# Lab Mid Term Exam

**Class:** **BSCS-8B**

**Subject:** **Data Science**

**Lab Instructor:** **Muhammad Imran**

**Exam Day:** **Friday (25-March-2022)**

**Exam Duration:** **2 Hours**

**Total Marks:** **50**

---

**Note: You are required to perform all the tasks in R Studio and submit a PDF file containing your code, output and explanation. ALSO Submit the Screenshot of your working console for every question (Do not crop the Screenshot). Late submissions will not be graded.**

**Exam contains 5 questions and all are compulsory.**

## Task 1: 2.5 Marks

Create a Vector and save 6 values 1 ,2, 3,4,5,6 your id digits, Assign names to each instance and display length and type of vector.

## Task 2: 2.5 Marks

Create a matrix "YourName" and store 9 values row wise with 3 rows and name the row and column names.
Multiply Complete Matrix with 100 and Display the output.

## Task 3: 15 Marks

- Read the dataset provided to you, "DataSet_MidTerm".
- Display the structure and summary of the dataset and explain it.
- Identify the categorical and numerical attributes in the dataset.

- Plot a categorical graph using the categorical attribute from the dataset provided to you.
- Plot a numerical graph using any numerical attribute from the dataset.
- Plot 2X numerical graph using multiple numerical attribute from the dataset, Set the name of the plot according to the attributes that you are using and also set the names of 'x' and 'y' axis accordingly.
- Plot histogram of any attribute from the dataset, the color of histogram should be green.
- Increase the number of breaks in histogram to 10 and visualize it.

**Task 4:** **15 Marks**

- Display the top 5 entries of the dataset you used in "Task 3".
- Display the last 5 entries of the dataset you used in "Task 3".
- Find the Summary and Five number summary of the dataset and explain the difference between them.
- Plot the Box plot of the variable "Report Epi Week" from the dataset and explain the output also explain the outliers in it.

**Task 5:** **15 Marks**

- Encode the character variables/categorical variables to numerical values.
- Fill the missing values with mean of that variable.
- Split the dataset in to testing and training sets with ratio of 80/20.

---------- **End of Paper** ----------

---------- ☺ **Best of Luck** ☺ ----------