

Task C

Khurush Khushrov Bengali (A0268410J)

Dataset Used: Amazon Books Dataset

<https://www.kaggle.com/datasets/chhavidhankhar11/amazon-books-dataset/data>

The following dataset has been chosen to analyze and visualize popular books and artists from Amazon Books Dataset.

Description: Highlighted fields are used in visualization.

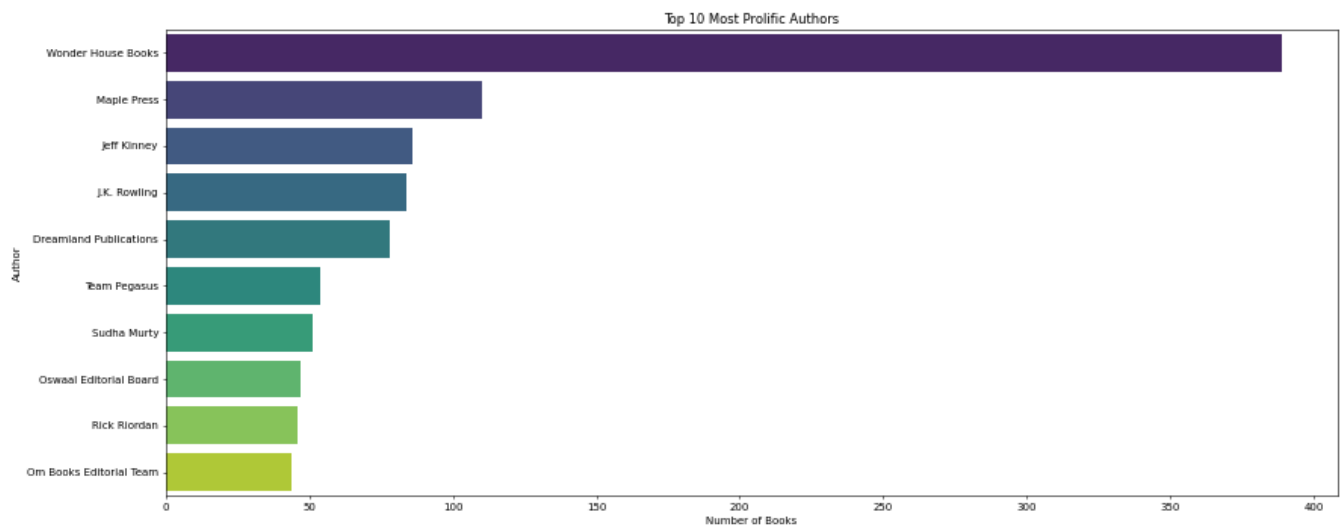
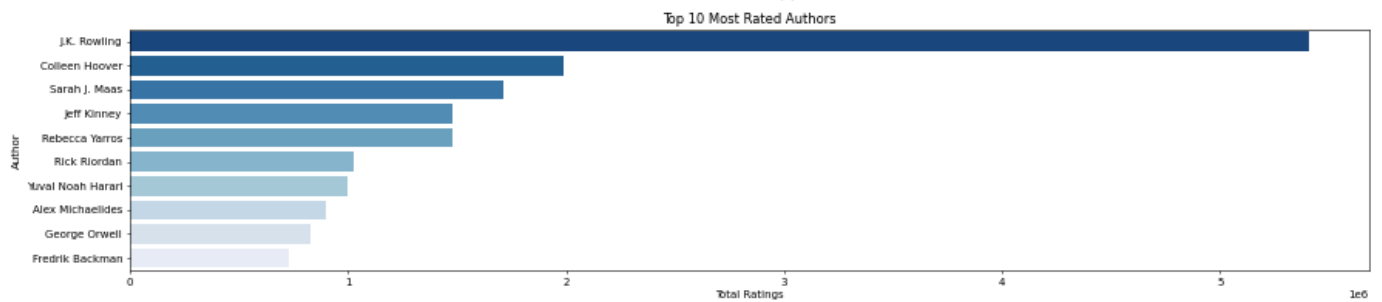
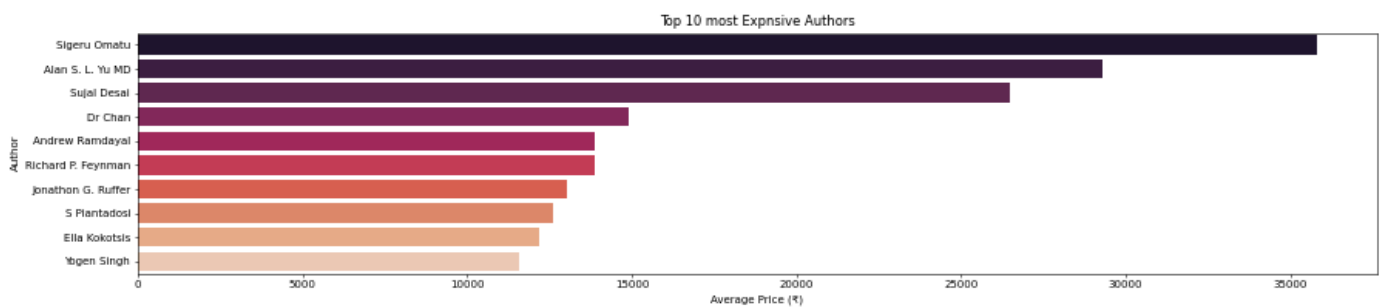
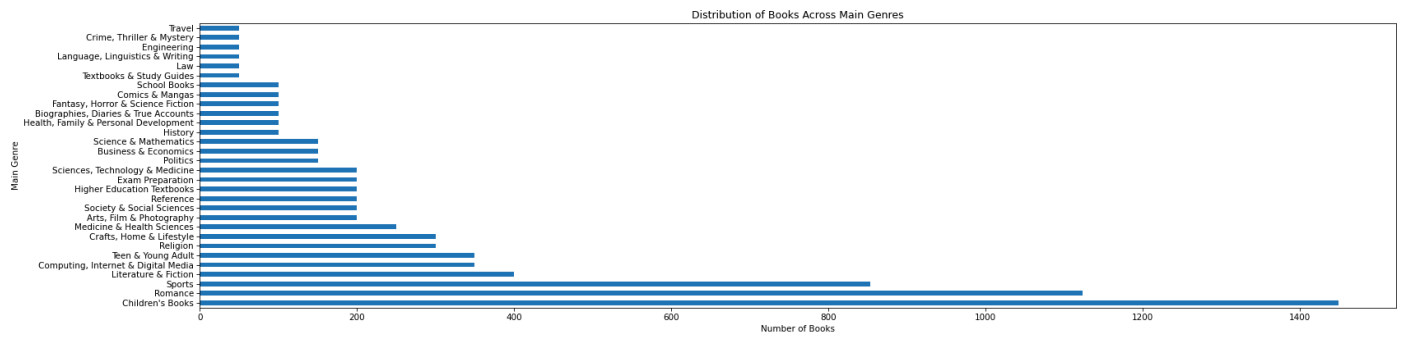
Dimension / Measure	Description	Data Type	NOIR
Title	The title of the book.	Qualitative	Nominal
Author	Name of the author or publication house.		
Main Genre	The main genre the book belongs to.		
Sub Genre	The specific sub-genre of the book.		
Type	Indicates the format of the book.		
URLs	Provides the link to the book's page on Amazon for further details and purchase options.	Quantitative	Ratio
Price	The price of the book.		
Rating	The average rating of the book given by users.		
No. of People Rated	Indicates the count of users who have rated the book.		
Price	The price of the book.		

Following are the queries for this task:

1. Distribution of books across main genre
2. Top 10 most rated authors.
3. Top 10 most expensive authors.
4. Top 10 most prolific authors.
5. Word cloud for books and genre.

Visualizations:

Source code: <https://github.com/khurushbengali/CS5346-Task-C/tree/main>



[illegible][illegible]

Visual Encodings Used:

Query	Chart Type	Visual Encodings
Distribution of Books Across Main Genres	Horizontal Bar Chart	Main Genre (y-axis) Number of Books (x-axis)
Top 10 Most Expensive Authors	Horizontal Bar Chart	Author (y-axis) Average Price (x-axis)

Top 10 Most Rated Authors	Horizontal Bar Chart	Author (y-axis) Total Ratings (x-axis)
Top 10 Most Prolific Authors	Horizontal Bar Chart	Author (y-axis) Number of Books (x-axis)
Word Cloud for Book Titles	Word Cloud	Size of word (frequency in dataset) Font color and style
Word Cloud for Genres	Word Cloud	Size of word (frequency in dataset) Font color and style

Findings and Observations:

1. Based on the distribution of books across the main genre, we find literature, sports, romance and children books in high majority.
2. Top 10 list of most expensive, most rated and most prolific authors have some common entries among them. This can indicate an author's overall success.
3. In both word clouds, terms related to genre literature, sports, romance and children books are in majority and hence in bigger font sizes.