



Audible Data Cleaning & Analysis Project

"Uncovering Insights from the World of Audible"

CASE STUDIE PERSENTATION

INTRODUCTION

- **Audible** is Amazon's audiobook and spoken-word entertainment platform.
- It allows users to listen to book 📖 instead of reading which is useful for busy people.
- Offers a huge library of audiobook, podcast, and original shows.

JEFF BEZOS

FOUNDER

Case study Overview

- *This Project analysis **Audible** dataset with a focus on data cleaning and business insights.*
- *Objective : Prepare raw data for meaningful analysis using Excel tools.*
- **Process followed:**

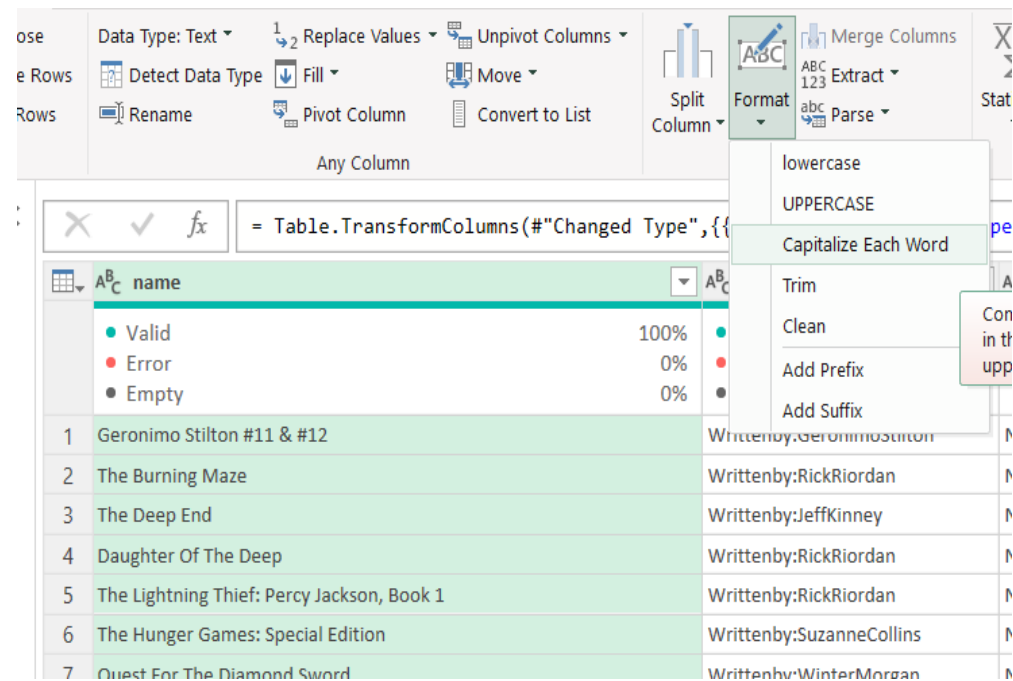


- **Tools Used :**
 - *Microsoft Excel (Power Query , Pivot Tables , Formulas , Charts).*

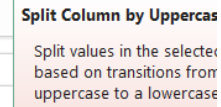
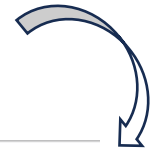
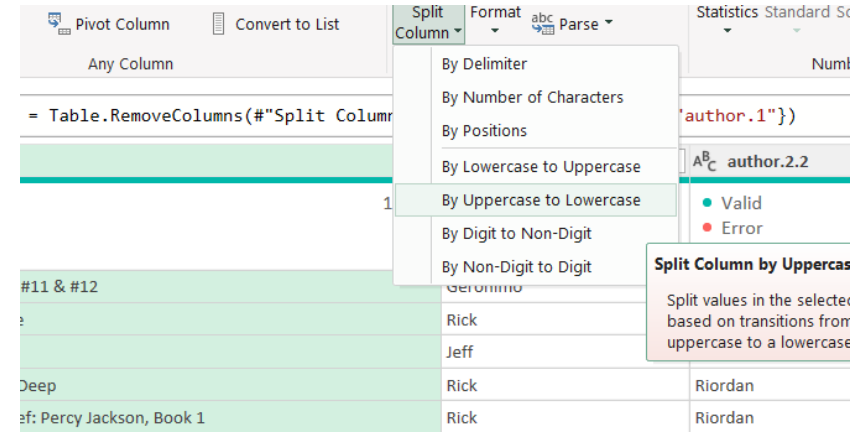
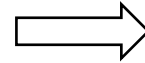
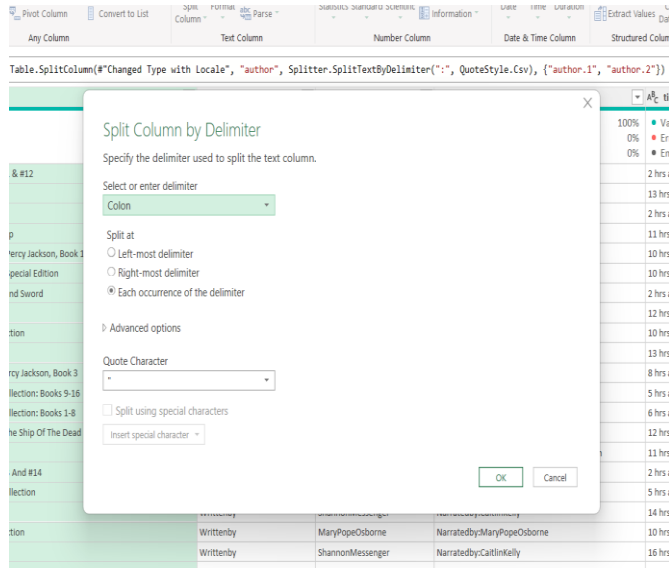
Content

Question : *Standardize the name column to ensure consistent title casing.*

- **Task :** Ensure all books names follow Title Case.
- **Step Used :** *Power Query → Transform → Format → Capitalize Each Word.*



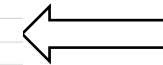
Question : *Separate combined names in the author column if there are multiple authors.*



Steps: Power Query ➡ Split column ➡ By Delimiter

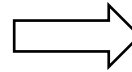
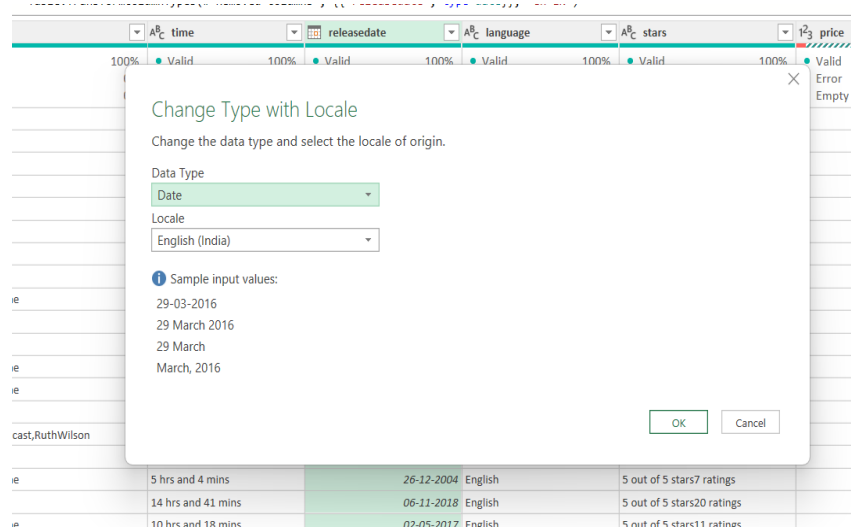
“Result: multiple author separated into different column.”

author.2.1	author.2.2	author.2.3
Geronimo	Stilton	
Rick	Riordan	
Jeff	Kinney	
Rick	Riordan	
Rick	Riordan	
Suzanne	Collins	
Winter	Morgan	
Rick	Riordan	
Mary	Pope	Osborne
Rick	Riordan	
Rick	Riordan	
Mary	Pope	Osborne
Mary	Pope	Osborne
Rick	Riordan	
Phillip	Pullman	
Geronimo	Stilton	
Mary	Pope	Osborne
Shannon	Messenger	
Mary	Pope	Osborne
Shannon	Messenger	
Rick	Riordan	
Christopher	Paolini	



author.1	author.2	author.3
Writtenby	GeronimoStilton	Ni
Writtenby	RickRiordan	Ni
Writtenby	JeffKinney	Ni
Writtenby	RickRiordan	Ni
Writtenby	RickRiordan	Ni
Writtenby	SuzanneCollins	Ni
Writtenby	WinterMorgan	Ni
Writtenby	RickRiordan	Ni
Writtenby	MaryPopeOsborne	Ni
Writtenby	RickRiordan	Ni
Writtenby	RickRiordan	Ni
Writtenby	MaryPopeOsborne	Ni
Writtenby	MaryPopeOsborne	Ni
Writtenby	MaryPopeOsborne	Ni
Writtenby	RickRiordan	Ni
Writtenby	PhilipPullman	Ni

Question : *Ensure all entries in the release date column follow a consistent date format (DD-MM-YYYY) .*



removed Columns", [{"releasedate", type date}], "en-I

	releasedate	language
100%	Valid	100%
0%	Error	0%
0%	Empty	0%
	04-08-2008	English
	01-05-2018	English
	06-11-2020	English
15	05-10-2021	English
	13-01-2010	English
15	30-10-2018	English
	25-11-2014	English
15	02-05-2017	English
15	02-05-2017	English
15	24-09-2019	English
	14-01-2010	English
	24-08-2011	English
	27-09-2011	English
15	03-10-2017	English
15	24-06-2021	English
	08-02-2008	English
	26-12-2004	English
15	06-11-2018	English
15	02-05-2017	English
15	06-11-2018	English
15	06-10-2020	English
15	14-04-2009	English
	11-04-2013	English
	01-02-2011	English

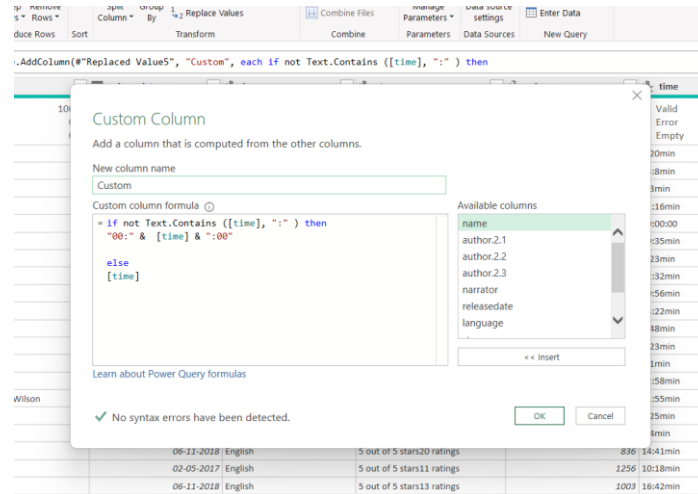
Question : Convert the time column from text format to a duration format that excel recognizes.

- First Used Replaced value to remove extra text like “hr”, “min”, “sec” for a clean numeric format in time column.
- Created a Custom column to rearrange values into proper time format.
- After Custom column we split the custom column and Add the column hr , min ,sec and Divided to 24 hr.
- Converted the column into Duration.
- Verified transformation through Applied Steps panel.
- “Now the column is in proper Duration format which can be used for further analysis like average time , total time ,etc.

All Properties

APPLIED STEPS

- Replaced Value1
- Replaced Value2
- Reordered Columns
- Replaced Value3
- Replaced Value4
- Replaced Value5
- Added Custom
- Replaced Value6
- Renamed Columns
- Split Column by Delimiter1
- Changed Type2
- Renamed Columns1
- Replaced Value7
- Multiplied Column
- ✕ Multiplied Column1
- Inserted Sum
- ✕ Removed Columns1
- Divided Column
- Divided Column1
- Divided Column2
- Changed Type3
- Renamed Columns2



0%	• Empty	0%	• Empty	0%	• Empty
04-08-2008	English	5 out of 5 stars34 ratings			
01-05-2018	English	4.5 out of 5 stars41 ratings			
06-11-2					
05-10-2					
13-01-2					
30-10-2					
25-11-2					
02-05-2					
02-05-2					
24-09-2					
14-01-2					
24-08-2					
27-09-2011	English	5 out of 5 stars7 ratings			
03-10-2017	English	5 out of 5 stars11 ratings			
24-06-2021	English	4 out of 5 stars2 ratings			

Divide

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

Value

24

	Duration column
%	Valid 100%
%	Error 0%
%	Empty 0%
	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
	0.10:18:00
	0.16:42:00
	0.12:12:00
	0.23:29:00
	0.06:46:00
	0.02:37:00

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

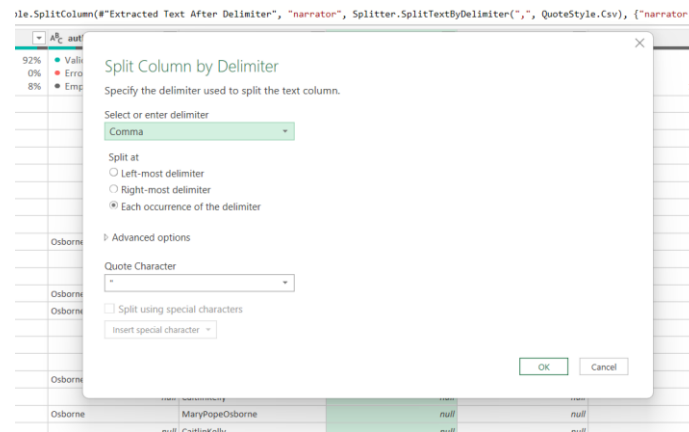
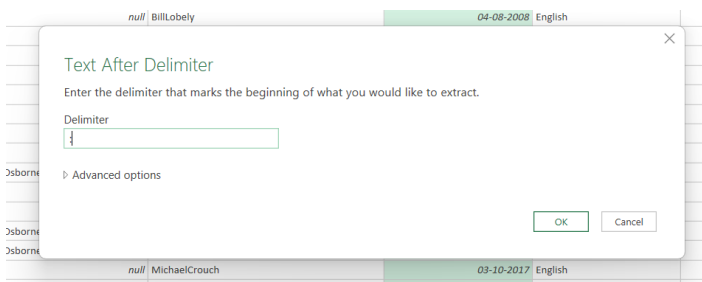
- *Selected the Price column in Power Query .*
- *Changed the Data type to Decimal Number.*
- *Used Replaced Errors \Rightarrow 0 to handle invalid values.*
- *Verified that now all values are valid & in Decimal format.*

1.2 price	
Valid	100%
Error	0%
Empty	0%
468	
820	
410	
615	
820	
656	
233	
820	
1256	
820	
820	
1206	
1206	
820	
1093	
467	
1206	
836	
1256	
1003	
615	
957	
683	
469	

Question : *Convert text rating in the star column to numeric values.*

Question : *Split the narrated by column into multiple column if multiple narrators are listed.*

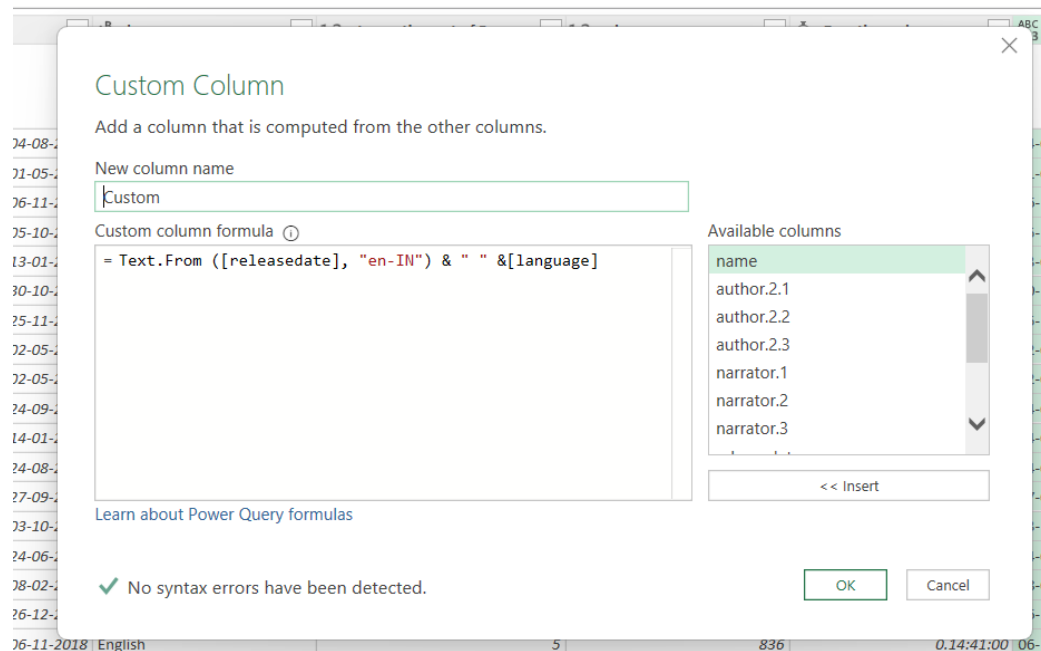
- *The Narrated By column had multiple narrators in one cell.*
- *Applied Split Column By Delimiter (comma) in power query.*
- *Each narrator was separated into new columns.*
- *This made the data cleaner and easier to analyze.*



A ^B C narrator.1	A ^B C narrator.2	A ^B C narrator.3
BillLobley		
RobbieDaymond		
DanRussell		
SoneelaNankani		
JesseBernstein		
TatianaMaslany		
LukeDaniels		
RobbieDaymond		
MaryPopeOsborne		
RobbieDaymond		
JesseBernstein		
MaryPopeOsborne		
MaryPopeOsborne		
MichaelCrouch		
PhilipPullman	fullcast	RuthWilson
BillLobley		
MaryPopeOsborne		
CaitlinKelly		
MaryPopeOsborne		
CaitlinKelly		
RobbieDaymond		
GerrardDoyle		
GerryO'Brien		
BillLobley		

Question : Merge the released column and language column into a single new column named released Info with the format “DD-MM-YYYY, Language”.

- **Ensure Released date column set into date type.**
- **Formatted Released as “DD-MM-YYYY”.**
- **Added Custom column to merge Released + Language.**
- **Final format : 25-12-2020 English.**



ABC 123	Released Info
%	Valid 100%
%	Error 0%
%	Empty 0%
00	04-08-2008 English
00	01-05-2018 English
00	06-11-2020 English
00	05-10-2021 English
00	13-01-2010 English
00	30-10-2018 English
00	25-11-2014 English
00	02-05-2017 English
00	02-05-2017 English
00	24-09-2019 English
00	14-01-2010 English
00	24-08-2011 English
00	27-09-2011 English
00	03-10-2017 English
00	24-06-2021 English
00	08-02-2008 English
00	26-12-2004 English
00	06-11-2018 English
00	02-05-2017 English
00	06-11-2018 English
00	06-10-2020 English
00	14-04-2009 English
00	11-04-2013 English
00	01-02-2011 English
00	01-01-2010 English

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

- ***Changed Price column data type to Currency.***
- ***Ensure values are sorted as valid numeric currency.***
- ***All prices standardized to two decimal places.***
- ***Achieved data consistency & accuracy.***

Round

Specify how many decimal places to round to.

Decimal Places

2

OK Cancel

pe}})			
	▼	\$ price	▼
100%	Valid	100%	Valid
0%	Error	0%	Error
0%	Empty	0%	Empty
5		468.00	
4		820.00	
4		410.00	
4		615.00	
4		820.00	
5		656.00	
5		233.00	
5		820.00	
5		1,256.00	
5		820.00	
4		820.00	
5		1,206.00	
5		1,206.00	
5		820.00	
4		1,093.00	
4		467.00	
5		1,206.00	
5		836.00	
5		1,256.00	
5		1,003.00	
5		615.00	
4		957.00	



CONCLUSION

- *The dataset was successfully cleaned and standardized using Power Query Editor in Excel.*
 - *Inconsistent formats across names, authors, dates, time, and prices were standardized for **accuracy**.*
 - *Columns were reformatted into numeric, text, and duration types for better usability.*
 - *Complex fields such as multiple authors and narrators were split into separate columns to improve clarity.*
 - *New combined fields like Release Info (Date + Language) were created to enhance **readability**.*
 - *All currency and rating values were converted into consistent numeric formats.*
-
- ✅ *Overall, the dataset is now **structured, consistent, and analysis-ready**, ensuring reliable insights and decision-making.*

thanks!

SUBMITTED BY: *Gajendra Singh Panwar*