**VIT-AP UNIVERSITY, ANDHRA PRADESH**

**CSE2047 – Data Analytics - Lab Sheet : 2**

**Academic year:** 2020-2021  **Branch/ Class:** B.Tech/M.Tech

**Semester:** Fall  **Date:**

**Faculty Name:** Prof. S.Gopikrishnan  **School:** SCOPE

**Student name:** Valiveti Manikanta bhuvanesh **Reg. no.: 19BCD7088**

**LAB 2**

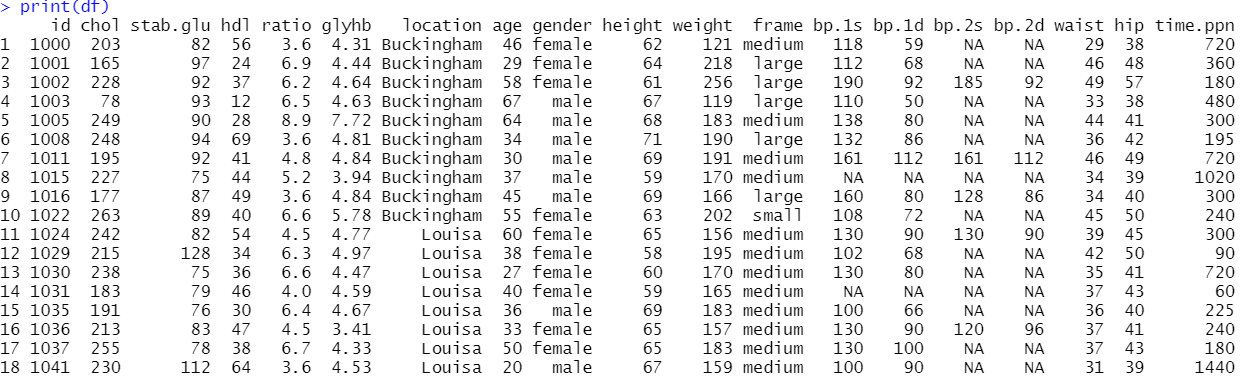
1. **USE DIABETES.CSV**

df<-read.csv("Diabetes\_Updated.csv")



1. **DISPLAY THE DATAFRAME**

print(df)



1. **HOW MANY ROWS AND COLUMNS ARE THERE?**

dim(df)



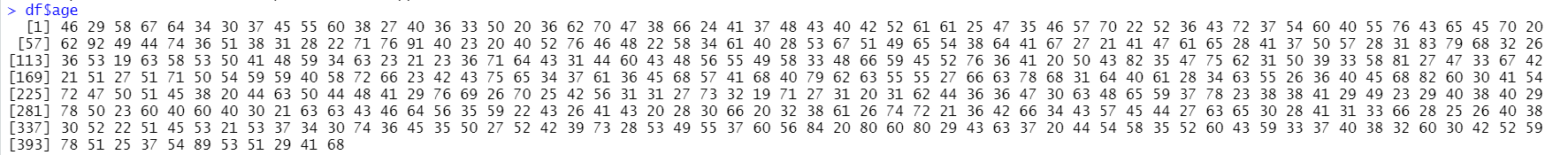
1. **FIND OUT THE COLUMNS NAMES IN THE DATAFRAME**

colnames(df)



1. **ACCESS THE AGE COLUMN.**

df$age



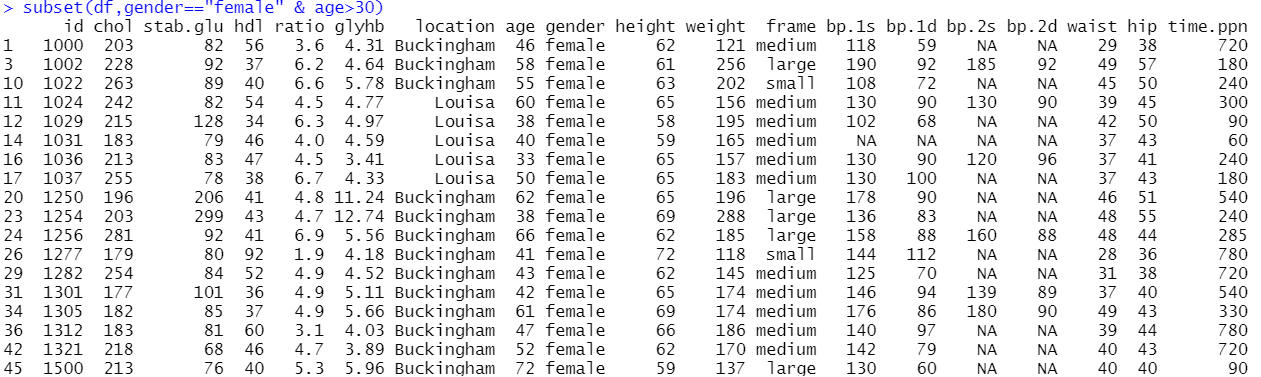
1. **DISPLAY THE NUMBER OF PEOPLE WHOSE AGE IS GREATER THAN 40.**

dim(subset(df,age>40))[1]



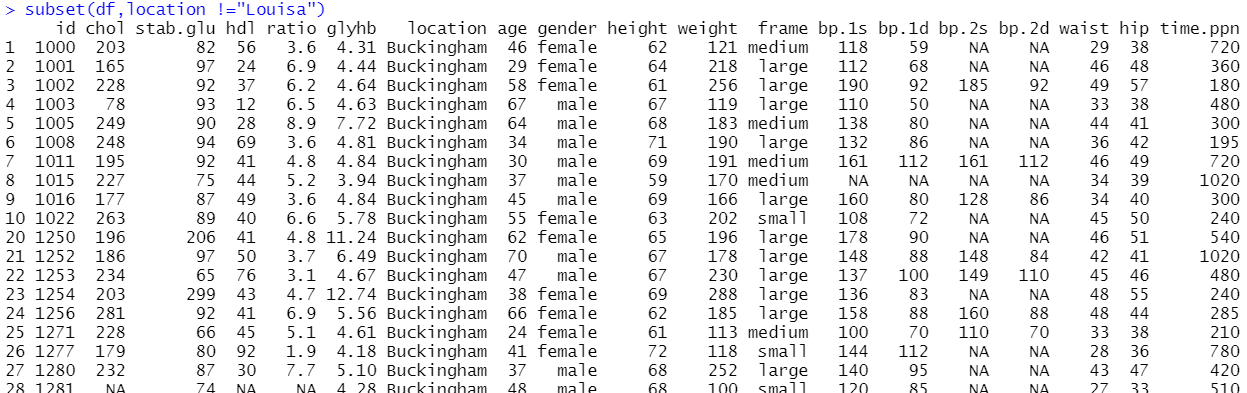
1. **FIND OUT THE FEMALE DIABETIC PATIENTS OF AGE > 30**

subset(df,gender=="female" & age>30)



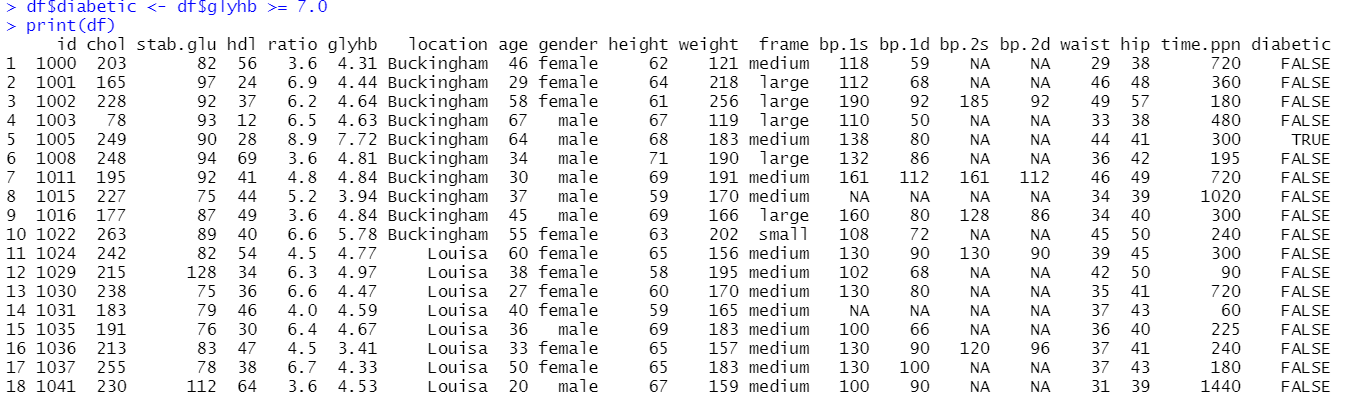
1. **FIND OUT THE DETAILS OF PATIENTS WHO ARE NOT FROM LOUSIA.**

subset(df,location !="Louisa")



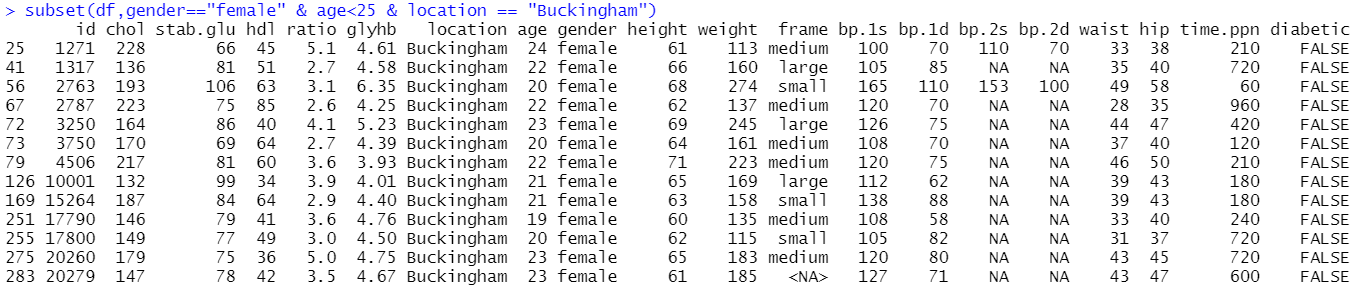
1. **IF GLUCOSE LEVELS IN BLOOD IS > 7, DIAGNOSE AS DIABETIC BY ADDING A COLUMN TO THE DATA FRAME.**

df$diabetic <- df$glyhb >= 7.0



1. **WHICH FEMALE SUBJECTS FROM BUCKINGHAM ARE UNDER THE AGE OF 25?**

subset(df,gender=="female" & age<25 & location == "Buckingham")



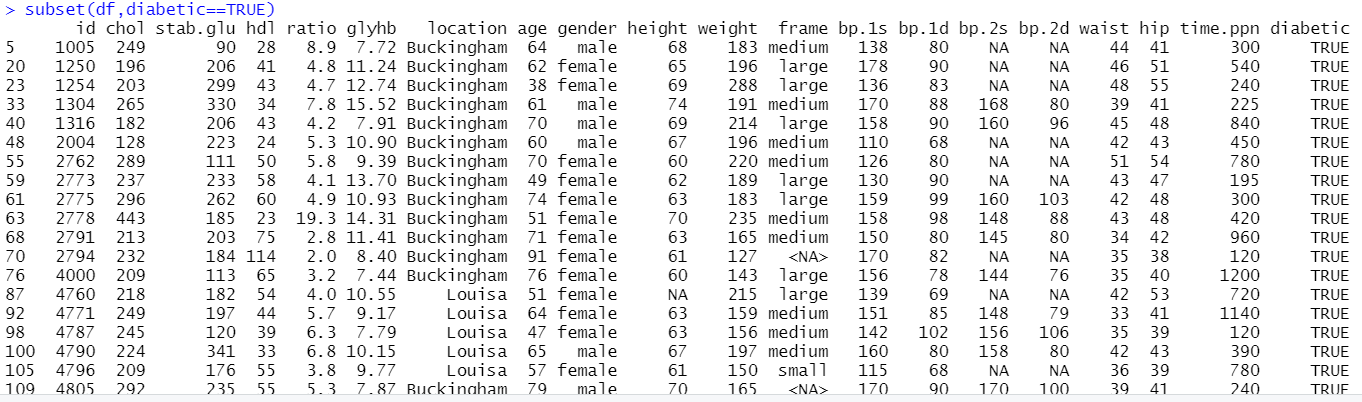
1. **WHAT IS THEIR AVERAGE GLYHB?**

mean(df$glyhb,na.rm = TRUE)



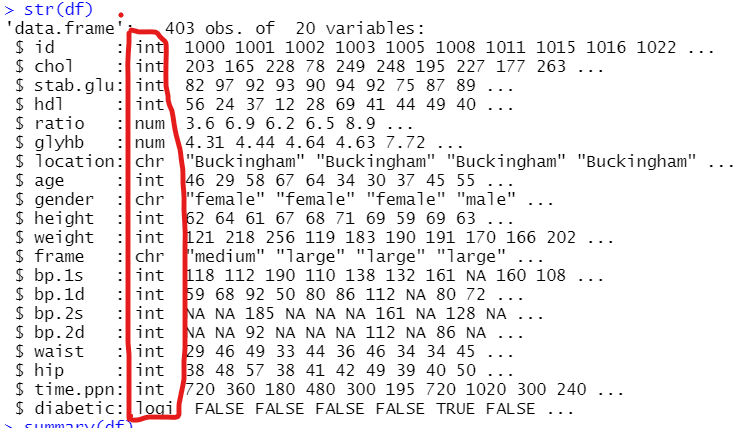
1. **ARE ANY OF THEM DIABETIC?**

subset(df,diabetic==TRUE)



1. **FIND OUT EACH COLUMN TYPE IN THE DATAFRAME**

str(df)



1. **PRODUCE THE SUMMARY OF THE DATAFRAME.**

summary(df)

