HOMEWORK ASSIGNMENT: WEEK 2

Assignment 01:

1. For the given input file (Module_4_Ex1.txt), find the maximum scorer in each gender and three age categories: less than 20, 20 to 50, greater than 50.

   Hint: You would like to use a custom partitioner for it

2. For the given input file (Module_4_Ex2.txt) find out the number of records, where second field is either less than 10 or more than 50 without writing the output to HDFS.

   Hint: You might like to just profile the data with counter

3. Given you have a stores location file (store_locations.txt) and a stores sales file (store_sales.txt)

   The store locations file has two fields (store location, store name) and store sales file has two fields (store location, sales amount)

   Join (Equi-Join) the two files based on store location and produce the following fields in output:
   <store location>, <store name>,<sales amount>

4. Repeat exercise 3 with Map Side Join.