**Assignment-1**

**Submission date: 19/01/2020**

Perform the following tasks:

Prepare data set as follows using Python:

-create an unstructured text file with USN (4JN18CS..), Name(random valid Indian name-don’t duplicate), Group-1elective (randomly any one subject in group of 2), Group-2 elective (randomly any one subject in group of 2), Marks scored for Group-1 elective chosen (random 0-100), Marks scored for Group-2 elective chosen (random 0-100)

Write SparkContext based Pyspark code using RDD to answer following questions:

1. Which elective was most chosen
2. Which combination of electives was most chosen
3. Which elective gave good performance (use mean and standard deviation)
4. Which combination of electives gave good performance (use mean and standard deviation)
5. Which range of marks was highest in each elective and overall (assume range of class interval 5 like 50-55 etc)

Submit the code and databricks snapshot in a word file and upload to google class