# Poorvi_Raut_CS513_Midterm_Question2.R

Owner

2023-03-28

```
#knowledge Discovery and Data Mining (CS 513) Midterm Question 2:EDA Analysis
#Course : CS 513-A
# First Name : Poorvi
#Last Name : Raut
# ID : 20009560
# Purpose : Midterm: Exploratory Data Analysis (EDA)

#clearing object environment
rm(list = ls())
#get working directory
getwd()
```

```
## [1] "C:/Users/Owner/Desktop/Spring 2023/CS 513 KDD/PoorviRaut_CS513_Midterm"
```

```
#Load the "CS513_targeting_num.csv" from canvas into R and perform the EDA analysis
dataSet<-read.csv("/Users/Owner/Desktop/Spring 2023/CS 513 KDD/CS513_targeting_num.csv",na.string
= "?" )
View(dataSet)
#I. Summarizing each numerical column (e.g., min, max, mean)
summary(dataSet)
```

```
##   Customer.ID         Age          Gender              Income
##  Min.   : 1.00   Min.   :22.0   Length:80          Min.   : 1291
##  1st Qu.:20.75   1st Qu.:31.0   Class :character   1st Qu.: 3890
##  Median :40.50   Median :41.0   Mode  :character   Median : 6653
##  Mean   :40.50   Mean   :41.1                      Mean   : 7900
##  3rd Qu.:60.25   3rd Qu.:50.0                      3rd Qu.: 9215
##  Max.   :80.00   Max.   :65.0                      Max.   :20703
##                                                    NA's   :3
##    Purchase
##  Length:80
##  Class :character
##  Mode  :character
##
##
##
##
```

```
#II.    Identifying missing values
is.na(dataSet)
```

```
##       Customer.ID   Age Gender Income Purchase
##  [1,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [2,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [3,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [4,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [5,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [6,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [7,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [8,]       FALSE FALSE  FALSE  FALSE    FALSE
##  [9,]       FALSE FALSE  FALSE   TRUE    FALSE
## [10,]       FALSE FALSE  FALSE  FALSE    FALSE
## [11,]       FALSE FALSE  FALSE  FALSE    FALSE
## [12,]       FALSE FALSE  FALSE  FALSE    FALSE
## [13,]       FALSE FALSE  FALSE  FALSE    FALSE
## [14,]       FALSE FALSE  FALSE  FALSE    FALSE
## [15,]       FALSE FALSE  FALSE  FALSE    FALSE
## [16,]       FALSE FALSE  FALSE  FALSE    FALSE
## [17,]       FALSE FALSE  FALSE  FALSE    FALSE
## [18,]       FALSE FALSE  FALSE   TRUE    FALSE
## [19,]       FALSE FALSE  FALSE  FALSE    FALSE
## [20,]       FALSE FALSE  FALSE  FALSE    FALSE
## [21,]       FALSE FALSE  FALSE  FALSE    FALSE
## [22,]       FALSE FALSE  FALSE  FALSE    FALSE
## [23,]       FALSE FALSE  FALSE  FALSE    FALSE
## [24,]       FALSE FALSE  FALSE  FALSE    FALSE
## [25,]       FALSE FALSE  FALSE  FALSE    FALSE
## [26,]       FALSE FALSE  FALSE  FALSE    FALSE
## [27,]       FALSE FALSE  FALSE  FALSE    FALSE
## [28,]       FALSE FALSE  FALSE  FALSE    FALSE
## [29,]       FALSE FALSE  FALSE  FALSE    FALSE
## [30,]       FALSE FALSE  FALSE  FALSE    FALSE
## [31,]       FALSE FALSE  FALSE  FALSE    FALSE
## [32,]       FALSE FALSE  FALSE  FALSE    FALSE
## [33,]       FALSE FALSE  FALSE  FALSE    FALSE
## [34,]       FALSE FALSE  FALSE  FALSE    FALSE
## [35,]       FALSE FALSE  FALSE  FALSE    FALSE
## [36,]       FALSE FALSE  FALSE  FALSE    FALSE
## [37,]       FALSE FALSE  FALSE  FALSE    FALSE
## [38,]       FALSE FALSE  FALSE  FALSE    FALSE
## [39,]       FALSE FALSE  FALSE  FALSE    FALSE
## [40,]       FALSE FALSE  FALSE  FALSE    FALSE
## [41,]       FALSE FALSE  FALSE  FALSE    FALSE
## [42,]       FALSE FALSE  FALSE  FALSE    FALSE
## [43,]       FALSE FALSE  FALSE  FALSE    FALSE
## [44,]       FALSE FALSE  FALSE  FALSE    FALSE
## [45,]       FALSE FALSE  FALSE  FALSE    FALSE
## [46,]       FALSE FALSE  FALSE  FALSE    FALSE
## [47,]       FALSE FALSE  FALSE  FALSE    FALSE
## [48,]       FALSE FALSE  FALSE  FALSE    FALSE
## [49,]       FALSE FALSE  FALSE  FALSE    FALSE
## [50,]       FALSE FALSE  FALSE  FALSE    FALSE
## [51,]       FALSE FALSE  FALSE  FALSE    FALSE
## [52,]       FALSE FALSE  FALSE  FALSE    FALSE
```

```
## [53,]       FALSE FALSE  FALSE  FALSE    FALSE
## [54,]       FALSE FALSE  FALSE  FALSE    FALSE
## [55,]       FALSE FALSE  FALSE  FALSE    FALSE
## [56,]       FALSE FALSE  FALSE  FALSE    FALSE
## [57,]       FALSE FALSE  FALSE   TRUE    FALSE
## [58,]       FALSE FALSE  FALSE  FALSE    FALSE
## [59,]       FALSE FALSE  FALSE  FALSE    FALSE
## [60,]       FALSE FALSE  FALSE  FALSE    FALSE
## [61,]       FALSE FALSE  FALSE  FALSE    FALSE
## [62,]       FALSE FALSE  FALSE  FALSE    FALSE
## [63,]       FALSE FALSE  FALSE  FALSE    FALSE
## [64,]       FALSE FALSE  FALSE  FALSE    FALSE
## [65,]       FALSE FALSE  FALSE  FALSE    FALSE
## [66,]       FALSE FALSE  FALSE  FALSE    FALSE
## [67,]       FALSE FALSE  FALSE  FALSE    FALSE
## [68,]       FALSE FALSE  FALSE  FALSE    FALSE
## [69,]       FALSE FALSE  FALSE  FALSE    FALSE
## [70,]       FALSE FALSE  FALSE  FALSE    FALSE
## [71,]       FALSE FALSE  FALSE  FALSE    FALSE
## [72,]       FALSE FALSE  FALSE  FALSE    FALSE
## [73,]       FALSE FALSE  FALSE  FALSE    FALSE
## [74,]       FALSE FALSE  FALSE  FALSE    FALSE
## [75,]       FALSE FALSE  FALSE  FALSE    FALSE
## [76,]       FALSE FALSE  FALSE  FALSE    FALSE
## [77,]       FALSE FALSE  FALSE  FALSE    FALSE
## [78,]       FALSE FALSE  FALSE  FALSE    FALSE
## [79,]       FALSE FALSE  FALSE  FALSE    FALSE
## [80,]       FALSE FALSE  FALSE  FALSE    FALSE
```

```
print("Number of Missing Values")
```

```
## [1] "Number of Missing Values"
```

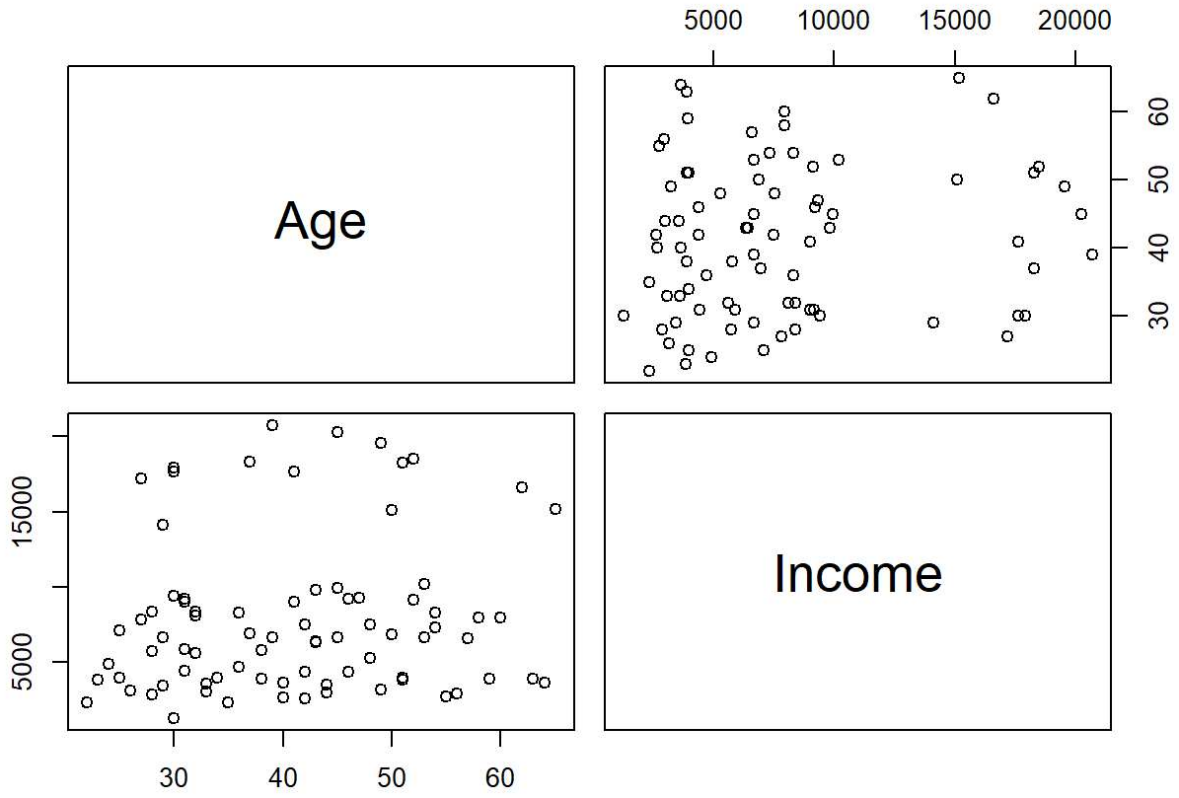
```
print(sum(is.na(dataSet)))
```

```
## [1] 3
```

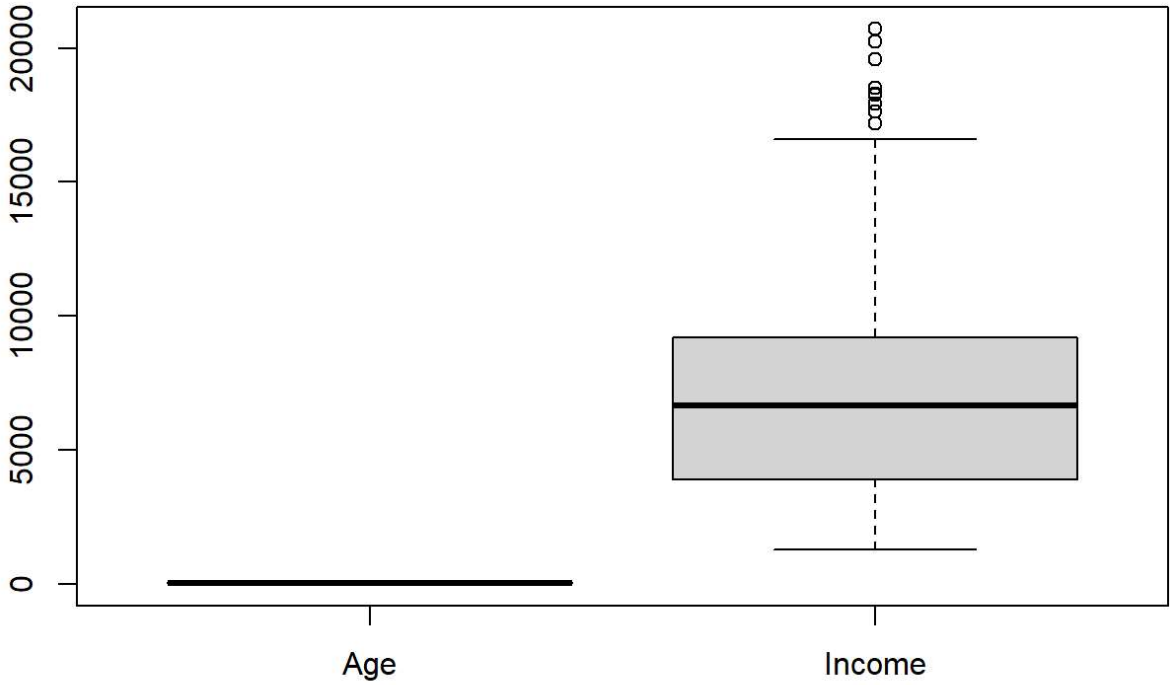
```
View(dataSet)
#III.   Replacing the numerical missing values with the “median” of the corresponding columns
median(dataSet$Income,na.rm=TRUE)
```

```
## [1] 6653
```

```
dataSet[is.na(dataSet$Income ),"Income"]<-median(dataSet$Income,na.rm=TRUE)
View(dataSet)
#IV.    Displaying the scatter plot of “Age”, and “Income”
pairs(dataSet[,c(2,4)])
```




```
#V. Show the box plots for columns: “Age” and “Income”
boxplot(dataSet[,c(2,4)])
```




```
##End of Question##
```