German Credit Card Data

Sreekanth Settur

7 January 2019

Presenting the summary of German credit dataset. Currently considered for analysis

This is downloaded from AzureML studio for analysis and peforming predictions

```
##   A11            X6         A34           A43          X1169
##  A11:273   Min.   : 4.00   A30: 40   A43    :279   Min.   :  250
##  A12:269   1st Qu.:12.00   A31: 49   A40    :234   1st Qu.: 1368
##  A13: 63   Median :18.00   A32:530   A42    :181   Median : 2320
##  A14:394   Mean   :20.92   A33: 88   A41    :103   Mean   : 3273
##            3rd Qu.:24.00   A34:292   A49    : 97   3rd Qu.: 3972
##            Max.   :72.00             A46    : 50   Max.   :18424
##                                      (Other): 55
##   A65       A75             X4          A93        A101          X4.1
##  A61:603   A71: 62   Min.   :1.000   A91: 50   A101:906   Min.   :1.000
##  A62:103   A72:172   1st Qu.:2.000   A92:310   A102: 41   1st Qu.:2.000
##  A63: 63   A73:339   Median :3.000   A93:547   A103: 52   Median :3.000
##  A64: 48   A74:174   Mean   :2.972   A94: 92              Mean   :2.844
##  A65:182   A75:252   3rd Qu.:4.000                        3rd Qu.:4.000
##                      Max.   :4.000                        Max.   :4.000
##
##    A121          X67           A143       A152           X2
##  A121:281   Min.   :19.00   A141:139   A151:179   Min.   :1.000
##  A122:232   1st Qu.:27.00   A142: 47   A152:712   1st Qu.:1.000
##  A123:332   Median :33.00   A143:813   A153:108   Median :1.000
##  A124:154   Mean   :35.51                         Mean   :1.406
##             3rd Qu.:42.00                         3rd Qu.:2.000
##             Max.   :75.00                         Max.   :4.000
##
##    A173           X1           A192       A201          X1.1
##  A171: 22   Min.   :1.000   A191:596   A201:962   Min.   :1.0
##  A172:200   1st Qu.:1.000   A192:403   A202: 37   1st Qu.:1.0
##  A173:629   Median :1.000                         Median :1.0
##  A174:148   Mean   :1.155                         Mean   :1.3
##             3rd Qu.:1.000                         3rd Qu.:2.0
##             Max.   :2.000                         Max.   :2.0
##
```

Checking if data has any NA values.

As per below snapshot, there are no missing values in the dataset.

```
##   A11    X6   A34   A43 X1169   A65   A75    X4   A93  A101  X4.1  A121
##     0     0     0     0     0     0     0     0     0     0     0     0
##   X67  A143  A152    X2  A173    X1  A192  A201  X1.1
##     0     0     0     0     0     0     0     0     0
```

listing columns in original data

Printing column names after providing appropriate names. This is done reference to data dictionary

```
##  [1] "ExistingAccStatus"      "DurationInMonths"
##  [3] "CreditHistory"          "Purpose"
##  [5] "CreditAmount"           "SavingsAccStatus"
##  [7] "DurationInCurrentComp"  "InterestRate"
##  [9] "StatusAndGender"        "guarantors"
## [11] "DurationInCurrentHouse" "Property"
## [13] "AgeInYears"             "OthrInstallmantPlans"
## [15] "Housing"                "CreditsAtThisBK"
## [17] "Profession"             "PeopleLiabled"
## [19] "Telephone"              "ForeignWorker"
## [21] "CustomerClass"
```

Summary of German credit card data

```
##  ExistingAccStatus DurationInMonths CreditHistory    Purpose
##  A11:273           Min.   : 4.00    A30: 40       A43    :279
##  A12:269           1st Qu.:12.00    A31: 49       A40    :234
##  A13: 63           Median :18.00    A32:530       A42    :181
##  A14:394           Mean   :20.92    A33: 88       A41    :103
##                    3rd Qu.:24.00    A34:292       A49    : 97
##                    Max.   :72.00                  A46    : 50
##                                                   (Other): 55
##   CreditAmount   SavingsAccStatus DurationInCurrentComp  InterestRate
##  Min.   :  250   A61:603          A71: 62               Min.   :1.000
##  1st Qu.: 1368   A62:103          A72:172               1st Qu.:2.000
##  Median : 2320   A63: 63          A73:339               Median :3.000
##  Mean   : 3273   A64: 48          A74:174               Mean   :2.972
##  3rd Qu.: 3972   A65:182          A75:252               3rd Qu.:4.000
##  Max.   :18424                                         Max.   :4.000
##
##  StatusAndGender guarantors DurationInCurrentHouse Property
##  A91: 50         A101:906   Min.   :1.000          A121:281
##  A92:310         A102: 41   1st Qu.:2.000          A122:232
##  A93:547         A103: 52   Median :3.000          A123:332
##  A94: 92                    Mean   :2.844          A124:154
##                             3rd Qu.:4.000
##                             Max.   :4.000
##
##    AgeInYears    OthrInstallmantPlans Housing    CreditsAtThisBK
##  Min.   :19.00   A141:139             A151:179   Min.   :1.000
##  1st Qu.:27.00   A142: 47             A152:712   1st Qu.:1.000
##  Median :33.00   A143:813             A153:108   Median :1.000
##  Mean   :35.51                                   Mean   :1.406
##  3rd Qu.:42.00                                   3rd Qu.:2.000
##  Max.   :75.00                                   Max.   :4.000
##
##  Profession PeopleLiabled   Telephone  ForeignWorker CustomerClass
##  A171: 22   Min.   :1.000   A191:596   A201:962      Min.   :1.0
##  A172:200   1st Qu.:1.000   A192:403   A202: 37      1st Qu.:1.0
##  A173:629   Median :1.000                            Median :1.0
##  A174:148   Mean   :1.155                            Mean   :1.3
##             3rd Qu.:1.000                            3rd Qu.:2.0
##             Max.   :2.000                            Max.   :2.0
##
```

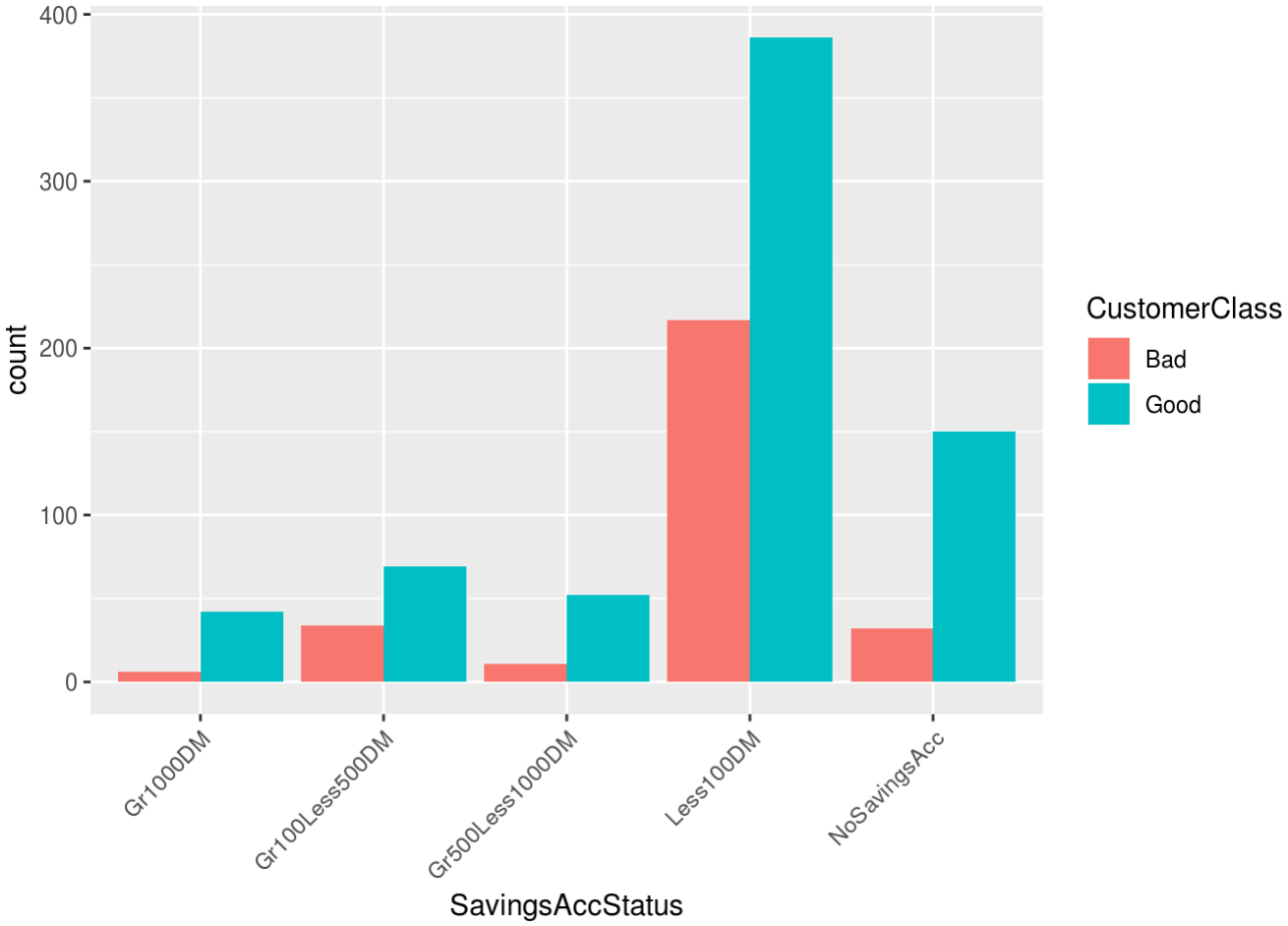
Creating a new copy of German Credit data by expanding with actual values.

This is done reference to data dictionary provided.

```
## 'data.frame':    999 obs. of  23 variables:
##  $ ExistingAccStatus     : Factor w/ 4 levels "LessThan0DM",..: 2 4 1 1 4 4 2 4 2 2 ...
##  $ DurationInMonths      : int  48 12 42 24 36 24 36 12 30 12 ...
##  $ CreditHistory         : Factor w/ 5 levels "CriticalAccount",..: 5 1 5 2 5 5 5 5 1 5 ...
##  $ Purpose               : Factor w/ 10 levels "business","Domestic Appliances",..: 8 3 4 5
3 4 6 8 5 5 ...
##  $ CreditAmount          : int  5951 2096 7882 4870 9055 2835 6948 3059 5234 1295 ...
##  $ SavingsAccStatus      : Factor w/ 5 levels "Gr1000DM","Gr100Less500DM",..: 4 4 4 4 5 3 4
1 4 4 ...
##  $ DurationInCurrentComp : Factor w/ 5 levels "Gr1YrLess4Yrs",..: 1 2 2 1 1 3 1 2 5 4 ...
##  $ InterestRate          : Factor w/ 4 levels "1","2","3","4": 2 2 2 3 2 3 2 2 4 3 ...
##  $ guarantors            : Factor w/ 3 levels "Co-Applicant",..: 3 3 2 3 3 3 3 3 3 3 ...
##  $ DurationInCurrentHouse: int  2 3 4 4 4 4 2 4 2 1 ...
##  $ Property              : Factor w/ 4 levels "LifeInsurance",..: 4 4 1 2 2 1 3 4 3 3 ...
##  $ AgeInYears            : int  22 49 45 53 35 53 35 61 28 25 ...
##  $ OthrInstallmantPlans  : Factor w/ 3 levels "Bank","None",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Housing               : Factor w/ 3 levels "ForFree","Own",..: 2 2 1 1 1 2 3 2 2 3 ...
##  $ CreditsAtThisBK       : Factor w/ 4 levels "1","2","3","4": 1 1 1 2 1 1 1 1 2 1 ...
##  $ Profession            : Factor w/ 4 levels "Officer","Official",..: 2 4 2 2 4 2 1 4 1 2
...
##  $ PeopleLiabled         : Factor w/ 2 levels "1","2": 1 2 2 2 2 1 1 1 1 1 ...
##  $ Telephone             : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 2 1 1 1 ...
##  $ ForeignWorker         : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ CustomerClass         : Factor w/ 2 levels "Bad","Good": 1 2 2 1 2 2 2 2 1 1 ...
##  $ Marital_Status        : Factor w/ 3 levels "Married","Separated",..: 1 3 3 3 3 3 3 2 1 1
...
##  $ Gender                : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 2 2 2 2 1 ...
##  $ integerClass          : int  2 1 1 2 1 1 1 1 2 2 ...
```

Plotting SavingsAccStatus

```
library(ggplot2)
ggplot2::ggplot(data = GermanCreditData,aes(SavingsAccStatus, fill = CustomerClass)) +
  stat_count(position = position_dodge()) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

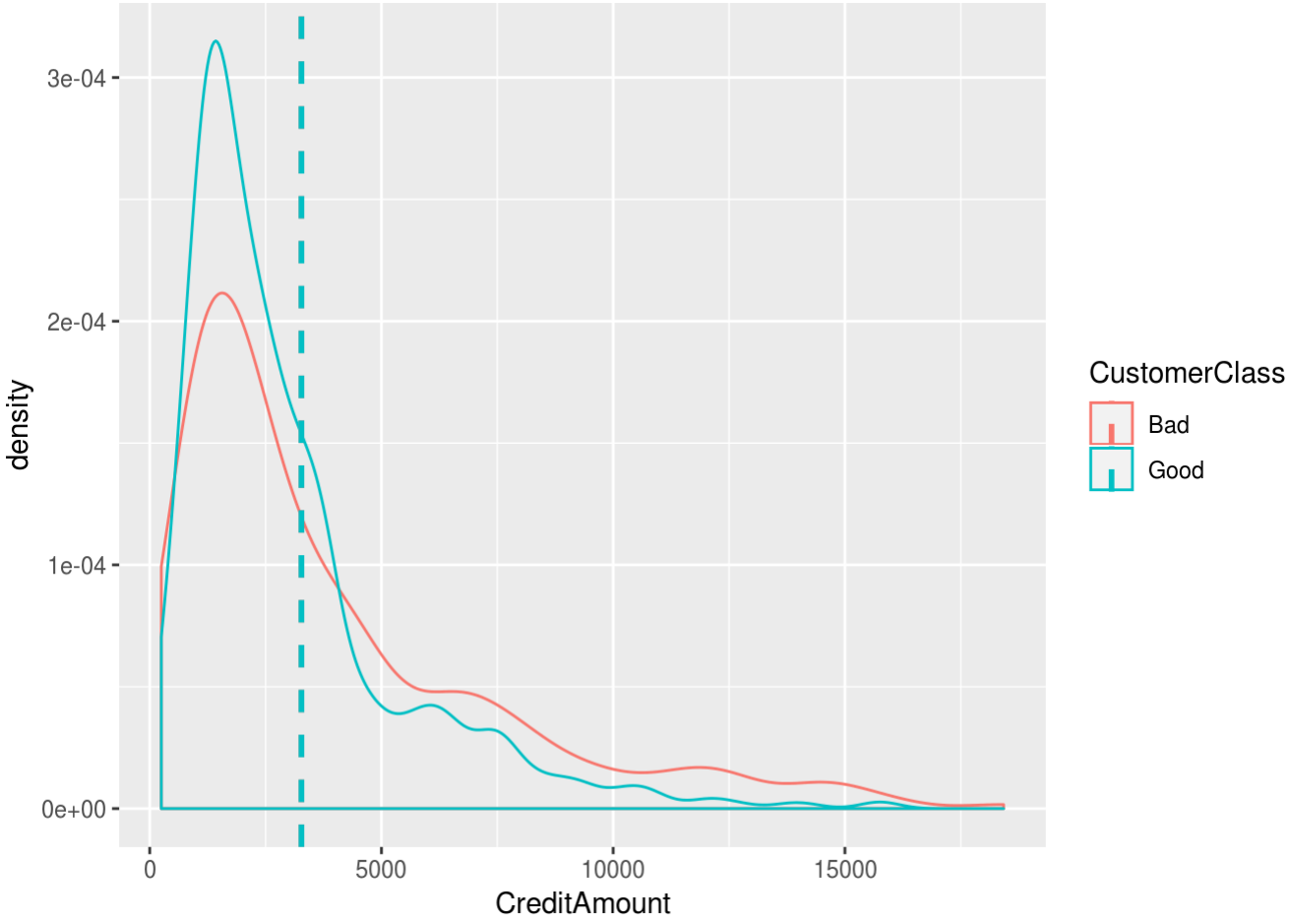


Performed density plot on credit amount to understand the spread of credit amount.

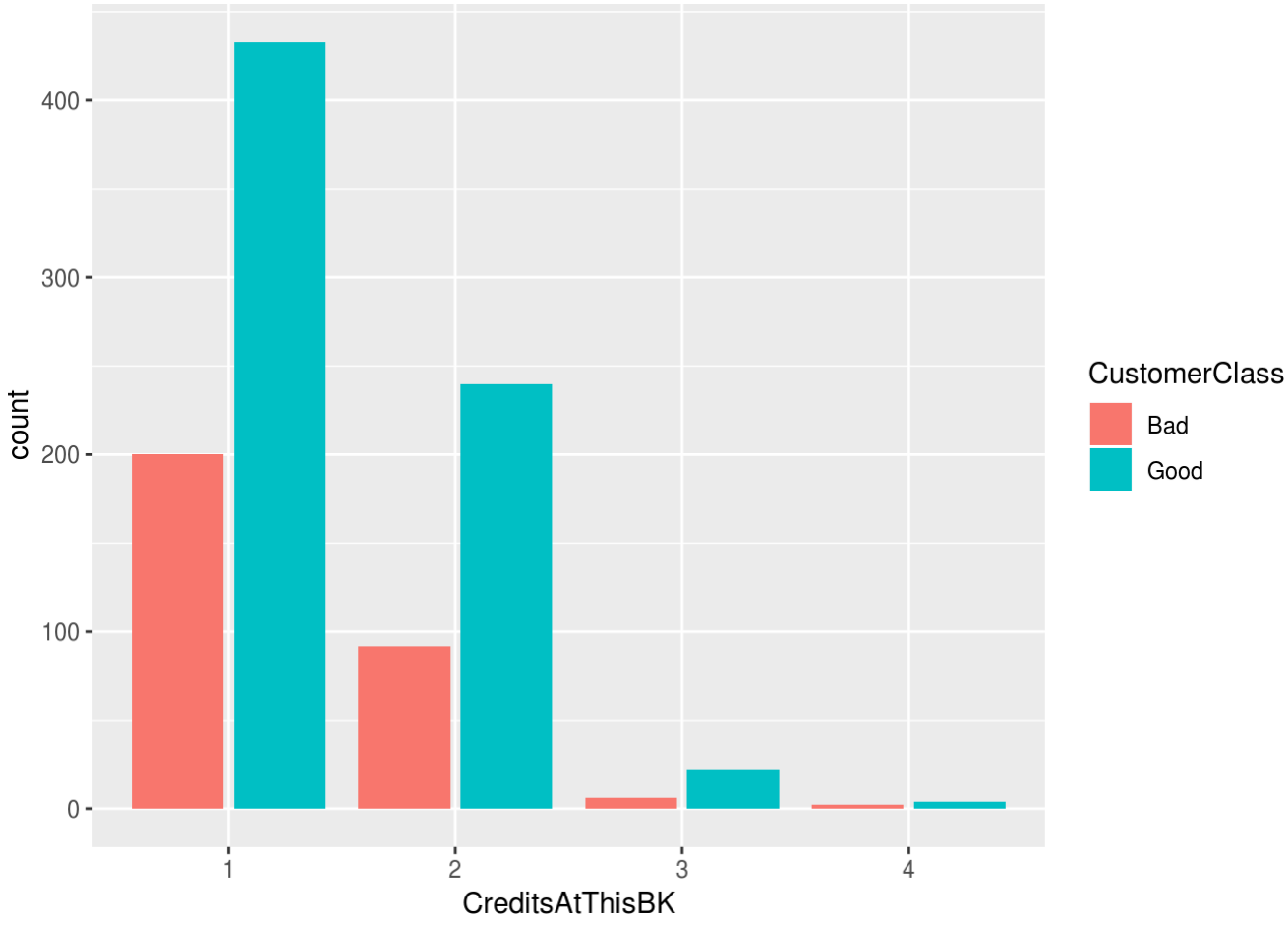
Also specified the mean for credit amount, to better understand the skewness of plot.

It seems that, credit amount for both “Good” and “Bad” seems to be similar and has same mean.

```
ggplot2::ggplot(data = GermanCreditData,aes(CreditAmount, col = CustomerClass)) +
  geom_density() +
  geom_vline(data = GermanCreditData,
             aes(xintercept=mean(CreditAmount), color = CustomerClass),
             linetype="dashed", size=1)
```

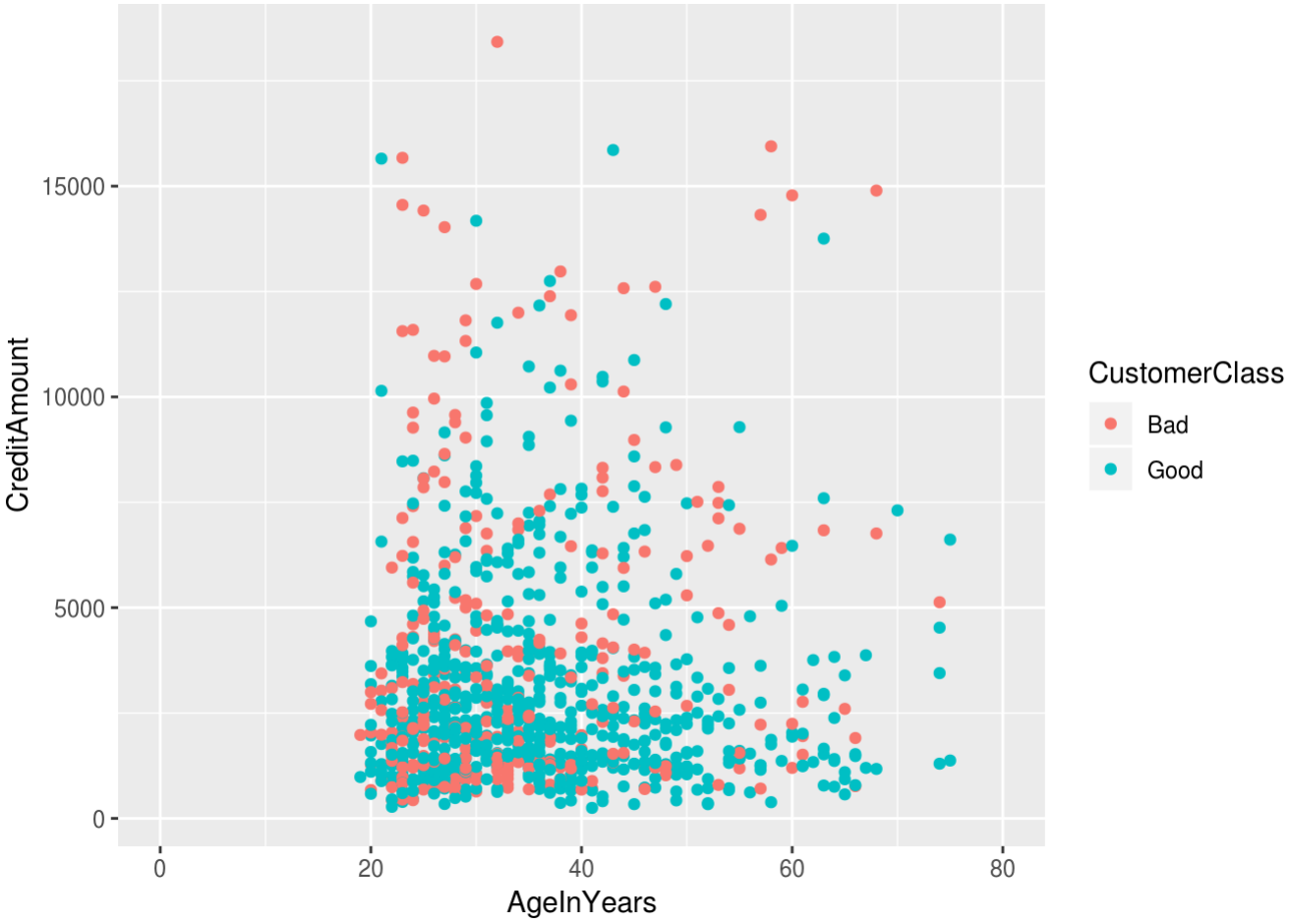



```
library(ggplot2)
ggplot2::ggplot(data = GermanCreditData, aes(CreditsAtThisBK, fill = CustomerClass)) +
 stat_count(position = position_dodge2())
```



Plotting bi-variate with Age and credit amount

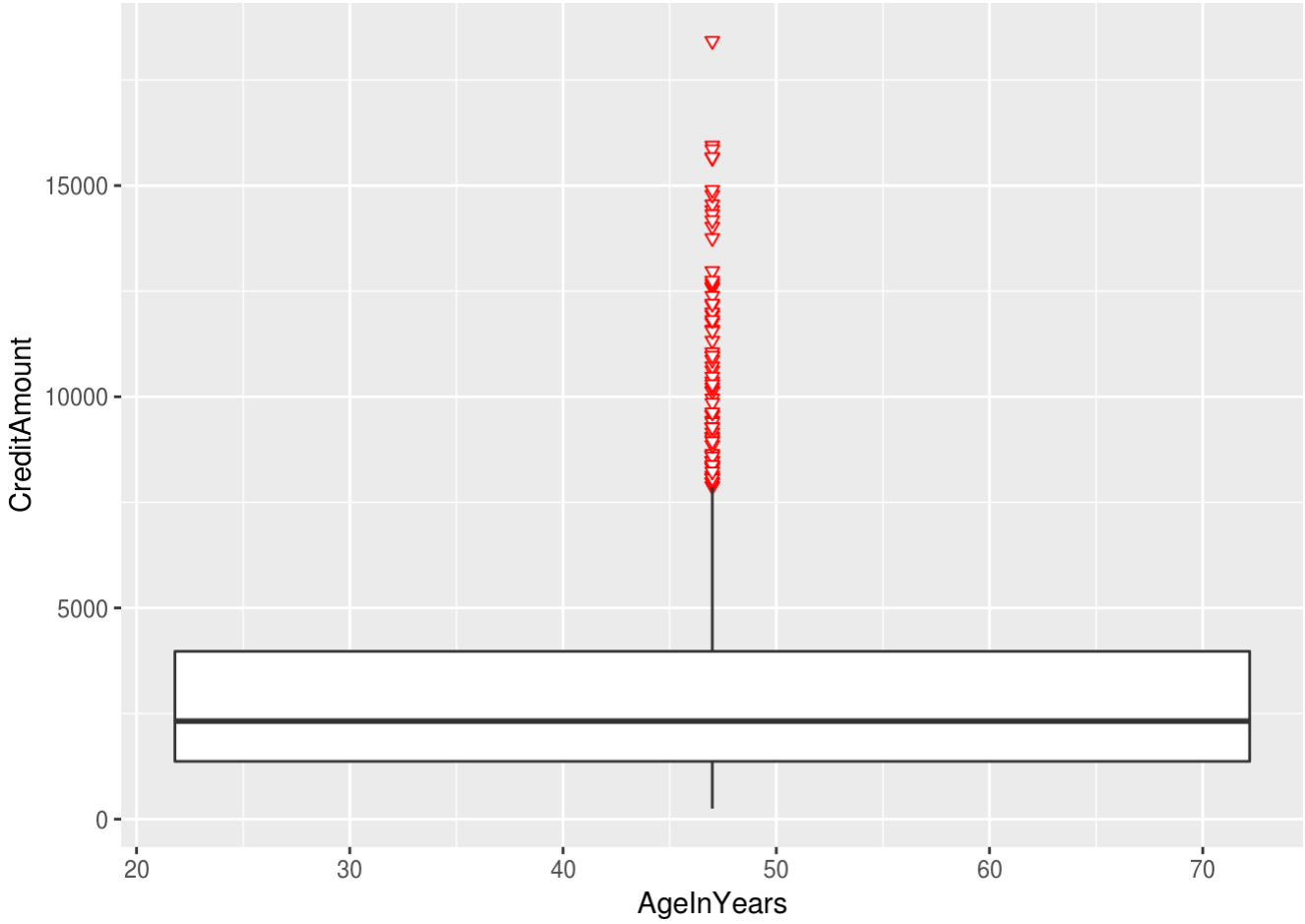
With the below plot, we can say, “Good” and “Bad” applicants are uniformly spread irrespective of Age



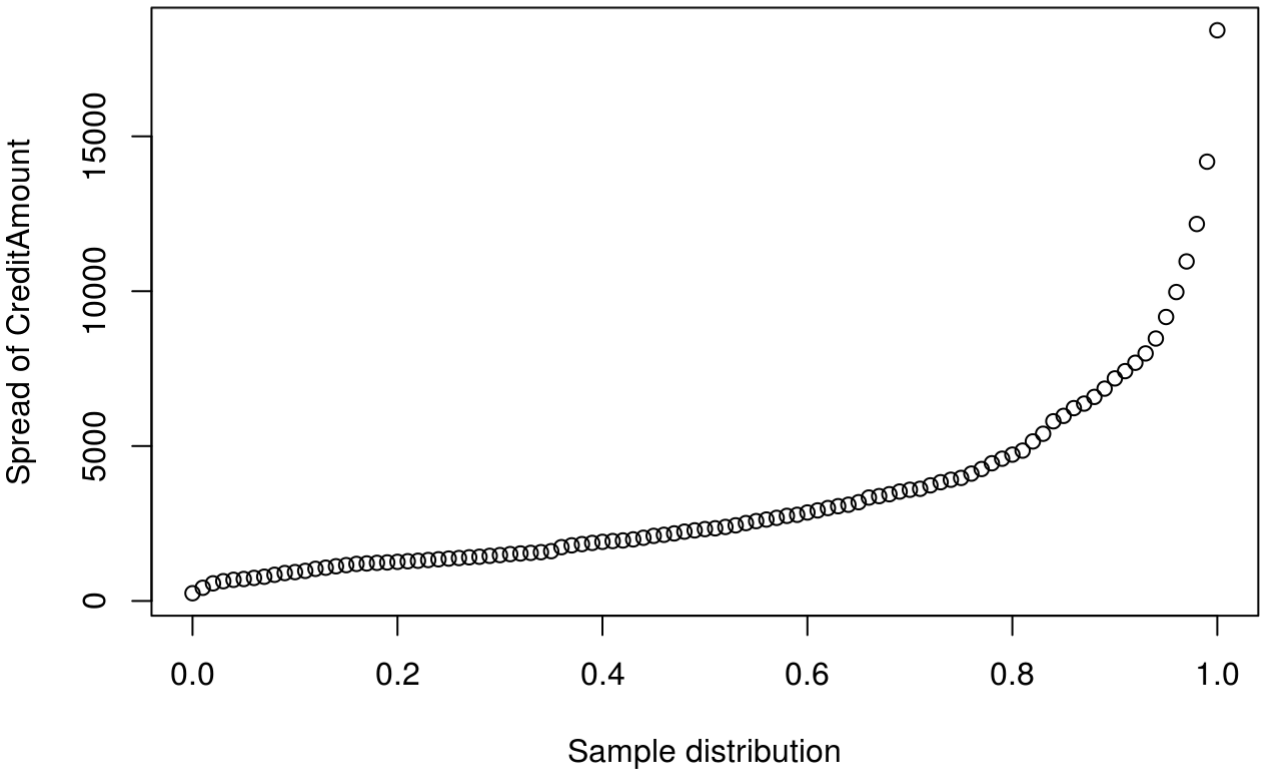
Box Plotting for credit amount and trimming at cutoff point

```
ggplot2::ggplot(data = GermanCreditData , aes(AgeInYears, CreditAmount)) +
  geom_boxplot(outlier.shape = 25, outlier.colour = "Red")
```

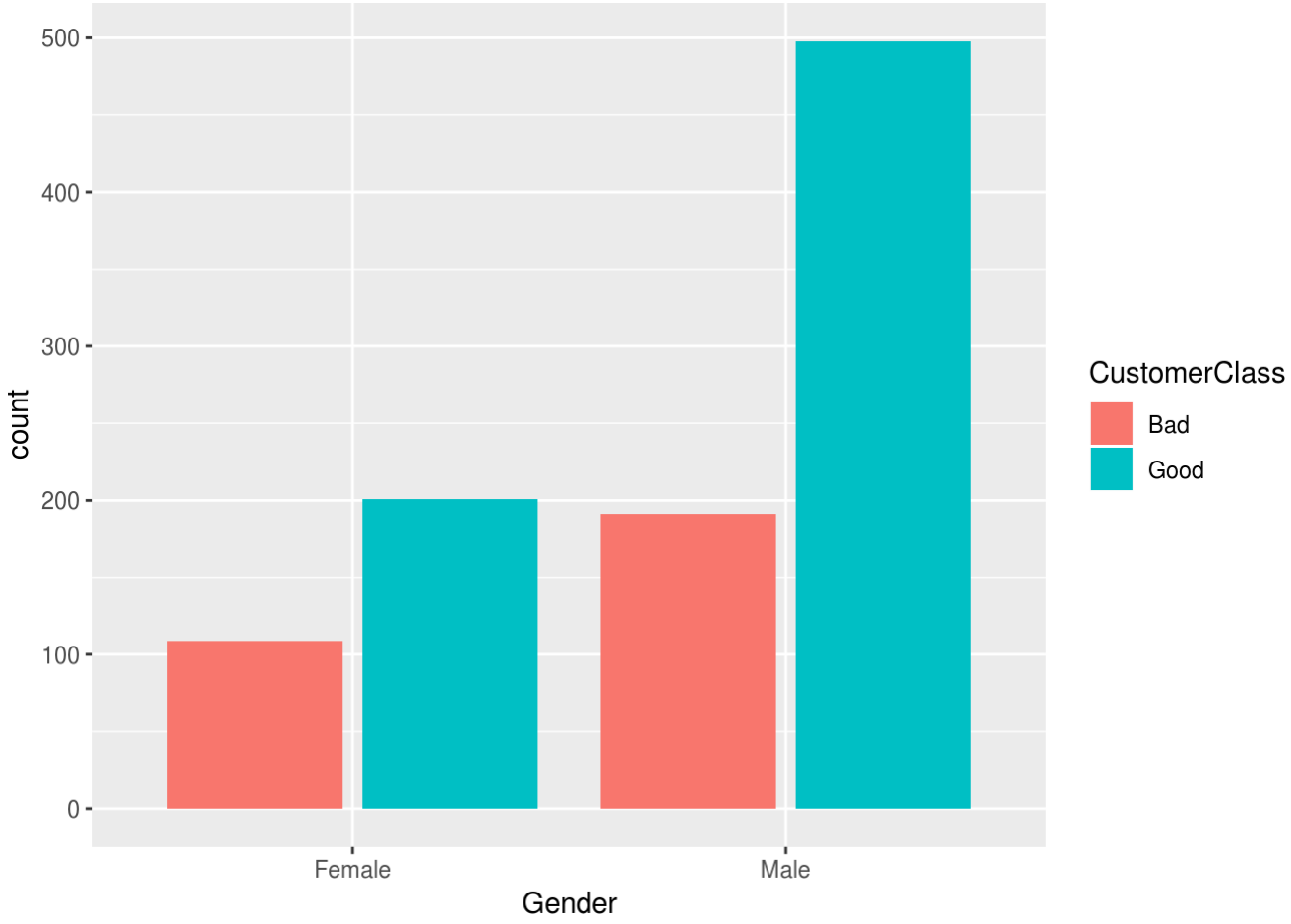
```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



Understand the quantile spread of CreditAmount



Understanding customer class based on gender



Populating the number of rows having age more than 60yrs

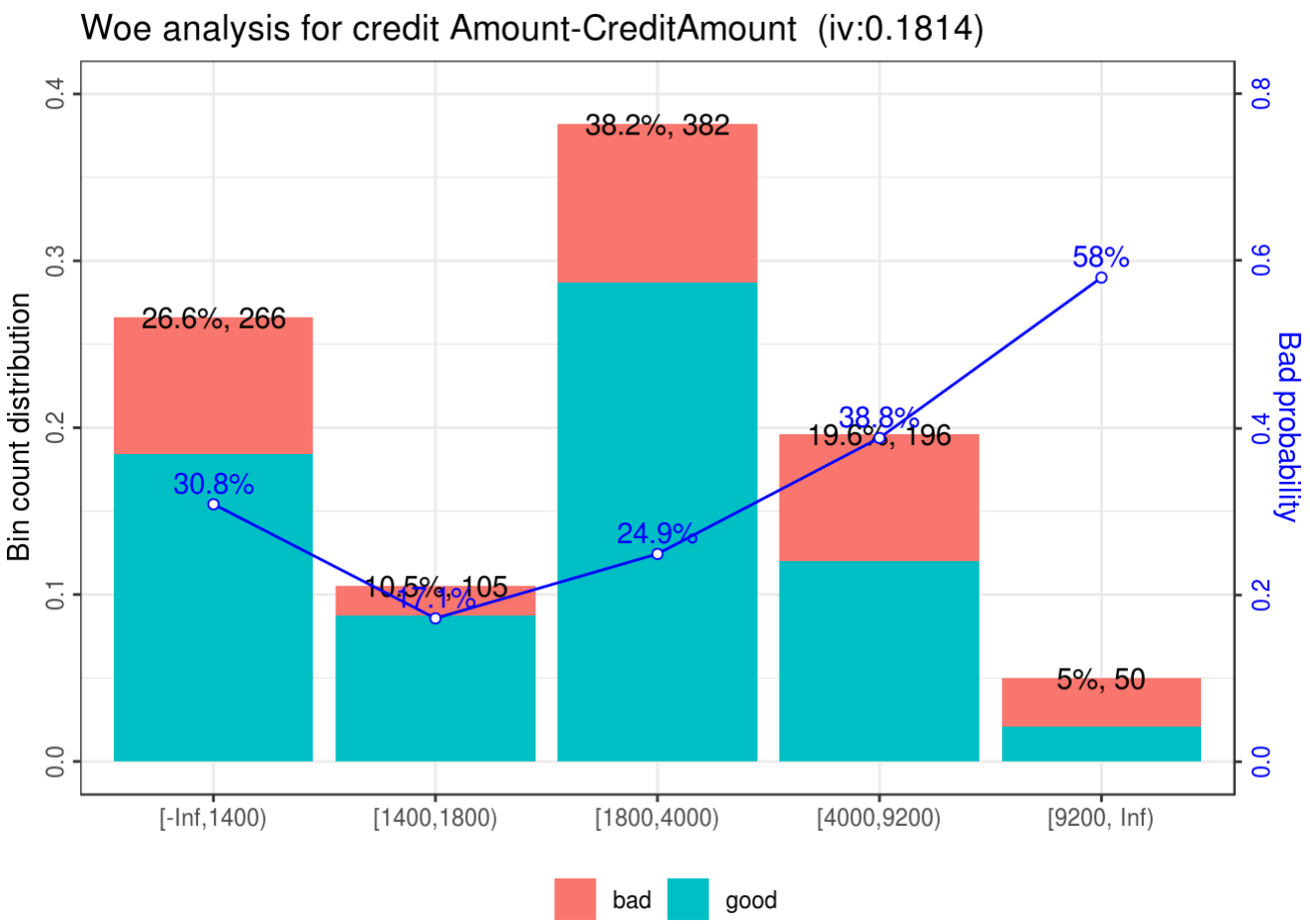
Starting with WoE Analysis

```
## Warning in check_y(dt, y, positive): The positive value in "CustomerClass"
## was replaced by 1 and negative value by 0.
```

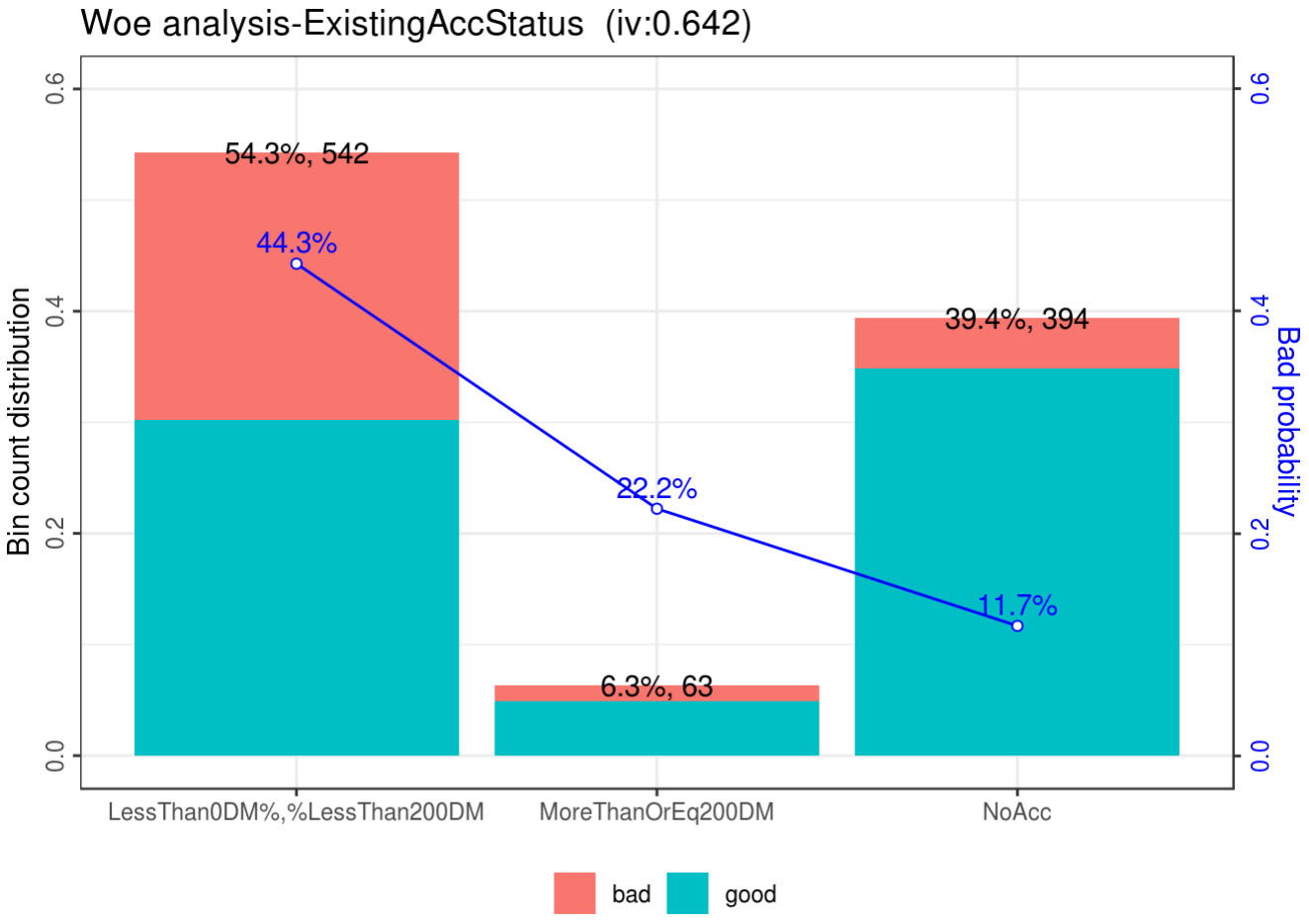
visualize woe bins

```
scorecard::woebin_plot(GermanCreditWoebins$CreditAmount, title = "Woe analysis for credit Amou
nt")
```

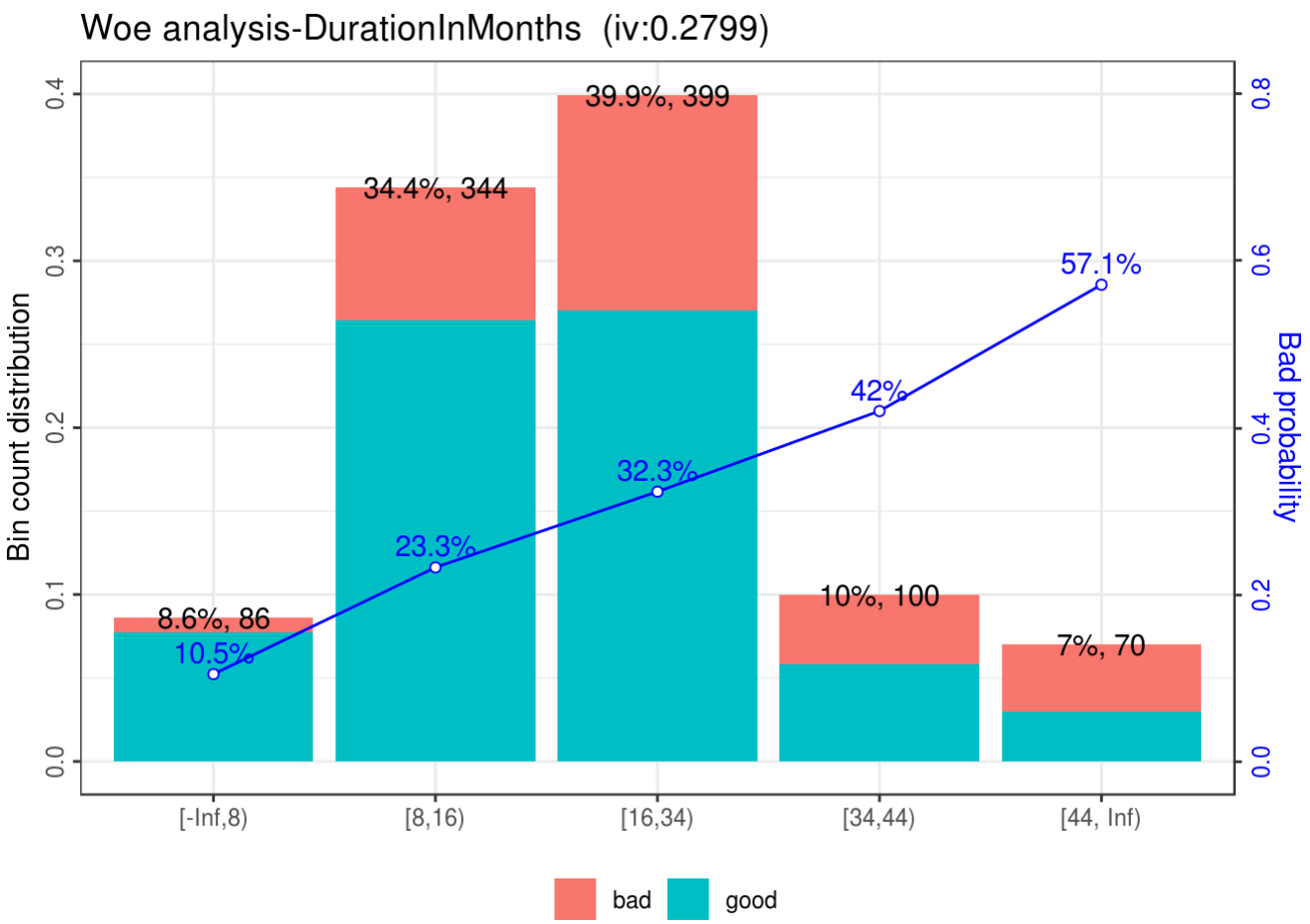
```
## $CreditAmount
```



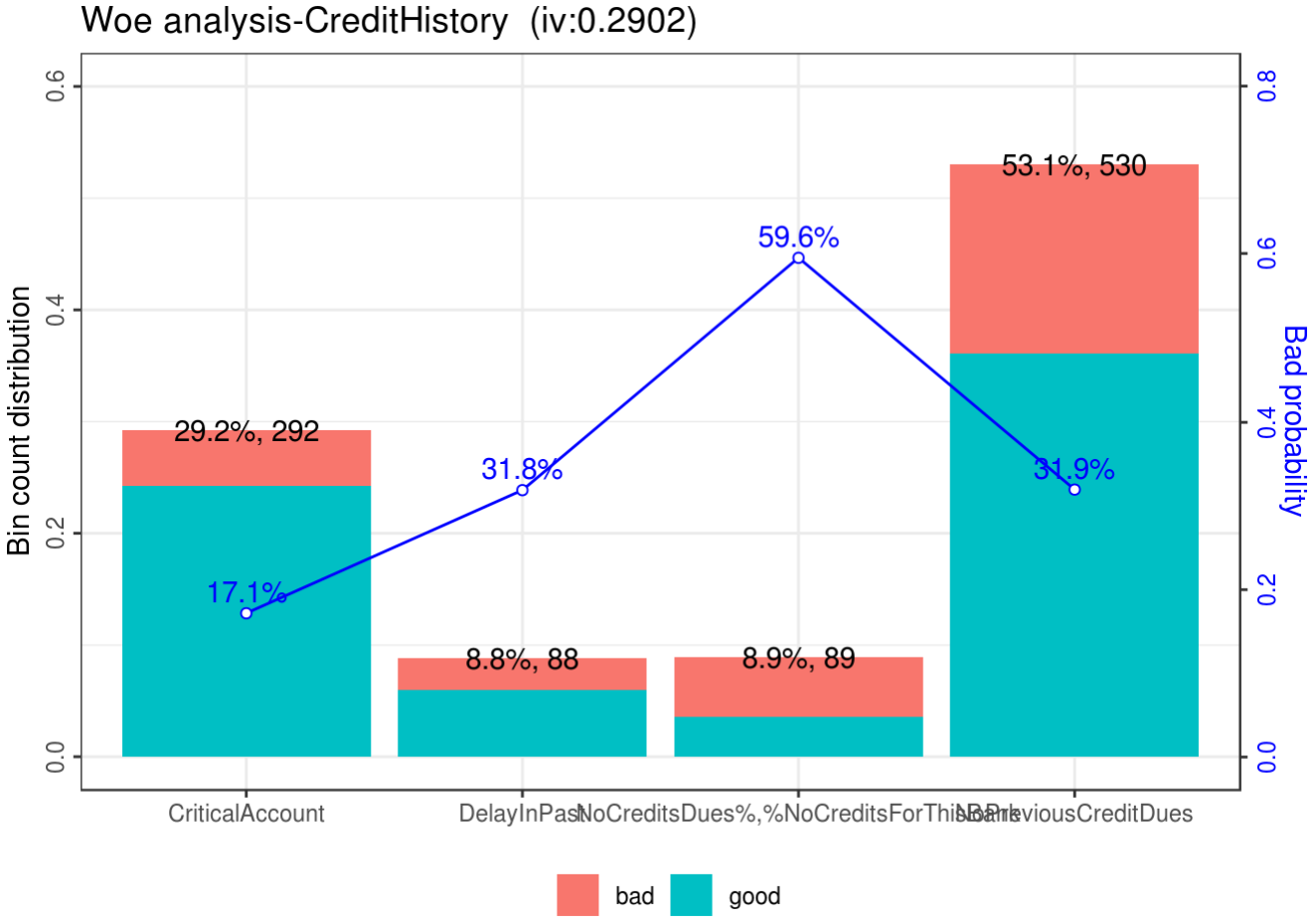

```
## $ExistingAccStatus
```



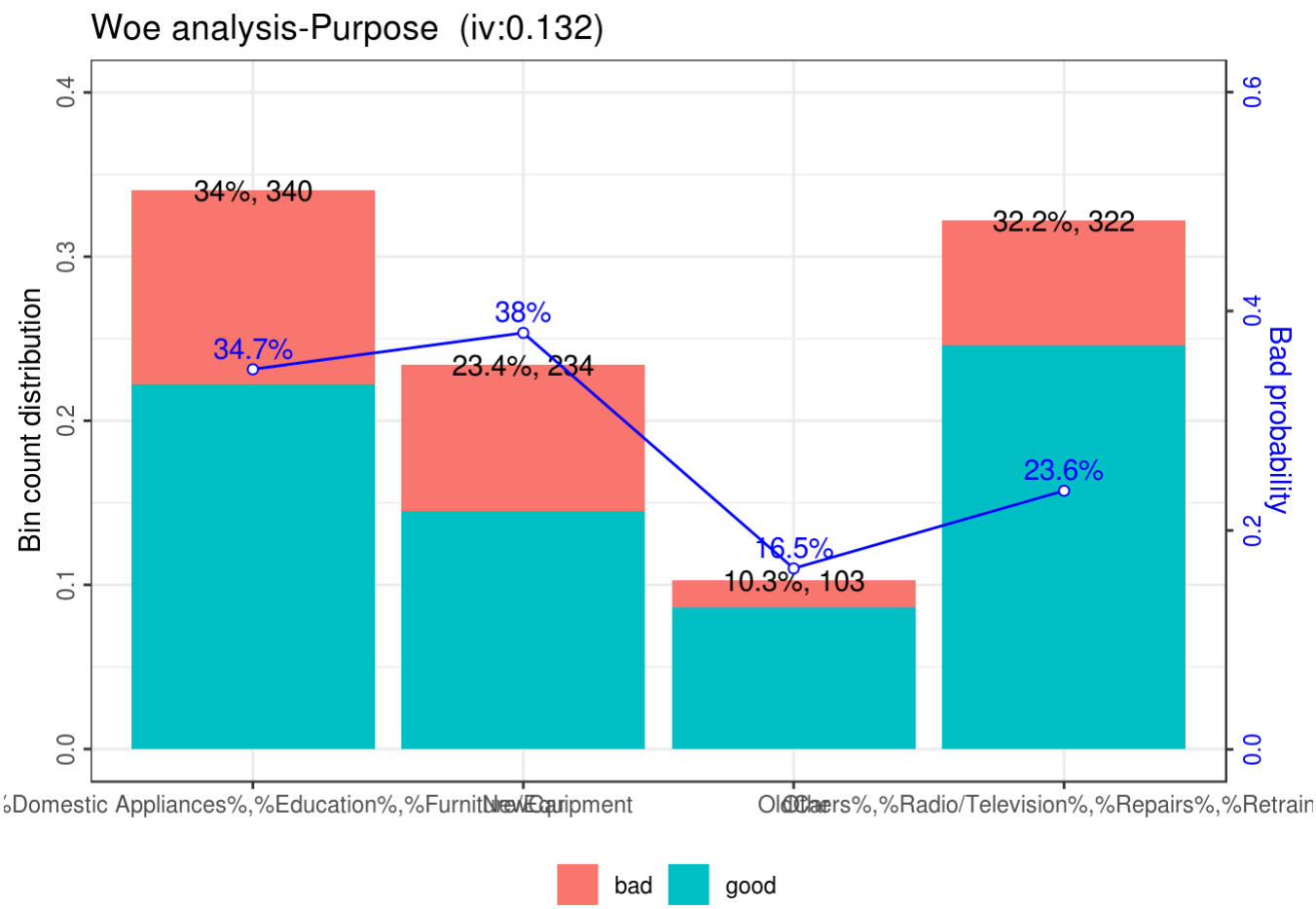

```
## $DurationInMonths
```



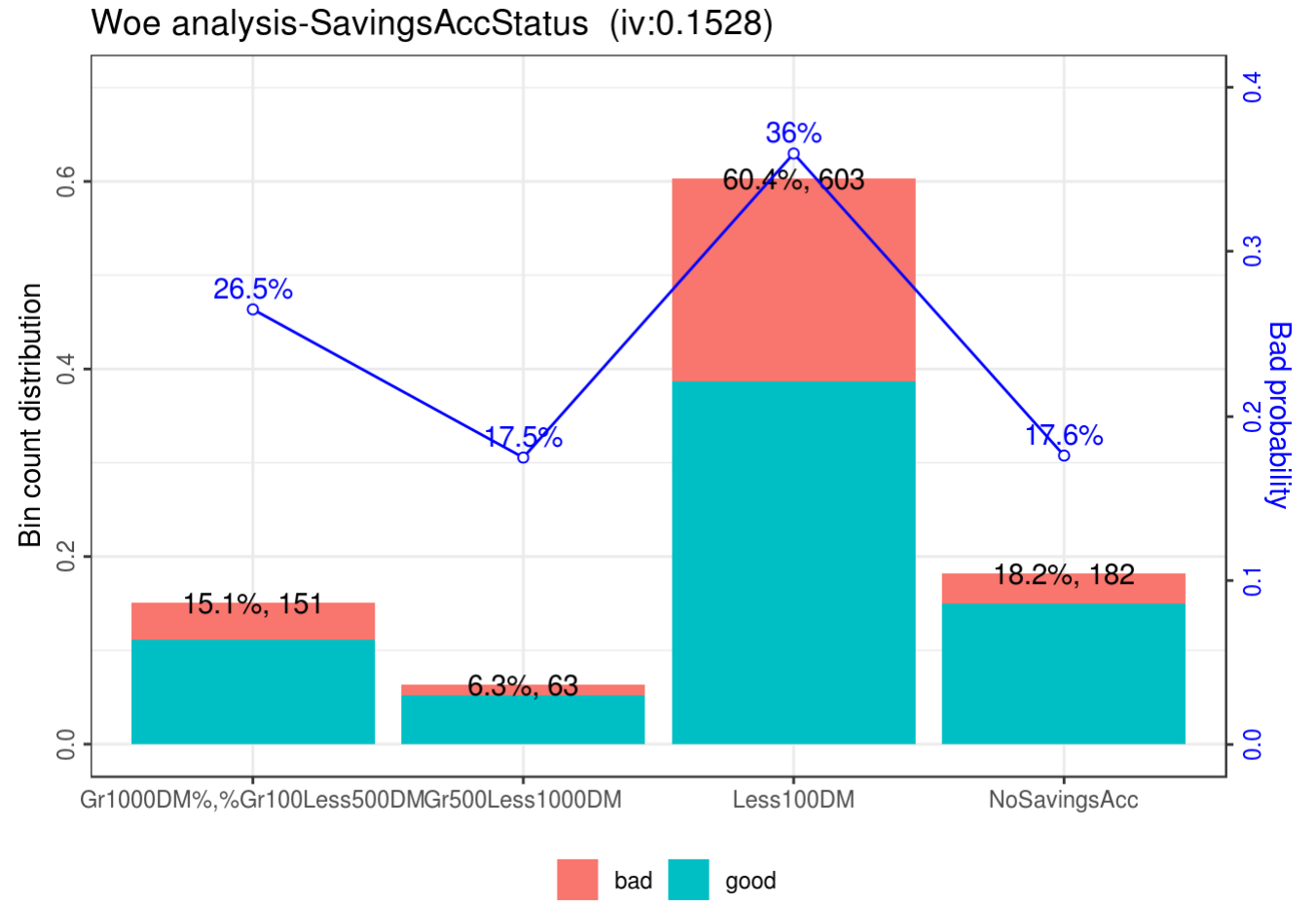

```
## $CreditHistory
```



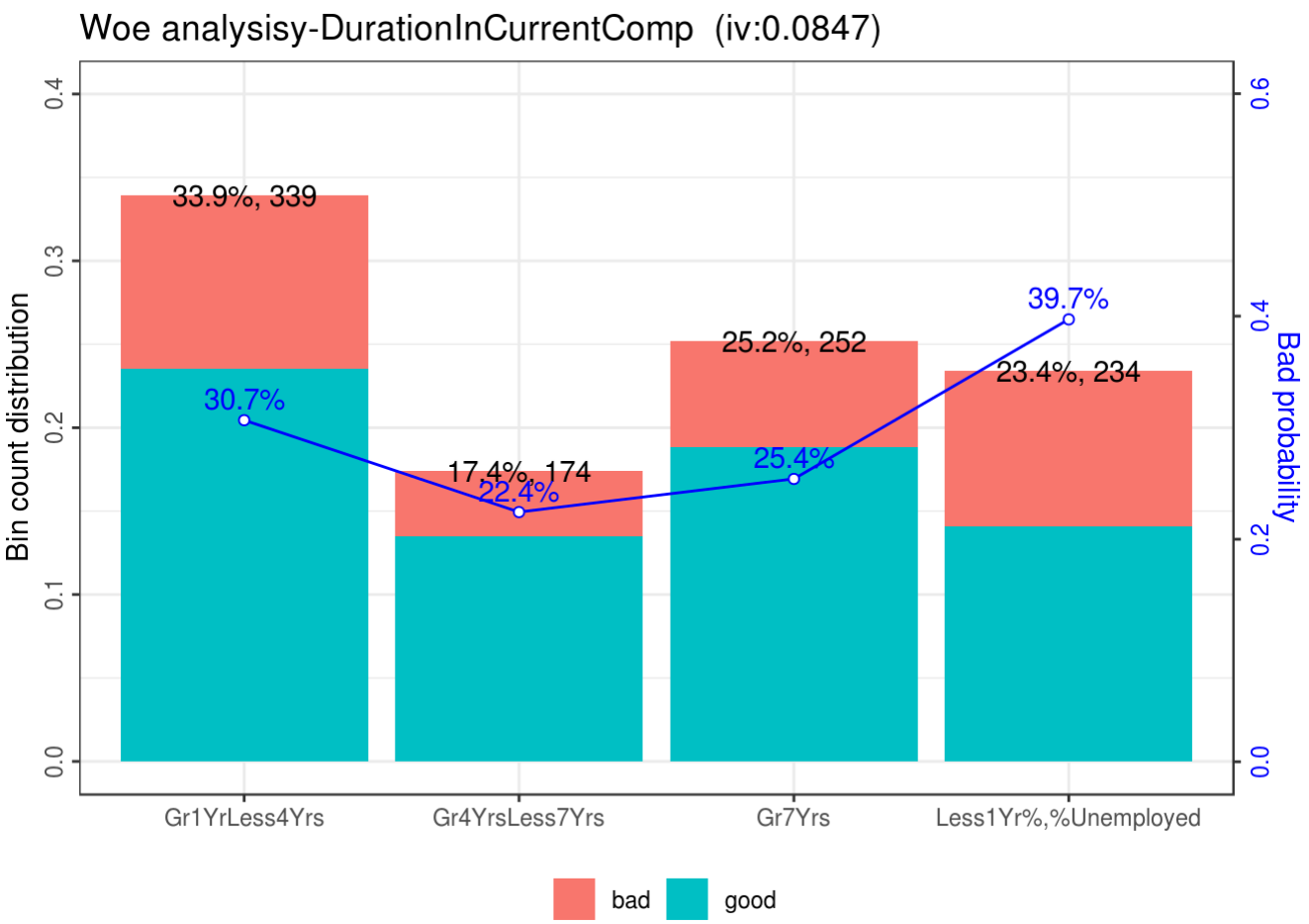

```
## $Purpose
```



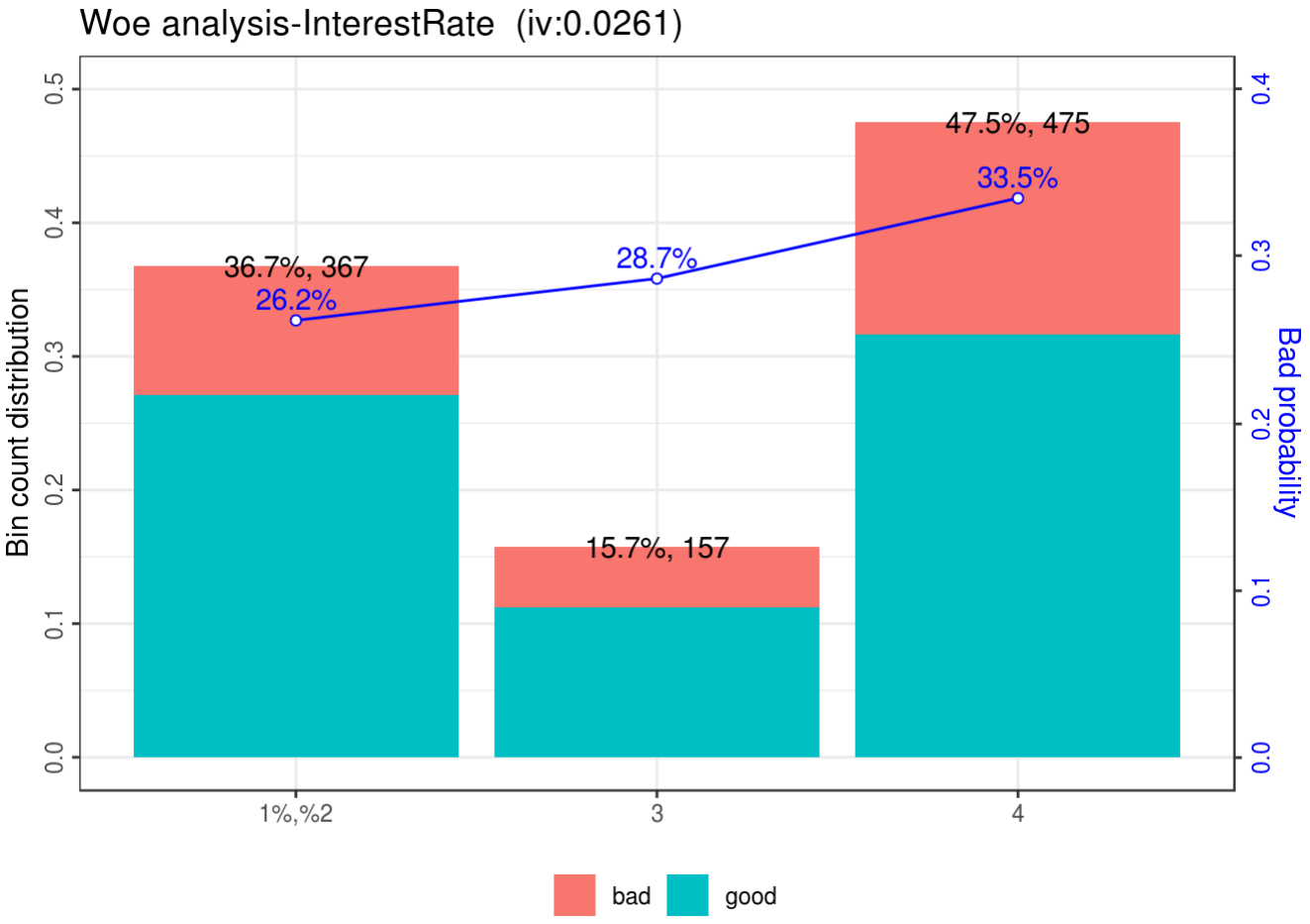

```
## $SavingsAccStatus
```



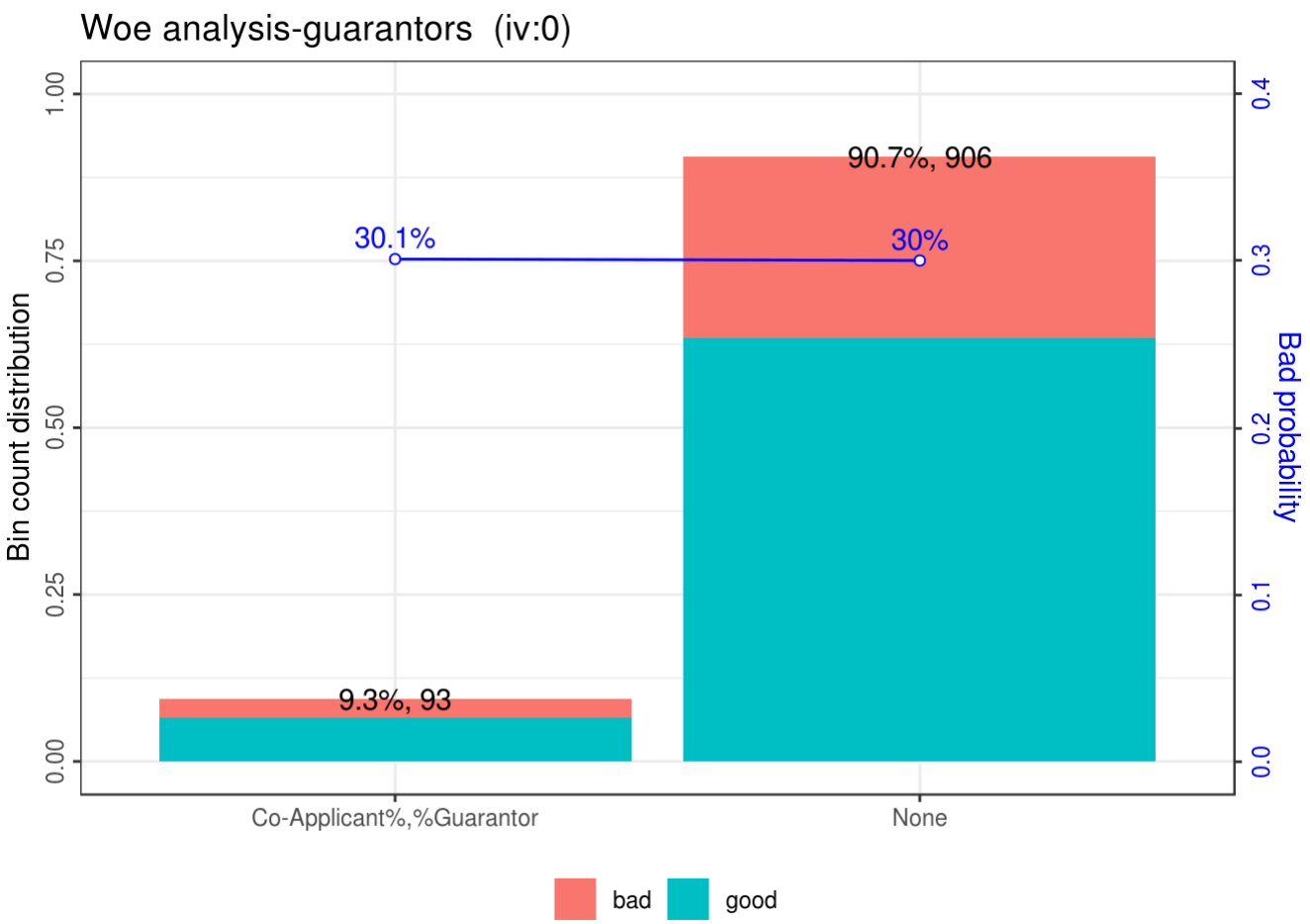

```
## $DurationInCurrentComp
```



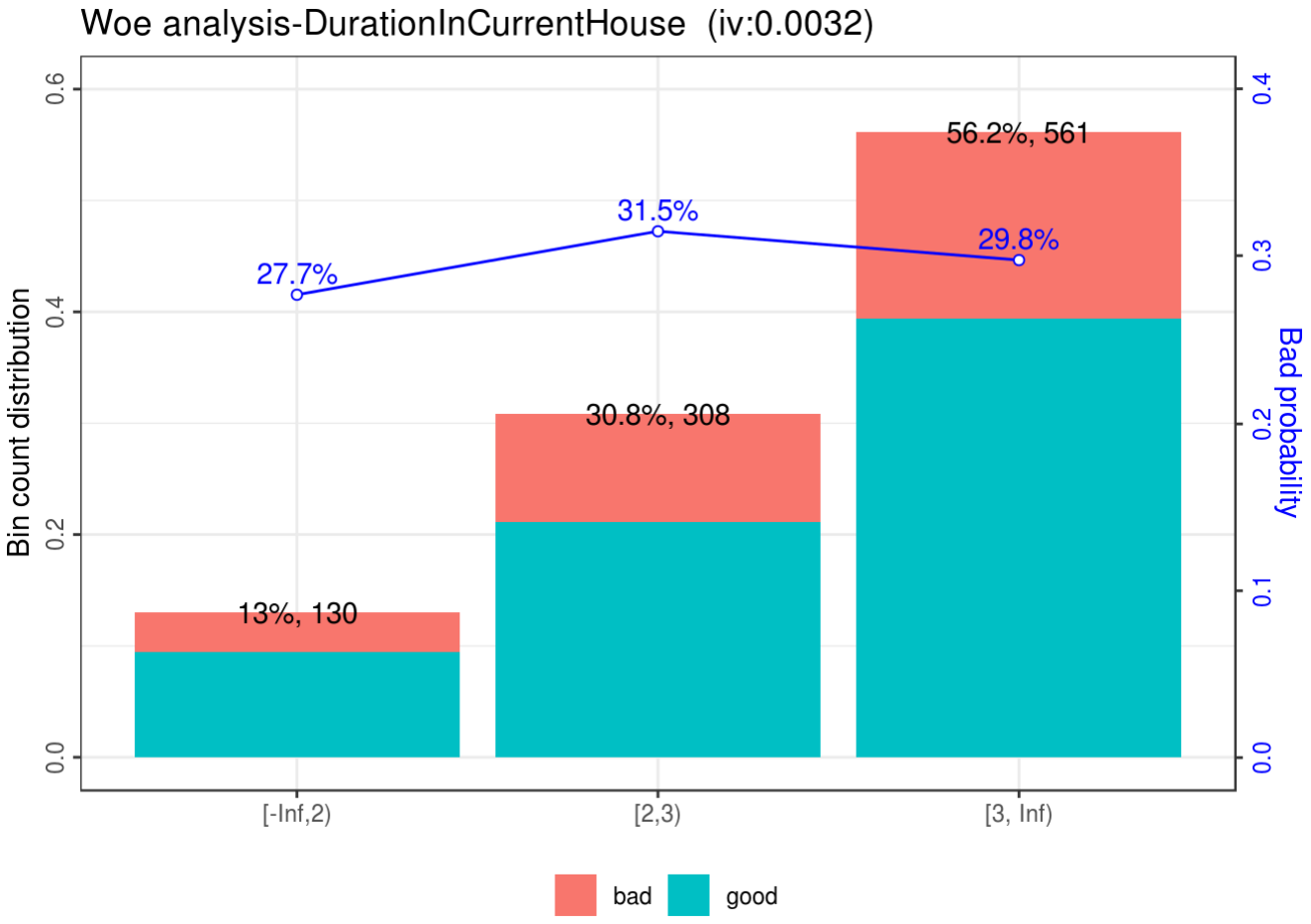

```
## $InterestRate
```



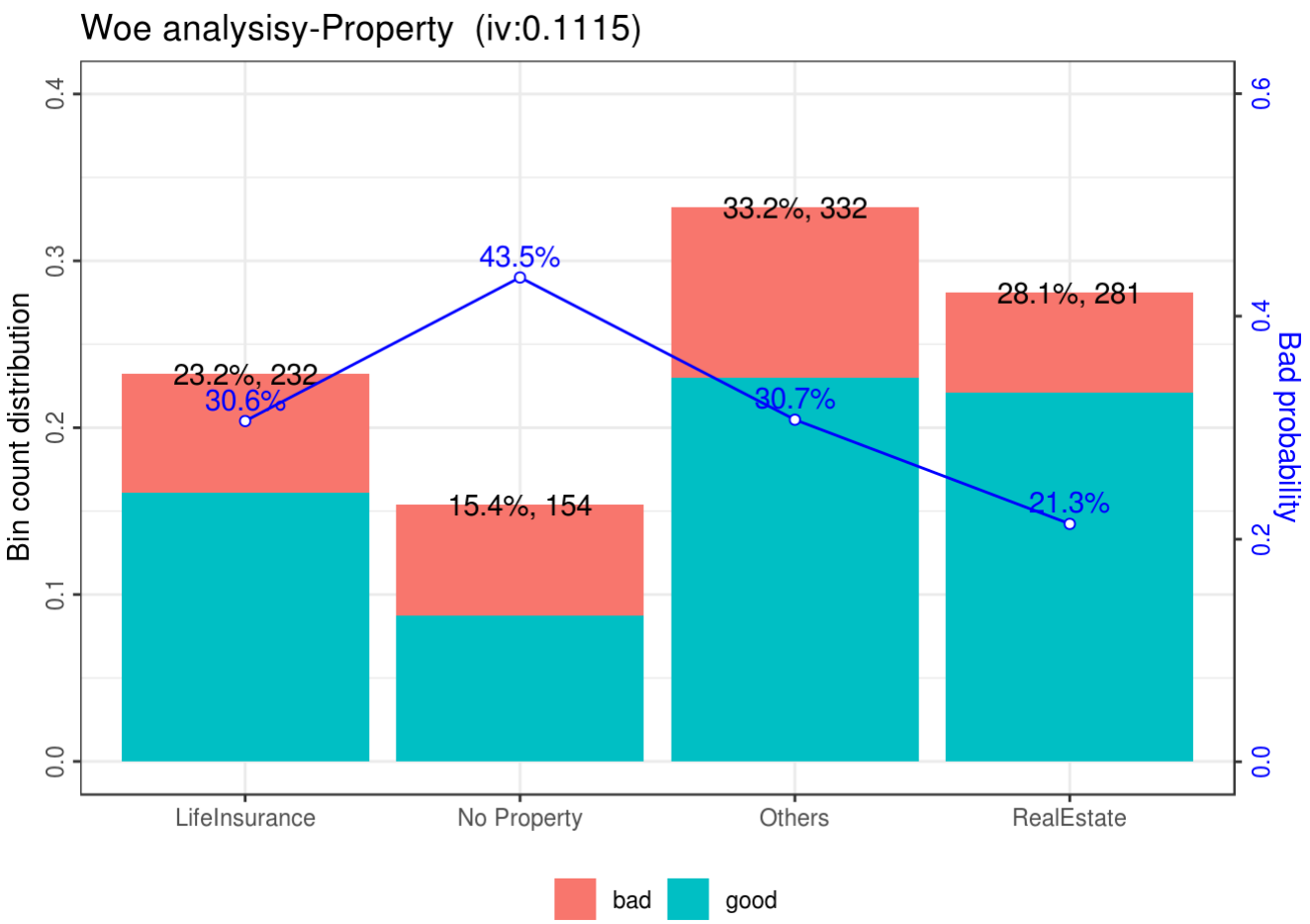

```
## $guarantors
```



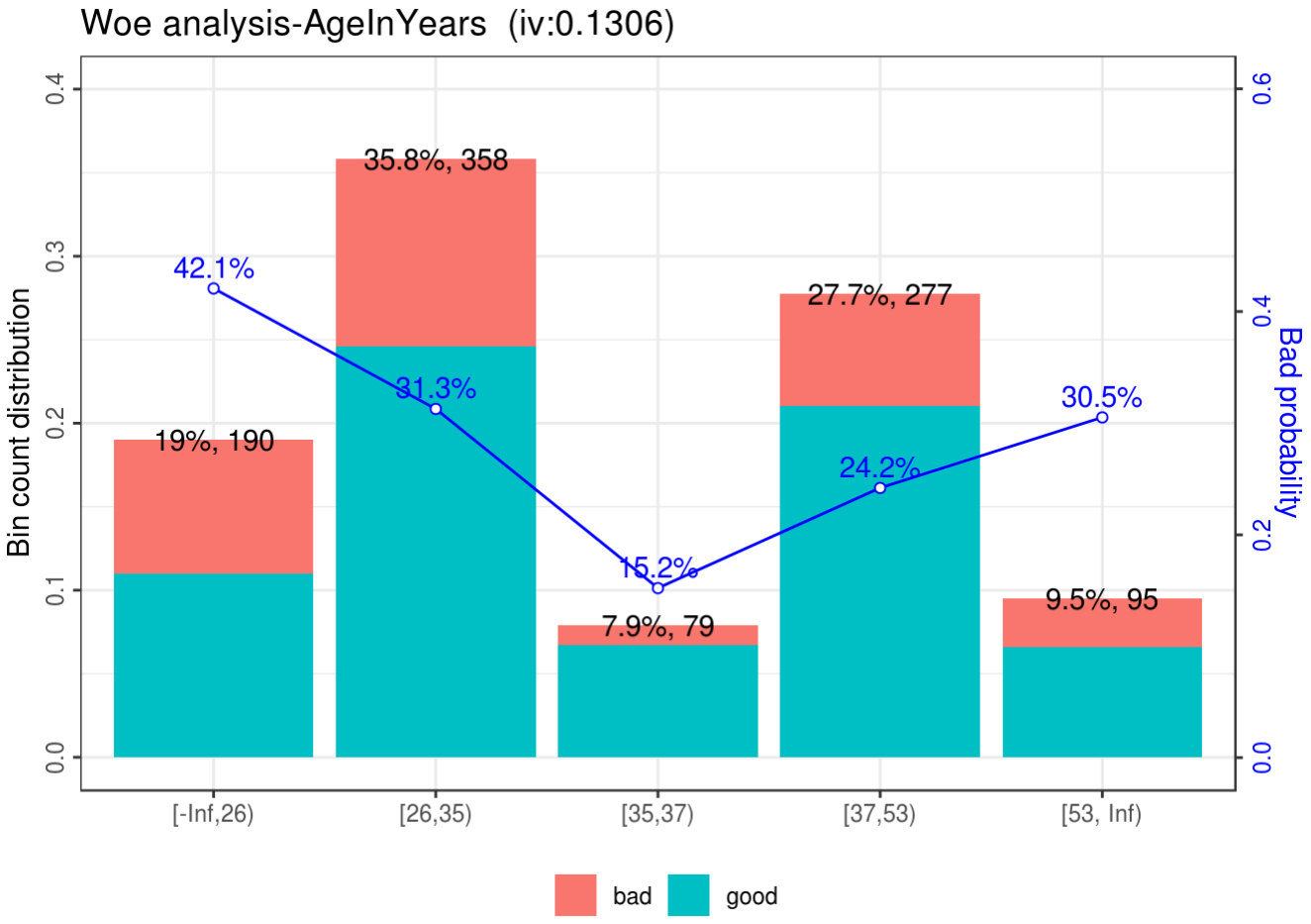

```
## $DurationInCurrentHouse
```



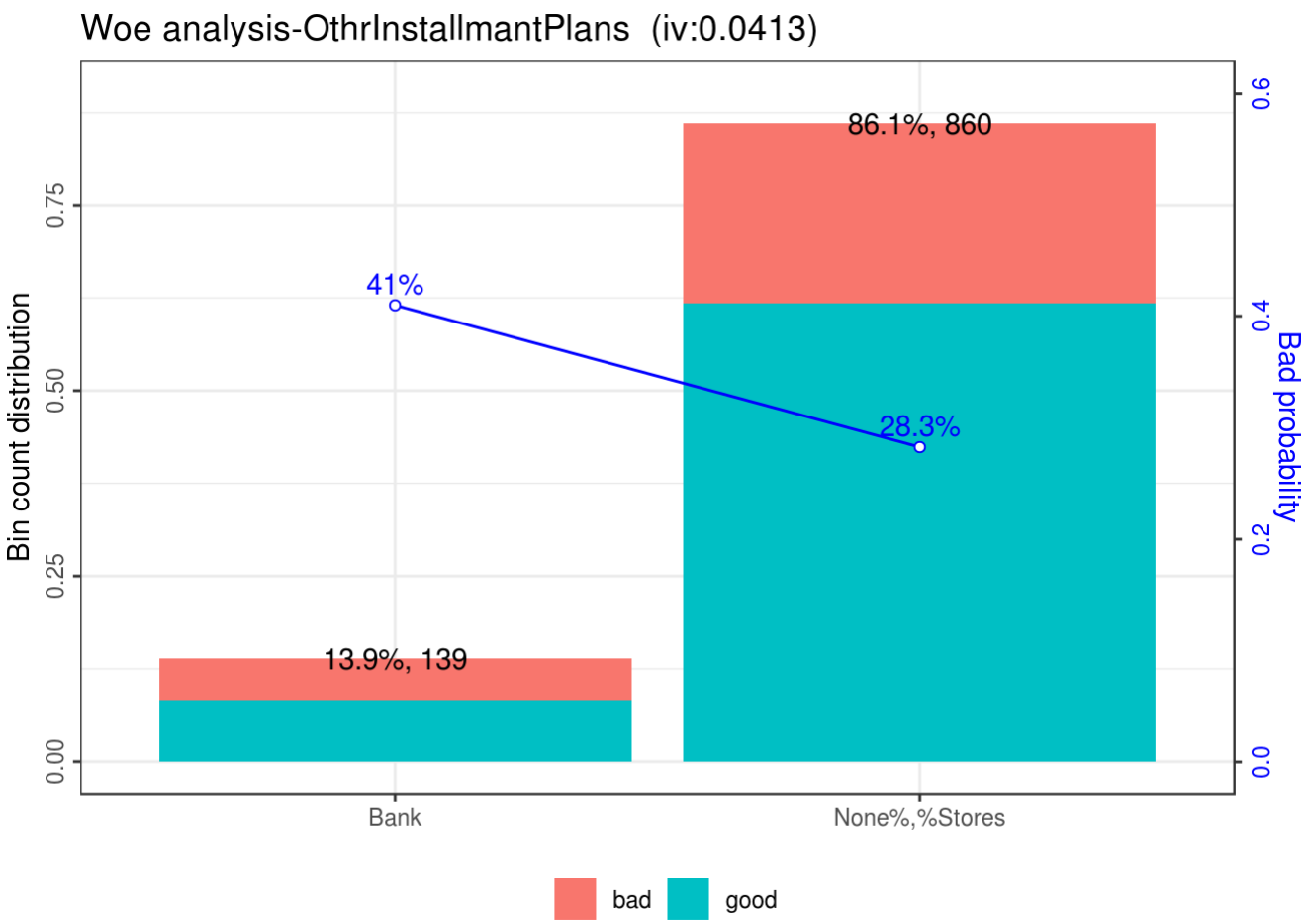

```
## $Property
```



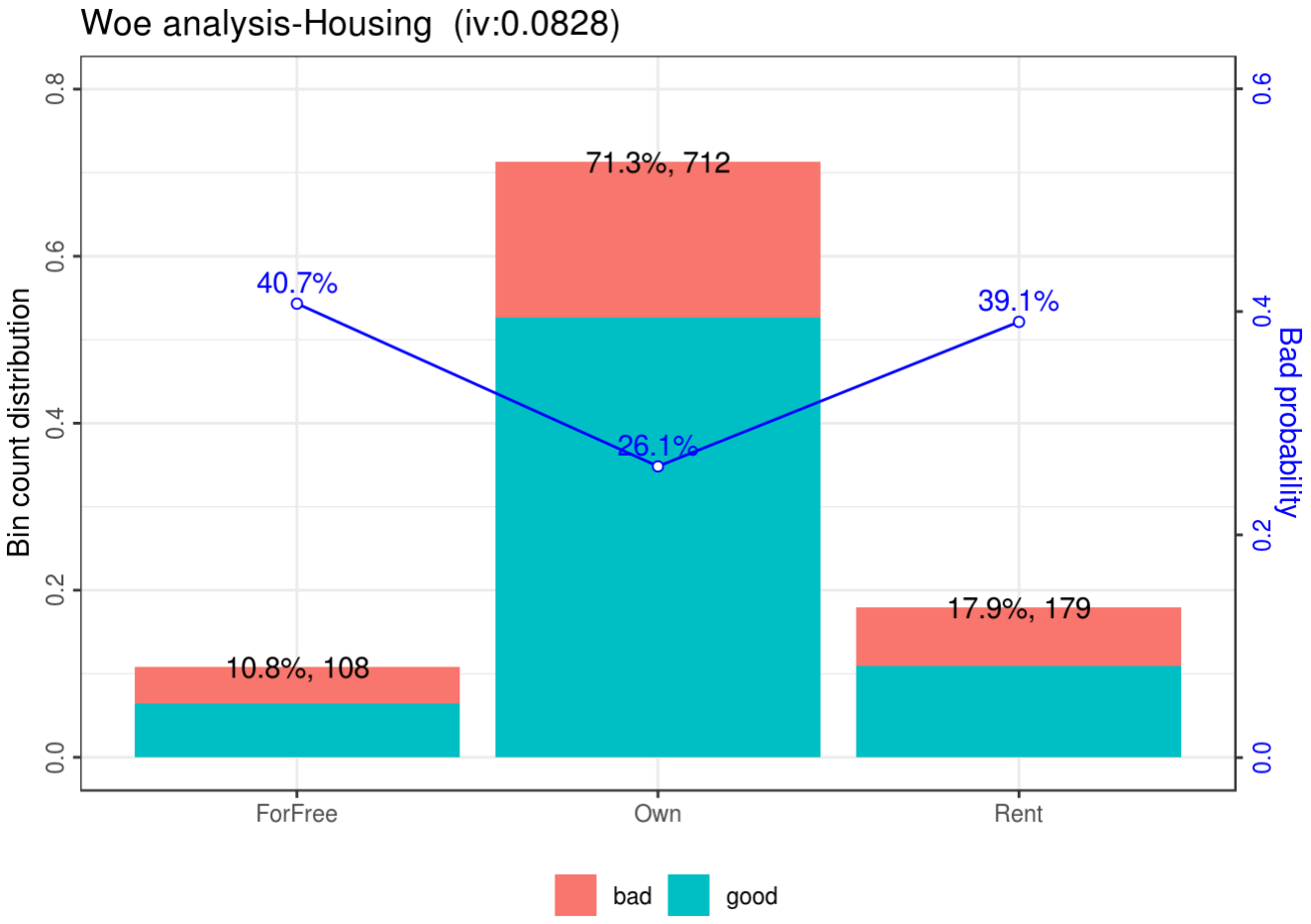

```
## $AgeInYears
```



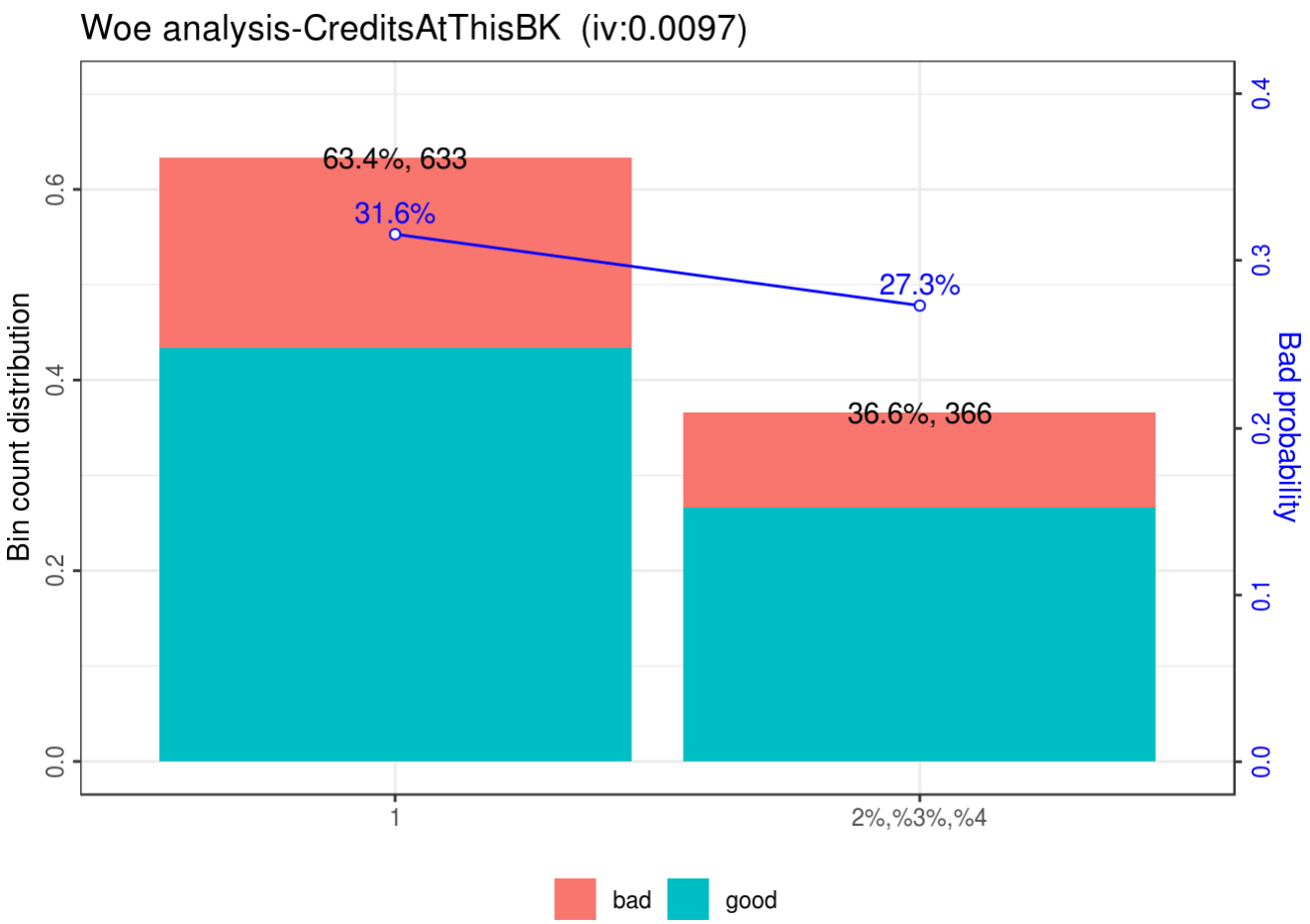

```
## $OthrInstallmantPlans
```



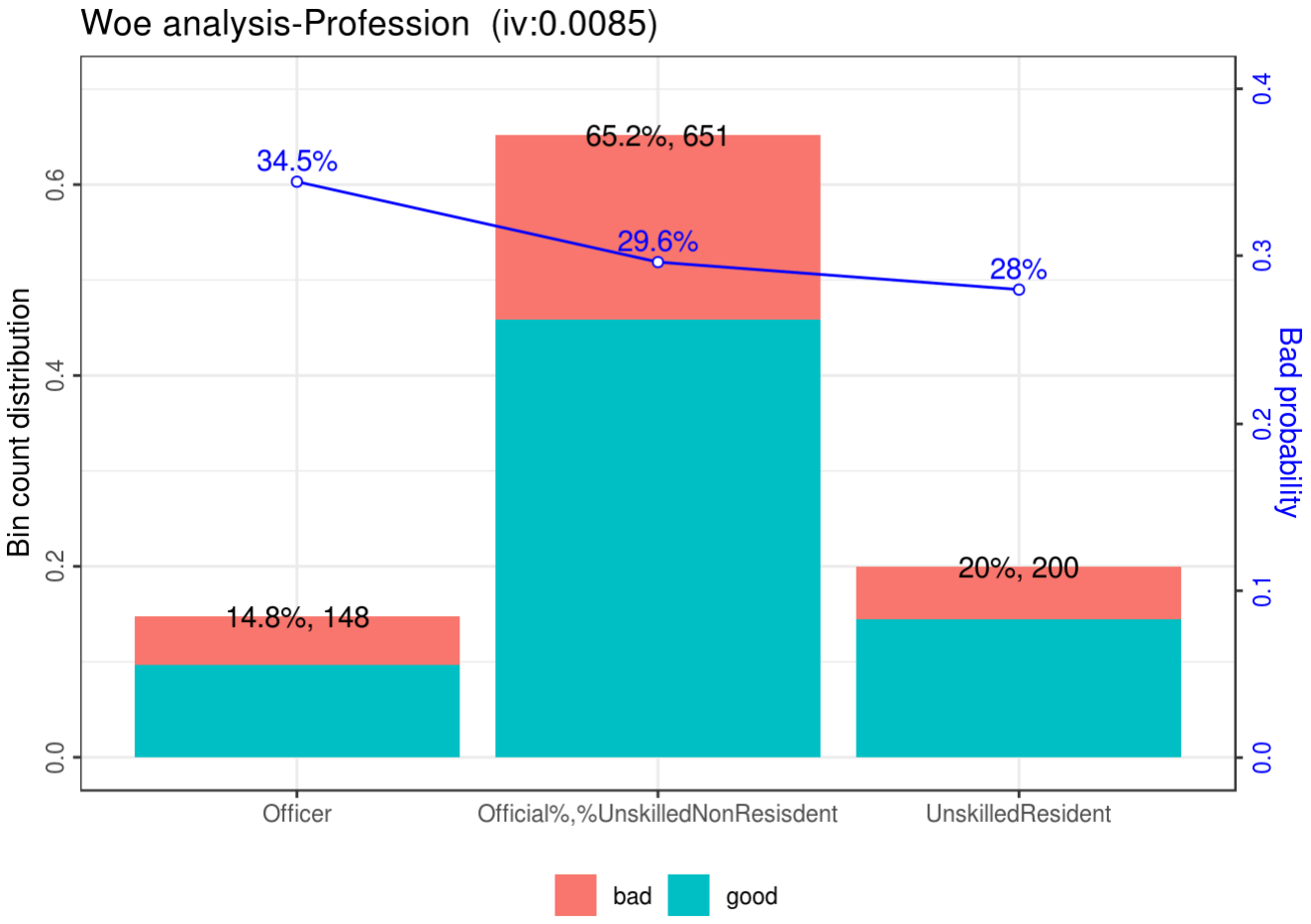

```
## $Housing
```



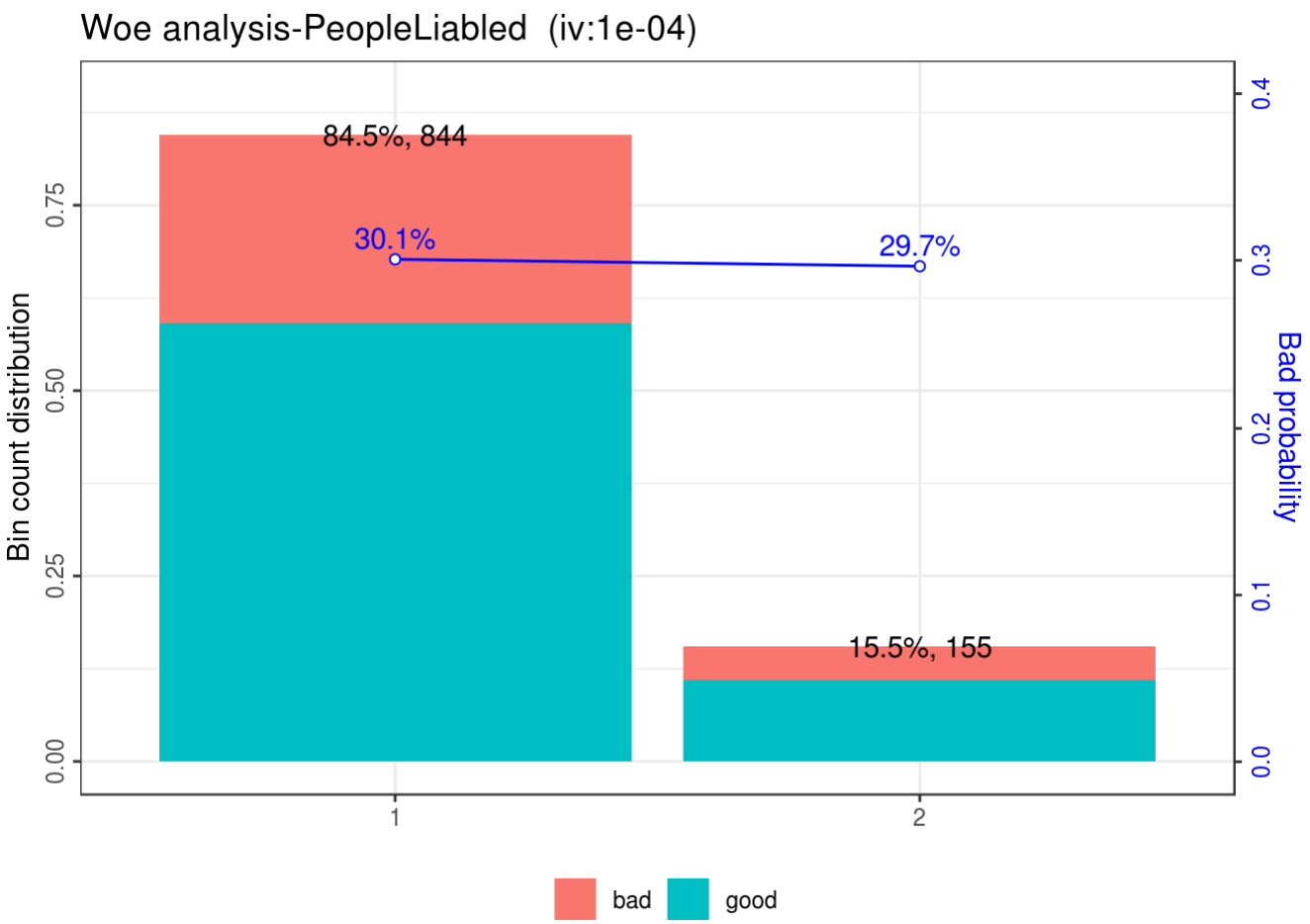

```
## $CreditsAtThisBK
```



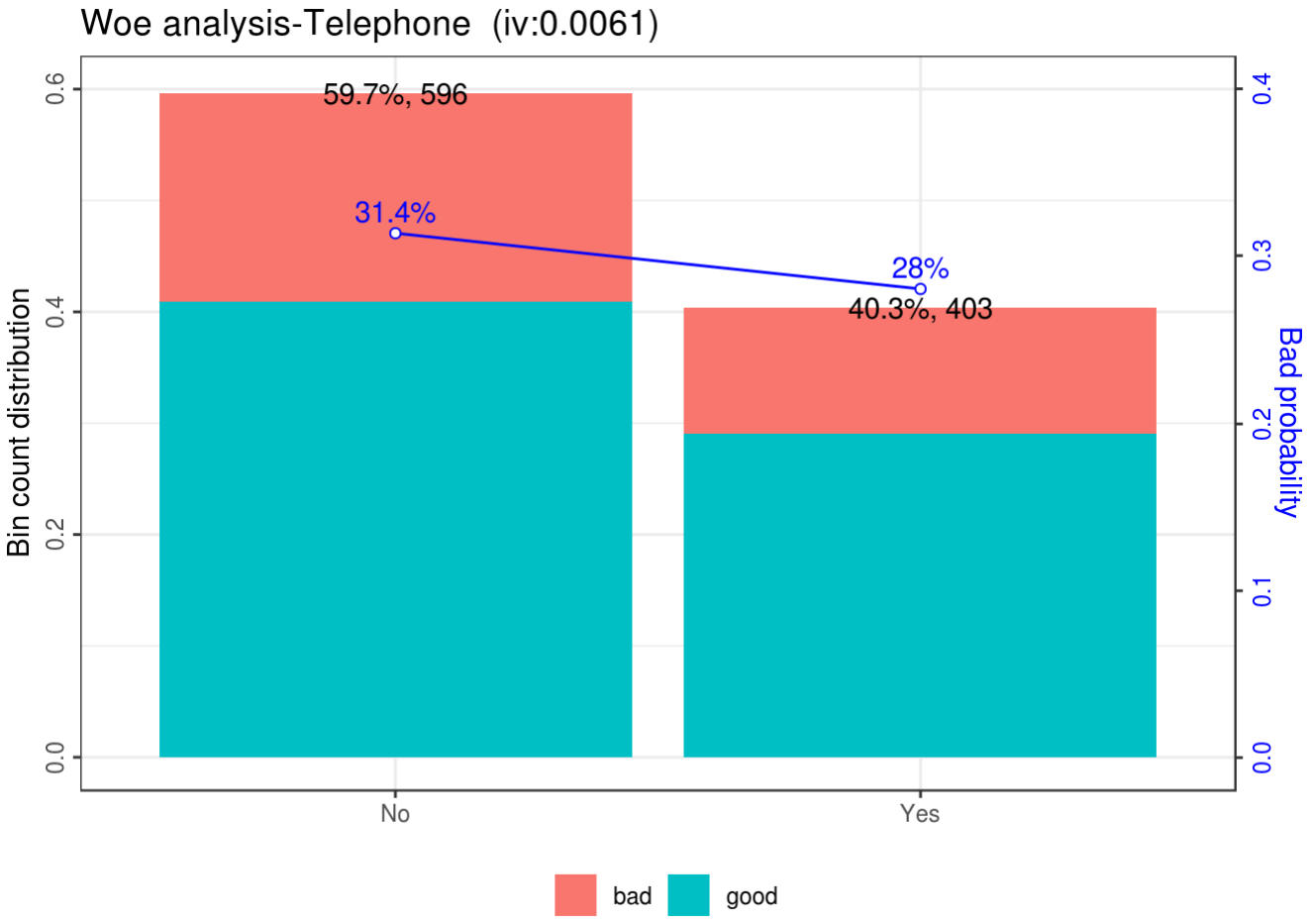

```
## $Profession
```




```
## $PeopleLiabled
```

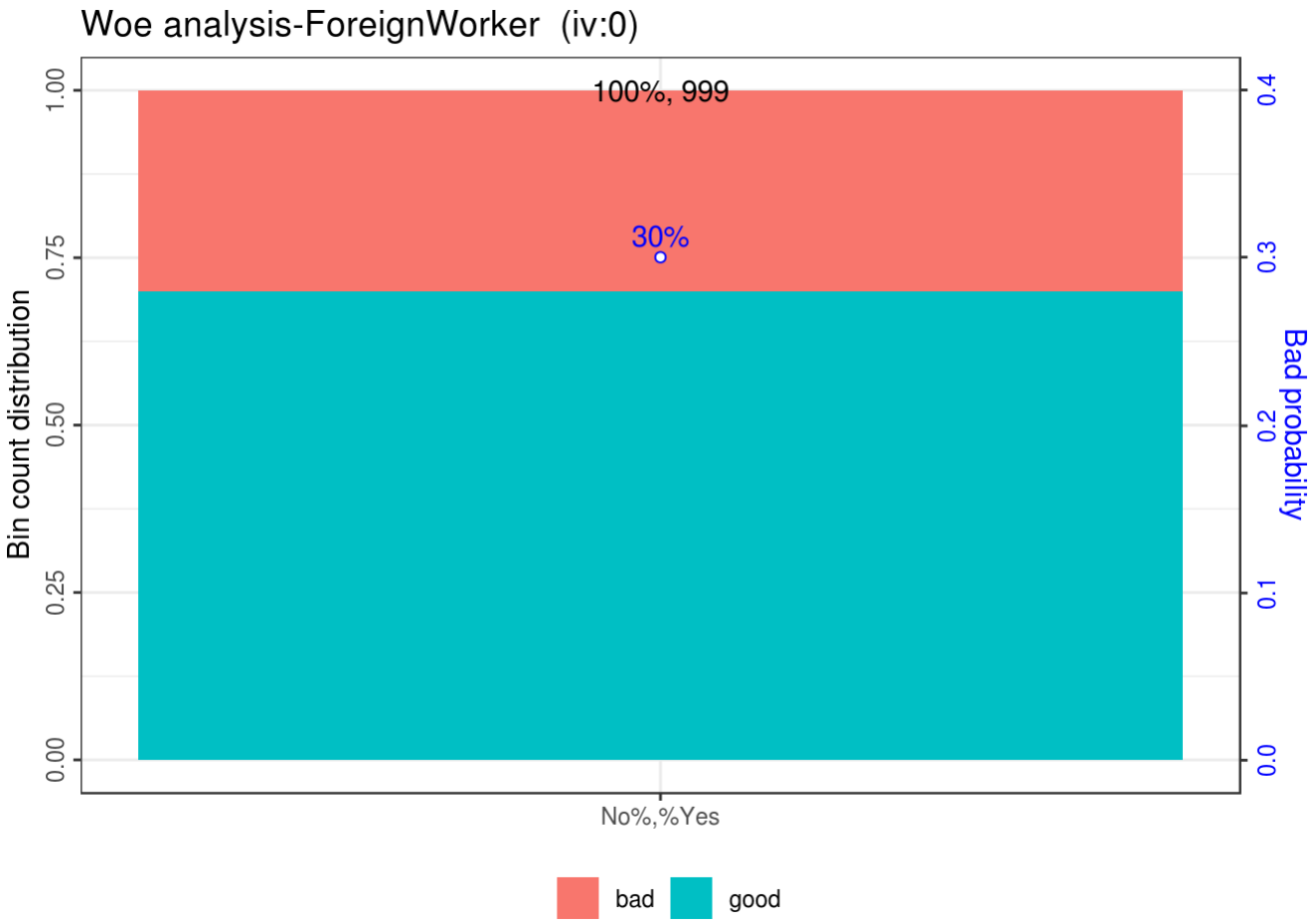



```
## $Telephone
```

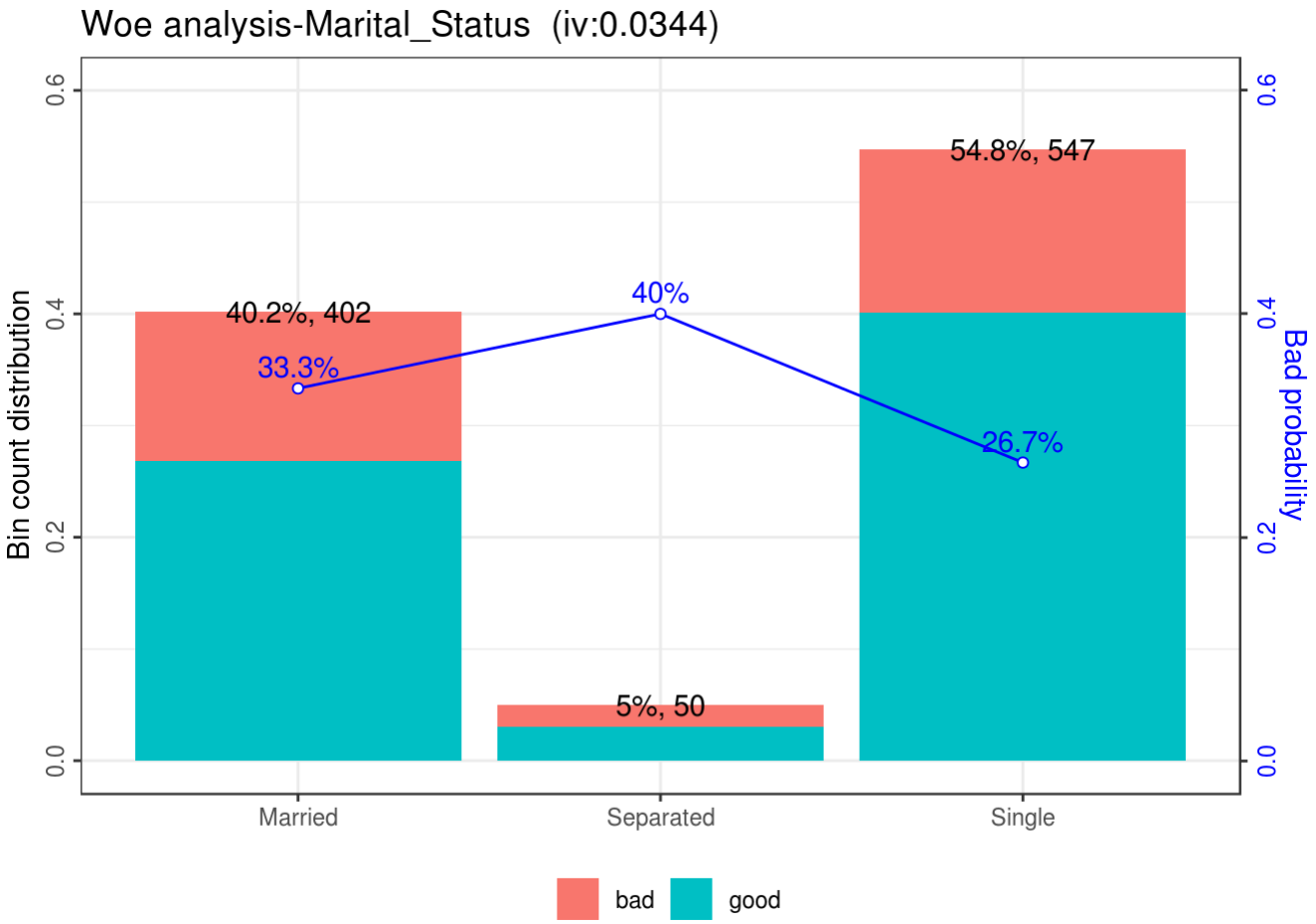



```
## $ForeignWorker
```

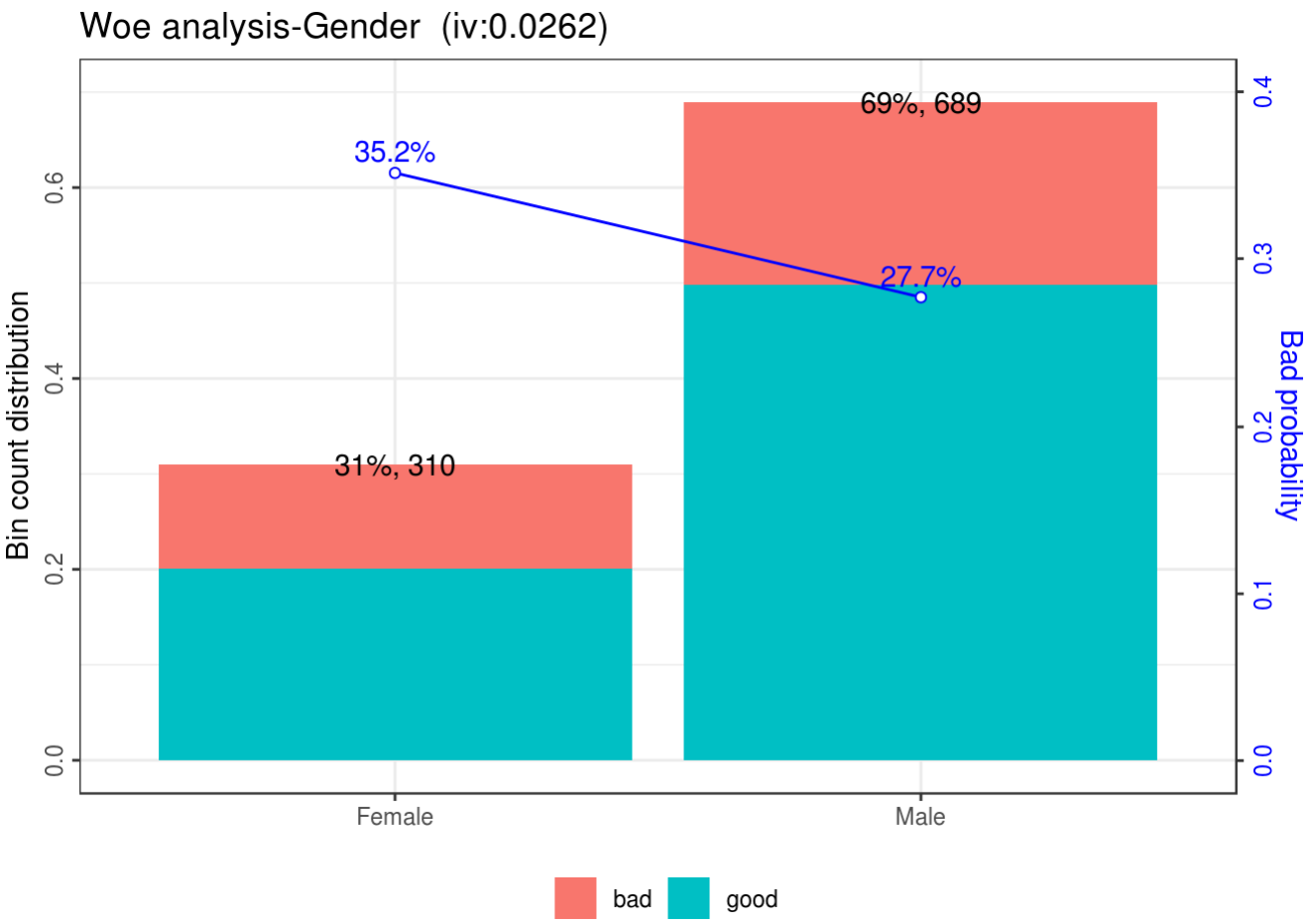
```
## geom_path: Each group consists of only one observation. Do you need to
## adjust the group aesthetic?
```




```
## $Marital_Status
```



```
## $Gender
```



Expanding Woe

```
##  CustomerClass ExistingAccStatus_woe DurationInMonths_woe
##  Bad :300      Min.   :-1.1777       Min.   :-1.30071
##  Good:699      1st Qu.:-1.1777       1st Qu.:-0.34805
##                Median : 0.6161       Median : 0.10726
##                Mean   :-0.1559       Mean   :-0.05719
##                3rd Qu.: 0.6161       3rd Qu.: 0.10726
##                Max.   : 0.6161       Max.   : 1.13355
##  CreditHistory_woe   Purpose_woe      CreditAmount_woe
##  Min.   :-0.73105   Min.   :-0.7753   Min.   :-0.72967
##  1st Qu.:-0.73105   1st Qu.:-0.3287   1st Qu.:-0.25974
##  Median : 0.08689   Median : 0.2139   Median : 0.03765
##  Mean   :-0.05039   Mean   :-0.0293   Mean   :-0.03115
##  3rd Qu.: 0.08689   3rd Qu.: 0.2139   3rd Qu.: 0.03765
##  Max.   : 1.23264   Max.   : 0.3578   Max.   : 1.16864
##  SavingsAccStatus_woe DurationInCurrentComp_woe InterestRate_woe
##  Min.   :-0.70748     Min.   :-0.39584          Min.   :-0.191902
##  1st Qu.:-0.17478     1st Qu.:-0.23169          1st Qu.:-0.191902
##  Median : 0.26993     Median : 0.03067          Median :-0.065968
##  Mean   :-0.03546     Mean   :-0.01633          Mean   :-0.005251
##  3rd Qu.: 0.26993     3rd Qu.: 0.03067          3rd Qu.: 0.159030
##  Max.   : 0.26993     Max.   : 0.42971          Max.   : 0.159030
##  guarantors_woe       DurationInCurrentHouse_woe  Property_woe
##  Min.   :-3.786e-04   Min.   :-0.1139076         Min.   :-0.45795
##  1st Qu.:-3.786e-04   1st Qu.:-0.0124888         1st Qu.:-0.45795
##  Median :-3.786e-04   Median :-0.0124888         Median : 0.02714
##  Mean   :-2.780e-07   Mean   :-0.0006488         Mean   :-0.02149
##  3rd Qu.:-3.786e-04   3rd Qu.: 0.0687211         3rd Qu.: 0.03276
##  Max.   : 3.686e-03   Max.   : 0.0687211         Max.   : 0.58465
##  AgeInYears_woe     OthrInstallmantPlans_woe  Housing_woe
##  Min.   :-0.87392   Min.   :-0.085939        Min.   :-0.19369
##  1st Qu.:-0.29655   1st Qu.:-0.085939        1st Qu.:-0.19369
##  Median : 0.05904   Median :-0.085939        Median :-0.19369
##  Mean   :-0.02763   Mean   :-0.006889        Mean   :-0.01489
##  3rd Qu.: 0.05904   3rd Qu.:-0.085939        3rd Qu.: 0.40302
##  Max.   : 0.52741   Max.   : 0.482200        Max.   : 0.47117
##  CreditsAtThisBK_woe Profession_woe      PeopleLiabled_woe
##  Min.   :-0.132458   Min.   :-0.098593   Min.   :-1.684e-02
##  1st Qu.:-0.132458   1st Qu.:-0.018311   1st Qu.: 3.080e-03
##  Median : 0.073448   Median :-0.018311   Median : 3.080e-03
##  Mean   :-0.001989   Mean   :-0.001599   Mean   :-1.042e-05
##  3rd Qu.: 0.073448   3rd Qu.:-0.018311   3rd Qu.: 3.080e-03
##  Max.   : 0.073448   Max.   : 0.202983   Max.   : 3.080e-03
##  Telephone_woe       ForeignWorker_woe Marital_Status_woe
##  Min.   :-0.096625   Min.   :0         Min.   :-0.164487
##  1st Qu.:-0.096625   1st Qu.:0         1st Qu.:-0.164487
##  Median : 0.063262   Median :0         Median :-0.164487
##  Mean   :-0.001237   Mean   :0         Mean   :-0.006567
##  3rd Qu.: 0.063262   3rd Qu.:0         3rd Qu.: 0.152721
##  Max.   : 0.063262   Max.   :0         Max.   : 0.440403
##    Gender_woe        integerClass_woe
##  Min.   :-0.112458   Min.   :0
##  1st Qu.:-0.112458   1st Qu.:0
##  Median :-0.112458   Median :0
##  Mean   :-0.004976   Mean   :0
##  3rd Qu.: 0.233911   3rd Qu.:0
##  Max.   : 0.233911   Max.   :0
```

Computing IV values to compute attribute significance

```
library(scorecard)
IVTable <- scorecard::iv(GermanCreditWoeDataSet, y = 'CustomerClass', positive = "Bad|2")
```

```
## Warning in rmcol_datetime_unique1(dt): There are 2 columns have only one unique values, whi
ch are removed from input dataset.
##  (ColumnNames: ForeignWorker_woe, integerClass_woe)
```

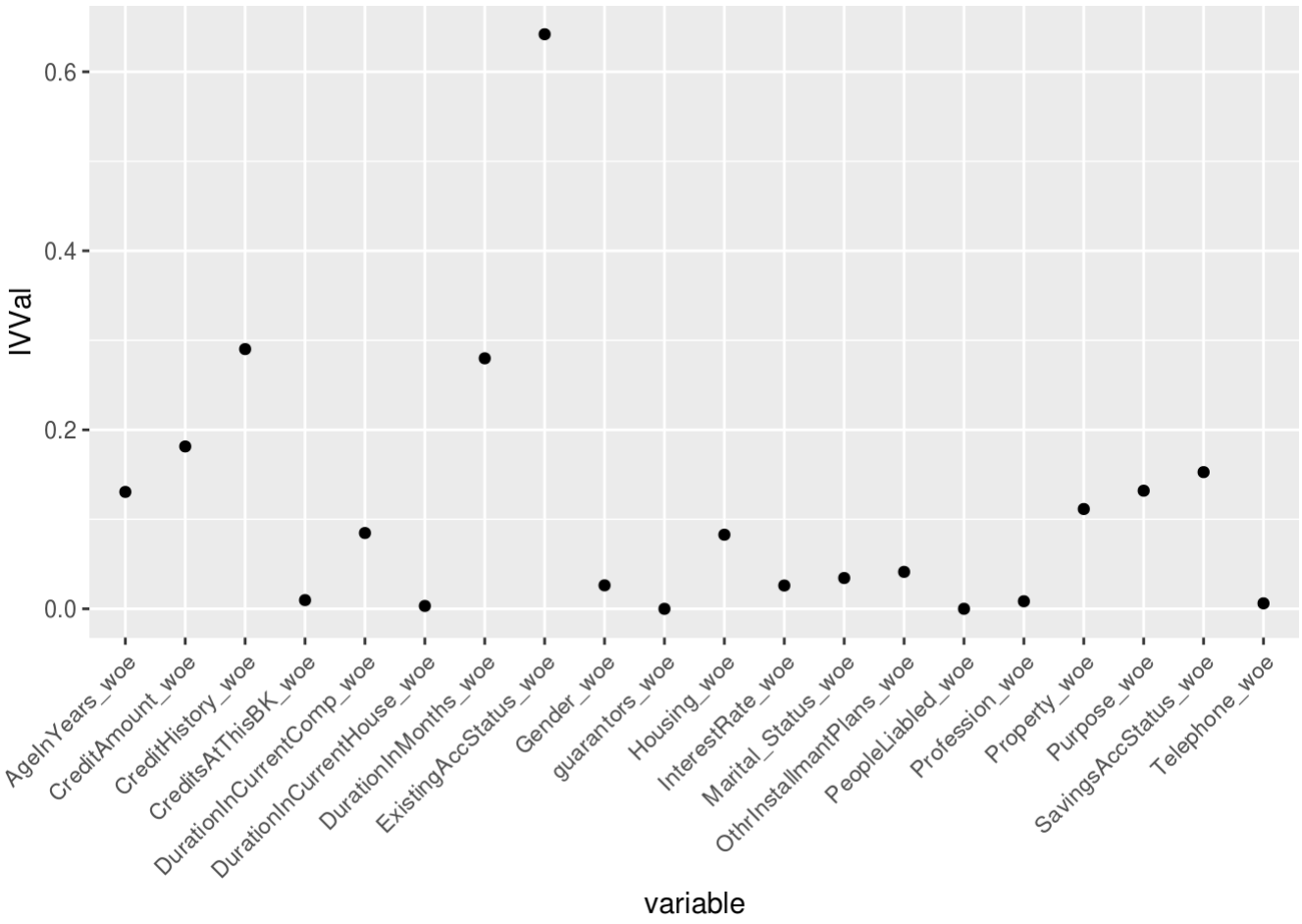
```
## Warning in check_y(dt, y, positive): The positive value in "CustomerClass"
## was replaced by 1 and negative value by 0.
```

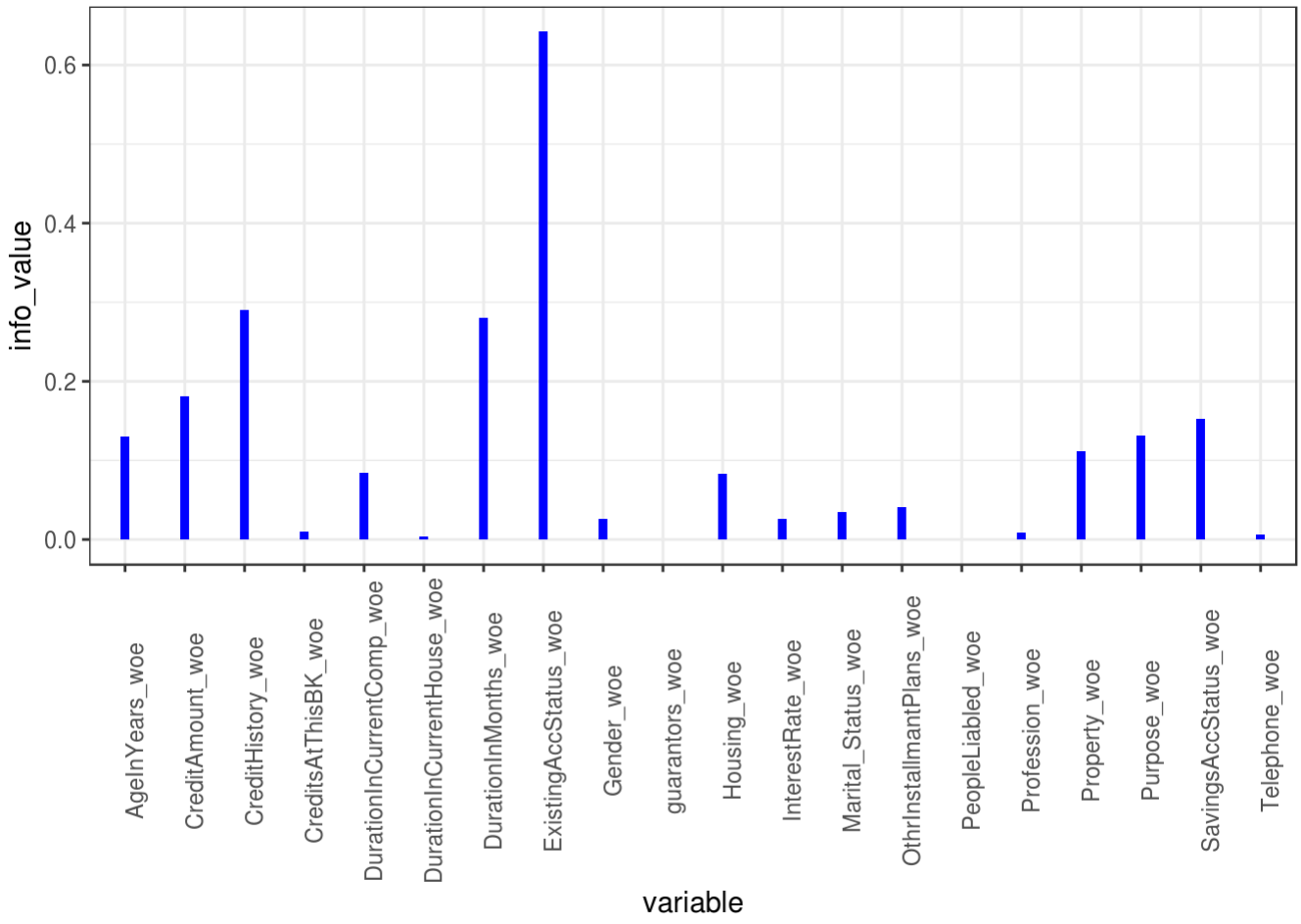
```
IVTable
```

```
##                       variable   info_value
##  1:      ExistingAccStatus_woe 6.419639e-01
##  2:          CreditHistory_woe 2.902378e-01
##  3:       DurationInMonths_woe 2.799110e-01
##  4:           CreditAmount_woe 1.814455e-01
##  5:       SavingsAccStatus_woe 1.527731e-01
##  6:                Purpose_woe 1.319848e-01
##  7:             AgeInYears_woe 1.306340e-01
##  8:               Property_woe 1.115337e-01
##  9:  DurationInCurrentComp_woe 8.472945e-02
## 10:                Housing_woe 8.282120e-02
## 11:   OthrInstallmantPlans_woe 4.129782e-02
## 12:         Marital_Status_woe 3.443217e-02
## 13:                 Gender_woe 2.624778e-02
## 14:           InterestRate_woe 2.605847e-02
## 15:        CreditsAtThisBK_woe 9.720873e-03
## 16:             Profession_woe 8.463893e-03
## 17:              Telephone_woe 6.109542e-03
## 18: DurationInCurrentHouse_woe 3.212187e-03
## 19:          PeopleLiabled_woe 5.186145e-05
## 20:             guarantors_woe 1.395409e-06
```

```
IVVal <- IVTable$info_value
```

plotting IV values





Buidling randomforest to understand the attribute importence

Approach - I

```
## Warning in `[<-.data.table`(x, j = name, value = value): Adding new column
## 'integerClass' then assigning NULL (deleting it).
```

```
## Classes 'data.table' and 'data.frame':   999 obs. of  23 variables:
##  $ CustomerClass             : Factor w/ 2 levels "Bad","Good": 1 2 2 1 2 2 2 2 1 1 ...
##  $ ExistingAccStatus_woe     : num  0.616 -1.178 0.616 0.616 -1.178 ...
##  $ DurationInMonths_woe      : num  1.134 -0.348 0.523 0.107 0.523 ...
##  $ CreditHistory_woe         : num  0.0869 -0.731 0.0869 0.0837 0.0869 ...
##  $ Purpose_woe               : num  -0.329 0.214 0.214 0.358 0.214 ...
##  $ CreditAmount_woe          : num  0.389 -0.26 0.389 0.389 0.389 ...
##  $ SavingsAccStatus_woe      : num  0.27 0.27 0.27 0.27 -0.699 ...
##  $ DurationInCurrentComp_woe : num  0.0307 -0.3958 -0.3958 0.0307 0.0307 ...
##  $ InterestRate_woe          : num  -0.192 -0.192 -0.192 -0.066 -0.192 ...
##  $ guarantors_woe            : num  -0.000379 -0.000379 0.003686 -0.000379 -0.000379 ...
##  $ DurationInCurrentHouse_woe: num  0.0687 -0.0125 -0.0125 -0.0125 -0.0125 ...
##  $ Property_woe              : num  -0.4579 -0.4579 0.0271 0.5847 0.5847 ...
##  $ AgeInYears_woe            : num  0.5274 -0.2965 -0.2965 0.0235 -0.8739 ...
##  $ OthrInstallmantPlans_woe  : num  -0.0859 -0.0859 -0.0859 -0.0859 -0.0859 ...
##  $ Housing_woe               : num  -0.194 -0.194 0.471 0.471 0.471 ...
##  $ CreditsAtThisBK_woe       : num  0.0734 0.0734 0.0734 -0.1325 0.0734 ...
##  $ Profession_woe            : num  -0.0183 -0.0986 -0.0183 -0.0183 -0.0986 ...
##  $ PeopleLiabled_woe         : num  0.00308 -0.01684 -0.01684 -0.01684 -0.01684 ...
##  $ Telephone_woe             : num  0.0633 0.0633 0.0633 0.0633 -0.0966 ...
##  $ ForeignWorker_woe         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Marital_Status_woe        : num  0.153 -0.164 -0.164 -0.164 -0.164 ...
##  $ Gender_woe                : num  0.234 -0.112 -0.112 -0.112 -0.112 ...
##  $ integerClass_woe          : num  0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
set.seed(1234)
split_indices <- sample.split(RandomForestData$CustomerClass, SplitRatio = 0.70)

str(RandomForestData)
```

```
## Classes 'data.table' and 'data.frame':   999 obs. of  23 variables:
##  $ CustomerClass             : Factor w/ 2 levels "Bad","Good": 1 2 2 1 2 2 2 2 1 1 ...
##  $ ExistingAccStatus_woe     : num  0.616 -1.178 0.616 0.616 -1.178 ...
##  $ DurationInMonths_woe      : num  1.134 -0.348 0.523 0.107 0.523 ...
##  $ CreditHistory_woe         : num  0.0869 -0.731 0.0869 0.0837 0.0869 ...
##  $ Purpose_woe               : num  -0.329 0.214 0.214 0.358 0.214 ...
##  $ CreditAmount_woe          : num  0.389 -0.26 0.389 0.389 0.389 ...
##  $ SavingsAccStatus_woe      : num  0.27 0.27 0.27 0.27 -0.699 ...
##  $ DurationInCurrentComp_woe : num  0.0307 -0.3958 -0.3958 0.0307 0.0307 ...
##  $ InterestRate_woe          : num  -0.192 -0.192 -0.192 -0.066 -0.192 ...
##  $ guarantors_woe            : num  -0.000379 -0.000379 0.003686 -0.000379 -0.000379 ...
##  $ DurationInCurrentHouse_woe: num  0.0687 -0.0125 -0.0125 -0.0125 -0.0125 ...
##  $ Property_woe              : num  -0.4579 -0.4579 0.0271 0.5847 0.5847 ...
##  $ AgeInYears_woe            : num  0.5274 -0.2965 -0.2965 0.0235 -0.8739 ...
##  $ OthrInstallmantPlans_woe  : num  -0.0859 -0.0859 -0.0859 -0.0859 -0.0859 ...
##  $ Housing_woe               : num  -0.194 -0.194 0.471 0.471 0.471 ...
##  $ CreditsAtThisBK_woe       : num  0.0734 0.0734 0.0734 -0.1325 0.0734 ...
##  $ Profession_woe            : num  -0.0183 -0.0986 -0.0183 -0.0183 -0.0986 ...
##  $ PeopleLiabled_woe         : num  0.00308 -0.01684 -0.01684 -0.01684 -0.01684 ...
##  $ Telephone_woe             : num  0.0633 0.0633 0.0633 0.0633 -0.0966 ...
##  $ ForeignWorker_woe         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Marital_Status_woe        : num  0.153 -0.164 -0.164 -0.164 -0.164 ...
##  $ Gender_woe                : num  0.234 -0.112 -0.112 -0.112 -0.112 ...
##  $ integerClass_woe          : num  0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
RandomForest_train <- RandomForestData[split_indices, ]
RandomForest_test <- RandomForestData[!split_indices, ]
this_mtry <- floor(sqrt(ncol(RandomForest_train)))
```

RandomForest modelling

```
RandomForestModel <- randomForest(CustomerClass ~ . ,
                                  data = RandomForest_train[,-13],
                                  ntree = 50, mtry = this_mtry, sampsize = 10,
                                  importance = TRUE)
RandomForestModel$confusion
```

```
##      Bad Good class.error
## Bad    4  206   0.9809524
## Good   0  489   0.0000000
```

Approch-II

```
## dummies-1.5.6 provided by Decision Patterns
```

```
##
## Attaching package: 'dummies'
```

```
## The following object is masked from 'package:lme4':
##
##     dummy
```

```
## 'data.frame':    999 obs. of  72 variables:
##  $ ExistingAccStatusLessThan0DM       : int  0 0 1 1 0 0 0 0 0 0 ...
##  $ ExistingAccStatusLessThan200DM     : int  1 0 0 0 0 0 1 0 1 1 ...
##  $ ExistingAccStatusMoreThanOrEq200DM : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ExistingAccStatusNoAcc             : int  0 1 0 0 1 1 0 1 0 0 ...
##  $ DurationInMonths                   : int  48 12 42 24 36 24 36 12 30 12 ...
##  $ CreditHistoryCriticalAccount       : int  0 1 0 0 0 0 0 0 1 0 ...
##  $ CreditHistoryDelayInPast           : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ CreditHistoryNoCreditsDues         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditHistoryNoCreditsForThisBank  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditHistoryNoPreviousCreditDues  : int  1 0 1 0 1 1 1 1 0 1 ...
##  $ Purposebusiness                    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeDomestic Appliances         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeEducation                   : int  0 1 0 0 1 0 0 0 0 0 ...
##  $ PurposeFurniture/Equipment         : int  0 0 1 0 0 1 0 0 0 0 ...
##  $ PurposeNewCar                      : int  0 0 0 1 0 0 0 0 1 1 ...
##  $ PurposeOldCar                      : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ PurposeOthers                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeRadio/Television            : int  1 0 0 0 0 0 0 1 0 0 ...
##  $ PurposeRepairs                     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeRetraining                  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditAmount                       : num  5500 2096 5500 4870 5500 ...
##  $ SavingsAccStatusGr1000DM           : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ SavingsAccStatusGr100Less500DM     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ SavingsAccStatusGr500Less1000DM    : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ SavingsAccStatusLess100DM          : int  1 1 1 1 0 0 1 0 1 1 ...
##  $ SavingsAccStatusNoSavingsAcc       : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ DurationInCurrentCompGr1YrLess4Yrs : int  1 0 0 1 1 0 1 0 0 0 ...
##  $ DurationInCurrentCompGr4YrsLess7Yrs: int  0 1 1 0 0 0 0 1 0 0 ...
##  $ DurationInCurrentCompGr7Yrs        : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ DurationInCurrentCompLess1Yr       : int  0 0 0 0 0 0 0 0 0 0 1 ...
##  $ DurationInCurrentCompUnemployed    : int  0 0 0 0 0 0 0 0 0 1 0 ...
##  $ InterestRate1                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ InterestRate2                      : int  1 1 1 0 1 0 1 1 0 0 ...
##  $ InterestRate3                      : int  0 0 0 1 0 1 0 0 0 1 ...
##  $ InterestRate4                      : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ guarantorsCo-Applicant             : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ guarantorsGuarantor                : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ guarantorsNone                     : int  1 1 0 1 1 1 1 1 1 1 ...
##  $ DurationInCurrentHouse             : int  2 3 4 4 4 4 2 4 2 1 ...
##  $ PropertyLifeInsurance              : int  0 0 1 0 0 1 0 0 0 0 ...
##  $ PropertyNo Property                : int  0 0 0 1 1 0 0 0 0 0 ...
##  $ PropertyOthers                     : int  0 0 0 0 0 0 1 0 1 1 ...
##  $ PropertyRealEstate                 : int  1 1 0 0 0 0 0 1 0 0 ...
##  $ AgeInYears                         : int  22 49 45 53 35 53 35 61 28 25 ...
##  $ OthrInstallmantPlansBank           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ OthrInstallmantPlansNone           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ OthrInstallmantPlansStores         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HousingForFree                     : int  0 0 1 1 1 0 0 0 0 0 ...
##  $ HousingOwn                         : int  1 1 0 0 0 1 0 1 1 0 ...
##  $ HousingRent                        : int  0 0 0 0 0 0 1 0 0 1 ...
##  $ CreditsAtThisBK1                   : int  1 1 1 0 1 1 1 1 0 1 ...
##  $ CreditsAtThisBK2                   : int  0 0 0 1 0 0 0 0 1 0 ...
##  $ CreditsAtThisBK3                   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditsAtThisBK4                   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ProfessionOfficer                  : int  0 0 0 0 0 0 1 0 1 0 ...
##  $ ProfessionOfficial                 : int  1 0 1 1 0 1 0 0 0 1 ...
##  $ ProfessionUnskilledNonResisdent    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ProfessionUnskilledResident        : int  0 1 0 0 1 0 0 1 0 0 ...
##  $ PeopleLiabled1                     : int  1 0 0 0 0 1 1 1 1 1 ...
##  $ PeopleLiabled2                     : int  0 1 1 1 1 0 0 0 0 0 ...
```

```
##  $ TelephoneNo                       : int  1 1 1 1 0 1 0 1 1 1 ...
##  $ TelephoneYes                      : int  0 0 0 0 1 0 1 0 0 0 ...
##  $ ForeignWorkerNo                   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ForeignWorkerYes                  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ CustomerClassBad                  : int  1 0 0 1 0 0 0 0 1 1 ...
##  $ CustomerClassGood                 : int  0 1 1 0 1 1 1 1 0 0 ...
##  $ Marital_StatusMarried             : int  1 0 0 0 0 0 0 0 1 1 ...
##  $ Marital_StatusSeparated           : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Marital_StatusSingle              : int  0 1 1 1 1 1 1 0 0 0 ...
##  $ GenderFemale                      : int  1 0 0 0 0 0 0 0 0 1 ...
##  $ GenderMale                        : int  0 1 1 1 1 1 1 1 1 0 ...
##  $ integerClass                      : int  2 1 1 2 1 1 1 1 2 2 ...
##  - attr(*, "dummies")=List of 18
##   ..$ ExistingAccStatus    : int  1 2 3 4
##   ..$ CreditHistory        : int  6 7 8 9 10
##   ..$ Purpose              : int  11 12 13 14 15 16 17 18 19 20
##   ..$ SavingsAccStatus     : int  22 23 24 25 26
##   ..$ DurationInCurrentComp: int  27 28 29 30 31
##   ..$ InterestRate         : int  32 33 34 35
##   ..$ guarantors           : int  36 37 38
##   ..$ Property             : int  40 41 42 43
##   ..$ OthrInstallmantPlans : int  45 46 47
##   ..$ Housing              : int  48 49 50
##   ..$ CreditsAtThisBK      : int  51 52 53 54
##   ..$ Profession           : int  55 56 57 58
##   ..$ PeopleLiabled        : int  59 60
##   ..$ Telephone            : int  61 62
##   ..$ ForeignWorker        : int  63 64
##   ..$ CustomerClass        : int  65 66
##   ..$ Marital_Status       : int  67 68 69
##   ..$ Gender               : int  70 71
```

Data preparation for RandomForest Approach-II

Splitting data by 70:30

```
## 'data.frame':    999 obs. of  72 variables:
##  $ ExistingAccStatusLessThan0DM       : int  0 0 1 1 0 0 0 0 0 0 ...
##  $ ExistingAccStatusLessThan200DM     : int  1 0 0 0 0 0 1 0 1 1 ...
##  $ ExistingAccStatusMoreThanOrEq200DM : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ExistingAccStatusNoAcc             : int  0 1 0 0 1 1 0 1 0 0 ...
##  $ DurationInMonths                   : int  48 12 42 24 36 24 36 12 30 12 ...
##  $ CreditHistoryCriticalAccount       : int  0 1 0 0 0 0 0 0 1 0 ...
##  $ CreditHistoryDelayInPast           : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ CreditHistoryNoCreditsDues         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditHistoryNoCreditsForThisBank  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditHistoryNoPreviousCreditDues  : int  1 0 1 0 1 1 1 1 0 1 ...
##  $ Purposebusiness                    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeEducation                   : int  0 1 0 0 1 0 0 0 0 0 ...
##  $ PurposeNewCar                      : int  0 0 0 1 0 0 0 0 1 1 ...
##  $ PurposeOldCar                      : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ PurposeOthers                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeRepairs                     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeRetraining                  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditAmount                       : num  5500 2096 5500 4870 5500 ...
##  $ SavingsAccStatusGr1000DM           : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ SavingsAccStatusGr100Less500DM     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ SavingsAccStatusGr500Less1000DM    : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ SavingsAccStatusLess100DM          : int  1 1 1 1 0 0 1 0 1 1 ...
##  $ SavingsAccStatusNoSavingsAcc       : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ DurationInCurrentCompGr1YrLess4Yrs : int  1 0 0 1 1 0 1 0 0 0 ...
##  $ DurationInCurrentCompGr4YrsLess7Yrs: int  0 1 1 0 0 0 0 1 0 0 ...
##  $ DurationInCurrentCompGr7Yrs        : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ DurationInCurrentCompLess1Yr       : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ DurationInCurrentCompUnemployed    : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ InterestRate1                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ InterestRate2                      : int  1 1 1 0 1 0 1 1 0 0 ...
##  $ InterestRate3                      : int  0 0 0 1 0 1 0 0 0 1 ...
##  $ InterestRate4                      : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ guarantorsGuarantor                : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ guarantorsNone                     : int  1 1 0 1 1 1 1 1 1 1 ...
##  $ DurationInCurrentHouse             : int  2 3 4 4 4 4 2 4 2 1 ...
##  $ PropertyLifeInsurance              : int  0 0 1 0 0 1 0 0 0 0 ...
##  $ PropertyOthers                     : int  0 0 0 0 0 0 1 0 1 1 ...
##  $ PropertyRealEstate                 : int  1 1 0 0 0 0 0 1 0 0 ...
##  $ AgeInYears                         : int  22 49 45 53 35 53 35 61 28 25 ...
##  $ OthrInstallmantPlansBank           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ OthrInstallmantPlansNone           : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ OthrInstallmantPlansStores         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HousingForFree                     : int  0 0 1 1 1 0 0 0 0 0 ...
##  $ HousingOwn                         : int  1 1 0 0 0 1 0 1 1 0 ...
##  $ HousingRent                        : int  0 0 0 0 0 0 1 0 0 1 ...
##  $ CreditsAtThisBK1                   : int  1 1 1 0 1 1 1 1 0 1 ...
##  $ CreditsAtThisBK2                   : int  0 0 0 1 0 0 0 0 1 0 ...
##  $ CreditsAtThisBK3                   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CreditsAtThisBK4                   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ProfessionOfficer                  : int  0 0 0 0 0 0 1 0 1 0 ...
##  $ ProfessionOfficial                 : int  1 0 1 1 0 1 0 0 0 1 ...
##  $ ProfessionUnskilledNonResisdent    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ProfessionUnskilledResident        : int  0 1 0 0 1 0 0 1 0 0 ...
##  $ PeopleLiabled1                     : int  1 0 0 0 0 1 1 1 1 1 ...
##  $ PeopleLiabled2                     : int  0 1 1 1 1 0 0 0 0 0 ...
##  $ TelephoneNo                        : int  1 1 1 1 0 1 0 1 1 1 ...
##  $ TelephoneYes                       : int  0 0 0 0 1 0 1 0 0 0 ...
##  $ ForeignWorkerNo                    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ ForeignWorkerYes                   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ CustomerClassBad                   : int  1 0 0 1 0 0 0 0 1 1 ...
```

```
##  $ CustomerClassGood                : int  0 1 1 0 1 1 1 1 0 0 ...
##  $ Marital_StatusMarried            : int  1 0 0 0 0 0 0 0 1 1 ...
##  $ Marital_StatusSeparated          : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Marital_StatusSingle             : int  0 1 1 1 1 1 1 0 0 0 ...
##  $ GenderFemale                     : int  1 0 0 0 0 0 0 0 0 1 ...
##  $ GenderMale                       : int  0 1 1 1 1 1 1 1 1 0 ...
##  $ integerClass                     : Factor w/ 2 levels "1","2": 2 1 1 2 1 1 1 1 2 2 ...
##  $ PurposeDomestic_Appliances       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ PurposeRadio_Television          : int  1 0 0 0 0 0 0 1 0 0 ...
##  $ PurposeFurniture_Equipment       : int  0 0 1 0 0 1 0 0 0 0 ...
##  $ guarantorsCo_Applicant           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Property_NoProperty              : int  0 0 0 1 1 0 0 0 0 0 ...
##  - attr(*, "dummies")=List of 18
##   ..$ ExistingAccStatus    : int  1 2 3 4
##   ..$ CreditHistory        : int  6 7 8 9 10
##   ..$ Purpose              : int  11 12 13 14 15 16 17 18 19 20
##   ..$ SavingsAccStatus     : int  22 23 24 25 26
##   ..$ DurationInCurrentComp: int  27 28 29 30 31
##   ..$ InterestRate         : int  32 33 34 35
##   ..$ guarantors           : int  36 37 38
##   ..$ Property             : int  40 41 42 43
##   ..$ OthrInstallmantPlans : int  45 46 47
##   ..$ Housing              : int  48 49 50
##   ..$ CreditsAtThisBK      : int  51 52 53 54
##   ..$ Profession           : int  55 56 57 58
##   ..$ PeopleLiabled        : int  59 60
##   ..$ Telephone            : int  61 62
##   ..$ ForeignWorker        : int  63 64
##   ..$ CustomerClass        : int  65 66
##   ..$ Marital_Status       : int  67 68 69
##   ..$ Gender               : int  70 71
```

RandomForest model building with dummy values

```
##     1   2 class.error
## 1 489   0   0.0000000
## 2  36 174   0.1714286
```

Performing predictions on randomforest model

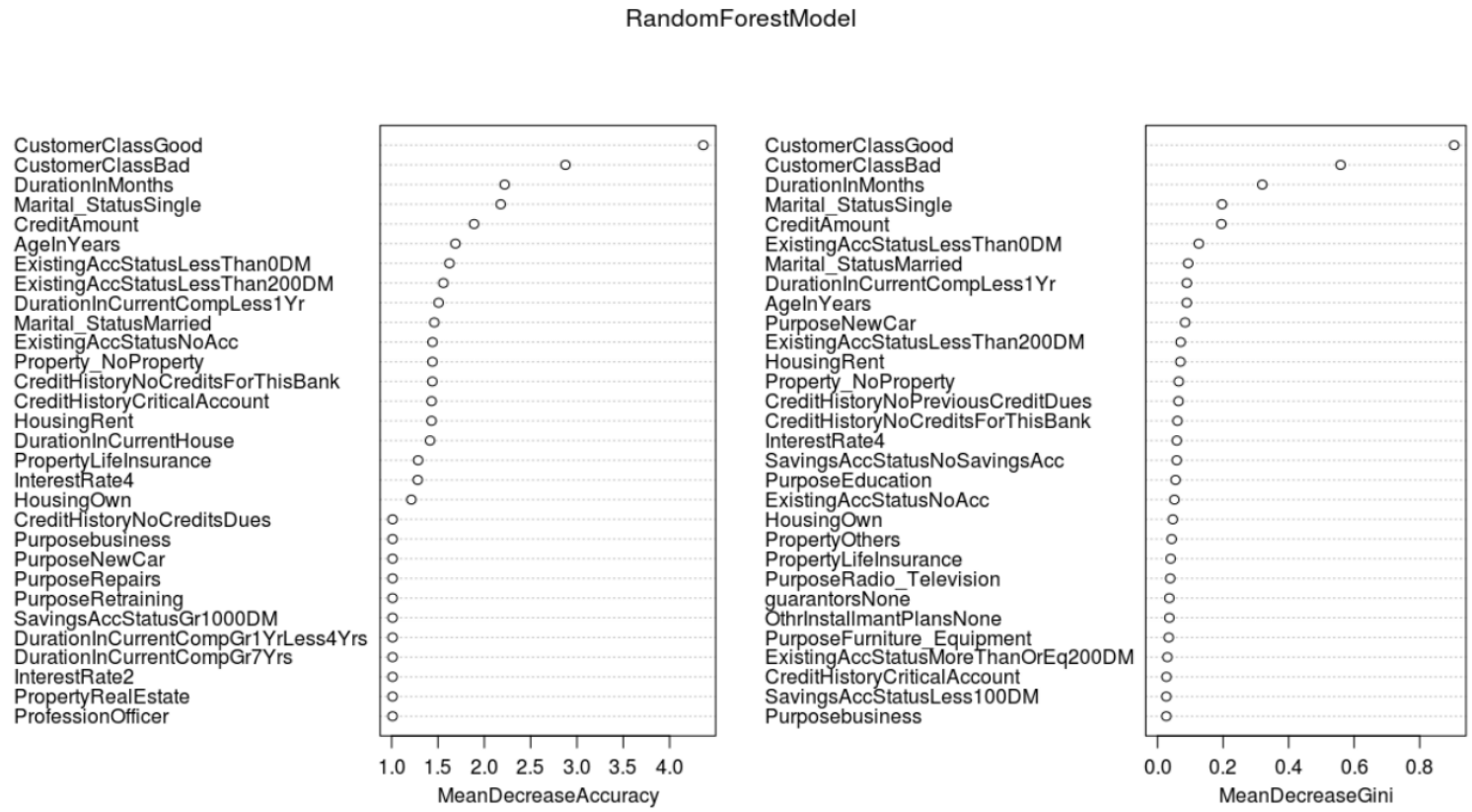
```
##                  actual_response
## predicted_response Bad Good
##               Bad   67    0
##               Good  23  210
```

```
##                                          1          2
## ExistingAccStatusLessThan0DM         1.5142101  1.6788677
## ExistingAccStatusLessThan200DM       1.8940508  1.4004179
## ExistingAccStatusMoreThanOrEq200DM  -1.0101525  0.0000000
## ExistingAccStatusNoAcc               1.0101525  1.4427841
## DurationInMonths                     1.5544654  2.6776110
## CreditHistoryCriticalAccount         1.3540122  1.0101525
## CreditHistoryDelayInPast             0.0000000  0.0000000
## CreditHistoryNoCreditsDues           1.0101525  1.0101525
## CreditHistoryNoCreditsForThisBank    1.4340145  1.2846891
## CreditHistoryNoPreviousCreditDues    1.0101525 -1.0101525
## Purposebusiness                     -1.0101525  1.0101525
## PurposeEducation                    -0.2028698  1.0101525
## PurposeNewCar                        1.0101525  1.0101525
## PurposeOldCar                        0.0000000  0.0000000
## PurposeOthers                        0.0000000  0.0000000
## PurposeRepairs                       1.0101525 -1.0101525
## PurposeRetraining                    1.0101525  1.0101525
## CreditAmount                         1.5026967  1.5143760
## SavingsAccStatusGr1000DM             1.0101525  1.0101525
## SavingsAccStatusGr100Less500DM       0.0000000  0.0000000
## SavingsAccStatusGr500Less1000DM      0.0000000  0.0000000
## SavingsAccStatusLess100DM            1.0101525  1.0101525
## SavingsAccStatusNoSavingsAcc        -1.4315823 -1.4277815
## DurationInCurrentCompGr1YrLess4Yrs  -1.0101525  1.0101525
## DurationInCurrentCompGr4YrsLess7Yrs  0.0000000  0.0000000
## DurationInCurrentCompGr7Yrs         -1.0101525  1.0101525
## DurationInCurrentCompLess1Yr         1.3492796  1.4225632
## DurationInCurrentCompUnemployed      0.0000000  0.0000000
## InterestRate1                        0.0000000  0.0000000
## InterestRate2                        1.0101525  1.0101525
## InterestRate3                       -1.0101525 -1.0101525
## InterestRate4                        0.7721697  1.4457858
## guarantorsGuarantor                  0.0000000  0.0000000
## guarantorsNone                      -1.0101525  0.0000000
## DurationInCurrentHouse               0.1533691  0.6312548
## PropertyLifeInsurance                1.0101525 -0.3374588
## PropertyOthers                      -1.1669206  1.1918276
## PropertyRealEstate                   1.0101525 -1.0101525
## AgeInYears                           1.4297771  1.2317872
## OthrInstallmantPlansBank            -1.0101525  0.0000000
## OthrInstallmantPlansNone             1.0101525 -0.8094678
## OthrInstallmantPlansStores           0.0000000  0.0000000
## HousingForFree                       0.0000000  0.0000000
## HousingOwn                           0.9392981  1.3214332
## HousingRent                          1.3564976  1.4083358
## CreditsAtThisBK1                    -1.0101525  1.0101525
## CreditsAtThisBK2                    -1.0101525  1.0101525
## CreditsAtThisBK3                     0.0000000  0.0000000
## CreditsAtThisBK4                    -1.0101525  0.0000000
## ProfessionOfficer                   -1.0101525  1.0101525
## ProfessionOfficial                   1.0101525 -1.0101525
## ProfessionUnskilledNonResisdent      0.0000000  0.0000000
## ProfessionUnskilledResident          0.0000000  0.0000000
## PeopleLiabled1                       0.0000000  1.0101525
## PeopleLiabled2                      -1.0101525  1.0101525
## TelephoneNo                          0.0000000  1.0101525
## TelephoneYes                         1.0101525 -1.0101525
## ForeignWorkerNo                      0.0000000 -1.0101525
## ForeignWorkerYes                     0.0000000  0.0000000
## CustomerClassBad                     2.8660307  2.8834187
```

```
## CustomerClassGood                    4.2459474  4.4359775
## Marital_StatusMarried                1.5523206  0.9078255
## Marital_StatusSeparated              0.0000000  0.0000000
## Marital_StatusSingle                 1.7708734  1.5596659
## GenderFemale                         0.0000000  0.0000000
## GenderMale                           1.0101525  1.0101525
## PurposeDomestic_Appliances           0.0000000  0.0000000
## PurposeRadio_Television              1.2309149 -0.5588230
## PurposeFurniture_Equipment           1.2309149 -1.0101525
## guarantorsCo_Applicant              -1.0101525  0.0000000
## Property_NoProperty                 -1.2694554  1.3993410
##                                    MeanDecreaseAccuracy MeanDecreaseGini
## ExistingAccStatusLessThan0DM                  1.6236543      0.125777778
## ExistingAccStatusLessThan200DM                1.5582794      0.070349206
## ExistingAccStatusMoreThanOrEq200DM           -1.0101525      0.030000000
## ExistingAccStatusNoAcc                        1.4408663      0.051428571
## DurationInMonths                              2.2198722      0.319460317
## CreditHistoryCriticalAccount                  1.4311758      0.027142857
## CreditHistoryDelayInPast                      0.0000000      0.000000000
## CreditHistoryNoCreditsDues                    1.0101525      0.012000000
## CreditHistoryNoCreditsForThisBank             1.4389479      0.060714286
## CreditHistoryNoPreviousCreditDues             0.0000000      0.064000000
## Purposebusiness                               1.0101525      0.026666667
## PurposeEducation                              0.2775637      0.054857143
## PurposeNewCar                                 1.0101525      0.084000000
## PurposeOldCar                                 0.0000000      0.000000000
## PurposeOthers                                 0.0000000      0.000000000
## PurposeRepairs                                1.0101525      0.020000000
## PurposeRetraining                             1.0101525      0.006666667
## CreditAmount                                  1.8894624      0.194825397
## SavingsAccStatusGr1000DM                      1.0101525      0.003333333
## SavingsAccStatusGr100Less500DM                0.0000000      0.000000000
## SavingsAccStatusGr500Less1000DM               0.0000000      0.000000000
## SavingsAccStatusLess100DM                     1.0101525      0.026666667
## SavingsAccStatusNoSavingsAcc                 -1.4433757      0.058444444
## DurationInCurrentCompGr1YrLess4Yrs            1.0101525      0.007619048
## DurationInCurrentCompGr4YrsLess7Yrs           0.0000000      0.000000000
## DurationInCurrentCompGr7Yrs                   1.0101525      0.006666667
## DurationInCurrentCompLess1Yr                  1.5070388      0.089333333
## DurationInCurrentCompUnemployed               0.0000000      0.000000000
## InterestRate1                                 0.0000000      0.000000000
## InterestRate2                                 1.0101525      0.026666667
## InterestRate3                                -1.0101525      0.006666667
## InterestRate4                                 1.2815365      0.058666667
## guarantorsGuarantor                           0.0000000      0.000000000
## guarantorsNone                               -1.0101525      0.036000000
## DurationInCurrentHouse                        1.4142136      0.020952381
## PropertyLifeInsurance                         1.2856487      0.040000000
## PropertyOthers                               -1.1461365      0.043047619
## PropertyRealEstate                            1.0101525      0.016666667
## AgeInYears                                    1.6884859      0.089142857
## OthrInstallmantPlansBank                     -1.0101525      0.012000000
## OthrInstallmantPlansNone                     -0.7207500      0.036000000
## OthrInstallmantPlansStores                    0.0000000      0.000000000
## HousingForFree                                0.0000000      0.000000000
## HousingOwn                                    1.2125667      0.046666667
## HousingRent                                   1.4307534      0.069571429
## CreditsAtThisBK1                             -1.0101525      0.010000000
## CreditsAtThisBK2                             -1.0101525      0.024000000
## CreditsAtThisBK3                              0.0000000      0.000000000
## CreditsAtThisBK4                             -1.0101525      0.020000000
## ProfessionOfficer                             1.0101525      0.021333333
```

```
## ProfessionOfficial                    1.0101525      0.016000000
## ProfessionUnskilledNonResisdent       0.0000000      0.000000000
## ProfessionUnskilledResident           0.0000000      0.000000000
## PeopleLiabled1                        1.0101525      0.002000000
## PeopleLiabled2                        1.0101525      0.023047619
## TelephoneNo                           1.0101525      0.006666667
## TelephoneYes                          1.0101525      0.004000000
## ForeignWorkerNo                      -1.0101525      0.020000000
## ForeignWorkerYes                      0.0000000      0.000000000
## CustomerClassBad                      2.8762697      0.558666667
## CustomerClassGood                     4.3628269      0.905238095
## Marital_StatusMarried                 1.4603292      0.093333333
## Marital_StatusSeparated               0.0000000      0.000000000
## Marital_StatusSingle                  2.1772599      0.196666667
## GenderFemale                          0.0000000      0.000000000
## GenderMale                            1.0101525      0.007619048
## PurposeDomestic_Appliances            0.0000000      0.000000000
## PurposeRadio_Television               1.0101525      0.038666667
## PurposeFurniture_Equipment            0.3183214      0.034095238
## guarantorsCo_Applicant               -1.0101525      0.020000000
## Property_NoProperty                   1.4400461      0.064666667
```

Interpreting variable importance



RandomForestVariableImportance

Building decision tree using rpart

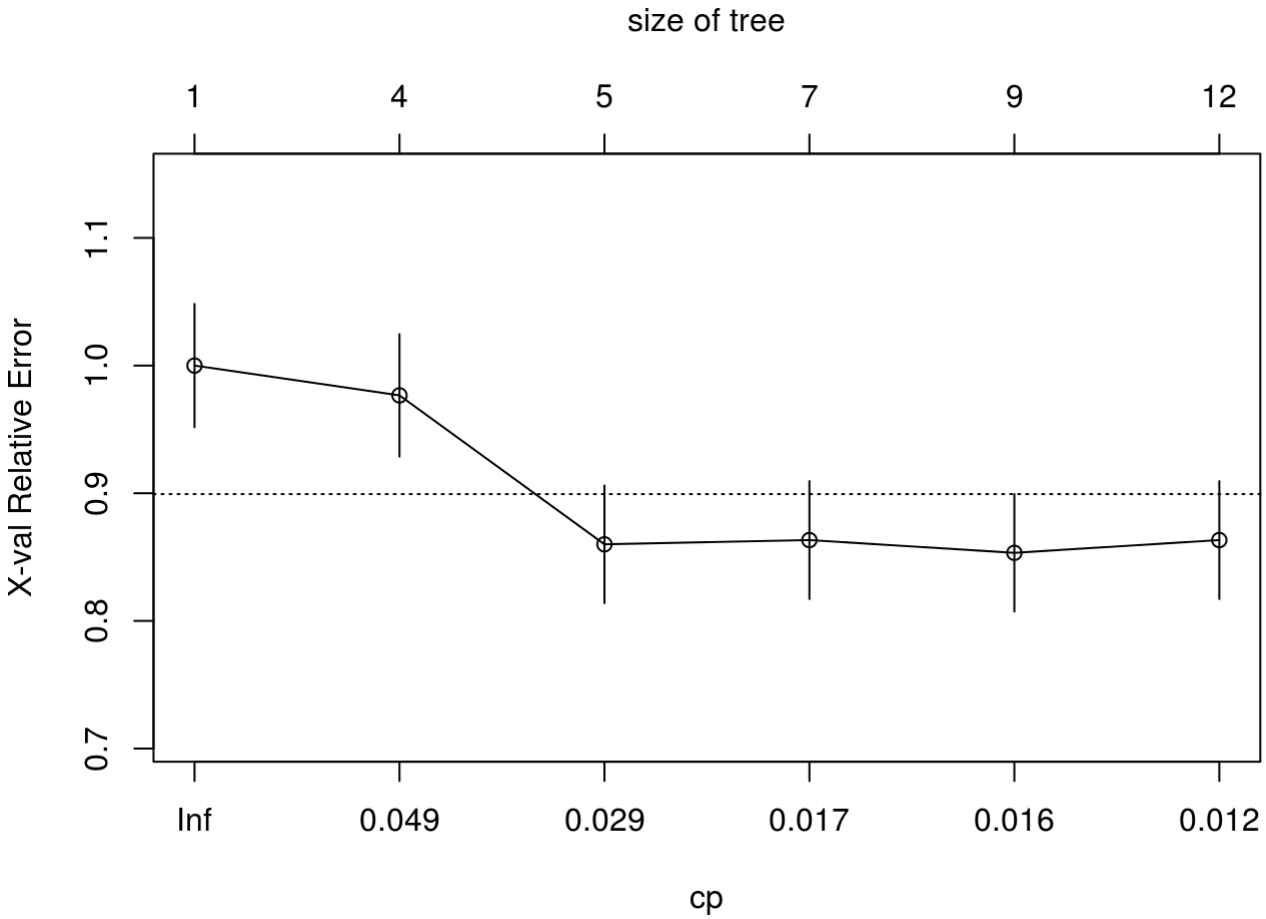
```
## Classes 'data.table' and 'data.frame':   999 obs. of  22 variables:
##  $ ExistingAccStatus     : Factor w/ 4 levels "LessThan0DM",..: 2 4 1 1 4 4 2 4 2 2 ...
##  $ DurationInMonths      : int  48 12 42 24 36 24 36 12 30 12 ...
##  $ CreditHistory         : Factor w/ 5 levels "CriticalAccount",..: 5 1 5 2 5 5 5 5 1 5 ...
##  $ Purpose               : Factor w/ 10 levels "business","Domestic Appliances",..: 8 3 4 5
3 4 6 8 5 5 ...
##  $ CreditAmount          : int  5951 2096 7882 4870 9055 2835 6948 3059 5234 1295 ...
##  $ SavingsAccStatus      : Factor w/ 5 levels "Gr1000DM","Gr100Less500DM",..: 4 4 4 4 5 3 4
1 4 4 ...
##  $ DurationInCurrentComp : Factor w/ 5 levels "Gr1YrLess4Yrs",..: 1 2 2 1 1 3 1 2 5 4 ...
##  $ InterestRate          : Factor w/ 4 levels "1","2","3","4": 2 2 2 3 2 3 2 2 4 3 ...
##  $ guarantors            : Factor w/ 3 levels "Co-Applicant",..: 3 3 2 3 3 3 3 3 3 3 ...
##  $ DurationInCurrentHouse: Factor w/ 4 levels "1","2","3","4": 2 3 4 4 4 4 2 4 2 1 ...
##  $ Property              : Factor w/ 4 levels "LifeInsurance",..: 4 4 1 2 2 1 3 4 3 3 ...
##  $ AgeInYears            : int  22 49 45 53 35 53 35 61 28 25 ...
##  $ OthrInstallmantPlans  : Factor w/ 3 levels "Bank","None",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Housing               : Factor w/ 3 levels "ForFree","Own",..: 2 2 1 1 1 2 3 2 2 3 ...
##  $ CreditsAtThisBK       : Factor w/ 4 levels "1","2","3","4": 1 1 1 2 1 1 1 1 2 1 ...
##  $ Profession            : Factor w/ 4 levels "Officer","Official",..: 2 4 2 2 4 2 1 4 1 2
...
##  $ PeopleLiabled         : Factor w/ 2 levels "1","2": 1 2 2 2 2 1 1 1 1 1 ...
##  $ Telephone             : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 2 1 1 1 ...
##  $ ForeignWorker         : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ CustomerClass         : Factor w/ 2 levels "Bad","Good": 1 2 2 1 2 2 2 2 1 1 ...
##  $ Marital_Status        : Factor w/ 3 levels "Married","Separated",..: 1 3 3 3 3 3 3 2 1 1
...
##  $ Gender                : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 2 2 2 2 1 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

Splitting data into 70:30

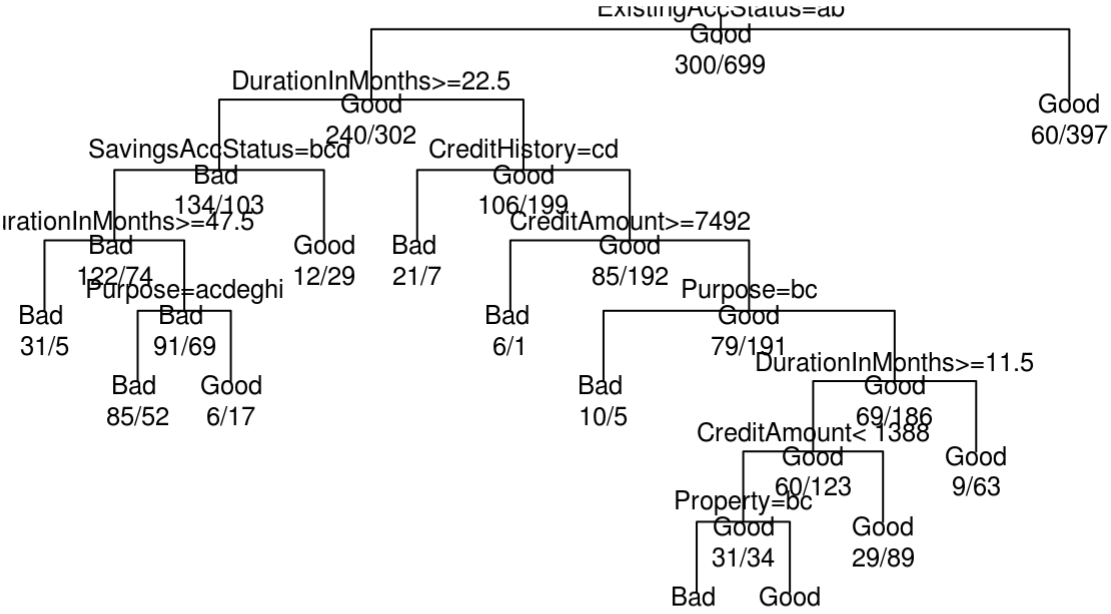
```
## Classes 'data.table' and 'data.frame':   999 obs. of  22 variables:
##  $ ExistingAccStatus     : Factor w/ 4 levels "LessThan0DM",..: 2 4 1 1 4 4 2 4 2 2 ...
##  $ DurationInMonths      : int  48 12 42 24 36 24 36 12 30 12 ...
##  $ CreditHistory         : Factor w/ 5 levels "CriticalAccount",..: 5 1 5 2 5 5 5 5 1 5 ...
##  $ Purpose               : Factor w/ 10 levels "business","Domestic Appliances",..: 8 3 4 5
3 4 6 8 5 5 ...
##  $ CreditAmount          : int  5951 2096 7882 4870 9055 2835 6948 3059 5234 1295 ...
##  $ SavingsAccStatus      : Factor w/ 5 levels "Gr1000DM","Gr100Less500DM",..: 4 4 4 4 5 3 4
1 4 4 ...
##  $ DurationInCurrentComp : Factor w/ 5 levels "Gr1YrLess4Yrs",..: 1 2 2 1 1 3 1 2 5 4 ...
##  $ InterestRate          : Factor w/ 4 levels "1","2","3","4": 2 2 2 3 2 3 2 2 4 3 ...
##  $ guarantors            : Factor w/ 3 levels "Co-Applicant",..: 3 3 2 3 3 3 3 3 3 3 ...
##  $ DurationInCurrentHouse: Factor w/ 4 levels "1","2","3","4": 2 3 4 4 4 4 2 4 2 1 ...
##  $ Property              : Factor w/ 4 levels "LifeInsurance",..: 4 4 1 2 2 1 3 4 3 3 ...
##  $ AgeInYears            : int  22 49 45 53 35 53 35 61 28 25 ...
##  $ OthrInstallmantPlans  : Factor w/ 3 levels "Bank","None",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Housing               : Factor w/ 3 levels "ForFree","Own",..: 2 2 1 1 1 2 3 2 2 3 ...
##  $ CreditsAtThisBK       : Factor w/ 4 levels "1","2","3","4": 1 1 1 2 1 1 1 1 2 1 ...
##  $ Profession            : Factor w/ 4 levels "Officer","Official",..: 2 4 2 2 4 2 1 4 1 2
...
##  $ PeopleLiabled         : Factor w/ 2 levels "1","2": 1 2 2 2 2 1 1 1 1 1 ...
##  $ Telephone             : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 2 1 1 1 ...
##  $ ForeignWorker         : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ CustomerClass         : Factor w/ 2 levels "Bad","Good": 1 2 2 1 2 2 2 2 1 1 ...
##  $ Marital_Status        : Factor w/ 3 levels "Married","Separated",..: 1 3 3 3 3 3 3 2 1 1
...
##  $ Gender                : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 2 2 2 2 1 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

Rpart algorithm for decision tree

```
##
## Classification tree:
## rpart(formula = CustomerClass ~ ., data = DecisionTreeData, method = "class",
##     control = rpartCtrl)
##
## Variables actually used in tree construction:
## [1] CreditAmount      CreditHistory     DurationInMonths  ExistingAccStatus
## [5] Property          Purpose           SavingsAccStatus
##
## Root node error: 300/999 = 0.3003
##
## n= 999
##
##         CP nsplit rel error  xerror     xstd
## 1 0.051667      0   1.00000 1.00000 0.048294
## 2 0.046667      3   0.84000 0.97667 0.047966
## 3 0.018333      4   0.79333 0.86000 0.046112
## 4 0.016667      6   0.75667 0.86333 0.046170
## 5 0.015556      8   0.72333 0.85333 0.045995
## 6 0.010000     11   0.67667 0.86333 0.046170
```



Visualizing tree model



Tree pruning