

George_Smith_HW3_IST707

George Smith

7/25/2021

Introduction

The purpose of this report is to conduct association rule discovery on the bank dataset. This will be completed by first conducting the necessary preprocessing steps, association rule discovery, and setting PEP on the right hand side of the rules to view the rules that are generated. Once these rules are generated the top 5 most “interesting” rules will be discussed in detail.

#Load the libraries

```
library(arules)
```

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

```
##
```

```
## Attaching package: 'arules'
```

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

```
##
```

```
##      abbreviate, write
```

```
library(arulesViz)
```

load the dataset

```
bank <- read.csv("C:/Users/GeorgeSmith/Documents/bankdata_csv_all (1).csv")
length(which(is.na(bank)))
```

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

The dataset has no blank values

view the data

```
str(bank)
```

```
## 'data.frame':    600 obs. of  12 variables:
## $ id           : chr  "ID12101" "ID12102" "ID12103" "ID12104" ...
## $ age          : int   48 40 51 23 57 57 22 58 37 54 ...
## $ sex          : chr   "FEMALE" "MALE" "FEMALE" "FEMALE" ...
## $ region       : chr   "INNER_CITY" "TOWN" "INNER_CITY" "TOWN" ...
## $ income       : num  17546 30085 16575 20375 50576 ...
## $ married      : chr   "NO" "YES" "YES" "YES" ...
## $ children     : int    1 3 0 3 0 2 0 0 2 2 ...
## $ car          : chr   "NO" "YES" "YES" "NO" ...
## $ save_act     : chr   "NO" "NO" "YES" "NO" ...
## $ current_act  : chr   "NO" "YES" "YES" "YES" ...
## $ mortgage     : chr   "NO" "YES" "NO" "NO" ...
## $ pep          : chr   "YES" "NO" "NO" "NO" ...
```

We see 600 observations of 12 variables including id, age, sex, region, income, married, children, car, save_act, current act, mortgage, and pep

preprocess

remove ID fields

```
bank <- bank[,-1]
```

discretize or convert fields to nominal

using factor

```
bank$sex <- factor(bank$sex)
bank$region <- factor(bank$region)
bank$married <- factor(bank$married)
bank$children <- factor(bank$children)
bank$car <- factor(bank$car)
bank$save_act <- factor(bank$save_act)
bank$current_act <- factor(bank$current_act)
bank$mortgage <- factor(bank$mortgage)
bank$pep <- factor(bank$pep)
# using cut
bank$age <- cut(bank$age, breaks = c(10,20,30,40,50,60,70),
               labels=c("teens", "twenties", "thirties", "fourties", "fifties",
                        "sixties"))
min_inc <- min(bank$income) - 1
max_inc <- max(bank$income)
bins <- 5
```

```
width <- (max_inc - min_inc)/bins
bank$income <- cut(bank$income, breaks=seq(min_inc, max_inc, width))
str(bank)
```

```
## 'data.frame':    600 obs. of  11 variables:
## $ age          : Factor w/ 6 levels "teens","twenties",...: 4 3 5 2 5 5 2 5 3 5 ...
## $ sex          : Factor w/ 2 levels "FEMALE","MALE": 1 2 1 1 1 1 2 2 1 2 ...
## $ region       : Factor w/ 4 levels "INNER_CITY","RURAL",...: 1 4 1 4 2 4 2 4 3 4 ...
## $ income       : Factor w/ 5 levels "(5.01e+03,1.66e+04]",...: 2 3 1 2 4 3 1 2 2 2 ...
## $ married      : Factor w/ 2 levels "NO","YES": 1 2 2 2 2 2 1 2 2 2 ...
## $ children     : Factor w/ 4 levels "0","1","2","3": 2 4 1 4 1 3 1 1 3 3 ...
## $ car          : Factor w/ 2 levels "NO","YES": 1 2 2 1 1 1 1 2 2 2 ...
## $ save_act     : Factor w/ 2 levels "NO","YES": 1 1 2 1 2 2 1 2 1 2 ...
## $ current_act  : Factor w/ 2 levels "NO","YES": 1 2 2 2 1 2 2 2 1 2 ...
## $ mortgage     : Factor w/ 2 levels "NO","YES": 1 2 1 1 1 1 1 1 1 1 ...
## $ pep          : Factor w/ 2 levels "NO","YES": 2 1 1 1 1 2 2 1 1 1 ...
```

All fields have been converted to Factors

any missing values?

```
length(which(is.na(bank)))
```

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

No missing values were identified in the data set

check for incomplete rows

```
nrow(bank[!complete.cases(bank),])
```

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

NO incomplete rows were identified in the data set

check for complete rows

```
nrow(bank[complete.cases(bank),])
```

```
## [1] 600
```

600 complete rows were identified in the data set

generate rules and explore data; note the low support level at this point

```
rules <- apriori(bank, parameter = list(supp=0.001, conf = 0.8))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.8      0.1    1 none FALSE              TRUE      5   0.001      1
## maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 0
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[33 item(s), 600 transaction(s)] done [0.00s].
## sorting and recoding items ... [33 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 8 9 10

## Warning in apriori(bank, parameter = list(supp = 0.001, conf = 0.8)): Mining
## stopped (maxlen reached). Only patterns up to a length of 10 returned!

## done [0.12s].
## writing ... [1110713 rule(s)] done [0.21s].
## creating S4 object ... done [0.70s].
```

Rounding rules to 2 digits

```
options(digits=2)
```

get summary info about all rules

```
summary(rules)
```

```
## set of 1110713 rules
##
## rule length distribution (lhs + rhs):sizes
##      2      3      4      5      6      7      8      9     10
##    10     332    5339   42275  162634  323259  338686  186474  51704
```

```
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.0      7.0      8.0     7.5     8.0     10.0
##
## summary of quality measures:
##      support      confidence      coverage      lift      count
## Min.      :0.002    Min.      :0.80    Min.      :0.00    Min.      : 1.1    Min.      : 1
## 1st Qu.:0.002    1st Qu.:1.00    1st Qu.:0.00    1st Qu.: 1.5    1st Qu.: 1
## Median :0.002    Median :1.00    Median :0.00    Median : 2.0    Median : 1
## Mean   :0.003    Mean   :0.99    Mean   :0.00    Mean   : 2.6    Mean   : 2
## 3rd Qu.:0.003    3rd Qu.:1.00    3rd Qu.:0.00    3rd Qu.: 2.9    3rd Qu.: 2
## Max.   :0.285    Max.   :1.00    Max.   :0.35    Max.   :17.1    Max.   :171
##
## mining info:
## data ntransactions support confidence
## bank          600    0.001      0.8
```

sort the rules to view most relevant first (confidence)

```
rules <- sort(rules, by="confidence", decreasing=TRUE)
inspect(rules[1:20])
```

```
##      lhs                                rhs      support confidence coverage lift count
## [1] {income=(5.15e+04,6.31e+04)} => {save_act=YES} 0.0583          1 0.0583 1.4 35
## [2] {income=(3.99e+04,5.15e+04)} => {save_act=YES} 0.1183          1 0.1183 1.4 71
## [3] {region=SUBURBAN,
##      income=(5.15e+04,6.31e+04)} => {age=sixties} 0.0067          1 0.0067 6.7 4
## [4] {region=SUBURBAN,
##      income=(5.15e+04,6.31e+04)} => {pep=YES} 0.0067          1 0.0067 2.2 4
## [5] {region=SUBURBAN,
##      income=(5.15e+04,6.31e+04)} => {save_act=YES} 0.0067          1 0.0067 1.4 4
## [6] {income=(5.15e+04,6.31e+04),
##      children=3} => {age=sixties} 0.0033          1 0.0033 6.7 2
## [7] {income=(5.15e+04,6.31e+04),
##      children=3} => {region=RURAL} 0.0033          1 0.0033 6.2 2
## [8] {income=(5.15e+04,6.31e+04),
##      children=3} => {pep=YES} 0.0033          1 0.0033 2.2 2
## [9] {income=(5.15e+04,6.31e+04),
##      children=3} => {sex=FEMALE} 0.0033          1 0.0033 2.0 2
## [10] {income=(5.15e+04,6.31e+04),
##      children=3} => {mortgage=NO} 0.0033          1 0.0033 1.5 2
## [11] {income=(5.15e+04,6.31e+04),
##      children=3} => {save_act=YES} 0.0033          1 0.0033 1.4 2
## [12] {income=(5.15e+04,6.31e+04),
##      children=3} => {current_act=YES} 0.0033          1 0.0033 1.3 2
## [13] {income=(5.15e+04,6.31e+04),
##      children=2} => {age=sixties} 0.0233          1 0.0233 6.7 14
## [14] {income=(5.15e+04,6.31e+04),
##      current_act=NO} => {age=sixties} 0.0083          1 0.0083 6.7 5
## [15] {income=(5.15e+04,6.31e+04),
##      pep=NO} => {age=sixties} 0.0083          1 0.0083 6.7 5
```

```
## [16] {age=sixties,
##       income=(5.15e+04,6.31e+04]} => {save_act=YES}      0.0517      1  0.0517  1.4   31
## [17] {region=RURAL,
##       income=(5.15e+04,6.31e+04]} => {pep=YES}           0.0083      1  0.0083  2.2    5
## [18] {region=RURAL,
##       income=(5.15e+04,6.31e+04]} => {mortgage=NO}       0.0083      1  0.0083  1.5    5
## [19] {region=RURAL,
##       income=(5.15e+04,6.31e+04]} => {save_act=YES}       0.0083      1  0.0083  1.4    5
## [20] {region=RURAL,
##       income=(5.15e+04,6.31e+04]} => {current_act=YES}    0.0083      1  0.0083  1.3    5
```

sort the rules to view based on lift

```
rules <- sort(rules, by="lift", decreasing=TRUE)
inspect(rules[1:20])
```

```
##      lhs                                rhs      support confidence coverage lift count
## [1] {age=sixties,
##      region=SUBURBAN,
##      children=2,
##      current_act=NO} => {income=(5.15e+04,6.31e+04]} 0.0017      1  0.0017   17    1
## [2] {age=sixties,
##      region=SUBURBAN,
##      children=2,
##      mortgage=YES}   => {income=(5.15e+04,6.31e+04]} 0.0033      1  0.0033   17    2
## [3] {age=sixties,
##      region=SUBURBAN,
##      children=2,
##      car=YES}        => {income=(5.15e+04,6.31e+04]} 0.0033      1  0.0033   17    2
## [4] {age=sixties,
##      region=SUBURBAN,
##      children=1,
##      car=NO}         => {income=(5.15e+04,6.31e+04]} 0.0017      1  0.0017   17    1
## [5] {age=sixties,
##      region=SUBURBAN,
##      children=1,
##      mortgage=NO}    => {income=(5.15e+04,6.31e+04]} 0.0017      1  0.0017   17    1
## [6] {age=sixties,
##      region=SUBURBAN,
##      children=1,
##      current_act=YES} => {income=(5.15e+04,6.31e+04]} 0.0017      1  0.0017   17    1
## [7] {age=sixties,
##      region=SUBURBAN,
##      married=NO,
##      mortgage=YES}   => {income=(5.15e+04,6.31e+04]} 0.0017      1  0.0017   17    1
## [8] {age=sixties,
##      region=SUBURBAN,
##      married=NO,
##      pep=YES}        => {income=(5.15e+04,6.31e+04]} 0.0017      1  0.0017   17    1
## [9] {age=sixties,
##      region=SUBURBAN,
```

```

##      married=NO,
##      save_act=YES}    => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [10] {region=SUBURBAN,
##      children=2,
##      current_act=NO,
##      mortgage=YES}    => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [11] {region=SUBURBAN,
##      children=2,
##      current_act=NO,
##      pep=YES}         => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [12] {sex=FEMALE,
##      region=SUBURBAN,
##      children=2,
##      mortgage=YES}    => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [13] {age=sixties,
##      region=RURAL,
##      married=NO,
##      children=3}      => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [14] {age=sixties,
##      married=NO,
##      children=3,
##      pep=YES}         => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [15] {age=sixties,
##      sex=FEMALE,
##      married=NO,
##      children=3}      => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [16] {age=sixties,
##      married=NO,
##      children=3,
##      car=NO}          => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [17] {age=sixties,
##      region=RURAL,
##      married=NO,
##      children=2}      => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [18] {age=sixties,
##      region=RURAL,
##      children=2,
##      pep=YES}         => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [19] {age=sixties,
##      sex=FEMALE,
##      region=RURAL,
##      children=1}      => {income=(5.15e+04,6.31e+04]}  0.0017          1  0.0017  17    1
## [20] {age=sixties,
##      region=RURAL,
##      married=NO,
##      pep=YES}         => {income=(5.15e+04,6.31e+04]}  0.0033          1  0.0033  17    2

```

“minlen” is to avoid empty LHS items

```
rules2 <- apriori(data = bank, parameter=list(supp=0.001,conf=0.08,minlen=3))
```

```
## Apriori
```

```
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##      0.08      0.1      1 none FALSE              TRUE        5   0.001      3
## maxlen target  ext
##      10 rules TRUE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##      0.1 TRUE TRUE  FALSE TRUE      2      TRUE
##
## Absolute minimum support count: 0
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[33 item(s), 600 transaction(s)] done [0.00s].
## sorting and recoding items ... [33 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 7 8 9 10

## Warning in apriori(data = bank, parameter = list(supp = 0.001, conf = 0.08, :
## Mining stopped (maxlen reached). Only patterns up to a length of 10 returned!

## done [0.12s].
## writing ... [2651598 rule(s)] done [0.40s].
## creating S4 object ... done [0.95s].
```

```
rules2 <- sort(rules2, by="lift", decreasing=TRUE)
inspect(rules2[1:20])
```

	lhs	rhs	support	confidence	coverage	lift	count
## [1]	{age=sixties, region=SUBURBAN, children=2, current_act=NO}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [2]	{age=sixties, region=SUBURBAN, children=2, mortgage=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0033	1	0.0033	17	2
## [3]	{age=sixties, region=SUBURBAN, children=2, car=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0033	1	0.0033	17	2
## [4]	{age=sixties, region=SUBURBAN, children=1, car=NO}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [5]	{age=sixties, region=SUBURBAN, children=1, mortgage=NO}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [6]	{age=sixties, region=SUBURBAN, children=1,						

##	current_act=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [7]	{age=sixties,						
##	region=SUBURBAN,						
##	married=NO,						
##	mortgage=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [8]	{age=sixties,						
##	region=SUBURBAN,						
##	married=NO,						
##	pep=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [9]	{age=sixties,						
##	region=SUBURBAN,						
##	married=NO,						
##	save_act=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [10]	{region=SUBURBAN,						
##	children=2,						
##	current_act=NO,						
##	mortgage=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [11]	{region=SUBURBAN,						
##	children=2,						
##	current_act=NO,						
##	pep=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [12]	{sex=FEMALE,						
##	region=SUBURBAN,						
##	children=2,						
##	mortgage=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [13]	{age=sixties,						
##	region=RURAL,						
##	married=NO,						
##	children=3}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [14]	{age=sixties,						
##	married=NO,						
##	children=3,						
##	pep=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [15]	{age=sixties,						
##	sex=FEMALE,						
##	married=NO,						
##	children=3}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [16]	{age=sixties,						
##	married=NO,						
##	children=3,						
##	car=NO}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [17]	{age=sixties,						
##	region=RURAL,						
##	married=NO,						
##	children=2}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [18]	{age=sixties,						
##	region=RURAL,						
##	children=2,						
##	pep=YES}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [19]	{age=sixties,						
##	sex=FEMALE,						
##	region=RURAL,						
##	children=1}	=> {income=(5.15e+04,6.31e+04]}	0.0017	1	0.0017	17	1
## [20]	{age=sixties,						

```
##      region=RURAL,
##      married=NO,
##      pep=YES}      => {income=(5.15e+04,6.31e+04]}  0.0033      1  0.0033  17  2
```

if we want to target items to generate rules (for example, pep=YES)

```
rules3 <- apriori(data = bank, parameter=list(supp=0.001,conf=0.08,minlen=2),
                  appearance = list(default="lhs", rhs="pep=YES"),
                  control=list(verbose=F))
rules3 <- sort(rules3, decreasing = TRUE, by="lift")
inspect(rules3[1:20])
```

##	lhs	rhs	support	confidence	coverage	lift	count
## [1]	{region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04]} => {pep=YES}	0.0067	1	0.0067	2.2	4	
## [2]	{income=(5.15e+04,6.31e+04],						
##	children=3} => {pep=YES}	0.0033	1	0.0033	2.2	2	
## [3]	{region=RURAL,						
##	income=(5.15e+04,6.31e+04]} => {pep=YES}	0.0083	1	0.0083	2.2	5	
## [4]	{age=fifties,						
##	income=(5.15e+04,6.31e+04]} => {pep=YES}	0.0067	1	0.0067	2.2	4	
## [5]	{income=(5.15e+04,6.31e+04],						
##	children=2} => {pep=YES}	0.0233	1	0.0233	2.2	14	
## [6]	{income=(5.15e+04,6.31e+04],						
##	children=1} => {pep=YES}	0.0150	1	0.0150	2.2	9	
## [7]	{income=(5.15e+04,6.31e+04],						
##	current_act=NO} => {pep=YES}	0.0083	1	0.0083	2.2	5	
## [8]	{age=fourties,						
##	children=1} => {pep=YES}	0.0533	1	0.0533	2.2	32	
## [9]	{age=sixties,						
##	region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04]} => {pep=YES}	0.0067	1	0.0067	2.2	4	
## [10]	{region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04],						
##	children=2} => {pep=YES}	0.0050	1	0.0050	2.2	3	
## [11]	{region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04],						
##	children=1} => {pep=YES}	0.0017	1	0.0017	2.2	1	
## [12]	{region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04],						
##	current_act=NO} => {pep=YES}	0.0017	1	0.0017	2.2	1	
## [13]	{region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04],						
##	married=NO} => {pep=YES}	0.0017	1	0.0017	2.2	1	
## [14]	{region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04],						
##	mortgage=YES} => {pep=YES}	0.0033	1	0.0033	2.2	2	
## [15]	{region=SUBURBAN,						
##	income=(5.15e+04,6.31e+04],						

```
##      car=YES}                                => {pep=YES}  0.0033          1  0.0033  2.2    2
## [16] {sex=MALE,
##      region=SUBURBAN,
##      income=(5.15e+04,6.31e+04]} => {pep=YES}  0.0033          1  0.0033  2.2    2
## [17] {sex=FEMALE,
##      region=SUBURBAN,
##      income=(5.15e+04,6.31e+04]} => {pep=YES}  0.0033          1  0.0033  2.2    2
## [18] {region=SUBURBAN,
##      income=(5.15e+04,6.31e+04],
##      car=NO}                                => {pep=YES}  0.0033          1  0.0033  2.2    2
## [19] {region=SUBURBAN,
##      income=(5.15e+04,6.31e+04],
##      mortgage=NO}                            => {pep=YES}  0.0033          1  0.0033  2.2    2
## [20] {region=SUBURBAN,
##      income=(5.15e+04,6.31e+04],
##      married=YES}                            => {pep=YES}  0.0050          1  0.0050  2.2    3
```

We can view the top 20 strongest rules by confidence interval for pep = yes # we can set the left hand side to be “pep=YES” and find its antecedents

```
rules4 <- apriori(data = bank, parameter=list(supp=0.001,conf=0.08,minlen=2),
                  appearance = list(default="rhs", lhs="pep=YES"),
                  control=list(verbose=F))
rules4 <- sort(rules4, decreasing = TRUE, by="lift")
rules4 <- sort(rules4, decreasing = TRUE, by="confidence")
inspect(rules4[1:5])
```

```
##      lhs      rhs      support confidence coverage lift count
## [1] {pep=YES} => {current_act=YES} 0.35    0.77      0.46    1.02 211
## [2] {pep=YES} => {mortgage=NO}    0.30    0.66      0.46    1.02 182
## [3] {pep=YES} => {save_act=YES}   0.30    0.65      0.46    0.95 179
## [4] {pep=YES} => {married=YES}    0.26    0.56      0.46    0.85 154
## [5] {pep=YES} => {sex=MALE}      0.24    0.53      0.46    1.05 144
```

#discussion

Based on the above analysis I found the 5 most interesting rules to be:

1. LHS values of region=SUBURBAN,income=(5.15e+04,6.31e+04],mortgage=NO, and a RHS of PEP = yes . This rule has support confidence and lift values of .0033, 1, and 2.2 respectively. I found this rule interesting as it seems strange that a suburban resident with no mortgage would be looking to invest in a PEP. This led me to question maybe this resident as potentially paid their mortgage in full, and is looking to invest additional funds. The support value pf .0033 means represents the fraction of instances that contain the discussed itemset. The confidence score of 1 means that every instances of LHS the RHS has ocured. the lift value of 2.2 is a measyre of dependent or correlated events. A value greater than 1 represents a meaningful event.
2. LHS values of income=(5.15e+04,6.31e+04], children=1 and RHS of PEP = yes. This rule has support confidence and lift values of .0150 , 1, and 2.2 respectively. This rule is interesting as I would think customers with 1 child would be less likely to invest in a PEP than those with no children. As having no children would allow for additional investment income.

3.LHS values of region=SUBURBAN, income=(5.15e+04,6.31e+04], car=NO and RHS of PEP = yes. This rule has support confidence and lift values of .033 , 1, 2.2 respectively. This rule is interesting as I would

think being a suburban resident and not having a car would be a result of finances being allocated elsewhere, however these customers are investing in PEP.

4. LHS values of sex=FEMALE, region=SUBURBAN,income=(5.15e+04,6.31e+04] and RHS of PEP = yes. This rule has support confidence and lift values of .0033, 1, and 2.2 respectively. I found this interesting as I would think females would be less likely to invest in a PEP.
5. LHS values of region=SUBURBAN,income=(5.15e+04,6.31e+04],married=NO and RHS values of PEP = yes. This rule has support confidence and lift values of .0017, 1, and 2.2 respectively. This rule is interesting as I would think customers who are not married would spend more money on themselves rather than looking to invest in PEP.

conclusion

In conclusion, after conducting the necessary preprocessing steps, association rule discovery, and setting PEP on the right hand side of the rules to view the rules that are generated. I was able to successfully analyze the bank data set using association rules to identify the top 5 most interesting rules based on my personal opinion.