



DATA CLEANING PROJECT



Project Introduction

Audible, a leading platform for audiobooks and spoken-word content, aims to enhance its user experience and optimize its offerings through data-driven decision-making. This project focuses on cleaning and standardizing Audible's dataset using Excel Power Query Editor to ensure data consistency and accuracy. The goal is to prepare the dataset for insightful analysis, helping Audible make informed decisions related to pricing, content availability, and customer satisfaction.



PROBLEMS



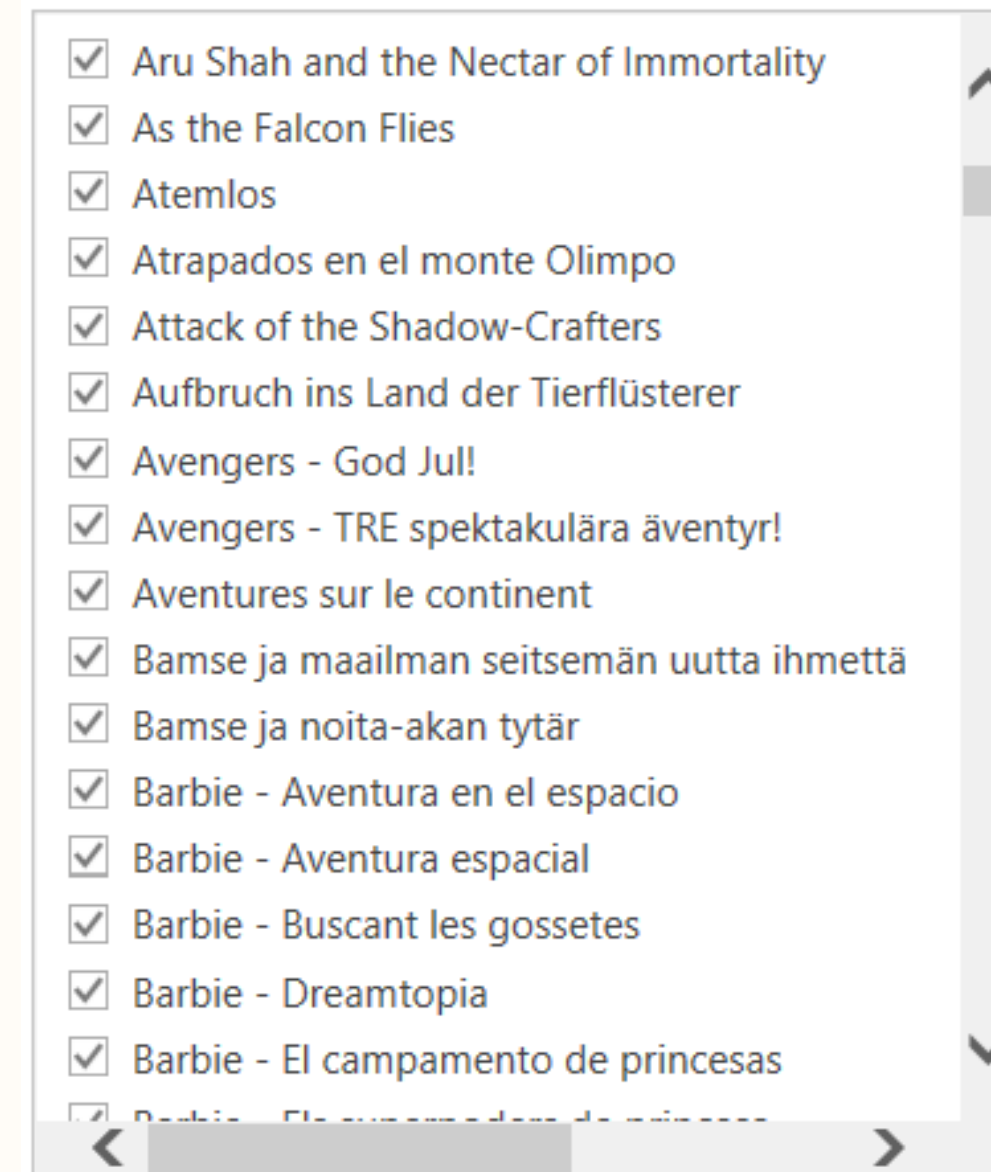
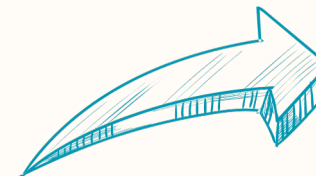
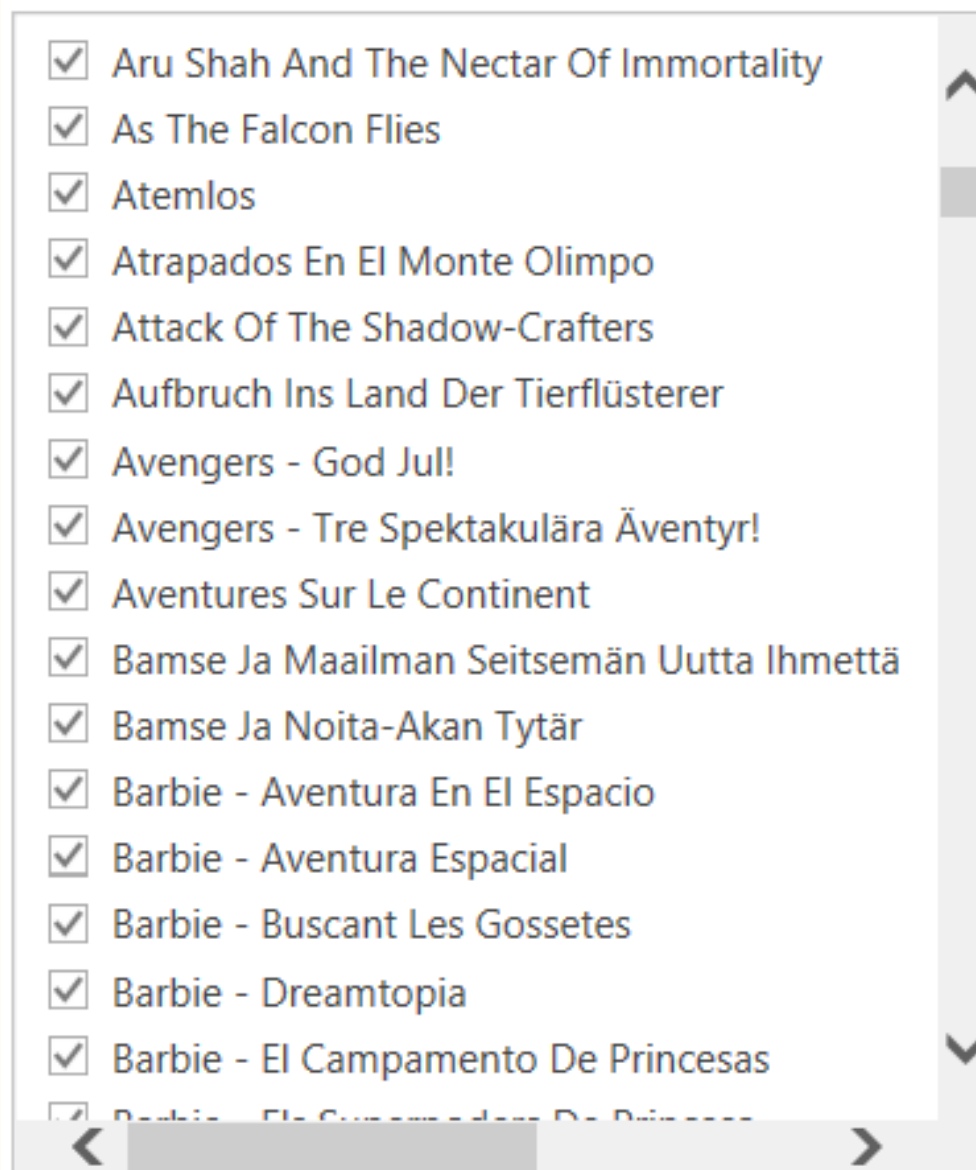
- 1 Standardize the name column to ensure consistent title casing.
- 2 Separate combined first and last names in the author column if they are currently combined.
- 3 Ensure all entries in the releasedate column follow a consistent date format (DD-MM-YYYY).
- 4 Convert the time column from text format 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.



1

STANDARDIZE THE NAME COLUMN TO ENSURE CONSISTENT TITLE CASING.

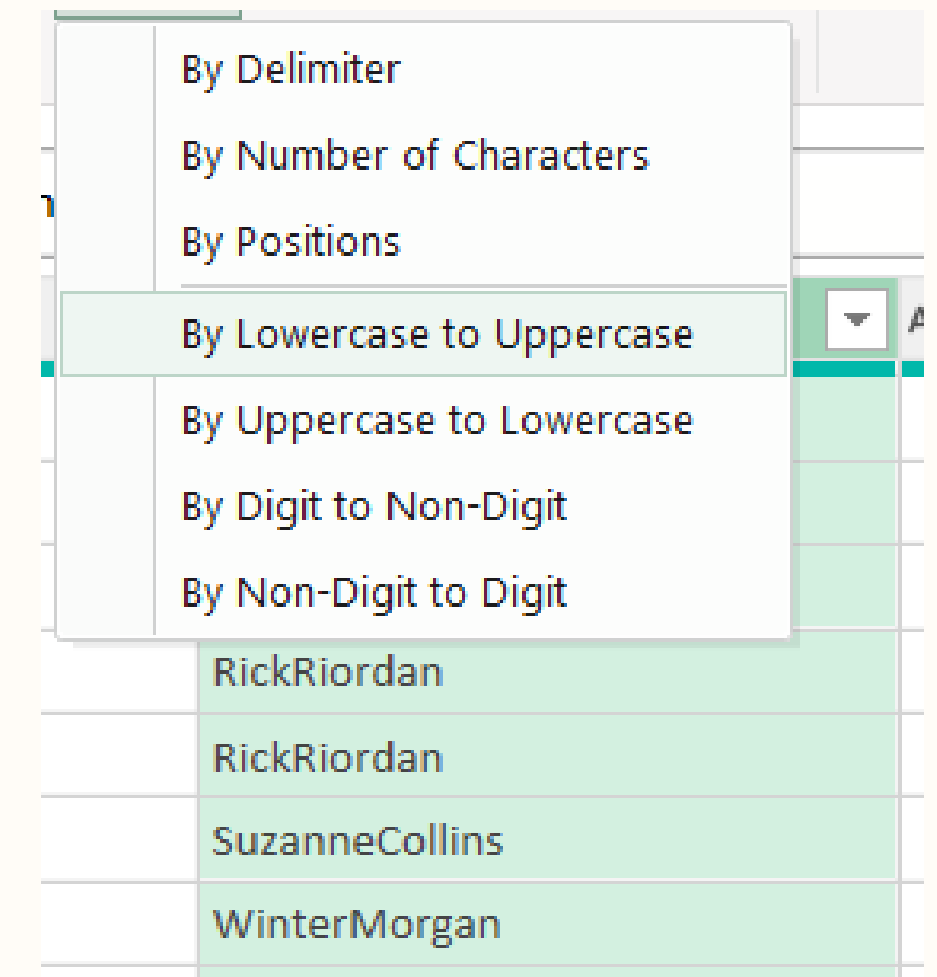
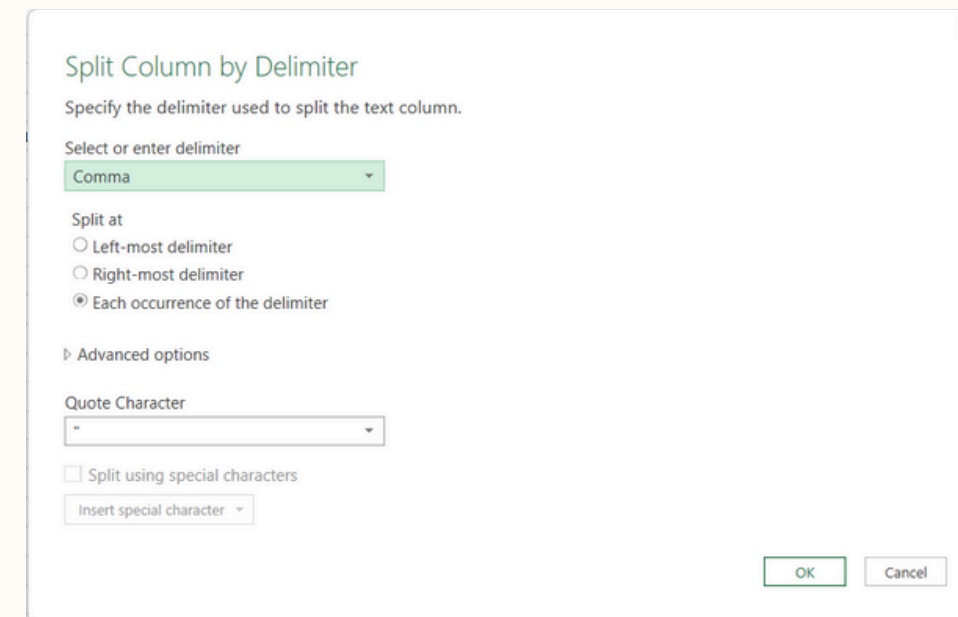
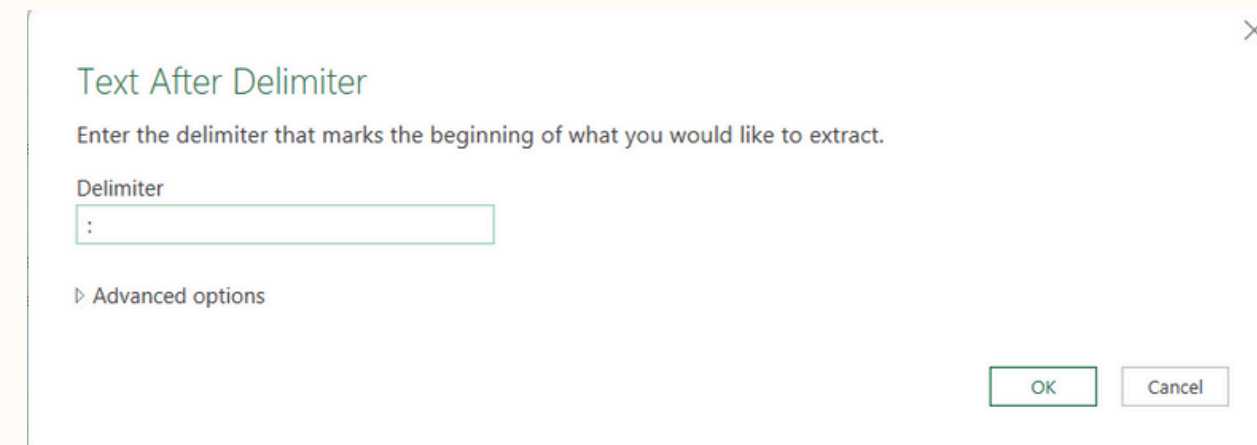
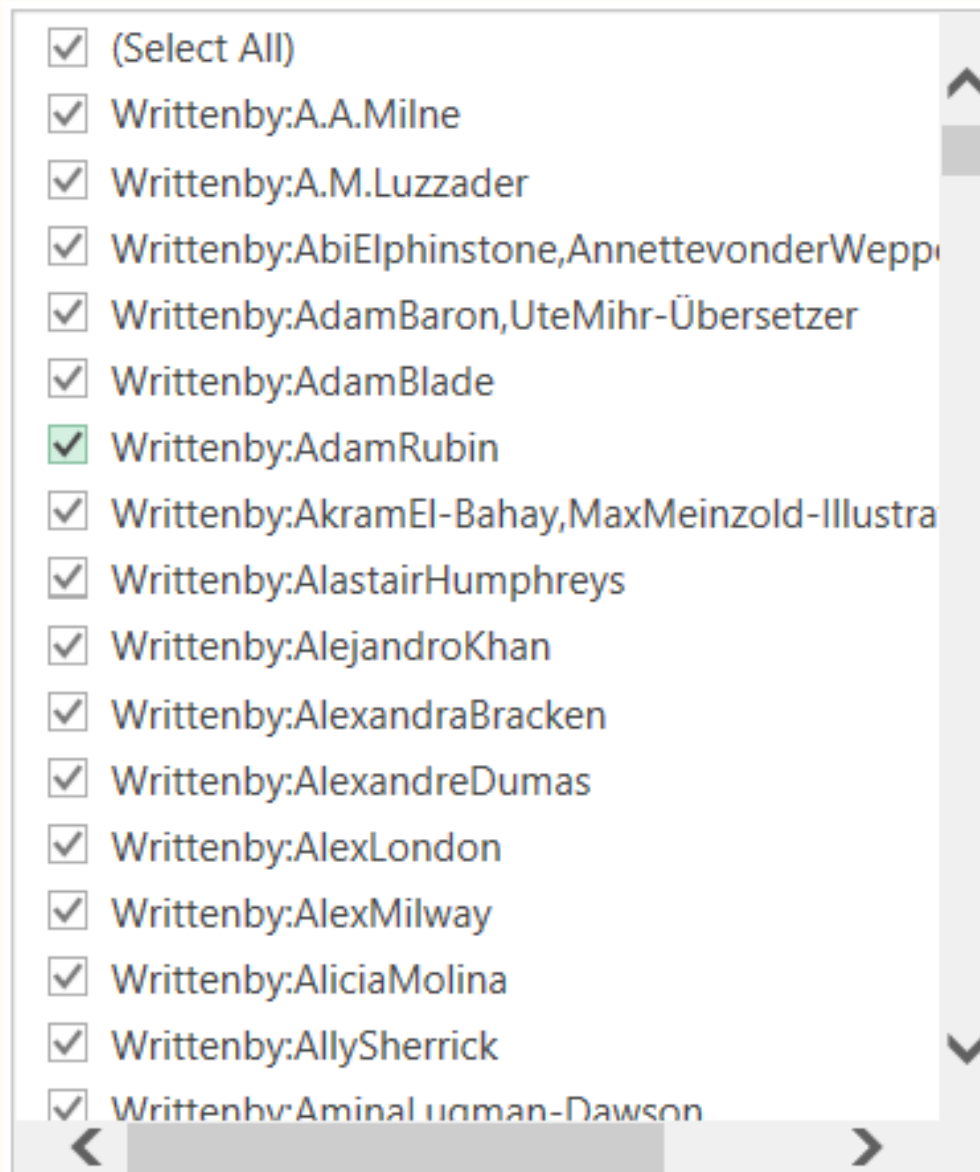
To standardize the name column we use:
Transform>>>Format>>>Capitalize Each Word



2

SEPARATE COMBINED FIRST AND LAST NAMES IN THE AUTHOR COLUMN IF THEY ARE CURRENTLY COMBINED.

We have three to four names in author column so we need to separate them all. First we separate the “Written by” from the column by using : as Delimiter.



Then we separate all the author names one by one. For that we use to split the column of all the author names by comma and lowercase to uppercase.

2

SEPARATE COMBINED FIRST AND LAST NAMES IN THE AUTHOR COLUMN IF THEY ARE CURRENTLY COMBINED.

To get the first name we merge the two columns with space.
And we need to follow the same steps for the remaining names to get the final result.

Merge Columns

Choose how to merge the selected columns.

Separator

Space

New column name (optional)

Author1

AB _C Author1	AB _C Author2	AB _C Author3
Bertrand Fichou	Nora Thullin	Catherinede Lasa
M.G.Leonard	Sam Sedgman	Elisa Paganelli-illustrator
Andrew Peterson-editor	Jonathan Rogers	N.D.Wilson
Mac Kenzie Cadenhead	Sean Ryan	Anke Albrecht-Übersetzer
Rohini Nilekani	Reshma Thapa-Gurung	Lavina Mahbubani
АлександрЦыпкин	АллаЛихачёва	АлександрЕгоров
Marvel	Hanna Graf	Ida Johansson
Jaume Cabré	Queralt Armengol-ilustrador	Concha Cardeñoso-traductor
Jaume Cabré	Romina Martí-ilustrador	Concha Cardeñoso-traductor
Mac Kenzie Cadenhead	Sean Ryan	Anke Albrecht-Übersetzer
Olivier Cohen	Yassen Vodenitcharov	Jean-Loup Pagesy

3

ENSURE ALL ENTRIES IN THE RELEASEDATE COLUMN FOLLOW A CONSISTENT DATE FORMAT (DD-MM-YYYY).

To standardize the entries in releasedate column we need to select the column>>>right click on it>>>change type>>>using locale>>>select the proper format

AB _C releasedate
04-08-08
01-05-18
06-11-20
05-10-21
13-01-10
30-10-18
25-11-14
02-05-17
02-05-17
24-09-19

Change Type with Locale

Change the data type and select the locale of origin.

Data Type

Locale


i Sample input values:
29/03/2016
29 March 2016
29 March
March 2016

releasedate
04/08/2008
01/05/2018
06/11/2020
05/10/2021
13/01/2010
30/10/2018
25/11/2014
02/05/2017
02/05/2017
24/09/2019

CONVERT THE TIME COLUMN FROM TEXT FORMAT TO A DURATION FORMAT THAT EXCEL RECOGNIZES.

From the time column, separate the hours and minutes using the custom column if-else formula to get the final result.

☒ (Select All)
☒ 1 hr
☒ 1 hr and 1 min
☒ 1 hr and 10 mins
☒ 1 hr and 11 mins
☒ 1 hr and 12 mins
☒ 1 hr and 13 mins
☒ 1 hr and 14 mins
☒ 1 hr and 15 mins
☒ 1 hr and 16 mins
☒ 1 hr and 17 mins
☒ 1 hr and 18 mins
☒ 1 hr and 19 mins
☒ 1 hr and 2 mins
☒ 1 hr and 20 mins
☒ 1 hr and 21 mins
☒ 1 hr and 22 mins
☐ 1 hr and 23 mins

 List may be incomplete. [Load more](#)

A ^B C time	1.2 time (standardized in mins)
2 hrs and 20 mins	140
13 hrs and 8 mins	788
2 hrs and 3 mins	123
11 hrs and 16 mins	676
10 hrs	600
10 hrs and 35 mins	635
2 hrs and 23 mins	143
12 hrs and 32 mins	752
10 hrs and 56 mins	656
13 hrs and 22 mins	802

4

TRANSFORMATION USING CUSTOM COLUMN

First, extract the minutes from the time column, then extract the hours, convert them into minutes, and summed both values.

Custom Column

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
Author2
Author3
narrator
time
releasedate

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

Custom Column

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
Author2
Author3
narrator
time
mins

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

5

ENSURE THE PRICE COLUMN IS IN A NUMERIC FORMAT, AND IDENTIFY ANY NON-NUMERIC VALUES.

First we need to convert non numeric that is “Free” value to 0. As now the column does not have any null or numeric values in it. Just simply change the type of column into currency to get the final result.

Replace Values

Replace one value with another in the selected columns.

Value To Find

Free

Replace With

0

▸ Advanced options

\$ price
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

☒ (Select All)

☒ 0

☒ 32

☒ 36

☒ 38

☒ 47

☒ 65

☒ 75

☒ 76

☒ 82

☒ 93

☒ 99

☒ 100


☒ 105

☒ 112

☒ 113

☒ 115

☒ 116

 List may be incomplete.

[Load more](#)

CONVERT TEXT RATINGS IN THE STARS COLUMN TO NUMERIC VALUES.

Split the column using 'space' as left most delimiter then replace all the null values from the new column to get the desired column.

A ^B _C stars
5 out of 5 stars34 ratings
4.5 out of 5 stars41 ratings
4.5 out of 5 stars38 ratings
4.5 out of 5 stars12 ratings
4.5 out of 5 stars181 ratings
5 out of 5 stars72 ratings
5 out of 5 stars11 ratings
5 out of 5 stars50 ratings
5 out of 5 stars5 ratings
5 out of 5 stars58 ratings

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Space

Split at

☒ Left-most delimiter

☐ Right-most delimiter

☐ Each occurrence of the delimiter

Advanced options

Quote Character

"

☐ Split using special characters

Insert special character

Replace Values

Replace one value with another in the selected columns.

Value To Find

Not

Replace With

Null

Advanced options

☒ (Select All)

☒ 3

☒ 3.5

☒ 4

☒ 4.5

☒ 5

☒ Not

! List may be incomplete.

7

SPLIT THE NARRATEDBY COLUMN INTO MULTIPLE COLUMNS IF MULTIPLE NARRATORS ARE LISTED.

Split the column by colon delimiter first then split it by comma delimiter with each occurrence and after that we have narrator's name in different columns.

☒ (Select All)

☒ Narratedby:A.J.Beckles

☒ Narratedby:AbbeHolmes

☒ Narratedby:AdamRubin

☒ Narratedby:AdamVerner

☒ Narratedby:AdnanKapadia

☒ Narratedby:AdolfoDuncan,AldoLumbía,MiguelC

☒ Narratedby:AdrianKowarzyk

☒ Narratedby:AimeeLilly

☒ Narratedby:AlanBennett

☒ Narratedby:AlanBochm

☒ Narratedby:AlastairHumphreys

☒ Narratedby:AlbertoFeixas

☒ Narratedby:AlexandraMori

☒ Narratedby:AlexBolte

☒ Narratedby:AlexJennings

☒ Narratedby:AlexWingfield

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Colon

Split at

- ☐ Left-most delimiter
- ☐ Right-most delimiter
- ☒ Each occurrence of the delimiter

Advanced options

Quote Character

"

☐ Split using special characters

Insert special character

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

"

☐ Split using special characters

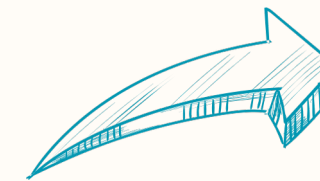
Insert special character

A ^B C narrator1	A ^B C narrator2	A ^B C narrator3
Philippullman	Fullcast	Ruthwilson
Monicarachelle	Jimd.Johnston	Katieotten
Lenforgione	Ashtonsundholm	Jadenrogers
Guychappelier	Frédéricsanchez	Mariezidi
Lenforgione	Winstonbromhead	I'Kemitchell
Jordanfisher	Daschapolanco	Jonieceabbott-Pratt
Davidwalliams	Peterserafinowicz	Morwennabanks
Katieleung	RebeccaLee	Alexandercapon
Jamesmurray	Kirbyheyborne	Inésdelcastillo
Chrissmith	Kariseyansen	Roymcmillan

MERGE THE RELEASEDATE AND LANGUAGE COLUMNS INTO A SINGLE NEW COLUMN NAMED RELEASEINFO WITH THE FORMAT "DD-MM- YYYY, LANGUAGE."

In the dataset, we have two different columns one with date and second with language we just simply need to merge these two columns using comma into a new column named as releaseinfo.

releasedate	language
04/08/2008	English
01/05/2018	English
06/11/2020	English
05/10/2021	English
13/01/2010	English
30/10/2018	English
25/11/2014	English
02/05/2017	English
02/05/2017	English
24/09/2019	English



releaseinfo
04/08/2008, English
01/05/2018, English
06/11/2020, English
05/10/2021, English
13/01/2010, English
30/10/2018, English
25/11/2014, English
02/05/2017, English
02/05/2017, English
24/09/2019, English

ENSURE ALL CURRENCY VALUES IN THE PRICE COLUMN ARE FORMATTED CONSISTENTLY WITH TWO DECIMAL PLACES.

In the dataset, we already have price column formatted with two decimal values. If it was not like that then go to
Transform>>>Rounding>>>Round>>>Select Decimal Places

\$ price
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

CLEANED VS. UNCLEANED DATA

Power Query Editor was used to clean and standardize the raw data for analysis.

name	author	narrator	time	releasedate	language	stars	price
Geronimo Stilton #	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 De	Writtenby:RickRiordan	Narratedby:SoneelaNankani	11 hrs and 16 mins	05/10/2021	English	4.5 out of 5 stars12 ratings	615
The Lightning Thief	Writtenby:RickRiordan	Narratedby:JesseBernstein	10 hrs	13/01/2010	English	4.5 out of 5 stars181 ratings	820
The Hunger Games	Writtenby:SuzanneCollins	Narratedby:TatianaMaslany	10 hrs and 35 mins	30/10/2018	English	5 out of 5 stars72 ratings	656
Quest for the Dian	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 Col	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 To	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 22 mins	24/09/2019	English	5 out of 5 stars58 ratings	820
The Titan's Curse:	Writtenby:RickRiordan	Narratedby:JesseBernstein	8 hrs and 48 mins	14/01/2010	English	4.5 out of 5 stars130 ratings	820
Magic Tree House	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	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	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,fullca	11 hrs and 55 mins	24/06/2021	English	4 out of 5 stars2 ratings	1,093.00
Geronimo Stilton #	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 25 mins	08/02/2008	English	4.5 out of 5 stars33 ratings	467
Magic Tree House	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 Col	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	10 hrs and 18 mins	02/05/2017	English	5 out of 5 stars11 ratings	1,256.00

name	Author1	Author2	Author3	narrator.2.1	narrator.2	narrator.2	narrator.1	time (standa	releaseinfo	stars	no of p	price
Geronimo Stilton #11 & #12	Geronimo Stilton			Billlobely				140	04/08/2008, English	5	34	468
The Burning Maze	Rick Riordan			Robbiedaymond				788	01/05/2018, English	4.5	41	820
The Deep End	Jeff Kinney			Danrussell				123	06/11/2020, English	4.5	38	410
Daughter Of The Deep	Rick Riordan			Soneelanankani				676	05/10/2021, English	4.5	12	615
The Lightning Thief: Percy Jackson, Book 1	Rick Riordan			Jessebernstein				600	13/01/2010, English	4.5	181	820
The Hunger Games: Special Edition	Suzanne Collins			Tatianamaslany				635	30/10/2018, English	5	72	656
Quest For The Diamond Sword	Winter Morgan			Lukedaniels				143	25/11/2014, English	5	11	233
The Dark Prophecy	Rick Riordan			Robbiedaymond				752	02/05/2017, English	5	50	820
Merlin Mission Collection	Mary Pope Osborne			Marypopeosborne				656	02/05/2017, English	5	5	1256
The Tyrant'S Tomb	Rick Riordan			Robbiedaymond				802	24/09/2019, English	5	58	820
The Titan'S Curse: Percy Jackson, Book 3	Rick Riordan			Jessebernstein				528	14/01/2010, English	4.5	130	820
Magic Tree House Collection: Books 9-16	Mary Pope Osborne			Marypopeosborne				323	24/08/2011, English	5	6	1206
Magic Tree House Collection: Books 1-8	Mary Pope Osborne			Marypopeosborne				361	27/09/2011, English	5	7	1206
Magnus Chase And The Ship Of The Dead	Rick Riordan			Michaelcrouch				778	03/10/2017, English	5	41	820
Northern Lights	Philip Pullman			Philippullman	Fullcast	Ruthwilson		715	24/06/2021, English	4	2	1093
Geronimo Stilton #13 And #14	Geronimo Stilton			Billlobely				145	08/02/2008, English	4.5	33	467
Magic Tree House Collection	Mary Pope Osborne			Marypopeosborne				304	26/12/2004, English	5	7	1206
Exile	Shannon Messenger			Caitlinkelly				881	06/11/2018, English	5	20	836
Merlin Mission Collection	Mary Pope Osborne			Marypopeosborne				618	02/05/2017, English	5	11	1256

DATA TRANSFORMATION STEPS

Sequential actions performed in Power Query Editor for data preparation.

APPLIED STEPS

Source	⚙
Promoted Headers	⚙
Replaced Value1	⚙
Filtered Rows7	
Changed Type	
Capitalized Each Word	
Extracted Text After Delimi...	⚙
Split Column by Delimiter	⚙
Changed Type1	
Filtered Rows	
Removed Columns	
Split Column by Character ...	
Merged Columns	⚙
Split Column by Character ...	
Merged Columns1	⚙
✕ Filtered Rows1	
Split Column by Character ...	
Merged Columns2	⚙
Changed Type with Locale	⚙
Filtered Rows2	
Added Custom	⚙
Reordered Columns	
Added Custom1	⚙
Filtered Rows3	

APPLIED STEPS

Reordered Columns1	
Inserted Multiplication	⚙
Reordered Columns2	
Inserted Addition	⚙
Reordered Columns3	
Renamed Columns	
Removed Columns1	
Split Column by Delimiter1	⚙
Changed Type2	
Filtered Rows4	
Replaced Value	⚙
Filtered Rows5	
Renamed Columns1	
Split Column by Character ...	
Removed Columns2	
Extracted Text Before Deli...	⚙
Changed Type3	
Renamed Columns2	
Filtered Rows6	
Changed Type4	
Filtered Rows8	
Capitalized Each Word1	
Merged Columns3	⚙

APPLIED STEPS

Replaced Value	⚙
Filtered Rows5	
Renamed Columns1	
Split Column by Character ...	
Removed Columns2	
Extracted Text Before Deli...	⚙
Changed Type3	
Renamed Columns2	
Filtered Rows6	
Changed Type4	
Filtered Rows8	
Capitalized Each Word1	
Merged Columns3	⚙
Capitalized Each Word2	
Split Column by Delimiter2	⚙
Changed Type5	
Filtered Rows9	
Removed Columns3	
Filtered Rows10	
Split Column by Delimiter3	⚙
Changed Type6	
Filtered Rows11	
Capitalized Each Word3	
Filtered Rows12	⚙
✕ Renamed Columns3	



THANK YOU

 RAHUL NAGRA