Name: Jie Wang

Email: [JWANG326@Syr.edu](mailto:JWANG326@Syr.edu)

Assignment: HW3

Topic: Explore bank data

Models/methods: Association rule mining

Date: July, 25,2022

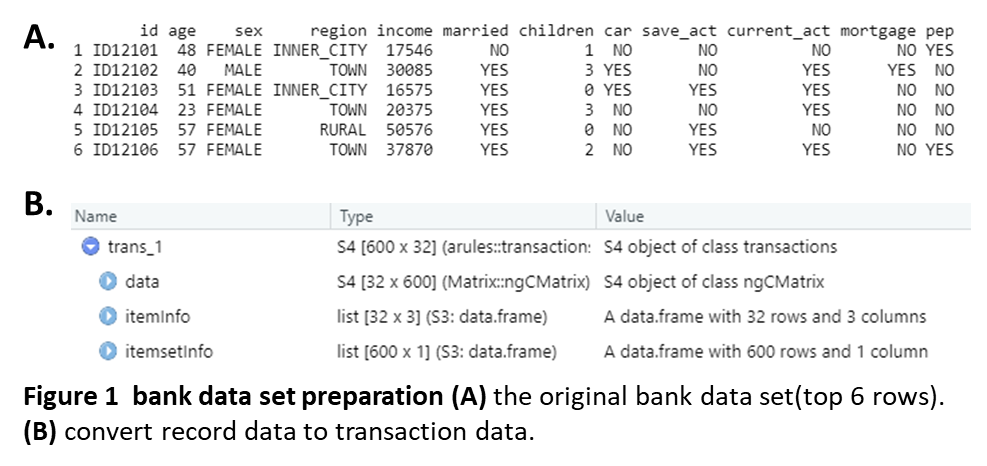
# Explore Bank Data

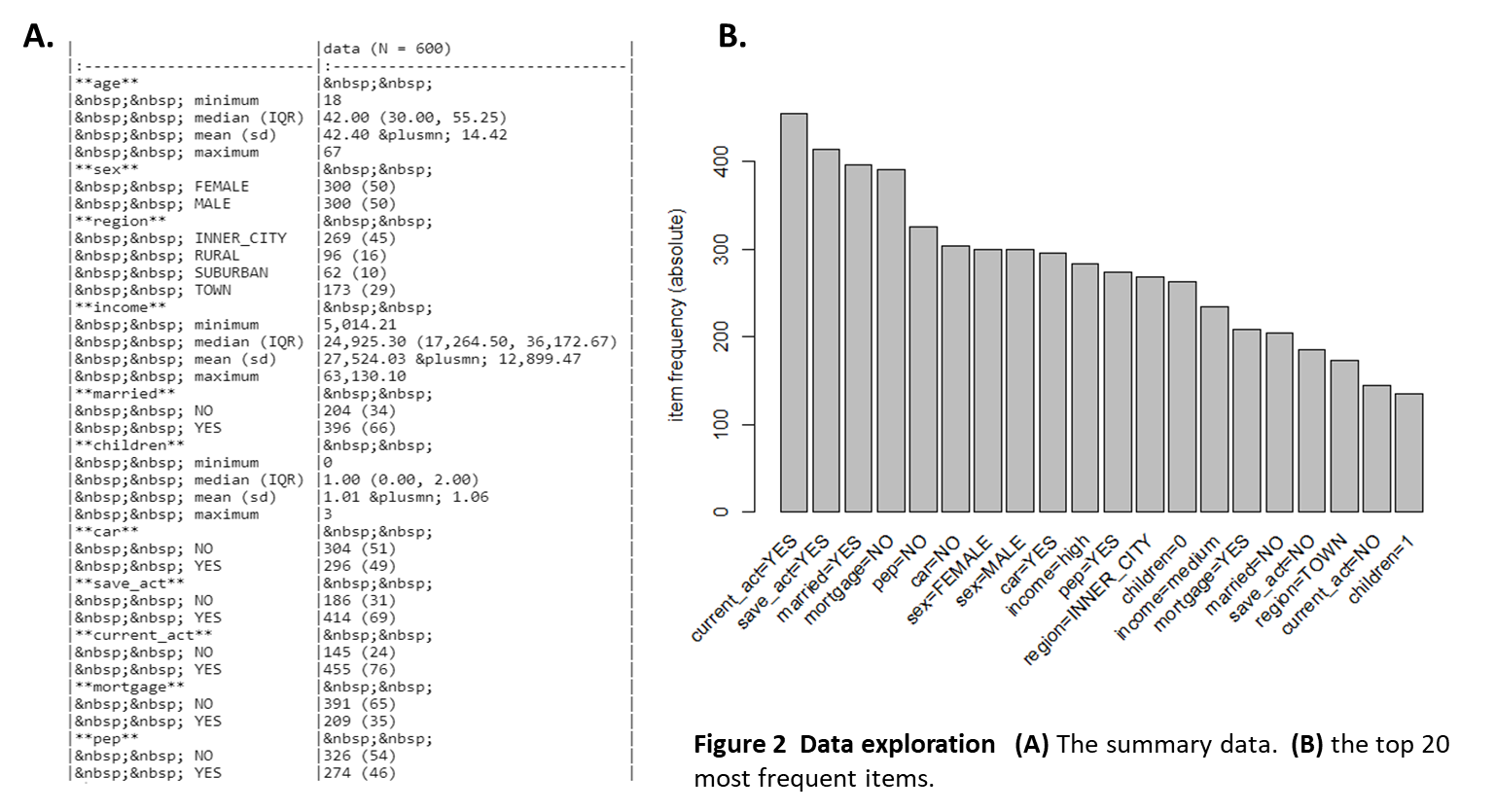
The Personal Equity Plan (PEP) was designed to encourage investment by individuals. The income from a PEP was tax-free, so long as the invested funds remained in the plan.

In determining whether a person will obtain PEP, typically, multiple factors need to be considered, including custom’s age, sex, living region, annal income, marital status, number of children, bank account status, and the mortgage etc.

Marketing department of a financial firm has records on customers’ demographics and bank information in order to determine they will want to obtain the new PEP. This project is focusing on exploring the bank data set and try to identify the potential interesting rules underlying the data set, and assistant the bank make decision.

# Analysis

Data preparation and cleaning: original dataset is a record data set, which contains variables ‘id, ‘age’, ’sex’, ‘region’, ‘income’, ‘married’, ‘children’, ‘car’, ’save\_act’, ’current\_act’, ’mortgage', and ‘pep’. 'age’ and ‘income’ are numeric variables, and rest of the variables are nominal variables (**Figure 1A**). Following methods were used to prepare and clean the data: 1)remove ‘id’ column.2) convert numeric field to nominal using discretization, including ‘age’ and ‘income’ contributes. 3) convert record data to transaction data (**Figure 1B**).

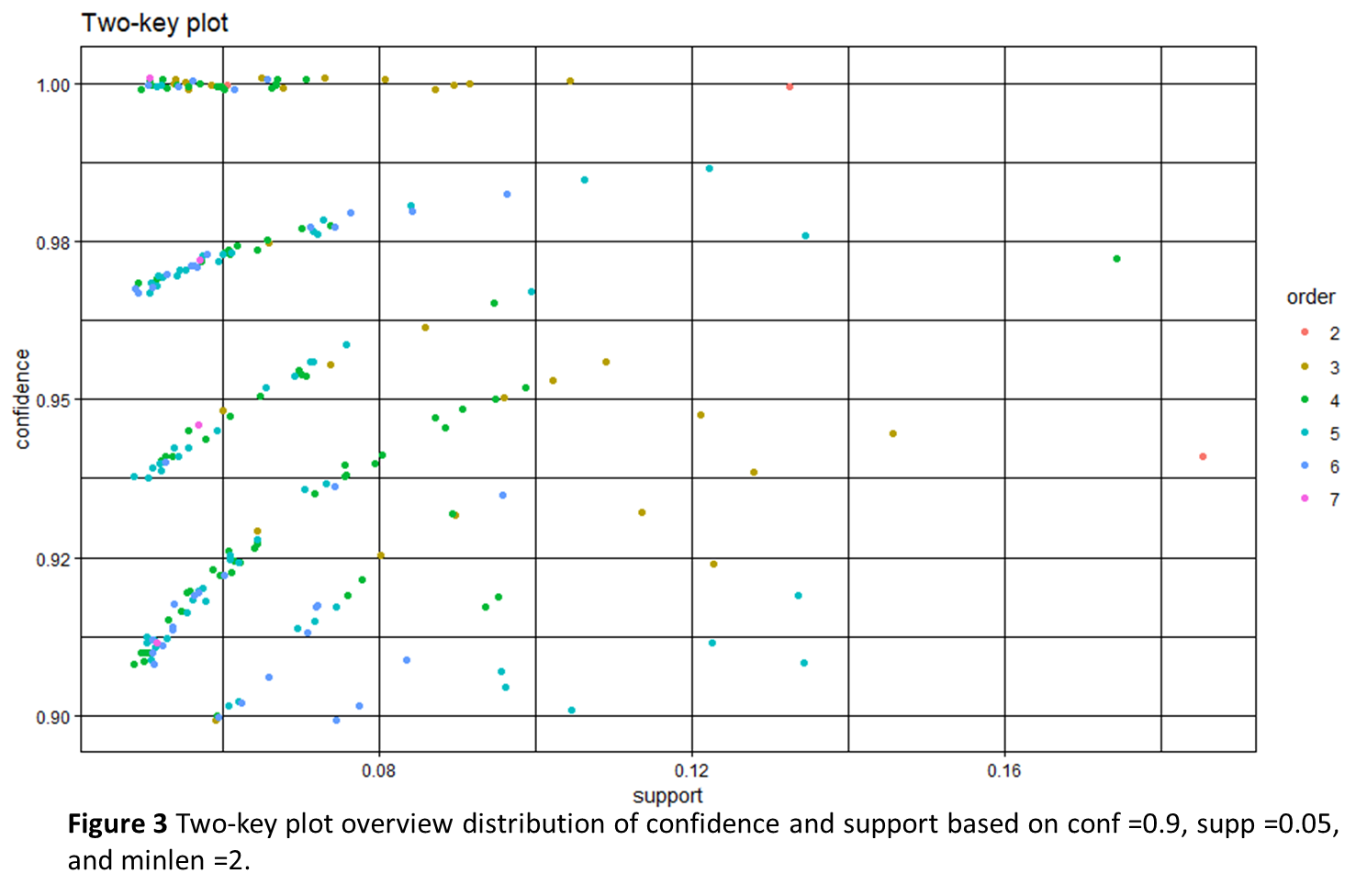
Data exploration: to explore the record data set, R summary function were used. This data would provide data distribution **(Figure 2A)** In addition, before making any rulesitemFrequencyPlot() was used as well to explore the data (**Figure 2B**).

Methods: “tidyverse”, “arule”, and “aruleViz” libraries and base R were used for analysis and data visualization. plot() function will be used for data visualization. **1)** read.transactions() was use to read the data and convert record data set to transaction data set (as mentioned above). **2)** itemFrequencyPlot() was used to mine rules to explore the data before make any rules (as mentioned above,Figure 2B). **3)** mine rules with the association rule algorithm and get the summary information about the rules, during mining, the parameters were adjust based on the results, more details were mentioned in below. **4)** set PEP as the right hand side (rhs) of the rules, and generate the roles, from where, sort the rules based on the support, confidence, and lift, respectively, with the decreasing is True. **5)** create network visualization with interactive use plot() function.

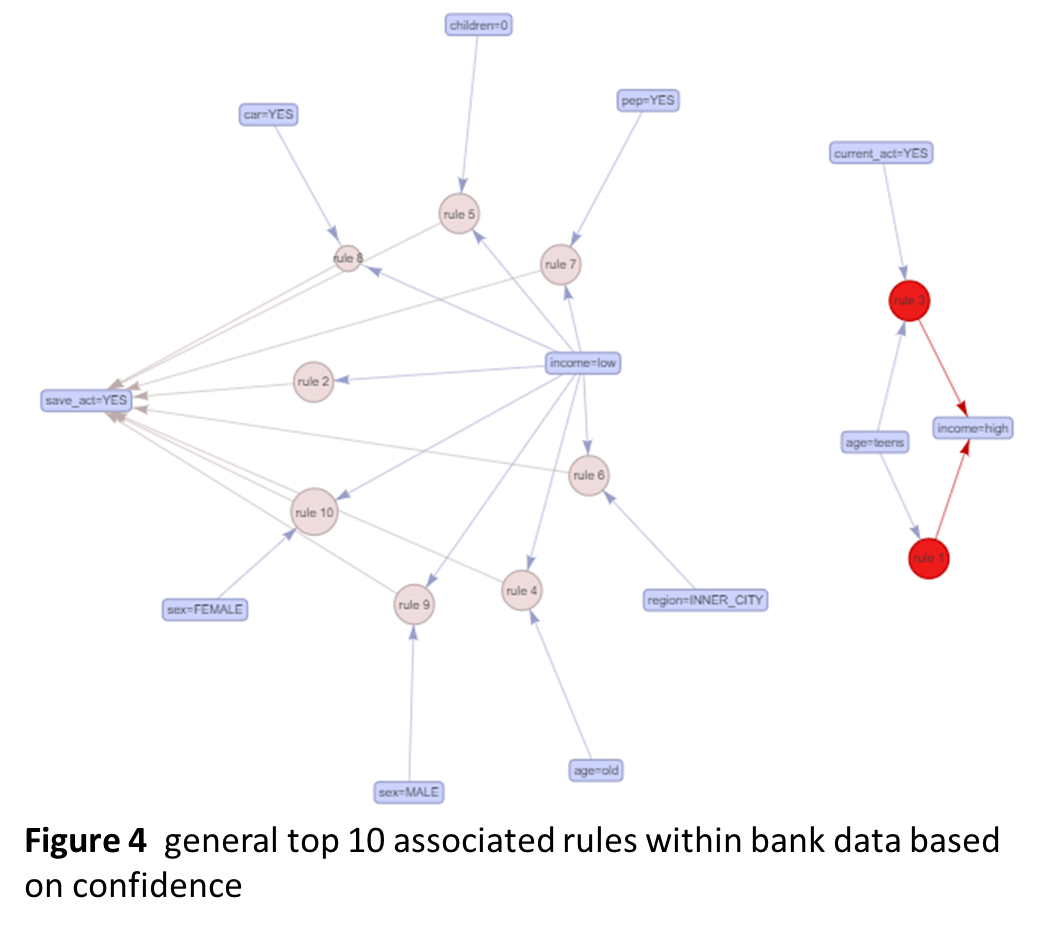
Analysis goals and Parameters: The goal of this project is to identify the potential interesting rules in this band data set, to this end, some parameters were set up when mining the association rules, including supp (support), conf (confidence), minlen (min length, to avoid the empty left hand side (lhs) items). Parameters were adjusted to gain meaningful association rules. First, minlen was set as 2. 1) when supp was set as 0.001, conf = 0.9, there were 896167 rules were generated, the data pool is too massive to explore the interesting rules. 2) when supp was set as 0.01, and conf = 0.9, there were 15064 rules were generated. 3) when supp was set as 0.1, and conf = 0.9, there were 19 rules were generated, includes three rules with ‘pep = NO’ on rhs , and no ‘pep = Yes’. 4) to explore the decision on PEP, supp was set as 0.05, and conf = 0.9, there are 219 rules were generated, which is good for further exploration.

# Results

## The general rules within the bank data

To get a first overview of the distribution of confidence and support, a two-key plot were generated based on the parameters conf = 0.9, supp =0.05, and minlen =2. As shown in **Figure 3**, total 219 rules were generated, and all the rules were showed at once, with the x axis represent rule support, y axis represent rule confidence, and color represent the length of the rules. From this figure, it is obvious that most of the rules (confidence greater than 0.9) were with support less than 0.08, very few rules have support greater than 0.1, and those rules contain less items. 

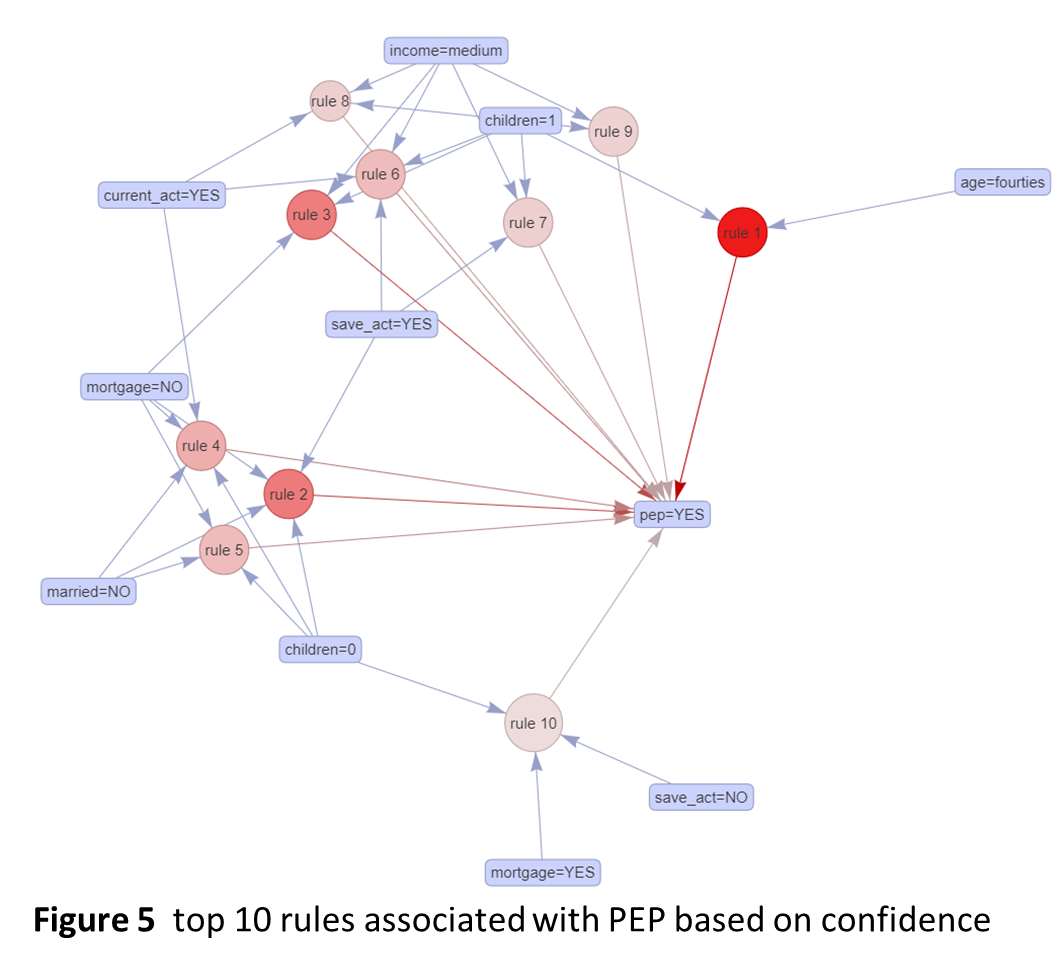
## The top 10 general rules with great confidence

To explore the most confidence rules in 219 rules mentioned above, the rules were sorted with parameter ‘confidence’. Top 10 rules were visualized in **Figure 4**, including {age=teens} => {income=high}, {income=low} => {save\_act=YES}, {age=teens, current\_act=YES}=> {income=high}, {age=old, income=low} => {save\_act=YES}, {income=low, children=0} => {save\_act=YES}, {region=INNER\_CITY, income=low} => {save\_act=YES}, {income=low, pep=YES}=> {save\_act=YES}, {income=low, car=YES} => {save\_act=YES}, {sex=MALE, income=low}=> {save\_act=YES}, and {sex=FEMALE, income=low} => {save\_act=YES}. Interestingly, the data showed that people who has low income, especially those old people, or people has no children, or people who own a car, or people live in inner city, would more likely have a saving account.

## Identify the population who would more likely buy PEP

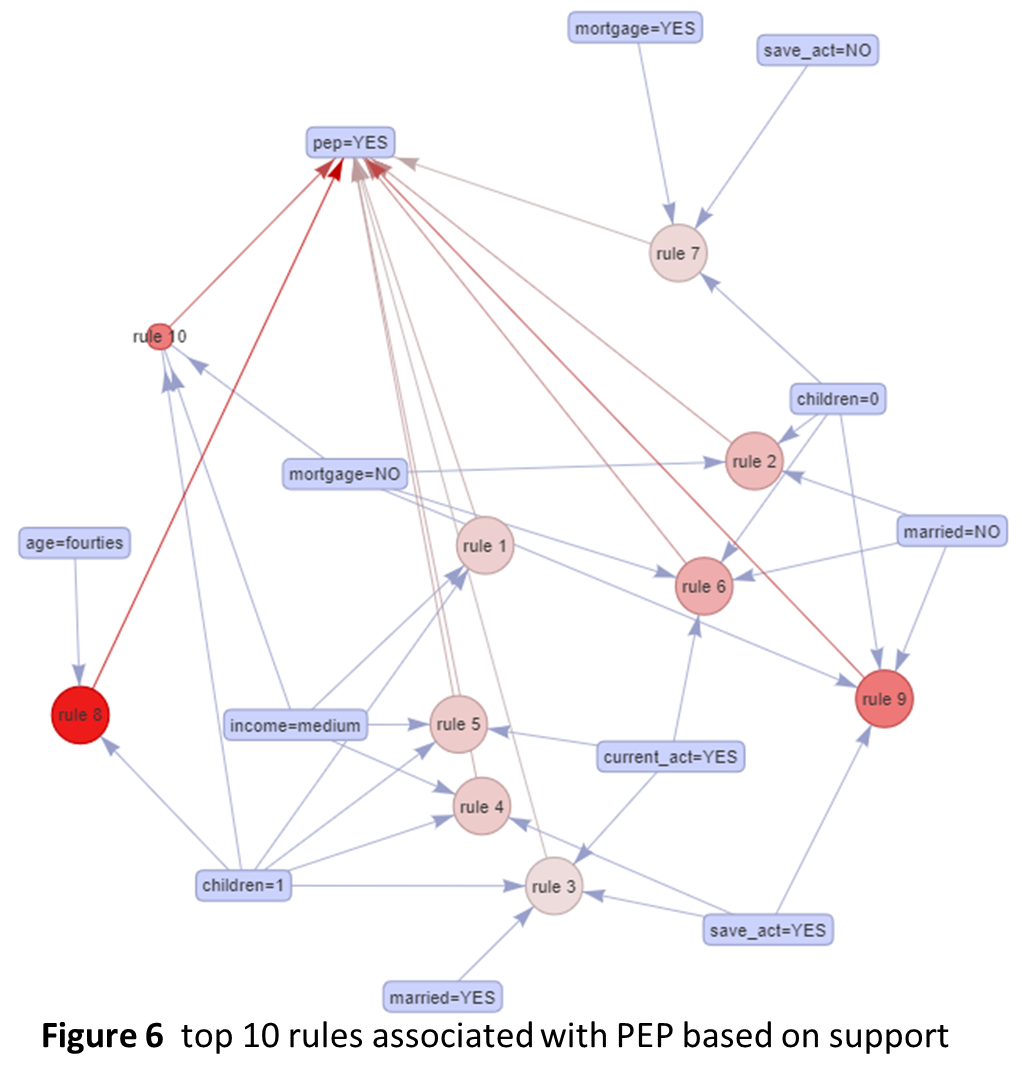
### Rules associated with PEP by confidence

**Figure 4** shows top 10(rule 1-10) most confidence rules with rhs is pep = YES, among those ten rules, 5 most interesting rules including rule 1: {age=fourties, children=1} => {pep=YES} (support: 0.053; confidence:1.00, and lift: 2.2), rule 2: {married=NO, children=0, save\_act=YES, mortgage=NO} => {pep=YES} (support: 0.052; confidence:0.97, and lift: 2.1), rule 3: {income=medium, children=1, mortgage=NO}=> {pep=YES} (support: 0.050; confidence:0.97, and lift: 2.1), rule 7: {income=medium, children=1, save\_act=YES} => {pep=YES} (support: 0.063 ; confidence: 0.93, and lift: 2.0), and rule 8: {income=medium, children=1, current\_act=YES} => {pep=YES} (support: 0.063 ; confidence: 0.93, and lift: 2.0), and rule 10: {children=0, save\_act=NO, mortgage=YES} => {pep=YES} (support: 0.057; confidence:0.92, and lift: 2.0). These data supported that 1) people who are fourties with one child are most likely to buy PEP, people around this age have very mature plan on the family finance, in addition, given that only one child, no big cash flow is needed, therefore, they would more likely buy financing products, such as PEP. 2) People who is not married, no children, no mortgage, with a saving account are more likely buy PEP, those people seem like have no big family expenses, and it would likely buy PEP. 3) people who have medium income, no mortgage, but with one child also would more likely buy PEP, probably they have limited expenses on the child and family, therefore, people who has medium income, they have a bit economic thought, and would like to give it a try. 4) It is very iterating that rule 7 and rule 8 indicated that people with medium income, 1 child with either current account or saving account would likely buy PEP, it is possible that holding account type do not affect their decision on PEP. 5) At last, in rule 10, people with no children, no save account, but has mortgage would as well buy pe PEP p, it is possible, those people have no family and kids’ education burden, and they are more flexible with their money, as on one side they borrow money from the bank, and on the other side, they also invest the money in the market.



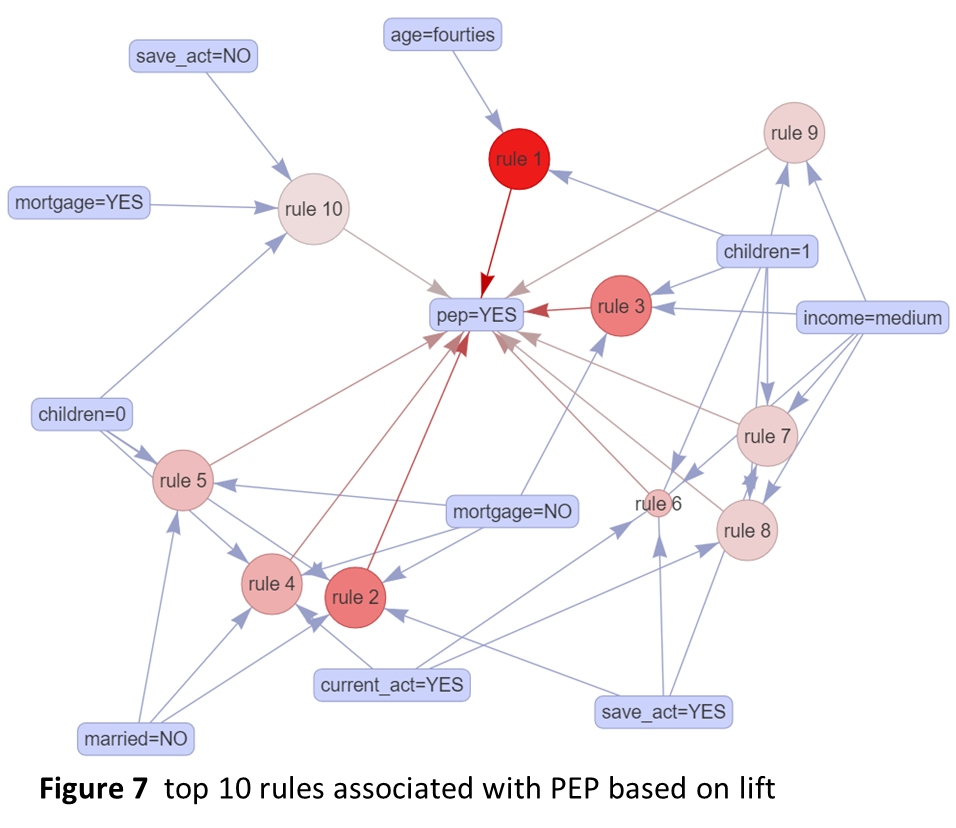
### Rules associated with PEP by support

In addition, top 10 rules with high support were shown in **Figure 6**, the support value is range from 0.082 to 0.050, with confidence values vary from 1.00 to 0.92. as shown in Figure 5, it is observed that most the rules were shared with those in top 10 confidence rules showed in above. Besides that, there are two rules are more interesting, including rule 2: {married=NO, children=0, mortgage=NO} => {pep=YES} (support: 0.075; confidence:0.94, and lift: 2.1), rule 3: {married=YES, children=1, save\_act=YES, current\_act=YES} => {pep=YES} (support: 0.073; confidence: 0.92, and lift: 2.0). people who have no life partner and children and no mortgage are more likely to buy PEP, one possible is that those people are at their young age, not start their own family yet, they are more sensitive to the financial market and bold, and more likely to invest market. People who married, and have start their family with one child, and have both saving account and current account also like to buy PEP, this population, they don’t have much burden on the kid, since one kid only, they own two different accounts, indicating they would like to put their money in multiple baskets, and therefore, they are more likely to buy PEP.



### Rules associated with PEP by lift

At last, top 10 rules ordered by lift value were showed in **Figure 6,** it is not surprise that those rules are almost consistent with the rules ordered by confidence. And all the rules showed here with lift value greater than 1, indicating that those rules are significant. It is interesting that people who has high income seems has no interest in PEP, since no item ‘income= high’ was showed in lfs with high lift, confidence, or support. It is possible that those PEP product is designed to target on the middle-class family or below.



# Conclusions

The aim of this project is to identify the population who would be most likely to buy PEP, people with different conditions (such as age, sex, living region, annal income, marital status, number of children, bank account status, and the mortgage etc.) have different attitude to the money market, such as PEP.

This study showed that marital status, number of children, and income may most likely affect their decision on PEP. In general, people with one child or no child, people who has medium income, and people who has no mortgage are most possible to buy PEP. In addition, regarding of age, it seems like that people who are 40~50 years old would like to consider investing PEP. However, sex, living region or their car status have very limited effect on their decision on PEP.

Taken together, PEP is more likely target on middle class family, and the decision on PEP is mostly based on household balance, less family burden would lead people invest PEP.