

- How did you develop the final blocker? What blocker did you start with? What problems did you see? Then how did you revise it to come up with the next blocker? In short, explain the \*development process\*, from the first blocker all the way to the final blocker (that you submit in the Jupyter file).

Ans: Started with equivalence blocking based on 'Runtime' (Only 1.62% of values were missing, so it was a good candidate). Then did overlap blocking with overlap size = 2 on 'Title' column. This returned 251 matching pairs. This should have near to 0 false positives (w.r.t. to correct matching)

- If you use Magellan, then did you use the debugger? If so, where in the process? And what did you find? Was it useful, in what way? If you do not use Magellan, you can skip this question.

Ans: Did not use the debugger.

- How much time did it take for you to do the whole blocking process?

Ans: Blocking process alone: 10 minutes. Preparing table (refining) for blocking + getting familiar with Jupyter notebook system took about 2-3 hours

- Report the size of table A, the size of table B, the total number of tuple pairs in the Catersian product of A and B, and the total number of tuple pairs in the table C.

Ans: Size of table A:  $7415 * 10$

Size of table B:  $5008 * 10$

Size of A \* B: 37134320

No. of tuple pairs in table C: 251

- Did you have to do any cleaning or additional information extraction on tables A and B?

Ans: Had to do some data refinement (standardisation) in Table B (had to make 'Runtime' column format consistent with that of table A)

- Did you run into any issues using Magellan (such as scalability?). Provide feedback on Magellan. Is there anything you want to see in Magellan (and is not there)? If you do not use Magellan, you can skip this question.

Ans: Did not run into any issues as such. With the relatively small scale of data in this project (7415 rows in table A and 5008 rows in table B) it's not fair to comment much about scalability issues.

- Any other feedback is appreciated.

Ans: Good work!