DATA SCIENCE MID PROJECT

**DATASET DESCRIPTION:**

Cardiovascular diseases (CVDs) are a major global cause of death, including conditions like coronary heart disease and cerebrovascular disease. The annual death toll from CVDs is 17.9 million, emphasizing the critical need to comprehend and forecast factors that contribute to heart attacks for public health. This dataset consists of 1319 samples, each defined by eight input variables—age, gender, heart rate, systolic and diastolic blood pressure, blood sugar level, CK-MB, and Test-Troponin. The ninth variable, 'class,' functions as the output, indicating the presence (positive) or absence (negative) of a heart attack. This dataset serves as a valuable tool for constructing models that can predict and elucidate the risk factors associated with cardiovascular events, contributing to advancements in preventive healthcare strategies.

**CODES – CONSOLE – DETAILS**

1. **IMPORT CSV FILE**

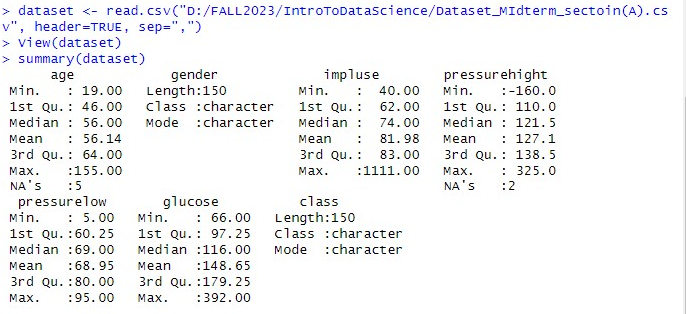
dataset <- read.csv("D:/FALL2023/IntroToDataScience/Dataset\_MIdterm\_sectoin(A).csv", header=TRUE, sep=",")



This code is used to import external Excel files (in CSV format) into R.

1. **SUMMARY KEYWORD**

summary(dataset)

 This code is used to show us a summary of the given dataset.

1. **MEAN VALUE OF AGE**

mean\_age\_value <- mean(dataset$age, na.rm = TRUE)

print(mean\_value)

A close up of a text

Description automatically generated

This code is used to calculate the mean value of the age.

1. **MEAN VALUE OF IMPULSE**

mean\_impulse\_value <- mean(dataset$impluse, na.rm = TRUE)

print(mean\_value)



This code is used to calculate the mean value of the impulse.

1. **MEAN VALUE OF PRESSURE HIGH**

mean\_ph\_value <- mean(dataset$pressurehight, na.rm = TRUE)

print(mean\_value)



This code is used to calculate the mean value of the high pressure.

1. **MEAN VALUE OF PRESSURE LOW**

mean\_pl­\_value <- mean(dataset$pressurelow, na.rm = TRUE)

print(mean\_value)



This code is used to calculate the mean value of pressure low.

1. **MEAN VALUE OF GLUCOSE**

Mean\_glucose\_value <- mean(dataset$glucose, na.rm = TRUE)

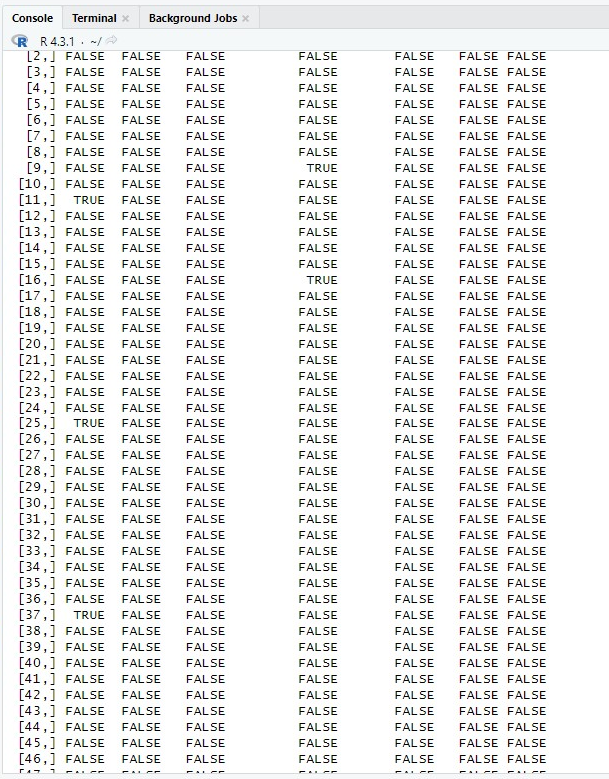
print(mean\_value)



This code is used to calculate the value of glucose.

1. **MEDIAN**
2. **FINDING MISSING VALUES**

is.na(dataset)

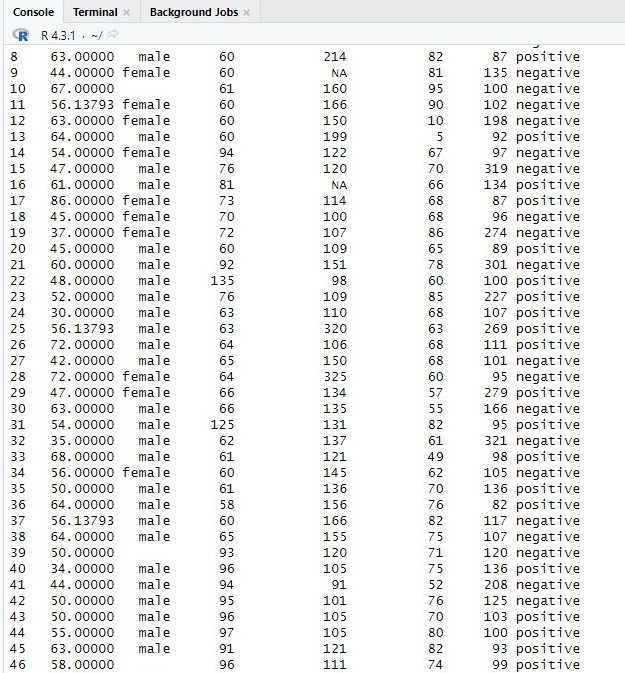


This code is used to identify the missing values in a dataset. The ones that has TRUE written in it is a missing value.

1. **FIXING MISSING VALUE**

dataset[is.na(dataset$age), "age"] <- mean\_age\_value

print(dataset)



This code is used to replace the missing value in the age column with the average mean value of age.

1. **HISTOGRAM**

barplot(table(dataset$gender), main = "Gender Distribution", xlab = "Gender", ylab = "Count", col = c("cyan", "pink"))

This code is

1. **MEDIAN FOR PRESSURE HIGH**

median\_ph\_value <- median(dataset$pressurehight, na.rm = TRUE)

print(median\_ph\_value)



This code is used to find the median of pressure high.

1. **MISSING VALUE**

is.na(dataset$pressurehight)

which(is.na(dataset$pressurehight))

dataset[is.na(dataset$pressurehight), "pressurehight"] <- median\_ph\_value

print(dataset)

A screenshot of a computer

Description automatically generated

This code is used to first detect the missing value, then replace it with the median value of pressure high.

1. **AGE RANGE**

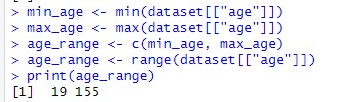
min\_age <- min(dataset[["age"]])

max\_age <- max(dataset[["age"]])

age\_range <- c(min\_age, max\_age)

age\_range <- range(dataset[["age"]])

print(age\_range)



This code is used to find the range of age.

1. **IMPULSE RANGE**

min\_impluse <- min(dataset[["impluse"]])

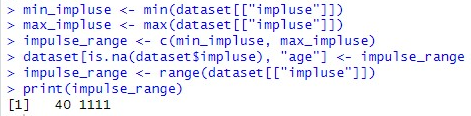
max\_impluse <- max(dataset[["impluse"]])

impulse\_range <- c(min\_impluse, max\_impluse)

dataset[is.na(dataset$impluse), "age"] <- impulse\_range

impulse\_range <- range(dataset[["impluse"]])

print(impulse\_range)



This code is used to find the range of impulse.

1. BOX PLOT FOR AGE

hist(dataset$age,main="Age Distribution", xlab="Age", xlim = c(0,300),ylim=c(0,30), breaks=30)

boxplot(dataset$age, main = "Age Distribution", ylab = "Age")

