

## Book Recommender

### Intro to Topic

We will be using the Kaggle Dataset [Goodreads Best Books Ever With Recommendations](#) to create a book recommendation dashboard. The data set includes ratings, number of reviews, book description and genre.

### Inspiration

We were inspired by the [Anime Recommender](#) done previously, and wanted to do one based on books.

There are two other examples on book recommenders that we've found:

- Toward Data Science, [How Did We Build Book Recommender Systems in an Hour](#)
- The Clever Programmer, [Book Recommendation System](#)

Then there are some Tableau Public dashboards we've found:

- [Goodreads](#) by Think Design Collaborative (image a, below)
- [Top Books from past 100 years on Goodreads Shelves](#) by Joshua Preston (image b, below)
- [Trends on Goodreads](#) by Alex Austin (image c, below)
- [Goodreads](#) by Kayla (image d, below)

### Color Theme

Our color theme is based on book page color and shelving, such as the image below.



## Intended Models (“What are we predicting?”)

We would like to recommend books based on genre, ratings, and review with an unsupervised learning model, as well as using natural language processing.

Using natural language processing, we will likely generate a recommender based on similar book description sentiment scores and/or a book generator using TD-IDF.

## Roles and Responsibilities

Data Cleaning: Cody

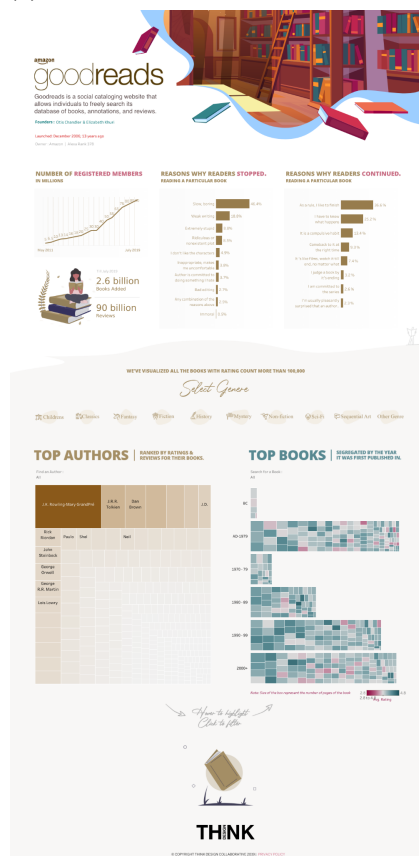
Tableau Dashboards: Cody and Lyndah

Unsupervised Learning Model: Ebrahim

Natural Language Processing: Sabrina

Images:

(a)



(b)

## Top books from past 100 years on Goodreads shelves



Below are the books with the most ratings on Goodreads based on the top 200 books published each year from 1915 to 2015 that are on readers' bookshelves.

Discover the most popular books in this community of 40 million readers; find top authors and their most acclaimed work; and maybe add a few books to your TBR.

Explore the DATA

SEARCH  
All

AVG RATING  
0 to 5

DECADE  
All



4.37  
Avg. Rating

### Rank (by # of user ratings)

- 1 The Hunger Games (The Hunger Games, #1) (3806017 ratings) 2008
- 2 Harry Potter and the Sorcerer's Stone (Harry Potter, #1) (3520776 ratings) 1997
- 3 Twilight (Twilight, #1) (3162081 ratings) 2005
- 4 To Kill a Mockingbird (2568431 ratings) 1960
- 5 The Great Gatsby (2224822 ratings) 1925
- 6 The Hobbit (1807431 ratings) 1937
- 7 The Catcher in the Rye (1738988 ratings) 1951
- 8 The Fault in Our Stars (1706537 ratings) 2012
- 9 Catching Fire (The Hunger Games, #2) (1704700 ratings) 2009
- 10 Divergent (Divergent, #1) (1634889 ratings) 2011
- 11 Animal Farm (1613557 ratings) 1945
- 12 The Diary of a Young Girl (1609947 ratings) 1947

### Authors (by # of books)

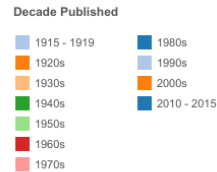
[Click to View their books](#)

Enid Blyton	116
Agatha Christie	115
H.P. Lovecraft	83
Nora Roberts	81
Stephen King	77
P.G. Wodehouse	76
Louis L'Amour	62
Carolyn Keene	61
Edgar Rice Burrou..	61
Robert A. Heinlein	60

### Average Rating (for top rated books)



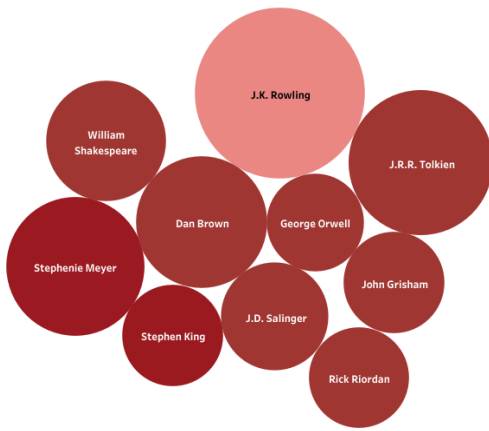
COMPARE AUTHORS  
(Scatterplot will be filtered)



Created by: Joshua Preston, <http://reignofbooks.com>  
Source: [https://www.goodreads.com/book/popular\\_by\\_date/](https://www.goodreads.com/book/popular_by_date/)

(c)

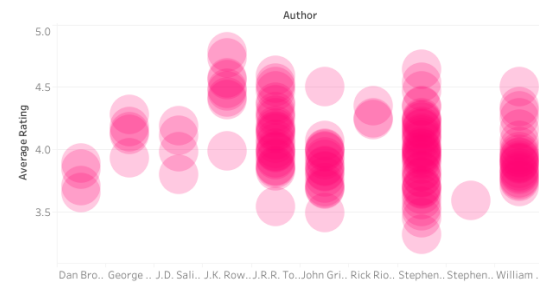
Most Popular Authors on Goodreads



What's the average length of a book by one of most popular authors?



Average Ratings for Most Popular Authors by Book



How does page count relate to a book's rating?

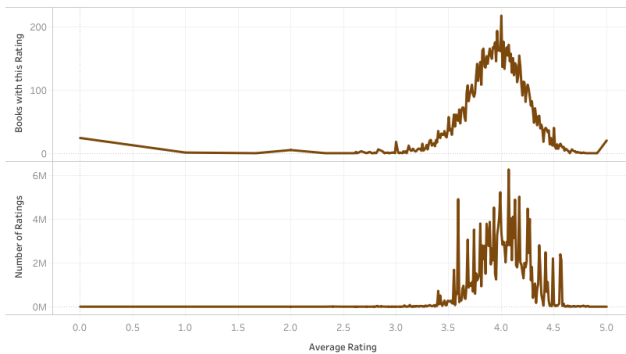


(d)

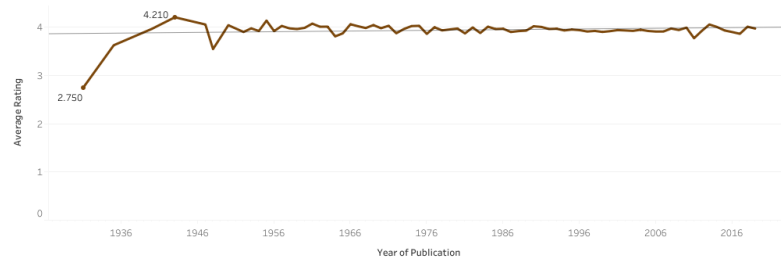
Average Rating Overall

Average Rating	3.93
Num of Books	11,112
Num of Ratings	199,531,931

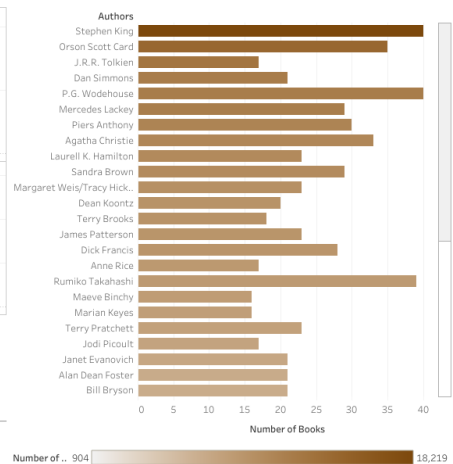
Most Common Ratings



Average Ratings by Publication Year from 1931-2019



Authors (>15 Books) with Most Pages Written



Number of Books (Per Author) From 15

How To Use

You can filter the authors displayed by number of books they've written. Here, we start with authors who have written at least 15 books, although you can raise or lower that number by clicking on the left end of the slider and adjusting accordingly.

Note: The visualization is ordered by authors who have written the most pages, not who have written the most books, but you can adjust this by clicking the down arrow next to the author name and either sorting alphabetically by author name or by data field: number of records to filter.