



Amazon top 50 bestselling BOOKS

BY FATMA BEN REJEB

Why this topic ?



Books have always been a source of inspiration and knowledge for me. As an avid reader, I'm fascinated by how stories and ideas shape our lives and cultures.

Why this dataset ?

When I came across this dataset of Amazon's Top 50 Bestselling Books from 2009 to 2019, I knew I had to explore it. Not only does it reflect my passion for books, but it was also a clean dataset with a mix of numerical and categorical variables—perfect for diving into meaningful analysis





Today, I'll take you on a journey through this dataset, uncovering trends, patterns, and the secrets behind the books we love.

Overview of Books, Authors, and Genres in the Bestseller Dataset

	Name	Author	Genre
count	550	550	550
unique	351	248	2
top	Publication Manual of the American Psychological Association, 6th Edition	Jeff Kinney	Non Fiction
freq	10	12	310

Overview of Books, Authors, and Genres in the Bestseller Dataset

Total Books: The dataset contains 550 books in total, providing a broad view of Amazon's Top 50 Bestsellers from 2009 to 2019.

Unique Books: Out of the 550 books, 351 are unique titles, indicating that some books have appeared multiple times on the bestseller list over the years.

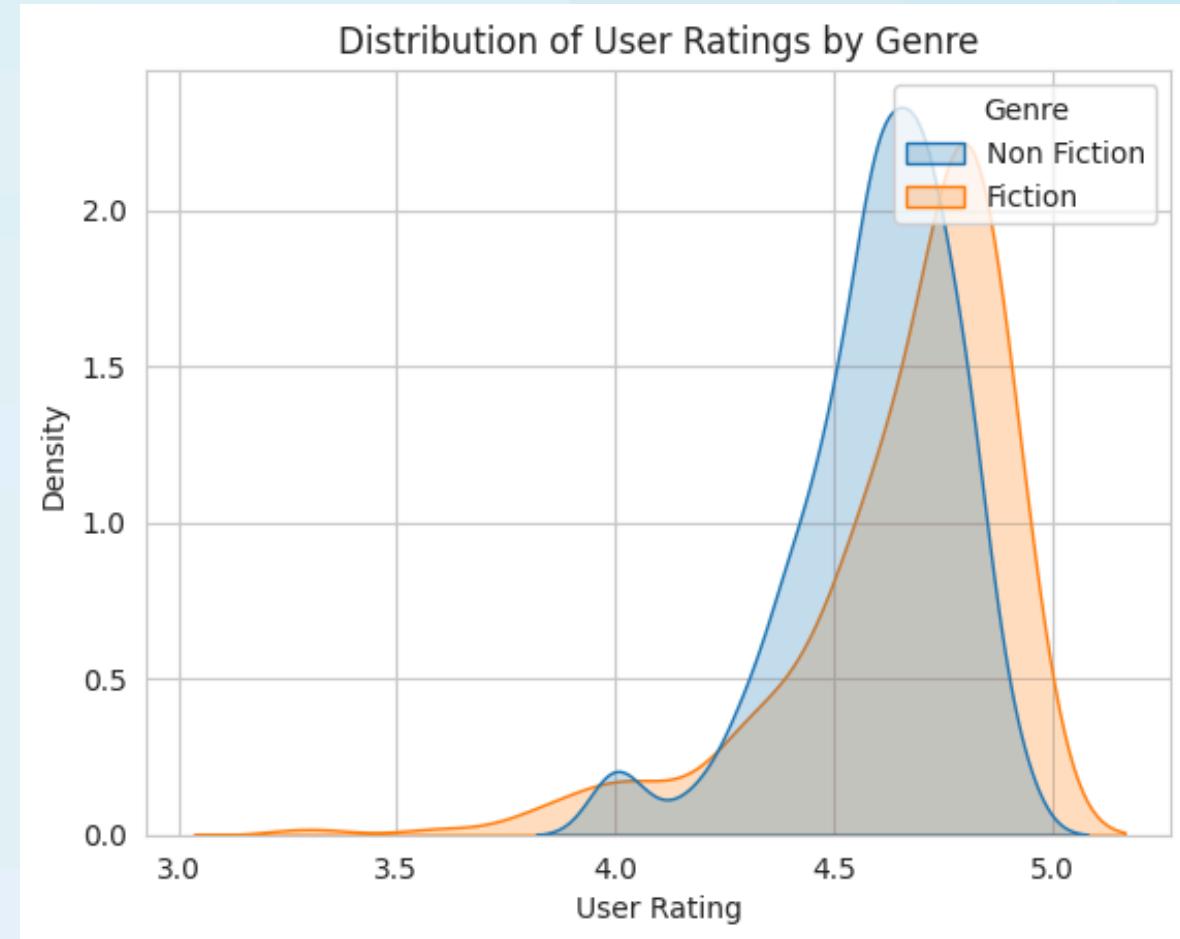
Top Author: Jeff Kinney is the most frequent author, contributing 12 books to the bestseller list, showcasing his strong presence in the market.

Genres: The dataset includes two main genres: Non-Fiction (310 books) and Fiction (240 books), with Non-Fiction being the dominant genre in the dataset.

The Story Behind the Stars: User Ratings by Genre

	Mean	Median	Mode
User Rating	4.618364	4.7	4.8

The close alignment of mean, median, and mode suggests a consistent and positive trend in customer ratings, with little variability. Readers tend to rate both fiction and non-fiction books highly.



The Power of Reviews: A Yearly Breakdown

This graph highlights how reader engagement with bestselling books has evolved over the years. The sharp rise in reviews through 2014 indicates a period of increased interest in books, potentially driven by digital accessibility and blockbuster releases. Although the growth stabilizes and fluctuates in later years, the recovery in 2019 suggests renewed enthusiasm for literary content.



What the Numbers Say: Insights from Book Reviews

Mean (11,953): On average, a book has 11,953 reviews, reflecting significant customer engagement and popularity for Amazon's bestsellers.

Median (8,580): Half of the books have fewer than 8,580 reviews, and the other half have more. The median is much lower than the mean, indicating that a few books with exceptionally high review counts are skewing the average upward.

Mode (8,580): Interestingly, 8,580 is the most common number of reviews, which likely corresponds to a specific subset of popular books.

Insight: The right-skewed distribution suggests that while some books attract a massive number of reviews (blockbusters), many others still have substantial but fewer reviews. This trend is likely influenced by factors such as genre (fiction vs. non-fiction), marketing, and long-term popularity

Exploring Book Prices: From Affordable Reads to Luxury Purchases

The price distribution of Amazon's bestselling books reveals a market heavily concentrated on affordability, with most books priced below \$20. However, the presence of a few high-priced outliers suggests a diverse market catering to both casual readers and niche audiences. This trend aligns with the idea that accessible pricing drives sales and popularity while specialty books occupy a smaller but notable segment.



Understanding Variability in Bestselling Books

	Range	Variance	Standard Deviation	IQR
User Rating	1.6	5.152009e-02	0.226980	0.30
Reviews	87804.0	1.376195e+08	11731.132017	13195.25
Price	105.0	1.175546e+02	10.842262	9.00

The variability in user ratings, reviews, and prices reveals the unique dynamics of bestselling books. Ratings are consistently high, reflecting strong reader satisfaction. However, the number of reviews varies greatly, suggesting differences in engagement levels and audience size. Prices, while generally clustered in an affordable range, show a wider spread, hinting at a market catering to both mass and niche audiences.

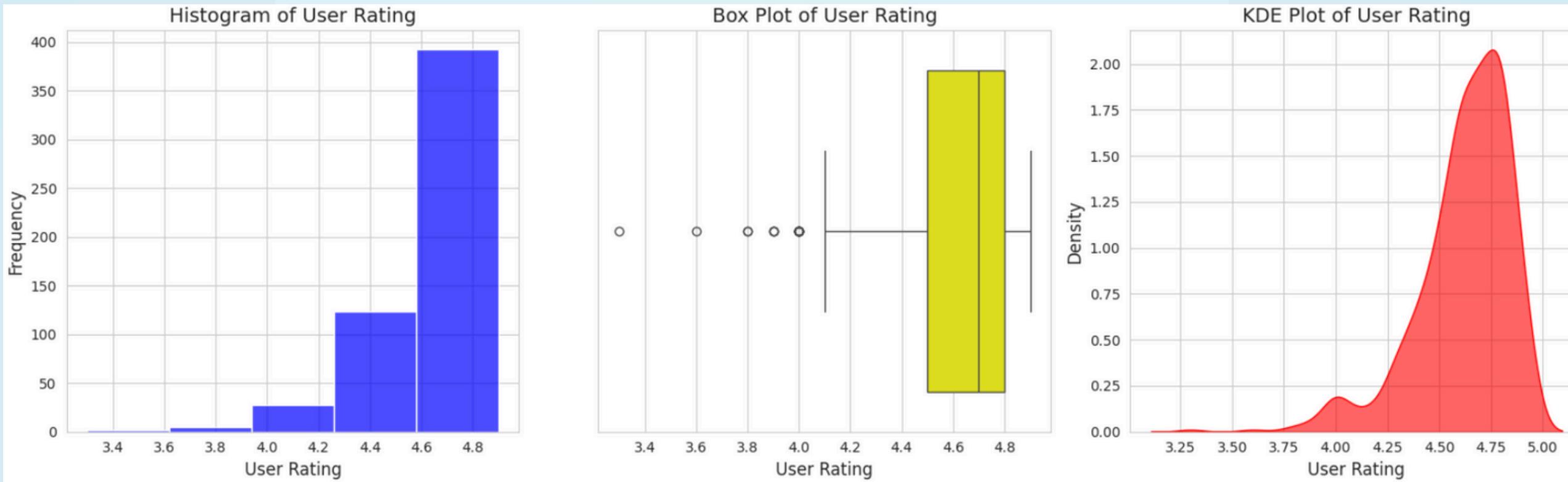
Skewed and Peaked: What the Data Tells Us About Bestsellers

The analysis of skewness and kurtosis highlights intriguing trends in Amazon's bestselling books.

Ratings are consistently high with little variability, reflecting strong customer satisfaction. Reviews show a blockbuster effect, where a small number of books dominate reader engagement. Prices are primarily affordable, though a few high-priced outliers create significant asymmetry in the distribution. These trends showcase the diversity and dynamics of the bestselling book market

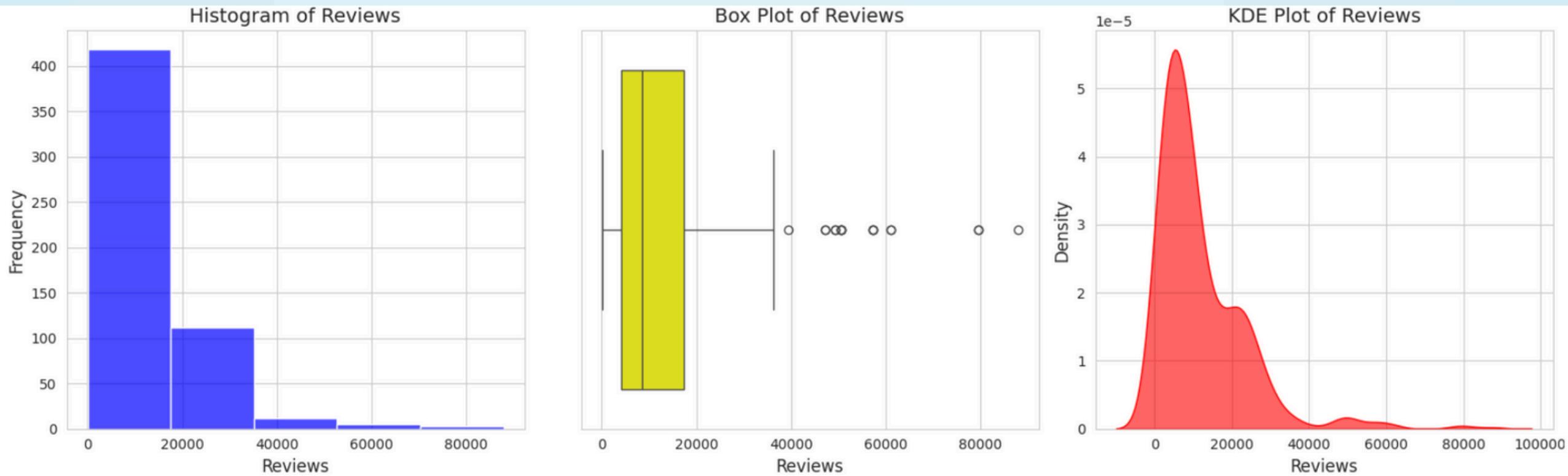
	Skewness	Kurtosis
User Rating	-1.502125	3.465375
Reviews	2.421597	8.809313
Price	3.685057	22.43352

User Rating Distribution: A Deep Dive into Rating Patterns



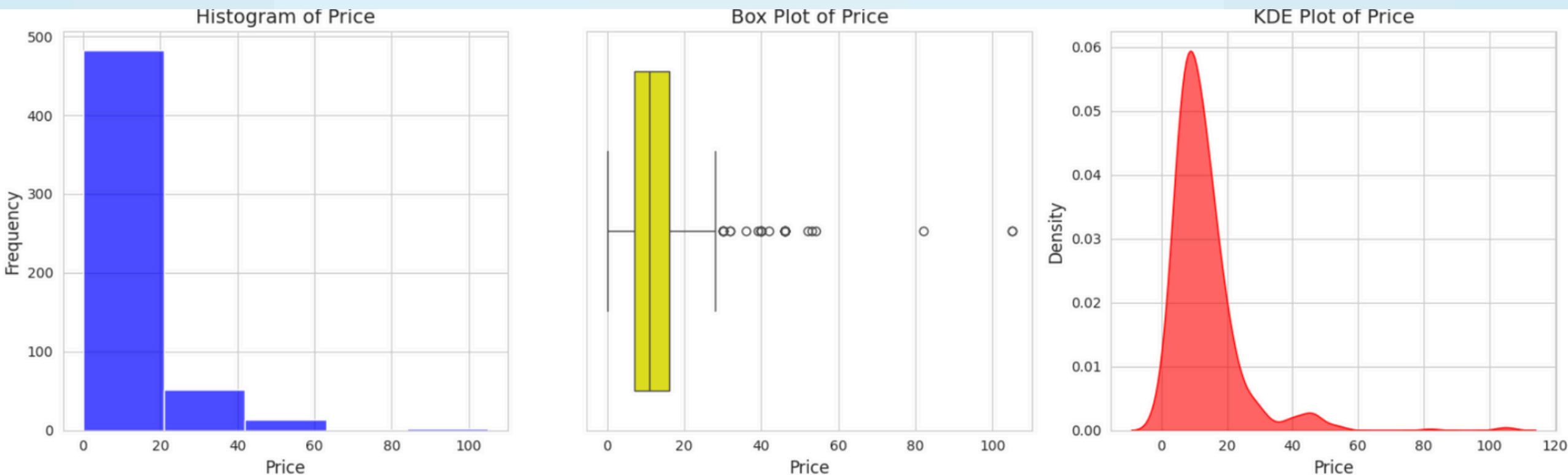
The analysis of user ratings through histograms, boxplots, and KDE plots reveals that most ratings are concentrated around higher values, reflecting the popularity and positive reception of bestselling books. However, a few lower-rated books act as outliers, creating slightly heavier tails in the distribution. This suggests that while Amazon's bestsellers are generally well-received, a small subset of books have a more polarizing impact, attracting lower ratings that stand out in comparison.

Reviews Distribution: A Few Blockbusters Drive the Majority



The distribution of reviews is right-skewed, with most books attracting a modest number of reviews, while a small number of bestsellers dominate the review count. The high kurtosis indicates a leptokurtic distribution, where extreme outliers (books with an exceptionally high number of reviews) create heavy tails. This suggests that a few books achieve massive engagement, while the majority still receive significant but much lower engagement.

Price Trends in Bestselling Books: Affordable Majority, Expensive Exceptions



The distribution of book prices is highly right-skewed, with most books falling within a lower price range, but a few expensive books drive the distribution toward higher values. The extremely high kurtosis indicates a severely leptokurtic distribution, with very heavy tails created by these expensive outliers. This suggests that while most books are priced affordably, a select few high-priced books significantly influence the overall price distribution.

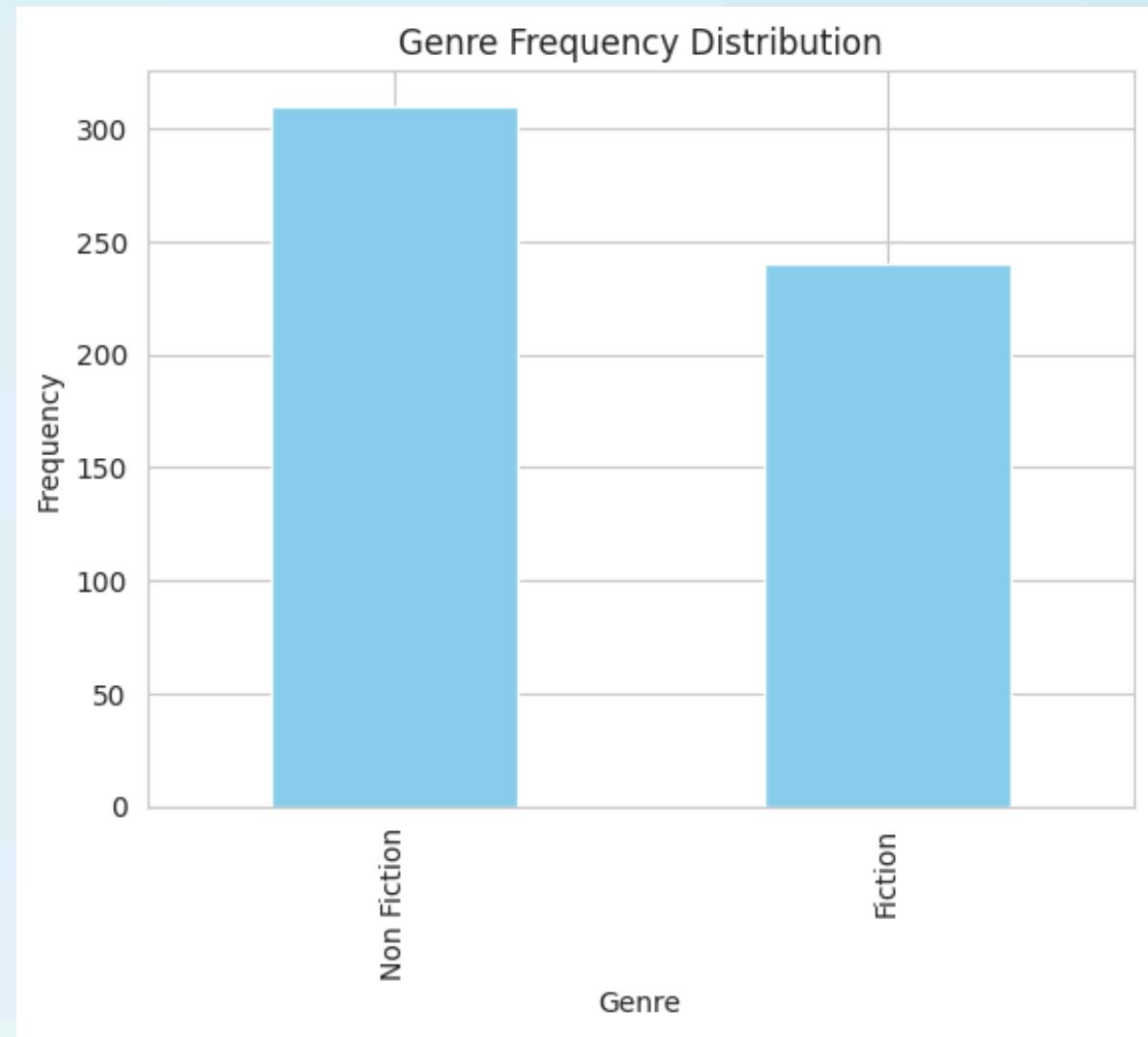
Price and Popularity: Analyzing the Correlation Between Price and Reviews

The scatter plot of Price vs. Number of Reviews shows a random distribution of data points, suggesting that there is no apparent correlation between the two variables. This means that the number of reviews a book has does not seem to influence its price. In other words, a book with a high number of reviews is not necessarily more expensive, nor is a low-priced book guaranteed to have fewer reviews. This lack of correlation highlights that price and reviews do not follow a clear relationship in this dataset.



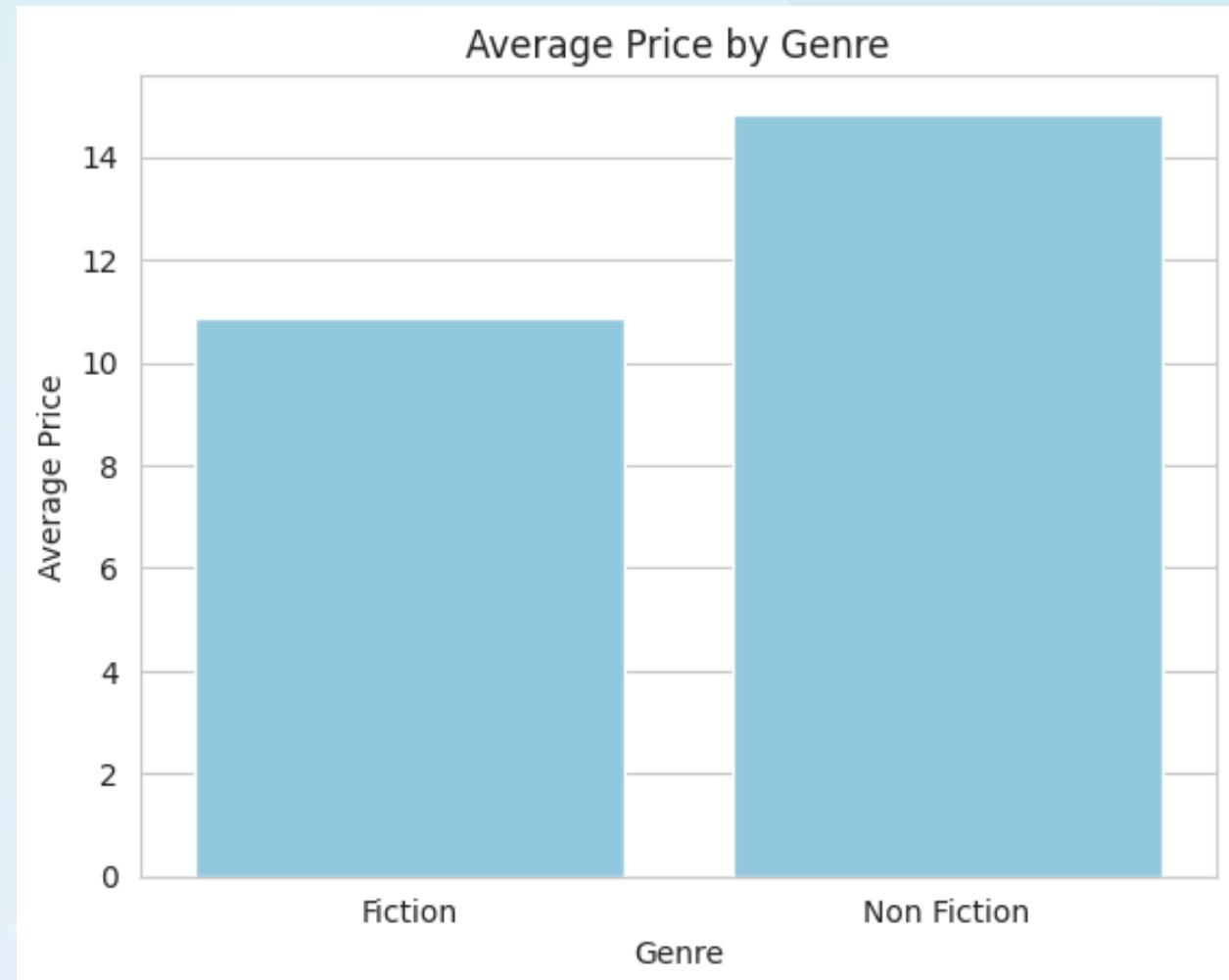
Exploring Bestseller Genres: Fiction vs. Non-Fiction

The frequency distribution of genres reveals that non-fiction books dominate the dataset, with 310 out of the 550 books classified as non-fiction. This shows that a significant portion of the bestselling books on Amazon during this period falls under the non-fiction category, which may reflect reader preferences for informational or educational content over fictional narratives. The remaining books are categorized as fiction, highlighting a diverse yet slightly lesser representation compared to non-fiction.

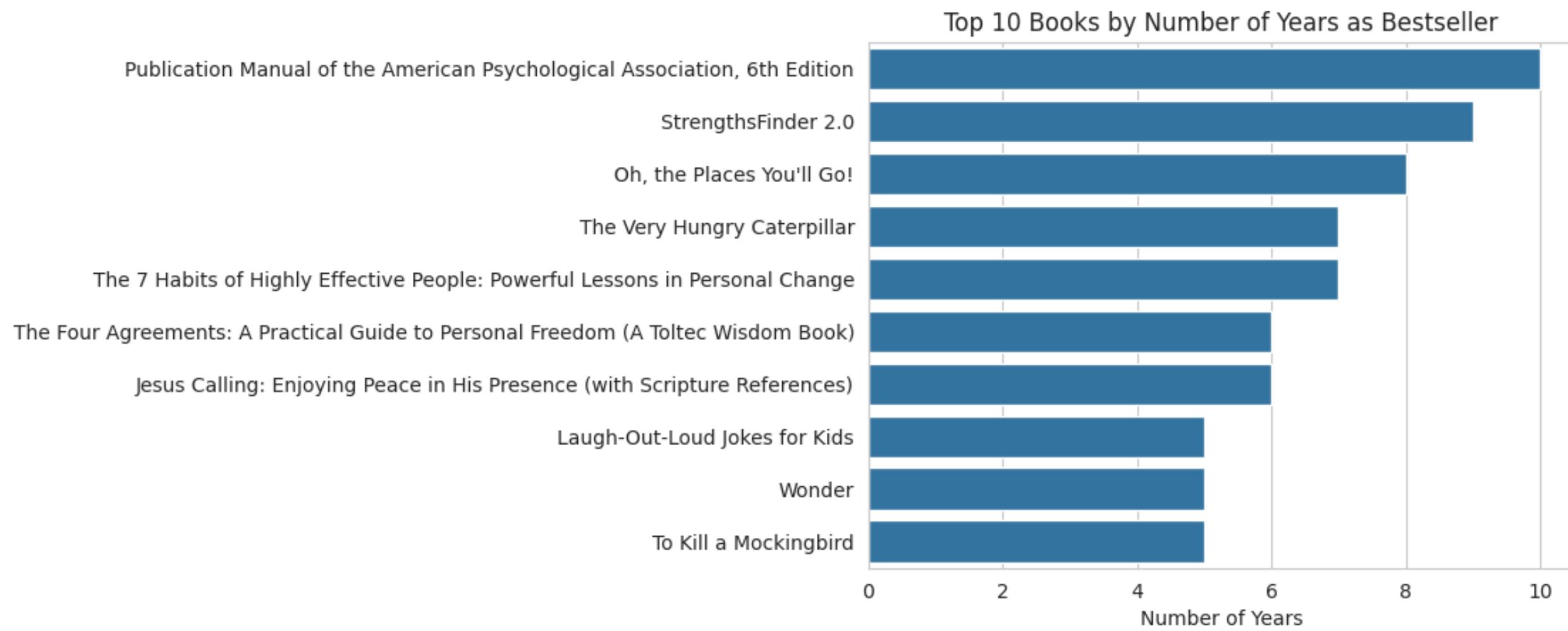


Exploring Genre-Based Price Differences: Fiction vs. Non-Fiction

The analysis of average prices by genre reveals a clear distinction between fiction and non-fiction books. Fiction books have a lower average price of 10.5, while non-fiction books are priced higher on average at 15. This suggests that non-fiction books, which often include specialized content, such as educational, self-help, or reference materials, may be priced higher due to their perceived value and market demand. In contrast, fiction books may have a more accessible price point, appealing to a wider audience.



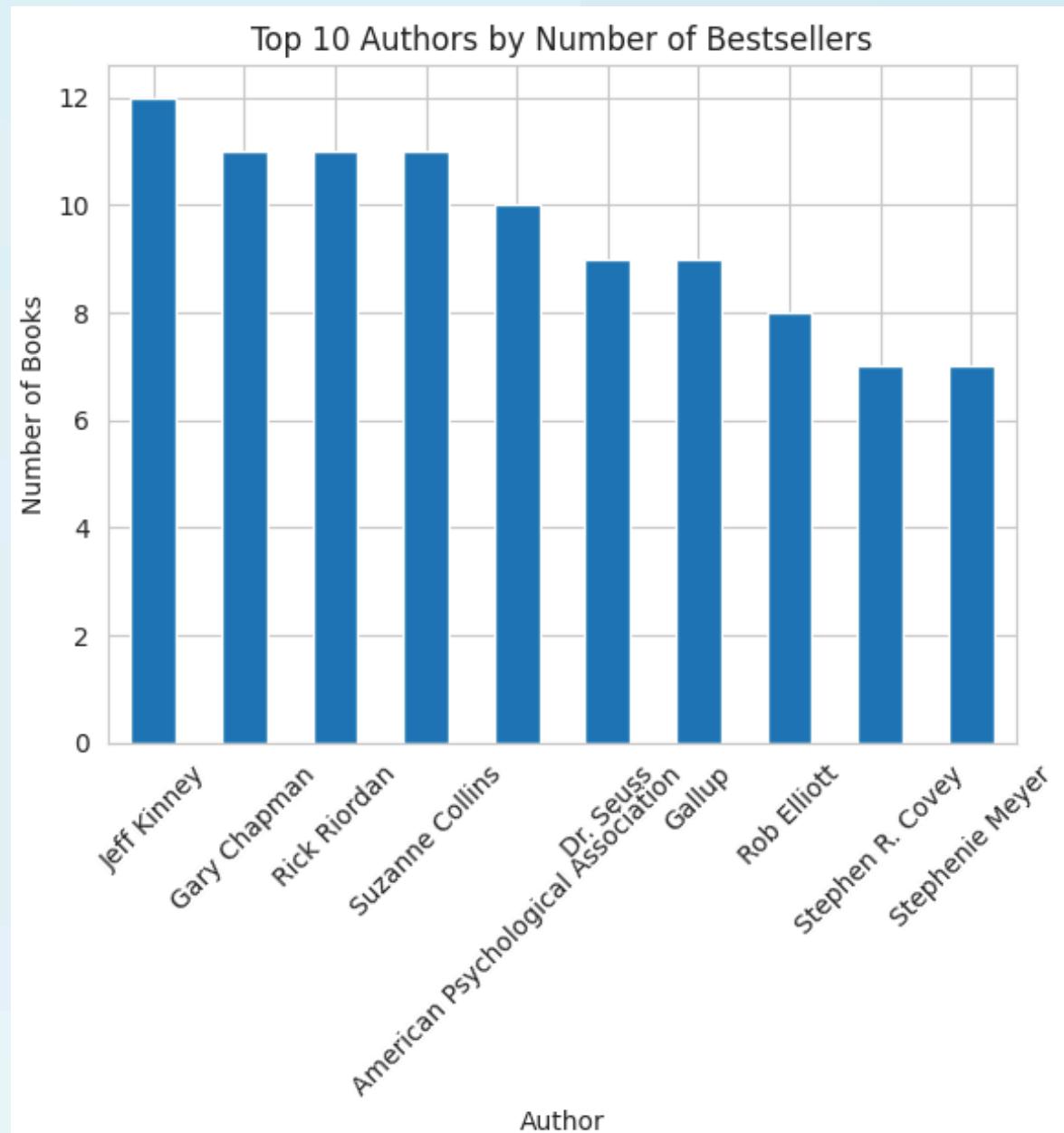
Top 10 Books: Dominating the Bestseller List for the Longest



This chart shows the top 10 books with the most years on Amazon's Bestseller List, highlighting their long-lasting popularity. These books have consistently captured readers' attention, staying relevant and in demand over time.

Most Successful Authors: Top 10 by Bestseller Count

This chart showcases the top 10 authors with the most books on Amazon's Bestseller List, highlighting their continued success and influence in the book market



Real-World Implications of the Bestseller Trends

This analysis provides valuable insights that can help libraries, booksellers, and publishers make informed decisions. Libraries can use these trends to curate their collections, prioritizing popular genres like Non-Fiction and bestselling authors to meet reader demand. Publishers can identify pricing strategies and factors that contribute to high engagement, such as affordable Fiction or specialized Non-Fiction content. Additionally, understanding user ratings and review patterns can help anyone interested in the book industry, including authors, to tailor their work to resonate with readers and achieve long-term success in a competitive market.



Code Implementation and Dataset Source

Dataset source

Google Colab code

Thank You