

# Amazon Best Selling Books Analysis with Python



Amazon officially launched as an online bookseller on July 16, 1995. All 50 U.S. states and 45 different nations received books from the startup retailer within a month.

In 1996, Bezos personally delivered his company's one-millionth order, to a customer in Japan who had purchased a Windows NT manual and a Princess Diana biography.

Seattle-born Amazon eventually morphed into an e-commerce colossus, selling everything from groceries to furniture to live ladybugs, and helping to revolutionize the way people shop.

And till today, Amazon is still the world's largest bookstore.

In this project, I will be presenting insights derived from the decade of analysis performed on the Amazon best-selling books data that was scrapped from Amazon's website.

These books were best-selling books on Amazon starting from 2009 to 2019.

## DATA GATHERING

The best-selling book data on Amazon was scrapped from the Amazon website using Beautiful soup and Selenium. A for loop and a function were used to get the details of each page of the best-selling book URLs. After the data was scrapped, it was saved to a text file, converted to a data frame, and finally saved to a CSV file.

## CATEGORIZATION OF THE BOOKS

For the successful analysis of the best-selling books on Amazon, I needed to analyze the books based on types. And it happened that Amazon does not have a definite categorization of books on their website as one book can be found in different departments which is a way they used to group their books.

The different departments they had to include Arts & Photography, Biographies & Memoirs, Books on CD, Business & Money, Children's Books, Christian Books & Bibles, Comics & Graphic Novels, e.t.c. And one book can be found in more than one category.

Due to this problem, I had to use information from [Goodreads](#), the world's largest site for readers, and book recommendations to group the books into two distinct categories.

On the CSV file, I added a new column called **genre** to the scrapped data and I categorized the book based on **Fiction** and **Non-Fiction**.

## DATA CLEANING

After assessing our data, I noticed there were missing values. To preserve the quality of the scrapped data, and to prevent deleting data that could affect our analysis, I assessed the CSV file where the data was saved, and I found those missing values and manually filled them up since I could access the data from Amazon websites where it was scrapped from.

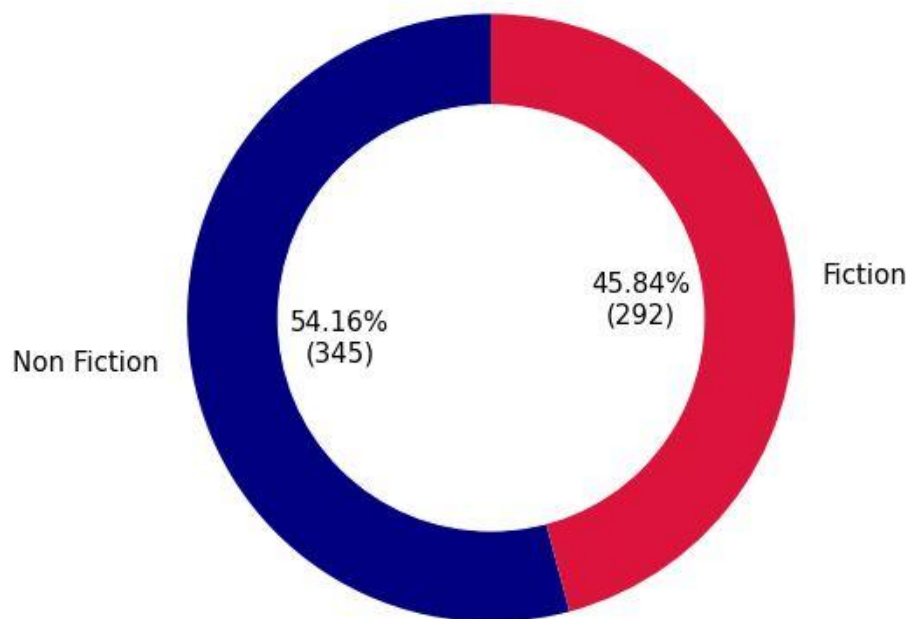
After the thorough checking and filling of missing values, the only missing values left were that of books that weren't available on the Amazon website.

Now we have quality data and I went ahead to visualize this data to derive insights from this analysis.

## DISTRIBUTION OF GENRE FROM ALL UNIQUE BOOKS FROM 2009 TO 2019

How many Fiction and Non-Fiction Books were Best Sellers on Amazon between 2009 and 2019?

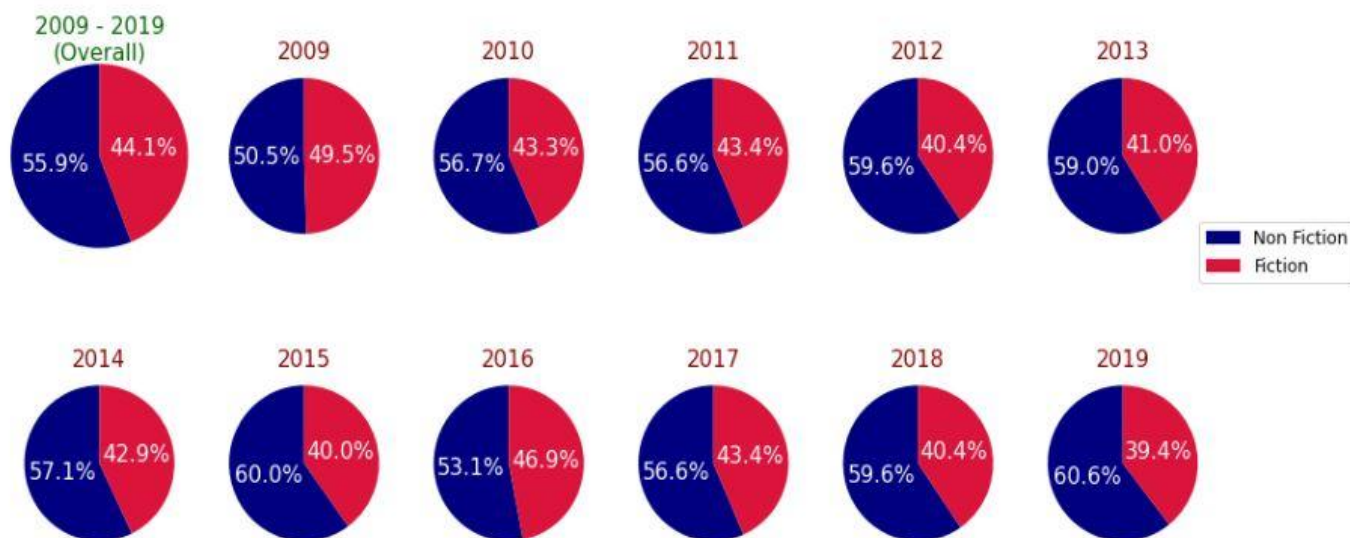
Book buyers on Amazon purchased more Non-Fiction books on Amazon as the pie chart tells us that we had more Non-Fiction books that were best-sellers between 2009 and 2019.



## DISTRIBUTION OF GENRE FOR THE UNIQUE YEARS

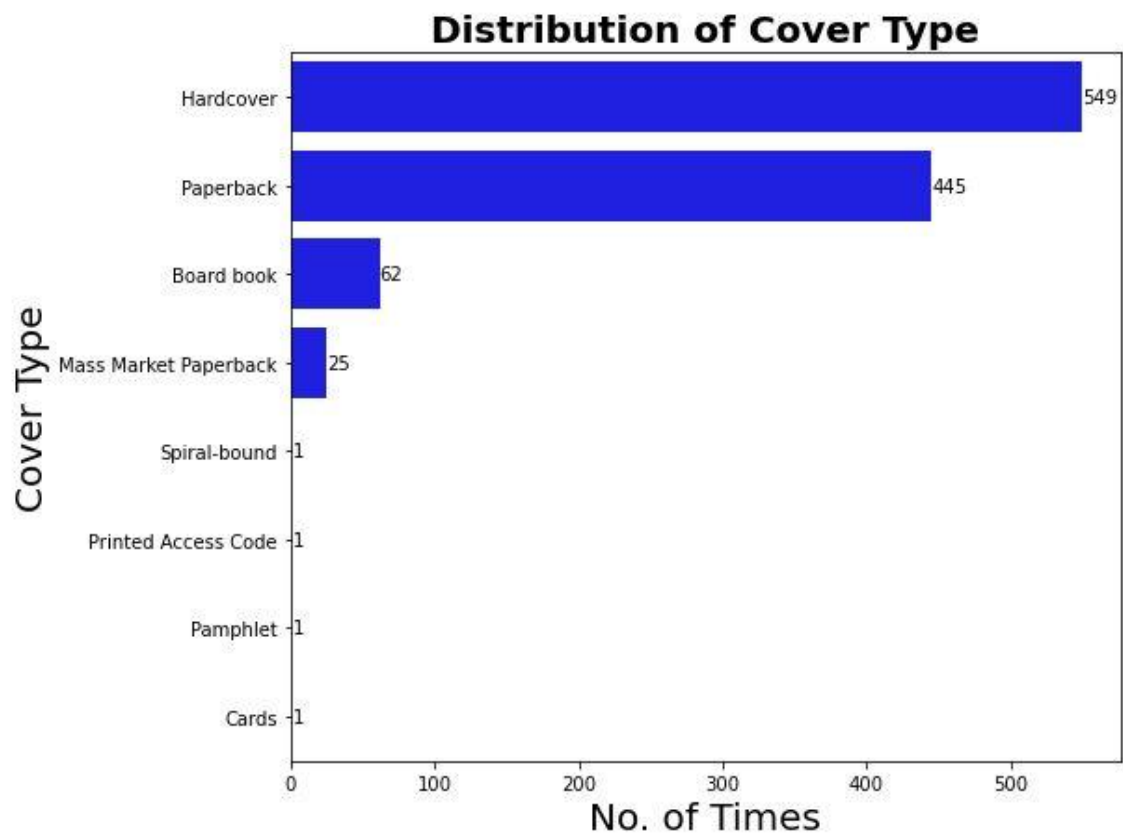
How many Fiction and Non-Fiction Books were sold each year?

2019 had the highest number of best-selling books which were Non-Fiction while 2009 had more Fiction books that were best sellers.



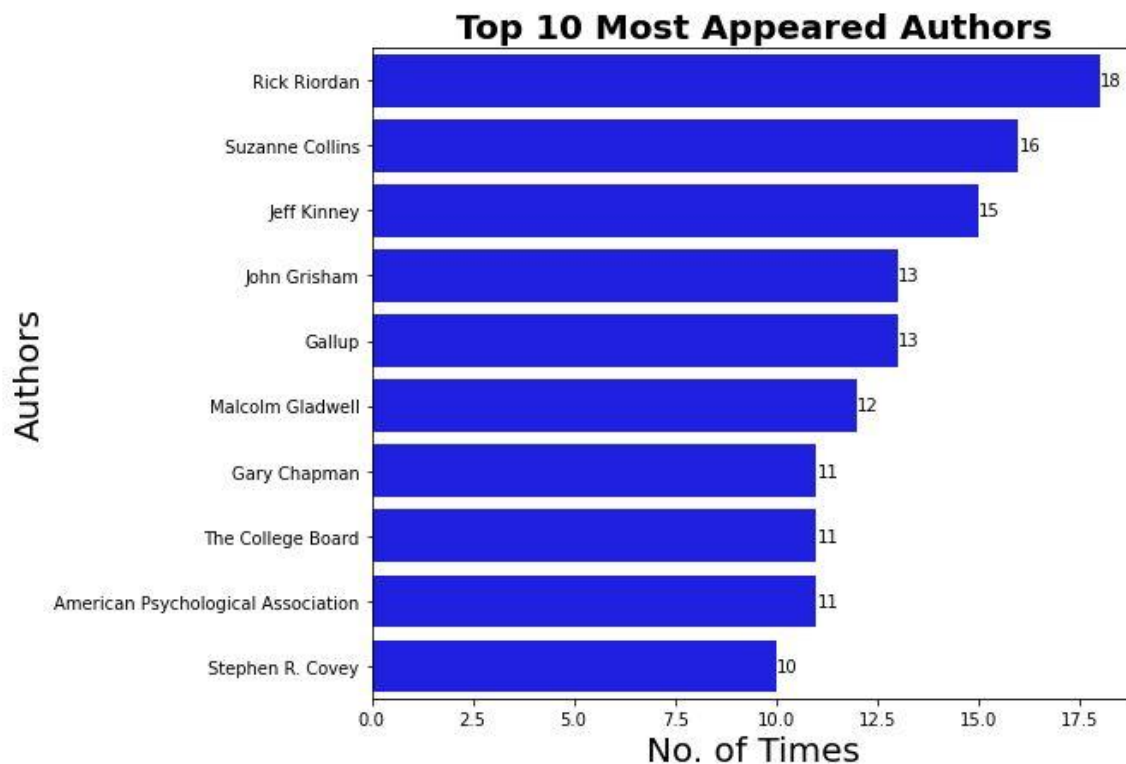
## DISTRIBUTION OF COVER TYPES

In the total number of best-selling books, the majority of the books, a total of 549 books had a hardcover type, followed by 445 books that had paperbacks.



## THE TOP 10 AUTHORS

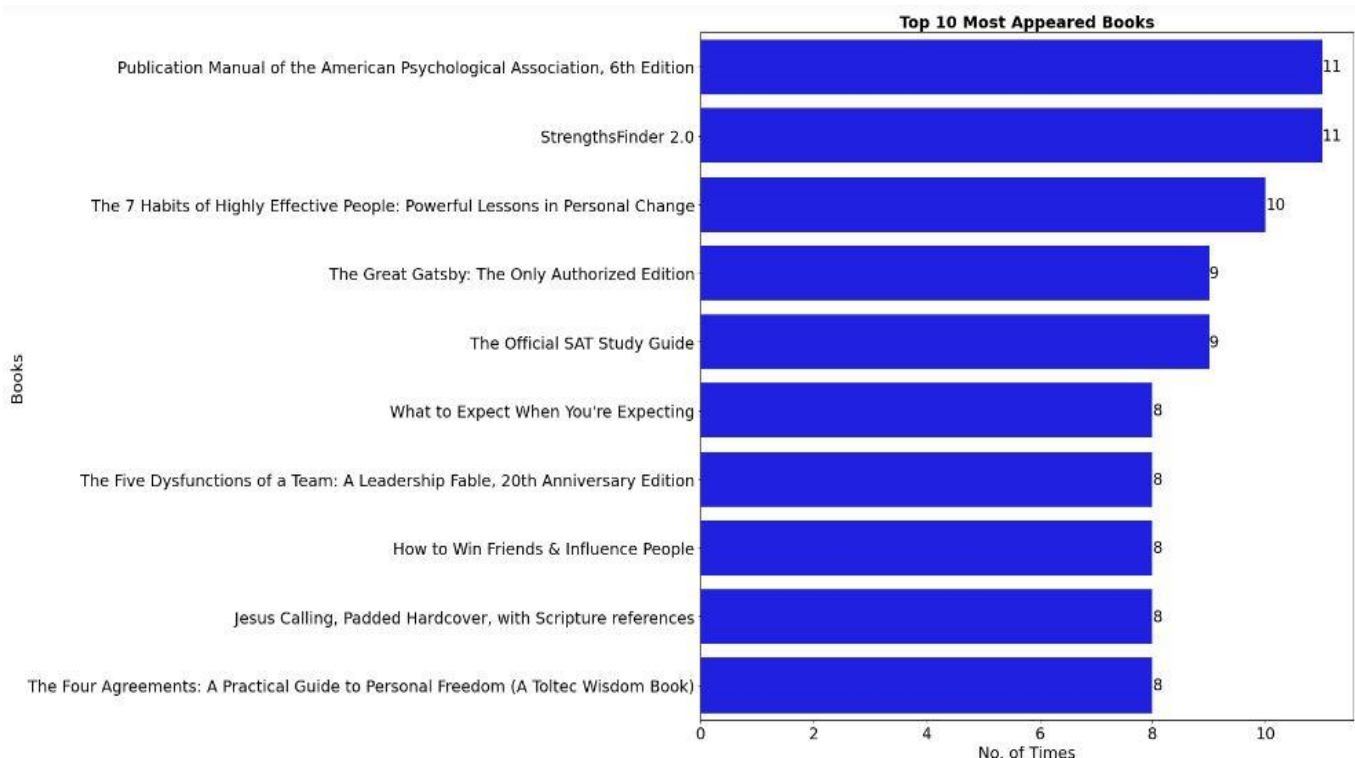
Generally, regardless of genre, Rick Riordan, Suzanne Collins, and Jeff Kinney were top best-selling authors from the top 10 most occurring authors.



## THE TOP 10 BOOKS

Based on the number of times the book appeared as a best-selling book from 2009 to 2019

Generally, regardless of genre, Publication Manual of the American Psychology Association, 6th Edition, StrengthsFinder 2.0, 7 Habits of Highly Effective People were the top best-selling books among the 10 most occurring books.

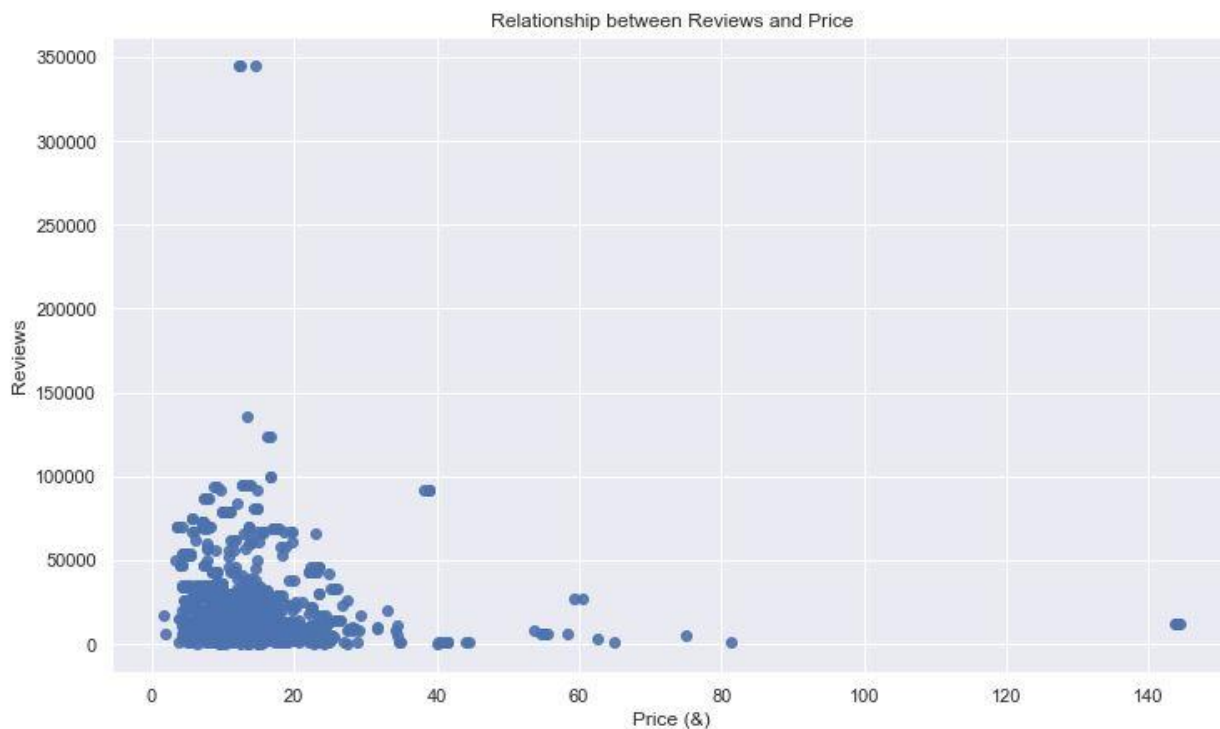




## RELATIONSHIP BETWEEN REVIEWS AND PRICE

Do low-priced books get more reviews and vice versa?

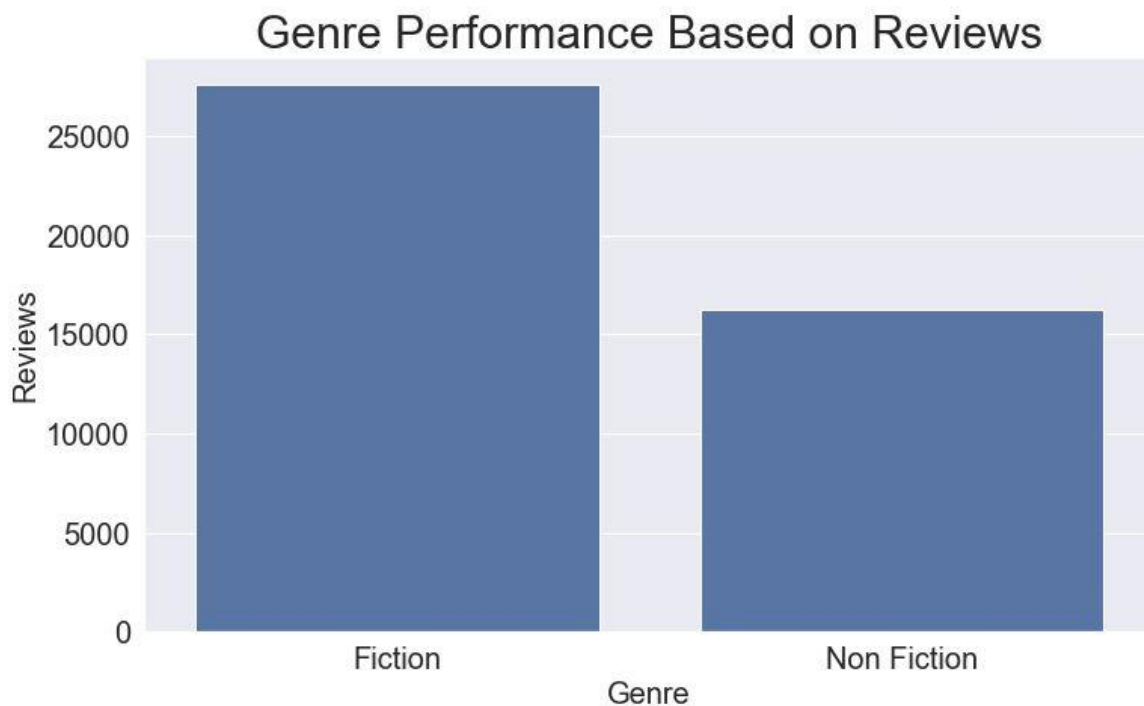
Reviews have no relationship with Price, but we also notice outliers in the book price. These are books with a book price of about \$140.



## GENRE PERFORMANCE BASED ON REVIEWS

Overall did Fiction books have more reviews than Non-Fiction books?

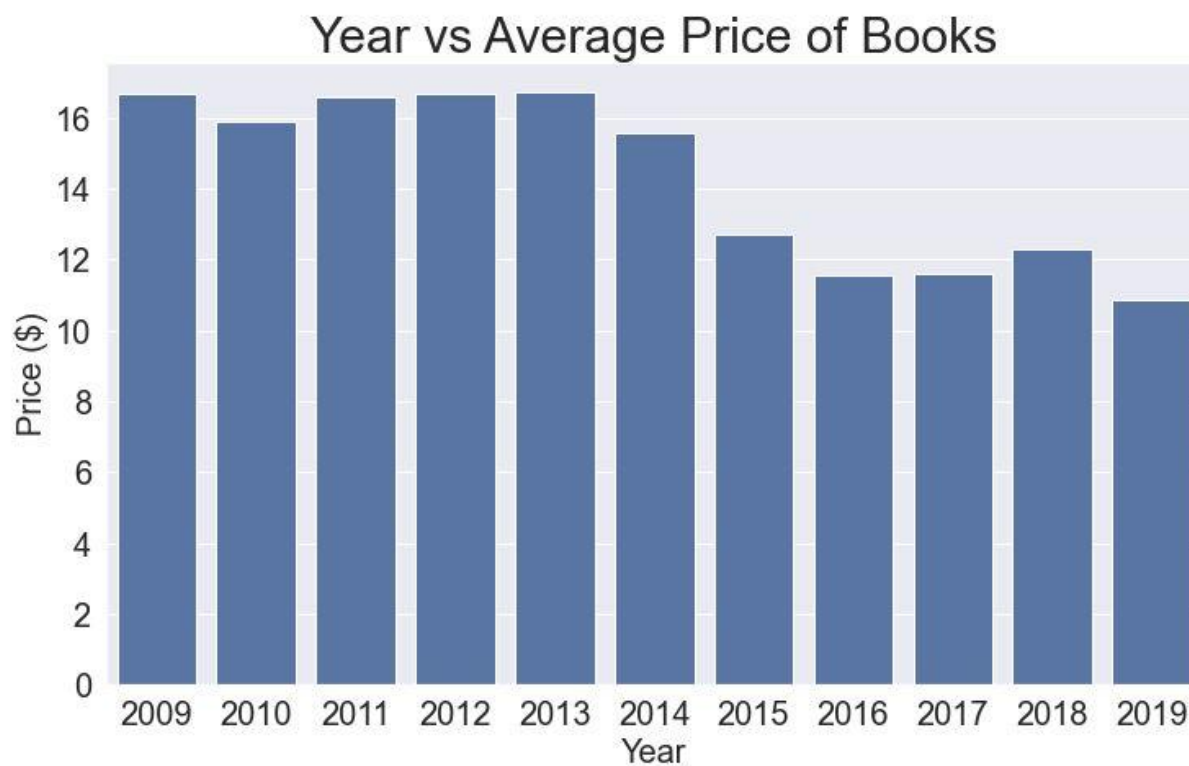
The bar graph shows that between 2009 and 2019, Fiction best selling books had more reviews than Non-Fiction books.



## RELATIONSHIP BETWEEN PRICE AND YEAR

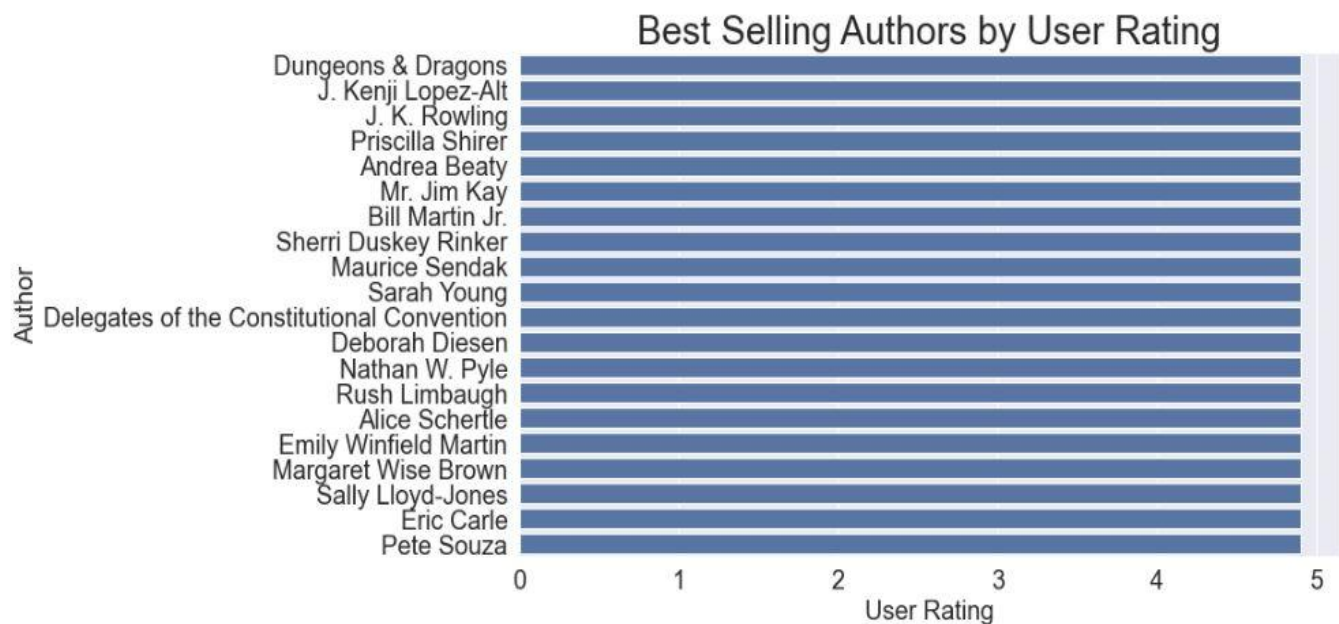
Were the costs of books increasing or decreasing with the years?

2013, 2009, 2012, and 2011 were the years with the highest average price of books. The average cost of books started decreasing in 2014, increased in 2018, and then reduced again in 2019.



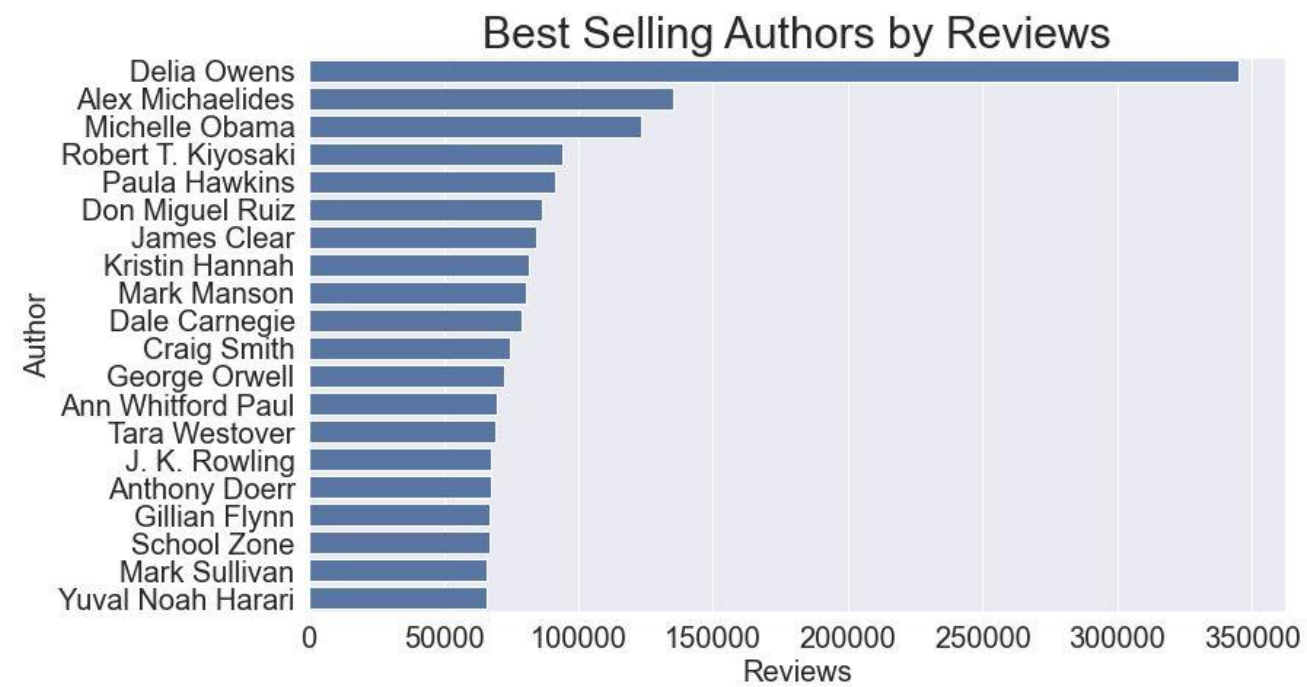
TOP 20 AUTHORS WITH THE HIGHEST USER RATINGS

The top 20 authors with higher user ratings had more than 4.5 stars in User Ratings.



TOP 20 AUTHORS WITH THE HIGHEST REVIEWS

Generally, Delia Owens, Alex Michaelides, Michelle Obama, Robert Kiyosaki, and Paula Hawkins were the top 5 in the top 20 best-selling authors with a higher number of reviews. But we notice that Delia Owens is the outlier that had the maximum number of reviews which is about 340,000 reviews which is very much higher than other authors' reviews.

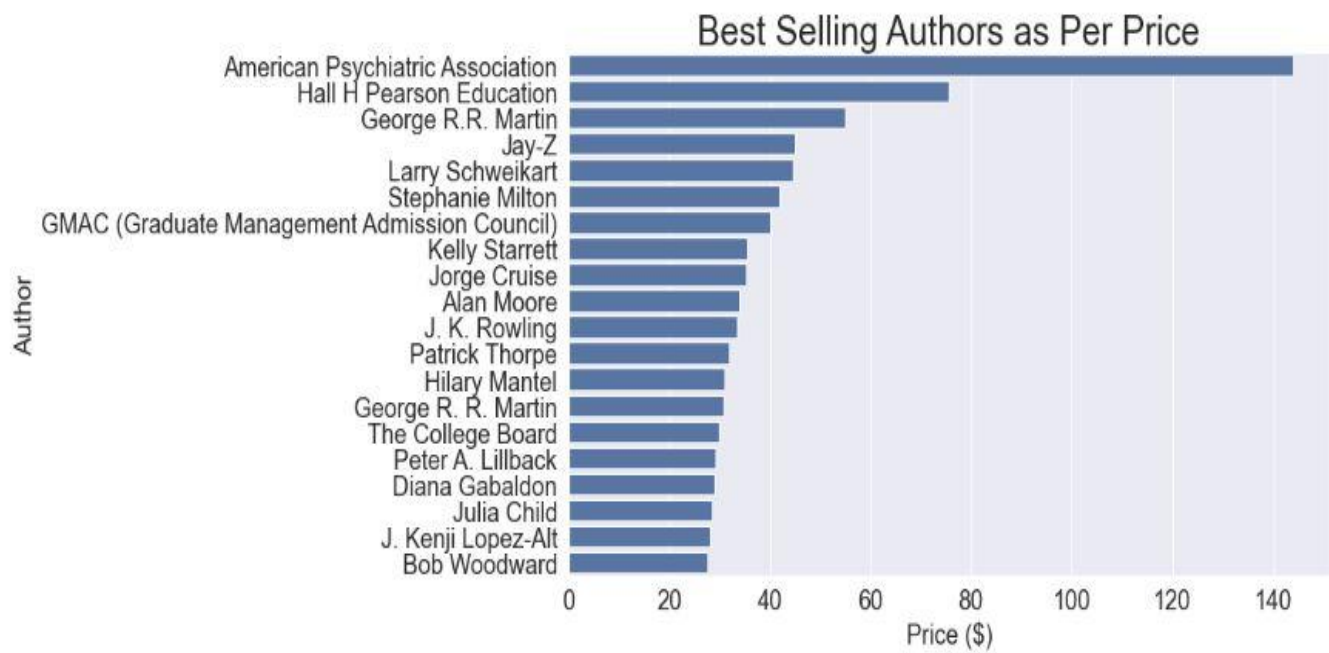




### TOP 20 AUTHORS BASED ON BOOK PRICE

The top 5 authors with the highest-priced books are American Psychiatric Association, Hall H Pearson Education, George R.R Martin, and Jay-Z.

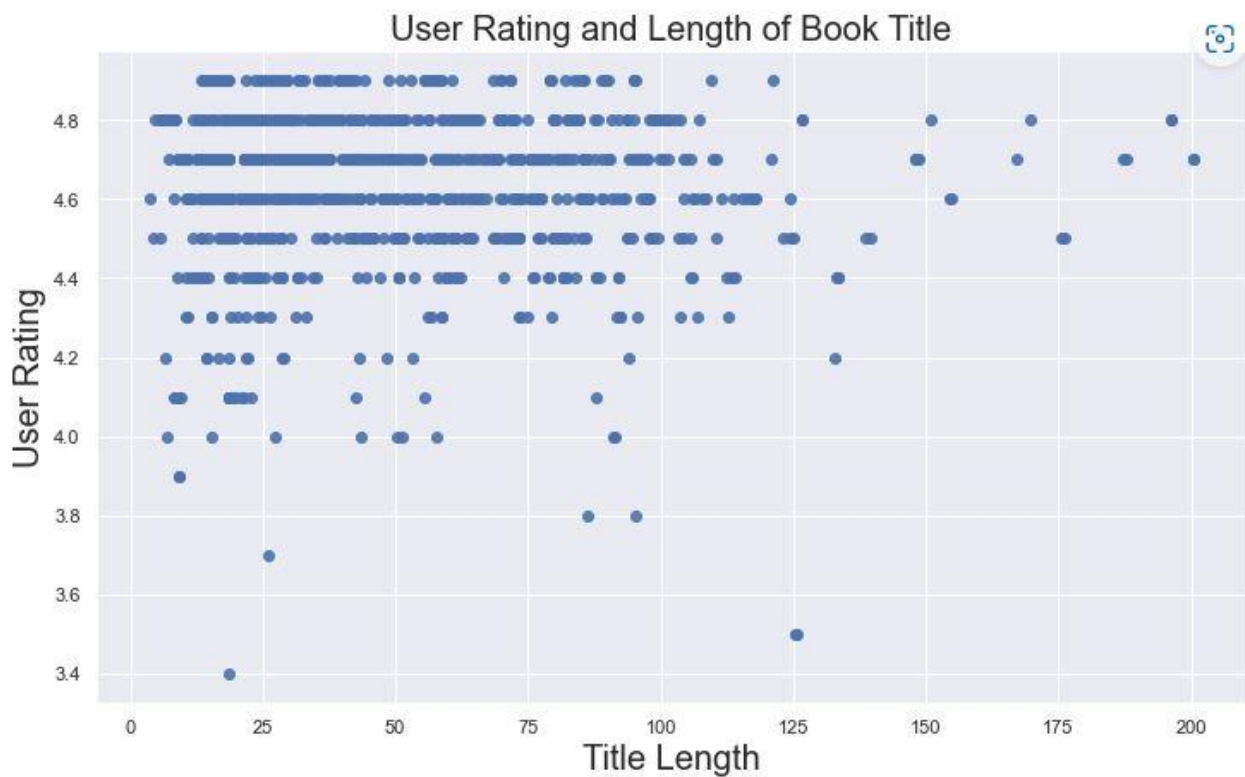
American Psychiatric Association is an outlier having a maximum average price of 144 dollars.



### RELATIONSHIP BETWEEN LENGTH OF BOOK TITLE AND USER RATINGS

Did Short titled books get more ratings than long-titled books?

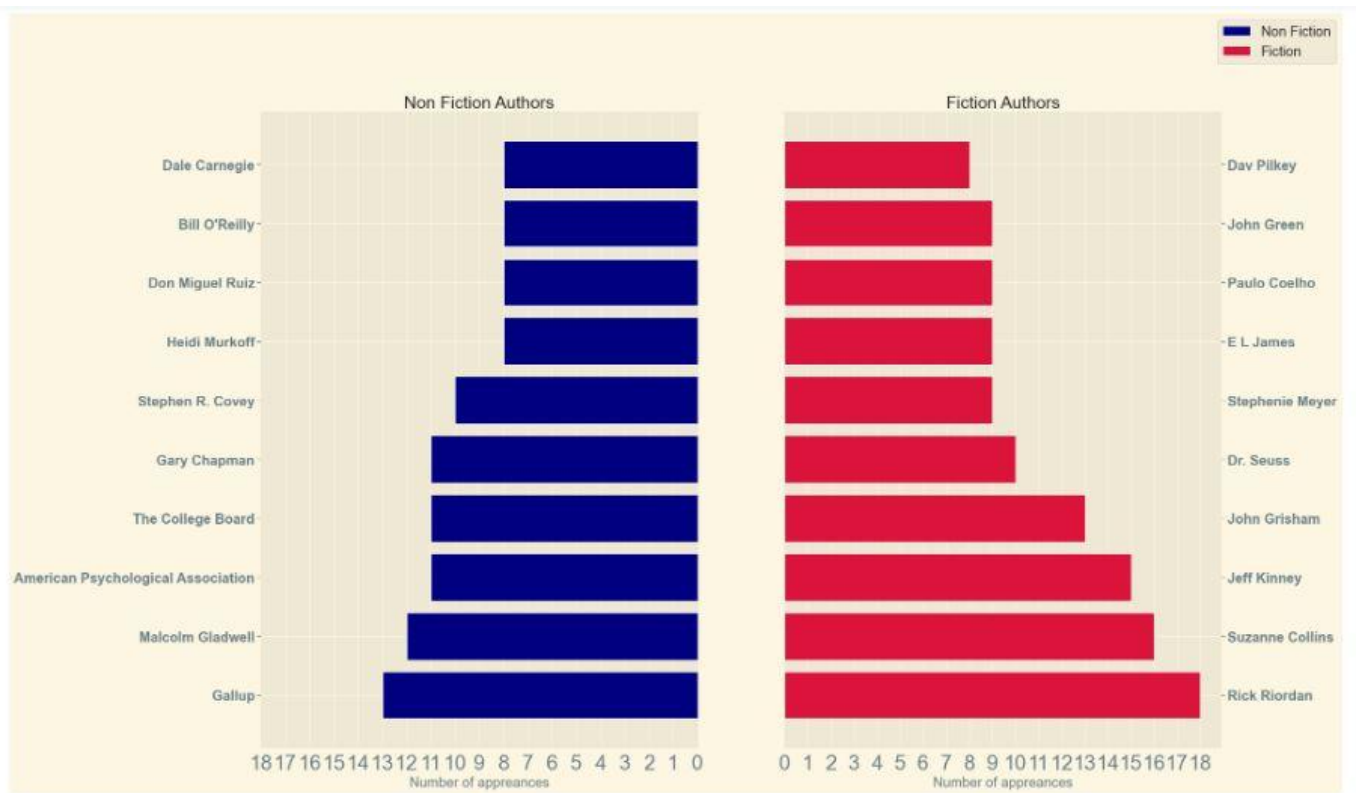
There is no correlation between user rating and the length of books. But we could also see that books with shorter titles have more user ratings.



## TOP 10 BEST-SELLING AUTHORS BASED ON GENRE AND NUMBER OF APPEARANCES

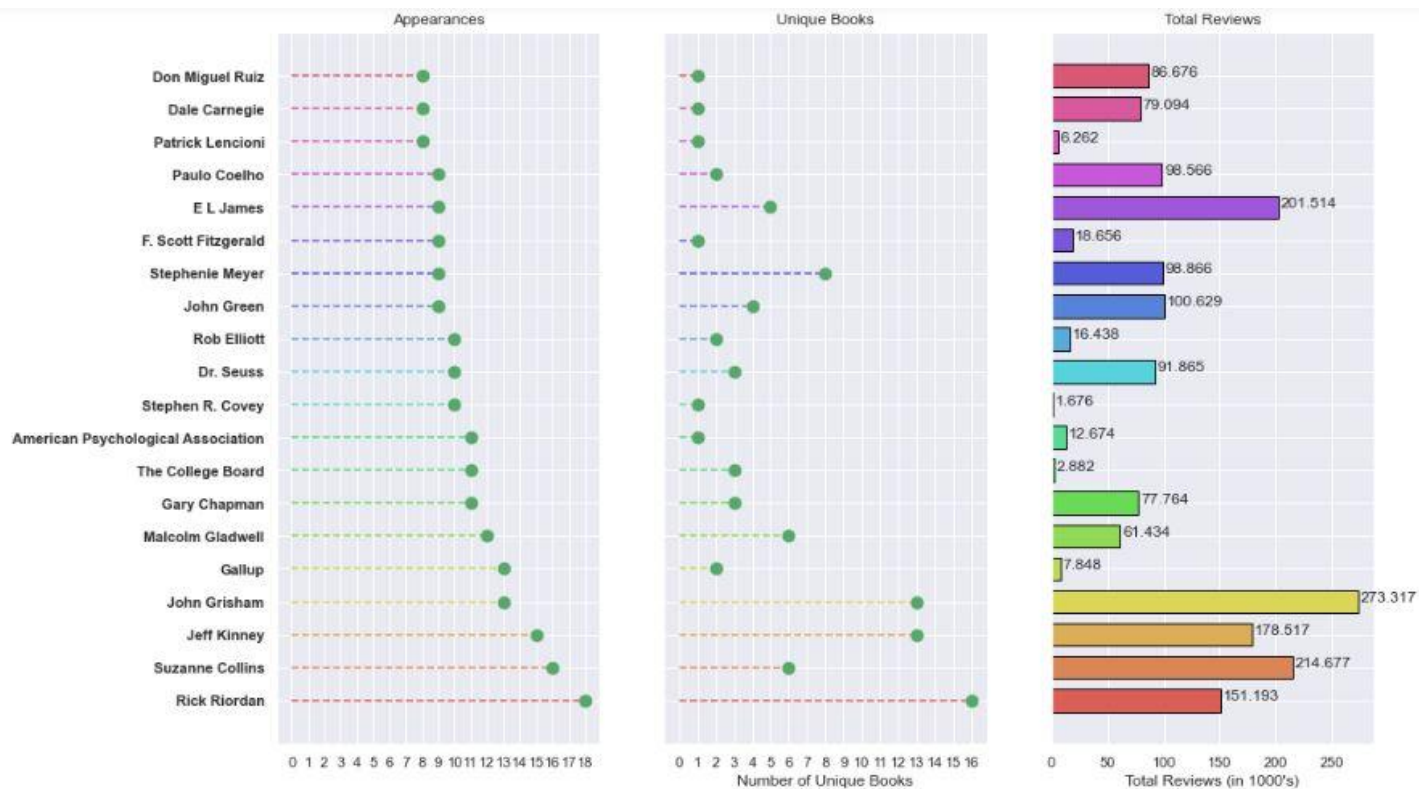
When it comes to the number of times their books were listed as a best-selling book between 2009 and 2019, in the top ten filtered authors,

- the top five best-selling authors in Fiction were Rick Riordan, Suzanne Collins, Jeff Kinney, Dr.Seuss, and Stephen Meyer.
- the top five best-selling authors in Non-Fiction were Gallup, Malcolm Gladwell, American Psychological Association, The College Board, and Gary Chapman.



## OVERALL BEST-SELLING AUTHORS BASED ON THEIR APPEARANCES (FICTION AND NON-FICTION COMBINED)

Here regardless of genre, Fiction writers like Rick Riordan, Suzanne Collins, Jeff Kinney, and John Grisham were the best-selling or sought-after authors.



## TOP 10 BOOKS BASED ON USER RATINGS

The top ten books based on user ratings had more than 4.5 stars in their user ratings.



## TOP 10 BOOKS BASED ON REVIEWS

The top best-selling books based on user ratings had When the Crawdads Sing, The Silent Patient, Becoming, and The Nightingale: A Novel as the top 4 books.

When The Crawdads Sing is a top outlier with the highest average number of reviews of about 344811 reviews.

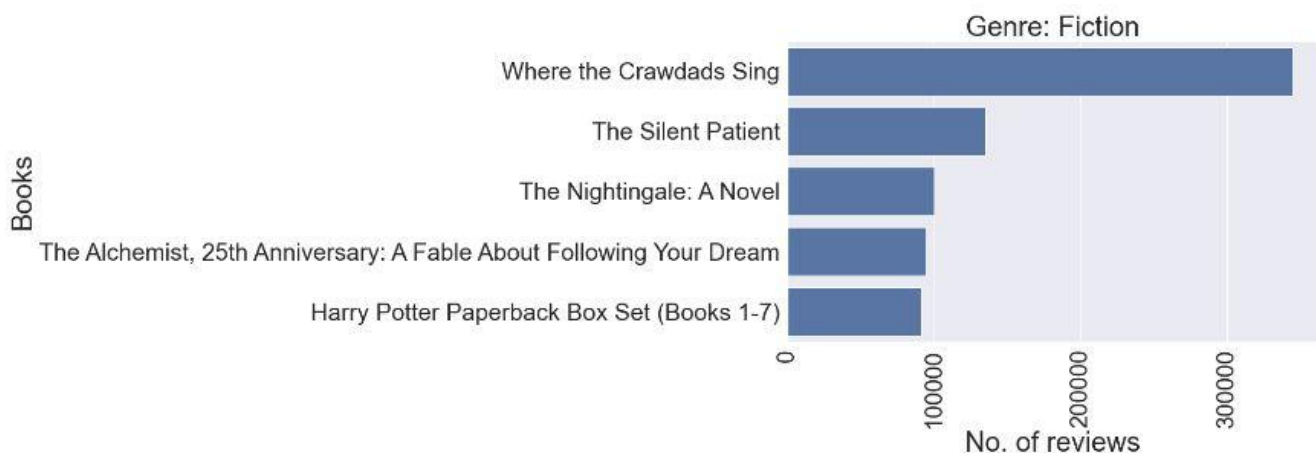


## TOP 10 BEST-SELLING BOOKS BY GENRE AND NUMBER OF REVIEWS

The top five Fiction Books with the highest number of reviews were *Where the Crawdads Sing*, *The Silent Patient*, *The Nightingale: A Novel*, *The Alchemist*, and *25th Anniversary: A Fable About Following Your Dreams*.

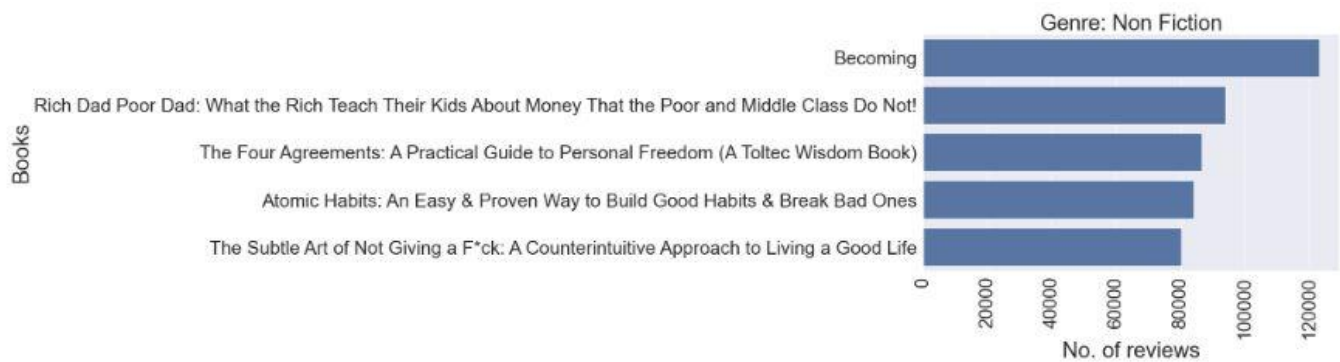
While in the Non-Fiction books, the Top five were *Becoming*, *Rich Dad Poor Dad: What the Rich Teach Their Kids About Money That the Poor and Middle Class Do Not!*, *The Four Agreements: A Practical Guide to Personal Freedom (A Toltec Wisdom Book)*, *Atomic Habits, An Easy & Proven Way to Build Habits & Break Bad Ones*, and *The Subtle Art of Not Giving a Fuuck: A Counterintuitive Approach to Living A Good Life*.

### For Fiction Books



### For Non-Fiction

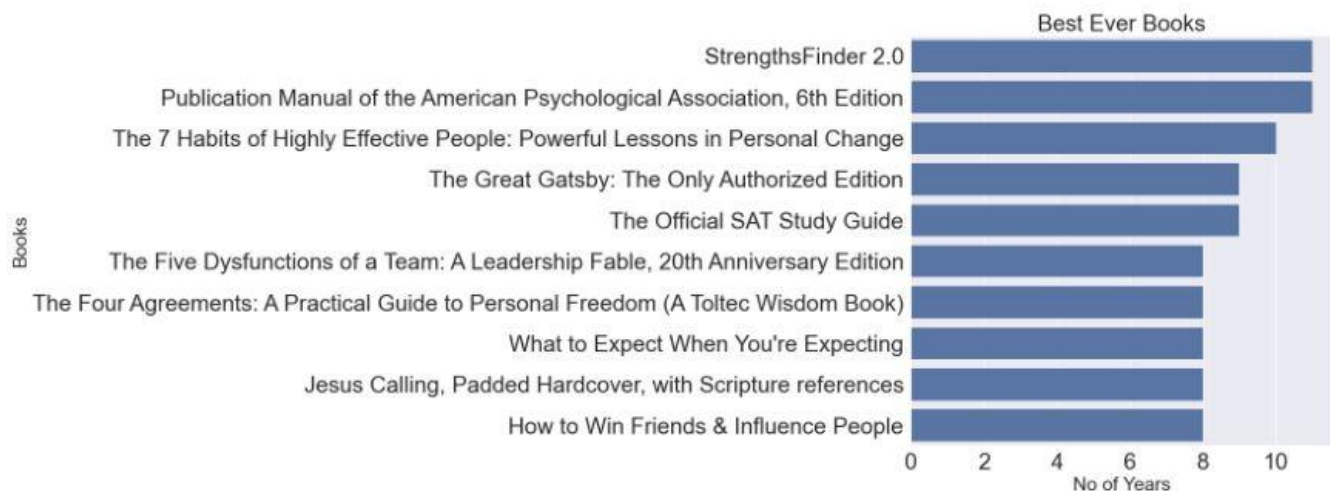




## GENERALLY, BEST-SELLING BOOKS

These are still the top 10 best books based on the number of times they appeared as best-selling books between 2009 and 2019.

In the overall top best-selling books based on appearance, self-development books were the most occurring books. These self-development books include StrengthsFinder 2.0, The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change, The Five Dysfunctions of a Team: A Leadership Fable, 20th Anniversary Edition, The Four Agreements: A Practical Guide to Personal Freedom (A Toltec Wisdom Book), and How to Win Friends and Influence People.



## A WORD CLOUD OF THE TOP BEST-SELLING BOOKS EVER based on the number of appearances

- ## CONCLUSION

8. The average price of books started decreasing in 2014, increased in 2018, and then reduced again in 2019.
9. The top 20 authors with higher user ratings had more than 4.5 stars in User Ratings.
10. Delia Owens, Alex Michaelides, Michelle Obama, Robert Kiyosaki, and Paula Hawkins were the top 5 in the top 20 best-selling authors with a higher number of reviews. Delia Owens had the maximum number of reviews which was about 340,000 reviews and was very much higher than other authors' reviews.
11. The top 5 authors with the highest-priced books are American Psychiatric Association, Hall H Pearson Education, George R.R Martin, and Jay-Z.

American Psychiatric Association is an outlier having a maximum average price of 144 dollars.

12. Based on Genre and number of book appearances:
  - the top five best-selling authors in Fiction were Rick Riordan, Suzanne Collins, Jeff Kinney, Dr. Seuss, and Stephen Meyer.
  - the top five best-selling authors in Non-Fiction were Gallup, Malcolm Gladwell, American Psychological Association, The College Board, and Gary Chapman.
13. Fiction writers like Rick Riordan, Suzanne Collins, Jeff Kinney, and John Grisham were best-selling or sought-after authors.
14. The top 10 books based on user ratings had more than 4.5 stars in their user ratings.
15. The top best-selling books based on user ratings had When the Crawdads Sing, The Silent Patient, Becoming, and The Nightingale: A Novel as the top 4 books.

Meanwhile When The Crawdads Sing is the top outlier with the highest average number of reviews of about 344811 reviews.

16. Based on Genre and Number of Reviews:
  - The top five Fiction Books with the highest number of reviews were Where the Crawdads Sing, The Silent Patient, The Nightingale: A Novel, The Alchemist, and 25th Anniversary: A Fable About Following Your Dreams.
  - While in the Non-Fiction books, the Top five were Becoming, Rich Dad Poor Dad: What the Rich Teach Their Kids About Money That the Poor and Middle Class Do Not!, The Four Agreements: A Practical Guide to Personal Freedom (A Tolle Wisdom Book), Atomic Habits, An Easy & Proven Way to Build Habits & Break Bad Ones, and The Subtle Art of Not Giving a Fuck: A Counterintuitive Approach to Living A Good Life.
17. The overall top best-selling books based on appearance were self-development books.