

StackOverflow

You are given a CSV file (/user/adampap/stackoverflow) with information about StackOverflow posts. Each line in the provided text file has the following format:
<postTypeId>,<id>,[<acceptedAnswer>],[<parentId>],<score>,[<tag>]

A short explanation of the comma-separated fields follows:

- <postTypeId>: Type of the post. Type 1 = question, type 2 = answer.
- <id>: Unique id of the post (regardless of type).
- <acceptedAnswer>: For a question: id of the accepted answer post; an empty string indicates that the issue has not been solved yet. For an answer: missing, indicated by an empty string.
- <parentId>: For an answer: id of the corresponding question. For a question: missing, indicated by an empty string.
- <score>: The StackOverflow score (based on user votes).
- <tag>: The tag indicates the programming language that the post is about, in case it's a question, or missing in case it's an answer.

The expected outcome:

the number of solved questions (sorting column)					
tag	the average number of answers per solved question				
Java	3	2.0		1	0.0
PHP	2	1.5		0	NaN
Python	1	3.0		1	2.0
C#	0	NaN		2	3.5

the average number of answers per unsolved question

the number of unsolved questions