



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.

The screenshot shows the Power Query Editor interface. A context menu is open over a column header, with the 'Format' option selected. Under 'Format', the 'Capitalize Each Word' option is highlighted with a yellow background and a red border. A red circle with the number '1' is positioned above the menu. To the right, the main Power Query window displays a table with columns for 'name', 'time', and 'releasedate'. The 'name' column contains book titles like 'Geronimo Stilton #11 & #12', 'The Burning Maze', etc. The 'Applied Steps' pane at the bottom lists the following steps: Source, Promoted Headers, Changed Type, Cleaned Text, and Capitalized Each Word (which is highlighted with a green background). The 'Applied Steps' pane has a small icon of a person's head with a gear.

A ^B name	A ^B time	A ^B releasedate
Geronimo Stilton #11 & #12	2 hrs and 20 mins	04-08-08
The Burning Maze		
The Deep End		
Daughter Of The Deep		
The Lightning Thief: Percy Jackson, Book 1		
The Hunger Games: Special Edition		
Quest For The Diamond Sword		
The Dark Prophecy		
Merlin Mission Collection		

Task 02



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

A	B	C	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

The dialog box has a title 'Text After Delimiter'. Below it is a descriptive message: 'Enter the delimiter that marks the beginning of what you would like to extract.' A 'Delimiter' input field contains a colon (:). There is also an 'Advanced options' link and 'OK' and 'Cancel' buttons at the bottom right.

- 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 occurrence of the delimiter to separate all authors and rename columns as Author 1 , Author 2 , Author 3.

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

"

Applied Steps

- Source
- Promoted Headers
- Changed Type
- Cleaned Text
- Capitalized Each Word
- Extracted Text After Delimiter
- Split Column by Delimiter

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

A ^B C author.1	A ^B C author.2	A ^B C author.3
● Valid ● Error ● Empty	100% 0% 0%	100% 0% 0%
Disney	Disney	Disney
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	BrianIlskov	null
AndyGriffiths	TerryDenton	null
AndrewPeterson-editor	JonathanRogers	N.D.Wilson
MacKenzieCadenhead	SeanRyan	AnkeAlbrecht-Übersetzer

Data After Cleaning

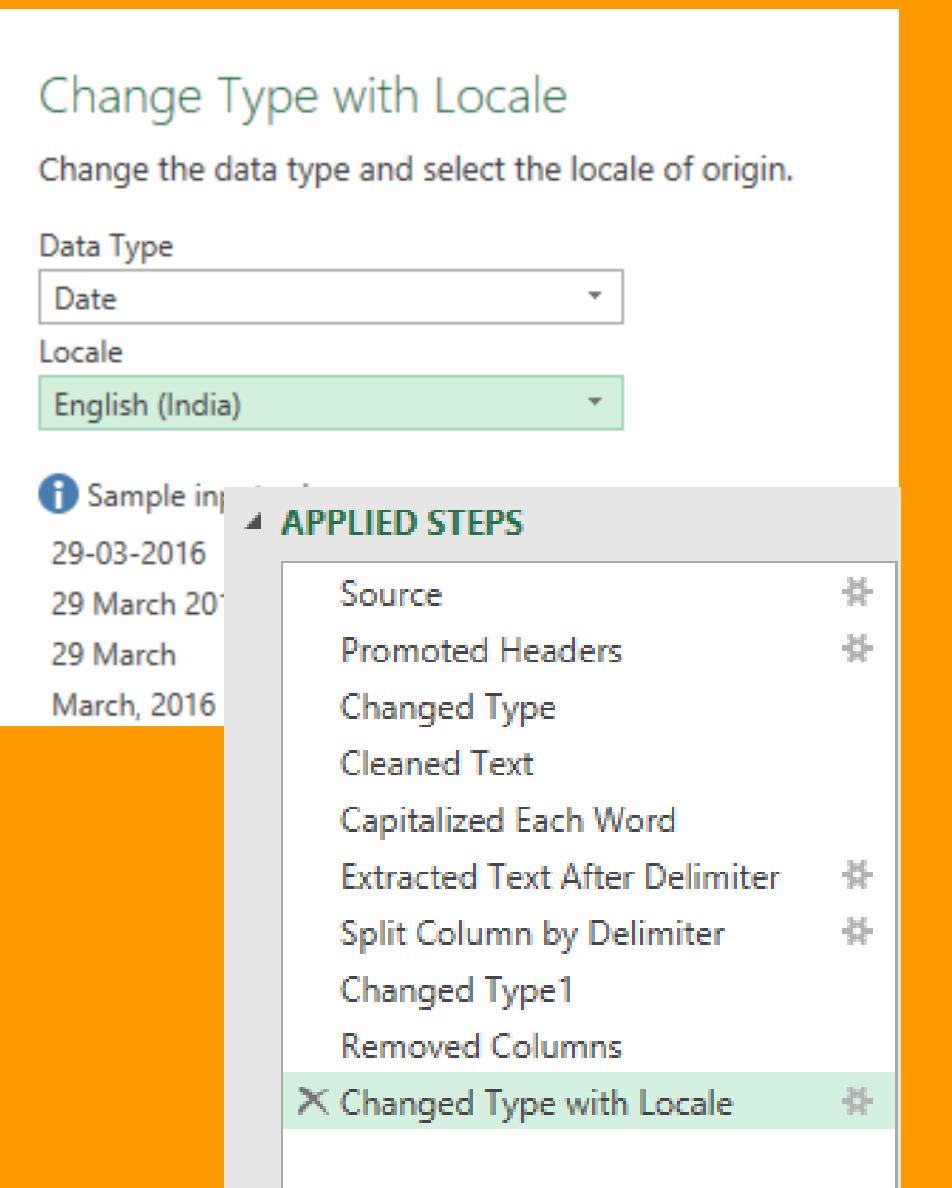
Task 03

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

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



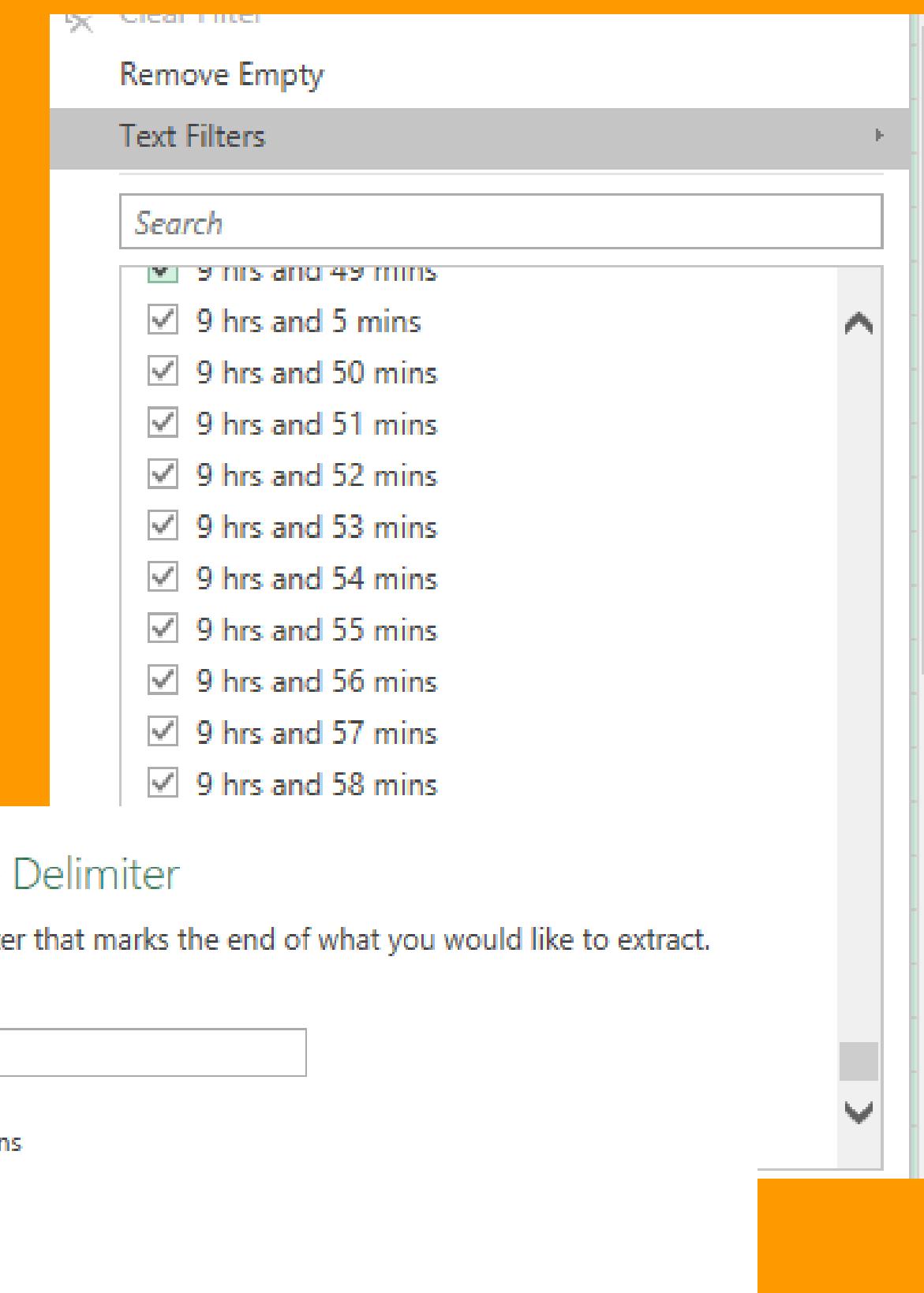
	releasedate	language
0%	● Valid	100%
0%	● Error	0%
0%	● Empty	0%
	04-Aug-08	English
	01-May-18	English
	06-Nov-20	English
	05-Oct-21	English
	13-Jan-10	English
	30-Oct-18	English
	25-Nov-14	English
	02-May-17	English
	02-May-17	English
	24-Sep-19	English
	14-Jan-10	English
	24-Aug-11	English
	27-Sep-11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



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”.



The screenshot shows the 'Text Filters' dialog box in Excel. At the top right are 'Clear Filter' and 'Remove Empty' buttons. Below is a 'Text Filters' section with a 'Search' input field containing '9 hrs and 49 mins'. A list of checked items follows, ranging from '9 hrs and 49 mins' to '9 hrs and 58 mins'. At the bottom left is a 'Text Before Delimiter' section with a 'Delimiter' input field containing 'hr'.

Text Filter Item
9 hrs and 49 mins
9 hrs and 5 mins
9 hrs and 50 mins
9 hrs and 51 mins
9 hrs and 52 mins
9 hrs and 53 mins
9 hrs and 54 mins
9 hrs and 55 mins
9 hrs and 56 mins
9 hrs and 57 mins
9 hrs and 58 mins

Text Before Delimiter
Enter the delimiter that marks the end of what you would like to extract.
Delimiter
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.

The screenshot shows two overlapping dialog boxes. The top dialog is titled "Text Filters" and contains a "Search" field and a list of checked items: 15 mins, 16, 16 mins, 17, 17 mins, 18 mins, 19 mins, 2, 2 mins, 20, 20 mins, 21 mins, and 22. The bottom dialog is titled "Replace Errors" and has the instruction "Enter the value which will replace errors in the selected columns." Below it is a "Value" input field containing the number "0".

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.

Last Characters
20 mins
8 mins
3 mins
16 mins
10 hrs
35 mins
23 mins
32 mins
56 mins
22 mins
48 mins

Add Conditional Column

Add a conditional column that is computed from the other columns or values.

New column name: MinutesCleaned2

If: MinutesCleaned contains ABC 123 Then: ABC 123

Else: MinutesCleaned

Output: 0

Add Clause

Else: MinutesCleaned

 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.

A screenshot of the Microsoft Power Query ribbon. The 'Hour_Minute' tab is selected. In the bottom-left corner, there is a dropdown menu with various data type options: Decimal Number, Currency, Whole Number, Percentage, Date/Time, Date, Time, Date/Time/Timezone, Duration, Text, True/False, Binary, and Using Locale... The 'Duration' option is highlighted with a mouse cursor.

A B	C	D
A	B	C
1.2	Decimal Number	100%
\$	Currency	0%
123	Whole Number	0%
%	Percentage	0%
Date/Time		0.02:20:00
Date		0.13:08:00
Time		0.02:03:00
Date/Time/Timezone		0.11:16:00
Duration		0.10:00:00
A B	Text	0.10:35:00
X V	True/False	0.02:23:00
Binary		0.12:32:00
Using Locale...		0.10:56:00
		0.13:22:00

1 2 3	Hour_Cleaned	1 2 3	MinutesCleaned
● Valid	100%	● Valid	100%
● Error	0%	● Error	0%
● Empty	0%	● Empty	0%
	2		20
	13		8
	2		3
	11		16
			0
			35
			23

Merge Columns

Choose how to merge the selected columns.

Separator

Colon

New column name (optional)

Hour_Minute

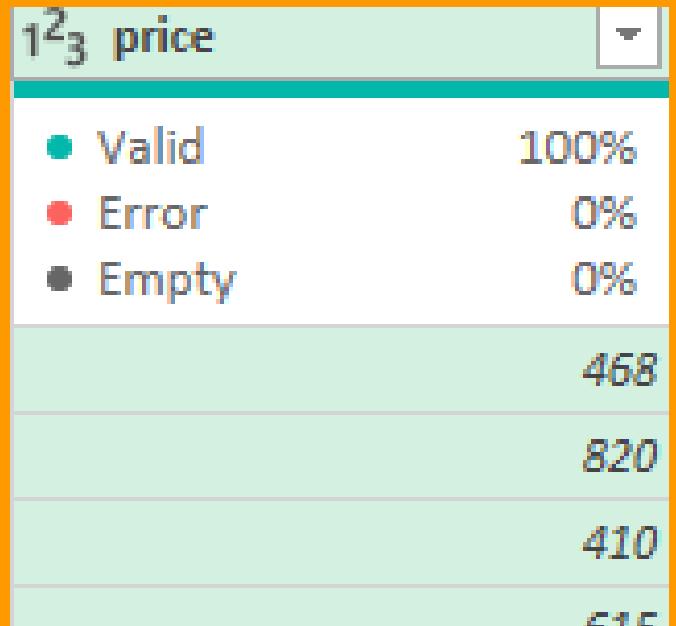
Task 05



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



Replace Values

Replace one value with another in the selected columns.

Value To Find: Free

Replace With: 0

Advanced options

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Replaced Value
- Changed Type1

A	B	C	price
			- %
			< 1%
			- %
			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

📌 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.

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

--Custom--

stars |

Split at

Left-most delimiter

Right-most delimiter

Each occurrence of the delimiter

A ^B Language	A ^B stars
● Valid 100%	● Valid 100%
● Error 0%	● Error 0%
● Empty 0%	● Empty 0%
English	5 out of 5 stars 34 ratings
	4.5 out of 5 stars 41 ratings
	4.5 out of 5 stars 38 ratings
	4.5 out of 5 stars 12 ratings
	4.5 out of 5 stars 181 ratings
	5 out of 5 stars 72 ratings

Replace Values

Replace one value with another in the selected columns.

Value To Find

ratings

Replace With

< Advanced options

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

Text After Delimiter

Enter the delimiter that marks the beginning of what you would like to extract.

Delimiter

:

Advanced options

A C narrator.1	A C narrator.2	A C narrator.3
Valid 100%	Valid 7%	Valid 3%
Error 0%	Error 0%	Error 0%
Empty 0%	Empty 93%	Empty 97%
many narrators	null	null
CaitlinKelly	null	null
RobbieDaymond	null	null
GerrardDoyle	null	null
GerryO'Brien	null	null
BillLobley		
JesseBernstein		
PeterDennis		

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



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.

Merge Columns

Choose how to merge the selected columns.

Separator: Comma

New column name (optional): releasedate_with_language

APPLIED STEPS

- Source *
- Promoted Headers *
- Changed Type
- Cleaned Text
- Capitalized Each Word
- Extracted Text After Delimiter *
- Split Column by Delimiter *
- Changed Type1
- Cleaned Text1
- X Merged Columns *

A	B	C	releasedate_with_language
			100%
			0%
			0%
	04-08-08,English		
	01-05-18,English		
	06-11-20,English		
	05-10-21,English		
	13-01-10,English		
	30-10-18,English		
	25-11-14,English		
	02-05-17,English		
	02-05-17,English		
	24-09-19,English		
	14-01-10,English		
	24-08-11,English		
	27-09-11,English		
	03-10-17,English		



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.

The screenshot shows a data editor interface. On the left, there is a table with a single column labeled '\$ price'. The table has three rows: 'Valid' (100%), 'Error' (0%), and 'Empty' (0%). Below the table are several numerical values: 468.00, 820.00, 410.00, 615.00, 820.00, 656.00, 233.00, 820.00, 1,256.00, and 820.00. On the right, a vertical dropdown menu titled 'Data Type: Currency' is open. The options listed are: Decimal Number, Currency (which is highlighted with a yellow background), Whole Number, Percentage, Date/Time, Date, Time, Date/Time/Timezone, Duration, Text, True/False, and Binary. The 'Currency' option is the one currently selected.



audible

Thankyou

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

