FML_NIHARIKA

Niharika

2024-02-02

Data set source

https://www.kaggle.com/datasets/marslinoedward/bank-customer-churn-prediction

##Import the dataset into R

```
d <- read.csv("C:/Users/rajes/OneDrive/Desktop/NIharika/customer churn.csv")
```

```
head(d)
```

```
##   RowNumber CustomerId CreditScore Geography Gender Age Tenure   Balance
## 1         1   15634602         619    France Female  42      2      0.00
## 2         2   15647311         608     Spain Female  41      1  83807.86
## 3         3   15619304         502    France Female  42      8 159660.80
## 4         4   15701354         699    France Female  39      1      0.00
## 5         5   15737888         850     Spain Female  43      2 125510.82
## 6         6   15574012         645     Spain   Male  44      8 113755.78
##   NumOfProducts HasCrCard IsActiveMember EstimatedSalary Exited
## 1             1         1              1       101348.88      1
## 2             1         0              1       112542.58      0
## 3             3         1              0       113931.57      1
## 4             2         0              0        93826.63      0
## 5             1         1              1        79084.10      0
## 6             2         1              0       149756.71      1
```

##Print out descriptive statistics for a selection of quantitative and categorical variables.

```
summary(d)
```

```
##    RowNumber       CustomerId        CreditScore     Geography
##  Min.   :    1   Min.   :15565701   Min.   :350.0   Length:10000
##  1st Qu.: 2501   1st Qu.:15628528   1st Qu.:584.0   Class :character
##  Median : 5000   Median :15690738   Median :652.0   Mode  :character
##  Mean   : 5000   Mean   :15690941   Mean   :650.5
##  3rd Qu.: 7500   3rd Qu.:15753234   3rd Qu.:718.0
##  Max.   :10000   Max.   :15815690   Max.   :850.0
##     Gender               Age            Tenure          Balance
##  Length:10000       Min.   :18.00   Min.   : 0.000   Min.   :     0
##  Class :character   1st Qu.:32.00   1st Qu.: 3.000   1st Qu.:     0
##  Mode  :character   Median :37.00   Median : 5.000   Median : 97199
```

```
##                     Mean   :38.92   Mean   : 5.013   Mean   : 76486
##                     3rd Qu.:44.00   3rd Qu.: 7.000   3rd Qu.:127644
##                     Max.   :92.00   Max.   :10.000   Max.   :250898
##  NumOfProducts    HasCrCard      IsActiveMember   EstimatedSalary
##  Min.   :1.00   Min.   :0.0000   Min.   :0.0000   Min.   :    11.58
##  1st Qu.:1.00   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 51002.11
##  Median :1.00   Median :1.0000   Median :1.0000   Median :100193.91
##  Mean   :1.53   Mean   :0.7055   Mean   :0.5151   Mean   :100090.24
##  3rd Qu.:2.00   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:149388.25
##  Max.   :4.00   Max.   :1.0000   Max.   :1.0000   Max.   :199992.48
##      Exited
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2037
##  3rd Qu.:0.0000
##  Max.   :1.0000
```

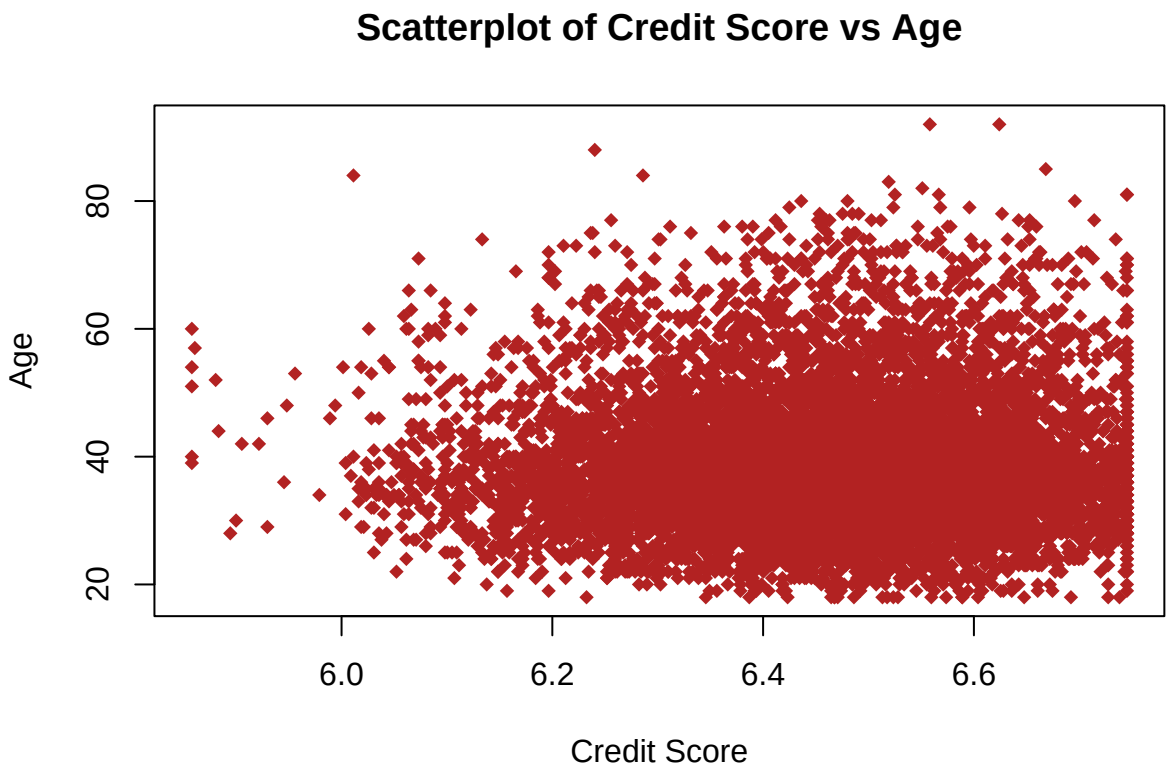
##transform at least one variable. It doesn't matter what the transformation is.

```
d$CreditScore <- log(d$CreditScore)
head(d$Age)
```

```
## [1] 42 41 42 39 43 44
```

##Plot at least one quantitative variable, and one scatterplot

```
plot(d$CreditScore, d$ Age ,main = "Scatterplot of Credit Score vs Age", xlab = "Credit Score", ylab =
```


Scatterplot of Credit Score vs Age
Age
20
40
60
80
6.0
6.2
6.4
6.6
Credit Score