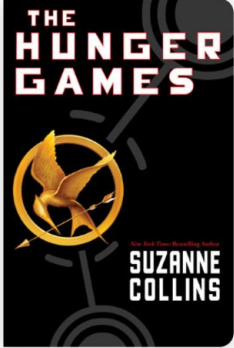


# Predicting Average Rating

From Goodreads

# Webscrape - Extracting Data

goodreads Home My Books Browse Community Search books Sign in



THE HUNGER GAMES  
SUZANNE COLLINS

**The Hunger Games #1**  
**The Hunger Games**  
Suzanne Collins  
4.34 9,425,596 ratings · 242,291 reviews  
The first novel in the worldwide bestselling series by Suzanne Collins!  
Winning means fame and fortune. Losing means certain death. The Hunger Games have begun. . .

Show more

Genres Young Adult Dystopia Fiction Fantasy Science Fiction Romance Adventure ...more

374 pages, Hardcover  
First published September 14, 2008

Literary awards [Locus Award Nominee for Best Young Adult Book \(2009\)](#), [Georgia Peach Book Award \(2009\)](#), [Buxtehuder Bulle \(2009\)](#), [Golden Duck Award for Young Adult \(Hal Clement Award\) \(2009\)](#), [Grand Prix de l'Imaginaire Nominee for Roman jeunesse](#)

Show more

Original title The Hunger Games

Want to read  
Shop this series  
Rate this book

Top 3 Genres

Title

Author

Avg Rating & Number of Ratings

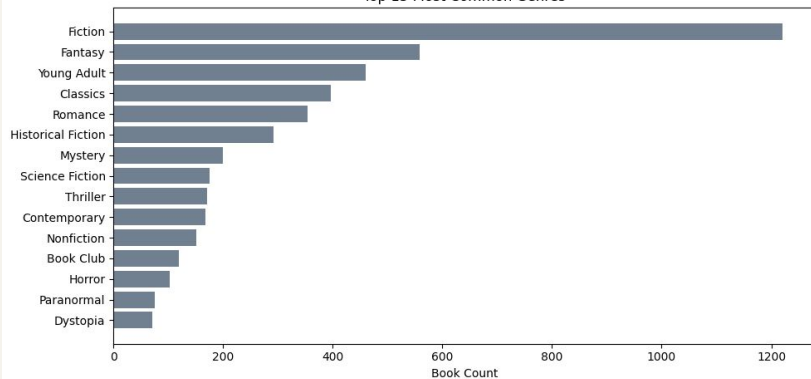
Number of Pages

Publication Year

# Exploratory Analysis

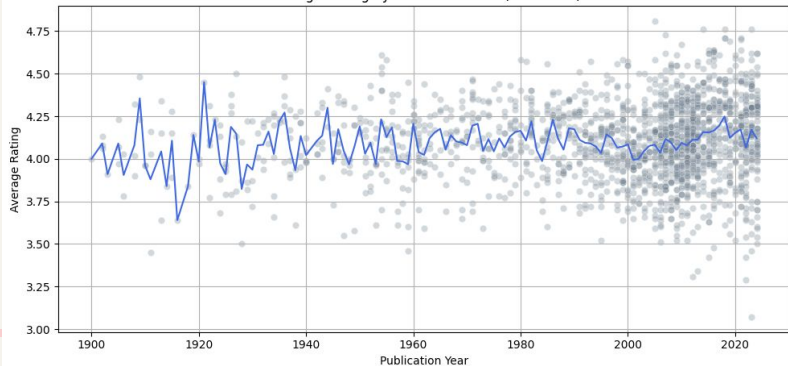
1996 Rows, 7 Columns

Top 15 Most Common Genres



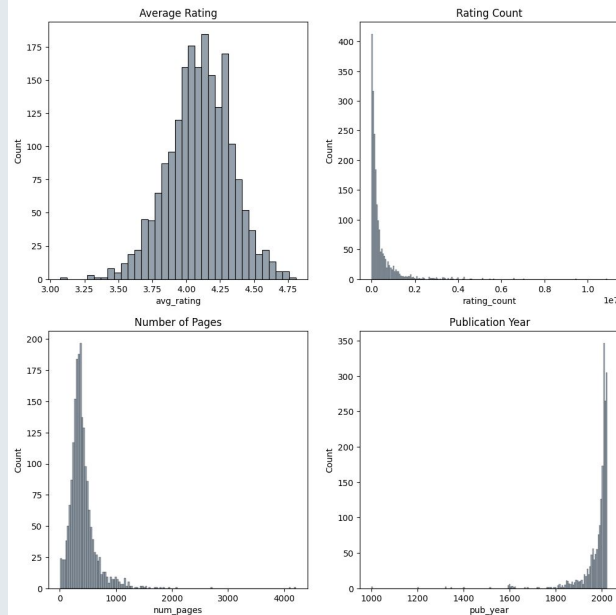
- *Fiction* dominates the dataset, followed by *Fantasy*, *Young Adult*, and *Classics*

Average Rating by Publication Year (Post-1900)



- Ratings remain **stable** between 4.0 and 4.3
- There is **higher variance** in older publication years (1900–1950)
- Post-2000, there's a **much denser concentration of books** with ratings centered around 4.1


Histograms of Numerical Features




- Avg rating: Nearly normal distribution centered at ~4.1
- Rating\_count, num\_pages, and pub\_year: All highly skewed
- Log transformations may be needed for modeling due to skew

# Project Proposal


---



**Objective:** Analyze Goodreads book data to explore how features like publication year, page count, and genre relate to average book ratings



**Approach:** Engineer additional features such as book age, log-transformed variables, top genre indicators, author popularity, and interaction terms to enhance model performance



**Next Steps:** Build a regression model (e.g., linear regression, random forest) to predict average ratings, and analyze which features (e.g., genre, length, popularity) most influence reader satisfaction