

# Scorecard Preparation For CredX Bank customers

Sreekanth Settur

02/01/2019

## Reading data

```
## Application.ID           Age           Gender
## Min.      :1.004e+05   Min.      :-3.00   F      :16837
## 1st Qu.:2.484e+08   1st Qu.:37.00   M      :54456
## Median :4.976e+08   Median :45.00   NA's:    2
## Mean      :4.990e+08   Mean      :44.94
## 3rd Qu.:7.496e+08   3rd Qu.:53.00
## Max.      :1.000e+09   Max.      :65.00
##
## Marital.Status..at.the.time.of.application. No.of.dependents
## Married:60730                               Min.      :1.000
## Single :10559                               1st Qu.:2.000
## NA's    :    6                               Median :3.000
##                                              Mean      :2.865
##                                              3rd Qu.:4.000
##                                              Max.      :5.000
##                                              NA's      :3
##      Income           Education           Profession
## Min.      :-0.5   Bachelor      :17697   SAL      :40439
## 1st Qu.:14.0   Masters      :23970   SE      :14307
## Median :27.0   Others      :   121   SE_PROF:16535
## Mean      :27.2   Phd      : 4549   NA's     :   14
## 3rd Qu.:40.0   Professional:24839
## Max.      :60.0   NA's      :   119
##
##      Type.of.residence No.of.months.in.current.residence
## Company provided      : 1630   Min.      :   6.00
## Living with Parents: 1818   1st Qu.:   6.00
## Others                :   199   Median : 11.00
## Owned                 :14243   Mean      : 34.56
## Rented                :53397   3rd Qu.: 60.00
## NA's                  :    8   Max.      :126.00
##
## No.of.months.in.current.company Performance.Tag
## Min.      :   3.00   Min.      :0.0000
## 1st Qu.: 16.00   1st Qu.:0.0000
## Median : 34.00   Median :0.0000
## Mean      : 33.96   Mean      :0.0422
## 3rd Qu.: 51.00   3rd Qu.:0.0000
## Max.      :133.00   Max.      :1.0000
##                      NA's      :1425
```

# Finding percentage of missing values in performance tag attribute of demographic data

```
Predictive_missing <- length(which(is.na(Demographic_data$Performance.Tag) == TRUE))
missingPercent <- (Predictive_missing/nrow(Demographic_data))*100
round(missingPercent,2)
```

```
## [1] 2
```

# Finding the number of missing values in whole dataset

```
colSums(is.na(Demographic_data) == TRUE)
```

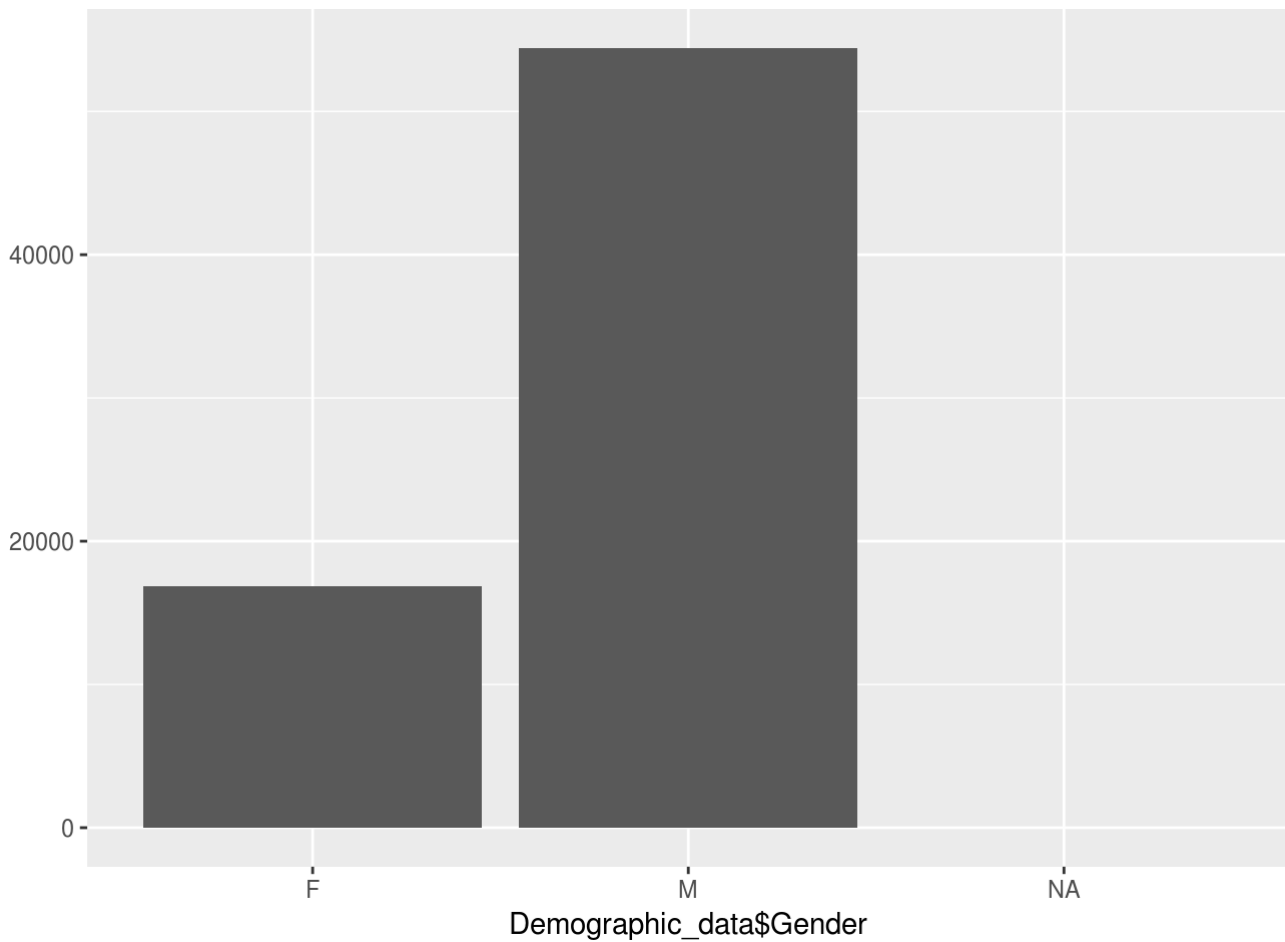
```
##               Application.ID
##                      0
##                   Age
##                      0
##                  Gender
##                      2
## Marital.Status..at.the.time.of.application.
##                      6
##             No.of.dependents
##                      3
##                   Income
##                      0
##                   Education
##                   119
##                   Profession
##                   14
##             Type.of.residence
##                      8
## No.of.months.in.current.residence
##                      0
## No.of.months.in.current.company
##                      0
##                   Performance.Tag
##                   1425
```

## Datacleaning and factoring attributes

## Plotting data univariate analysis

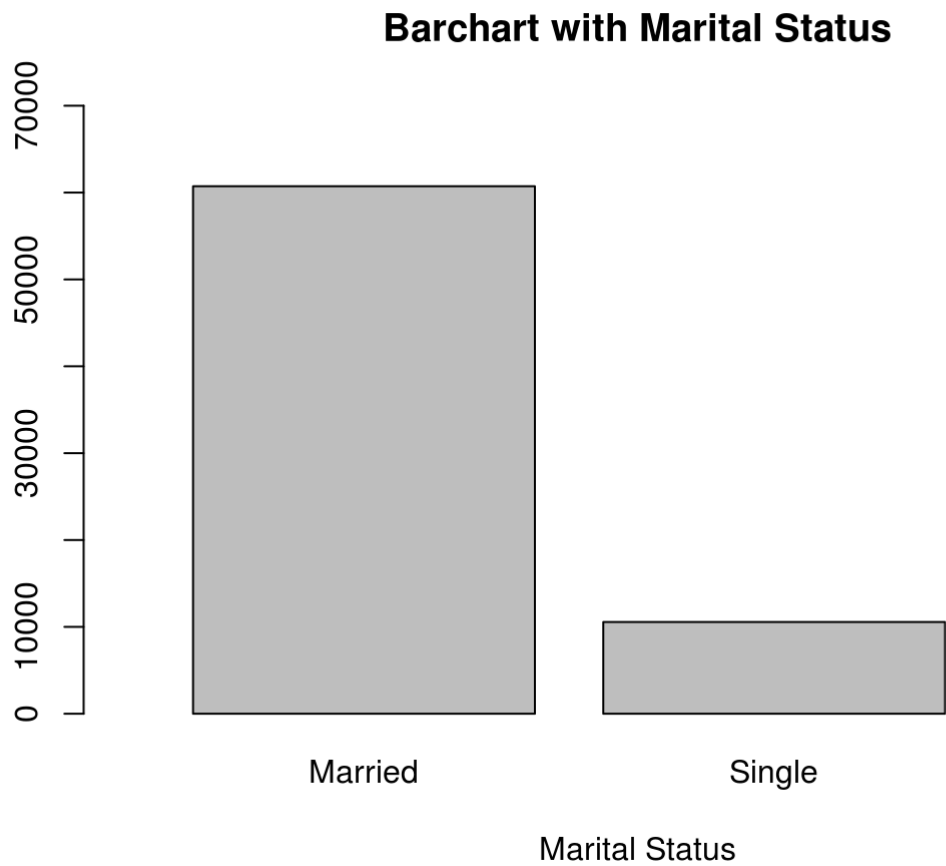
## Univariate on gender

```
library(ggplot2)
qplot(Demographic_data$Gender)
```

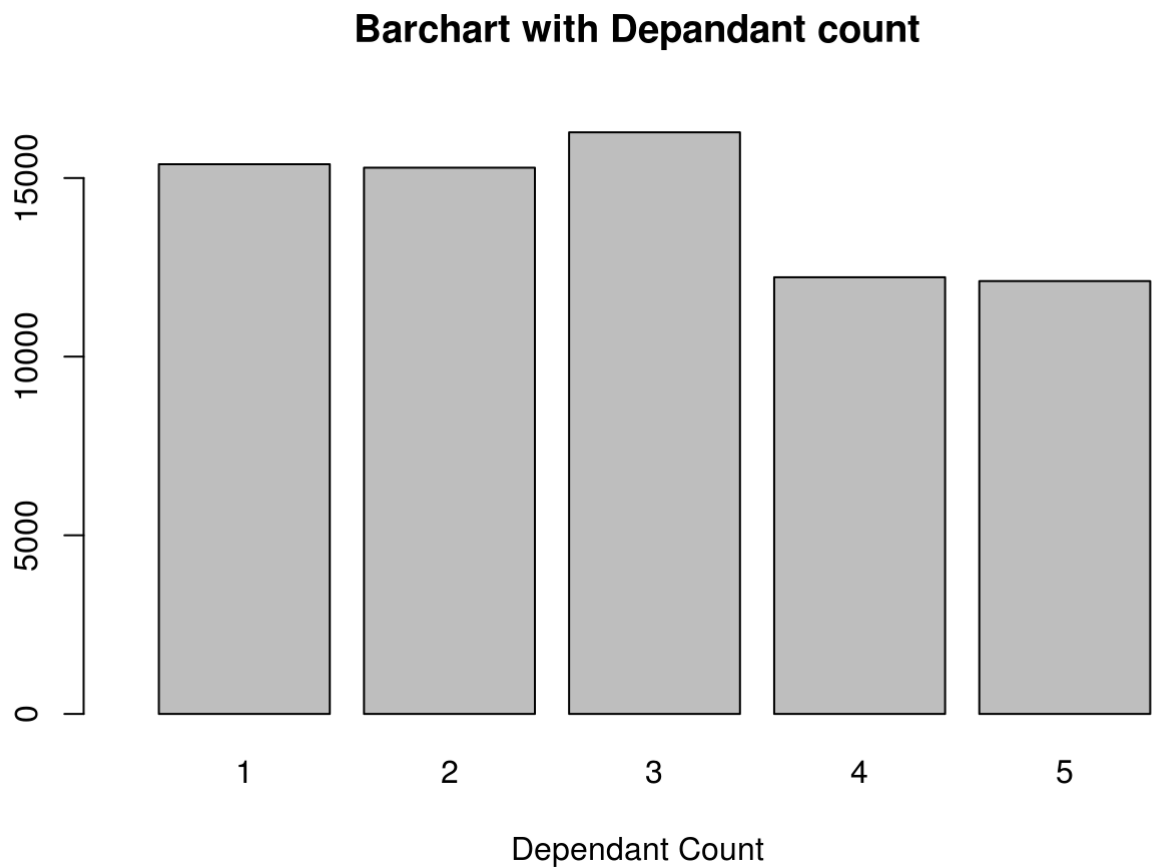


```
#plot(Demographic_data$Gender, xlab= "Gender",  
#     main = "Barchart with Gender details")
```

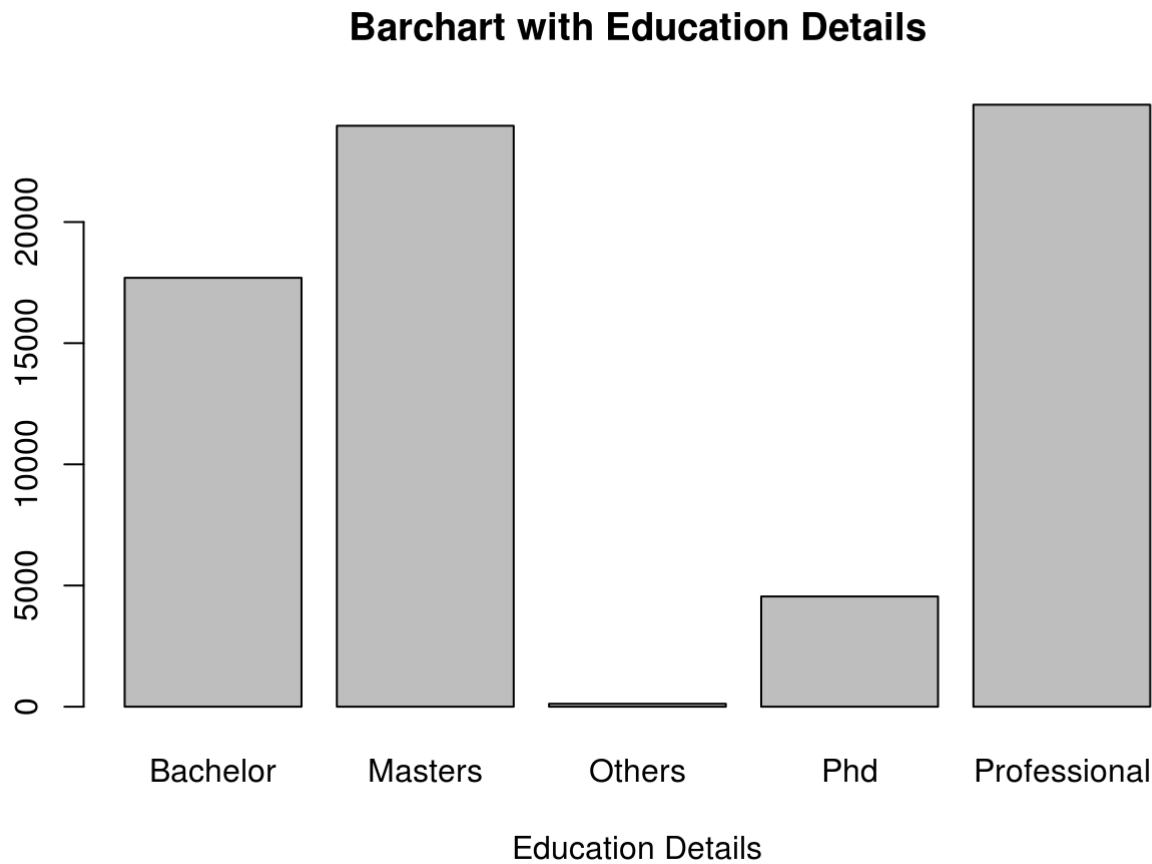
```
plot(as.factor(Demographic_data$Marital_Status), xlab = "Marital Status",  
     main = "Barchart with Marital Status", xlim = c(0, 3), ylim = c(0, 70000))
```



```
plot(Demographic_data$Dependents, xlab = "Dependant Count", main = "Barchart with Dep  
andant count", xlim = c(0,6), ylim = c(0, 17000))
```

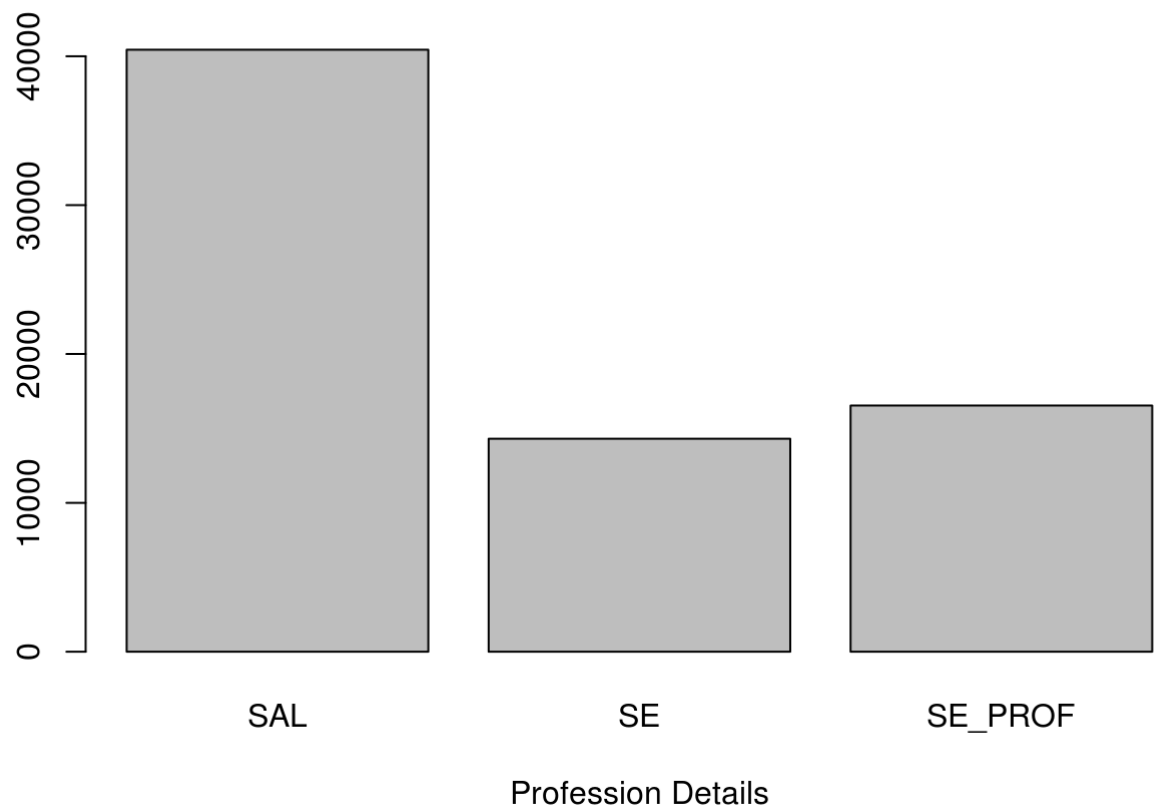


```
plot(as.factor(Demographic_data$Education), xlab = "Education Details", main = "Barchart with Education Details")
```



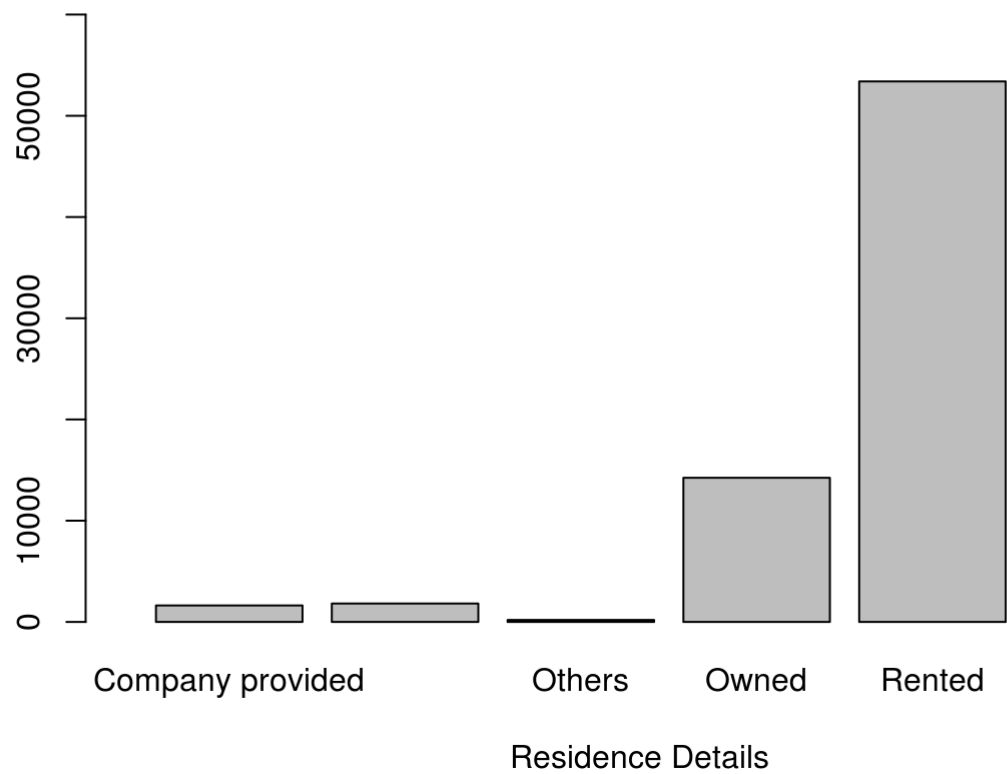
```
plot(as.factor(Demographic_data$Profession), xlab = "Profession Details", main = "Bar chart with Profession Details")
```

Barchart with Profession Details

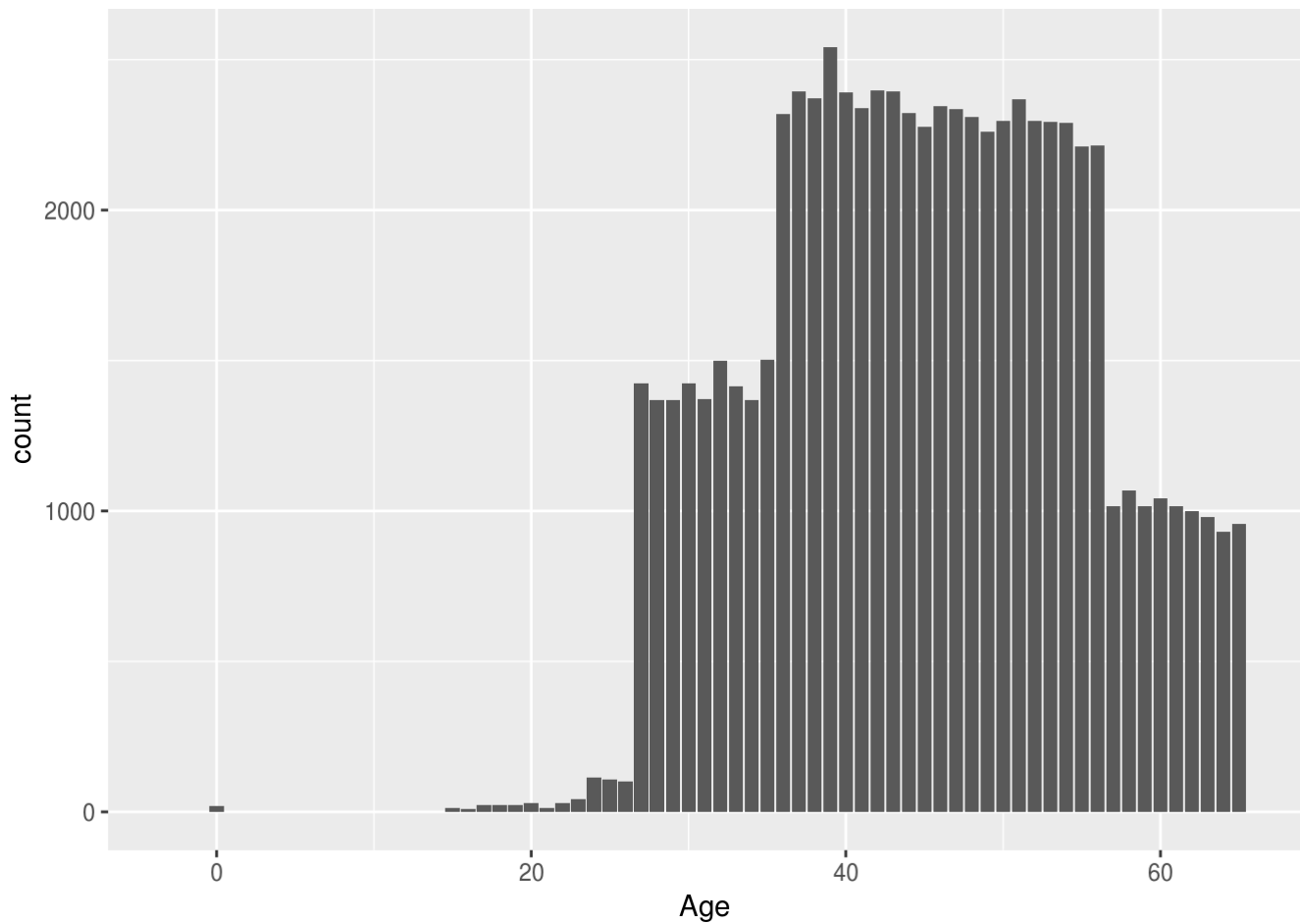


```
plot(as.factor(Demographic_data$Residence), xlab = "Residence Details", main = "Barchart with Residence Details", xlim = c(0,7), ylim = c(0, 60000))
```

Barchart with Residence Details

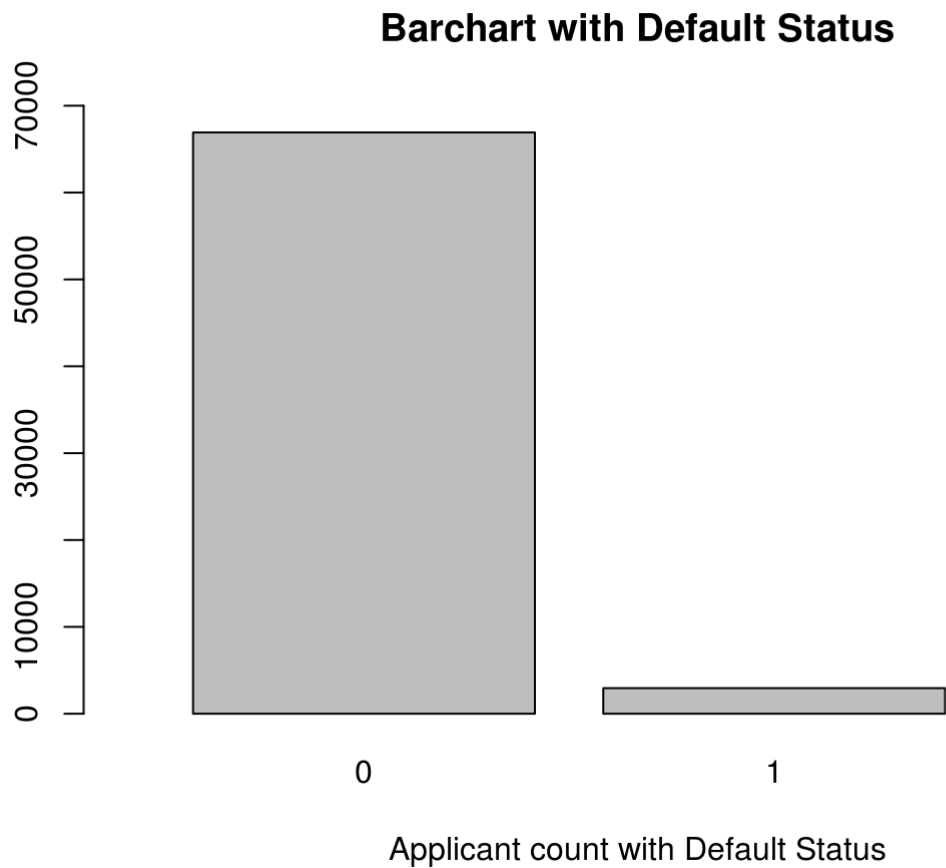


```
ggplot2::ggplot(data = Demographic_data, aes(x=Age)) + stat_count()
```

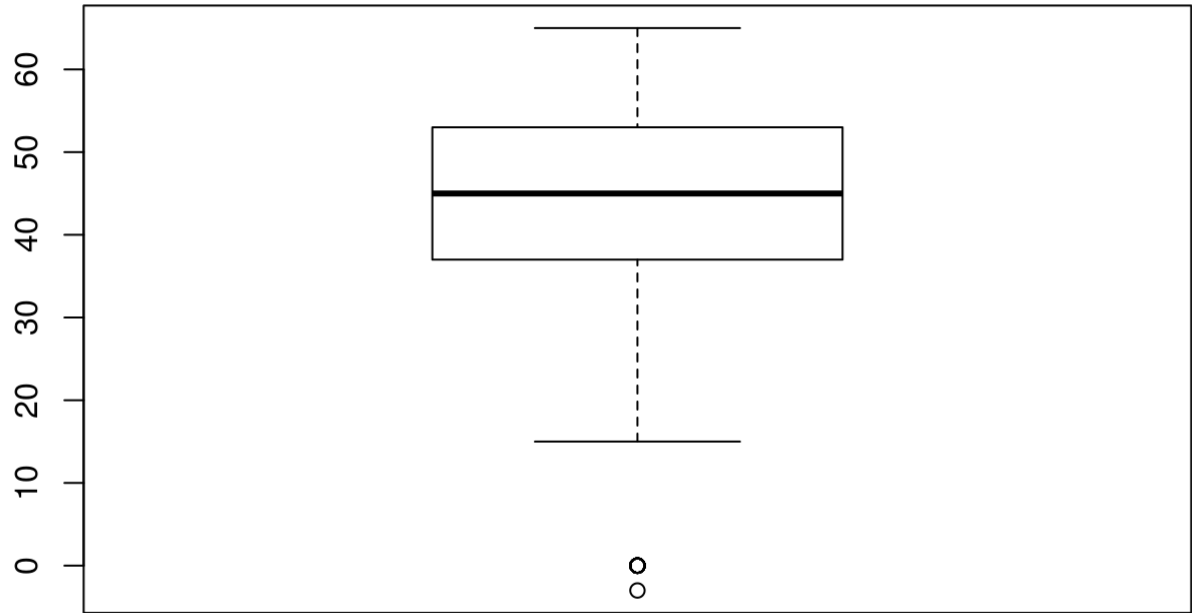


```
#plot(Demographic_data$Age, type = "l", xlab = "Residence Details",  
#      main = "Barchart with Residence Details")
```

```
plot(as.factor(Demographic_data$PerformanceTag), xlab = "Applicant count with Default  
Status",  
      main = "Barchart with Default Status", xlim = c(0,3), ylim = c(0, 70000))
```

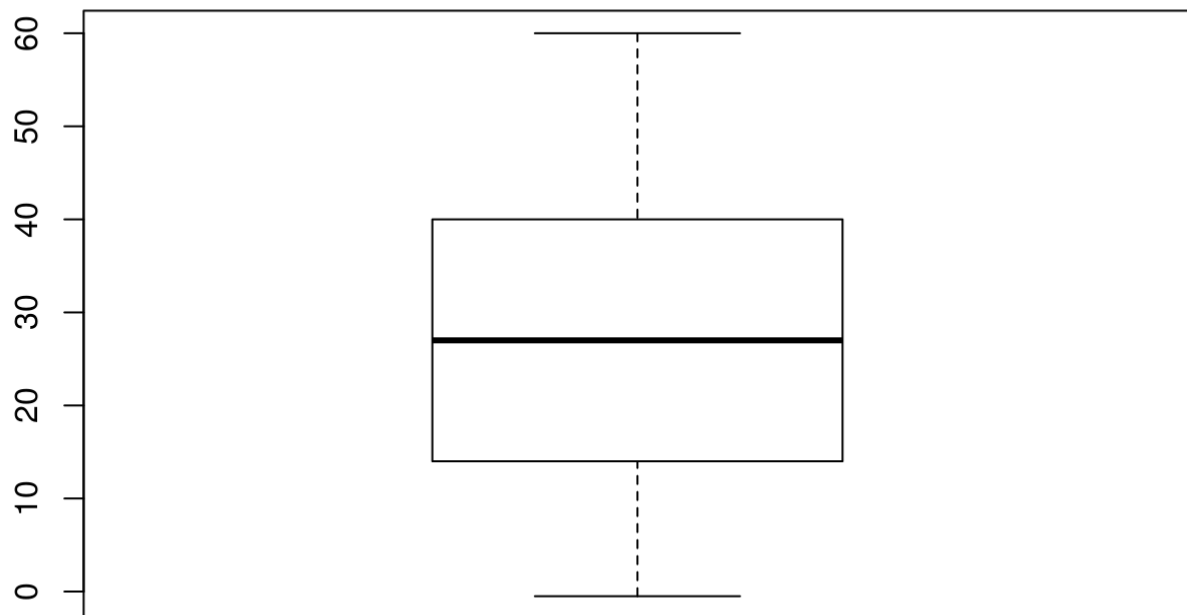


```
boxplot(Demographic_data$Age)
```





```
boxplot(Demographic_data$Income)
```



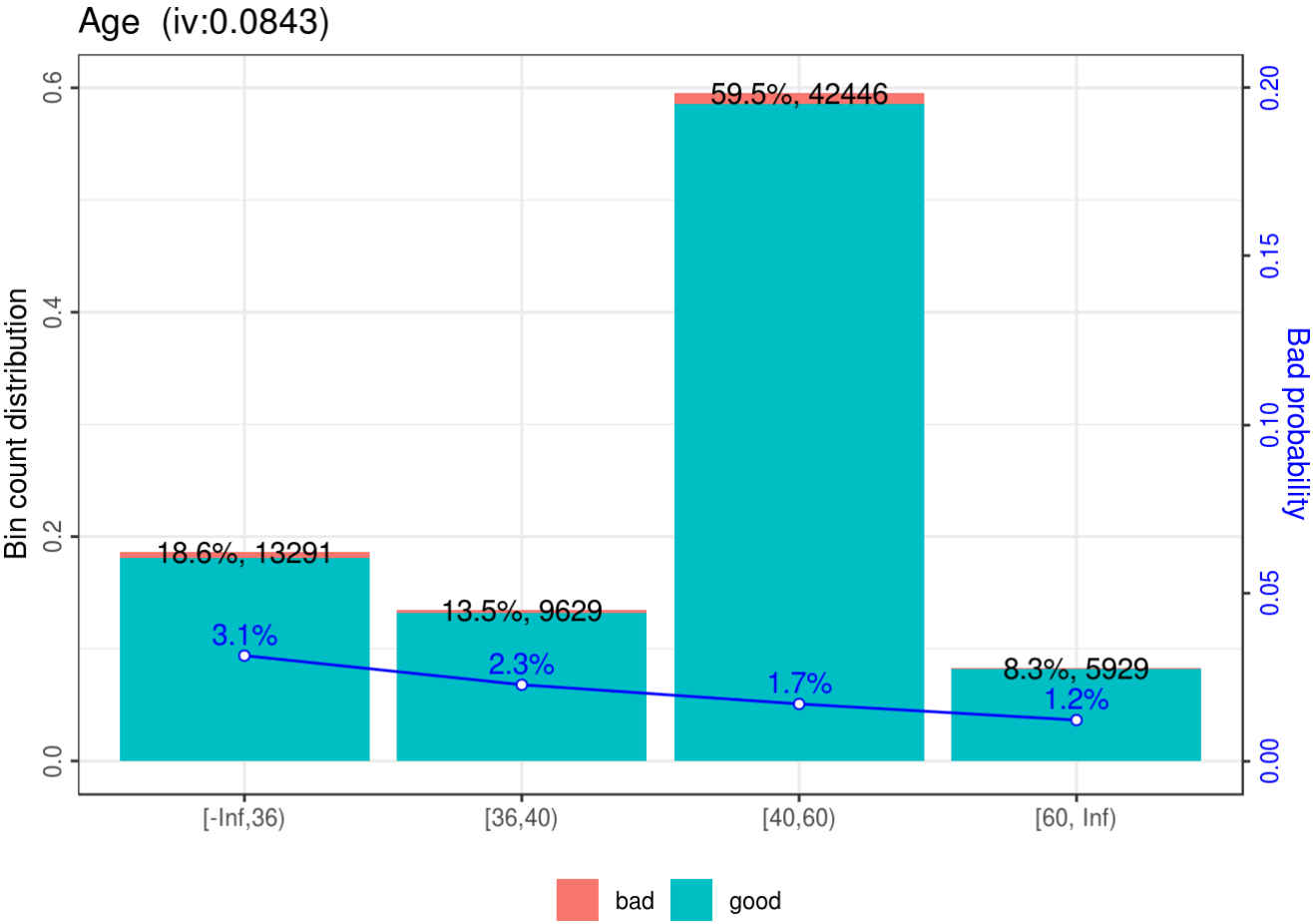
## Building Woebins

```
library(scorecard)
Demographic_data_woebins = woebin(dt = Demographic_data[, -c(12,13)],
                                   y = "Status_flag",
                                   special_values=c("NA"))
```

```
## Warning in check_special_values(special_values, xs): The special_values
## should be a list. Make sure special values are exactly the same in all
## variables if special_values is a vector.
```

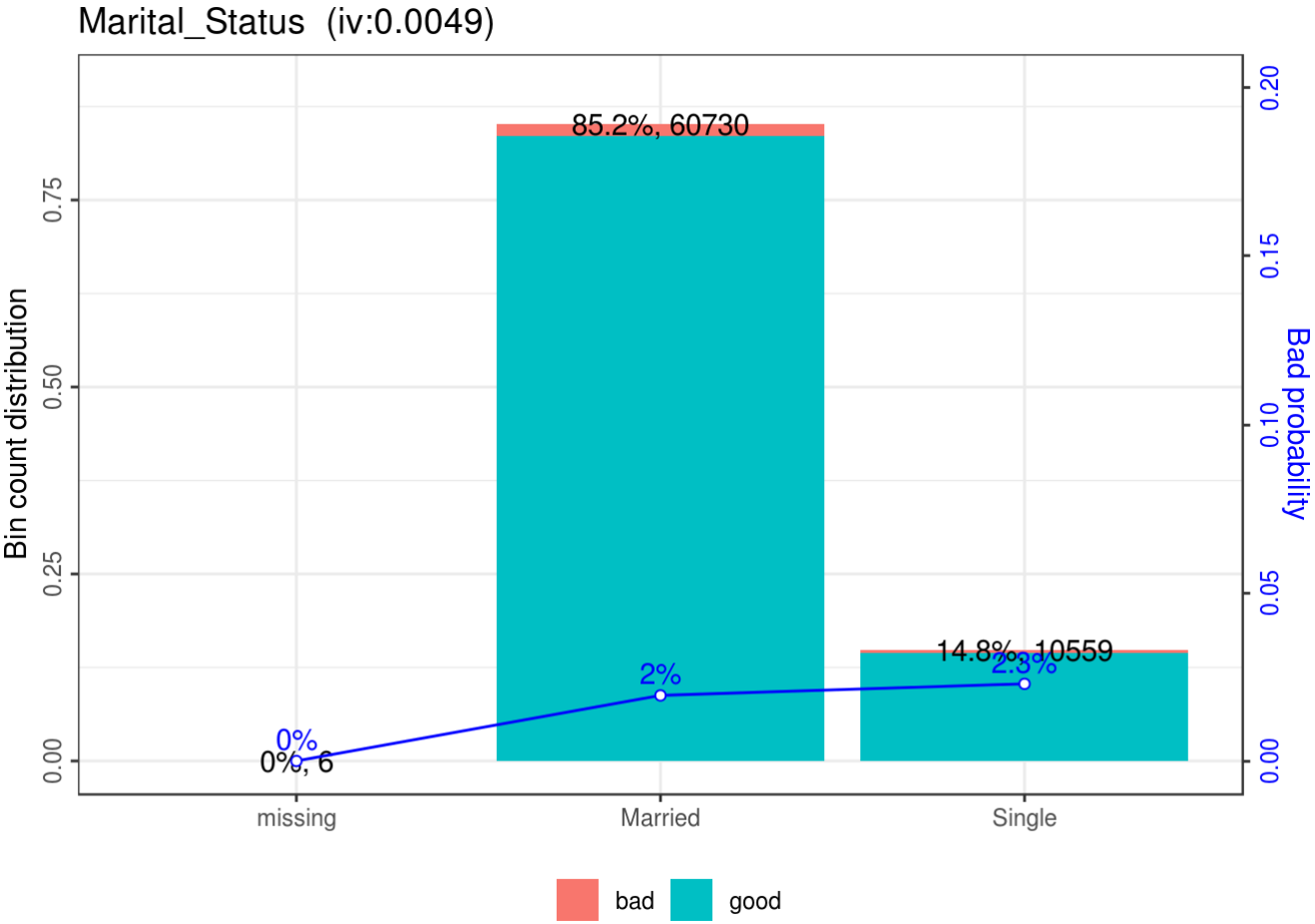
```
library(scorecard)
woebin_plot(Demographic_data_woebins$Age)
```

```
## $Age
```



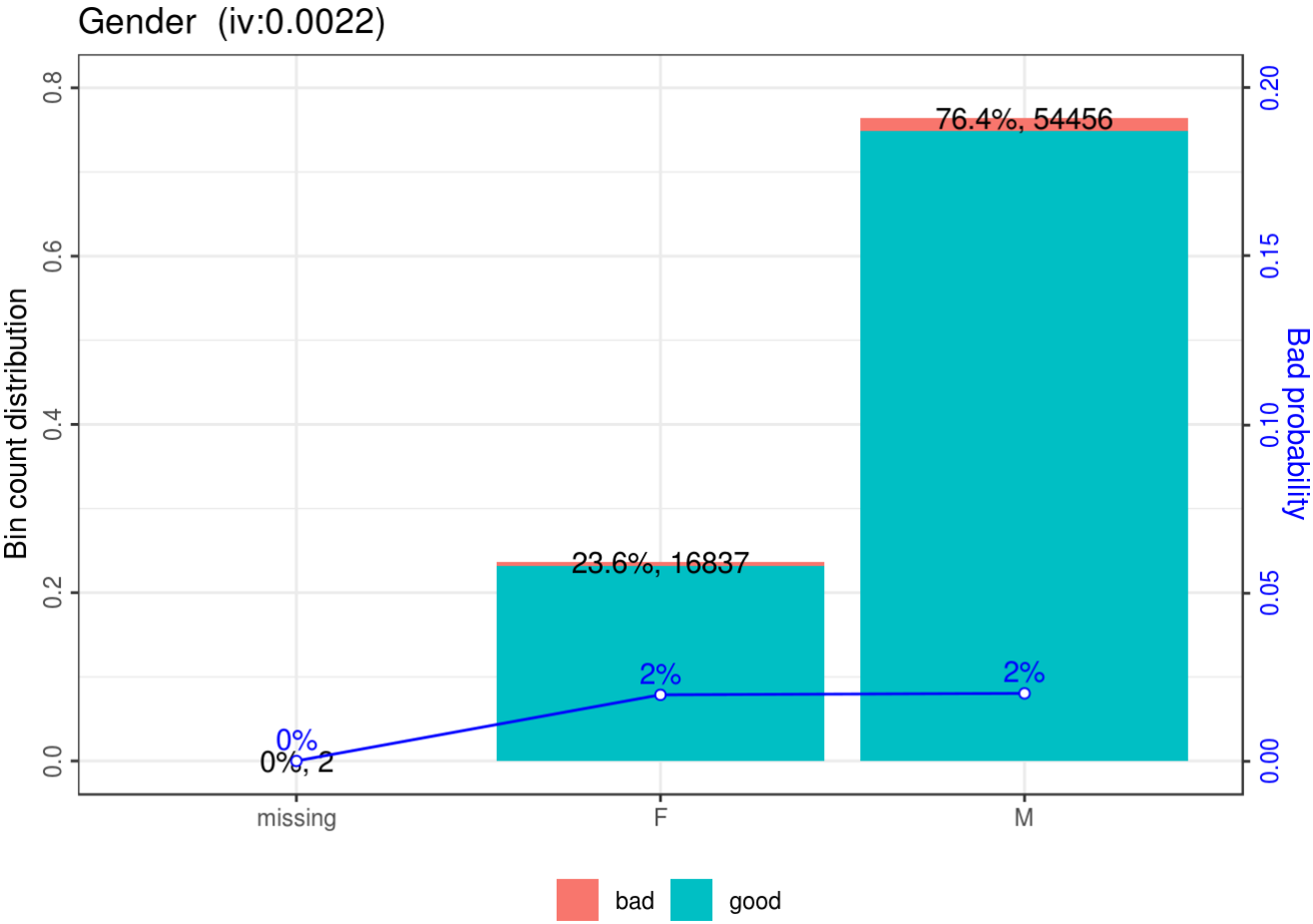
```
woebin_plot(Demographic_data_woebins$Marital_Status)
```

```
## $Marital_Status
```



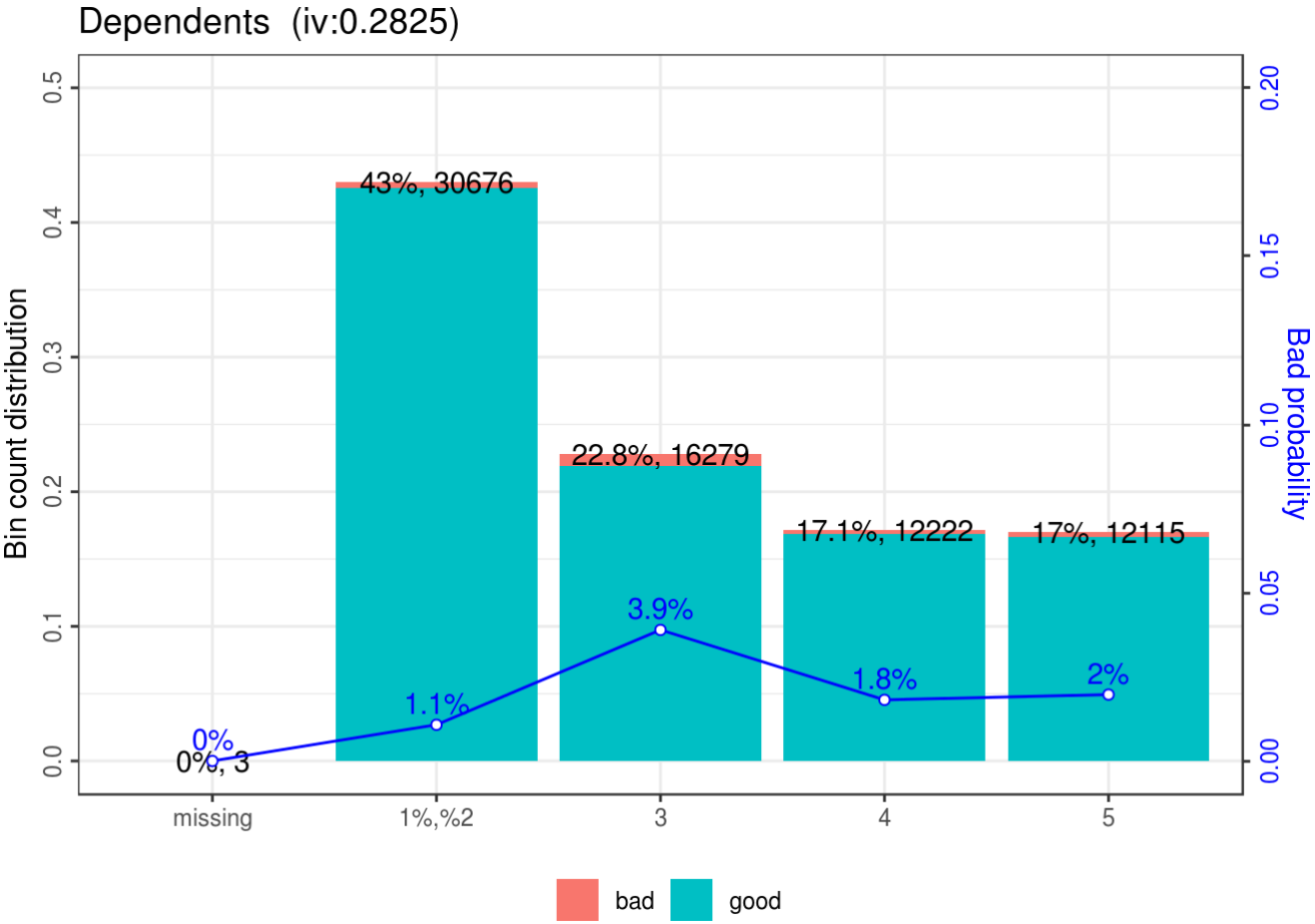
```
woebin_plot(Demographic_data_woebins$Gender)
```

```
## $Gender
```



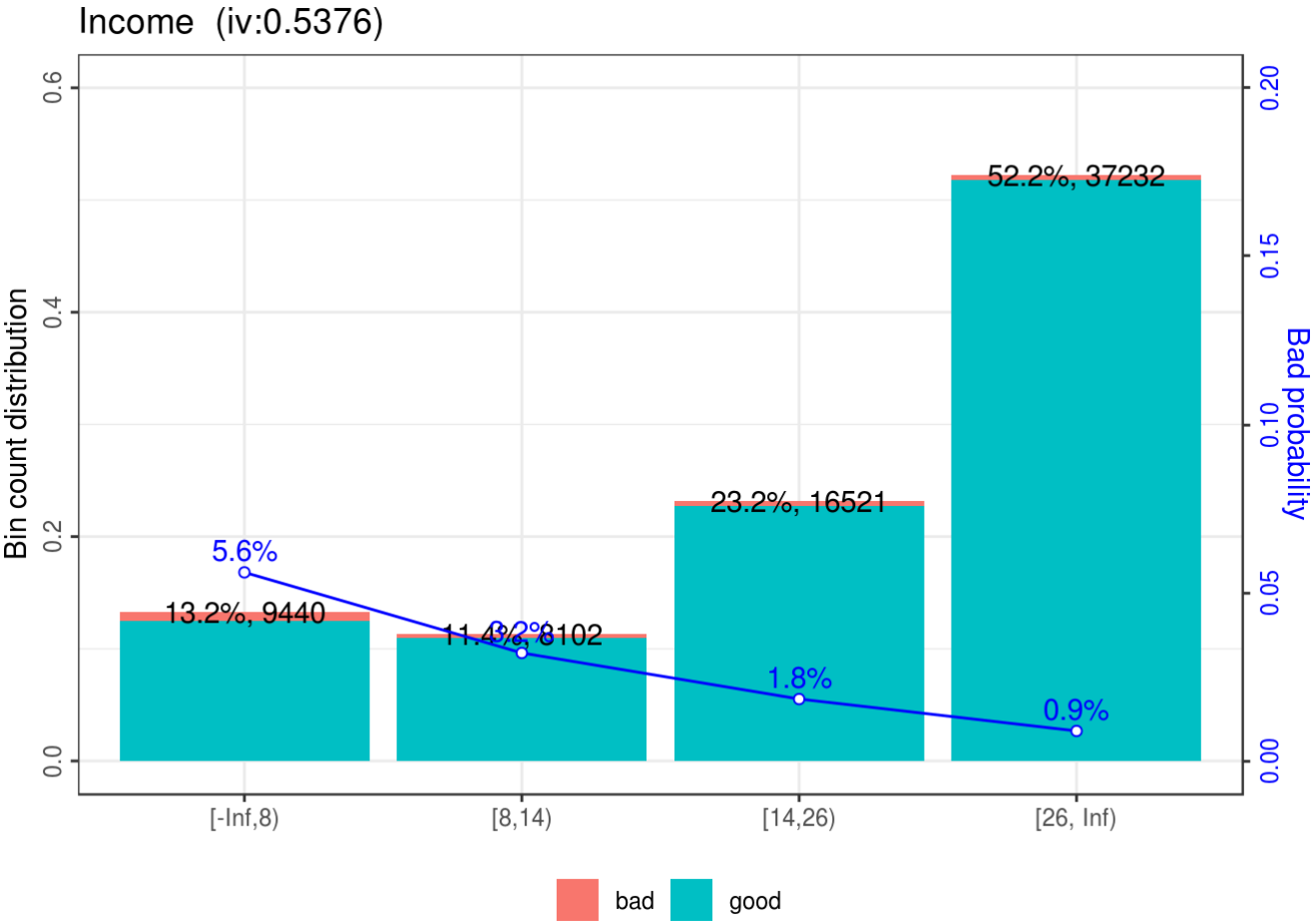
```
woebin_plot(Demographic_data_woebins$Dependents)
```

```
## $Dependents
```



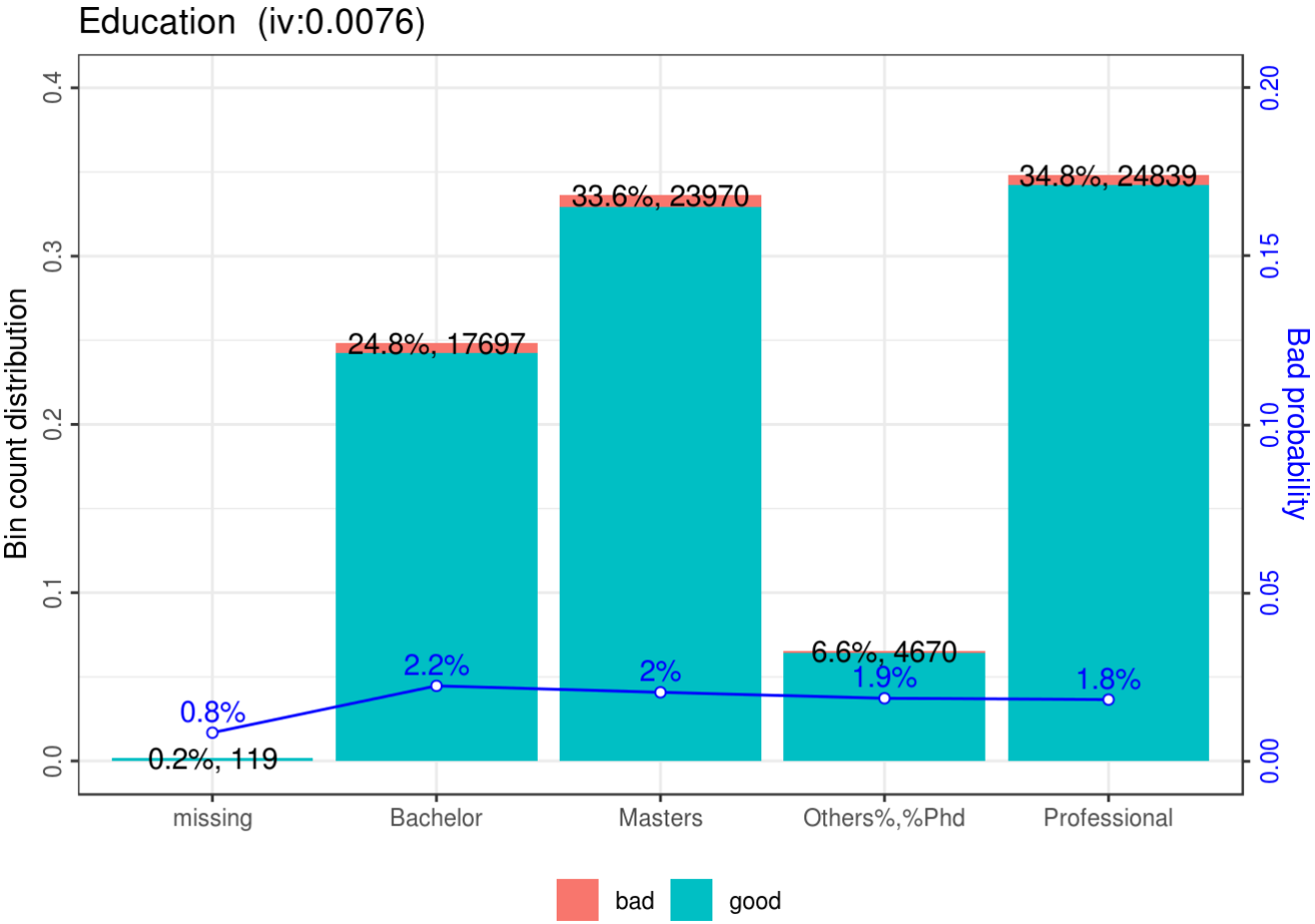
```
woebin_plot(Demographic_data_woebins$Income)
```

```
## $Income
```



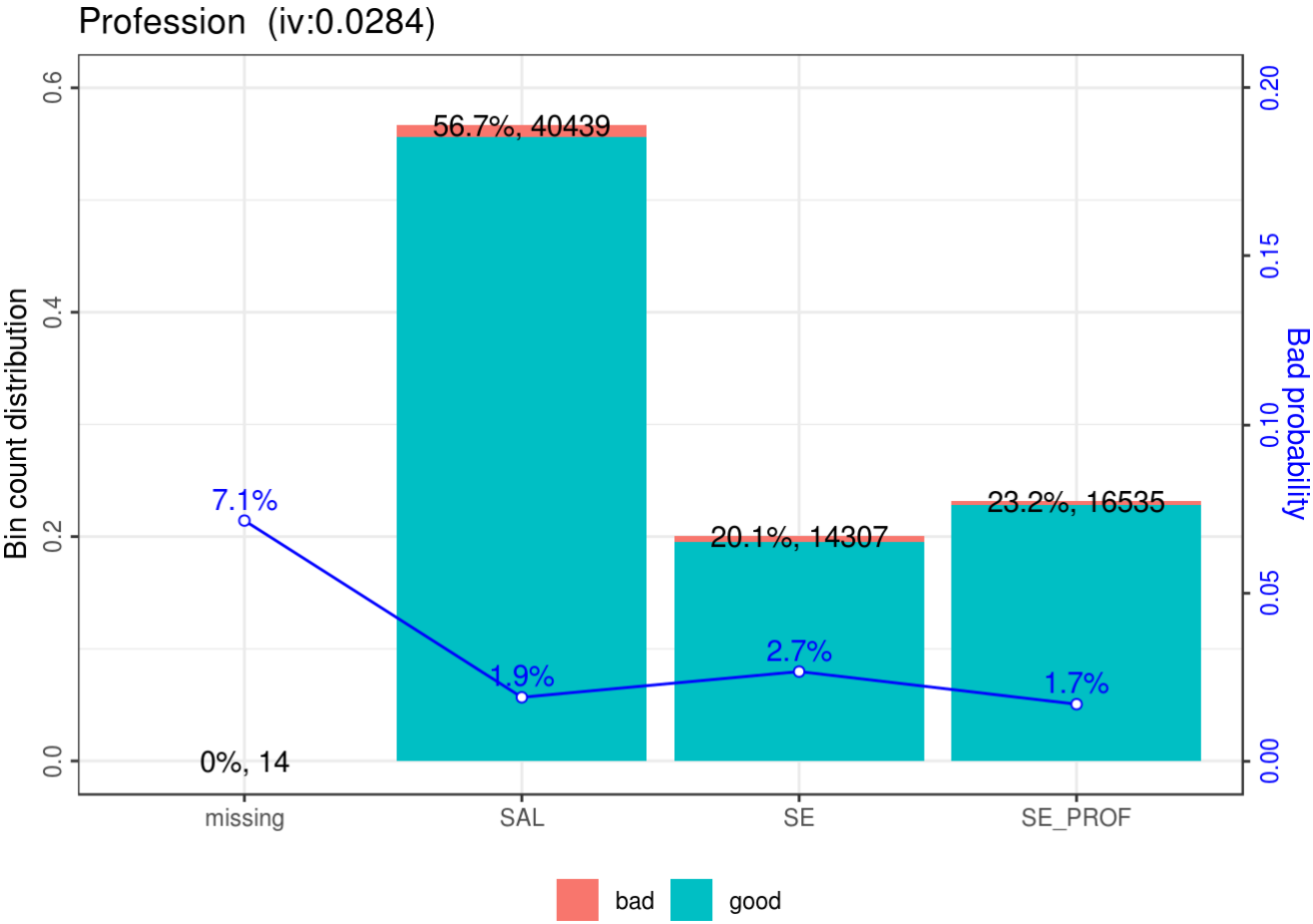
```
woebin_plot(Demographic_data_woebins$Education)
```

```
## $Education
```



```
woebin_plot(Demographic_data_woebins$Profession)
```

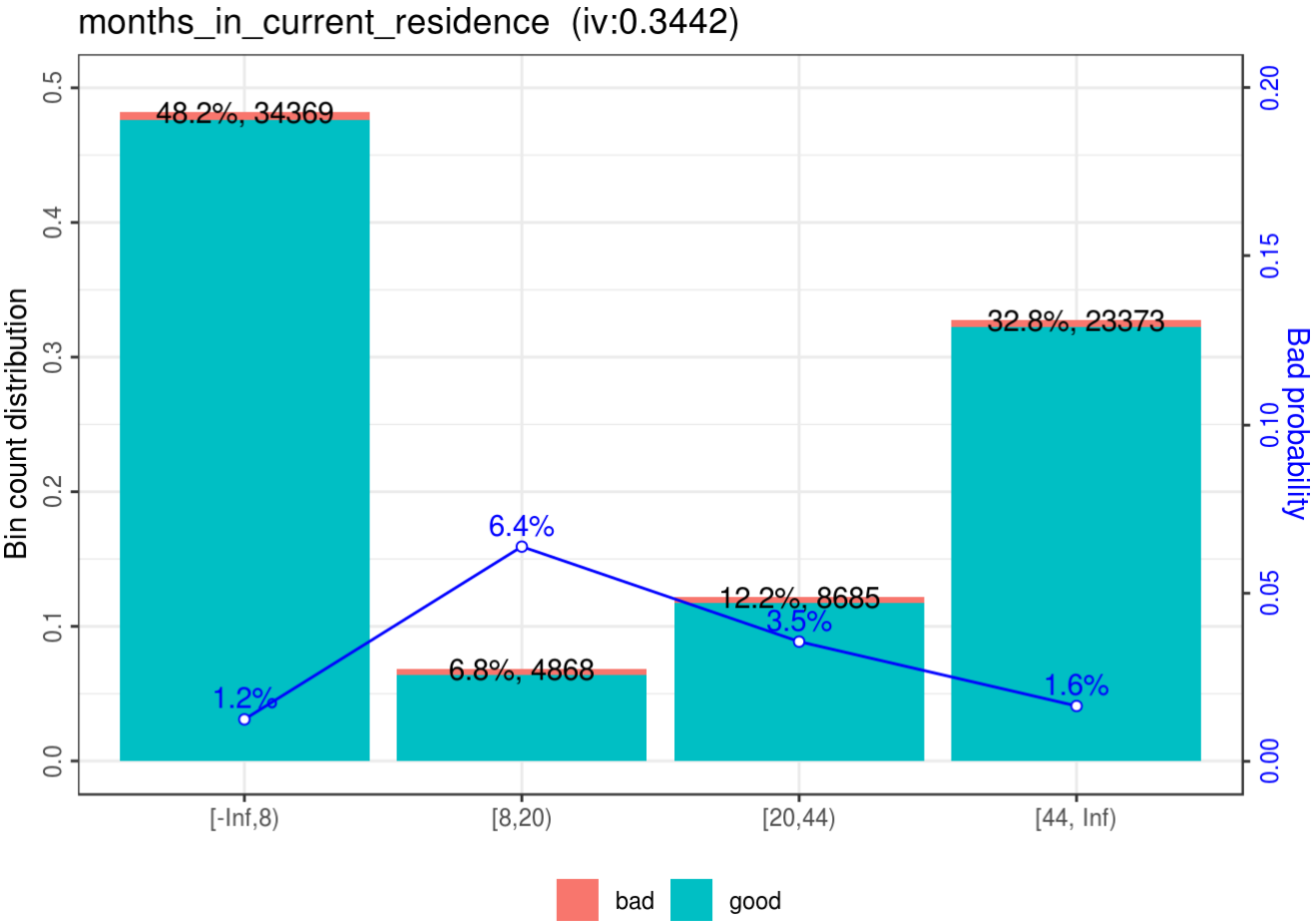
```
## $Profession
```



```
woebin_plot(Demographic_data_woebins$months_in_current_residence)
```

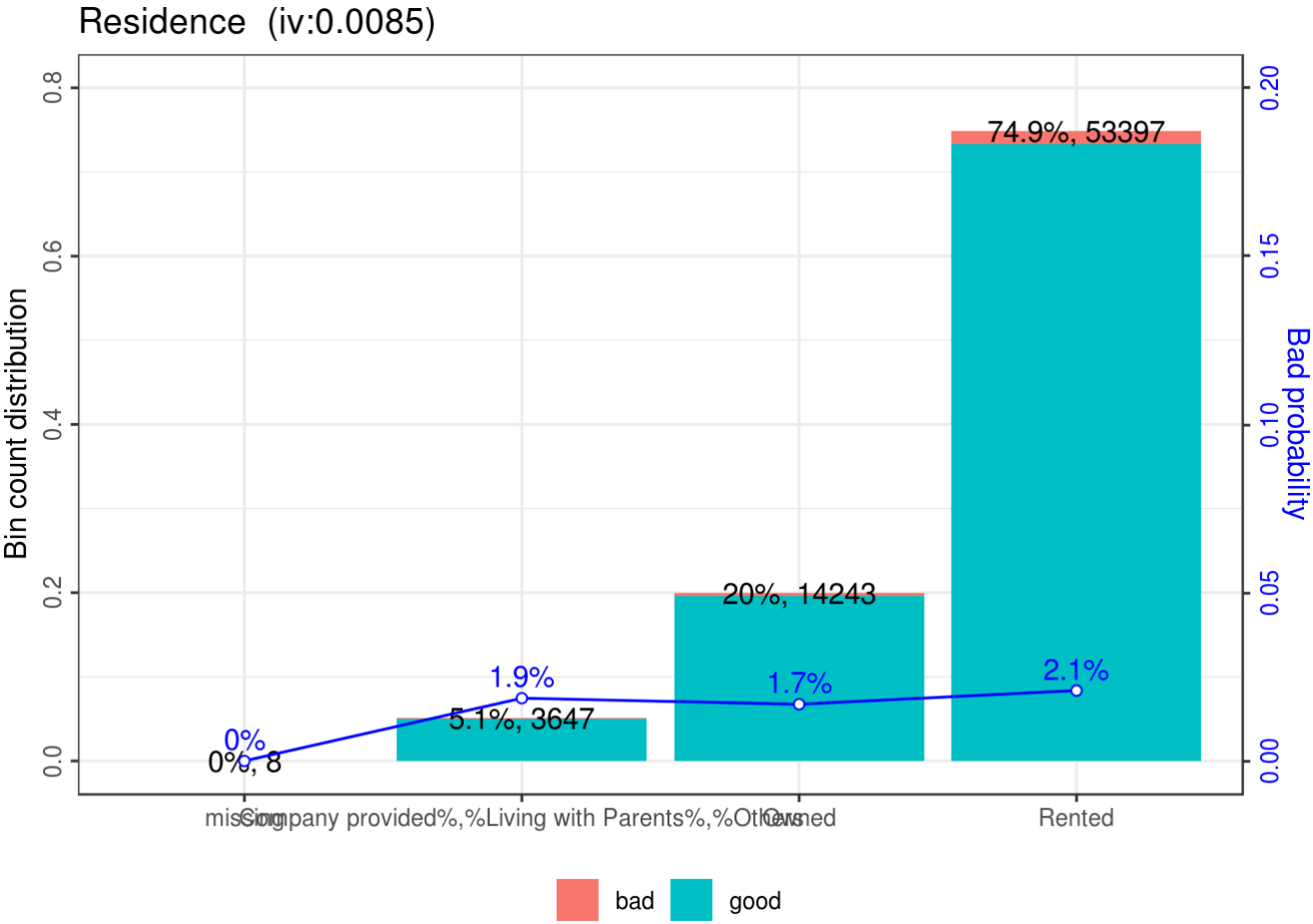
```
## $months_in_current_residence
```





```
woebin_plot(Demographic_data_woebins$Residence)
```

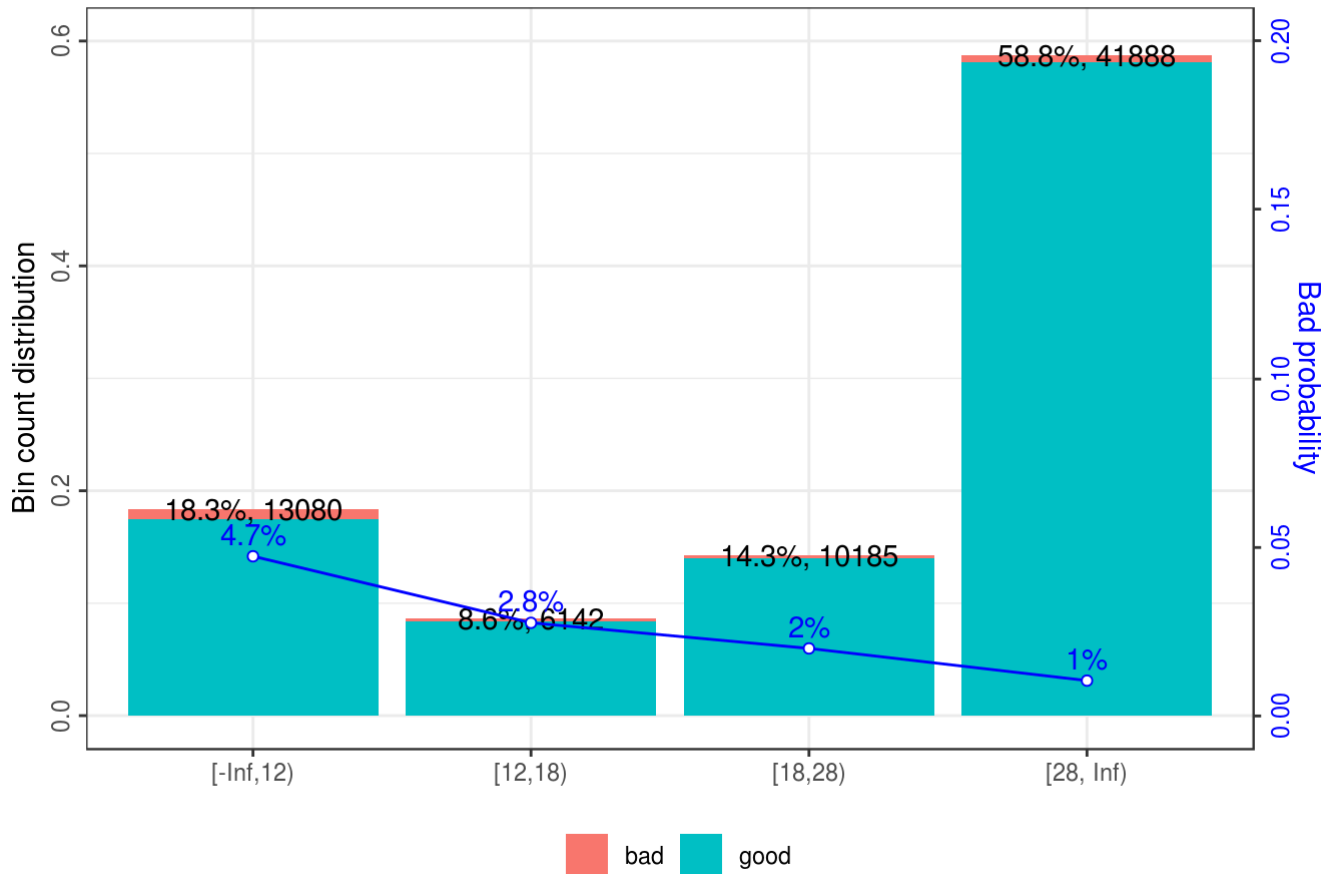
```
## $Residence
```



```
woebin_plot(Demographic_data_woebins$months_in_current_company)
```

```
## $months_in_current_company
```

## months\_in\_current\_company (iv:0.4291)



```
Demographic_Woedataset = woebin_ply(Demographic_data[, -c(12,13)],
                                     bins = Demographic_data_woebins)
```

```
## Woe transforming on 71295 rows and 11 columns in 0: 0:12
```

```
Demographic_Woedataset$Status_flag <- Demographic_data$Status_flag
Demographic_Woedataset$Application.ID_woe <- Demographic_data$Application.ID
summary(Demographic_Woedataset)
```

```
## Status_flag      Application.ID_woe      Age_woe
## Length:71295     Min.      :1.004e+05      Min.      : -0.50626
## Class :character  1st Qu.:2.484e+08      1st Qu.: -0.16858
## Mode  :character  Median :4.976e+08      Median : -0.16858
##                               Mean  :4.990e+08      Mean  : -0.03949
##                               3rd Qu.:7.496e+08      3rd Qu.: 0.12733
##                               Max.  :1.000e+09      Max.  : 0.46011
##   Gender_woe      Marital_Status_woe      Dependents_woe
## Min.      : -0.017591      Min.      : -0.026942      Min.      : -0.62956
## 1st Qu.: 0.004512      1st Qu.: -0.026942      1st Qu.: -0.62956
## Median : 0.004512      Median : -0.026942      Median : -0.09821
## Mean  : -0.000619      Mean  : -0.002164      Mean  : -0.13337
## 3rd Qu.: 0.004512      3rd Qu.: -0.026942      3rd Qu.: -0.01404
## Max.  : 3.188573      Max.  : 2.089960      Max.  : 2.78311
##   Income_woe      Education_woe      Profession_woe
## Min.      : -0.81837      Min.      : -0.878220      Min.      : -0.17254
## 1st Qu.: -0.81837      1st Qu.: -0.093408      1st Qu.: -0.05611
## Median : -0.81837      Median : 0.020880      Median : -0.05611
## Mean  : -0.25023      Mean  : -0.003695      Mean  : -0.01318
## 3rd Qu.: -0.08432      3rd Qu.: 0.020880      3rd Qu.: -0.05611
## Max.  : 1.06841      Max.  : 0.112773      Max.  : 1.32752
## Residence_woe      months_in_current_residence_woe
## Min.      : -0.174618      Min.      : -0.4879
## 1st Qu.: -0.071561      1st Qu.: -0.4879
## Median : 0.045803      Median : -0.2050
## Mean  : -0.004038      Mean  : -0.1484
## 3rd Qu.: 0.045803      3rd Qu.: -0.2050
## Max.  : 1.802278      Max.  : 1.2044
## months_in_current_company_woe
## Min.      : -0.6645
## 1st Qu.: -0.6645
## Median : -0.6645
## Mean  : -0.1996
## 3rd Qu.: 0.3274
## Max.  : 0.8885
```

```
names(Demographic_Woedataset)[2] <- paste("Application.ID")
names(Demographic_Woedataset)[3] <- paste("Age")
names(Demographic_Woedataset)[4] <- paste("Gender")
names(Demographic_Woedataset)[5] <- paste("Marital_Status")
names(Demographic_Woedataset)[6] <- paste("Dependents")
names(Demographic_Woedataset)[7] <- paste("Income")
names(Demographic_Woedataset)[8] <- paste("Education")
names(Demographic_Woedataset)[9] <- paste("Profession")
names(Demographic_Woedataset)[10] <- paste("Residence")
names(Demographic_Woedataset)[11] <- paste("months_in_current_residence")
names(Demographic_Woedataset)[12] <- paste("months_in_current_company")
```

## Reading CreditBureau Data

```
CreditHistory = read.csv("Credit Bureau data.csv", na.strings = c(" ", "", NA))
summary(CreditHistory)
```

```

## Application.ID      No.of.times.90.DPD.or.worse.in.last.6.months
## Min.      :1.004e+05  Min.      :0.0000
## 1st Qu.:2.484e+08  1st Qu.:0.0000
## Median :4.976e+08  Median :0.0000
## Mean    :4.990e+08  Mean    :0.2703
## 3rd Qu.:7.496e+08  3rd Qu.:0.0000
## Max.    :1.000e+09  Max.    :3.0000
##
## No.of.times.60.DPD.or.worse.in.last.6.months
## Min.      :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean    :0.4305
## 3rd Qu.:1.0000
## Max.    :5.0000
##
## No.of.times.30.DPD.or.worse.in.last.6.months
## Min.      :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean    :0.5772
## 3rd Qu.:1.0000
## Max.    :7.0000
##
## No.of.times.90.DPD.or.worse.in.last.12.months
## Min.      :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean    :0.4503
## 3rd Qu.:1.0000
## Max.    :5.0000
##
## No.of.times.60.DPD.or.worse.in.last.12.months
## Min.      :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean    :0.6555
## 3rd Qu.:1.0000
## Max.    :7.0000
##
## No.of.times.30.DPD.or.worse.in.last.12.months
## Min.      :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean    :0.8009
## 3rd Qu.:1.0000
## Max.    :9.0000
##
## Avgas.CC.Utilization.in.last.12.months
## Min.      : 0.0
## 1st Qu.: 8.0
## Median : 15.0
## Mean    : 29.7
## 3rd Qu.: 46.0
## Max.    :113.0
## NA's     :1058
## No.of.trades.opened.in.last.6.months

```

```

## Min.      : 0.000
## 1st Qu.: 1.000
## Median : 2.000
## Mean    : 2.298
## 3rd Qu.: 3.000
## Max.    :12.000
## NA's    :1
## No.of.trades.opened.in.last.12.months
## Min.      : 0.000
## 1st Qu.: 2.000
## Median : 5.000
## Mean    : 5.827
## 3rd Qu.: 9.000
## Max.    :28.000
##
## No.of.PL.trades.opened.in.last.6.months
## Min.      :0.000
## 1st Qu.:0.000
## Median :1.000
## Mean    :1.207
## 3rd Qu.:2.000
## Max.    :6.000
##
## No.of.PL.trades.opened.in.last.12.months
## Min.      : 0.000
## 1st Qu.: 0.000
## Median : 2.000
## Mean    : 2.397
## 3rd Qu.: 4.000
## Max.    :12.000
##
## No.of.Inquiries.in.last.6.months..excluding.home...auto.loans.
## Min.      : 0.000
## 1st Qu.: 0.000
## Median : 1.000
## Mean    : 1.764
## 3rd Qu.: 3.000
## Max.    :10.000
##
## No.of.Inquiries.in.last.12.months..excluding.home...auto.loans.
## Min.      : 0.000
## 1st Qu.: 0.000
## Median : 3.000
## Mean    : 3.535
## 3rd Qu.: 5.000
## Max.    :20.000
##
## Presence.of.open.home.loan Outstanding.Balance Total.No.of.Trades
## Min.      :0.00000 Min.      : 0 Min.      : 0.000
## 1st Qu.:0.00000 1st Qu.: 211532 1st Qu.: 3.000
## Median :0.00000 Median : 774992 Median : 6.000
## Mean    :0.2564 Mean    :1249163 Mean    : 8.187
## 3rd Qu.:1.0000 3rd Qu.:2920796 3rd Qu.:10.000
## Max.    :1.0000 Max.    :5218801 Max.    :44.000
## NA's    :272 NA's    :272
## Presence.of.open.auto.loan Performance.Tag
## Min.      :0.00000 Min.      :0.0000
## 1st Qu.:0.00000 1st Qu.:0.0000

```

```
## Median :0.00000      Median :0.0000
## Mean   :0.08462      Mean    :0.0422
## 3rd Qu.:0.00000      3rd Qu.:0.0000
## Max.    :1.00000      Max.    :1.0000
##                                     NA's    :1425
```

## Counting NA values

```
colSums(is.na(CreditHistory) == TRUE)
```

```
##                                     Application.ID
##                                     0
##      No.of.times.90.DPD.or.worse.in.last.6.months
##                                     0
##      No.of.times.60.DPD.or.worse.in.last.6.months
##                                     0
##      No.of.times.30.DPD.or.worse.in.last.6.months
##                                     0
##      No.of.times.90.DPD.or.worse.in.last.12.months
##                                     0
##      No.of.times.60.DPD.or.worse.in.last.12.months
##                                     0
##      No.of.times.30.DPD.or.worse.in.last.12.months
##                                     0
##      Avgas.CC.Utilization.in.last.12.months
##                                     1058
##      No.of.trades.opened.in.last.6.months
##                                     1
##      No.of.trades.opened.in.last.12.months
##                                     0
##      No.of.PL.trades.opened.in.last.6.months
##                                     0
##      No.of.PL.trades.opened.in.last.12.months
##                                     0
##      No.of.Inquiries.in.last.6.months..excluding.home...auto.loans.
##                                     0
##      No.of.Inquiries.in.last.12.months..excluding.home...auto.loans.
##                                     0
##      Presence.of.open.home.loan
##                                     272
##      Outstanding.Balance
##                                     272
##      Total.No.of.Trades
##                                     0
##      Presence.of.open.auto.loan
##                                     0
##      Performance.Tag
##      1425
```

```

names(CreditHistory)[2] <- paste("DPD90InLast6Months")
names(CreditHistory)[3] <- paste("DPD60InLast6Months")
names(CreditHistory)[4] <- paste("DPD30InLast6Months")
names(CreditHistory)[5] <- paste("DPD90InLast12Months")
names(CreditHistory)[6] <- paste("DPD60InLast12Months")
names(CreditHistory)[7] <- paste("DPD30InLast12Months")
names(CreditHistory)[8] <- paste("CCUtilizationIn12Months")
names(CreditHistory)[9] <- paste("TradesInLast6Months")
names(CreditHistory)[10] <- paste("TradesInLast12Months")
names(CreditHistory)[11] <- paste("PLTradesInLast6Months")
names(CreditHistory)[12] <- paste("PLTradesInLast12Months")
names(CreditHistory)[13] <- paste("InquiriesInLast6Months")
names(CreditHistory)[14] <- paste("InquiriesInLast12Months")
names(CreditHistory)[15] <- paste("PresenceOfOpenHomeLoan")
names(CreditHistory)[16] <- paste("OutstandingBalance")
names(CreditHistory)[17] <- paste("TotalTrades")
names(CreditHistory)[18] <- paste("PresenceOfOpenLoan")
names(CreditHistory)[19] <- paste("PerformanceTag")

```

Considering all the NA values in performencetag as rejected entries.

So, out of given dataset we have 2% of rejected applicants.

Further on, only Status and Status\_flag will be used to identify the rejected class instead of performencetag attribute

```

CreditHistory_CopyData <- CreditHistory
CreditHistory$PerformanceTag[which(is.na(CreditHistory$PerformanceTag) == TRUE)] <- 2

CreditHistory$Status = ifelse(CreditHistory$PerformanceTag == 2, "Rejected", "Approved")
CreditHistory$Status_flag = ifelse(CreditHistory$PerformanceTag == 2, "1", "0")

```

Below code written using woe package in R

Generating woe bins for Demographic data, by excluding PerformanceTag and Status



This will be replaced by actual data, instead of WoE value.

This also treats the missing values

```
CreditHistory_woebins = woebin(CreditHistory[, -c(19,20)],  
                               y = "Status_flag",  
                               special_values=c("NA"))
```

```
## Warning in check_special_values(special_values, xs): The special_values  
## should be a list. Make sure special values are exactly the same in all  
## variables if special_values is a vector.
```

```
## Binning on 71295 rows and 19 columns in 0: 0:10
```

Replacing the actual data with its respective WoE value using already created WoE bins for Credit bureau data

```
CreditBureau_Woedataset = woebin_ply(CreditHistory[, -c(19,20)], bins = CreditHistory_woebins)
```

```
## Woe transforming on 71295 rows and 18 columns in 0: 0:20
```

```
CreditBureau_Woedataset$Status_flag = CreditHistory$Status_flag  
CreditBureau_Woedataset$Application.ID_woe <- CreditHistory$Application.ID  
  
summary(CreditBureau_Woedataset)
```

```

## Status_flag      Application.ID_woe  DPD90InLast6Months_woe
## Length:71295     Min.      :1.004e+05    Min.      :-1.7033
## Class :character  1st Qu.:2.484e+08    1st Qu.: -1.7033
## Mode  :character  Median :4.976e+08    Median : -1.7033
##                               Mean  :4.990e+08    Mean  : -0.9949
##                               3rd Qu.:7.496e+08    3rd Qu.: -1.7033
##                               Max.  :1.000e+09    Max.   : 1.3714
## DPD60InLast6Months_woe DPD30InLast6Months_woe DPD90InLast12Months_woe
## Min.      :-3.013      Min.      :-4.6267      Min.      :-3.0870
## 1st Qu.: -3.013      1st Qu.: -4.6267      1st Qu.: -3.0870
## Median : -3.013      Median : -4.6267      Median : -3.0870
## Mean     :-1.939      Mean     :-3.0677      Mean     :-1.9010
## 3rd Qu.:  0.147      3rd Qu.: -0.5753      3rd Qu.:  0.3213
## Max.     :  2.070      Max.     :  2.4262      Max.     :  1.8983
## DPD60InLast12Months_woe DPD30InLast12Months_woe
## Min.      :-4.3562      Min.      :-5.209
## 1st Qu.: -4.3562      1st Qu.: -5.209
## Median : -4.3562      Median : -5.209
## Mean     :-2.6959      Mean     :-3.346
## 3rd Qu.: -0.8565      3rd Qu.: -2.123
## Max.     :  2.3511      Max.     :  2.630
## CCUtilizationIn12Months_woe TradesInLast6Months_woe
## Min.      :-3.2161      Min.      :-2.2546
## 1st Qu.: -3.1962      1st Qu.: -1.3796
## Median : -1.7759      Median : -1.3796
## Mean     :-1.0512      Mean     :-0.5284
## 3rd Qu.:  0.8416      3rd Qu.:  0.8241
## Max.     :  0.8416      Max.     :  3.8817
## TradesInLast12Months_woe PLTradesInLast6Months_woe
## Min.      :-4.4222      Min.      :-1.7245
## 1st Qu.: -2.5987      1st Qu.: -1.7245
## Median : -1.0320      Median :  0.1021
## Mean     :-1.3192      Mean     :-0.4880
## 3rd Qu.:  0.8932      3rd Qu.:  0.6700
## Max.     :  0.8932      Max.     :  0.6700
## PLTradesInLast12Months_woe InquiriesInLast6Months_woe
## Min.      :-4.8803      Min.      :-1.5641
## 1st Qu.: -4.8803      1st Qu.: -1.5641
## Median :  0.5848      Median : -0.5406
## Mean     :-1.5399      Mean     :-0.3859
## 3rd Qu.:  0.5848      3rd Qu.:  0.7238
## Max.     :  0.5848      Max.     :  0.7238
## InquiriesInLast12Months_woe PresenceOfOpenHomeLoan_woe
## Min.      :-3.5548      Min.      :-1.724082
## 1st Qu.: -3.5548      1st Qu.:  0.003206
## Median : -0.1806      Median :  0.003206
## Mean     :-1.1661      Mean     :-0.003384
## 3rd Qu.:  0.7577      3rd Qu.:  0.003206
## Max.     :  0.7577      Max.     :  0.003206
## OutstandingBalance_woe TotalTrades_woe  PresenceOfOpenLoan_woe
## Min.      :-4.3181      Min.      :-5.2389      Min.      :0
## 1st Qu.: -2.6759      1st Qu.: -3.4378      1st Qu.:0
## Median : -0.4298      Median : -1.5040      Median :0
## Mean     :-1.0891      Mean     :-1.2068      Mean     :0
## 3rd Qu.:  0.6985      3rd Qu.:  0.6664      3rd Qu.:0
## Max.     :  0.6985      Max.     :  0.6664      Max.     :0

```

```

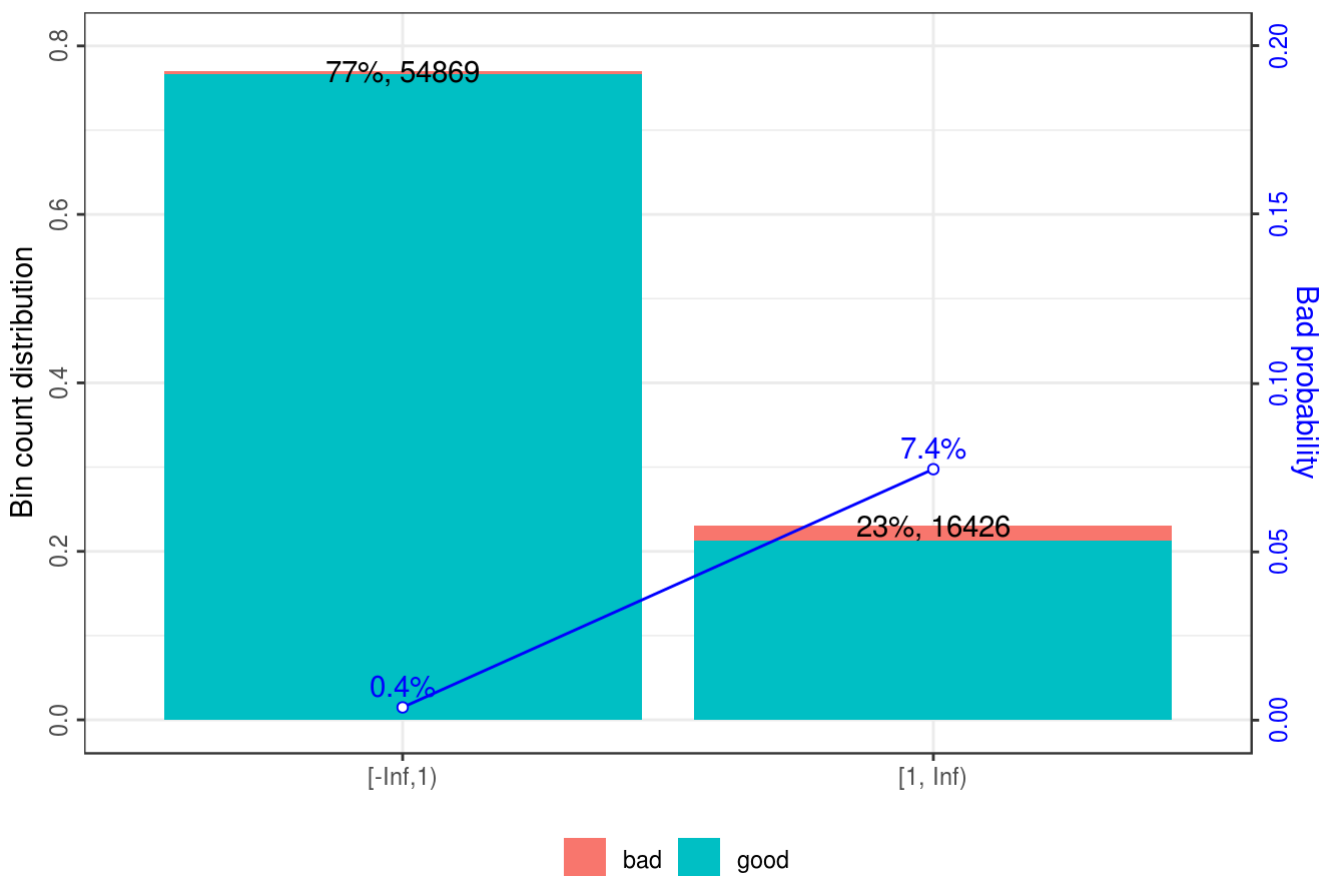
names(CreditBureau_Woedataset)[2] <- paste("Application.ID")
names(CreditBureau_Woedataset)[3] <- paste("DPD90InLast6Months")
names(CreditBureau_Woedataset)[4] <- paste("DPD60InLast6Months")
names(CreditBureau_Woedataset)[5] <- paste("DPD30InLast6Months")
names(CreditBureau_Woedataset)[6] <- paste("DPD90InLast12Months")
names(CreditBureau_Woedataset)[7] <- paste("DPD60InLast12Months")
names(CreditBureau_Woedataset)[8] <- paste("DPD30InLast12Months")
names(CreditBureau_Woedataset)[9] <- paste("CCUtilizationIn12Months")
names(CreditBureau_Woedataset)[10] <- paste("TradesInLast6Months")
names(CreditBureau_Woedataset)[11] <- paste("TradesInLast12Months")
names(CreditBureau_Woedataset)[12] <- paste("PLTradesInLast6Months")
names(CreditBureau_Woedataset)[13] <- paste("PLTradesInLast12Months")
names(CreditBureau_Woedataset)[14] <- paste("InquiriesInLast6Months")
names(CreditBureau_Woedataset)[15] <- paste("InquiriesInLast12Months")
names(CreditBureau_Woedataset)[16] <- paste("PresenceOfOpenHomeLoan")
names(CreditBureau_Woedataset)[17] <- paste("OutstandingBalance")
names(CreditBureau_Woedataset)[18] <- paste("TotalTrades")
names(CreditBureau_Woedataset)[19] <- paste("PresenceOfOpenLoan")

```

```
woebin_plot(CreditHistory_woebins$DPD90InLast6Months)
```

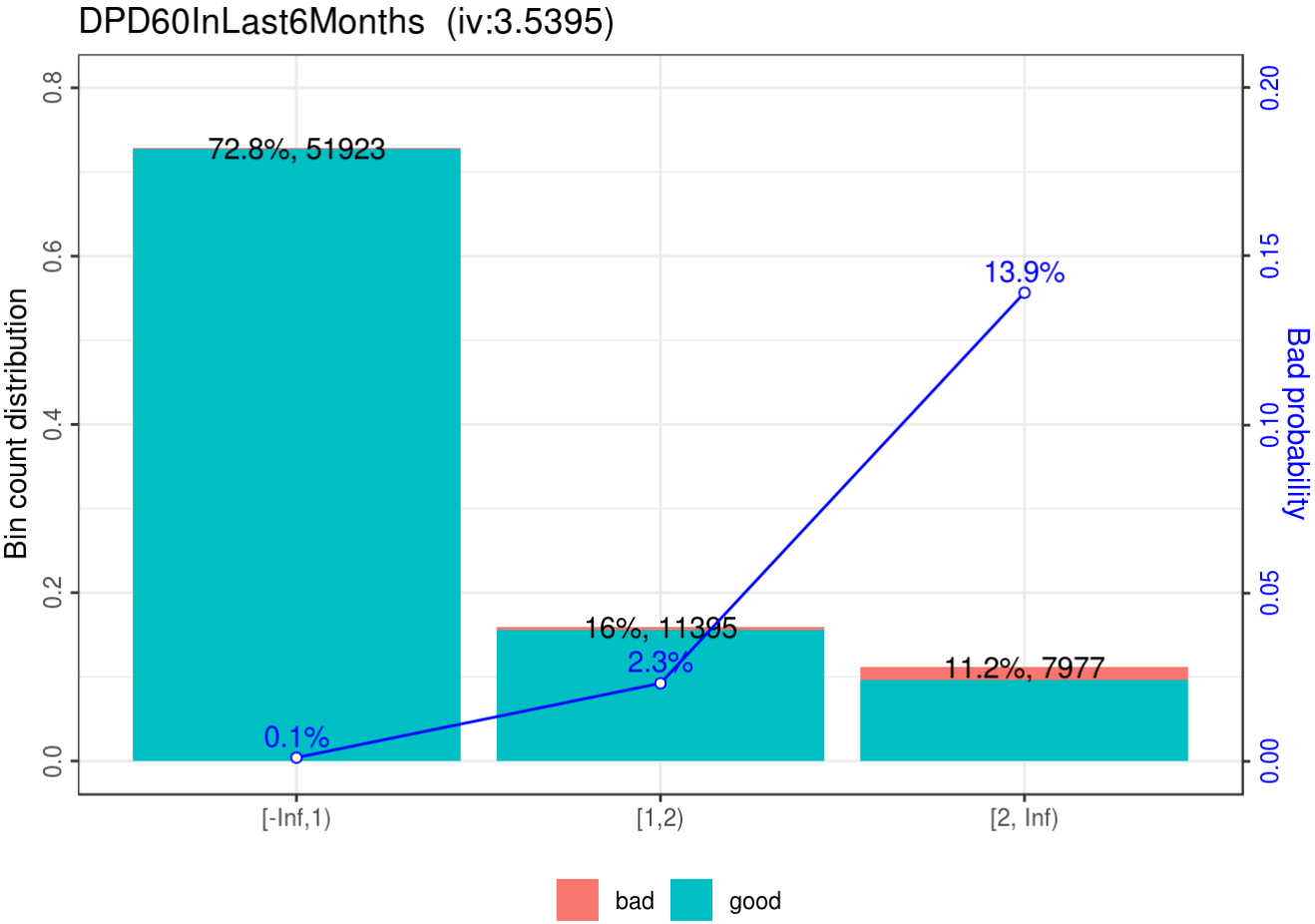
```
## $DPD90InLast6Months
```

DPD90InLast6Months (iv:1.9676)



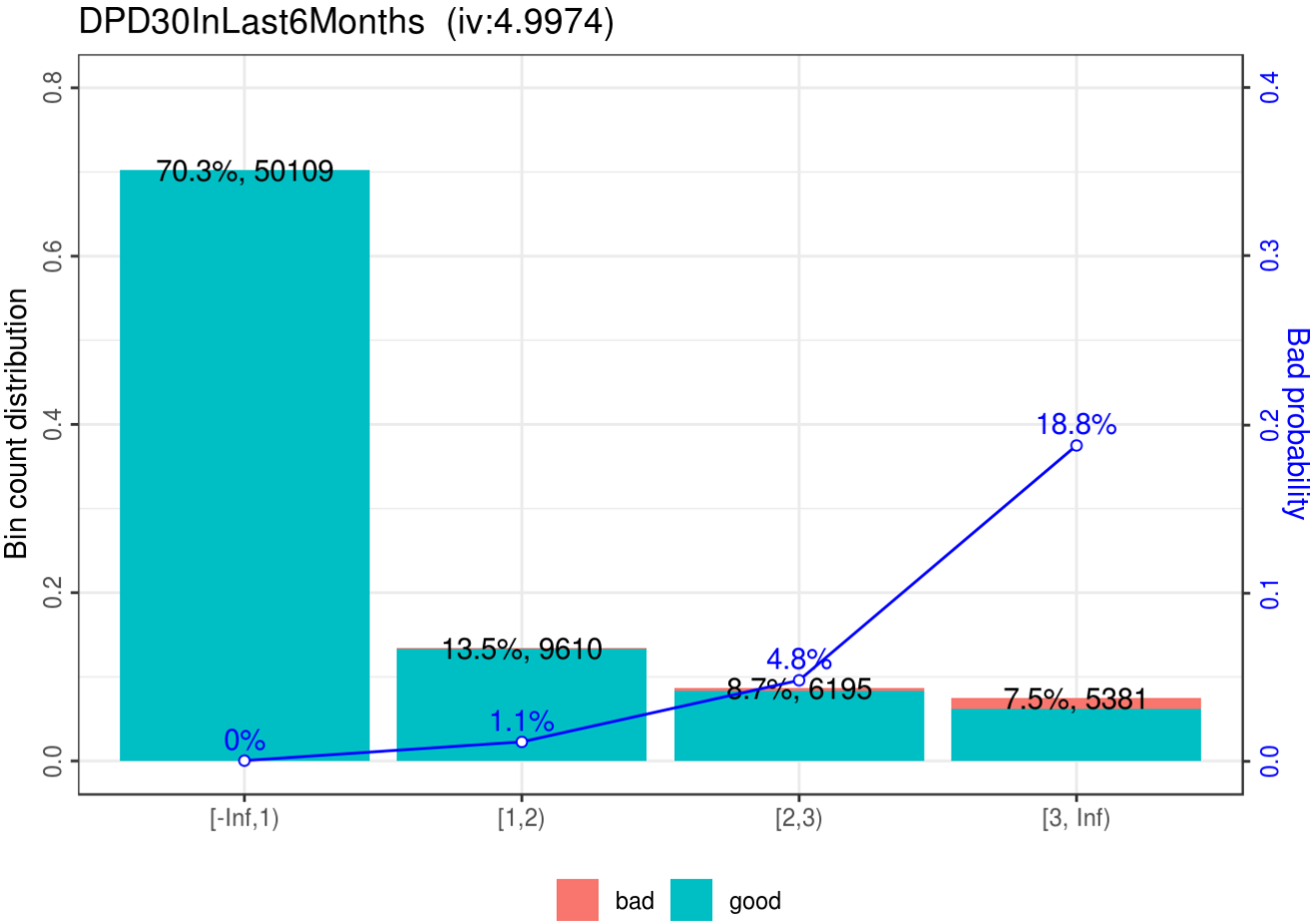
```
woebin_plot(CreditHistory_woebins$DPD60InLast6Months)
```

```
## $DPD60InLast6Months
```



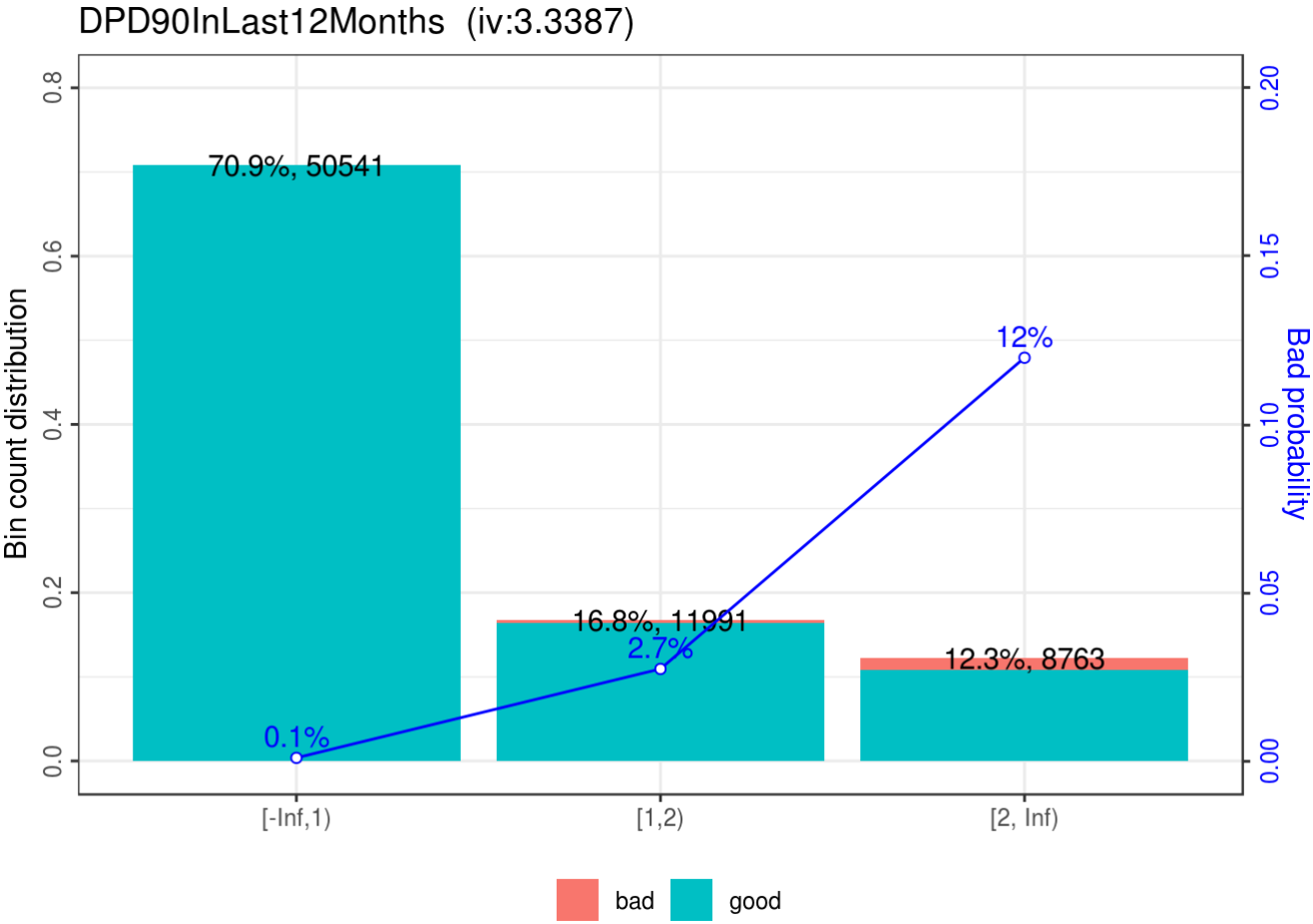
```
woebin_plot(CreditHistory_woebins$DPD30InLast6Months)
```

```
## $DPD30InLast6Months
```



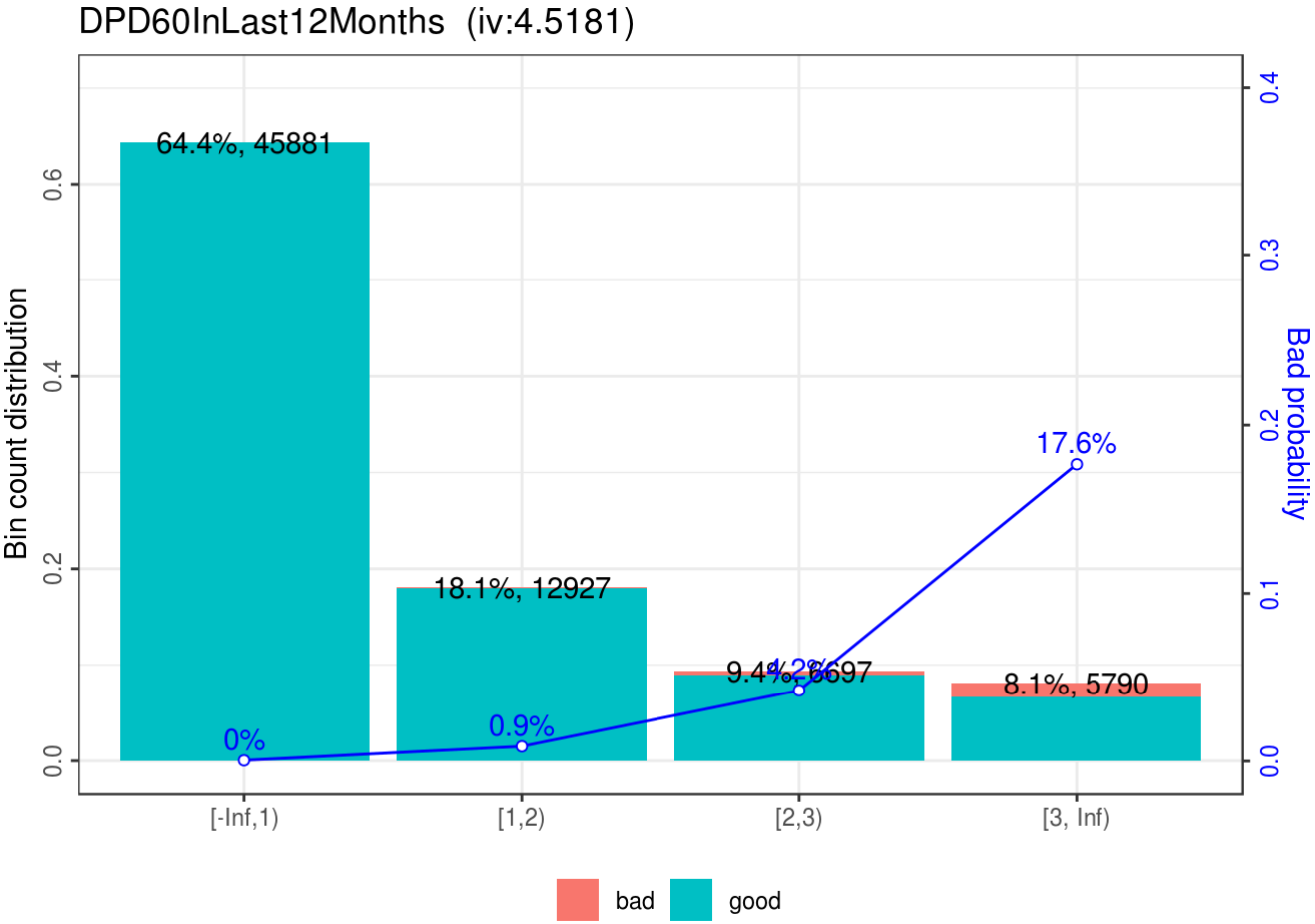
```
woebin_plot(CreditHistory_woebins$DPD90InLast12Months)
```

```
## $DPD90InLast12Months
```



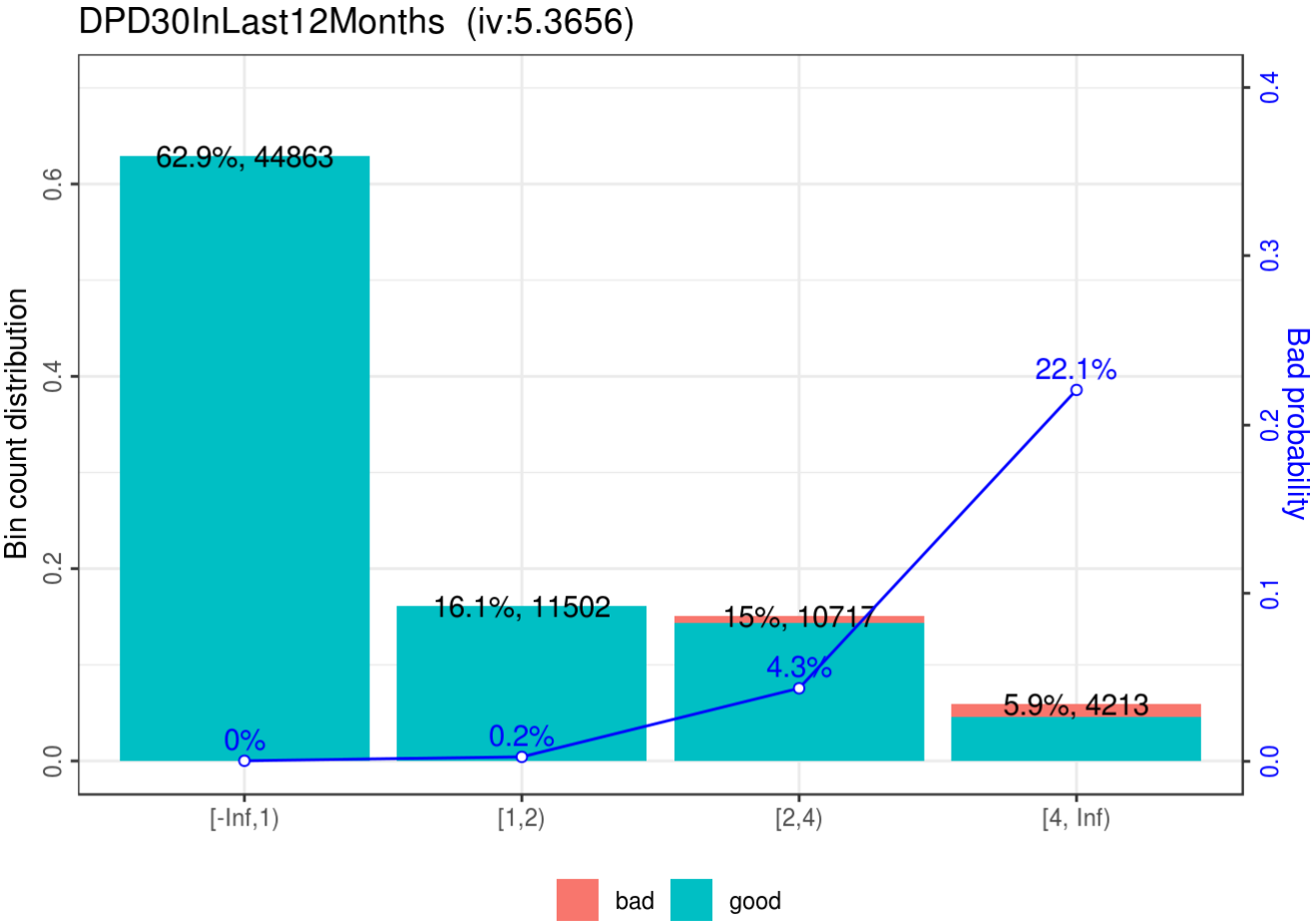
```
woebin_plot(CreditHistory_woebins$DPD60InLast12Months)
```

```
## $DPD60InLast12Months
```



```
woebin_plot(CreditHistory_woebins$DPD30InLast12Months)
```

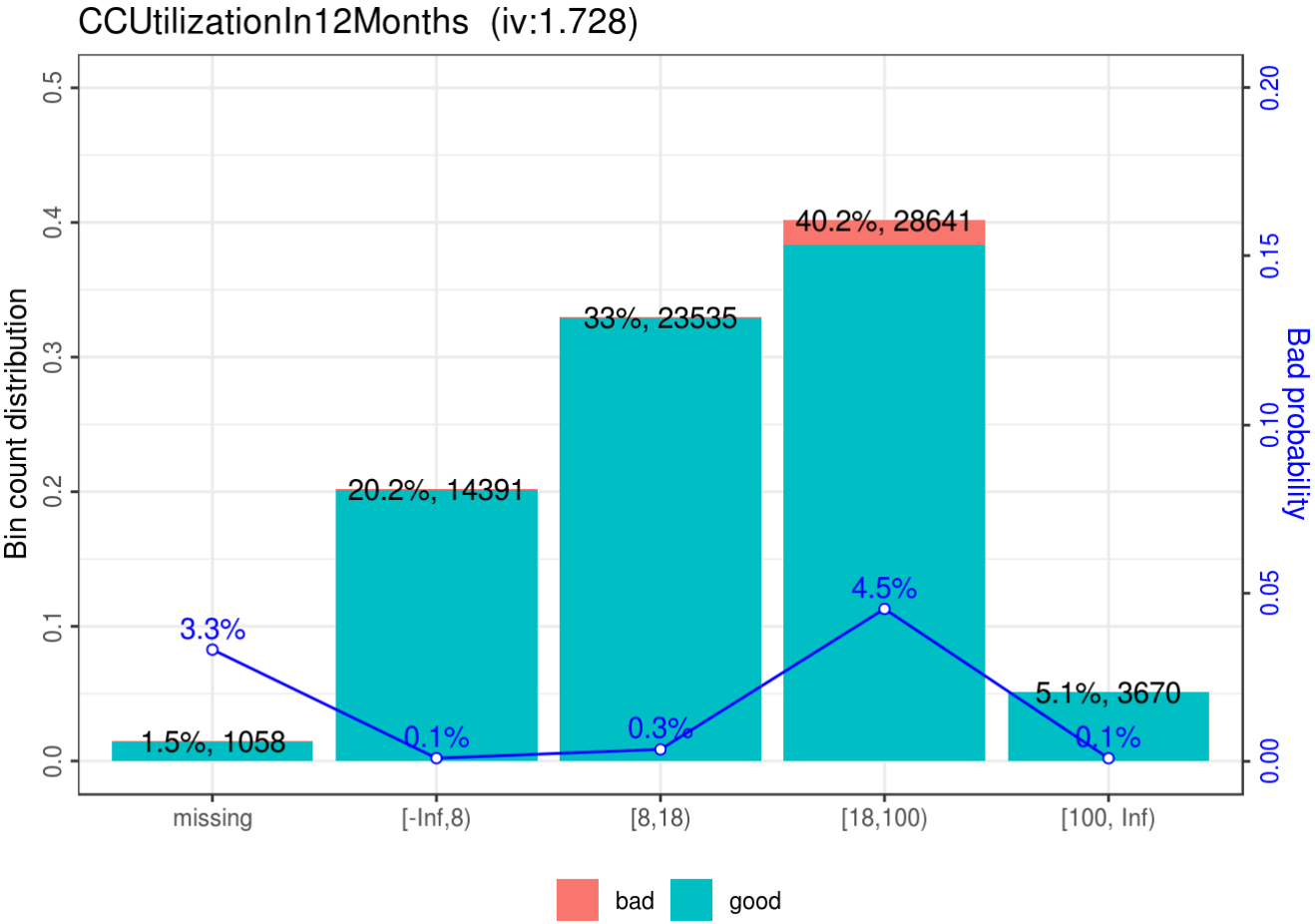
```
## $DPD30InLast12Months
```



```
woebin_plot(CreditHistory_woebins$CCUtilizationIn12Months)
```

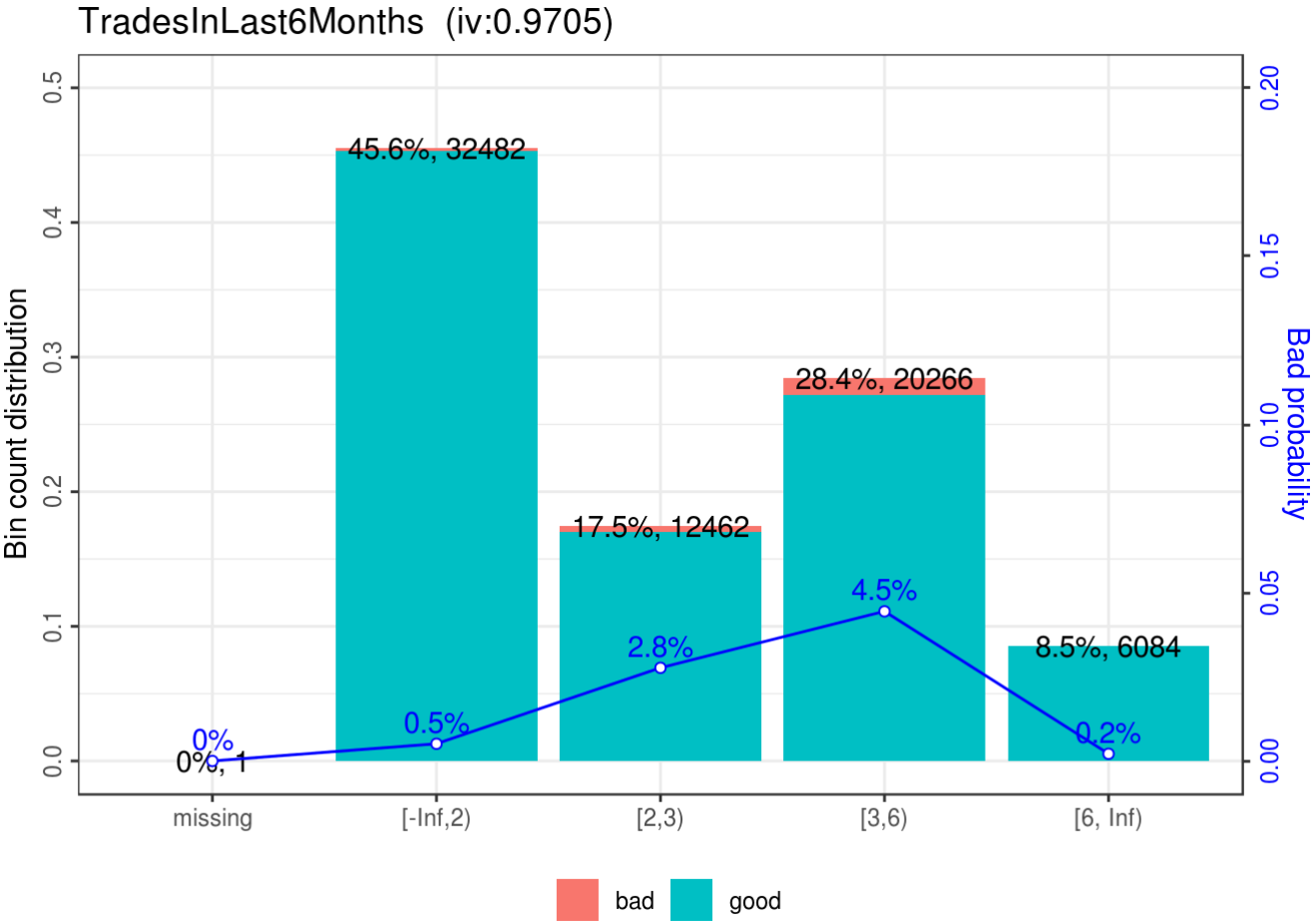
```
## $CCUtilizationIn12Months
```





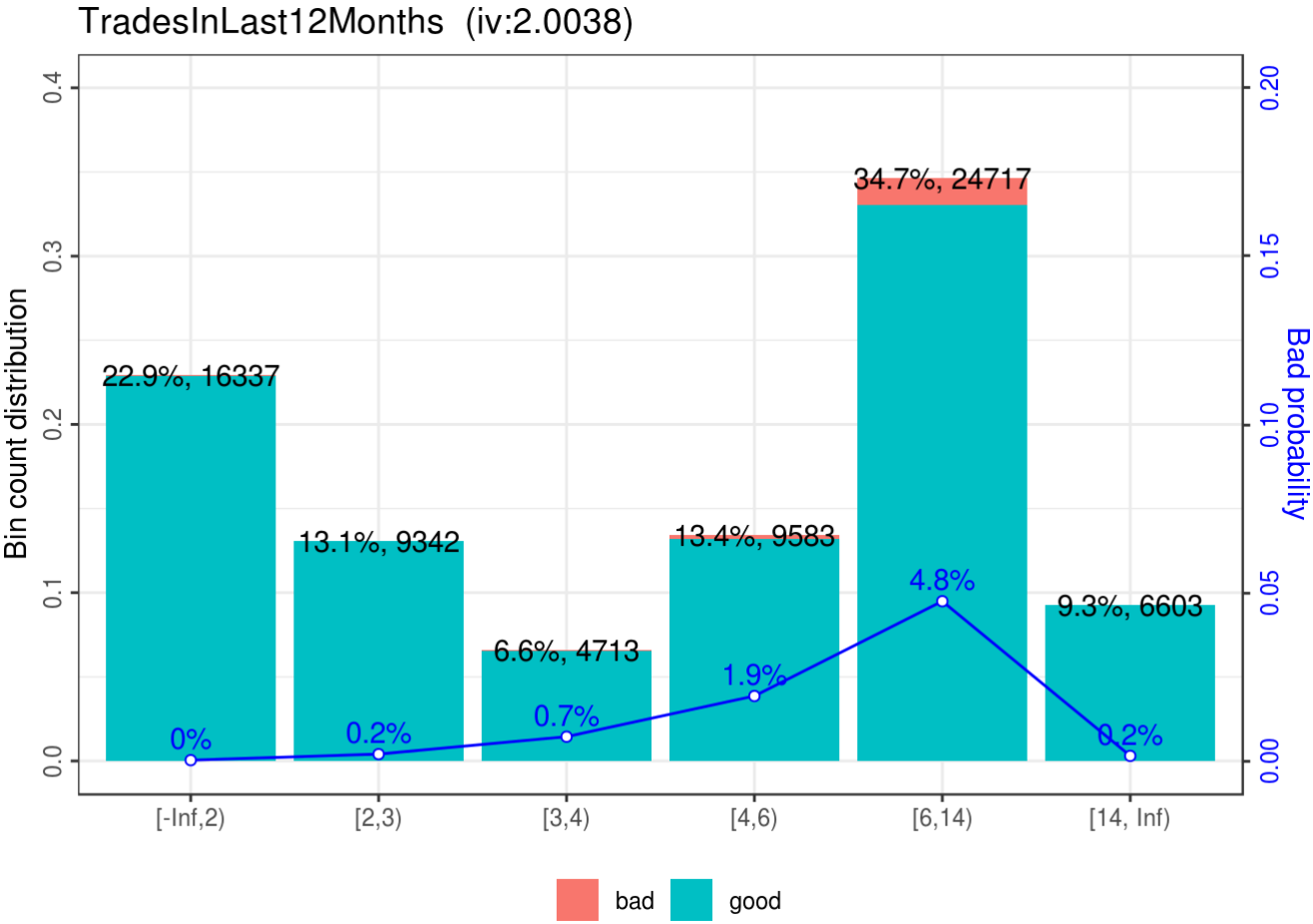
```
woebin_plot(CreditHistory_woebins$TradesInLast6Months)
```

```
## $TradesInLast6Months
```



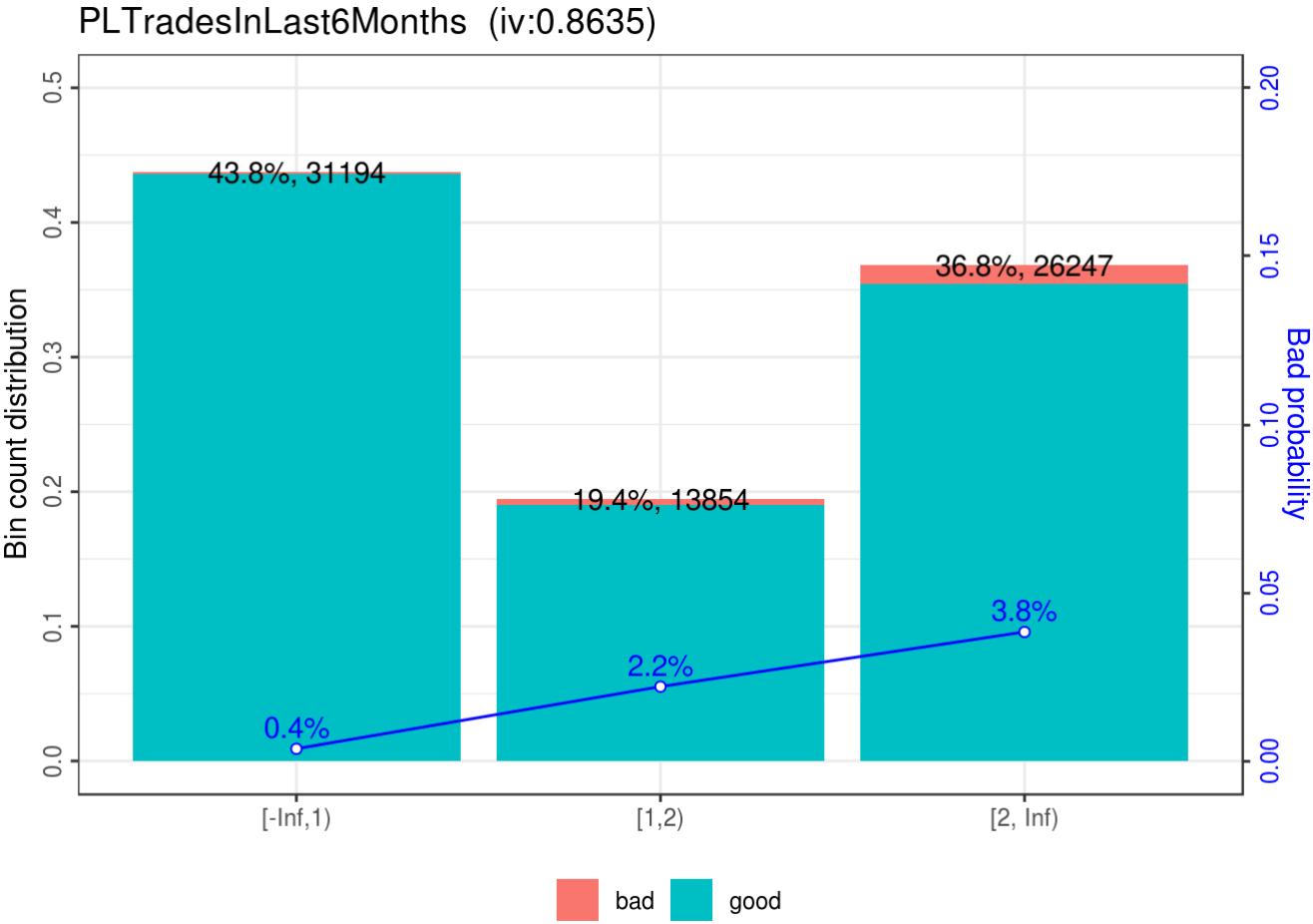
```
woebin_plot(CreditHistory_woebins$TradesInLast12Months)
```

```
## $TradesInLast12Months
```



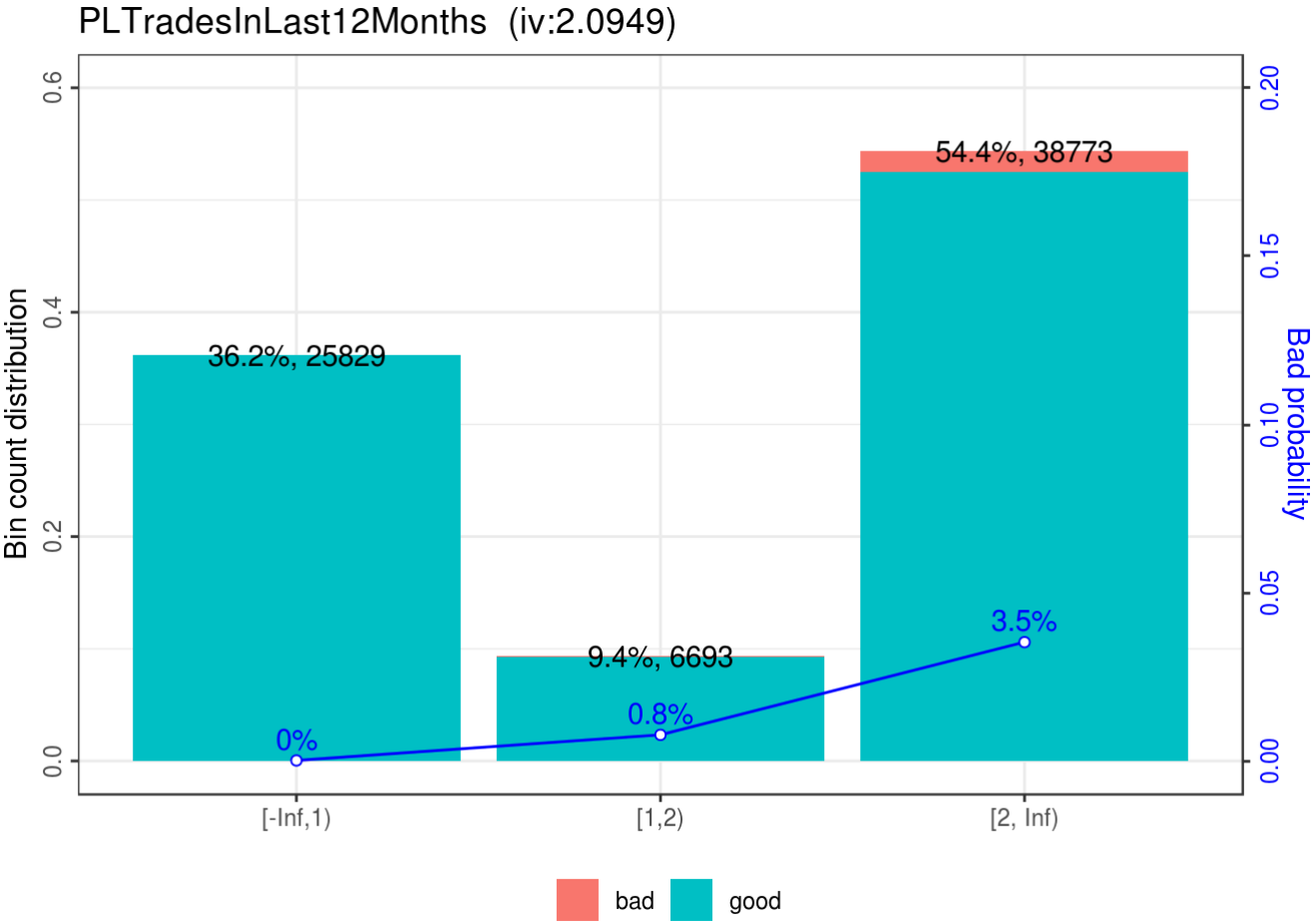
```
woebin_plot(CreditHistory_woebins$PLTradesInLast6Months)
```

```
## $PLTradesInLast6Months
```



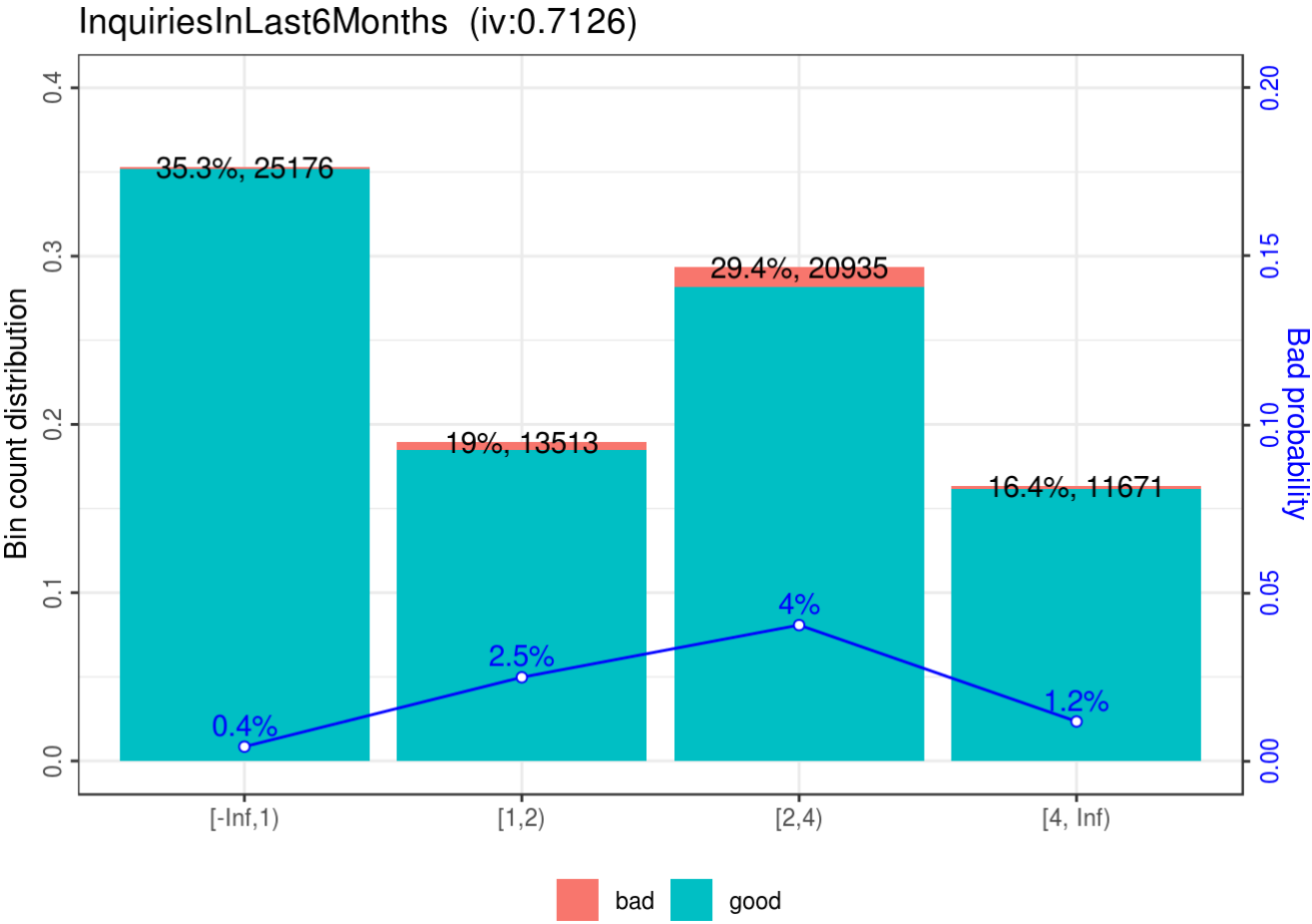
```
woebin_plot(CreditHistory_woebins$PLTradesInLast12Months)
```

```
## $PLTradesInLast12Months
```



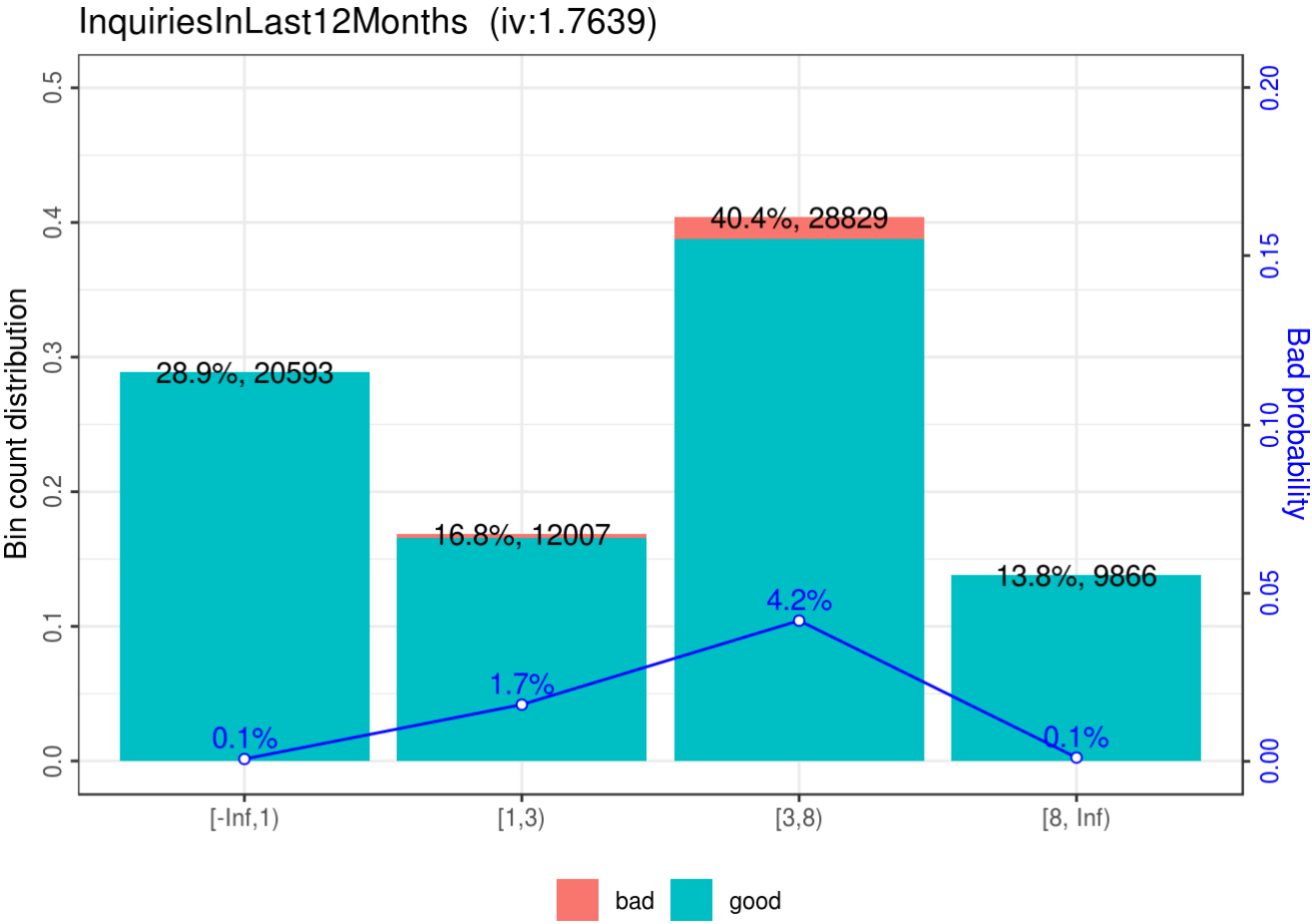
```
woebin_plot(CreditHistory_woebins$InquiriesInLast6Months)
```

```
## $InquiriesInLast6Months
```



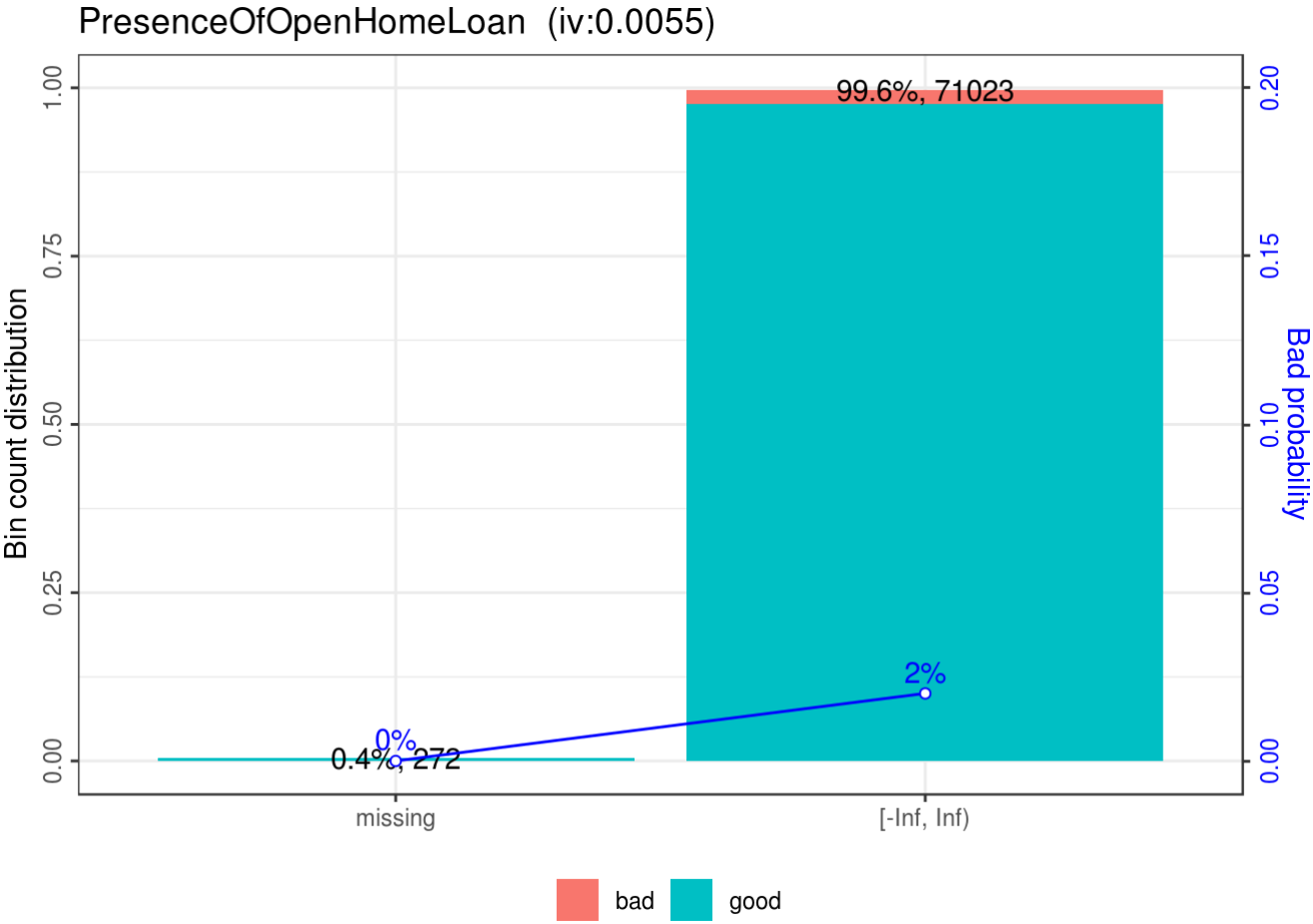
```
woebin_plot(CreditHistory_woebins$InquiriesInLast12Months)
```

```
## $InquiriesInLast12Months
```



```
woebin_plot(CreditHistory_woebins$PresenceOfOpenHomeLoan)
```

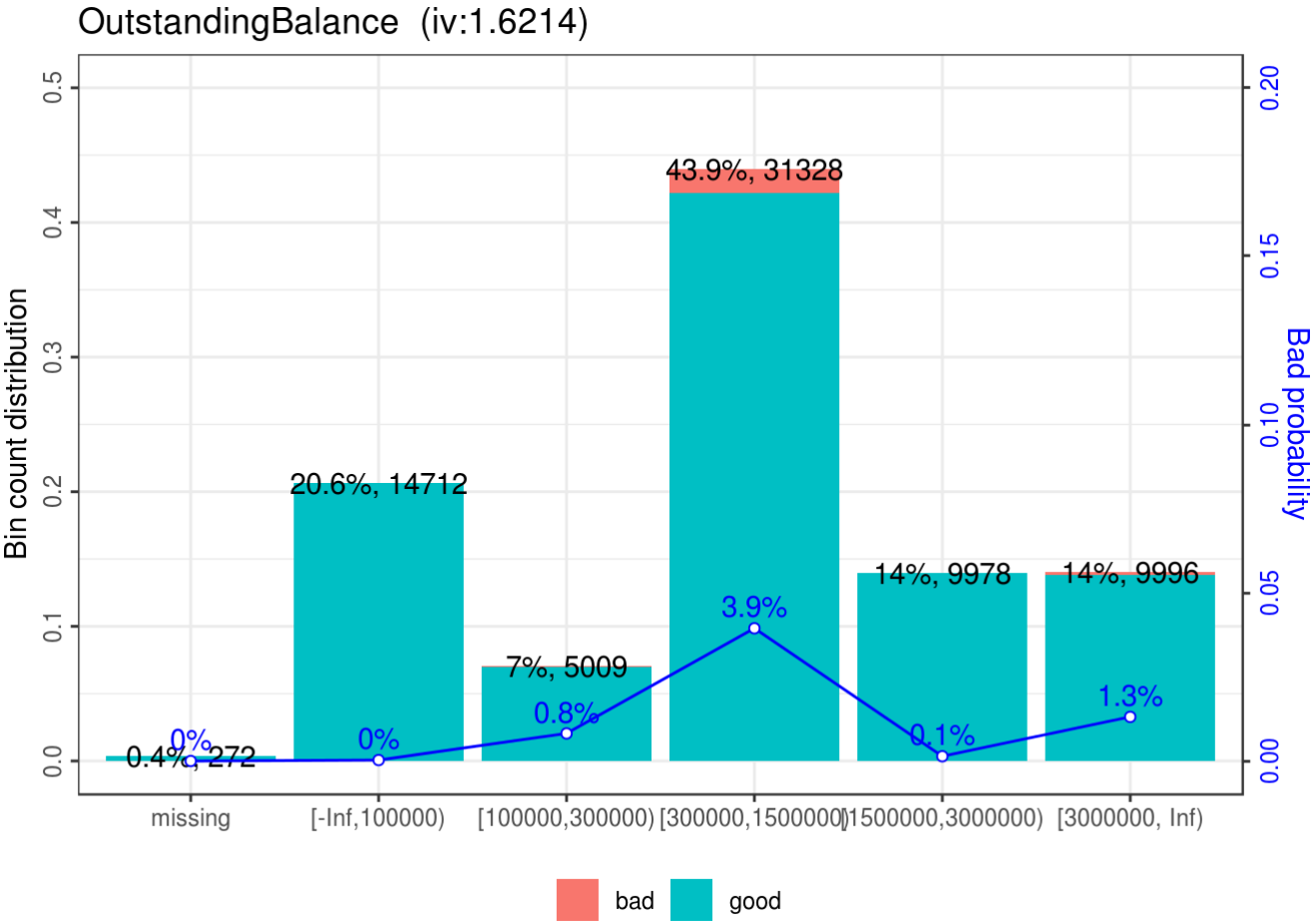
```
## $PresenceOfOpenHomeLoan
```



```
woebin_plot(CreditHistory_woebins$OutstandingBalance)
```

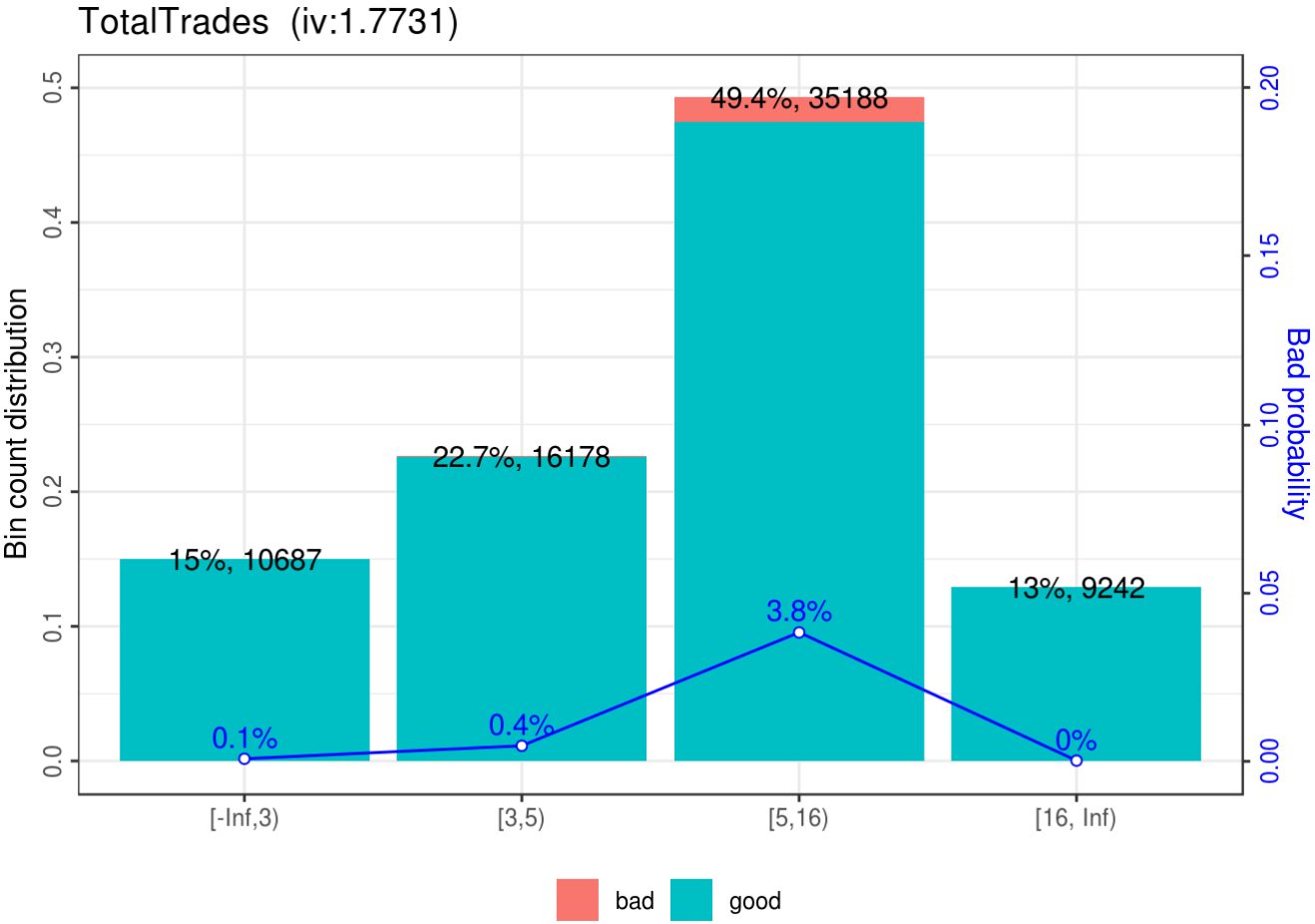
```
## $OutstandingBalance
```





```
woebin_plot(CreditHistory_woebins$TotalTrades)
```

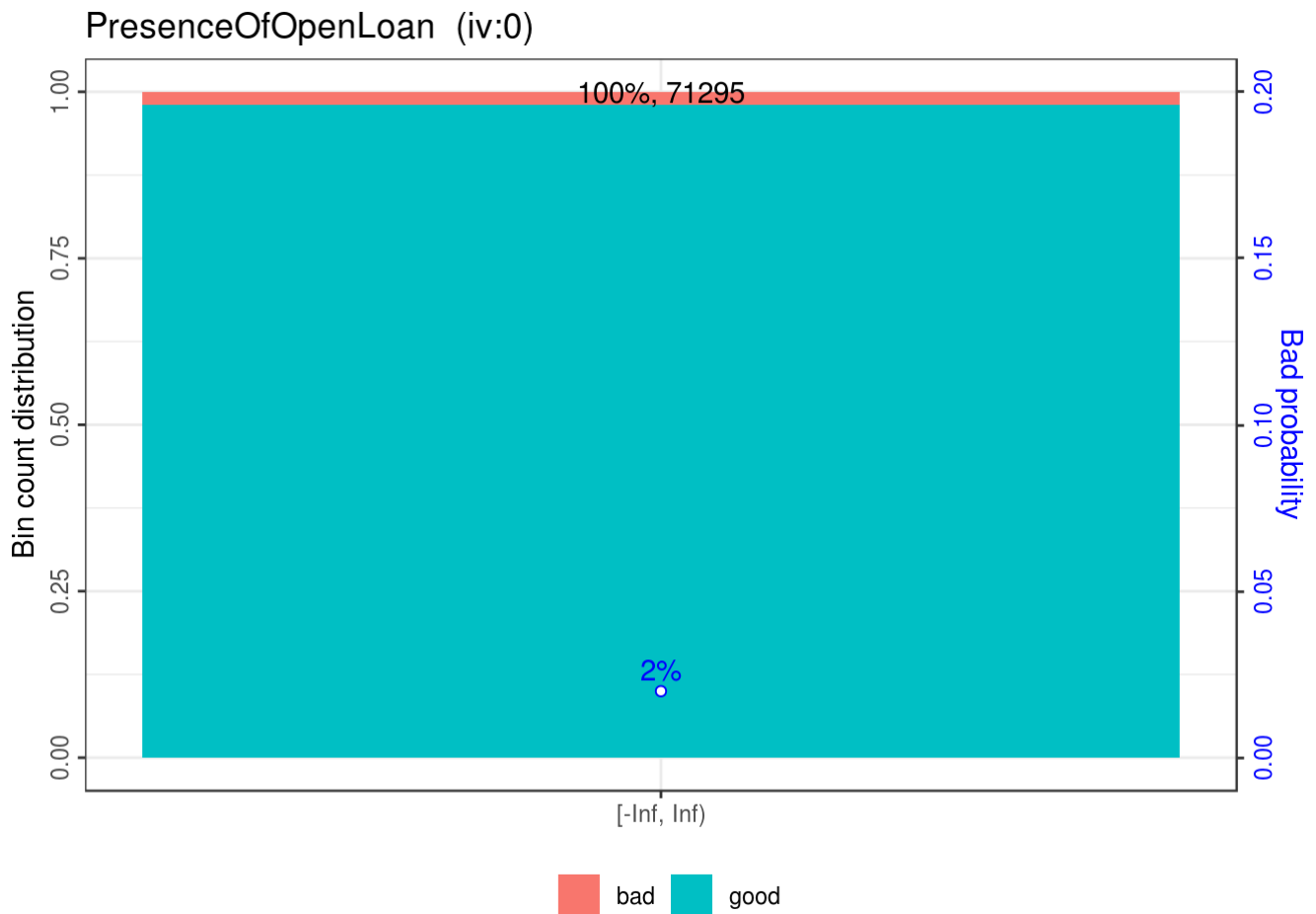
```
## $TotalTrades
```



```
woebin_plot(CreditHistory_woebins$PresenceOfOpenLoan)
```

```
## $PresenceOfOpenLoan
```

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

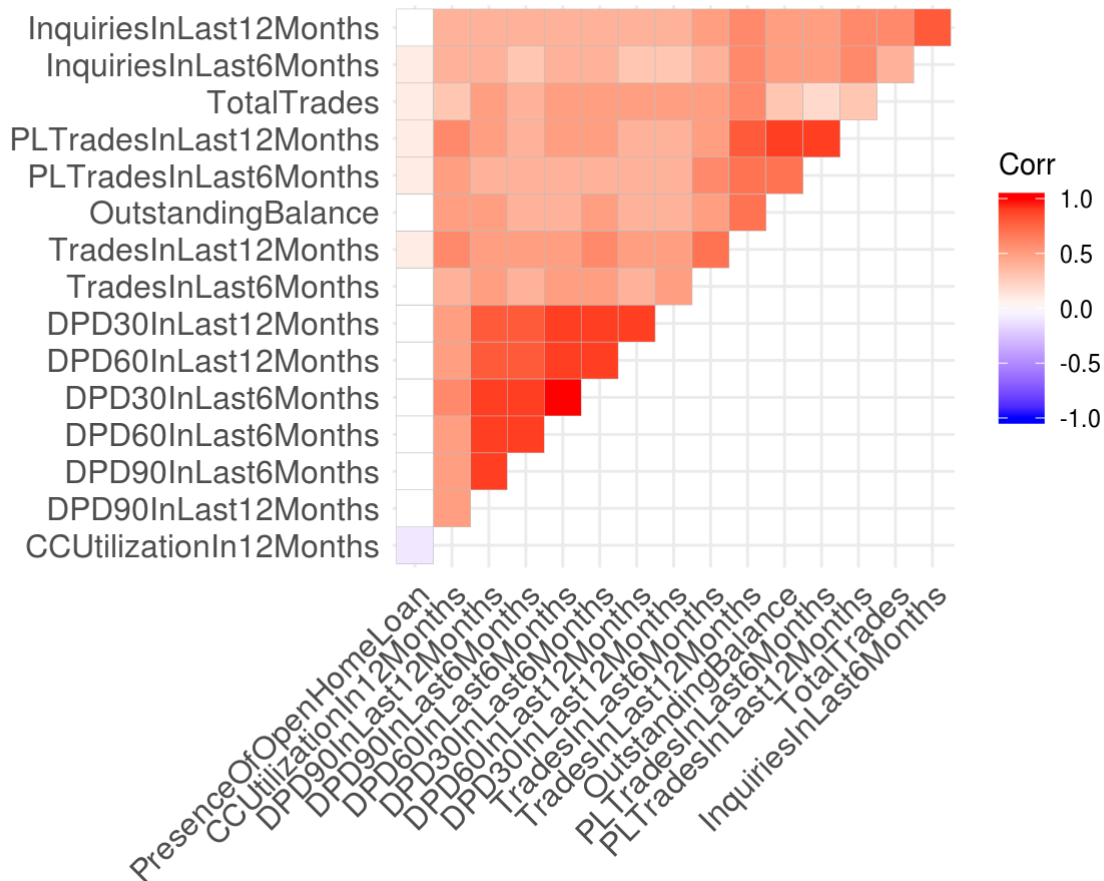


## Performing correlation for CreditBureau Data

### Plotting correlation matrix for credit bureau data

```
library(ggcorrplot)
CreditScoreCorr <- round(cor(x = CreditBureau_Woedataset[, -c(1,2,19)]), 1)
ggcorrplot(CreditScoreCorr, hc.order = TRUE, type = "upper", lab = FALSE, show.legend = TRUE,
            title = "Correlation Matrix For Credit Bureau Data")
```

## Correlation Matrix For Credit Bureau Data



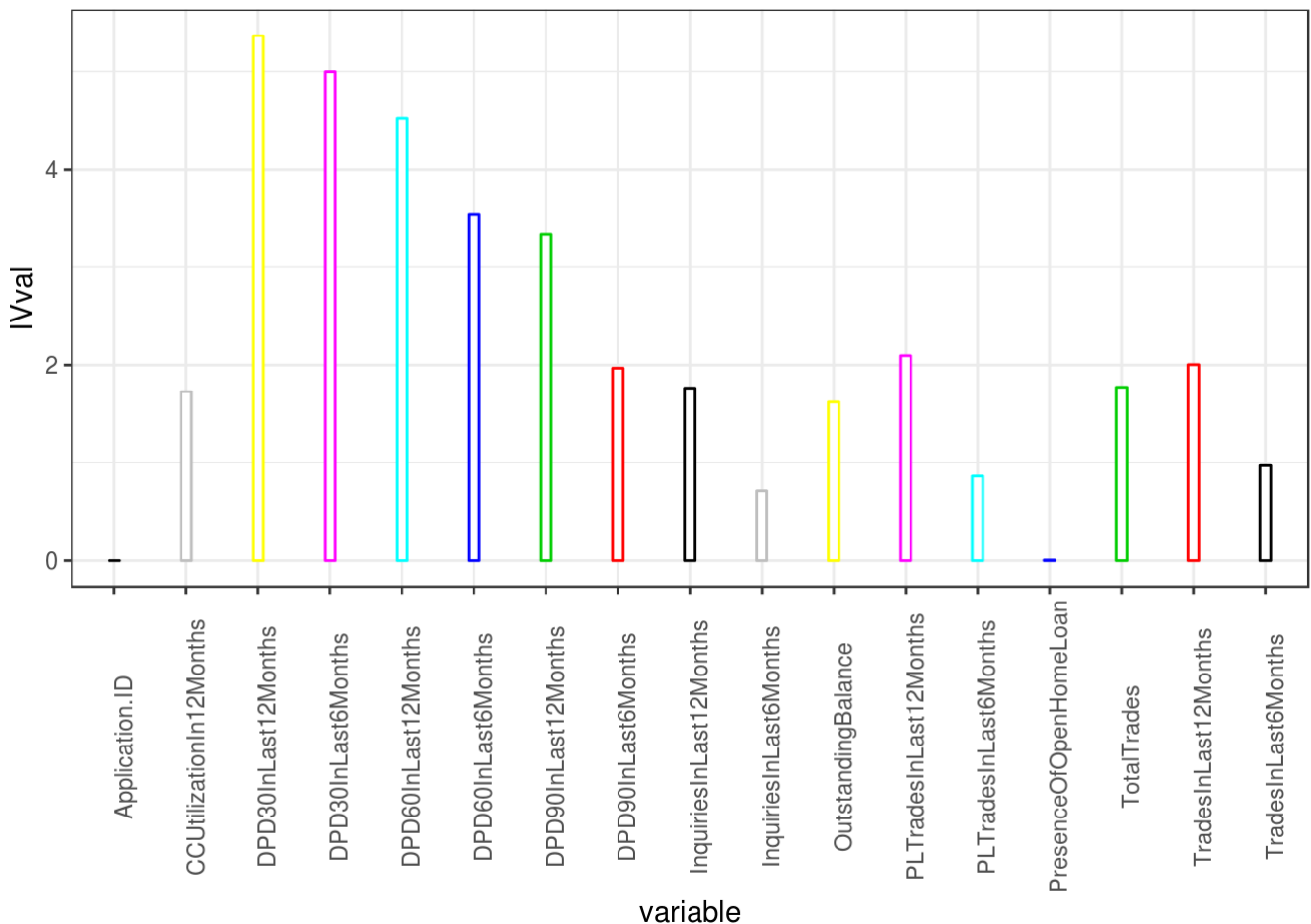
## IV value calculation for Credit bureau data

```
CreditHistory_data_IV = iv(CreditBureau_Woedataset,
                           y = "Status_flag",
                           positive="Bad|1")
```

```
## Warning in rmcol_datetime_unique1(dt): There are 1 columns have only one unique values, which are removed from input dataset.
## (ColumnNames: PresenceOfOpenLoan)
```

```
IVval <- CreditHistory_data_IV$info_value
```

```
ggplot(CreditHistory_data_IV, aes(x = variable, y = IVval)) +
  geom_bar(width = .15, stat = "identity", color = as.factor(CreditHistory_data_IV$info_value), fill = "white") +
  theme_bw() +
  theme(plot.title = element_text(size = 10)) +
  theme(axis.text.x = element_text(angle = 90))
```



## Removing duplicate Applicant IDs from both demographic and credit bureau data

```
DemographicWoeDataRowIndices <- which(duplicated(Demographic_Woedataset$Application.ID) == TRUE)
Demographic_Woedataset <- Demographic_Woedataset[!DemographicWoeDataRowIndices, ]

CreditBureauWoeDataRowIndices <- which(duplicated(CreditBureau_Woedataset$Application.ID) == TRUE)
CreditBureau_Woedataset <- CreditBureau_Woedataset[!CreditBureauWoeDataRowIndices, ]
```

## Merging demographic and credit bureau data, by mapping Application ID and their respective status

```
ApplicantData <- merge(Demographic_Woedataset, CreditBureau_Woedataset,
  by = c('Application.ID', 'Status_flag'))
```

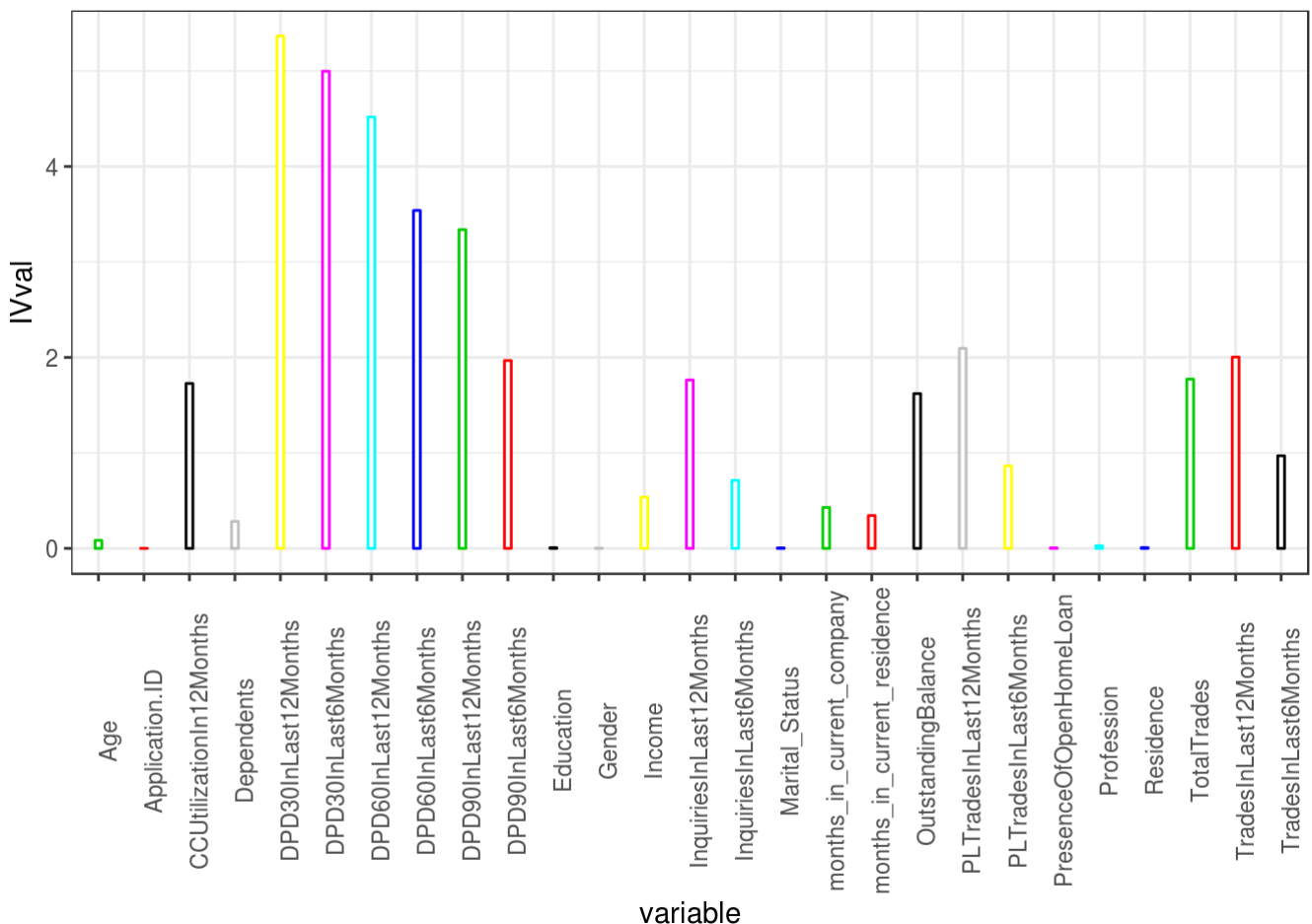
## Collecting significant attributes for applicant data

```
Applicant_data_IV = iv(ApplicantData,
                        y = "Status_flag",
                        positive="Bad|1")
```

```
## Warning in rmcol_datetime_unique1(dt): There are 1 columns have only one unique values, which are removed from input dataset.
## (ColumnNames: PresenceOfOpenLoan)
```

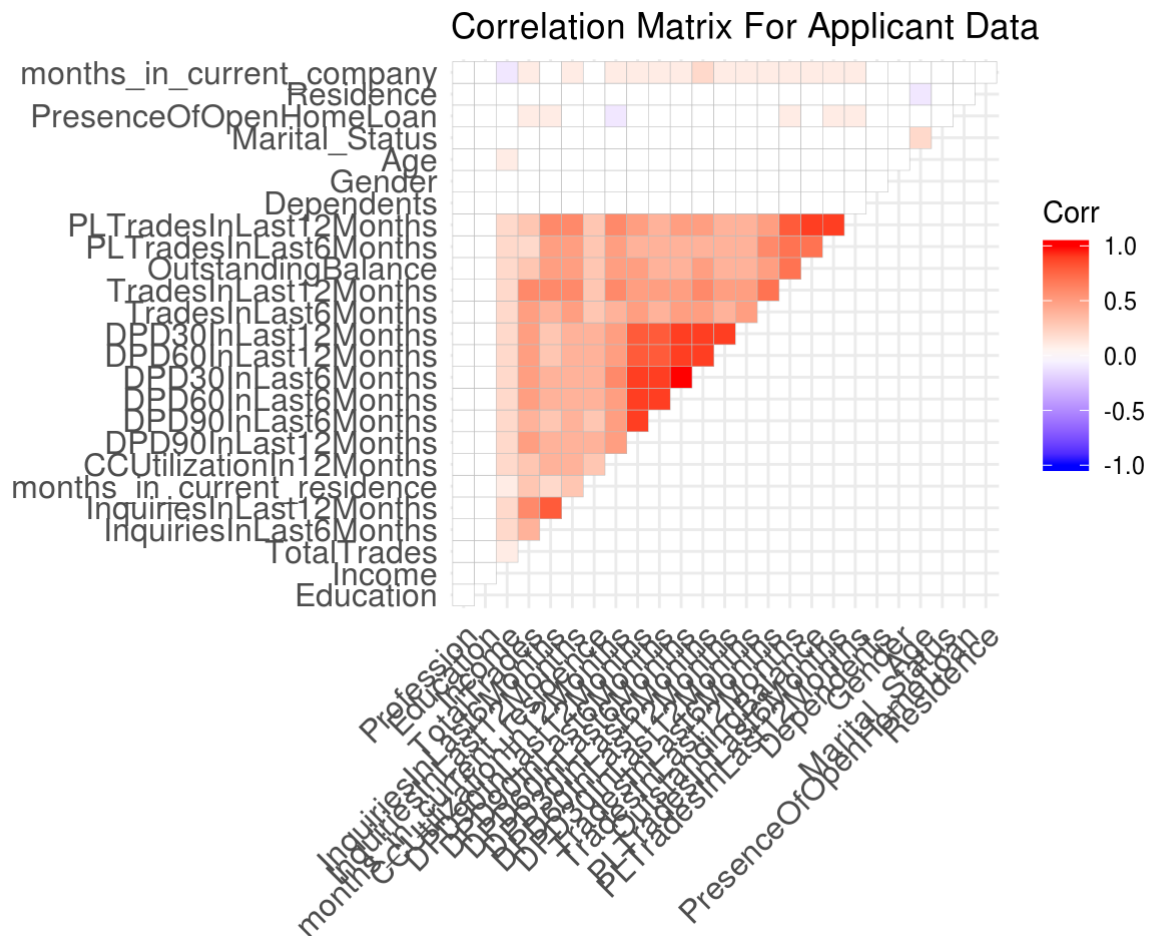
```
IVval <- Applicant_data_IV$info_value
```

```
ggplot(Applicant_data_IV, aes(x = variable, y = IVval)) +
  geom_bar(width = .15, stat = "identity", color = as.factor(Applicant_data_IV$info_value), fill = "white") +
  theme_bw() +
  theme(plot.title = element_text(size = 10)) +
  theme(axis.text.x = element_text(angle = 90))
```



Removing Application.ID, Status\_flag, PresenceOfOpenLoan for correlation computation

```
corr <- round(cor(x = ApplicantData[, -c(1,2,29)]), 1)
ggcorrplot(corr, hc.order = TRUE, type = "upper", lab = FALSE, show.legend = TRUE,
            title = "Correlation Matrix For Applicant Data")
```



```
Applicant_data_IV = iv(ApplicantData, y = 'Status_flag', positive="Bad|1")
```

```
## Warning in rmcol_datetime_unique1(dt): There are 1 columns have only one unique values, which are removed from input dataset.
## (ColumnNames: PresenceOfOpenLoan)
```

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Loading required package: RSQLite
```

```
sqldf::sqldf("select * from Applicant_data_IV where info_value > 0.3")
```

```
##          variable info_value
## 1      DPD30InLast12Months  5.3657049
## 2      DPD30InLast6Months   4.9975168
## 3      DPD60InLast12Months  4.5181862
## 4      DPD60InLast6Months   3.5395373
## 5      DPD90InLast12Months  3.3387324
## 6      PLTradesInLast12Months 2.0949671
## 7      TradesInLast12Months  2.0039044
## 8      DPD90InLast6Months   1.9676705
## 9      TotalTrades          1.7732614
## 10     InquiriesInLast12Months 1.7640799
## 11     CCUtilizationIn12Months 1.7279856
## 12     OutstandingBalance     1.6214047
## 13     TradesInLast6Months    0.9705820
## 14     PLTradesInLast6Months  0.8634414
## 15     InquiriesInLast6Months 0.7127013
## 16     Income                 0.5376751
## 17     months_in_current_company 0.4291018
## 18     months_in_current_residence 0.3441866
```

## Building logistic regression using merged data

```
library(car)
```

```
## Loading required package: carData
```

```
library(MASS)
library(arm)
```

```
## Loading required package: Matrix
```

```
## Loading required package: lme4
```

```
##
## arm (Version 1.10-1, built: 2018-4-12)
```

```
## Working directory is /home/sree/Dive/Training/upGrad/BFSI_Domain/BFSI_CapstoneProject/ReadyToCommit
```

```
##
## Attaching package: 'arm'
```

```
## The following object is masked from 'package:car':
##
##      logit
```



```
library(caTools)
library(e1071)
library(ggcorrplot)
library(caret)
```

```
## Loading required package: lattice
```

```
library(ROCR)
```

```
## Loading required package: gplots
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##      lowess
```

```
library(InformationValue)
```

```
##
## Attaching package: 'InformationValue'
```

```
## The following objects are masked from 'package:caret':
##
##      confusionMatrix, precision, sensitivity, specificity
```

## making a copy of data, specific for building logistic regression model

```
LogisticData <- ApplicantData

colnames(LogisticData)
```

```
## [1] "Application.ID"           "Status_flag"
## [3] "Age"                      "Gender"
## [5] "Marital_Status"          "Dependents"
## [7] "Income"                   "Education"
## [9] "Profession"               "Residence"
## [11] "months_in_current_residence" "months_in_current_company"
## [13] "DPD90InLast6Months"      "DPD60InLast6Months"
## [15] "DPD30InLast6Months"      "DPD90InLast12Months"
## [17] "DPD60InLast12Months"     "DPD30InLast12Months"
## [19] "CCUtilizationIn12Months" "TradesInLast6Months"
## [21] "TradesInLast12Months"    "PLTradesInLast6Months"
## [23] "PLTradesInLast12Months"  "InquiriesInLast6Months"
## [25] "InquiriesInLast12Months" "PresenceOfOpenHomeLoan"
## [27] "OutstandingBalance"      "TotalTrades"
## [29] "PresenceOfOpenLoan"
```

```
set.seed(1366)
LogisticData$Status_flag = as.factor(LogisticData$Status_flag)
split_indices <- sample.split(LogisticData$Status_flag, SplitRatio = 0.70)

LogisticData_train <- LogisticData[split_indices, ]
LogisticData_test <- LogisticData[!split_indices, ]
```

## Model Building - Logistic Regression

```
BayesianLogit <- bayesglm(Status_flag ~ DPD30InLast12Months + DPD30InLast6Months +
                          DPD60InLast12Months + DPD60InLast6Months + DPD90InLast1
2Months +
                          PLTradesInLast12Months + TradesInLast12Months + DPD90In
Last6Months +
                          TotalTrades + InquiriesInLast12Months + CCUtilizationIn
12Months +
                          OutstandingBalance + TradesInLast6Months + PLTradesInLa
st6Months +
                          InquiriesInLast6Months + Income + months_in_current_com
pany +
                          months_in_current_residence , data = LogisticData_train,
family = 'binomial')

summary(BayesianLogit)
```

```
##
## Call:
## bayesglm(formula = Status_flag ~ DPD30InLast12Months + DPD30InLast6Months +
##   DPD60InLast12Months + DPD60InLast6Months + DPD90InLast12Months +
##   PLTradesInLast12Months + TradesInLast12Months + DPD90InLast6Months +
##   TotalTrades + InquiriesInLast12Months + CCUtilizationIn12Months +
##   OutstandingBalance + TradesInLast6Months + PLTradesInLast6Months +
##   InquiriesInLast6Months + Income + months_in_current_company +
##   months_in_current_residence, family = "binomial", data = LogisticData_train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4721  -0.0645  -0.0097  -0.0045   4.3613
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.16336    0.09006  -46.228 < 2e-16 ***
## DPD30InLast12Months    0.50397    0.05189   9.712 < 2e-16 ***
## DPD30InLast6Months     0.29904    0.06770   4.417 9.99e-06 ***
## DPD60InLast12Months    0.50481    0.05222   9.667 < 2e-16 ***
## DPD60InLast6Months   -0.36425    0.06209  -5.867 4.45e-09 ***
## DPD90InLast12Months    0.29345    0.04333   6.772 1.27e-11 ***
## PLTradesInLast12Months -0.05212    0.10748  -0.485  0.62772
## TradesInLast12Months    0.09846    0.08962   1.099  0.27193
## DPD90InLast6Months   -0.30759    0.04739  -6.491 8.55e-11 ***
## TotalTrades          -0.02616    0.07663  -0.341  0.73280
## InquiriesInLast12Months  0.36231    0.05050   7.174 7.26e-13 ***
## CCUtilizationIn12Months  0.04168    0.04657   0.895  0.37070
## OutstandingBalance     0.21024    0.07905   2.660  0.00782 **
## TradesInLast6Months   -0.01379    0.06270  -0.220  0.82590
## PLTradesInLast6Months  -0.05320    0.07355  -0.723  0.46954
## InquiriesInLast6Months  0.03065    0.05642   0.543  0.58692
## Income                0.40225    0.04930   8.159 3.37e-16 ***
## months_in_current_company  0.42005    0.05421   7.749 9.27e-15 ***
## months_in_current_residence 0.02696    0.05797   0.465  0.64190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 9776.7  on 49903  degrees of freedom
## Residual deviance: 5621.4  on 49885  degrees of freedom
## AIC: 5659.4
##
## Number of Fisher Scoring iterations: 11
```

```
stepAIC(BayesianLogit)
```

```

## Start: AIC=5659.42
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
##   DPD60InLast6Months + DPD90InLast12Months + PLTradesInLast12Months +
##   TradesInLast12Months + DPD90InLast6Months + TotalTrades +
##   InquiriesInLast12Months + CCUtilizationIn12Months + OutstandingBalance +
##   TradesInLast6Months + PLTradesInLast6Months + InquiriesInLast6Months +
##   Income + months_in_current_company + months_in_current_residence
##
##
##      Df Deviance    AIC
## - TradesInLast6Months      1   5621.4 5657.4
## - TotalTrades              1   5621.5 5657.5
## - months_in_current_residence 1   5621.6 5657.6
## - InquiriesInLast6Months    1   5621.6 5657.6
## - PLTradesInLast12Months    1   5621.7 5657.7
## - PLTradesInLast6Months     1   5621.9 5657.9
## - CCUtilizationIn12Months   1   5622.2 5658.2
## - TradesInLast12Months      1   5622.7 5658.7
## <none>                      5621.4 5659.4
## - OutstandingBalance        1   5628.9 5664.9
## - DPD30InLast6Months        1   5641.4 5677.4
## - DPD60InLast6Months        1   5655.6 5691.6
## - DPD90InLast6Months        1   5660.8 5696.8
## - DPD90InLast12Months       1   5671.1 5707.1
## - months_in_current_company  1   5681.7 5717.7
## - Income                    1   5689.0 5725.0
## - InquiriesInLast12Months   1   5690.3 5726.3
## - DPD30InLast12Months       1   5723.8 5759.8
## - DPD60InLast12Months       1   5725.0 5761.0
##
## Step: AIC=5657.48
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
##   DPD60InLast6Months + DPD90InLast12Months + PLTradesInLast12Months +
##   TradesInLast12Months + DPD90InLast6Months + TotalTrades +
##   InquiriesInLast12Months + CCUtilizationIn12Months + OutstandingBalance +
##   PLTradesInLast6Months + InquiriesInLast6Months + Income +
##   months_in_current_company + months_in_current_residence
##
##
##      Df Deviance    AIC
## - TotalTrades              1   5621.6 5655.6
## - months_in_current_residence 1   5621.6 5655.6
## - PLTradesInLast12Months    1   5621.7 5655.7
## - InquiriesInLast6Months    1   5621.7 5655.7
## - CCUtilizationIn12Months   1   5622.3 5656.3
## - PLTradesInLast6Months     1   5622.3 5656.3
## - TradesInLast12Months      1   5622.7 5656.7
## <none>                      5621.5 5657.5
## - OutstandingBalance        1   5629.0 5663.0
## - DPD30InLast6Months        1   5641.4 5675.4
## - DPD60InLast6Months        1   5655.7 5689.7
## - DPD90InLast6Months        1   5660.9 5694.9
## - DPD90InLast12Months       1   5671.2 5705.2
## - months_in_current_company  1   5681.8 5715.8
## - Income                    1   5689.0 5723.0
## - InquiriesInLast12Months   1   5690.4 5724.4
## - DPD30InLast12Months       1   5723.9 5757.9
## - DPD60InLast12Months       1   5725.1 5759.1
##

```

```

## Step: AIC=5655.62
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
##   DPD60InLast6Months + DPD90InLast12Months + PLTradesInLast12Months +
##   TradesInLast12Months + DPD90InLast6Months + InquiriesInLast12Months +
##   CCUtilizationIn12Months + OutstandingBalance + PLTradesInLast6Months +
##   InquiriesInLast6Months + Income + months_in_current_company +
##   months_in_current_residence
##
##              Df Deviance    AIC
## - months_in_current_residence 1  5621.8 5653.8
## - InquiriesInLast6Months      1  5621.8 5653.8
## - PLTradesInLast12Months      1  5621.8 5653.8
## - CCUtilizationIn12Months     1  5622.4 5654.4
## - PLTradesInLast6Months       1  5622.4 5654.4
## - TradesInLast12Months        1  5622.8 5654.8
## <none>                        5621.6 5655.6
## - OutstandingBalance          1  5629.0 5661.0
## - DPD30InLast6Months          1  5641.6 5673.6
## - DPD60InLast6Months          1  5655.8 5687.8
## - DPD90InLast6Months          1  5661.1 5693.1
## - DPD90InLast12Months         1  5671.4 5703.4
## - months_in_current_company   1  5682.0 5714.0
## - Income                      1  5689.2 5721.2
## - InquiriesInLast12Months     1  5690.5 5722.5
## - DPD30InLast12Months         1  5724.1 5756.1
## - DPD60InLast12Months         1  5725.2 5757.2
##
## Step: AIC=5653.83
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
##   DPD60InLast6Months + DPD90InLast12Months + PLTradesInLast12Months +
##   TradesInLast12Months + DPD90InLast6Months + InquiriesInLast12Months +
##   CCUtilizationIn12Months + OutstandingBalance + PLTradesInLast6Months +
##   InquiriesInLast6Months + Income + months_in_current_company
##
##              Df Deviance    AIC
## - InquiriesInLast6Months      1  5622.0 5652.0
## - PLTradesInLast12Months      1  5622.0 5652.0
## - CCUtilizationIn12Months     1  5622.6 5652.6
## - PLTradesInLast6Months       1  5622.7 5652.7
## - TradesInLast12Months        1  5623.0 5653.0
## <none>                        5621.8 5653.8
## - OutstandingBalance          1  5629.2 5659.2
## - DPD30InLast6Months          1  5641.8 5671.8
## - DPD60InLast6Months          1  5656.1 5686.1
## - DPD90InLast6Months          1  5661.3 5691.3
## - DPD90InLast12Months         1  5671.7 5701.7
## - months_in_current_company   1  5682.2 5712.2
## - Income                      1  5689.5 5719.5
## - InquiriesInLast12Months     1  5690.8 5720.8
## - DPD30InLast12Months         1  5724.4 5754.4
## - DPD60InLast12Months         1  5725.4 5755.4
##
## Step: AIC=5652.1
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
##   DPD60InLast6Months + DPD90InLast12Months + PLTradesInLast12Months +
##   TradesInLast12Months + DPD90InLast6Months + InquiriesInLast12Months +
##   CCUtilizationIn12Months + OutstandingBalance + PLTradesInLast6Months +
##   Income + months_in_current_company

```

```

##
##
## Df Deviance AIC
## - PLTradesInLast12Months 1 5622.3 5650.3
## - CCUtilizationIn12Months 1 5622.9 5650.9
## - PLTradesInLast6Months 1 5622.9 5650.9
## - TradesInLast12Months 1 5623.3 5651.3
## <none> 5622.1 5652.1
## - OutstandingBalance 1 5629.5 5657.5
## - DPD30InLast6Months 1 5642.1 5670.1
## - DPD60InLast6Months 1 5656.4 5684.4
## - DPD90InLast6Months 1 5661.6 5689.6
## - DPD90InLast12Months 1 5672.0 5700.0
## - months_in_current_company 1 5682.7 5710.7
## - Income 1 5689.6 5717.6
## - InquiriesInLast12Months 1 5703.7 5731.7
## - DPD30InLast12Months 1 5724.9 5752.9
## - DPD60InLast12Months 1 5725.8 5753.8
##
## Step: AIC=5650.37
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
## DPD60InLast6Months + DPD90InLast12Months + TradesInLast12Months +
## DPD90InLast6Months + InquiriesInLast12Months + CCUtilizationIn12Months +
## OutstandingBalance + PLTradesInLast6Months + Income + months_in_current_compan
y
##
## Df Deviance AIC
## - CCUtilizationIn12Months 1 5623.2 5649.2
## - TradesInLast12Months 1 5623.3 5649.3
## - PLTradesInLast6Months 1 5623.4 5649.4
## <none> 5622.4 5650.4
## - OutstandingBalance 1 5630.8 5656.8
## - DPD30InLast6Months 1 5642.3 5668.3
## - DPD60InLast6Months 1 5656.7 5682.7
## - DPD90InLast6Months 1 5661.7 5687.7
## - DPD90InLast12Months 1 5672.1 5698.1
## - months_in_current_company 1 5683.1 5709.1
## - Income 1 5689.7 5715.7
## - InquiriesInLast12Months 1 5703.7 5729.7
## - DPD30InLast12Months 1 5725.2 5751.2
## - DPD60InLast12Months 1 5726.1 5752.1
##
## Step: AIC=5649.21
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
## DPD60InLast6Months + DPD90InLast12Months + TradesInLast12Months +
## DPD90InLast6Months + InquiriesInLast12Months + OutstandingBalance +
## PLTradesInLast6Months + Income + months_in_current_company
##
## Df Deviance AIC
## - PLTradesInLast6Months 1 5624.2 5648.2
## - TradesInLast12Months 1 5624.2 5648.2
## <none> 5623.2 5649.2
## - OutstandingBalance 1 5631.7 5655.7
## - DPD30InLast6Months 1 5643.4 5667.4
## - DPD60InLast6Months 1 5657.6 5681.6
## - DPD90InLast6Months 1 5663.0 5687.0
## - DPD90InLast12Months 1 5673.2 5697.2
## - months_in_current_company 1 5683.9 5707.9
## - Income 1 5690.9 5714.9

```

```

## - InquiriesInLast12Months      1   5704.7 5728.7
## - DPD30InLast12Months          1   5726.3 5750.3
## - DPD60InLast12Months          1   5727.0 5751.0
##
## Step:  AIC=5648.26
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
##      DPD60InLast6Months + DPD90InLast12Months + TradesInLast12Months +
##      DPD90InLast6Months + InquiriesInLast12Months + OutstandingBalance +
##      Income + months_in_current_company
##
##              Df Deviance    AIC
## - TradesInLast12Months      1   5624.6 5646.6
## <none>                      5624.3 5648.3
## - OutstandingBalance        1   5632.4 5654.4
## - DPD30InLast6Months        1   5644.6 5666.6
## - DPD60InLast6Months        1   5658.8 5680.8
## - DPD90InLast6Months        1   5664.1 5686.1
## - DPD90InLast12Months       1   5674.5 5696.5
## - months_in_current_company 1   5685.1 5707.1
## - Income                    1   5691.8 5713.8
## - InquiriesInLast12Months   1   5705.6 5727.6
## - DPD30InLast12Months       1   5727.0 5749.0
## - DPD60InLast12Months       1   5728.0 5750.0
##
## Step:  AIC=5646.61
## Status_flag ~ DPD30InLast12Months + DPD30InLast6Months + DPD60InLast12Months +
##      DPD60InLast6Months + DPD90InLast12Months + DPD90InLast6Months +
##      InquiriesInLast12Months + OutstandingBalance + Income + months_in_current_comp
any
##
##              Df Deviance    AIC
## <none>                      5624.6 5646.6
## - OutstandingBalance        1   5635.8 5655.8
## - DPD30InLast6Months        1   5645.2 5665.2
## - DPD60InLast6Months        1   5659.2 5679.2
## - DPD90InLast6Months        1   5664.6 5684.6
## - DPD90InLast12Months       1   5675.0 5695.0
## - months_in_current_company 1   5685.4 5705.4
## - Income                    1   5692.2 5712.2
## - InquiriesInLast12Months   1   5706.0 5726.0
## - DPD30InLast12Months       1   5727.5 5747.5
## - DPD60InLast12Months       1   5728.3 5748.3

```

```
##
## Call: bayesglm(formula = Status_flag ~ DPD30InLast12Months + DPD30InLast6Months +
##      DPD60InLast12Months + DPD60InLast6Months + DPD90InLast12Months +
##      DPD90InLast6Months + InquiriesInLast12Months + OutstandingBalance +
##      Income + months_in_current_company, family = "binomial",
##      data = LogisticData_train)
##
## Coefficients:
##      (Intercept)          DPD30InLast12Months
##      -4.1304          0.5046
##      DPD30InLast6Months      DPD60InLast12Months
##      0.3029          0.5052
##      DPD60InLast6Months      DPD90InLast12Months
##      -0.3663          0.2954
##      DPD90InLast6Months      InquiriesInLast12Months
##      -0.3097          0.3698
##      OutstandingBalance      Income
##      0.2008          0.4018
##      months_in_current_company
##      0.4214
##
## Degrees of Freedom: 49903 Total (i.e. Null); 49893 Residual
## Null Deviance: 9777
## Residual Deviance: 5625 AIC: 5647
```

```
model_2 <- bayesglm(formula = Status_flag ~ DPD30InLast12Months + DPD30InLast6Months
+
      DPD60InLast12Months + DPD60InLast6Months + DPD90InLast12Month
s +
      DPD90InLast6Months + InquiriesInLast12Months + OutstandingBal
ance +
      Income + months_in_current_company , family = "binomial",
data = LogisticData_train)

summary(model_2)
```



```
##
## Call:
## bayesglm(formula = Status_flag ~ DPD30InLast12Months + DPD30InLast6Months +
##   DPD60InLast12Months + DPD60InLast6Months + DPD90InLast12Months +
##   DPD90InLast6Months + InquiriesInLast12Months + OutstandingBalance +
##   Income + months_in_current_company, family = "binomial",
##   data = LogisticData_train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4734  -0.0660  -0.0102  -0.0048   4.3670
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.13041    0.08110  -50.932  < 2e-16 ***
## DPD30InLast12Months    0.50456    0.05182   9.737  < 2e-16 ***
## DPD30InLast6Months    0.30287    0.06760   4.480 7.46e-06 ***
## DPD60InLast12Months    0.50525    0.05222   9.675  < 2e-16 ***
## DPD60InLast6Months   -0.36626    0.06212  -5.896 3.72e-09 ***
## DPD90InLast12Months    0.29537    0.04332   6.818 9.24e-12 ***
## DPD90InLast6Months   -0.30970    0.04731  -6.546 5.92e-11 ***
## InquiriesInLast12Months  0.36984    0.04851   7.624 2.47e-14 ***
## OutstandingBalance    0.20080    0.06361   3.157  0.0016 **
## Income            0.40181    0.04924   8.160 3.34e-16 ***
## months_in_current_company 0.42137    0.05417   7.779 7.32e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 9776.7  on 49903  degrees of freedom
## Residual deviance: 5624.6  on 49893  degrees of freedom
## AIC: 5646.6
##
## Number of Fisher Scoring iterations: 11
```

```
model_3 <- bayesglm(formula = Status_flag ~ DPD30InLast12Months + DPD30InLast6Months
+
      DPD60InLast12Months + DPD60InLast6Months + DPD90InLast12Month
s +
      DPD90InLast6Months + InquiriesInLast12Months +
      Income + months_in_current_company , family = "binomial",
data = LogisticData_train)

summary(model_3)
```

```
##
## Call:
## bayesglm(formula = Status_flag ~ DPD30InLast12Months + DPD30InLast6Months +
##   DPD60InLast12Months + DPD60InLast6Months + DPD90InLast12Months +
##   DPD90InLast6Months + InquiriesInLast12Months + Income + months_in_current_comp
any,
##   family = "binomial", data = LogisticData_train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4613  -0.0651  -0.0114  -0.0059   4.3879
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.05247    0.07531  -53.812  < 2e-16 ***
## DPD30InLast12Months    0.50516    0.05178   9.756  < 2e-16 ***
## DPD30InLast6Months    0.30748    0.06736   4.565 5.00e-06 ***
## DPD60InLast12Months    0.50655    0.05221   9.703  < 2e-16 ***
## DPD60InLast6Months   -0.36803    0.06217  -5.920 3.22e-09 ***
## DPD90InLast12Months    0.29875    0.04335   6.891 5.53e-12 ***
## DPD90InLast6Months   -0.31241    0.04728  -6.608 3.89e-11 ***
## InquiriesInLast12Months  0.37505    0.04838   7.752 9.07e-15 ***
## Income            0.40698    0.04918   8.275  < 2e-16 ***
## months_in_current_company 0.42442    0.05414   7.839 4.54e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 9776.7  on 49903  degrees of freedom
## Residual deviance: 5635.9  on 49894  degrees of freedom
## AIC: 5655.9
##
## Number of Fisher Scoring iterations: 11
```

Found all significant attributes for building scorecard of applicant

using logisitic regression. Will be computing scorecard using model\_3

```
final_logit_model <- model_3
ScoreCard_logit_model <- model_3
```

Removing Status\_flag while performing prediction

```
predictions_logit <- predict(final_logit_model, newdata = LogisticData_test[, -2],
                             type = "response")

predictions_logit <- round(predictions_logit, 2)
summary(predictions_logit)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.02034 0.00000 0.57000
```

```
actual_response <- factor(ifelse(LogisticData_test$Status_flag == "1", "Rejected", "Approved"))
```

## Function to derive optimal cutoff value

```
perform_fn <- function(cutoff)
{
  predicted_attrition <- factor(ifelse(predictions_logit >= cutoff, "Rejected", "Approved"))
  conf <- caret::confusionMatrix(predicted_attrition, actual_response, positive = "Rejected")
  acc <- conf$overall[1]
  sens <- conf$byClass[1]
  spec <- conf$byClass[2]
  out <- t(as.matrix(c(sens, spec, acc)))
  colnames(out) <- c("sensitivity", "specificity", "accuracy")
  return(out)
}

s = seq(.01, .80, length=100)
OUT = matrix(0, 100, 3)

for(i in 1:100)
{
  OUT[i,] = perform_fn(s[i])
}
```

```
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :  
## Levels are not in the same order for reference and data. Refactoring data  
## to match.  
  
## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
```

```
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.

## Warning in confusionMatrix.default(predicted_attrition, actual_response, :
## Levels are not in the same order for reference and data. Refactoring data
## to match.
```

## Plot graph for “Sensitivity”, “Specificity”, “Accuracy”

```
dev.off()
```

```
## null device  
##          1
```

```
plot(s, OUT[,1],xlab="Cutoff",ylab="Value",cex.lab=1.5,cex.axis=1.5,ylim=c(0,1),type="l",lwd=2,axes=FALSE,col=2)  
axis(1,seq(0,1,length=5),seq(0,1,length=5),cex.lab=1.5)  
axis(2,seq(0,1,length=5),seq(0,1,length=5),cex.lab=1.5)  
lines(s,OUT[,2],col="darkgreen",lwd=2)  
lines(s,OUT[,3],col=4,lwd=2)  
box()  
legend(0,.50,col=c(2,"darkgreen",4,"darkred"),lwd=c(2,2,2,2),c("Sensitivity","Specificity","Accuracy"))
```

## Obtained cutoff for optimal values of sensitivity, specificity and accuracy of model

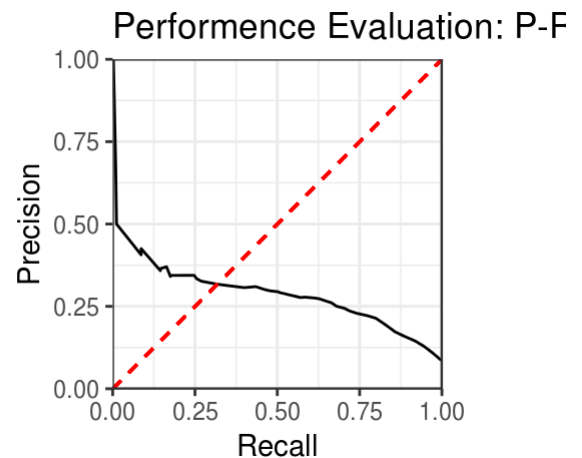
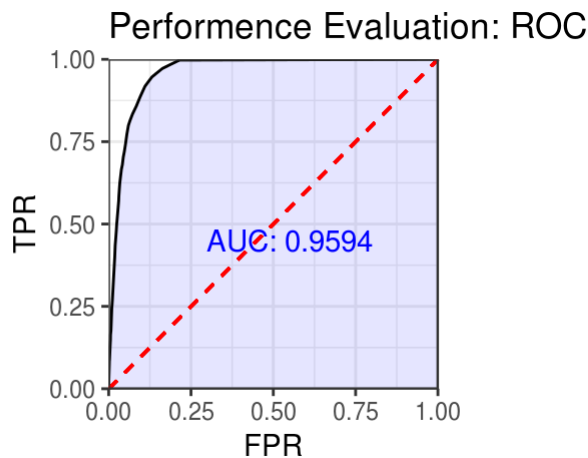
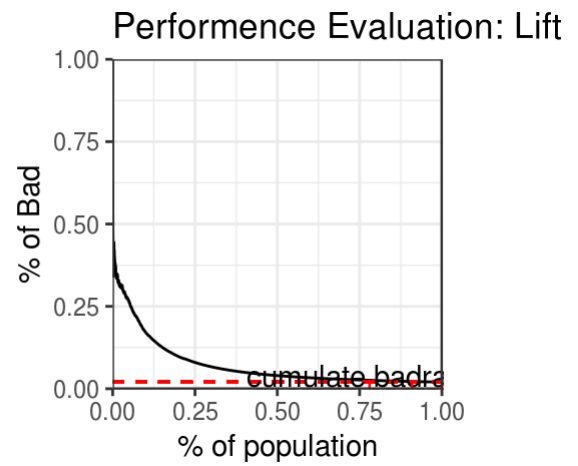
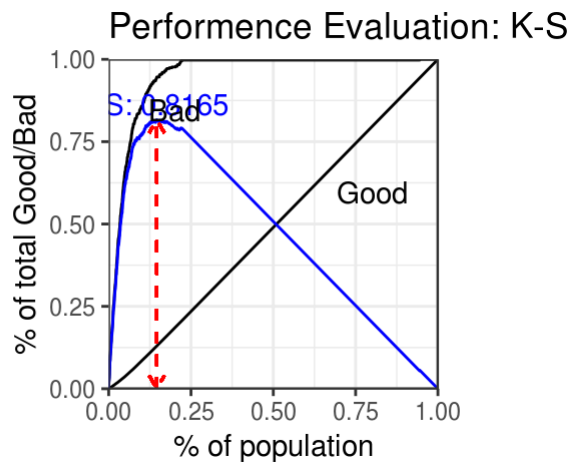
```
cutoff <- s[which(abs(OUT[,1]-OUT[,2])<0.02)]  
test_cutoff_attrition <- factor(ifelse(predictions_logit >=cutoff, "Rejected", "Approved"))  
conf_final <- caret::confusionMatrix(test_cutoff_attrition, actual_response, positive = "Rejected")  
conf_final
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction Approved Rejected
##   Approved    18961      49
##   Rejected     1999     379
##
##           Accuracy : 0.9042
##           95% CI : (0.9002, 0.9082)
##   No Information Rate : 0.98
##   P-Value [Acc > NIR] : 1
##
##           Kappa : 0.2445
## Mcnemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.88551
##           Specificity : 0.90463
##           Pos Pred Value : 0.15938
##           Neg Pred Value : 0.99742
##           Prevalence : 0.02001
##           Detection Rate : 0.01772
##   Detection Prevalence : 0.11118
##           Balanced Accuracy : 0.89507
##
##           'Positive' Class : Rejected
##
```

## generaring KS, lift chart, ROC curve and positive rate

```
perf_eva(LogisticData_test$Status_flag, predictions_logit,
         title = "Performence Evaluation", show_plot = TRUE,
         type = c("ks","lift","roc","pr"))
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```



```
## $KS
## [1] 0.8165
##
## $AUC
## [1] 0.9594
##
## $Gini
## [1] 0.9189
##
## $pic
## TableGrob (2 x 2) "arrange": 4 grobs
##       z      cells      name      grob
## pks    1 (1-1,1-1) arrange gtable[layout]
## plift  2 (1-1,2-2) arrange gtable[layout]
## proc   3 (2-2,1-1) arrange gtable[layout]
## ppr    4 (2-2,2-2) arrange gtable[layout]
```

## Scorecard building using logistic Regression

```
library(car)
library(MASS)
library(arm)
library(caTools)
library(scorecard)
```



# Building score card for demographic data and credit bureau data

```
CreditBureau_Scorecard <- scorecard::scorecard(CreditHistory_woebins, ScoreCard_logit_model)
Demographic_scorecard <- scorecard::scorecard(Demographic_data_woebins, ScoreCard_logit_model)
```

## Generating scores for individual attributes for demographic data and credit bureau data

```
Demographic_score = scorecard_ply(LogisticData, Demographic_scorecard, only_total_score = FALSE, print_step=0)
```

```
## Warning in rmcol_datetime_unique1(dt): There are 1 columns have only one unique values, which are removed from input dataset.
## (ColumnNames: PresenceOfOpenLoan)
```

```
CreditBureau_score = scorecard_ply(LogisticData, CreditBureau_Scorecard, only_total_score = FALSE, print_step=0)
```

```
## Warning in rmcol_datetime_unique1(dt): There are 1 columns have only one unique values, which are removed from input dataset.
## (ColumnNames: PresenceOfOpenLoan)
```

## Calculating Total score of applicant

```
Demographic_score <- Demographic_score$Income_points + Demographic_score$months_in_current_company_points
Creditbureau_score <- CreditBureau_score$DPD30InLast12Months_points +
  CreditBureau_score$DPD30InLast6Months_points +
  CreditBureau_score$DPD60InLast12Months_points +
  CreditBureau_score$DPD60InLast6Months_points +
  CreditBureau_score$DPD90InLast12Months_points +
  CreditBureau_score$DPD90InLast6Months_points +
  CreditBureau_score$InquiriesInLast12Months_points
```

## Visualizing score generated

```
ApplicantData$Total_Score <- Demographic_score + Creditbureau_score

# Histogram presentation of Applicant scores
qplot(ApplicantData$Total_Score, geom = "histogram",
      main = "Histogram of Applicant ScoreCard",
      fill = ApplicantData$Status_flag)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

