Evaluation Lab 3 - MR

Implementing Counter
1) Download an attachment in STAGING_AREA location
2) Extract it
3) Pull it to LABS_AREA/counter
4) Study the code
5) Create counties folder in your home on HDFS
6) Put all counties_*.csv files from local linux file system to counties folder on HDFS
7) Run the application with command as yarn jar average.jar <Main Class>
8) See the output of the job on terminal window and confirm whether counters you created are working or not
Study the code

Code/Dataset:-
data/Average.rar

Implement following business requirement for Average application above where
1) Business wants to know how many records are Good records and
how many are bad.
2) It should notify client about the number of good and bad records
3) Bad records should not be processed
4) According to business Good records are the records where the row will have
    i) Only 12 fields
    ii) The 10th value (field) in every row will have a numerical value
5) If above criterias are not matched then the records should be treated as Bad records
6) Automate the workflow where building the java application will also be the part of an automation process.
Please refer attached MR_Workflow.rar for the same.

Note - Dataset is already shared with participants