OPSM 324

2023-04-26

Loading Libraries

```
library(caret)
```

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

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

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(kknn)
```

```
##
## Attaching package: 'kknn'
```

```
## The following object is masked from 'package:caret':
##
##     contr.dummy
```

```
library(fields)
```

```
## Loading required package: spam
```

```
## Spam version 2.9-1 (2022-08-07) is loaded.
## Type 'help( Spam)' or 'demo( spam)' for a short introduction
## and overview of this package.
## Help for individual functions is also obtained by adding the
## suffix '.spam' to the function name, e.g. 'help( chol.spam)'.
```

```
##
## Attaching package: 'spam'
```

```
## The following objects are masked from 'package:base':
##
##     backsolve, forwardsolve
```

```
## Loading required package: viridis
```

```
## Loading required package: viridisLite
```

```
##
## Try help(fields) to get started.
```

```
library(caret)
library(rpart)
library(rpart.plot)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(ggplot2)
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:fields':
##
##     describe
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(lattice)
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

#Loading dataset

```
getwd()
```

```
## [1] "C:/Users/adity/OneDrive/Desktop"
```

```
setwd("C:/Users/adity/OneDrive/Desktop/SEM - 6")
#Reading Data
dataset <- read.csv("Deans Dilemma Datasheet.csv", stringsAsFactors =  TRUE)
View(dataset)
describe(dataset)
```

```
##                     vars   n      mean        sd    median   trimmed      mad
## SlNo                   1 391    196.00    113.02    196.00    196.00   145.29
## Gender*                2 391      1.68      0.47      2.00      1.72     0.00
## Gender.B               3 391      0.32      0.47      0.00      0.28     0.00
## Percent_SSC            4 391     64.65     10.96     64.50     64.76    12.60
## Board_SSC*             5 391      2.23      0.87      3.00      2.28     0.00
## Board_CBSE             6 391      0.29      0.45      0.00      0.24     0.00
## Board_ICSE             7 391      0.20      0.40      0.00      0.12     0.00
## Percent_HSC            8 391     63.80     11.42     63.00     63.34    13.34
## Board_HSC*             9 391      2.39      0.85      3.00      2.48     0.00
## Stream_HSC*           10 391      2.34      0.56      2.00      2.36     0.00
## Percent_Degree        11 391     62.98      8.92     63.00     62.91     8.90
## Course_Degree*        12 391      3.85      1.61      4.00      3.81     1.48
## Degree_Engg           13 391      0.09      0.29      0.00      0.00     0.00
## Experience_Yrs        14 391      0.48      0.67      0.00      0.36     0.00
## Entrance_Test*        15 391      5.63      2.43      7.00      6.00     0.00
## S.TEST                16 391      0.83      0.38      1.00      0.91     0.00
## Percentile_ET         17 391     54.93     31.17     62.00     56.87    25.20
## Percent_MBA           18 391     61.67      5.85     61.01     61.45     6.39
## S.TEST.SCORE          19 391     54.93     31.17     62.00     56.87    25.20
## Specialization_MBA*   20 391      1.47      0.56      1.00      1.42     0.00
## Marks_Communication   21 391     60.54      8.82     58.00     59.68     8.90
## Marks_Projectwork     22 391     68.36      7.15     69.00     68.60     7.41
## Marks_BOCA            23 391     64.38      9.58     63.00     64.08    11.86
## Placement*            24 391      1.80      0.40      2.00      1.87     0.00
## Placement_B           25 391      0.80      0.40      1.00      0.87     0.00
## Salary                26 391 219078.26 138311.65 240000.00 217011.50 88956.00
##                        min       max     range  skew kurtosis      se
## SlNo                  1.00    391.00    390.00  0.00    -1.21    5.72
## Gender*               1.00      2.00      1.00 -0.75    -1.45    0.02
## Gender.B              0.00      1.00      1.00  0.75    -1.45    0.02
## Percent_SSC          37.00     87.20     50.20 -0.06    -0.72    0.55
## Board_SSC*            1.00      3.00      2.00 -0.45    -1.53    0.04
## Board_CBSE            0.00      1.00      1.00  0.93    -1.14    0.02
## Board_ICSE            0.00      1.00      1.00  1.52     0.31    0.02
## Percent_HSC          40.00     94.70     54.70  0.29    -0.67    0.58
## Board_HSC*            1.00      3.00      2.00 -0.83    -1.13    0.04
## Stream_HSC*           1.00      3.00      2.00 -0.12    -0.72    0.03
## Percent_Degree       35.00     89.00     54.00  0.05     0.24    0.45
## Course_Degree*        1.00      7.00      6.00  0.00    -1.08    0.08
## Degree_Engg           0.00      1.00      1.00  2.76     5.63    0.01
## Experience_Yrs        0.00      3.00      3.00  1.27     1.17    0.03
## Entrance_Test*        1.00      9.00      8.00 -1.23    -0.35    0.12
## S.TEST                0.00      1.00      1.00 -1.74     1.02    0.02
## Percentile_ET         0.00     98.69     98.69 -0.74    -0.69    1.58
## Percent_MBA          50.83     77.89     27.06  0.34    -0.52    0.30
## S.TEST.SCORE          0.00     98.69     98.69 -0.74    -0.69    1.58
## Specialization_MBA*   1.00      3.00      2.00  0.70    -0.56    0.03
## Marks_Communication  50.00     88.00     38.00  0.74    -0.25    0.45
## Marks_Projectwork    50.00     87.00     37.00 -0.26    -0.27    0.36
## Marks_BOCA           50.00     96.00     46.00  0.29    -0.85    0.48
## Placement*            1.00      2.00      1.00 -1.48     0.19    0.02
## Placement_B           0.00      1.00      1.00 -1.48     0.19    0.02
## Salary                0.00 940000.00 940000.00  0.24     1.74 6994.72
```

```
str(dataset)
```

```
## 'data.frame':    391 obs. of  26 variables:
##  $ SlNo               : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Gender             : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 1 2 2 1 ...
##  $ Gender.B           : int  0 0 0 0 0 0 1 0 0 1 ...
##  $ Percent_SSC        : num  62 76.3 72 60 61 ...
##  $ Board_SSC          : Factor w/ 3 levels "CBSE","ICSE",..: 3 2 3 1 1 2 3 2 1 1 ...
##  $ Board_CBSE         : int  0 0 0 1 1 0 0 0 1 1 ...
##  $ Board_ICSE         : int  0 1 0 0 0 1 0 1 0 0 ...
##  $ Percent_HSC        : num  88 75.3 78 63 55 ...
##  $ Board_HSC          : Factor w/ 3 levels "CBSE","ISC","Others": 3 3 3 1 2 1 3 2 1 1 ...
##  $ Stream_HSC         : Factor w/ 3 levels "Arts","Commerce",..: 2 3 2 1 3 2 3 2 2 1 ...
##  $ Percent_Degree     : num  52 75.5 66.6 58 54 ...
##  $ Course_Degree      : Factor w/ 7 levels "Arts","Commerce",..: 7 3 4 5 4 2 6 5 2 5 ...
##  $ Degree_Engg        : int  0 0 1 0 1 0 0 0 0 0 ...
##  $ Experience_Yrs     : int  0 1 0 0 1 0 2 0 0 1 ...
##  $ Entrance_Test      : Factor w/ 9 levels "","CAT","G-MAT",..: 7 7 1 7 7 1 1 7 7 1 ...
##  $ S.TEST             : int  1 1 0 1 1 0 0 1 1 0 ...
##  $ Percentile_ET      : num  55 86.5 0 75 66 ...
##  $ Percent_MBA        : num  58.8 66.3 52.9 57.8 59.4 ...
##  $ S.TEST.SCORE       : num  55 86.5 0 75 66 ...
##  $ Specialization_MBA : Factor w/ 3 levels "Marketing & Finance",..: 2 1 1 1 2 1 2 1 1 2
...
##  $ Marks_Communication: int  50 69 50 54 52 53 63 74 65 50 ...
##  $ Marks_Projectwork  : int  65 70 61 66 65 70 56 72 76 59 ...
##  $ Marks_BOCA         : int  74 75 59 62 67 53 50 50 70 77 ...
##  $ Placement          : Factor w/ 2 levels "Not Placed","Placed": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Placement_B        : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Salary             : int  270000 200000 240000 250000 180000 300000 260000 235000 42500
0 240000 ...
```

```
#Median salary of all the students in the dataset
median(dataset$Salary)
```

```
## [1] 240000
```

```
#Percentage of students who were placed
options(digits=4)
mytable=prop.table(table(dataset$Placement))*100
mytable[2]
```

```
## Placed
##   79.8
```

```
#created a dataframe called placed
placed=dataset[which(dataset$Placement=='Placed'),]
View(placed)

#created a dataframe called Not placed
not_placed=dataset[which(dataset$Placement=='Not Placed'),]
View(not_placed)
```

```
#Median salary of placed students
median(placed$Salary)
```

```
## [1] 260000
```

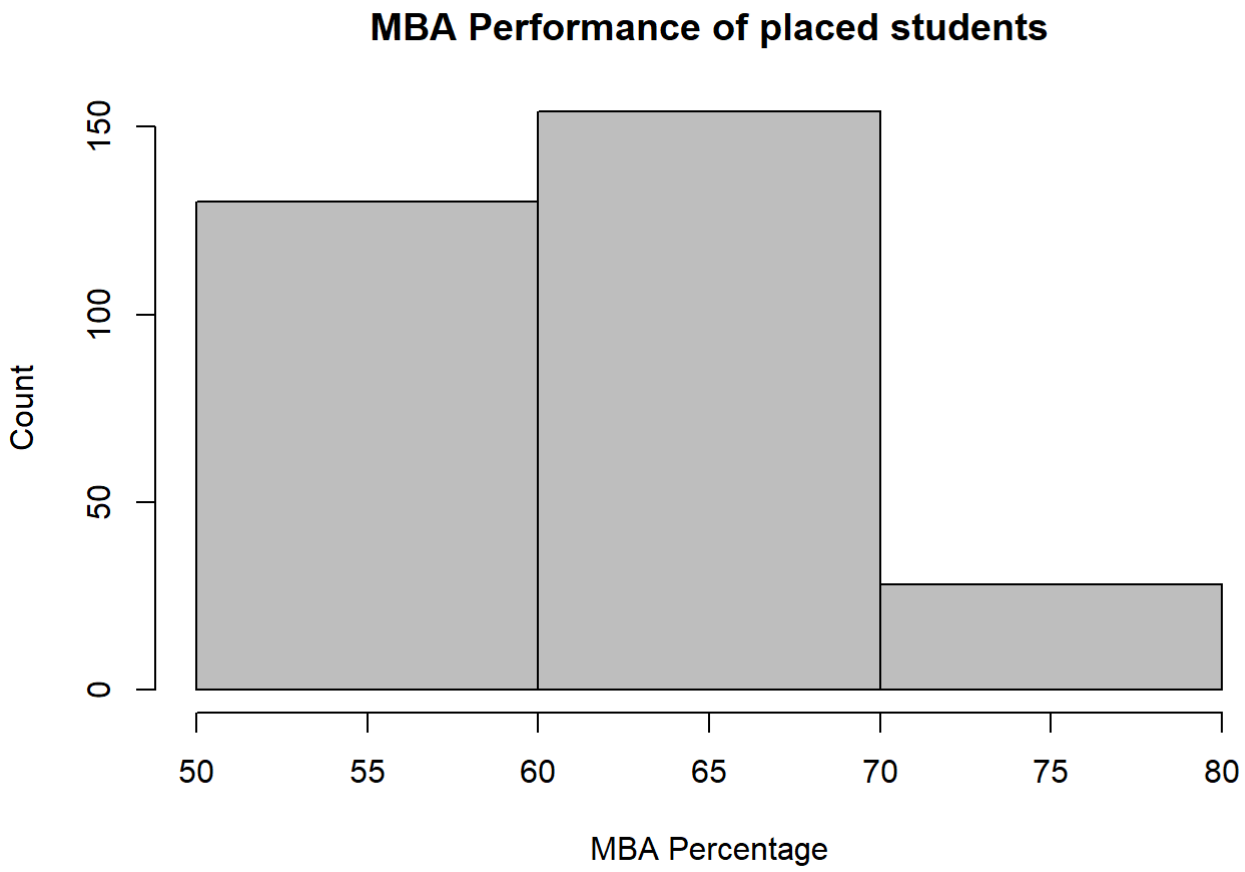
```
#Salary Comparison
mytable <- aggregate(Salary~Gender ,data=placed,mean)
mytable
```

```
##   Gender Salary
## 1      F 253068
## 2      M 284242
```

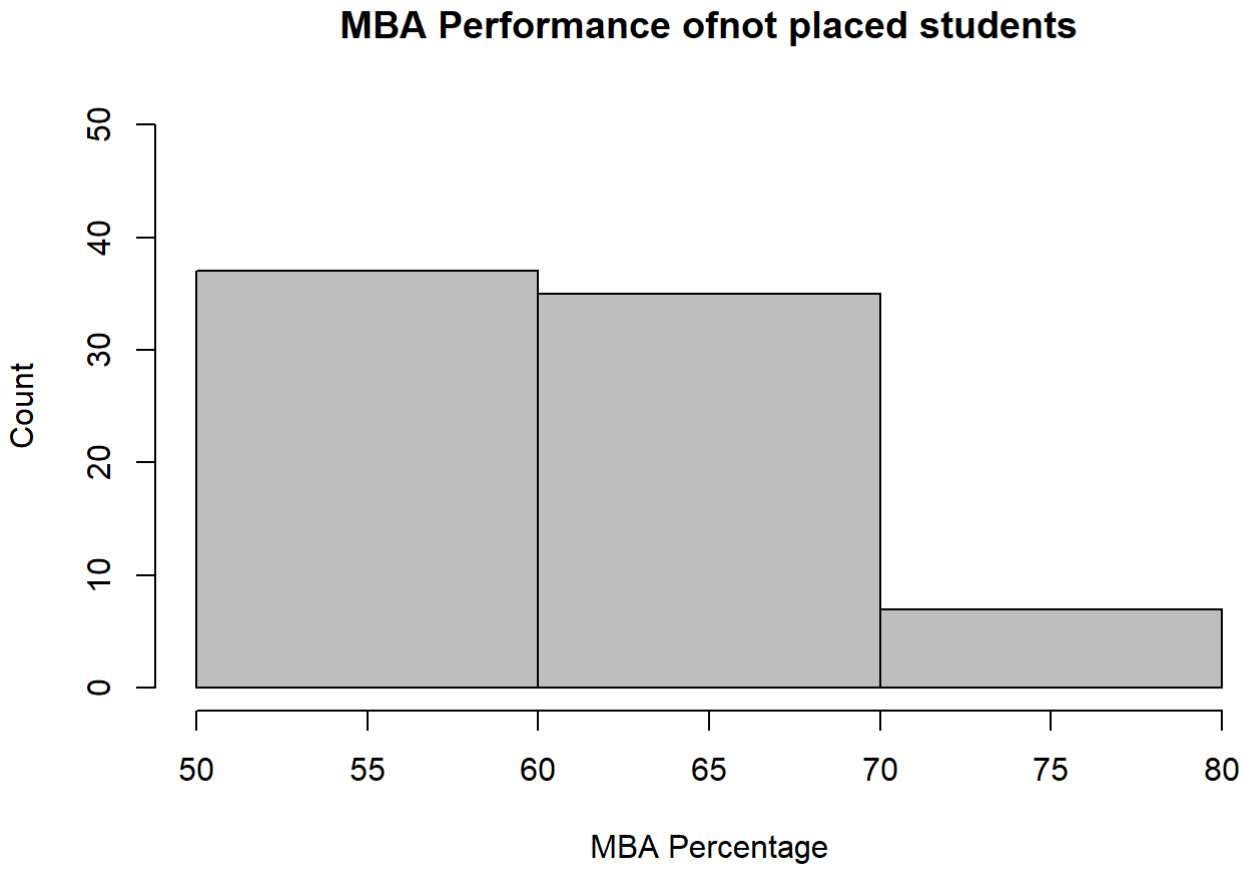
```
#boxplots, comparing the distribution of salaries of males and females who were placed
bwplot(placed$Gender~placed$Salary ,data=placed,xlab="Salary",ylab="Gender",main = "Compariso
n of Salaries of Males and Females")
```

```
#MBA performance of the students who were placed and Not Placed students
par(mfrow=c(1,2))
```

```
hist(placed$Percent_MBA ,main="MBA Performance of placed students",
     xlab="MBA Percentage",ylab="Count",
     xlim=c(50,80),ylim=c(0,150),breaks=3,col="grey")
```




```
hist(not_placed$Percent_MBA ,main="MBA Performance ofnot placed students",
     xlab="MBA Percentage",ylab="Count",
     xlim=c(50,80),ylim=c(0,50),breaks=3,col="grey")
```

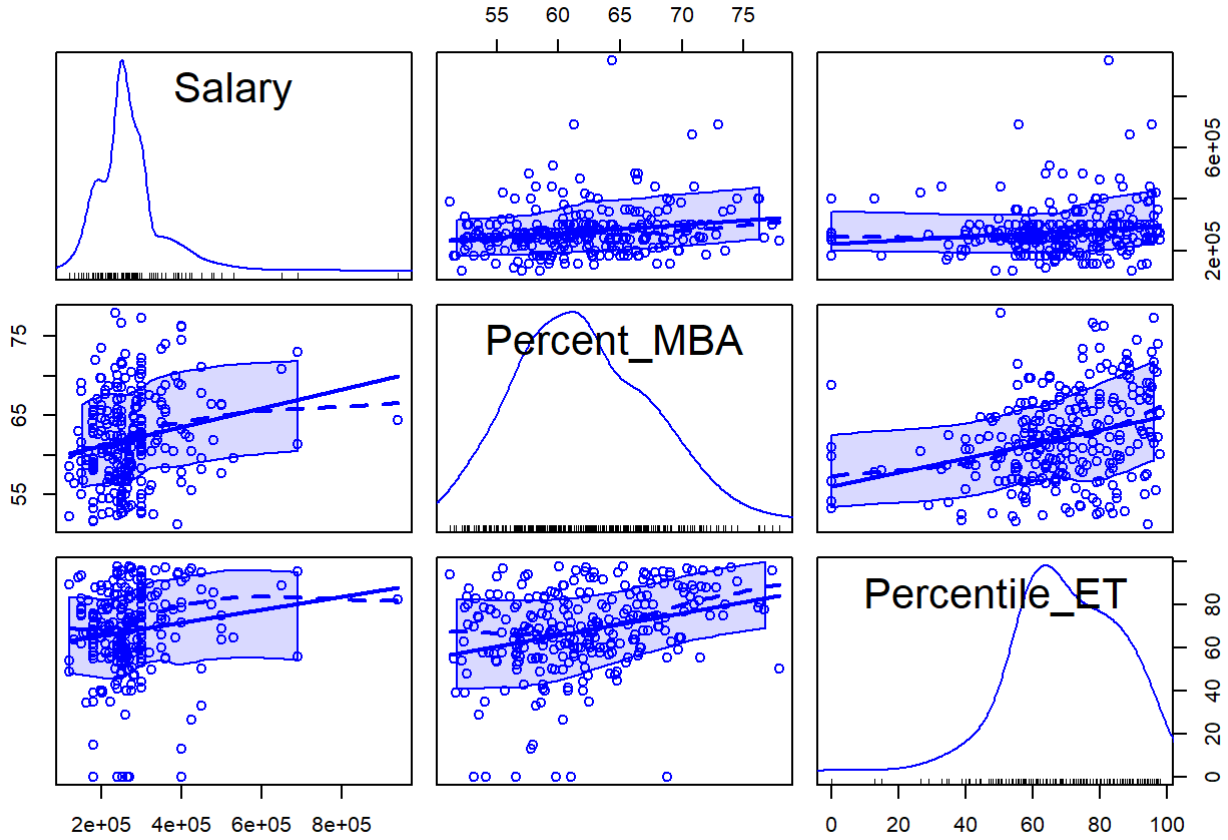


Datafram called placed_entrance, representing students who were placed after the MBA.

```
# Students who gave some MBA entrance test before admission into the MBA program.
placed_entrance <- placed[which(placed$S.TEST==1),]
View(placed_entrance)
```

Scatter Plot

```
scatterplotMatrix(formula= ~Salary+Percent_MBA+Percentile_ET ,data=placed_entrance)
```



#Logistic Regression Model with all variables except 'Salary'

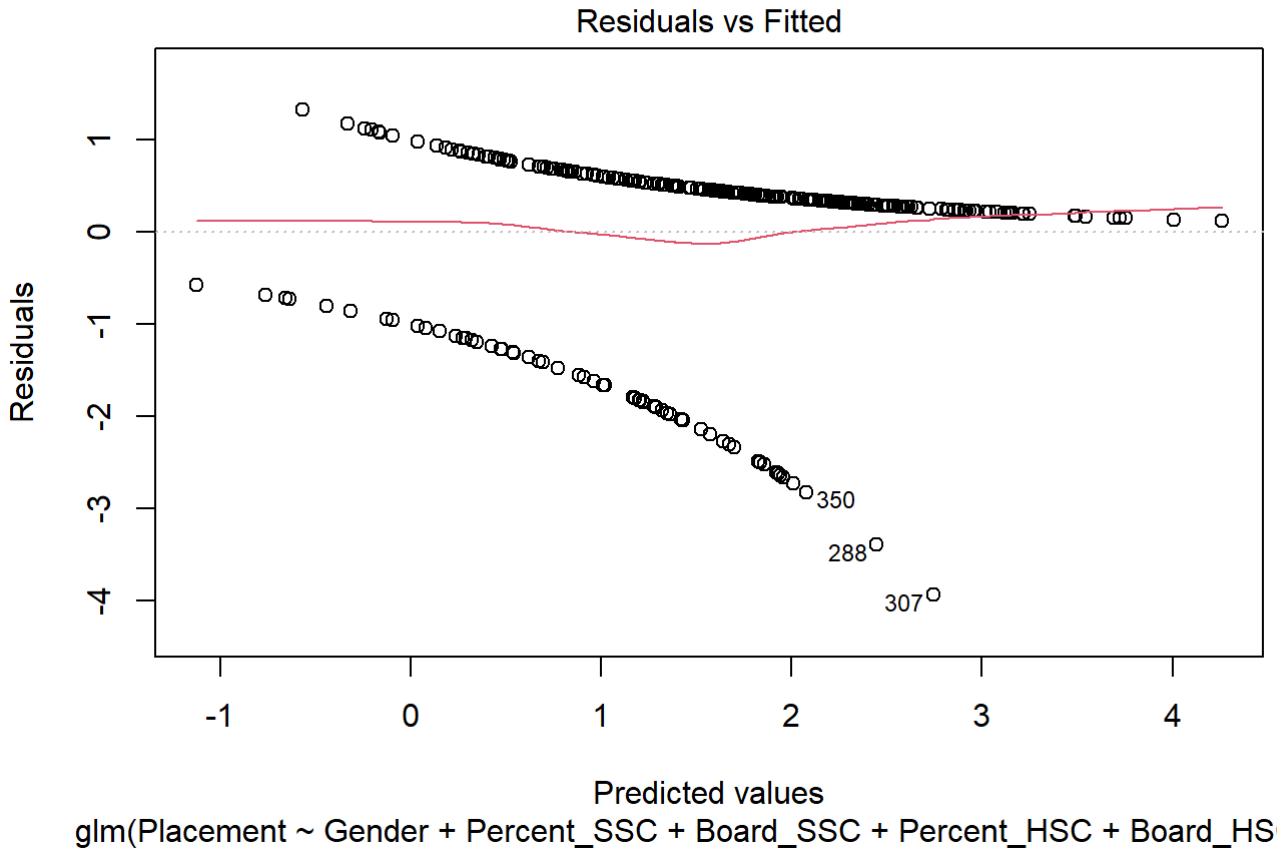
```
set.seed(123)
dataset$Placement <- as.factor(dataset$Placement)
dataset$S.TEST <- as.factor(dataset$S.TEST)
random_sample <- createDataPartition(dataset$Placement, p = 0.8, list = FALSE)
training_dataset  <- dataset[random_sample, ]
testing_dataset <- dataset[-random_sample, ]
glm_model <- glm(Placement ~ Gender+Percent_SSC+Board_SSC+Percent_HSC+Board_HSC+Stream_HSC+Pe
rcent_Degree+
                   Degree_Engg+Course_Degree+Experience_Yrs+S.TEST+Percentile_ET+Percent_MBA+
Specialization_MBA
                 +Marks_Communication+Marks_Projectwork+Marks_BOCA, data = training_dataset,
family = "binomial")
varImp(glm_model, scaled = TRUE)
```

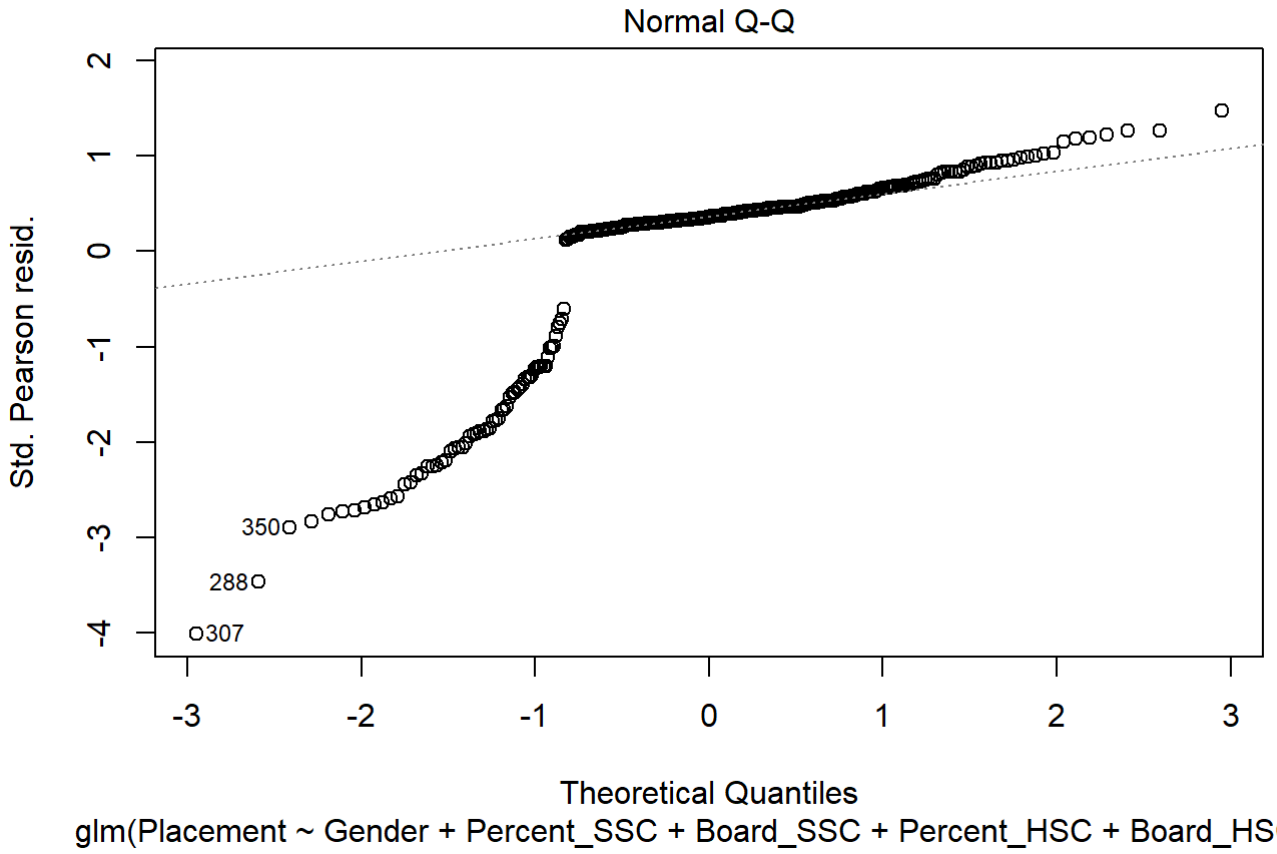
```
##                                  Overall
## GenderM                          1.65339
## Percent_SSC                      3.03948
## Board_SSCICSE                    0.24447
## Board_SSCOthers                  0.07922
## Percent_HSC                      1.59577
## Board_HSCISC                     1.39728
## Board_HSCOthers                  1.31208
## Stream_HSCCommerce               1.27928
## Stream_HSCScience                0.83983
## Percent_Degree                   1.10764
## Degree_Engg                      0.24437
## Course_DegreeCommerce            1.34457
## Course_DegreeComputer Applications 0.02527
## Course_DegreeManagement          0.93335
## Course_DegreeOthers              0.13710
## Course_DegreeScience             0.55993
## Experience_Yrs                   0.47828
## S.TEST1                          1.14957
## Percentile_ET                    1.89924
## Percent_MBA                      0.05535
## Specialization_MBAMarketing & HR 1.01371
## Specialization_MBAMarketing & IB 0.12974
## Marks_Communication              2.05664
## Marks_Projectwork                2.16731
## Marks_BOCA                       0.14374
```

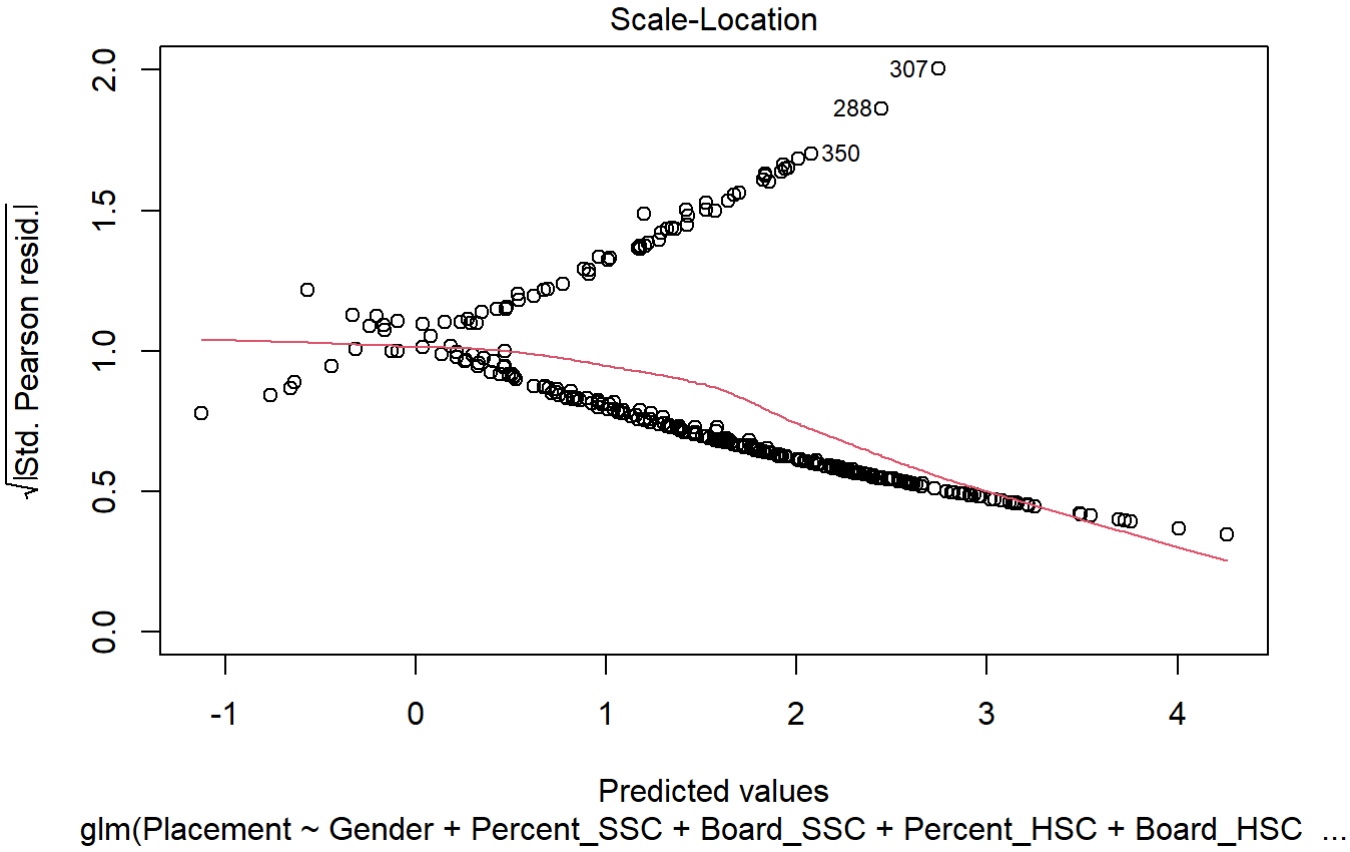
```
summary(glm_model)
```

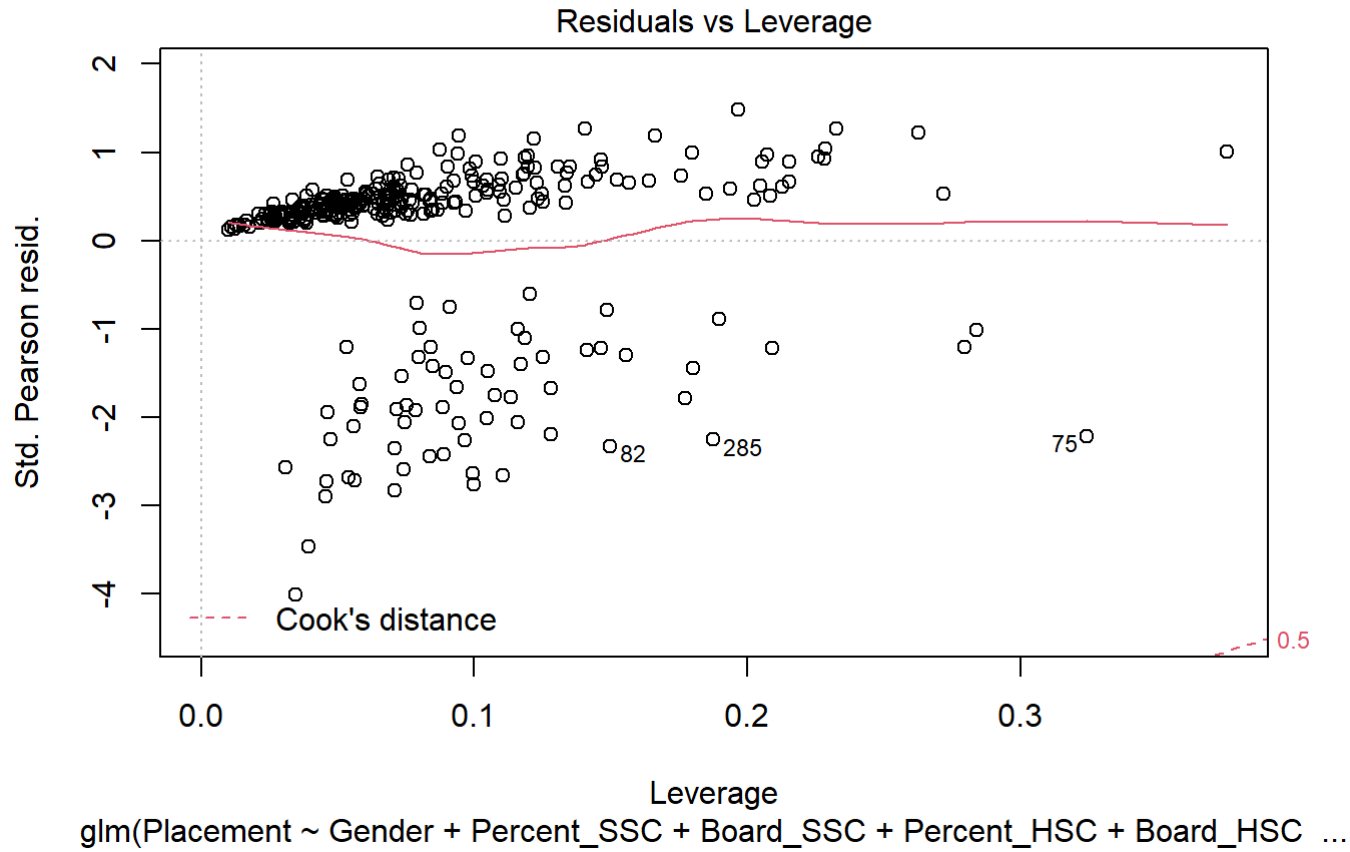
```
##
## Call:
## glm(formula = Placement ~ Gender + Percent_SSC + Board_SSC +
##     Percent_HSC + Board_HSC + Stream_HSC + Percent_Degree + Degree_Engg +
##     Course_Degree + Experience_Yrs + S.TEST + Percentile_ET +
##     Percent_MBA + Specialization_MBA + Marks_Communication +
##     Marks_Projectwork + Marks_BOCA, family = "binomial", data = training_dataset)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.369   0.290   0.481   0.669   1.426
##
## Coefficients: (1 not defined because of singularities)
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -2.88004    2.35720   -1.22   0.2218
## GenderM                            0.58217    0.35211    1.65   0.0983 .
## Percent_SSC                        0.05996    0.01973    3.04   0.0024 **
## Board_SSCICSE                      0.14474    0.59207    0.24   0.8069
## Board_SSCOthers                    0.04050    0.51127    0.08   0.9369
## Percent_HSC                       -0.02920    0.01830   -1.60   0.1105
## Board_HSCISC                      -0.94626    0.67722   -1.40   0.1623
## Board_HSCOthers                   -0.71657    0.54614   -1.31   0.1895
## Stream_HSCCommerce                -1.18312    0.92483   -1.28   0.2008
## Stream_HSCScience                 -0.82385    0.98096   -0.84   0.4010
## Percent_Degree                     0.02516    0.02272    1.11   0.2680
## Degree_Engg                        0.26665    1.09115    0.24   0.8069
## Course_DegreeCommerce              1.35945    1.01107    1.34   0.1788
## Course_DegreeComputer Applications -0.02729    1.07969   -0.03   0.9798
## Course_DegreeEngineering                NA         NA      NA       NA
## Course_DegreeManagement            0.89970    0.96395    0.93   0.3506
## Course_DegreeOthers                0.21186    1.54528    0.14   0.8910
## Course_DegreeScience               0.65662    1.17268    0.56   0.5755
## Experience_Yrs                     0.12516    0.26169    0.48   0.6324
## S.TEST1                           -0.81449    0.70851   -1.15   0.2503
## Percentile_ET                      0.01818    0.00957    1.90   0.0575 .
## Percent_MBA                        0.00254    0.04590    0.06   0.9559
## Specialization_MBAMarketing & HR   0.34144    0.33683    1.01   0.3107
## Specialization_MBAMarketing & IB   0.10753    0.82878    0.13   0.8968
## Marks_Communication               -0.05529    0.02688   -2.06   0.0397 *
## Marks_Projectwork                  0.05313    0.02451    2.17   0.0302 *
## Marks_BOCA                         0.00290    0.02016    0.14   0.8857
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.55  on 313  degrees of freedom
## Residual deviance: 279.56  on 288  degrees of freedom
## AIC: 331.6
##
## Number of Fisher Scoring iterations: 5
```

```
plot(glm_model)
```


Residuals vs Fitted
Residuals
Predicted values
glm(Placement ~ Gender + Percent_SSC + Board_SSC + Percent_HSC + Board_HSC ...
350
288
307


Normal Q-Q
Std. Pearson resid.
Theoretical Quantiles
glm(Placement ~ Gender + Percent_SSC + Board_SSC + Percent_HSC + Board_HSC ...
350
288
307






```
pred <- predict(glm_model, testing_dataset, type = "response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from a rank-deficient fit may be misleading
```

```
pred <- ifelse(pred>0.7, "Placed", "Not Placed")
pred <- as.factor(pred)
confusionMatrix(pred, testing_dataset$Placement, mode = "prec_recall", positive = "Placed")
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   Not Placed Placed
##   Not Placed          4     15
##   Placed             11     47
##
##                Accuracy : 0.662
##                  95% CI : (0.546, 0.766)
##     No Information Rate : 0.805
##     P-Value [Acc > NIR] : 0.999
##
##                   Kappa : 0.022
##
##  Mcnemar's Test P-Value : 0.556
##
##               Precision : 0.810
##                  Recall : 0.758
##                      F1 : 0.783
##              Prevalence : 0.805
##          Detection Rate : 0.610
##    Detection Prevalence : 0.753
##       Balanced Accuracy : 0.512
##
##        'Positive' Class : Placed
##
```

#Logistic Regression with important variables (BEST MODEL)

```
glm_model_2 <- glm(Placement ~Gender+Percent_SSC+Board_SSC+Board_HSC+Stream_HSC+Percent_Degre
e
                   +Degree_Engg+S.TEST+Percentile_ET+Marks_Projectwork+Marks_BOCA, data = tra
ining_dataset, family = "binomial")
varImp(glm_model_2, scaled = TRUE)
```

```
##                     Overall
## GenderM            1.953042
## Percent_SSC        2.010077
## Board_SSCICSE      0.359607
## Board_SSCOthers    0.100943
## Board_HSCISC       1.218609
## Board_HSCOthers    1.287429
## Stream_HSCCommerce 0.726519
## Stream_HSCScience  0.308430
## Percent_Degree     0.005853
## Degree_Engg        1.015739
## S.TEST1            0.990720
## Percentile_ET      1.715258
## Marks_Projectwork  1.866876
## Marks_BOCA         0.163085
```

```
summary(glm_model_2)
```

```
##
## Call:
## glm(formula = Placement ~ Gender + Percent_SSC + Board_SSC +
##     Board_HSC + Stream_HSC + Percent_Degree + Degree_Engg + S.TEST +
##     Percentile_ET + Marks_Projectwork + Marks_BOCA, family = "binomial",
##     data = training_dataset)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.316   0.355   0.529   0.714   1.262
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -3.384405   2.001040   -1.69    0.091 .
## GenderM             0.641388   0.328405    1.95    0.051 .
## Percent_SSC         0.033215   0.016524    2.01    0.044 *
## Board_SSCICSE      -0.199589   0.555021   -0.36    0.719
## Board_SSCOthers    -0.048410   0.479580   -0.10    0.920
## Board_HSCISC       -0.792364   0.650220   -1.22    0.223
## Board_HSCOthers    -0.664188   0.515902   -1.29    0.198
## Stream_HSCCommerce -0.523030   0.719912   -0.73    0.468
## Stream_HSCScience  -0.237040   0.768538   -0.31    0.758
## Percent_Degree      0.000114   0.019400    0.01    0.995
## Degree_Engg        -0.583771   0.574726   -1.02    0.310
## S.TEST1            -0.632397   0.638321   -0.99    0.322
## Percentile_ET       0.014783   0.008619    1.72    0.086 .
## Marks_Projectwork   0.042042   0.022520    1.87    0.062 .
## Marks_BOCA          0.002841   0.017421    0.16    0.870
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.55  on 313  degrees of freedom
## Residual deviance: 292.71  on 299  degrees of freedom
## AIC: 322.7
##
## Number of Fisher Scoring iterations: 4
```

```
pred2 <- predict(glm_model_2, testing_dataset, type = "link")
pred2 <- ifelse(pred2>0.7, "Placed", "Not Placed")
pred2 <- as.factor(pred2)
confusionMatrix(pred2, testing_dataset$Placement, mode = "prec_recall", positive = "Placed")
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   Not Placed Placed
##   Not Placed          4      8
##   Placed             11     54
##
##                Accuracy : 0.753
##                  95% CI : (0.642, 0.844)
##     No Information Rate : 0.805
##     P-Value [Acc > NIR] : 0.900
##
##                   Kappa : 0.149
##
##  Mcnemar's Test P-Value : 0.646
##
##               Precision : 0.831
##                  Recall : 0.871
##                      F1 : 0.850
##              Prevalence : 0.805
##          Detection Rate : 0.701
##    Detection Prevalence : 0.844
##       Balanced Accuracy : 0.569
##
##        'Positive' Class : Placed
##
```

#Model with Only SSC Percentage.

```
glm_model_3 <- glm(Placement ~Percent_SSC, data = training_dataset, family = "binomial")
varImp(glm_model_3, scaled = TRUE)
```

```
##             Overall
## Percent_SSC   2.705
```

```
summary(glm_model_3)
```

```
##
## Call:
## glm(formula = Placement ~ Percent_SSC, family = "binomial", data = training_dataset)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.125   0.500   0.609   0.713   1.001
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.8726     0.8227   -1.06   0.2889
## Percent_SSC   0.0352     0.0130    2.70   0.0068 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.55  on 313  degrees of freedom
## Residual deviance: 309.99  on 312  degrees of freedom
## AIC: 314
##
## Number of Fisher Scoring iterations: 4
```

```
pred3 <- predict(glm_model_3, testing_dataset, type = "link")
pred3 <- ifelse(pred3>0.7, "Placed", "Not Placed")
pred3 <- as.factor(pred3)
confusionMatrix(pred3, testing_dataset$Placement, mode = "prec_recall", positive = "Placed")
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   Not Placed Placed
##   Not Placed          1      2
##   Placed             14     60
##
##                Accuracy : 0.792
##                  95% CI : (0.685, 0.876)
##     No Information Rate : 0.805
##     P-Value [Acc > NIR] : 0.67548
##
##                   Kappa : 0.049
##
##  Mcnemar's Test P-Value : 0.00596
##
##               Precision : 0.811
##                  Recall : 0.968
##                      F1 : 0.882
##              Prevalence : 0.805
##          Detection Rate : 0.779
##    Detection Prevalence : 0.961
##       Balanced Accuracy : 0.517
##
##        'Positive' Class : Placed
##
```

#Model for students being getting into the program with 60% in SSC.

```
newdata4 <-  data.frame(Percent_SSC = c(60))
pred4 <- predict(glm_model_3, newdata4, type="response")
pred4
```

```
##      1
## 0.7755
```

#Model for students being getting into the program with 80% in SSC.

```
newdata6 <-  data.frame(Percent_SSC = c(80))
pred6 <- predict(glm_model_3, newdata6, type="response")
pred6
```

```
##      1
## 0.8747
```

#Logistic Regression for the specified Variables.

```
glm_model_7 <- glm(Placement ~Gender+Percent_SSC+Percentile_ET+Marks_Projectwork+Marks_Commun
ication, data = training_dataset, family = "binomial")
varImp(glm_model_7, scaled = TRUE)
```

```
##                     Overall
## GenderM               1.405
## Percent_SSC           3.071
## Percentile_ET         2.039
## Marks_Projectwork     2.185
## Marks_Communication   2.269
```

```
summary(glm_model_7)
```

```
##
## Call:
## glm(formula = Placement ~ Gender + Percent_SSC + Percentile_ET +
##     Marks_Projectwork + Marks_Communication, family = "binomial",
##     data = training_dataset)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.182   0.381   0.545   0.718   1.350
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -2.84299    1.73956   -1.63   0.1022
## GenderM              0.44486    0.31668    1.40   0.1601
## Percent_SSC          0.04803    0.01564    3.07   0.0021 **
## Percentile_ET        0.00955    0.00469    2.04   0.0415 *
## Marks_Projectwork    0.04831    0.02211    2.18   0.0289 *
## Marks_Communication -0.04765    0.02100   -2.27   0.0233 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.55  on 313  degrees of freedom
## Residual deviance: 295.20  on 308  degrees of freedom
## AIC: 307.2
##
## Number of Fisher Scoring iterations: 4
```

```
pred7 <- predict(glm_model_7, testing_dataset, type = "link")
pred7 <- ifelse(pred7>0.7, "Placed", "Not Placed")
pred7 <- as.factor(pred7)
confusionMatrix(pred7, testing_dataset$Placement, mode = "prec_recall", positive = "Placed")
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction   Not Placed Placed
##   Not Placed          3      6
##   Placed             12     56
##
##                Accuracy : 0.766
##                  95% CI : (0.656, 0.855)
##     No Information Rate : 0.805
##     P-Value [Acc > NIR] : 0.843
##
##                   Kappa : 0.122
##
##  Mcnemar's Test P-Value : 0.239
##
##               Precision : 0.824
##                  Recall : 0.903
##                      F1 : 0.862
##              Prevalence : 0.805
##          Detection Rate : 0.727
##    Detection Prevalence : 0.883
##       Balanced Accuracy : 0.552
##
##        'Positive' Class : Placed
##
```