IST 565 Data Mining

Olivieri

HW 3

Association Rule Discovery for PEP

## Explanation

Association rules are structured in a left-hand-side and a right-hand-side:

{lhs} -> {rhs}

A common use of association rules reveals purchasing patterns of customers. For example, a grocery store may find the following rules in their data analysis:

{beer} -> {diapers}

{beer} -> {condiments, hot dog buns}

{beer} -> {diapers, paper towels}

{root vegetables, other vegetables} -> {beef stock}

The first rule shows that people who purchased beer also often purchased diapers in the same transaction. The last rule shows that people who purchased various vegetables often purchased beef stock as well. Customer behavior can be hypothesized from the resulting rules. The last rule suggests customers who purchased certain vegetables may be making soup. Marketing plans can be derived from these rules.

## Effectiveness of Rules

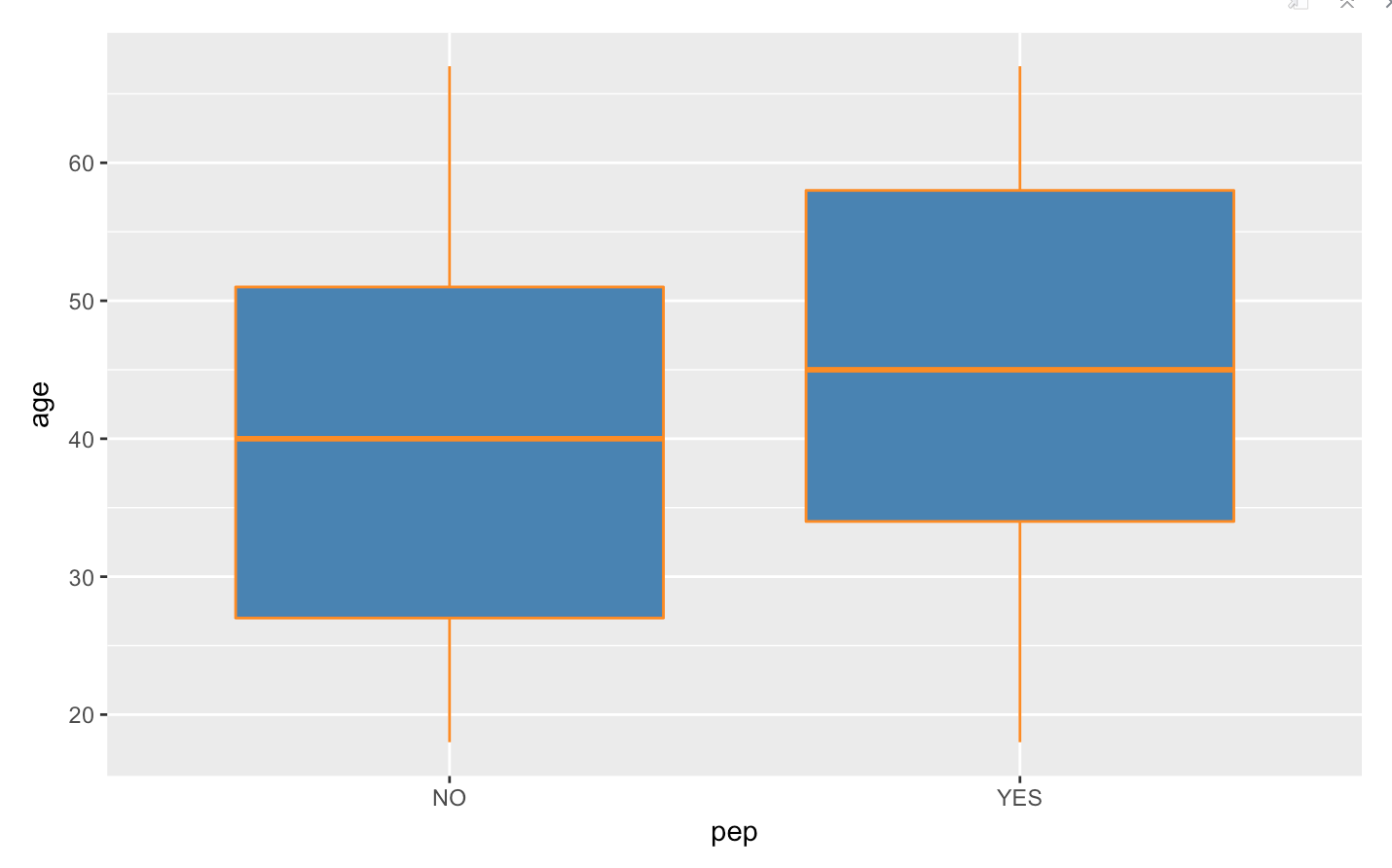
A rule can be evaluated for its possible effectiveness based on three main characteristics: support, confidence, and lift. Support is the total number of times items were purchased together out of all of the transactions. From the examples above, the rule {beer} -> {diapers} appears twice out of four transactions. This rule’s support count is 2. Confidence is a measure of how often a left-hand-side results in a right-hand-side. In the rules above, beer is purchased three times, but only results in diapers being purchased twice. The {beer} -> {diapers} confidence is 2/3. Lift is a calculation of correlation: out of all of the transactions, what is the probability that the combination of purchased items (a rule) will occur. The equation for lift is: lift (XY) = support(xy) / support(x) \* support(y). Without getting too deep into the math, the higher a resultant lift value the better. Lift should be above 1 in all rules.

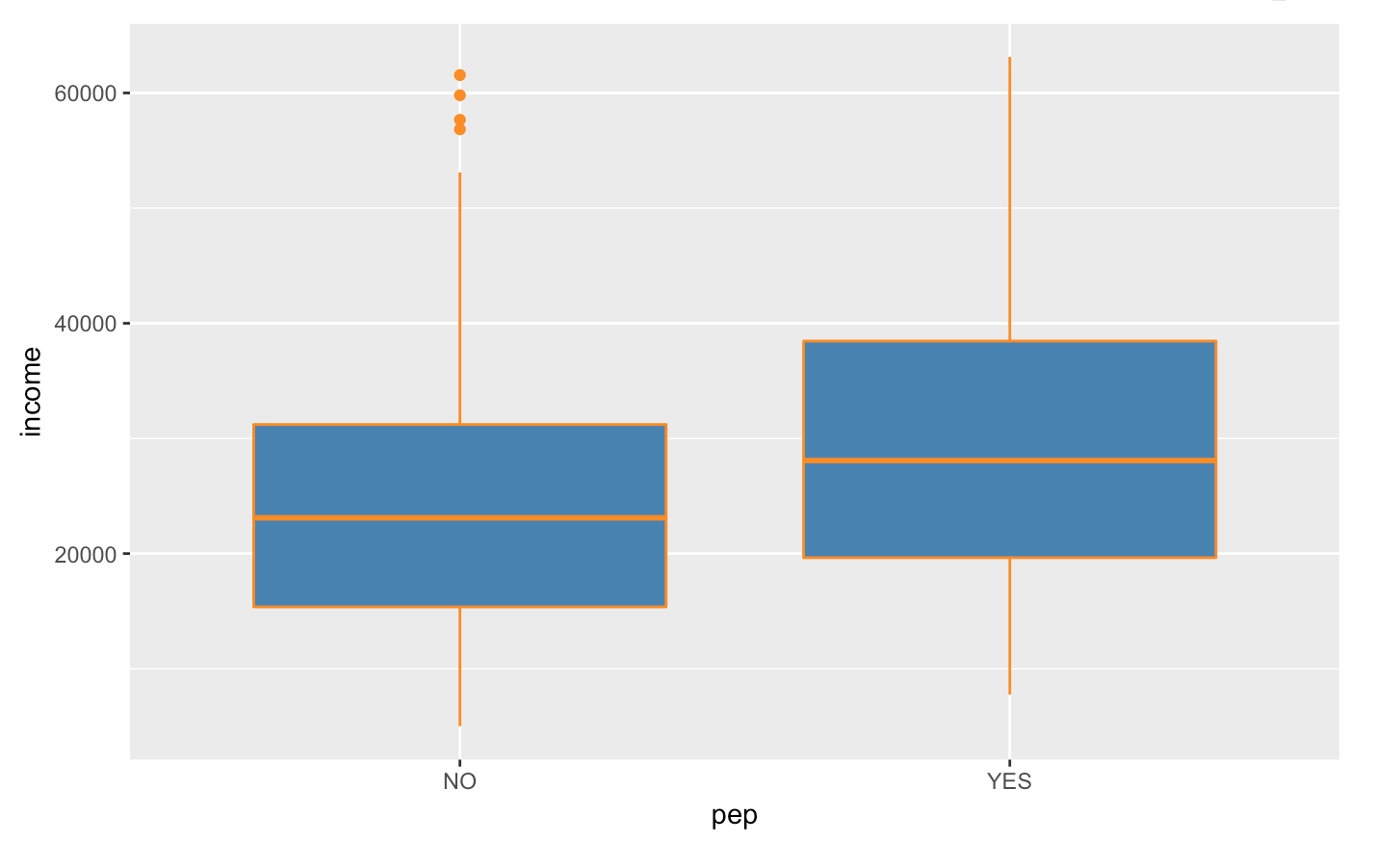
## Data Processing

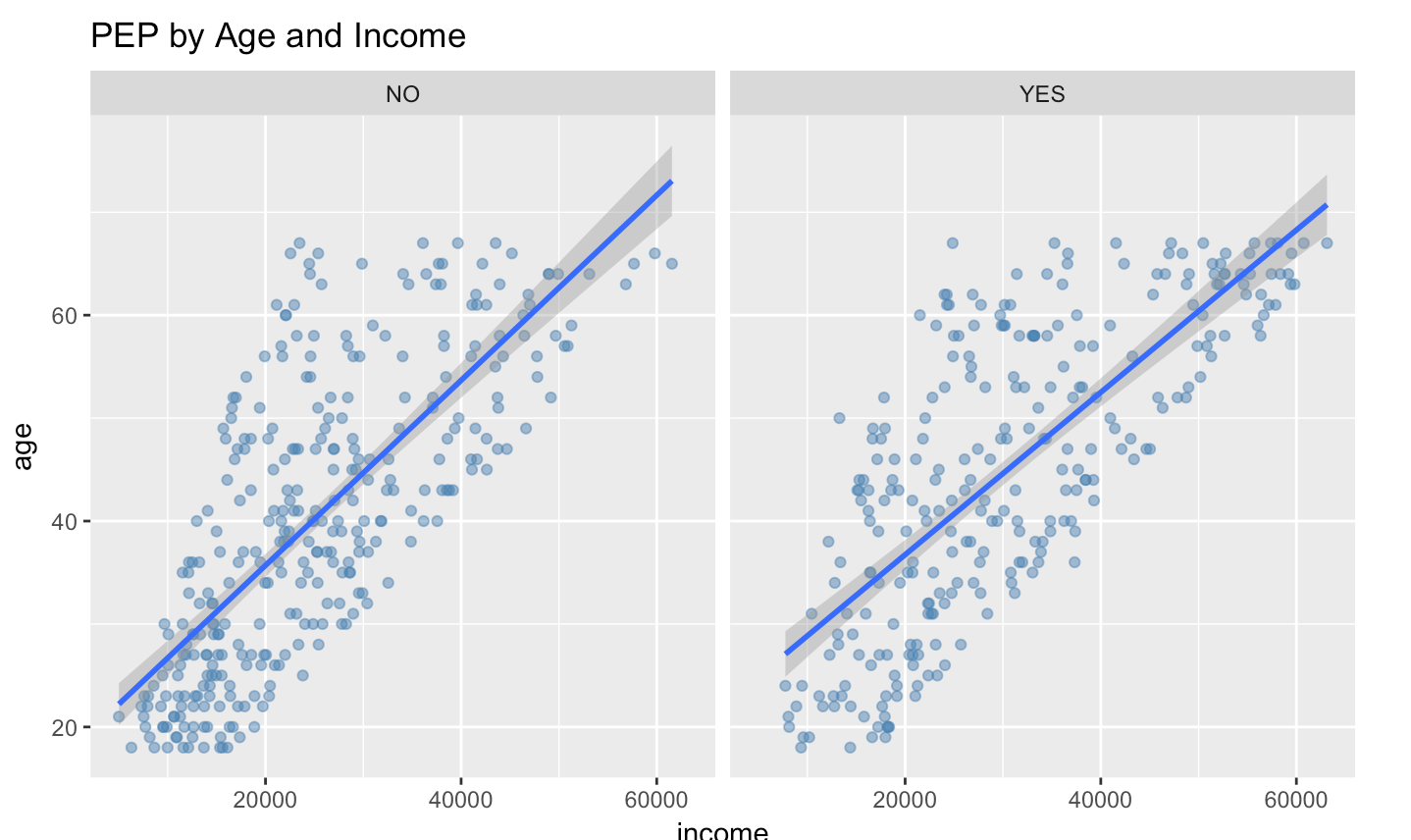
The bank data provided was brought into R Studio and processed to allow for the Rules Association analysis. The bank customer id was removed and record data was transformed into transactional data by normalizing plain-language results into variables which the programming language R can use in the analyses.

## Basic Customer Information

Before creating association rules, the data provided can reveal overall demographics of the bank’s customers. These demographics analyses can also provide business insight. For example, customers who purchased a PEP tend to be slightly older and have a higher income than those who did not purchase a PEP.







## Association Rules

Running the apriori algorithm in R resulted in interesting rule sets which also show demographics of the bank’s customers regardless of whether they purchased a PEP or not. The results were ordered for confidence, support, and any rules which had a low lift were removed. The remaining rules showed actionable insights. For example, if a bank customer lived in a rural region and had a high income, they tended to have a savings account. And if a customer is older and had a high income they also tended to have a savings account:

{region=RURAL, income=high} => {save\_act=YES}

{age=older, income=high} => {save\_act=YES}

This rule, while not applicable to PEPs, still offers the bank opportunities to cross-sell rural or older individuals of higher income a savings account.

Ordering the results by highest confidence and highest support yields these top 25 results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| lhs | rhs | support | conf | lift |
| {married=YES, mortgage=NO, age=older} | {save\_act=YES} | 0.11 | 0.82 | 1.19 |
| {married=NO, age=young} | {current\_act=YES} | 0.09 | 0.82 | 1.08 |
| {married=NO, car=NO, children=low} | {current\_act=YES} | 0.09 | 0.82 | 1.08 |
| {sex=FEMALE, married=YES, age=middle} | {children=low} | 0.09 | 0.82 | 1.24 |
| {age=older, income=high, children=low} | {current\_act=YES} | 0.08 | 0.82 | 1.08 |
| {car=NO, age=older, children=low} | {current\_act=YES} | 0.08 | 0.82 | 1.08 |
| {region=INNER\_CITY, save\_act=YES, age=young} | {current\_act=YES} | 0.08 | 0.82 | 1.08 |
| {sex=FEMALE, age=young, children=low} | {current\_act=YES} | 0.08 | 0.82 | 1.08 |
| {sex=FEMALE, region=INNER\_CITY, married=YES} | {current\_act=YES} | 0.12 | 0.82 | 1.08 |
| {married=NO, age=older} | {current\_act=YES} | 0.09 | 0.82 | 1.08 |
| {current\_act=NO, mortgage=NO, pep=NO} | {married=YES} | 0.07 | 0.82 | 1.25 |
| {car=YES, pep=YES, age=middle} | {children=low} | 0.07 | 0.82 | 1.24 |
| {sex=MALE, married=NO, save\_act=YES} | {current\_act=YES} | 0.09 | 0.82 | 1.09 |
| {current\_act=YES, mortgage=YES, pep=YES} | {children=low} | 0.09 | 0.82 | 1.24 |
| {mortgage=NO, pep=YES, income=medium} | {current\_act=YES} | 0.09 | 0.82 | 1.09 |
| {married=NO, car=NO} | {current\_act=YES} | 0.14 | 0.82 | 1.09 |
| {car=NO, mortgage=NO, pep=NO} | {married=YES} | 0.15 | 0.82 | 1.25 |
| {married=NO, save\_act=YES} | {current\_act=YES} | 0.19 | 0.82 | 1.09 |
| {sex=FEMALE, region=INNER\_CITY, car=NO} | {current\_act=YES} | 0.09 | 0.83 | 1.09 |
| {current\_act=YES, mortgage=YES, pep=NO} | {save\_act=YES} | 0.12 | 0.83 | 1.20 |
| {sex=FEMALE,region=INNER\_CITY,save\_act=YES} | {current\_act=YES} | 0.12 | 0.83 | 1.09 |
| {sex=FEMALE, married=NO, save\_act=YES} | {current\_act=YES} | 0.10 | 0.83 | 1.09 |
| {mortgage=NO, age=older, children=low} | {current\_act=YES} | 0.10 | 0.83 | 1.09 |
| {sex=FEMALE, age=young, income=low} | {current\_act=YES} | 0.11 | 0.83 | 1.09 |
| {region=TOWN, married=NO} | {current\_act=YES} | 0.08 | 0.83 | 1.09 |

## PEP Rules

An analysis showing rules in which a PEP occurs in the right-hand-side of a rule (a rule resulting in a PEP purchase) offers the following top five rule sets ordered by confidence, lift, and support. The PEP rule which has the highest support, high confidence, and highest lift is:

{married=NO, mortgage=NO, children=low} => {pep=YES}, sup .117, conf .92, lift 2.0

Support for this rule is .117, the confidence is .92, and lift is 2.0. While the support seems low, lower support numbers are typical in larger datasets which can derive many variations of rules. The confidence is a measure of how often the left-hand-side results in a PEP purchase, meaning that if a customer is married with no mortgage and no children, he or she purchases a PEP 92% of the time. A lift of 2.0 is high, meaning this rule has good probability of working as a marketing tool.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| lhs | rhs | support | conf | lift |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| {married=NO, mortgage=NO, income=high} | {pep=YES} | 0.040 | 0.96 | 2.1 |
| {married=NO, mortgage=NO, children=low} | {pep=YES} | 0.117 | 0.92 | 2.0 |
| {age=older, income=high, children=medium} | {pep=YES} | 0.035 | 0.91 | 2.0 |
| {income=high, children=medium} | {pep=YES} | 0.042 | 0.89 | 2.0 |
| {save\_act=YES, income=high, children=medium} | {pep=YES} | 0.042 | 0.89 | 2.0 |

## Summary

The data provided can offer interesting insights. For example, many of the left-hand-side rules have “{children = medium}”. These are families or individuals who have children, but fewer than three. Behavior derived from this analysis might indicate that small families are planning for the future by purchasing savings accounts, mortgages, or PEPs. The highest-rated PEP rule from above shows that non-married customers who have no mortgage and few-to-no children also purchase a PEP. Which could indicate that younger customers are more interested in financial planning than intuition would suggest.

The basic demographics analysis done before the Rules Association analysis shows that most PEP customers are older and have a higher income. The PEP rules themselves show that younger, non-married customers with no mortgage often purchase a PEP. This demographic may have been previously underestimated.

The overarching goal of rule association analysis is to segment customers into actionable demographics; i.e. identify customers who may be open to bank product opportunities. The following graph shows customer segments who did and did not purchase PEPs. The actionable segments are those customers who have not yet purchased PEPs but also inhabit the same space as other customers who have. Customers who have purchased PEPs are in orange while customers who have not are in blue. The PEP Rules from above allow the chance to capture these segments.

