Market basket analysis

Sugarbayar Enkhbayar

2/10/2023

Introduction

I used market basket analysis with student performance data from 2 schools in Portugal to find out what factors influence whether students pass or not final exam. In other words, I selected student pass or fail data as rhs. And I selected student other profile information as lhs.

Packages

```
library(readr)
library(tidyverse)
library(arules)
library(arulesViz)
```

Data preparetion

Firstly, I imported data from machine learning repository website. I removed X rownumber variable from data. Then, I added column names to values to make result of association rule easier to interpret.

```
df=read.csv("df.csv")
df=df %>% select(-X)
data=df
for (i in 1:31) {
  x=colnames(df)[i]
  df[,i]<-paste0(x,'-',df[,i])
}
head(df)
```

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

Exploratory Data Analysis

Variable identification | Missing values treatment | Central Tendency | Visualization of EDA

Age, absences variables has many unique values and It is hard to explain. So it was removed from original data. Then, df1 is our final dataset to use in market basket analysis. We have 28 variables as lefthandside, and 1 variable as righthandside.

Left hand sides:

- 1.school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
- 2.sex - student's sex (binary: 'F' - female or 'M' - male)
- 3.address - student's home address type (binary: 'U' - urban or 'R' - rural)
- 4.famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
- 5.Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
- 6.Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- 7.Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- 8.Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
- 9.Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
- 10.reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
- 11.guardian - student's guardian (nominal: 'mother', 'father' or 'other')
- 12.traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
- 13.studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
- 14.failures - number of past class failures (numeric: n if 1<=n<3, else 4)
- 15.schoolsup - extra educational support (binary: yes or no)
- 16.famsup - family educational support (binary: yes or no)
- 17.paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
- 18.activities - extra-curricular activities (binary: yes or no)
- 19.nursery - attended nursery school (binary: yes or no)
- 20.higher - wants to take higher education (binary: yes or no)
- 21.internet - Internet access at home (binary: yes or no)
- 22.romantic - with a romantic relationship (binary: yes or no)
- 23.famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
- 24.freetime - free time after school (numeric: from 1 - very low to 5 - very high)
- 25.goout - going out with friends (numeric: from 1 - very low to 5 - very high)
- 26.Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
- 27.Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
- 28.health - current health status (numeric: from 1 - very bad to 5 - very good)

Right hand sides:

- 1.final - (success, fail)

```
df<-df %>% select(-age,-absences)
write.csv(df,"df1.csv",row.names = F)
colnames(df)
```

```
##  [1] "school"     "sex"        "address"    "famsize"    "Pstatus"
##  [6] "Medu"       "Fedu"       "Mjob"       "Fjob"       "reason"
## [11] "guardian"   "traveltime" "studytime"  "failures"   "schoolsup"
## [16] "famsup"     "paid"       "activities" "nursery"    "higher"
## [21] "internet"   "romantic"   "famrel"     "freetime"   "goout"
## [26] "Dalc"       "Walc"       "health"     "final"
```

Manually categorised

In this section, we do market basket analysis with manually categorised data. Data has 650 itemsets and 126 items. Top 5 most frequent items are paid-no,schoolsup-no, higher-yes, Pstatus-T,failures-0.

- Paid-no: no extra paid classes with in the course subject.
- Schoolsup-no: no extra educational support
- Higher-yes: wants to take higher education
- Pstatus-T: living together with parents
- Failures-0: most students didn't fail any lesson

I removed rare observations that has higher itemFrequency than 0.05. After it, I have 87 items. Let's see most high itemFrequency 15 itemsets as barchart.

```
trans=read.transactions("df1.csv",format='basket',sep=",",skip=0)
summary(trans)
```

```
## transactions as itemMatrix in sparse format with
##  650 rows (elements/itemsets/transactions) and
##  126 columns (items) and a density of 0.2301587
##
## most frequent items:
##      paid-no schoolsup-no   higher-yes    Pstatus-T   failures-0      (Other)
##          610          581          580          569          549        15961
##
## element (itemset/transaction) length distribution:
## sizes
##  29
## 650
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      29      29      29      29      29      29
##
## includes extended item information - examples:
##           labels
## 1     activities
## 2  activities-no
## 3 activities-yes
```

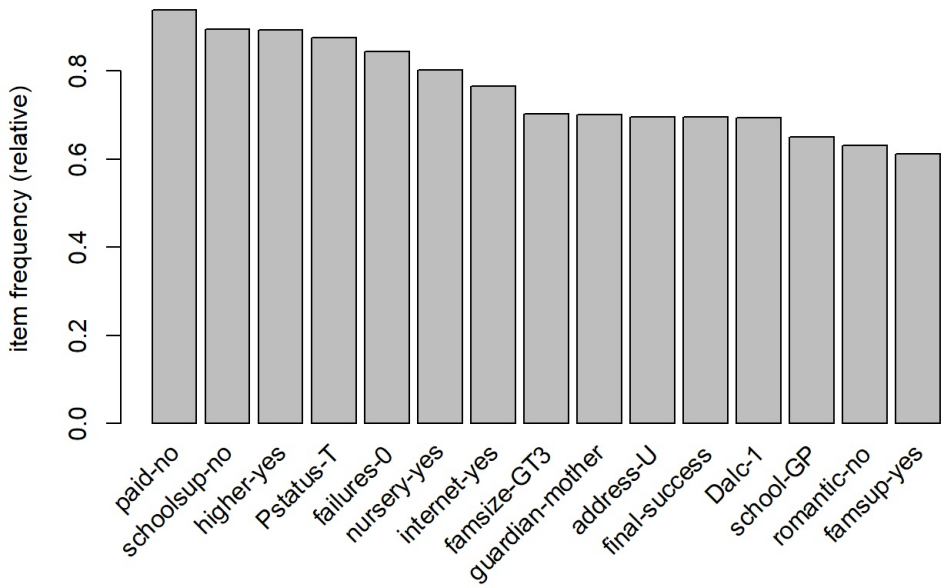
```
# cleaning the data from rare observations
trans1=trans[,itemFrequency(trans)>0.05]
trans1
```

```
## transactions in sparse format with
##  650 transactions (rows) and
##  87 items (columns)
```

```
# we can get all levels in dataset and their frequency
sort(itemFrequency(trans1, type="relative"),decreasing = T)
```

```
##           paid-no      schoolsup-no        higher-yes         Pstatus-T
##        0.93846154        0.89384615        0.89230769        0.87538462
##        failures-0       nursery-yes      internet-yes       famsize-GT3
##        0.84461538        0.80153846        0.76615385        0.70307692
##   guardian-mother         address-U     final-success            Dalc-1
##        0.70000000        0.69538462        0.69538462        0.69384615
##         school-GP       romantic-no        famsup-yes             sex-F
##        0.65076923        0.63076923        0.61230769        0.58923077
##        Fjob-other      traveltime-1     activities-no          famrel-4
##        0.56461538        0.56307692        0.51384615        0.48769231
##    activities-yes       studytime-2     reason-course             sex-M
##        0.48461538        0.46923077        0.43846154        0.40923077
##        Mjob-other         famsup-no        freetime-3          health-5
##        0.39692308        0.38615385        0.38615385        0.38307692
##            Walc-1      romantic-yes         school-MS     traveltime-2
##        0.38000000        0.36769231        0.34769231        0.32769231
##       studytime-1            Fedu-2           goout-3         address-R
##        0.32615385        0.32153846        0.31538462        0.30307692
##        final-fail       famsize-LE3            Medu-2     Fjob-services
##        0.30307692        0.29538462        0.28615385        0.27846154
##          famrel-5        freetime-4            Medu-4            Fedu-1
##        0.27692308        0.27384615        0.26923077        0.26769231
##   guardian-father       internet-no            Walc-2       reason-home
##        0.23538462        0.23230769        0.23076923        0.22923077
##           goout-2            Medu-1 reason-reputation           goout-4
##        0.22307692        0.22000000        0.22000000        0.21692308
##            Medu-3     Mjob-services      Mjob-at_home            Fedu-3
##        0.21384615        0.20923077        0.20769231        0.20153846
##            Fedu-4        nursery-no          health-3            Dalc-2
##        0.19692308        0.19692308        0.19076923        0.18615385
##            Walc-3           goout-5          health-4        freetime-2
##        0.18461538        0.16923077        0.16615385        0.16461538
##          famrel-3       studytime-3          health-1            Walc-4
##        0.15538462        0.14923077        0.13846154        0.13384615
##         Pstatus-A          health-2      Mjob-teacher      reason-other
##        0.12307692        0.12000000        0.11076923        0.11076923
##        failures-1         higher-no        freetime-5     schoolsup-yes
##        0.10769231        0.10615385        0.10461538        0.10461538
##      traveltime-3           goout-1       Mjob-health        freetime-1
##        0.08307692        0.07384615        0.07384615        0.06923077
##            Walc-5            Dalc-3      Fjob-at_home     guardian-other
##        0.06923077        0.06615385        0.06461538        0.06307692
##          paid-yes      Fjob-teacher       studytime-4
##        0.06000000        0.05538462        0.05384615
```

```
itemFrequencyPlot(trans1,type='relative',topN=15)
```



RHS - student who passed the final exam | RHS - student who failed the final exam

In order to see profile of students who are passed final exam, we have to use apriori method. I choose minimum support level 0.1, and minimum confidence level 0.5. Also, righthandside is 'final-success'. After training model, I sorted result of model by decreasing lift.

- Model has 116034 rules
- rule length distribution is up to 10
- Model's support is 0.1 and confidence is 0.5
- We can see top 15 result of our model. let's explain first one. If we see somebody who successfully passed final exam, she lives in urban,she didn't fail any exams, no romantic relationship, study in GP school, no extra educational support, and she is female. support is 0.13. And this case's frequency of features combination in dataset is 13 percentage. Confidence is 1. It means that for sure (100%). Also, lift is 1.43 and it is higher than 1. It means that we see features appearing together more often than separately.
- Consider the second lhs. If we see somebody who successfully passed final exam, she is female, she has no extra educational support, she study in GP school, she wants to take higher education, her family support her, she didn't fail any course, she has no extra-curricular activities.

```
rules<-apriori(data=trans1, parameter=list(supp=0.1,conf = 0.5),
                      appearance=list(default="lhs", rhs='final-success'), control=list(verbose=F))
rules.clean<-rules
rules.final<-sort(rules.clean, by="lift", decreasing=TRUE)
summary(rules.final)
```

```
## set of 116034 rules
##
## rule length distribution (lhs + rhs):sizes
##     1     2     3     4     5     6     7     8     9    10
##     1    64   842  4486 13303 24825 30556 24688 12990  4279
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   6.000   7.000   6.956   8.000  10.000
##
## summary of quality measures:
##     support         confidence        coverage           lift
##  Min.   :0.1000   Min.   :0.5000   Min.   :0.1000   Min.   :0.719
##  1st Qu.:0.1077   1st Qu.:0.8068   1st Qu.:0.1277   1st Qu.:1.160
##  Median :0.1215   Median :0.8523   Median :0.1462   Median :1.226
##  Mean   :0.1360   Mean   :0.8440   Mean   :0.1625   Mean   :1.214
##  3rd Qu.:0.1477   3rd Qu.:0.8902   3rd Qu.:0.1769   3rd Qu.:1.280
##  Max.   :0.6954   Max.   :1.0000   Max.   :1.0000   Max.   :1.438
##      count
##  Min.   : 65.00
##  1st Qu.: 70.00
##  Median : 79.00
##  Mean   : 88.43
##  3rd Qu.: 96.00
##  Max.   :452.00
##
## mining info:
##    data ntransactions support confidence
##  trans1           650     0.1        0.5
##
call
##  apriori(data = trans1, parameter = list(supp = 0.1, conf = 0.5), appearance = list(default = "lhs", rhs = "fi
nal-success"), control = list(verbose = F))
```

```
rules.final<-sort(rules.final, by="lift", decreasing=TRUE)
inspect(head(rules.final,15))
```

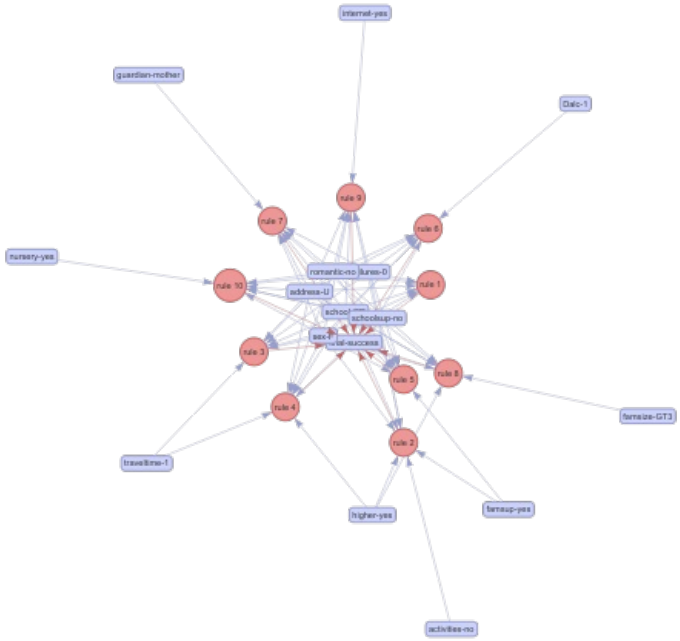
```
##      lhs                   rhs               support confidence  coverage     lift count
## [1]  {address-U,
##       failures-0,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1323077          1 0.1323077 1.438053    86
## [2]  {activities-no,
##       failures-0,
##       famsup-yes,
##       higher-yes,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1046154          1 0.1046154 1.438053    68
## [3]  {address-U,
##       failures-0,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
```

```
##       sex-F,
##       traveltime-1}     => {final-success} 0.1015385          1 0.1015385 1.438053    66
## [4]  {address-U,
##       higher-yes,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F,
##       traveltime-1}     => {final-success} 0.1046154          1 0.1046154 1.438053    68
## [5]  {address-U,
##       failures-0,
##       famsup-yes,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1000000          1 0.1000000 1.438053    65
## [6]  {address-U,
##       Dalc-1,
##       failures-0,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1092308          1 0.1092308 1.438053    71
## [7]  {address-U,
##       failures-0,
##       guardian-mother,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1015385          1 0.1015385 1.438053    66
## [8]  {address-U,
##       famsize-GT3,
##       higher-yes,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1015385          1 0.1015385 1.438053    66
## [9]  {address-U,
##       failures-0,
##       internet-yes,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1123077          1 0.1123077 1.438053    73
## [10] {address-U,
##       failures-0,
##       nursery-yes,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1061538          1 0.1061538 1.438053    69
## [11] {address-U,
##       failures-0,
##       Pstatus-T,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1200000          1 0.1200000 1.438053    78
## [12] {address-U,
##       failures-0,
##       higher-yes,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1307692          1 0.1307692 1.438053    85
## [13] {address-U,
##       failures-0,
##       paid-no,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1292308          1 0.1292308 1.438053    84
## [14] {address-U,
##       higher-yes,
##       Pstatus-T,
##       romantic-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1230769          1 0.1230769 1.438053    80
```

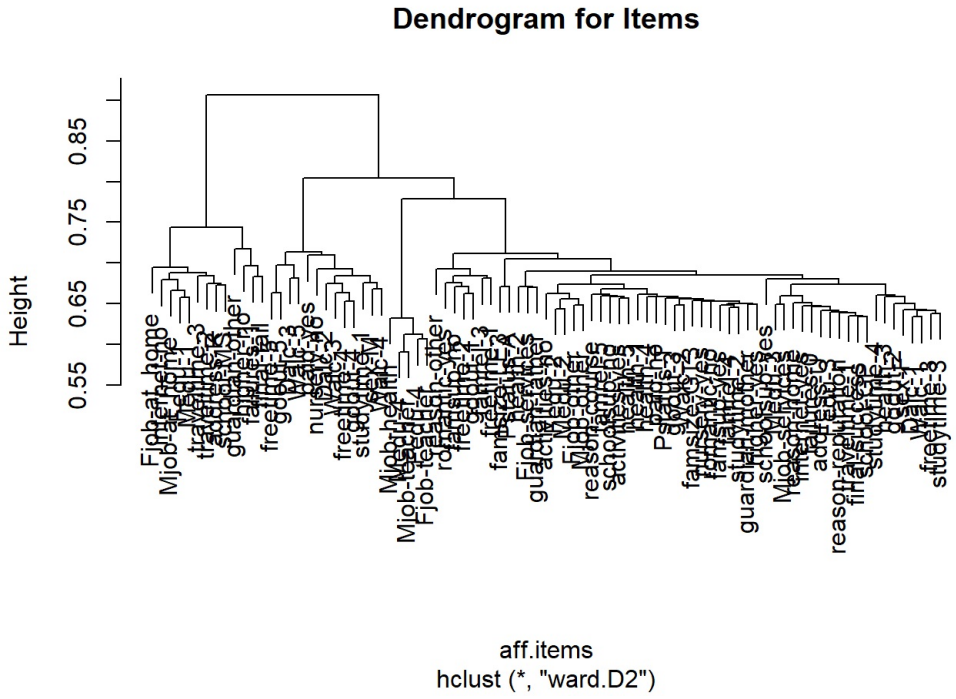
```
## [15] {activities-no,
##       failures-0,
##       famsup-yes,
##       higher-yes,
##       paid-no,
##       school-GP,
##       schoolsup-no,
##       sex-F}            => {final-success} 0.1000000          1 0.1000000 1.438053    65
```

```
plot(head(rules.final,10), method="graph", engine="htmlwidget") #
```

Select by id




```
aff.items<-dissimilarity(trans1, which="items", method="affinity")
hc<-hclust(aff.items, method="ward.D2")
plot(hc, main="Dendrogram for Items")
```



Automatically categorised

In order to do market basket analysis with automatically categorised data, we need to convert all datas to factor.

RHS - student who passed the final exam | RHS - student who failed the final exam

We have 452 dataset and we can see summary of them. I removed rare observations that has higher itemFrequency than 0.05. I choose minimum support level 0.1, and minimum confidence level 0.5.

- We have 36813 rules
- rule length distributions are from 1 to 5
- 2nd lhs: If we see who successfully passed final exam, who wants to take higher education
- 3rd lhs: If we see who successfully passed final exam, who didn't fail any exams
- 5th lhs: If we see who successfully passed final exam, who wants to take higher education and who didn't fail any exam.

```
library(arulesCBA)
trans2=df
for (i in 1:29) {
  trans2[,i]<-factor(trans2[,i],levels = unique(trans2[,i]))
}
trans21=trans2 %>% filter(final=='final-success')
data.disc=discretizeDF.supervised(data=trans21,final~.,methods = 'chi2')
summary(data.disc)
```

```
##        school       sex           address           famsize         Pstatus
##  school-GP:338   sex-F:281   address-U:335   famsize-GT3:317   Pstatus-A: 56
##  school-MS:114   sex-M:171   address-R:117   famsize-LE3:135   Pstatus-T:396
##
##
##
##      Medu         Fedu               Mjob               Fjob
##  Medu-4:151   Fedu-4:107   Mjob-at_home : 78   Fjob-teacher : 33
##  Medu-1: 75   Fedu-1: 95   Mjob-health  : 36   Fjob-other   :261
##  Medu-3: 94   Fedu-2:142   Mjob-other   :176   Fjob-services:117
##  Medu-2:127   Fedu-3:103   Mjob-services: 98   Fjob-health  : 16
##  Medu-0:  5   Fedu-0:  5   Mjob-teacher : 64   Fjob-at_home : 25
##                reason              guardian          traveltime
##  reason-course    :176   guardian-mother:310   traveltime-2:137
##  reason-other     : 44   guardian-father:118   traveltime-1:273
##  reason-home      :116   guardian-other : 24   traveltime-3: 33
##  reason-reputation:116                         traveltime-4:  9
##
##       studytime         failures           schoolsup          famsup
##  studytime-2:221   failures-0:433   schoolsup-yes: 45   famsup-no :169
##  studytime-3: 82   failures-3:  1   schoolsup-no :407   famsup-yes:283
##  studytime-1:121   failures-1: 15
##  studytime-4: 28   failures-2:  3
##
##       paid              activities         nursery           higher
##  paid-no :425   activities-no :224   nursery-yes:365   higher-yes:435
##  paid-yes: 27   activities-yes:228   nursery-no : 87   higher-no : 17
##
##
##
##         internet           romantic        famrel          freetime
##  internet-no : 90   romantic-no :300   famrel-4:234   freetime-3:184
##  internet-yes:362   romantic-yes:152   famrel-5:120   freetime-2: 87
##                                        famrel-3: 68   freetime-4:115
##                                        famrel-1: 11   freetime-1: 29
##                                        famrel-2: 19   freetime-5: 37
##      goout         Dalc         Walc          health              final
##  goout-4: 99   Dalc-1:334   Walc-1:183   health-3: 88   final-success:452
##  goout-3:150   Dalc-2: 77   Walc-3: 83   health-5:165   final-fail   :  0
##  goout-2:110   Dalc-5:  8   Walc-2:110   health-1: 65
##  goout-1: 29   Dalc-3: 22   Walc-4: 51   health-2: 55
##  goout-5: 64   Dalc-4: 11   Walc-5: 25   health-4: 79
```

```
data.trans<-transactions(data.disc)
trans2<-data.trans[, itemFrequency(data.trans)>0.05]
data.ass<-mineCARs(final~ ., transactions=trans2, support=0.1, confidence=0.5)
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE           FALSE       5     0.1      1
##  maxlen target  ext
##       5  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 45
##
## set item appearances ...[83 item(s)] done [0.00s].
## set transactions ...[83 item(s), 452 transaction(s)] done [0.00s].
## sorting and recoding items ... [83 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 done [0.22s].
## writing ... [36813 rule(s)] done [0.02s].
## creating S4 object  ... done [0.02s].
```

```
data.ass=sort(data.ass,by='support',decreasing = TRUE)
summary(data.ass)
```

```
## set of 36813 rules
##
## rule length distribution (lhs + rhs):sizes
##     1     2     3     4     5
##     1    69  1171  7845 27727
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   5.000   5.000   4.718   5.000   5.000
##
## summary of quality measures:
##     support         confidence    coverage           lift       count
##  Min.   :0.1018   Min.   :1    Min.   :0.1018   Min.   :1   Min.   : 46.00
##  1st Qu.:0.1173   1st Qu.:1    1st Qu.:0.1173   1st Qu.:1   1st Qu.: 53.00
##  Median :0.1438   Median :1    Median :0.1438   Median :1   Median : 65.00
##  Mean   :0.1756   Mean   :1    Mean   :0.1756   Mean   :1   Mean   : 79.35
##  3rd Qu.:0.1969   3rd Qu.:1    3rd Qu.:0.1969   3rd Qu.:1   3rd Qu.: 89.00
##  Max.   :1.0000   Max.   :1    Max.   :1.0000   Max.   :1   Max.   :452.00
##
## mining info:
##          data ntransactions support confidence
##  transactions           452     0.1        0.5
##
call
##  apriori(data = transactions, parameter = parameter, appearance = list(rhs = vars$class_items, lhs = vars$feat
ure_items), control = control)
```

```
inspect(head(data.ass,20))
```

```
##      lhs                           rhs                     support confidence  coverage lift count
## [1]  {}                         => {final=final-success} 1.0000000          1 1.0000000    1   452
## [2]  {higher=higher-yes}        => {final=final-success} 0.9623894          1 0.9623894    1   435
## [3]  {failures=failures-0}      => {final=final-success} 0.9579646          1 0.9579646    1   433
## [4]  {paid=paid-no}             => {final=final-success} 0.9402655          1 0.9402655    1   425
## [5]  {failures=failures-0,
##       higher=higher-yes}        => {final=final-success} 0.9269912          1 0.9269912    1   419
## [6]  {failures=failures-0,
##       paid=paid-no}             => {final=final-success} 0.9048673          1 0.9048673    1   409
## [7]  {paid=paid-no,
##       higher=higher-yes}        => {final=final-success} 0.9048673          1 0.9048673    1   409
## [8]  {schoolsup=schoolsup-no}   => {final=final-success} 0.9004425          1 0.9004425    1   407
## [9]  {Pstatus=Pstatus-T}        => {final=final-success} 0.8761062          1 0.8761062    1   396
## [10] {failures=failures-0,
##       paid=paid-no,
##       higher=higher-yes}        => {final=final-success} 0.8761062          1 0.8761062    1   396
## [11] {failures=failures-0,
##       schoolsup=schoolsup-no}   => {final=final-success} 0.8650442          1 0.8650442    1   391
## [12] {schoolsup=schoolsup-no,
##       higher=higher-yes}        => {final=final-success} 0.8628319          1 0.8628319    1   390
## [13] {schoolsup=schoolsup-no,
##       paid=paid-no}             => {final=final-success} 0.8517699          1 0.8517699    1   385
## [14] {Pstatus=Pstatus-T,
##       higher=higher-yes}        => {final=final-success} 0.8429204          1 0.8429204    1   381
## [15] {Pstatus=Pstatus-T,
##       failures=failures-0}      => {final=final-success} 0.8407080          1 0.8407080    1   380
## [16] {failures=failures-0,
##       schoolsup=schoolsup-no,
##       higher=higher-yes}        => {final=final-success} 0.8340708          1 0.8340708    1   377
## [17] {Pstatus=Pstatus-T,
##       paid=paid-no}             => {final=final-success} 0.8230088          1 0.8230088    1   372
## [18] {failures=failures-0,
##       schoolsup=schoolsup-no,
##       paid=paid-no}             => {final=final-success} 0.8230088          1 0.8230088    1   372
## [19] {schoolsup=schoolsup-no,
##       paid=paid-no,
##       higher=higher-yes}        => {final=final-success} 0.8163717          1 0.8163717    1   369
## [20] {Pstatus=Pstatus-T,
##       failures=failures-0,
##       higher=higher-yes}        => {final=final-success} 0.8119469          1 0.8119469    1   367
```

Conclusion

Student with the following characteristics has successfully passed the final exam.

- female
- lives in urban
- no romantic relationship
- study in GP school
- wants to take higher education
- her family support him
- didn’t fail any exams

However, Students with the following characteristics failed to pass the final exam.

- study in MS school* no extra paid classes with in the course subject
- no extra educational support