

# 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
## # 307,501 more rows
## # 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_TY
PE
## Length:307511          Length:307511          Length:307511          Length:307511
## Class :character        Class :character        Class :character        Class :characte
r
## Mode  :character        Mode  :character        Mode  :character        Mode  :characte
r
##
##
##
##
## REGION_POPULATION_RELATIVE          DAYS_BIRTH          DAYS_EMPLOYED          DAYS_REGISTR
ATION
## Min.   :0.00029          Min.   : -25229          Min.   : -17912          Min.   : -246
72
## 1st Qu.:0.01001          1st Qu.: -19682          1st Qu.:  -2760          1st Qu.:  -74
80
## Median :0.01885          Median : -15750          Median :  -1213          Median :  -45
04
## Mean   :0.02087          Mean   : -16037          Mean   :   63815          Mean   :  -49
86
## 3rd Qu.:0.02866          3rd Qu.: -12413          3rd Qu.:  -289          3rd Qu.:  -20
10
## 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_TY
PE
## Length:307511    Length:307511    Length:307511    Length:307511
## Class :character  Class :character  Class :character  Class :characte
r
## Mode  :character  Mode  :character  Mode  :character  Mode  :characte
r
##
##
##
## REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED  DAYS_REGISTR
ATION
## Min.   :1.000              Min.   :-25229  Min.   :-17912  Min.   :-246
72
## 1st Qu.:1.010              1st Qu.: -19682  1st Qu.: -2760  1st Qu.: -74
80
## Median :1.019              Median :-15750  Median : -1213  Median : -45
04
## Mean   :1.021              Mean   :-16037  Mean   : 63815  Mean   : -49
```

```

86
## 3rd Qu.:1.029          3rd Qu.: -12413    3rd Qu.:  -289    3rd Qu.: -20
10
## 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_RE
GION
## 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

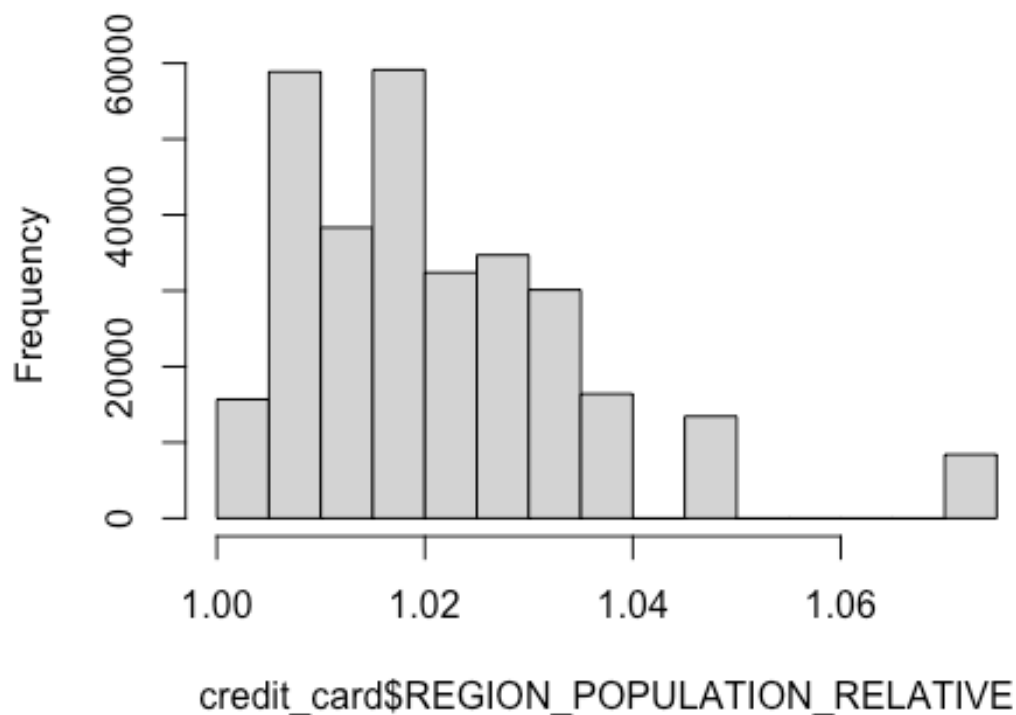
```

```
## 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)
```

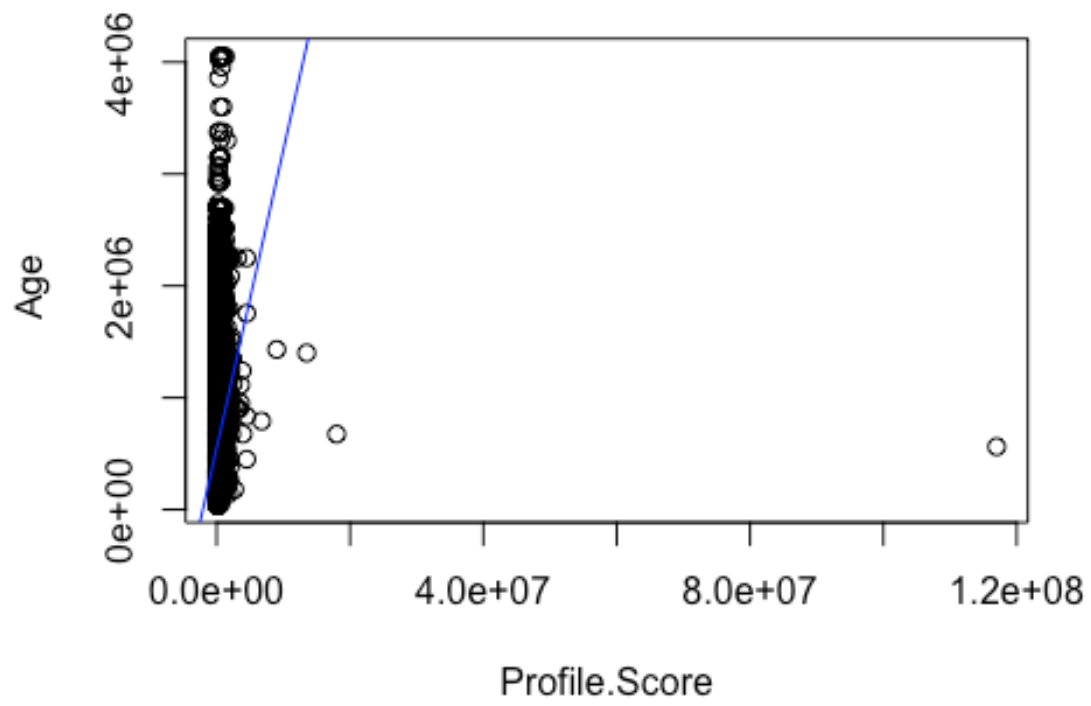
## stogram of credit\_card\$REGION\_POPULATION\_REL



#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")
```

**Profile.Score VS Age**



Filename: assignment.1.docx  
Directory: /Users/varshamekala/Library/Containers/com.microsoft.Word/Data/Documents  
Template: /Users/varshamekala/Library/Group Containers/UBF8T346G9.Office/User Content.localized/Templates.localized/Normal.dotm  
Title: fml  
Subject:  
Author: Varsha  
Keywords:  
Comments:  
Creation Date: 9/9/23 10:25:00 PM  
Change Number: 2  
Last Saved On: 9/9/23 10:25:00 PM  
Last Saved By: Mekala, Varsha Karunya  
Total Editing Time: 0 Minutes  
Last Printed On: 9/9/23 10:25:00 PM  
As of Last Complete Printing  
Number of Pages: 7  
Number of Words: 1,386  
Number of Characters: 14,312 (approx.)