

# AUDIBLE DATA CLEANING PROJECT



Microsoft Excel using power query Editor

**Presented By : Divya Kaushik**

# Uncleaned Data

name	author	narrator	time	releasedate	language
Geronimo Stilton #11 & #12	Writtenby:GeronimoStilton	Narratedby:BillLobely	2 hrs and 20 mins	04-08-2008	English
The Burning Maze	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 8 mins	01-05-2018	English
The Deep End	Writtenby:JeffKinney	Narratedby:DanRussell	2 hrs and 3 mins	06-11-2020	English
Daughter of the Deep	Writtenby:RickRiordan	Narratedby:SoneelaNankani	11 hrs and 16 mins	05-10-2021	English
The Lightning Thief: Percy Jackson, Book 1	Writtenby:RickRiordan	Narratedby:JesseBernstein	10 hrs	13-01-2010	English
The Hunger Games: Special Edition	Writtenby:SuzanneCollins	Narratedby:TatianaMaslany	10 hrs and 35 mins	30-10-2018	English
Quest for the Diamond Sword	Writtenby:WinterMorgan	Narratedby:LukeDaniels	2 hrs and 23 mins	25-11-2014	English
The Dark Prophecy	Writtenby:RickRiordan	Narratedby:RobbieDaymond	12 hrs and 32 mins	02-05-2017	English
Merlin Mission Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	10 hrs and 56 mins	02-05-2017	English
The Tyrant's Tomb	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 22 mins	24-09-2019	English
The Titan's Curse: Percy Jackson, Book 3	Writtenby:RickRiordan	Narratedby:JesseBernstein	8 hrs and 48 mins	14-01-2010	English
Magic Tree House Collection: Books 9-16	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	5 hrs and 23 mins	24-08-2011	English
Magic Tree House Collection: Books 1-8	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	6 hrs and 1 min	27-09-2011	English
Magnus Chase and the Ship of the Dead	Writtenby:RickRiordan	Narratedby:MichaelCrouch	12 hrs and 58 mins	03-10-2017	English
Northern Lights	Writtenby:PhilipPullman	Narratedby:PhilipPullman,fullcast,RuthWilson	11 hrs and 55 mins	24-06-2021	English
Geronimo Stilton #13 and #14	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 25 mins	08-02-2008	English
Magic Tree House Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	5 hrs and 4 mins	26-12-2004	English
Exile	Writtenby:ShannonMessenger	Narratedby:CaitlinKelly	14 hrs and 41 mins	06-11-2018	English
Merlin Mission Collection	Writtenby:MaryPopeOsborne	Narratedby:MaryPopeOsborne	10 hrs and 18 mins	02-05-2017	English
Neverseen	Writtenby:ShannonMessenger	Narratedby:CaitlinKelly	16 hrs and 42 mins	06-11-2018	English
The Tower of Nero	Writtenby:RickRiordan	Narratedby:RobbieDaymond	12 hrs and 12 mins	06-10-2020	English
Eldest: The Inheritance Cycle, Book 2	Writtenby:ChristopherPaolini	Narratedby:GerrardDoyle	23 hrs and 29 mins	14-04-2009	English
Artemis Fowl	Writtenby:EoinColfer	Narratedby:GerryO'Brien	6 hrs and 46 mins	11-04-2013	English
Geronimo Stilton 22 & 24	Writtenby:GeronimoStilton	Narratedby:BillLobley	2 hrs and 37 mins	01-02-2011	English
Percy Jackson and the Battle of the Labyrinth	Writtenby:RickRiordan	Narratedby:JesseBernstein	10 hrs and 32 mins	21-01-2010	English
Winnie-the-Pooh	Writtenby:A.A.Milne	Narratedby:PeterDennis	2 hrs and 46 mins	28-12-2003	English

# Uncleaned Data

time	releasedate	language	stars	price
2 hrs and 20 mins	04-08-2008	English	5 out of 5 stars34 ratings	468
13 hrs and 8 mins	01-05-2018	English	4.5 out of 5 stars41 ratings	820
2 hrs and 3 mins	06-11-2020	English	4.5 out of 5 stars38 ratings	410
11 hrs and 16 mins	05-10-2021	English	4.5 out of 5 stars12 ratings	615
10 hrs	13-01-2010	English	4.5 out of 5 stars181 ratings	820
10 hrs and 35 mins	30-10-2018	English	5 out of 5 stars72 ratings	656
2 hrs and 23 mins	25-11-2014	English	5 out of 5 stars11 ratings	233
12 hrs and 32 mins	02-05-2017	English	5 out of 5 stars50 ratings	820
10 hrs and 56 mins	02-05-2017	English	5 out of 5 stars5 ratings	1256
13 hrs and 22 mins	24-09-2019	English	5 out of 5 stars58 ratings	820
8 hrs and 48 mins	14-01-2010	English	4.5 out of 5 stars130 ratings	820
5 hrs and 23 mins	24-08-2011	English	5 out of 5 stars6 ratings	1206
6 hrs and 1 min	27-09-2011	English	5 out of 5 stars7 ratings	1206
12 hrs and 58 mins	03-10-2017	English	5 out of 5 stars41 ratings	820
11 hrs and 55 mins	24-06-2021	English	4 out of 5 stars2 ratings	1093
2 hrs and 25 mins	08-02-2008	English	4.5 out of 5 stars33 ratings	467
5 hrs and 4 mins	26-12-2004	English	5 out of 5 stars7 ratings	1206
14 hrs and 41 mins	06-11-2018	English	5 out of 5 stars20 ratings	836
10 hrs and 18 mins	02-05-2017	English	5 out of 5 stars11 ratings	1256
16 hrs and 42 mins	06-11-2018	English	5 out of 5 stars13 ratings	1003
12 hrs and 12 mins	06-10-2020	English	5 out of 5 stars79 ratings	615
23 hrs and 29 mins	14-04-2009	English	4.5 out of 5 stars47 ratings	957
6 hrs and 46 mins	11-04-2013	English	5 out of 5 stars27 ratings	683
2 hrs and 37 mins	01-02-2011	English	4.5 out of 5 stars25 ratings	469
10 hrs and 32 mins	21-01-2010	English	5 out of 5 stars101 ratings	820
2 hrs and 46 mins	28-12-2003	English	5 out of 5 stars15 ratings	468

[Dataset Link](#)

**Total Count:87490**

# Tasks

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.



# Cleaned Data

A	B	C	D	E	F	G	H	I	J
name	Author 1	Author 2	Author 3	Narrator 1	Narrator 2	Narrator 3	time	releasedate	language
Atomic Habits	James Clear	Not Applicable	Not Applicable	James Clear	Not Applicable	Not Applicable	5 hrs and 35 mins	18-10-2018	English
Atomic Habits	James Clear	Not Applicable	Not Applicable	James Clear	Not Applicable	Not Applicable	5 hrs and 35 mins	18-10-2018	English
Ikigai	Héctor García	Francesc Miralles	Not Applicable	Naoko Mori	Not Applicable	Not Applicable	3 hrs and 23 mins	18-09-2017	English
The Psychology Of Money	Morgan Housel	Not Applicable	Not Applicable	Chris Hill	Not Applicable	Not Applicable	5 hrs and 48 mins	08-09-2020	English
The Psychology Of Money	Morgan Housel	Not Applicable	Not Applicable	Chris Hill	Not Applicable	Not Applicable	5 hrs and 48 mins	08-09-2020	English
Sapiens	Yuval Noah Harari	Not Applicable	Not Applicable	Derek Perkins	Not Applicable	Not Applicable	15 hrs and 18 mins	30-04-2015	English
Rich Dad Poor Dad	Robert T.Kiyosaki	Not Applicable	Not Applicable	Tim Wheeler	Not Applicable	Not Applicable	6 hrs and 9 mins	15-06-2012	English
Life'S Amazing Secrets	Gaur Gopal Das	Not Applicable	Not Applicable	Gaur Gopal Das	Not Applicable	Not Applicable	6 hrs and 25 mins	10-12-2019	English
The Subtle Art Of Not Giving A F*Ck	Mark Manson	Not Applicable	Not Applicable	Roger Wayne	Not Applicable	Not Applicable	5 hrs and 17 mins	13-09-2016	English
Becoming	Michelle Obama	Not Applicable	Not Applicable	Michelle Obama	Not Applicable	Not Applicable	19 hrs and 3 mins	13-11-2018	English
How To Win Friends And Influence People	Dale Carnegie	Not Applicable	Not Applicable	Shernaz Patel	Not Applicable	Not Applicable	8 hrs and 5 mins	08-11-2018	English
Can'T Hurt Me	David Goggins	Not Applicable	Not Applicable	David Goggins	Adam Skolnick	Not Applicable	13 hrs and 37 mins	28-11-2018	English
21 Lessons For The 21St Century	Yuval Noah Harari	Not Applicable	Not Applicable	Derek Perkins	Not Applicable	Not Applicable	11 hrs and 41 mins	30-08-2018	English
Elon Musk	Ashlee Vance	Not Applicable	Not Applicable	Fred Sanders	Not Applicable	Not Applicable	13 hrs and 23 mins	28-04-2016	English
The Power Of Habit	Charles Duhigg	Not Applicable	Not Applicable	Mike Chamberlain	Not Applicable	Not Applicable	10 hrs and 53 mins	05-04-2012	English
The Monk Who Sold His Ferrari	Robin Sharma	Not Applicable	Not Applicable	Robin Sharma	Not Applicable	Not Applicable	2 hrs and 40 mins	05-07-2007	English
Wings Of Fire	APJAbdul Kalam	Arun Tiwari	Not Applicable	Girish Karnad	Not Applicable	Not Applicable	1 hr and 15 mins	01-04-2020	English
Wings Of Fire	APJAbdul Kalam	Arun Tiwari	Not Applicable	Girish Karnad	Not Applicable	Not Applicable	1 hr and 15 mins	01-04-2020	English
Zero To One	Peter Thiel	Blake Masters	Not Applicable	Blake Masters	Not Applicable	Not Applicable	4 hrs and 50 mins	04-06-2015	English
A Promised Land	Barack Obama	Not Applicable	Not Applicable	Barack Obama	Not Applicable	Not Applicable	29 hrs and 10 mins	17-11-2020	English
The Compound Effect	Darren Hardy	Not Applicable	Not Applicable	Darren Hardy	Not Applicable	Not Applicable	4 hrs and 44 mins	22-01-2019	English
The Compound Effect	Darren Hardy	Not Applicable	Not Applicable	Darren Hardy	Not Applicable	Not Applicable	4 hrs and 44 mins	22-01-2019	English
Think Like A Monk	Jay Shetty	Not Applicable	Not Applicable	Jay Shetty	Not Applicable	Not Applicable	10 hrs and 54 mins	08-09-2020	English
Raavan	Amish Tripathi	Not Applicable	Not Applicable	Sagar Arya	Not Applicable	Not Applicable	11 hrs and 32 mins	04-10-2019	English
Homo Deus	Yuval Noah Harari	Not Applicable	Not Applicable	Derek Perkins	Not Applicable	Not Applicable	14 hrs and 53 mins	08-09-2016	English
Harry Potter And The Philosopher'S Stone, Book 1	J.K.Rowling	Not Applicable	Not Applicable	Stephen Fry	Not Applicable	Not Applicable	9 hrs and 33 mins	20-11-2015	English
Harry Potter And The Philosopher'S Stone, Book 1	J.K.Rowling	Not Applicable	Not Applicable	Stephen Fry	Not Applicable	Not Applicable	9 hrs and 33 mins	20-11-2015	English
The Girl In Room 105	Chetan Bhagat	Not Applicable	Not Applicable	Siddhanta Pinto	Not Applicable	Not Applicable	8 hrs and 23 mins	12-11-2018	English

# Cleaned Data

[Click here to see](#)

F	G	H	I	J	K	L	M	N
Narrator 2	Narrator 3	time	releasedate	language	stars	price	Duration in mins	Ratings
Not Applicable	Not Applicable	5 hrs and 35 mins	18-10-2018	English	5	820	335	12573
Not Applicable	Not Applicable	5 hrs and 35 mins	18-10-2018	English	5	820	335	12569
Not Applicable	Not Applicable	3 hrs and 23 mins	18-09-2017	English	4.5	615	203	6981
Not Applicable	Not Applicable	5 hrs and 48 mins	08-09-2020	English	4.5	668	348	5973
Not Applicable	Not Applicable	5 hrs and 48 mins	08-09-2020	English	4.5	668	348	5972
Not Applicable	Not Applicable	15 hrs and 18 mins	30-04-2015	English	5	957	918	5500
Not Applicable	Not Applicable	6 hrs and 9 mins	15-06-2012	English	4.5	837	369	5324
Not Applicable	Not Applicable	6 hrs and 25 mins	10-12-2019	English	4.5	1005	385	4184
Not Applicable	Not Applicable	5 hrs and 17 mins	13-09-2016	English	4.5	1181	317	3461
Not Applicable	Not Applicable	19 hrs and 3 mins	13-11-2018	English	4.5	957	1143	2924
Not Applicable	Not Applicable	8 hrs and 5 mins	08-11-2018	English	4.5	155	485	2494
Adam Skolnick	Not Applicable	13 hrs and 37 mins	28-11-2018	English	5	836	817	2277
Not Applicable	Not Applicable	11 hrs and 41 mins	30-08-2018	English	4.5	957	701	2236
Not Applicable	Not Applicable	13 hrs and 23 mins	28-04-2016	English	4.5	820	803	2161
Not Applicable	Not Applicable	10 hrs and 53 mins	05-04-2012	English	4.5	820	653	1869
Not Applicable	Not Applicable	2 hrs and 40 mins	05-07-2007	English	4.5	351	160	1838
Not Applicable	Not Applicable	1 hr and 15 mins	01-04-2020	English	5	75	75	1827
Not Applicable	Not Applicable	1 hr and 15 mins	01-04-2020	English	5	75	75	1823
Not Applicable	Not Applicable	4 hrs and 50 mins	04-06-2015	English	4.5	615	290	1807
Not Applicable	Not Applicable	29 hrs and 10 mins	17-11-2020	English	4.5	1708	1750	1636
Not Applicable	Not Applicable	4 hrs and 44 mins	22-01-2019	English	4.5	608	284	1564
Not Applicable	Not Applicable	4 hrs and 44 mins	22-01-2019	English	4.5	608	284	1563
Not Applicable	Not Applicable	10 hrs and 54 mins	08-09-2020	English	4.5	569	654	1539
Not Applicable	Not Applicable	11 hrs and 32 mins	04-10-2019	English	4.5	342	692	1491
Not Applicable	Not Applicable	14 hrs and 53 mins	08-09-2016	English	4.5	957	893	1490
Not Applicable	Not Applicable	9 hrs and 33 mins	20-11-2015	English	5	999	573	1425
Not Applicable	Not Applicable	9 hrs and 33 mins	20-11-2015	English	5	999	573	1423
Not Applicable	Not Applicable	8 hrs and 23 mins	12-11-2018	English	4	568	503	1412

# Applied Steps

Query Settings

PROPERTIES

Name

audible\_uncleaned\_\_1

All Properties

APPLIED STEPS

Source

Navigation

Changed Type

Capitalized Each Word-name

author-Extracted Text After Delimiter

multiple author-Split Column by Delimiter

Changed Type1

Filtered Rows

Author4 null-Removed Columns

author2-null-not applicable-Replaced Value

author 3-null-NA-Replaced Value

author1-Split lowercase to uppercase Transition

Merged author1.1,1.2,1.3 with separator space

Trimmed extra space from Author 1

split author 2 with lower to uppercase

merged Author 2.1,2.2 column with space

trim author 2

split author 3 lowercase to uppercase

merges author 3.1,3.2 to 3

trimmed Author 3

Narrator-Extracted Text After Delimiter

narrator-Split Column by Delimiter



Query Settings

PROPERTIES

Name

audible\_uncleaned\_\_1

All Properties

APPLIED STEPS

Narrator-Extracted Text After Delimiter

✕ narrator-Split Column by Delimiter

Removed narrator 4 containing null,blank values

narrator 1-Split Column by Character Transition

Merged Columns-narrator 1.1,1.2,1.3 to Narrator 1

Trimmed extra space-Narrator 1

Narrator 2-Replaced Value-null to NA

Split narrator 2 lower to uppercase

Merged narrator 2.1,2.2,2.3 to Narrator 2 with space ...

Trimmed Text-Narrator 2-remove extra space

Extract hours from time column

hours\*60=hours in minutes -Inserted Multiplication

Removed previous Hours column

extract minutes from time column by using statemm...

Renamed Columns1

add column minutes and hours-mins-Inserted Additi...

Renamed -Addtions

Removed Columns-hours and mins

Filtered Rows1

Replaced Value-price null to 0

Added Ratings-extract using formula from stars

Inserted Text Before Delimiter-stars



# Formulas

## Custom Column

Add a column that is computed from the other columns.

New column name

Minutes

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
    null
```

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

Available columns

name  
Author 1  
Author 2  
Author 3  
Narrator 1  
Narrator 2  
time

## Custom Column

Add a column that is computed from the other columns.

New column name

Ratings

Custom column formula ⓘ

```
= if Text.Contains([stars], "ratings") then
    Number.FromText(Text.BeforeDelimiter
(Text.AfterDelimiter([stars], "stars"), " ratings"))
else
    0
```

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

Available columns

name  
Author 1  
Author 2  
Author 3  
Narrator 1  
Narrator 2  
time

<< Insert

OK

Cancel



# Formulas

### Custom Column

Add a column that is computed from the other columns.

New column name

Custom column formula ⓘ

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

Available columns

- name
- Author 1
- Author 2
- Author 3
- Narrator 1
- Narrator 2
- time

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK Cancel

### Multiply

Enter a number by which to multiply each value in the column.

Value

ABC  
123

OK Cancel

Queries [1]	
23 audible_uncleaned1	4.5

Extracted Text Before Deli...	⚙
Replaced Value2	⚙
Replaced Value3	⚙
Filtered Rows2	
Replaced Value4	⚙
Filtered Rows3	
Replaced Value5	⚙
Changed Type5	
✕ Calculated Median	

# Explanation of Steps

1. **Capitalize the Name Column:** Transform all entries in the "name" column to have each word capitalized for consistent formatting.
2. **Extract Author Names:** Clean the author names by removing any unwanted text, keeping only the relevant name information..
3. **Separate Multiple Authors into Different Columns:** Split entries containing multiple authors into separate columns for each author. Remove columns that contain only null values.
4. **Replace Null Values with "Not Applicable":** For columns where an author may not be present, replace null values with "Not Applicable" to clearly indicate missing information.
5. **Split and Merge Author Names:** For each author, split the first, middle, and last names into separate columns if available. Then merge these parts back into a single column with spaces in between, ensuring extra spaces are trimmed.

# Explanation of Steps

6. **Repeat for All Authors:** Apply the splitting, merging, and trimming process for every author column to standardize the format across all author names.
7. **Standardize Narrator Column:** Repeat the process used for authors in the narrator column, ensuring each narrator's name is split, formatted, and merged similarly.
8. **Convert Time to Minutes:** Extract the hours and minutes from the "time" column, then convert the total time to minutes using an appropriate formula to standardize the time format.
9. **Replace Null Values in Price Column:** Replace any null values in the "price" column with 0 to indicate no cost where the price is missing.

Merge the releasedate and language into a single new column named releasedate info with the format DD-MM-YYYY,language.

Changed Type

✖

Inserted Merged Column

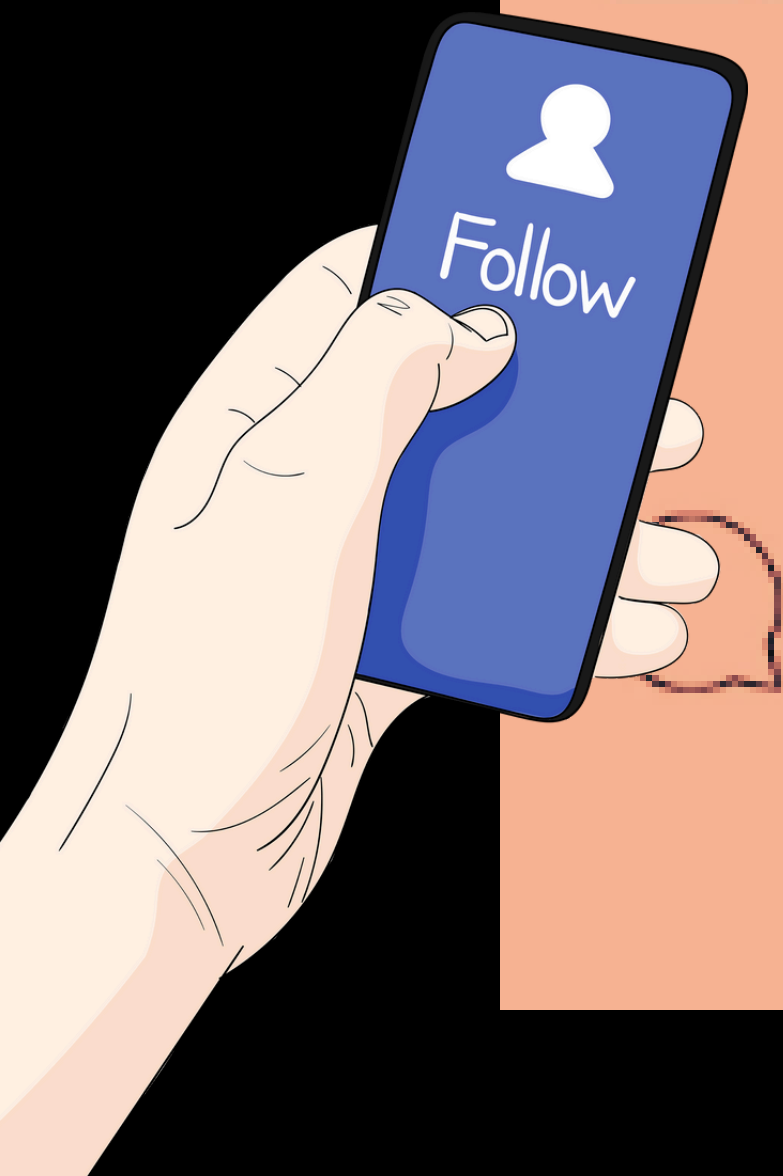
⚙



	releasedate info
73	18-10-2018,English
59	18-10-2018,English
81	18-09-2017,English
73	08-09-2020,English
72	08-09-2020,English
00	30-04-2015,English
24	15-06-2012,English
84	10-12-2019,English
51	13-09-2016,English
24	13-11-2018,English
94	08-11-2018,English
77	28-11-2018,English
86	30-08-2018,English



# HAVE SOMETHING ELSE TO ADD?



Leave a comment to  
share your thoughts!