EXPERIMENT -3

**Aim:** Given a case study of predicting diabetics from a data set. You are expected to explore the domain, types of attributes, what does each attribute signifies and their ranges etc.

**Theory:**

Attributes:

1. Number of times pregnant

2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test

3. Diastolic blood pressure (mm Hg)

4. Triceps skin fold thickness (mm)

5. 2-Hour serum insulin (mu U/ml)

6. Body mass index (weight in kg/(height in m)^2)

7. Diabetes pedigree function

8. Age (years)

9. Class variable (0 or 1)

R Code

diabetes <- read.csv("~/diabetes.csv", stringsAsFactors = FALSE)

View(diabetes)

#structure of the data

str(diabetes)

#summary of data

summary(diabetes)

# attributes of data

attributes(diabetes)

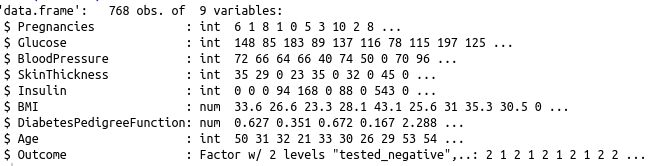
colSums(is.na(diabetes))

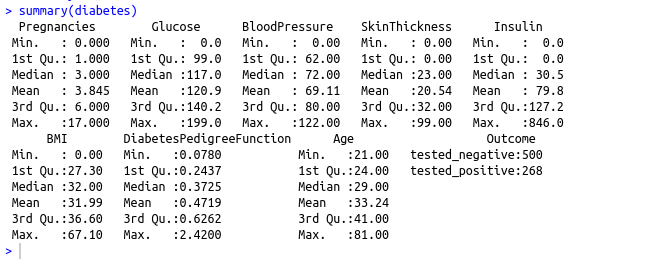
#no of rows

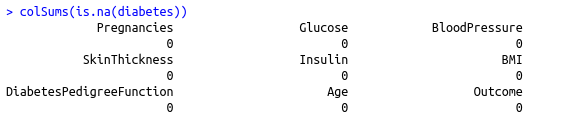
nrow(diabetes)

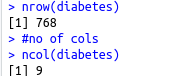
#no of cols

ncol(diabetes)









**Conclusion: Hence we have successfully explored the dataset**