

Problem Statement:

Congratulations! Now that you have created the output file containing the best authors for every year. Let's open the file and have a look at it. When we write the output to a file from Apache Flink, it doesn't add the headers on its own. So, in this exercise, you need to add headers to the output CSV file. Modify the same code you have done for the earlier exercise to come up with the headers.

Input & output criteria:

The Dataset contains a single file: books.csv. The file contains the following columns:

Id, Name, Authors, ISBN, Rating, PublishYear, PublishMonth, PublishDay, Publisher, RatingDist5, RatingDist4, RatingDist3, RatingDist2, RatingDist1, RatingDistTotal, CountsOfReview, Language, pagesNumber

You have to read this file from your respective S3 account and then write the output into another CSV file in local file storage. The Output file should contain the following columns, in the same order:

Year, BestAuthor, TotalRatings, Rating

Program argument(s):

The flink program needs to be provided with the input & output file paths. A sample program argument will be like this:

```
--input s3:///Users/mayukh/books.csv --output file:///Users/mayukh/books_output.csv
```

Input path will be the S3 path, where you have uploaded the file earlier. The output path can be any path in your local file system.

Good Luck!