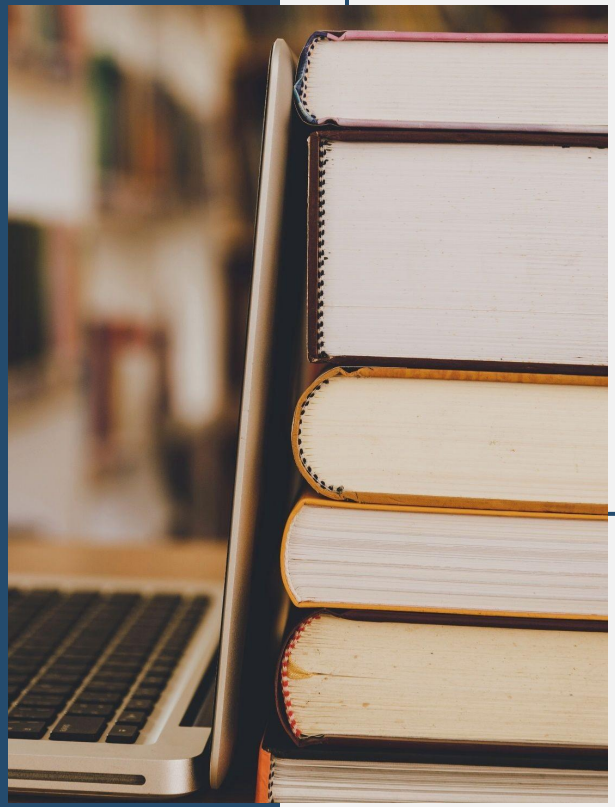


JUMIA BOOK RECOMMENDATION SYSTEM

Authors:

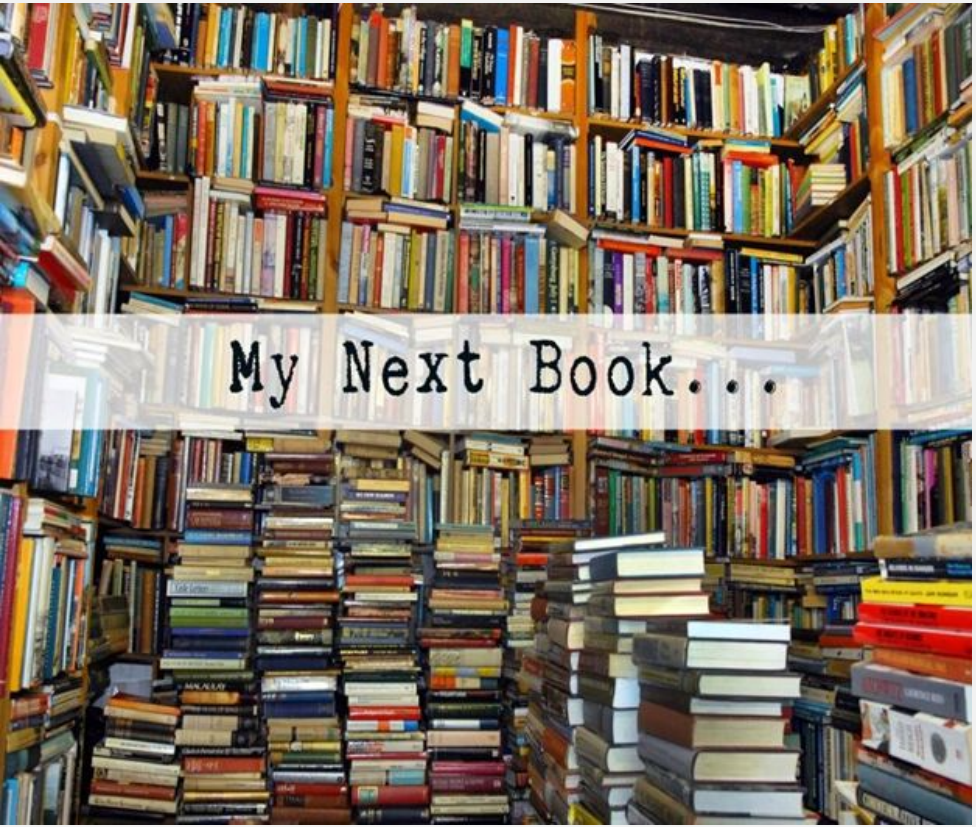
- Alice Wamuyu
- Eva Moisasi
- Fridah Kimathi
- Nicholus Magak

OVERVIEW

The project aims to develop a machine learning model that can provide personalized book recommendations to users based on their past ratings.

The model will analyze the user’s ratings on books to identify patterns and preferences and recommend books that the user is likely to enjoy but has not read yet.

Problem Statement



- Jumia a fast-growing online retailer, is currently experiencing low sales on their book section.
- The cause is overwhelming options for customers and people sticking to specific authors.

OBJECTIVE

GOAL

To enable Jumia to effectively provide its customers with book recommendations that are tailored to the users.

SPECIFIC OBJECTIVE

1. Develop a machine learning model that can accurately predict the rating a user will give to books they haven't read based on their previous ratings.
2. To identify factors influencing a user’s book preferences by analyzing the relationship amongst the individual features on the dataset
3. To evaluate the performance of the recommendation system using appropriate metrics and compare it with other models.

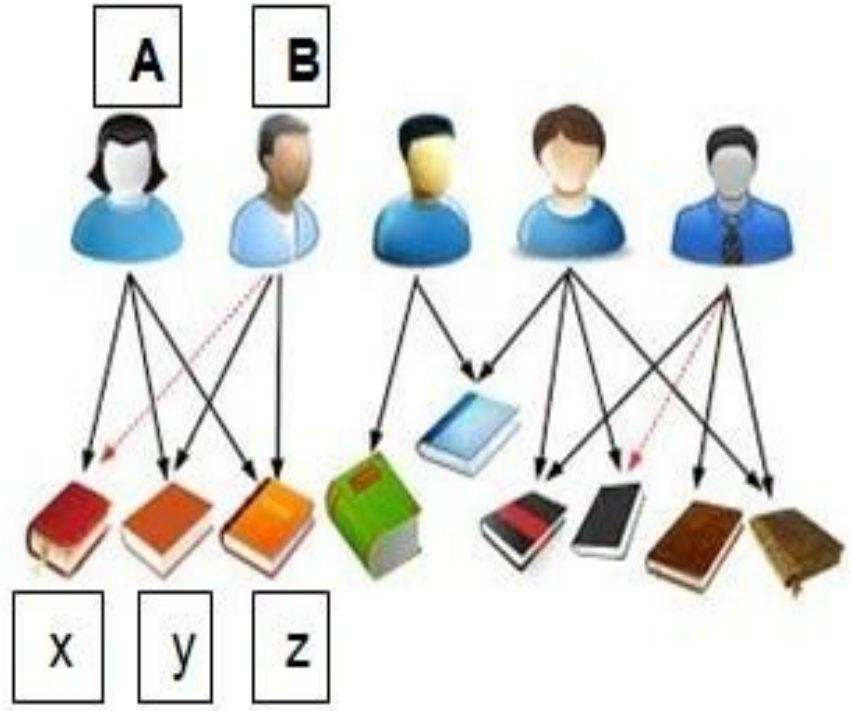
Research Questions

i. How can we accurately match users with books that they will enjoy?

ii. What are the most important factors in determining a user’s book preference?

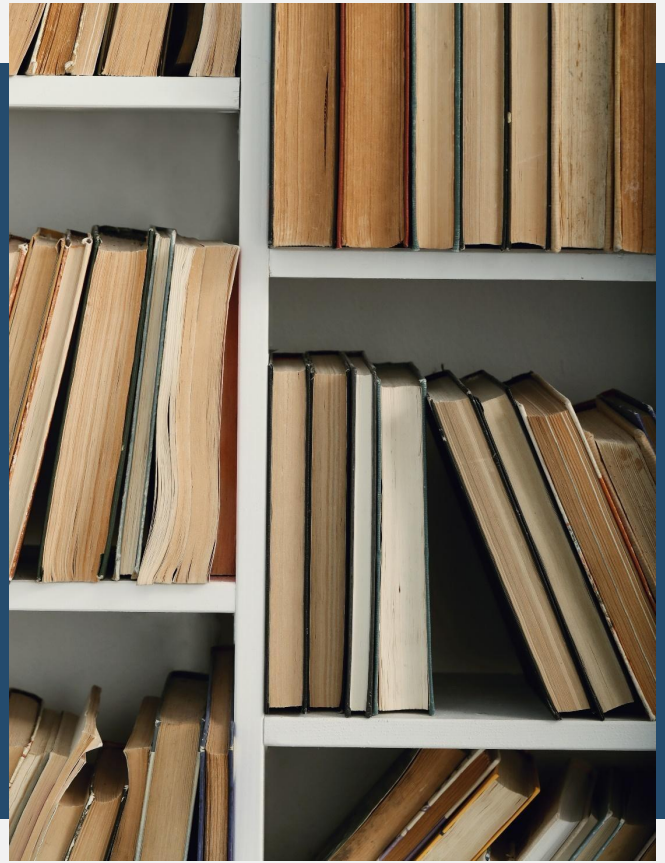
iii. How can we measure the effectiveness of the book recommendation system?

Data

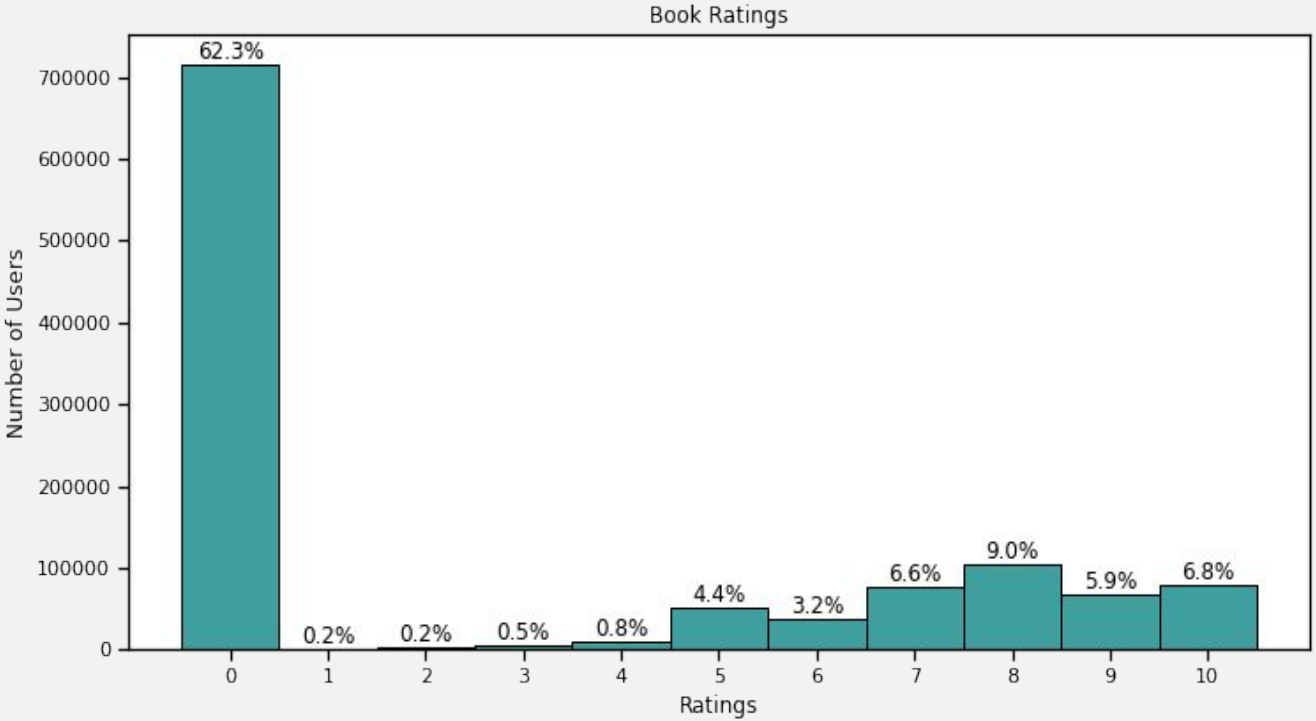


- The Book Recommendation Data used in this project is from Kaggle.
- It comprises of three files; books.csv, rating.csv and users.csv

Exploratory Data Analysis

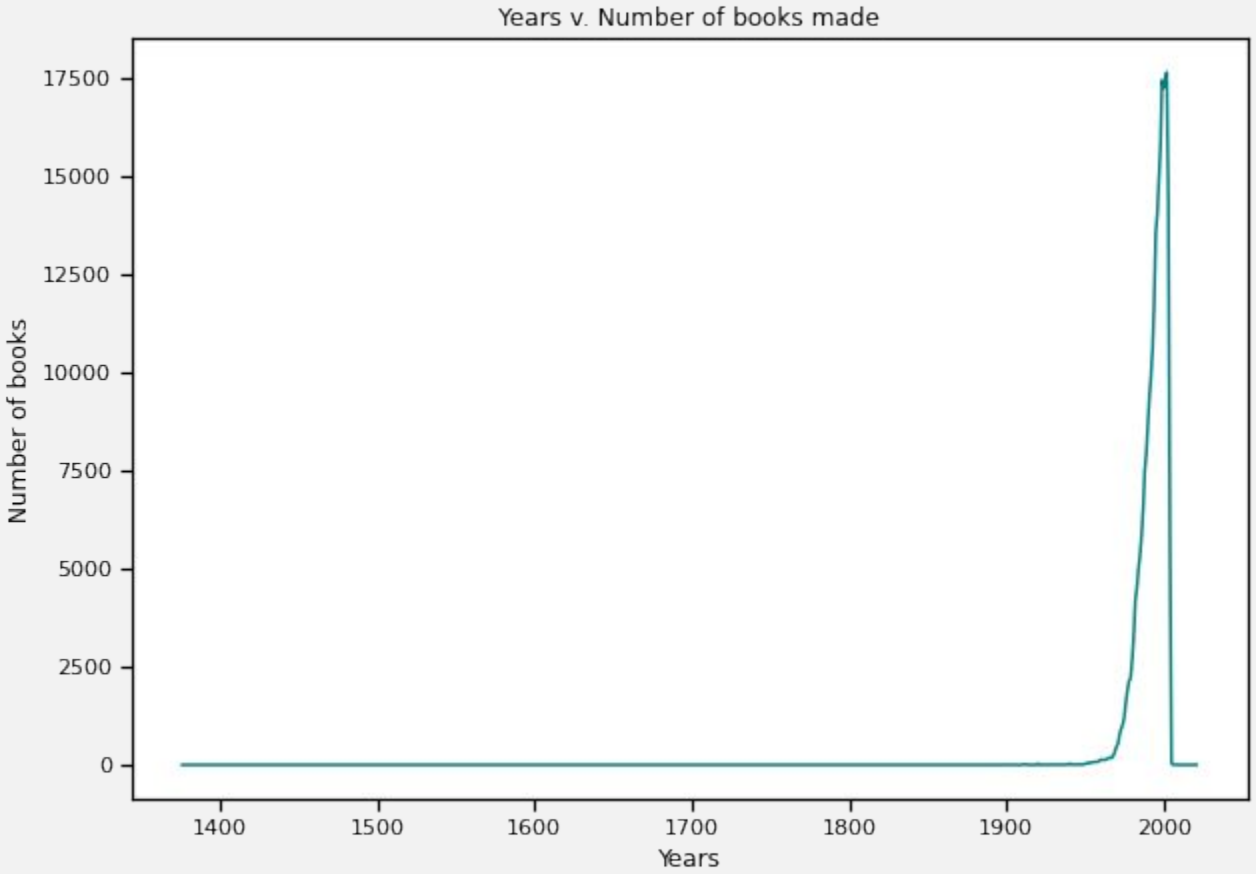


How is Rating Distributed?



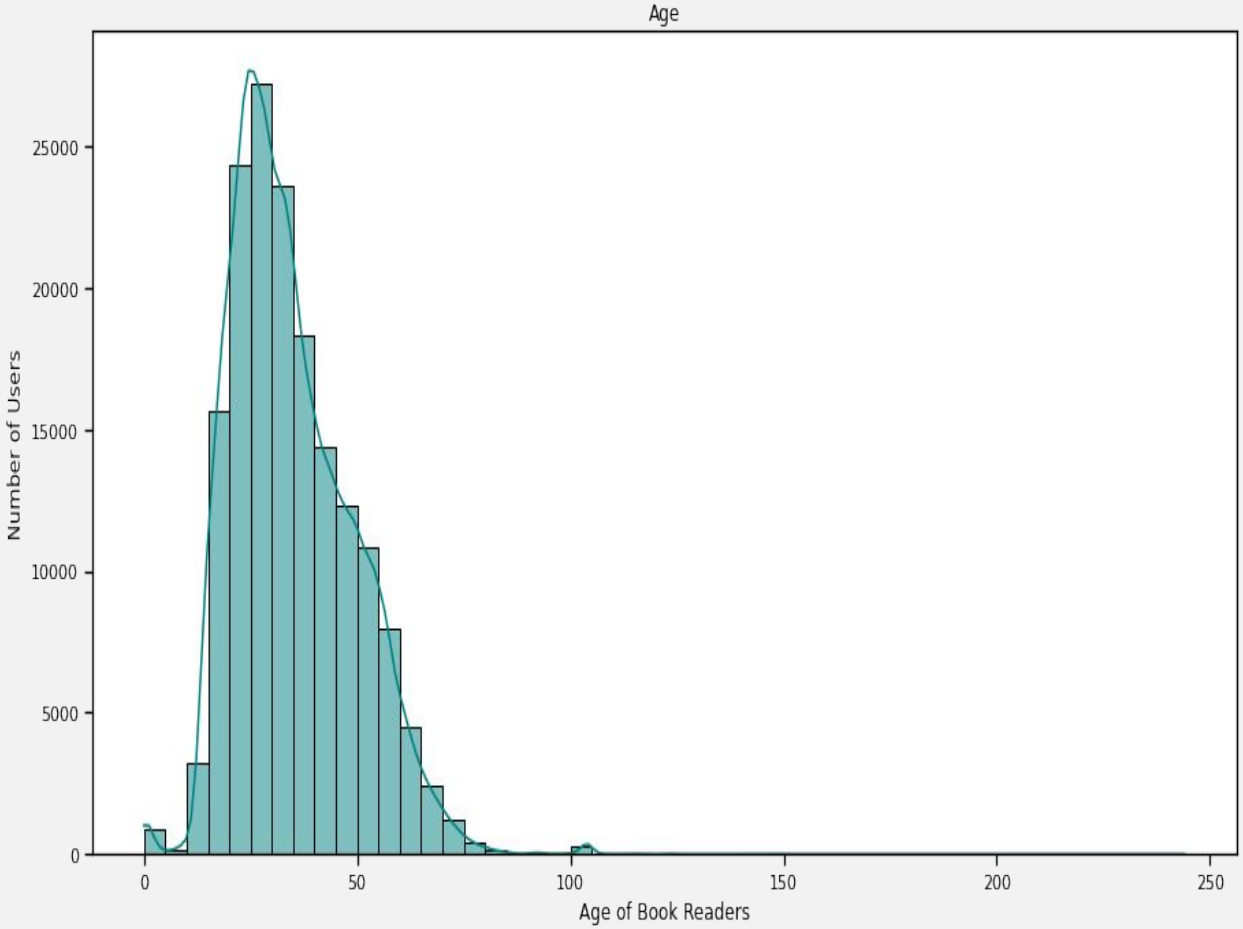
- The book rating ranges from 0 to 10
- Most books have a rating of 0
- The ratings between 1-4 have very few books

What is the trend in book production over time?



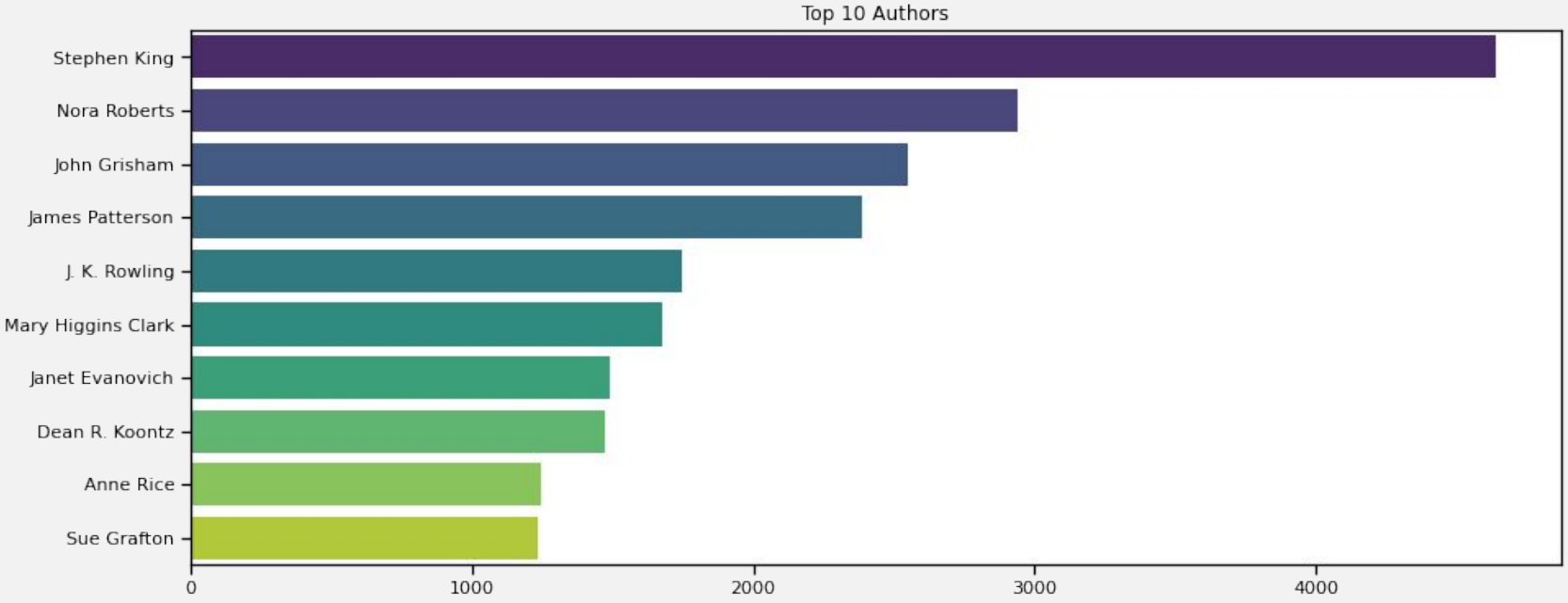
- Most books in the dataset were published from 1950 to 2021

What observations can be made from the age column in Users_df?



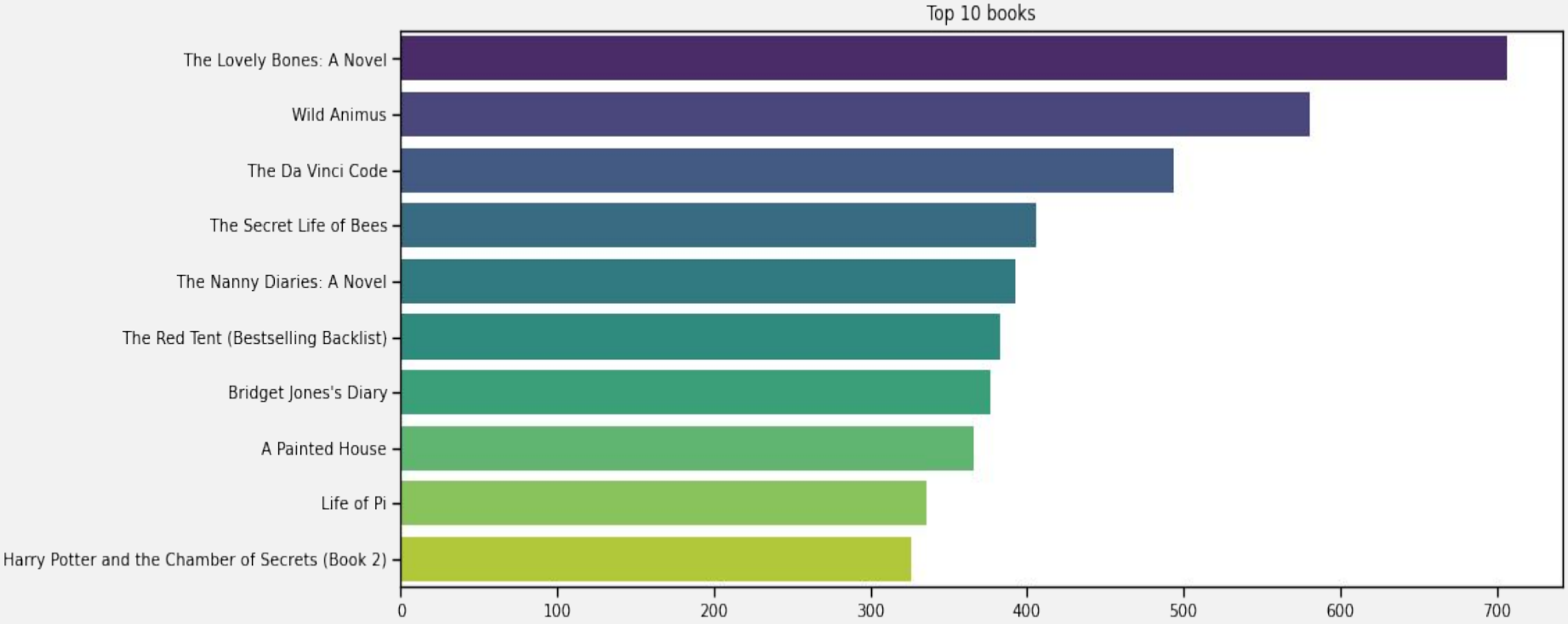
- Most readers are below 50 years of age
- The Age range from 0 To 250.
- We have outliers in the Age column.

Checking for author popularity



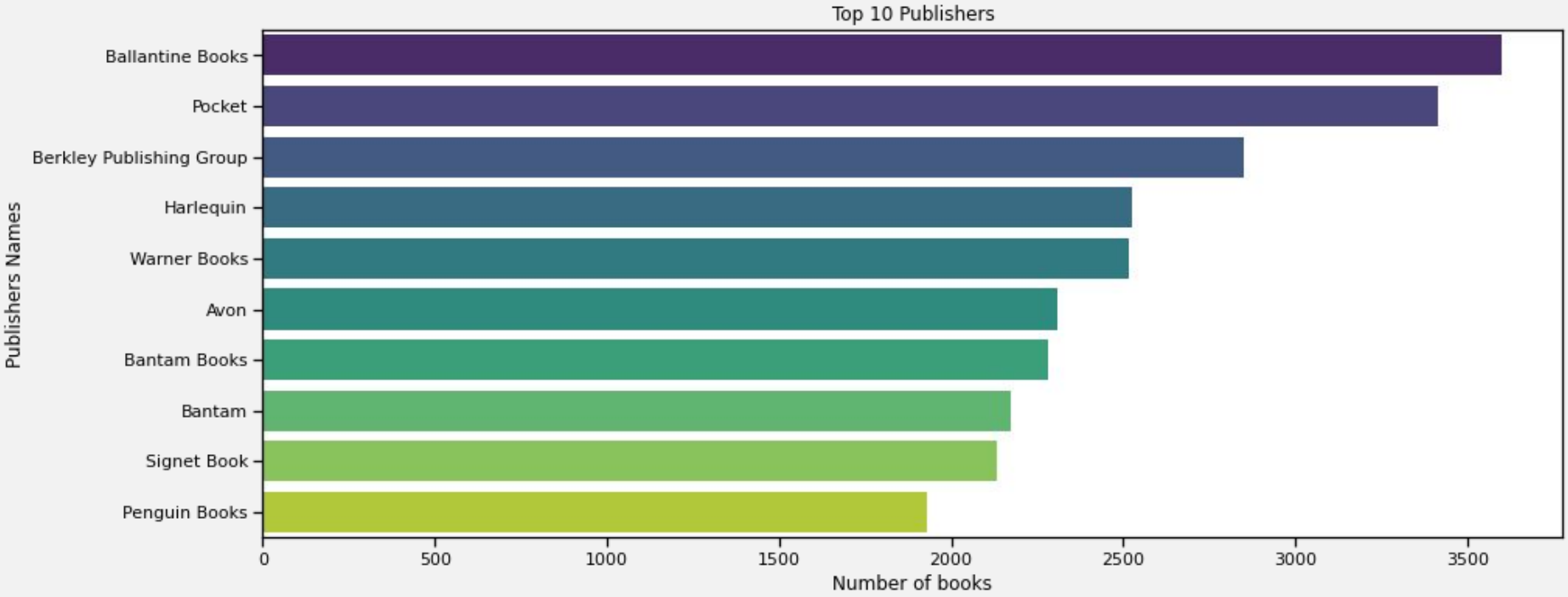
- Stephen King is the most popular author with 5,000 copies sold

Checking for the most popular books



- **The Lovely Bones: A Novel is the most popular book with about 700 sold copies**

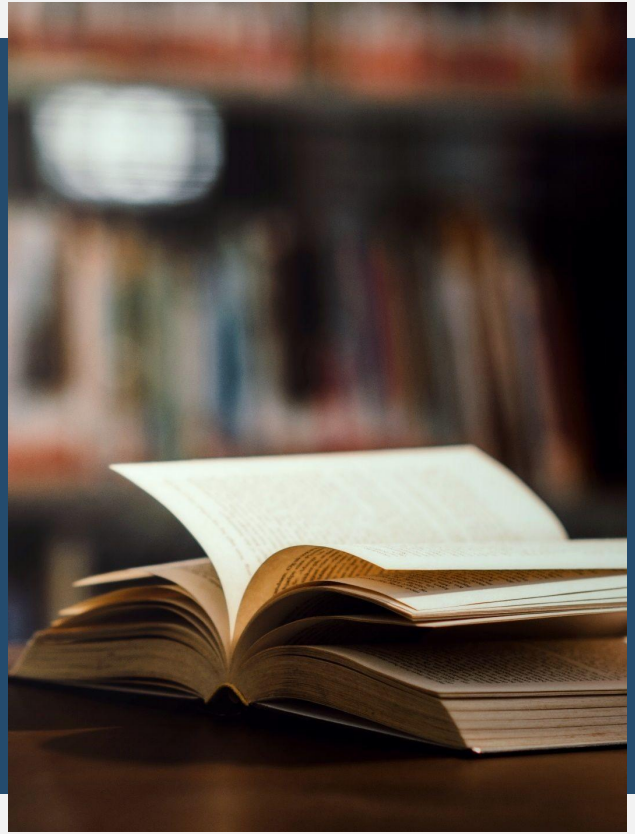
Checking for the most popular publishers in the dataset



- Ballantine Books is the most popular publisher
- They have sold about 3500 copies

Modelling

1. Popularity Based Recommendation System
2. Model Based Recommendation System



Modeling

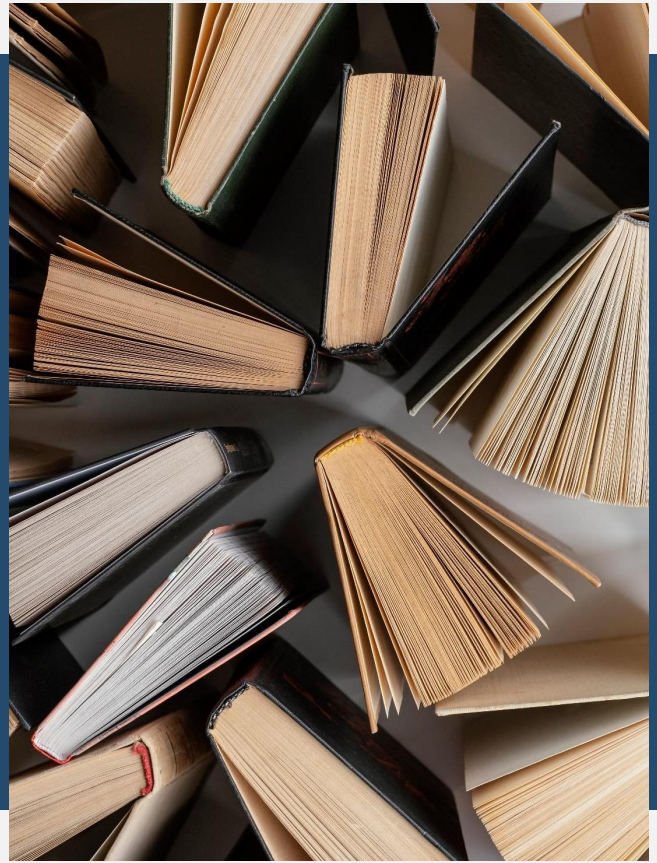
- Recommender systems are important for providing personalized recommendations to users.
- Some recommender systems lack a performance metric, such as popularity-based systems that rely on the popularity of items.
- Popularity-based systems can be used as an initial model when individual user data is unavailable to learn more about the user's preferences and provide more personalized recommendations later on.
- An RMSE of 1.47 for a book recommender system indicates that the system is providing fairly accurate recommendations.
- Removing implicit data can improve the accuracy and fairness of the system's recommendations, and other factors such as user satisfaction and system scalability should also be considered when evaluating a recommender system's performance.

Result and Evaluation

Evaluation

- The project includes multiple recommender system implementations
- The first is a popularity-based recommendation system that recommends the most positively rated books among the largest possible audience
- This is a simple implementation but can work well when user and book data is unavailable
- The second implementation is a model-based collaborative filtering recommendation system
- Matrix factorization with SVD is used to compare the latent features of books and users and produce a matrix for better book recommendations

Conclusion and Recommendation



Conclusions

- Personalized recommendations are more effective than generic ones for a book recommender system
- Collaborative filtering is used for building book recommender systems by analyzing user behavior and preferences
- Data quality is critical for optimal performance of a book recommender system
- The Root Mean Square Metric (RMSE) is used to evaluate and improve the system's performance

Recommendations

- Ensure data quality. Ensure that the data used to train and test the model is accurate, relevant and up-to-date
- Incorporate more types of data, such as book genres and authors, to enhance the recommendation system's effectiveness.
- Regularly evaluate the performance of the book recommender system using appropriate metrics to identify areas for improvement and optimize the system for better performance.

Thank you