



# Audible Data Cleaning Project

Presented by Md Suhail Alam

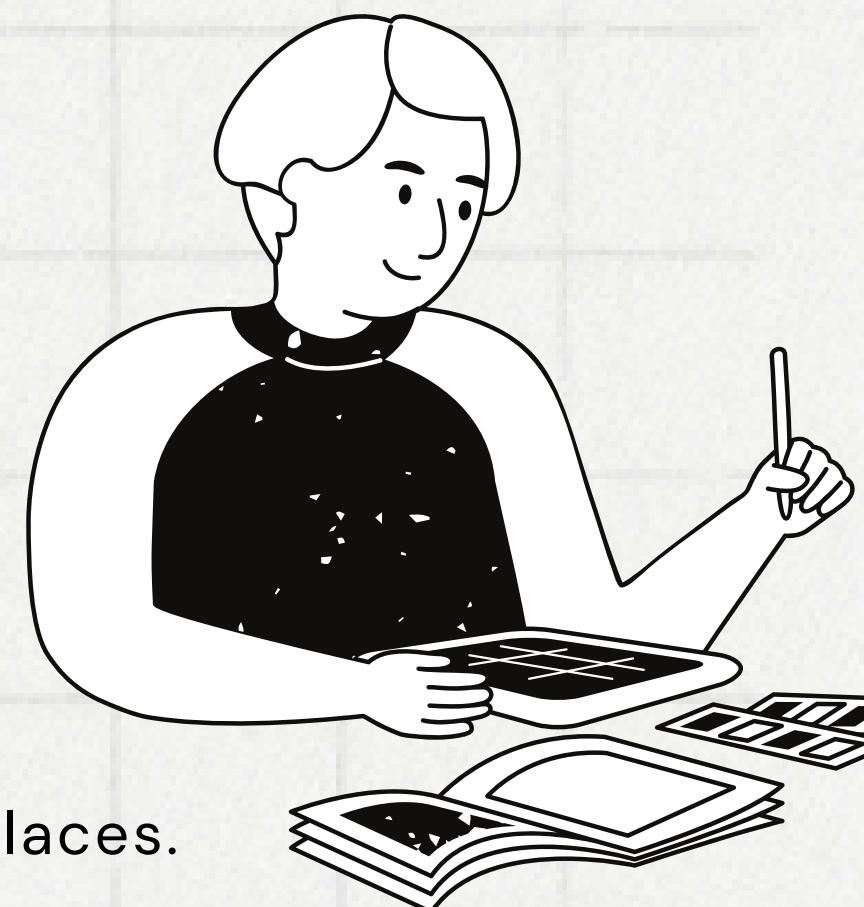
# Questions

- Standardized the Name Column: Converted all names to title case for consistent formatting.
- Separated Combined Names: Split the combined first and last names in the "Author" column into two separate columns.
- Ensured Consistent Date Format: Converted all dates in the "ReleaseDate" column to the DD-MM-YYYY format.
- Converted Text to Duration: Transformed the "Time" column from text format to a recognized duration format in Excel.
- Validated Numeric Price Values: Converted the "Price" column to numeric format, identifying and addressing any non-numeric values.
- Converted Text Ratings: Transformed text-based ratings in the "Stars" column into numeric values.
- Handled Multiple Narrators: Split the "NarratedBy" column into multiple columns where multiple narrators were listed.
- Merged Columns: Combined the "ReleaseDate" and "Language" columns into a new column called "ReleaseInfo" in the format "DD-MM-YYYY, Language."
- Formatted Currency Values: Ensured all values in the "Price" column were consistently formatted with two decimal places.



# Summary

- This project involved cleaning and transforming an Audible dataset using Power Query Editor in Microsoft Excel. The primary objective was to standardize the dataset, ensure data consistency, and prepare it for further analysis or dashboard creation.
- The dataset contained inconsistencies in text formatting, combined fields, unstructured durations, and currency formatting. Each of these issues was addressed using a series of data transformation steps in Power Query, applying both built-in and custom M language functions.
- Key Objectives:
  - Ensure consistency in naming conventions, date formats, and numeric fields.
  - Convert unstructured text durations into Excel-recognized duration format.
  - Prepare the dataset for easy analysis and visualization.
- Tasks Performed:
  - Standardized the Name column using title case.
  - Split combined author names into separate First Name and Last Name columns.
  - Standardized the ReleaseDate format to DD-MM-YYYY.
  - Converted time entries (e.g., "2 hrs and 20 mins") into a duration format.
  - Ensured all entries in the Price column were numeric and formatted to two decimal places.
  - Converted text-based ratings in the Stars column into numeric values.
  - Split multiple narrators listed in the NarratedBy column into separate columns.
  - Created a new column ReleaseInfo by merging ReleaseDate and Language in the format: "DD-MM-YYYY, Language".
- Outcome:
  - A clean, structured, and analysis-ready dataset was produced, enabling easier reporting, visualization, and insights generation.

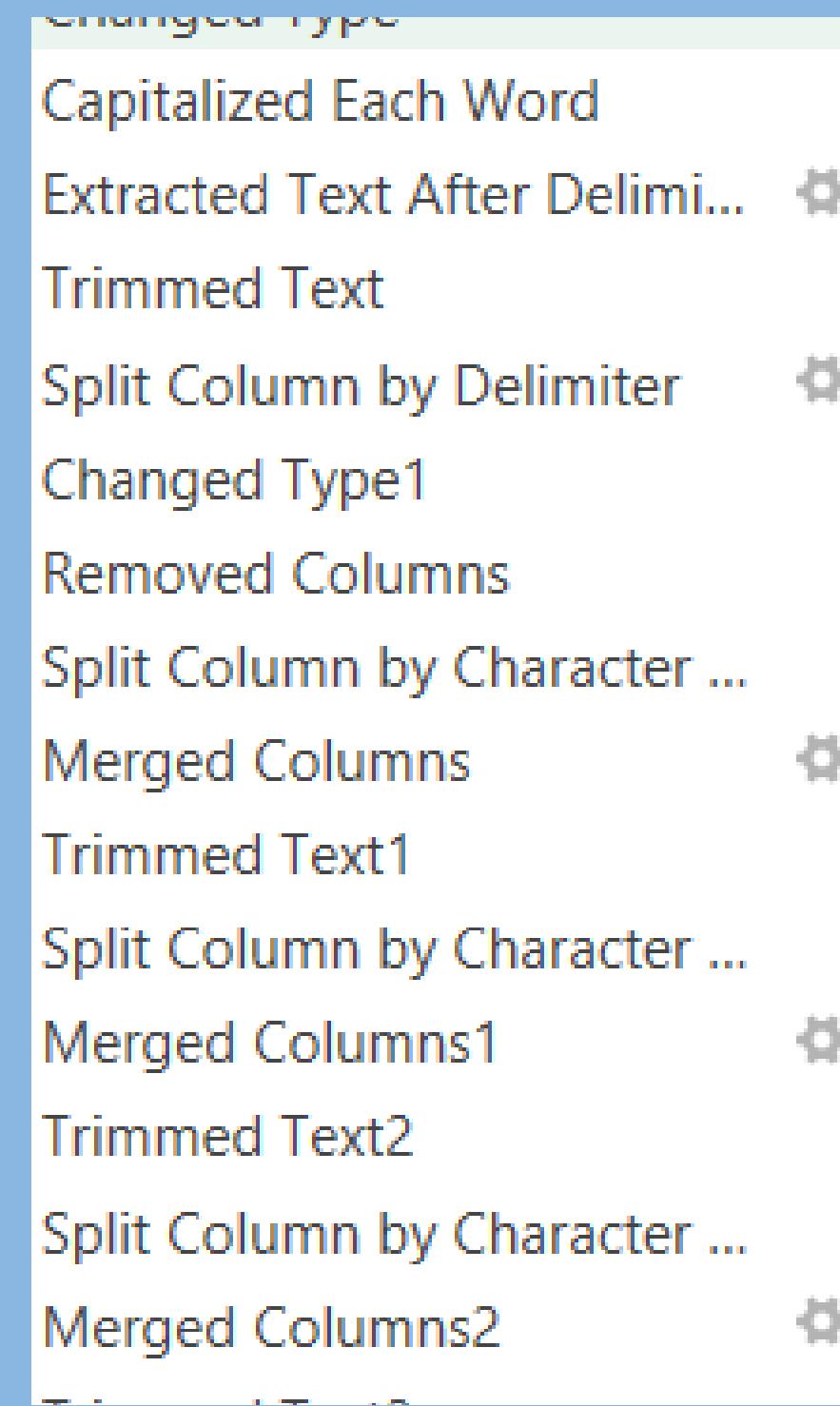




# Cleaning process

O1. ♦

## Standardize the name column to ensure consistent title casing.



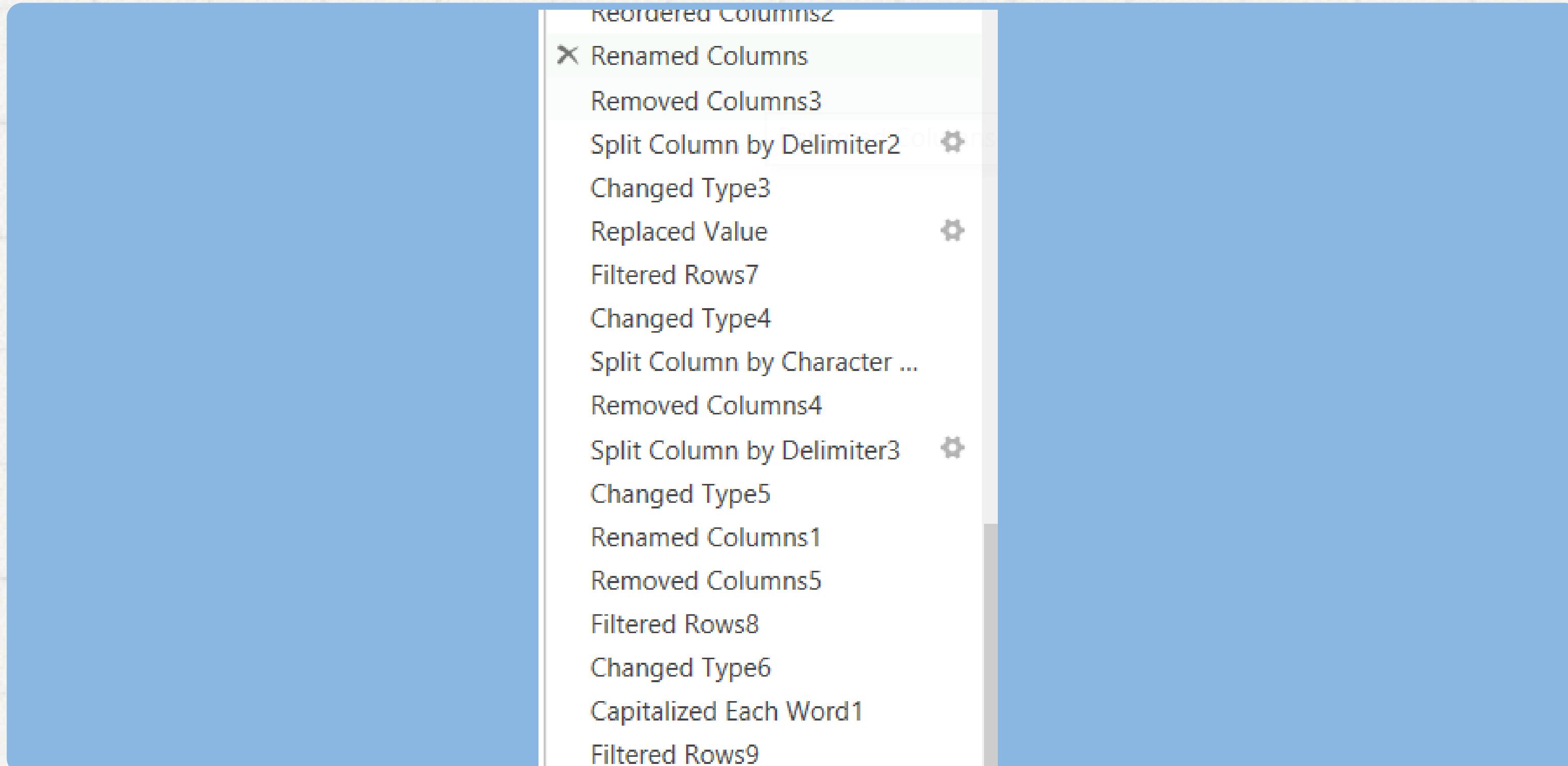
02. ♦

**Separate combined first and last names in the author column if they are currently combined.**

- Merged Columns1
- Trimmed Text2
- Split Column by Character ...
- Merged Columns2
- Trimmed Text3
- Filtered Rows
- Extracted Text After Delimi...
- Split Column by Delimiter1
- Changed Type2
- Split Column by Character ...
- Merged Columns3
- Trimmed Text4
- Filtered Rows1
- Split Column by Character ...
- Merged Columns4
- Filtered Rows2
- Trimmed Text5
- Split Column by Character ...

03. ♦

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



A screenshot of a data processing interface, likely from a tool like Microsoft Power BI or similar. The interface shows a list of operations performed on a dataset. The operations listed are:

- Reordered Columns
- Renamed Columns
- Removed Columns
- Split Column by Delimiter
- Changed Type
- Replaced Value
- Filtered Rows
- Changed Type
- Split Column by Character
- Removed Columns
- Split Column by Delimiter
- Changed Type
- Renamed Columns
- Removed Columns
- Filtered Rows
- Changed Type
- Capitalized Each Word
- Filtered Rows



# Adding a “mins”Custom Column

## Custom Column

Add a column that is computed from the other columns.

New column name

mins

Custom column formula ⓘ

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

Available columns

name
Author-1 Full Name
Author-2 Full Name
Author-3 Full Name
Narrator-1
Narrator-2
Narrator-3
...

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

# Adding a “hr” Custom Column



## Custom Column

Add a column that is computed from the other columns.

New column name

hrs

Custom column formula ⓘ

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

[Learn about Power Query formulas](#)

Available columns

name
Author-1 Full Name
Author-2 Full Name
Author-3 Full Name
Narrator-1
Narrator-2
Narrator-3

[<< Insert](#)

# Cleaned Dataset

Name	Author-1 Full Name	Author-2	Author-3	Narrator-1	Narrator-2	Narrator-3	time	Standardize Time	releasedate	language	stars.1	No.of people	price	ReleaseInfo
The Geronimo Stilton	Geronimo Stilton			Bill Lobley			2 hrs and 20 mins	140	04-08-2008	English	5	34	468	04-08-2008, Eng
Burning Maze	Rick Riordan			Robbie Daymond			13 hrs and 8 mins	788	01-05-2018	English	4.5	41	820	01-05-2018, Eng
Deep End	Jeff Kinney			Dan Russell			2 hrs and 3 mins	123	06-11-2020	English	4.5	38	410	06-11-2020, Eng
daughter Of The	Rick Riordan			Soneela Nankani			11 hrs and 16 mins	676	05-10-2021	English	4.5	12	615	05-10-2021, Eng
Lightning Thie	Rick Riordan			Jesse Bernstein			10 hrs	600	13-01-2010	English	4.5	181	820	13-01-2010, Eng
Hunger Game	Suzanne Collins			Tatiana Maslany			10 hrs and 35 mins	635	30-10-2018	English	5	72	656	30-10-2018, Eng
st For The Dia	Winter Morgan			Luke Daniels			2 hrs and 23 mins	143	25-11-2014	English	5	11	233	25-11-2014, Eng
Dark Prophe	Rick Riordan			Robbie Daymond			12 hrs and 32 mins	752	02-05-2017	English	5	50	820	02-05-2017, Eng
lin Mission Co	Mary Pope Osborne			Mary Pope Osborne			10 hrs and 56 mins	656	02-05-2017	English	5	5	1256	02-05-2017, Eng
Tyrant'S Tom	Rick Riordan			Robbie Daymond			13 hrs and 22 mins	802	24-09-2019	English	5	58	820	24-09-2019, Eng
Titan'S Curse:	Rick Riordan			Jesse Bernstein			8 hrs and 48 mins	528	14-01-2010	English	4.5	130	820	14-01-2010, Eng
gic Tree House	Mary Pope Osborne			Mary Pope Osborne			5 hrs and 23 mins	323	24-08-2011	English	5	6	1206	24-08-2011, Eng
gic Tree House	Mary Pope Osborne			Mary Pope Osborne			6 hrs and 1 min	361	27-09-2011	English	5	7	1206	27-09-2011, Eng
gnus Chase An	Rick Riordan			Michael Crouch			12 hrs and 58 mins	778	03-10-2017	English	5	41	820	03-10-2017, Eng
thern Lights	Philip Pullman			Philip Pullman	fullcast	Ruth Wilson	11 hrs and 55 mins	715	24-06-2021	English	4	2	1093	24-06-2021, Eng
onimo Stilton	Geronimo Stilton			Bill Lobley			2 hrs and 25 mins	145	08-02-2008	English	4.5	33	467	08-02-2008, Eng
gic Tree House	Mary Pope Osborne			Mary Pope Osborne			5 hrs and 4 mins	304	26-12-2004	English	5	7	1206	26-12-2004, Eng
e	Shannon Messenger			Caitlin Kelly			14 hrs and 41 mins	881	06-11-2018	English	5	20	836	06-11-2018, Eng
lin Mission Co	Mary Pope Osborne			Mary Pope Osborne			10 hrs and 18 mins	618	02-05-2017	English	5	11	1256	02-05-2017, Eng
erseen	Shannon Messenger			Caitlin Kelly			16 hrs and 42 mins	1002	06-11-2018	English	5	13	1003	06-11-2018, Eng
Tower Of Ner	Rick Riordan			Robbie Daymond			12 hrs and 12 mins	732	06-10-2020	English	5	79	615	06-10-2020, Eng
est: The Inherit	Christopher Paolini			Gerrard Doyle			23 hrs and 29 mins	1409	14-04-2009	English	4.5	47	957	14-04-2009, Eng
emis Fowl	Eoin Colfer			Gerry O'Brien			6 hrs and 46 mins	406	11-04-2013	English	5	27	683	11-04-2013, Eng



# Thank you very much!

[Link of Cleaned Audible  
Dataset](#)

[https://docs.google.com/spreadsheets/d/1jqOivhT6UyP28lx6mDtRSJC5leimjsN/edit?  
usp=sharing&ouid=110419981726357331255&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1jqOivhT6UyP28lx6mDtRSJC5leimjsN/edit?usp=sharing&ouid=110419981726357331255&rtpof=true&sd=true)