

PROJECT DOCUMENTATION

Audible Dataset- Data Cleaning, Analysis & Visualization

1. PROJECT OVERVIEW

The project analyzes and explores the collection of audio books available across 36 languages, to comprehend the customer preferences with respect to audio duration and book price. The basic and advanced features of MS Excel, Power Query editor, and Power BI platforms are utilized for the data cleaning, analysis and visualization.

2. TOOLS USED

- Microsoft Excel
- Power BI

3. DATASET

- Source: Audible_uncleaned.csv
- Data contains: Name of book, author, narrator, release date, audio duration, star rating, count of ratings, price

4. STEPS FOLLOWED

Data cleaning in Excel:

- Converted the file extension from .csv to .xlsx, and saved in MS Excel
- Formatting is done for table with first column as headers
- Basic data proofreading using Remove Duplicates and Spell check options
- Capitalized column headers and modified titles
- Loaded the data to Power Query Editor for subsequent data transformations
- Cleaned the names of author and narrator in Power Query using (i) Split columns by delimiter (by space), (ii) Capitalize each word, and (iii) Merge columns by separator options
- Capitalized each word for Language column
- The null values were replaced with numerical values: (i) 0, wherever applicable and (ii) the average of book price per language was considered for the null values per language in the column 'Price'. This was executed using AVERAGEIF function in Excel.
- For 'Release date' column, the date formats were cleaned and standardized using multiple features in Excel- Text to Columns option, Concatenate function, Find and Replace option.
- Time duration was converted to minutes (numerical format) using Text to columns feature and mathematical functions in Excel.

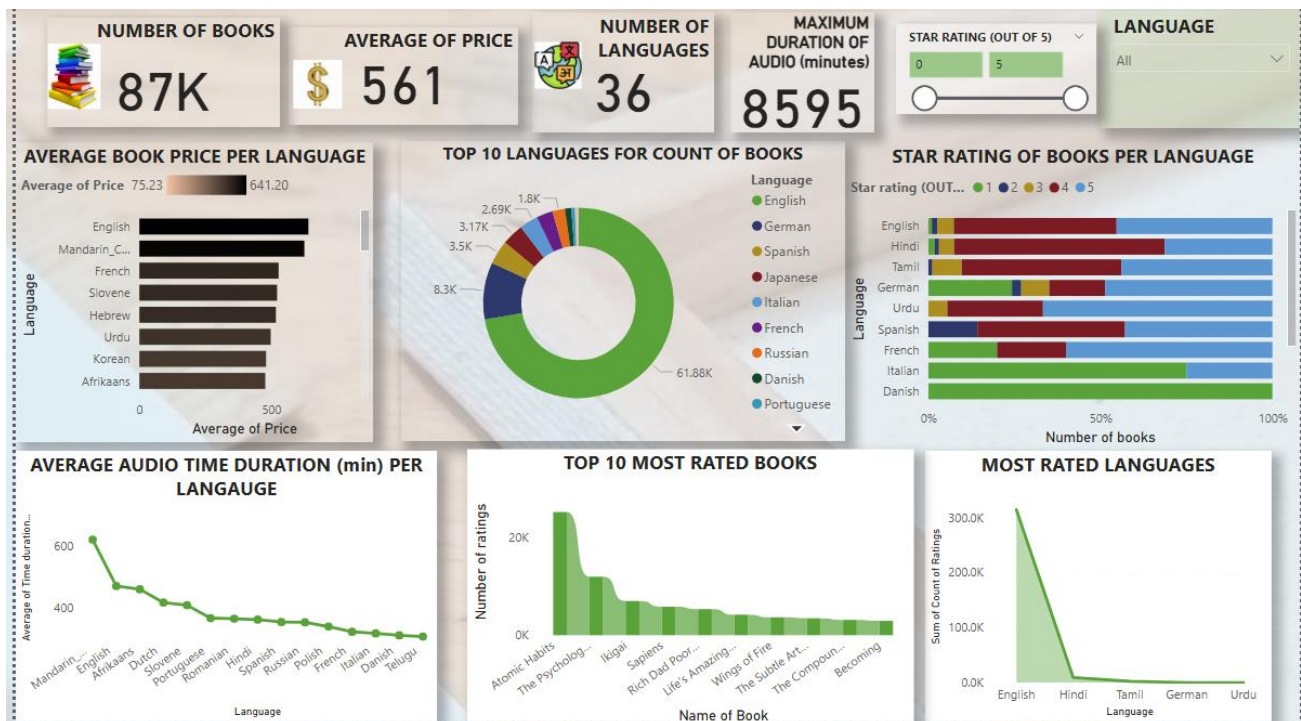
Data Visualization:

- Imported the pre-processed data into Power BI
- Built Power BI dashboard using charts (Stacked bar chart, area chart, line chart, clustered bar chart, ribbon chart, cards with new measures) and slicers.

5. KEY INSIGHTS

- The language with highest number of books (61.8K) and customer ratings across 36 languages, is English.
- The 'Atomic Habits' is the most rated book (12569 customer ratings).
- Mandarin language has the highest average audio duration for the audio books.
- The audio books available in English are the top accepted books with highest count of 5-star ratings given, though the average book price is highest for the same language.
- English, Spanish and Portuguese languages offer few cheaper audio book options (\$11 to \$18).

6. SCREENSHOTS



7. FILES INCLUDED

- Final_AUDIBLE_Pre-processed_Aswathy.xlsx – Cleaned data and basic analysis
- Visualisation_Audible.pbix – Power BI dashboard
- Project Documentation.pdf – Project description

8. HOW TO USE

- Open Final_AUDIBLE_Pre-processed_Aswathy.xlsx to view the cleaned data
- Open Visualisation_Audible.pbix in Power BI desktop to explore the visuals