Annotating datasets:

dataset$Infection<-factor(dataset$Infection,levels = c("yes","no","marginal"),labels = c(1,2,3))

dataset

dataset$RiskLevel<-factor(dataset$RiskLevel,levels = c("high risk","low risk","mid risk"),labels = c(1,2,3))

dataset

summery of the structure of data set:

str(dataset)

column name of data set:

names(dataset)

Descriptive Statistics Using summary() Function:

Summary(dataset)

Counting number of Null values in each column:

colSums(is.na(dataset))

Finding the specific row number of Null values.

which(is.na(dataset$Age))

which(is.na(dataset$Infection))

which(is.na(dataset$Smoking))

which(is.na(dataset$SystolicBP))

which(is.na(dataset$DiastolicBP ))

which(is.na(dataset$BS))

which(is.na(dataset$BodyTemp))

which(is.na(dataset$HeartRate))

which(is.na(dataset$RiskLevel))

Remove null values from data set.

remove<-na.omit(dataset)

Filling missing values with mean:

filled\_data\_mean <- dataset

filled\_data\_mean$Age <- ifelse(is.na(filled\_data\_mean$Age), mean(filled\_data\_mean$Age, na.rm = TRUE), filled\_data\_mean$Age)

filled\_data\_mean$Infection<- ifelse(is.na(filled\_data\_mean$Infection), mean(filled\_data\_mean$Infection, na.rm = TRUE), filled\_data\_mean$Infection)

filled\_data\_mean$Smoking<- ifelse(is.na(filled\_data\_mean$Smoking), mean(filled\_data\_mean$Smoking, na.rm = TRUE), filled\_data\_mean$Smoking)

filled\_data\_mean$SystolicBP<- ifelse(is.na(filled\_data\_mean$SystolicBP), mean(filled\_data\_mean$SystolicBP, na.rm = TRUE), filled\_data\_mean$SystolicBP)

filled\_data\_mean$DiastolicBP<- ifelse(is.na(filled\_data\_mean$DiastolicBP), mean(filled\_data\_mean$DiastolicBP, na.rm = TRUE), filled\_data\_mean$DiastolicBP)

filled\_data\_mean$BS<- ifelse(is.na(filled\_data\_mean$BS), mean(filled\_data\_mean$BS, na.rm = TRUE), filled\_data\_mean$BS)

filled\_data\_mean$BodyTemp<- ifelse(is.na(filled\_data\_mean$BodyTemp), mean(filled\_data\_mean$BodyTemp, na.rm = TRUE), filled\_data\_mean$BodyTemp)

filled\_data\_mean$HeartRate<- ifelse(is.na(filled\_data\_mean$HeartRate), mean(filled\_data\_mean$HeartRate, na.rm = TRUE), filled\_data\_mean$HeartRate)

filled\_data\_mean$RiskLevel<- ifelse(is.na(filled\_data\_mean$RiskLevel), mean(filled\_data\_mean$RiskLevel, na.rm = TRUE), filled\_data\_mean$RiskLevel)

print(filled\_data\_mean)

fill missing values with the median:

filled\_data\_median <- dataset

filled\_data\_median$Age <- ifelse(is.na(filled\_data\_median$Age), median(filled\_data\_median$Age, na.rm = TRUE), filled\_data\_median$Age)

filled\_data\_median$Infection <- ifelse(is.na(filled\_data\_median$Infection), median(filled\_data\_median$Infection, na.rm = TRUE), filled\_data\_median$Infection)

filled\_data\_median$Smoking <- ifelse(is.na(filled\_data\_median$Smoking), median(filled\_data\_median$Smoking, na.rm = TRUE), filled\_data\_median$Smoking)

filled\_data\_median$SystolicBP <- ifelse(is.na(filled\_data\_median$SystolicBP), median(filled\_data\_median$SystolicBP, na.rm = TRUE), filled\_data\_median$SystolicBP)

filled\_data\_median$DiastolicBP <- ifelse(is.na(filled\_data\_median$DiastolicBP), median(filled\_data\_median$DiastolicBP, na.rm = TRUE), filled\_data\_median$DiastolicBP)

filled\_data\_median$BS <- ifelse(is.na(filled\_data\_median$BS), median(filled\_data\_median$BS, na.rm = TRUE), filled\_data\_median$BS)

filled\_data\_median$BodyTemp <- ifelse(is.na(filled\_data\_median$BodyTemp), median(filled\_data\_median$BodyTemp, na.rm = TRUE), filled\_data\_median$BodyTemp)

filled\_data\_median$HeartRate <- ifelse(is.na(filled\_data\_median$HeartRate), median(filled\_data\_median$HeartRate, na.rm = TRUE), filled\_data\_median$HeartRate)

filled\_data\_median$RiskLevel <- ifelse(is.na(filled\_data\_median$RiskLevel), median(filled\_data\_median$RiskLevel, na.rm = TRUE), filled\_data\_median$RiskLevel)

print(filled\_data\_median)

Filling missing values with mode: