

Audible Data Cleaning Project

Dataset Link:

<https://drive.google.com/file/d/1yjyozaSrwShoaROq-TDuSgC5HNLLmrTE/view?usp=sharing>

Case Study Description:

Utilize Power Query Editor in Excel to clean and standardize an Audible dataset. Tasks include formatting columns, ensuring data consistency, and preparing the dataset for analysis.

1. Standardize the name column to ensure consistent title casing.
2. Separate combined names in the author column if there are multiple authors.
3. Ensure all entries in the releasedate column follow a consistent date format (DD-MM-YYYY).
4. Convert the time column from text format to a duration format that Excel recognizes.
5. Ensure the price column is in a numeric format, and identify any non-numeric values.
6. Convert text ratings in the stars column to numeric values.
7. Split the narratedby column into multiple columns if multiple narrators are listed.
8. Merge the releasedate and language columns into a single new column named releaseinfo with the format "DD-MM-YYYY, Language."
9. Ensure all currency values in the price column are formatted consistently with two decimal places.

Submission Instructions:

1. Submit the link to your cleaned dataset.
2. Attach the presentation with the screenshots and explanations.

Points to keep in mind:

1. You can use tools like Google Slides, PowerPoint or even tools like Figma, Canva, Pitch (or any other tool of your choice) to create the slide. You can either directly save it as a PDF, or export it as image and add it to your deck.
2. There's no recommended template or format for the slide. Feel free to use your own creativity to showcase your approach and thought process.

Finally, don't worry if you seem a little confused or feel challenged at the moment. Productive struggle is part of the learning process, and by the end of the week - you should have a lot more clarity on how to solve this.

