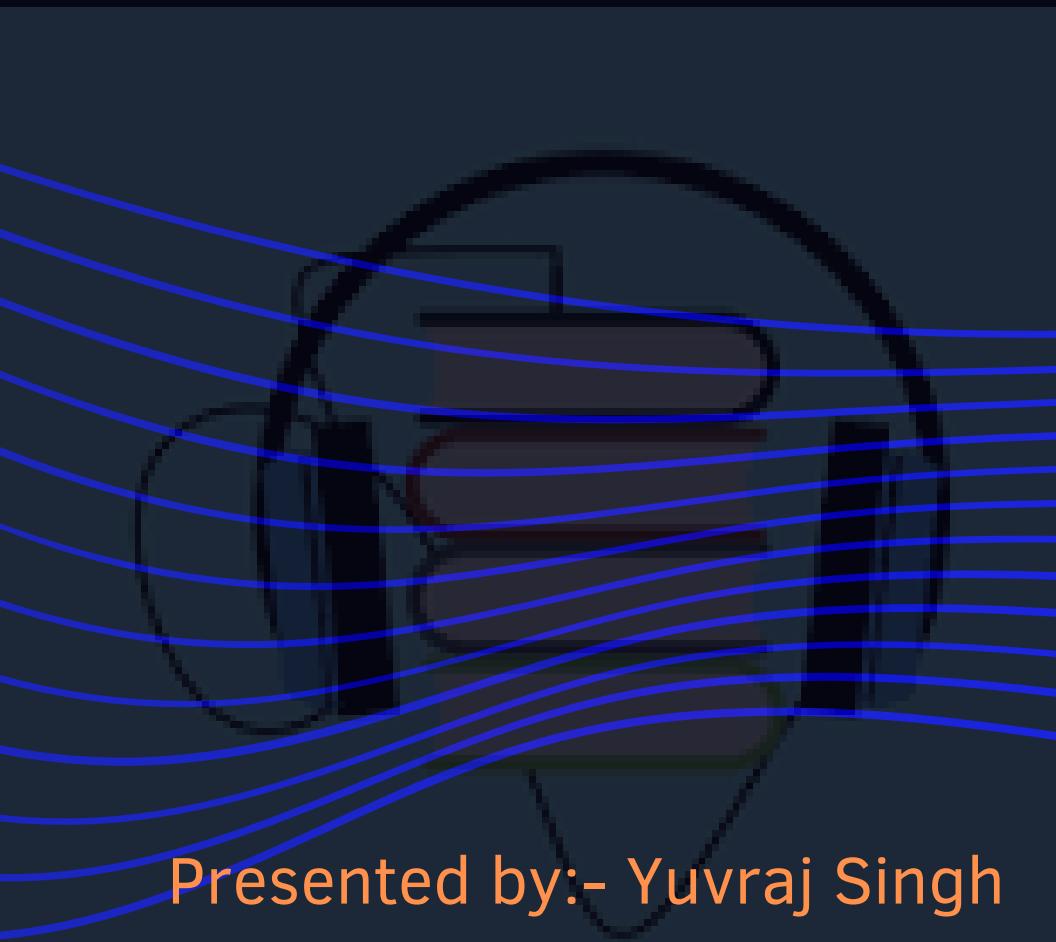


# Data Cleaning Project

With Power Editor



audible



Presented by:- Yuvraj Singh

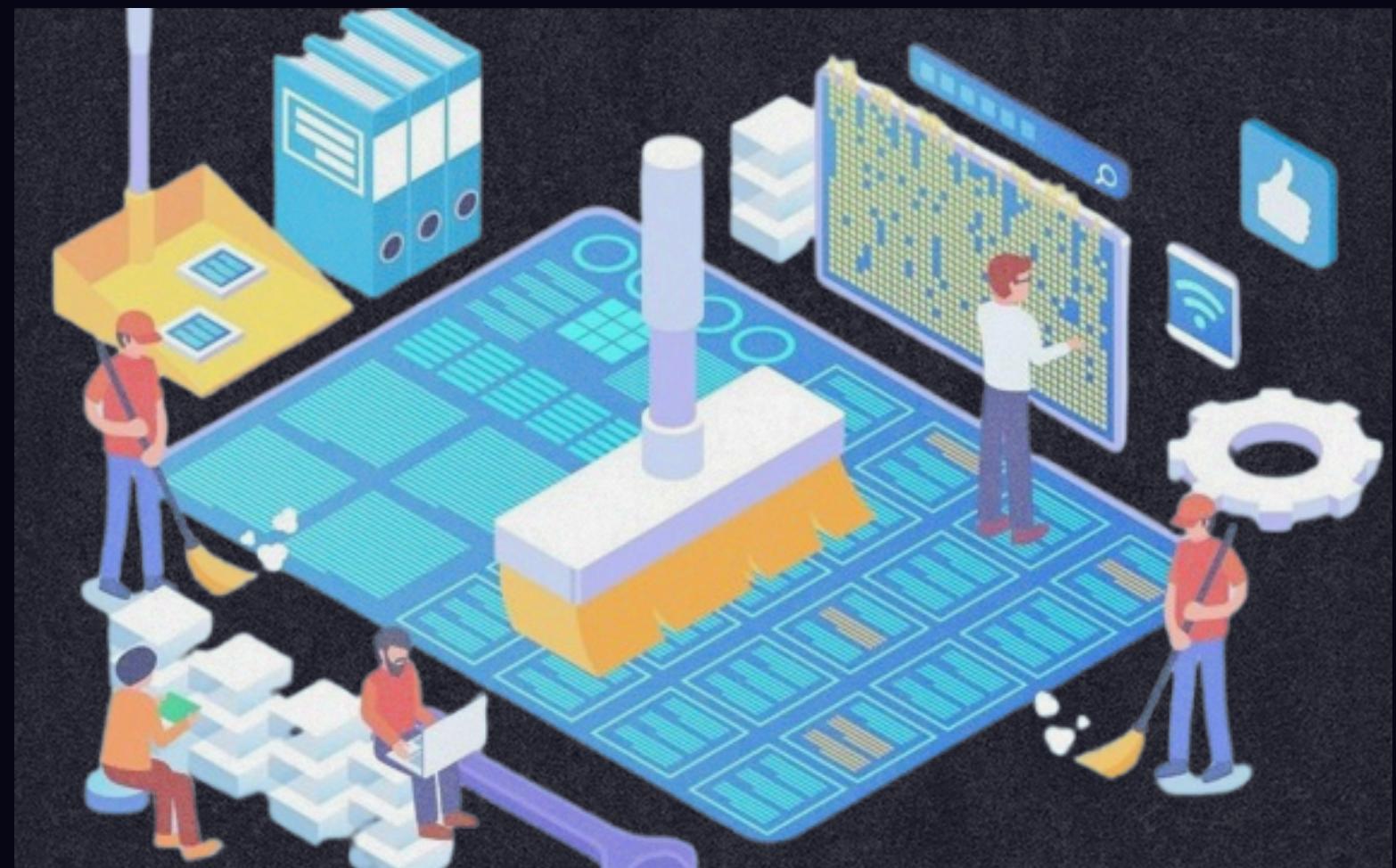
# About us

Audible is an Amazon company that produces and sells audiobooks, podcasts, and spoken-word content. It offers a vast library across genres, allowing users to listen anytime, anywhere through its app.



# Project Description

I will make use of the Power in this project. Query Editor in Excel to clean and standardize a dataset for Audible. The Among the tasks are reformatting. columns, ensuring consistency in the data, and preparing the dataset for further analysis.



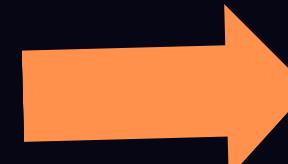
# Q 1. STANDARDIZE THE NAME COLUMN TO ENSURE CONSISTENT TITLE CASING.

Step-1



The screenshot shows a Microsoft Excel interface with a table containing book information. The 'Format' context menu is open, and the 'Capitalise Each Word' option is selected. A tooltip provides a description of the function.

name="Table1"]}	ARC
Writtenby:RickRiordan	Narratedby:SoneelaNankani
Writtenby:RickRiordan	Narratedby:JesseBernstein
Writtenby:SuzanneCollins	Narratedby:TatianaMaslany
Writtenby:WinterMorgan	Narratedby:LukeDaniels



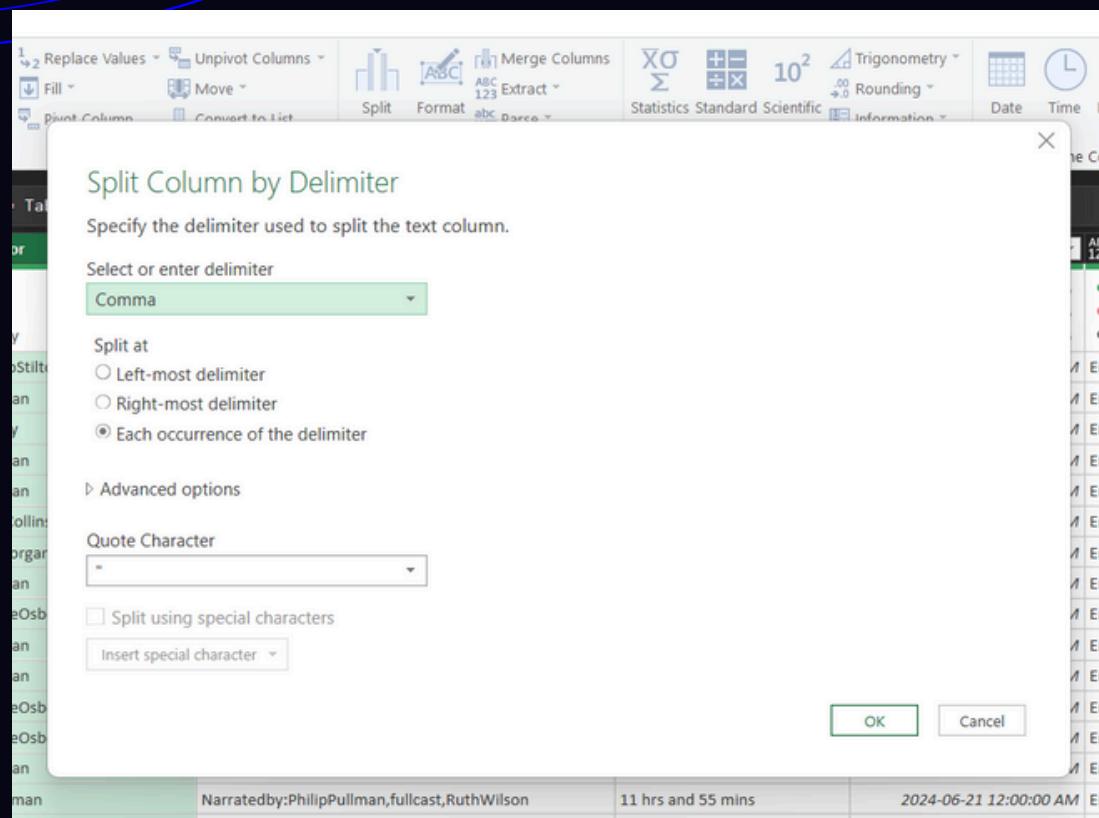
Step-2

	A B C name
1	Geronimo Stilton #11 & #12
2	The Burning Maze
3	The Deep End
4	Daughter Of The Deep
5	The Lightning Thief: Percy Jackson, Book 1
6	The Hunger Games: Special Edition
7	Quest For The Diamond Sword
8	The Dark Prophecy
9	Merlin Mission Collection
10	The Tyrantâ€™S Tomb

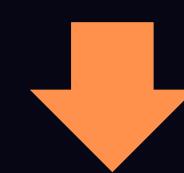
- 1.SELECT THE NAME COLUMN.
- 2.GO TO TRANSFORM → FORMAT → CAPITALIZE EACH WORD.
- 3.TITLES ARE NOW FORMATTED IN PROPER TITLE CASE.

## Q 2. SEPARATE COMBINED FIRST AND LAST NAMES IN THE AUTHOR COLUMN IF THEY ARE CURRENTLY COMBINED.

Step-1



Step-2

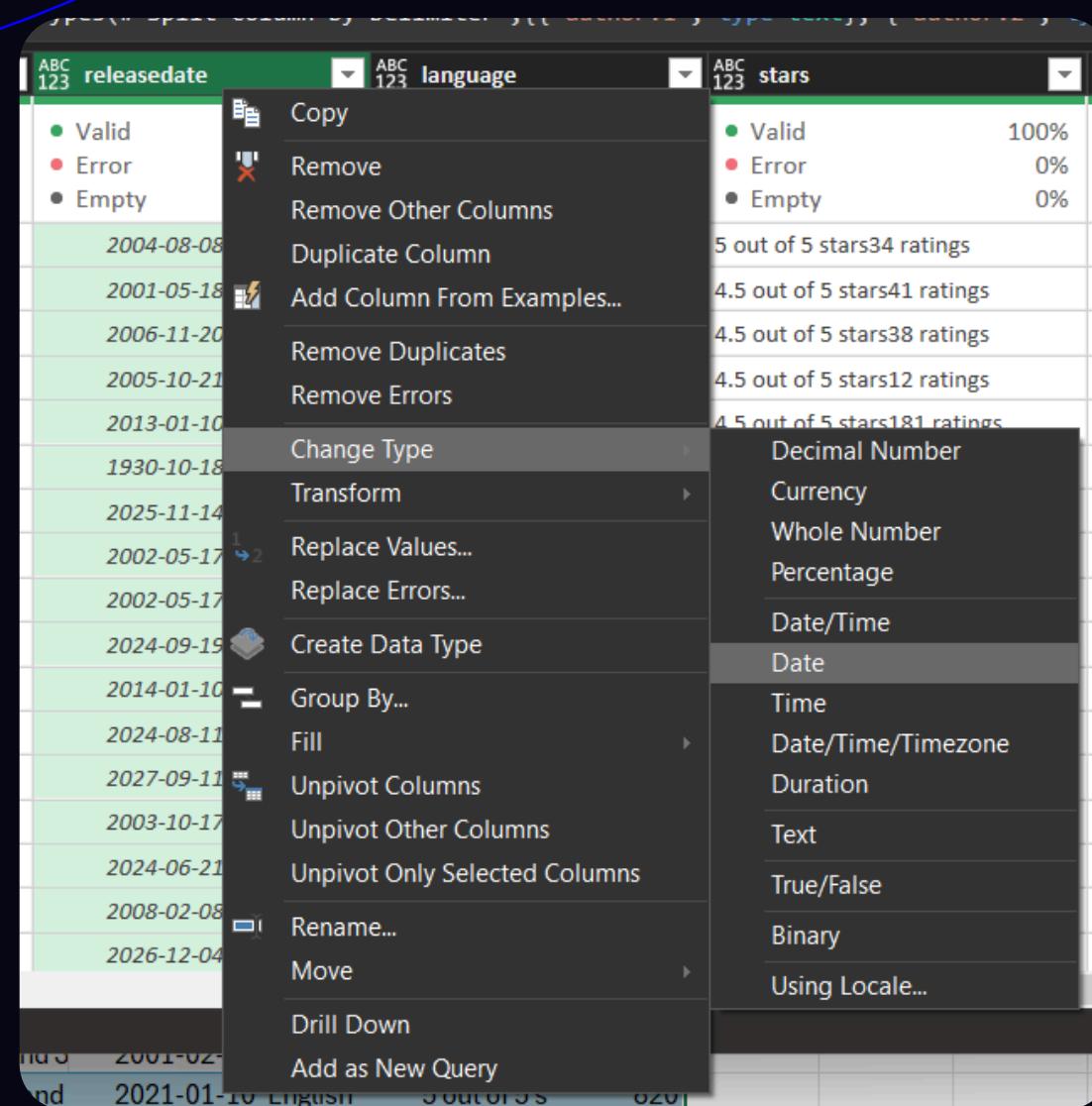


fx	= Table.TransformColumnTypes(#"Split Column by Delimiter",{{"author.1", type text}, {"aut	A <sup>B</sup> <sub>C</sub> author.1	A <sup>B</sup> <sub>C</sub> author.2	A <sup>B</sup> <sub>C</sub> author.3	A <sup>B</sup> <sub>C</sub> author.4
		● Valid ● Error ● Empty	100% 0% 0%	● Valid ● Error ● Empty	21% 0% 79%
		GeronimoStilton		null	null
		RickRiordan		null	null
		JeffKinney		null	null
		RickRiordan		null	null
		RickRiordan		null	null
		SuzanneCollins		null	null

1. Go to Transform → Split Column → By Delimiter.
2. In the popup, choose Comma ( , ) as the delimiter and click OK.
3. Author names are split into separate columns.

# Q 3. ENSURE ALL ENTRIES IN THE RELEASEDATE COLUMN FOLLOW A CONSISTENT DATE FORMAT (DD-MM-YYYY).

Step-1



Step-2

releasedate	ABC 123
2004-08-08	English
2001-05-18	English
2006-11-20	English
2005-10-21	English
2013-01-10	English
1930-10-18	English
2025-11-14	English
2002-05-17	English
2002-05-17	English
2024-09-19	English
2014-01-10	English
2024-08-11	English
2027-09-11	English
2003-10-17	English
2024-06-21	English
2008-02-08	English
2026-12-04	English

1. Select the releasedate column Go to Transform → Date → Date Only.

2. The column now shows only the date in a consistent format.

# Q 4. CONVERT THE TIME COLUMN FROM TEXT FORMAT TO A DURATION FORMAT THAT EXCEL RECOGNIZES.

1.Duplicated the time Column .

2.Added a Conditional Column: If time does not contain:  
“hr”,Output “0 hr and ”.

3.Merged the conditional column with the original time Column  
(no separator).

4.Added another Conditional Column: If the merged column does  
not contain “and”, Output “and 00 min”

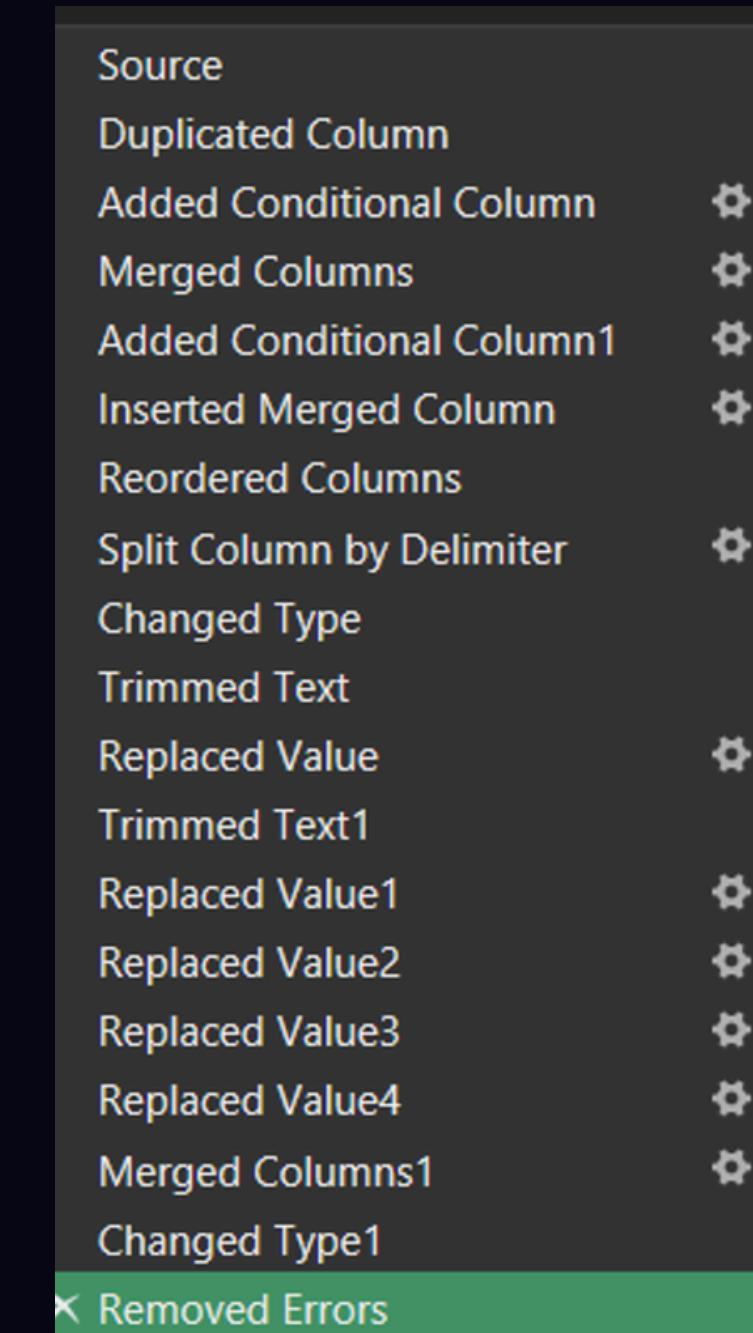
5.Merged this with the previous result to complete the  
formatting.

6.The final column now displays time values in a consistent  
format, such as 2 hr and 30 mins.

7.Once the time data was standardized, we applied Split Column  
by Delimiter to extract hours and minutes into separate columns.

8.After splitting into two columns, we merged them using a colon  
(:) to format the time as hr:min:sec.

## Step-1



## Step-2

Time in format		
%	Valid	100%
%	Error	0%
%	Empty	0%
	02:20:00	
	13:08:00	
	02:03:00	
	11:16:00	
	10:00:00	
	10:35:00	
	02:23:00	
	12:32:00	
	10:56:00	
	13:22:00	
	08:48:00	
	05:23:00	
	06:01:00	
	12:58:00	
	11:55:00	
	02:25:00	
	05:04:00	
	14:41:00	

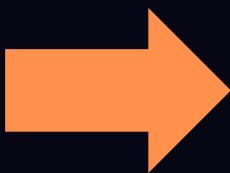
# Q 5. ENSURE THE PRICE COLUMN IS IN A NUMERIC FORMAT, AND IDENTIFY ANY NON-NUMERIC VALUES.

## Step-1



First Changed the Price column to Text to identify nonnumeric values.

The screenshot shows the Power BI ribbon with the 'Transform' tab selected. In the center, there is a dropdown menu labeled 'Data Type: Any'. Below the dropdown, a list of data types is visible, including 'Decimal Number', 'Currency', 'Whole Number', 'Percentage', 'Date/Time', 'Date', 'Time', 'Date/Time/Timezone', 'Duration', 'Text' (which is highlighted in blue), 'True/False', and 'Binary'. The number '1' is displayed above the 'Text' option, and the number '2' is displayed above the 'True/False' option.



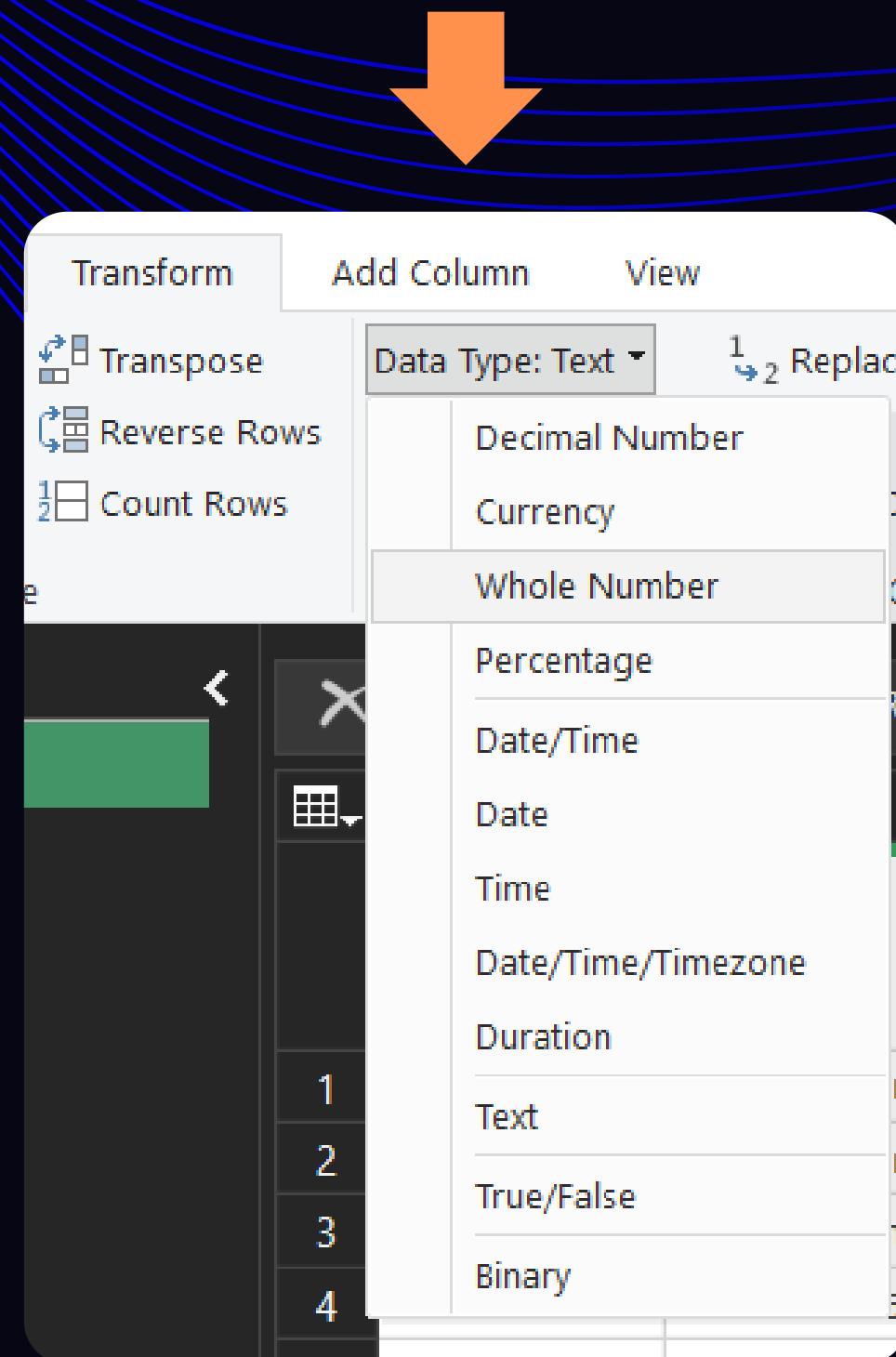
## Step-2

Replaced “Free” with “0”.

The screenshot shows a 'Replace Values' dialog box. At the top, it says 'Replace Values' and 'Replace one value with another in the selected columns.' Below this, there are two input fields: 'Value To Find' containing 'Free' and 'Replace With' containing '0'. There is also a 'Advanced options' link. At the bottom right are 'OK' and 'Cancel' buttons. In the background, a portion of a Power BI report is visible, showing a table with columns like 'and 48 mins', '2014-01-10', 'English', '4.5 out of 5 stars', '130 ratings', '820', 'and 23 mins', '2024-08-11', 'English', '5 out of 5 stars', '6 ratings', and '1206'.



## Step-3



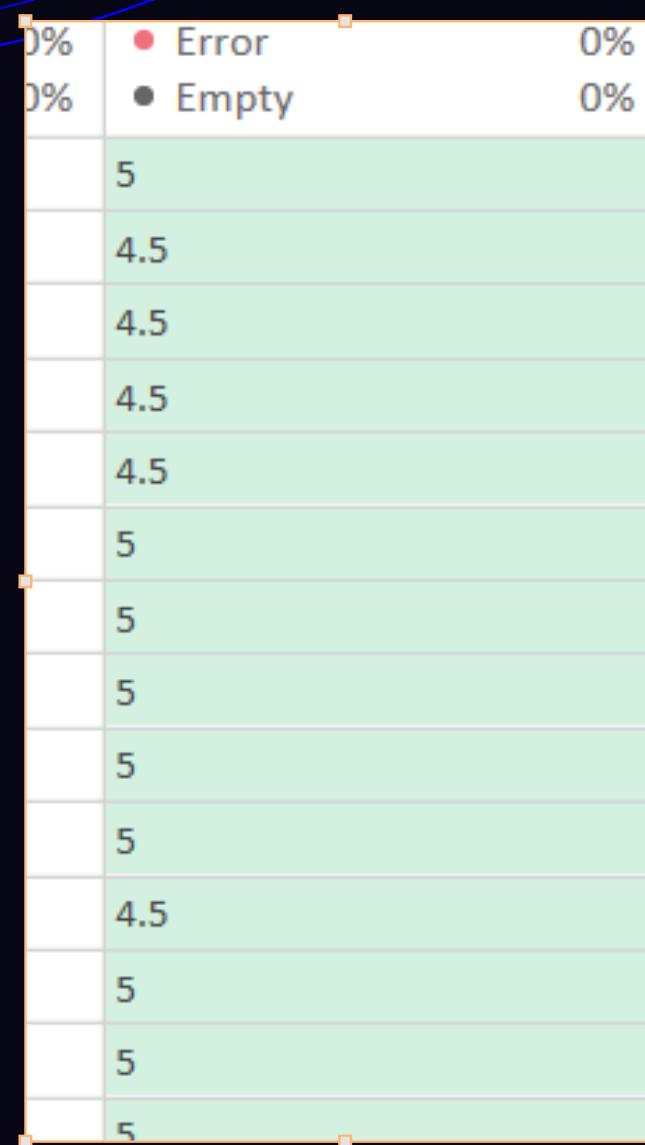
## Step-4

1	2	3	price
100%	● Valid		100%
0%	● Error		0%
0%	● Empty		0%
			468
			820
			410
			615
			820
			656
			233
			820
			1256
			820
			820
			1206

Converted the column to Whole Number for proper numeric formatting.

# Q 6. CONVERT TEXT RATINGS IN THE STARS COLUMN TO NUMERIC VALUES.

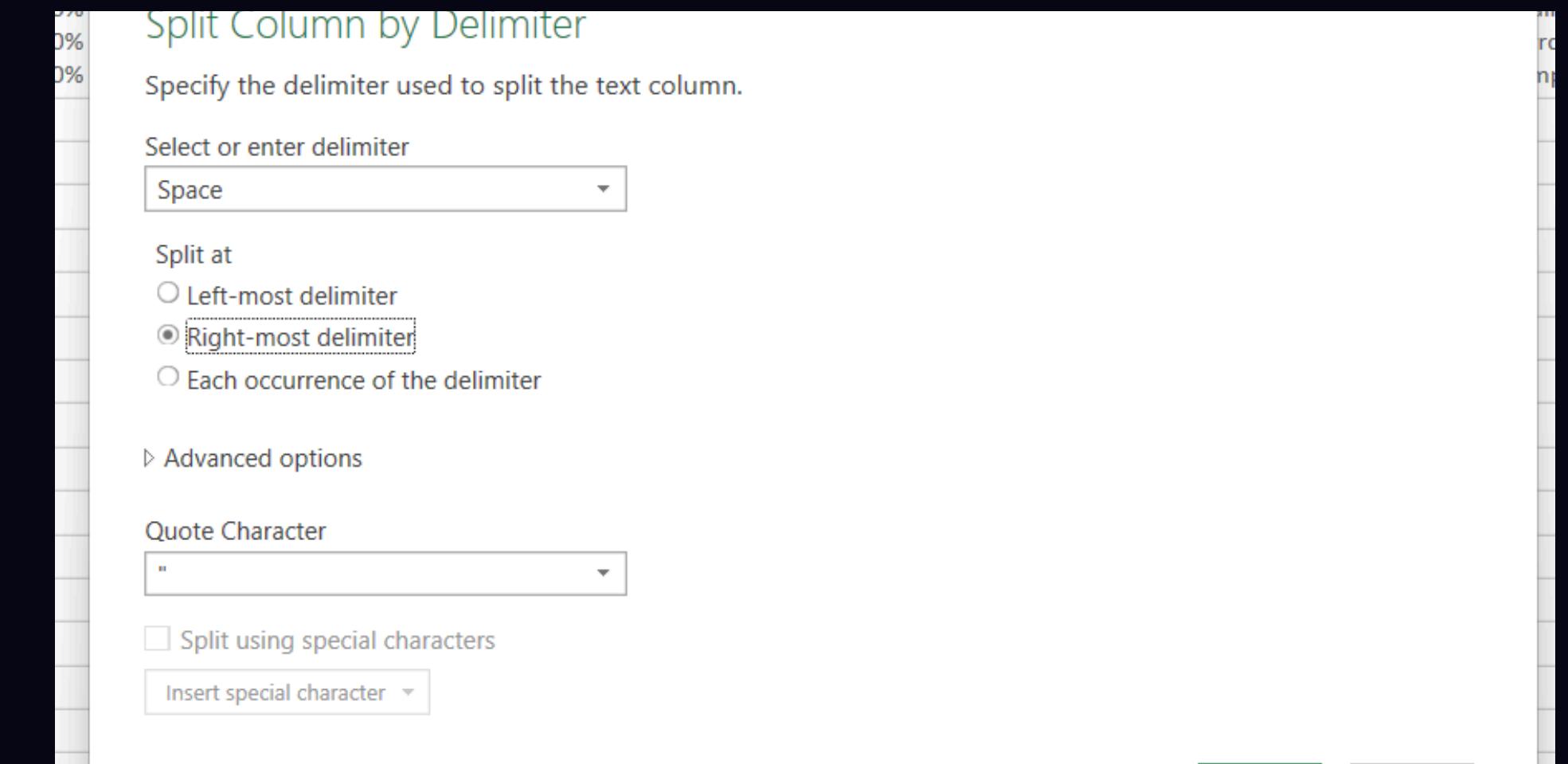
Step-1



1. Selected Split Column from the Transform tab.

2. Chose By Delimiter and entered “out” as the delimiter.

Step-2

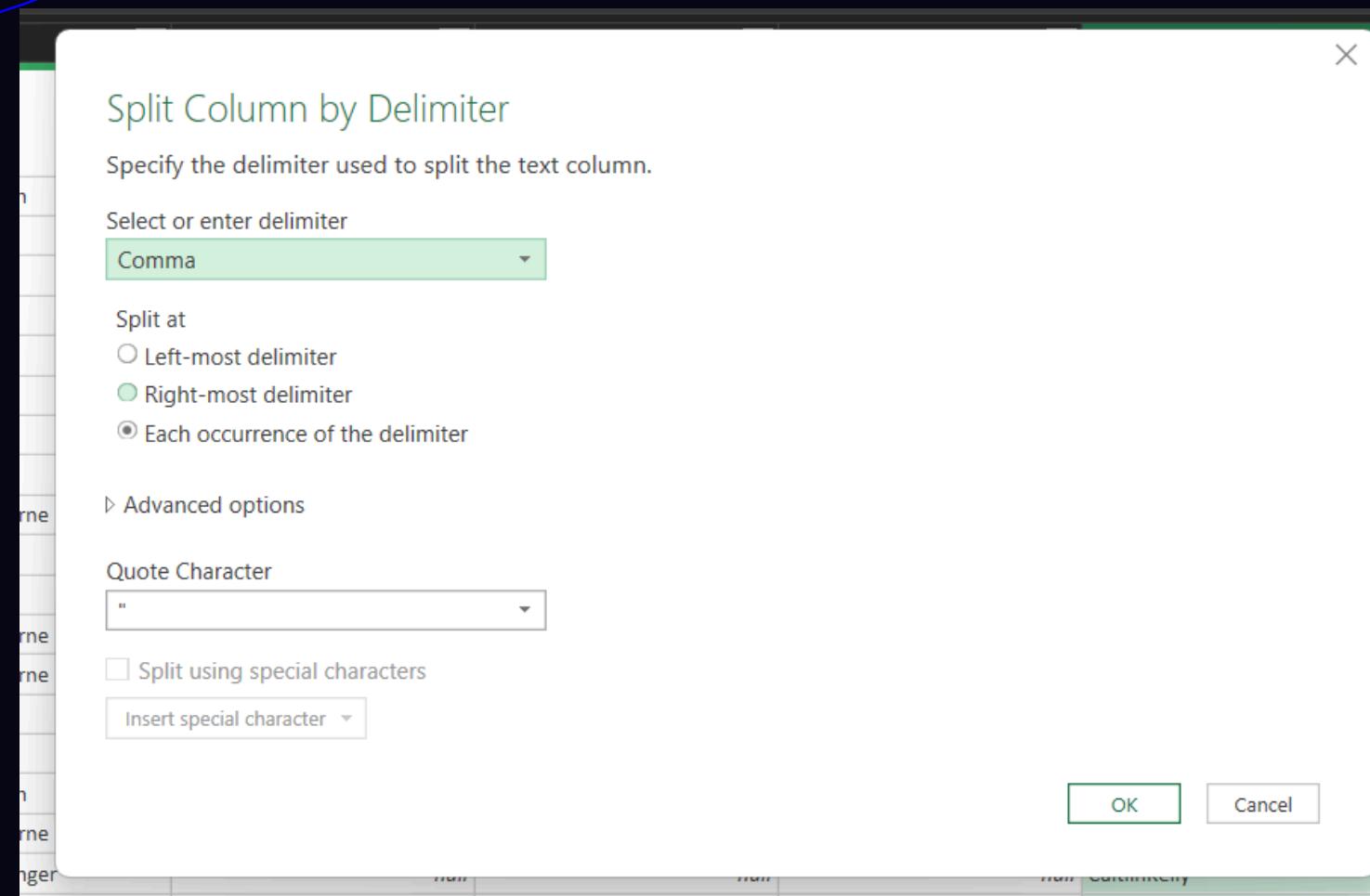


3. Selected Left-most delimiter under “Split at” and clicked OK.

4. Extracted the numeric star rating from the text.

# Q 7. SPLIT THE NARRATEDBY COLUMN INTO MULTIPLE COLUMNS IF MULTIPLE NARRATORS ARE LISTED.

Step-1



Step-2



A <sup>B</sup> C narrator.1	A <sup>B</sup> C narrator.2	A <sup>B</sup> C narrator.3	A <sup>B</sup> C narrator.4
BillLobley	null	null	null
RobbieDaymond	null	null	null
DanRussell	null	null	null
SoneelaNankani	null	null	null
JesseBernstein	null	null	null
TatianaMaslany	null	null	null
LukeDaniels	null	null	null
RobbieDaymond	null	null	null
MaryPopeOsborne	null	null	null
RobbieDaymond	null	null	null
JesseBernstein	null	null	null
MaryPopeOsborne	null	null	null
MaryPopeOsborne	null	null	null
MichaelCrouch	null	null	null
PhilipPullman	fullcast	RuthWilson	null
BillLobley	null	null	null

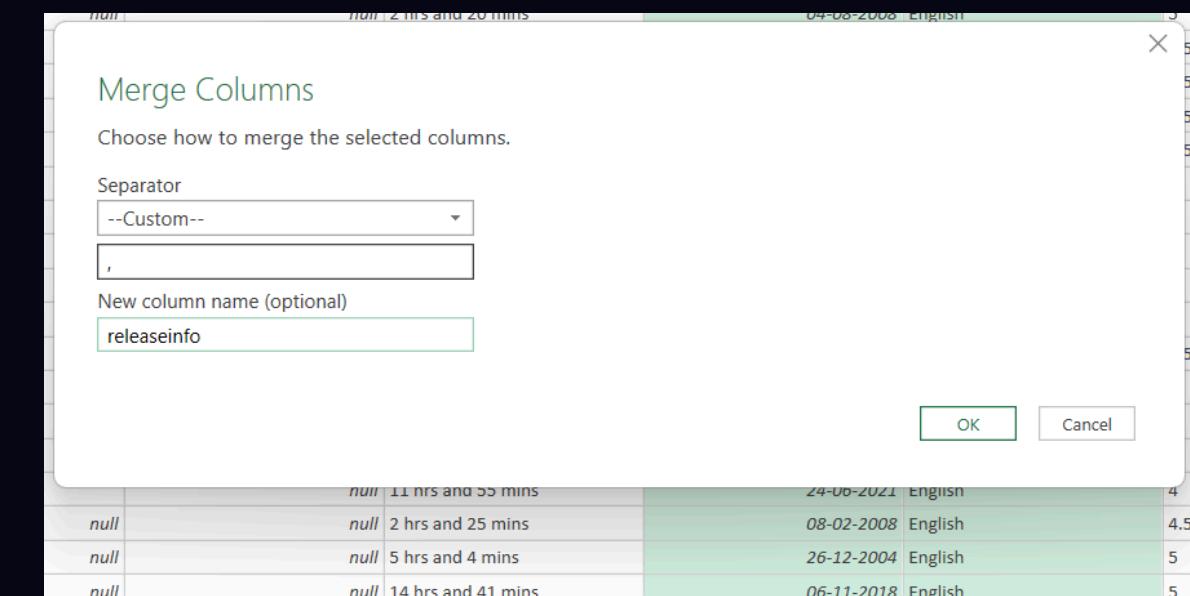
1. Selected the Went to narratedby column and removed “Narratedby:” from all entries.
2. Went to Transform → Split Column → By Delimiter.

3. Chose Comma ( , ) as the delimiter and split at each occurrence
4. Narrators are now separated into individual columns.narrte dby

## Q 8. MERGE THE RELEASE-DATE AND LANGUAGE COLUMNS INTO A SINGLE NEW COLUMN NAMED RELEASE-INFO WITH THE FORMAT "DD-MM-YYYY, LANGUAGE."

A screenshot of a data analysis interface showing a table with two columns selected: 'releasedate' and 'language'. A context menu is open over these columns, with the 'Merge Columns' option highlighted.

1. Selected both releasedate and language column



2. Merge Columns with comma ( , ) as the separator.

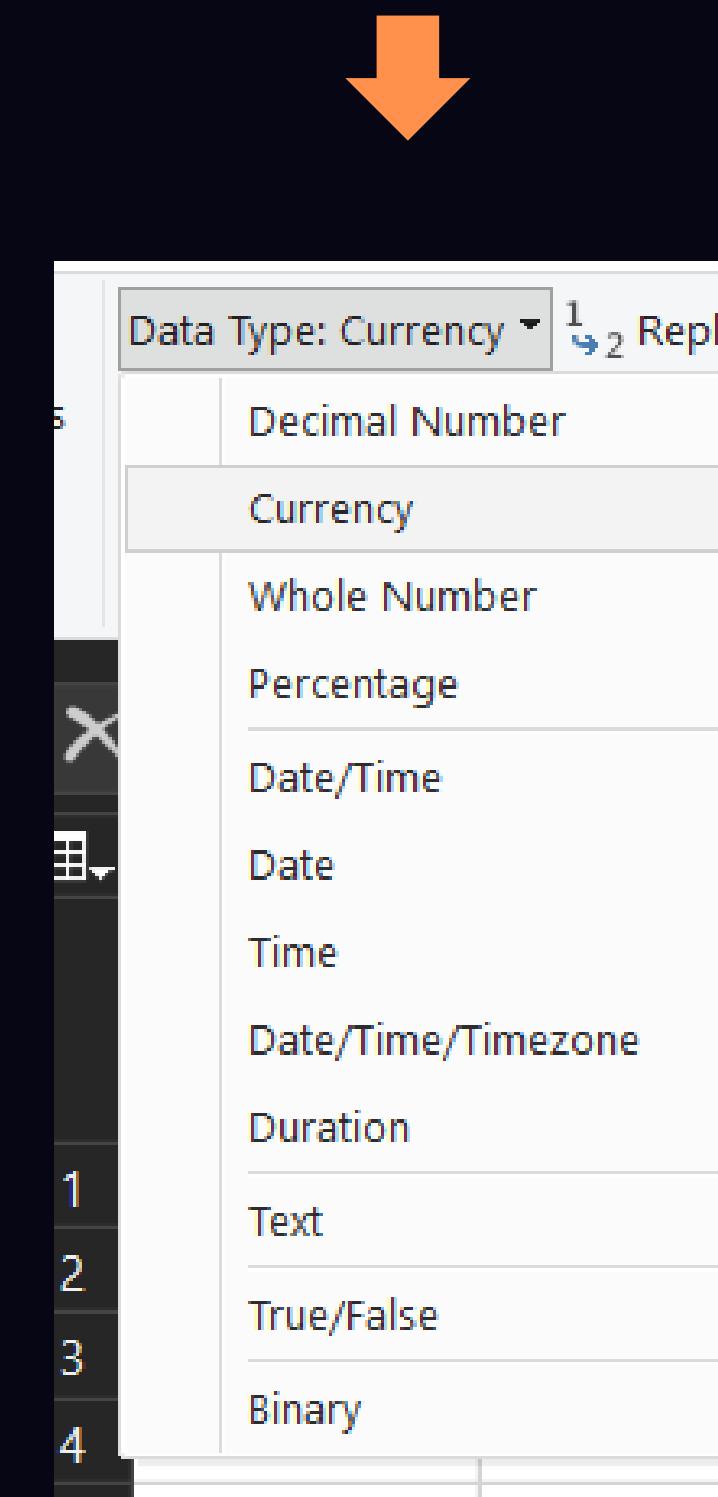
3. Named the new column releaseinfo and clicked OK

A screenshot of the final data table showing the merged 'releaseinfo' column. The first few rows are: '04-08-2008,English', '01-05-2018,English', '06-11-2020,English', etc. The top of the table includes a status bar showing 100% Valid, 0% Error, and 0% Empty.

## Q 9. ENSURE ALL CURRENCY VALUES IN THE PRICE COLUMN ARE FORMATTED CONSISTENTLY WITH TWO DECIMAL PLACES.

1. Selected the Price column.
2. Went to Transform → Data Type → selected Currency.

Step-1



Step-2

	\$	price
6	● Valid	100%
6	● Error	0%
6	● Empty	0%
		468.00
		820.00
		410.00
		615.00
		820.00
		656.00
		233.00
		820.00
		1,256.00
		820.00
		820.00
		1,206.00
		1,206.00
		820.00
		1,093.00
		467.00
		1,206.00
		836.00



**audible**  
an **amazon** company

**“Thanks a bunch”**