



audible

Case Study

What You Need to Know A...

Presented by: Shivani Awasthi

Table of Contents

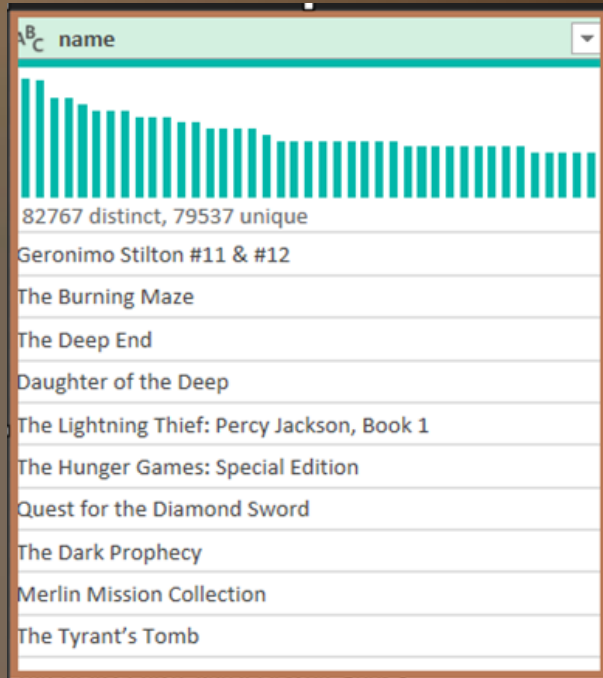
1. Standardize the name column to ensure consistent title casing.
2. Separate combined names in the author column if there are multiple authors.
3. Ensure all entries in the releasedate column follow a consistent date format (DD-MM-YYYY).
4. Convert the time column from text format to a duration format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



Question 1

Problem: Standardize the name column to ensure consistent title casing.

Resolution: It is already in title case. Power Query Editor has changed it.



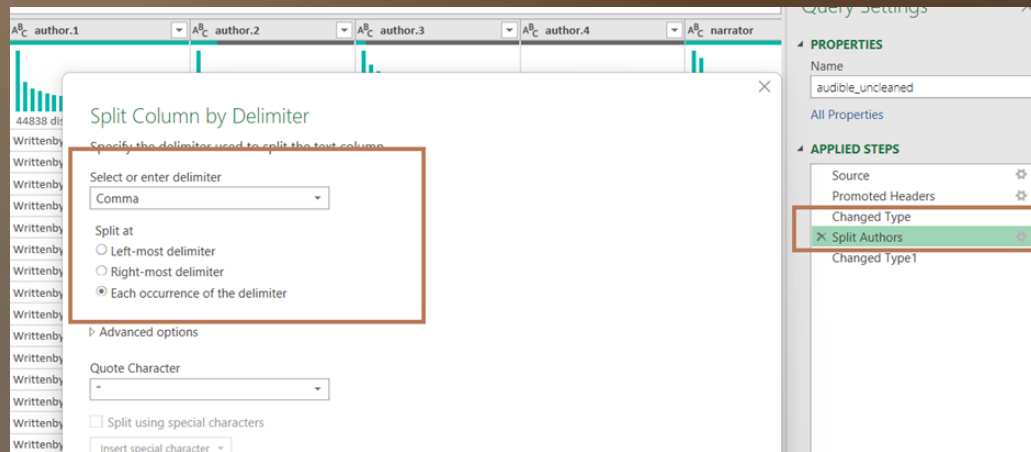
LISTENING ON
audible



Question 2

Problem: Separate combined names in the author column if there are multiple authors.

Resolution: I have used Split by delimiter options to separate the values and then renamed the Author columns.



author 1	author 2	author 3	author 4	narrator
44838 distinct, 32977 unique	9155 distinct, 7871 unique	1999 distinct, 1805 unique	2 distinct, 0 unique	29717 distinct, 20676 unique
Writtenby:JulieSykes	null	null	null	NarratedBy:KristinAther...
Writtenby:FranciscoDiazValladar...	null	null	null	NarratedBy:NuriaSamsó
Writtenby:Mattel	null	null	null	NarratedBy:VanessaPérez
Writtenby:Vegetta777	Willyrex	null	null	NarratedBy:ÁlexdePorrat
Writtenby:RoshaniChokshi	null	null	null	NarratedBy:SoneelaNank
Writtenby:SophieSchoenwald	NadineReitz-Illustrator	null	null	NarratedBy:BerndReheus
Writtenby:HollyRivers	null	null	null	NarratedBy:SusieTrayling

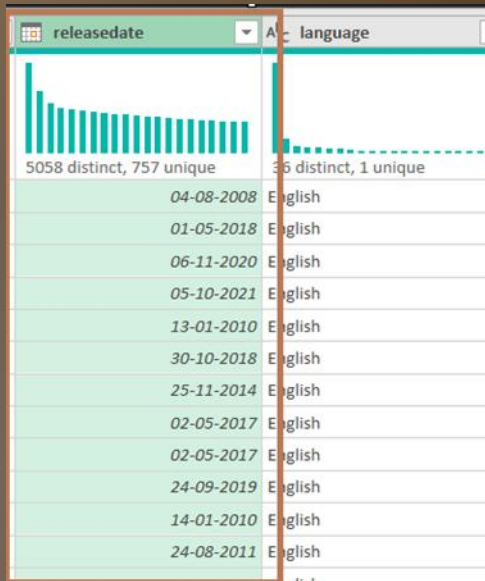
LISTENING ON
audible



Question 3

Problem: Ensure all entries in the releasedate column follow a consistent date format (DD-MM-YYYY).

Resolution: Power Query Editor has already changed it to the required format i.e. DD-MM-YYYY.



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
14-01-2010	English
24-08-2011	English

LISTENING ON
audible

Question 4

Problem: Convert the time column from text format to a duration format that Excel recognizes.

Resolution: I have added a custom column, using below M query and then changed the datatype to Duration.

M Query :

```
let
    txt = Text.Lower([time]),
    hrs = if Text.Contains(txt, "hr") then
        Number.FromText(Text.BeforeDelimiter(txt, "hr")) else 0,
    minsPart = Text.AfterDelimiter(txt, "and "),
    mins = if Text.Contains(minsPart, "min") then
        Number.FromText(Text.BeforeDelimiter(minsPart, "min")) else
0
in
    #duration(0,hrs, mins, 0)
```

Continued on next page....



Question 4

Resolution:

Custom Column

Add a column that is computed from the other columns.

New column name
standardtimeduration

Custom column formula ⓘ

```
= let  
    txt = Text.Lower([time]),  
    hrs = if Text.Contains(txt, "hr") then Number.FromText  
          (Text.BeforeDelimiter(txt, "hr")) else 0,  
    minsPart = Text.AfterDelimiter(txt, "and " ),  
    mins = if Text.Contains(minsPart, "min") then  
            Number.FromText(Text.BeforeDelimiter(minsPart, "min"))  
          else 0  
in  
    #duration(0,hrs, mins, 0)
```

Available columns

- name
- author 1
- author 2
- author 3
- author 4
- narrator
- time

<< Insert

✓ No syntax errors have been detected.

OK Cancel

Properties

Name
audible

All Properties

Applied Steps

- Source
- Promoted Headers
- Changed Type
- Split Authors
- Changed Type1
- Renamed author Columns
- Create standardtime column
- Changed stdtime type to Dura...
- Reordered Columns
- Replaced Errors
- Create ratings Column
- Changed ratings DataType
- Reordered ratings Column
- Split narrators
- Renamed Narrator Columns
- Create releaseinfo Column
- Reordered releaseinfo
- Changed price Type

standardtimeduration

400 distinct, 229 unique

0.02:20:00

0.13:08:00

0.02:03:00

0.11:16:00

0.10:00:00

0.10:35:00

0.02:23:00

0.12:32:00

0.10:56:00

0.13:22:00

0.08:48:00

0.05:23:00

0.06:01:00

0.12:58:00

0.11:55:00

0.02:25:00

0.05:04:00

0.14:41:00

Properties

Name
audible

All Properties

Applied Steps

- Source
- Promoted Headers
- Changed Type
- Split Authors
- Changed Type1
- Renamed author Columns
- Create standardtime column
- Changed stdtime type to Duration
- Reordered Columns
- Replaced Errors
- Changed price Type

LISTENING ON
audible



Question 5

Problem: Ensure the price column is in a numeric format, and identify any non-numeric values.

Resolution: I've looked into the data for Price column and found that a value "free" is present in a few places so used the option "Replace errors" to replace the value "free" with amount "0" to get all the values in numeric format.

The screenshot shows the Microsoft Power Query Editor interface. The 'Transform' tab is active, and the 'Replace Values' button is highlighted in the ribbon. A 'Replace Errors' dialog box is open, prompting the user to enter a value to replace errors in the selected columns. The 'Value' field contains '0'. The background shows a data table with columns: time, standardtime, releasedate, language, stars, and price. The 'price' column is selected, and its data is visible in the table. The 'Query Settings' pane on the right shows the 'APPLIED STEPS' list, with 'Replaced Errors' highlighted.

time	standardtime	releasedate	language	stars	price
1 and 20 mins	0.02:20:00	04-08-2008	English	5 out of 5 stars34 ratings	468
2 and 8 mins	0.12:08:00	02-05-2018	English	4.5 out of 5 stars41 ratings	820
3 and 3 mins					410
4 and 16 mins					615
5					820
6 and 35 mins					656
7 and 23 mins					239
8 and 32 mins					820
9 and 56 mins					1256
10 and 22 mins					820
11 and 48 mins					820
12 and 23 mins					1206
13 and 1 min	0.06:01:00	27-09-2011	English	5 out of 5 stars7 ratings	1206
14 and 58 mins	0.12:58:00	03-10-2017	English	5 out of 5 stars41 ratings	820

LISTENING ON
audible



Question 6

Problem: Convert text ratings in the stars column to numeric values.

Resolution: (i). Created New Custom Column using below M query. Then change the datatype to Number in step (ii).

M Query:

```
try Number.FromText(Text.Start([stars], 1)) otherwise null
```

The screenshot shows the Power Query Editor interface. A 'Custom Column' dialog box is open, displaying the formula: `= try Number.FromText(Text.Start([stars], 1)) otherwise null`. The 'Available columns' list on the right includes 'name', 'author 1', 'author 2', 'author 3', 'author 4', 'narrator', and 'time'. The 'Query Settings' pane on the right shows the 'APPLIED STEPS' list, where 'Create ratings Column' is highlighted. The background shows a table with columns: 'releasedate', 'language', 'stars', 'price', and 'ratings'.

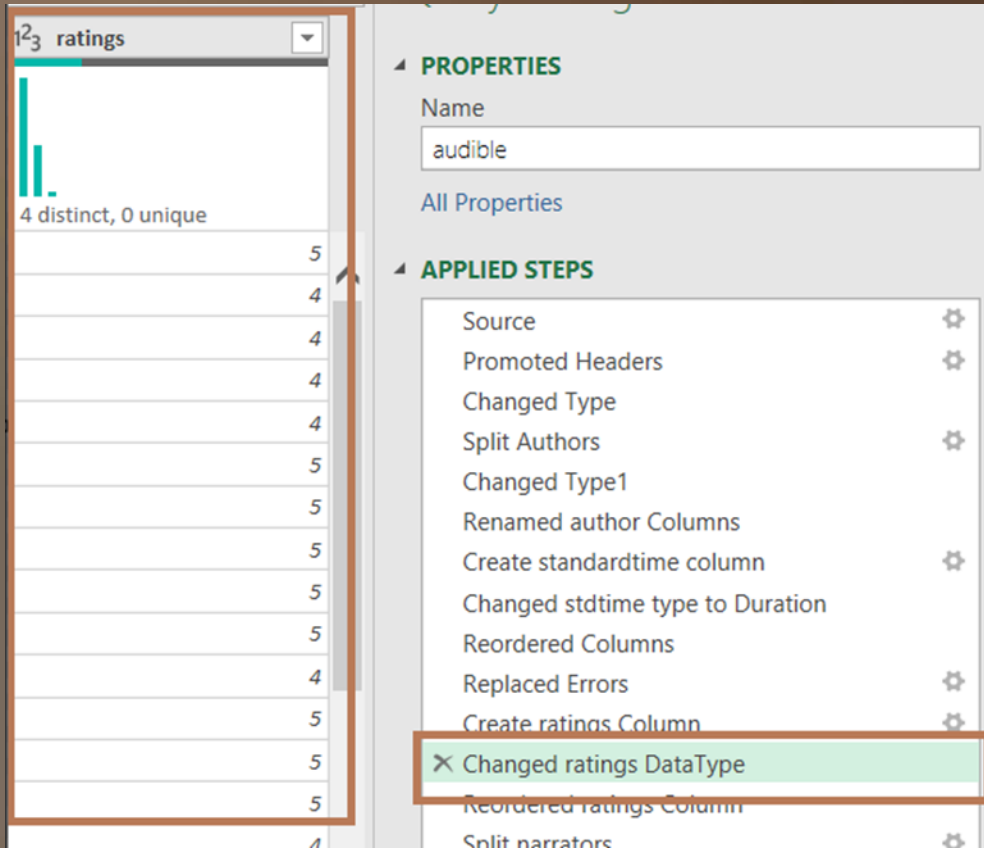
LISTENING ON
audible



Continued on next page.....

Question 6

Resolution: (ii). Changed Datatype of ratings



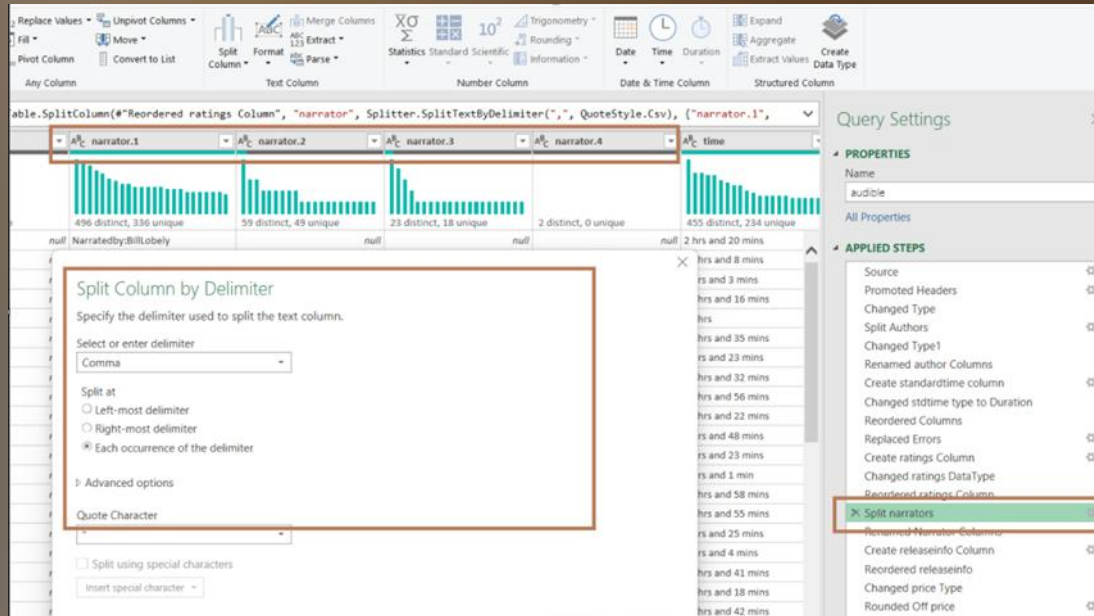
LISTENING ON
audible



Question 7

Problem: Split the narrated by column into multiple columns if multiple narrators are listed.

Resolution: (i). Used Split by delimiter option to separate the values and then rename the headers in step (ii).








LISTENING ON
audible



Continued on next page.....

Question 7

Resolution: (ii). renamed narrator columns.

A narrator 1	A narrator 2	A narrator 3	A narrator 4	A time
				
Narratedby:BillLobley		null	null	2 hrs and 20 mins
Narratedby:RobbieDaymond		null	null	13 hrs and 8 mins
Narratedby:DanRussell		null	null	2 hrs and 3 mins
Narratedby:SoneelaNankani		null	null	11 hrs and 16 mins
Narratedby:JesseBernstein		null	null	10 hrs
Narratedby:TatianaMaslany		null	null	10 hrs and 35 mins
Narratedby:LukeDaniels		null	null	2 hrs and 23 mins
Narratedby:RobbieDaymond		null	null	12 hrs and 32 mins
Narratedby:MaryPopeOsborne		null	null	10 hrs and 56 mins
Narratedby:RobbieDaymond		null	null	13 hrs and 22 mins
Narratedby:JesseBernstein		null	null	8 hrs and 48 mins
Narratedby:MaryPopeOsborne		null	null	5 hrs and 23 mins
Narratedby:MaryPopeOsborne		null	null	6 hrs and 1 min
Narratedby:MichaelCrouch		null	null	12 hrs and 58 mins
Narratedby:PhilipPullman	fullcast	RuthWilson		11 hrs and 55 mins
Narratedby:BillLobley		null	null	2 hrs and 25 mins

LISTENING ON
audible



PROPERTIES

Name

audible

All Properties

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Split Authors
- Changed Type1
- Renamed author Columns
- Create standardtime column
- Changed stdtime type to Duration
- Reordered Columns
- Replaced Errors
- Create ratings Column
- Changed ratings DataType
- Reordered ratings Column
- Split narrators
- Renamed Narrator Columns

Question 8

Problem: Merge the releasedate and language columns into a single new column named releaseinfo with the format "DD-MM-YYYY, Language."

Resolution: Used Merge Columns options in Add Column tab and created a new merged column named "releaseinfo".

LISTENING ON
audible



The screenshot shows the Power BI Desktop interface. The 'Add Column' tab is active, displaying various transformation options. The 'Merge Columns' option is highlighted. Below the 'Merge Columns' button, the 'From Text' section is expanded, showing the 'Merge Columns' dialog box. The dialog box prompts the user to 'Choose how to merge the selected columns.' The 'Separator' is set to 'Comma'. The 'New column name (optional)' is set to 'releaseinfo'. The background shows a data table with columns: releasedate, language, stars, ratings, price, and releaseinfo. The 'releaseinfo' column contains the merged data in the format 'DD-MM-YYYY, Language'.

releasedate	language	stars	ratings	price	releaseinfo
04-08-2008	English	5 out of 5 stars7 ratings	5	1206	04-08-2008,English
01-05-2018	English	5 out of 5 stars41 ratings	5	820	01-05-2018,English
06-11-2020	English	4 out of 5 stars2 ratings	4	1093	06-11-2020,English
05-10-2021	English	4.5 out of 5 stars33 ratings	4	467	05-10-2021,English
13-01-2010	English	5 out of 5 stars7 ratings	5	1206	13-01-2010,English
30-10-2018	English	5 out of 5 stars28 ratings	5	820	30-10-2018,English
25-11-2014	English	5 out of 5 stars7 ratings	5	1206	25-11-2014,English
02-05-2017	English	5 out of 5 stars7 ratings	5	1206	02-05-2017,English
02-09-2019	English	5 out of 5 stars7 ratings	5	1206	02-09-2019,English
24-09-2019	English	5 out of 5 stars7 ratings	5	1206	24-09-2019,English
14-01-2010	English	5 out of 5 stars7 ratings	5	1206	14-01-2010,English
24-08-2011	English	5 out of 5 stars7 ratings	5	1206	24-08-2011,English
27-09-2011	English	5 out of 5 stars7 ratings	5	1206	27-09-2011,English
03-10-2017	English	5 out of 5 stars7 ratings	5	1206	03-10-2017,English
24-06-2021	English	5 out of 5 stars7 ratings	5	1206	24-06-2021,English
08-02-2008	English	5 out of 5 stars7 ratings	5	1206	08-02-2008,English
26-12-2004	English	5 out of 5 stars7 ratings	5	1206	26-12-2004,English

Question 9

Problem: Ensure all currency values in the price column are formatted consistently with two decimal places.

Resolution: Changed the datatype to Currency and then from Transform Tab used Round, added value as 2.

The screenshot shows a data transformation tool interface. A 'Round' dialog box is open, prompting the user to specify how many decimal places to round to. The 'Decimal Places' field is set to 2. The background shows a data table with columns: language, releaseinfo, stars, ratings, and price. The 'price' column contains values like 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, 820.00, 1,093.00, 467.00, 1,206.00, 836.00, 1,256.00, and 1,003.00. The 'APPLIED STEPS' panel on the right shows a list of transformations, with 'Rounded Off price' highlighted at the bottom.

language	releaseinfo	stars	ratings	price
English	04-08-2008,English	5 out of 5 stars	34 ratings	468.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	820.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	410.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	615.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	820.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	656.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	233.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	820.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	1,256.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	820.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	820.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	1,206.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	1,206.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	820.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	1,093.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	467.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	1,206.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	836.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	1,256.00
English	04-08-2008,English	5 out of 5 stars	34 ratings	1,003.00

Round

Specify how many decimal places to round to.

Decimal Places

2

OK Cancel

PROPERTIES

Name

audible

All Properties

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Split Authors
- Changed Type1
- Renamed author Columns
- Create standardtime column
- Changed stdtime type to Duration
- Reordered Columns
- Replaced Errors
- Create ratings Column
- Changed ratings DataType
- Reordered ratings Column
- Split narrators
- Renamed Narrator Columns
- Create releaseinfo Column
- Reordered releaseinfo
- Changed price Type
- Rounded Off price

LISTENING ON

audible





Thank You

Presented by: Shivani Awasthi
Mentored by: Ayushi Jain

