Diego Valdes

IST 707

Home Work 3

4/20/2019

**Introduction**

The financial industry is consistently adapting to the changes in its customer base. They have been at the front of adapting technology (direct deposit, web banking), redesigning their branch locations (bank cafes, personal banking stations, drive up tellers, ATMs), and adding new services and products. While all of these changes have been widely adopted by customers, and expected, the end result of all of these changes has been for one purpose—to make money.

A bank is not a gigantic piggy bank for its customer base; it’s a business, and like all businesses its purpose is to make money. Banks due that by providing products like loans, checking/savings accounts, and investment products. Investment products, which generally are long term sources of revenue, provide banks with revenue from fees and from the dividends of the investment. Unlike loans, the principle should increase, and thus be a sustaining form of revenue.

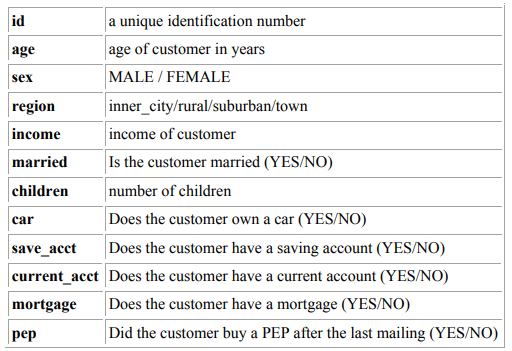
A Personal Equity Plan, (PEP) is an investment plan that generally requires a minimum amount to start. They may have limits to the amount that can be contributed through the year, and there are many options for how the funds can be invested. It is beneficial to customers because the risk of having funds in the market comes with the possibility of a greater return on investment than having equity in a more secure account such as a savings or CD. This also means that a PEP will attract a customer that doesn’t mind added risk with their finances.

**Analysis**

**About the Data**

The dataset (bankdata\_csv\_all.csv) had no missing data or NA values. To prepare it for analysis, the first col (id) was removed because it functions as a primary key and would have no impact on any analysis. A series of columns were turned into factor (sex, region, married, children, car, save\_act, current\_act, mortgage, pep) to treat them as classifications. Column ‘income’ was converted to numeric.

The original dataset has a total of 14 variables.



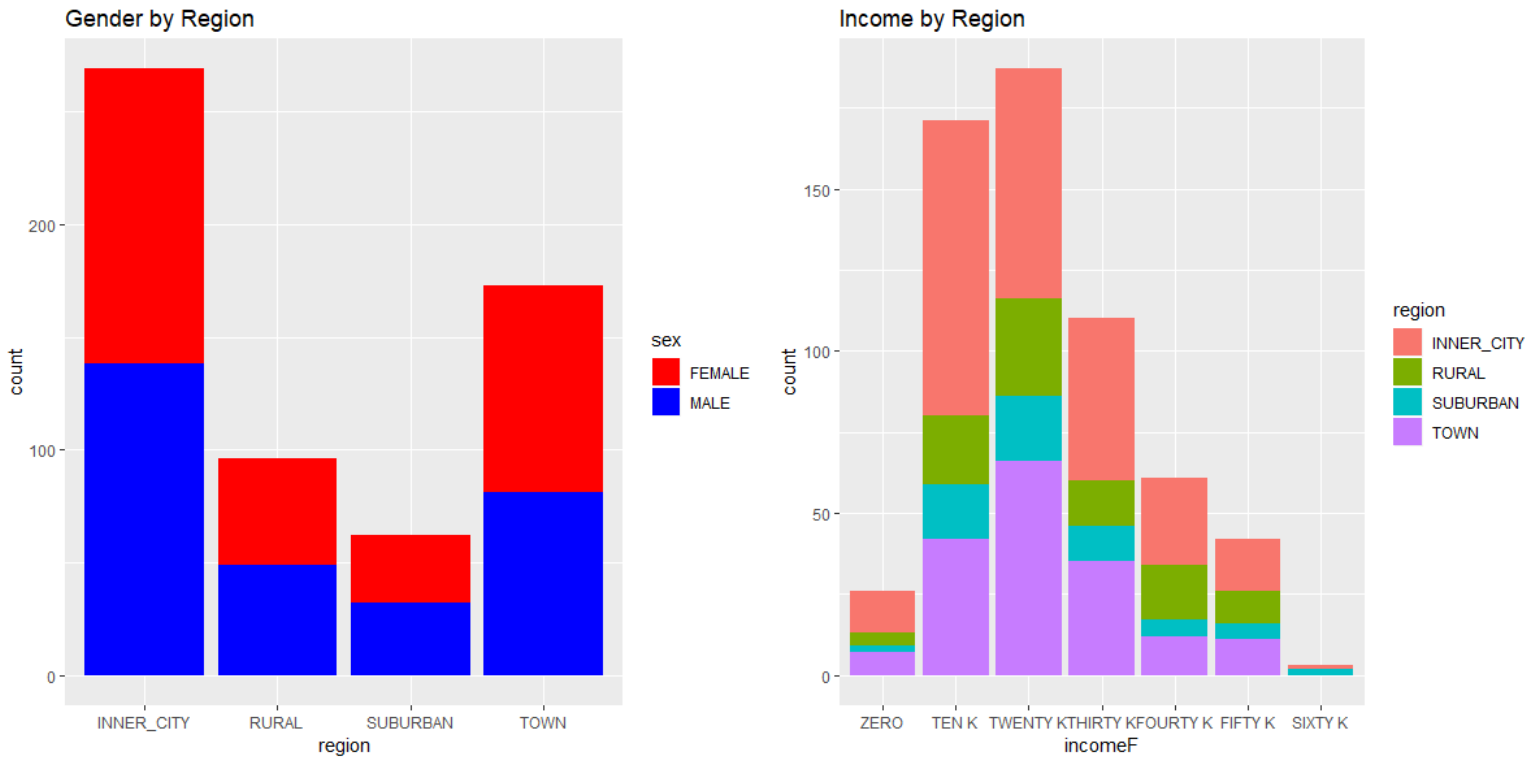
Three columns were added (ageF, incomeF, childrenInt) were added. Both ‘ageF’ and ‘incomeF’ were factor versions of the columns ‘age’ and ‘income’. This was done to use these columns as classifications when aggregating data. IncomeF has seven levels based on the range of incomes in the data set in increments of 10,000, ie; 0 – ‘ZERO’, 10,000 – ‘TEN K’, … The column ‘ageF’ classifies ages in increments of 10, ie: 0 – ‘child’, 10 – ‘teens’, … The column childrenInt was added for graphing and exploring results.

Two additional datasets were generated for graphing (includes only customers with accounts) and for rule mining (excluding age(factor) and income(numeric)).

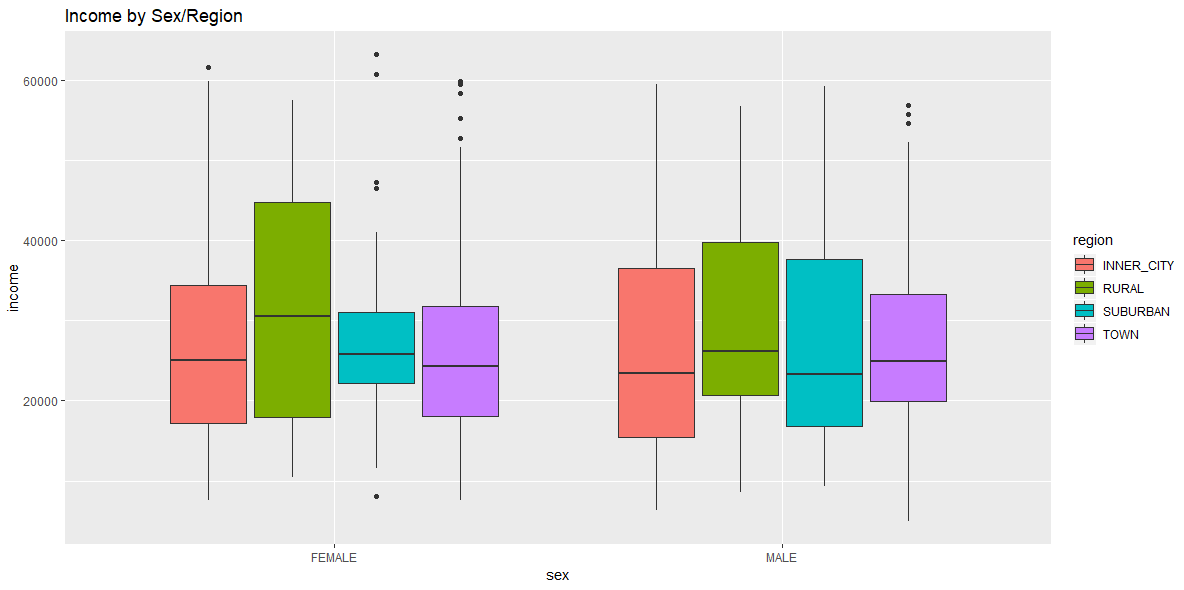
Most customers reside in either the inner city or the town and the distribution of males is mostly equal, with 300 of each split in the regions. The mean income for the sample is approximately twenty-seven thousand dollars, higher than the median which gives the distribution a right tale.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Male** | | | | | |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** |
| **5014** | **16805** | **24449** | **27217** | **36611** | **59409** |

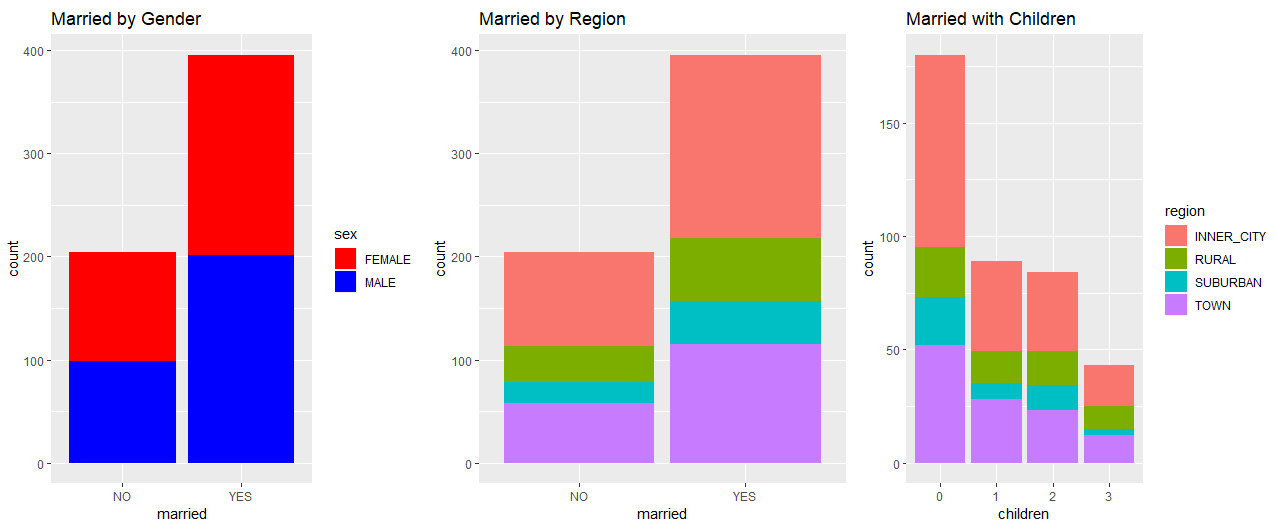
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Female** | | | | | |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** |
| **7549** | **17711** | **25449** | **27831** | **34873** | **63130** |



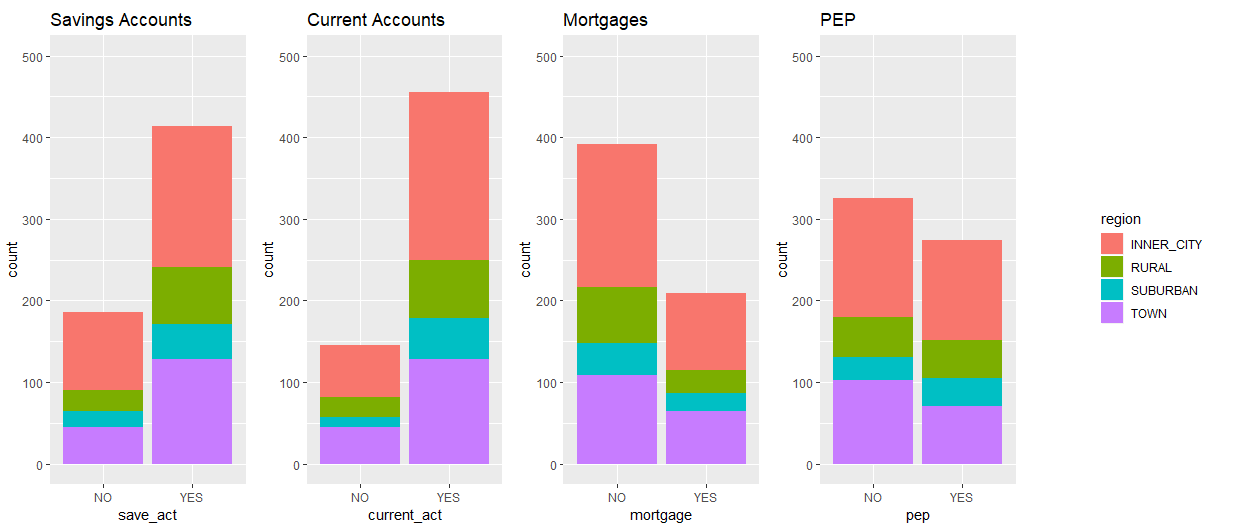
Examining income by region and gender reveals more outliers for females than males. There is a much greater disparity in pay in the suburban region for the genders than in any other region.



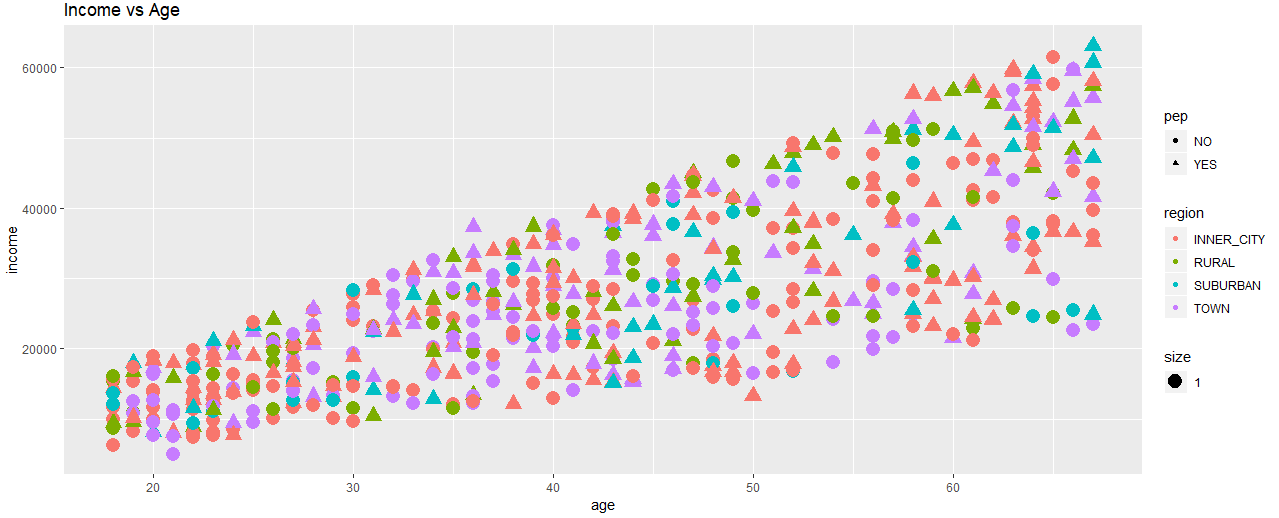
The number of married and unmarried people in the sample is evenly split among gender. That remains consistent through the regions and the ratio of unmarried to married people also remains consistent through the regions. Most customers have at least one child, with a heavy drop after 2 children.



Most customers have a current account, but unclear if this includes checking accounts or other accounts not included in the dataset. Savings account is the most common product bank customers are using. Data that could be useful is how many customers are homeowners.

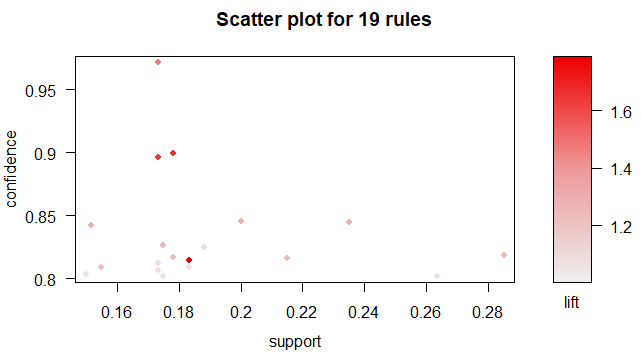


There is a correlation between age and income, the older a customer the more likely they are to be higher on the income ladder. Adding a third and fourth variable to the plot didn’t reveal anything significant to investigate.

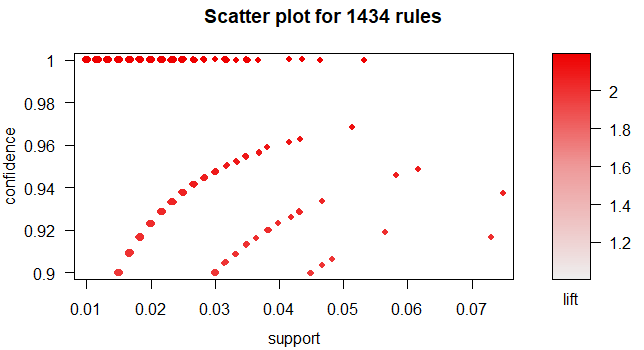


**Results**

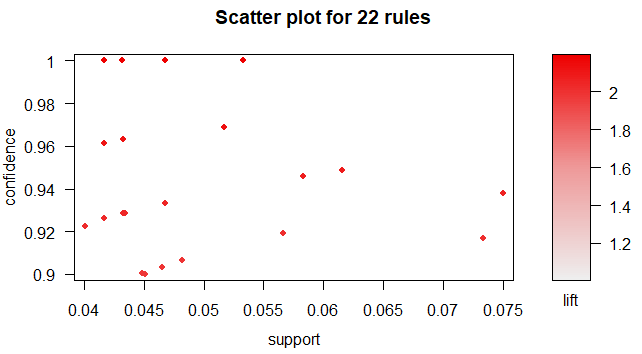
Two approaches were taken to mining for rules using the apriori function. First, adjusting support and confidence to get a target of 20 strong rules. The initial values for support started at 0.001 and confidence of 1. This produced over 1,000,000 rules. Support was adjusted up to 0.1, which produced 1 rule. Confidence was then adjusted down up to 0.8, consistently checking the number of rules at each increment. The final values that produced the target number of rules (19) were support of 0.15 and confidence of 0.8.



The second approach was to set the right hand side (RHS) to be instances were PEP was ‘YES’. The target set of rules, like the first attempt, was seeking approximately 20 strong rules. Starting values were support at 0.001 and confidence at 1. After much experimentation adjusting support, it became clear that a good value for the RHS condition would be 0.01.



This was due to a lower value produced too many rules, and a higher value didn’t produce enough. Confidence was consistently raised up to 1, which still produced over 1000 rules. Raising confidence after this point by increments of 0.01 produced the final values for support, 0.04, and confidence, 0.9, to produce 22 rules.



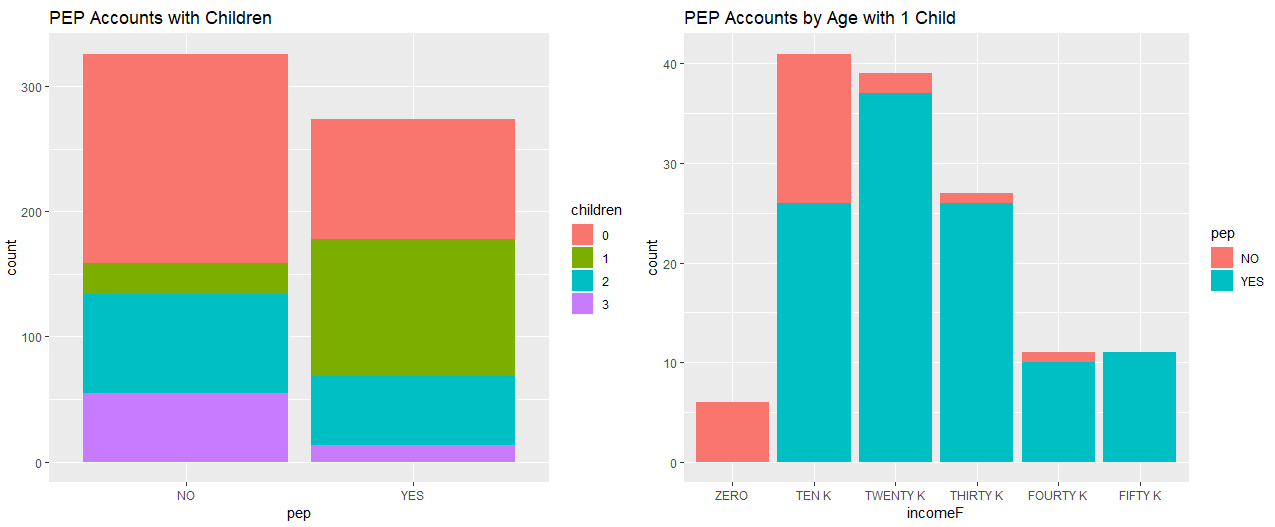
When examining the 41 rules produced, the ones that are of most interest are those that can assist to target customers for new products (PEP in this instance) or track potential customers for existing products.

|  |  |  |  |
| --- | --- | --- | --- |
| **Rule** | **Supp** | **Conf** | **Lift** |
| {children=1} => {pep=YES} | 0.18 | 0.81 | 1.8 |
| {children=1,incomeF=THIRTY K} => {pep=YES} | 0.043 | 0.96 | 2.1 |
| {children=1,incomeF=TWENTY K} => {pep=YES} | 0.062 | 0.95 | 2.1 |
| {married=YES,children=1,save\_act=YES,current\_act=YES} => {pep=YES} | 0.073 | 0.92 | 2 |
| {married=YES,children=0,save\_act=YES} => {pep=NO} | 0.18 | 0.9 | 1.7 |
| {car=NO,save\_act=YES,mortgage=NO} => {current\_act=YES} | 0.17 | 0.81 | 1.1 |

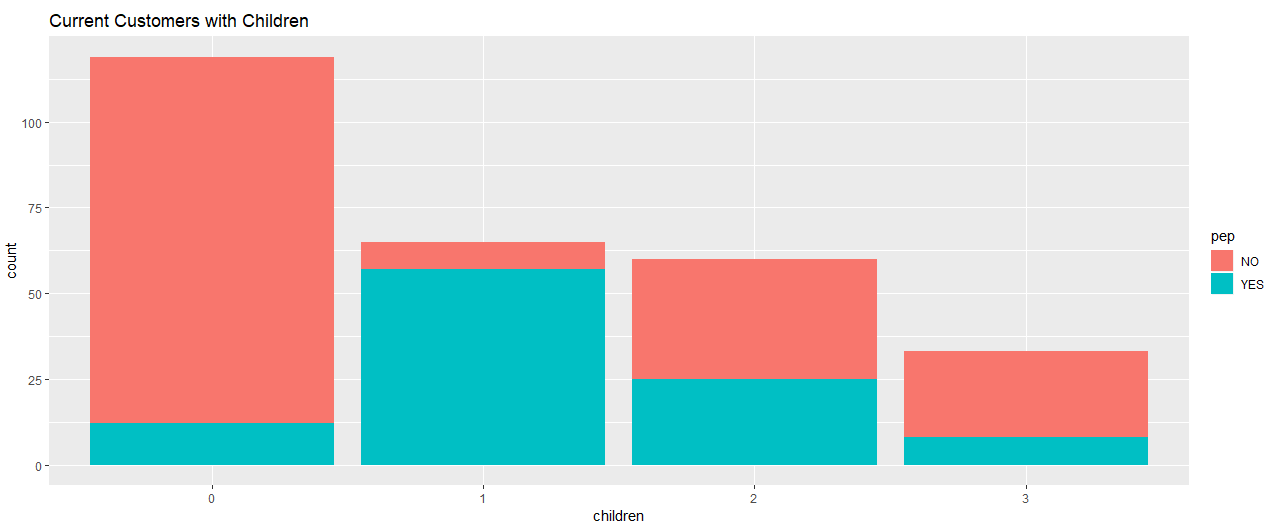
The first rule ({children=1} => {pep=YES}) represents that if a customer has one child, they are likely to have enrolled in a PEP. This is interesting because it is only one child that makes the PEP more likely. Of the 19 rules produced during this analysis, none of them had more than one child.

The rules ({children=1,incomeF=THIRTY K} => {pep=YES}) and ({children=1,incomeF=TWENTY K} => {pep=YES}) show that income is also a factor in a customer’s choice to enroll in a PEP. More specifically, customers on the lower half of the income spectrum.

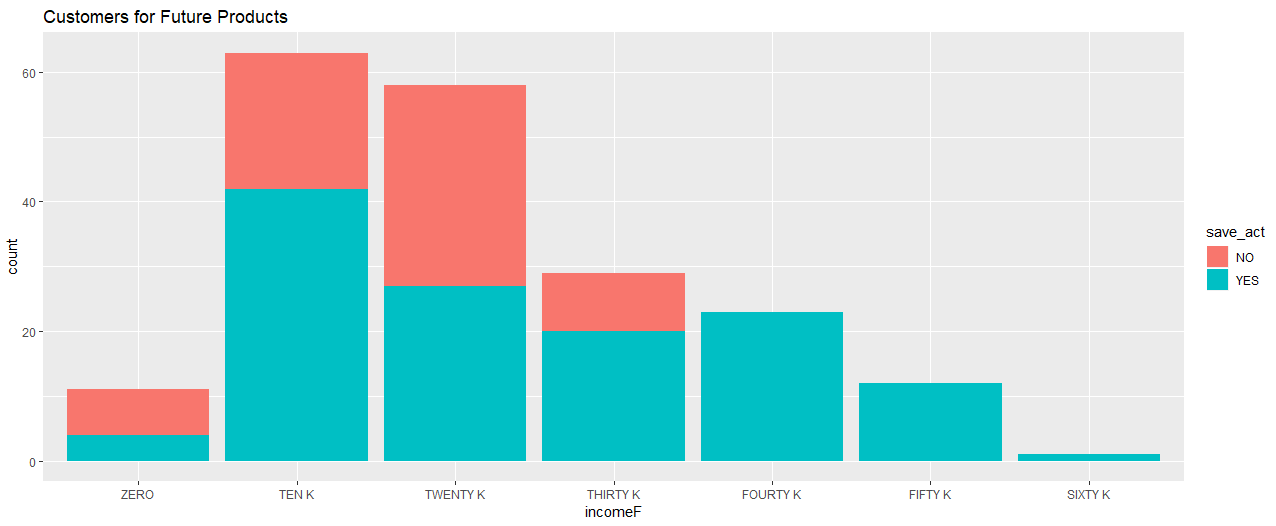
Looking at these two rules visually, it can be seen that most PEP customers do have 1 child, and among customers with 1 child, the higher there income, the more likely they are to have a PEP.



Two interesting rules that are connected are ({married=YES,children=1,save\_act=YES,current\_act=YES} => {pep=YES}) and ({married=YES,children=0,save\_act=YES} => {pep=NO}) because they can be used to target customers for PEPs today and tomorrow. Having one child is one of the best ways to determine the likelihood of opening a PEP. With the addition of marriage, a savings account, and a current account, both lift and confidence rise. The customers that are most similar, but have no child, are not likely to have a PEP. However, these are customers the bank should monitor for when that variable changes, they will move from a segment that is not likely to enroll in a PEP (0.9 confidence) to the most likely (0.92 confidence).



The last rule ({car=NO,save\_act=YES,mortgage=NO} => {current\_act=YES}) shows customers that may be targeted for future products like home and car loans. They already have a relationship with the bank, and because they have a savings account, the bank will have financial data to access possible future loans.



**Conclusion**

The bank has many opportunities to expand its business within its customers. There are specific attributes that make a customer more likely to enroll in a PEP. These customers tend to be married, with one child, and have an active account with the bank. Their income also plays a role, as the more earning power they have, the more likely they are to have an account. Instead of blindly targeting all customers, the marketing department can target customers that fit these attributes to target marketing for PEP.

The biggest potential lies in the possibility to target customers for products they will want in the future. There are two segments that were identified. One of the rules identified customers who are not car or home owners. Loans are a key product a bank can offer its customers. Identifying customers that are absent of products (car loans, mortgages, savings accounts) are good targets for the marketing department. Even if customers aren’t ready for applying for a loan, the occasional ad will keep the bank in mind when they are ready.

Another segment are customers that will go from not likely to most likely to invest in a PEP is identified as part of this analysis. Those are current customers with active accounts who are married. This segment is one of the least likely to invest in a PEP, but once they add a child to their family they become one of the highest likely for that product. This is the prime point to target this segment, as with time and the growth of their family, they become less likely to invest in the product.

The main takeaway from this analysis is that the bank has enough data to distinguish which customers to target for specific products, which customers may be interested in current products, and which customers will be interested in a product in the future.