Problem Statement:

You have been given a dataset of books released over various years. The CSV file contains various information about the books: it's title, author, release year, the rating statistics, publisher and several other attributes.

You have to come up with the best author's name for every year. The criteria for selection of an author is that the author's books should get at least 10,000 ratings in that particular year. The author, whose books in that year got a collectively highest rating(weighted rating of all the books written by the author in that year) will be nominated as the best author for that year.

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 (So, please upload the file into S3 before running the program) 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!