# LAB: 4

SUBMITTED BY: Kush Bhardwaj (110070548)

SUBMITTED TO: Prof. Shafaq Khan

# Part 1: Data Exploration

1. Import the diabetes datasets.

Graphical user interface, text, application, email

Description automatically generated

2. Summarize that diabetes dataset and explain the output.

Graphical user interface, text, application

Description automatically generated

The output gives the overall summary for each column including the min value, the lower quadrille (the value that cuts off the first 25% of the data when it is sorted in ascending order), median, mean, top quadrille, max value, and total number of NA values. This summary helps the user to understand the dataset more easily.

3. Show the structure and dimension of diabetes dataset and explain it.

Text

Description automatically generated with medium confidence

For the given data frames, total number of values are 768 inside 9 columns(variables). Each variable has different datatype such as int, num, chr.

4. Show the first 10 rows of the diabetes dataset.

Table

Description automatically generated

5. Show the column names of diabetes dataset.

Graphical user interface, text, application

Description automatically generated

# Part 2: Data Pre-Processing

6. What is the class variable in diabetes dataset?

Graphical user interface, application

Description automatically generated

7. Change the class type of the class variable of diabetes dataset to factor. Show the output after the conversion.

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated with low confidence

8. Find which columns contain missing values in diabetes dataset. What are the total missing values for each column?

Graphical user interface, application

Description automatically generated

9. Replace the missing values in diabetes by 0. Check what if the missing values was replaced successfully.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

# Part 3: Data Visualization

10. Create a scatter plot. The plot should show the relationship between any two variables of your choice. Add labels, title, and color to the plot.

Graphical user interface, text, application

Description automatically generated