

# credit\_german.R

DeepsikhaSarma

Sun Jan 13 21:57:50 2023

```
#Load Libraries
loadlibraries <- c("Hmisc","dplyr","ggplot2","caret","pROC","ROCR","MASS","Information","ggthemes",
                  "visNetwork","rpart","sparkline","data.table")
installlib <- loadlibraries[!loadlibraries %in% installed.packages()]
for(libs in installlib) install.packages(libs, dependencies = TRUE)
sapply(loadlibraries, require, character = TRUE)
```

```
## Loading required package: Hmisc
```

```
## Warning: package 'Hmisc' was built under R version 3.5.2
```

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

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##   format.pval, units
```

```
## Loading required package: dplyr
```

```
## Warning: package 'dplyr' was built under R version 3.5.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:Hmisc':  
##  
##   src, summarize
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
## Loading required package: caret
```

```
##  
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:survival':  
##  
##   cluster
```

```
## Loading required package: pROC
```

```
## Type 'citation("pROC")' for a citation.
```

```
##  
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':  
##  
##   cov, smooth, var
```

```
## Loading required package: ROCR
```

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

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

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

```
## Loading required package: MASS
```

```
##  
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':  
##  
##      select
```

```
## Loading required package: Information
```

```
## Loading required package: ggthemes
```

```
## Loading required package: visNetwork
```

```
## Warning: package 'visNetwork' was built under R version 3.5.2
```

```
## Loading required package: rpart
```

```
##  
## Attaching package: 'rpart'
```

```
## The following object is masked from 'package:survival':  
##  
##      solder
```

```
## Loading required package: sparkline
```

```
## Warning: package 'sparkline' was built under R version 3.5.2
```

```
## Loading required package: data.table
```

```
##  
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':  
##  
##      between, first, last
```

```
##      Hmisc      dplyr      ggplot2      caret      pROC      ROCR
##      TRUE       TRUE       TRUE       TRUE       TRUE       TRUE
##      MASS Information ggthemes visNetwork rpart  sparkline
##      TRUE       TRUE       TRUE       TRUE       TRUE       TRUE
## data.table
##      TRUE
```

```
#Load the file
setwd("C:/Users/dwaip/Desktop/")
df <- read.csv("german_credit.csv", stringsAsFactors = F)
prop.table(table(df$Creditability))
```

```
##
##      0      1
## 0.3 0.7
```

```
#Checking the Data
describe(df)
```

```
## df
##
## 21 Variables      1000 Observations
## -----
## Creditability
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0        2      0.63      700      0.7      0.4204
##
## -----
## Account.Balance
##      n missing distinct      Info      Mean      Gmd
##    1000      0        4      0.899      2.577      1.373
##
## Value      1      2      3      4
## Frequency  274  269   63  394
## Proportion 0.274 0.269 0.063 0.394
## -----
## Duration.of.Credit..month.
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1000      0        33      0.985      20.9      12.98        6        9
##      .25      .50      .75      .90      .95
##      12      18      24      36      48
##
## lowest :  4  5  6  7  8, highest: 47 48 54 60 72
## -----
## Payment.Status.of.Previous.Credit
##      n missing distinct      Info      Mean      Gmd
##    1000      0        5      0.825      2.545      1.126
##
## Value      0      1      2      3      4
## Frequency   40   49  530   88  293
## Proportion 0.040 0.049 0.530 0.088 0.293
## -----
## Purpose
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1000      0        10      0.957      2.828      2.852        0        0
##      .25      .50      .75      .90      .95
##      1      2      3      9      9
##
## Value      0      1      2      3      4      5      6      8      9     10
## Frequency  234  103  181  280  12   22   50   9   97   12
## Proportion 0.234 0.103 0.181 0.280 0.012 0.022 0.050 0.009 0.097 0.012
## -----
## Credit.Amount
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1000      0      923        1      3271      2773      709.0      934.7
##      .25      .50      .75      .90      .95
##    1365.5  2319.5  3972.2  7179.4  9162.7
##
## lowest :   250   276   338   339   343, highest: 15653 15672 15857 15945 18424
## -----
## Value.Savings.Stocks
```

```
##          n missing distinct      Info      Mean      Gmd
##      1000          0          5      0.773      2.105      1.55
##
## Value          1      2      3      4      5
## Frequency      603     103     63     48     183
## Proportion 0.603 0.103 0.063 0.048 0.183
## -----
## Length.of.current.employment
##          n missing distinct      Info      Mean      Gmd
##      1000          0          5      0.934      3.384      1.343
##
## Value          1      2      3      4      5
## Frequency       62     172     339     174     253
## Proportion 0.062 0.172 0.339 0.174 0.253
## -----
## Instalment.per.cent
##          n missing distinct      Info      Mean      Gmd
##      1000          0          4      0.873      2.973      1.2
##
## Value          1      2      3      4
## Frequency      136     231     157     476
## Proportion 0.136 0.231 0.157 0.476
## -----
## Sex...Marital.Status
##          n missing distinct      Info      Mean      Gmd
##      1000          0          4      0.805      2.682      0.7236
##
## Value          1      2      3      4
## Frequency       50     310     548     92
## Proportion 0.050 0.310 0.548 0.092
## -----
## Guarantors
##          n missing distinct      Info      Mean      Gmd
##      1000          0          3      0.254      1.145      0.2676
##
## Value          1      2      3
## Frequency      907     41     52
## Proportion 0.907 0.041 0.052
## -----
## Duration.in.Current.address
##          n missing distinct      Info      Mean      Gmd
##      1000          0          4      0.895      2.845      1.205
##
## Value          1      2      3      4
## Frequency      130     308     149     413
## Proportion 0.130 0.308 0.149 0.413
## -----
## Most.valuable.available.asset
##          n missing distinct      Info      Mean      Gmd
##      1000          0          4      0.925      2.358      1.166
##
## Value          1      2      3      4
```

```

## Frequency      282   232   332   154
## Proportion 0.282 0.232 0.332 0.154
## -----
## Age..years.
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    1000      0      53    0.999    35.54    12.38      22      23
##      .25      .50      .75      .90      .95
##      27      33      42      52      60
##
## lowest : 19 20 21 22 23, highest: 67 68 70 74 75
## -----
## Concurrent.Credits
##      n missing distinct      Info      Mean      Gmd
##    1000      0      3    0.458    2.675    0.5427
##
## Value      1      2      3
## Frequency   139    47   814
## Proportion 0.139 0.047 0.814
## -----
## Type.of.apartment
##      n missing distinct      Info      Mean      Gmd
##    1000      0      3    0.629    1.928    0.4855
##
## Value      1      2      3
## Frequency   179   714   107
## Proportion 0.179 0.714 0.107
## -----
## No.of.Credits.at.this.Bank
##      n missing distinct      Info      Mean      Gmd
##    1000      0      4    0.709    1.407    0.5428
##
## Value      1      2      3      4
## Frequency   633   333    28      6
## Proportion 0.633 0.333 0.028 0.006
## -----
## Occupation
##      n missing distinct      Info      Mean      Gmd
##    1000      0      4    0.739    2.904    0.6413
##
## Value      1      2      3      4
## Frequency    22   200   630   148
## Proportion 0.022 0.200 0.630 0.148
## -----
## No.of.dependents
##      n missing distinct      Info      Mean      Gmd
##    1000      0      2    0.393    1.155    0.2622
##
## Value      1      2
## Frequency   845   155
## Proportion 0.845 0.155
## -----
## Telephone

```

```
##          n missing distinct      Info      Mean      Gmd
##      1000          0          2      0.722      1.404      0.4821
##
## Value          1      2
## Frequency      596      404
## Proportion 0.596 0.404
## -----
## Foreign.Worker
##          n missing distinct      Info      Mean      Gmd
##      1000          0          2      0.107      1.037      0.07133
##
## Value          1      2
## Frequency      963      37
## Proportion 0.963 0.037
## -----
```

```
str(df)
```

```
## 'data.frame':  1000 obs. of  21 variables:
## $ Creditability          : int  1 1 1 1 1 1 1 1 1 1 ...
## $ Account.Balance        : int  1 1 2 1 1 1 1 1 4 2 ...
## $ Duration.of.Credit..month. : int  18 9 12 12 12 10 8 6 18 24 ...
## $ Payment.Status.of.Previous.Credit: int  4 4 2 4 4 4 4 4 2 ...
## $ Purpose                : int  2 0 9 0 0 0 0 0 3 3 ...
## $ Credit.Amount          : int  1049 2799 841 2122 2171 2241 3398 1361 1098 3758
## ...
## $ Value.Savings.Stocks    : int  1 1 2 1 1 1 1 1 1 3 ...
## $ Length.of.current.employment : int  2 3 4 3 3 2 4 2 1 1 ...
## $ Instalment.per.cent     : int  4 2 2 3 4 1 1 2 4 1 ...
## $ Sex...Marital.Status    : int  2 3 2 3 3 3 3 3 2 2 ...
## $ Guarantors              : int  1 1 1 1 1 1 1 1 1 1 ...
## $ Duration.in.Current.address : int  4 2 4 2 4 3 4 4 4 4 ...
## $ Most.valuable.available.asset : int  2 1 1 1 2 1 1 1 3 4 ...
## $ Age..years.             : int  21 36 23 39 38 48 39 40 65 23 ...
## $ Concurrent.Credits      : int  3 3 3 3 1 3 3 3 3 3 ...
## $ Type.of.apartment       : int  1 1 1 1 2 1 2 2 2 1 ...
## $ No.of.Credits.at.this.Bank : int  1 2 1 2 2 2 2 1 2 1 ...
## $ Occupation              : int  3 3 2 2 2 2 2 2 1 1 ...
## $ No.of.dependents        : int  1 2 1 2 1 2 1 2 1 1 ...
## $ Telephone               : int  1 1 1 1 1 1 1 1 1 1 ...
## $ Foreign.Worker          : int  1 1 1 2 2 2 2 2 1 1 ...
```

```
names(df)
```



```
## [1] "Creditability"
## [2] "Account.Balance"
## [3] "Duration.of.Credit..month."
## [4] "Payment.Status.of.Previous.Credit"
## [5] "Purpose"
## [6] "Credit.Amount"
## [7] "Value.Savings.Stocks"
## [8] "Length.of.current.employment"
## [9] "Instalment.per.cent"
## [10] "Sex...Marital.Status"
## [11] "Guarantors"
## [12] "Duration.in.Current.address"
## [13] "Most.valuable.available.asset"
## [14] "Age..years."
## [15] "Concurrent.Credits"
## [16] "Type.of.apartment"
## [17] "No.of.Credits.at.this.Bank"
## [18] "Occupation"
## [19] "No.of.dependents"
## [20] "Telephone"
## [21] "Foreign.Worker"
```

```
nrow(df)
```

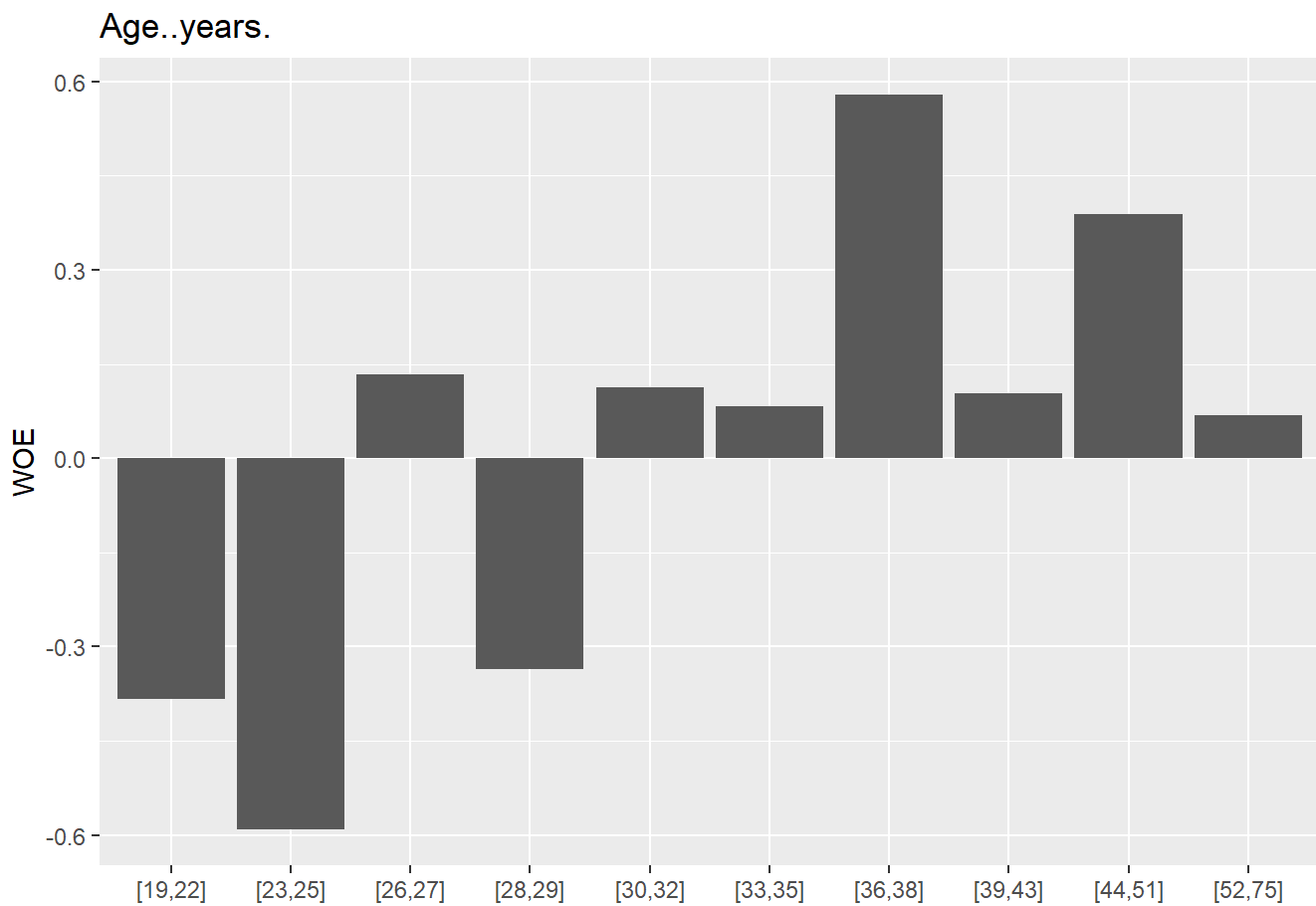
```
## [1] 1000
```

```
# Splitting the data for model
#Splitting the data into train and test
dt = sort(sample(nrow(df), nrow(df)*.7))
train<-df[dt,]
test<-df[-dt,]
rm(list=ls()[! ls() %in% c("train","test","df")])

#Information Value ::
IV <- create_infotables(data=df, y="Creditability", bins=10, parallel=FALSE)
IV_Value = data.frame(IV$Summary)
# Age and Amount IV tables
print(IV$Tables$Age..years., row.names=FALSE)
```

##	Age..years.	N	Percent	WOE	IV
##	[19,22]	57	0.057	-0.38299225	0.008936486
##	[23,25]	133	0.133	-0.59025276	0.059810652
##	[26,27]	99	0.099	0.13353139	0.061527484
##	[28,29]	80	0.080	-0.33647224	0.071140977
##	[30,32]	112	0.112	0.11316409	0.072542056
##	[33,35]	106	0.106	0.08223810	0.073246954
##	[36,38]	93	0.093	0.57981850	0.100581254
##	[39,43]	104	0.104	0.10289442	0.101659196
##	[44,51]	111	0.111	0.38817361	0.117001296
##	[52,75]	105	0.105	0.06899287	0.117494102

```
Age = data.frame(IV$Tables$Age..years.)
plot_infotables(IV, "Age..years.")
```

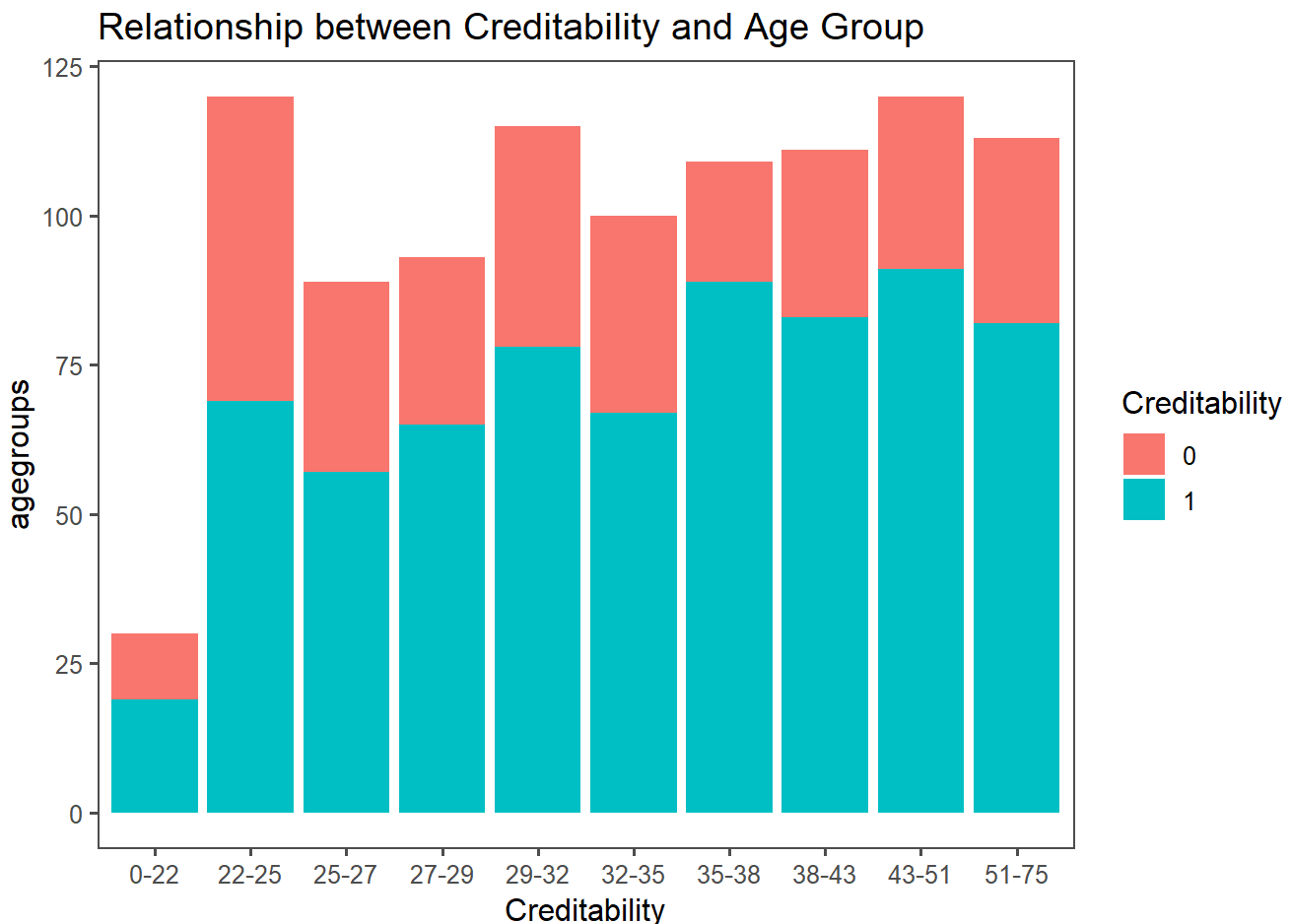


```

#Age Vs Creditability
agebreaks <- c(0,22,25,27,29,32,35,38,43,51,95)
agelabels <- c("0-22","22-25","25-27","27-29","29-32","32-35","35-38","38-43",
               "43-51","51-75")
setDT(df)[ , agegroups := cut(Age..years.,
                              breaks = agebreaks,
                              right = FALSE,
                              labels = agelabels)]

default_cred<- ggplot(df, aes(agegroups, fill = factor(Creditability))) +
  geom_bar() + theme_few() + xlab("Creditability") + ylab("agegroups") +
  scale_fill_discrete(name = "Creditability") +theme_few()+ ggtitle("Relationship between Credit
ability and Age Group")
default_cred

```



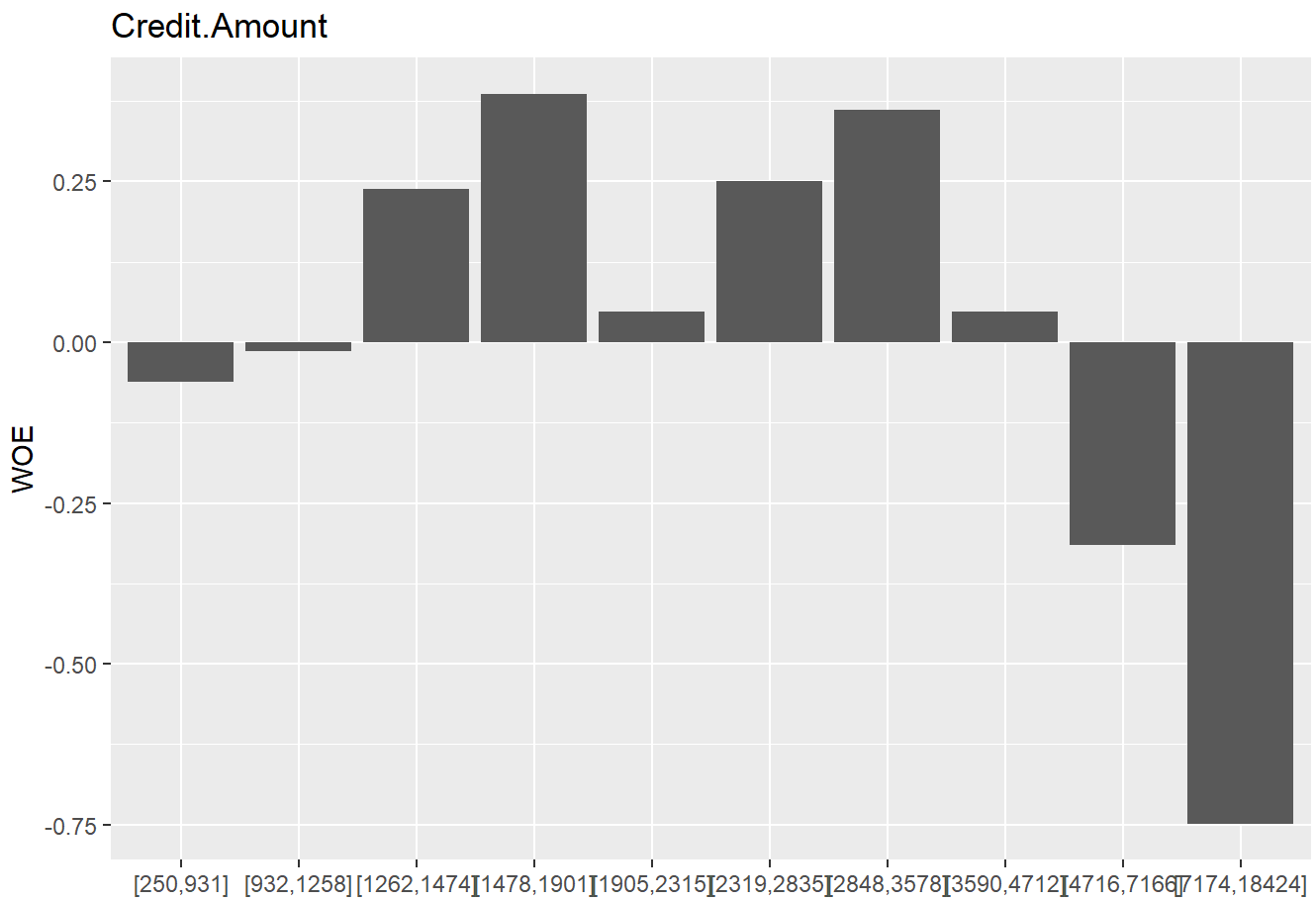
```

# Amount IV tables
print(IV$Tables$Credit.Amount, row.names=FALSE)

```

##	Credit.Amount	N	Percent	WOE	IV
##	[250,931]	99	0.099	-0.06177736	0.0003824313
##	[932,1258]	99	0.099	-0.01438874	0.0004029866
##	[1262,1474]	99	0.099	0.23789141	0.0057272229
##	[1478,1901]	102	0.102	0.38665578	0.0197204795
##	[1905,2315]	100	0.100	0.04808619	0.0199494614
##	[2319,2835]	100	0.100	0.25131443	0.0259331383
##	[2848,3578]	100	0.100	0.36101335	0.0379669164
##	[3590,4712]	100	0.100	0.04808619	0.0381958983
##	[4716,7166]	100	0.100	-0.31508105	0.0486985998
##	[7174,18424]	101	0.101	-0.74820696	0.1117617577

```
Amount = data.frame(IV$Tables$Credit.Amount)  
plot_infotables(IV, "Credit.Amount")
```

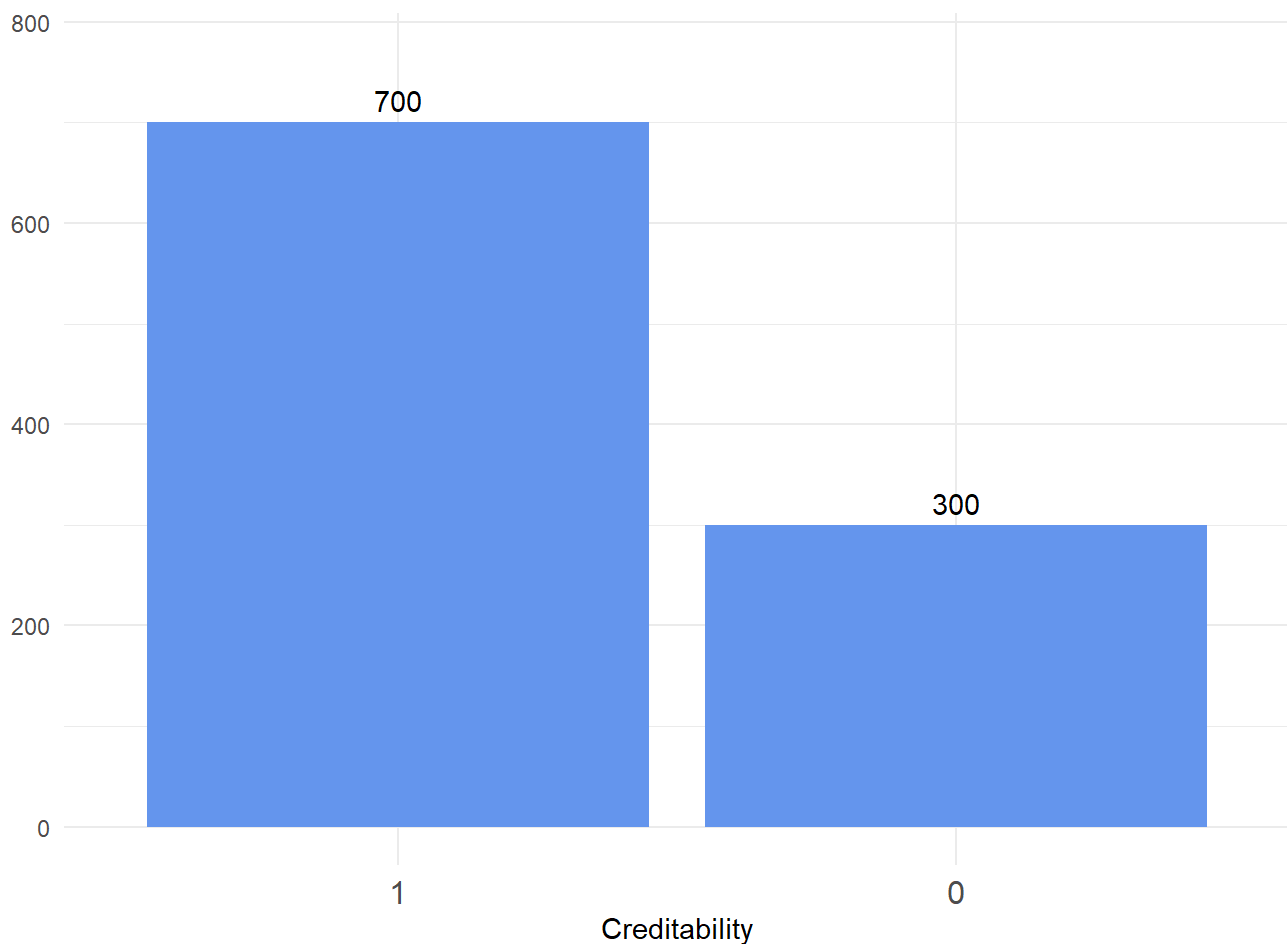


```
# Changing to factor
df <- as.data.frame(df)
F=c(1,2,4,5,7,8,9,10,11,12,13,15,16,17,18,19,20)
for(i in F) df[,i]=as.factor(df[,i])

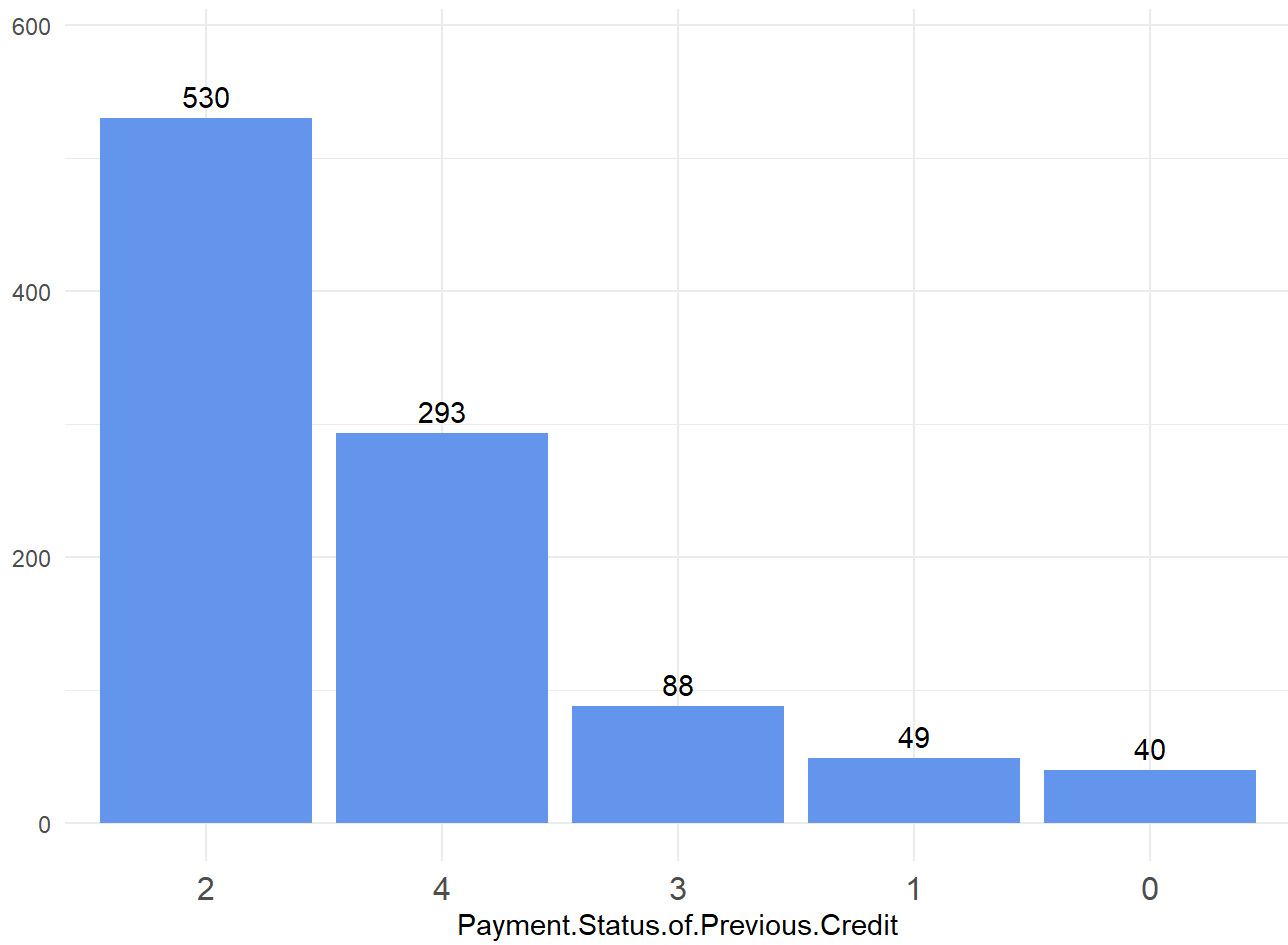
plot.categoric <- function(cols, df){
  for (col in cols) {
    order.cols <- names(sort(table(df[,col]), decreasing = TRUE))
    num.plot <- qplot(df[,col]) +
      geom_bar(fill = 'cornflowerblue') +
      geom_text(aes(label = ..count..), stat='count', vjust=-0.5) +
      theme_minimal() +
      scale_y_continuous(limits = c(0,max(table(df[,col]))*1.1)) +
      scale_x_discrete(limits = order.cols) +
      xlab(col) +
      theme(axis.text.x = element_text(angle = 0, size=12))

    print(num.plot)
  }
}

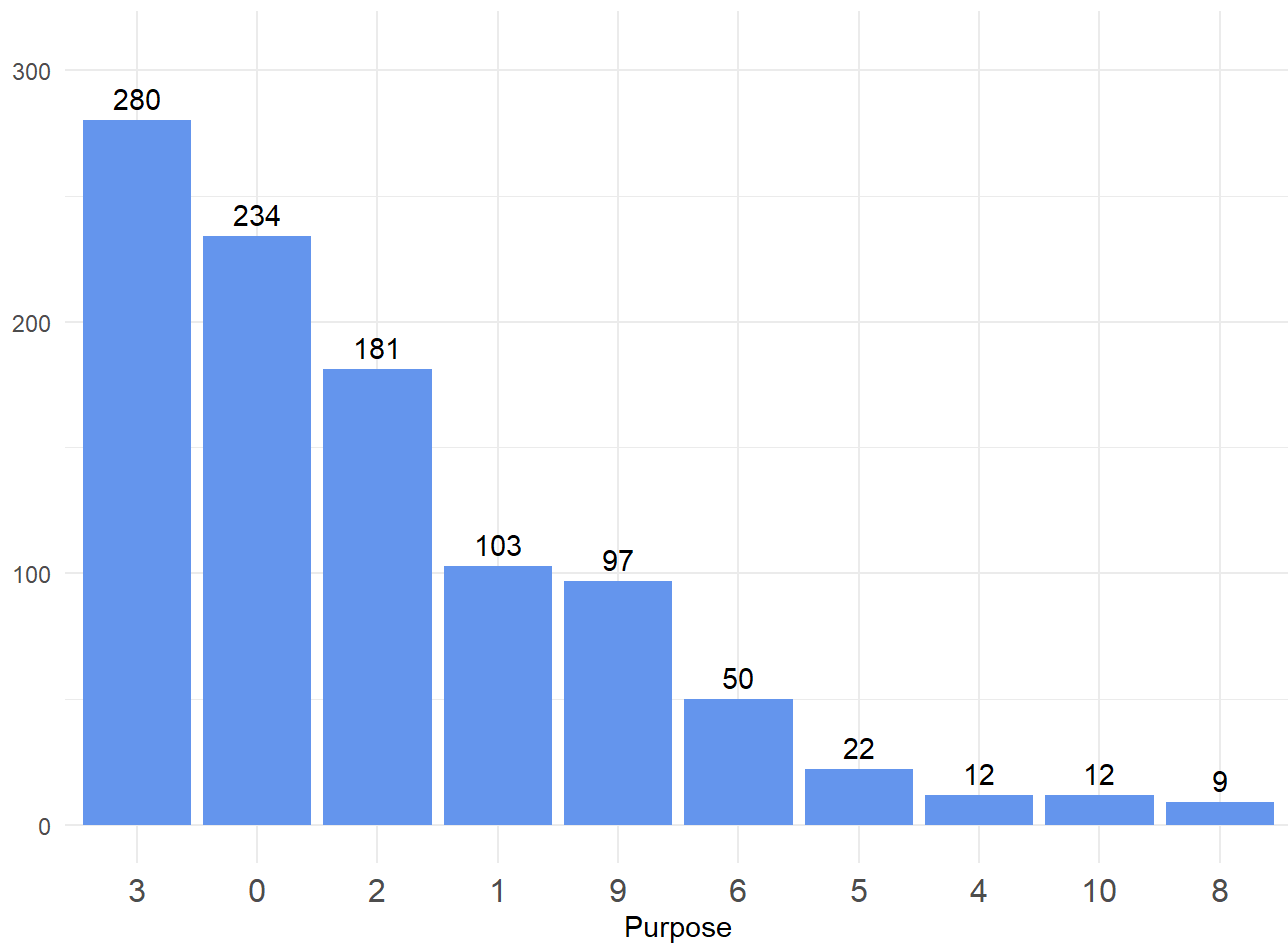
plot.categoric("Creditability", df)
```



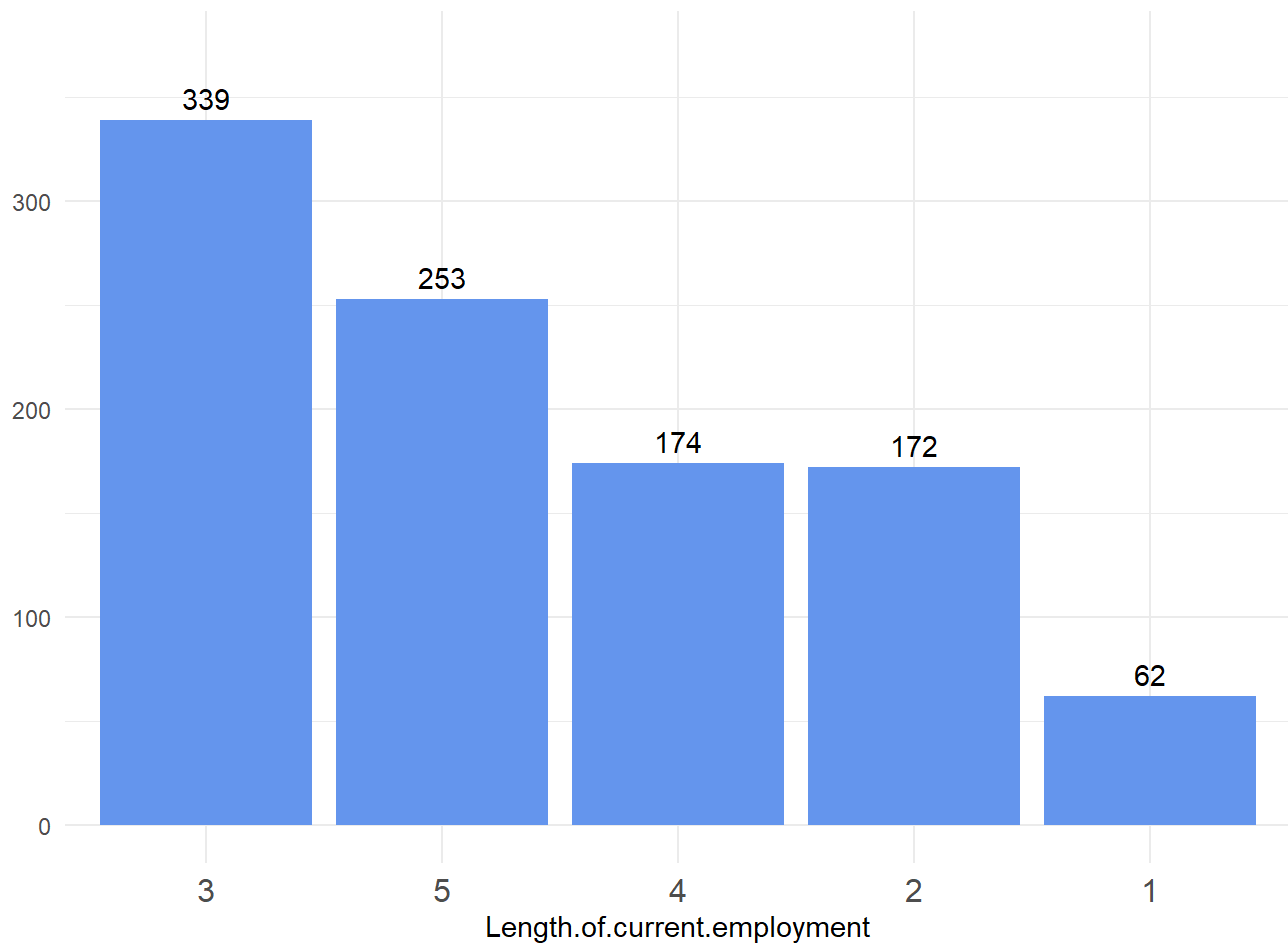
```
plot.categoric("Payment.Status.of.Previous.Credit",df)
```



```
plot.categoric("Purpose",df)
```

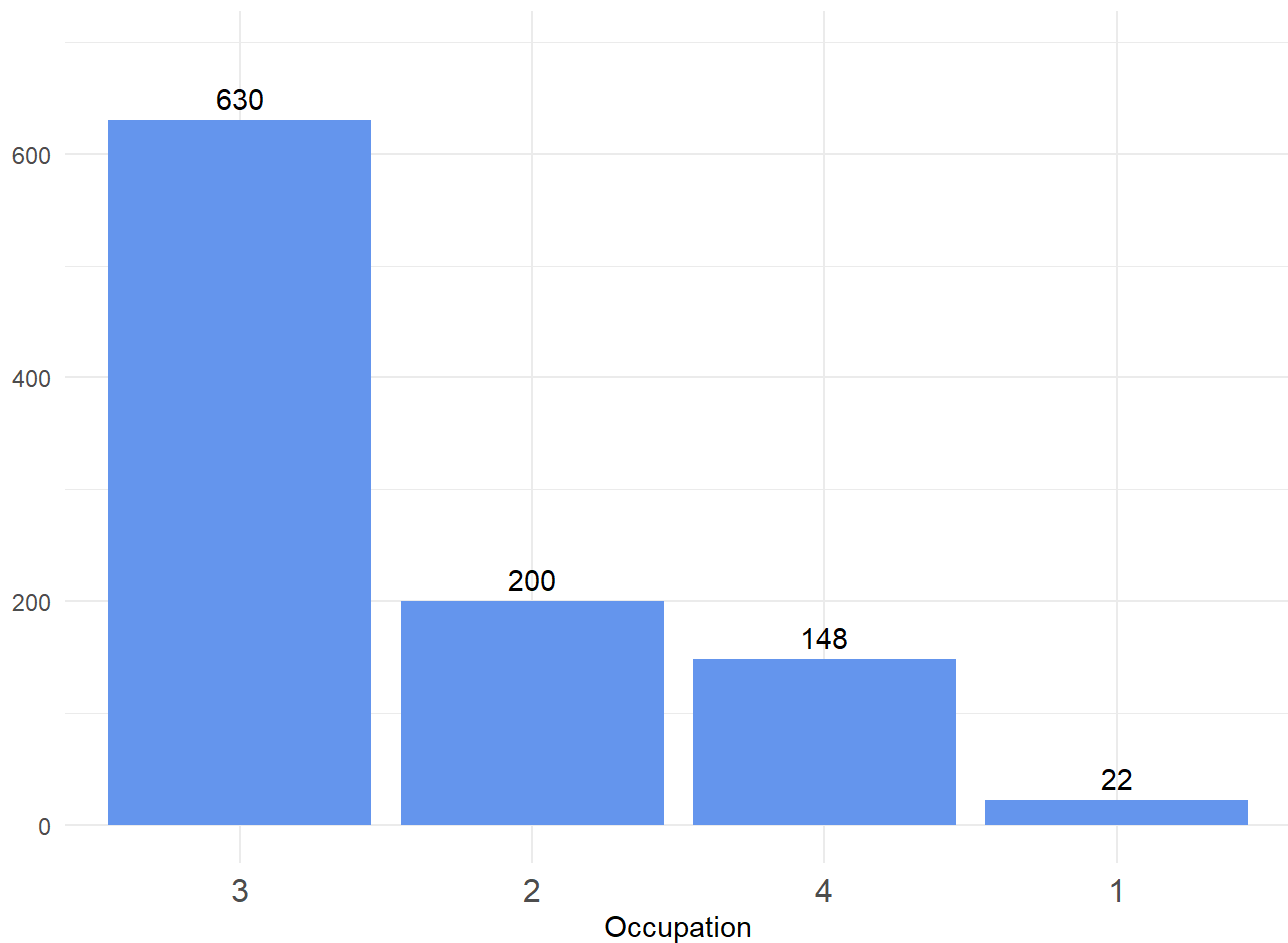


```
plot.categoric("Length.of.current.employment" ,df)
```



```
plot.categoric("Occupation" ,df)
```





```
# Variables
```

```
var = names(train[,!(names(train) %in% c("Creditability"))])
```

```
var
```

```
## [1] "Account.Balance"
## [2] "Duration.of.Credit..month."
## [3] "Payment.Status.of.Previous.Credit"
## [4] "Purpose"
## [5] "Credit.Amount"
## [6] "Value.Savings.Stocks"
## [7] "Length.of.current.employment"
## [8] "Instalment.per.cent"
## [9] "Sex...Marital.Status"
## [10] "Guarantors"
## [11] "Duration.in.Current.address"
## [12] "Most.valuable.available.asset"
## [13] "Age..years."
## [14] "Concurrent.Credits"
## [15] "Type.of.apartment"
## [16] "No.of.Credits.at.this.Bank"
## [17] "Occupation"
## [18] "No.of.dependents"
## [19] "Telephone"
## [20] "Foreign.Worker"
```

```
variables<-reformulate(var, response = 'Creditability')
variables
```

```
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##   Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##   Sex...Marital.Status + Guarantors + Duration.in.Current.address +
##   Most.valuable.available.asset + Age..years. + Concurrent.Credits +
##   Type.of.apartment + No.of.Credits.at.this.Bank + Occupation +
##   No.of.dependents + Telephone + Foreign.Worker
```

```
#Logistic model
# Null Model First
fit1 <- glm(variables, data = train, family = binomial)
coef(fit1)
```

```
##              (Intercept)              Account.Balance
##              -3.982328e+00              6.002988e-01
##      Duration.of.Credit..month. Payment.Status.of.Previous.Credit
##              -3.004832e-02              3.973217e-01
##              Purpose              Credit.Amount
##              3.974563e-02              -5.997753e-05
##      Value.Savings.Stocks      Length.of.current.employment
##              2.112958e-01              2.001166e-01
##      Instalment.per.cent      Sex...Marital.Status
##              -2.043488e-01              2.525005e-01
##      Guarantors      Duration.in.Current.address
##              2.875814e-01              -4.080617e-02
##      Most.valuable.available.asset      Age..years.
##              -2.239649e-01              1.043115e-02
##      Concurrent.Credits      Type.of.apartment
##              2.545348e-01              4.207910e-01
##      No.of.Credits.at.this.Bank      Occupation
##              -2.223407e-01              1.906004e-02
##      No.of.dependents      Telephone
##              -6.603314e-01              1.332803e-01
##      Foreign.Worker
##              1.327527e+00
```

```
step_mod <- fit1 %>% stepAIC(trace = TRUE)
```

```
## Start: AIC=707.02
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##   Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##   Sex...Marital.Status + Guarantors + Duration.in.Current.address +
##   Most.valuable.available.asset + Age..years. + Concurrent.Credits +
##   Type.of.apartment + No.of.Credits.at.this.Bank + Occupation +
##   No.of.dependents + Telephone + Foreign.Worker
##
##
```

	Df	Deviance	AIC
## - Occupation	1	665.04	705.04
## - Duration.in.Current.address	1	665.22	705.22
## - Telephone	1	665.36	705.36
## - Age..years.	1	666.09	706.09
## - Purpose	1	666.18	706.18
## - No.of.Credits.at.this.Bank	1	666.33	706.33
## - Credit.Amount	1	666.64	706.64
## - Guarantors	1	667.01	707.01
## <none>		665.02	707.02
## - Sex...Marital.Status	1	668.33	708.33
## - Concurrent.Credits	1	668.69	708.69
## - Most.valuable.available.asset	1	669.17	709.17
## - Foreign.Worker	1	669.27	709.27
## - Type.of.apartment	1	669.45	709.45
## - Instalment.per.cent	1	669.51	709.51
## - Length.of.current.employment	1	670.50	710.50
## - No.of.dependents	1	670.57	710.57
## - Duration.of.Credit..month.	1	673.47	713.47
## - Value.Savings.Stocks	1	674.88	714.88
## - Payment.Status.of.Previous.Credit	1	679.40	719.40
## - Account.Balance	1	721.56	761.56

```
##
## Step: AIC=705.04
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##   Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##   Sex...Marital.Status + Guarantors + Duration.in.Current.address +
##   Most.valuable.available.asset + Age..years. + Concurrent.Credits +
##   Type.of.apartment + No.of.Credits.at.this.Bank + No.of.dependents +
##   Telephone + Foreign.Worker
##
##
```

	Df	Deviance	AIC
## - Duration.in.Current.address	1	665.25	703.25
## - Telephone	1	665.46	703.46
## - Age..years.	1	666.11	704.11
## - Purpose	1	666.18	704.18
## - No.of.Credits.at.this.Bank	1	666.36	704.36
## - Credit.Amount	1	666.67	704.67
## - Guarantors	1	667.02	705.02
## <none>		665.04	705.04
## - Sex...Marital.Status	1	668.33	706.33
## - Concurrent.Credits	1	668.74	706.74

```

## - Foreign.Worker          1  669.27 707.27
## - Most.valuable.available.asset 1  669.27 707.27
## - Type.of.apartment      1  669.45 707.45
## - Instalment.per.cent    1  669.55 707.55
## - Length.of.current.employment 1  670.60 708.60
## - No.of.dependents       1  670.65 708.65
## - Duration.of.Credit..month. 1  673.60 711.60
## - Value.Savings.Stocks    1  674.88 712.88
## - Payment.Status.of.Previous.Credit 1  679.45 717.45
## - Account.Balance         1  721.59 759.59
##
## Step: AIC=703.25
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##   Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##   Sex...Marital.Status + Guarantors + Most.valuable.available.asset +
##   Age..years. + Concurrent.Credits + Type.of.apartment + No.of.Credits.at.this.Bank +
##   No.of.dependents + Telephone + Foreign.Worker
##
##               Df Deviance   AIC
## - Telephone          1  665.65 701.65
## - Age..years.        1  666.19 702.19
## - Purpose            1  666.44 702.44
## - No.of.Credits.at.this.Bank 1  666.67 702.67
## - Credit.Amount      1  666.88 702.88
## - Guarantors         1  667.22 703.22
## <none>                665.25 703.25
## - Sex...Marital.Status 1  668.66 704.66
## - Concurrent.Credits  1  668.84 704.84
## - Foreign.Worker      1  669.59 705.59
## - Instalment.per.cent 1  669.86 705.86
## - Most.valuable.available.asset 1  669.92 705.92
## - Type.of.apartment  1  670.07 706.07
## - Length.of.current.employment 1  670.65 706.65
## - No.of.dependents    1  670.93 706.93
## - Duration.of.Credit..month. 1  673.85 709.85
## - Value.Savings.Stocks 1  674.94 710.94
## - Payment.Status.of.Previous.Credit 1  679.56 715.56
## - Account.Balance     1  723.00 759.00
##
## Step: AIC=701.65
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
##   Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
##   Sex...Marital.Status + Guarantors + Most.valuable.available.asset +
##   Age..years. + Concurrent.Credits + Type.of.apartment + No.of.Credits.at.this.Bank +
##   No.of.dependents + Foreign.Worker
##
##               Df Deviance   AIC
## - Age..years.        1  666.84 700.84
## - Credit.Amount      1  666.97 700.97
## - No.of.Credits.at.this.Bank 1  666.98 700.98

```

```

## - Purpose 1 666.99 700.99
## - Guarantors 1 667.53 701.53
## <none> 665.65 701.65
## - Sex...Marital.Status 1 669.23 703.23
## - Concurrent.Credits 1 669.37 703.37
## - Foreign.Worker 1 669.90 703.90
## - Instalment.per.cent 1 670.08 704.08
## - Most.valuable.available.asset 1 670.11 704.11
## - Type.of.apartment 1 670.38 704.38
## - Length.of.current.employment 1 671.15 705.15
## - No.of.dependents 1 671.53 705.53
## - Duration.of.Credit..month. 1 674.86 708.86
## - Value.Savings.Stocks 1 675.60 709.60
## - Payment.Status.of.Previous.Credit 1 679.87 713.87
## - Account.Balance 1 724.56 758.56
##
## Step: AIC=700.84
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
## Payment.Status.of.Previous.Credit + Purpose + Credit.Amount +
## Value.Savings.Stocks + Length.of.current.employment + Instalment.per.cent +
## Sex...Marital.Status + Guarantors + Most.valuable.available.asset +
## Concurrent.Credits + Type.of.apartment + No.of.Credits.at.this.Bank +
## No.of.dependents + Foreign.Worker
##
## Df Deviance AIC
## - Credit.Amount 1 667.93 699.93
## - No.of.Credits.at.this.Bank 1 667.93 699.93
## - Purpose 1 668.17 700.17
## - Guarantors 1 668.64 700.64
## <none> 666.84 700.84
## - Sex...Marital.Status 1 670.11 702.11
## - Concurrent.Credits 1 670.61 702.61
## - Instalment.per.cent 1 670.96 702.96
## - Most.valuable.available.asset 1 671.28 703.28
## - Foreign.Worker 1 671.28 703.28
## - No.of.dependents 1 672.27 704.27
## - Length.of.current.employment 1 673.59 705.59
## - Type.of.apartment 1 673.77 705.77
## - Duration.of.Credit..month. 1 676.96 708.96
## - Value.Savings.Stocks 1 677.33 709.33
## - Payment.Status.of.Previous.Credit 1 681.45 713.45
## - Account.Balance 1 725.45 757.45
##
## Step: AIC=699.93
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
## Payment.Status.of.Previous.Credit + Purpose + Value.Savings.Stocks +
## Length.of.current.employment + Instalment.per.cent + Sex...Marital.Status +
## Guarantors + Most.valuable.available.asset + Concurrent.Credits +
## Type.of.apartment + No.of.Credits.at.this.Bank + No.of.dependents +
## Foreign.Worker
##
## Df Deviance AIC

```

```

## - No.of.Credits.at.this.Bank      1  669.16 699.16
## - Purpose                          1  669.24 699.24
## - Guarantors                       1  669.76 699.76
## <none>                             667.93 699.93
## - Instalment.per.cent              1  671.02 701.02
## - Sex...Marital.Status              1  671.14 701.14
## - Concurrent.Credits                1  671.78 701.78
## - Foreign.Worker                    1  672.10 702.10
## - Most.valuable.available.asset     1  673.33 703.33
## - No.of.dependents                  1  673.34 703.34
## - Length.of.current.employment      1  674.65 704.65
## - Type.of.apartment                 1  674.67 704.67
## - Value.Savings.Stocks              1  678.12 708.12
## - Payment.Status.of.Previous.Credit 1  682.52 712.52
## - Duration.of.Credit..month.        1  690.69 720.69
## - Account.Balance                   1  726.22 756.22
##
## Step:  AIC=699.16
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Purpose + Value.Savings.Stocks +
##   Length.of.current.employment + Instalment.per.cent + Sex...Marital.Status +
##   Guarantors + Most.valuable.available.asset + Concurrent.Credits +
##   Type.of.apartment + No.of.dependents + Foreign.Worker
##
##                                     Df Deviance   AIC
## - Purpose                          1  670.27 698.27
## - Guarantors                       1  671.06 699.06
## <none>                             669.16 699.16
## - Instalment.per.cent              1  672.05 700.05
## - Sex...Marital.Status              1  672.16 700.16
## - Foreign.Worker                    1  673.34 701.34
## - Concurrent.Credits                1  673.57 701.57
## - Most.valuable.available.asset     1  674.61 702.61
## - No.of.dependents                  1  675.11 703.11
## - Length.of.current.employment      1  675.48 703.48
## - Type.of.apartment                 1  675.96 703.96
## - Value.Savings.Stocks              1  679.82 707.82
## - Payment.Status.of.Previous.Credit 1  682.88 710.88
## - Duration.of.Credit..month.        1  692.15 720.15
## - Account.Balance                   1  727.03 755.03
##
## Step:  AIC=698.27
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Value.Savings.Stocks +
##   Length.of.current.employment + Instalment.per.cent + Sex...Marital.Status +
##   Guarantors + Most.valuable.available.asset + Concurrent.Credits +
##   Type.of.apartment + No.of.dependents + Foreign.Worker
##
##                                     Df Deviance   AIC
## - Guarantors                       1  672.13 698.13
## <none>                             670.27 698.27
## - Instalment.per.cent              1  672.97 698.97

```

```
## - Sex...Marital.Status      1  673.28 699.28
## - Foreign.Worker            1  674.22 700.22
## - Concurrent.Credits        1  674.37 700.37
## - Most.valuable.available.asset 1  676.09 702.09
## - No.of.dependents          1  676.28 702.28
## - Length.of.current.employment 1  676.68 702.68
## - Type.of.apartment         1  677.38 703.38
## - Value.Savings.Stocks      1  680.71 706.71
## - Payment.Status.of.Previous.Credit 1  683.25 709.25
## - Duration.of.Credit..month. 1  692.15 718.15
## - Account.Balance           1  729.67 755.67
##
## Step: AIC=698.13
## Creditability ~ Account.Balance + Duration.of.Credit..month. +
##   Payment.Status.of.Previous.Credit + Value.Savings.Stocks +
##   Length.of.current.employment + Instalment.per.cent + Sex...Marital.Status +
##   Most.valuable.available.asset + Concurrent.Credits + Type.of.apartment +
##   No.of.dependents + Foreign.Worker
##
##              Df Deviance   AIC
## <none>              672.13 698.13
## - Instalment.per.cent      1  674.70 698.70
## - Sex...Marital.Status      1  675.61 699.61
## - Concurrent.Credits        1  676.06 700.06
## - Foreign.Worker            1  677.27 701.27
## - No.of.dependents          1  677.94 701.94
## - Length.of.current.employment 1  678.44 702.44
## - Type.of.apartment         1  679.00 703.00
## - Most.valuable.available.asset 1  679.10 703.10
## - Value.Savings.Stocks      1  681.70 705.70
## - Payment.Status.of.Previous.Credit 1  684.93 708.93
## - Duration.of.Credit..month. 1  693.40 717.40
## - Account.Balance           1  730.26 754.26
```

```
coef(step_mod)
```

```
##              (Intercept)              Account.Balance
##              -3.58905354              0.59895730
##      Duration.of.Credit..month. Payment.Status.of.Previous.Credit
##              -0.03634973              0.33246362
##      Value.Savings.Stocks      Length.of.current.employment
##              0.20400179              0.20619594
##      Instalment.per.cent              Sex...Marital.Status
##              -0.13800721              0.25247993
##      Most.valuable.available.asset      Concurrent.Credits
##              -0.26872553              0.25742637
##      Type.of.apartment              No.of.dependents
##              0.48860130              -0.66183015
##      Foreign.Worker
##              1.41141604
```

```
# Results
probabilities <- fit1 %>% predict(train, type = "response")
train$prob = probabilities
g1 <- roc(Creditability ~ train$prob,data = train)
print(coords(g1, "best"))
```

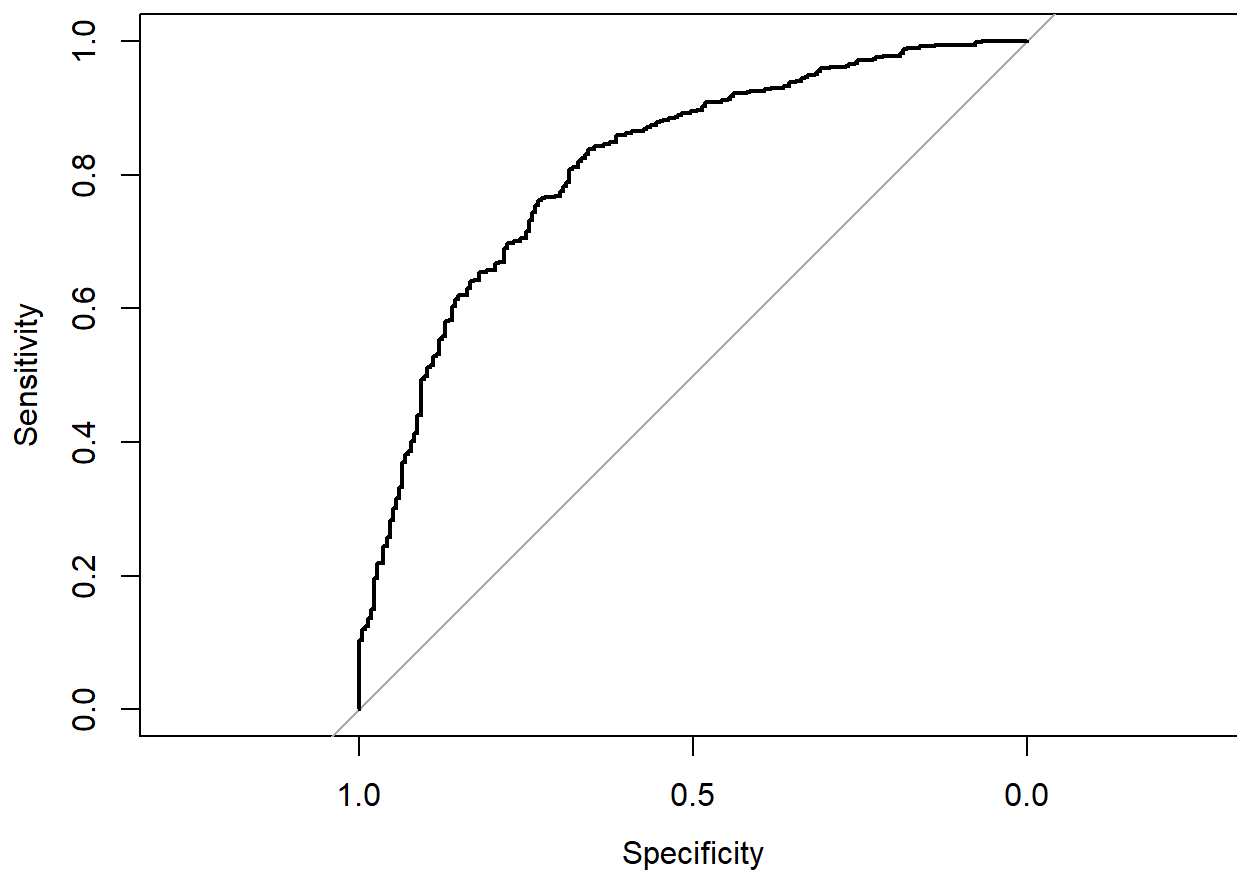
```
## threshold specificity sensitivity
## 0.5787590 0.6574074 0.8388430
```

```
train$prediction=ifelse(probabilities>=coords(g1, "best")[1],1,0)
train$prediction <- as.factor(train$prediction)
train$Creditability<-as.factor(train$Creditability)
confusionMatrix(train$prediction,train$Creditability)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0    1
##           0 142  78
##           1  74 406
##
##           Accuracy : 0.7829
##           95% CI : (0.7504, 0.8129)
##           No Information Rate : 0.6914
##           P-Value [Acc > NIR] : 4.119e-08
##
##           Kappa : 0.4937
##           Mcnemar's Test P-Value : 0.8077
##
##           Sensitivity : 0.6574
##           Specificity : 0.8388
##           Pos Pred Value : 0.6455
##           Neg Pred Value : 0.8458
##           Prevalence : 0.3086
##           Detection Rate : 0.2029
##           Detection Prevalence : 0.3143
##           Balanced Accuracy : 0.7481
##
##           'Positive' Class : 0
##
```

```
# Make predictions :Test Data
probabilities <- fit1 %>% predict(test, type = "response")
test$prob = probabilities
g1 <- roc(Creditability ~ train$prob,data = train)
plot(g1)
```





```
test$prediction=ifelse(probabilities>=0.6279582,1,0)
test$prediction <- as.factor(test$prediction)
test$Creditability<-as.factor(test$Creditability)
confusionMatrix(test$prediction,test$Creditability)
```

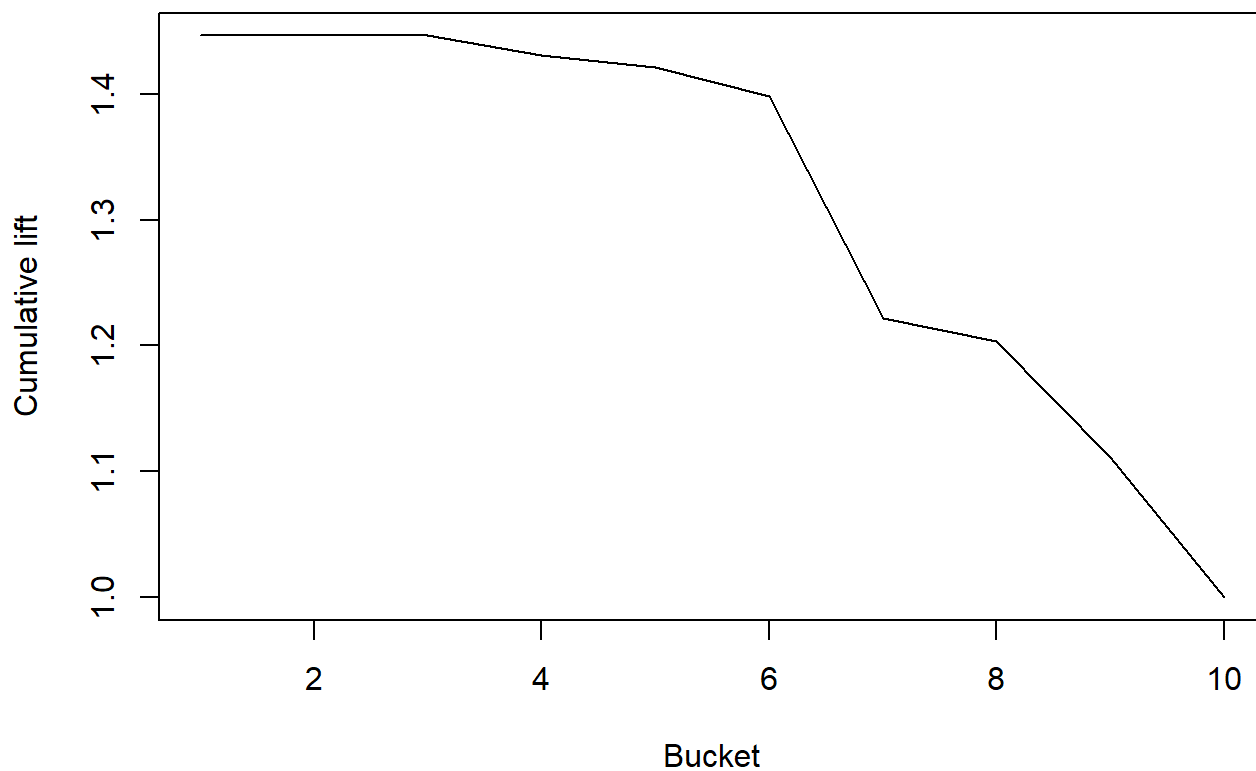
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##           0  49  42
##           1  35 174
##
##           Accuracy : 0.7433
##           95% CI : (0.69, 0.7918)
##       No Information Rate : 0.72
##       P-Value [Acc > NIR] : 0.2024
##
##           Kappa : 0.3792
##  McNemar's Test P-Value : 0.4941
##
##           Sensitivity : 0.5833
##           Specificity : 0.8056
##       Pos Pred Value : 0.5385
##       Neg Pred Value : 0.8325
##           Prevalence : 0.2800
##       Detection Rate : 0.1633
##  Detection Prevalence : 0.3033
##       Balanced Accuracy : 0.6944
##
##       'Positive' Class : 0
##
```

```
# Logistic model Ends Here
lift <- function(depvar, predcol, groups=10) {
  if(!require(dplyr)){
    install.packages("dplyr")
    library(dplyr)}
  if(is.factor(depvar)) depvar <- as.integer(as.character(depvar))
  if(is.factor(predcol)) predcol <- as.integer(as.character(predcol))
  helper = data.frame(cbind(depvar, predcol))
  helper[, "bucket"] = ntile(-helper[, "predcol"], groups)
  gaintable = helper %>% group_by(bucket) %>%
    summarise_at(vars(depvar), funs(total = n(),
                                     totalresp=sum(., na.rm = TRUE))) %>%
    mutate(Cumresp = cumsum(totalresp),
           Gain=Cumresp/sum(totalresp)*100,
           Cumlift=Gain/(bucket*(100/groups)))
  return(gaintable)
}

dt = lift(train$Creditability , train$prediction, groups = 10)
dt
```

```
## # A tibble: 10 x 6
##   bucket total totalresp Cumresp  Gain Cumlift
##   <int> <int>    <int>   <int> <dbl> <dbl>
## 1     1     70      70     70  14.5  1.45
## 2     2     70     140    140  28.9  1.45
## 3     3     70     210    210  43.4  1.45
## 4     4     67     277    277  57.2  1.43
## 5     5     67     344    344  71.1  1.42
## 6     6     62     406    406  83.9  1.40
## 7     7      8     414    414  85.5  1.22
## 8     8     52     466    466  96.3  1.20
## 9     9     18     484    484 100    1.11
## 10    10      0     484    484 100     1
```

```
graphics::plot(dt$bucket, dt$Cumlift, type="l", ylab="Cumulative lift", xlab="Bucket")
```



```
# Random Forest Model
control <- rfeControl(functions=rffuncs, method="cv", number=10)
## Training
modFit <- train(variables,method="rf",data=train)
modFit
```

```
## Random Forest
##
## 700 samples
## 20 predictor
## 2 classes: '0', '1'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 700, 700, 700, 700, 700, 700, ...
## Resampling results across tuning parameters:
##
##  mtry  Accuracy   Kappa
##    2    0.7497986 0.3202479
##   11    0.7444289 0.3569663
##   20    0.7391200 0.3521568
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 2.
```

```
#Decision Trees
set.seed(1)
TreeModel <- rpart(Creditability ~ ., data = train)
visTree(TreeModel, main = "Credibility lassification Tree", width = "100%")
```

## Credibility lassification Tree


[Export as png](#)

```
# library(rpart.plot)
# prp(TreeModel, type = 2, extra = 1)
```