



Group 14
Capstone Project

AMAZON BOOKS RECOMMENDATION SYSTEM

Presented by: All members

OVERVIEW

- Introduction
- Business and Data understanding
- Modelling techniques
- Visualizations
- Findings and conclusions





BUSINESS UNDERSTANDING

Problem Statement

- Customers prefer the convenience of online purchases.
- Online retailers offer an extensive range of products.
- However navigating through endless catalogs can overwhelm customers.
- Hence, a growing demand for recommendation systems to streamline the buying process.
- Here, we shall be focusing on the Amazon books dataset reflecting the evolving reading habits.
- Such a system would benefit both retailers and customers.
- Providing tailored suggestions for ideal and trending books.

DATA UNDERSTANDING

- Data is sourced from <https://amazon-reviews-2023.github.io/> in JSONL format, ideal for unstructured or time-produced data.
- The dataset comprises Amazon book listings, originally spanning 1996 to 2023 with 4 million rows.
- For ease of use, we trimmed it down to the most 300k entries from the year 2023.
- Vital metadata was missing from the data (price, book title).
 - This was obtained by merging the data with the metadata.
- The data contains the following features/ columns in the dataset:

text	Text body of the user review.
images	Links to images (comma-separated if multiple).
asin(product key)	Unique identifier for the product.
parent_asin	Identifier for the parent product (applicable for variations).
user_id	Unique identifier for the reviewer.
timestamp	Date and time of the review.
helpful_vote	Number of helpful votes received by the review.
verified_purchase	Indicates whether the reviewer purchased the product (True/False).
main_Category	Main category (domain) to which the product belongs (e.g., Electronics, Clothing).
title_y	Name of the product as mentioned in the review.
price	Price of the product in US dollars.

VISUALIZATIONS

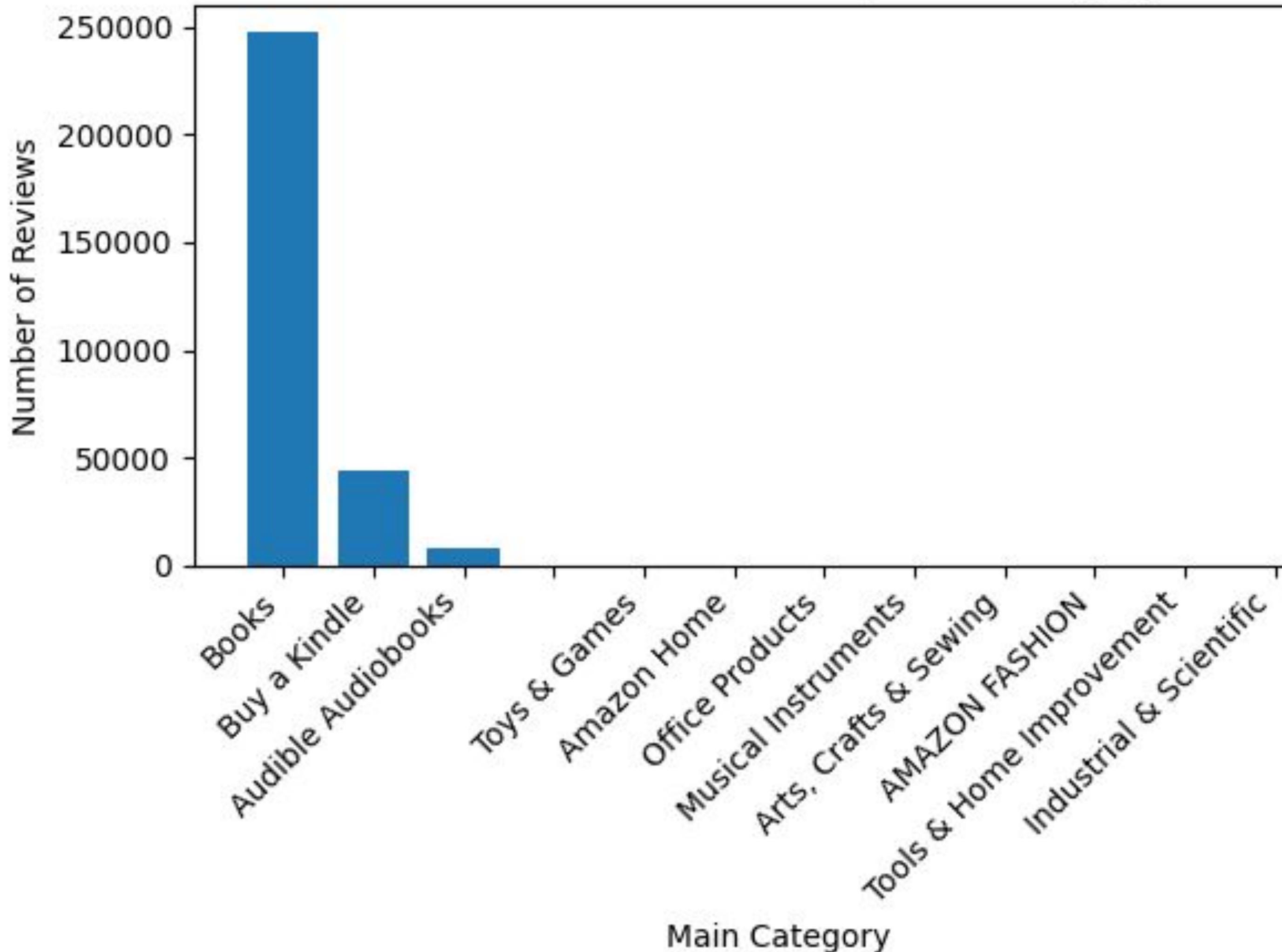
BARGRAPH



REVIEWS BY MAIN CATEGORY

The plot indicates that the majority of books fall into the categories of Books, Buy a Kindle, and Audiobooks. Other categories contain fewer than 30 books each, with 182 books remaining uncategorized. Due to the inaccurate representation of book genres in the Category column, it will be removed from the analysis.

Distribution of Reviews by Main Category



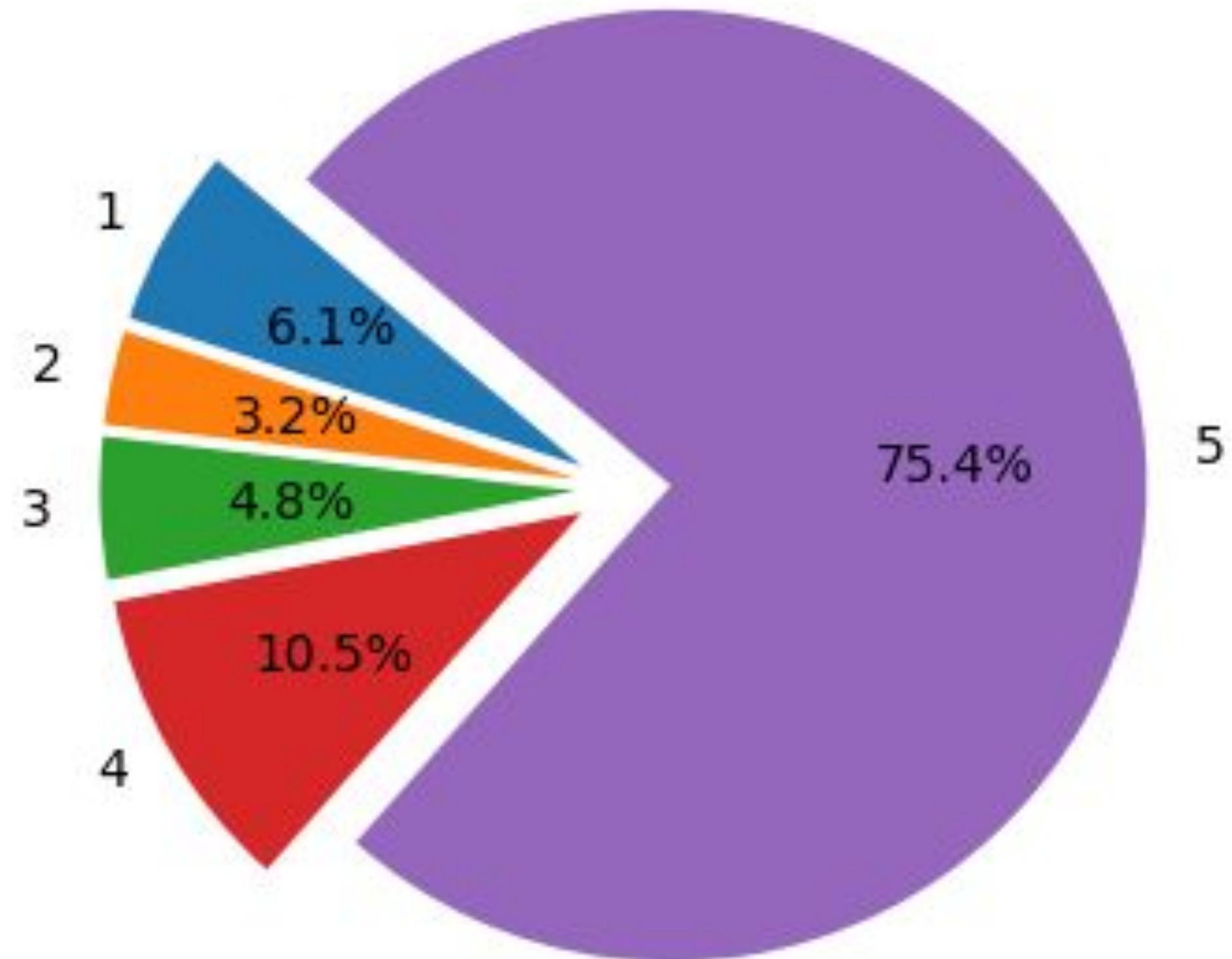
PIECHART



DISTRIBUTION OF RATINGS

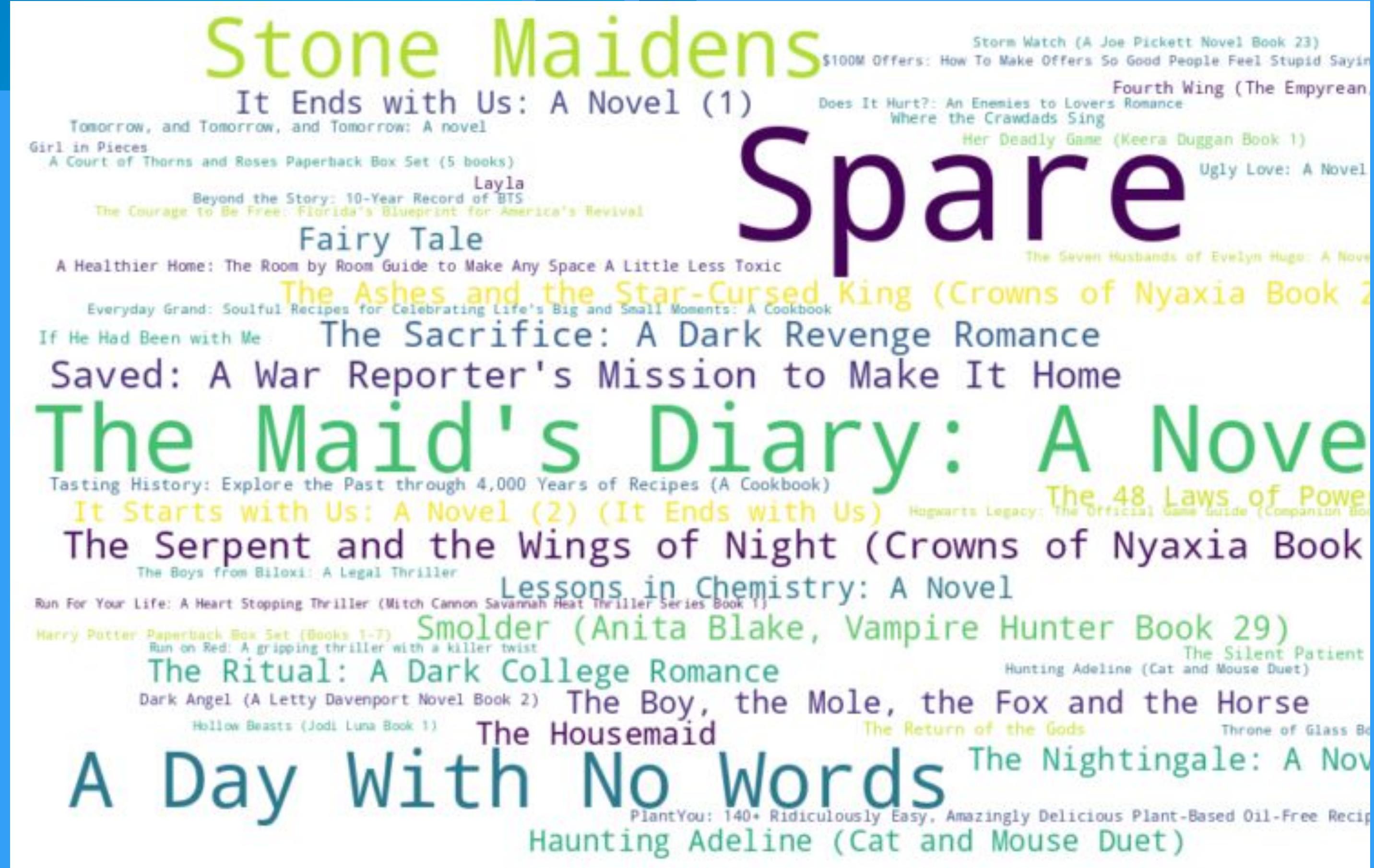
Books with a rating of 5 take up 74.8% of the data. This necessitates for more features to be used in recommendation, that is title for rating and rating text.

Distribution of Ratings



WORDCLOUD

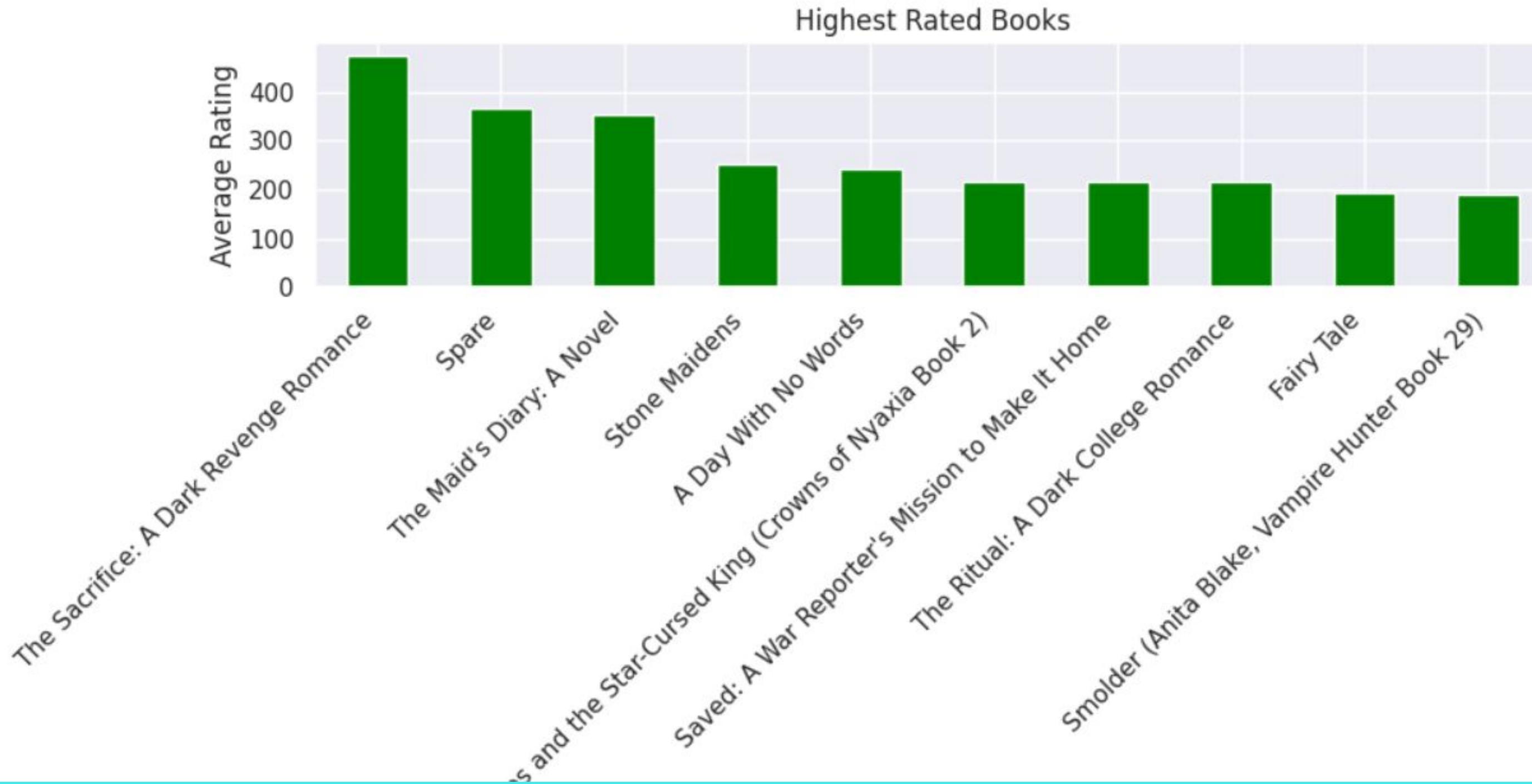
 MOST PURCHASED BOOKS
This wordcloud visualizations displays the most purchased books in the platform.



BARGRAPH

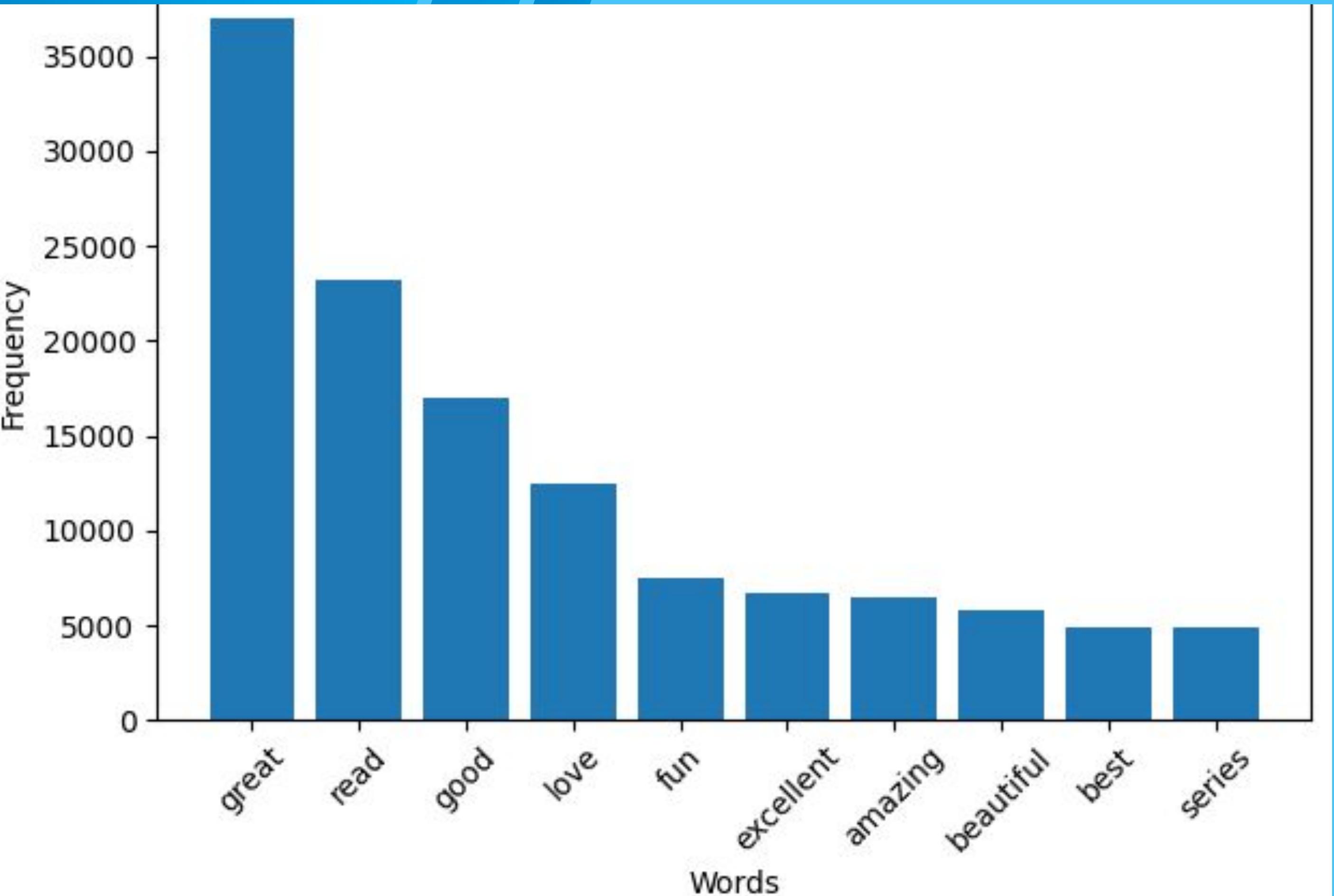


TOP 10 RATED BOOKS



BARGRAPH

MOST COMMON WORDS



WORDCLOUD



This wordcloud visualization displays the books with common word 'great' on reviews.

Fatigues 301 for Kids: The ABC of Harry - The Professional Polar Bear's Handbook (Expanded In My Mind)

[Just a Kid from Seampoodie to Vietnam](#)

Wild Card (A Sinatra Thriller)

123 Counting Sticker Book (My Little World)

Emeril Lagasse Power Air Fryer 360 Cookbook 2022: The Ultimate Everyday Delicious Days of Power Air Fryer 360 Recipes
Beanies Tracker Third Edition

Perfect Sous Vide At Home Cookbook: Compatible with Anova & Best Sous Vide Cookers - 101 Restaurant-Quality Recipes Your Family Will Love

[11 Spooky Writing Prompts for Kids](#) | [Smart Writing Questions](#) | [Creative Writing Prompts](#) | [Opinion Writing Prompts](#) | [Descriptive Writing Prompts](#) | [Narrative Writing Prompts](#)

Is East Africa

Weight Loss Journal for Women: Simple 10-Week Food and Fitness Logbook / Motivational Diet and Exercise Planner with Daily Health and Beauty Affirmations / Weight Loss and Fitness Tracker

Captivating Expanded Edition: Unveiling the Mystery of a Stream's

Lucie Rose's Resolution 5

The Megalodon Mix-Up (A Charlie Rhodes Cozy Mystery Book 4)

The Megalodon Mix-Up (A Charlie Rhodes Cozy Mystery Book 1)

Dear Mr. Bear: Women, Motorcycles and the Rapture of the Road, New

Latin/Latin Theory: Entering the Latin

1001 All-Natural Secrets to Pest Control (If They Are FLYING CRAWLING BURROWING OR SNEAKING IN THIS BOOK HAS THE SOLUTION)

Specialty: Shari's Once Was a Host Called Tranquility

Ancient Civilizations: Lamentations and Magic book 1

Dead Fall: A Thriller (The Scot Harvath Series Book 2)

Anatomy for the Artist

[Toddler's First Coloring Book Ages 1-4: Amazing Coloring Book for Toddlers and Kids Ages 1, 2, 3 & 4](#) | Over 100 simple pictures for coloring, doodling ... fruits and everyday objects to color for kids

our and Ready to Explore: Taking on Chores and Reading Galore

THE RIVER'S EDGE a gripping crime thriller full of stunning twists (JACKMAN & EVANS Book 10)

Writing Gatsby: The Real Story of the Writing of the Greatest American Novel

[How to Train and Tame: A totally addictive W! fantasy romance \(part of Western Romance\)](#)

CONCLUSION & RECOMMENDATIONS

CONCLUSION

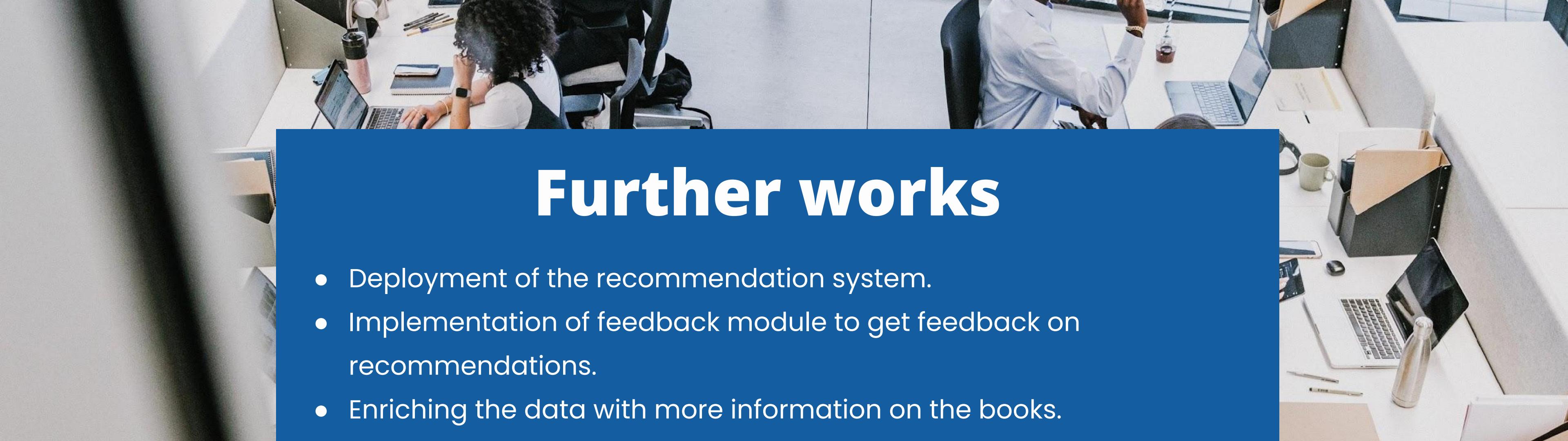
- The comments on the rating of the books provided valuable insights to build a recommendation system.
- The analysis of data may be affected by ambiguity where words have multiple meanings.
- After conducting thorough analysis, it is evident that the Hybrid Recommendation system excels beyond compare.



RECOMMENDATION

1. Enhance the metadata to include data on the genre and author of books to create a more comprehensive recommendation system.
2. Carry out further analysis of the books not highly recommended to understand what makes them unique.
3. Include vital meta data on the books purchased data to reduce time spent on data analysis.





Further works

- Deployment of the recommendation system.
- Implementation of feedback module to get feedback on recommendations.
- Enriching the data with more information on the books.

OUR TEAM

Our team is made up of meticulous data analysts, including one lady and four gentlemen.



Jacky



Johnmark



Nicholas



Peter



Mark



Group 14
Capstone Project

Thanks For Listening!

