**Syracuse University**

**IST-707 Assignment 3**

ThulasiRam RuppaKrishnan

IST 707Section: 35

Professor Ami Gates

Contents

[Introduction 3](#_Toc7036565)

[Analysis and Models 5](#_Toc7036566)

[**About the data** 5](#_Toc7036567)

[**Models** 8](#_Toc7036568)

[Results 17](#_Toc7036569)

[**Model II – No RHS restrictions** 17](#_Toc7036570)

[**Model II – pep=NO in RHS.** 18](#_Toc7036571)

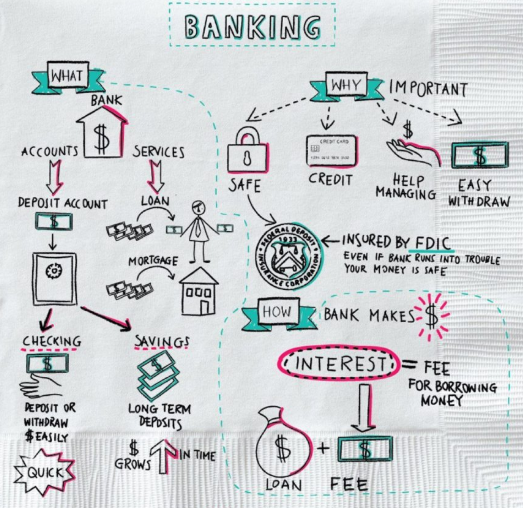
[**Model II – pep=YES in RHS.** 21](#_Toc7036572)

[**Interesting Facts** 23](#_Toc7036573)

[Conclusion 24](#_Toc7036574)

## **Introduction**

A **bank** is a financial institution that accepts deposits from the public and creates credit. Lending activities can be performed either directly or indirectly through capital markets. Due to their importance in the financial stability of a country, banks are highly regulated in most countries.



Activities undertaken by banks include personal banking, corporate banking, investment banking, private banking, transaction banking, insurance, consumer finance, foreign exchange trading, commodity trading, trading in equities, futures and options trading and money market trading.

Retail banking, also known as consumer banking, is the provision of services by a bank to the general public, rather than to companies, corporations or other banks, which are often described as wholesale banking. Banking services which are regarded as retail include provision of savings and transactional accounts, mortgages, personal loans, debit cards, and credit cards. Retail banking is also distinguished from investment banking or commercial banking. It may also refer to a division or department of a bank which deals with individual customers.

In the U.S., the term commercial bank is used for a normal bank to distinguish it from an investment bank.

Typical retail banking services offered by banks include:

* Current accounts
* Savings accounts
* Debit cards, ATM Cards and Credit cards
* Traveler's cheques
* Mortgages
* Home equity loans
* Personal loans
* Certificates of deposit/Term deposits
* Investment plan

A **Personal Equity Plan (PEP)** was a form of tax-privileged investment account introduced in the U.K. that allowed people over the age of 18 to invest in shares of British companies. Investors received both income and capital gains free of tax. Shares held in a PEP are held on trust for investors, who retain the beneficial ownership of the shares. While they are held in the PEP no income tax is payable on dividends, and no capital gains tax on disposals.

Today, marketing and analytics plays a major role in selling a product. Banking industry are the leading firms to apply various techniques in analytics to succeed in selling their product. Bank uses its customer profile and their banking history to target a specific group of people who can potentially buy a product. Based on the response, Banks can refine the customer profile models to improve their targeted outreach and yield better response. Market segmentation involves grouping your various customers into segments that have common needs or will respond similarly to a marketing action. Each segment will respond to a different marketing mix strategy, with each offering alternate growth and profit opportunities. Consists of a group of customers who share a similar set of needs and wants.



For a successful marketing, it needs the following

* Ensure the product is ready and it meets the customer demand
* Price the product according to the customer needs and competitive in the market
* Place ads, brochures and information to generate interests
* How, when, where and whom needs to be contacted for effective marketing

From the customer point of view, making an investment plan involves more than just choosing a few stocks to put money in. it depends on the financial situation and future goals of the customer.

## **Analysis and Models**

### **About the data**

**Table 1.1** represents the sample data collected in a banking industry which is used for advertising customers about one of the banking products called Personal Equity Plan (PEP). This dataset contains demographic information about the customers who were reached about the plan along with the action resulted from this communication. The action includes whether the customer enrolled in the PEP plan or not.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| id | age | sex | region | income | married | children | car | save\_act | current\_act | mortgage | pep |
| ID12101 | 48 | FEMALE | INNER\_CITY | 17546 | NO | 1 | NO | NO | NO | NO | YES |
| ID12102 | 40 | MALE | TOWN | 30085.1 | YES | 3 | YES | NO | YES | YES | NO |
| ID12103 | 51 | FEMALE | INNER\_CITY | 16575.4 | YES | 0 | YES | YES | YES | NO | NO |
| ID12104 | 23 | FEMALE | TOWN | 20375.4 | YES | 3 | NO | NO | YES | NO | NO |
| ID12105 | 57 | FEMALE | RURAL | 50576.3 | YES | 0 | NO | YES | NO | NO | NO |
| ID12106 | 57 | FEMALE | TOWN | 37869.6 | YES | 2 | NO | YES | YES | NO | YES |
| ID12107 | 22 | MALE | RURAL | 8877.07 | NO | 0 | NO | NO | YES | NO | YES |
| ID12108 | 58 | MALE | TOWN | 24946.6 | YES | 0 | YES | YES | YES | NO | NO |
| ID12109 | 37 | FEMALE | SUBURBAN | 25304.3 | YES | 2 | YES | NO | NO | NO | NO |
| ID12110 | 54 | MALE | TOWN | 24212.1 | YES | 2 | YES | YES | YES | NO | NO |
| ID12111 | 66 | FEMALE | TOWN | 59803.9 | YES | 0 | NO | YES | YES | NO | NO |
| ID12112 | 52 | FEMALE | INNER\_CITY | 26658.8 | NO | 0 | YES | YES | YES | YES | NO |
| ID12113 | 44 | FEMALE | TOWN | 15735.8 | YES | 1 | NO | YES | YES | YES | YES |
| ID12114 | 66 | FEMALE | TOWN | 55204.7 | YES | 1 | YES | YES | YES | YES | YES |
| ID12115 | 36 | MALE | RURAL | 19474.6 | YES | 0 | NO | YES | YES | YES | NO |
| ID12116 | 38 | FEMALE | INNER\_CITY | 22342.1 | YES | 0 | YES | YES | YES | YES | NO |

**Table 1.1 (Customer demographics and PEP action)**

#### **Demographic Information**

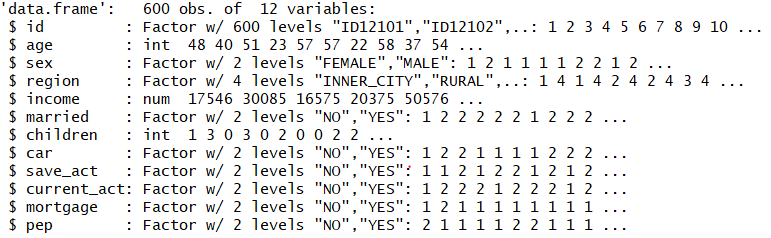
The data contains of a number of the following fields

|  |  |
| --- | --- |
| **Name** | **Description** |
| 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\_acct | Does the customer have a saving account (YES/NO) |
| current\_acct | 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) |

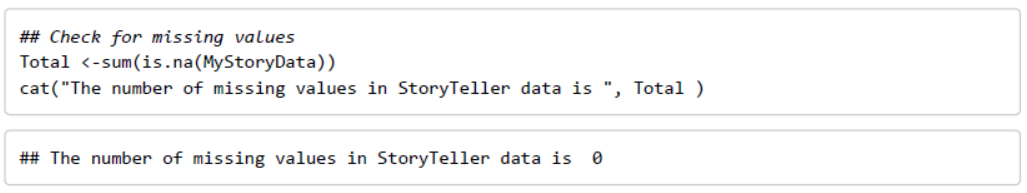
**Table 2.2**

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

#### **Structure of the original Dataset**



#### **Missing Values Check**

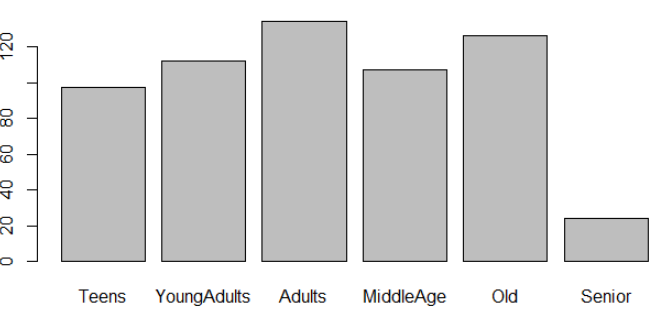


#### **Discretization**

Discretization of continuous variable such as age and income are required to perform further analysis.

Age is discretized into 6 buckets as shown in Fig 1.1

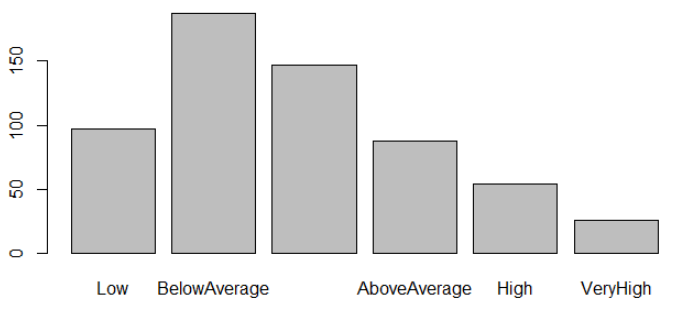
MyStoryData$age\_Discrete <- cut(MyStoryData$age, breaks = c(0,25,35,45,55,65,Inf),labels=c("Teens","YoungAdults","Adults","MiddleAge","Old","Senior"))





**Fig 1.1**

Income is discretized into 6 bins of equal width as shown in Fig 1.2

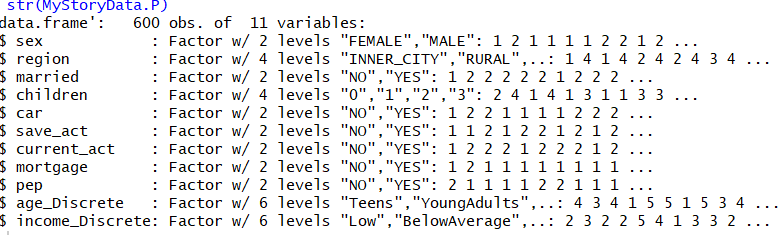




**Fig 1.2**

#### **Structure of the converted dataset**

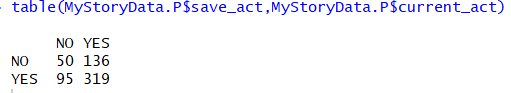
After discretization and required data type conversion, the final structure of the data consists of all factor variable suitable for performing association rule



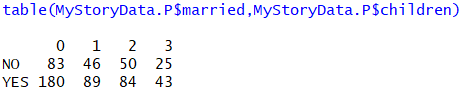
#### **Observation on some of the data distribution**

Frequency distribution of related variable are analyzed to see if they can be grouped together to reduce number of rules available for further analysis

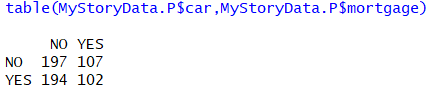
* Current and savings account



* Married and number of children



* Owning a Car and mortgage



### **Models**

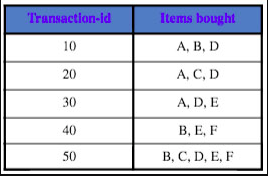
Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. In contrast with sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions.

#### **Algorithm**

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets if those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database:

To select interesting rules from the set of all possible rules, constraints on various measures of significance and interest are used. The best-known constraints are minimum thresholds on support and confidence.

In the below sample dataset, the support confidence and the lift measures for the rule {A} ==> {D} is defined as follows



**Table 2.3**

**Support**

Support is an indication of how frequently the itemset appears in the dataset.

{A} ==> {D}

Support: P(AD) / N= 3/5 = 0.6

{D} ==> {A}

Support: P(DA) / N= 3/5 = 0.6

**Confidence**

Confidence is an indication of how often the rule has been found to be true.

{A} ==> {D}

Confidence: P(AD)/P(A) = (3/5)/(3/5) =0.6/(0.6)= 1

{D} ==> {A}

Confidence: P(DA)/P(D) = (3/5)/(4/5) =0.6/(0.8)= 0.75

**Lift**

The lift of a rule is defined as the ratio of the observed support to that expected if A and B were independent.

{A} ==> {D}

Lift: P(AD)/P(A)P(D) = (3/5)/(3/5)\*(4/5) =0.6/(0.6\*0.8)= 1.25

{D} ==> {A}

Lift: P(DA)/P(D)P(A) = (3/5)/(3/5)\*(4/5) =0.6/(0.6\*0.8)= 1.25

If the rule had a lift of 1, it would imply that the probability of occurrence of the antecedent and that of the consequent are independent of each other. When two events are independent of each other, no rule can be drawn involving those two events.

If the lift is > 1, then the occurrences are dependent on one another, and makes those rules potentially useful for predicting the consequent in future data sets.

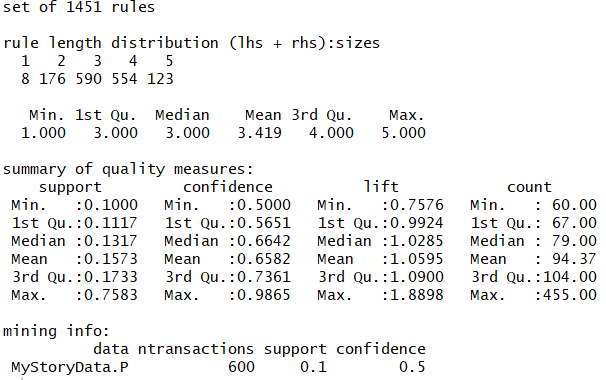
If the lift is < 1, then the items are substitute to each other. This means that presence of one item has negative effect on presence of another item and vice versa.

The value of lift is that it considers both the support of the rule and the overall data set.

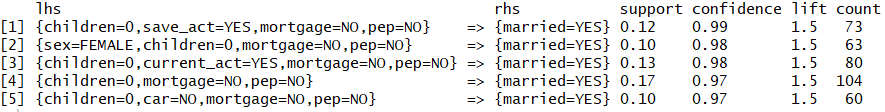
#### **Model 1: List all rules from original dataset with support 0.1, confidence 0.5**

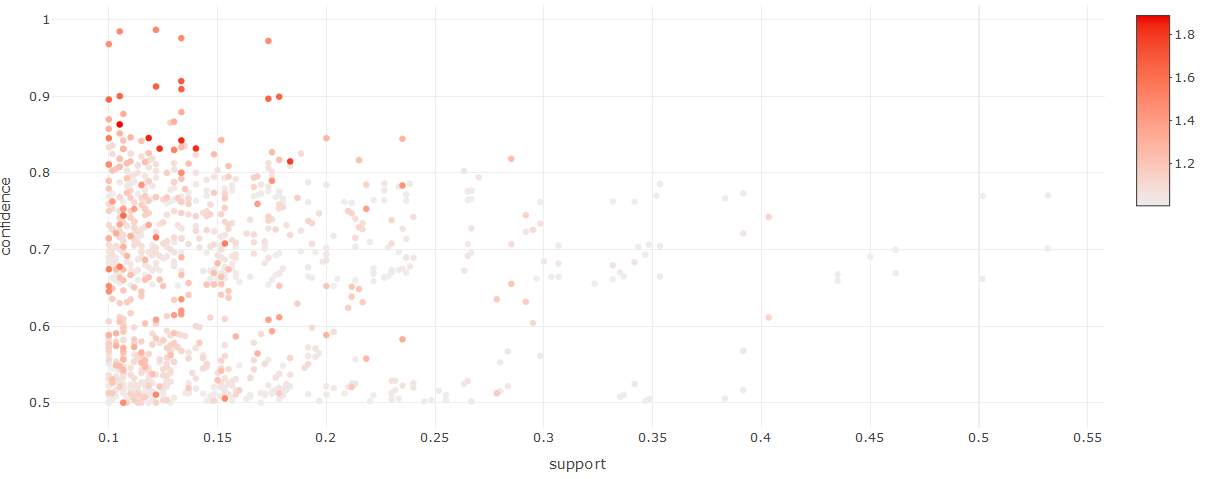
Initial analysis on association rule using all variable with a support of 0.1 and confidence 0.5 is generated. A total of 1451 rules are observed. Fig 2.1 showing scatter plot of all 1451 rules with support in x-axis and confidence in y-axis. The shade of the color is darker for higher lift and lighter for lower lift.

myRules = apriori(MyStoryData.P, parameter = list(supp = 0.1, conf = 0.5, maxlen =6))



**Inspect first 5 rules**





**Fig 2.1 Model-1 scatter plot**

#### **Model 2: All rules from transformed dataset with support 0.1, confidence 0.5**

To reduce the number of rules, transformations are applied on married and children attribute to merge them into a single attribute. similarly, the savings\_act and current\_act are merged into one attribute.

**Transformation**

MyStoryData.P$married\_children <- paste("Married=",MyStoryData.P$married ,";Children=", ifelse(as.integer(as.character(MyStoryData.P$children))>0,"YES","NO"))

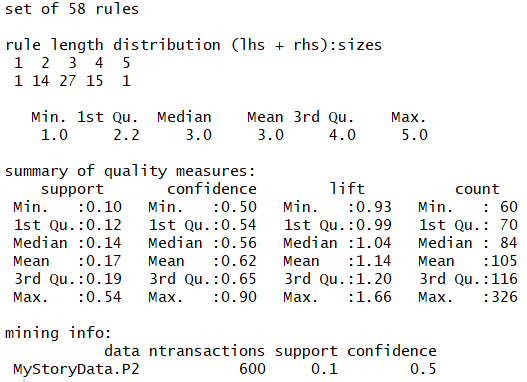
MyStoryData.P$Account <- paste(ifelse(paste(MyStoryData.P$save\_act, MyStoryData.P$current\_act)=="NO NO","NO","YES"))

Apriori algorithms with different parameters are applied on the transformed dataset and the observations are noted

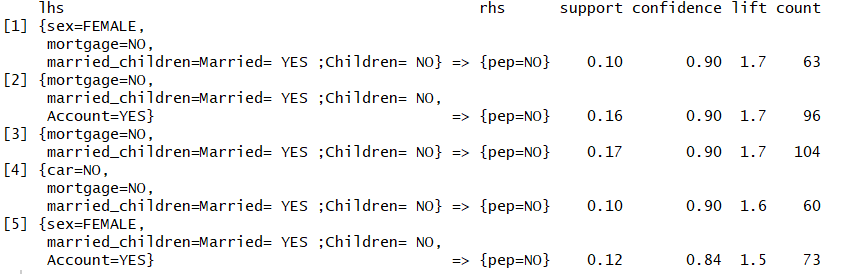
1. All the rules with a support of more than 0.1, confidence more than 0.5 and pep=NO on the RHS are observed. The data is visualized using scatter plot (Fig 1.4) and graph (Fig 1.5) to show the relations of each variable with respect to each other

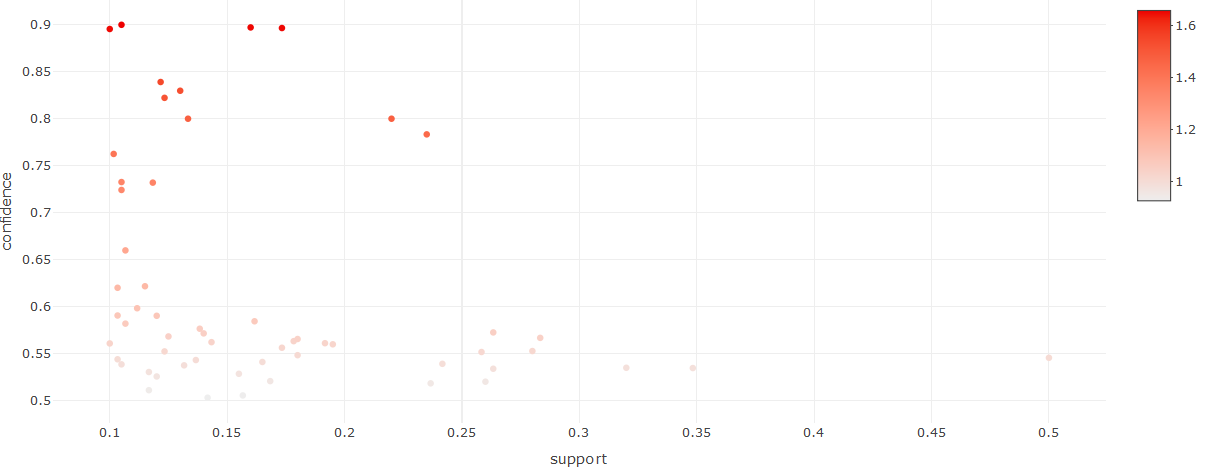
myRules = apriori(MyStoryData.P2, parameter = list(supp = 0.1, conf = 0.5, maxlen =6),appearance = list(rhs = c("pep=NO")))

**Summary**

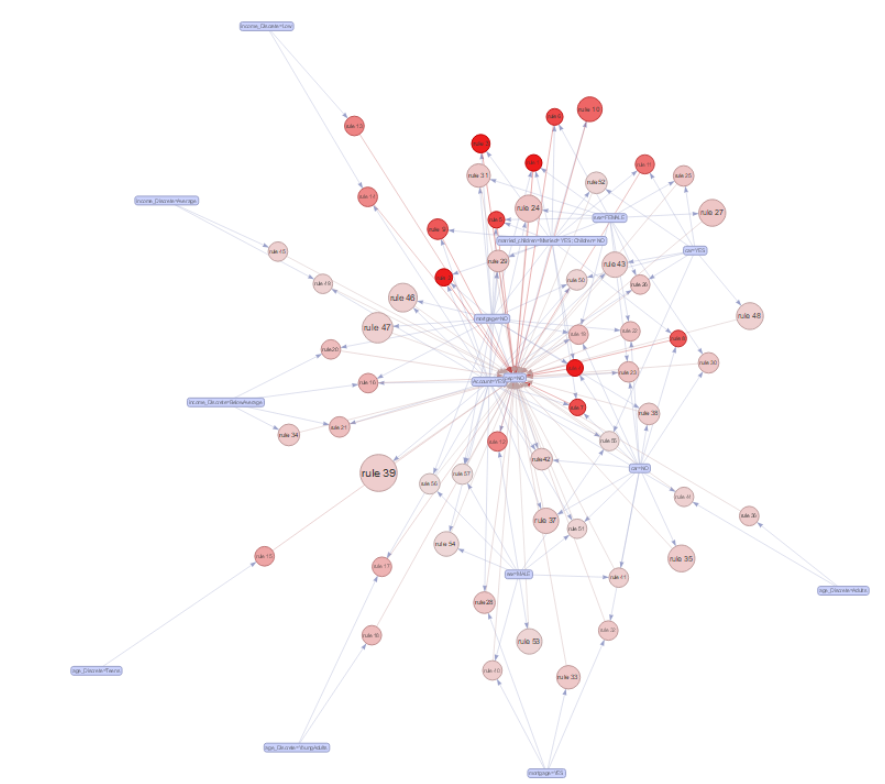


**Top 5 rules**





**Fig 1.4 Model II scatter plot of pep=NO on RHS**

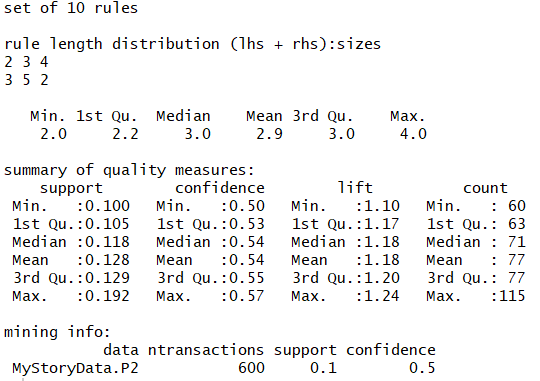


**Fig 1.5 Model II Graph of rule relations with pep=NO on RHS**

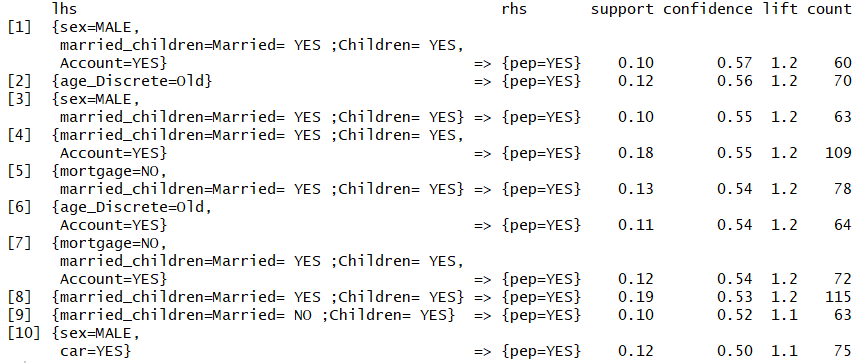
1. All the rules with a support of more than 0.1, confidence more than 0.5 and pep=YES on the RHS are observed. The data is visualized using scatter plot (Fig 1.6) and graph (Fig 1.7) to show the relations of each variable with respect to each other

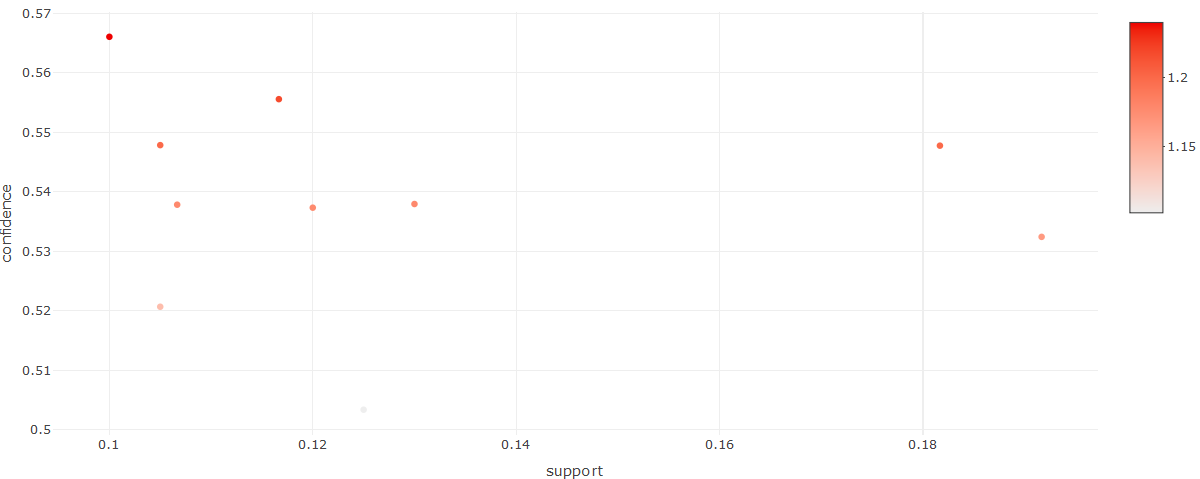
myRules = apriori(MyStoryData.P2, parameter = list(supp = 0.1, conf = 0.5, maxlen =6),appearance = list(rhs = c("pep=YES")))

**Summary**

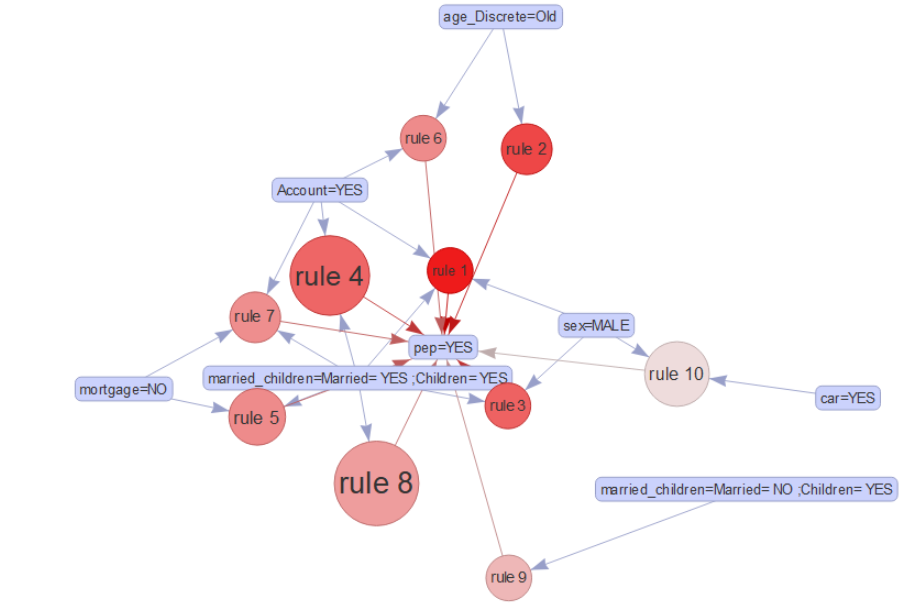


**Top 10 rules**





**Fig 1.6 Model II scatter plot of pep=YES on RHS**

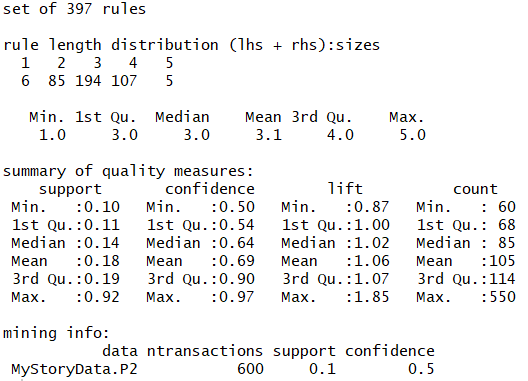


**Fig 1.7 Model II Graph of rule relations with pep=YES on RHS**

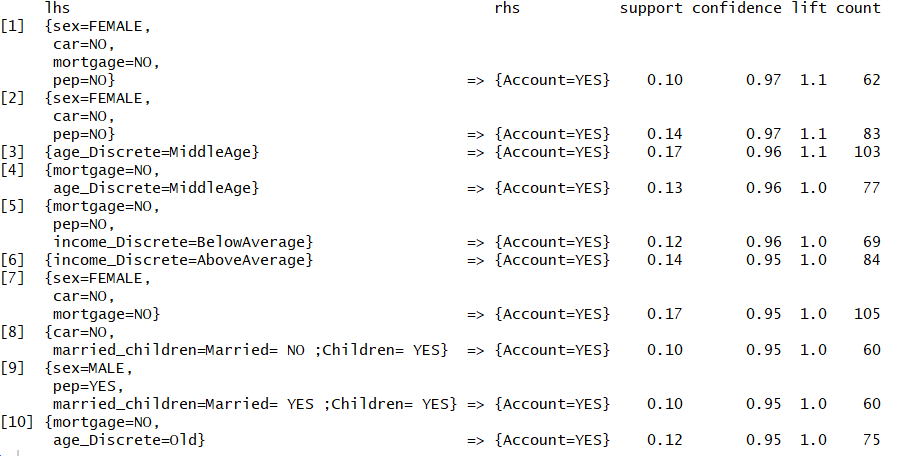
1. All the rules with a support of more than 0.1, confidence more than 0.5 and without any restriction on the RHS are observed. The data is visualized using scatter plot (Fig 1.8) and graph (Fig 1.9) to show the relations of each variable with respect to each other

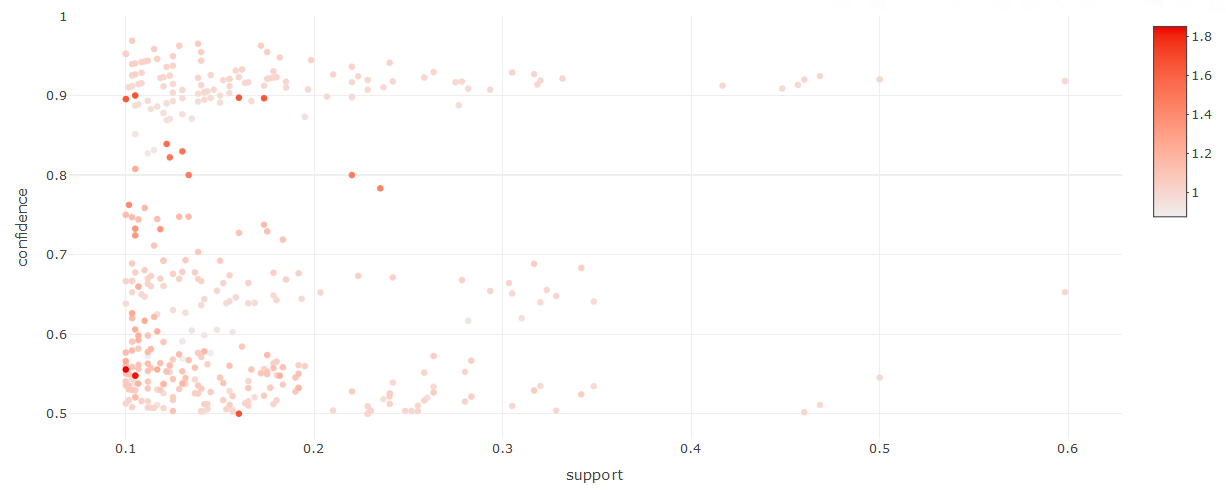
myRules = apriori(MyStoryData.P2, parameter = list(supp = 0.1, conf = 0.5, maxlen =6))

**Summary**

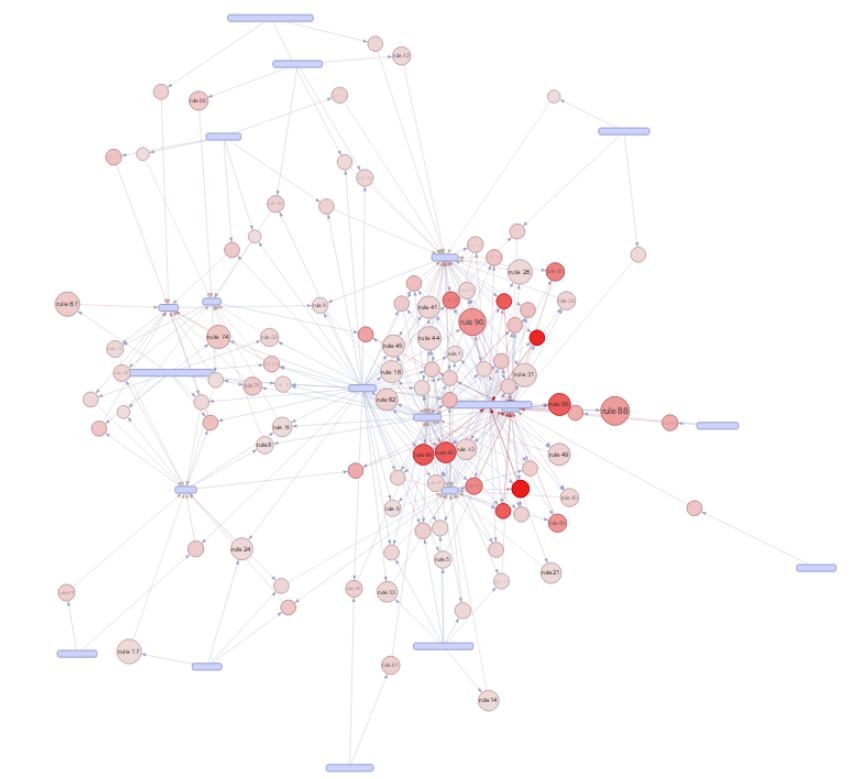


**Top 10 rules**





**Fig 1.8 Model II scatter plot**



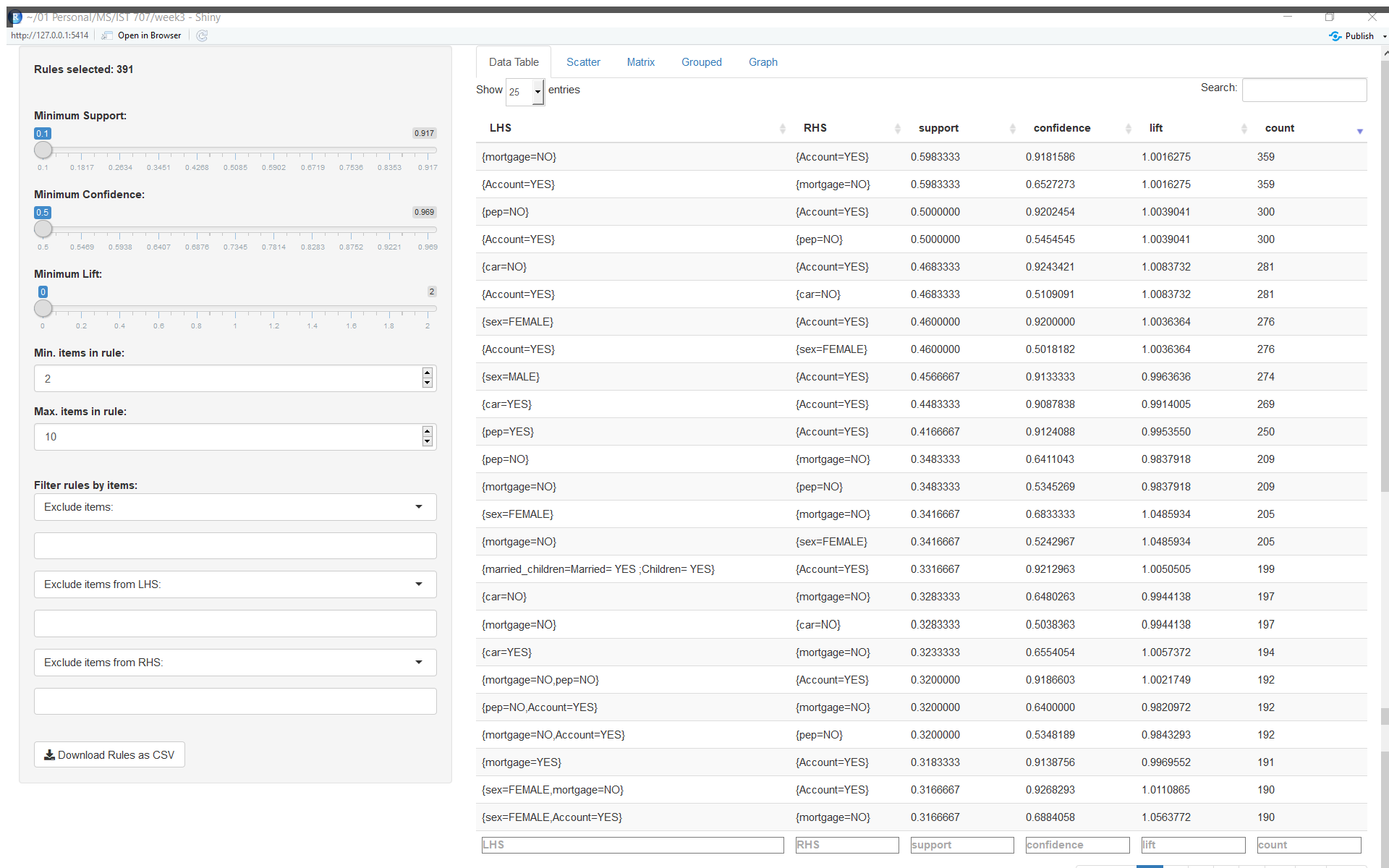
**Fig 1.9 Model II Graph of rule relations**

## **Results**

### **Model II – No RHS restrictions**

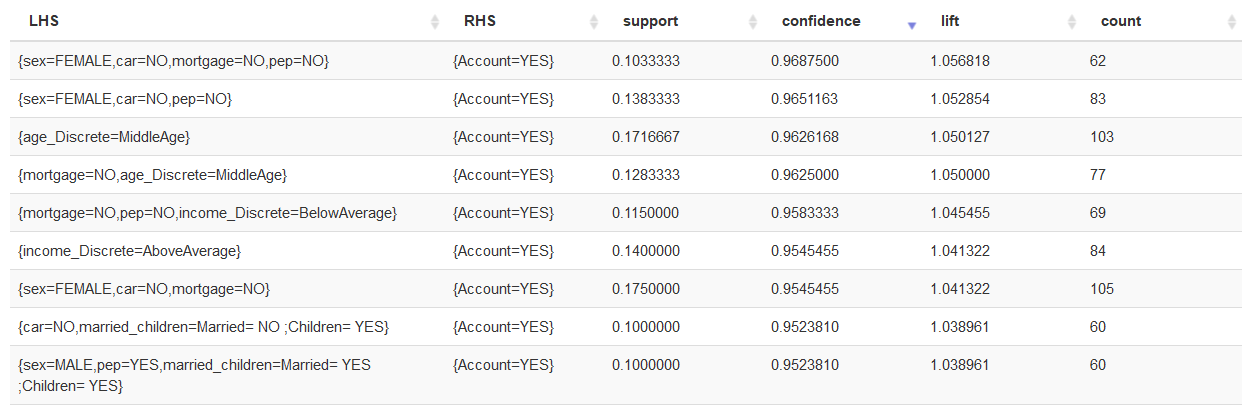
391 rules have been observed in model II with no RHS restrictions. The tables 3.1, 3.2 and 3.3 illustrates the rules with high support, high confidence and high lift respectively.

**Rules with high support**



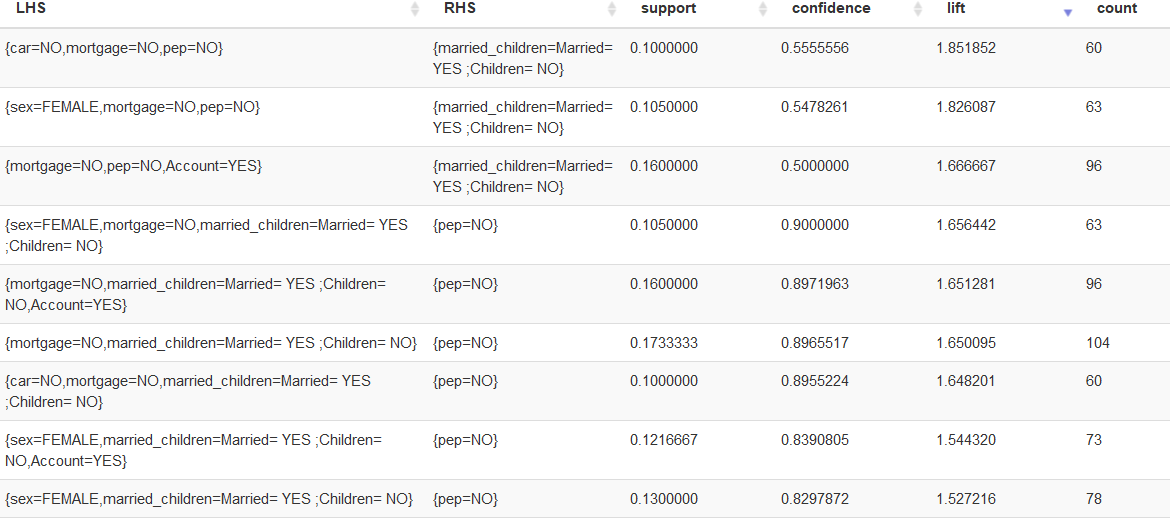
**Table 3.1 Top rules with high support**

**Rules with high confidence**



**Table 3.2 Top rules with high confidence**

**Rules with high lift**



**Table 3.3 Top rules with high lift**

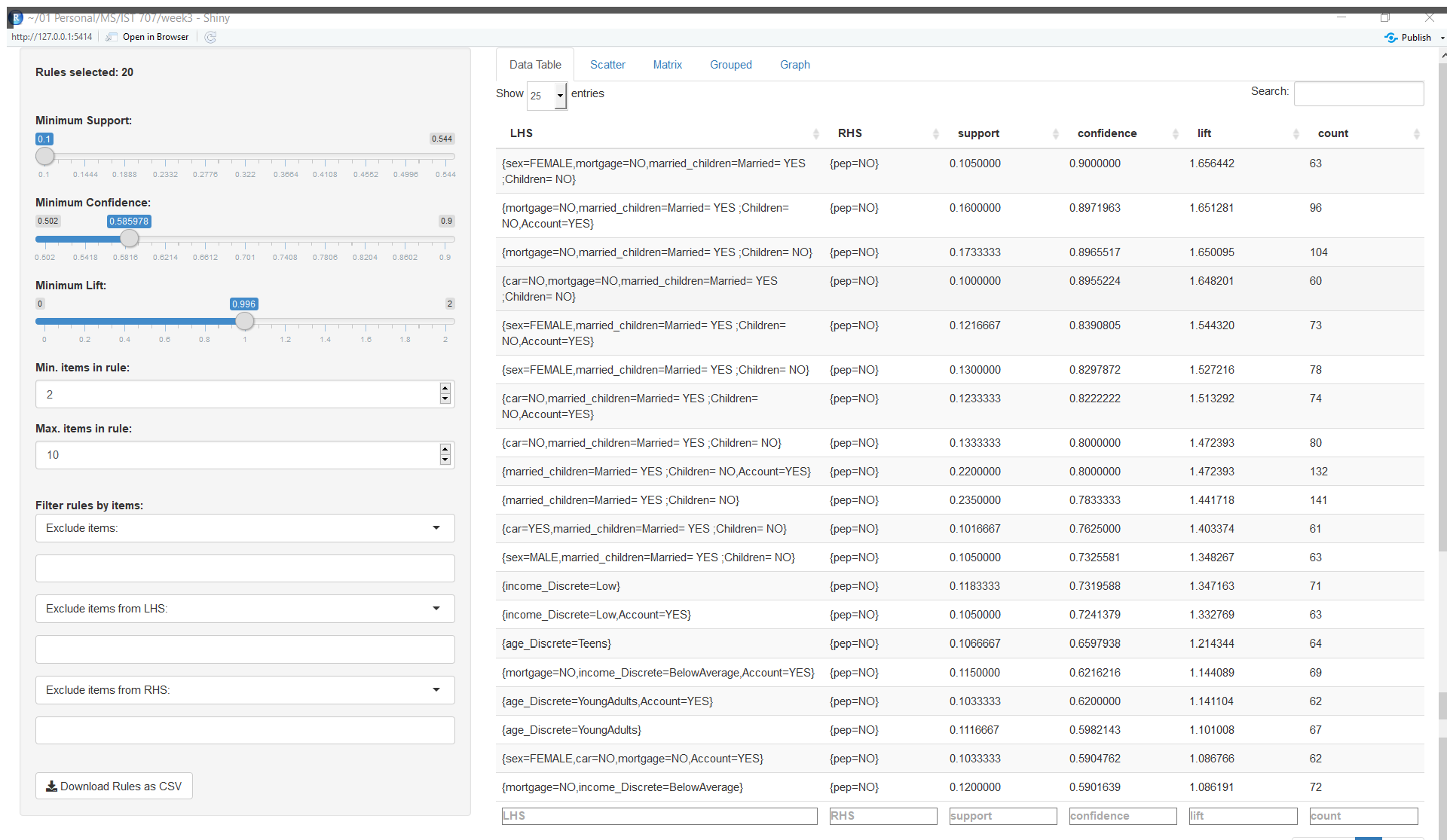
### **Model II – pep=NO in RHS.**

Top 20 rules which are having high confidence in supporting pep=NO is identified and are shown in Table 3.4.

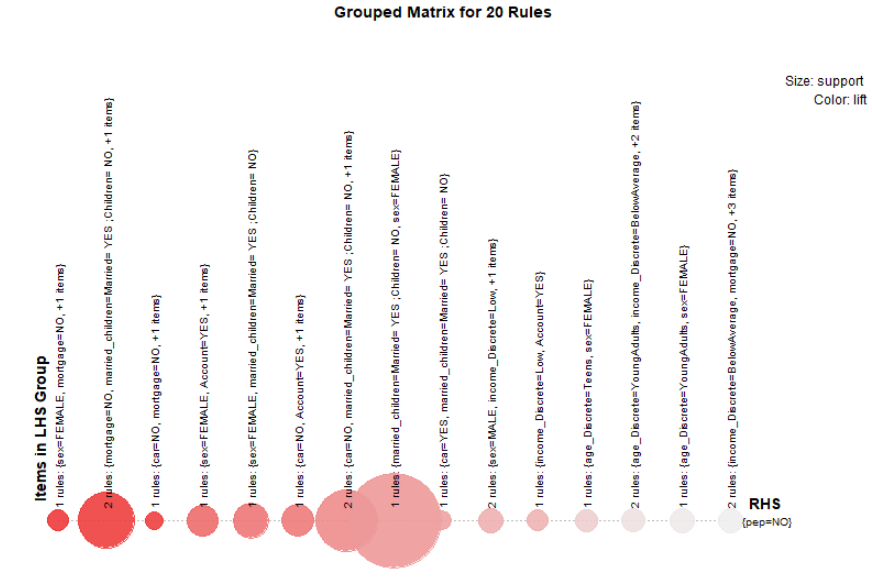
Fig 3.1 and 3.2 illustrates the relations between all 20 rules visually using a grouped matrix and a network diagram

Here the top most rule is {sex=Female,mortage=NO,married=YES,Children=NO} 🡺 {pep=NO} as shown in Table 3.1 has a support = 0.105, confidence = 0.9 and lift =1.656

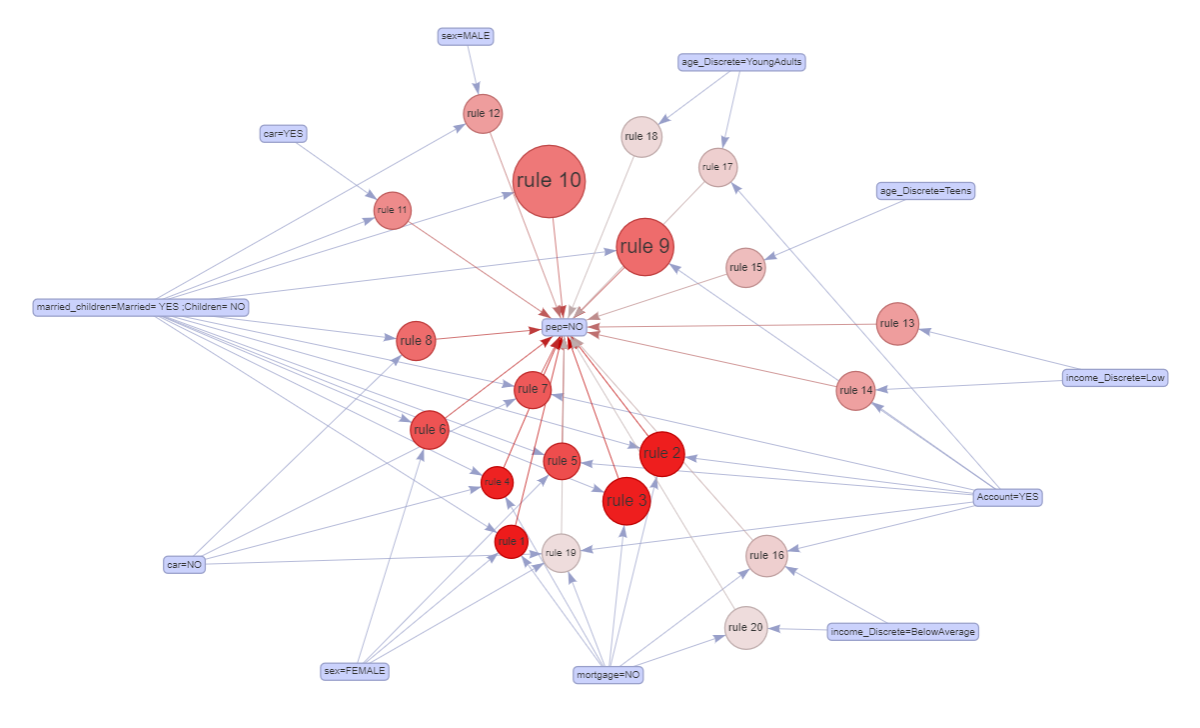
There is a strong relation between the itemset sex=Female,mortgage=NO,married=YES,Children=NO and pep=NO with a high confidence and high lift

****

**Table 3.4 All rules with pep=NO in RHS.**



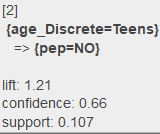
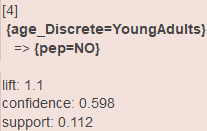
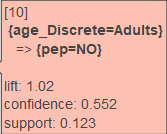
**Fig 3.1 Grouped Matrix with pep=NO in RHS**



**Fig 3.2 Graph of rule relations with pep=NO on RHS**

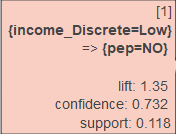
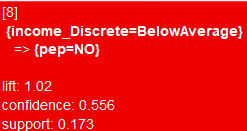
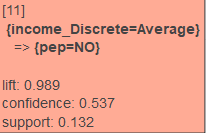
Age Association with pep=NO

Confidence and lift increase with the decrease in age bracket for pep=NO showing negative correlation between Age and pep=No items



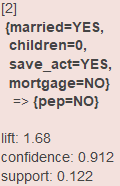
Income Association with pep=NO

Confidence and lift increase with the decrease in Income range for pep=NO showing negative correlation between Income and pep=No items



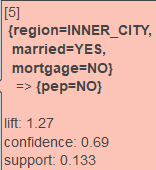
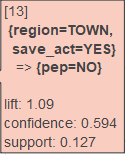
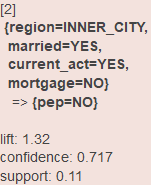
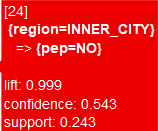
Marital status Association with pep=NO

High confidence and lift for Married=Yes, Children=0, Save\_act=Yes, Mortgage=No for Pep=NO



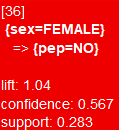
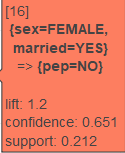
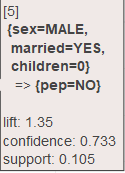
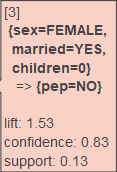
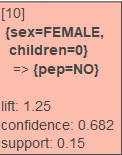
Region Association with pep=NO

Region=INNER\_CITY and region=TOWN showing strong association for pep=NO



Sex Association with pep=NO

Sex=FEMALE showing strong association for pep=NO than sex=MALE

****

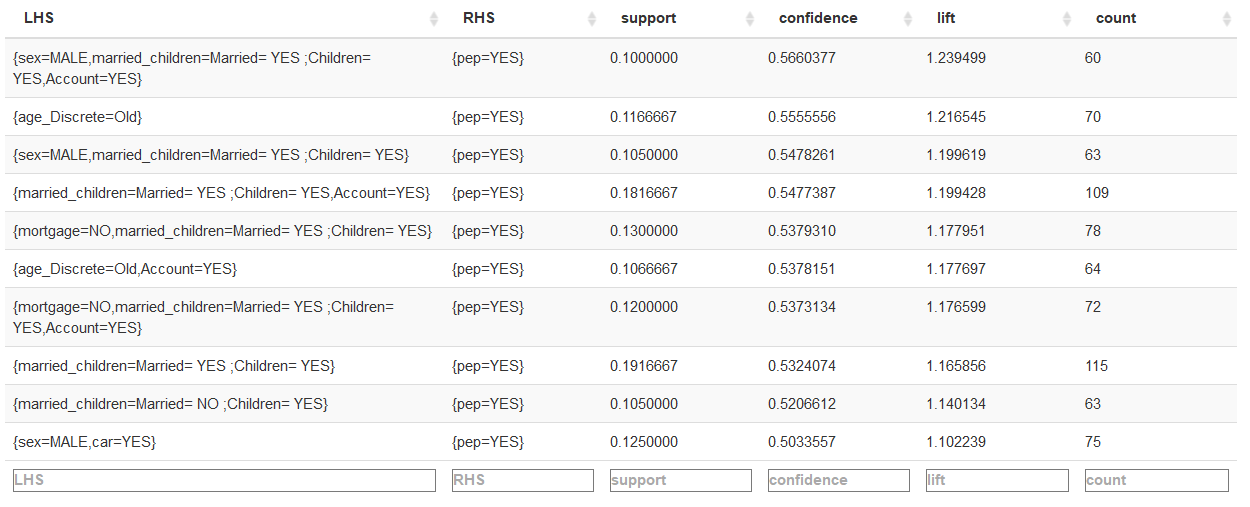
### **Model II – pep=YES in RHS.**

Top 10 rules which are having high confidence in supporting pep=YES is identified and are shown in Table 3.5.

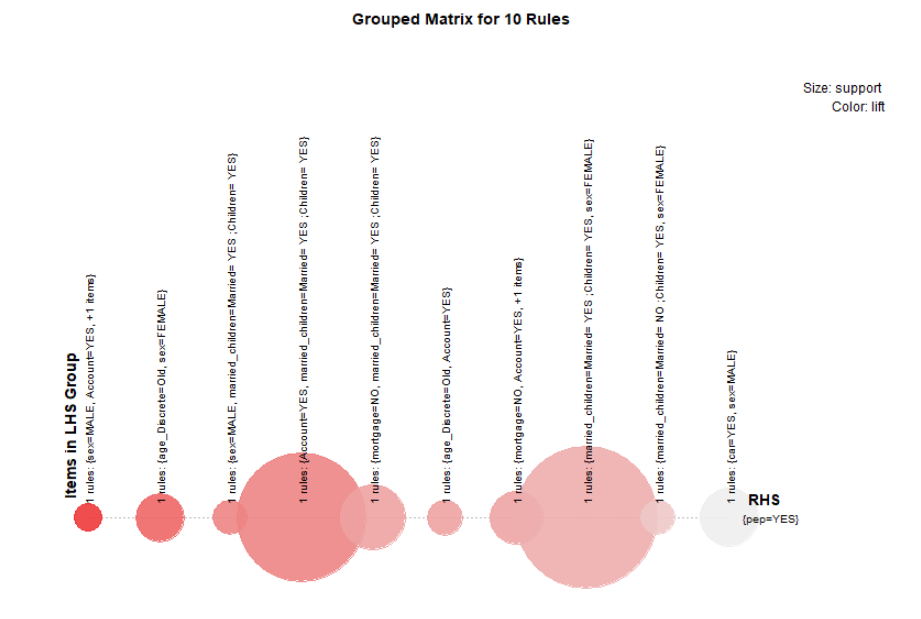
Fig 3.3 and 3.4 illustrates the relations between all 20 rules visually using a grouped matrix and a network diagram

Here the top most rule is {sex=Male,married=YES,Children=YES,Account=YES } 🡺 {pep=YES} as shown in Table 3.5 has a support = 0.1, confidence = 0.566 and lift =1.239

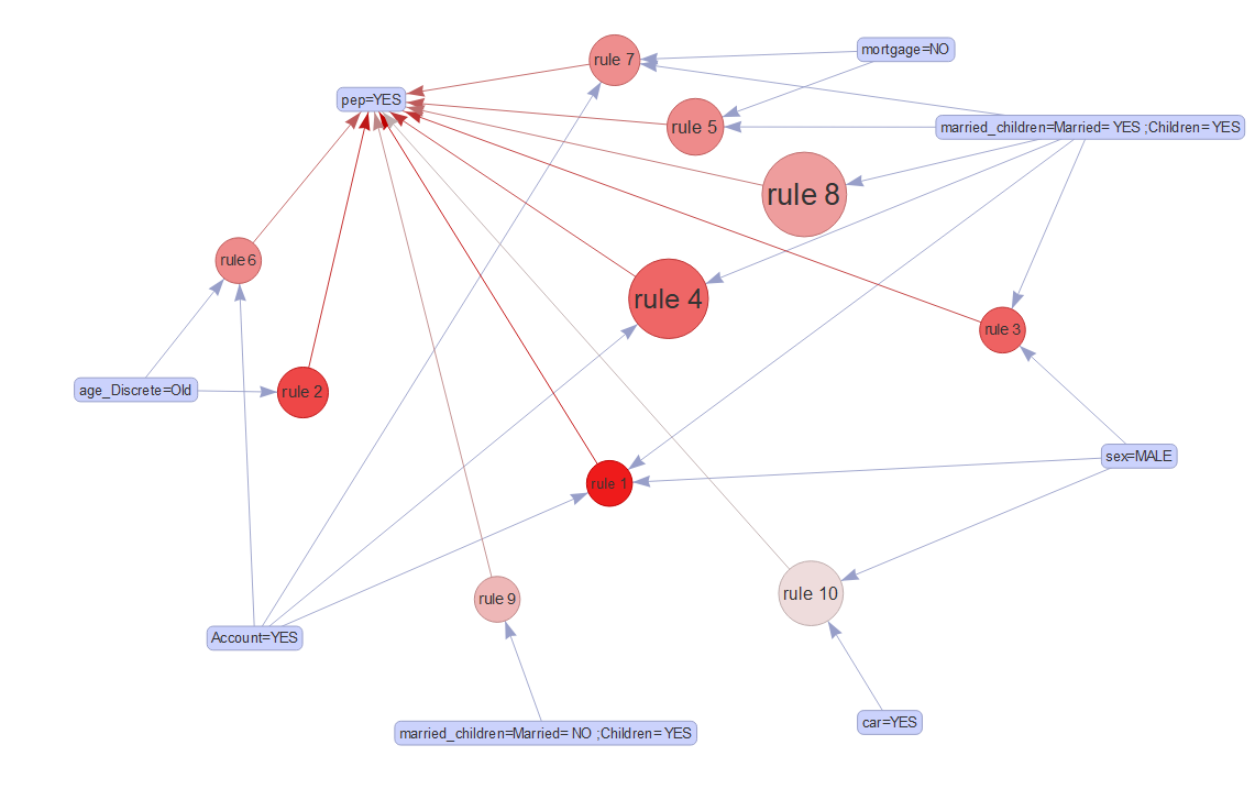
There is a strong relation between the itemset sex=Male,married=YES,Children=YES,Account=YES and pep=YES with a high confidence and high lift



**Table 3.5 All rules with pep=YES in RHS.**



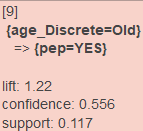
**Fig 3.3 Grouped Matrix with pep=YES in RHS**



**Fig 3.4 Graph of rule relations with pep=YES on RHS**

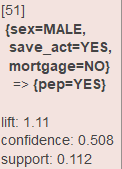
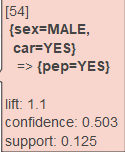
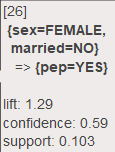
Age Association with pep=YES

Age\_Discrete=Old showing strong association with pep=YES



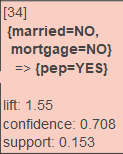
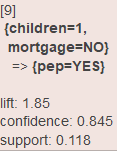
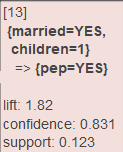
Sex Association with pep=YES

{Sex=Female,Married=No}, {Sex=Male, Mortgage=No,Save\_act=YES}, {Sex=Male, Car=YES} showing strong association with pep=YES

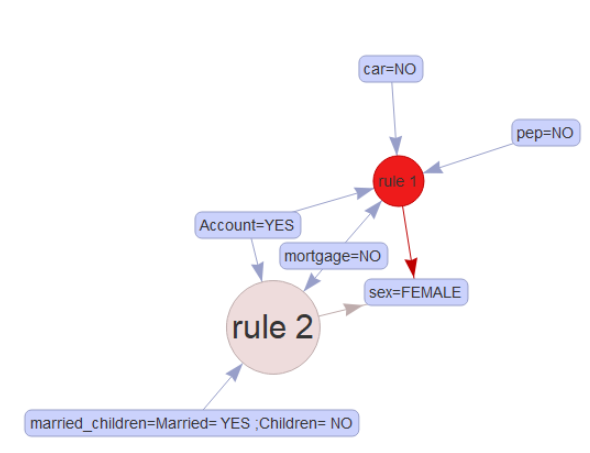
Marital status Association with pep=YES

Confidence and lift increase with married=NO to Children=1 for pep=YES

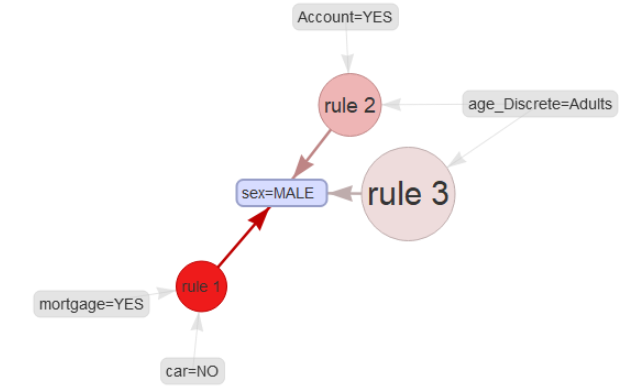
### **Interesting Facts**

Fig 3.5 exhibiting strong association for female with car=No, mortgage=No but with Account=Yes. Also, more confidence and lift has been noted for females with married=Yes, children=No but with Account=Yes



**Fig 3.5**

Fig 3.6 exhibiting strong association for sex=male with car=No, mortgage=Yes but with Account=Yes. Also, more confidence and lift has been noted for sex=male with age\_Discrete=Adult and with Account=YES



**Fig 3.6**

## **Conclusion**

Based on the above findings, bank can utilize the following inferences to model its customer profile and outperform the targeted outreach for marketing a new banking product to its customer.

**Segmentation**-

1. **Determine which kinds of customers exist.**

Demographic

* Teens and Young Adults whose ages are in between 18 to 35
* Middle Aged and Old customer whose ages are in between 35 to 55
* Male and Female customers
* Low to average income customers
* Average to high-income customers
* Customers who has checking or savings account

Geographic

* Customers living in inner city and town
* Customers living in other parts

Behavioral

* Married men with mortgage
* Married men with no mortgage
* Men with no car
* Married men with children
* Married women with children
* Married women with no children
* Married women with no mortgage
* Married women with not current/savings account

1. **Select which ones we are best off trying to serve**

Demographic

* Teens and Young Adults whose age is in between 18 to 35 are less likely to enroll in PEP product and hence a product like an investment plan is not suitable for this group
* Middle aged and Old customers are more likely to enroll in investment plan and can be targeted for future opportunities in selling a product
* Male customers are likely to invest than female customers
* Low income customers are less likely to enroll in the investment plan than the high-income customers

Geographic

* Customers from inner city and town and less likely to enroll in the investment plan

Behavioral

* Married men with children are more likely to enroll in PEP product or any other investment plan whereas married women with children are less likely to enroll in the investment plan

1. **Implement segmentation by optimizing products/services for that segment.**

* Target married women with children who has no account for opening new checking or savings account
* Target men with no cars for car loads
* Target married men with no mortgage for investment plans
* Target married men with account for mortgages
* Target high income and old customers for investment plans