 0 0 0 0 ...

$ poutcome : chr "unknown" "unknown" "unknown" "unknown" ...

$ y : chr "no" "no" "no" "no" ...

Call:

glm(formula = factor(y) ~ ., family = binomial, data = bank)

Deviance Residuals:

Min 1Q Median 3Q Max

-5.7286 -0.3744 -0.2530 -0.1502 3.4288

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.536e+00 1.837e-01 -13.803 < 2e-16 \*\*\*

age 1.127e-04 2.205e-03 0.051 0.959233

jobblue-collar -3.099e-01 7.267e-02 -4.264 2.01e-05 \*\*\*

jobentrepreneur -3.571e-01 1.256e-01 -2.844 0.004455 \*\*

jobhousemaid -5.040e-01 1.365e-01 -3.693 0.000221 \*\*\*

jobmanagement -1.653e-01 7.329e-02 -2.255 0.024130 \*

jobretired 2.524e-01 9.722e-02 2.596 0.009436 \*\*

jobself-employed -2.983e-01 1.120e-01 -2.664 0.007726 \*\*

jobservices -2.238e-01 8.406e-02 -2.662 0.007763 \*\*

jobstudent 3.821e-01 1.090e-01 3.505 0.000457 \*\*\*

jobtechnician -1.760e-01 6.893e-02 -2.554 0.010664 \*

jobunemployed -1.767e-01 1.116e-01 -1.583 0.113456

jobunknown -3.133e-01 2.335e-01 -1.342 0.179656

maritalmarried -1.795e-01 5.891e-02 -3.046 0.002318 \*\*

maritalsingle 9.250e-02 6.726e-02 1.375 0.169066

educationsecondary 1.835e-01 6.479e-02 2.833 0.004618 \*\*

educationtertiary 3.789e-01 7.532e-02 5.031 4.88e-07 \*\*\*

educationunknown 2.505e-01 1.039e-01 2.411 0.015915 \*

defaultyes -1.668e-02 1.628e-01 -0.102 0.918407

balance 1.283e-05 5.148e-06 2.493 0.012651 \*

housingyes -6.754e-01 4.387e-02 -15.395 < 2e-16 \*\*\*

loanyes -4.254e-01 5.999e-02 -7.091 1.33e-12 \*\*\*

contacttelephone -1.634e-01 7.519e-02 -2.173 0.029784 \*

contactunknown -1.623e+00 7.317e-02 -22.184 < 2e-16 \*\*\*

day 9.969e-03 2.497e-03 3.993 6.53e-05 \*\*\*

monthaug -6.939e-01 7.847e-02 -8.842 < 2e-16 \*\*\*

monthdec 6.911e-01 1.767e-01 3.912 9.17e-05 \*\*\*

monthfeb -1.473e-01 8.941e-02 -1.648 0.099427 .

monthjan -1.262e+00 1.217e-01 -10.367 < 2e-16 \*\*\*

monthjul -8.308e-01 7.740e-02 -10.733 < 2e-16 \*\*\*

monthjun 4.536e-01 9.367e-02 4.843 1.28e-06 \*\*\*

monthmar 1.590e+00 1.199e-01 13.265 < 2e-16 \*\*\*

monthmay -3.991e-01 7.229e-02 -5.521 3.36e-08 \*\*\*

monthnov -8.734e-01 8.441e-02 -10.347 < 2e-16 \*\*\*

monthoct 8.814e-01 1.080e-01 8.159 3.37e-16 \*\*\*

monthsep 8.741e-01 1.195e-01 7.314 2.58e-13 \*\*\*

duration 4.194e-03 6.453e-05 64.986 < 2e-16 \*\*\*

campaign -9.078e-02 1.014e-02 -8.955 < 2e-16 \*\*\*

pdays -1.027e-04 3.061e-04 -0.335 0.737268

previous 1.015e-02 6.503e-03 1.561 0.118476

poutcomeother 2.035e-01 8.986e-02 2.265 0.023543 \*

poutcomesuccess 2.291e+00 8.235e-02 27.821 < 2e-16 \*\*\*

poutcomeunknown -9.179e-02 9.347e-02 -0.982 0.326093

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 32631 on 45210 degrees of freedom

Residual deviance: 21562 on 45168 degrees of freedom

AIC: 21648

Number of Fisher Scoring iterations: 6

> exp(coef(logit))

(Intercept) age jobblue-collar jobentrepreneur jobhousemaid jobmanagement

0.07921118 1.00011272 0.73354041 0.69969989 0.60410838 0.84765765

jobretired jobself-employed jobservices jobstudent jobtechnician jobunemployed

1.28706269 0.74205191 0.79947733 1.46541095 0.83860411 0.83802016

jobunknown maritalmarried maritalsingle educationsecondary educationtertiary educationunknown

0.73105661 0.83572682 1.09691056 1.20144892 1.46073758 1.28464040

defaultyes balance housingyes loanyes contacttelephone contactunknown

0.98345715 1.00001284 0.50896077 0.65352685 0.84927322 0.19726311

day monthaug monthdec monthfeb monthjan monthjul

1.01001878 0.49961996 1.99595838 0.86301696 0.28316690 0.43570251

monthjun monthmar monthmay monthnov monthoct monthsep

1.57400386 4.90321221 0.67091594 0.41753015 2.41436771 2.39661674

duration campaign pdays previous poutcomeother poutcomesuccess

1.00420250 0.91321697 0.99989732 1.01020406 1.22565868 9.88537129

poutcomeunknown

0.91229351

> table(bank$y)

no yes

39922 5289

> table(probo)

probo

FALSE TRUE

42396 2815

> confusion

no yes

FALSE 38940 3456

TRUE 982 1833

> Accuracy

[1] 0.901838

> Error

[1] 0.09816195

**ROC Curve**

