Reader Preferences and Book Recommendations

By Melanie Ackerman

Abstract

The goal of this project is to propose the use of clustering algorithms to segment reader audiences of Penguin Random House (PRH) books. By defining groups of readers, this will aid decision-makers at PRH in content and marketing choices to appeal to as wide an audience as possible while maintaining integrity to the author and genre. This data will also inform a personalized recommendation system that readers can use to find their next read within PRH's book catalog. I used publicly available data from goodreads.com to produce summary visualizations that introduce this opportunity.

Design & Algorithms

While I propose the use of clustering algorithms in this project, this presentation serves to introduce the concept and only contains descriptive statistics. Visualizations include the distribution of average ratings, top 10 books by rating, top 10 most popular books (by number of ratings), genre breakdowns, and top books within genres.

Data

I used a dataset with information from goodreads.com that contains ratings for 10,000 popular books, books marked "to read" by users, book metadata (e.g. author, year, etc.), and Goodreads tags/shelves/genres. The data is publicly available at this link:
https://github.com/zygmuntz/goodbooks-10k

Tools

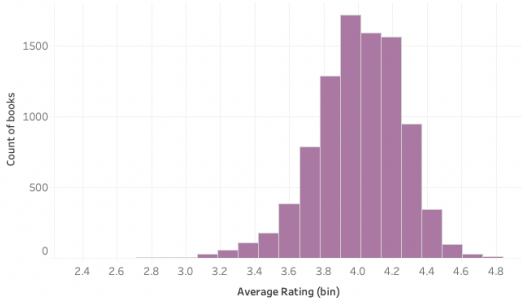I used Microsoft Excel and Python to clean the data and Tableau to produce the visualizations.

Communication

I produced slides containing the following visuals (as mentioned in the Design & Algorithms section above):
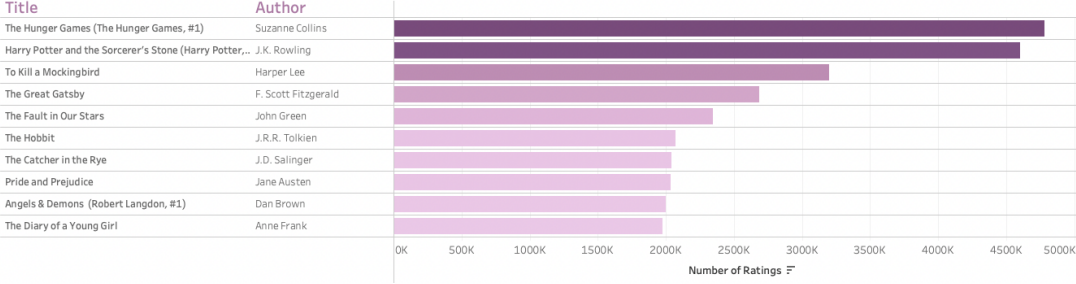
Top 10 Books by Rating

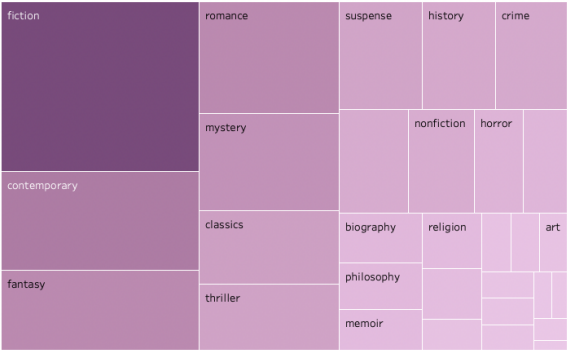| Title	Author	Avg. Rating
The Complete Calvin and Hobbes	Bill Watterson	4.82
Words of Radiance (The Stormlight Archive, #2)	Brandon Sanderson	4.77
Mark of the Lion Trilogy	Francine Rivers	4.76
It's a Magical World: A Calvin and Hobbes Collection	Bill Watterson	4.75
Harry Potter Boxset (Harry Potter, #1-7)	J.K. Rowling	4.74
There's Treasure Everywhere: A Calvin and Hobbes Collection	Bill Watterson	4.74
Harry Potter Collection (Harry Potter, #1-6)	J.K. Rowling	4.73
The Authoritative Calvin and Hobbes: A Calvin and Hobbes Tr..	Bill Watterson	4.73
The Indispensable Calvin and Hobbes	Bill Watterson	4.73
A Court of Mist and Fury (A Court of Thorns and Roses, #2)	Sarah J. Maas	4.72
Attack of the Deranged Mutant Killer Monster Snow Goons	Bill Watterson	4.72

Avg. Ratings Distribution

Top 10 Most Popular Books

Title	Author
The Hunger Games (The Hunger Games, #1)	Suzanne Collins
Harry Potter and the Sorcerer's Stone (Harry Potter,...	J.K. Rowling
To Kill a Mockingbird	Harper Lee
The Great Gatsby	F. Scott Fitzgerald
The Fault in Our Stars	John Green
The Hobbit	J.R.R. Tolkien
The Catcher in the Rye	J.D. Salinger
Pride and Prejudice	Jane Austen
Angels & Demons (Robert Langdon, #1)	Dan Brown
The Diary of a Young Girl	Anne Frank

All Genres

Fiction Sub-Genres

Top 10 Most Popular Books by Genre

Genre: fiction
Sub-Genre: fiction