FML Assignment - 1

Varsha Karunya Mekala

2023-09-09

#importing the dataset by defining the path

library(readxl)  
credit\_card <- read\_excel("/Users/varshamekala/Desktop/fml/application\_data.xlsx")  
print(credit\_card)

## # A tibble: 307,511 × 40  
## SK\_ID\_CURR TARGET NAME\_CONTRACT\_TYPE CODE\_GENDER FLAG\_OWN\_CAR FLAG\_OWN\_REALTY  
## <dbl> <dbl> <chr> <chr> <chr> <chr>   
## 1 100002 1 Cash loans M N Y   
## 2 100003 0 Cash loans F N N   
## 3 100004 0 Revolving loans M Y Y   
## 4 100006 0 Cash loans F N Y   
## 5 100007 0 Cash loans M N Y   
## 6 100008 0 Cash loans M N Y   
## 7 100009 0 Cash loans F Y Y   
## 8 100010 0 Cash loans M Y Y   
## 9 100011 0 Cash loans F N Y   
## 10 100012 0 Revolving loans M N Y   
## # ℹ 307,501 more rows  
## # ℹ 34 more variables: CNT\_CHILDREN <dbl>, AMT\_INCOME\_TOTAL <dbl>,  
## # AMT\_CREDIT <dbl>, AMT\_ANNUITY <dbl>, AMT\_GOODS\_PRICE <dbl>,  
## # NAME\_TYPE\_SUITE <chr>, NAME\_INCOME\_TYPE <chr>, NAME\_EDUCATION\_TYPE <chr>,  
## # NAME\_FAMILY\_STATUS <chr>, NAME\_HOUSING\_TYPE <chr>,  
## # REGION\_POPULATION\_RELATIVE <dbl>, DAYS\_BIRTH <dbl>, DAYS\_EMPLOYED <dbl>,  
## # DAYS\_REGISTRATION <dbl>, DAYS\_ID\_PUBLISH <dbl>, FLAG\_MOBIL <dbl>, …

#descriptive statistics for a selection of quantitative and categorical variables

summary(credit\_card)

## SK\_ID\_CURR TARGET NAME\_CONTRACT\_TYPE CODE\_GENDER   
## Min. :100002 Min. :0.00000 Length:307511 Length:307511   
## 1st Qu.:189146 1st Qu.:0.00000 Class :character Class :character   
## Median :278202 Median :0.00000 Mode :character Mode :character   
## Mean :278180 Mean :0.08073   
## 3rd Qu.:367142 3rd Qu.:0.00000   
## Max. :456255 Max. :1.00000   
##   
## FLAG\_OWN\_CAR FLAG\_OWN\_REALTY CNT\_CHILDREN AMT\_INCOME\_TOTAL   
## Length:307511 Length:307511 Min. : 0.0000 Min. : 25650   
## Class :character Class :character 1st Qu.: 0.0000 1st Qu.: 112500   
## Mode :character Mode :character Median : 0.0000 Median : 147150   
## Mean : 0.4171 Mean : 168798   
## 3rd Qu.: 1.0000 3rd Qu.: 202500   
## Max. :19.0000 Max. :117000000   
##   
## AMT\_CREDIT AMT\_ANNUITY AMT\_GOODS\_PRICE NAME\_TYPE\_SUITE   
## Min. : 45000 Min. : 1616 Min. : 40500 Length:307511   
## 1st Qu.: 270000 1st Qu.: 16524 1st Qu.: 238500 Class :character   
## Median : 513531 Median : 24903 Median : 450000 Mode :character   
## Mean : 599026 Mean : 27109 Mean : 538396   
## 3rd Qu.: 808650 3rd Qu.: 34596 3rd Qu.: 679500   
## Max. :4050000 Max. :258026 Max. :4050000   
## NA's :12 NA's :278   
## NAME\_INCOME\_TYPE NAME\_EDUCATION\_TYPE NAME\_FAMILY\_STATUS NAME\_HOUSING\_TYPE   
## Length:307511 Length:307511 Length:307511 Length:307511   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## REGION\_POPULATION\_RELATIVE DAYS\_BIRTH DAYS\_EMPLOYED DAYS\_REGISTRATION  
## Min. :0.00029 Min. :-25229 Min. :-17912 Min. :-24672   
## 1st Qu.:0.01001 1st Qu.:-19682 1st Qu.: -2760 1st Qu.: -7480   
## Median :0.01885 Median :-15750 Median : -1213 Median : -4504   
## Mean :0.02087 Mean :-16037 Mean : 63815 Mean : -4986   
## 3rd Qu.:0.02866 3rd Qu.:-12413 3rd Qu.: -289 3rd Qu.: -2010   
## Max. :0.07251 Max. : -7489 Max. :365243 Max. : 0   
##   
## DAYS\_ID\_PUBLISH FLAG\_MOBIL FLAG\_EMP\_PHONE FLAG\_WORK\_PHONE   
## Min. :-7197 Min. :0 Min. :0.0000 Min. :0.0000   
## 1st Qu.:-4299 1st Qu.:1 1st Qu.:1.0000 1st Qu.:0.0000   
## Median :-3254 Median :1 Median :1.0000 Median :0.0000   
## Mean :-2994 Mean :1 Mean :0.8199 Mean :0.1994   
## 3rd Qu.:-1720 3rd Qu.:1 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. : 0 Max. :1 Max. :1.0000 Max. :1.0000   
##   
## FLAG\_CONT\_MOBILE FLAG\_PHONE FLAG\_EMAIL OCCUPATION\_TYPE   
## Min. :0.0000 Min. :0.0000 Min. :0.00000 Length:307511   
## 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:0.00000 Class :character   
## Median :1.0000 Median :0.0000 Median :0.00000 Mode :character   
## Mean :0.9981 Mean :0.2811 Mean :0.05672   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.0000 Max. :1.00000   
##   
## CNT\_FAM\_MEMBERS REGION\_RATING\_CLIENT REGION\_RATING\_CLIENT\_W\_CITY  
## Min. : 1.000 Min. :1.000 Min. :1.000   
## 1st Qu.: 2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median : 2.000 Median :2.000 Median :2.000   
## Mean : 2.153 Mean :2.052 Mean :2.032   
## 3rd Qu.: 3.000 3rd Qu.:2.000 3rd Qu.:2.000   
## Max. :20.000 Max. :3.000 Max. :3.000   
## NA's :2   
## WEEKDAY\_APPR\_PROCESS\_START HOUR\_APPR\_PROCESS\_START REG\_REGION\_NOT\_LIVE\_REGION  
## Length:307511 Min. : 0.00 Min. :0.00000   
## Class :character 1st Qu.:10.00 1st Qu.:0.00000   
## Mode :character Median :12.00 Median :0.00000   
## Mean :12.06 Mean :0.01514   
## 3rd Qu.:14.00 3rd Qu.:0.00000   
## Max. :23.00 Max. :1.00000   
##   
## REG\_REGION\_NOT\_WORK\_REGION LIVE\_REGION\_NOT\_WORK\_REGION REG\_CITY\_NOT\_LIVE\_CITY  
## Min. :0.00000 Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000   
## Median :0.00000 Median :0.00000 Median :0.00000   
## Mean :0.05077 Mean :0.04066 Mean :0.07817   
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000   
## Max. :1.00000 Max. :1.00000 Max. :1.00000   
##   
## REG\_CITY\_NOT\_WORK\_CITY LIVE\_CITY\_NOT\_WORK\_CITY ORGANIZATION\_TYPE   
## Min. :0.0000 Min. :0.0000 Length:307511   
## 1st Qu.:0.0000 1st Qu.:0.0000 Class :character   
## Median :0.0000 Median :0.0000 Mode :character   
## Mean :0.2305 Mean :0.1796   
## 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000   
##

sd(credit\_card$HOUR\_APPR\_PROCESS\_START)

## [1] 3.265832

var(credit\_card$HOUR\_APPR\_PROCESS\_START)

## [1] 10.66566

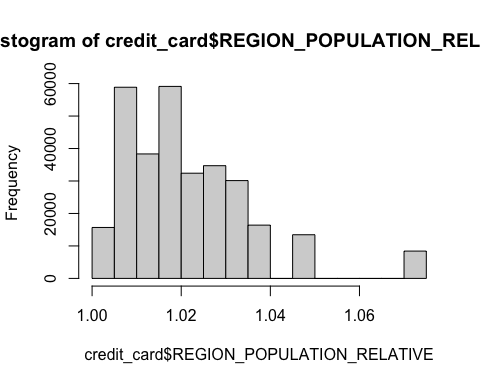
#transformation of one variable

credit\_card$REGION\_POPULATION\_RELATIVE<-(credit\_card$REGION\_POPULATION\_RELATIVE+1.0)  
summary(credit\_card)

## SK\_ID\_CURR TARGET NAME\_CONTRACT\_TYPE CODE\_GENDER   
## Min. :100002 Min. :0.00000 Length:307511 Length:307511   
## 1st Qu.:189146 1st Qu.:0.00000 Class :character Class :character   
## Median :278202 Median :0.00000 Mode :character Mode :character   
## Mean :278180 Mean :0.08073   
## 3rd Qu.:367142 3rd Qu.:0.00000   
## Max. :456255 Max. :1.00000   
##   
## FLAG\_OWN\_CAR FLAG\_OWN\_REALTY CNT\_CHILDREN AMT\_INCOME\_TOTAL   
## Length:307511 Length:307511 Min. : 0.0000 Min. : 25650   
## Class :character Class :character 1st Qu.: 0.0000 1st Qu.: 112500   
## Mode :character Mode :character Median : 0.0000 Median : 147150   
## Mean : 0.4171 Mean : 168798   
## 3rd Qu.: 1.0000 3rd Qu.: 202500   
## Max. :19.0000 Max. :117000000   
##   
## AMT\_CREDIT AMT\_ANNUITY AMT\_GOODS\_PRICE NAME\_TYPE\_SUITE   
## Min. : 45000 Min. : 1616 Min. : 40500 Length:307511   
## 1st Qu.: 270000 1st Qu.: 16524 1st Qu.: 238500 Class :character   
## Median : 513531 Median : 24903 Median : 450000 Mode :character   
## Mean : 599026 Mean : 27109 Mean : 538396   
## 3rd Qu.: 808650 3rd Qu.: 34596 3rd Qu.: 679500   
## Max. :4050000 Max. :258026 Max. :4050000   
## NA's :12 NA's :278   
## NAME\_INCOME\_TYPE NAME\_EDUCATION\_TYPE NAME\_FAMILY\_STATUS NAME\_HOUSING\_TYPE   
## Length:307511 Length:307511 Length:307511 Length:307511   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## REGION\_POPULATION\_RELATIVE DAYS\_BIRTH DAYS\_EMPLOYED DAYS\_REGISTRATION  
## Min. :1.000 Min. :-25229 Min. :-17912 Min. :-24672   
## 1st Qu.:1.010 1st Qu.:-19682 1st Qu.: -2760 1st Qu.: -7480   
## Median :1.019 Median :-15750 Median : -1213 Median : -4504   
## Mean :1.021 Mean :-16037 Mean : 63815 Mean : -4986   
## 3rd Qu.:1.029 3rd Qu.:-12413 3rd Qu.: -289 3rd Qu.: -2010   
## Max. :1.073 Max. : -7489 Max. :365243 Max. : 0   
##   
## DAYS\_ID\_PUBLISH FLAG\_MOBIL FLAG\_EMP\_PHONE FLAG\_WORK\_PHONE   
## Min. :-7197 Min. :0 Min. :0.0000 Min. :0.0000   
## 1st Qu.:-4299 1st Qu.:1 1st Qu.:1.0000 1st Qu.:0.0000   
## Median :-3254 Median :1 Median :1.0000 Median :0.0000   
## Mean :-2994 Mean :1 Mean :0.8199 Mean :0.1994   
## 3rd Qu.:-1720 3rd Qu.:1 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. : 0 Max. :1 Max. :1.0000 Max. :1.0000   
##   
## FLAG\_CONT\_MOBILE FLAG\_PHONE FLAG\_EMAIL OCCUPATION\_TYPE   
## Min. :0.0000 Min. :0.0000 Min. :0.00000 Length:307511   
## 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:0.00000 Class :character   
## Median :1.0000 Median :0.0000 Median :0.00000 Mode :character   
## Mean :0.9981 Mean :0.2811 Mean :0.05672   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.0000 Max. :1.00000   
##   
## CNT\_FAM\_MEMBERS REGION\_RATING\_CLIENT REGION\_RATING\_CLIENT\_W\_CITY  
## Min. : 1.000 Min. :1.000 Min. :1.000   
## 1st Qu.: 2.000 1st Qu.:2.000 1st Qu.:2.000   
## Median : 2.000 Median :2.000 Median :2.000   
## Mean : 2.153 Mean :2.052 Mean :2.032   
## 3rd Qu.: 3.000 3rd Qu.:2.000 3rd Qu.:2.000   
## Max. :20.000 Max. :3.000 Max. :3.000   
## NA's :2   
## WEEKDAY\_APPR\_PROCESS\_START HOUR\_APPR\_PROCESS\_START REG\_REGION\_NOT\_LIVE\_REGION  
## Length:307511 Min. : 0.00 Min. :0.00000   
## Class :character 1st Qu.:10.00 1st Qu.:0.00000   
## Mode :character Median :12.00 Median :0.00000   
## Mean :12.06 Mean :0.01514   
## 3rd Qu.:14.00 3rd Qu.:0.00000   
## Max. :23.00 Max. :1.00000   
##   
## REG\_REGION\_NOT\_WORK\_REGION LIVE\_REGION\_NOT\_WORK\_REGION REG\_CITY\_NOT\_LIVE\_CITY  
## Min. :0.00000 Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000   
## Median :0.00000 Median :0.00000 Median :0.00000   
## Mean :0.05077 Mean :0.04066 Mean :0.07817   
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000   
## Max. :1.00000 Max. :1.00000 Max. :1.00000   
##   
## REG\_CITY\_NOT\_WORK\_CITY LIVE\_CITY\_NOT\_WORK\_CITY ORGANIZATION\_TYPE   
## Min. :0.0000 Min. :0.0000 Length:307511   
## 1st Qu.:0.0000 1st Qu.:0.0000 Class :character   
## Median :0.0000 Median :0.0000 Mode :character   
## Mean :0.2305 Mean :0.1796   
## 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000   
##

#one quantitative variable

hist(credit\_card$REGION\_POPULATION\_RELATIVE)

 #scatterplot

x<-(credit\_card$AMT\_INCOME\_TOTAL)  
y<-(credit\_card$AMT\_CREDIT)  
# Corrected code  
plot(x, y, main = "Profile.Score VS Age ", xlab = "Profile.Score", ylab = "Age ")  
abline(lm(y ~ x), col = "blue")

