

# Predicting Book Demand on Overdrive

Sam Blass

Web-scraping and Linear Regression Module  
26 Jan 2022



# Business case

Help libraries **purchase the right number of ebooks** by predicting a book's demand on Overdrive with popularity data from Goodreads



**Popularity**

- **The Hunger Games (The Hunger Games, #1)**  
by Suzanne Collins  
★★★★★ 4.32 avg rating — 7,200,216 ratings  
Vote For This Book score: 3,220,038, and 32,799 people voted  
Want to Read ▼  
Rate this book  
★★★★★
- **Harry Potter and the Order of the Phoenix (Harry Potter, #5)**  
by J.K. Rowling  
★★★★★ 4.50 avg rating — 2,859,005 ratings  
Vote For This Book score: 2,806,037, and 28,683 people voted  
Want to Read ▼  
Rate this book  
★★★★★



HOLDS INFORMATION

Wait time: About 10 weeks

Library copies: 31

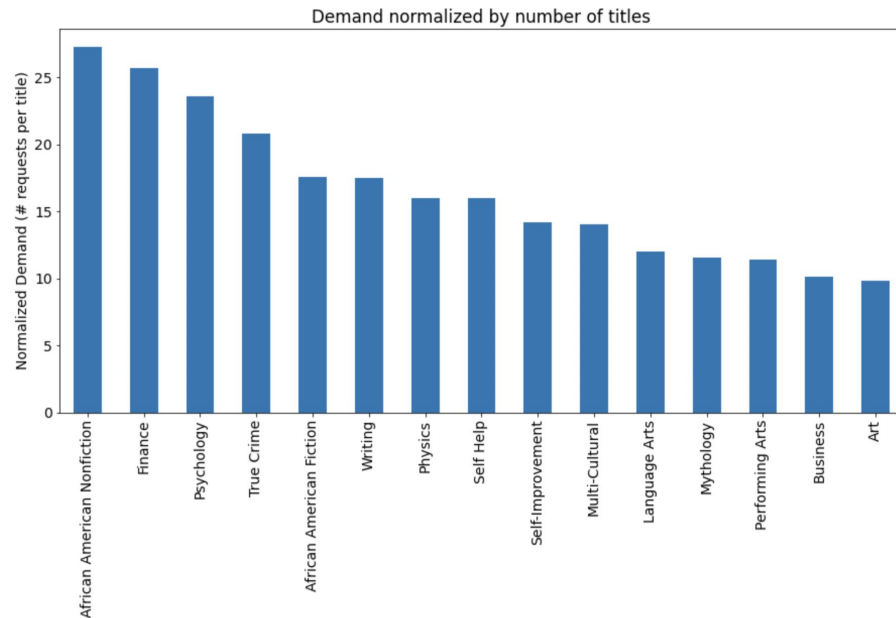
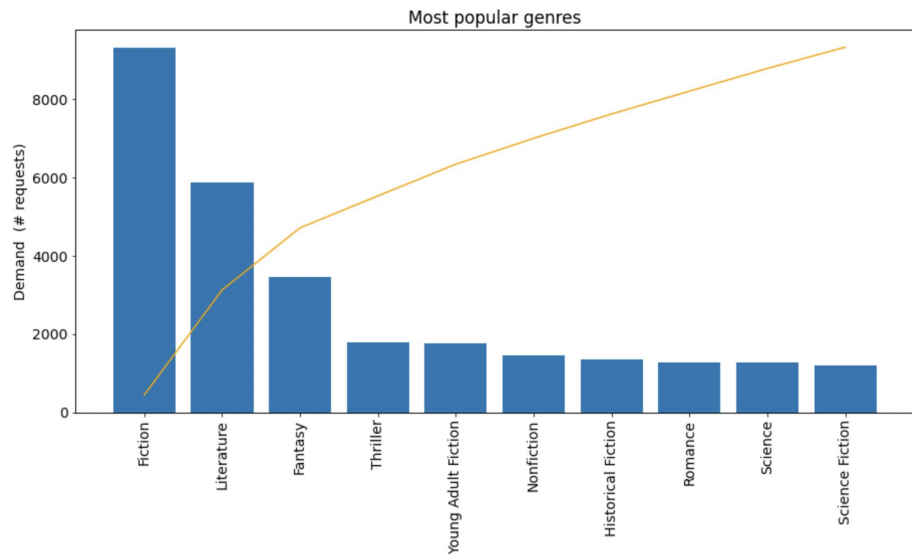
People waiting in total: 141

People waiting per copy: 5

**Demand**

# Most in-demand genres

Which genres have greatest demand? Which genres should libraries target?



# Features that predict demand

- Weak correlations for all features
- “Number of ratings” has time-dependency
- Normalizing by book age improves correlation from 15% to 30%
- Older books have less demand



**The Hunger Games (The Hunger Games, #1)**  
by Suzanne Collins

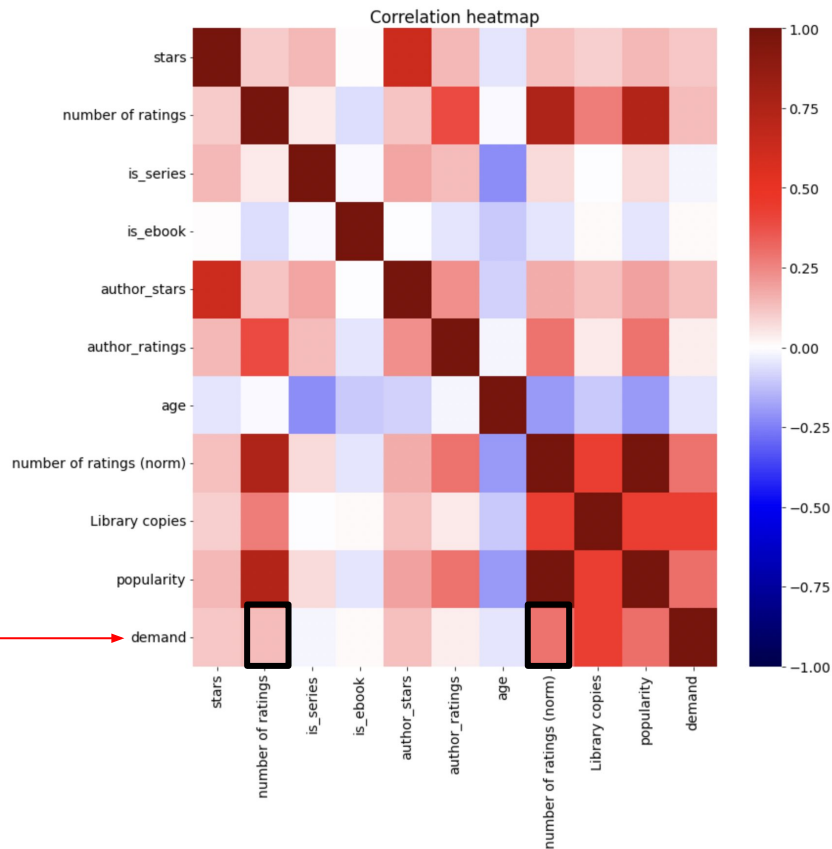
★★★★★ 4.32 avg rating

7,200,216 ratings

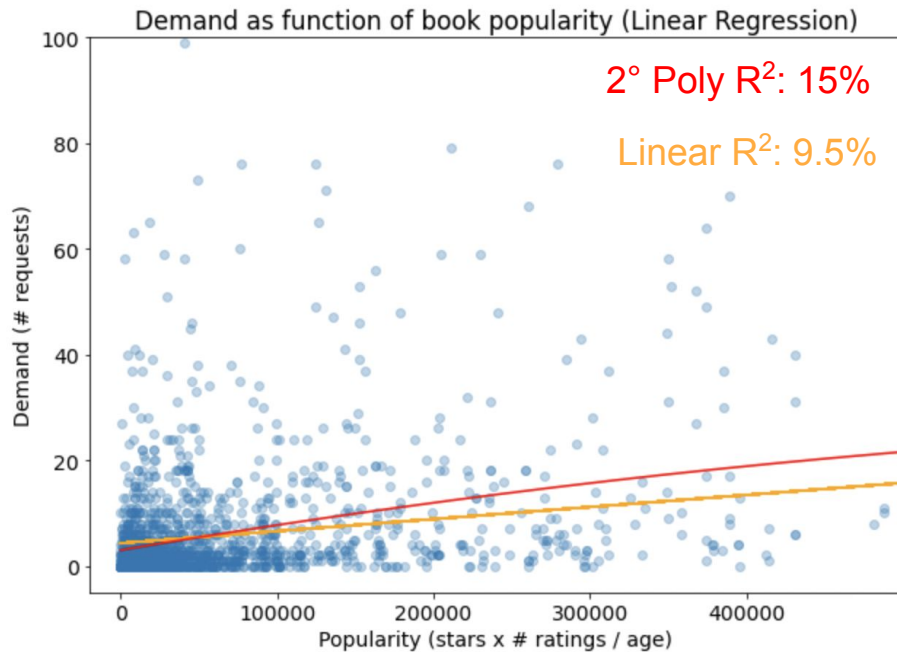
Vote For This Book

score: 3,220,038, and 32,799 people voted

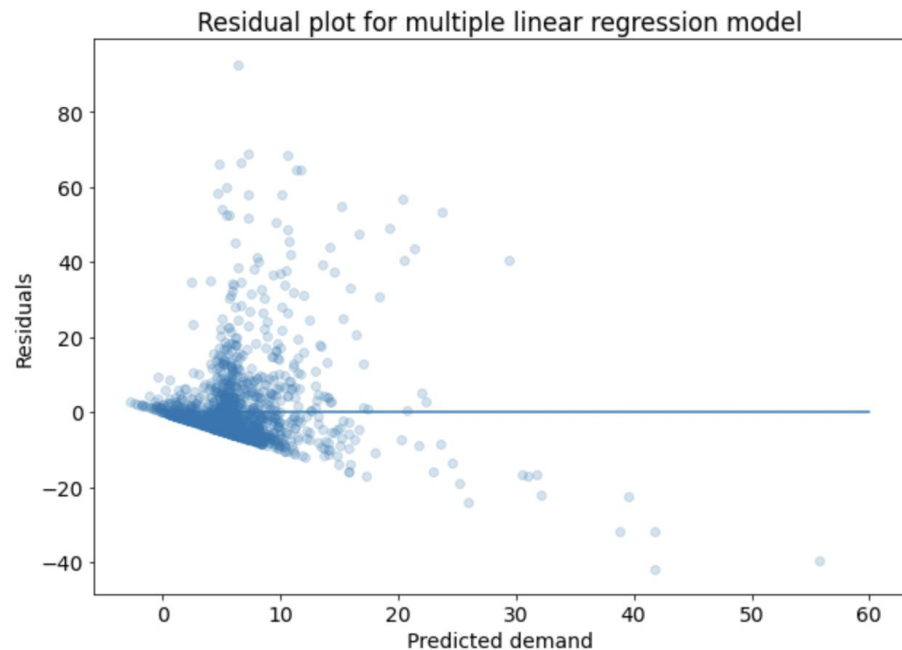
$$\text{Popularity} = \frac{\text{stars} * \text{ratings}}{\text{age}}$$



# Linear Regression results



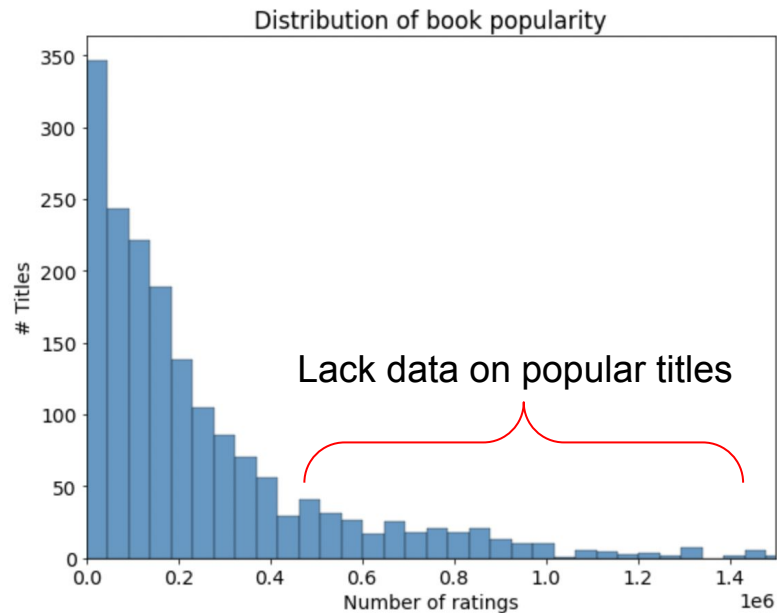
Combining two time-dependent features into one



Heteroskedasticity + unbalanced

# Why the model fails

- Missing historical demand info
- Author popularity measurement is time-dependent
- Lack of data for popular titles (power law)
- Errors in scraping (some old books are re-published so appear newer)
- Customers may select books by browsing rather than searching



**“80-20 rule”**

20% of books have 80% of ratings

# Conclusions and Recommendations

- Results can help libraries meet demand by assessing which genres of books have most demand
- Model needs historical information on demand, title popularity, and author popularity