**Programming for Big Data**

**Assignment 3**

**Due date : 03-05-24**

Provided is a dataset of students and you have to apply preprocessing on data by performing following steps

* first you have to fill a dataframe with the provided dataset
* assign headings (Attribute\_1: student\_id

Attribute\_2: Test\_Score

Attribute\_3: Hours\_Studied

Attribute\_4: GPA

Attribute\_5: Attendance

Attribute\_6: Scholarship%

Attribute\_7: Extracurricular\_Activities)

* Show total number of missing values in all columns
* Then replace missing values in hours\_studied by the median of this column.
* GPA should be filled by the most frequent value
* Apply normalization to handel Attendance that should be considered 6 days/week instead of 5 days/week
* draw a bar plot between student\_id, hours\_studied and gpa
* draw an area/density plot to show the progress of student gpa w.r.t hours\_studied