

# DATA CLEANING PROJECT

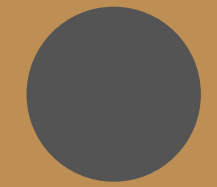
WITH POWER QUERY EDITOR

Presented by : Kinjan chauhan



# Overview

- Introduction about company
- Project description
- Data cleaning steps
- Link of clean data set



# 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.

Audible operates the Audiobook Creation Exchange, which enables individual authors or publishers to work with professional actors and producers to create audiobooks, which are then distributed to Amazon and iTunes. Currently, the service is available to residents of the United States, United Kingdom, Canada, and Ireland.



# 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.

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



# Data cleaning steps

Step 1: Open Excel worksheet ----> Data tab ----> Get data -----> From file ----> From Text/CSV

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H
	name	author	narrator	time	releasedate	language	stars	price
1	Geronimo Stilton #11 & #12	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 20 mins	04-08-2008	English	5 out of 5 stars34 ratings	468
2	The Burning Maze	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 8 mins	01-05-2018	English	4.5 out of 5 stars41 ratings	820
3	The Deep End	Writtenby:JeffKinney	Narratedby:DanRussell	2 hrs and 3 mins	06-11-2020	English	4.5 out of 5 stars38 ratings	410
4	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
5	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
6	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
7	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
8	The Dark Prophecy	Writtenby:RickRiordan	Narratedby:RobbieDaymond	12 hrs and 32 mins	02-05-2017	English	5 out of 5 stars50 ratings	820
9	Merlin Mission Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	10 hrs and 56 mins	02-05-2017	English	5 out of 5 stars5 ratings	1256
10	The Tyrant's Tomb	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 22 mins	24-09-2019	English	5 out of 5 stars58 ratings	820
11	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
12	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	1206
13	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	1206
14	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
15	Northern Lights	Writtenby:PhilipPullman	Narratedby:PhilipPullman,fullcast,RuthWilso	11 hrs and 55 mins	24-06-2021	English	4 out of 5 stars2 ratings	1093
16	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
17	Magic Tree House Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	5 hrs and 4 mins	26-12-2004	English	5 out of 5 stars7 ratings	1206
18	Exile	Writtenby:ShannonMessenger	Narratedby:CaitlinKelly	14 hrs and 41 mins	06-11-2018	English	5 out of 5 stars20 ratings	836
19	Merlin Mission Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	10 hrs and 18 mins	02-05-2017	English	5 out of 5 stars11 ratings	1256
20	Neverseen	Writtenby:ShannonMessenger	Narratedby:CaitlinKelly	16 hrs and 42 mins	06-11-2018	English	5 out of 5 stars13 ratings	1003
21	The Tower of Nero	Writtenby:RickRiordan	Narratedby:RobbieDaymond	12 hrs and 12 mins	06-10-2020	English	5 out of 5 stars79 ratings	615
22	Eldest: The Inheritance Cycle, Book 2	Writtenby:ChristopherPaolini	Narratedby:GerrardDoyle	23 hrs and 29 mins	14-04-2009	English	4.5 out of 5 stars47 ratings	957
23	Artemis Fowl	Writtenby:EoinColfer	Narratedby:GerryO'Brien	6 hrs and 46 mins	11-04-2013	English	5 out of 5 stars27 ratings	683
24	Geronimo Stilton 22 & 24	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 37 mins	01-02-2011	English	4.5 out of 5 stars25 ratings	469
25	Percy Jackson and the Battle of the Labyrinth	Writtenby:RickRiordan	Narratedby:JesseBernstein	10 hrs and 32 mins	21-01-2010	English	5 out of 5 stars101 ratings	820
26	Winnie-the-Pooh	Writtenby:A.A.Milne	Narratedby:PeterDennis	2 hrs and 46 mins	28-12-2003	English	5 out of 5 stars15 ratings	468
27	Magic Tree House Collection: Books 25-32	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	7 hrs and 9 mins	01-09-2020	English	5 out of 5 stars3 ratings	1256
28	Geronimo Stilton #20 and #21	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 32 mins	01-10-2010	English	5 out of 5 stars18 ratings	469
29	The Red Pyramid	Writtenby:RickRiordan	Narratedby:JaneCollingwood,JosephMay	13 hrs and 44 mins	24-01-2013	English	4.5 out of 5 stars37 ratings	820
30	Mystery of the Griever's Mark	Writtenby:WinterMorgan	Narratedby:LukeDaniels	1 hr and 53 mins	25-11-2014	English	5 out of 5 stars9 ratings	233
31	Geronimo Stilton #17	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 28 mins	13-07-2009	English	5 out of 5 stars8 ratings	468
32	The House at Pooh Corner	Writtenby:A.A.Milne	Narratedby:PeterDennis	3 hrs and 12 mins	05-04-2005	English	5 out of 5 stars5 ratings	374

audible\_uncleaned (1) Sheet1

Uncleaned data with errors

Step 2 : The steps applied to clean and standradise the data set are below.

This data cleaning task includes : Capitalized each words on columns , Extract & splitting columns by delimiters, Merging & trimming columns , Replacing the values with another value in columns, Extract time by custom column , Multiplication & addition on columns

PROPERTIES

Name

audible\_uncleaned

All Properties

APPLIED STEPS

Source

Promoted Headers

Replaced Value

Changed Type

Capitalized Each Word

Extracted Text After Delimiter

Trimmed Text

Split Column by Delimiter

Changed Type1

Filtered Rows

Split Column by Character Transition

Merged Columns

Split Column by Character Transition1

Merged Columns1

Removed Columns

Split Column by Character Transition2

Filtered Rows1

Merged Columns2

APPLIED STEPS

Filtered Rows1

Merged Columns2

Changed Type2

extract minutes by custom column

Filtered Rows2

Renamed Columns

extract hrs by custom column

Filtered Rows3

Multiplied Column

Inserted Addition

Renamed Columns1

Removed Columns1

Reordered Columns

Removed Columns2

Split Column by Delimiter1

Changed Type3

Replaced Value1

Changed Type4

Filtered Rows4

Inserted Text Between Delimiters

Reordered Columns1

Filtered Rows5

APPLIED STEPS

Replaced Value1

Changed Type4

Filtered Rows4

Inserted Text Between Delimiters

Reordered Columns1

Filtered Rows5

Removed Columns3

Changed Type5

narratoe column Extracted Text After Deli...

Renamed Columns2

Split Column by Delimiter2

Split Column by Character Transition3

Trimmed Text1

Merged Columns3

Split Column by Character Transition4

Trimmed Text2

Merged Columns4

Split Column by Character Transition5

Trimmed Text3

Merged Columns5

Removed Columns4

Merged Columns6

# Custom formulas

## Custom Column

Add a column that is computed from the other columns.

New column name

Custom

Custom column formula ⓘ

```
= if Text.Contains([time],"Less") 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
```

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

Available columns

name

Author name .1

Author name -2

Author name -3

narrator

time

releasedate

<< Insert

## Custom Column

Add a column that is computed from the other columns.

New column name

Custom

Custom column formula ⓘ

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

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

Available columns

name

Author name .1

Author name -2

Author name -3

narrator

time

releasedate

<< Insert

The formula above extracts the “minutes” value form the time column.

The formula above extracts the “hr” value form the time column.



# Cleaned dataset with no errors

	A	B	C	D	E	F	G	H	I	J	K	L
1	name	Author name .1	Author name - 2	Author name - 3	Narrator full name -1	Narrator full name -2	Narrator full name-3	Time( standradize in Min)	Release info	Stars (Out of 5)	Ratings( No. of people rated)	price
2	Geronimo Stilton #11 & #12	Geronimo Stilton			Bill Lobely			140	04-08-2008, English	5	34	468
3	The Burning Maze	Rick Riordan			Robbie Daymond			788	01-05-2018, English	4.5	41	820
4	The Deep End	Jeff Kinney			Dan Russell			123	06-11-2020, English	4.5	38	410
5	Daughter Of The Deep	Rick Riordan			Soneela Nankani			676	05-10-2021, English	4.5	12	615
6	The Lightning Thief: Percy Jackson, Book 1	Rick Riordan			Jesse Bernstein			600	13-01-2010, English	4.5	181	820
7	The Hunger Games: Special Edition	Suzanne Collins			Tatiana Maslany			635	30-10-2018, English	5	72	656
8	Quest For The Diamond Sword	Winter Morgan			Luke Daniels			143	25-11-2014, English	5	11	233
9	The Dark Prophecy	Rick Riordan			Robbie Daymond			752	02-05-2017, English	5	50	820
10	Merlin Mission Collection	Mary Pope Osborne			Mary Pope Osborne			656	02-05-2017, English	5	5	1256
11	The Tyrant'S Tomb	Rick Riordan			Robbie Daymond			802	24-09-2019, English	5	58	820
12	The Titan'S Curse: Percy Jackson, Book 3	Rick Riordan			Jesse Bernstein			528	14-01-2010, English	4.5	130	820
13	Magic Tree House Collection: Books 9-16	Mary Pope Osborne			Mary Pope Osborne			323	24-08-2011, English	5	6	1206
14	Magic Tree House Collection: Books 1-8	Mary Pope Osborne			Mary Pope Osborne			361	27-09-2011, English	5	7	1206
15	Magnus Chase And The Ship Of The Dead	Rick Riordan			Michael Crouch			778	03-10-2017, English	5	41	820
16	Northern Lights	Philip Pullman			Philip Pullman	fullcast	Ruth Wilson	715	24-06-2021, English	4	2	1093
17	Geronimo Stilton #13 And #14	Geronimo Stilton			Bill Lobley			145	08-02-2008, English	4.5	33	467
18	Magic Tree House Collection	Mary Pope Osborne			Mary Pope Osborne			304	26-12-2004, English	5	7	1206
19	Exile	Shannon Messenger			Caitlin Kelly			881	06-11-2018, English	5	20	836
20	Merlin Mission Collection	Mary Pope Osborne			Mary Pope Osborne			618	02-05-2017, English	5	11	1256
21	Neverseen	Shannon Messenger			Caitlin Kelly			1002	06-11-2018, English	5	13	1003
22	The Tower Of Nero	Rick Riordan			Robbie Daymond			732	06-10-2020, English	5	79	615
23	Eldest: The Inheritance Cycle, Book 2	Christopher Paolini			Gerrard Doyle			1409	14-04-2009, English	4.5	47	957
24	Artemis Fowl	Eoin Colfer			Gerry O'Brien			406	11-04-2013, English	5	27	683
25	Geronimo Stilton 22 & 24	Geronimo Stilton			Bill Lobley			157	01-02-2011, English	4.5	25	469
26	Percy Jackson And The Battle Of The Labyrint	Rick Riordan			Jesse Bernstein			632	21-01-2010, English	5	101	820
27	Winnie-The-Pooh	A.A.Milne			Peter Dennis			166	28-12-2003, English	5	15	468
28	Magic Tree House Collection: Books 25-32	Mary Pope Osborne			Mary Pope Osborne			429	01-09-2020, English	5	3	1256
29	Geronimo Stilton #20 And #21	Geronimo Stilton			Bill Lobley			152	01-10-2010, English	5	18	469
30	The Red Pyramid	Rick Riordan			Jane Collingwood	Joseph May		824	24-01-2013, English	4.5	37	820
31	Mystery Of The Griefer'S Mark	Winter Morgan			Luke Daniels			113	25-11-2014, English	5	9	233
32	Geronimo Stilton #17	Geronimo Stilton			Bill Lobley			148	13-07-2009, English	5	8	468

Developer   Help   Power Pivot   **Table Design**   Query

Wrap Text   Merge & Center   General   Conditional Formatting   Format as Table   Cell Styles   Insert   Delete   Format   AutoSum   Fill   Clear   Sort & Filter   Find & Select   Add-ins   Analyze Data

Number   Styles   Cells   Editing

Author name .1	Author name .2
Geronimo Stilton	
Rick Riordan	
Jeff Kinney	
Rick Riordan	
Rick Riordan	
Suzanne Collins	
Winter Morgan	
Rick Riordan	
Mary Pope Osborne	

**Queries & Connections**

Queries | Connections

1 query

audible\_uncleaned  
87,489 rows loaded.

## Link of cleaned dataset

<https://docs.google.com/spreadsheets/d/1jPKRNt2Qpa35sEW3ESJkSmukBZzGmckK/edit?usp=sharing&oid=115470209051493735344&rtpof=true&sd=true>

no errors





# Thank you

Presented by : Kinjan chauhan

