# Introduction

This research is to explore the bank data using association rule. The dataset contains attributes on each person's demographics and banking information to determine they will want to obtain the new PEP (Personal Equity Plan). By using the association rule, the researcher can identify what are the factors that would lead to higher chances of obtaining the new PEP.

We will review how the researcher performs analysis, what the result is, and what conclusions we get from the research. The analysis includes support, confidence, and lift values from the association rule analysis. An explanation of the pattern and why the researcher believe it is interesting based on the business objectives of the company. It will also include recommendations based on the discovered rule that might help the company to better understand the behavior of its customers or to develop a business opportunity.

# Analysis and Models

## Libraries

The researcher will use three libraries in performing the data analysis of this project. The three libraries are 1. Readr 2. Dplyr 3. Arules. Readr is used to read the CSV file of the bank data. Dplyr is a grammar of data manipulation, providing a consistent set of verbs that help the researchers solve the most common data manipulation challenges. The arules package for R provides the infrastructure for representing, manipulating, and analyzing transaction data and patterns using frequent itemsets and association rules. We will utilize all three libraries to analyze the bank data and apply the association rules concept in the research.

## Discretization

Discretization is the process of transferring continuous functions, models, variables, and equations into discrete counterparts. This process is usually carried out as a first step toward making them suitable for numerical evaluation and implementation on digital computers. We discretize age by customized bin and income by the equal-width bin. Based on the age ranges, we labeled the groups into children, teens, twenties, thirties, forties, fifties, old. Based on the income ranges, we categorized them into thee bins.

bd**$**age <- **cut**(bd**$**age, breaks = **c**(0,10,20,30,40,50,60,Inf),labels=**c**("child","teens","twenties","thirties","fourties","fifties","old"))

min\_income <- **min**(bd**$**income)  
max\_income <- **max**(bd**$**income)  
bins = 3   
width=(max\_income **-** min\_income)**/**bins;  
bd**$**income = **cut**(bd**$**income, breaks=**seq**(min\_income, max\_income, width))

## Numeric to Nominal

We converted children's data from numeric to nominal. In statistics, nominal data (also known as nominal scale) is a type of data that is used to label variables without providing any quantitative value. In this case, the children (1, 2, 3, 4…) served as nominal data. It has more category value than a numeric value.

## Data Conversion and Preprocess Data

In this step, for arules package to find the data pattern and apply association rules on the data, we converted "Yes" to "[variable\_name]=YES”. Also, we applied a function to convert categorical data to factor and discretize numeric variables.

bd**$**children=**factor**(bd**$**children)

bd**$**married=dplyr**::recode**(bd**$**married, YES="married=YES", NO="married=NO")  
bd**$**car=dplyr**::recode**(bd**$**car, YES="car=YES", NO="car=NO")  
bd**$**save\_act=dplyr**::recode**(bd**$**save\_act, YES="save\_act=YES", NO="save\_act=NO")  
bd**$**current\_act=dplyr**::recode**(bd**$**current\_act, YES="current\_act=YES", NO="current\_act=NO")  
bd**$**mortgage=dplyr**::recode**(bd**$**mortgage, YES="mortgage=YES", NO="mortgage=NO")  
bd**$**pep=dplyr**::recode**(bd**$**pep, YES="pep=YES", NO="pep=NO")

bd <- bd **%>%**   
 **select**(**-**id) **%>%**   
 **mutate\_if**(is.character, **funs**(as.factor)) **%>%**   
 **mutate\_if**(is.numeric, **funs**(discretize))

## Generate Rules and Explore

In this step, we used the apriori package to generate rules. The apriori() generates the most relevant set of rules from a given transaction data. The researcher utilized the package to mine frequent itemsets, association rules, or association hyperedges using the Apriori algorithm. The Apriori algorithm employs a level-wise search for frequent itemsets.

## Top 5 Rules

This section presents the top five rules that are generated by apriori package. The results provide insights of support, confidence, lift, count matrix based on the item sets on the left-hand side, and the right-hand side.

**inspect**(myRules[1**:**5])

## lhs rhs support confidence lift count  
## [1] {age=teens} => {income=(5.01e+03,2.44e+04]} 0.061666667 1.0000000 2.112676 37  
## [2] {income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.133333333 1.0000000 1.449275 80  
## [3] {age=twenties} => {income=(5.01e+03,2.44e+04]} 0.186666667 0.9411765 1.988401 112  
## [4] {age=teens,   
## region=SUBURBAN} => {income=(5.01e+03,2.44e+04]} 0.006666667 1.0000000 2.112676 4  
## [5] {age=teens,   
## region=SUBURBAN} => {car=car=NO} 0.006666667 1.0000000 1.973684 4

## Summary of Rules

This section summarizes the statistics of each quality measures including support, confidence, lift, and count. Based on a set of 2346 rules, we can find min, 1st quartile, median, mean, 3rd quartile, and a max of each measure.

**summary**(myRules)

## set of 2346 rules  
##   
## rule length distribution (lhs + rhs):sizes  
## 2 3 4   
## 3 109 2234   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 4.000 4.000 3.951 4.000 4.000   
##   
## summary of quality measures:  
## support confidence lift count   
## Min. :0.001667 Min. :0.9000 Min. :1.187 Min. : 1.000   
## 1st Qu.:0.003333 1st Qu.:1.0000 1st Qu.:1.449 1st Qu.: 2.000   
## Median :0.008333 Median :1.0000 Median :1.840 Median : 5.000   
## Mean :0.014397 Mean :0.9858 Mean :1.902 Mean : 8.638   
## 3rd Qu.:0.018333 3rd Qu.:1.0000 3rd Qu.:2.113 3rd Qu.: 11.000   
## Max. :0.186667 Max. :1.0000 Max. :9.677 Max. :112.000   
##   
## mining info:  
## data ntransactions support confidence  
## bd 600 0.001 0.9

## Sorting Rules by Measures

Below is a preview of the rules. We present the rules sorted by three measures (support, confidence, and lift). By reviewing all three measures, we can find interesting rules and target those rules for further analysis

### sort top 30 rules by support

myRules<-**sort**(myRules, by="support", decreasing=TRUE)  
**inspect**(myRules[1**:**30])

## lhs rhs support confidence lift count  
## [1] {age=twenties} => {income=(5.01e+03,2.44e+04]} 0.18666667 0.9411765 1.988401 112  
## [2] {children=0,   
## mortgage=mortgage=NO,   
## pep=pep=NO} => {married=married=YES} 0.17333333 0.9719626 1.472671 104  
## [3] {age=twenties,   
## current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.14500000 0.9456522 1.997857 87  
## [4] {income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.13333333 1.0000000 1.449275 80  
## [5] {age=twenties,   
## mortgage=mortgage=NO} => {income=(5.01e+03,2.44e+04]} 0.12666667 0.9382716 1.982264 76  
## [6] {age=twenties,   
## pep=pep=NO} => {income=(5.01e+03,2.44e+04]} 0.12166667 0.9240506 1.952220 73  
## [7] {age=twenties,   
## married=married=YES} => {income=(5.01e+03,2.44e+04]} 0.12166667 0.9480519 2.002927 73  
## [8] {age=twenties,   
## save\_act=save\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.11333333 0.9315068 1.967972 68  
## [9] {age=twenties,   
## car=car=NO} => {income=(5.01e+03,2.44e+04]} 0.10833333 0.9558824 2.019470 65  
## [10] {income=(4.38e+04,6.31e+04],   
## current\_act=current\_act=YES} => {save\_act=save\_act=YES} 0.10500000 1.0000000 1.449275 63  
## [11] {age=twenties,   
## region=INNER\_CITY} => {income=(5.01e+03,2.44e+04]} 0.10333333 0.9538462 2.015168 62  
## [12] {age=twenties,   
## current\_act=current\_act=YES,   
## mortgage=mortgage=NO} => {income=(5.01e+03,2.44e+04]} 0.09833333 0.9516129 2.010450 59  
## [13] {age=twenties,   
## sex=MALE} => {income=(5.01e+03,2.44e+04]} 0.09666667 0.9508197 2.008774 58  
## [14] {age=twenties,   
## married=married=YES,   
## current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.09500000 0.9661017 2.041060 57  
## [15] {children=0,   
## mortgage=mortgage=YES,   
## pep=pep=NO} => {save\_act=save\_act=YES} 0.09500000 0.9500000 1.376812 57  
## [16] {children=0,   
## save\_act=save\_act=YES,   
## mortgage=mortgage=YES} => {pep=pep=NO} 0.09500000 0.9193548 1.692064 57  
## [17] {age=twenties,   
## current\_act=current\_act=YES,   
## pep=pep=NO} => {income=(5.01e+03,2.44e+04]} 0.09333333 0.9180328 1.939506 56  
## [18] {income=(4.38e+04,6.31e+04],   
## mortgage=mortgage=NO} => {save\_act=save\_act=YES} 0.09166667 1.0000000 1.449275 55  
## [19] {age=twenties,   
## save\_act=save\_act=YES,   
## current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.09166667 0.9482759 2.003400 55  
## [20] {income=(4.38e+04,6.31e+04],   
## pep=pep=YES} => {save\_act=save\_act=YES} 0.09000000 1.0000000 1.449275 54  
## [21] {age=twenties,   
## sex=FEMALE} => {income=(5.01e+03,2.44e+04]} 0.09000000 0.9310345 1.966974 54  
## [22] {age=twenties,   
## married=married=YES,   
## mortgage=mortgage=NO} => {income=(5.01e+03,2.44e+04]} 0.09000000 0.9310345 1.966974 54  
## [23] {income=(4.38e+04,6.31e+04],   
## married=married=YES} => {save\_act=save\_act=YES} 0.08833333 1.0000000 1.449275 53  
## [24] {age=twenties,   
## car=car=NO,   
## current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.08833333 0.9464286 1.999497 53  
## [25] {age=twenties,   
## married=married=YES,   
## pep=pep=NO} => {income=(5.01e+03,2.44e+04]} 0.08833333 0.9464286 1.999497 53  
## [26] {age=twenties,   
## children=0} => {income=(5.01e+03,2.44e+04]} 0.08500000 0.9622642 2.032952 51  
## [27] {age=old,   
## income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.08166667 1.0000000 1.449275 49  
## [28] {income=(2.44e+04,4.38e+04],   
## children=1} => {pep=pep=YES} 0.08166667 0.9245283 2.024515 49  
## [29] {age=twenties,   
## region=INNER\_CITY,   
## current\_act=current\_act=YES} => {income=(5.01e+03,2.44e+04]} 0.08166667 0.9423077 1.990791 49  
## [30] {age=twenties,   
## car=car=YES} => {income=(5.01e+03,2.44e+04]} 0.07833333 0.9215686 1.946976 47

### sort top 30 rules by confidence

myRules<-**sort**(myRules, by="confidence", decreasing=TRUE)  
**inspect**(myRules[1**:**30])

## lhs rhs support confidence lift count  
## [1] {age=teens} => {income=(5.01e+03,2.44e+04]} 0.061666667 1 2.112676 37  
## [2] {income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.133333333 1 1.449275 80  
## [3] {age=teens,   
## region=SUBURBAN} => {income=(5.01e+03,2.44e+04]} 0.006666667 1 2.112676 4  
## [4] {age=teens,   
## region=SUBURBAN} => {car=car=NO} 0.006666667 1 1.973684 4  
## [5] {age=teens,   
## region=SUBURBAN} => {current\_act=current\_act=YES} 0.006666667 1 1.318681 4  
## [6] {age=teens,   
## children=3} => {income=(5.01e+03,2.44e+04]} 0.006666667 1 2.112676 4  
## [7] {age=teens,   
## children=3} => {pep=pep=NO} 0.006666667 1 1.840491 4  
## [8] {age=teens,   
## children=3} => {current\_act=current\_act=YES} 0.006666667 1 1.318681 4  
## [9] {age=teens,   
## region=RURAL} => {income=(5.01e+03,2.44e+04]} 0.011666667 1 2.112676 7  
## [10] {age=teens,   
## children=2} => {income=(5.01e+03,2.44e+04]} 0.020000000 1 2.112676 12  
## [11] {age=teens,   
## children=1} => {income=(5.01e+03,2.44e+04]} 0.013333333 1 2.112676 8  
## [12] {age=teens,   
## children=1} => {current\_act=current\_act=YES} 0.013333333 1 1.318681 8  
## [13] {age=teens,   
## current\_act=current\_act=NO} => {income=(5.01e+03,2.44e+04]} 0.006666667 1 2.112676 4  
## [14] {age=teens,   
## current\_act=current\_act=NO} => {car=car=NO} 0.006666667 1 1.973684 4  
## [15] {age=teens,   
## current\_act=current\_act=NO} => {married=married=YES} 0.006666667 1 1.515152 4  
## [16] {age=teens,   
## region=TOWN} => {income=(5.01e+03,2.44e+04]} 0.015000000 1 2.112676 9  
## [17] {age=teens,   
## save\_act=save\_act=NO} => {income=(5.01e+03,2.44e+04]} 0.026666667 1 2.112676 16  
## [18] {age=teens,   
## married=married=NO} => {income=(5.01e+03,2.44e+04]} 0.020000000 1 2.112676 12  
## [19] {age=teens,   
## married=married=NO} => {current\_act=current\_act=YES} 0.020000000 1 1.318681 12  
## [20] {age=teens,   
## mortgage=mortgage=YES} => {income=(5.01e+03,2.44e+04]} 0.023333333 1 2.112676 14  
## [21] {age=teens,   
## children=0} => {income=(5.01e+03,2.44e+04]} 0.021666667 1 2.112676 13  
## [22] {age=teens,   
## region=INNER\_CITY} => {income=(5.01e+03,2.44e+04]} 0.028333333 1 2.112676 17  
## [23] {age=teens,   
## pep=pep=YES} => {income=(5.01e+03,2.44e+04]} 0.018333333 1 2.112676 11  
## [24] {age=teens,   
## car=car=YES} => {income=(5.01e+03,2.44e+04]} 0.025000000 1 2.112676 15  
## [25] {age=teens,   
## sex=MALE} => {income=(5.01e+03,2.44e+04]} 0.035000000 1 2.112676 21  
## [26] {age=teens,   
## sex=FEMALE} => {income=(5.01e+03,2.44e+04]} 0.026666667 1 2.112676 16  
## [27] {age=teens,   
## car=car=NO} => {income=(5.01e+03,2.44e+04]} 0.036666667 1 2.112676 22  
## [28] {age=teens,   
## pep=pep=NO} => {income=(5.01e+03,2.44e+04]} 0.043333333 1 2.112676 26  
## [29] {age=teens,   
## mortgage=mortgage=NO} => {income=(5.01e+03,2.44e+04]} 0.038333333 1 2.112676 23  
## [30] {age=teens,   
## married=married=YES} => {income=(5.01e+03,2.44e+04]} 0.041666667 1 2.112676 25

### sort top 30 rules by lift

myRules<-**sort**(myRules, by="lift", decreasing=TRUE)  
**inspect**(myRules[1**:**30])

## lhs rhs support confidence lift count  
## [1] {age=teens,   
## children=2,   
## pep=pep=YES} => {region=SUBURBAN} 0.001666667 1 9.677419 1  
## [2] {age=twenties,   
## region=SUBURBAN,   
## income=(2.44e+04,4.38e+04]} => {children=3} 0.001666667 1 8.823529 1  
## [3] {age=old,   
## region=SUBURBAN,   
## children=1} => {income=(4.38e+04,6.31e+04]} 0.003333333 1 7.500000 2  
## [4] {age=old,   
## sex=MALE,   
## region=SUBURBAN} => {income=(4.38e+04,6.31e+04]} 0.006666667 1 7.500000 4  
## [5] {age=fifties,   
## region=SUBURBAN,   
## children=1} => {income=(4.38e+04,6.31e+04]} 0.001666667 1 7.500000 1  
## [6] {age=old,   
## children=3,   
## pep=pep=YES} => {income=(4.38e+04,6.31e+04]} 0.006666667 1 7.500000 4  
## [7] {region=RURAL,   
## children=3,   
## pep=pep=YES} => {income=(4.38e+04,6.31e+04]} 0.013333333 1 7.500000 8  
## [8] {age=fifties,   
## children=3,   
## pep=pep=YES} => {income=(4.38e+04,6.31e+04]} 0.005000000 1 7.500000 3  
## [9] {children=3,   
## current\_act=current\_act=NO,   
## pep=pep=YES} => {income=(4.38e+04,6.31e+04]} 0.005000000 1 7.500000 3  
## [10] {age=old,   
## children=2,   
## mortgage=mortgage=YES} => {income=(4.38e+04,6.31e+04]} 0.010000000 1 7.500000 6  
## [11] {region=SUBURBAN,   
## income=(4.38e+04,6.31e+04],   
## car=car=YES} => {age=old} 0.005000000 1 6.666667 3  
## [12] {income=(4.38e+04,6.31e+04],   
## children=2,   
## current\_act=current\_act=NO} => {age=old} 0.006666667 1 6.666667 4  
## [13] {region=TOWN,   
## income=(4.38e+04,6.31e+04],   
## children=2} => {age=old} 0.008333333 1 6.666667 5  
## [14] {income=(4.38e+04,6.31e+04],   
## children=2,   
## pep=pep=NO} => {age=old} 0.001666667 1 6.666667 1  
## [15] {region=INNER\_CITY,   
## income=(4.38e+04,6.31e+04],   
## children=1} => {age=old} 0.008333333 1 6.666667 5  
## [16] {region=INNER\_CITY,   
## income=(4.38e+04,6.31e+04],   
## current\_act=current\_act=NO} => {age=old} 0.005000000 1 6.666667 3  
## [17] {sex=MALE,   
## region=TOWN,   
## income=(4.38e+04,6.31e+04]} => {age=old} 0.011666667 1 6.666667 7  
## [18] {income=(4.38e+04,6.31e+04],   
## children=3} => {region=RURAL} 0.013333333 1 6.250000 8  
## [19] {age=old,   
## income=(4.38e+04,6.31e+04],   
## children=3} => {region=RURAL} 0.006666667 1 6.250000 4  
## [20] {age=fifties,   
## income=(4.38e+04,6.31e+04],   
## children=3} => {region=RURAL} 0.005000000 1 6.250000 3  
## [21] {age=fourties,   
## income=(4.38e+04,6.31e+04],   
## children=3} => {region=RURAL} 0.001666667 1 6.250000 1  
## [22] {income=(4.38e+04,6.31e+04],   
## children=3,   
## current\_act=current\_act=NO} => {region=RURAL} 0.005000000 1 6.250000 3  
## [23] {income=(4.38e+04,6.31e+04],   
## married=married=NO,   
## children=3} => {region=RURAL} 0.005000000 1 6.250000 3  
## [24] {income=(4.38e+04,6.31e+04],   
## children=3,   
## pep=pep=YES} => {region=RURAL} 0.013333333 1 6.250000 8  
## [25] {income=(4.38e+04,6.31e+04],   
## children=3,   
## car=car=YES} => {region=RURAL} 0.006666667 1 6.250000 4  
## [26] {sex=MALE,   
## income=(4.38e+04,6.31e+04],   
## children=3} => {region=RURAL} 0.005000000 1 6.250000 3  
## [27] {sex=FEMALE,   
## income=(4.38e+04,6.31e+04],   
## children=3} => {region=RURAL} 0.008333333 1 6.250000 5  
## [28] {income=(4.38e+04,6.31e+04],   
## children=3,   
## car=car=NO} => {region=RURAL} 0.006666667 1 6.250000 4  
## [29] {income=(4.38e+04,6.31e+04],   
## children=3,   
## mortgage=mortgage=NO} => {region=RURAL} 0.013333333 1 6.250000 8  
## [30] {income=(4.38e+04,6.31e+04],   
## married=married=YES,   
## children=3} => {region=RURAL} 0.008333333 1 6.250000 5

## Redundancies

A rule is redundant if a more general rule with the same or higher confidence exists. That is, a more specific rule is redundant if it is only equally or even less predictive than a more general rule. A rule is more general if it has the same RHS but one or more items removed from the LHS. In this step, we remove those redundancies and prune the results.

subset.matrix <- **is.subset**(myRules, myRules)  
subset.matrix[**lower.tri**(subset.matrix, diag=T)] <- NA

## Warning in `[<-`(`\*tmp\*`, as.vector(i), value = NA): x[.] <- val: x is  
## "ngTMatrix", val not in {TRUE, FALSE} is coerced; NA |--> TRUE.

redundant <- **colSums**(subset.matrix, na.rm=T) **>=** 1  
myRules.pruned <- myRules[**!**redundant]  
myRules<-myRules.pruned

## Targeting Items

In this step, we are going to target at five interesting rules by specifying parameters.

Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: region=SUBURBAN

myRules<-**apriori**(data=bd, parameter=**list**(supp=0.001,conf = 0.9),   
 appearance = **list**(default="lhs",rhs="region=SUBURBAN"),  
 control = **list**(verbose=F))  
myRules<-**sort**(myRules, decreasing=TRUE,by="confidence")  
**inspect**(myRules[1**:**5])

## lhs rhs support confidence lift count  
## [1] {age=teens,   
## children=2,   
## pep=pep=YES} => {region=SUBURBAN} 0.001666667 1 9.677419 1  
## [2] {age=teens,   
## income=(5.01e+03,2.44e+04],   
## children=2,   
## pep=pep=YES} => {region=SUBURBAN} 0.001666667 1 9.677419 1  
## [3] {age=teens,   
## sex=FEMALE,   
## children=2,   
## pep=pep=YES} => {region=SUBURBAN} 0.001666667 1 9.677419 1  
## [4] {age=teens,   
## children=2,   
## car=car=NO,   
## pep=pep=YES} => {region=SUBURBAN} 0.001666667 1 9.677419 1  
## [5] {age=teens,   
## children=2,   
## mortgage=mortgage=NO,   
## pep=pep=YES} => {region=SUBURBAN} 0.001666667 1 9.677419 1

Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: children=3

myRules<-**apriori**(data=bd, parameter=**list**(supp=0.001,conf = 0.9),   
 appearance = **list**(default="lhs",rhs="children=3"),  
 control = **list**(verbose=F))  
myRules<-**sort**(myRules, decreasing=TRUE,by="confidence")  
**inspect**(myRules[1**:**5])

## lhs rhs support confidence lift count  
## [1] {age=twenties,   
## region=SUBURBAN,   
## income=(2.44e+04,4.38e+04]} => {children=3} 0.001666667 1 8.823529 1  
## [2] {age=teens,   
## region=RURAL,   
## married=married=YES,   
## save\_act=save\_act=YES} => {children=3} 0.001666667 1 8.823529 1  
## [3] {age=teens,   
## married=married=NO,   
## save\_act=save\_act=NO,   
## pep=pep=NO} => {children=3} 0.001666667 1 8.823529 1  
## [4] {age=fifties,   
## region=SUBURBAN,   
## current\_act=current\_act=NO,   
## pep=pep=NO} => {children=3} 0.001666667 1 8.823529 1  
## [5] {age=fifties,   
## region=SUBURBAN,   
## current\_act=current\_act=NO,   
## mortgage=mortgage=NO} => {children=3} 0.001666667 1 8.823529 1

Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: children=2

myRules<-**apriori**(data=bd, parameter=**list**(supp=0.001,conf = 0.9),   
 appearance = **list**(default="lhs",rhs="children=2"),  
 control = **list**(verbose=F))  
myRules<-**sort**(myRules, decreasing=TRUE,by="confidence")  
**inspect**(myRules[1**:**5])

## lhs rhs support confidence lift count  
## [1] {age=teens,   
## region=RURAL,   
## current\_act=current\_act=NO} => {children=2} 0.001666667 1 4.477612 1  
## [2] {age=teens,   
## region=TOWN,   
## current\_act=current\_act=NO} => {children=2} 0.001666667 1 4.477612 1  
## [3] {age=teens,   
## sex=FEMALE,   
## married=married=NO} => {children=2} 0.003333333 1 4.477612 2  
## [4] {age=old,   
## region=SUBURBAN,   
## married=married=NO} => {children=2} 0.003333333 1 4.477612 2  
## [5] {age=fifties,   
## region=SUBURBAN,   
## income=(5.01e+03,2.44e+04]} => {children=2} 0.001666667 1 4.477612 1

Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: save\_act=save\_act=YES

myRules<-**apriori**(data=bd, parameter=**list**(supp=0.001,conf = 0.9),

appearance = **list**(default="lhs",rhs="save\_act=save\_act=YES"),

control = **list**(verbose=F))

myRules<-**sort**(myRules, decreasing=TRUE,by="confidence")

**inspect**(myRules[1**:**5])

## lhs rhs support confidence lift count

## [1] {income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.13333333 1 1.449275 80

## [2] {region=SUBURBAN,

## income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.01833333 1 1.449275 11

## [3] {income=(4.38e+04,6.31e+04],

## children=3} => {save\_act=save\_act=YES} 0.01333333 1 1.449275 8

## [4] {children=3,

## pep=pep=YES} => {save\_act=save\_act=YES} 0.02166667 1 1.449275 13

## [5] {age=old,

## income=(4.38e+04,6.31e+04]} => {save\_act=save\_act=YES} 0.08166667 1 1.449275 49

Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: pep=pep=YES

myRules<-**apriori**(data=bd, parameter=**list**(supp=0.001,conf = 0.9),   
 appearance = **list**(default="lhs",rhs="pep=pep=YES"),  
 control = **list**(verbose=F))  
myRules<-**sort**(myRules, decreasing=TRUE,by="confidence")  
**inspect**(myRules[1**:**5])

## 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

Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: pep=pep=NO

myRules<-**apriori**(data=bd, parameter=**list**(supp=0.001,conf = 0.9),   
 appearance = **list**(default="lhs",rhs="pep=pep=NO"),  
 control = **list**(verbose=F))  
myRules<-**sort**(myRules, decreasing=TRUE,by="confidence")  
**inspect**(myRules[1**:**5])

## lhs rhs support confidence lift count  
## [1] {age=teens,   
## children=3} => {pep=pep=NO} 0.006666667 1 1.840491 4  
## [2] {children=3,   
## save\_act=save\_act=NO} => {pep=pep=NO} 0.036666667 1 1.840491 22  
## [3] {age=teens,   
## region=SUBURBAN,   
## children=1} => {pep=pep=NO} 0.003333333 1 1.840491 2  
## [4] {age=teens,   
## region=SUBURBAN,   
## married=married=NO} => {pep=pep=NO} 0.001666667 1 1.840491 1  
## [5] {age=teens,   
## region=RURAL,   
## children=3} => {pep=pep=NO} 0.001666667 1 1.840491 1

# Results

In association rules analysis, it’s critical to understand those three measurements -the support, confidence, and lifts. These three measures can be used to decide the relative strength of the rules.

* Support measure gives an idea of how frequent an itemset is in all the transactions. It is a measure of how frequently the itemset 𝑋 appears in the dataset as the antecedent/left-hand-side.
* Confidence measure defines the likeliness of occurrence of consequent in the group given that the group already has the antecedents. It is a measure of how frequently the association rule {𝑋}⇒{𝑌} is true in the dataset.
* Lift controls for the support (frequency) of consequent while calculating the conditional probability of occurrence of {Y} given {X}. Lift is also the factor by which

the co-occurrence of A and B exceeds the expected probability of A and B co-occurring, had they been independent. So, the higher the lift, the higher the chance of A and B occurring together.

Below presents the formula for those three measures.

A screenshot of a cell phone

Description automatically generated

With all the steps mentioned above, we concluded five interesting rules to the researcher. We will discuss three measurements for each rule, why it’s interesting based on the business objectives of the company, and the recommendations on developing a business opportunity.

Filter: Support: 0.001 Confidence: 0.9 Sort by: confidence RHS: region=SUBURBAN

Interesting Rule:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| LHS | RHS | SUPPORT | CONFIDENCE | LIFT | COUNT |
| age=teens, children=2, pep=pep=YES | region=SUBURBAN | 0.001666667 | 1 | 9.677419 | 1 |

For the selected rule in this filtering group, the researcher identified support to be 0.001666667; confidence to be 1; lift to be 9.677419. It’s interesting to see that there’s a strong relationship between region suburban and age teens. According to the analysis of this rule, most of the customers who live in suburban are teens with 2 children. In this case, the business should identify what financial products that could be more attractive to this group. For example, will they look for a mortgage type of product soon? Are they still living with their parents? What type of loan products can the bank supply to them to build their future? Those questions should help the bank plan growing opportunities.

Filter: Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: children=3

Interesting Rule:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| LHS | RHS | SUPPORT | CONFIDENCE | LIFT | COUNT |
| age=teens,  region=RURAL,  married=married=YES,  save\_act=save\_act=YES | children=3 | 0.001666667 | 1 | 8.823529 | 1 |

The second interesting rule is the customers who have 3 children. For the selected rule in this filtering group, the researcher identified support to be 0.001666667; confidence to be 1; lift to be 8.823529. According to the analysis, the customers in this category are either married with a savings account (possibly higher income) or they are not married and without a saving account. In this case, we can target the customers with two product lines – credit/loans and savings. Credits and loans mean the bank would lend the money to the customers who might have financial needs while they work hard to support three children. Another opportunity can be providing a savings account with higher interest rates. Those who have three children but also have savings could potentially have a higher income. If they have extra money to put aside for the kids, they would be looking for higher returns in saving.

Filter: Support: 0.001 Confidence: 0.9 Sort by: confidence rhs: children=2

Interesting Rule:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| LHS | RHS | SUPPORT | CONFIDENCE | LIFT | COUNT |
| age=teens,  region=RURAL,  current\_act=current\_act=NO | children=2 | 0.001666667 | 1 | 4.477612 | 1 |

The third interesting rule is the customers who have 2 children. For the selected rule in this filtering group, the researcher identified support to be 0.001666667; confidence to be 1; lift to be 4.477612. According to the analysis, the customers in this category usually don’t have an account with the bank. This area could be an exploding market. We need to further understand why the customers in this category don't have an account with the bank. What can potentially be a financial product to support them? Mobile banking products could be an idea. The younger generation might lack interest in going to the bank physically. The opportunity will be how the bank supplies services physically and virtually.

Filter: Support: 0.001 Confidence: 0.9 Sort by: confidence RHS: save\_act=save\_act=YES

Interesting Rule:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| LHS | RHS | SUPPORT | CONFIDENCE | LIFT | COUNT |
| income=(4.38e+04,6.31e+04] | save\_act=save\_act=YES | 0.13333333 | 1 | 1.449275 | 80 |

The fourth interesting rule is the customers who have saving accounts or not. For the selected rule in this filtering group, the researcher identified support to be 0.13333333; confidence to be 1; lift to be 1.449275. It's predictable that people who have higher incomes could have a savings account. Others like those who have 3 children and in the older group would have a savings account as well. To build customers in this group, we can consider a package of saving incentives. For those who have 3 children, the purpose of saving might be for the children to go to college in the future. Bundling tax benefits with savings would increase the interests of the customers.

Filter: Support: 0.001 Confidence: 0.9 Sort by: confidence RHS: pep=pep=YES

Filter: Support: 0.001 Confidence: 0.9 Sort by: confidence RHS: pep=pep=NO

Interesting Rule:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| LHS | RHS | SUPPORT | CONFIDENCE | LIFT | COUNT |
| income=(4.38e+04,6.31e+04],  children=3 | pep=pep=YES | 0.013333333 | 1 | 2.189781 | 1 |
| age=teens,  region=SUBURBAN,  children=1 | pep=pep=NO | 0.003333333 | 1 | 1.840491 | 2 |

Last but not least, we focused on if the customers have PEP or not. A personal equity plan (PEP) was an investment plan introduced in the U.K. that allowed people over the age of 18 to invest in shares of British companies. For the selected rule in this filtering group, the researcher identified 1. (pep=YES) support to be 0.013333333; confidence to be 1; lift to be 2.189781. 2. For the selected rule in this filtering group, the researcher identified support to be 0.003333333; confidence to be 1; lift to be 1.840491. It was done through an approved plan, qualifying unit trust, or investment trust. Investors received both income and capital gains free of tax. We found that those who have PEP loans are usually in the higher-income category and with 1 ~ 3 children. Those who currently have a PEP usually have some money to set aside for investment. Those customers who don't have PEP usually are at their younger age and with kids. In other words, they might have their money tied up with current expenses. However, a business will always look for opportunities to serve the clients. What are the other benefits that PEP can provide to this category of customers? For teens, they might not be eligible to purchase PEP. How about the customers who have 3 kids but with no saving accounts? The bank can develop a product specifically for this group.

# Conclusions

With the association rule, we can help the bank understand its customers and demographics. When running the association rules, we can use different arguments to find the rule that best interests us. We can also focus on a product line and research why they own a specific product or not. What could potentially be the reason? With that said, the analysis was able to locate the demographic info for those who own PEP and those who don’t. The bank can further analyze and develop several PEP products for them.

# Reference

Analysis, and Presentation of Strong Rules - Discovery - 1991