

**DYAL SINGH COLLEGE
PRACTICAL ASSIGNMENT**

**COURSE: B.Sc. (Hons.) Computer Science VI Semester
SUBJECT: Introduction to Data Science**

Time Duration: - 4 Hours

Max. Marks: 25

GUIDELINES

1. Assignment should be handwritten with your examination roll number on each page. First page should have the following details:
Examination Roll. No.: _____
Name of Student: _____
Course: _____
Semester: _____
Subject: _____
Unique Paper Code: _____
2. Rename the scanned copy of your assignment with your name_examinationrollno.
3. Send the scanned copy of your assignment on the reply all option of the same mail on which you received the assignment or send to both the given email ids- mrinaliyadav@dsc.cu.ac.in and goel.anita@dsc.du.ac.in .
4. Upload the assignment on the given dropbox link- <https://www.dropbox.com/request/OoV1fpof6TOdeDIKa1Zb>.

Question:

From the Bollywood dataset given as .csv file, write R script for the following:

- i) Find total number of missing values and replace NA with median value of the same column.
- ii) Find the average opening collection according to the column “Genre”. Create a suitable plot to visualize the result. x-label: “Average”; y-label: “Genre” and title: “Exam roll number”.
- iii) Create a subset of the dataset for all the “Super Hit” movies.
- iv) Write a R function to return the name of movies whose “Total Collection” is more than the average Total Collection.

Write a R script to simulate a sample of 20 random numbers in the range- 10 to 500 and store it in a vector. Create a scatter plot for the vector, in blue color and draw vertical lines in red and green color corresponding to mean and median, respectively.