audible
an amazon company

Stories that speak to you

audible
an amazon company

Blog Products About us contact Login

Cleaning & Standardizing Audible Dataset Using Power Query Editor in Excel

Listen anytime, anywhere

audible
an amazon company

Project Objective

To clean and standardize the Audible dataset using Power Query Editor in Excel.
Ensure data consistency and proper formatting for seamless analysis.
Prepare a refined dataset ready for insights and visualization

The listening never has to stop

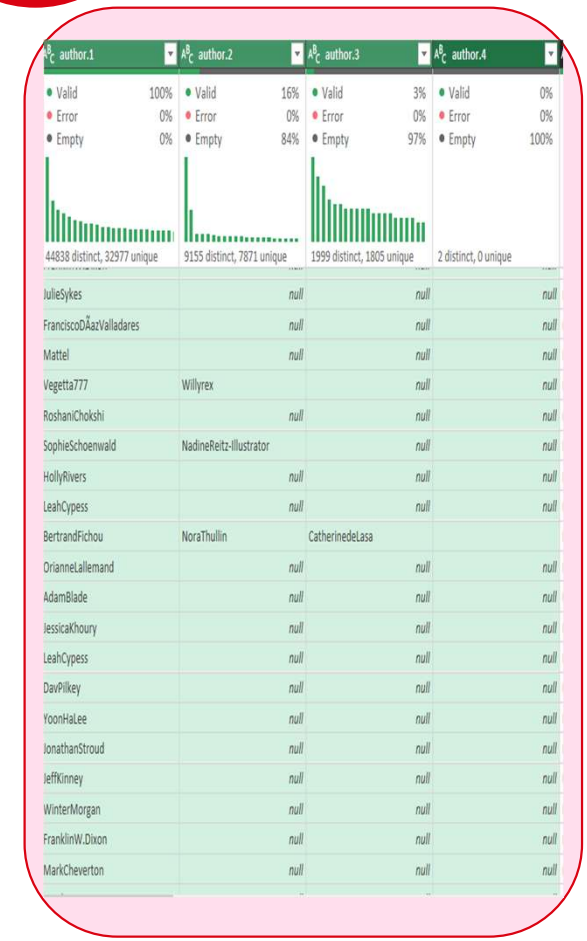
Tools Used

🛠 Microsoft Excel
With built-in Power Query Editor
📁 Audible Dataset
(CSV/Excel format)

Standardize the name column to ensure consistent title casing.

Standardize the Name Column to Title Case in Power Query
Objective: To ensure consistency and professionalism in the dataset by converting all names (e.g., book titles, authors, narrators) to Title Case—where the first letter of each word is capitalized.
Importance: Prevents duplicates due to case mismatches (e.g., "john doe" ≠ "John Doe")
Steps:-
Load your data into Power Query:
Select your data → Go to Data → Click From Table/Range
Select the 'Name' column:
This could be the Book Title, Author, or Narrator column
Apply Text Proper (Title Case):
Go to Transform tab
Click on Format → Choose Capitalize Each Word
john doe -> John Doe, THE POWER OF NOW -> The Power Of Now
Click "Close & Load" to return the cleaned data to Excel.``

audible
an amazon company

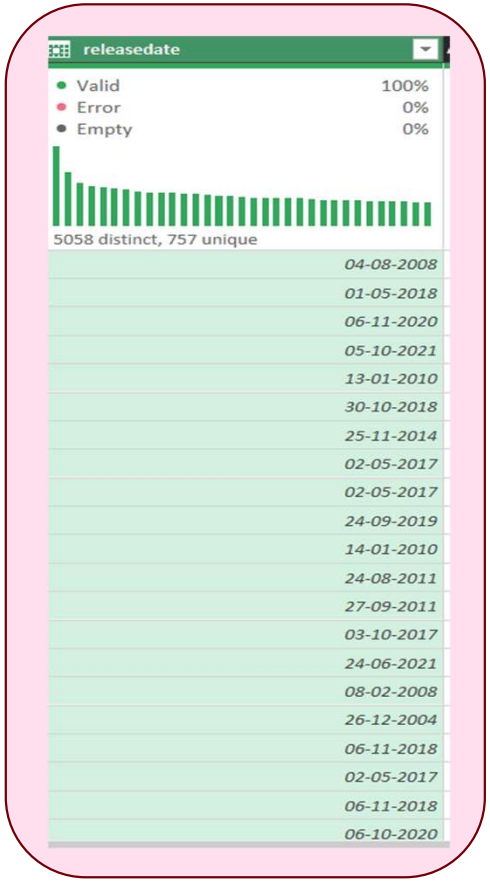


author.1	author.2	author.3	author.4
Valid 100%	Valid 16%	Valid 3%	Valid 0%
Error 0%	Error 0%	Error 0%	Error 0%
Empty 0%	Empty 84%	Empty 97%	Empty 100%
44838 distinct, 32977 unique	9155 distinct, 7871 unique	1999 distinct, 1805 unique	2 distinct, 0 unique
JulieSykes	null	null	null
FranciscoDÅazValladares	null	null	null
Mattel	null	null	null
Vegetta777	Willyrex	null	null
RoshaniChokshi	null	null	null
SophieSchoenwald	NadineReitz-Illustrator	null	null
HollyRivers	null	null	null
LeahCypess	null	null	null
BertrandFichou	NoraThullin	CatherinedeLasa	
OrianneLallemand	null	null	null
AdamBlade	null	null	null
JessicaKhoury	null	null	null
LeahCypess	null	null	null
DavPilkey	null	null	null
YoonHaLee	null	null	null
JonathanStroud	null	null	null
JeffKinney	null	null	null
WinterMorgan	null	null	null
FranklinW.Dixon	null	null	null
MarkCheverton	null	null	null

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

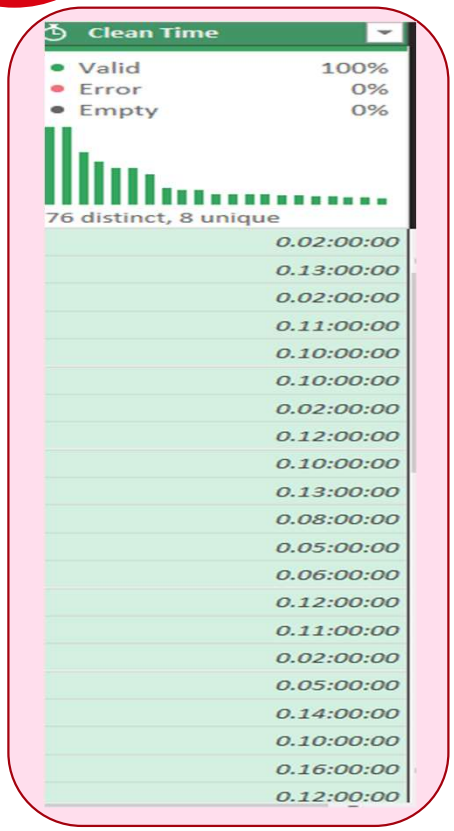
Separate Combined Names in the Author Column (Multiple Authors)
Objective:Enable individual-level analysis (e.g., count books per author)
Support filtering, grouping, and pivoting by individual author
Prepare the dataset for normalization or relational modeling
Steps:-
Load into Power Query
Select your dataset → Data tab → From Table/Range
Select the Author Column
Standardize Separators:- If authors are joined by different connectors like & or and first replace them with a common delimiter like comma
Transform → Replace Values
Go to Home → Split Column → By Delimiter
Choose delimiter:- Comma (,)
Choose: Split into Rows *(for better analysis)* or Split into Columns
Result:-
Clean, standardized author names
Better for filtering, analysis, or relational database structure

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



Standardize the release date Column to Format: DD-MM-YYYY in Power Query
Objective:- To ensure that all values in the Release Date Column follow a uniform date format which is essential for:
Accurate filtering, sorting, and grouping by date
Consistent time-based analysis (e.g., monthly/yearly trends)
Preventing calculation errors caused by inconsistent formats
Steps:-
Load Your Data into Power Query
Select your data → go to Data tab → click From Table/Range
Select the Release Date Column
Ensure Column is in Date Format
Go to Transform tab → Click on Data Type → choose Date
Result:- A clean, consistent Release Date column in DD-MM-YYYY format
Ready for timeline visualizations, filters, or trend analysis

audible
an amazon company



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

Convert Time Column from Text to Excel Duration Format

Objective: To convert time values like "10 hrs and 22 min" or "45 min" from text format to Excel-recognized duration format so that:

Time-based calculations (e.g., total listening time) can be performed

Steps:- Load Data into Power Query

Select your dataset → go to Data → click Get & Transform Data → From Table/Range

Select the Duration Column

Let's assume the column is named Listing Time

Replace Text for Easier Parsing

Go to Transform → Replace Values: Replace " hrs and " with :

Replace " min" with `` So "10 hrs and 15 min" becomes "10:15"

Change Data Type to Duration

Select the column Go to Transform → Data Type → choose Duration

Result:- Transforms messy text like "1 hr and 20 min" into clean duration

Enables accurate time-based insights

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

Ensure the price Column is in Numeric Format and Identify Non-Numeric Values
Objective:- Convert the price column to pure numeric format
Detect and flag any non-numeric or invalid entries
Steps:- Load Data into Power Query Select your dataset → Go to Data → Click From Table/Range Remove Currency Symbols (Optional)
If price values contain symbols like $ remove them: Select the price column → Transform → Replace Values: Attempt to Change Data Type to Decimal Number
Go to Transform → Data Type → Select Decimal Number Identify Non-Numeric Entries
To catch the errors:
Go to Home → Click the "Remove Errors" dropdown → Choose "Keep Errors"
This filters only rows with invalid entries.
Result:- Ensures data consistency for numeric operations
Flags invalid or manually entered values for review
Prepares dataset for financial dashboards or KPIs

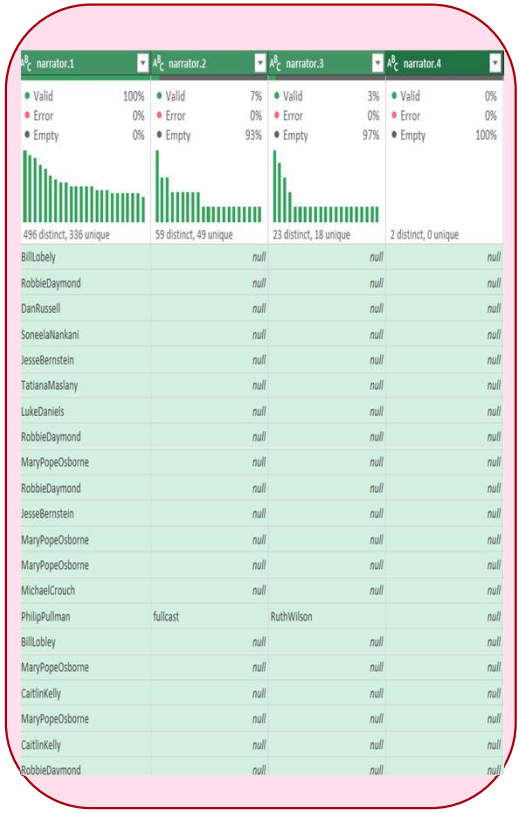
audible
an amazon company

1.2 rating out of 5	123 total review
Valid 23%	Valid 23%
Error 0%	Error 0%
Empty 77%	Empty 77%
6 distinct, 1 unique	38 distinct, 16 unique
5	34
4.5	41
4.5	38
4.5	12
4.5	181
5	72
5	11
5	50
5	5
5	58
4.5	130
5	6
5	7
5	41
4	2
4.5	33
5	7
5	20
5	11
5	13
5	79

Convert text ratings in the Stars column to numeric values.

Convert Text Ratings in the Stars Column to Numeric Values
Objective:- To convert textual star ratings to pure numeric values
Quantitative analysis, Preparing for visualizations, KPIs, and performance metrics
Steps:- Load Your Data into Power Query
Select your data → Go to Data → Click From Table/Range
Select the Stars Column Extract the Numeric Part
Go to: Add Column → Extract → Text Before Delimiter Delimiter: “out”
This will keep only the numeric part before the text
Change Data Type to Decimal Number Select the new Extracted Rating column
Go to Transform → Data Type → Decimal Number
Rename Column (Optional) Rename Extracted Rating -> Stars rating
Remove Original Column (Optional)
Result:- Enables accurate averaging, ranking, and comparison
Supports integration with charts, filters, KPIs, dashboards.

audible
an amazon company



narrator.1	narrator.2	narrator.3	narrator.4
Valid 100% Error 0% Empty 0%	Valid 7% Error 0% Empty 93%	Valid 3% Error 0% Empty 97%	Valid 0% Error 0% Empty 100%
496 distinct, 336 unique	59 distinct, 49 unique	23 distinct, 18 unique	2 distinct, 0 unique
BillLobely	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
MaryPopeOsborne	null	null	null
CaitlinKelly	null	null	null
MaryPopeOsborne	null	null	null
CaitlinKelly	null	null	null
RobbieDaymond	null	null	null

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

Split the Narrated by column into Multiple Columns (for Multiple Narrators)

Objective:- Analyzing performance by individual narrators

Filtering or grouping based on single narrators

Ensuring data is clean, structured, and analysis-ready

Steps:- Load Data into Power Query

Select your data → Go to Data tab → Click From Table/Range

Select the Narrated by Column

Standardize the Separator Replace all with a common delimiter, e.g., comma

Go to: Transform → Replace Values: Split Column by Delimiter Go to:

Home → Split Column → By Delimiter

Choose Comma (,) Choose Split into Columns

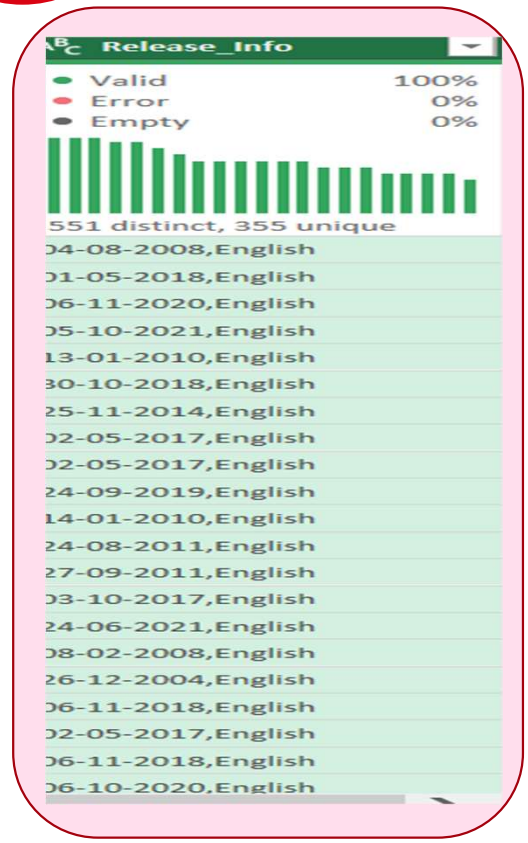
Name the columns: Narrator 1, Narrator 2

Now each narrator will appear in a separate column.

Result:- Enables individual analysis (e.g., how many audiobooks a narrator has done)

Ensures data normalization and readability

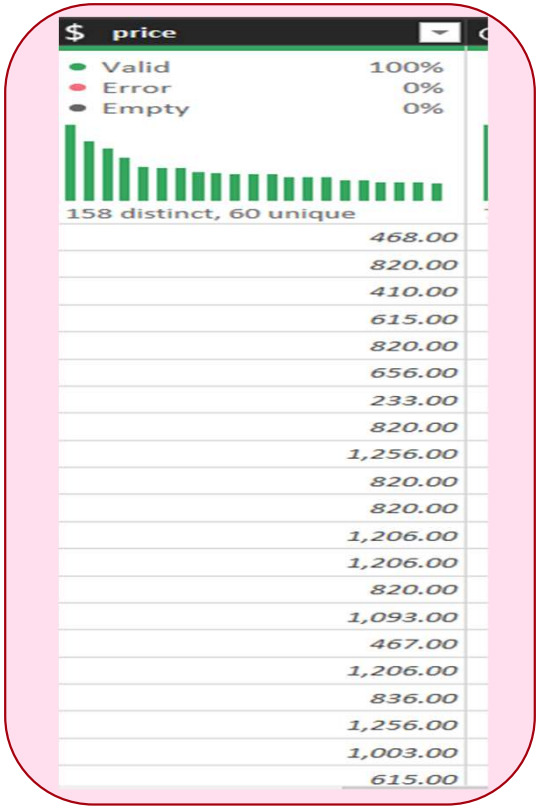
audible
an amazon company



Merge the release date and language columns into a single new column named release info with the format "DD-MM-YYYY, Language.“

Merge release date and language columns into a new columns release info
Objective: To create a new column that combines the release date and language in a readable format:"DD-MM-YYYY, Language“
This helps: Improve clarity when analyzing or displaying data
Combine key metadata into a single field for reports or dashboards
Simplify export or summaries (e.g., “01-01-2023, English”)
Steps:-Load Data into Power Query
Select your dataset → Go to Data → From Table/Range
Ensure Release Date in date format , Select Release Date Column , Go to Transform → Change Data Type → Date
Add Column → Format → Format Date → Choose Custom
Add Column → Custom Column
Rename the Column to Release Info
Result:- Merged Release Date and language into a single readable column Ensures consistency and simplifies display or export
Output is clean and well-formatted for reporting

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



Format Currency Values in the Price Column with Two Decimal Places
Objective:- Numerically consistent (e.g., ₹199 → ₹199.00)
Formatted uniformly for reporting and dashboards
Ready for accurate sorting, filtering, aggregation, and comparison
Steps:- Load Data into Power Query
Select your dataset → go to Data → From Table/Range
Select the Price Column Ensure It's a Numeric Type
Go to Transform tab → Click on Data Type → Select Decimal Number
If the column contains currency symbols like ₹ or $, first remove them:
Go to Transform → Replace Values Replace "₹“ with "“ Replace "$“ with "“
Then change data type again to Decimal Number
Result:- Ensures clean and consistent presentation of prices
Prevents issues in financial calculations or pivot tables
Creates a professional, analysis-ready dataset.

Here is the link of audible case study clean data set.

https://1drv.ms/x/c/cf7e597e0fbefd38/EchUICXgPWlDpE_Jb1oEPfUBxcE7-98zEWaA82QNdTXcbA?e=O3qSqd

Audible Case Study – Data Cleaning & Analysis using Excel

This project involves cleaning, transforming, and analyzing an Audible dataset using **Microsoft Excel**, with a focus on preparing the data for insights through the **Power Query Editor**, formulas, and dashboards.

Outcome:- A clean and analysis-ready dataset suitable for generating business insights related to audiobooks, narrators, pricing trends, and more.

Key Objectives: Clean and standardize messy data
Format and transform columns (e.g., time, price, release date)
Split combined values and normalize text
Merge relevant fields for improved readability
Prepare data for analysis and reporting

audible
an amazon company

Thank You!

We Respect your valuable time with **audible (an amazon company)**
If you have any questions, please reach us