**Ryan Timbrook**

**NetID: RTIMBROO**

**Assignment Topic: Association Rules**

**Term: Summer, 2019**

Table of Contents

[1 Introduction 3](#_Toc14881659)

[1.1 Purpose 3](#_Toc14881660)

[1.2 Scope 3](#_Toc14881661)

[2 Analysis and Models 4](#_Toc14881662)

[2.1.1 Dataset Info 4](#_Toc14881663)

[2.1.2 Banking Dataset, Data Exploration & Cleaning 5](#_Toc14881664)

[2.1.3 Banking Dataset, Data Transformations 5](#_Toc14881665)

[2.2 Models 7](#_Toc14881666)

[2.2.1 Model ‘pep=YES Item Frequency’ Details 8](#_Toc14881667)

[2.2.2 Model ‘pep=YES Item Frequency’ Parameters 9](#_Toc14881668)

[2.2.3 Model ‘pep=NO Item Frequency’ Details 9](#_Toc14881669)

[2.2.4 Model ‘pep=NO Item Frequency’ Parameters 10](#_Toc14881670)

[2.2.5 Model ‘arules\_supp0.002\_conf0.9\_mil2\_mal4’ Details 11](#_Toc14881671)

[2.2.6 Model ‘arules\_supp0.002\_conf0.9\_mil2\_mal4’ Parameters 12](#_Toc14881672)

[2.2.7 Model ‘arules\_lhs-pep=YES\_supp0.0001\_conf0.15\_mil2’ Details 13](#_Toc14881673)

[2.2.8 Model ‘arules\_lhs-pep=YES\_supp0.0001\_conf0.15\_mil2’ Parameters 13](#_Toc14881674)

[2.2.9 Model ‘arules\_lhs-pep=NO\_supp0.0001\_conf0.15\_mil2’ Details 14](#_Toc14881675)

[2.2.10 Model ‘arules\_rhs-pep=NO\_supp0.0001\_conf0.15\_mil2’ Parameters 14](#_Toc14881676)

[2.2.11 Model ‘arules\_rhs-PEP\_supp0.002\_conf0.\_mil2’ Details 15](#_Toc14881677)

[2.2.12 Model ‘arules\_rhs-PEP\_supp0.002\_conf0.\_mil2’ Parameters 16](#_Toc14881678)

[3 Results 17](#_Toc14881679)

[3.1 Exploratory Model Results, 20 Strong Rules 17](#_Toc14881680)

[3.1.1 Exploratory Model Visualizations, Strong Rules 17](#_Toc14881681)

[3.2 RHS PEP Model Results, Top 5 Most Interesting 18](#_Toc14881682)

[4 Conclusions 21](#_Toc14881683)

# Introduction

## Purpose

Provide insights and suggestions on the kinds of potential buyers the financial institution (client) should target in their new ‘Personal Equity Plan’ (PEP) product launch using association rule data mining techniques.

## Scope

The marketing department of a financial firm keeps records on customers, including demographic information and, number of type of accounts. When launching a new product, such as a "Personal Equity Plan" (PEP), a direct mail piece, advertising the product, is sent to existing customers, and a record kept as to whether that customer responded and bought the product.

Perform Association Rule discovery on the clients banking dataset. Experiment with different parameters and preprocessing that identifies 20-30 strong rules. A strong rule is one that has high lift and confidence while at the same time having relatively good support.

Target rule generation of the PEP class to understand the types of customers who have bought PEP in the past and those who have not. Identify the rules this targeting creates and select the top 5 most ‘interesting’ rules. Provide the quality measures of these rules along with explaining their patterns. Make recommendations based on the discovery that provides the client with potential business opportunities.

# Analysis and Models

The marketing department of a financial firm keeps records on customers, including demographic information and, number of type of accounts. When launching a new product, such as a "Personal Equity Plan" (PEP), a direct mail piece, advertising the product, is sent to existing customers, and a record kept as to whether that customer responded and bought the product. Based on this store of prior experience, the managers decide to use data mining techniques to build customer profile models.

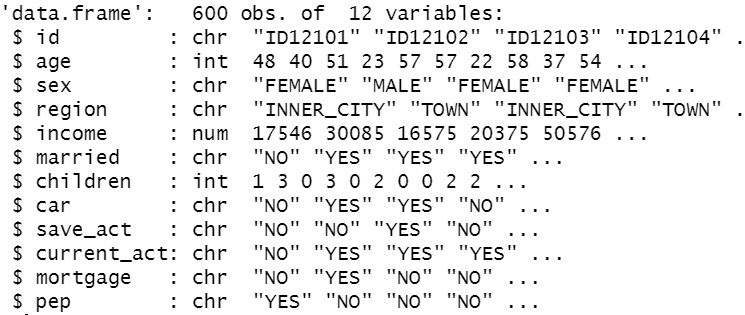
The data contains of a number of the following fields:

|  |  |
| --- | --- |
| id | a unique identification number |
| age | age of customer in years |
| sex | MALE / FEMALE |
| region | inner\_city/rural/suburban/town |
| income | income of customer |
| married | Is the customer married (YES/NO) |
| children | number of children |
| car | Does the customer own a car (YES/NO) |
| save\_act | Does the customer have a saving account (YES/NO) |
| current\_act | Does the customer have a current account (YES/NO) |
| mortgage | Does the customer have a mortgage (YES/NO) |
| pep | Did the customer buy a PEP after the last mailing (YES/NO) |

Each record is a customer description where the "pep" field indicates whether or not that customer bought a PEP after the last mailing.

### Dataset Info

The original dataset contains 600 observations with 12 variables in record format. In order to use the Apriori algorithm, this dataset needed to be transformed to a transactional format. See Data Transformation 2.1.3 for details.



### Banking Dataset, Data Exploration & Cleaning

There were no missing values from this dataset.

### Banking Dataset, Data Transformations

Preprocessing steps to convert data into transactional format before it can be used in the Apriori Algorithm for Association Rule discovery.

All numeric variables were converted to nominal through discretization or transformation into factors.

The id field was removed from the dataset prior to converting it from a record type to a transaction type using the function: *as(dataset,’transactions’)*

Specifically:

**Age**: was discretized to seven bins with labels

("CHILD","TEENS","TWENTIES","THIRTIES","FORTIES","FIFTIES","SENIORS")

Frequency Table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CHILD** | **TEENS** | **TWENTIES** | **THIRTIES** | **FORTIES** | **FIFTIES** | **SENIORS** |
| **0** | 37 | 119 | 125 | 128 | 101 | 90 |

**Income**: was discretized to three equal width bines with labels

('LOW','MID','HIGH')

Frequency Table:

|  |  |  |
| --- | --- | --- |
| **LOW** | **MID** | **HIGH** |
| **284** | 235 | 80 |

**Children**: was changed to a factor with four levels

(‘0’, ‘1’, ‘2’, ‘3’)

Frequency Table:

|  |  |  |  |
| --- | --- | --- | --- |
| **0** | **1** | **2** | **3** |
| **263** | 135 | 134 | 68 |

**Married**: was changed to a factor with two levels

(‘NO, ‘YES)

Frequency Table:

|  |  |
| --- | --- |
| **NO** | **YES** |
| **204** | 396 |

**Car**: was changed to a factor with two levels

(‘NO, ‘YES)

Frequency Table:

|  |  |
| --- | --- |
| **NO** | **YES** |
| **304** | 296 |

**Save\_act**: was changed to a factor with two levels

(‘NO, ‘YES)

Frequency Table:

|  |  |
| --- | --- |
| **NO** | **YES** |
| **186** | 414 |

**Current\_act**: was changed to a factor with two levels

(‘NO, ‘YES)

Frequency Table:

|  |  |
| --- | --- |
| **NO** | **YES** |
| **145** | 455 |

**Mortgage**: was changed to a factor with two levels

(‘NO, ‘YES)

Frequency Table:

|  |  |
| --- | --- |
| **NO** | **YES** |
| **391** | 209 |

**Sex**: was changed to a factor with two levels

(‘FEMAL’, ‘MALE’)

Frequency Table:

|  |  |
| --- | --- |
| **FEMAL** | **MALE** |
| **300** | 300 |

**Region**: was changed to a factor with two levels

(‘INNER\_CITY, ‘RURAL’,’SUBURBAN’,’TOWN’)

Frequency Table:

|  |  |  |  |
| --- | --- | --- | --- |
| **INNER\_CITY** | **RURAL** | **SUBURBAN** | **TOWN** |
| **269** | 96 | 62 | 173 |

**Pep**: was changed to a factor with two levels

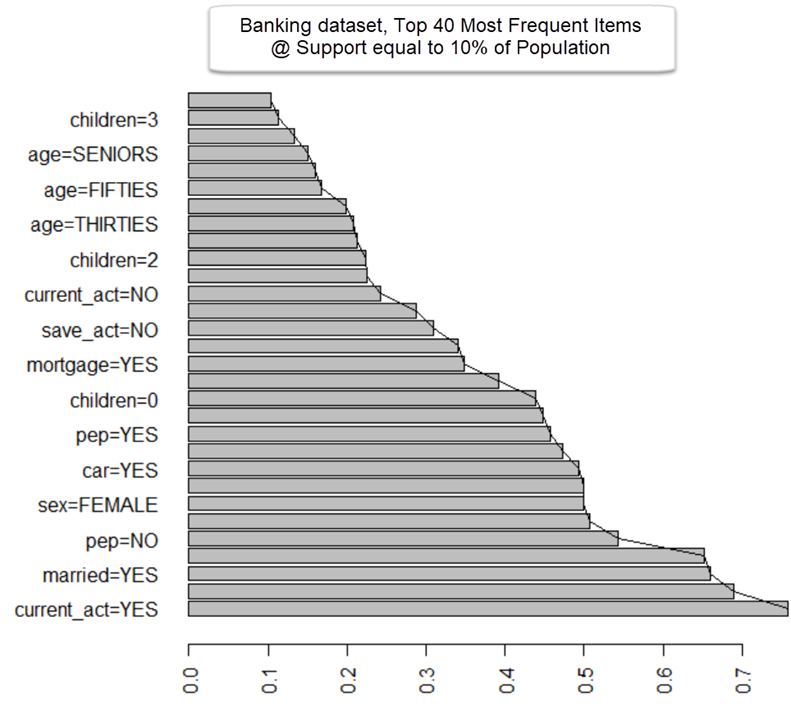
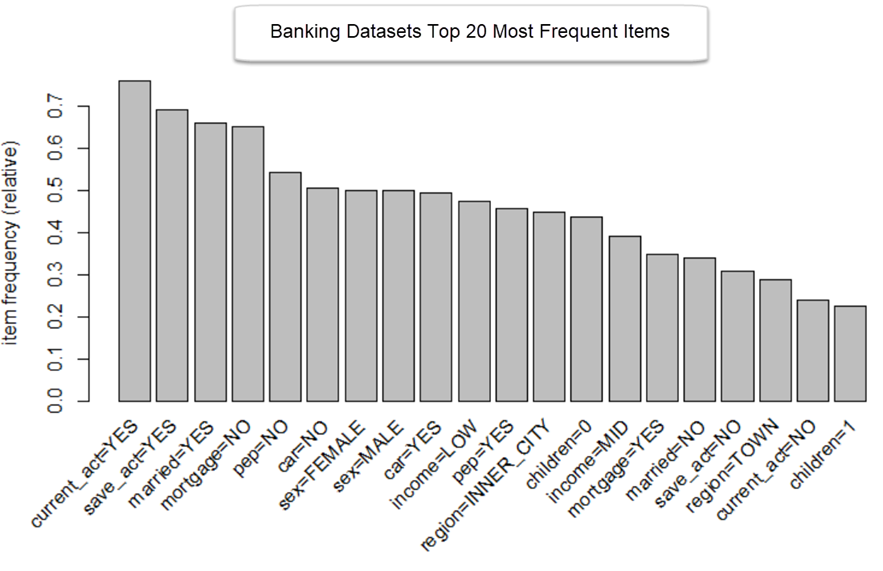
(‘NO, ‘YES)

Frequency Table:

|  |  |
| --- | --- |
| **NO** | **YES** |
| **326** | 274 |

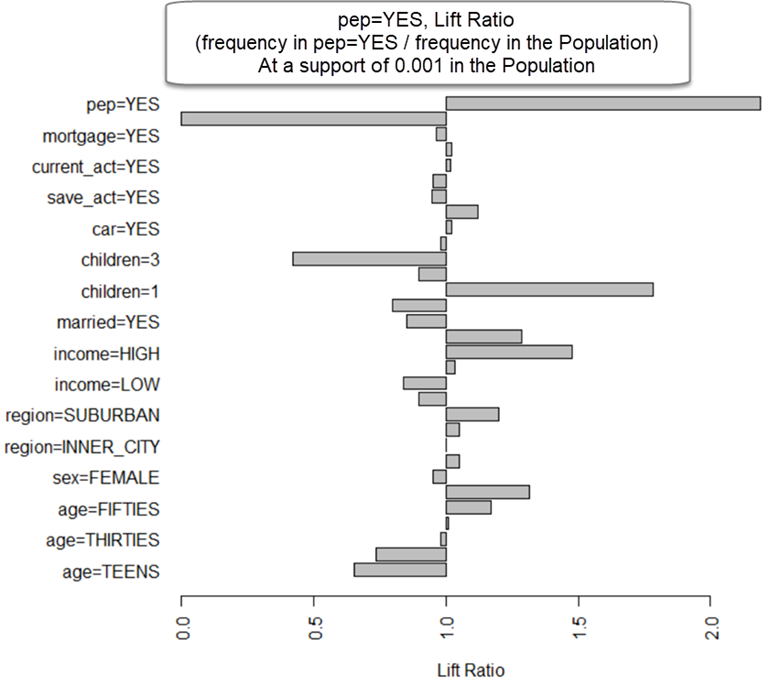
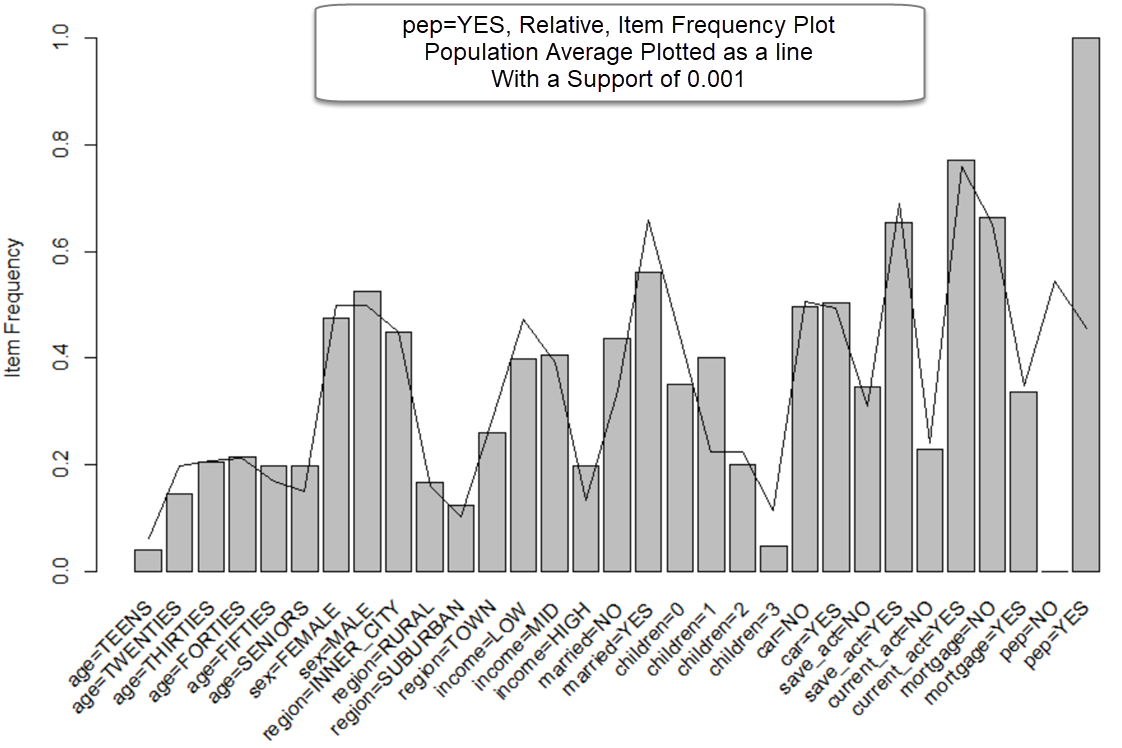
## Models

The banking dataset was transformed into a transaction object to be modeled using the Apriori Algorithm. The structure of the dataset as a transaction object showed it to be of 600 transactions (rows) and 32 items (columns) in sparse format.



### Model ‘pep=YES Item Frequency’ Details

Exploration of the frequency of items that cooccur with pep=YES.



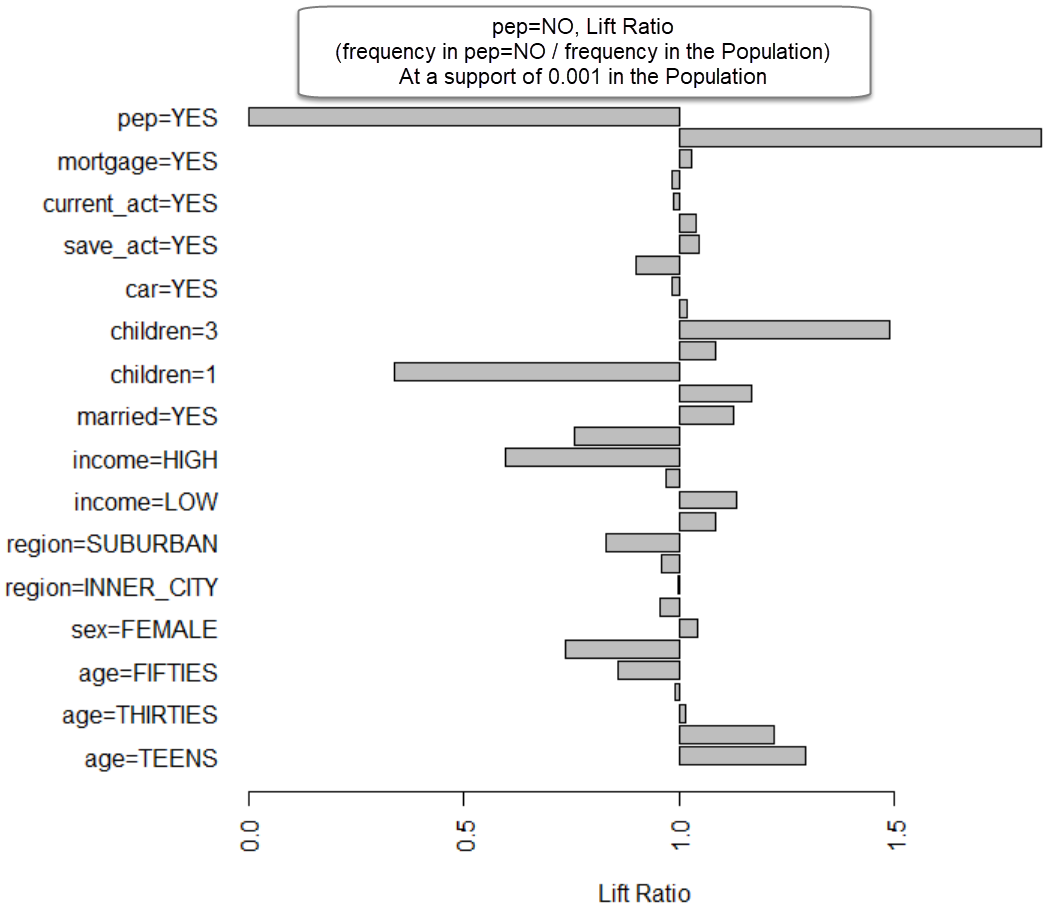
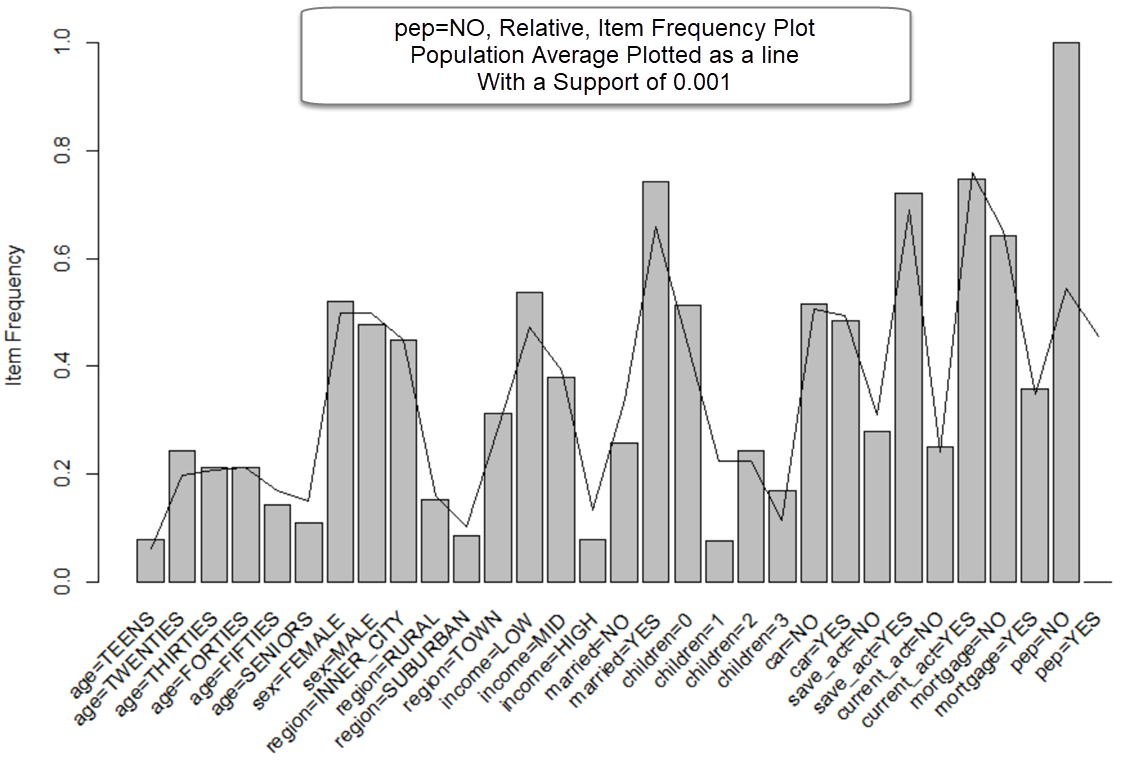
### Model ‘pep=YES Item Frequency’ Parameters

Banking transaction dataset filtered for those transactions where pep=YES is in the transaction record.

For the visualizations, a population support of 0.001 was configured.

### Model ‘pep=NO Item Frequency’ Details

Exploration of the frequency of items that cooccur with pep=NO for comparison to pep=YES.



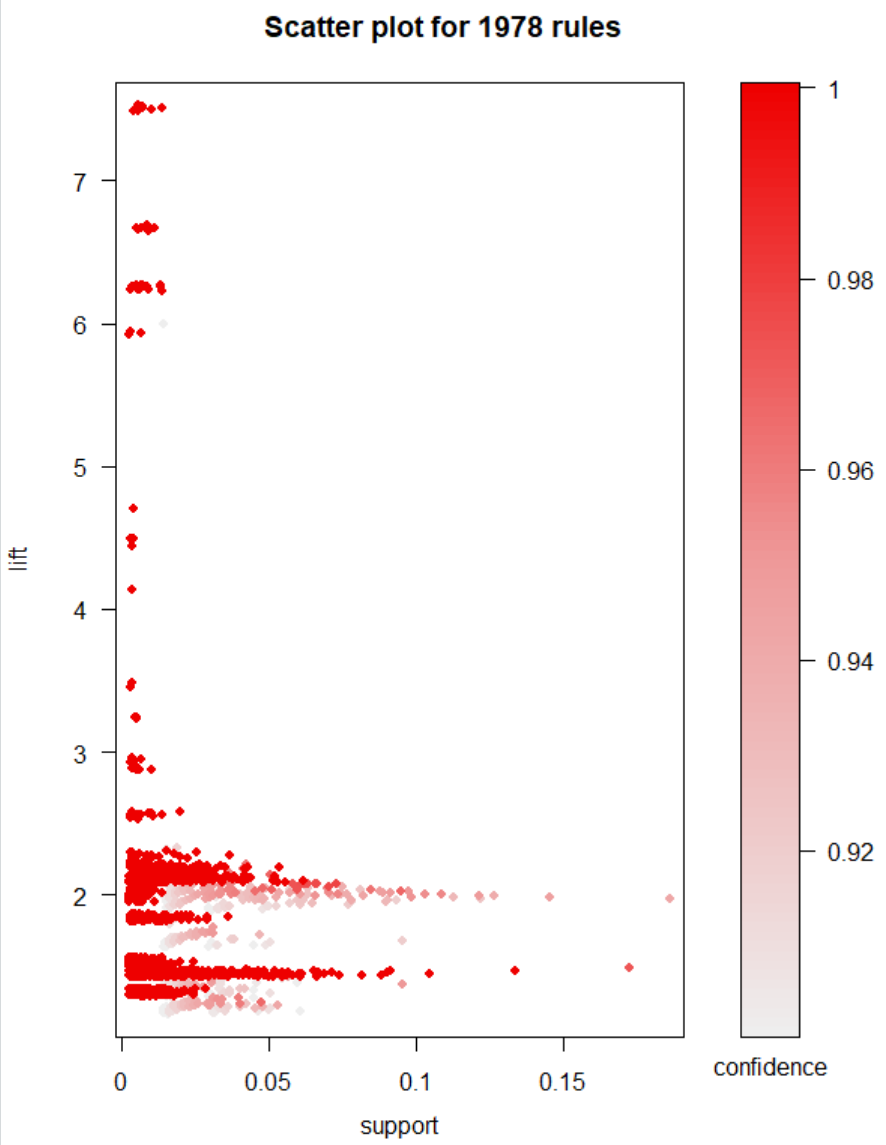
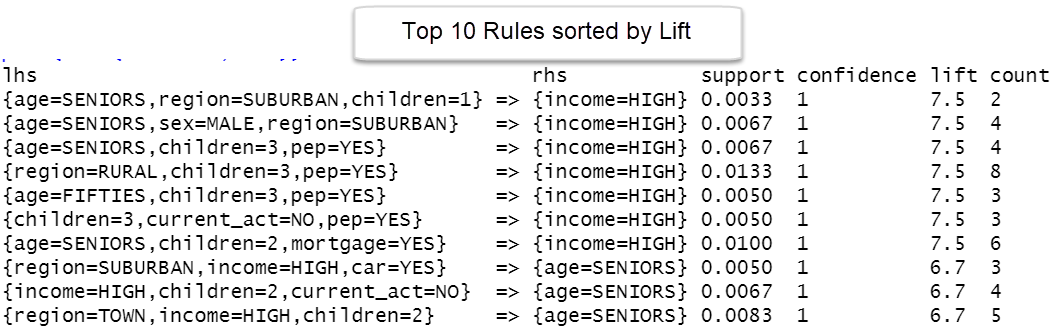
### Model ‘pep=NO Item Frequency’ Parameters

Banking transaction dataset filtered for those transactions where pep=NO is in the transaction record.

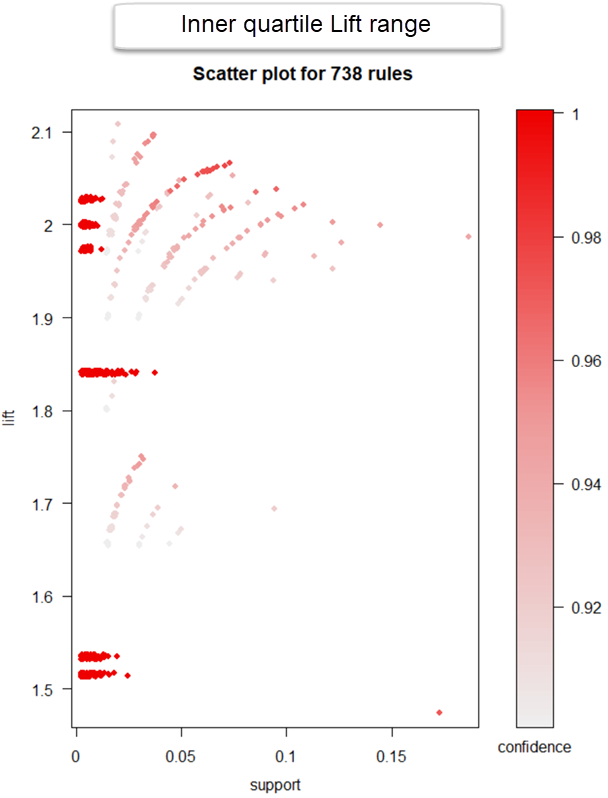
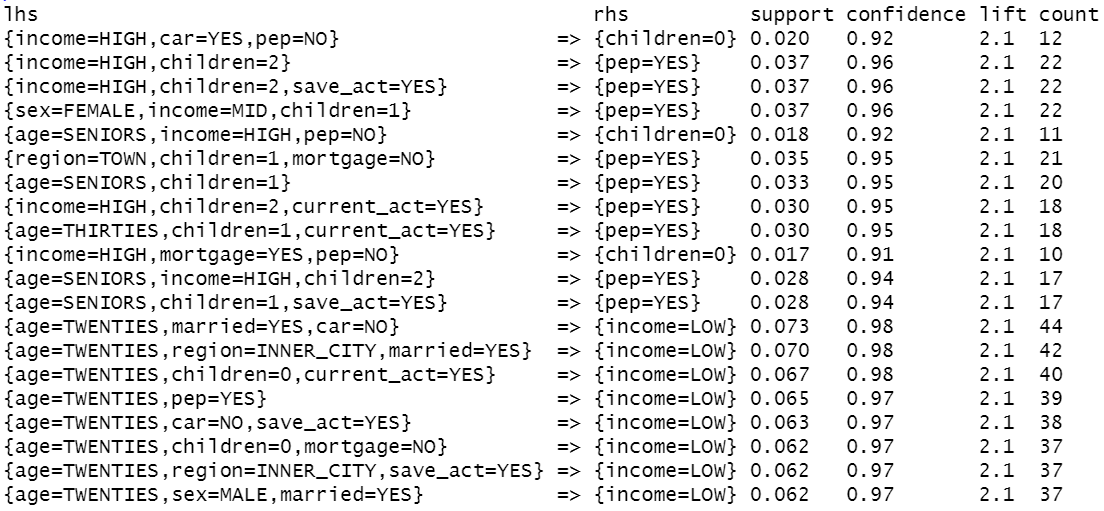
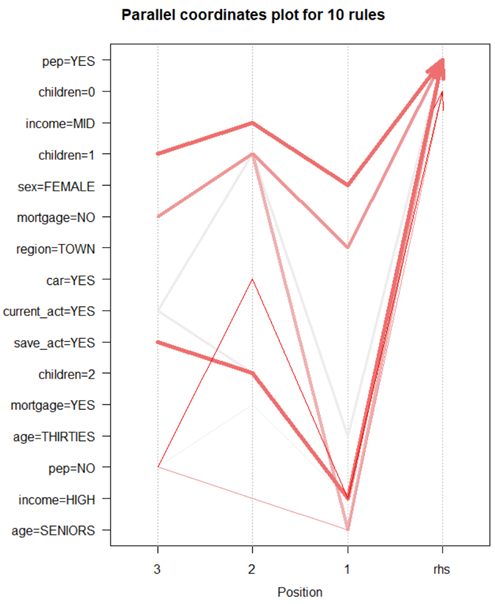
For the visualizations, a population support of 0.001 was configured.

### Model ‘arules\_supp0.002\_conf0.9\_mil2\_mal4’ Details

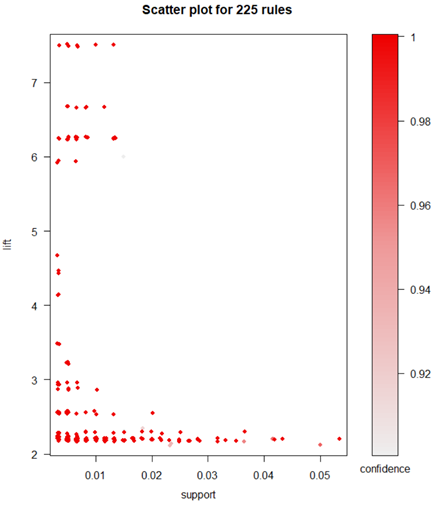
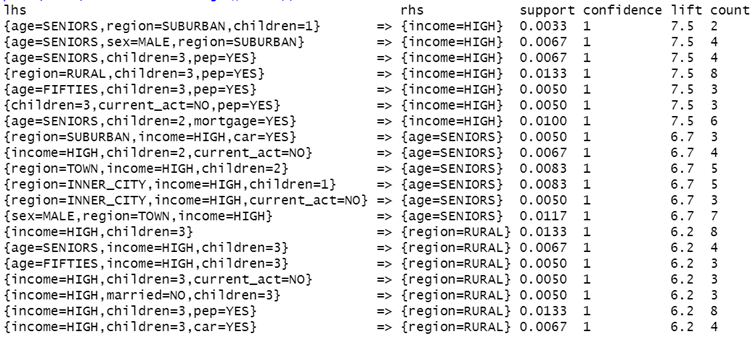
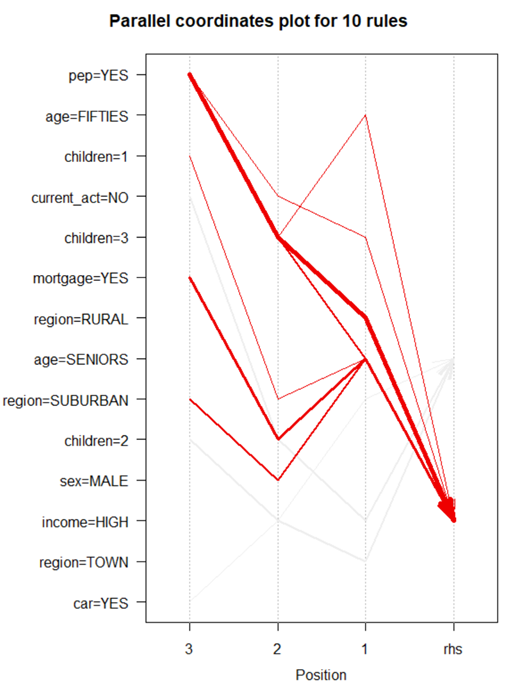
Initial rule set creation of 1978 rules.



To further narrow down the rule sets from this model, the model was narrowed by focusing on the middle range of rules based on their quality lift scores. This model’s middle lift range was found to be between 1.4 and 2.1. The resulting rule set size after narrowing was 738 rules.



Looking at the top 25% highest Lift scores (Lift > 2.1) for this model, the rule set was trimmed down to 225 rules.



### Model ‘arules\_supp0.002\_conf0.9\_mil2\_mal4’ Parameters

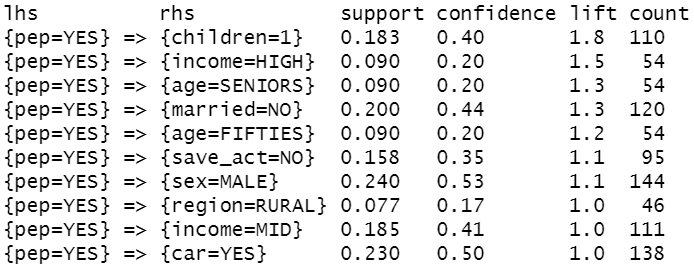
tBdRules = apriori(rBankData, parameter = list(supp = 0.002, conf = 0.9, minlen=2, maxlen=4, aval=TRUE),)

### Model ‘arules\_lhs-pep=YES\_supp0.0001\_conf0.15\_mil2’ Details

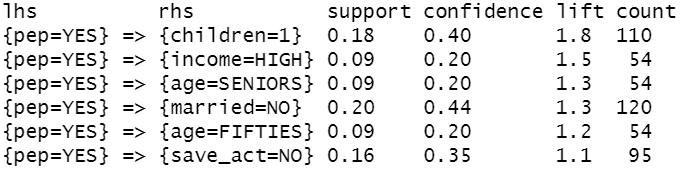
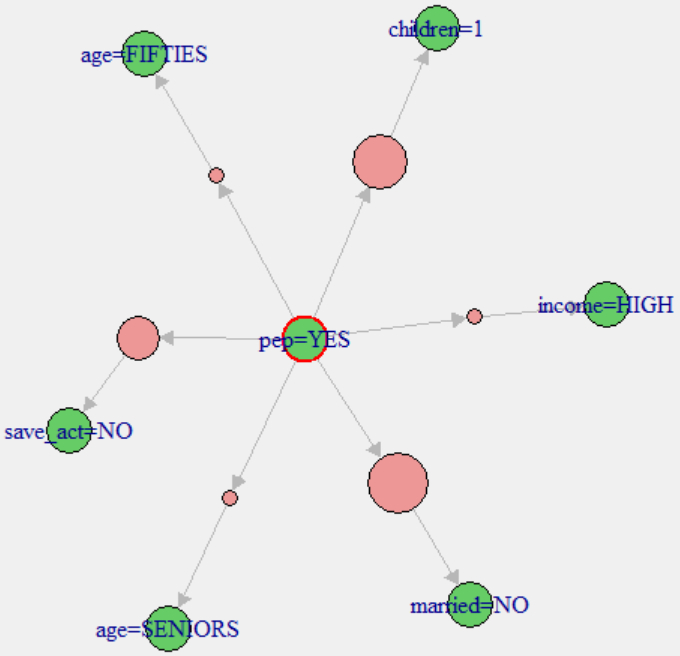
Creation of a rule set where ‘pep=YES’ is the lhs to see which attributes have the highest lift for someone who has responded in the past and purchased the PEP product.

Returns a rule set of 25 rules.

Top 10 rules:



Further sub-setting the rules set to those with a lift quality greater than 1.1 resulted in a new rule set of 6 transactions:



### Model ‘arules\_lhs-pep=YES\_supp0.0001\_conf0.15\_mil2’ Parameters

tBdRules\_pepYes = apriori(rBankData, parameter = list(supp = 0.0001, conf = 0.15, minlen=2 ,aval=TRUE),

appearance = list(lhs='pep=YES', default='rhs'),

control = list(verbose=T))

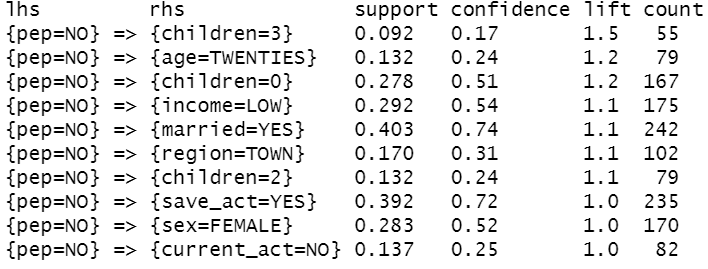
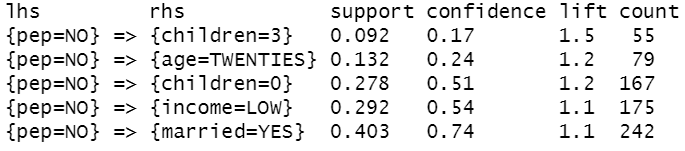
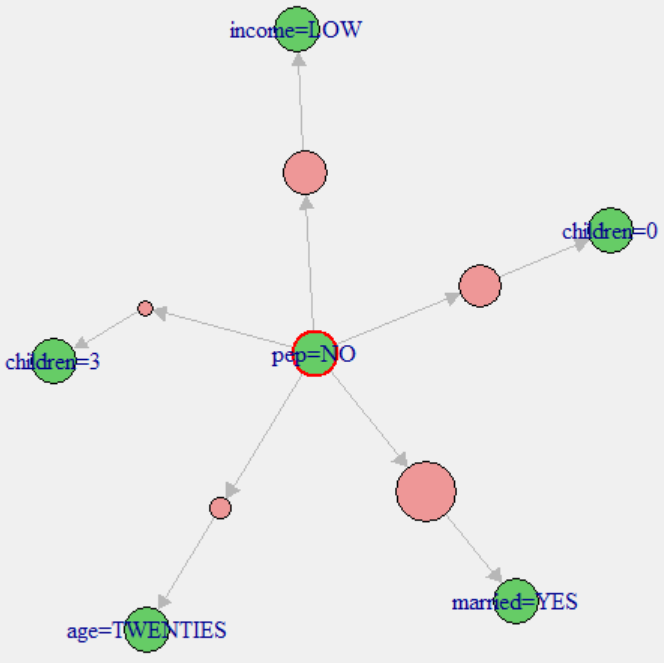
The confidence level was lowered to .15. At higher levels no rules were being returned.

### Model ‘arules\_lhs-pep=NO\_supp0.0001\_conf0.15\_mil2’ Details

Creation of a rule set where ‘pep=NO’ is the lhs to see which attributes have the highest lift for someone who has NOT responded in the past and purchased the PEP product.

Returns a rule set of 23 rules.

Top 10 rules: Top rules with a meaningful lift:



### Model ‘arules\_rhs-pep=NO\_supp0.0001\_conf0.15\_mil2’ Parameters

tBdRules\_pepNO = apriori(rBankData, parameter = list(supp = 0.0001, conf = 0.15, minlen=2 ,aval=TRUE),

appearance = list(lhs='pep=NO', default='rhs'),

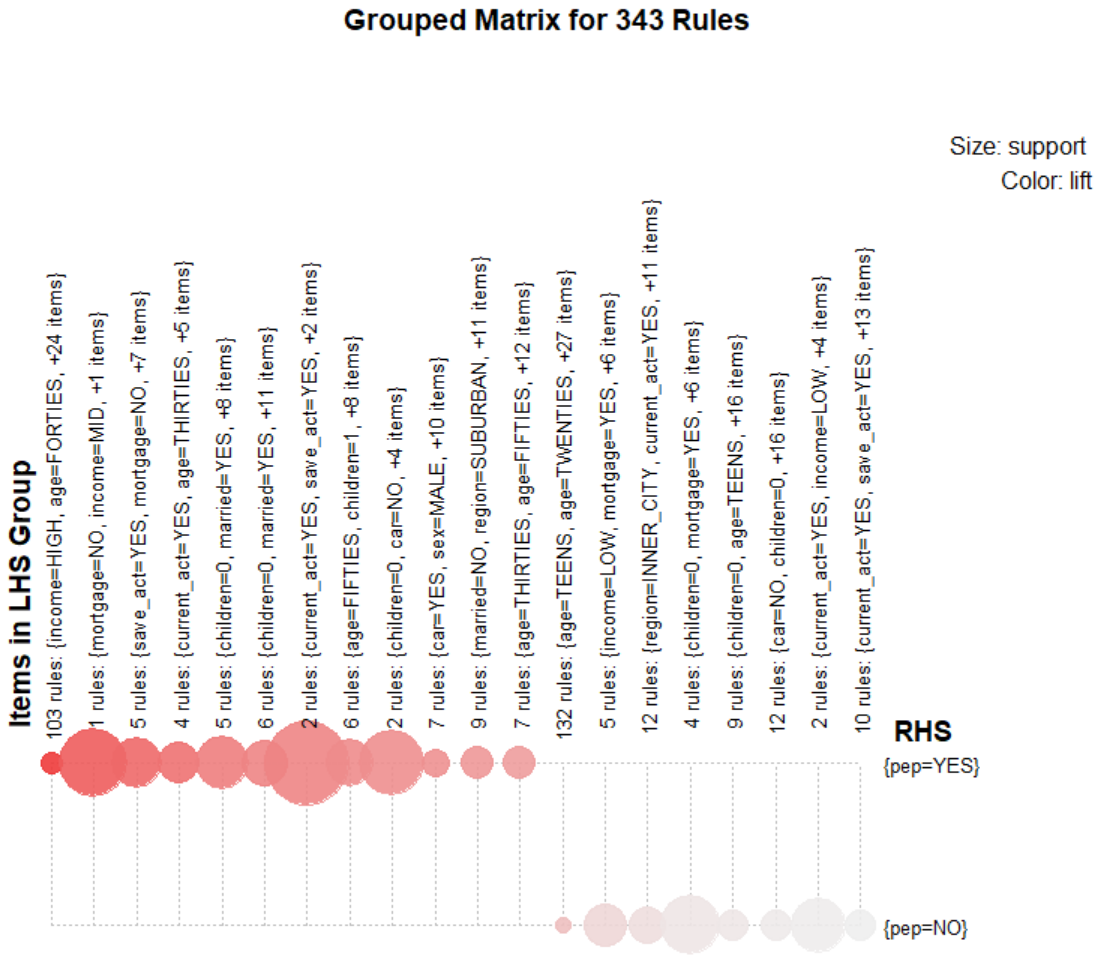
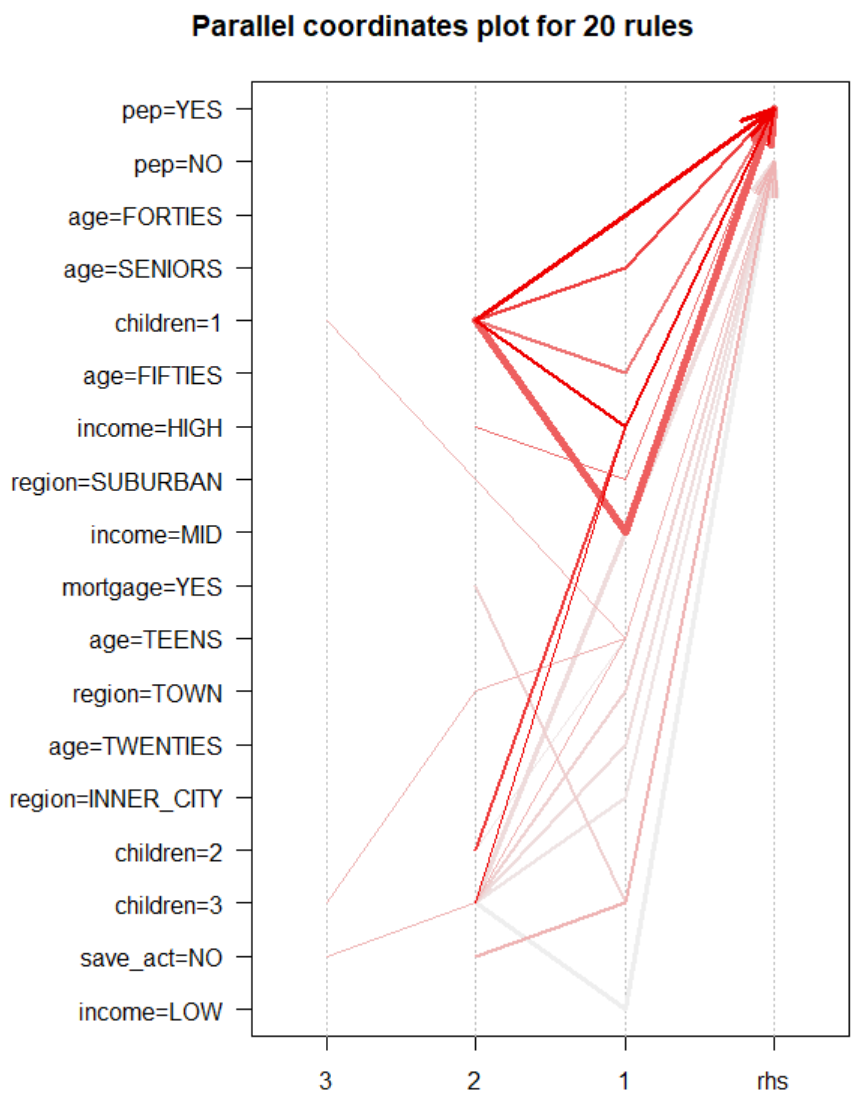
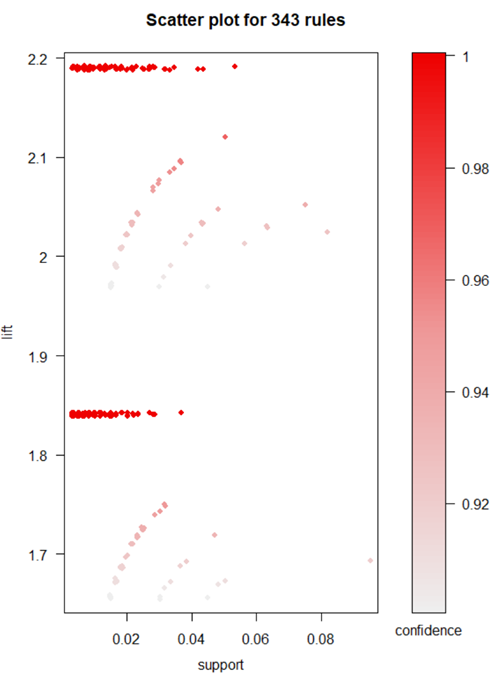
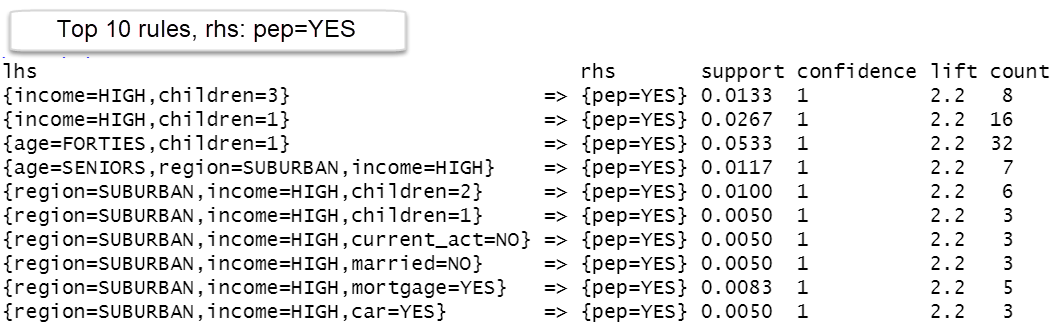
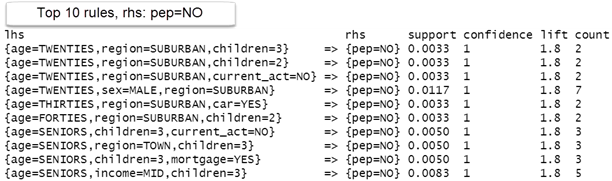
control = list(verbose=T))

The confidence level was lowered to .15. At higher levels no rules were being returned.

### Model ‘arules\_rhs-PEP\_supp0.002\_conf0.\_mil2’ Details

Creation of a rule set where ‘pep=YES’ and ‘pep=NO’ is the rhs. Identify which attributes customers are likely to have if they purchase PEP plans and those they are likely to have if they do not purchase PEP plans.

A rule set of 343 rules were generated.



### Model ‘arules\_rhs-PEP\_supp0.002\_conf0.\_mil2’ Parameters

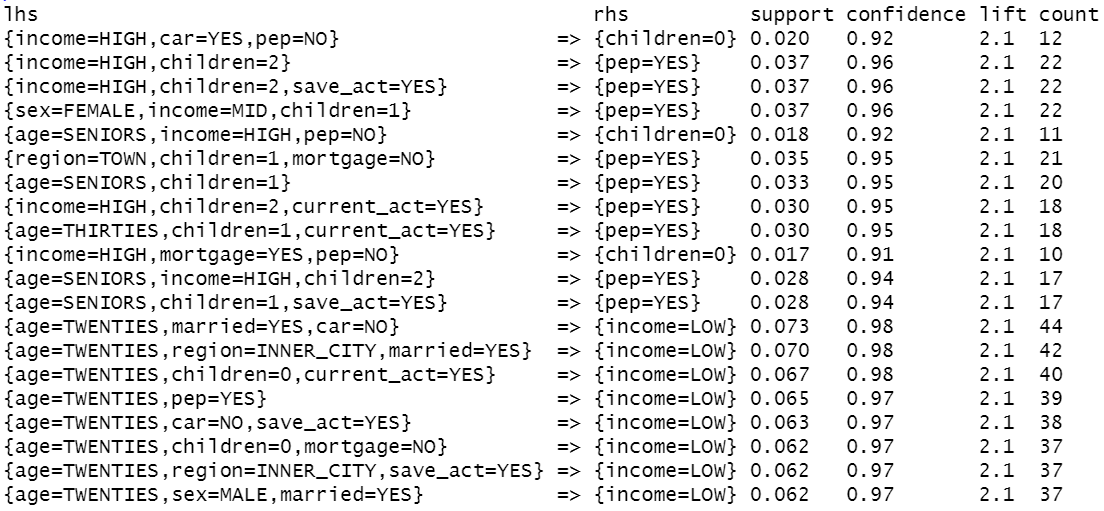
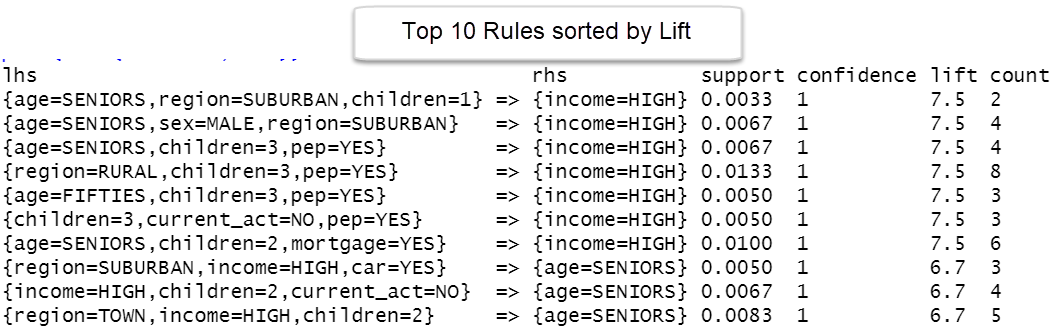
pepRhsRules = apriori(rBankData, parameter=list(sup=0.002, conf=0.9, minlen=2, maxlen=4),

appearance = list(rhs= c("pep=YES","pep=NO"), default='lhs'),

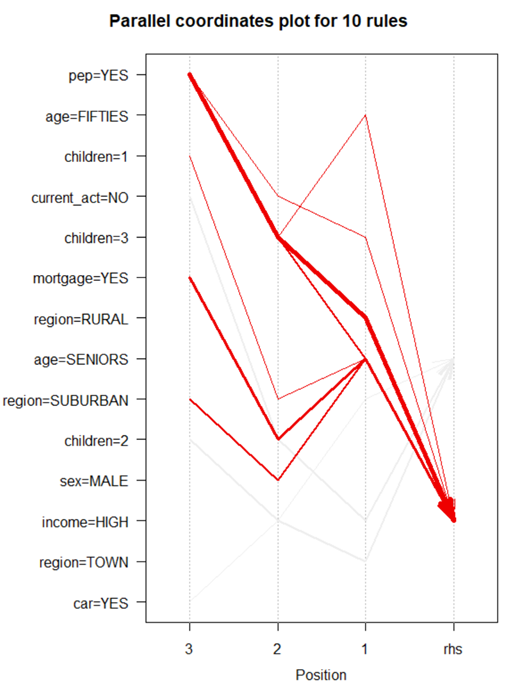
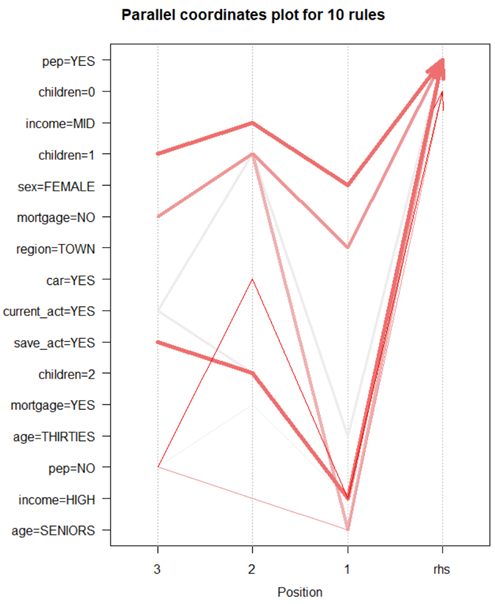
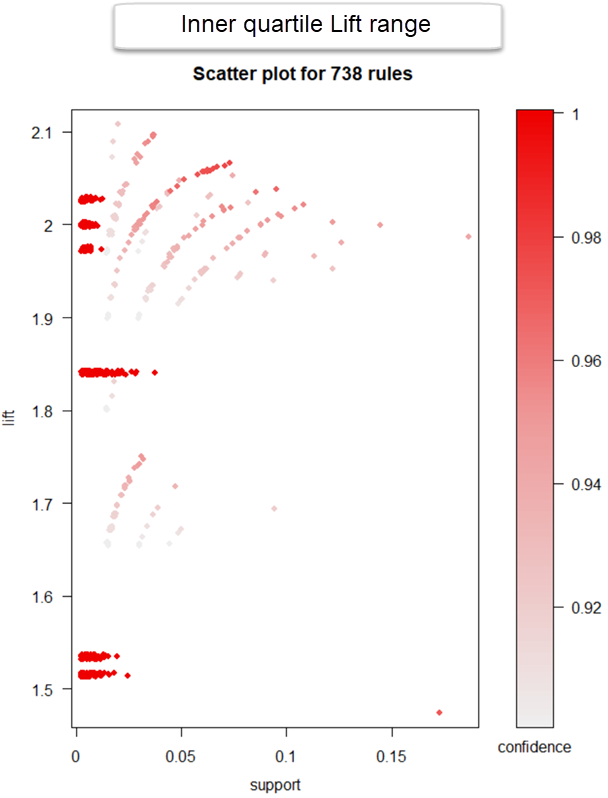
control = list(verbose=T))

# Results

## Exploratory Model Results, 20 Strong Rules



### Exploratory Model Visualizations, Strong Rules



Save\_act=YES, children=1 or more, income=HIGH

Region=SUBURBAN, age=SENORS|FIFTIES

## RHS PEP Model Results, Top 5 Most Interesting

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Rule** | **Support** | **Conf** | **Lift** | **Count** |
| 1 | {age=FORTIES,children=1}  => {pep=YES} | 0.0533 | 1 | 2.2 | 32 |
| 2 | {age=SENIORS,region=SUBURBAN,income=HIGH}  => {pep=YES} | 0.0117 | 1 | 2.2 | 7 |
| 3 | {age=SENIORS,income=HIGH,children=1}  => {pep=YES} | 0.0200 | 1 | 2.2 | 12 |
| 4 | {income=HIGH,children=1}  => {pep=YES} | 0.0267 | 1 | 2.2 | 16 |
| 5 | {age=FORTIES,married=NO,children=1}  => {pep=YES} | 0.0217 | 1 | 2.2 | 13 |

Rule 1: Customers who are in their forties and have one child are highly likely to buy PEP products.

Rule 2: Customers who are in their Senior years, who live in the Suburbs and a high income are highly likely to buy PEP products.

Rule 3: Customers who are in their Senior years, who have a high income and one child are highly likely to buy PEP products.

Rule 4: Customers with high incomes and have one child are highly likely to buy PEP products.

Rule 5: Customers in their forties, who are not married, and have one child are highly likely to buy PEP products.

Frequency of items within pep=YES in RHS. Restricting maxlen parameter to 4.

|  |  |
| --- | --- |
| Item | Frequency |
| save\_act=NO | 0 |
| save\_act=YES | 5% |
| current\_act=YES | 5% |
| current\_act=NO | 9% |
| car=NO | 5% |
| car=YES | 8% |
| married=NO | 10% |
| married=YES | 7% |
| income=LOW | 2% |
| income=MID | 11% |
| income=HIGH | 45% |
| children=0 | 0.6% |
| children=1 | 63% |
| children=2 | 12% |
| children=3 | 10% |
| age=TEENS | 0% |
| age=TWENTIES | 0% |
| age=THIRTIES | 5% |
| age=FORTIES | 13% |
| age=FIFTIES | 1% |
| age=SENIORS | 15% |
| mortgage=NO | 8% |
| mortgage=YES | 5% |
| sex=MALE | 9% |
| sex=FEMALE | 5% |
| region=INNER\_CITY | 3% |
| region=RURAL | 9% |
| region=SUBURBAN | 12% |
| region=TOWN | 8% |

# Conclusions

It’s recommended that the financial institution focus its new PEP product marketing at potential customers who are in their senior years or forties, have one child or possibly two, receive a high income and live in the suburbs.