



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.
- Offer's 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:**

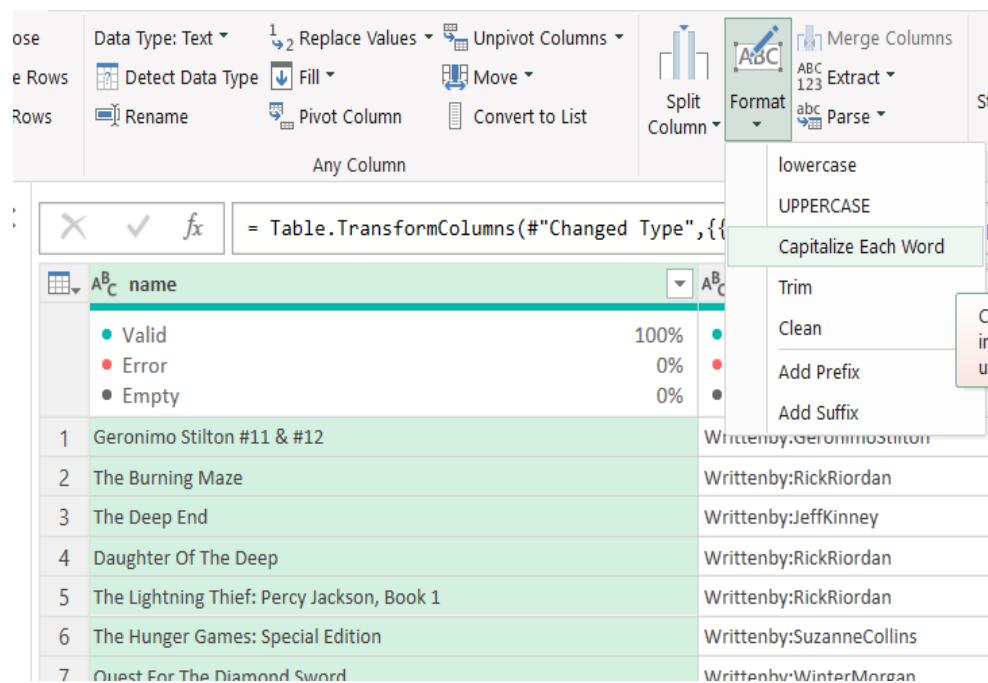


- **Tools Used :**
 - *Microsoft Excel (Power Query , Pivot Tables , Formulas , Charts).*

Content

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

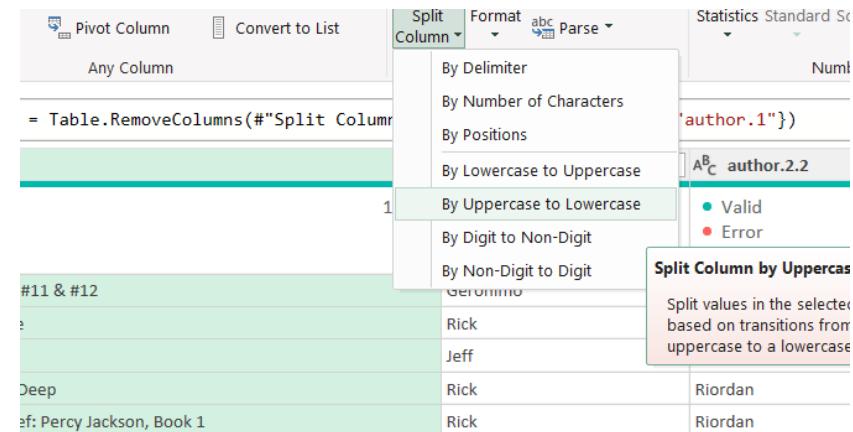
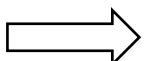
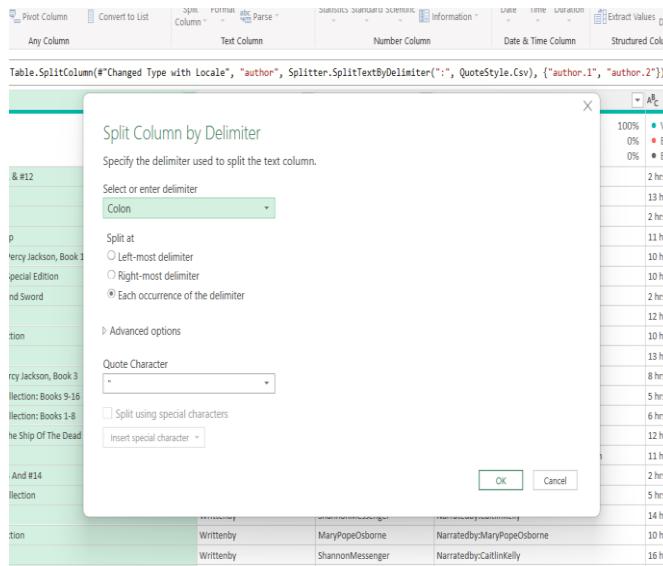
- **Task :** Ensure all books names follow Title Case.
- **Step Used :** Power Query → Transform → Format → Capitalize Each Word.



The screenshot shows the Microsoft Power Query Editor interface. A context menu is open over the 'name' column, specifically under the 'Format' section. The 'Capitalize Each Word' option is selected. Below the table, the formula bar shows the formula used for the transformation: `= Table.TransformColumns(#"Changed Type", {{"name", Capitalize Each Word}})`.

| | name | Writtenby |
|---|--|---------------------------|
| 1 | Geronimo Stilton #11 & #12 | Writtenby:geronimostilton |
| 2 | The Burning Maze | Writtenby:RickRiordan |
| 3 | The Deep End | Writtenby:JeffKinney |
| 4 | Daughter Of The Deep | Writtenby:RickRiordan |
| 5 | The Lightning Thief: Percy Jackson, Book 1 | Writtenby:RickRiordan |
| 6 | The Hunger Games: Special Edition | Writtenby:SuzanneCollins |
| 7 | Quest For The Diamond Sword | Writtenby:WinterMorgan |

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



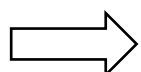
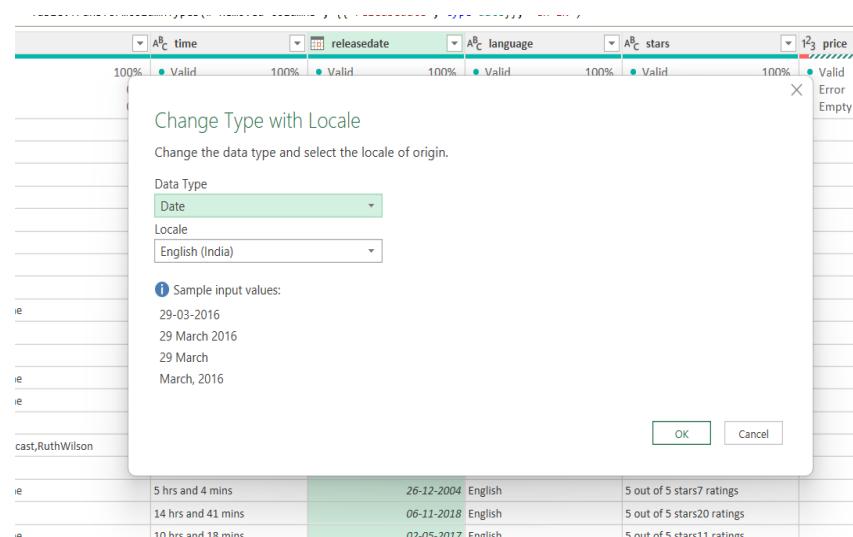
| | A _C author.1 | A _C author.2 | A _C author.3 |
|-------------|-------------------------|-------------------------|-------------------------|
| i0% | ● Valid | 100% | ● Valid |
| 0% | ● Error | 0% | ● Error |
| 0% | ● Empty | 0% | ● Empty |
| Geronimo | Stilton | | |
| Rick | Riordan | | |
| Jeff | Kinney | | |
| Rick | Riordan | | |
| Suzanne | Collins | | |
| Winter | Morgan | | |
| Rick | Riordan | | |
| Mary | Pope | Osborne | |
| Rick | Riordan | | |
| Rick | Riordan | | |
| Mary | Pope | Osborne | |
| Mary | Pope | Osborne | |
| Rick | Riordan | | |
| Philip | Pullman | | |
| Geronimo | Stilton | | |
| Mary | Pope | Osborne | |
| Shannon | Messenger | | |
| Mary | Pope | Osborne | |
| Shannon | Messenger | | |
| Rick | Riordan | | |
| Christopher | Paolini | | |

| | A _C author.1 | A _C author.2 | A _B author.3 |
|-----------|-------------------------|-------------------------|-------------------------|
| 0% | ● Valid | 100% | ● Valid |
| 0% | ● Error | 0% | ● Error |
| 0% | ● Empty | 0% | ● Empty |
| Writtenby | Geronimo | Stilton | N/A |
| Writtenby | Rick | Riordan | N/A |
| Writtenby | Jeff | Kinney | N/A |
| Writtenby | Rick | Riordan | N/A |
| Writtenby | Rick | Riordan | N/A |
| Writtenby | Suzanne | Collins | N/A |
| Writtenby | Winter | Morgan | N/A |
| Writtenby | Rick | Riordan | N/A |
| Writtenby | Mary | Pope | Osborne |
| Writtenby | Rick | Riordan | N/A |
| Writtenby | Mary | Pope | Osborne |
| Writtenby | Rick | Riordan | N/A |
| Writtenby | Philip | Pullman | N/A |
| Writtenby | Geronimo | Stilton | N/A |
| Writtenby | Mary | Pope | Osborne |
| Writtenby | Shannon | Messenger | N/A |
| Writtenby | Mary | Pope | Osborne |
| Writtenby | Shannon | Messenger | N/A |
| Writtenby | Rick | Riordan | N/A |
| Writtenby | Christopher | Paolini | N/A |

Steps: Power Query → Split column → By Delimiter

"Result: multiple author separated into different column."

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



| | 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 |
| | 30-10-2018 | English |
| | 25-11-2014 | English |
| 15 | 02-05-2017 | English |
| | 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 |
| | 24-06-2021 | English |
| | 08-02-2008 | English |
| | 26-12-2004 | English |
| 15 | 06-11-2018 | English |
| | 02-05-2017 | English |
| 15 | 06-11-2018 | English |
| | 06-10-2020 | English |
| 15 | 14-04-2009 | English |
| | 11-04-2013 | English |
| | 01-02-2011 | English |

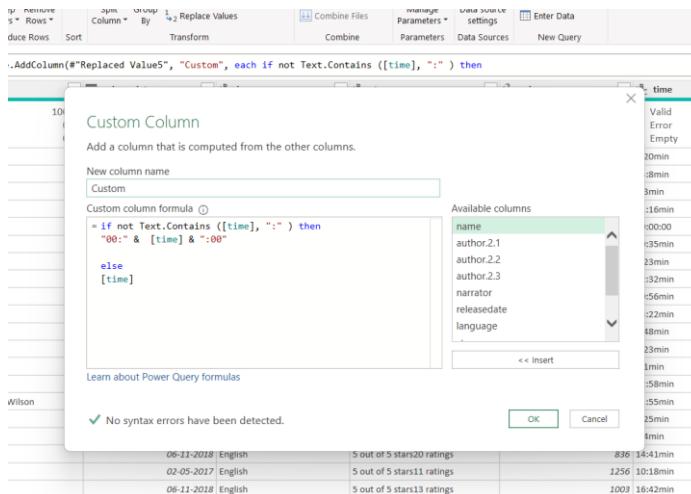
Question : Convert the time column from text format to a duration format that excel recognizes.

- 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 stars | 34 ratings | | |
| 01-05-2018 | English | 4.5 out of 5 stars | 41 ratings | | |
| 06-11-2018 | | | | | |
| 05-10-2018 | | | | | |
| 13-01-2018 | | | | | |
| 30-10-2017 | | | | | |
| 25-11-2017 | | | | | |
| 02-05-2018 | | | | | |
| 02-05-2018 | | | | | |
| 24-09-2018 | | | | | |
| 14-01-2018 | | | | | |
| 24-08-2018 | | | | | |
| 27-09-2017 | English | 5 out of 5 stars | 7 ratings | | |
| 03-10-2017 | English | 5 out of 5 stars | 41 ratings | | |
| 24-06-2021 | English | 4 out of 5 stars | 2 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 |
| | 0.10:22:00 |

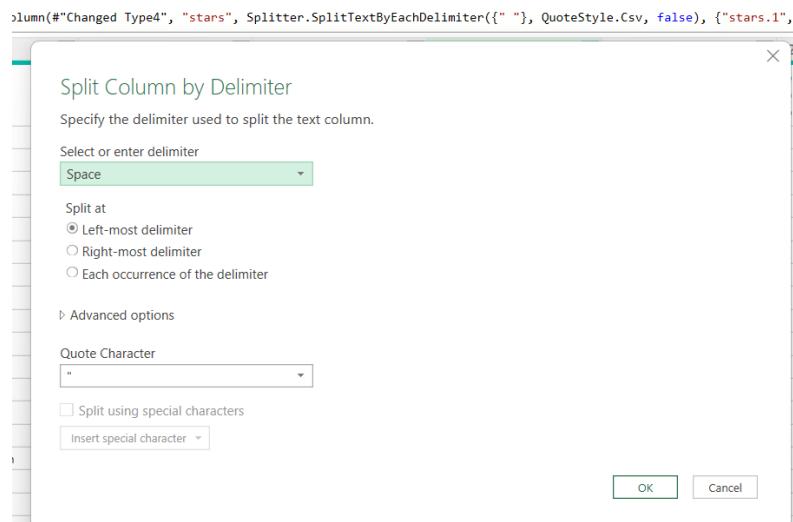
Question : Ensure the price column is in a numeric format , and identify any non-numeric values.

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

| 1.2 price | |
|-----------|-------|
| %) | Valid |
| %) | Error |
| %) | Empty |
| | 100% |
| | 0% |
| | 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.

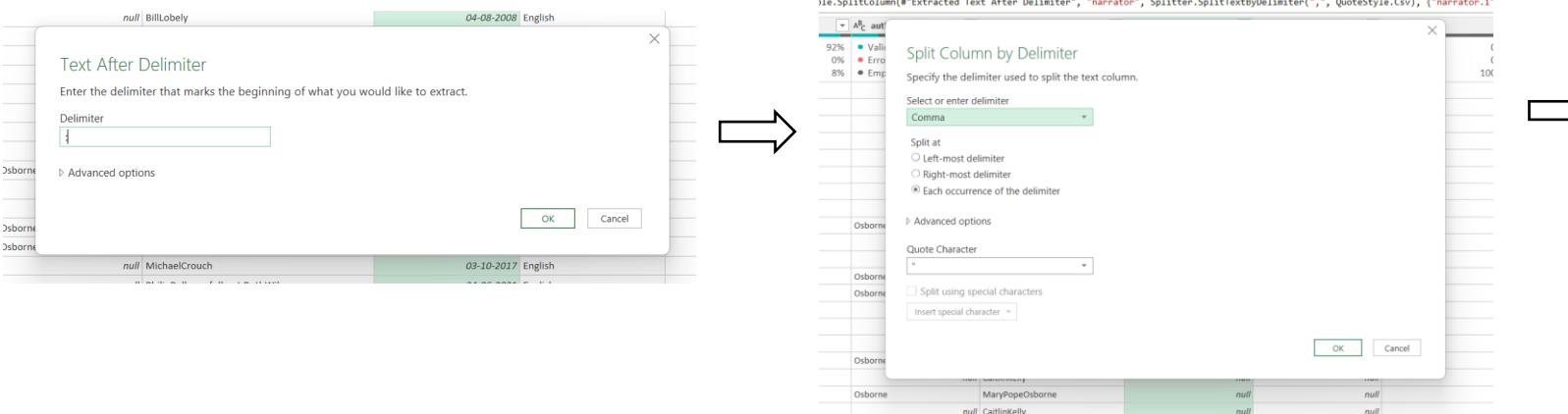
- *Used Split column by Delimiter to separate numeric rating from text.*
 - *Removed extra text column (like “out of star 5 rating”).*
 - *Changed column data type to decimal number (for accuracy with 4.5 etc.).*
 - *Final result Clean numeric star rating .*



| 1.2 stars rating out of 5 | | |
|---------------------------|----|---------|
| 100% | 0% | 0% |
| | | ● Valid |
| | | ● Error |
| | | ● Empty |
| | | 5 |
| | | 4 |
| | | 4 |
| | | 4 |
| | | 4 |
| | | 5 |
| | | 5 |
| | | 5 |
| | | 5 |
| | | 5 |
| | | 4 |
| | | 5 |
| | | 5 |
| | | 4 |
| | | 4 |
| | | 5 |
| | | 5 |
| | | 5 |
| | | 5 |
| | | 4 |

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

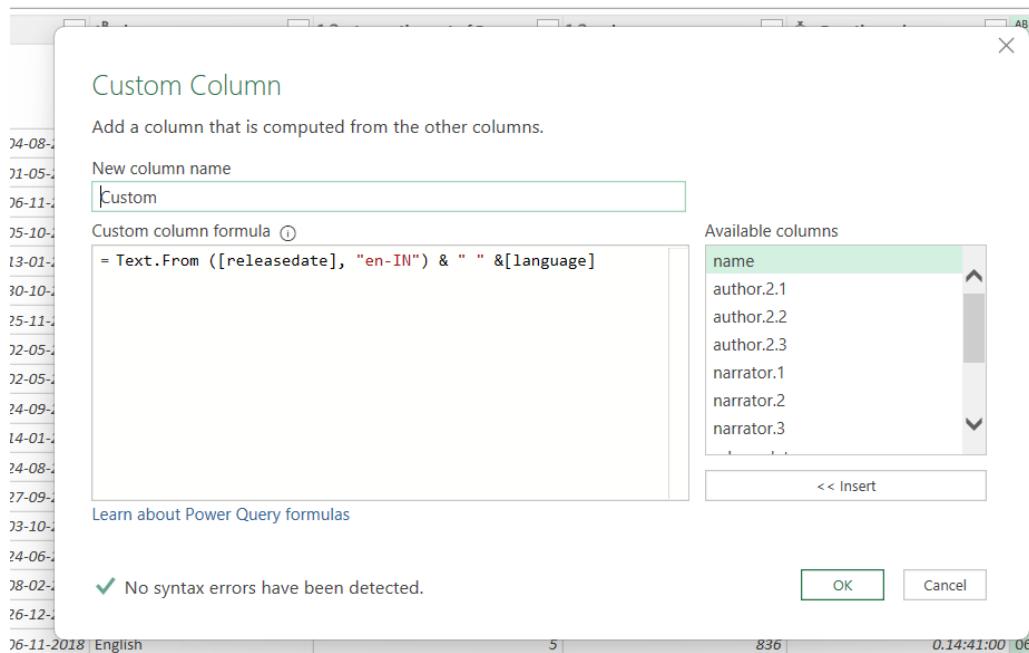
- **The Narrated By column had multiple narrators in one cell.**
- **Applied Split Column By Delimiter (comma)in power query.**
- **Each narrator was separated into new columns.**
- **This made the data cleaner and easier to analyze.**




A screenshot of a spreadsheet showing the results of the column split. The original "narrator" column has been split into three new columns: "narrator.1", "narrator.2", and "narrator.3". The data in "narrator.1" includes names like BillLobley, RobbieDaymond, DanRussell, etc. The data in "narrator.2" includes names like SoneelaNankani, JesseBernstein, TatianaMaslany, etc. The data in "narrator.3" includes names like LukeDaniels, RobbieDaymond, MaryPopeOsborne, etc. The last row shows "fulcast" and "RuthWilson".

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 | Released Info | 123 |
|----|--------------------|---------------|-----|
| % | ● 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 | 21-01-2010 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.**

| | | | | | | | | | |
|------------|---------|----|---------|----|---------|------------|------------|----|---------|
| 0% | • Empty | 0% | • Empty | 0% | • Empty | 0% | • Empty | 0% | • Empty |
| 04-08-2008 | English | | | 5 | 468 | 0.02:20:00 | 04-08-2008 | | |
| 01-05-2018 | English | | | 4 | 820 | 0.13:08:00 | 01-05-2018 | | |
| 06-11- | | | | | | | | | |
| 05-10- | | | | | | | | | |
| 13-01- | | | | | | | | | |
| 30-10- | | | | | | | | | |
| 25-11- | | | | | | | | | |
| 02-05- | | | | | | | | | |
| 02-05- | | | | | | | | | |
| 24-09- | | | | | | | | | |
| 14-01- | | | | | | | | | |
| 24-08- | | | | | | | | | |
| 27-09-2011 | English | | | 5 | 1206 | 0.06:01:00 | 27-09-2011 | | |
| 03-10-2017 | English | | | 5 | 820 | 0.12:58:00 | 03-10-2017 | | |
| 24-06-2021 | English | | | 4 | 1093 | 0.11:55:00 | 24-06-2021 | | |

Round

Specify how many decimal places to round to.

Decimal Places

OK Cancel

| pe}}) | \$ price | 100% | Valid | 100% |
|-------|----------|------|---------|------|
| 0% | 0% | 0% | • Valid | 100% |
| 0% | 0% | 0% | • Error | 0% |
| 0% | 0% | 0% | • Empty | 0% |
| 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!
A large, stylized orange arrow that curves upwards and to the right, ending in a small arrowhead, positioned below the word "thanks!"

SUBMITTED BY: *Gajendra Singh Panwar*