IST 707 – Applied Machine Learning
Professor Wallace
Spring 22

Assignment 3

Dixitha Kasturi
dkasturi@syr.edu

Topic: Association rules mining.
Dataset: Bank transactions
Outcome: PEP analysis

Exploratory Data analysis:

```
str(bank)
```

```
## spec_tbl_df [600 x 12] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ id         : chr [1:600] "ID12101" "ID12102" "ID12103" "ID12104" ...
##  $ age        : num [1:600] 48 40 51 23 57 57 22 58 37 54 ...
##  $ sex        : chr [1:600] "FEMALE" "MALE" "FEMALE" "FEMALE" ...
##  $ region     : chr [1:600] "INNER_CITY" "TOWN" "INNER_CITY" "TOWN" ...
##  $ income     : num [1:600] 17546 30085 16575 20375 50576 ...
##  $ married    : chr [1:600] "NO" "YES" "YES" "YES" ...
##  $ children   : num [1:600] 1 3 0 3 0 2 0 0 2 2 ...
##  $ car        : chr [1:600] "NO" "YES" "YES" "NO" ...
##  $ save_act   : chr [1:600] "NO" "NO" "YES" "NO" ...
##  $ current_act: chr [1:600] "NO" "YES" "YES" "YES" ...
##  $ mortgage   : chr [1:600] "NO" "YES" "NO" "NO" ...
##  $ pep        : chr [1:600] "YES" "NO" "NO" "NO" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   id = col_character(),
##   ..   age = col_double(),
##   ..   sex = col_character(),
##   ..   region = col_character(),
##   ..   income = col_double(),
##   ..   married = col_character(),
##   ..   children = col_double(),
##   ..   car = col_character(),
##   ..   save_act = col_character(),
##   ..   current_act = col_character(),
##   ..   mortgage = col_character(),
##   ..   pep = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
describe(bank)
```

```
## bank
##
##  12  Variables      600  Observations
## --------------------------------------------------------------------------------
## id
##        n  missing distinct
##      600        0      600
##
## lowest : ID12101 ID12102 ID12103 ID12104 ID12105
## highest: ID12696 ID12697 ID12698 ID12699 ID12700
## --------------------------------------------------------------------------------
## age
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      600        0       50    0.999     42.4    16.65    20.00    22.90
##      .25      .50      .75      .90      .95
```

```
##    30.00    42.00    55.25    63.00    65.00
##
## lowest : 18 19 20 21 22, highest: 63 64 65 66 67
## --------------------------------------------------------------------------------
## sex
##        n  missing distinct
##      600        0        2
##
## Value      FEMALE   MALE
## Frequency     300    300
## Proportion    0.5    0.5
## --------------------------------------------------------------------------------
## region
##        n  missing distinct
##      600        0        4
##
## Value      INNER_CITY      RURAL   SUBURBAN       TOWN
## Frequency         269         96         62        173
## Proportion      0.448      0.160      0.103      0.288
## --------------------------------------------------------------------------------
## income
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      600        0      599        1    27524    14482    10620    12644
##      .25      .50      .75      .90      .95
##    17265    24925    36173    47254    52663
##
## lowest :  5014.21  6294.21  7304.20  7549.38  7606.25
## highest: 59803.90 59805.60 60747.50 61554.60 63130.10
## --------------------------------------------------------------------------------
## married
##        n  missing distinct
##      600        0        2
##
## Value        NO  YES
## Frequency   204  396
## Proportion 0.34 0.66
## --------------------------------------------------------------------------------
## children
##        n  missing distinct     Info     Mean      Gmd
##      600        0        4    0.892    1.012    1.142
##
## Value          0     1     2     3
## Frequency    263   135   134    68
## Proportion 0.438 0.225 0.223 0.113
## --------------------------------------------------------------------------------
## car
##        n  missing distinct
##      600        0        2
##
## Value         NO   YES
## Frequency    304   296
## Proportion 0.507 0.493
## --------------------------------------------------------------------------------
## save_act
```

```
##        n  missing distinct
##      600        0        2
##
## Value        NO  YES
## Frequency   186  414
## Proportion 0.31 0.69
## ---------------------------------------------------------------------------------
## current_act
##        n  missing distinct
##      600        0        2
##
## Value         NO   YES
## Frequency    145   455
## Proportion 0.242 0.758
## ---------------------------------------------------------------------------------
## mortgage
##        n  missing distinct
##      600        0        2
##
## Value         NO   YES
## Frequency    391   209
## Proportion 0.652 0.348
## ---------------------------------------------------------------------------------
## pep
##        n  missing distinct
##      600        0        2
##
## Value         NO   YES
## Frequency    326   274
## Proportion 0.543 0.457
## ---------------------------------------------------------------------------------
```

Despite the formatting not being the way we would want it to for association rules, there are no missing values in any of the columns. So there is no data cleaning. We do the following transformations:

a) Each ID is distinct, so removing it would make best sense.
b) Age and income are continuous variables, if stored precisely. Here despite having discrete values for age we still consider it continuous, we further want to put the values into bins or properly discretize them for easy understanding of the age and income levels and rules.
   Based on the age range -> kids, teens, twenties etc were discretized
   Income was discretised based on the maximum and minimum values as these can change with new customers, 5 levels were decided based on the levels
c) sex(2),region(4),married(2),children(4),car(2),save_act(2),current_act(2),mortgage(2), pep(2), all have distinct values as mentioned in the brackets. This shows the repetitions in values. Factorizing it would make better sense and would be easier to generate rules as association rules work better on factor data

```
# Data transformation:

bank <- bank%>% select_if(!names(.) %in% c('id')) #removing id column

#Discretizing age and income
#using mutate for age
bank <- bank %>% mutate( age_levels = case_when( 1<= age & age<= 10~ "Kids",
                                                11<=age & age<=19~ "Teens",
                                                20<=age & age<=29 ~ "Twenties",
                                                30<=age & age<=39 ~ "Thirties",
                                                40<=age & age<=49 ~ "Forties",
```



```
                                                50<=age & age<=59 ~ "Fifties",
                                                60<=age & age<= 69 ~ "Sixties"

))

#using cut,

min <- min(bank$income)
max <- max(bank$income)
bins<- seq(min, max,(max -min)/5)
bank$income_levels <- cut(bank$income,breaks =bins , labels = c("i1","i2","i3","i4","i5"))


#converting varibales into factors:
bank$sex <- as.factor(bank$sex )
bank$region <- as.factor(bank$region)
bank$married  <- as.factor(bank$married)
bank$children  <- as.factor(bank$children)
bank$car <- as.factor(bank$car)
bank$save_act <- as.factor(bank$save_act)
bank$current_act <- as.factor(bank$current_act)
bank$mortgage <- as.factor(bank$mortgage)
bank$pep <- as.factor(bank$pep)
bank$age_levels <- as.factor(bank$age_levels)
```

ASSOCIATION RULES:

Association rules help in understanding any underlying patterns that would lead to the target or are likely to cause the target. We use Apriori algorithm which considers the prior knowledge of data or patterns that are present to form rules which can be used as a firm reference for future use cases.

After transforming the data we need to convert it into transaction format with only factor variables as inputs as rules can be mined only from nominal data.

````
```{r}
library(arules)
library(arulesViz)


bank_transactions <- as(select(bank, col = -c('age','income')),"transactions")
rules <- apriori(bank_transactions, parameter = list(supp = 0.13, conf = 0.8, minlen = 2))
sorted_rules <- sort(rules,  by="confidence", decreasing=TRUE)
inspect(sorted_rules)
````

Support 0.01 and conf - 0.8 gave us to many rules to inspect.

Support of 0.1 and confidence = 0.8 gave us 95 rules. Having low support or confidence values generated way too many rules.

A general observation was that People with no children and mortgage ,have no pep and either have current account or savings or neither, are likely to be married. Other factors like gender, and region are also likely to say if the person is married or not. So plans can be suggested accordingly to them. Which we will see further through pep analysis.

People with no mortgage, no pep , coming from inner city region, unmarried females are likely to get a current account in addition to a savings account if they have one.
````

```
##      lhs                     rhs                 support confidence  coverage     lift count
## [1]  {children=0,
##       current_act=YES,
##       mortgage=NO,
##       pep=NO}             => {married=YES}     0.1333333  0.9756098 0.1366667 1.478197    80
## [2]  {children=0,
##       mortgage=NO,
##       pep=NO}             => {married=YES}     0.1733333  0.9719626 0.1783333 1.472671   104
## [3]  {married=YES,
##       children=0,
##       save_act=YES,
##       current_act=YES}    => {pep=NO}          0.1333333  0.9195402 0.1450000 1.692405    80
## [4]  {married=YES,
##       children=0,
##       current_act=YES,
##       mortgage=NO}        => {pep=NO}          0.1333333  0.9090909 0.1466667 1.673173    80
## [5]  {married=YES,
##       children=0,
##       save_act=YES}       => {pep=NO}          0.1783333  0.8991597 0.1983333 1.654895   107
## [6]  {married=YES,
##       children=0,
##       mortgage=NO}        => {pep=NO}          0.1733333  0.8965517 0.1933333 1.650095   104
## [7]  {children=0,
##       car=NO,
##       pep=NO}             => {married=YES}     0.1333333  0.8791209 0.1516667 1.332001    80
## [8]  {sex=FEMALE,
##       children=0,
##       pep=NO}             => {married=YES}     0.1300000  0.8666667 0.1500000 1.313131    78
## [9]  {age_levels=Sixties} => {save_act=YES}    0.1383333  0.8469388 0.1633333 1.227448    83
```

```
## [10] {save_act=YES,
##       mortgage=NO,
##       pep=NO}            => {married=YES}     0.2000000  0.8450704 0.2366667 1.280410    120
## [11] {children=0,
##       pep=NO}            => {married=YES}     0.2350000  0.8443114 0.2783333 1.279260    141
## [12] {save_act=YES,
##       current_act=YES,
##       mortgage=NO,
##       pep=NO}            => {married=YES}     0.1516667  0.8425926 0.1800000 1.276655     91
## [13] {children=1,
##       save_act=YES}      => {pep=YES}         0.1333333  0.8421053 0.1583333 1.844026     80
## [14] {region=INNER_CITY,
##       mortgage=NO,
##       pep=NO}            => {married=YES}     0.1333333  0.8333333 0.1600000 1.262626     80
## [15] {children=1,
##       current_act=YES}   => {pep=YES}         0.1400000  0.8316832 0.1683333 1.821204     84
## [16] {sex=FEMALE,
##       married=YES,
##       children=0}        => {pep=NO}          0.1300000  0.8297872 0.1566667 1.527216     78
## [17] {sex=MALE,
##       mortgage=NO,
##       pep=NO}            => {married=YES}     0.1300000  0.8297872 0.1566667 1.257253     78
## [18] {children=0,
##       current_act=YES,
##       pep=NO}            => {married=YES}     0.1750000  0.8267717 0.2116667 1.252684    105
## [19] {married=NO,
##       save_act=YES}      => {current_act=YES} 0.1883333  0.8248175 0.2283333 1.087671    113
## [20] {car=NO,
##       mortgage=NO,
##       pep=NO}            => {married=YES}     0.1483333  0.8240741 0.1800000 1.248597     89
## [21] {married=NO,
##       car=NO}            => {current_act=YES} 0.1400000  0.8235294 0.1700000 1.085973     84
## [22] {mortgage=NO,
##       pep=NO}            => {married=YES}     0.2850000  0.8181818 0.3483333 1.239669    171
## [23] {children=0,
##       save_act=YES,
##       pep=NO}            => {married=YES}     0.1783333  0.8167939 0.2183333 1.237567    107
## [24] {current_act=YES,
##       mortgage=NO,
##       pep=NO}            => {married=YES}     0.2150000  0.8164557 0.2633333 1.237054    129
## [25] {children=1}        => {pep=YES}         0.1833333  0.8148148 0.2250000 1.784266    110
## [26] {region=INNER_CITY,
##       mortgage=NO,
##       pep=NO}            => {current_act=YES} 0.1300000  0.8125000 0.1600000 1.071429     78
## [27] {save_act=YES,
##       mortgage=NO,
##       pep=YES}           => {current_act=YES} 0.1733333  0.8125000 0.2133333 1.071429    104
## [28] {car=YES,
##       mortgage=NO,
##       pep=NO}            => {married=YES}     0.1366667  0.8118812 0.1683333 1.230123     82
## [29] {car=NO,
##       pep=YES}           => {current_act=YES} 0.1833333  0.8088235 0.2266667 1.066580    110
## [30] {sex=FEMALE,
##       mortgage=NO,
```

```
##       pep=NO}            => {married=YES}     0.1550000  0.8086957 0.1916667 1.225296     93
## [31] {car=NO,
##       save_act=YES,
##       mortgage=NO}       => {current_act=YES} 0.1733333  0.8062016 0.2150000 1.063123    104
## [32] {region=INNER_CITY,
##       save_act=YES,
##       mortgage=NO}       => {current_act=YES} 0.1500000  0.8035714 0.1866667 1.059655     90
## [33] {car=NO,
##       mortgage=NO}       => {current_act=YES} 0.2633333  0.8020305 0.3283333 1.057623    158
## [34] {sex=FEMALE,
##       region=INNER_CITY} => {current_act=YES} 0.1750000  0.8015267 0.2183333 1.056958    105
## [35] {sex=FEMALE,
##       married=NO}        => {current_act=YES} 0.1400000  0.8000000 0.1750000 1.054945     84
## [36] {married=YES,
##       children=0,
##       car=NO}            => {pep=NO}          0.1333333  0.8000000 0.1666667 1.472393     80
## [37] {region=INNER_CITY,
##       save_act=YES,
##       pep=NO}            => {married=YES}     0.1333333  0.8000000 0.1666667 1.212121     80
## [38] {region=INNER_CITY,
##       save_act=YES,
##       pep=NO}            => {current_act=YES} 0.1333333  0.8000000 0.1666667 1.054945     80
```

PEP Analysis :

I divided my analysis/ rule mining to people signing up for PEP and people not signing up for pep. This helps in understanding who the target audience should be and if some extra targets were to be made outside the rules/patterns ,who can be excluded from the marketing list as they are highly likely not to signup for PEP. From a business POV understanding which factors to focus on and which factors to avoid while marketing, would help in marketing costs and efforts.

Targets for marketing : who sign up for PEP

After adjusting the support and confidence levels, support = 0.08 and confidence = 0.6 gave some good results to understand. Tried out various levels like support = 0.05, 0.01 , confidence = 0.9, 0.75(these generated less or no rules)

Rule 2:
{married=YES, children=1,save_act=YES} => {pep=YES} : This rule appears 9.5% frequently in our data. People who have one children, a savings account and are married are 87.69% likely to buy the PEP. The lift of 1.92 suggests that People who have one children, a savings account and are married buy pep more than the expected value. This suggests that targeting such clients would result in increase in pep sales.

Rule 3 :
{children=1, save_act=YES, mortgage=NO} => {pep=YES} : This rule appears 8% frequently in our data. People who have one children, a savings account and have no mortgage are 87.27% likely to buy the PEP. The lift of 1.91 suggests that People who have one children, a savings account and have no mortgage buy pep more than the expected value. This suggests that targeting such clients would result in increase in pep sales.

A similar rule of **{children=1, save_act=YES, current_act=NO} => {pep=YES},**this category of people were 10.5% frequent with a likelihood of 86.3% with overall account of 63(from the 600 observations)

Rules 16,17,18,19:
{married=NO,save_act=YES,current_act=YES,mortgage=NO} => {pep =YES}
:support = 8.8% confidence = 75.7% lift = 1.65 count= 53

{married=NO,save_act=YES,mortgage=NO} =>{pep = YES}
support = 10.67% confidence = 74.42% lift= 1.63 count =64

{married=NO,current_act=YES,mortgage=NO}=> {pep = YES}
support = 12.17% confidence = 71.6% lift = 1.57 count = 73

{married=NO,mortgage=NO}=> {pep = YES}
support = 15.3% confidence = 70.76% lift = 1.55 count= 92

All 4 of these rules had closer numbers with 2 conditions in common, the person not being married and having no mortage, in addition to this, if the person had savings account, current account or both, was highly likely in buying or signing up for pep. So when a new client has the above mentioned qualities, pep should be suggested to them and they will most likely get it. A general observation from all rules is that people with children are more likely to have pep and there is one case where the people who are not married and have no children would get pep, this could be because single people are making investments as they might not have a lot of financial responsibilities now and are looking at gaining profits through their investment.

```
pep_yes_rules <- apriori(bank_transactions,
             parameter=list(supp=0.08, conf=0.60,minlen = 3),
             control=list(verbose=F),
             appearance=list(default="lhs",rhs= c("pep=YES")))

sorted_pep_yes_rules <- sort(pep_yes_rules,by = "lift",descending = TRUE)
inspect(sorted_pep_yes_rules)
```

```
##      lhs                     rhs          support confidence   coverage     lift count
## [1]  {children=1,
##       income_levels=i2}   => {pep=YES} 0.08000000  0.9600000 0.08333333 2.102190    48
## [2]  {married=YES,
##       children=1,
##       save_act=YES}       => {pep=YES} 0.09500000  0.8769231 0.10833333 1.920270    57
## [3]  {children=1,
##       save_act=YES,
##       mortgage=NO}        => {pep=YES} 0.08000000  0.8727273 0.09166667 1.911082    48
## [4]  {children=1,
##       save_act=YES,
##       current_act=YES}    => {pep=YES} 0.10500000  0.8630137 0.12166667 1.889811    63
## [5]  {married=YES,
##       children=1,
##       current_act=YES}    => {pep=YES} 0.09333333  0.8615385 0.10833333 1.886581    56
## [6]  {children=1,
##       mortgage=NO}        => {pep=YES} 0.11833333  0.8452381 0.14000000 1.850886    71
## [7]  {children=1,
##       save_act=YES}       => {pep=YES} 0.13333333  0.8421053 0.15833333 1.844026    80
## [8]  {children=1,
##       current_act=YES,
##       mortgage=NO}        => {pep=YES} 0.09500000  0.8382353 0.11333333 1.835552    57
## [9]  {sex=FEMALE,
##       children=1}         => {pep=YES} 0.09166667  0.8333333 0.11000000 1.824818    55
## [10] {children=1,
##       current_act=YES}    => {pep=YES} 0.14000000  0.8316832 0.16833333 1.821204    84
## [11] {married=YES,
##       children=1}         => {pep=YES} 0.12333333  0.8314607 0.14833333 1.820717    74
## [12] {children=1,
##       car=YES}            => {pep=YES} 0.09166667  0.8208955 0.11166667 1.797581    55
## [13] {children=1,
##       car=NO}             => {pep=YES} 0.09166667  0.8088235 0.11333333 1.771146    55
## [14] {sex=MALE,
##       children=1}         => {pep=YES} 0.09166667  0.7971014 0.11500000 1.745478    55
## [15] {region=INNER_CITY,
##       children=1}         => {pep=YES} 0.08500000  0.7846154 0.10833333 1.718136    51
## [16] {married=NO,
```

```
##       save_act=YES,
##       current_act=YES,
##       mortgage=NO}        => {pep=YES} 0.08833333  0.7571429 0.11666667 1.657977    53
## [17] {married=NO,
##       save_act=YES,
##       mortgage=NO}        => {pep=YES} 0.10666667  0.7441860 0.14333333 1.629604    64
## [18] {married=NO,
##       current_act=YES,
##       mortgage=NO}        => {pep=YES} 0.12166667  0.7156863 0.17000000 1.567196    73
## [19] {married=NO,
##       mortgage=NO}        => {pep=YES} 0.15333333  0.7076923 0.21666667 1.549691    92
## [20] {married=NO,
##       children=0}         => {pep=YES} 0.09500000  0.6867470 0.13833333 1.503826    57
## [21] {region=INNER_CITY,
##       married=NO}         => {pep=YES} 0.09500000  0.6263736 0.15166667 1.371621    57
## [22] {married=NO,
##       car=NO,
##       current_act=YES}    => {pep=YES} 0.08500000  0.6071429 0.14000000 1.329510    51
```

<u>Targets to avoid in marketing : who donot sign up for PEP</u>

After adjusting the support and confidence levels, support = 0.1 and confidence = 0.8 gave some good results to understand On the contrary the people who didnot get pep are:

{married=YES,children=0,save_act=YES,current_act=YES} => {pep=NO}
support = 13.33% confidence = 91.95% lift= 1.692405 count = 80

{married=YES,children=0,save_act=YES,mortgage=NO} =>{pep=NO}
support= 12.16% confidence = 91.25% lift= 1.679448 count = 73

{married=YES,children=0,current_act=YES,mortgage = NO} => {pep=NO}
support = 13.33% confidence = 90.9% lift = 1.673173 count = 80

People who are married, have savings or current account or both , have no mortgage and have no children all have more or less close likeliness(around 91%) of not getting pep. Such customers should not be sent a followup email and some other plans should be suggested instead. Another observation was, females with no children but who are married are less likely to get the pep plan.So people with no children should can be safely avoided as potential buyers of pep plan.

```{r}

pep_no_rules <- apriori(bank_transactions,
                parameter=list(supp=0.1, conf=0.80,minlen = 3),
                control=list(verbose=F),
                appearance=list(default="lhs",rhs= c("pep=NO")))

sorted_pep_no_rules <- sort(pep_no_rules,by = "lift",descending = TRUE)

inspect(sorted_pep_no_rules)
```

```
##      lhs                  rhs        support confidence  coverage     lift count
## [1]  {married=YES,
##       children=0,
##       save_act=YES,
##       current_act=YES} => {pep=NO} 0.1333333  0.9195402 0.1450000 1.692405    80
## [2]  {married=YES,
##       children=0,
##       save_act=YES,
##       mortgage=NO}     => {pep=NO} 0.1216667  0.9125000 0.1333333 1.679448    73
## [3]  {married=YES,
##       children=0,
##       current_act=YES,
##       mortgage=NO}     => {pep=NO} 0.1333333  0.9090909 0.1466667 1.673173    80
## [4]  {sex=FEMALE,
##       married=YES,
##       children=0,
##       mortgage=NO}     => {pep=NO} 0.1050000  0.9000000 0.1166667 1.656442    63
## [5]  {married=YES,
##       children=0,
##       save_act=YES}    => {pep=NO} 0.1783333  0.8991597 0.1983333 1.654895   107
## [6]  {married=YES,
##       children=0,
##       mortgage=NO}     => {pep=NO} 0.1733333  0.8965517 0.1933333 1.650095   104
## [7]  {married=YES,
##       children=0,
##       car=NO,
##       mortgage=NO}     => {pep=NO} 0.1000000  0.8955224 0.1116667 1.648201    60
## [8]  {sex=FEMALE,
##       married=YES,
##       children=0,
##       current_act=YES} => {pep=NO} 0.1000000  0.8450704 0.1183333 1.555344    60
## [9]  {sex=FEMALE,
##       married=YES,
##       children=0}      => {pep=NO} 0.1300000  0.8297872 0.1566667 1.527216    78
## [10] {married=YES,
##       children=0,
##       car=NO,
##       current_act=YES} => {pep=NO} 0.1000000  0.8108108 0.1233333 1.492290    60
## [11] {married=YES,
##       children=0,
##       car=NO}          => {pep=NO} 0.1333333  0.8000000 0.1666667 1.472393    80
```