

audible



AUDIBLE DATA CLEANING PROJECT

Presented by N. Hari Krishna



Introduction

"Welcome to our presentation on Audible data cleaning using Power Query. We used this tool to clean and organize Audible's data, ensuring accuracy and consistency for further analysis."

Audible is a leading provider of spoken-word entertainment and audiobooks, offering a vast selection of titles across various genres. It allows users to listen to books, podcasts, and other audio content on the go through its app and website.



Clean and standardize

Remove inconsistencies and standardize data formats.

Formatting column

Properly format columns to maintain uniformity.



Project objectives



Ensure Data Consistency

Verify and correct data to ensure consistency across the dataset.

Prepare Dataset for Analysis

Organize and structure data for accurate and efficient analysis.

Step - 1

Standardized names, separated combined author names, and split the 'narrated by' column for clarity and consistency."



Step - 3

Formatted the price column numerically, merged release date and language into 'Release Info,' and ensured currency values have two decimal points."



Step - 2

"Standardized release dates, converted time to duration format, and changed star ratings to numeric values."



Uncleaned data set

name	author	narrator	time	releasedate	language	stars	price
Geronimo Stilton #11 & #12	Writtenby:GeronimoStilton	Narratedby:BillLobely	2 hrs and 20 mins	04-08-2008	English	5 out of 5 stars34 ratings	468
The Burning Maze	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 8 mins	01-05-2018	English	4.5 out of 5 stars41 ratings	820
The Deep End	Writtenby:JeffKinney	Narratedby:DanRussell	2 hrs and 3 mins	06-11-2020	English	4.5 out of 5 stars38 ratings	410
Daughter of the Deep	Writtenby:RickRiordan	Narratedby:SoneelaNankani	11 hrs and 16 mins	05-10-2021	English	4.5 out of 5 stars12 ratings	615
The Lightning Thief: Percy Jackson, Book 1	Writtenby:RickRiordan	Narratedby:JesseBernstein	10 hrs	13-01-2010	English	4.5 out of 5 stars181 ratings	820
The Hunger Games: Special Edition	Writtenby:SuzanneCollins	Narratedby:TatianaMaslany	10 hrs and 35 mins	30-10-2018	English	5 out of 5 stars72 ratings	656
Quest for the Diamond Sword	Writtenby:WinterMorgan	Narratedby:LukeDaniels	2 hrs and 23 mins	25-11-2014	English	5 out of 5 stars11 ratings	233
The Dark Prophecy	Writtenby:RickRiordan	Narratedby:RobbieDaymond	12 hrs and 32 mins	02-05-2017	English	5 out of 5 stars50 ratings	820
Merlin Mission Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	10 hrs and 56 mins	02-05-2017	English	5 out of 5 stars5 ratings	1,256.00
The Tyrant's Tomb	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 22 mins	24-09-2019	English	5 out of 5 stars58 ratings	820
The Titan's Curse: Percy Jackson, Book 3	Writtenby:RickRiordan	Narratedby:JesseBernstein	8 hrs and 48 mins	14-01-2010	English	4.5 out of 5 stars130 ratings	820
Magic Tree House Collection: Books 9-16	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	5 hrs and 23 mins	24-08-2011	English	5 out of 5 stars6 ratings	1,206.00
Magic Tree House Collection: Books 1-8	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	6 hrs and 1 min	27-09-2011	English	5 out of 5 stars7 ratings	1,206.00
Magnus Chase and the Ship of the Dead	Writtenby:RickRiordan	Narratedby:MichaelCrouch	12 hrs and 58 mins	03-10-2017	English	5 out of 5 stars41 ratings	820
Northern Lights	Writtenby:PhilipPullman	Narratedby:PhilipPullman,fullcast,RuthWilson	11 hrs and 55 mins	24-06-2021	English	4 out of 5 stars2 ratings	1,093.00
Geronimo Stilton #13 and #14	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 25 mins	08-02-2008	English	4.5 out of 5 stars33 ratings	467
Magic Tree House Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	5 hrs and 4 mins	26-12-2004	English	5 out of 5 stars7 ratings	1,206.00
Exile	Writtenby:ShannonMessenger	Narratedby:CaitlinKelly	14 hrs and 41 mins	06-11-2018	English	5 out of 5 stars20 ratings	836
Merlin Mission Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	10 hrs and 18 mins	02-05-2017	English	5 out of 5 stars11 ratings	1,256.00

[Link](#) 

Step - 1

- **Select Name Column → Go to Transform Tab → Select Format → Choose Capitalize Each Word**
- **Transform Tab → Select Extract → Choose After Delimiter [:]**
- **In the Transform tab, we selected the Extract option and chose "After Delimiter [:]".**
This extracted all text following the specified delimiter in the selected column.
- **Transform Tab → Select Split → Fill Null with "Not Applicable"**
- **After selecting Split in the Transform tab, we divided the text into separate columns based on the delimiter [,]. Then, we filled any null values with "Not Applicable" to ensure data completeness.**
- **Select Author 1 Column → Select Transform Tab → Select Split → Split by Lower Case to Upper Case → Merge New Columns as Author.1 → Trim Author.1 Column**
- **We selected the Author 1 column and used the Transform tab to split the data by changing from lowercase to uppercase, creating new columns. We then merged these columns into a single column named Author.1 and trimmed any extra spaces to ensure a clean and consistent format.**
- **Repeat the same process for author.2 and author.3 columns as well as**

Step - 1

- **Transform Tab → Select Extract → Choose After Delimiter [:]**
- **In the Transform tab, we selected the Extract option and chose "After Delimiter [:]". This extracted all text following the specified delimiter in the selected column.**
- **Transform Tab → Select Split → Fill Null with "Not Applicable"**
- **After selecting Split in the Transform tab, we divided the text into separate columns based on the delimiter [,]. Then, we filled any null values with "Not Applicable" to ensure data completeness.**
- **Select narrator 1 Column → Select Transform Tab → Select Split → Split by Lower Case to Upper Case → Merge New Columns as Author.1 → Trim Author.1 Column**
- **We selected the Author 1 column and used the Transform tab to split the data by changing from lowercase to uppercase, creating new columns. We then merged these columns into a single column named Author.1 and trimmed any extra spaces to ensure a clean and consistent format.**
- **Repeat the same process for narrator 2 and narrator 3 columns as well as**

Step - 2

Custom column formula ⓘ

```
= if Text.Contains([time],"hrs")
  then Number.FromText(Text.BeforeDelimiter([time], " hrs"))

else if Text.Contains([time],"hr")
  then Number.FromText(Text.BeforeDelimiter([time]," hr"))

else 0
```

Custom column formula ⓘ

```
= if Text.Contains([time],"min") and not Text.Contains([time]
,"and") then
  Number.FromText(Text.BeforeDelimiter([time]," min"))

else if Text.Contains([time],"min")
  then Number.FromText(Text.BeforeDelimiter
(Text.AfterDelimiter([time],"and "),"min"))

else 0
```

- **Select Released Date Column → Change Data Type from Text to Date Format (DD-MM-YYYY)**
- **We selected the Released Date column and changed its data type from text to date format (DD-MM-YYYY). This ensures that all dates are consistently formatted for accurate analysis.**
- **Select Time Column → Apply Formula to Extract Hours and Minutes → Convert Hours and Minutes to Minutes → Add Both Columns → Create New Column Named Duration in Minutes**
- **We selected the Time column and changed its data type from text to duration format. We then applied a formula to extract hours and minutes, converted them to minutes, summed them, and created a new column named Duration in Minutes.**

Step - 2

Custom column formula ⓘ

```
= if Text.Contains([stars],"ratings")
  then Number.FromText(Text.BeforeDelimiter
    (Text.AfterDelimiter([stars],"stars"),"ratings"))

  else null
```

- **Select Ratings Column → Create Custom Column → Use Formula to Extract Numeric Ratings**
- **We selected the Ratings column and created a custom column using a formula to convert text ratings into numeric values, ensuring proper data analysis.**
- **Replace Null Values with 0 → Extract Numeric Values in Stars → Convert Text Format to Numeric Format**
- **We replaced null values with 0 in the Ratings column, extracted numeric values in stars, and converted the text format to numeric format to facilitate accurate data analysis.**



Step - 3

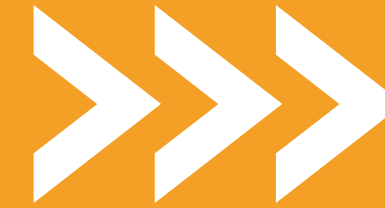
- **Select Prices Column → Change Data Type to Currency Format → Set Decimal Places to 2**
- **We selected the Prices column, changed its data type to currency format, and set it to display with 2 decimal points for consistent financial representation.**
- **Select Release Date and Language Columns → Merge Columns → Separate with Comma → Name New Column as Release Info**
- **We selected the Release Date and Language columns, merged them with a comma separator, and named the new column Release Info for consolidated information.**

Applied steps

- Source
- Promoted Headers
- Changed Type
- Capitalized Each Word
- Extracted Text After Delimi...
- Split Column by Delimiter
- Replaced Value
- Split Column by Character ...
- Merged Columns
- Trimmed Text
- Split Column by Character ...
- Merged Columns1
- Filtered Rows
- Split Column by Character ...
- Removed Columns
- Merged Columns2
- Filtered Rows1
- Extracted Text After Delimi...
- Split Column by Delimiter1
- Changed Type1
- Replaced Value1
- Removed Columns1



- Removed Columns1
- Split Column by Character ...
- Merged Columns3
- Trimmed Text1
- Split Column by Character ...
- ✕ Merged Columns4
- Split Column by C Merged Colum
- Merged Columns5
- Trimmed Text2
- Trimmed Text3
- Added Custom
- Reordered Columns
- Renamed Columns
- Added Custom1
- Inserted Addition
- Renamed Columns1
- Removed Columns2
- Changed Type2
- Removed Columns3
- Replaced Errors
- Changed Type3
- Added Custom2



- Replaced Errors
- Changed Type3
- Added Custom2
- Extracted Text Before Deli...
- Replaced Value2
- Changed Type4
- Replaced Value3
- ✕ Merged Columns6

Cleaned data set

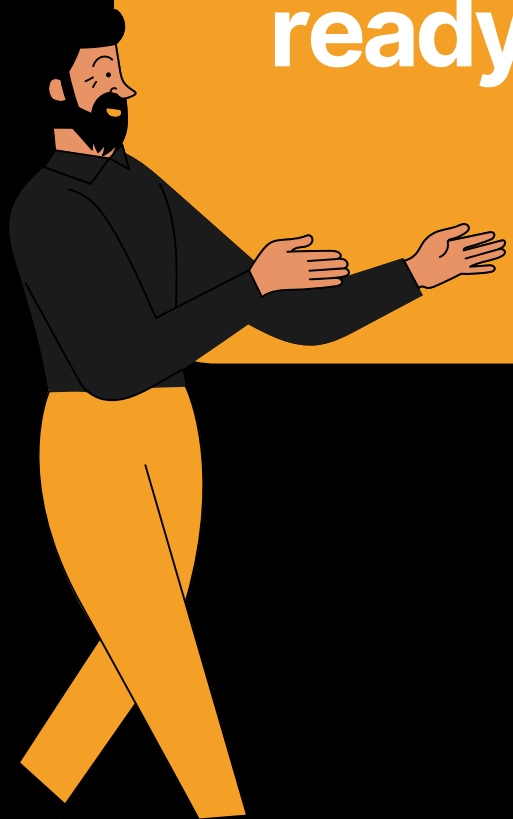


name	author.1	author.2	author.3	narrator.1	narrator.2	narrator.3	release info	stars	price	Duration	ratings
Atomic Habits	James Clear	Not applicable	Not applicable	James Clear	Not applicable	Not applicable	18-10-2018,English	5	820	40	12573
Atomic Habits	James Clear	Not applicable	Not applicable	James Clear	Not applicable	Not applicable	18-10-2018,English	5	820	40	12569
Ikigai	Héctor García	Francesc Miralles	Not applicable	Naoko Mori	Not applicable	Not applicable	18-09-2017,English	4.5	615	26	6981
The Psychology Of Money	Morgan Housel	Not applicable	Not applicable	Chris Hill	Not applicable	Not applicable	08-09-2020,English	4.5	668	53	5973
The Psychology Of Money	Morgan Housel	Not applicable	Not applicable	Chris Hill	Not applicable	Not applicable	08-09-2020,English	4.5	668	53	5972
Sapiens	Yuval Noah Harari	Not applicable	Not applicable	Derek Perkins	Not applicable	Not applicable	30-04-2015,English	5	957	33	5500
Rich Dad Poor Dad	Robert T.Kiyosaki	Not applicable	Not applicable	Tim Wheeler	Not applicable	Not applicable	15-06-2012,English	4.5	837	15	5324
Life'S Amazing Secrets	Gaur Gopal Das	Not applicable	Not applicable	Gaur Gopal Das	Not applicable	Not applicable	10-12-2019,English	4.5	1005	31	4184
The Subtle Art Of Not Giving A F*Ck	Mark Manson	Not applicable	Not applicable	Roger Wayne	Not applicable	Not applicable	13-09-2016,English	4.5	1181	22	3461
Becoming	Michelle Obama	Not applicable	Not applicable	Michelle Obama	Not applicable	Not applicable	13-11-2018,English	4.5	957	22	2924
How To Win Friends And Influence People	Dale Carnegie	Not applicable	Not applicable	Shernaz Patel	Not applicable	Not applicable	08-11-2018,English	4.5	155	13	2494
Can'T Hurt Me	David Goggins	Not applicable	Not applicable	David Goggins	Adam Skolnick	Not applicable	28-11-2018,English	5	836	50	2277
21 Lessons For The 21St Century	Yuval Noah Harari	Not applicable	Not applicable	Derek Perkins	Not applicable	Not applicable	30-08-2018,English	4.5	957	52	2236
Elon Musk	Ashlee Vance	Not applicable	Not applicable	Fred Sanders	Not applicable	Not applicable	28-04-2016,English	4.5	820	36	2161
The Power Of Habit	Charles Duhigg	Not applicable	Not applicable	Mike Chamberlain	Not applicable	Not applicable	05-04-2012,English	4.5	820	63	1869
The Monk Who Sold His Ferrari	Robin Sharma	Not applicable	Not applicable	Robin Sharma	Not applicable	Not applicable	05-07-2007,English	4.5	351	42	1838
Wings Of Fire	APJAbdul Kalam	Arun Tiwari	Not applicable	Girish Karnad	Not applicable	Not applicable	01-04-2020,English	5	75	16	1827
Wings Of Fire	APJAbdul Kalam	Arun Tiwari	Not applicable	Girish Karnad	Not applicable	Not applicable	01-04-2020,English	5	75	16	1823
Zero To One	Peter Thiel	Blake Masters	Not applicable	Blake Masters	Not applicable	Not applicable	04-06-2015,English	4.5	615	54	1807
A Promised Land	Barack Obama	Not applicable	Not applicable	Barack Obama	Not applicable	Not applicable	17-11-2020,English	4.5	1708	39	1636
The Compound Effect	Darren Hardy	Not applicable	Not applicable	Darren Hardy	Not applicable	Not applicable	22-01-2019,English	4.5	608	48	1564
The Compound Effect	Darren Hardy	Not applicable	Not applicable	Darren Hardy	Not applicable	Not applicable	22-01-2019,English	4.5	608	48	1563
Think Like A Monk	Jay Shetty	Not applicable	Not applicable	Jay Shetty	Not applicable	Not applicable	08-09-2020,English	4.5	569	64	1539
Raavan	Amish Tripathi	Not applicable	Not applicable	Sagar Arya	Not applicable	Not applicable	04-10-2019,English	4.5	342	43	1491
Homo Deus	Yuval Noah Harari	Not applicable	Not applicable	Derek Perkins	Not applicable	Not applicable	08-09-2016,English	4.5	957	67	1490
Harry Potter And The Philosopher'S Stone, Book 1	J.K.Rowling	Not applicable	Not applicable	Stephen Fry	Not applicable	Not applicable	20-11-2015,English	5	999	42	1425
Harry Potter And The Philosopher'S Stone, Book 1	J.K.Rowling	Not applicable	Not applicable	Stephen Fry	Not applicable	Not applicable	20-11-2015,English	5	999	42	1423
The Girl In Room 105	Chetan Bhagat	Not applicable	Not applicable	Siddhanta Datta	Not applicable	Not applicable	12-11-2018,English	4	568	31	1412

[Link](#)

Conclusion

In this project, we successfully cleaned and standardized Audible's dataset using Power Query. This involved ensuring consistent formatting, correcting data inconsistencies, and preparing the data for accurate analysis. The resulting clean and organized dataset is now ready for deeper insights and analysis, demonstrating the importance of thorough data preparation in achieving reliable results.



**Follow to
get more
information
like this**

FOLLOW US!