

student_performance_code

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
#reading csv
stud_perf <- read.csv("C:/Users/cmanz/OneDrive/Documents/Ryerson
stuff/cind820/student dataset/student-perf.csv", header = T, stringsAsFactors
= F, na.strings = c("", "NA"), sep = ";")
head(stud_perf)
```

##	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob
## 1	GP	F	18	U	GT3	A	4	4	at_home	teacher
## 2	GP	F	17	U	GT3	T	1	1	at_home	other
## 3	GP	F	15	U	LE3	T	1	1	at_home	other
## 4	GP	F	15	U	GT3	T	4	2	health	services
## 5	GP	F	16	U	GT3	T	3	3	other	other
## 6	GP	M	16	U	LE3	T	4	3	services	other

```
reputation
## guardian traveltime studytime failures schoolsup famsup paid activities
## 1 mother 2 2 0 yes no no no
## 2 father 1 2 0 no yes no no
## 3 mother 1 2 3 yes no yes no
## 4 mother 1 3 0 no yes yes yes
## 5 father 1 2 0 no yes yes no
## 6 mother 1 2 0 no yes yes yes
## nursery higher internet romantic famrel freetime goout Dalc Walc health
## 1 yes yes no no 4 3 4 1 1 3
## 2 no yes yes no 5 3 3 1 1 3
## 3 yes yes yes no 4 3 2 2 3 3
## 4 yes yes yes yes 3 2 2 1 1 5
## 5 yes yes no no 4 3 2 1 2 5
```

```
## 6      yes      yes      yes      no      5      4      2      1      2      5
## absences G1 G2 G3
## 1          6  5  6  6
## 2          4  5  5  6
## 3         10  7  8 10
## 4          2 15 14 15
## 5          4  6 10 10
## 6         10 15 15 15
```

#checking the datatypes of the attributes

```
str(stud_perf)
```

```
## 'data.frame':    1044 obs. of  33 variables:
## $ school      : chr  "GP" "GP" "GP" "GP" ...
## $ sex         : chr  "F" "F" "F" "F" ...
## $ age         : int   18 17 15 15 16 16 16 17 15 15 ...
## $ address     : chr  "U" "U" "U" "U" ...
## $ famsize     : chr  "GT3" "GT3" "LE3" "GT3" ...
## $ Pstatus     : chr  "A" "T" "T" "T" ...
## $ Medu        : int   4 1 1 4 3 4 2 4 3 3 ...
## $ Fedu        : int   4 1 1 2 3 3 2 4 2 4 ...
## $ Mjob        : chr  "at_home" "at_home" "at_home" "health" ...
## $ Fjob        : chr  "teacher" "other" "other" "services" ...
## $ reason      : chr  "course" "course" "other" "home" ...
## $ guardian    : chr  "mother" "father" "mother" "mother" ...
## $ traveltime  : int   2 1 1 1 1 1 1 2 1 1 ...
## $ studytime   : int   2 2 2 3 2 2 2 2 2 2 ...
## $ failures    : int   0 0 3 0 0 0 0 0 0 0 ...
## $ schoolsup   : chr  "yes" "no" "yes" "no" ...
## $ famsup      : chr  "no" "yes" "no" "yes" ...
## $ paid        : chr  "no" "no" "yes" "yes" ...
## $ activities  : chr  "no" "no" "no" "yes" ...
## $ nursery     : chr  "yes" "no" "yes" "yes" ...
## $ higher      : chr  "yes" "yes" "yes" "yes" ...
## $ internet    : chr  "no" "yes" "yes" "yes" ...
## $ romantic    : chr  "no" "no" "no" "yes" ...
## $ famrel      : int   4 5 4 3 4 5 4 4 4 5 ...
## $ freetime    : int   3 3 3 2 3 4 4 1 2 5 ...
## $ goout       : int   4 3 2 2 2 2 4 4 2 1 ...
## $ Dalc        : int   1 1 2 1 1 1 1 1 1 1 ...
## $ Walc        : int   1 1 3 1 2 2 1 1 1 1 ...
## $ health      : int   3 3 3 5 5 5 3 1 1 5 ...
## $ absences    : int   6 4 10 2 4 10 0 6 0 0 ...
## $ G1          : int   5 5 7 15 6 15 12 6 16 14 ...
## $ G2          : int   6 5 8 14 10 15 12 5 18 15 ...
## $ G3          : int   6 6 10 15 10 15 11 6 19 15 ...
```

#checking for missing values

```
sum(is.na(stud_perf))
```

```
## [1] 0
```

```
#Looking for correlation between numeric attributes except final grade(G3)
cor(stud_perf[, c('age', 'Medu', 'Fedu', 'traveltime', 'studytime',
'failures', 'famrel', 'freetime', 'goout', 'Dalc', 'Walc', 'health',
'absences',
'G1', 'G2')])
```

```
##          age          Medu          Fedu    traveltime
studytime
## age          1.000000000 -0.130196115 -0.1385207614  0.049215707 -
0.007870098
## Medu        -0.130196115  1.000000000  0.6420631457 -0.238180728
0.090616377
## Fedu        -0.138520761  0.642063146  1.0000000000 -0.196328161
0.033457874
## traveltime  0.049215707 -0.238180728 -0.1963281605  1.000000000 -
0.081328016
## studytime -0.007870098  0.090616377  0.0334578745 -0.081328016
1.000000000
## failures    0.282363566 -0.187769404 -0.1913904210  0.087177495 -
0.152023523
## famrel      0.007161921  0.015003618  0.0130659150 -0.012577522
0.012324093
## freetime    0.002645147  0.001054219  0.0021417298 -0.007402578 -
0.094429345
## goout       0.118510124  0.025614278  0.0300748764  0.049739783 -
0.072940739
## Dalc        0.133452990  0.001515097 -0.0001648393  0.109423016 -
0.159664641
## Walc        0.098291406 -0.029330541  0.0195239342  0.084292404 -
0.229073148
## health     -0.029129265 -0.013254090  0.0342882377 -0.029001978 -
0.063044459
## absences    0.153195647  0.059707676  0.0408288855 -0.022668699 -
0.075593669
## G1         -0.124121249  0.226100602  0.1958980209 -0.121053301
0.211313915
## G2         -0.119474744  0.224661748  0.1826339619 -0.140162973
0.183166702
##          failures          famrel          freetime          goout          Dalc
## age          0.28236357  0.007161921  0.002645147  0.11851012  0.1334529897
## Medu        -0.18776940  0.015003618  0.001054219  0.02561428  0.0015150967
## Fedu        -0.19139042  0.013065915  0.002141730  0.03007488 -0.0001648393
## traveltime  0.08717749 -0.012577522 -0.007402578  0.04973978  0.1094230162
## studytime -0.15202352  0.012324093 -0.094429345 -0.07294074 -0.1596646413
## failures    1.00000000 -0.053676457  0.102678757  0.07468331  0.1163357901
## famrel     -0.05367646  1.000000000  0.136900650  0.08061921 -0.0764826572
## freetime    0.10267876  0.136900650  1.000000000  0.32355575  0.1449791279
## goout       0.07468331  0.080619212  0.323555753  1.000000000  0.2531348291
## Dalc        0.11633579 -0.076482657  0.144979128  0.25313483  1.0000000000
## Walc        0.10743159 -0.100663375  0.130377028  0.39979373  0.6278138380
## health      0.04831102  0.104100776  0.081517225 -0.01373623  0.0655153422
```

```
## absences      0.09999785 -0.062170662 -0.032078736  0.05614214  0.1328671345
## G1            -0.37417487  0.036947274 -0.051984712 -0.10116347 -0.1509425374
## G2            -0.37717218  0.042053621 -0.068951886 -0.10841089 -0.1315764840
##              Walc      health      absences      G1      G2
## age           0.09829141 -0.02912927  0.15319565 -0.12412125 -0.11947474
## Medu          -0.02933054 -0.01325409  0.05970768  0.22610060  0.22466175
## Fedu          0.01952393  0.03428824  0.04082889  0.19589802  0.18263396
## traveltime    0.08429240 -0.02900198 -0.02266870 -0.12105330 -0.14016297
## studytime     -0.22907315 -0.06304446 -0.07559367  0.21131391  0.18316670
## failures      0.10743159  0.04831102  0.09999785 -0.37417487 -0.37717218
## famrel        -0.10066338  0.10410078 -0.06217066  0.03694727  0.04205362
## freetime      0.13037703  0.08151722 -0.03207874 -0.05198471 -0.06895189
## goout         0.39979373 -0.01373623  0.05614214 -0.10116347 -0.10841089
## Dalc          0.62781384  0.06551534  0.13286713 -0.15094254 -0.13157648
## Walc          1.00000000  0.10666944  0.13970313 -0.14240140 -0.12811435
## health        0.10666944  1.00000000 -0.02747860 -0.06047794 -0.08800109
## absences      0.13970313 -0.02747860  1.00000000 -0.09242463 -0.08933169
## G1            -0.14240140 -0.06047794 -0.09242463  1.00000000  0.85873875
## G2            -0.12811435 -0.08800109 -0.08933169  0.85873875  1.00000000
```

#graphing frequency distribution of final grade(G3)

```
library(epiDisplay)
```

```
## Warning: package 'epiDisplay' was built under R version 4.1.3
```

```
## Loading required package: foreign
```

```
## Loading required package: survival
```

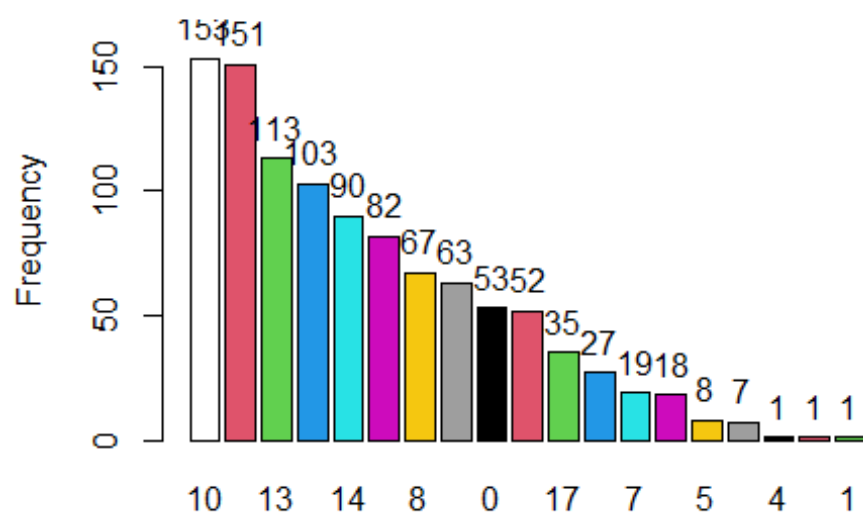
```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 4.1.2
```

```
## Loading required package: nnet
```

```
tab1(stud_perf$G3, sort.group = "decreasing", cum.percent = T)
```

Distribution of stud_perf\$G3



```
## stud_perf$G3 :
##      Frequency Percent  Cum. percent
## 10           153     14.7         14.7
## 11           151     14.5         29.1
## 13           113     10.8         39.9
## 12           103     9.9          49.8
## 14            90     8.6          58.4
## 15            82     7.9          66.3
## 8             67     6.4          72.7
## 9             63     6.0          78.7
## 0            53     5.1          83.8
## 16            52     5.0          88.8
## 17            35     3.4          92.1
## 18            27     2.6          94.7
## 7             19     1.8          96.6
## 6             18     1.7          98.3
## 5              8     0.8          99.0
## 19              7     0.7          99.7
## 4              1     0.1          99.8
## 20              1     0.1          99.9
## 1              1     0.1         100.0
## Total        1044    100.0         100.0
```

#assigning numeric values to Mjob(mother's job) and Fjob(father's job).

1 - at_home

2 - services

3 - other

```
# 4 - teacher
```

```
# 5 - health
```

```
stud_perf$Mjob[stud_perf$Mjob == 'at_home'] = 1
stud_perf$Mjob[stud_perf$Mjob == 'services'] = 2
stud_perf$Mjob[stud_perf$Mjob == 'other'] = 3
stud_perf$Mjob[stud_perf$Mjob == 'teacher'] = 4
stud_perf$Mjob[stud_perf$Mjob == 'health'] = 5
```

```
stud_perf$Fjob[stud_perf$Fjob == 'at_home'] = 1
stud_perf$Fjob[stud_perf$Fjob == 'services'] = 2
stud_perf$Fjob[stud_perf$Fjob == 'other'] = 3
stud_perf$Fjob[stud_perf$Fjob == 'teacher'] = 4
stud_perf$Fjob[stud_perf$Fjob == 'health'] = 5
```

```
head(stud_perf)
```

```
##   school sex age address famsize Pstatus Medu Fedu Mjob Fjob   reason
## 1    GP   F  18      U    GT3      A    4    4    1    4   course
## 2    GP   F  17      U    GT3      T    1    1    1    3   course
## 3    GP   F  15      U    LE3      T    1    1    1    3   other
## 4    GP   F  15      U    GT3      T    4    2    5    2   home
## 5    GP   F  16      U    GT3      T    3    3    3    3   home
## 6    GP   M  16      U    LE3      T    4    3    2    3 reputation
##   guardian traveltime studytime failures schoolsup famsup paid activities
## 1   mother          2          2         0      yes    no   no         no
## 2   father          1          2         0      no    yes  no         no
## 3   mother          1          2         3      yes    no  yes         no
## 4   mother          1          3         0      no    yes  yes         yes
## 5   father          1          2         0      no    yes  yes         no
## 6   mother          1          2         0      no    yes  yes         yes
##   nursery higher internet romantic famrel freetime goout Dalc Walc health
## 1    yes    yes      no      no      4          3    4    1    1    3
## 2    no    yes      yes      no      5          3    3    1    1    3
## 3    yes    yes      yes      no      4          3    2    2    3    3
## 4    yes    yes      yes      yes      3          2    2    1    1    5
## 5    yes    yes      no      no      4          3    2    1    2    5
## 6    yes    yes      yes      no      5          4    2    1    2    5
##   absences G1 G2 G3
## 1        6  5  6  6
## 2        4  5  5  6
## 3       10  7  8 10
## 4        2 15 14 15
## 5        4  6 10 10
## 6       10 15 15 15
```

```
#assigning binary values to yes or no attributes (schoolsup, famsup, paid,
activities, nursery, higher, internet, romantic)
```

```
stud_perf$schoolsup[stud_perf$schoolsup == 'yes'] = 1
stud_perf$schoolsup[stud_perf$schoolsup == 'no'] = 0
```

```

stud_perf$famsup[stud_perf$famsup == 'yes'] = 1
stud_perf$famsup[stud_perf$famsup == 'no'] = 0

stud_perf$paid[stud_perf$paid == 'yes'] = 1
stud_perf$paid[stud_perf$paid == 'no'] = 0

stud_perf$activities[stud_perf$activities == 'yes'] = 1
stud_perf$activities[stud_perf$activities == 'no'] = 0

stud_perf$nursery[stud_perf$nursery == 'yes'] = 1
stud_perf$nursery[stud_perf$nursery == 'no'] = 0

stud_perf$higher[stud_perf$higher == 'yes'] = 1
stud_perf$higher[stud_perf$higher == 'no'] = 0

stud_perf$internet[stud_perf$internet == 'yes'] = 1
stud_perf$internet[stud_perf$internet == 'no'] = 0

stud_perf$romantic[stud_perf$romantic == 'yes'] = 1
stud_perf$romantic[stud_perf$romantic == 'no'] = 0

```

```
head(stud_perf)
```

```

##   school sex age address famsize Pstatus Medu Fedu Mjob Fjob      reason
## 1    GP   F  18      U    GT3      A    4    4    1    4    course
## 2    GP   F  17      U    GT3      T    1    1    1    3    course
## 3    GP   F  15      U    LE3      T    1    1    1    3    other
## 4    GP   F  15      U    GT3      T    4    2    5    2    home
## 5    GP   F  16      U    GT3      T    3    3    3    3    home
## 6    GP   M  16      U    LE3      T    4    3    2    3 reputation
##   guardian traveltime studytime failures schoolsup famsup paid activities
## 1   mother          2          2          0          1          0          0
## 2   father          1          2          0          0          1          0
## 3   mother          1          2          3          1          0          1
## 4   mother          1          3          0          0          1          1
## 5   father          1          2          0          0          1          1
## 6   mother          1          2          0          0          1          1
##   nursery higher internet romantic famrel freetime goout Dalc Walc health
## 1         1      1          0          0      4          3          4          1          1          3
## 2         0      1          1          0      5          3          3          1          1          3
## 3         1      1          1          0      4          3          2          2          3          3
## 4         1      1          1          1      3          2          2          1          1          5
## 5         1      1          0          0      4          3          2          1          2          5
## 6         1      1          1          0      5          4          2          1          2          5
##   absences G1 G2 G3
## 1         6  5  6  6
## 2         4  5  5  6
## 3        10  7  8 10
## 4         2 15 14 15

```

```
## 5      4  6 10 10
## 6      10 15 15 15
```

#assigning pass or fail to the grades columns

#for G1

```
stud_perf$G1[stud_perf$G1 == 0] = 'Fail'
stud_perf$G1[stud_perf$G1 == 1] = 'Fail'
stud_perf$G1[stud_perf$G1 == 2] = 'Fail'
stud_perf$G1[stud_perf$G1 == 3] = 'Fail'
stud_perf$G1[stud_perf$G1 == 4] = 'Fail'
stud_perf$G1[stud_perf$G1 == 5] = 'Fail'
stud_perf$G1[stud_perf$G1 == 6] = 'Fail'
stud_perf$G1[stud_perf$G1 == 7] = 'Fail'
stud_perf$G1[stud_perf$G1 == 8] = 'Fail'
stud_perf$G1[stud_perf$G1 == 9] = 'Fail'
stud_perf$G1[stud_perf$G1 == 10] = 'Pass'
stud_perf$G1[stud_perf$G1 == 11] = 'Pass'
stud_perf$G1[stud_perf$G1 == 12] = 'Pass'
stud_perf$G1[stud_perf$G1 == 13] = 'Pass'
stud_perf$G1[stud_perf$G1 == 14] = 'Pass'
stud_perf$G1[stud_perf$G1 == 15] = 'Pass'
stud_perf$G1[stud_perf$G1 == 16] = 'Pass'
stud_perf$G1[stud_perf$G1 == 17] = 'Pass'
stud_perf$G1[stud_perf$G1 == 18] = 'Pass'
stud_perf$G1[stud_perf$G1 == 19] = 'Pass'
stud_perf$G1[stud_perf$G1 == 20] = 'Pass'
```

#for G2

```
stud_perf$G2[stud_perf$G2 == 0] = 'Fail'
stud_perf$G2[stud_perf$G2 == 1] = 'Fail'
stud_perf$G2[stud_perf$G2 == 2] = 'Fail'
stud_perf$G2[stud_perf$G2 == 3] = 'Fail'
stud_perf$G2[stud_perf$G2 == 4] = 'Fail'
stud_perf$G2[stud_perf$G2 == 5] = 'Fail'
stud_perf$G2[stud_perf$G2 == 6] = 'Fail'
stud_perf$G2[stud_perf$G2 == 7] = 'Fail'
stud_perf$G2[stud_perf$G2 == 8] = 'Fail'
stud_perf$G2[stud_perf$G2 == 9] = 'Fail'
stud_perf$G2[stud_perf$G2 == 10] = 'Pass'
stud_perf$G2[stud_perf$G2 == 11] = 'Pass'
stud_perf$G2[stud_perf$G2 == 12] = 'Pass'
stud_perf$G2[stud_perf$G2 == 13] = 'Pass'
stud_perf$G2[stud_perf$G2 == 14] = 'Pass'
stud_perf$G2[stud_perf$G2 == 15] = 'Pass'
stud_perf$G2[stud_perf$G2 == 16] = 'Pass'
stud_perf$G2[stud_perf$G2 == 17] = 'Pass'
stud_perf$G2[stud_perf$G2 == 18] = 'Pass'
stud_perf$G2[stud_perf$G2 == 19] = 'Pass'
stud_perf$G2[stud_perf$G2 == 20] = 'Pass'
```


#for G3

```
stud_perf$G3[stud_perf$G3 == 0] = 'Fail'
stud_perf$G3[stud_perf$G3 == 1] = 'Fail'
stud_perf$G3[stud_perf$G3 == 2] = 'Fail'
stud_perf$G3[stud_perf$G3 == 3] = 'Fail'
stud_perf$G3[stud_perf$G3 == 4] = 'Fail'
stud_perf$G3[stud_perf$G3 == 5] = 'Fail'
stud_perf$G3[stud_perf$G3 == 6] = 'Fail'
stud_perf$G3[stud_perf$G3 == 7] = 'Fail'
stud_perf$G3[stud_perf$G3 == 8] = 'Fail'
stud_perf$G3[stud_perf$G3 == 9] = 'Fail'
stud_perf$G3[stud_perf$G3 == 10] = 'Pass'
stud_perf$G3[stud_perf$G3 == 11] = 'Pass'
stud_perf$G3[stud_perf$G3 == 12] = 'Pass'
stud_perf$G3[stud_perf$G3 == 13] = 'Pass'
stud_perf$G3[stud_perf$G3 == 14] = 'Pass'
stud_perf$G3[stud_perf$G3 == 15] = 'Pass'
stud_perf$G3[stud_perf$G3 == 16] = 'Pass'
stud_perf$G3[stud_perf$G3 == 17] = 'Pass'
stud_perf$G3[stud_perf$G3 == 18] = 'Pass'
stud_perf$G3[stud_perf$G3 == 19] = 'Pass'
stud_perf$G3[stud_perf$G3 == 20] = 'Pass'
```

```
head(stud_perf)
```

```
##  school sex age address famsize Pstatus Medu Fedu Mjob Fjob      reason
## 1     GP  F  18      U    GT3      A    4    4    1    4    course
## 2     GP  F  17      U    GT3      T    1    1    1    3    course
## 3     GP  F  15      U    LE3      T    1    1    1    3    other
## 4     GP  F  15      U    GT3      T    4    2    5    2    home
## 5     GP  F  16      U    GT3      T    3    3    3    3    home
## 6     GP  M  16      U    LE3      T    4    3    2    3 reputation
##  guardian traveltime studytime failures schoolsup famsup paid activities
## 1  mother           2           2           0           1           0           0
## 2  father           1           2           0           0           1           0
## 3  mother           1           2           3           1           0           1
## 4  mother           1           3           0           0           1           1
## 5  father           1           2           0           0           1           1
## 6  mother           1           2           0           0           1           1
##  nursery higher internet romantic famrel freetime goout Dalc Walc health
## 1      1      1      0      0      4      3      4      1      1      3
## 2      0      1      1      0      5      3      3      1      1      3
## 3      1      1      1      0      4      3      2      2      3      3
## 4      1      1      1      1      3      2      2      1      1      5
## 5      1      1      0      0      4      3      2      1      2      5
## 6      1      1      1      0      5      4      2      1      2      5
##  absences  G1  G2  G3
## 1      6 Fail Fail Fail
## 2      4 Fail Fail Fail
## 3     10 Fail Fail Pass
```

```
## 4      2 Pass Pass Pass
## 5      4 Fail Pass Pass
## 6     10 Pass Pass Pass
```

#changing specific columns to numeric

```
stud_perf$Mjob <- as.numeric(as.character(stud_perf$Mjob))
stud_perf$Fjob <- as.numeric(as.character(stud_perf$Fjob))
stud_perf$schoolsup <- as.numeric(as.character(stud_perf$schoolsup))
stud_perf$famsup <- as.numeric(as.character(stud_perf$famsup))
stud_perf$paid <- as.numeric(as.character(stud_perf$paid))
stud_perf$activities <- as.numeric(as.character(stud_perf$activities))
stud_perf$nursery <- as.numeric(as.character(stud_perf$nursery))
stud_perf$higher <- as.numeric(as.character(stud_perf$higher))
stud_perf$internet <- as.numeric(as.character(stud_perf$internet))
stud_perf$romantic <- as.numeric(as.character(stud_perf$romantic))
```

```
str(stud_perf)
```

```
## 'data.frame': 1044 obs. of 33 variables:
## $ school : chr "GP" "GP" "GP" "GP" ...
## $ sex : chr "F" "F" "F" "F" ...
## $ age : int 18 17 15 15 16 16 16 17 15 15 ...
## $ address : chr "U" "U" "U" "U" ...
## $ famsize : chr "GT3" "GT3" "LE3" "GT3" ...
## $ Pstatus : chr "A" "T" "T" "T" ...
## $ Medu : int 4 1 1 4 3 4 2 4 3 3 ...
## $ Fedu : int 4 1 1 2 3 3 2 4 2 4 ...
## $ Mjob : num 1 1 1 5 3 2 3 3 2 3 ...
## $ Fjob : num 4 3 3 2 3 3 3 4 3 3 ...
## $ reason : chr "course" "course" "other" "home" ...
## $ guardian : chr "mother" "father" "mother" "mother" ...
## $ traveltime: int 2 1 1 1 1 1 1 2 1 1 ...
## $ studytime : int 2 2 2 3 2 2 2 2 2 2 ...
## $ failures : int 0 0 3 0 0 0 0 0 0 0 ...
## $ schoolsup : num 1 0 1 0 0 0 0 1 0 0 ...
## $ famsup : num 0 1 0 1 1 1 0 1 1 1 ...
## $ paid : num 0 0 1 1 1 1 0 0 1 1 ...
## $ activities: num 0 0 0 1 0 1 0 0 0 1 ...
## $ nursery : num 1 0 1 1 1 1 1 1 1 1 ...
## $ higher : num 1 1 1 1 1 1 1 1 1 1 ...
## $ internet : num 0 1 1 1 0 1 1 0 1 1 ...
## $ romantic : num 0 0 0 1 0 0 0 0 0 0 ...
## $ famrel : int 4 5 4 3 4 5 4 4 4 5 ...
## $ freetime : int 3 3 3 2 3 4 4 1 2 5 ...
## $ goout : int 4 3 2 2 2 2 4 4 2 1 ...
## $ Dalc : int 1 1 2 1 1 1 1 1 1 1 ...
## $ Walc : int 1 1 3 1 2 2 1 1 1 1 ...
## $ health : int 3 3 3 5 5 5 3 1 1 5 ...
## $ absences : int 6 4 10 2 4 10 0 6 0 0 ...
## $ G1 : chr "Fail" "Fail" "Fail" "Pass" ...
```

```
## $ G2      : chr "Fail" "Fail" "Fail" "Pass" ...
## $ G3      : chr "Fail" "Fail" "Pass" "Pass" ...

#normalizing the numeric attributes
minmaxNorm <- function(x) {
  (x - min(x)) / (max(x) - min(x))
}

studperf_norm1 <- as.data.frame(lapply(stud_perf[7:10], minmaxNorm))
head(studperf_norm1)

## Medu Fedu Mjob Fjob
## 1 1.00 1.00 0.00 0.75
## 2 0.25 0.25 0.00 0.50
## 3 0.25 0.25 0.00 0.50
## 4 1.00 0.50 1.00 0.25
## 5 0.75 0.75 0.50 0.50
## 6 1.00 0.75 0.25 0.50

studperf_norm2 <- as.data.frame(lapply(stud_perf[13:30], minmaxNorm))
head(studperf_norm2)

## traveltime studytime failures schoolsup famsup paid activities nursery
## higher
## 1 0.3333333 0.3333333 0 1 0 0 0 1
## 1
## 2 0.0000000 0.3333333 0 0 1 0 0 0
## 1
## 3 0.0000000 0.3333333 1 1 0 1 0 1
## 1
## 4 0.0000000 0.6666667 0 0 1 1 1 1
## 1
## 5 0.0000000 0.3333333 0 0 1 1 0 1
## 1
## 6 0.0000000 0.3333333 0 0 1 1 1 1
## 1
## internet romantic famrel freetime goout Dalc Walc health absences
## 1 0 0 0.75 0.50 0.75 0.00 0.00 0.5 0.08000000
## 2 1 0 1.00 0.50 0.50 0.00 0.00 0.5 0.05333333
## 3 1 0 0.75 0.50 0.25 0.25 0.50 0.5 0.13333333
## 4 1 1 0.50 0.25 0.25 0.00 0.00 1.0 0.02666667
## 5 0 0 0.75 0.50 0.25 0.00 0.25 1.0 0.05333333
## 6 1 0 1.00 0.75 0.25 0.00 0.25 1.0 0.13333333

#merging the normalized data frames side by side
studperf_norm <- cbind(studperf_norm1, studperf_norm2)
head(studperf_norm)

## Medu Fedu Mjob Fjob traveltime studytime failures schoolsup famsup paid
## 1 1.00 1.00 0.00 0.75 0.3333333 0.3333333 0 1 0 0
## 2 0.25 0.25 0.00 0.50 0.0000000 0.3333333 0 0 1 0
```

```
## 3 0.25 0.25 0.00 0.50 0.0000000 0.3333333 1 1 0 1
## 4 1.00 0.50 1.00 0.25 0.0000000 0.6666667 0 0 1 1
## 5 0.75 0.75 0.50 0.50 0.0000000 0.3333333 0 0 1 1
## 6 1.00 0.75 0.25 0.50 0.0000000 0.3333333 0 0 1 1
## activities nursery higher internet romantic famrel freetime goout Dalc
Walc
## 1 0 1 1 0 0 0.75 0.50 0.75 0.00
0.00
## 2 0 0 1 1 0 1.00 0.50 0.50 0.00
0.00
## 3 0 1 1 1 0 0.75 0.50 0.25 0.25
0.50
## 4 1 1 1 1 1 0.50 0.25 0.25 0.00
0.00
## 5 0 1 1 0 0 0.75 0.50 0.25 0.00
0.25
## 6 1 1 1 1 0 1.00 0.75 0.25 0.00
0.25
## health absences
## 1 0.5 0.08000000
## 2 0.5 0.05333333
## 3 0.5 0.13333333
## 4 1.0 0.02666667
## 5 1.0 0.05333333
## 6 1.0 0.13333333
```

#creating the train and test sets

```
train_ind <- sample(1:nrow(stud_perf), 0.7 * nrow(stud_perf))
train.perf <- stud_perf[train_ind, ]
test.perf <- stud_perf[-train_ind, ]
```

```
head(train.perf)
```

```
## school sex age address famsize Pstatus Medu Fedu Mjob Fjob reason
## 136 GP F 15 U GT3 T 4 4 2 1 course
## 888 MS F 16 R GT3 A 2 2 5 3 course
## 3 GP F 15 U LE3 T 1 1 1 3 other
## 884 MS F 19 U GT3 T 1 1 1 2 course
## 380 MS F 17 R GT3 T 3 1 1 3 reputation
## 533 GP F 16 U GT3 A 2 2 3 3 home
## guardian traveltime studytime failures schoolsup famsup paid
activities
## 136 mother 1 3 0 0 1 0
1
## 888 mother 1 2 0 0 0 0
0
## 3 mother 1 2 3 1 0 1
0
## 884 mother 1 3 1 0 0 0
1
```

```

## 380  mother      1      2      0      0      1      1
1
## 533  mother      1      1      1      0      0      0
0
##      nursery higher internet romantic famrel freetime goout Dalc Walc
health
## 136      1      1      1      1      4      3      3      1      1
5
## 888      0      1      0      1      3      3      2      1      1
3
## 3      1      1      1      0      4      3      2      2      3
3
## 884      1      0      0      1      5      3      1      1      1
3
## 380      0      1      1      0      4      5      4      2      3
1
## 533      1      1      0      0      5      3      4      1      1
5
##      absences  G1  G2  G3
## 136      0 Pass Fail Fail
## 888      2 Fail Pass Pass
## 3      10 Fail Fail Pass
## 884      6 Fail Fail Fail
## 380      17 Pass Pass Pass
## 533      12 Pass Pass Pass

```

```
head(test.perf)
```

```

##      school sex age address famsize Pstatus Medu Fedu Mjob Fjob      reason
## 4      GP  F  15      U      GT3      T      4      2      5      2      home
## 12     GP  F  15      U      GT3      T      2      1      2      3 reputation
## 14     GP  M  15      U      GT3      T      4      3      4      3      course
## 15     GP  M  15      U      GT3      A      2      2      3      3      home
## 17     GP  F  16      U      GT3      T      4      4      2      2 reputation
## 18     GP  F  16      U      GT3      T      3      3      3      3 reputation
##      guardian traveltime studytime failures schoolsup famsup paid activities
## 4      mother      1      3      0      0      1      1      1
## 12     father      3      3      0      0      1      0      1
## 14     mother      2      2      0      0      1      1      0
## 15     other      1      3      0      0      1      0      0
## 17     mother      1      3      0      0      1      1      1
## 18     mother      3      2      0      1      1      0      1
##      nursery higher internet romantic famrel freetime goout Dalc Walc health
## 4      1      1      1      1      3      2      2      1      1      5
## 12     1      1      1      0      5      2      2      1      1      4
## 14     1      1      1      0      5      4      3      1      2      3
## 15     1      1      1      1      4      5      2      1      1      3
## 17     1      1      1      0      3      2      3      1      2      2
## 18     1      1      0      0      5      3      2      1      1      4
##      absences  G1  G2  G3

```

```
## 4      2 Pass Pass Pass
## 12     4 Pass Pass Pass
## 14     2 Pass Pass Pass
## 15     0 Pass Pass Pass
## 17     6 Pass Pass Pass
## 18     4 Fail Pass Pass
```

#creating the regression model

```
glm_model <-
glm(as.factor(G3)~Medu+Fedu+Mjob+Fjob+traveltime+studytime+failures+schoolsup
+famsup+paid+activities+nursery+higher+internet+romantic+famrel+freetime+goou
t+Dalc+Walc+health+absences, family = "binomial", data = train.perf)
```

```
summary(glm_model)
```

```
##
```

```
## Call:
```

```
## glm(formula = as.factor(G3) ~ Medu + Fedu + Mjob + Fjob + traveltime +
##      studytime + failures + schoolsup + famsup + paid + activities +
##      nursery + higher + internet + romantic + famrel + freetime +
##      goout + Dalc + Walc + health + absences, family = "binomial",
##      data = train.perf)
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.4025   0.2850   0.4888   0.6459   2.0259
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.583249   0.828760   1.910   0.0561 .
## Medu         0.167085   0.131077   1.275   0.2024
## Fedu        -0.010040   0.125115  -0.080   0.9360
## Mjob        -0.140075   0.098510  -1.422   0.1550
## Fjob         0.075935   0.126669   0.599   0.5489
## traveltime   0.008526   0.137288   0.062   0.9505
## studytime    0.063604   0.131067   0.485   0.6275
## failures    -1.022376   0.151596  -6.744 1.54e-11 ***
## schoolsup   -0.593333   0.294108  -2.017   0.0437 *
## famsup       0.036415   0.211855   0.172   0.8635
## paid        -0.521985   0.239145  -2.183   0.0291 *
## activities  -0.018107   0.201106  -0.090   0.9283
## nursery     -0.023617   0.250879  -0.094   0.9250
## higher       1.013577   0.329481   3.076   0.0021 **
## internet     0.276891   0.252936   1.095   0.2736
## romantic    -0.515460   0.204148  -2.525   0.0116 *
## famrel       0.115461   0.108562   1.064   0.2875
## freetime     0.049562   0.107503   0.461   0.6448
## goout       -0.426451   0.102462  -4.162 3.15e-05 ***
## Dalc        -0.048767   0.133205  -0.366   0.7143
## Walc         0.172679   0.108159   1.597   0.1104
```

```

## health      -0.158335   0.074217  -2.133   0.0329 *
## absences    -0.024439   0.014340  -1.704   0.0883 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 804.12  on 729  degrees of freedom
## Residual deviance: 661.01  on 707  degrees of freedom
## AIC: 707.01
##
## Number of Fisher Scoring iterations: 4

#confusion matrix for the regression model
predicted <- predict(glm_model, test.perf, type = "response")

predicted_class <- ifelse(predicted >= 0.5, 1, 0)
ConfusionMatrix <- table(actual = test.perf$G3, predicted = predicted_class)
ConfusionMatrix

##      predicted
## actual    0    1
## Fail    13   42
## Pass    14  245

#finding accuracy, precision, recall, sensitivity and specificity using the
confusion matrix
#accuracy
acc <- sum(diag(ConfusionMatrix))/nrow(test.perf)

#precision
prec <- ConfusionMatrix[2,2]/sum(ConfusionMatrix[2,2]+ConfusionMatrix[2,1])

#recall
recall <- ConfusionMatrix[2,2]/sum(ConfusionMatrix[2,2]+ConfusionMatrix[1,2])

#sensitivity
sens <- ConfusionMatrix[1,1]/sum(ConfusionMatrix[1,1]+ConfusionMatrix[2,1])

#specificity
spec <- ConfusionMatrix[2,2]/sum(ConfusionMatrix[1,2]+ConfusionMatrix[2,2])

acc
## [1] 0.8216561

prec
## [1] 0.9459459

recall

```

```
## [1] 0.8536585

sens

## [1] 0.4814815

spec

## [1] 0.8536585

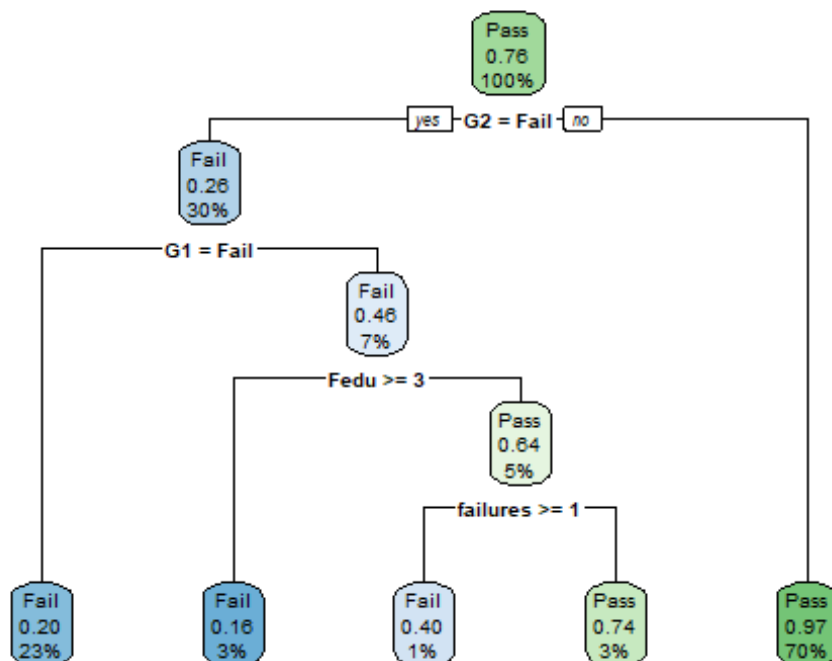
#finding f score
fscore <- (2*prec*recall)/(prec+recall)
fscore

## [1] 0.8974359

#creating a decision tree to predict G3
#install.packages("rpart.plot")
library(rpart)
library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 4.1.3

tree <- rpart(G3~., data = train.perf, method = 'class')
rpart.plot(tree)
```



```
#matrix for the decision tree
matrix_tree <- predict(tree, test.perf, type = 'class')
```



```

table_mat <- table(test.perf$G3, matrix_tree)
table_mat

##          matrix_tree
##          Fail Pass
##  Fail    47    8
##  Pass    15   244

#accuracy of the decision tree matrix
acc_tree <- sum(diag(table_mat))/sum(table_mat)

#precision
prec_tree <- table_mat[2,2]/sum(table_mat[2,2]+table_mat[2,1])

#recall
recall_tree <- table_mat[2,2]/sum(table_mat[2,2]+table_mat[1,2])

#sensitivity
sens_tree <- table_mat[1,1]/sum(table_mat[1,1]+table_mat[2,1])

#specificity
spec_tree <- table_mat[2,2]/sum(table_mat[1,2]+table_mat[2,2])

acc_tree
## [1] 0.9267516

prec_tree
## [1] 0.9420849

recall_tree
## [1] 0.968254

sens_tree
## [1] 0.7580645

spec_tree
## [1] 0.968254

```