Shivangi Mundhra
SUID - 842548148

Preprocessing the data

When we loaded the file, below was the structure of the dataset.

```
'data.frame':   600 obs. of  12 variables:
 $ id         : chr  "ID12101" "ID12102" "ID12103" "ID12104" ...
 $ age        : int  48 40 51 23 57 57 22 58 37 54 ...
 $ sex        : chr  "FEMALE" "MALE" "FEMALE" "FEMALE" ...
 $ region     : chr  "INNER_CITY" "TOWN" "INNER_CITY" "TOWN" ...
 $ income     : num  17546 30085 16575 20375 50576 ...
 $ married    : chr  "NO" "YES" "YES" "YES" ...
 $ children   : int  1 3 0 3 0 2 0 0 2 2 ...
 $ car        : chr  "NO" "YES" "YES" "NO" ...
 $ save_act   : chr  "NO" "NO" "YES" "NO" ...
 $ current_act: chr  "NO" "YES" "YES" "YES" ...
 $ mortgage   : chr  "NO" "YES" "NO" "NO" ...
 $ pep        : chr  "YES" "NO" "NO" "NO" ...
```

We remove the id field, convert the character fields to factors and discretize the numeric fields, age and income. Below is the resulting dataset after preprocessing. This data is now ready for association rule mining.

```
'data.frame':   600 obs. of  11 variables:
 $ age        : Factor w/ 3 levels "[18,35)","[35,49)",..: 2 2 3 1 3 3 1 3 2 3 ...
  ..- attr(*, "discretized:breaks")= num [1:4] 18 35 49 67
  ..- attr(*, "discretized:method")= chr "frequency"
 $ sex        : Factor w/ 2 levels "FEMALE","MALE": 1 2 1 1 1 1 2 2 1 2 ...
 $ region     : Factor w/ 4 levels "INNER_CITY","RURAL",..: 1 4 1 4 2 4 2 4 3 4 ...
 $ income     : Factor w/ 3 levels "[5.01e+03,2.03e+04)",..: 1 2 1 2 3 3 1 2 2 2 ...
  ..- attr(*, "discretized:breaks")= num [1:4] 5014 20254 31133 63130
  ..- attr(*, "discretized:method")= chr "frequency"
 $ married    : Factor w/ 2 levels "NO","YES": 1 2 2 2 2 2 1 2 2 2 ...
 $ children   : Factor w/ 4 levels "0","1","2","3": 2 4 1 4 1 3 1 1 3 3 ...
 $ car        : Factor w/ 2 levels "NO","YES": 1 2 2 1 1 1 1 2 2 2 ...
 $ save_act   : Factor w/ 2 levels "NO","YES": 1 1 2 1 2 2 1 2 1 2 ...
 $ current_act: Factor w/ 2 levels "NO","YES": 1 2 2 2 1 2 2 2 1 2 ...
 $ mortgage   : Factor w/ 2 levels "NO","YES": 1 2 1 1 1 1 1 1 1 1 ...
 $ pep        : Factor w/ 2 levels "NO","YES": 2 1 1 1 1 2 2 1 1 1 ...
```

Parameters and experiments to obtain strong rules

We convert the data into transaction format after it has been transformed, with only factor variables as inputs, because rules can only be mined from nominal data.

Shivangi Mundhra
SUID - 842548148

We try supp = 0.001, conf = 0.9 which results in too many records of rules to inspect. We try a couple different combinations.

Next, we set pep (YES) as the right-hand side (RHS) of the rules, and see what rules are generated.

```
```{r}
pepRules <- apriori(data = bankData, parameter = list(supp = 0.08, conf = 0.6, minlen = 3),
control = list(verbose = F), appearance = list(default = 'lhs', rhs = c('pep=YES')))
pepRulesSorted <- sort(pepRules, by = "lift", descending = TRUE)
inspect(pepRulesSorted)
```
```

Below shows a snapshot of results, the entire result set is in the code file.

```
     lhs                              rhs         support confidence coverage lift count
[1]  {married=YES,
      children=1,
      save_act=YES}                => {pep=YES}   0.095        0.88     0.108  1.9    57
[2]  {children=1,
      save_act=YES,
      mortgage=NO}                 => {pep=YES}   0.080        0.87     0.092  1.9    48
[3]  {children=1,
      save_act=YES,
      current_act=YES}             => {pep=YES}   0.105        0.86     0.122  1.9    63
[4]  {married=YES,
      children=1,
      current_act=YES}             => {pep=YES}   0.093        0.86     0.108  1.9    56
[5]  {children=1,
      mortgage=NO}                 => {pep=YES}   0.118        0.85     0.140  1.9    71
[6]  {children=1,
      save_act=YES}                => {pep=YES}   0.133        0.84     0.158  1.8    80
[7]  {children=1,
      current_act=YES,
      mortgage=NO}                 => {pep=YES}   0.095        0.84     0.113  1.8    57
[8]  {sex=FEMALE,
      children=1}                  => {pep=YES}   0.092        0.83     0.110  1.8    55
[9]  {children=1,
      current_act=YES}             => {pep=YES}   0.140        0.83     0.168  1.8    84
[10] {married=YES,
      children=1}                  => {pep=YES}   0.123        0.83     0.148  1.8    74
[11] {children=1,
      car=YES}                     => {pep=YES}   0.092        0.82     0.112  1.8    55
[12] {children=1,
      car=NO}                      => {pep=YES}   0.092        0.81     0.113  1.8    55
[13] {sex=MALE,
      children=1}                  => {pep=YES}   0.092        0.80     0.115  1.7    55
[14] {region=INNER_CITY,
      children=1}                  => {pep=YES}   0.085        0.78     0.108  1.7    51
[15] {income=[3.11e+04,6.31e+04],
      married=NO}                  => {pep=YES}   0.093        0.78     0.120  1.7    56
[16] {married=NO,
      save_act=YES,
```

5 interesting results

1. Customers with 1 child and both a saving and current account are buying PEP
    a. Support = 0.105, confidence = 0.86, lift = 1.9
    b. This should be a low hanging fruit. Company can easily focus on customers with both types of accounts to increase their client base.
2. Customers with a child and a saving account that do not have a mortgage will be more willing to buy PEP
    a. Support = 0.080, confidence = 0.87, lift = 1.9
3. Customers with a child and a saving account that do not have a mortgage will be more willing to buy PEP

  a. support = 0.095, confidence = 0.84, lift = 1.8

From the two examples above, the company can focus on those that have a kid and an account but no mortgage.

4. Customers that are not married and have no mortgage tend to buy PEP
   a. Support = 0.153, confidence = 0.71, lift = 1.5
   b. The company can target these customers
5. Individuals with medium income range and no mortgage tend to be interested as well
   a. Support = 0.145, confidence = 0.65, lift = 1.4