SimpleStatisticalAnalysis

Xiaosong Zhang

9/18/2016

STATS250 KEY FACTOR ANALYSIS

Please answer the following question: Other than GPAO, which variable(s) best predict the variable for GRD_PTS_PER_UNIT for the course STATS 250 (SUBJECT="STATS" and CATALOG_NBR=250)? In other words, what is the best predictor of a student's performance in STATS 250 other than the student's own GPA?

Data Importation

```r
require("dplyr")
```

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

```r
require("ggplot2")
```

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

```r
#sourceDir is use to source all .r file under same Dir all in once
sourceDir <- function(path, trace = TRUE, ...) {
  for (nm in list.files(path, pattern = "[.][RrSsQq]$")) {
    if(trace) cat(nm,":")
    source(file.path(path, nm), ...)
    if(trace) cat("\n")
  }
}
#Source all functions uder PLA-MOOC into R-Environment for future use.
sourceDir('PLA-MOOC')
```

```
## course.impact.R :
## course.pathways.barplots.R :
## course.pathways.treemaps.R :
## course.persistence.module.R :
## grade.penalty.module.R :
```

```
#loading data into Environment
student.course <- read.csv("PLA-MOOC/student.course.csv")
student.record <- read.csv("PLA-MOOC/student.record.csv")
```

Data Exploration

```
summary(student.record)
```

```
##                    MAJOR3_DESCR                    MAJOR2_DESCR
##   Mathematics BS        :    40    Political Science BA:  1575
##   Economics BS          :    31    Psychology BA       :  1174
##   Political Science BA:    30      English BA          :   998
##   Spanish BA            :    28    Economics BA        :   940
##   Economics BA          :    25    Communication BA    :   701
##   (Other)               :   372    (Other)             : 10748
##   NA's                  :138362    NA's                :122752
##                         MAJOR1_DESCR        HSGPA          LAST_ACT_ENGL_SCORE
##   Psychology BA               : 6806    Min.   : 0.000   Min.   : 3.00
##   Business Administration BBA: 5921    1st Qu.: 3.300   1st Qu.:26.00
##   Economics BA                : 5327    Median : 3.700   Median :29.00
##   Political Science BA        : 5219    Mean   : 3.127   Mean   :28.89
##   English BA                  : 4871    3rd Qu.: 3.900   3rd Qu.:32.00
##   (Other)                     :74406    Max.   :36.000   Max.   :36.00
##   NA's                        :36338    NA's   :13666    NA's   :64846
##   LAST_ACT_MATH_SCORE LAST_ACT_READ_SCORE LAST_ACT_SCIRE_SCORE
##   Min.   : 4.00       Min.   : 5.00       Min.   : 8.00
##   1st Qu.:26.00       1st Qu.:26.00       1st Qu.:25.00
##   Median :29.00       Median :29.00       Median :27.00
##   Mean   :28.93       Mean   :28.75       Mean   :27.58
##   3rd Qu.:32.00       3rd Qu.:32.00       3rd Qu.:31.00
##   Max.   :36.00       Max.   :36.00       Max.   :36.00
##   NA's   :64841       NA's   :74168       NA's   :74523
##   LAST_ACT_COMP_SCORE LAST_SATI_VERB_SCORE LAST_SATI_MATH_SCORE
##   Min.   : 6.00       Min.   :200.0        Min.   :200.0
##   1st Qu.:27.00       1st Qu.:580.0        1st Qu.:620.0
##   Median :29.00       Median :630.0        Median :680.0
##   Mean   :28.73       Mean   :629.1        Mean   :671.8
##   3rd Qu.:31.00       3rd Qu.:680.0        3rd Qu.:730.0
##   Max.   :36.00       Max.   :800.0        Max.   :800.0
##   NA's   :64852       NA's   :99777        NA's   :99777
##   LAST_SATI_TOTAL_SCORE SEX        STDNT_GROUP1     STDNT_GROUP2
##   Min.   : 470          F:68889    A   :    36    G      : 12477
##   1st Qu.:1220          M:69999    B   :   255    E      :  1668
##   Median :1310                     C   :   785    D      :  1370
##   Mean   :1301                     D   :  1639    B      :   115
##   3rd Qu.:1390                     E   :  9025    F      :    69
##   Max.   :1600                     F   :  1370    (Other):    99
##   NA's   :99777                    NA's:125778    NA's   :123090
##                              MAJOR1_DEPT
##   Psychology Department          :10818
##   Economics Department           : 6243
```

2

```
##   School of Business Administration: 5921
##   Political Science Department    : 5297
##   Electrical Engr & Computer Sci  : 5231
##   (Other)                         :69015
##   NA's                            :36363
##                          MAJOR2_DEPT
##   Political Science Department    :  1635
##   Psychology Department           :  1467
##   Economics Department            :  1427
##   English Language & Literature Dept:  1074
##   Romance Languages Department    :   897
##   (Other)                         :  9634
##   NA's                            :122754
##                          MAJOR3_DEPT         ANONID
##   Economics Department            :   56  Min.   :     1
##   Mathematics Department          :   43  1st Qu.: 53412
##   Romance Languages Department    :   40  Median :106831
##   Political Science Department    :   32  Mean   :106966
##   School of Music, Theatre and Dance:   21  3rd Qu.:160508
##   (Other)                         :  334  Max.   :214170
##   NA's                            :138362
##     ADMIT_TERM      MAJOR1_TERM      MAJOR2_TERM       MAJOR3_TERM
##   Min.   : 53.00  Min.   :  1    Min.   :  6.0   Min.   : 39.0
##   1st Qu.: 78.00  1st Qu.: 84    1st Qu.: 88.0   1st Qu.: 94.0
##   Median :102.00  Median :103    Median :107.0   Median :111.0
##   Mean   : 97.62  Mean   :101    Mean   :103.9   Mean   :107.6
##   3rd Qu.:118.00  3rd Qu.:121    3rd Qu.:123.0   3rd Qu.:127.0
##   Max.   :138.00  Max.   :135    Max.   :135.0   Max.   :135.0
##   NA's   :13666   NA's   :36338  NA's   :122752  NA's   :138362
```

```r
summary(student.course)
```

```
##      ANONID           SUBJECT        CATALOG_NBR     GRD_PTS_PER_UNIT
##   Min.   :     1  CHEM   :184512  Min.   :100.0   Min.   :0.000
##   1st Qu.: 53301  MATH   :127551  1st Qu.:115.0   1st Qu.:2.700
##   Median :106789  PSYCH  :122574  Median :201.0   Median :3.300
##   Mean   :106901  PHYSICS:114254  Mean   :193.9   Mean   :3.063
##   3rd Qu.:160553  ENGLISH:108447  3rd Qu.:241.0   3rd Qu.:3.700
##   Max.   :214170  ECON   : 78307  Max.   :489.0   Max.   :9.000
##                   (Other):591420
##       GPAO          DIV         ANON_INSTR_ID       TERM
##   Min.   : 0.000  E : 95444  Min.   :   1   Min.   : 63.0
##   1st Qu.: 2.915  H :228824  1st Qu.:1229   1st Qu.: 84.0
##   Median : 3.289  O : 28531  Median :2730   Median :106.0
##   Mean   : 3.224  P : 78531  Mean   :2488   Mean   :103.1
##   3rd Qu.: 3.600  S :569072  3rd Qu.:3604   3rd Qu.:122.0
##   Max.   :19.700  SS:326663  Max.   :4805   Max.   :135.0
##
```

```r
# = 36? it most likely to be a error
# UM admit students with HSGPA 0?? or it should be NAs?
filter(student.record, HSGPA>4.0)
```

```
##   MAJOR3_DESCR       MAJOR2_DESCR                 MAJOR1_DESCR HSGPA
```

```
## 1          <NA> Movement Science BS Physical Education T.E. BS   32
## 2          <NA>             <NA>                        <NA>   36
##   LAST_ACT_ENGL_SCORE LAST_ACT_MATH_SCORE LAST_ACT_READ_SCORE
## 1                  NA                  NA                  NA
## 2                  NA                  NA                  NA
##   LAST_ACT_SCIRE_SCORE LAST_ACT_COMP_SCORE LAST_SATI_VERB_SCORE
## 1                   NA                  NA                   NA
## 2                   NA                  NA                   NA
##   LAST_SATI_MATH_SCORE LAST_SATI_TOTAL_SCORE SEX STDNT_GROUP1 STDNT_GROUP2
## 1                   NA                    NA   F        <NA>         <NA>
## 2                   NA                    NA   M        <NA>         <NA>
##            MAJOR1_DEPT         MAJOR2_DEPT MAJOR3_DEPT ANONID
## 1 School Of Kinesiology School Of Kinesiology        <NA>  80226
## 2             <NA>                <NA>        <NA> 144677
##   ADMIT_TERM MAJOR1_TERM MAJOR2_TERM MAJOR3_TERM
## 1         68          93          93          NA
## 2         68          NA          NA          NA
```

```
# many of those who have HSGPA = 0 have LAST_ACT_MATH_SCORE close to Max
# Cloud be a evidence their HSGPA should be NA instead of 0.

summary(filter(student.record, HSGPA< 1 & LAST_ACT_MATH_SCORE >35 ))
```

```
##                      MAJOR3_DESCR                      MAJOR2_DESCR
##  Aerospace Engineering  BSE   :  0   Mathematics BS            :  6
##  American Culture BA          :  0   Biology BS                :  1
##  Anthropology BA              :  0   Biopsych, Cognit & Neurosci BS:  1
##  Anthropology-Zoology BS      :  0   Chemistry BSChem          :  1
##  Arab,Arm,Pers,Turk & Islamic S:  0   Computer Science BS       :  1
##  (Other)                      :  0   (Other)                   :  6
##  NA's                         :118   NA's                      :102
##                   MAJOR1_DESCR     HSGPA    LAST_ACT_ENGL_SCORE
##  Economics BA             :  6   Min.   :0   Min.   :21.00
##  Mathematics BS           :  5   1st Qu.:0   1st Qu.:29.00
##  Movement Science BS      :  4   Median :0   Median :33.00
##  Economics BS             :  3   Mean   :0   Mean   :31.81
##  Aerospace Engineering  BSE:  2   3rd Qu.:0   3rd Qu.:35.00
##  (Other)                  : 39   Max.   :0   Max.   :36.00
##  NA's                     : 59
##  LAST_ACT_MATH_SCORE LAST_ACT_READ_SCORE LAST_ACT_SCIRE_SCORE
##  Min.   :36          Min.   :19.00       Min.   :21.00
##  1st Qu.:36          1st Qu.:30.00       1st Qu.:31.00
##  Median :36          Median :32.00       Median :33.50
##  Mean   :36          Mean   :31.73       Mean   :32.52
##  3rd Qu.:36          3rd Qu.:34.00       3rd Qu.:35.00
##  Max.   :36          Max.   :36.00       Max.   :36.00
##                      NA's   :22          NA's   :22
##  LAST_ACT_COMP_SCORE LAST_SATI_VERB_SCORE LAST_SATI_MATH_SCORE
##  Min.   :25.00       Min.   :470.0        Min.   :620.0
##  1st Qu.:32.00       1st Qu.:630.0        1st Qu.:730.0
##  Median :33.00       Median :685.0        Median :770.0
##  Mean   :32.78       Mean   :675.3        Mean   :760.1
##  3rd Qu.:35.00       3rd Qu.:740.0        3rd Qu.:800.0
##  Max.   :36.00       Max.   :800.0        Max.   :800.0
```

```
##                     NA's   :46           NA's   :46
## LAST_SATI_TOTAL_SCORE SEX   STDNT_GROUP1 STDNT_GROUP2
## Min.   :1180         F:24   A  : 0       G     :24
## 1st Qu.:1380         M:94   B  : 0       A     : 0
## Median :1440                C  : 1       B     : 0
## Mean   :1435                D  : 0       C     : 0
## 3rd Qu.:1520                E  : 0       D     : 0
## Max.   :1590                F  : 0       (Other): 0
## NA's   :46                  NA's:117     NA's  :94
##              MAJOR1_DEPT                    MAJOR2_DEPT
## Economics Department  : 9   Mathematics Department    : 6
## School Of Kinesiology : 6   Biology Department        : 1
## Mathematics Department: 5   Chemistry Department      : 1
## Psychology Department : 5   Ecology & Evolutionary Biology: 1
## Physics Department    : 4   Economics Department      : 1
## (Other)               :30   (Other)                   : 6
## NA's                  :59   NA's                      :102
##                 MAJOR3_DEPT      ANONID        ADMIT_TERM
##                         : 0   Min.   :   317   Min.   : 78.0
## Aerospace Engineering   : 0   1st Qu.: 57943   1st Qu.:114.0
## American Culture Program: 0   Median :121338   Median :123.0
## Anthropology Department : 0   Mean   :113433   Mean   :121.4
## Asian Languages And Cultures: 0   3rd Qu.:173832   3rd Qu.:130.0
## (Other)                 : 0   Max.   :213172   Max.   :138.0
## NA's                    :118
##  MAJOR1_TERM    MAJOR2_TERM     MAJOR3_TERM
## Min.   :102.0  Min.   :114.0   Min.   : NA
## 1st Qu.:119.0  1st Qu.:119.0   1st Qu.: NA
## Median :127.0  Median :127.0   Median : NA
## Mean   :124.3  Mean   :125.8   Mean   :NaN
## 3rd Qu.:131.0  3rd Qu.:131.0   3rd Qu.: NA
## Max.   :135.0  Max.   :135.0   Max.   : NA
## NA's   :59     NA's   :102     NA's   :118
```