**Introduction**

As banks compete to gain competitive advantage, the need for managing big data and analytics becomes more relevant. Big Data has transformed the way traditional banks worked in the past and has been very helpful in informing decision-making. Through associated big data tools, banks can gain greater visibility into customers’ behaviors, assess the probability of risk and help small businesses. Big Data combines various data sources like the company, its channel partners, customers, suppliers, social media and even external data suppliers. Analytical tools solve this issue of storing, managing and analyzing complex and large data. With its increased accuracy and efficiency, banks are starting to realize Big Data’s value and are slowly adapting to this new change. United Kingdom banks in the 1990s used Big Data in marketing Personal Equity Plans.

A Personal Equity Plan, introduced in 1987, was an investment plan introduced in the United Kingdom that allowed people over the age of 18 to invest in shares of British companies. It was done through an approved plan, qualifying trust, or investment trust. Investors received both income and capital gains free of tax if the invested funds remained in the plan. The plan was designed to encourage investments. Many plans required a minimum investment. PEPs were discontinued in 1999 and replaced by Individual Savings Accounts. There was an annual contribution limit of 6,000 pounds. The investments made under self-select plans were directed by the individual, though a manager or firm was still needed to facilitate the plan, making the plan owner responsible for deciding where their funds should be applied. Managed PEPs, on the other hand, were overseen by a professional manager who put together investment portfolios for the funds.

**About the Data**

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 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\_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)

The data needs to be converted from record to transaction data for analysis. To do so:

* Ages are grouped in to decades.
* Income is put in to 3 bins.
* Children, sex, region, married, car, savings account, current account, mortgage, and pep are converted to factor data points.
* “YES” and “NO” transactions are changed to “[variable\_name]=YES” and “[variable\_name]=NO.”
* Remove id column.

**Analysis**

Sorted Rules by Confidence

|  |
| --- |
| lhs rhs support confidence lift count  [1] {age=teens} => {income=(5.01e+03,2.44e+04]} 0.0617 1 2.1 37  [2] {income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.1333 1 1.4 80  [3] {age=teens,  region=SUBURBAN} => {income=(5.01e+03,2.44e+04]} 0.0067 1 2.1 4  [4] {age=teens,  region=SUBURBAN} => {car=car=NO} 0.0067 1 2.0 4  [5] {age=teens,  region=SUBURBAN} => {current\_act=current\_act=YES} 0.0067 1 1.3 4  [6] {age=teens,  children=3} => {income=(5.01e+03,2.44e+04]} 0.0067 1 2.1 4  [7] {age=teens,  children=3} => {pep=pep=NO} 0.0067 1 1.8 4  [8] {age=teens,  children=3} => {current\_act=current\_act=YES} 0.0067 1 1.3 4  [9] {age=teens,  region=RURAL} => {income=(5.01e+03,2.44e+04]} 0.0117 1 2.1 7  [10] {age=teens,  children=2} => {income=(5.01e+03,2.44e+04]} 0.0200 1 2.1 12  [11] {age=teens,  children=1} => {income=(5.01e+03,2.44e+04]} 0.0133 1 2.1 8  [12] {age=teens,  children=1} => {current\_act=current\_act=YES} 0.0133 1 1.3 8  [13] {age=teens,  current\_act=current\_act=NO} => {income=(5.01e+03,2.44e+04]} 0.0067 1 2.1 4  [14] {age=teens,  current\_act=current\_act=NO} => {car=car=NO} 0.0067 1 2.0 4  [15] {age=teens,  current\_act=current\_act=NO} => {married=married=YES} 0.0067 1 1.5 4  [16] {age=teens,  region=TOWN} => {income=(5.01e+03,2.44e+04]} 0.0150 1 2.1 9  [17] {age=teens,  save\_act=save\_act=NO} => {income=(5.01e+03,2.44e+04]} 0.0267 1 2.1 16  [18] {age=teens,  married=married=NO} => {income=(5.01e+03,2.44e+04]} 0.0200 1 2.1 12  [19] {age=teens,  married=married=NO} => {current\_act=current\_act=YES} 0.0200 1 1.3 12  [20] {age=teens,  mortgage=mortgage=YES} => {income=(5.01e+03,2.44e+04]} 0.0233 1 2.1 14  [21] {age=teens,  children=0} => {income=(5.01e+03,2.44e+04]} 0.0217 1 2.1 13  [22] {age=teens,  region=INNER\_CITY} => {income=(5.01e+03,2.44e+04]} 0.0283 1 2.1 17  [23] {age=teens,  pep=pep=YES} => {income=(5.01e+03,2.44e+04]} 0.0183 1 2.1 11  [24] {age=teens,  car=car=YES} => {income=(5.01e+03,2.44e+04]} 0.0250 1 2.1 15  [25] {age=teens,  sex=MALE} => {income=(5.01e+03,2.44e+04]} 0.0350 1 2.1 21  [26] {age=teens,  sex=FEMALE} => {income=(5.01e+03,2.44e+04]} 0.0267 1 2.1 16  [27] {age=teens,  car=car=NO} => {income=(5.01e+03,2.44e+04]} 0.0367 1 2.1 22  [28] {age=teens,  pep=pep=NO} => {income=(5.01e+03,2.44e+04]} 0.0433 1 2.1 26  [29] {age=teens,  mortgage=mortgage=NO} => {income=(5.01e+03,2.44e+04]} 0.0383 1 2.1 23  [30] {age=teens,  married=married=YES} => {income=(5.01e+03,2.44e+04]} 0.0417 1 2.1 25 |

The rules have high confidence, high lift, and low support. The rules are pretty obvious and give little information.

Sorted Rules by Lift

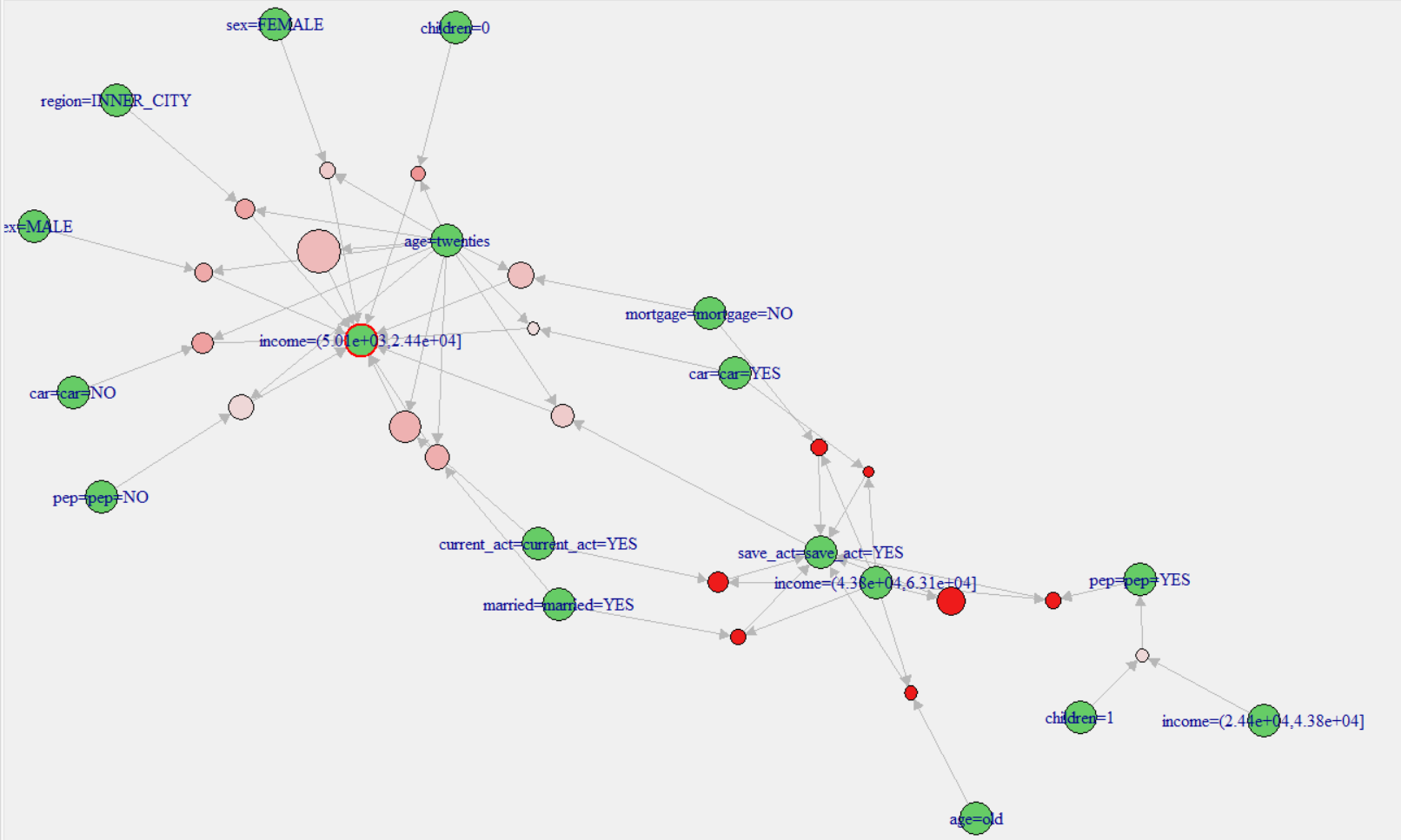
|  |
| --- |
| lhs rhs support confidence lift count  [1] {income=(4.38e+04,6.31e+04],  children=3} => {region=RURAL} 0.0133 1.00 6.2 8  [2] {income=(4.38e+04,6.31e+04],  pep=pep=NO} => {children=0} 0.0417 0.96 2.2 25  [3] {income=(4.38e+04,6.31e+04],  children=3} => {pep=pep=YES} 0.0133 1.00 2.2 8  [4] {income=(4.38e+04,6.31e+04],  children=1} => {pep=pep=YES} 0.0267 1.00 2.2 16  [5] {age=fourties,  children=1} => {pep=pep=YES} 0.0533 1.00 2.2 32  [6] {age=teens} => {income=(5.01e+03,2.44e+04]} 0.0617 1.00 2.1 37  [7] {age=teens,  region=SUBURBAN} => {income=(5.01e+03,2.44e+04]} 0.0067 1.00 2.1 4  [8] {age=teens,  children=3} => {income=(5.01e+03,2.44e+04]} 0.0067 1.00 2.1 4  [9] {age=teens,  region=RURAL} => {income=(5.01e+03,2.44e+04]} 0.0117 1.00 2.1 7  [10] {age=teens,  children=2} => {income=(5.01e+03,2.44e+04]} 0.0200 1.00 2.1 12  [11] {age=teens,  children=1} => {income=(5.01e+03,2.44e+04]} 0.0133 1.00 2.1 8  [12] {age=teens,  current\_act=current\_act=NO} => {income=(5.01e+03,2.44e+04]} 0.0067 1.00 2.1 4  [13] {age=teens,  region=TOWN} => {income=(5.01e+03,2.44e+04]} 0.0150 1.00 2.1 9  [14] {age=teens,  save\_act=save\_act=NO} => {income=(5.01e+03,2.44e+04]} 0.0267 1.00 2.1 16  [15] {age=teens,  married=married=NO} => {income=(5.01e+03,2.44e+04]} 0.0200 1.00 2.1 12  [16] {age=teens,  mortgage=mortgage=YES} => {income=(5.01e+03,2.44e+04]} 0.0233 1.00 2.1 14  [17] {age=teens,  children=0} => {income=(5.01e+03,2.44e+04]} 0.0217 1.00 2.1 13  [18] {age=teens,  region=INNER\_CITY} => {income=(5.01e+03,2.44e+04]} 0.0283 1.00 2.1 17  [19] {age=teens,  pep=pep=YES} => {income=(5.01e+03,2.44e+04]} 0.0183 1.00 2.1 11  [20] {age=teens,  car=car=YES} => {income=(5.01e+03,2.44e+04]} 0.0250 1.00 2.1 15  [21] {age=teens,  sex=MALE} => {income=(5.01e+03,2.44e+04]} 0.0350 1.00 2.1 21  [22] {age=teens,  sex=FEMALE} => {income=(5.01e+03,2.44e+04]} 0.0267 1.00 2.1 16  [23] {age=teens,  car=car=NO} => {income=(5.01e+03,2.44e+04]} 0.0367 1.00 2.1 22  [24] {age=teens,  pep=pep=NO} => {income=(5.01e+03,2.44e+04]} 0.0433 1.00 2.1 26  [25] {age=teens,  mortgage=mortgage=NO} => {income=(5.01e+03,2.44e+04]} 0.0383 1.00 2.1 23  [26] {age=teens,  married=married=YES} => {income=(5.01e+03,2.44e+04]} 0.0417 1.00 2.1 25  [27] {age=teens,  save\_act=save\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.0350 1.00 2.1 21  [28] {age=teens,  current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.0550 1.00 2.1 33  [29] {age=twenties,  region=RURAL} => {income=(5.01e+03,2.44e+04]} 0.0267 1.00 2.1 16  [30] {age=twenties,  children=1} => {income=(5.01e+03,2.44e+04]} 0.0400 1.00 2.1 24 |
|  |
| |  | | --- | | > | |

The rules have high lift, high confidence, and low support. These rules give very little useful information.

Rules Sorted by Support

|  |
| --- |
| lhs rhs support confidence lift count  [1] {age=twenties} => {income=(5.01e+03,2.44e+04]} 0.187 0.94 2.0 112  [2] {age=twenties,  current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.145 0.95 2.0 87  [3] {income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.133 1.00 1.4 80  [4] {age=twenties,  mortgage=mortgage=NO} => {income=(5.01e+03,2.44e+04]} 0.127 0.94 2.0 76  [5] {age=twenties,  pep=pep=NO} => {income=(5.01e+03,2.44e+04]} 0.122 0.92 2.0 73  [6] {age=twenties,  married=married=YES} => {income=(5.01e+03,2.44e+04]} 0.122 0.95 2.0 73  [7] {age=twenties,  save\_act=save\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.113 0.93 2.0 68  [8] {age=twenties,  car=car=NO} => {income=(5.01e+03,2.44e+04]} 0.108 0.96 2.0 65  [9] {income=(4.38e+04,6.31e+04],  current\_act=current\_act=YES} => {save\_act=save\_act=YES} 0.105 1.00 1.4 63  [10] {age=twenties,  region=INNER\_CITY} => {income=(5.01e+03,2.44e+04]} 0.103 0.95 2.0 62  [11] {age=twenties,  sex=MALE} => {income=(5.01e+03,2.44e+04]} 0.097 0.95 2.0 58  [12] {income=(4.38e+04,6.31e+04],  mortgage=mortgage=NO} => {save\_act=save\_act=YES} 0.092 1.00 1.4 55  [13] {income=(4.38e+04,6.31e+04],  pep=pep=YES} => {save\_act=save\_act=YES} 0.090 1.00 1.4 54  [14] {age=twenties,  sex=FEMALE} => {income=(5.01e+03,2.44e+04]} 0.090 0.93 2.0 54  [15] {income=(4.38e+04,6.31e+04],  married=married=YES} => {save\_act=save\_act=YES} 0.088 1.00 1.4 53  [16] {age=twenties,  children=0} => {income=(5.01e+03,2.44e+04]} 0.085 0.96 2.0 51  [17] {age=old,  income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.082 1.00 1.4 49  [18] {income=(2.44e+04,4.38e+04],  children=1} => {pep=pep=YES} 0.082 0.92 2.0 49  [19] {age=twenties,  car=car=YES} => {income=(5.01e+03,2.44e+04]} 0.078 0.92 1.9 47  [20] {income=(4.38e+04,6.31e+04],  car=car=YES} => {save\_act=save\_act=YES} 0.073 1.00 1.4 44  [21] {age=twenties,  save\_act=save\_act=NO} => {income=(5.01e+03,2.44e+04]} 0.073 0.96 2.0 44  [22] {sex=FEMALE,  income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.068 1.00 1.4 41  [23] {sex=MALE,  income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.065 1.00 1.4 39  [24] {age=twenties,  married=married=NO} => {income=(5.01e+03,2.44e+04]} 0.065 0.93 2.0 39  [25] {age=twenties,  pep=pep=YES} => {income=(5.01e+03,2.44e+04]} 0.065 0.97 2.1 39  [26] {age=teens} => {income=(5.01e+03,2.44e+04]} 0.062 1.00 2.1 37  [27] {region=SUBURBAN,  mortgage=mortgage=NO} => {current\_act=current\_act=YES} 0.060 0.90 1.2 36  [28] {income=(4.38e+04,6.31e+04],  car=car=NO} => {save\_act=save\_act=YES} 0.060 1.00 1.4 36  [29] {age=twenties,  mortgage=mortgage=YES} => {income=(5.01e+03,2.44e+04]} 0.060 0.95 2.0 36  [30] {age=teens,  current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.055 1.00 2.1 33 |
|  |
| |  | | --- | | > | |

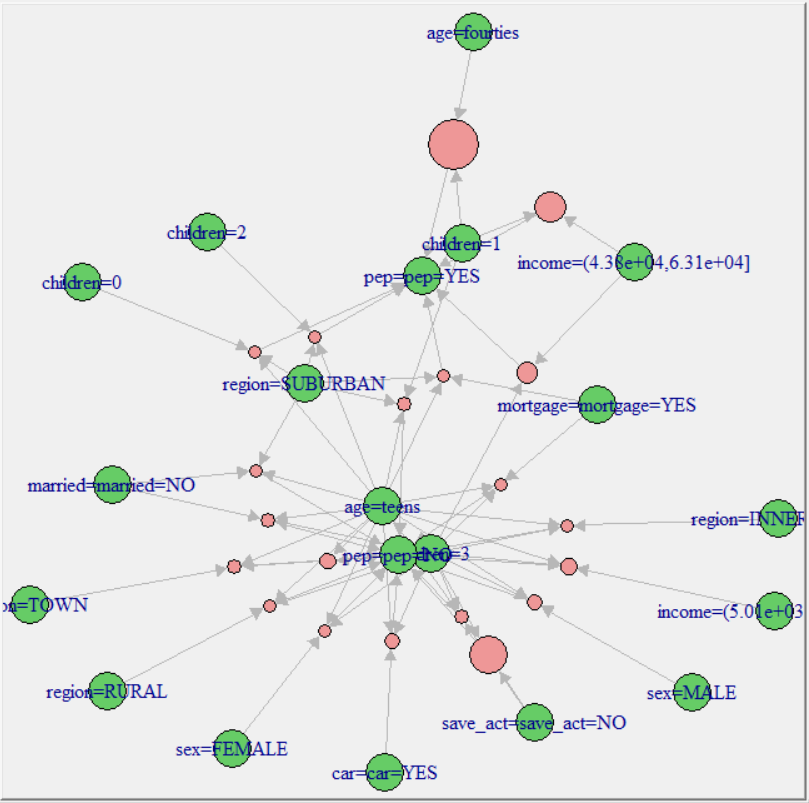
These rules provide more of a variety of information and are a little more useful. The rules have more support, for the most part an acceptable lift and confidence.



PEP Rules Sorted by Confidence

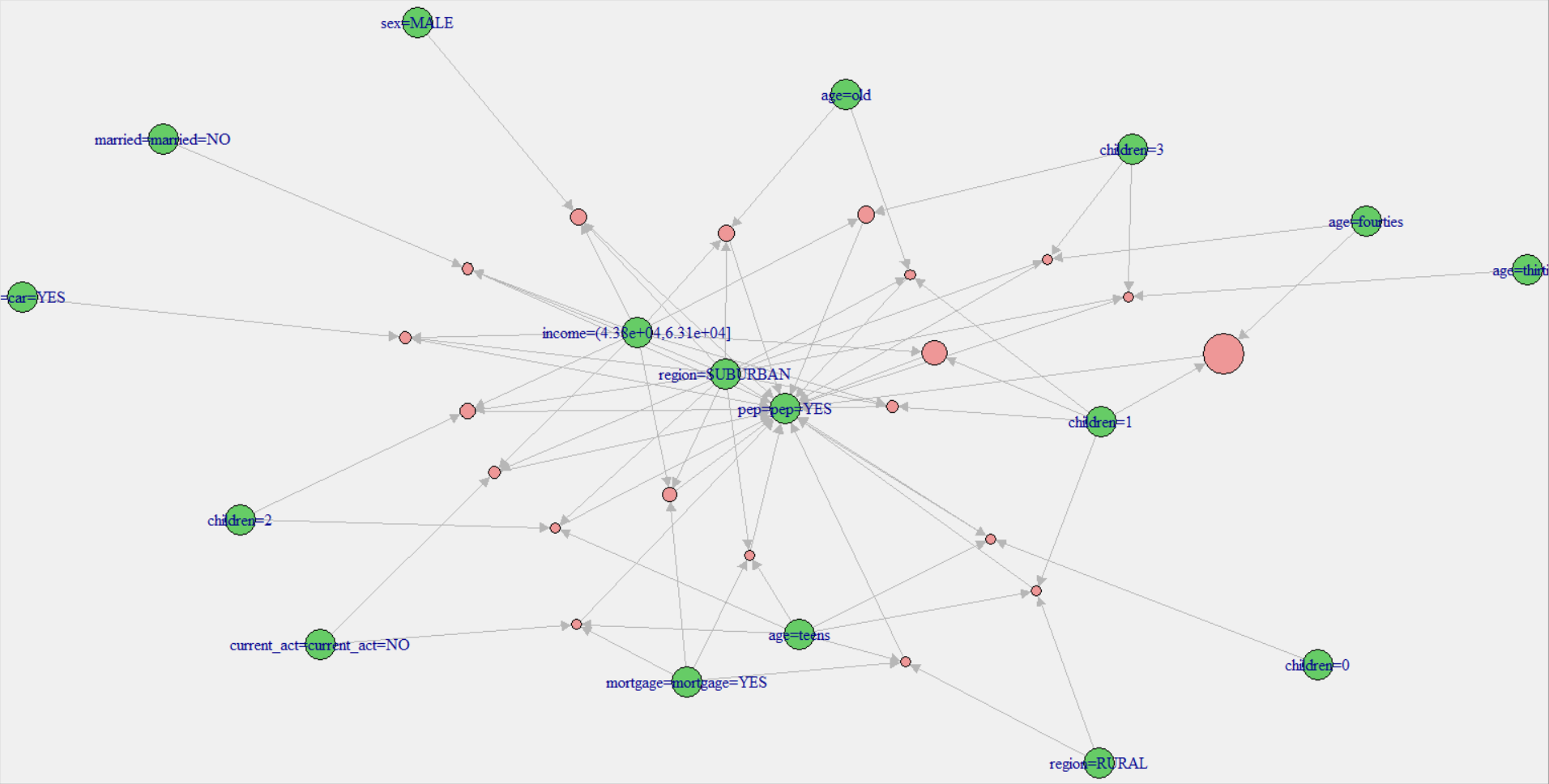
|  |
| --- |
| lhs rhs support confidence lift count  [1] {age=teens,children=3} => {pep=pep=NO} 0.006666667 1 1.840491 4  [2] {income=(4.38e+04,6.31e+04],children=3} => {pep=pep=YES} 0.013333333 1 2.189781 8  [3] {children=3,save\_act=save\_act=NO} => {pep=pep=NO} 0.036666667 1 1.840491 22  [4] {income=(4.38e+04,6.31e+04],children=1} => {pep=pep=YES} 0.026666667 1 2.189781 16  [5] {age=fourties,children=1} => {pep=pep=YES} 0.053333333 1 2.189781 32  [6] {age=teens,region=SUBURBAN,children=2} => {pep=pep=YES} 0.001666667 1 2.189781 1  [7] {age=teens,region=SUBURBAN,children=1} => {pep=pep=NO} 0.003333333 1 1.840491 2  [8] {age=teens,region=SUBURBAN,married=married=NO} => {pep=pep=NO} 0.001666667 1 1.840491 1  [9] {age=teens,region=SUBURBAN,mortgage=mortgage=YES} => {pep=pep=YES} 0.001666667 1 2.189781 1  [10] {age=teens,region=SUBURBAN,children=0} => {pep=pep=YES} 0.001666667 1 2.189781 1 |
|  |
| |  | | --- | |  | |

The rules do not have enough support to be useful.



PEP Rules Sorted by Lift

|  |
| --- |
| lhs rhs support confidence lift count  [1] {income=(4.38e+04,6.31e+04],children=3} => {pep=pep=YES} 0.013333333 1 2.189781 8  [2] {income=(4.38e+04,6.31e+04],children=1} => {pep=pep=YES} 0.026666667 1 2.189781 16  [3] {age=fourties,children=1} => {pep=pep=YES} 0.053333333 1 2.189781 32  [4] {age=teens,region=SUBURBAN,children=2} => {pep=pep=YES} 0.001666667 1 2.189781 1  [5] {age=teens,region=SUBURBAN,mortgage=mortgage=YES} => {pep=pep=YES} 0.001666667 1 2.189781 1  [6] {age=teens,region=SUBURBAN,children=0} => {pep=pep=YES} 0.001666667 1 2.189781 1  [7] {age=teens,region=RURAL,children=1} => {pep=pep=YES} 0.001666667 1 2.189781 1  [8] {age=teens,region=RURAL,mortgage=mortgage=YES} => {pep=pep=YES} 0.001666667 1 2.189781 1  [9] {age=teens,current\_act=current\_act=NO,mortgage=mortgage=YES} => {pep=pep=YES} 0.001666667 1 2.189781 1  [10] {age=thirties,region=SUBURBAN,children=3} => {pep=pep=YES} 0.001666667 1 2.189781 1 |
|  |
| |  | | --- | |  |   The rules have higher lift and adequate confidence but are still low on support. |



PEP Rules Sorted by Support

lhs rhs support confidence lift count

[1] {current\_act=current\_act=YES} => {pep=pep=NO} 0.4066667 0.5362637 0.9869885 244

[2] {married=married=YES} => {pep=pep=NO} 0.4033333 0.6111111 1.1247444 242

[3] {save\_act=save\_act=YES} => {pep=pep=NO} 0.3916667 0.5676329 1.0447230 235

[4] {current\_act=current\_act=YES} => {pep=pep=YES} 0.3516667 0.4637363 1.0154809 211

[5] {mortgage=mortgage=NO} => {pep=pep=NO} 0.3483333 0.5345269 0.9837918 209

[6] {mortgage=mortgage=NO} => {pep=pep=YES} 0.3033333 0.4654731 1.0192843 182

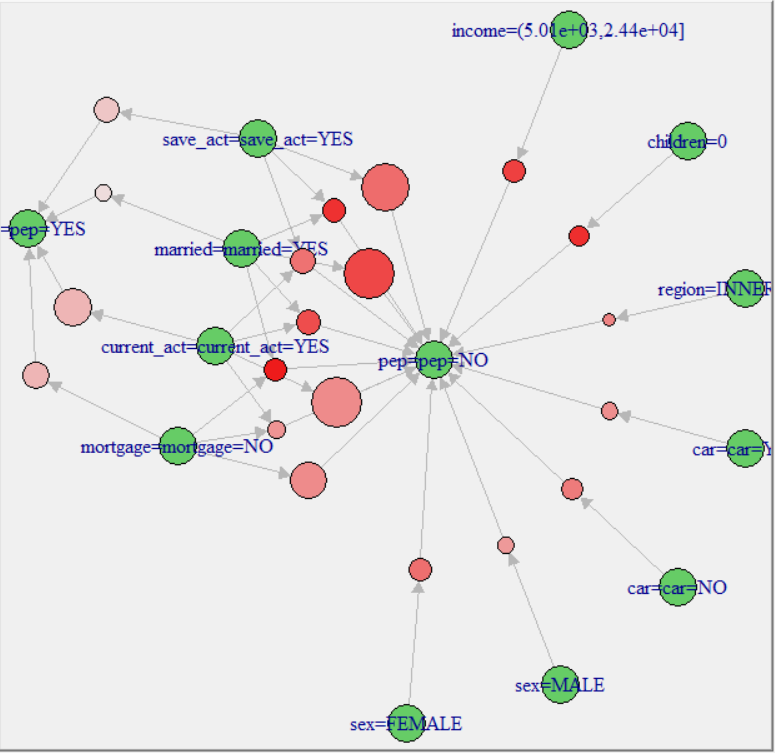
[7] {save\_act=save\_act=YES} => {pep=pep=YES} 0.2983333 0.4323671 0.9467894 179

[8] {save\_act=save\_act=YES,current\_act=current\_act=YES} => {pep=pep=NO} 0.2983333 0.5611285 1.0327519 179

[9] {married=married=YES,current\_act=current\_act=YES} => {pep=pep=NO} 0.2950000 0.6040956 1.1118323 177

[10] {income=(5.01e+03,2.44e+04]} => {pep=pep=NO} 0.2916667 0.6161972 1.1341052 175

The rules seem to be more relevant they have higher support and acceptable lift and confidence.



**Results**

5 Most “Interesting” Rules

1. Married individuals with no children and no mortgage will not have a PEP. This rule has a support of .173, confidence of .897, and lift of 1.65. The married individual would be more likely to get a PEP after they purchase a home.
2. Individuals in their thirties with one child are likely to have a PEP. This rule has a support of .013, confidence of 1, and lift of 2.19. This rule makes sense because the individual has responsibilities of raising a child and needs to invest money for the child’s future.
3. Individuals that are not married, do not have a mortgage, and make higher income are likely to have a PEP account. The rule has a support of .033, confidence of 1, and lift of 2.19. This is intuitive because individuals that make more money will want to save on taxes.
4. Teenagers living in the suburbs with 2 children are likely to have a PEP. This rule may seem surprising but makes sense because the amount to open a PEP is low, so all ages can open the account. A teenage with two children will want to start saving for their children’s future. The rule has a support of .002, confidence of 1, and lift of 2.19.
5. Individuals in their thirties with a car living in the suburbs are not likely to have a PEP account. This rule has a support of .003, confidence of 1, and lift of 1.84.

**Conclusion**

The data is helpful to determine the clientele that opens a PEP account. To do a better analysis the marketing firm should collect more data. Dates, credit rating, vacation plans, and satisfaction with the bank would be helpful to determine how to market to potential clients. The data provides enough information to do a high-level analysis. For the dataset, a confidence level of .75, a lift level of 1, and a support level of .001 are appropriate.

Next steps would be to run k-means cluster and properly segment the customers, determine the segments to target, and finalize the marketing strategy. A proper Marketing ROI analyses is necessary to see how effective the marketing campaign is.