Answer Key - De-duplicating Data Tutorial

After viewing the De-duplicating Data Tutorial, use the steps outlined to check the 'messy' dataset and resolve any de-duplication. 'Answers' can be found below.

For UID_C variable (Column B):

- There are three duplicate UID_C and one triplicate UID_C. First you will need to establish that these are unique respondents by comparing answers. Once you have done that, to distinguish between the records, you will need to add a digit after the matching UID. Then you will know that any 5 digit UID is due to its second appearance, and it won't look like the same person took the survey twice in your analysis.
    - For UID 36 and 40 change AS01 to AS011 and the other AS01 to AS012
    - For UID 2, 8, and 23 change CR84 for all three to CR841, CR842, and CR843
    - For UID 56 and 74 change GM72 to GM721 and the other GM72 to GM722
- If you are unable to establish that the responses are unique (I.e. all answers match) then you likely have a duplicate entry, where the same person responded twice or the survey was entered twice. You will want to delete exact matches (and make sure that you document your deletion from the dataset somewhere for reference related to your data cleaning methods).
    - For UID 45 and 101, delete row 101 as it is an exact duplicate