Audible Case Study

audible
an amazon company

Mentor: Ayushi Jain

Presented by: Sapna Bhambhwani

Problem:

Standardize the name column to ensure consistent title casing.
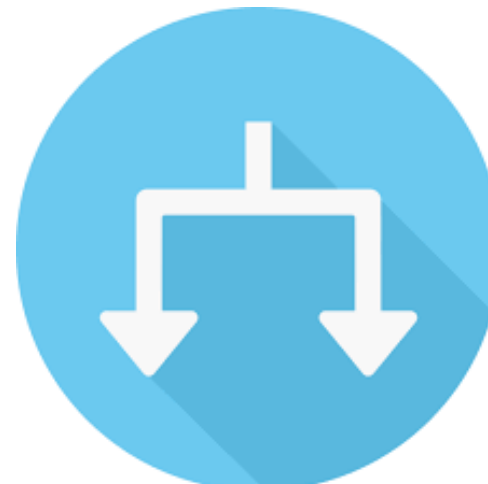
Solution:

The **Name** column has been reviewed and is **already in title case**, ensuring consistent formatting across all entries. No further transformation is required.

Problem:

Separate combined names in the author column if there are multiple authors.

Solution:

Applied the **"Split Column by Delimiter"** transformation in Power Query to divide combined author names into **separate columns**. This ensures each author is treated as an individual entity for accurate analysis, filtering, and reporting.

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Comma

Split at
- ○ Left-most delimiter
- ○ Right-most delimiter
- ◉ Each occurrence of the delimiter

▷ Advanced options

Quote Character

"

☐ Split using special characters

Insert special character

OK   Cancel

Replace Values   ·   Unpivot Columns
Fill ·   Move ·
Pivot Column   ·   Convert to List   ·   Split Column ·   Format
Any Column

ble.SplitColumn(#"Changed Type", "autho

author.1   ·   A♭C author.2
tenby:RickRiordan
tenby:PhilipPullman
tenby:GeronimoStilton
tenby:MaryPopeOsborne
tenby:ShannonMessenger   null

By Delimit
By Number
By Position
By Lowerc
By Upperc
By Digit to
By Non-Di

Problem:
Ensure all entries in the releasedate column follow a consistent date format (DD-MM-YYYY).

Solution:

The **ReleaseDate** column has been reviewed and all entries are **already standardized to the DD-MM-YYYY format**, ensuring consistency across the dataset. No further transformation is required.

Problem:
Convert the time column from text format to a duration format that Excel recognizes.
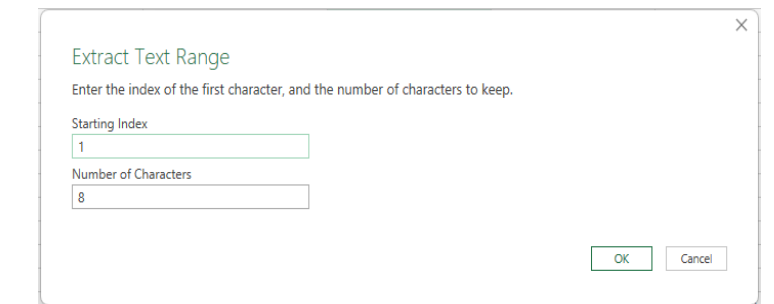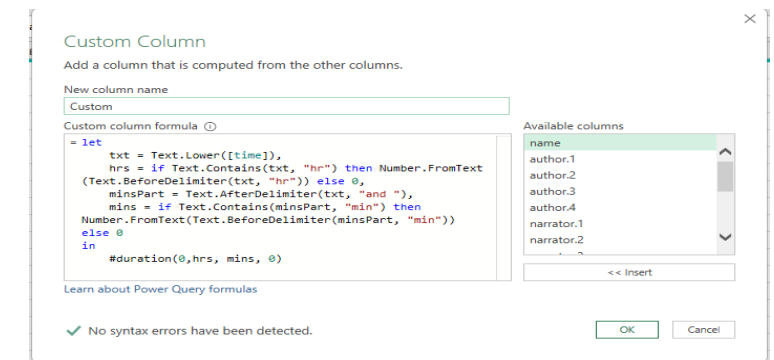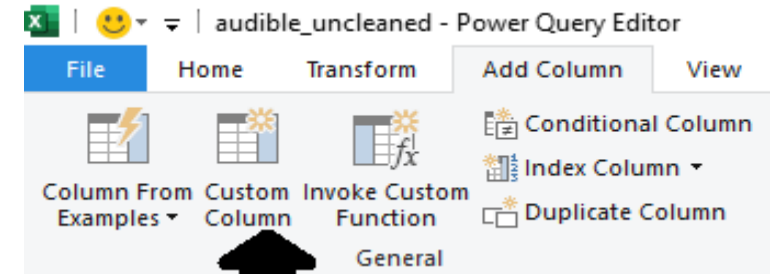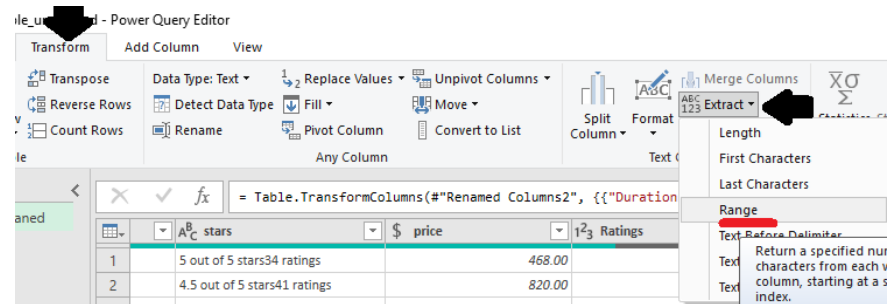
Solution:

Added a Custom Column in Power Query using an M query formula to convert the text into a recognizable duration format.

After conversion, used the "Extract Range" option to retrieve the time in hh:mm:ss format, allowing proper time calculations, filtering, and visualizations.
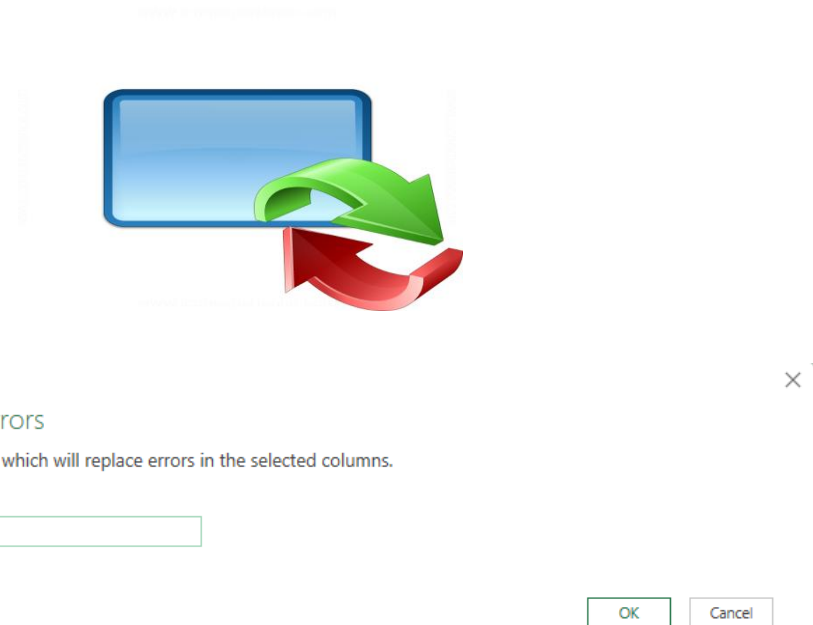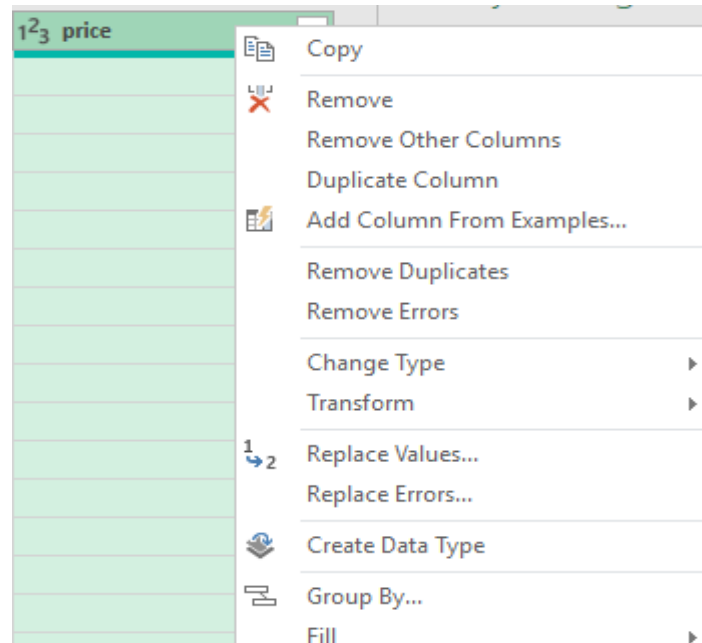
Problem:
Ensure the price column is in a numeric format and identify any non-numeric values.

Solution:

The column is already in numeric format; however, it was generating errors. To ensure smooth calculations, all error values were replaced with 0."
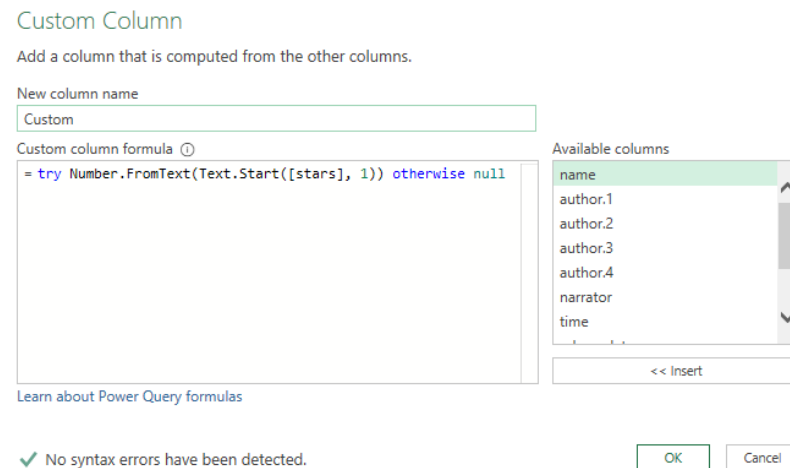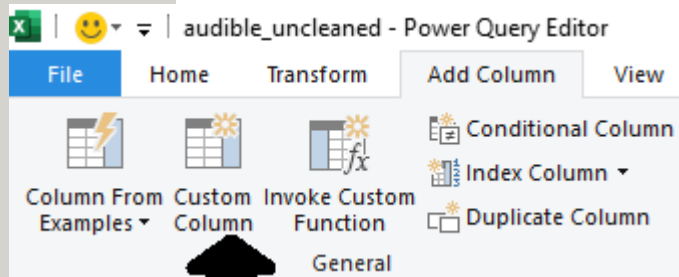
Problem:
Convert text ratings in the star's column to numeric values.

Solution:

Created a new custom column using the M query

Ratings = try Number.FromText(Text.Start([stars], 1)) otherwise null.

Then, changed the data type to Number and renamed the column to
Ratings for clarity and consistency.

Problem:

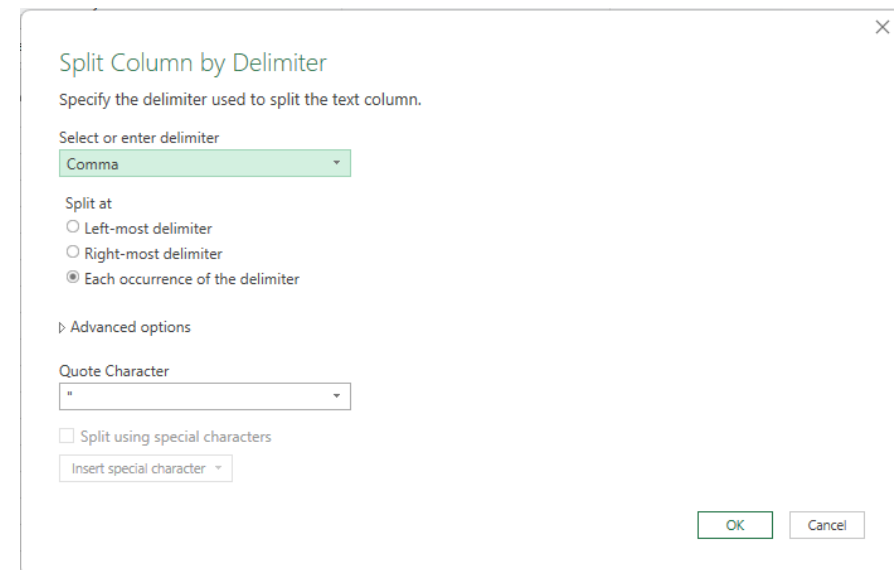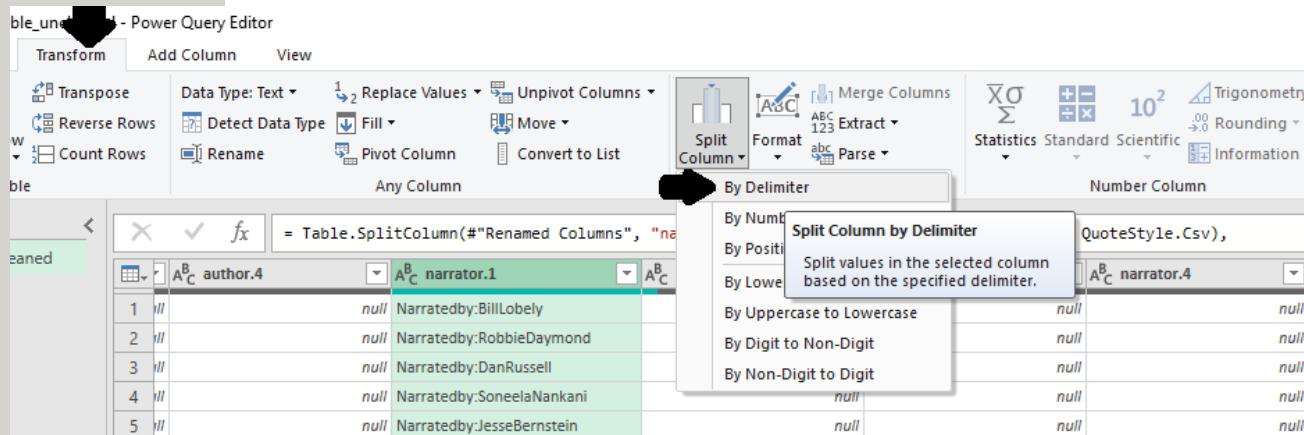Split the narratedby column into multiple columns if multiple narrators are listed.

Solution:

Used the **'Split Column by Delimiter'** option in Power Query to separate the combined values into individual fields for easier analysis and data handling.
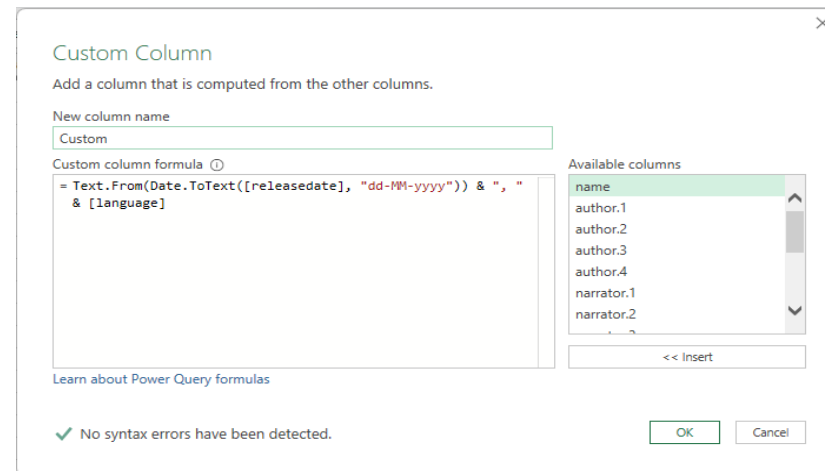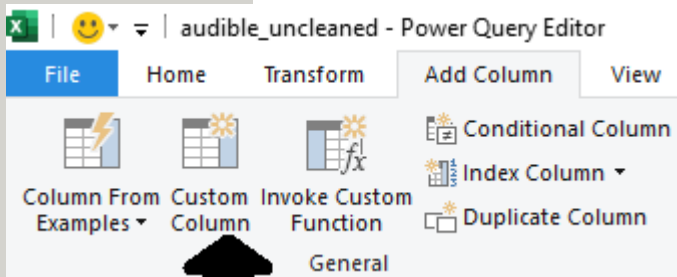
Problem:

Merge the releasedate and language columns into a single new column named releaseinfo with the format "DD-MM-YYYY, Language."

Solution:

Added a custom column using the following M query:

ReleaseInfo = Text.From(Date.ToText([releasedate], "dd-MM-yyyy")) & ", " & [language].

The resulting column was then renamed to ReleaseInfo to combine the release date and language into a single, readable format.
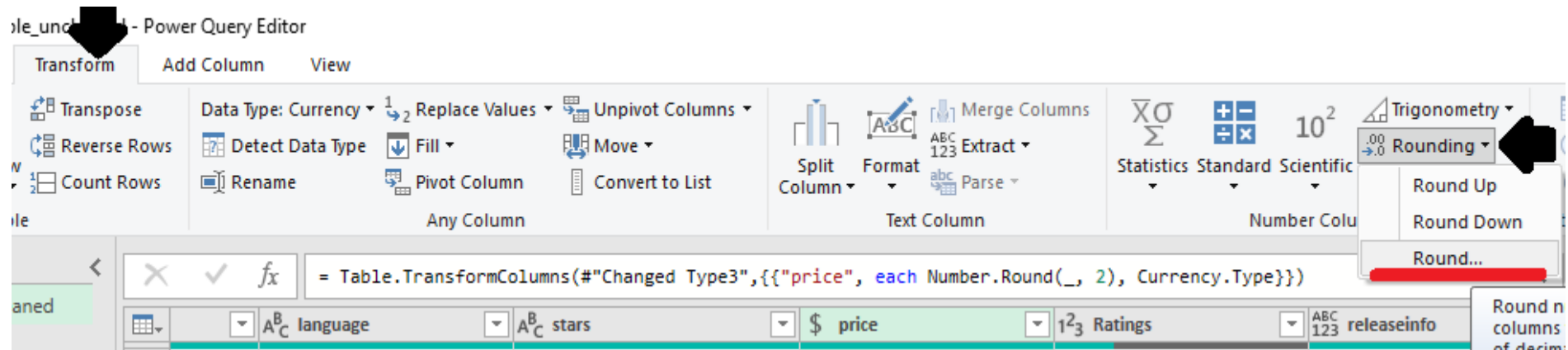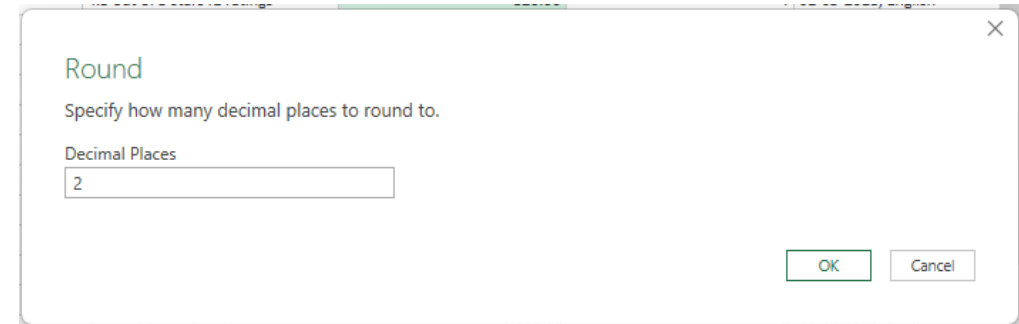

ABC 123	releaseinfo	▼
	04-08-2008, English	
	01-05-2018, English	
	06-11-2020, English	
	05-10-2021, English	
	13-01-2010, English	
	30-10-2018, English	
	25-11-2014, English	
	02-05-2017, English	
	02-05-2017, English	
	24-09-2019, English	
	14-01-2010, English	
	24-08-2011, English	
	27-09-2011, English	
	03-10-2017, English	
	24-06-2021, English	
	08-02-2008, English	
	26-12-2004, English	
	06-11-2018, English	
	02-05-2017, English	
	06-11-2018, English	
	06-10-2020, English	
	14-04-2009, English	

Custom Column

Add a column that is computed from the other columns.

New column name
Custom

Custom column formula ⓘ
= Text.From(Date.ToText([releasedate], "dd-MM-yyyy")) & ", " & [language]

Available columns
name
author.1
author.2
author.3
author.4
narrator.1
narrator.2

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK   Cancel

Problem:
Ensure all currency values in the price column are formatted consistently with two decimal places.

Solution :

Changed the data type to Currency, then used the 'Round' option from the Transform tab to round the values to 2 decimal places for consistency in financial data representation.

Conclusion

Using Power Query in Excel, the Audible dataset was cleaned and standardized by formatting text, splitting and merging columns, converting data types, and ensuring consistency across fields. The result is a structured, analysis-ready dataset suitable for reporting and insights.

Thanks