HW5A -Due date 6th Nov

MongoDB

|  |  |
| --- | --- |
| Instructions | Points |
| * Download “[Books 5-core (8,898,041 reviews)](http://snap.stanford.edu/data/amazon/productGraph/categoryFiles/reviews_Books_5.json.gz)” From the link “<http://jmcauley.ucsd.edu/data/amazon/>” and insert directly to MongoDB using mongoimport command | 0%  (\*required) |
| * Use ‘OR’ operation to show only the ‘summary’ and ‘overall’ columns with overall rating 1 or 3 or 5. | 20% (Student 1) |
| * Use ‘AND’ operation to show only the ‘reviewerName’ and ‘overall’ columns with overall rating **less than 3** and **greater than 1** and sort the result by the ascending alphabetic order of ‘reviewerName’. | 20% (student 2) |
| * Group by ‘reviewerName’ to show the **minimum** ‘overall’ rating they posted. | 15% (joint) |
| * Group by ‘Helpful’ to show the ‘total number’ of entries found for different ‘Helpful’ data, sorted by descending order of the ‘total number’. | 20% (student 3) |
| * Show ‘Review Text’ that contains at least one numeric value. | 15% (joint) |

*\*\*\* The dataset is very big, and queries will take time to execute. You may need to try several times to import the whole dataset. You also need to handle the errors generated due to memory outage. Search how to “index a column” and how to use “allowDiskUse” command to handle those errors.*