Exploring the World of Books: An Analysis of User Trends and Ratings

GROUP 8 - Hritik Mahajan, Kirthana Shri, Navya Pragathi, Shubhada Avahad

02/12/2023

Contents

Introduction & Problem Statement

The book industry has undergone a major transformation in recent years, due to advancements in technology and changes in consumer behavior. The widespread availability of e-books, audiobooks, and the ease of online purchasing has revolutionized the way people read and access literature. Despite the rise of e-books and audiobooks, physical books have maintained a strong following, with many readers preferring the traditional feel and experience of holding a physical book in their hands. As a result, the demand for books has skyrocketed, leading to a significant increase in sales and popularity. We are working with the books read by users and ratings provided by them on Amazon to understand these changes and the evolving reading habits of consumers. We have designed a case study to analyze three key datasets: Books, Users, and Ratings. This report is divided into three sections exploring these three datasets. These datasets include locations, age, name of the books, publisher name, ratings provided by the users and the ISBN. The three sections we explored are [1] Books Analysis [2] User profile analysis based on age and location [3] User reading habits based on ratings Analysis
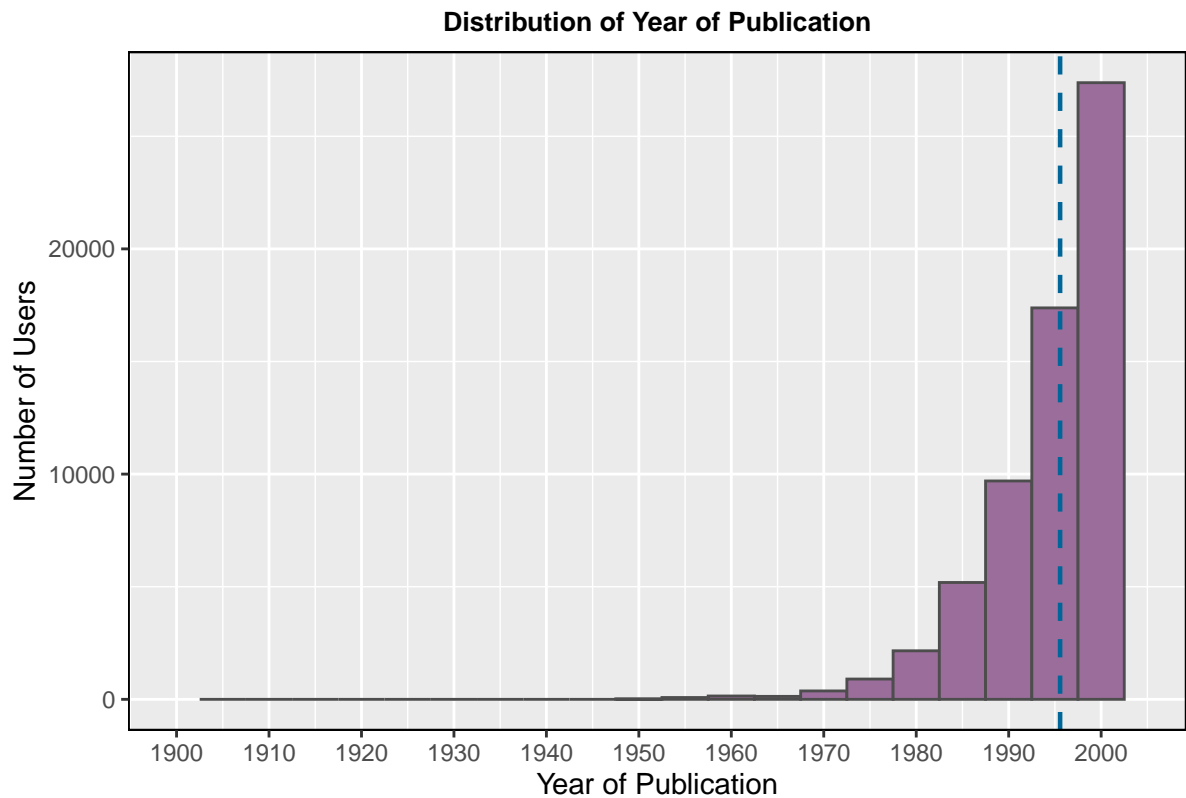
About the dataset

The Books dataset provides information about the books that are being purchased, including their publication, author, and publication year. This information can help publishers and retailers to better understand the types of books that are popular and in demand. The Users dataset gives an insight into the demographics of book consumers, including age. This information can help publishers and retailers to tailor their marketing and product offerings to better meet the needs of their target audience. The user location also provides information about the locations where books are being purchased and consumed. This information can be

used to identify areas with high demand for books, allowing publishers and retailers to better understand where to focus their efforts. It can also help to identify geographic trends and patterns, such as regional preferences for certain publishers or authors. The Ratings dataset can also be used to identify trends and patterns in user preferences. For example, it may be possible to identify specific authors that are particularly popular among certain demographics, such as young adults or older readers.

Section 1: Books Analysis

Q. What is the distribution of the publication years?

The information presented in the graph can be used to make inferences about the state of the publishing industry at the time.We can see an upward trend in the number of books published starting from the year 1970.This increase could be due to a variety of factors, such as advancements in technology that made it easier to publish and distribute books, an increase in the number of authors, or a growing demand for books from the general public. It's also possible that the increasing trend from the year 1970 could have been influenced by broader cultural, economic, and political factors or government policies. This could have had a positive impact on the publishing industry, potentially leading to more investment, innovation, and growth.

Distribution of Year of Publication

Q. What are the most popular books in the dataset?

Top 10 books : We can see that the book "Wild Animus" appeals to readers looking for something unique and experimental, while "The Lovely Bones: A Novel" is popular due to its emotional themes and powerful writing. On the other hand, "The Da Vinci Code" has captured the attention of readers with its intriguing storyline and exciting plot twists. These books demonstrate that different elements of a book, such as storyline, themes, and narrative style, can all play a significant role in determining its popularity.

Moreover, the diverse tastes of readers are highlighted by the popularity of these three books, as they cover a range of genres and styles. This finding highlights the need for publishers and authors to cater to a variety of reader preferences, as there is no one-size-fits-all formula for what makes a book popular.

Top 10 popular books in the dataset

Top 10 Prolific Book Authors
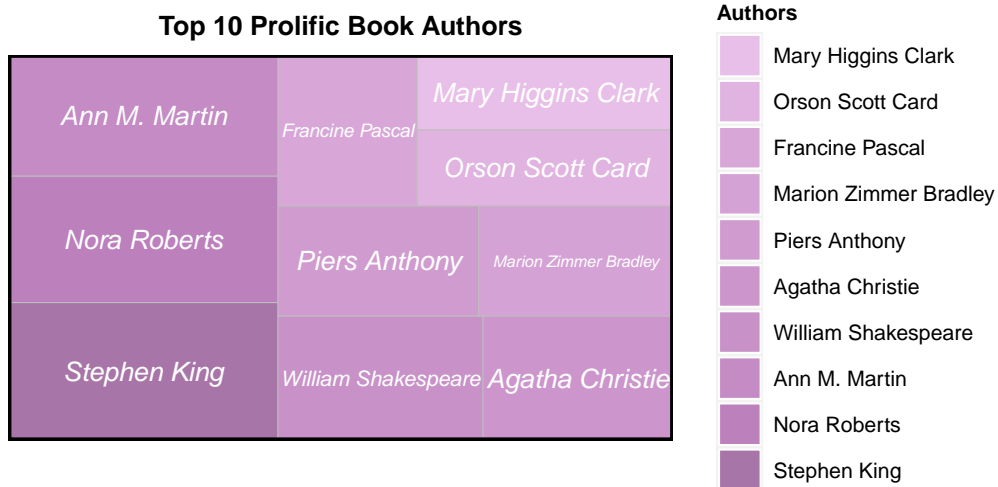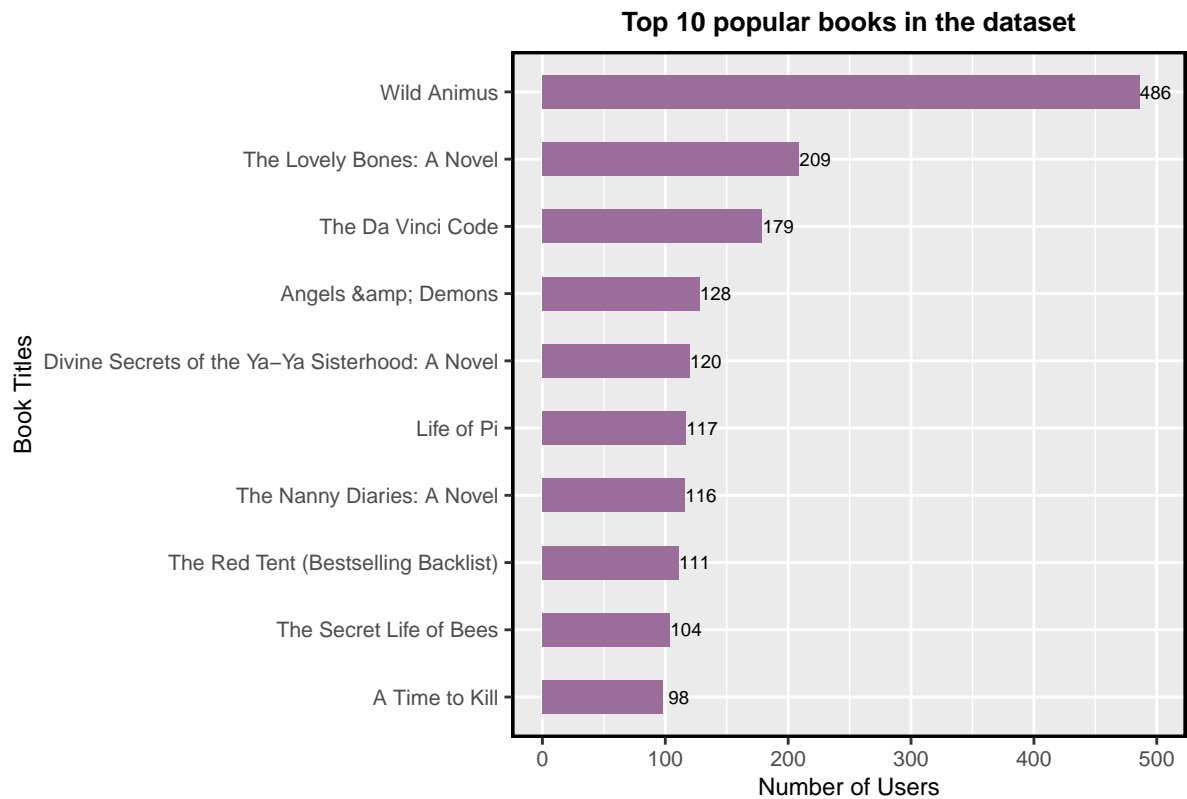
Top 10 authors: We can easily see the authors who have made the biggest contribution to the industry and compare the number of books written by each author. One of the key observations from the graph is the commitment and strong passion for writing of the top 10 authors. The large number of books written by each author suggests that they have a deep passion for writing and a strong commitment to their craft. These authors are not only able to write a high number of books, but they are also able to maintain a high

level of quality in their work, which is a testament to their talent and dedication and have been able to make a significant impact on the publishing industry.The above analysis also has a broader implication for society.They inspire new generations of writers and readers and help to keep the publishing industry vibrant and alive.

Q. What is the most used book title?

A book title provides a quick and visually appealing way to get a sense of themes and genres of a group of books. The word cloud identifies the most used book titles making it possible to determine the popularity of specific books and to understand what types of book titles are in demand. The top words suggest that there is a significant demand for books that focus on life, love, mystery, and classic literature, as well as guides.The word "life" being one of the top words indicates that there is a strong demand for books that explore various aspects of life, such as personal growth, self-discovery, and the human experience. The word "mystery" suggests that there is a significant interest in books that feature suspense, intrigue, and detective work. These trends are likely due to the universal appeal of this theme, as well as the ongoing interest in stories that touch on the human condition.

## Section 2: User profile Analysis based on age and location

Q. What is the number of users based on age category?

**Percentage of Reviewers based on Age category**
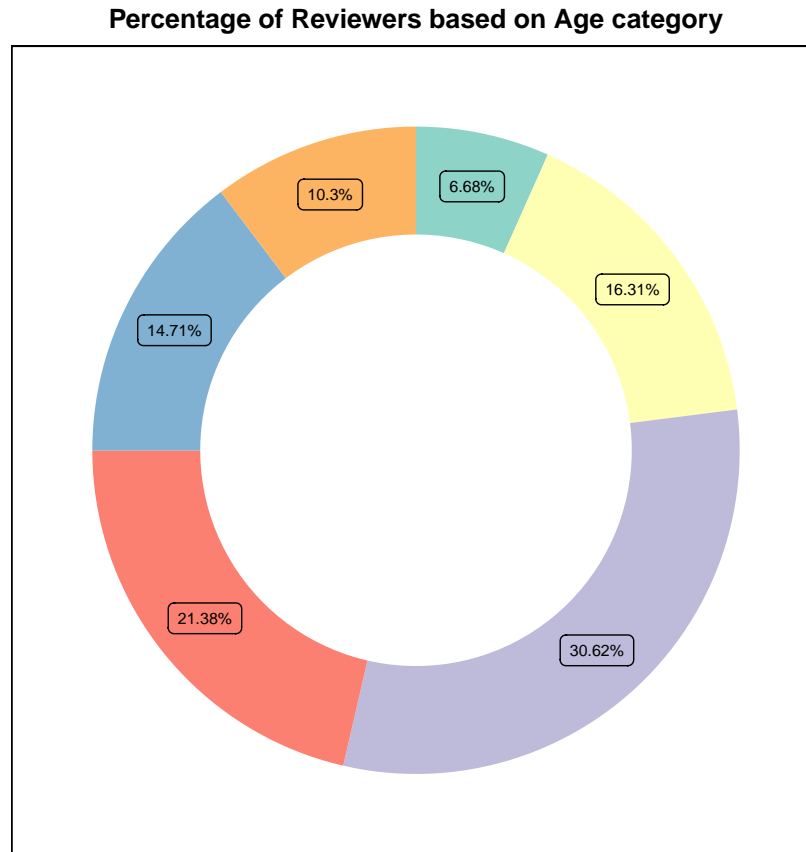
The 25 to 34 age group typically consists of young professionals who have more disposable income and leisure time to devote to reading and rating books. Additionally, many people in this age group are well-educated and may have intellectual pursuits, which could contribute to their increased engagement in book reading and rating.On the other hand, the lower number of users in the age group of less than 18 could be due to factors like they are still in school and have limited free time to devote to reading and rating books. Additionally, many children in this age group are not yet financially independent, so they may not have the means to purchase books and participate in book-related activities.

Q. What is the number of users based in the top 10 countries?

The high number of users in the United States, Canada, and Germany could be attributed to a variety of factors. These countries are home to large populations with a high level of literacy and education, which could contribute to their increased engagement in book reading and related activities. Additionally, these countries have strong economies and well-developed publishing industries, which could provide more access to books and resources for readers.On the other hand, the lower number of users in New Zealand and Italy could be due to a variety of reasons such as smaller populations.

Top 10 Countries User Distribution
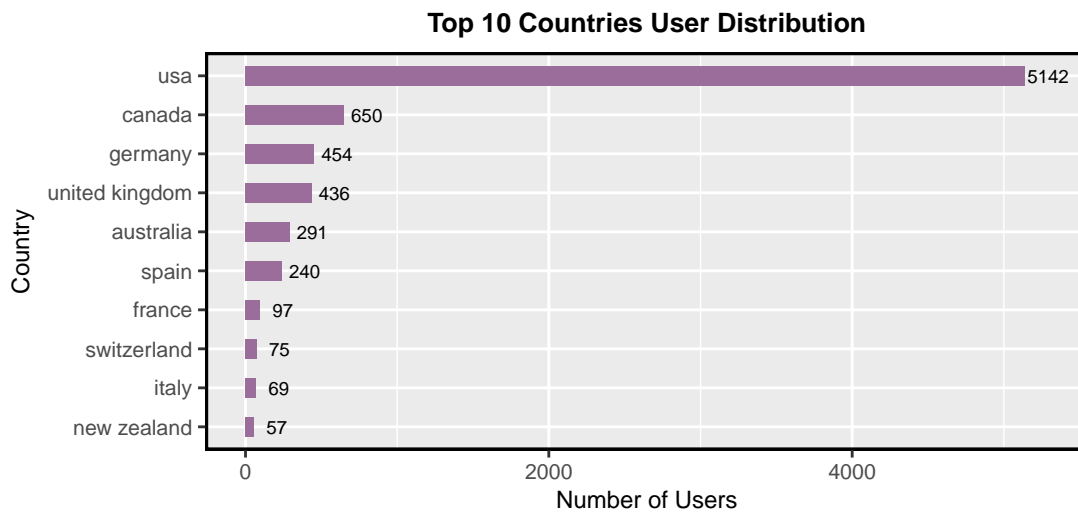
Q. What is the age distribution of users from top 5 countries actively participating in ratings?

Book Ratings by Age Category
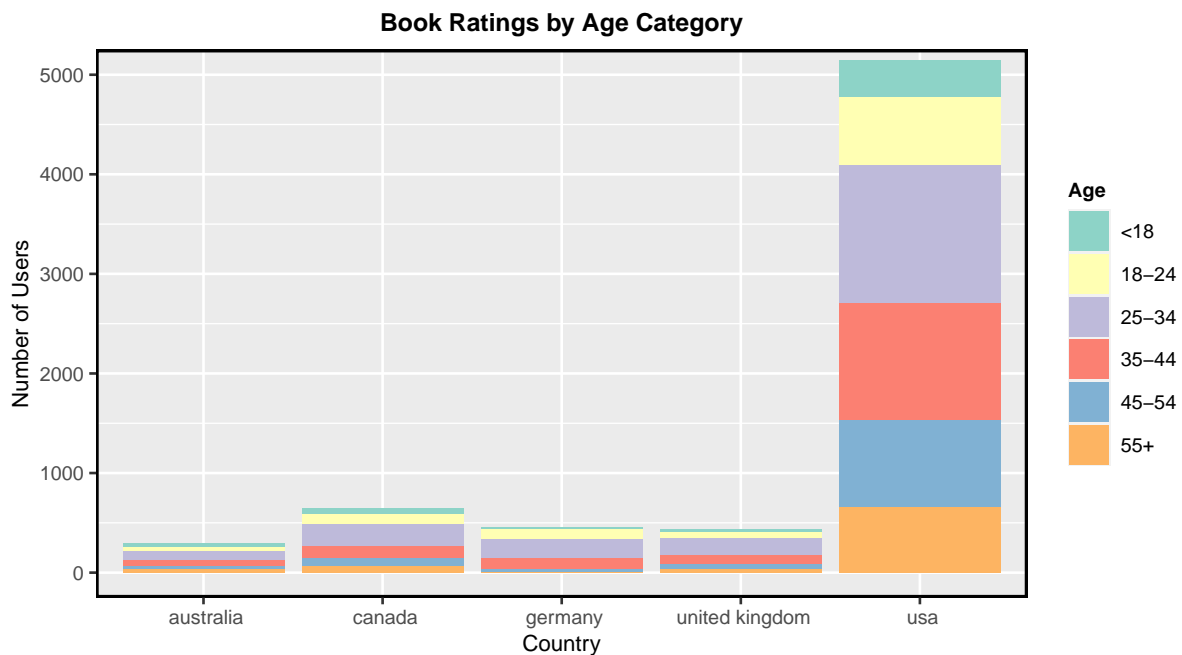
The age distribution of users varies among the different countries represented in the graph. This difference in age distribution could be due to several factors, such as cultural differences, different levels of access to technology, or even different stages of market penetration for the product.It's also worth noting that the age distribution of users in each country is generally skewed towards the age ranges between 25-34. This could be due to a variety of reasons, such as a greater affinity for technology and digital products among younger generations, or a greater willingness to engage in online activities such as rating and reviewing products. Additionally, the consistency in the age distribution of users across the top 5 countries suggests that the preferences and habits of book readers are relatively similar across different countries.

Section 3: User reading habits based on ratings Analysis

Q. What are the ratings of top 5 authors?

**Distribution of Ratings for top 5 Book Authors**
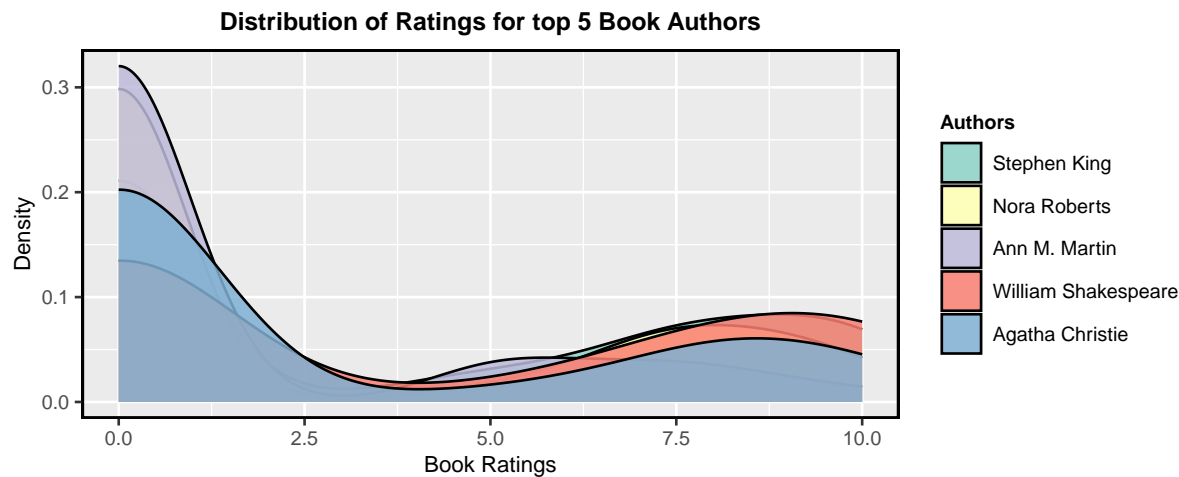
The plot displays the distribution of ratings for the top 5 authors, allowing us to see patterns and trends in the data, as well as get a sense of the overall spread of ratings for each author. William Shakespeare and Stephen King have a higher concentration of high ratings whereas Ann M. Martin and Agatha Christie have a higher proportion of low ratings.

Q. How is the user behavior regarding ratings throughout the years?

Book Ratings by Age Category
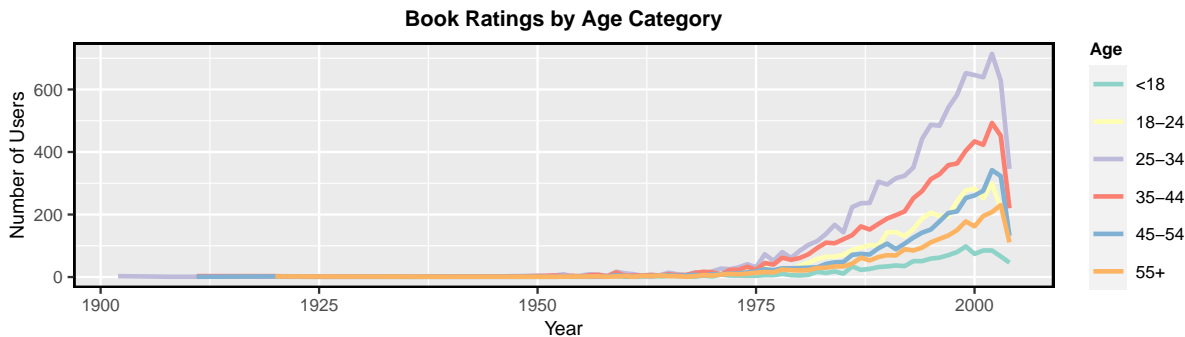
Book Ratings by age category: We can see an upward trend in the number of user ratings over the years, this could indicate that people are becoming more interested in reading and rating books, or that there has been an increase in access to technology and resources for rating books. Additionally, the graph also shows that the age group less than 18 years old is the lowest in terms of user ratings. This suggests that this age group is less interested in reading and rating books, or that they may have fewer resources (such as access to technology or the ability to purchase books) to engage with the publishing industry This analysis is valuable for publishers and authors, who can use it to understand the market and to tailor their work to meet the demands of this age group. about how to reach this age group and to encourage them to become more active in reading and rating books.

Rating distribution across user age groups
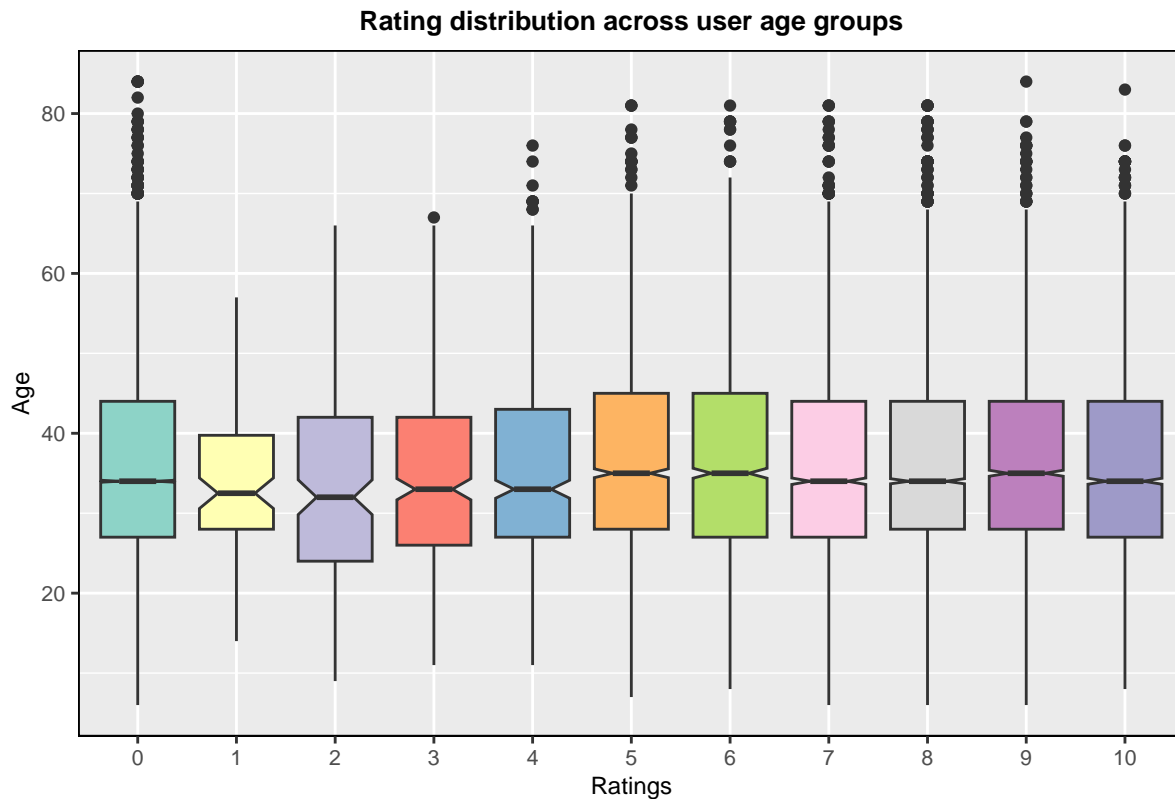
Rating distribution across age groups: The analysis of the relationship between user age and book ratings provides important insights into the reading habits of different age groups. This trend could be due to several factors. For example, younger users may be more optimistic and have a general tendency to rate things positively, while older users may have higher expectations and be more critical in their assessments. Another possible explanation is that younger users are more open to exploring new genres and authors, and as a result, are more likely to enjoy the books they read. In contrast, older users may have more established reading preferences and may be less likely to try new things, which could result in them being more critical of the books they read. It's also possible that the difference in ratings is due to a difference in reading experience.

Q. How do people from different age groups rate the top 5 authors?

The chart shows the flow of ratings for each author from one age group to another. The top 5 authors were Agatha Christie, Ann M. Martin, Nora Roberts, Stephen King and William Shakespeare. A higher proportion of older people rated William Shakespeare highly, while younger people rated Ann M. Martin more highly. The overall all-rounder with high average ratings from most of the age categories is Stephen King.

Alluvial Plot for Average Ratings of Books by Authors Across Different Age Groups
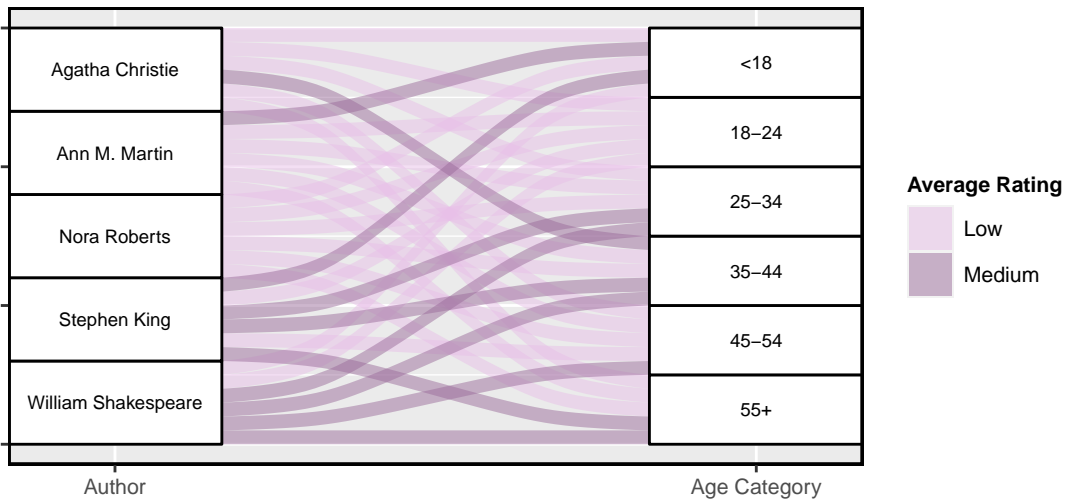

Q. From the top 5 books, what are the ratings from each category?

Top 5 Most–Rated Books: A Distribution of Ratings Breakdown
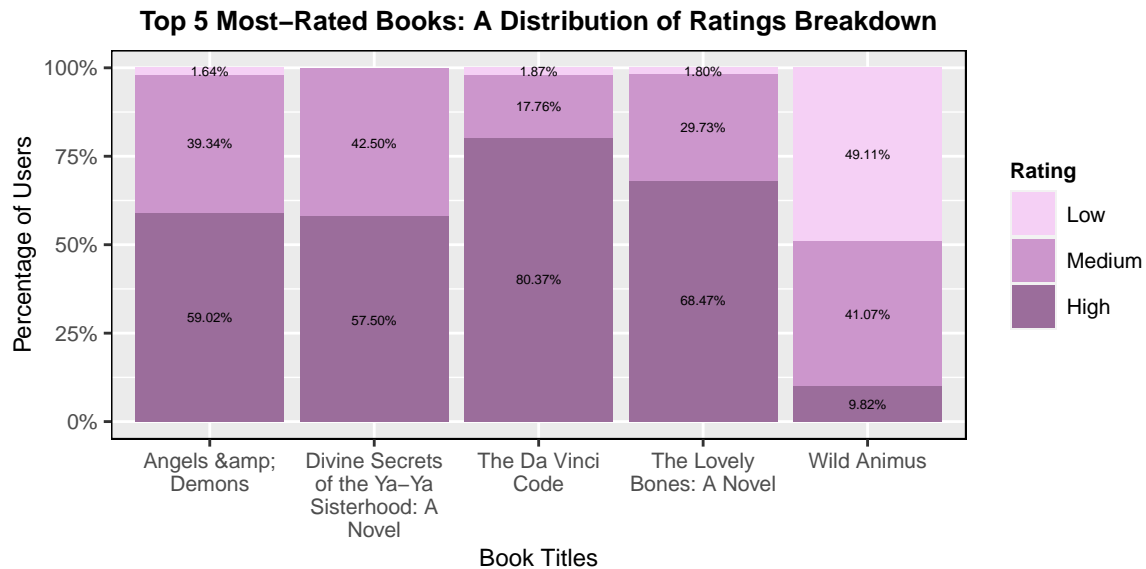
This information can provide valuable insights into the popularity of a book and its reception among the audience. Starting with Angels and Demons, Divine Secrets, the lovely bones and The Da Vinci Code the stacked bar graph shows that a high percentage of users rated it as high. This suggests that these books were not only popular but also well-received among the audience and was considered to be of high quality. The proportion of users who rated it as low is comparatively low. However, the graph tells a different story for, Wild Animus has the lowest proportion of users rating it as high. Many users rated the book low, indicating poor reception and low quality.

Q. From the top 5 countries, what are the ratings from each category?

Top 5 Countries and their Book Rating Distribution
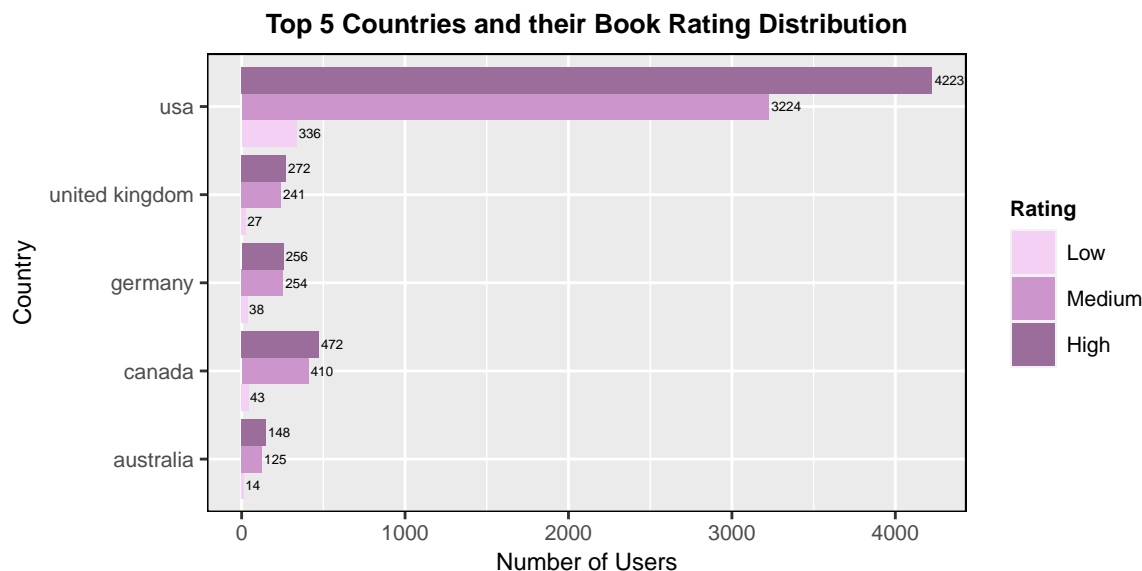
One of the key observations from the grouped bar chart is the relative proportion of low, medium, and high ratings in each country. USA has a high proportion of high ratings compared to other countries due to factors such as availability of high-quality books, strong publishing industry, or cultural factors. We could see that low proportion of low ratings in top 5 countries suggests well-developed publishing industries catering to diverse reader preferences due to large and diverse market opportunities.

Conclusion

The comprehensive analysis of book details, user demographics, and user ratings offered a profound understanding of the reading preferences and behaviors of users. Our findings shed light on the intricacies of reading habits, revealing that the most sought-after books were written by renowned authors,suggesting that the author's reputation plays a significant role in determining book popularity. These insights have significant implications for the publishing industry, library management, and readers alike.

Additionally, the study revealed that most users were in the age range of 25-34 and were in urban areas or in developed nations, suggesting that demographic factors, such as age and location, also influence book popularity. We could also observe how readers in different countries give ratings. The results showed that user ratings were positively correlated with the author and publisher of the book, as well as the user's age and location, further emphasizing the impact of these factors on book popularity these factors can be used to predict book popularity and help publishers make informed decisions about which books to promote and where to target their marketing efforts.

Future studies could build upon these findings by exploring the relationships between these factors and book popularity in more detail. For example, further research could examine the impact of marketing and advertising on book popularity or explore the relationship between book genre and reader preferences. Additionally, future studies could examine the impact of cultural differences and personal preferences on book popularity, by comparing the reading habits of users across different countries or demographic groups.

Reference

- https://www.kaggle.com/datasets/saurabhbagchi/books-dataset?resource=download