Analysis of Amazon Top 50 Bestselling Books (2009–2022)

1. Introduction: 3
2. Dataset Overview: 3
3. Data Preprocessing: 3
4. Exploratory Data Analysis (EDA): 4
4.1 Most Frequent Authors: 4
4.2 Genre Analysis: 4
4.3 Price Distribution: 5
4.4 User Ratings: 5
4.5 Relationship Between Reviews and Ratings: 6
4.6 Top Performing Books by Ratings and Reviews: 6

Source: **Kaggle – Amazon Top 50 Bestselling Books (2009–2022)**

Prepared By: Hani Bint E Rauf – master's in data science

Tools Used: Python 3.x, Pandas, Matplotlib, Seaborn, Jupyter Notebook

1. Introduction:

The aim of this report is to analyze the Amazon Top 50 Bestselling Books dataset spanning from 2009 to 2022. The analysis focuses on identifying publishing patterns, bestselling authors, reader preferences, and price trends over the years.

2. Dataset Overview:

The dataset contains annual lists of the top 50 bestselling books on Amazon for each year between 2009 and 2022.

Key Features:

- **Name:** Book title
- **Author:** Author(s) of the book
- **User Rating:** Average customer rating on Amazon (out of 5)
- **Reviews:** Number of customer reviews
- **Price:** Price of the book in USD
- **Year:** Year in which the book ranked in the top 50
- **Genre:** Fiction or Non-Fiction

3. Data Preprocessing:

The following preprocessing steps were applied:

- Loaded dataset and checked for missing values — none were found.
- Verified data types for each column and converted as needed (e.g., ensuring numerical columns are integers/floats).
- Removed duplicate entries based on title and year to avoid repeated counting.
- Standardized column names for consistency.

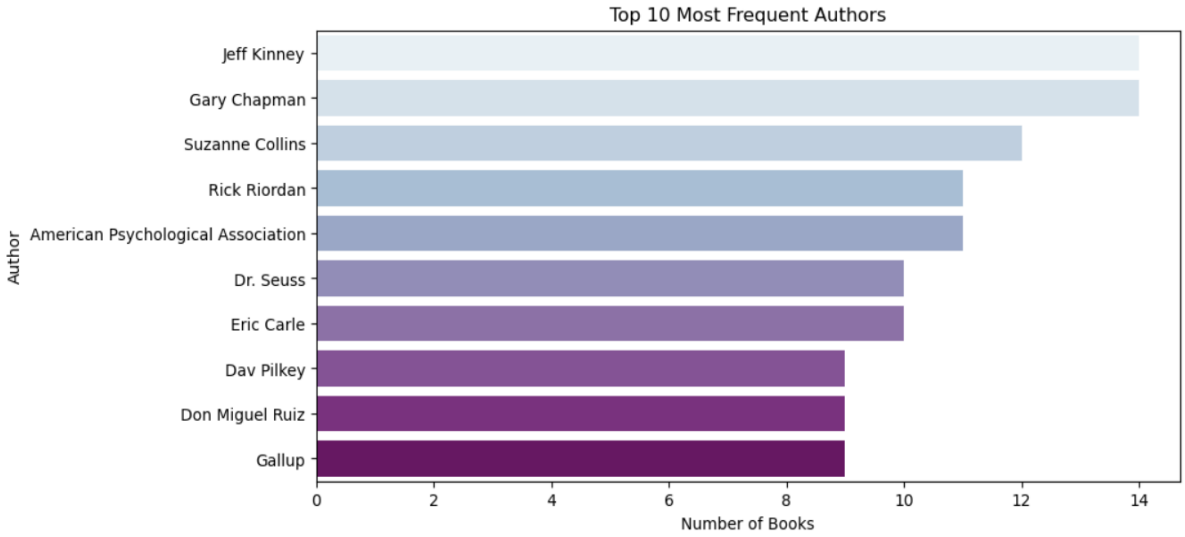
[4]:

	name	author	user_rating	reviews	price	year	genre
0	Act Like a Lady, Think Like a Man: What Men Re...	Steve Harvey	4.6	5013	17	2009	Non Fiction
1	Arguing with Idiots: How to Stop Small Minds a...	Glenn Beck	4.6	798	5	2009	Non Fiction
2	Breaking Dawn (The Twilight Saga, Book 4)	Stephenie Meyer	4.6	9769	13	2009	Fiction
3	Crazy Love: Overwhelmed by a Relentless God	Francis Chan	4.7	1542	14	2009	Non Fiction
4	Dead And Gone: A Sookie Stackhouse Novel (Sook...	Charlaine Harris	4.6	1541	4	2009	Fiction

4. Exploratory Data Analysis (EDA):

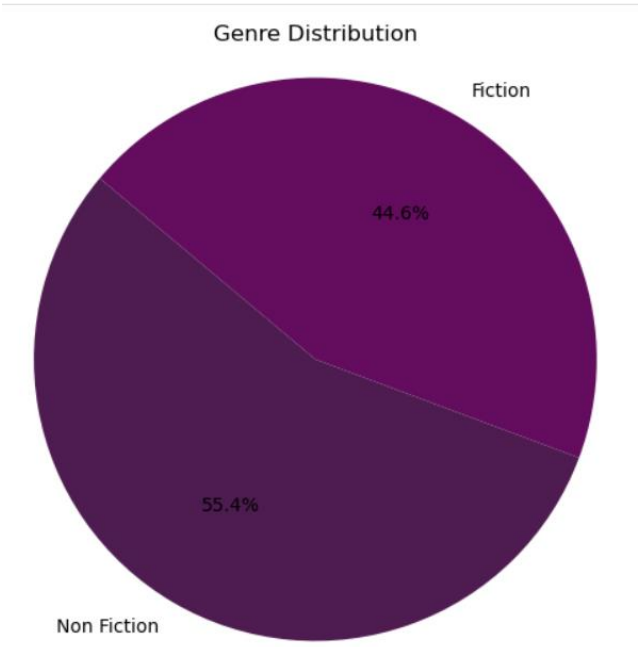
4.1 Most Frequent Authors:

I have identified authors who appeared most often in the top 50 list over the years. Authors like **Jeff Kinney**, **Gary Chapman**, and **Suzanne Collins** appear multiple times, showing consistent popularity over the years.



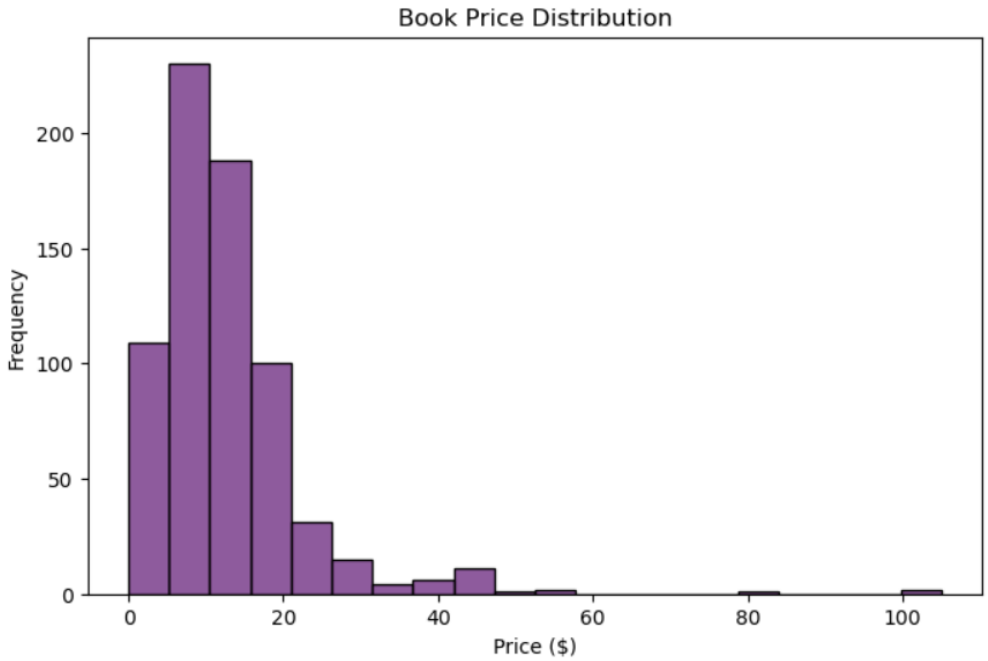
4.2 Genre Analysis:

Compared the proportion of Fiction vs. Non-Fiction books in the dataset. Non-fiction dominates the dataset, but fiction also has a strong presence, indicating reader interest in both storytelling and informative content.



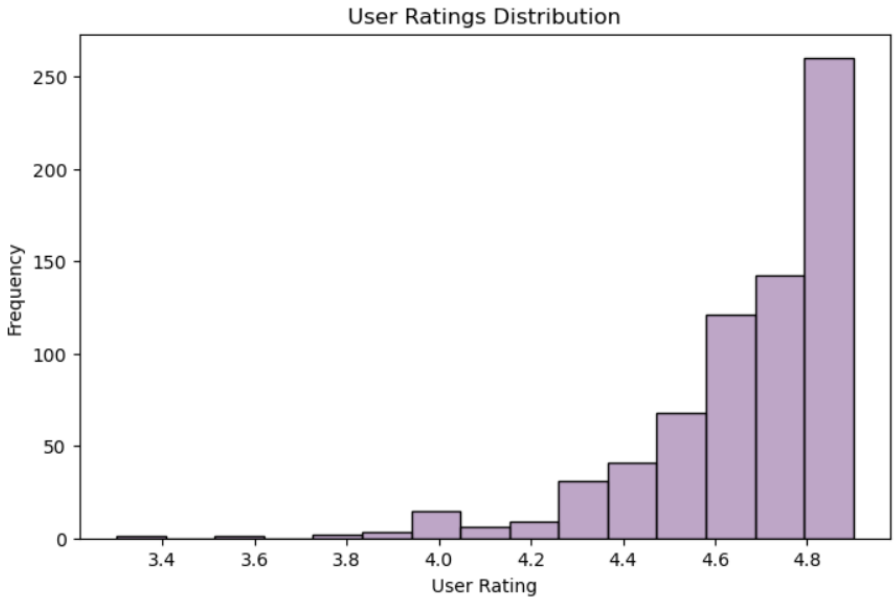
4.3 Price Distribution:

Most bestselling books are priced between 5-15. Outliers exist, but affordability seems linked to popularity.



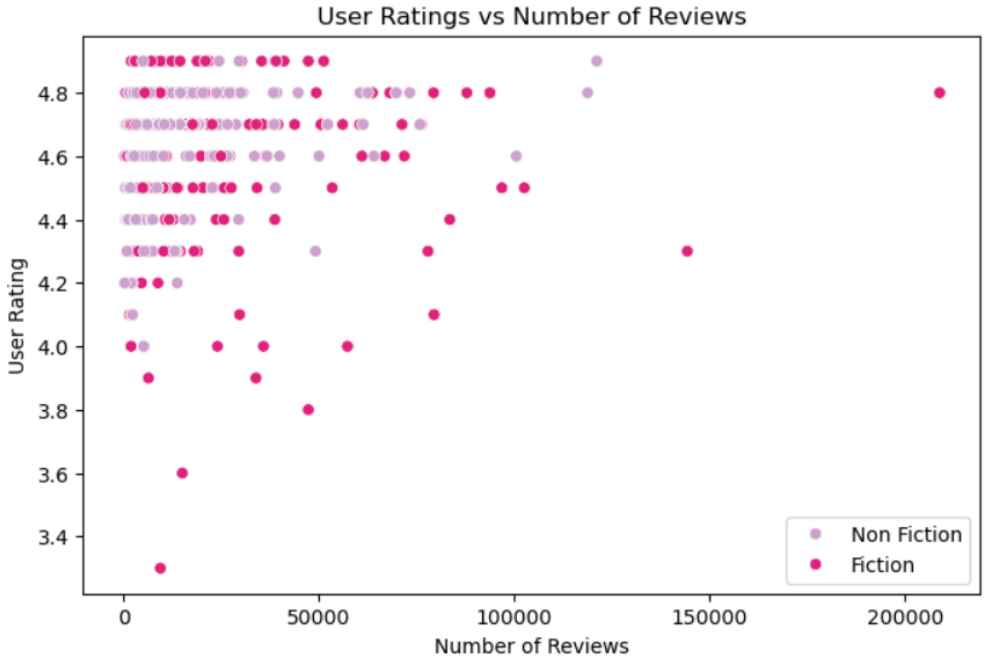
4.4 User Ratings:

I compared average user ratings over the years to check whether book quality perception changed. Almost all books have ratings above 4.0, but the number of reviews varies drastically, indicating that high ratings alone don't guarantee review volume.



4.5 Relationship Between Reviews and Ratings:

I checked if there's a relationship between the number of reviews and user ratings. Many books have very high ratings (above 4.5) but **low review counts**. Some books have both **high reviews & high ratings**, meaning they achieved **mass popularity**.



4.6 Top Performing Books by Ratings and Reviews:

I also identified the top 5 books based on ratings and reviews.

Top 5 Books by User Rating:

	Name	Author	User Rating
605	Brown Bear, Brown Bear, What Do You See?	Bill Martin Jr.	4.9
607	Call Us What We Carry: Poems	Amanda Gorman	4.9
457	Dog Man: Brawl of the Wild: From the Creator o...	Dav Pilkey	4.9
456	Dog Man and Cat Kid: From the Creator of Capta...	Dav Pilkey	4.9
223	Oh, the Places You'll Go!	Dr. Seuss	4.9

Top 5 Books by Number of Reviews:

	Name	Author	Reviews
699	Where the Crawdads Sing	Delia Owens	208917
689	The Midnight Library: A Novel	Matt Haig	144314
550	A Promised Land	Barack Obama	121109
553	Becoming	Michelle Obama	118767
589	The Silent Patient	Alex Michaelides	102544