

# PROJECT PRESENTATION

**Title: Cleaning and Standardizing an Audible Dataset  
using Power Query Editor in Excel**

**BY MOHIT SINGH**

# Project Overview

## OBJECTIVE

The aim of this project is to clean and standardize an Audible dataset using Power Query Editor in Excel to prepare it for further analysis. This involves ensuring data consistency, formatting columns, and handling various data quality issues.

# Dataset Description

**Data Source:** Audible dataset containing information about audiobooks.

**Columns:** The dataset includes columns such as Name, Author, Releasedate, Time, Price, Stars, Narratedby, and Language.

**Challenges:** The dataset has inconsistencies in formatting, combined fields (e.g., Author names), non-numeric values in numeric fields, and multiple values in single cells (e.g., Narratedby).

# Tools and Methodology.

**Tool Used:** Power Query Editor in Excel.

**Process Overview:** The process involved several steps to clean and standardize the data:

1. Standardizing the Name column.
2. Separating combined first and last names in the Author column.
3. Ensuring consistent date format in the Releasedate column.
4. Converting the Time column to a recognized duration format.
5. Ensuring the Price column is in a numeric format.
6. Converting text ratings to numeric values.
7. Splitting the Narratedby column if multiple narrators are listed.
8. Merging Releasedate and Language columns into a new Releaseinfo column.
9. Formatting the Price column consistently with two decimal places.

# RAW DATA

name	author	narrator	time	releasedate	language	stars	price
The Geronimo Stilton	Written by: Geronimo Stilton	Narrated by: Bill Lobley	2 hrs and 20 mins	04-08-2008	English	5 out of 5	468
The Burn	Written by: Rick Riordan	Narrated by: Robbie Daymond	13 hrs and 8 mins	01-05-2018	English	4.5 out of 5	820
The Deep	Written by: Jeff Kinney	Narrated by: Dan Russell	2 hrs and 3 mins	06-11-2020	English	4.5 out of 5	410
Daughter	Written by: Rick Riordan	Narrated by: Soneela Nankani	11 hrs and 16 mins	05-10-2021	English	4.5 out of 5	615
The Lightning Thief	Written by: Rick Riordan	Narrated by: Jesse Bernstein	10 hrs	13-01-2010	English	4.5 out of 5	820
The Hunger Games	Written by: Suzanne Collins	Narrated by: Tatiana Maslany	10 hrs and 35 mins	30-10-2018	English	5 out of 5	656
Quest for the Golden Compass	Written by: Philip Pullman	Narrated by: Luke Daniels	2 hrs and 23 mins	25-11-2014	English	5 out of 5	233
The Dark Prophecy	Written by: Rick Riordan	Narrated by: Robbie Daymond	12 hrs and 32 mins	02-05-2017	English	5 out of 5	820
Merlin	Written by: Mary Pope Osborne	Narrated by: Mary Pope Osborne	10 hrs and 56 mins	02-05-2017	English	5 out of 5	1,256.00
The Tyrant	Written by: Rick Riordan	Narrated by: Robbie Daymond	13 hrs and 22 mins	24-09-2019	English	5 out of 5	820
The Titan's Curse	Written by: Rick Riordan	Narrated by: Jesse Bernstein	8 hrs and 48 mins	14-01-2010	English	4.5 out of 5	820
Magic Tree House	Written by: Mary Pope Osborne	Narrated by: Mary Pope Osborne	5 hrs and 23 mins	24-08-2011	English	5 out of 5	1,206.00
Magic Tree House	Written by: Mary Pope Osborne	Narrated by: Mary Pope Osborne	6 hrs and 1 min	27-09-2011	English	5 out of 5	1,206.00
Magnus Chase	Written by: Rick Riordan	Narrated by: Michael Crouch	12 hrs and 58 mins	03-10-2017	English	5 out of 5	820
Northern Lights	Written by: Philip Pullman	Narrated by: Philip Pullman, full cast, Ruth Wilcock	11 hrs and 55 mins	24-06-2021	English	4 out of 5	1,093.00
Geronimo Stilton	Written by: Geronimo Stilton	Narrated by: Bill Lobley	2 hrs and 25 mins	08-02-2008	English	4.5 out of 5	467
Magic Tree House	Written by: Mary Pope Osborne	Narrated by: Mary Pope Osborne	5 hrs and 4 mins	26-12-2004	English	5 out of 5	1,206.00
Exile	Written by: Shannon Messenger	Narrated by: Caitlin Kelly	14 hrs and 41 mins	06-11-2018	English	5 out of 5	836
Merlin	Written by: Mary Pope Osborne	Narrated by: Mary Pope Osborne	10 hrs and 18 mins	02-05-2017	English	5 out of 5	1,256.00
Neverseen	Written by: Shannon Messenger	Narrated by: Caitlin Kelly	16 hrs and 42 mins	06-11-2018	English	5 out of 5	1,003.00
The Tower	Written by: Rick Riordan	Narrated by: Robbie Daymond	12 hrs and 12 mins	06-10-2020	English	5 out of 5	615
Eldest	Written by: Christopher Paolini	Narrated by: Gerrard Doyle	23 hrs and 29 mins	14-04-2009	English	4.5 out of 5	957
Artemis Fowl	Written by: Eoin Colfer	Narrated by: Gerry O'Brien	6 hrs and 46 mins	11-04-2013	English	5 out of 5	683
Geronimo Stilton	Written by: Geronimo Stilton	Narrated by: Bill Lobley	2 hrs and 37 mins	01-02-2011	English	4.5 out of 5	469
Percy Jackson	Written by: Rick Riordan	Narrated by: Jesse Bernstein	10 hrs and 32 mins	21-01-2010	English	5 out of 5	820
Winnie-the-Pooh	Written by: A.A. Milne	Narrated by: Peter Dennis	2 hrs and 46 mins	28-12-2003	English	5 out of 5	468
Magic Tree House	Written by: Mary Pope Osborne	Narrated by: Mary Pope Osborne	7 hrs and 9 mins	01-09-2020	English	5 out of 5	1,256.00
Geronimo Stilton	Written by: Geronimo Stilton	Narrated by: Bill Lobley	2 hrs and 22 mins	01-10-2010	English	5 out of 5	469

# 1. Standardizing the Name Column

**Problem:** The Name column had inconsistent casing.

**Solution:** Applied the "Capitalize Each Word" function.

**Steps:**

1. Selected the Name column.
2. Navigated to the "Transform" tab.
3. Used the "Capitalize Each Word" option to ensure consistent title casing.

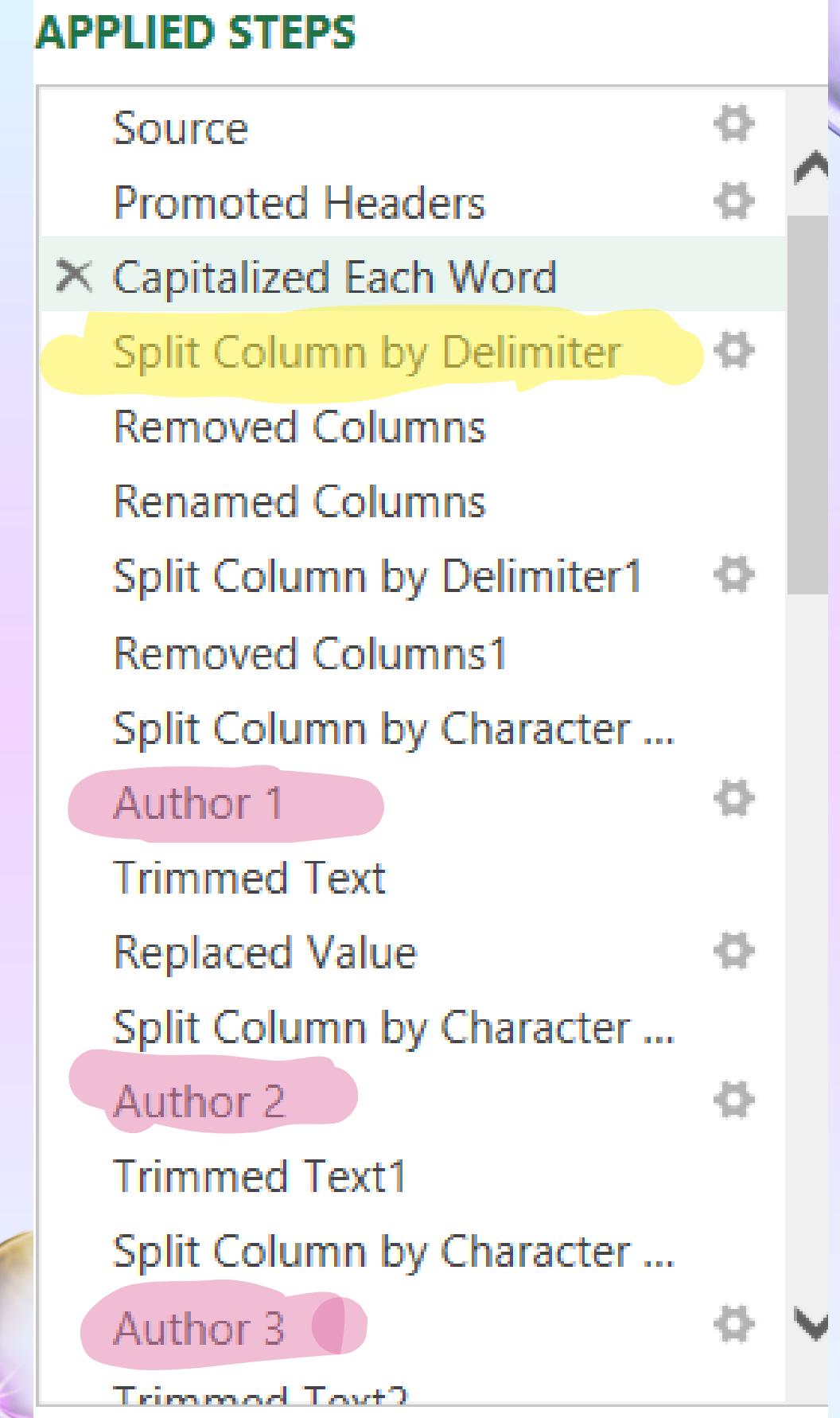
# 2. Separating Combined First and Last Names in the Author Column.

**Problem:** Author names were combined in a single column..

**Solution:** Split the column by delimiter (space).

**Steps:**

1. Selected the Author column.
2. Went to the "Transform" tab.
3. Chose "Split Column" > "By Delimiter" and selected space as the delimiter.



### **3. Ensuring Consistent Date Format in the Releasedate Column**

**Problem:** Inconsistent date formats in the Releasedate column..

**Solution:** Converted the column to Date type and ensured the format was "DD-MM-YYYY".

#### **Steps:**

1. Selected the Releasedate column.
2. Went to the "Transform" tab.
3. Set the Data Type to Date.
4. Used the "Format" option to ensure the date format was consistent.

## APPLIED STEPS

Source	
Promoted Headers	
<del>Capitalized Each Word</del>	
Split Column by Delimiter	
Removed Columns	
Renamed Columns	
Split Column by Delimiter1	
Removed Columns1	
Split Column by Character ...	
Author 1	
Trimmed Text	
Replaced Value	
Split Column by Character ...	
Author 2	
Trimmed Text1	
Split Column by Character ...	
Author 3	
Trimmed Text2	

## APPLIED STEPS

Trimmed Text2	
Split Column by Delimiter2	
Removed Columns2	
Renamed Columns1	
Split Column by Delimiter3	
Filtered Rows	
Removed Columns7	
Split Column by Character ...	
Narrator 1	
Trimmed Text3	
Split Column by Character ...	
Narrator 2	
Trimmed Text4	
Filtered Rows2	
Split Column by Character ...	
Narrator 3	
Trimmed Text5	

## APPLIED STEPS

Split Column by Character ...	
Narrator 3	
Trimmed Text5	
Filtered Rows1	
Hours	
Mnutes	
Duration	
Removed Columns3	
Replaced Value1	
Changed Type	
Ratings	
Reordered Columns	
Extracted Text Before Deli...	
Replaced Value2	
Changed Type1	
Replaced Value3	
Changed Type2	
<del>Inserted Merged Column</del>	
Inserted Merged Column	

These are all applied steps, now show you applied formula in next slide.

### Custom Column

Add a column that is computed from the other columns.

New column name  
Hours

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
Narrator 3

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

### 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 0`

Available columns

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

<< Insert

Learn about Power Query formulas

✖ Selected.

OK Cancel

### Custom Column

Add a column that is computed from the other columns.

New column name  
Duration

Custom column formula ⓘ  
`= [Hours]*60+[Minutes]`

Available columns

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

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK Cancel

Add a column that is computed from the other columns.

New column name

Ratings

Custom column formula (i)

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

Available columns

name
Author 1
Author 2
Author 3
Narrator 1
Narrator 2
Narrator 3
...

<< Insert

[Learn about Power Query formulas](#)

A	B	C	D
name	Author 1	Author 2	Author 3
The Tales Of Beedle The Bard	J.K.Rowling	Not Applicable	Not Applicable
Daisy Jones And The Six	Taylor Jenkins Reid	Not Applicable	Not Applicable
Tools Of Titans	Tim Ferriss	Not Applicable	Not Applicable
Dune	Frank Herbert	Not Applicable	Not Applicable
Measure What Matters	John Doerr	Not Applicable	Not Applicable
The Handmaid'S Tale	Margaret Atwood	Not Applicable	Not Applicable
Cosmos	Carl Sagan	Not Applicable	Not Applicable
The Testaments	Margaret Atwood	Not Applicable	Not Applicable
The Sandman: Act II	Neil Gaiman	Dirk Maggs	Not Applicable
Six Of Crows	Leigh Bardugo	Not Applicable	Not Applicable
Crooked Kingdom	Leigh Bardugo	Not Applicable	Not Applicable
Norse Mythology	Neil Gaiman	Not Applicable	Not Applicable
Sadie	Courtney Summers	Not Applicable	Not Applicable
The Christmas Pig	J.K.Rowling	Not Applicable	Not Applicable
Heads Will Roll	Kate Mc Kinnon	Emily Lynne	Not Applicable
Me Before You	Jojo Moyes	Not Applicable	Not Applicable
World War Z	Max Brooks	Not Applicable	Not Applicable
Assassin'S Creed: Gold	Anthony Del Col	Not Applicable	Not Applicable
Interview With The Robot	Lee Bacon	Not Applicable	Not Applicable
Sherlock Holmes: The Voice Of Treason	George Mann	Cavan Scott	Not Applicable
One Of Us Is Lying	Karen M.Mc Manus	Not Applicable	Not Applicable
Sleeping Giants	Sylvain Neuvel	Not Applicable	Not Applicable
Hyperion	Dan Simmons	Not Applicable	Not Applicable
Dracula [Audible Edition]	Bram Stoker	Not Applicable	Not Applicable
Dune Messiah	Frank Herbert	Not Applicable	Not Applicable
2034	Elliot Ackerman	Admiral James	Not Applicable

Narrator 1	Narrator 2	Narrator 3	time
Warwick Davis	Noma Dumezweni	Jason Isaacs	1 hr and 35 mins
Sara Arrington	Jennifer Beals	Arthur Bishop	9 hrs and 3 mins
Ray Porter	Kaleo Griffith	Tim Ferriss	22 hrs and 41 mins
Scott Brick	Orlagh Cassidy	Euan Morton	21 hrs and 2 mins
Jini Kim	Susan Wojcicki	Alex Garden	7 hrs and 56 mins
Elisabeth Moss	Bradley Whitford	Amy Landecker	11 hrs and 22 mins
LeVar Burton	Seth Mac Farlane	Neilde Grasse Tyson	14 hrs and 31 mins
Ann Dowd	Bryce Dallas Howard	Mae Whitman	13 hrs and 18 mins
Neil Gaiman	James McAvoy	Emma Corrin	13 hrs and 47 mins
Jay Snyder	Brandon Rubin	Fred Berman	15 hrs and 4 mins
Roger Clark	Jay Snyder	Elizabeth Evans	17 hrs and 58 mins
fullcast	Diana Rigg	Derek Jacobi	1 hr and 27 mins
Rebecca Soler	Fred Berman	Dan Bittner	7 hrs and 57 mins
Amaka Okafor	Rocco Padden	Gerran Howell	5 hrs and 37 mins
Kate McKinnon	Emily Lynne	Tim Gunn	4 hrs and 6 mins
Jo Hall	Anna Bentinck	Steve Crossley	16 hrs and 37 mins
Christopher Ragland	Rupert Farley	Nigel Pilkington	13 hrs and 56 mins
Riz Ahmed	John Chancer	Ray Fearon	4 hrs and 40 mins
Kevin T. Collins	Ellen Archer	Josh Hurley	3 hrs and 42 mins

releasedate	language	price	Duratin	stars	Ratings	releasedate&language
31-03-2020	English	652	95	4.5	1088	31-03-2020 English
05-03-2019	English	820	543	4.5	299	05-03-2019 English
02-06-2020	English	1506	1361	4.5	291	02-06-2020 English
31-12-2006	English	301	1262	4.5	268	31-12-2006 English
01-11-2018	English	820	476	4.5	262	01-11-2018 English
30-04-2019	English	957	682	4.5	222	30-04-2019 English
30-05-2017	English	1005	871	5	209	30-05-2017 English
10-09-2019	English	1093	798	4.5	156	10-09-2019 English
22-09-2021	English	1003	827	4.5	143	22-09-2021 English
29-09-2015	English	1003	904	4.5	137	29-09-2015 English
27-09-2016	English	1003	1078	4.5	93	27-09-2016 English
21-02-2019	English	615	87	4.5	88	21-02-2019 English
04-09-2018	English	134	477	4.5	84	04-09-2018 English
12-10-2021	English	799	337	5	81	12-10-2021 English
02-05-2019	English	1003	246	4	74	02-05-2019 English
01-01-2012	English	827	997	5	71	01-01-2012 English
13-11-2012	English	721	836	4.5	70	13-11-2012 English
27-02-2020	English	759	280	4.5	70	27-02-2020 English
02-01-2020	English	199	222	4.5	65	02-01-2020 English

# CONCLUSION

This project utilized Power Query Editor in Excel to effectively clean and standardize an Audible dataset. The process addressed formatting inconsistencies, separated combined fields, and ensured numeric values, preparing the dataset for accurate analysis.

# THANK YOU

info@wscubetech.com

+91-9024244886

[www.Ws Cube Tech.com](http://www.Ws Cube Tech.com)