- 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).
  We started with Attribute equivalence blocking on “released time” but we found that it did not achieve high precision, and so we added overlap blocker on top of the candidate set we got from the previous blocker.
  In general, we first applied Attribute equivalence blocking on “released time”, then we added overlap blocker on top of the candidate set we got from the previous blocker with matching ”Artist” attribute in two tables, getting an result C2. Next step, we tried to put limits on “TrackName” attribute in which we set “lev_sim(ltuple, rtuple) < 0.8”, getting an result C3. And finally we combined the result C2, and C3 to build our final candidate set.

- 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.
  We used the Magellan debugger when we tested the appropriate threshold value for setting level similarity between “TrackName” attribute across two tables. If in the debugger table we still find tuples pairs that should have been matched, then we continue to go down the rows until we are unable to find any more tuples pair matches. We recorded the bottom tuple pairs that have the least similarity score as our new threshold value.

- How much time did it take for you to do the whole blocking process?
  It takes about 15 mins for the blocking process in total.

- Report the size of table A, the size of table B, the total number of tuple pairs in the Cartesian product of A and B, and the total number of tuple pairs in the table C.
  size of table A: 3019
  size of table B: 3056
  total number of tuple pairs in the Cartesian product of A and B:9226064
  total number of tuple pairs in the table C: 383

- Did you have to do any cleaning or additional information extraction on tables A and B?
  Since we have done data cleaning in the previous stage, we did not have to any cleaning in current stage.
  We did not do additional information extraction either.

- 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.
  We found the rule-based blocking sometimes does not efficiently block the results. It will also block some tuple pairs that has similarity scores larger than threshold value in our case. This needs to be fixed.

- Any other feedback is appreciated.
  Overall, the Magellan debugger is really useful for debugging and it supports many blockers, making the whole process much easier.