audible

Audible Case Study

About Audible

Audible Inc., a subsidiary of Amazon, is the world's largest producer and distributor of audiobooks, spoken-word entertainment, and audio storytelling. Founded in 1995 and acquired by Amazon in 2008, Audible has revolutionized how people consume content by transforming the traditional reading experience into immersive audio.

Why Data Cleaning Matters?

In any data-driven business, raw datasets often contain inconsistencies, formatting issues, and clutter. These problems hinder effective analysis and can lead to inaccurate insights.

For Audible, clean data is crucial to:

- Understand customer preferences
- Analyze content performance
- Track market trends
- Improve recommendation systems

Our Objective 🎯

The primary objective of this case study is to transform a raw, inconsistent Audible dataset into a clean, structured, and analysis-ready format using Power Query in Microsoft Excel.

Original Audible dataset contains inconsistencies which make analysis difficult and unreliable. Using Power Query, the case study focuses on:

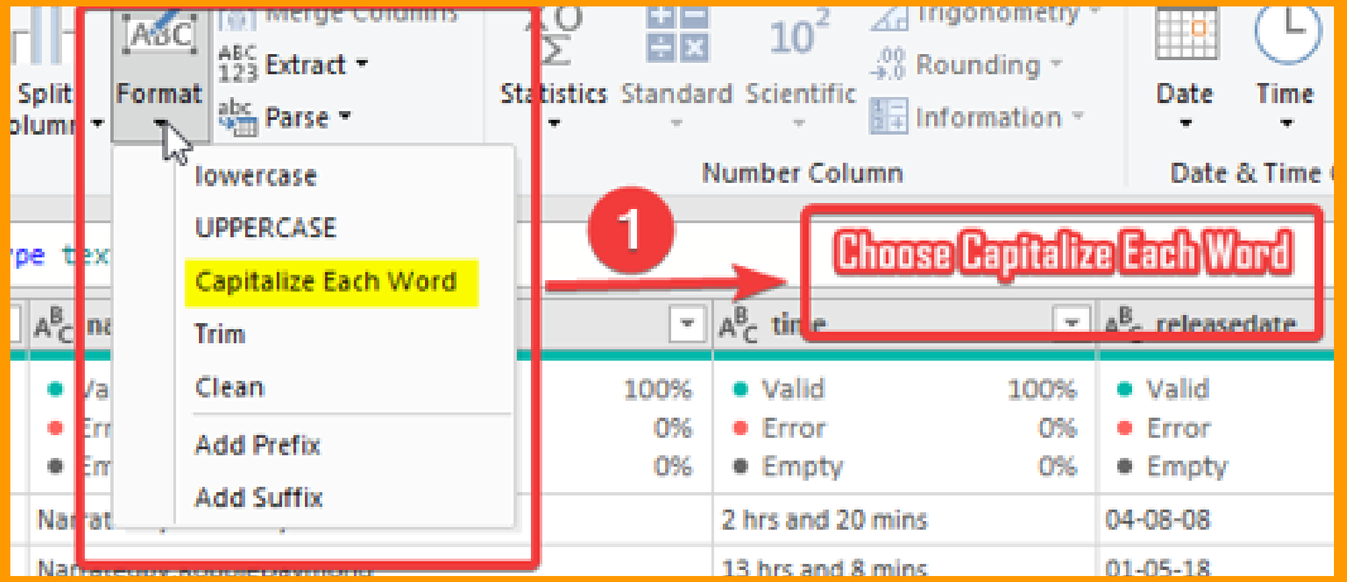
- Automating repetitive cleaning tasks
- Standardizing formats across key fields
- Converting unstructured text into Excel-recognized formats

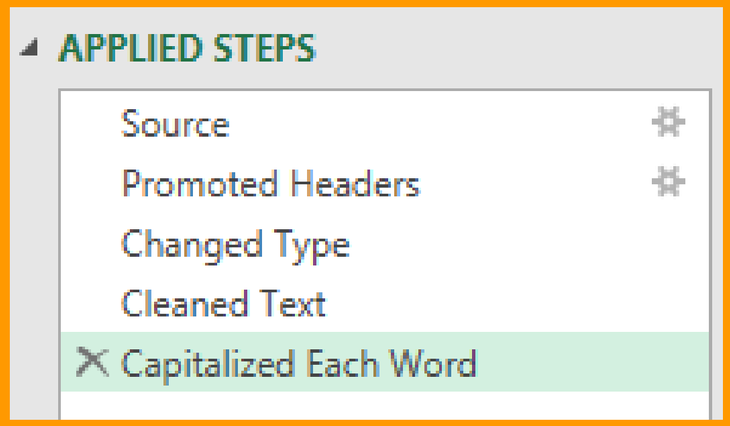
The goal is a clean, structured dataset ready for analysis and business decision-making

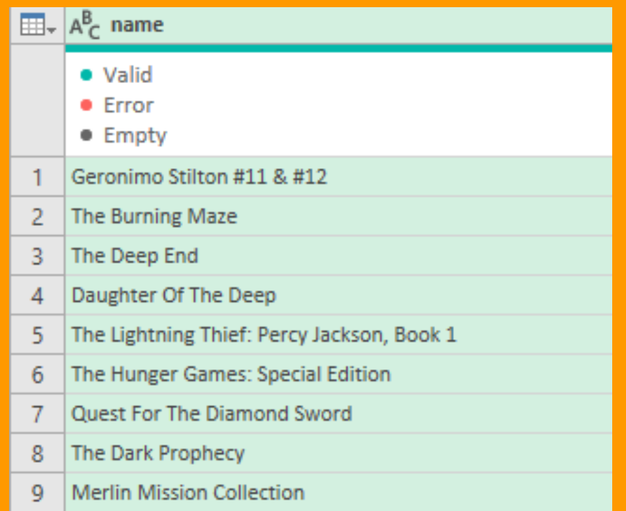
Task 01

Standardize the name column to ensure consistent title casing.

- Open the dataset in Power Query Editor.
- Navigate to the Name column.
- Go to Transform > Format > Capitalize Each Word.







📌 Outcome: Removes inconsistencies due to improper casing and prepares data for reliable sorting, filtering, and grouping.

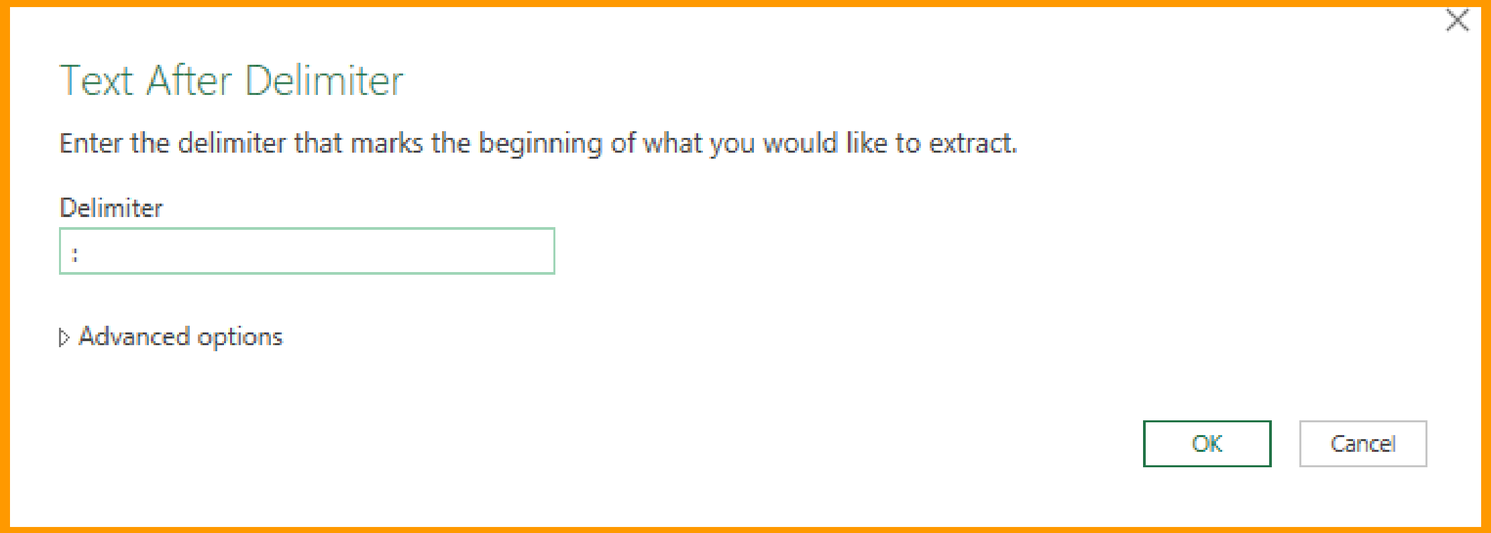
Task 02

Separate combined first and last names in the author column if they are currently combined.

author
Valid 100%
Error 0%
Empty 0%
Writtenby:GeronimoStilton
Writtenby:RickRiordan
Writtenby:JeffKinney
Writtenby:RickRiordan
Writtenby:RickRiordan
Writtenby:SuzanneCollins
Writtenby:WinterMorgan
Writtenby:RickRiordan
Writtenby:MaryPopeOsborne
Writtenby:RickRiordan
Writtenby:RickRiordan
Writtenby:MaryPopeOsborne

Data Before Cleaning

- Step 01:- To Clean This Column: Go to Home > Format > Extract> Text After Delimiter and Enter ":" as Delimiter

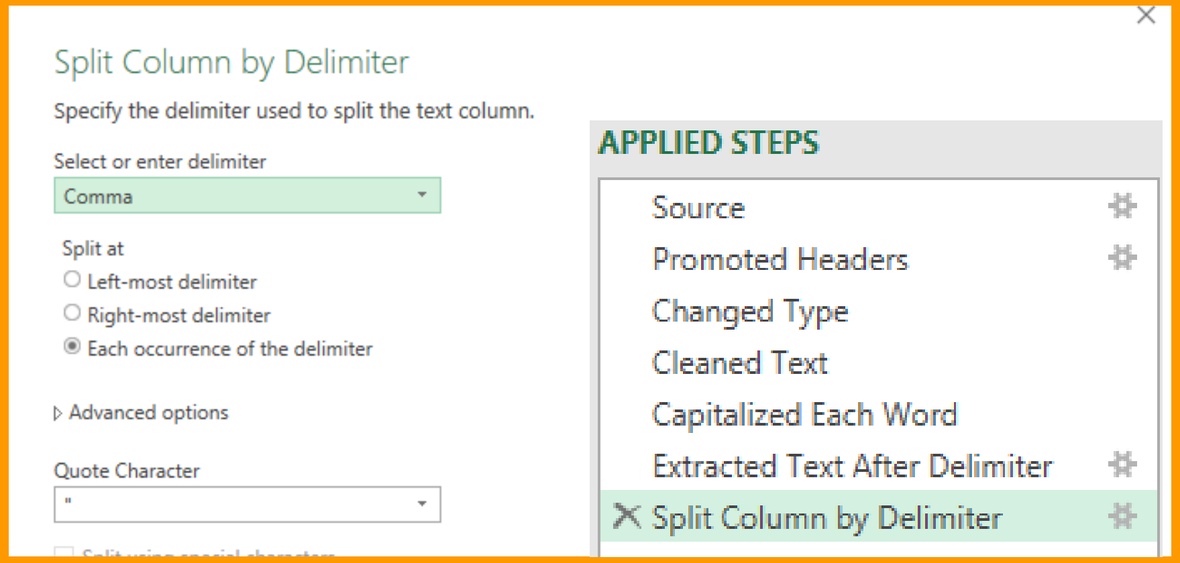


- Step 02:- Next head over to Split Column to Split Combined First and Last Names of Author Column based on Comma

Task 02

Separate combined first and last names in the author column if they are currently combined.

- Make sure to Split at each occurence of the delimiter to separate all authors and rename columns as Author 1 , Author 2 , Author 3.



📌 Outcome: Enables analysis on individual author contributions and popularity metrics.

author.1	author.2	author.3
Valid 100%	Valid 100%	Valid 0%
Error 0%	Error 0%	Error 0%
Empty 0%	Empty 0%	Empty 100%
StephieMorton	NicoleWong-illustrator	null
Mattel	NicolásOluchaSánchez	null
Sanrio	NeusCasanovaVico	null
TomFletcher	ShaneDevries-illustrator	null
ShannonMessenger	DorisAttwood-Übersetzer	null
Disney	ÅsaMalm	null
HumzaArshad	HenryWhite	null
GregJames	ChrisSmith	null
Mattel	NicolásOluchaSánchez	null
MalorieBlackman	DapoAdeola	null
ChristianMcKayHeidicker	JunyiWu-illustrator	null
JeffreyArcher	MariaNilsson	null
Disney	NilsWadström	null
AyanMalpani	AshishMalpani	null
RudyardKipling	LisaChurch-adaptation	null
Marvel	BrianIskov	null
AndyGriffiths	TerryDenton	null
AndrewPeterson-editor	JonathanRogers	N.D.Wilson
MacKenzieCadenhead	SeanRyan	AnkeAlbrecht-Übersetzer

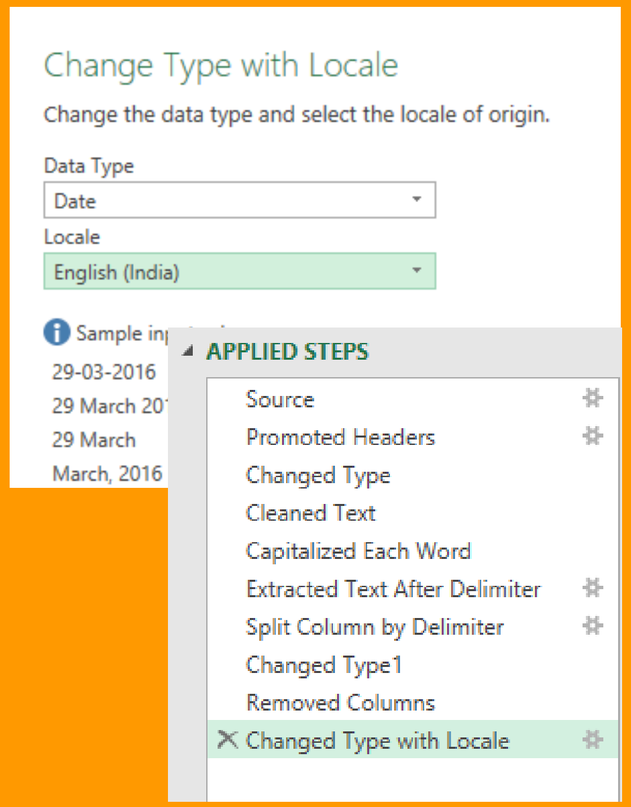
Data After Cleaning

Task 03

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

- Select the Releasedate Column and Change Data Type using Locale to Change it into Date Format

- Ensure to Change Date Format into India as the Values are in DD-MM-YYYY Format based on the Analysis



releasedate		language
0% Valid	100%	Valid
0% Error	0%	Error
0% Empty	0%	Empty
	04-Aug-08	English
	01-May-18	English
	06-Nov-20	English
	05-Oct-21	English
	13-Jan-10	English
	30-Oct-18	English
	25-Nov-14	English
	02-May-17	English
	02-May-17	English
	24-Sep-19	English
	14-Jan-10	English
	24-Aug-11	English
	27-Sep-11	English

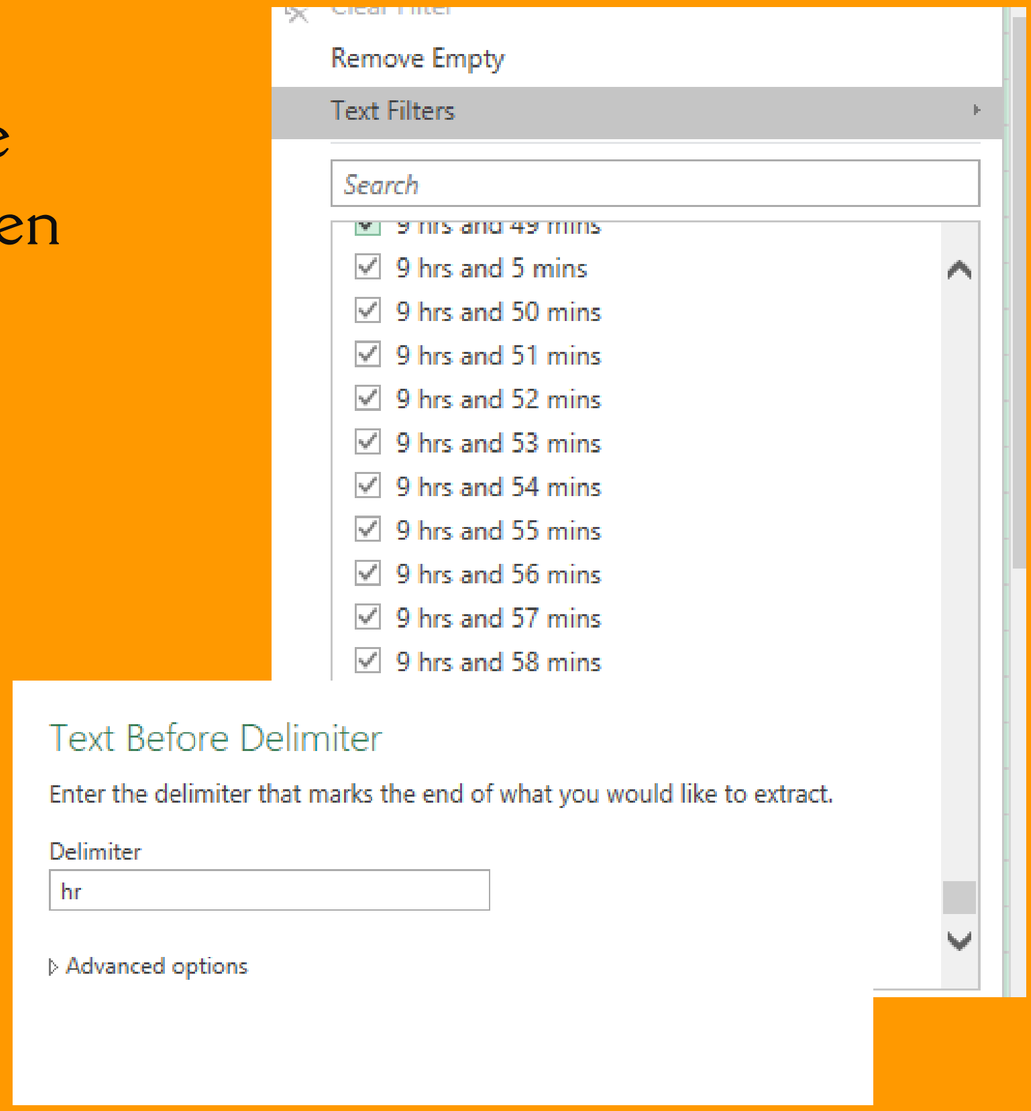
📌 Outcome: Provides consistency in timeline visuals, sorting, and date-based filtering.

Task 04

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

- To Convert Time Column into Duration Format We need to Convert Values into Numeric Format and then Convert it into Duration for further analysis.

- First We'll Replace Values with Text "Less than 1 minute" to "null" . Then We'll Separate Hours and Minute and Convert them into Numeric Values. To Extract Hours We'll Use Text Before Delimiter and Choose Delimiter as "Hr".



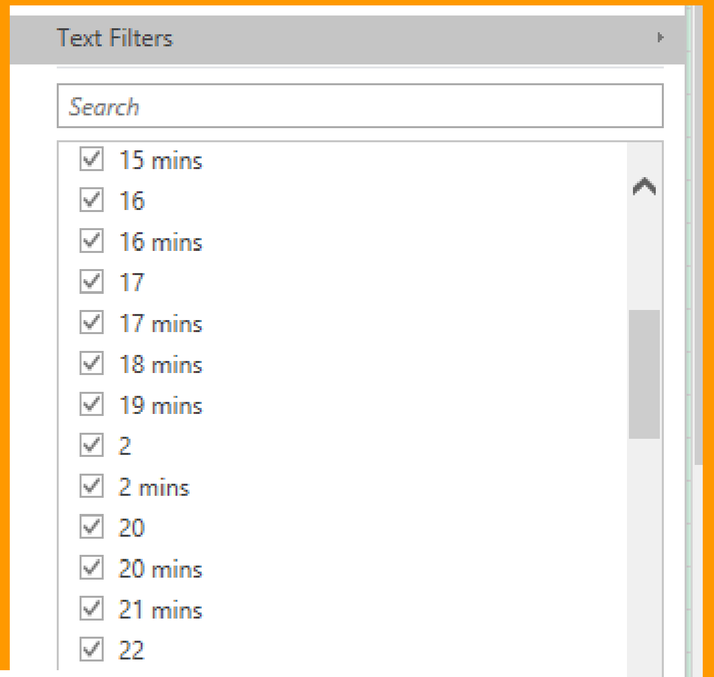
Task 04

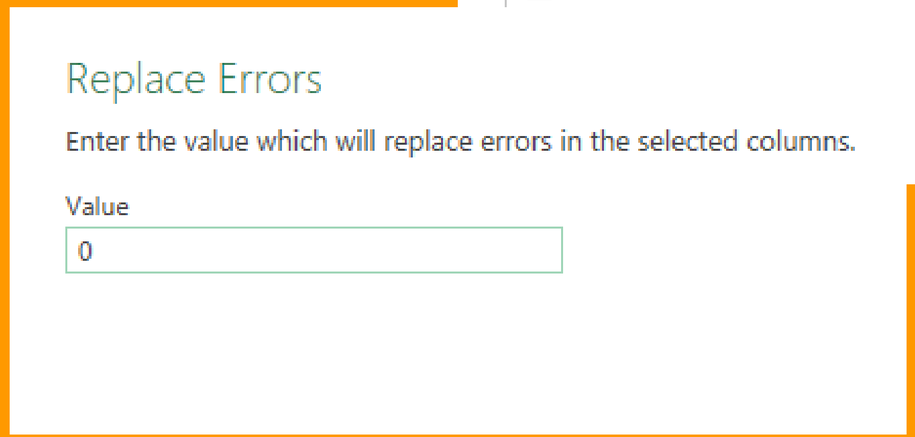
Convert the time column from text format to a duration format that Excel recognizes.

- Note that Hour Column is Still Not Cleaned as We Have Values Like “43 Minutes” in it. To Fix This We Convert the Column into Whole Numbers and the Remaining Errors are our values containing text “mins”.

- To fix this we’ll replace errors with “0” to showcase values contain no Hours in it.

- Now we head to clean Minutes Column.



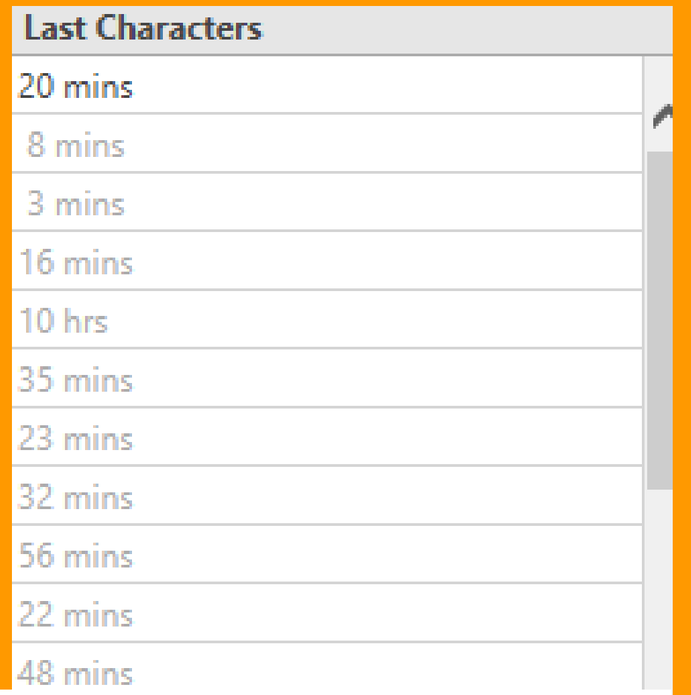


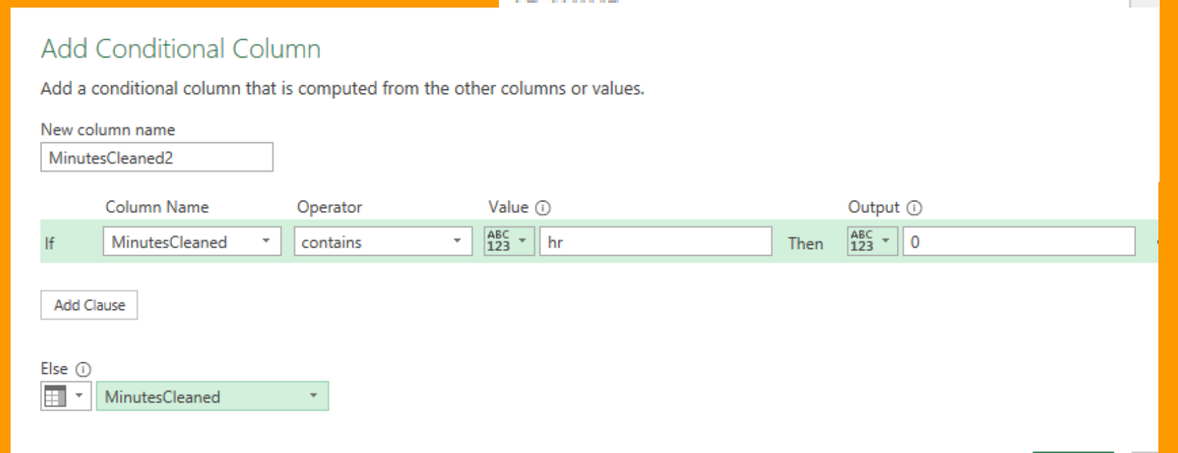
Task 04

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

- For Minutes Column We'll Use Column by Examples to Extract Minutes. You'll Notice We Also Have Some "Hrs" value in column which we'll clean using condition column.

- If column contain "Hrs" Replace with Null Else Display "Minutes". Once Done We'll Replace Mins with Blank Values and Trim it and Convert it to Numerical Values.



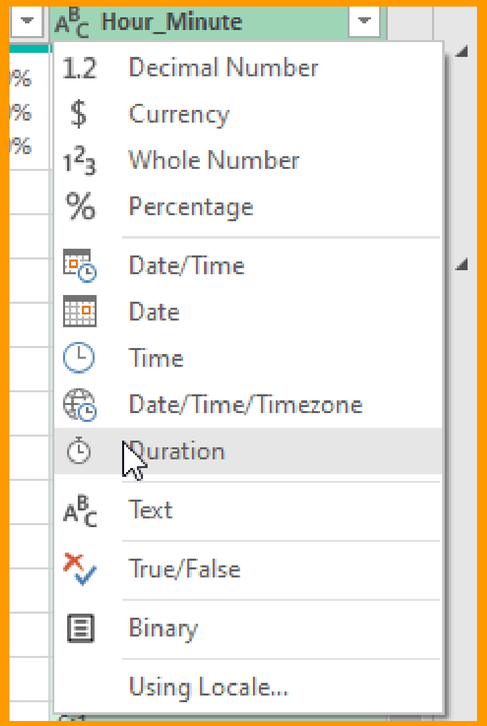


📌 Outcome: Makes time-based analysis (like average listening duration) possible.

Task 04

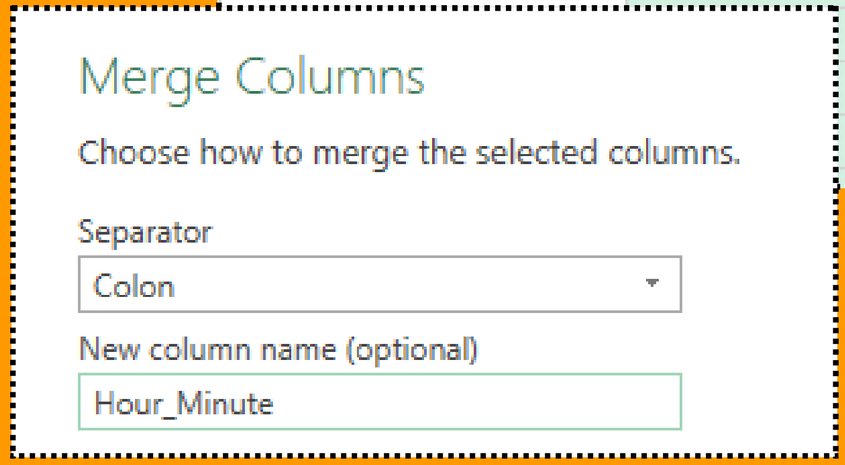
Convert the time column from text format to a duration format that Excel recognizes.

- Once we are done with Cleaning, Merge Hour and Minute Columns and Convert it into Duration Format. Make sure to choose Colon as a Separator.



time	Hour_Minute
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
2 hrs and 20 mins	0.02:20:00
13 hrs and 8 mins	0.13:08:00
2 hrs and 3 mins	0.02:03:00
11 hrs and 16 mins	0.11:16:00
10 hrs	0.10:00:00
10 hrs and 35 mins	0.10:35:00
2 hrs and 23 mins	0.02:23:00
12 hrs and 32 mins	0.12:32:00
10 hrs and 56 mins	0.10:56:00
13 hrs and 22 mins	0.13:22:00

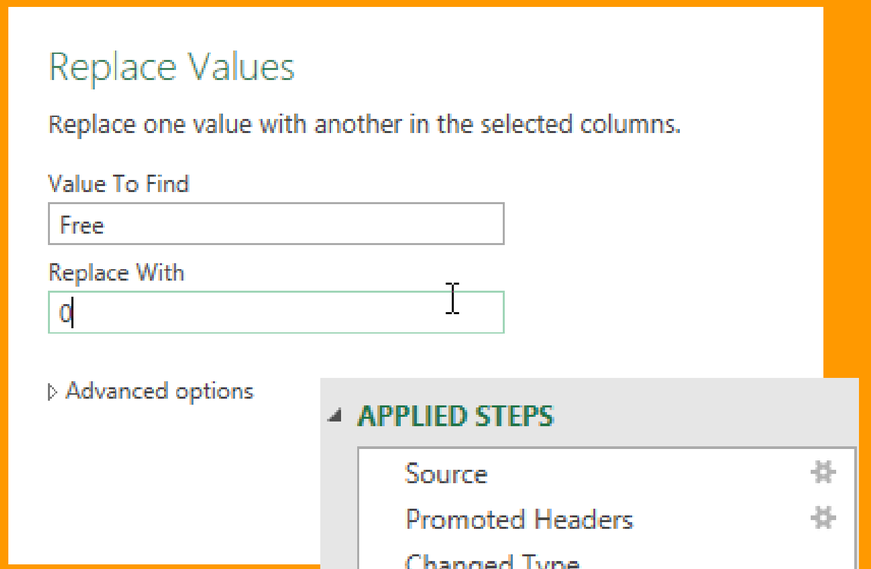
Hour Cleaned	MinutesCleaned
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
2	20
13	8
2	3
11	16
	0
	35
	23



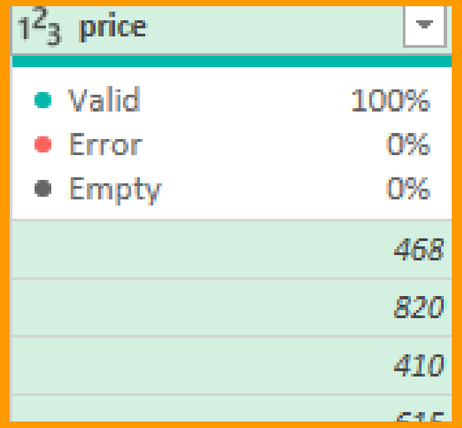
Task 05

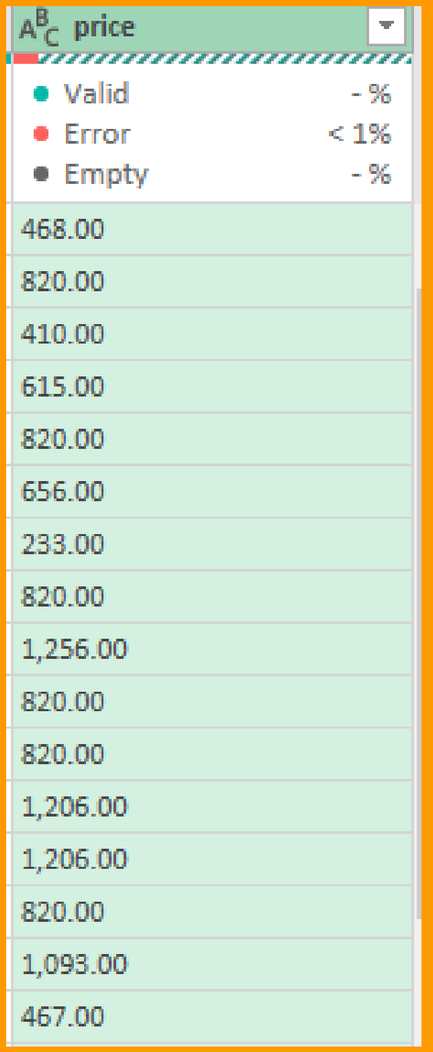
Ensure the price column is in a numeric format, and identify any non-numeric values.

- Upon Analysing the Column Contains Values Such as “Free” So We Need to Replace Them by “0” and then Convert it into Numeric Format



Data After Cleaning



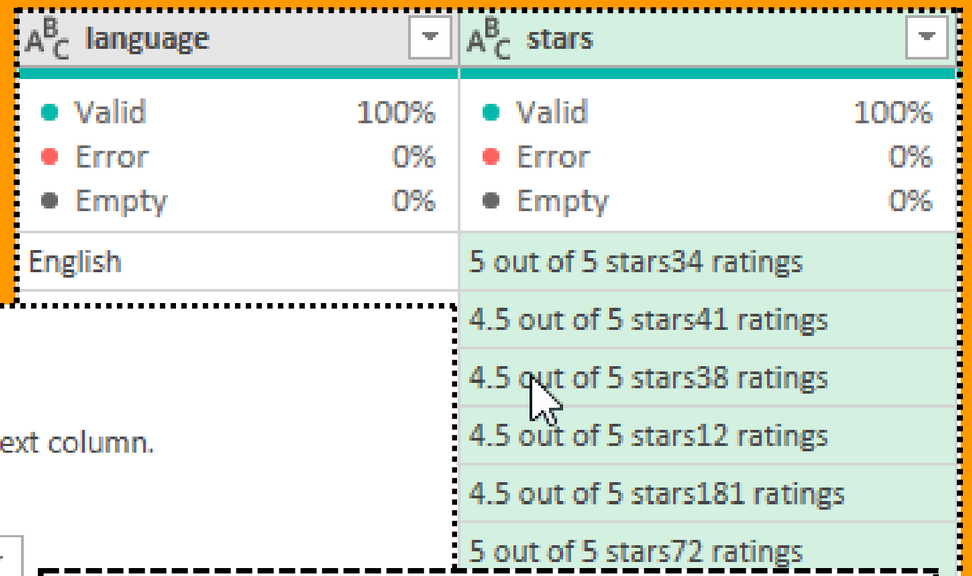


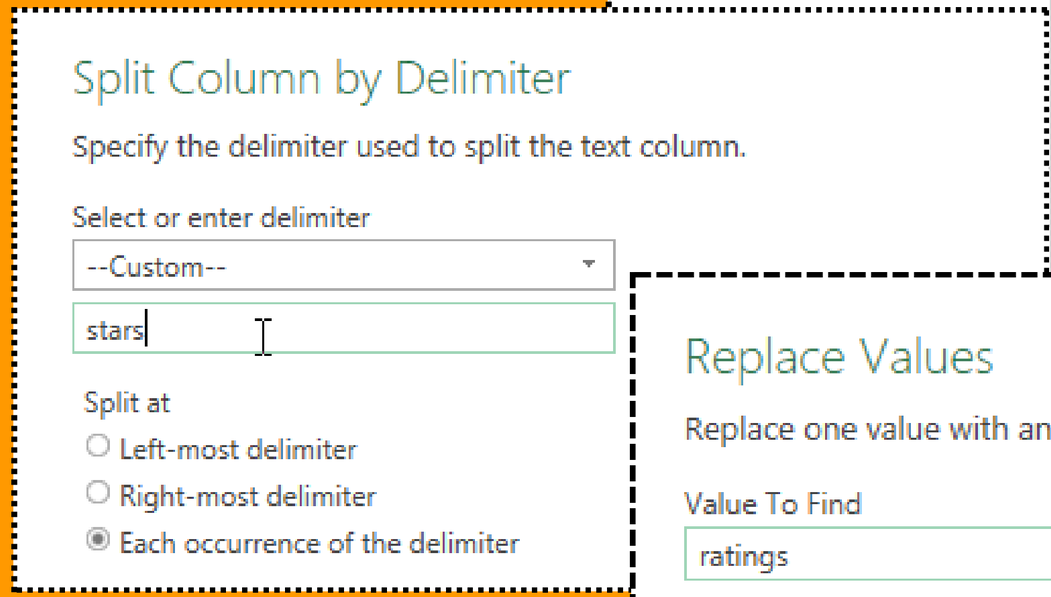
📌 Outcome: Prepares the column for accurate pricing summaries and visualizations.

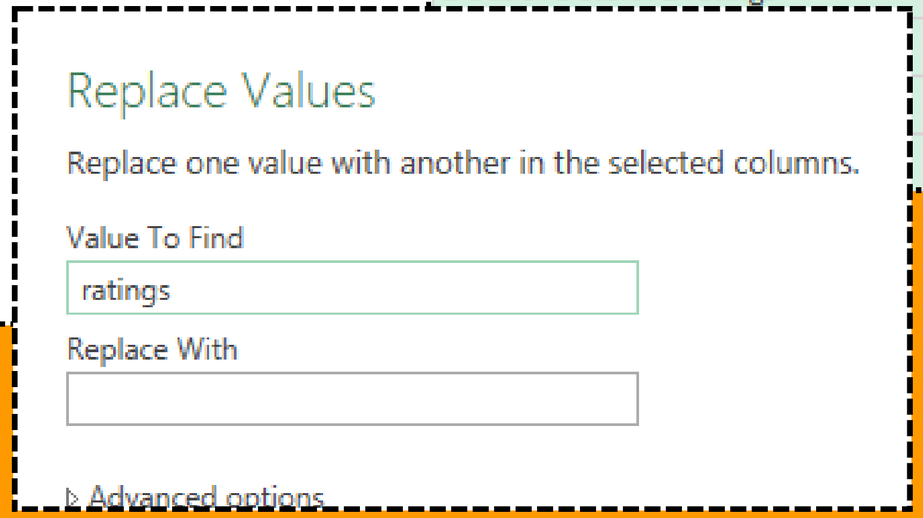
Task 06

Convert text ratings in the stars column to numeric values.

- To Extract Ratings We will Split Column using Delimiter “stars” and extract ratings. Then we will replace ratings with “blank” to acquire numerical values.
- Make sure to Change the Datatype to Whole Number after extracting ratings in numerical format.







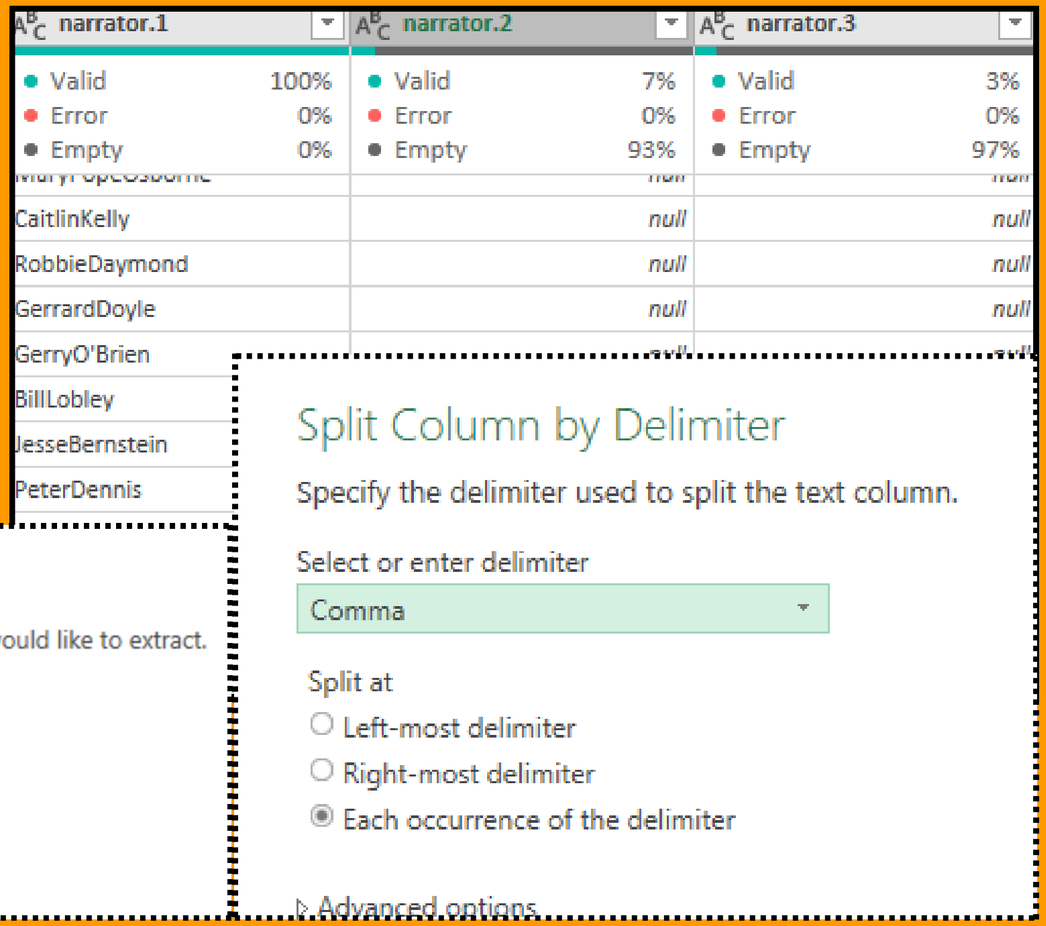
📌 Outcome:

Allows for rating-based filtering, sorting, and average rating analysis.

Task 07

Split the narrated by column into multiple columns if multiple narrators are listed.

- To Clean This Column: Go to Home > Format > Extract> Text After Delimiter and Enter ":" as Delimiter
- Next head over to Split Column to Split Narrators based on Comma
- Rename Columns as Narrator1,Narrator 2 for more clarity



📌 Outcome: Allows detailed narrator analysis (e.g., performance, preference).

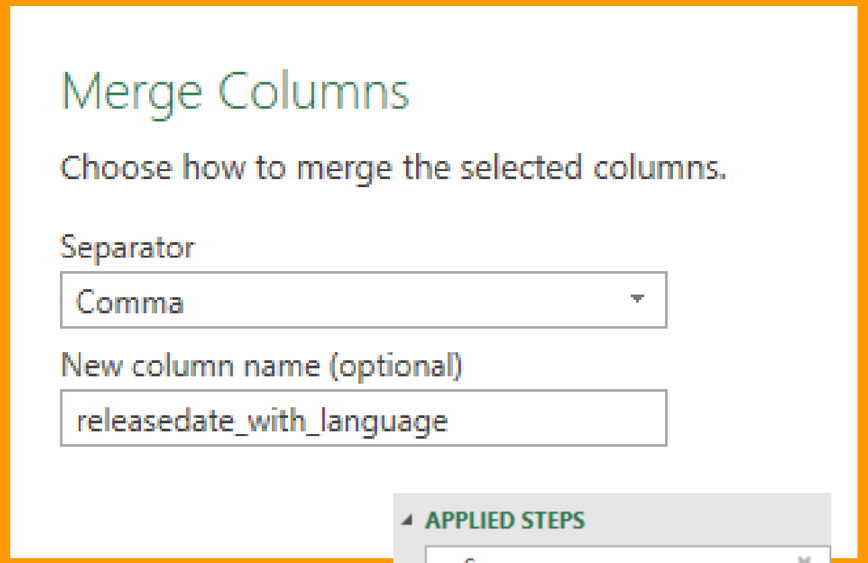
Task 08

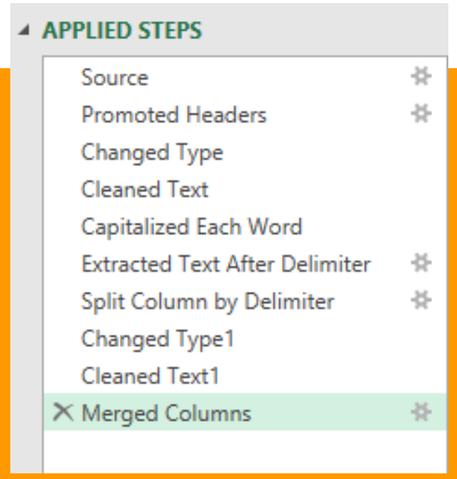
Merge the releasedate and language columns into a single new column named releaseinfo with the format "DD-MM-YYYY, Language."

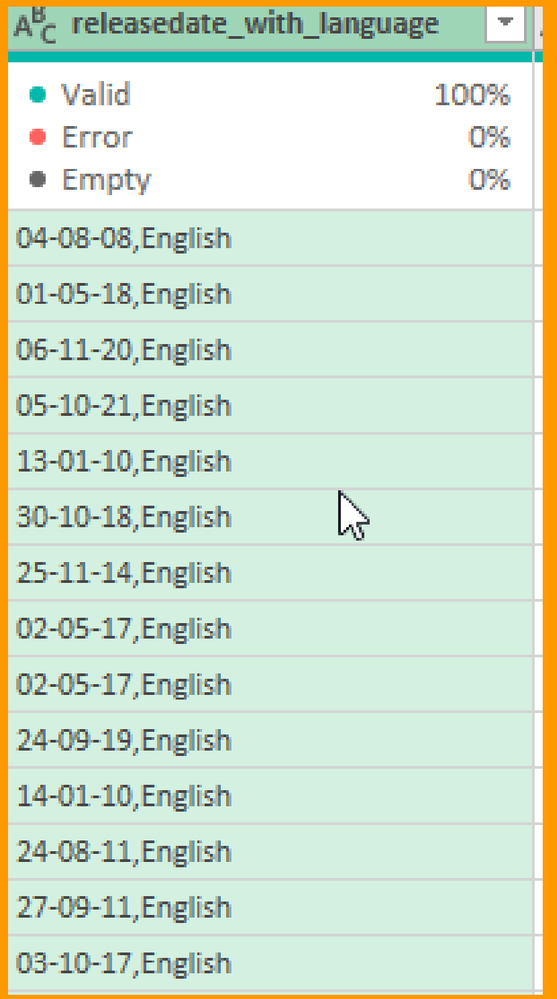
- Select releasedate and language columns and choose the Merge Columns Option using the Transform Tab.

- Make sure to Use Comma as a Separator.

📌 Outcome: Concise data representation that adds contextual clarity to reports.





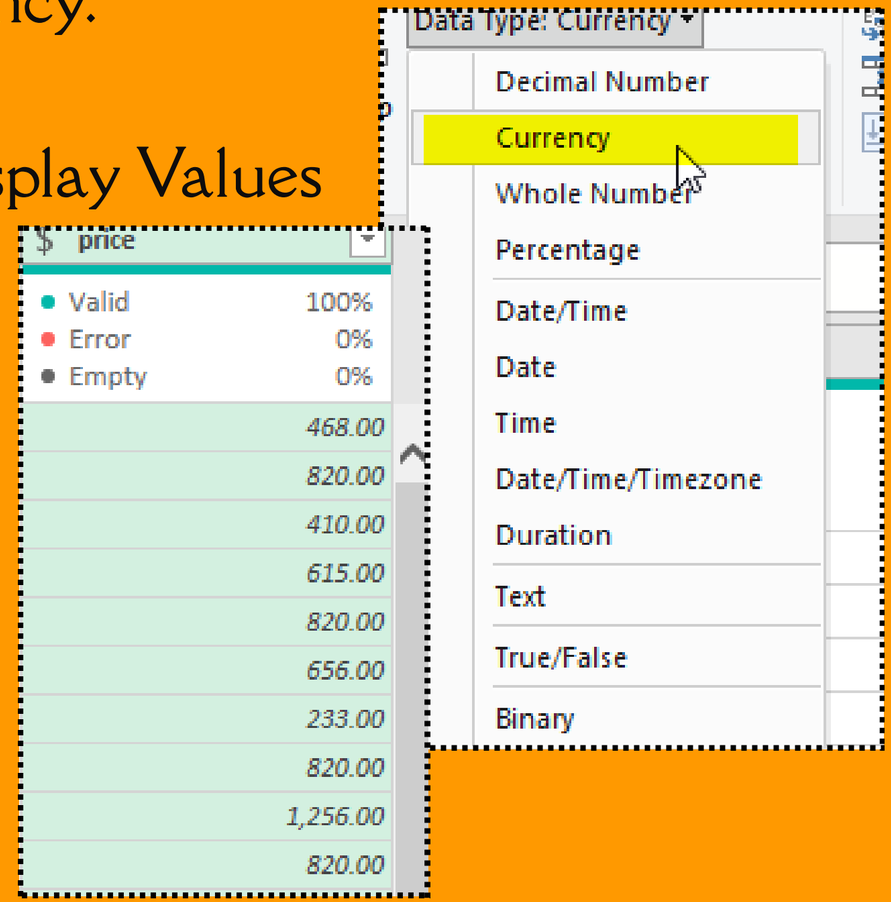


Task 09

Ensure all price entries are displayed with two decimal places.

- We have already cleaned the Price Column in Task 5. To ensure proper formatting we'll change the datatype into currency.

- Head over to Data Type and Choose Currency to Display Values with 02 Decimal Places



📌 Outcome:

Creates clean, professional appearance for financial data and ensures uniformity.

audible

Thankyou

This concludes my presentation. I appreciate your time and engagement.