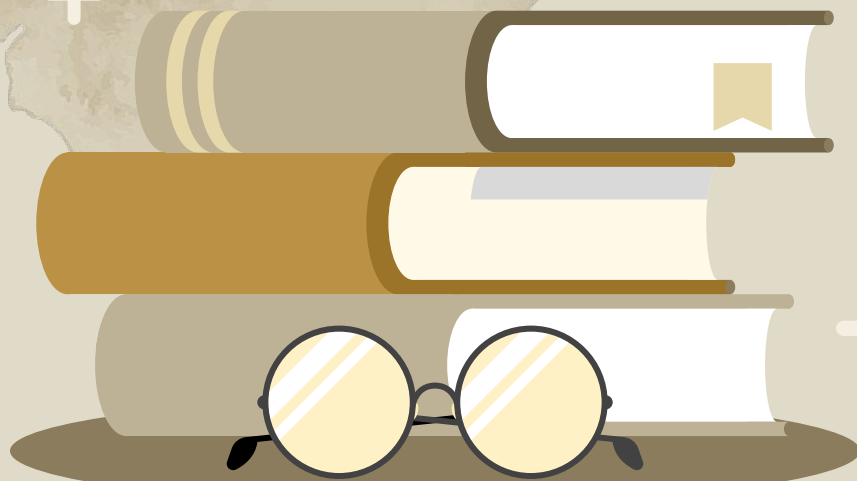


# GoodReads Recommender System



Shelley Wang

<https://github.com/shelleylwang>

# About Me

## The serious stuff:

- Economics B.A.
- Financial Modelling at Bloomberg Global Data
- Data Science at Flatiron

## The silly stuff:

- Rock climber
- Artist
- Reader



# Table of contents

**01.** Background + Research  
Question

**04.** Content-Based  
Recommender

**02.** Data Sources

**05.** Results

**03.** Data Prep

**06.** Next Steps

# 01

## Background + Research Question



# GoodReads

goodreads Search books

Home My Books Browse Community

**Pride and Prejudice**

Jane Austen, Anna Quindlen (Introduction)

★ ★ ★ ★ ☆ 4.29

4,239,050 ratings · 113,209 reviews

Since its immediate success in 1813, *Pride and Prejudice* has remained one of the most popular novels in the English language. Jane Austen called this brilliant work "her own darling child" and its vivacious heroine, Elizabeth Bennet, "as delightful a creature as ever appeared in print." The romantic clash between the opinionated Elizabeth and her proud beau, Mr. Darcy, is a splendid performance of civilized sparring. And

Show more

Genres Classics Romance Fiction Historical Fiction Historical

Literature Audiobook ...more

279 pages, Paperback

First published January 28, 1813

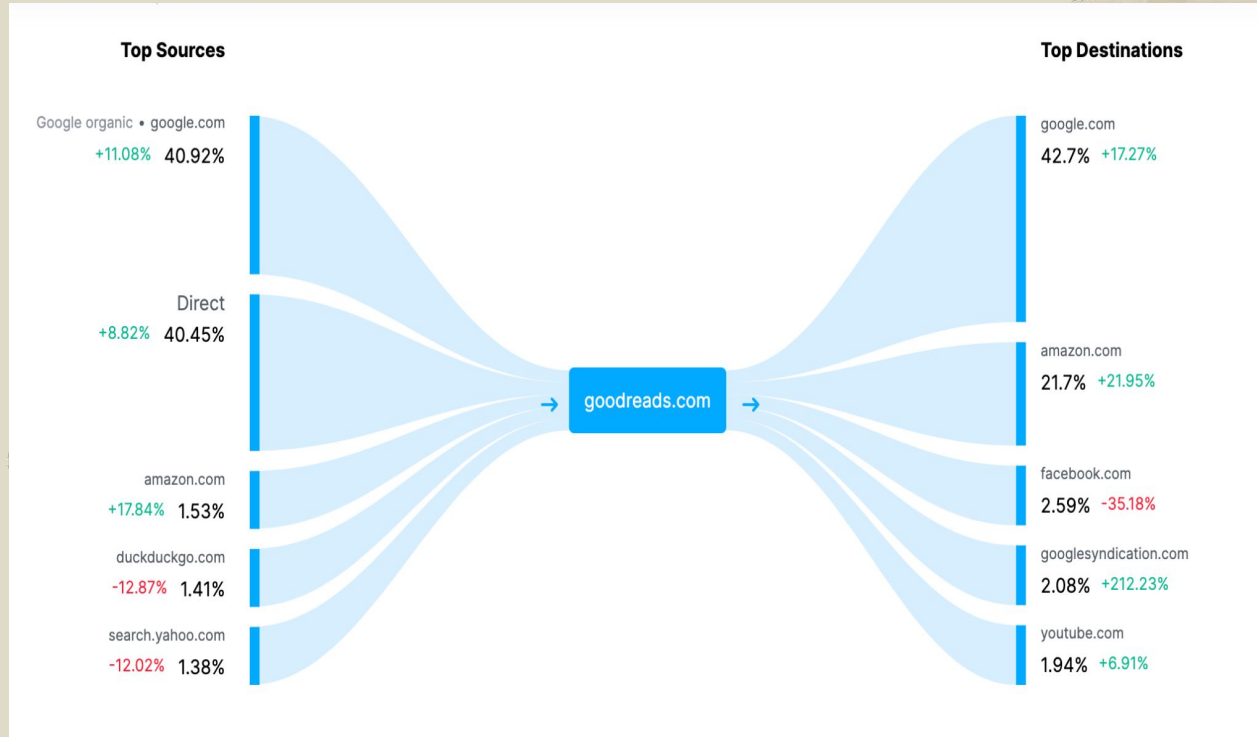
Want to read

Kindle \$2.99

☆☆☆☆

- User ratings and reviews site
- Popular and growing

# Critical to Book Sales



- Intermediary that drives book purchases
- Book sales industry is growing

The illustration features a brown book with a white label displaying the number '02'. A yellow quill pen with a brown nib lies in front of the book. To the right, a light brown map of Australia is shown. Three white plus signs are positioned above the map. The entire scene is set against a light beige background with a thin brown border.

02

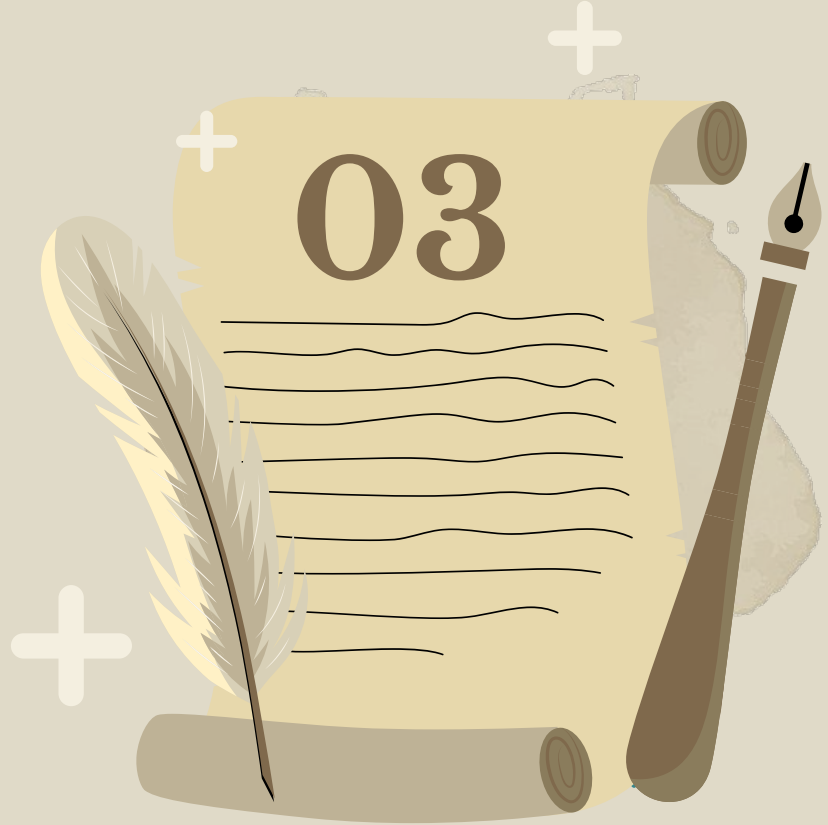
Data Sources

# 4 Datasets





# Data Prep



# Metadata



Book ID	Link	Average Rating	Description	ISBN	Author ID	Page #	Ratings Count	Title	Publication Year
---------	------	----------------	-------------	------	-----------	--------	---------------	-------	------------------

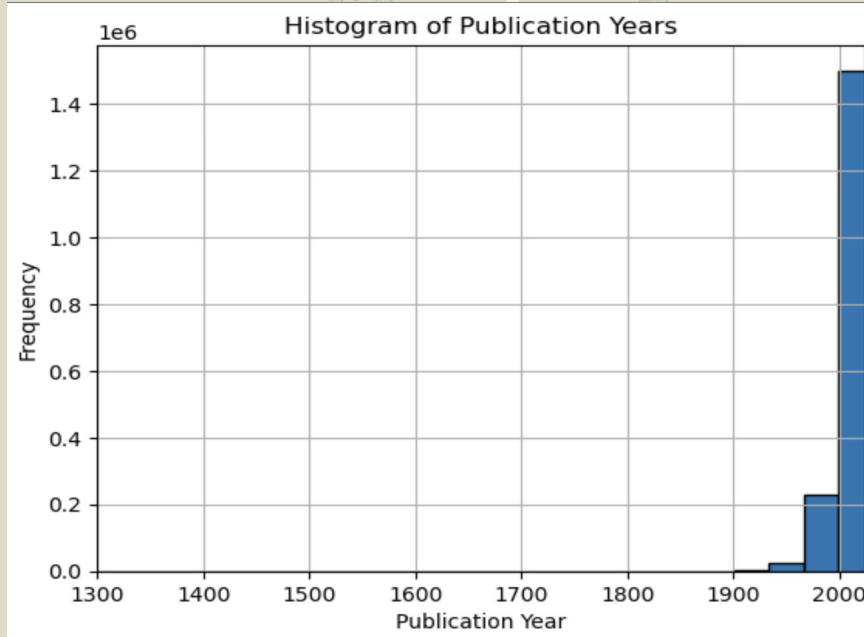
Book ID	Author Name
---------	-------------

Book ID	Genres
---------	--------

# Metadata (Cont)



Book ID	Link	Avg Rating	Descrip	ISBN	Author ID	Page #	Ratings Count	Title	<del>Publication Year</del>	Genre
---------	------	------------	---------	------	-----------	--------	---------------	-------	-----------------------------	-------



# Metadata (Cont)



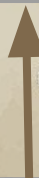
Book ID	Link	Avg Rating	Descrip	ISBN	Author ID	Page #	Ratings Count	Title	Genre
---------	------	------------	---------	------	-----------	--------	---------------	-------	-------



# Model Dataset



Book ID	Avg Rating	Description	Page #	Ratings Count	Genre
---------	------------	-------------	--------	---------------	-------



Book ID	User ID	User Review
---------	---------	-------------



# 04

## Content-Based Recommender+



# Processing Model Dataset



Scaling Numeric Columns  
(Normalization)



One-Hot Encoding Genres



NLP Text Data

# A Closer Look at NLP



Removing  
Stopwords,  
Punctuation,  
Digits



Lemmatizing



Tokenizing



# TF-IDF Vectorization

	action	actually	adventure	age
book_id				
1333909	0.0	0.0	0.00000	0.0
6066819	0.0	0.0	0.23184	0.0

# The Recommendation Function





05

Results

# Recommendations 1

“Two women friends, husbands, daughters”

Vinegar Hill

A wife and husband  
threatened

Heaviness  
of Things  
That Float

A woman with no  
family, a man like a  
son, upheaval and  
pain

The House  
on Swiss  
Ave

Two women with  
dark secrets,  
devoted to family

The  
Waiting  
Place

A pregnant woman  
and wife on the  
family farm

## Recommendations 2

“The Tale of Jemima Puddle-Duck”

```
graph TD; A["The Tale of Jemima Puddle-Duck"] --- B["The Tale of Benjamin Bunny"]; A --- C["The Tale of Mrs. Tiggy-Winkle"]; A --- D["Peter Rabbit Let's Cuddle"]; A --- E["The Tale of Ginger and Pickles"];
```

The Tale of  
Benjamin  
Bunny

The Tale of  
Mrs.  
Tiggy-Winkle

Peter  
Rabbit Let's  
Cuddle

The Tale of  
Ginger and  
Pickles

# Next Steps:

- Additional features available from GoodReads
- Applicable to other sites
- Collaborative Filtering



# Thank You

## Citations:

Mengting Wan, Julian McAuley, "Item Recommendation on Monotonic Behavior Chains", in RecSys'18. [bibtex]

Mengting Wan, Rishabh Misra, Ndapa Nakashole, Julian McAuley, "Fine-Grained Spoiler Detection from Large-Scale Review Corpora", in ACL'19. [bibtex]

SEMrush. "Goodreads.com - Overview." SEMrush,

<https://www.semrush.com/website/goodreads.com/overview/>

McLean, Kristen. "Book Sales in the U.S. Are Stronger Than Ever." Publishers Weekly, 19 Apr. 2022,

<https://www.publishersweekly.com/pw/by-topic/industry-news/bookselling/article/89040-book-sales-in-the-u-s-are-stronger-than-ever.html>