

Project presentation

Presented by

PRADEEP MAURYA



Audible Data Cleaning Project With Power Query Editor

Introduction

Audible is an American online audiobook and podcast service that allows users to purchase and stream audiobooks and other forms of spoken word content.



Project Description



Utilize Power Query Editor in Excel to clean and standardize an Audible dataset. Tasks include formatting columns, ensuring data consistency, and preparing the dataset for analysis.

1. Standardize the name column to ensure consistent title casing.
2. Separate combined first and last names in the author column if they are currently combined.
3. Ensure all entries in the releasedate column follow a consistent date format (DD-MM-YYYY).
4. Convert the time column from text format to a duration format that Excel recognizes.
5. Ensure the price column is in a numeric format, and identify any non-numeric values.
6. Convert text ratings in the stars column to numeric values.
7. Split the narratedby column into multiple columns if multiple narrators are listed.
8. Merge the releasedate and language columns into a single new column named releaseinfo with the format "DD-MM-YYYY, Language."
9. Ensure all currency values in the price column are formatted consistently with two decimal places.

APPLIED STEPS

01

APPLIED STEPS	
Source	⚙
Promoted Headers	⚙
Replaced Value1	⚙
Changed Type	
Capitalized Each Word	
Extracted Text After Delimi...	⚙
Trimmed Text	
Split Column by Delimiter	⚙
Changed Type1	
Removed Columns	
Split Column by Character ...	
Trimmed Text1	
Merged Columns	⚙
Split Column by Character ...	
Trimmed Text2	
Merged Columns1	⚙
Split Column by Character ...	
Trimmed Text3	
Merged Columns2	⚙
Trimmed Text4	
Merged Columns3	⚙
Extracted Text After Delimi...	⚙
Split Column by Delimiter1	⚙
Changed Type2	
Removed Columns1	

02

- Split Column by Character ...
- Trimmed Text5
- Merged Columns4
- Trimmed Text6
- Trimmed Text7
- Merged Columns5
- Split Column by Character ...
- Trimmed Text8
- Merged Columns6
- Renamed Columns
- Split Column by Character ...
- Trimmed Text9
- Merged Columns7
- Added Custom
- Reordered Columns
- Added Custom1
- Reordered Columns1
- Multiplied Column
- Inserted Addition
- Reordered Columns2
- Renamed Columns1
- Removed Columns2
- Split Column by Delimiter2
- Changed Type3
- Replaced Value
- Renamed Columns2

03

Split Column by Delimiter	
Changed Type3	
Replaced Value	
Renamed Columns2	
Removed Columns3	
Split Column by Character .	
Removed Columns4	
Changed Type4	
Renamed Columns3	
Capitalized Each Word1	
◀ Inserted Merged Column	

Uncleaned vs. Cleaned Data



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	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	Writtenby:RickRiordan	Narratedby:JesseBernstein	8 hrs and 48 mins	14-01-2010	English	4.5 out of 5 stars130 ratings	820

name	Author1	Author2	Author3	Narrator 1	Narrator 2	time	time std in min	releasedate	language	rat	no o	price	rele
Geronimo Stilton #11 & #12	Geronimo Stilton			Bill Lobely		2 hrs and 20 mins	140	04-08-2008	English	5	34	468	04-
The Burning Maze	Rick Riordan			Robbie Daymond		13 hrs and 8 mins	788	01-05-2018	English	4.5	41	820	01-
The Deep End	Jeff Kinney			Dan Russell		2 hrs and 3 mins	123	06-11-2020	English	4.5	38	410	06-
Daughter Of The Deep	Rick Riordan			Soneela Nankani		11 hrs and 16 mins	676	05-10-2021	English	4.5	12	615	05-
The Lightning Thief: Percy Jackson, Book 1	Rick Riordan			Jesse Bernstein		10 hrs	600	13-01-2010	English	4.5	181	820	13-
The Hunger Games: Special Edition	Suzanne Collins			Tatiana Maslany		10 hrs and 35 mins	635	30-10-2018	English	5	72	656	30-
Quest For The Diamond Sword	Winter Morgan			Luke Daniels		2 hrs and 23 mins	143	25-11-2014	English	5	11	233	25-
The Dark Prophecy	Rick Riordan			Robbie Daymond		12 hrs and 32 mins	752	02-05-2017	English	5	50	820	02-
Merlin Mission Collection	Mary Pope Osborne			Mary Pope Osborne		10 hrs and 56 mins	656	02-05-2017	English	5	5	1256	02-
The Tyrant'S Tomb	Rick Riordan			Robbie Daymond		13 hrs and 22 mins	802	24-09-2019	English	5	58	820	24-
The Titan'S Curse: Percy Jackson, Book 2	Rick Riordan			Jesse Bernstein		8 hrs and 48 mins	528	14-01-2010	English	4.5	130	820	14-
Magic Tree House Collection: Books 9-Mary Pope Osborne				Mary Pope Osborne		5 hrs and 23 mins	323	24-08-2011	English	5	6	1206	24-
Magic Tree House Collection: Books 1-Mary Pope Osborne				Mary Pope Osborne		6 hrs and 1 min	361	27-09-2011	English	5	7	1206	27-

Transformation using custom column

standardized the time format in minute with displayed custom formula

Add a column that is computed from the other columns.

New column name
mins

Custom column formula ⓘ

```
= if Text.Contains([time],"minute") then 1
else 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
```

Available columns

name
Author1.1
Author2
Author3
Narrator 1
Narrator 2
Narrator3

Add a column that is computed from the other columns.

New column name
hrs

Custom column formula ⓘ

```
= if Text.Contains([time],"hr") then
Number.FromText(Text.BeforeDelimiter([time],"hr"))
else 0
```

Available columns

name
Author1.1
Author2
Author3
Narrator 1
Narrator 2
Narrator3

<< Insert

[Learn about Power Query formulas](#)



Cleaned audible data set link

<https://docs.google.com/spreadsheets/d/1g5rCq8iEpWrea5dK7emp6AtTmAGDip-a/edit?usp=sharing&ouid=105516287096100407491&rtpof=true&sd=true>

Thank you
very much!

<https://www.linkedin.com/in/pradeep-maurya-307bb3286>