

CreditCardAttrition

2023-06-21

```
#Load the libraries required
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.2
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
```

```
## v ggplot2 3.4.2      v purrr   1.0.1
```

```
## v tibble  3.1.8      v dplyr  1.0.10
```

```
## v tidyr   1.3.0      v stringr 1.5.0
```

```
## v readr   2.1.3      v forcats 0.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
## Warning: package 'purrr' was built under R version 4.2.2
```

```
## Warning: package 'stringr' was built under R version 4.2.3
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
library(ggplot2)
```

```
#Load the dataset
```

```
credit <- read_csv("BankChurners.csv")
```

```
## Rows: 10127 Columns: 23
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr  (6): Attrition_Flag, Gender, Education_Level, Marital_Status, Income_Ca...
```

```
## dbl  (17): CLIENTNUM, Customer_Age, Dependent_count, Months_on_book, Total_Re...
```

```
##
```

```
## i Use 'spec()' to retrieve the full column specification for this data.
```

```
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Exploratory Data Analysis

```
#View the few top and last rows of the data
head(credit)
```

```
## # A tibble: 6 x 23
##   CLIEN~1 Attri~2 Custo~3 Gender Depen~4 Educa~5 Marit~6 Incom~7 Card_~8 Month~9
##   <dbl> <chr>      <dbl> <chr>      <dbl> <chr>      <chr> <chr> <chr>      <dbl>
## 1  7.69e8 Existi~    45 M          3 High S~ Married $60K -- Blue      39
## 2  8.19e8 Existi~    49 F          5 Gradua~ Single  Less t~ Blue      44
## 3  7.14e8 Existi~    51 M          3 Gradua~ Married $80K -- Blue      36
## 4  7.70e8 Existi~    40 F          4 High S~ Unknown Less t~ Blue      34
## 5  7.09e8 Existi~    40 M          3 Uneduc~ Married $60K -- Blue      21
## 6  7.13e8 Existi~    44 M          2 Gradua~ Married $40K -- Blue      36
## # ... with 13 more variables: Total_Relationship_Count <dbl>,
## #   Months_Inactive_12_mon <dbl>, Contacts_Count_12_mon <dbl>,
## #   Credit_Limit <dbl>, Total_Revolving_Bal <dbl>, Avg_Open_To_Buy <dbl>,
## #   Total_Amt_Chng_Q4_Q1 <dbl>, Total_Trans_Amt <dbl>, Total_Trans_Ct <dbl>,
## #   Total_Ct_Chng_Q4_Q1 <dbl>, Avg_Utilization_Ratio <dbl>,
## #   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Educac
## #   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Educac
```

```
tail(credit)
```

```
## # A tibble: 6 x 23
##   CLIEN~1 Attri~2 Custo~3 Gender Depen~4 Educa~5 Marit~6 Incom~7 Card_~8 Month~9
##   <dbl> <chr>      <dbl> <chr>      <dbl> <chr>      <chr> <chr> <chr>      <dbl>
## 1  7.14e8 Existi~    56 F          1 Gradua~ Single  Less t~ Blue      50
## 2  7.72e8 Existi~    50 M          2 Gradua~ Single  $40K -- Blue      40
## 3  7.11e8 Attrit~    41 M          2 Unknown Divorc~ $40K -- Blue      25
## 4  7.17e8 Attrit~    44 F          1 High S~ Married Less t~ Blue      36
## 5  7.17e8 Attrit~    30 M          2 Gradua~ Unknown $40K -- Blue      36
## 6  7.14e8 Attrit~    43 F          2 Gradua~ Married Less t~ Silver    25
## # ... with 13 more variables: Total_Relationship_Count <dbl>,
## #   Months_Inactive_12_mon <dbl>, Contacts_Count_12_mon <dbl>,
## #   Credit_Limit <dbl>, Total_Revolving_Bal <dbl>, Avg_Open_To_Buy <dbl>,
## #   Total_Amt_Chng_Q4_Q1 <dbl>, Total_Trans_Amt <dbl>, Total_Trans_Ct <dbl>,
## #   Total_Ct_Chng_Q4_Q1 <dbl>, Avg_Utilization_Ratio <dbl>,
## #   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Educac
## #   Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Educac
```

```
#Structure of the data
glimpse(credit)
```

```
## Rows: 10,127
## Columns: 23
## $ CLIENTNUM
## $ Attrition_Flag
## $ Customer_Age
## $ Gender
## $ Dependent_count
## $ Education_Level
## $ Marital_Status
## $ Income_Category
```

```
## $ Card_Category
## $ Months_on_book
## $ Total_Relationship_Count
## $ Months_Inactive_12_mon
## $ Contacts_Count_12_mon
## $ Credit_Limit
## $ Total_Revolving_Bal
## $ Avg_Open_To_Buy
## $ Total_Amt_Chng_Q4_Q1
## $ Total_Trans_Amt
## $ Total_Trans_Ct
## $ Total_Ct_Chng_Q4_Q1
## $ Avg_Utilization_Ratio
## $ Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Marital_Status_Income
## $ Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Marital_Status_Income
```

```
#Dropping the first and last two columns of the data
credit <- credit[, 2:21]
colnames(credit)
```

```
## [1] "Attrition_Flag"      "Customer_Age"
## [3] "Gender"              "Dependent_count"
## [5] "Education_Level"     "Marital_Status"
## [7] "Income_Category"     "Card_Category"
## [9] "Months_on_book"      "Total_Relationship_Count"
## [11] "Months_Inactive_12_mon" "Contacts_Count_12_mon"
## [13] "Credit_Limit"        "Total_Revolving_Bal"
## [15] "Avg_Open_To_Buy"      "Total_Amt_Chng_Q4_Q1"
## [17] "Total_Trans_Amt"      "Total_Trans_Ct"
## [19] "Total_Ct_Chng_Q4_Q1" "Avg_Utilization_Ratio"
```

```
#Searching for null values in all variables
```

```
#Create a function to search null values in a variable
null_values <- function(variable){
  sum(is.null(variable) | is.na(variable))
}
```

```
#Apply the function to all variables in the data
null_counts <- sapply(credit, null_values)
```

```
#Print the count of null values
null_counts
```

```
##      Attrition_Flag      Customer_Age      Gender
##      0                0                0
##      Dependent_count      Education_Level      Marital_Status
##      0                0                0
##      Income_Category      Card_Category      Months_on_book
##      0                0                0
##      Total_Relationship_Count      Months_Inactive_12_mon      Contacts_Count_12_mon
##      0                0                0
##      Credit_Limit      Total_Revolving_Bal      Avg_Open_To_Buy
```

```
##           0           0           0
##   Total_Amt_Chng_Q4_Q1   Total_Trans_Amt   Total_Trans_Ct
##           0           0           0
##   Total_Ct_Chng_Q4_Q1   Avg_Utilization_Ratio
##           0           0
```

There are no null values in the dataset

```
#Finding the duplicate rows
duplicate_rows <- credit[duplicated(credit) , ]

#Print the duplicate rows
duplicate_rows
```

```
## # A tibble: 0 x 20
## # ... with 20 variables: Attrition_Flag <chr>, Customer_Age <dbl>,
## #   Gender <chr>, Dependent_count <dbl>, Education_Level <chr>,
## #   Marital_Status <chr>, Income_Category <chr>, Card_Category <chr>,
## #   Months_on_book <dbl>, Total_Relationship_Count <dbl>,
## #   Months_Inactive_12_mon <dbl>, Contacts_Count_12_mon <dbl>,
## #   Credit_Limit <dbl>, Total_Revolving_Bal <dbl>, Avg_Open_To_Buy <dbl>,
## #   Total_Amt_Chng_Q4_Q1 <dbl>, Total_Trans_Amt <dbl>, ...
```

There are no duplicate rows in the dataset

```
#Statistical summaries for each variable
summary(credit)
```

```
##   Attrition_Flag   Customer_Age   Gender   Dependent_count
##   Length:10127    Min.   :26.00   Length:10127   Min.   :0.000
##   Class :character 1st Qu.:41.00   Class :character 1st Qu.:1.000
##   Mode  :character Median :46.00   Mode  :character Median :2.000
##                   Mean  :46.33                   Mean  :2.346
##                   3rd Qu.:52.00                   3rd Qu.:3.000
##                   Max.   :73.00                   Max.   :5.000
##   Education_Level   Marital_Status   Income_Category   Card_Category
##   Length:10127     Length:10127     Length:10127     Length:10127
##   Class :character Class :character Class :character Class :character
##   Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##   Months_on_book   Total_Relationship_Count   Months_Inactive_12_mon
##   Min.   :13.00    Min.   :1.000                Min.   :0.000
##   1st Qu.:31.00    1st Qu.:3.000                1st Qu.:2.000
##   Median :36.00    Median :4.000                Median :2.000
##   Mean   :35.93    Mean   :3.813                Mean   :2.341
##   3rd Qu.:40.00    3rd Qu.:5.000                3rd Qu.:3.000
##   Max.   :56.00    Max.   :6.000                Max.   :6.000
##   Contacts_Count_12_mon   Credit_Limit   Total_Revolving_Bal   Avg_Open_To_Buy
##   Min.   :0.000          Min.   : 1438   Min.   : 0           Min.   : 3
##   1st Qu.:2.000          1st Qu.: 2555   1st Qu.: 359         1st Qu.: 1324
```

```
## Median :2.000      Median : 4549      Median :1276      Median : 3474
## Mean   :2.455      Mean    : 8632      Mean    :1163      Mean    : 7469
## 3rd Qu.:3.000      3rd Qu.:11068     3rd Qu.:1784      3rd Qu.: 9859
## Max.   :6.000      Max.    :34516     Max.    :2517      Max.    :34516
## Total_Amt_Chng_Q4_Q1 Total_Trans_Amt Total_Trans_Ct Total_Ct_Chng_Q4_Q1
## Min.   :0.0000     Min.    : 510     Min.    : 10.00    Min.    :0.0000
## 1st Qu.:0.6310     1st Qu.: 2156     1st Qu.: 45.00    1st Qu.:0.5820
## Median :0.7360     Median : 3899     Median : 67.00    Median :0.7020
## Mean   :0.7599     Mean    : 4404     Mean    : 64.86    Mean    :0.7122
## 3rd Qu.:0.8590     3rd Qu.: 4741     3rd Qu.: 81.00    3rd Qu.:0.8180
## Max.   :3.3970     Max.    :18484     Max.    :139.00    Max.    :3.7140
## Avg_Utilization_Ratio
## Min.   :0.0000
## 1st Qu.:0.0230
## Median :0.1760
## Mean   :0.2749
## 3rd Qu.:0.5030
## Max.   :0.9990
```

Modelling

```
#Converting Y variable into 0s and 1s.
```

```
#0 - Existing Customer
```

```
#1 - Attrited Customer
```

```
credit$Attrition_Flag <- ifelse(credit$Attrition_Flag == "Existing Customer",0,1)
```

```
#Factor Analysis for Mixed Data
```

```
library(FactoMineR)
```

```
## Warning: package 'FactoMineR' was built under R version 4.2.3
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.2.3
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
famd_result <- FAMD(credit[,2:20], graph = FALSE)
```

```
#Summary of FAMD
```

```
summary(famd_result)
```

```
##
```

```
## Call:
```

```
## FAMD(base = credit[, 2:20], graph = FALSE)
```

```
##
```

```
##
```

```
## Eigenvalues
```

```
##          Dim.1  Dim.2  Dim.3  Dim.4  Dim.5
```

```
## Variance    3.459  2.152  1.822  1.623  1.332
```

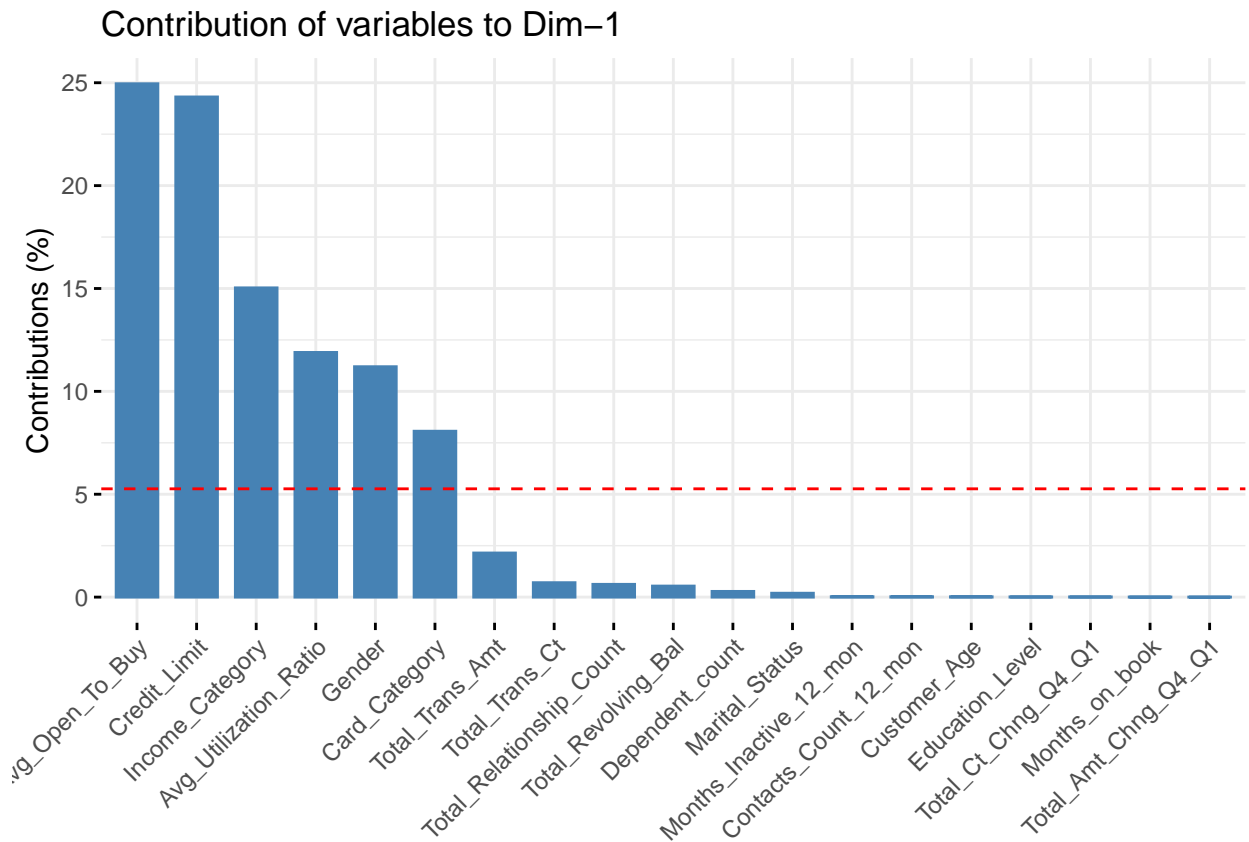
```

## % of var.          10.809  6.724  5.694  5.072  4.162
## Cumulative % of var. 10.809 17.533 23.226 28.299 32.460
##
## Individuals (the 10 first)
##
##          Dist    Dim.1    ctr    cos2    Dim.2    ctr    cos2
## 1          |  6.352 |  1.025  0.003  0.026 | -1.181  0.006  0.035
## 2          | 13.810 | -0.910  0.002  0.004 |  0.921  0.004  0.004
## 3          | 11.900 |  0.383  0.000  0.001 | -0.224  0.000  0.000
## 4          |  9.626 | -2.361  0.016  0.060 |  1.164  0.006  0.015
## 5          | 11.445 |  0.252  0.000  0.000 |  0.328  0.000  0.001
## 6          |  4.936 | -0.752  0.002  0.023 | -1.068  0.005  0.047
## 7          | 12.895 |  4.902  0.069  0.145 | -1.393  0.009  0.012
## 8          | 10.339 |  4.028  0.046  0.152 |  0.322  0.000  0.001
## 9          | 13.310 |  1.805  0.009  0.018 |  0.102  0.000  0.000
## 10         |  5.436 |  0.877  0.002  0.026 | -1.845  0.016  0.115
##
##          Dim.3    ctr    cos2
## 1          | -1.288  0.009  0.041 |
## 2          | -0.602  0.002  0.002 |
## 3          | -1.447  0.011  0.015 |
## 4          | -1.344  0.010  0.019 |
## 5          | -3.555  0.068  0.096 |
## 6          | -1.079  0.006  0.048 |
## 7          |  0.176  0.000  0.000 |
## 8          | -2.601  0.037  0.063 |
## 9          | -2.615  0.037  0.039 |
## 10         | -0.947  0.005  0.030 |
##
## Continuous variables (the 10 first)
##
##          Dim.1    ctr    cos2    Dim.2    ctr    cos2    Dim.3
## Customer_Age          | -0.024  0.017  0.001 | -0.321  4.800  0.103 |  0.865
## Dependent_count       |  0.098  0.277  0.010 |  0.116  0.625  0.013 | -0.205
## Months_on_book        | -0.014  0.005  0.000 | -0.313  4.552  0.098 |  0.863
## Total_Relationship_Count | -0.147  0.621  0.021 | -0.418  8.111  0.175 | -0.218
## Months_Inactive_12_mon | -0.026  0.020  0.001 | -0.113  0.591  0.013 |  0.119
## Contacts_Count_12_mon  |  0.026  0.019  0.001 | -0.262  3.198  0.069 | -0.110
## Credit_Limit          |  0.917 24.310  0.841 | -0.019  0.016  0.000 |  0.035
## Total_Revolving_Bal    | -0.136  0.536  0.019 |  0.191  1.688  0.036 |  0.057
## Avg_Open_To_Buy       |  0.929 24.951  0.863 | -0.036  0.060  0.001 |  0.030
## Total_Amt_Chng_Q4_Q1   |  0.006  0.001  0.000 |  0.123  0.705  0.015 | -0.131
##
##          ctr    cos2
## Customer_Age          41.025  0.747 |
## Dependent_count       2.305  0.042 |
## Months_on_book        40.866  0.745 |
## Total_Relationship_Count 2.610  0.048 |
## Months_Inactive_12_mon  0.772  0.014 |
## Contacts_Count_12_mon  0.664  0.012 |
## Credit_Limit          0.067  0.001 |
## Total_Revolving_Bal    0.177  0.003 |
## Avg_Open_To_Buy       0.049  0.001 |
## Total_Amt_Chng_Q4_Q1   0.946  0.017 |
##
## Categories (the 10 first)
##
##          Dim.1    ctr    cos2  v.test    Dim.2    ctr
## F          | -1.092  5.274  0.626 -62.629 |  0.368  1.550

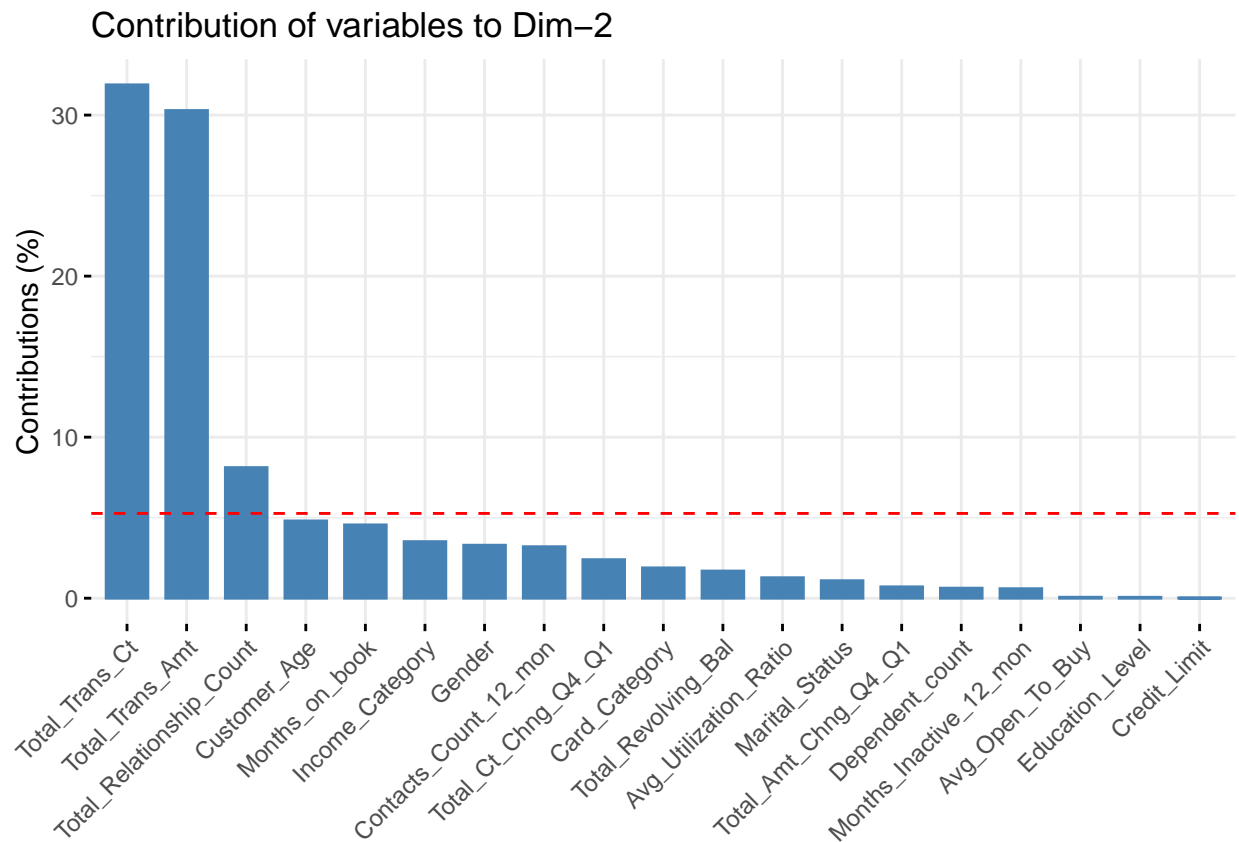
```

## M		1.227	5.925	0.626	62.629		-0.414	1.741
## College		0.053	0.002	0.000	0.949		0.005	0.000
## Doctorate		-0.064	0.002	0.000	-0.745		-0.206	0.041
## Graduate		-0.027	0.002	0.000	-0.964		0.008	0.000
## High School		0.006	0.000	0.000	0.159		-0.025	0.003
## Post-Graduate		0.060	0.002	0.000	0.752		0.094	0.010
## Uneducated		0.046	0.003	0.000	1.040		0.027	0.002
## Unknown		-0.035	0.002	0.000	-0.792		0.018	0.001
## Divorced		0.151	0.014	0.002	2.311		0.200	0.064
##		cos2	v.test	Dim.3	ctr		cos2	v.test
## F		0.071	26.777		0.189	0.568	0.019	14.922
## M		0.071	-26.777		-0.212	0.639	0.019	-14.922
## College		0.000	0.110		-0.126	0.048	0.002	-3.128
## Doctorate		0.002	-3.052		0.367	0.181	0.006	5.905
## Graduate		0.000	0.357		0.016	0.003	0.000	0.817
## High School		0.000	-0.865		-0.003	0.000	0.000	-0.115
## Post-Graduate		0.000	1.490		-0.284	0.124	0.004	-4.905
## Uneducated		0.000	0.757		0.034	0.005	0.000	1.067
## Unknown		0.000	0.507		0.008	0.000	0.000	0.251
## Divorced		0.003	3.866		-0.305	0.207	0.007	-6.422

```
#Variable's contribution to Dimension 1
fviz_contrib(famd_result,"var", axes=1)
```



```
#Contribution of Variables to Dimension 2
fviz_contrib(famd_result, "var", axes=2)
```



```
#Total unique values in income category
print(unique(credit$Income_Category))
```

```
## [1] "$60K - $80K" "Less than $40K" "$80K - $120K" "$40K - $60K"
## [5] "$120K +" "Unknown"
```

```
#Converting the categories in Income_Category to numbers
credit$Income_Category <- as.numeric(factor(credit$Income_Category, levels = c("Unknown","Less than $40K",
"$40K - $60K","$60K - $80K",
"$80K - $120K","$120K +"),
labels = c(0,1,2,3,4,5)))
```

```
#Total Unique Values in Gender Variable
print(unique(credit$Gender))
```

```
## [1] "M" "F"
```

```
#Converting Males to 0 and Females to 1
credit$Gender <- ifelse(credit$Gender == "M",0,1)
```



```
#Total Unique Values in Card Category  
print(unique(credit$Card_Category))
```

```
## [1] "Blue"      "Gold"      "Silver"    "Platinum"
```

```
#Assigning numbers to card categories  
#1 - Blue  
#2 - Silver  
#3- Gold  
#4 - Platinum
```

```
credit$Card_Category <- as.numeric(factor(credit$Card_Category, levels = c("Blue","Silver","Gold","Platinum"),  
                                         labels = c(1,2,3,4)))
```

```
#Convert the target variable to factor  
#credit$Attrition_Flag <- as.factor(credit$Attrition_Flag)
```

```
#Set the seed and split the data into training and testing parts  
set.seed(7)  
credit_index <- sample(nrow(credit), 0.6*nrow(credit), replace = FALSE)  
credit_train <- credit[credit_index, ]  
credit_test <- credit[-credit_index, ]
```

```
#Train simple neural network  
library(neuralnet)
```

```
## Warning: package 'neuralnet' was built under R version 4.2.2
```

```
##  
## Attaching package: 'neuralnet'
```

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

```
NN1 <- neuralnet(Attrition_Flag ~ Avg_Open_To_Buy + Credit_Limit + Income_Category + Avg_Utilization_Ratio +  
                  Gender + Card_Category, credit_train, hidden = 4, lifesign = "minimal",  
                  linear.output = FALSE, threshold = 0.1)
```

```
## hidden: 4 thresh: 0.1 rep: 1/1 steps:
```

```
##      92 error: 414.22505 time: 0.24 secs
```

```
#Plot the Simple Neural Network  
plot(NN1)
```

```
#Testing the Simple Neural Network  
predictions <- compute(NN1, credit_test[, -1])
```

```

#Accuracy
library(caret)

## Warning: package 'caret' was built under R version 4.2.3

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
## lift

results <- data.frame(actual = credit_test$Attrition_Flag, prediction = predictions$net.result)
results$prediction <- ifelse(results$prediction > 0.5, 1, 0)

actual <- as.factor(credit_test$Attrition_Flag)
neural_result <- as.factor(results$prediction)

confusionMatrix(neural_result, actual)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##           0 3410  632
##           1    8    1
##
##           Accuracy : 0.842
##           95% CI : (0.8304, 0.8531)
##       No Information Rate : 0.8437
##       P-Value [Acc > NIR] : 0.6289
##
##           Kappa : -0.0013
##
##  McNemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.99766
##           Specificity : 0.00158
##       Pos Pred Value : 0.84364
##       Neg Pred Value : 0.11111
##           Prevalence : 0.84374
##       Detection Rate : 0.84177
##   Detection Prevalence : 0.99778
##       Balanced Accuracy : 0.49962
##
##       'Positive' Class : 0
##

```