**Book Recommender System**

**Abstract:**

This project presents a smart book recommender system using two methods. First, it suggests the top 50 books based on their average ratings, focusing on popular and well-liked choices. Only books with over 250 ratings are considered. Second, it uses a clever technique called item-item collaborative filtering. This method suggests books similar to what a user likes, using their past ratings. We improve this by focusing on users who've rated over 200 books and books with over 50 ratings. By combining these two approaches, our system aims to provide a mix of popular and personalized book recommendations.

**Introduction:**

In the world of online book platforms, helping users find books they'll love is crucial. Our project introduces a smart way to do this – a book recommender system that's both popular and personalized.

The first part recommends top-rated books based on average ratings. We make sure to only consider books with a lot of ratings, at least 250. Then, we sort them by average rating to show the best ones.

The second part gets more personal. It suggests books similar to what a user likes, using a clever method called item-item collaborative filtering. To make it even better, we focus on users who've rated many books (more than 200) and books that are well-rated (more than 50 ratings).

By combining these two methods, our system aims to give users a well-rounded set of recommendations – some popular hits and some tailored just for them. In the next sections, we'll dive into the details of how this works and how we measure its success.

**Dataset**:

The dataset was taken from a very popular cite for publically available datasets, Kaggle. The dataset can be accessed using [Book Recommendation Dataset (kaggle.com)](https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset?resource=download).

The dataset has three CSV files.

* **Books.csv**
* **Users.csv**
* **Ratings.csv**

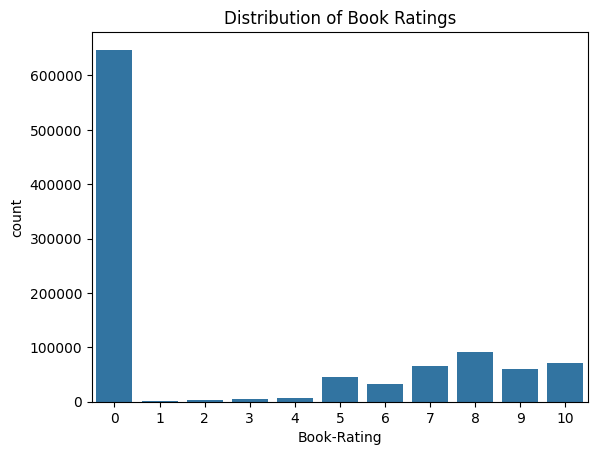
The **Books.csv** has following columns: ***[ISBN, Book-Title, Book-Author, Year-Of-Publication, Publisher, Image-URL-S, Image-URL-M, Image-URL-L]****.* Out of these columns, the one used are ***Book-Title, ISBN*** and ***Book-Author****.* The **Ratings.csv** has following columns: ***[User-ID, ISBN, Book-Rating].*** As it is apparent that the **books.csv** can be merged with **Ratings.csv** based on ***ISBN*** column. Users.csv has columns: ***[User-ID, Location, Age].*** It can be used to get additional user information but is not utilized in the recommender system.

**Data Analysis and Preprocessing:**

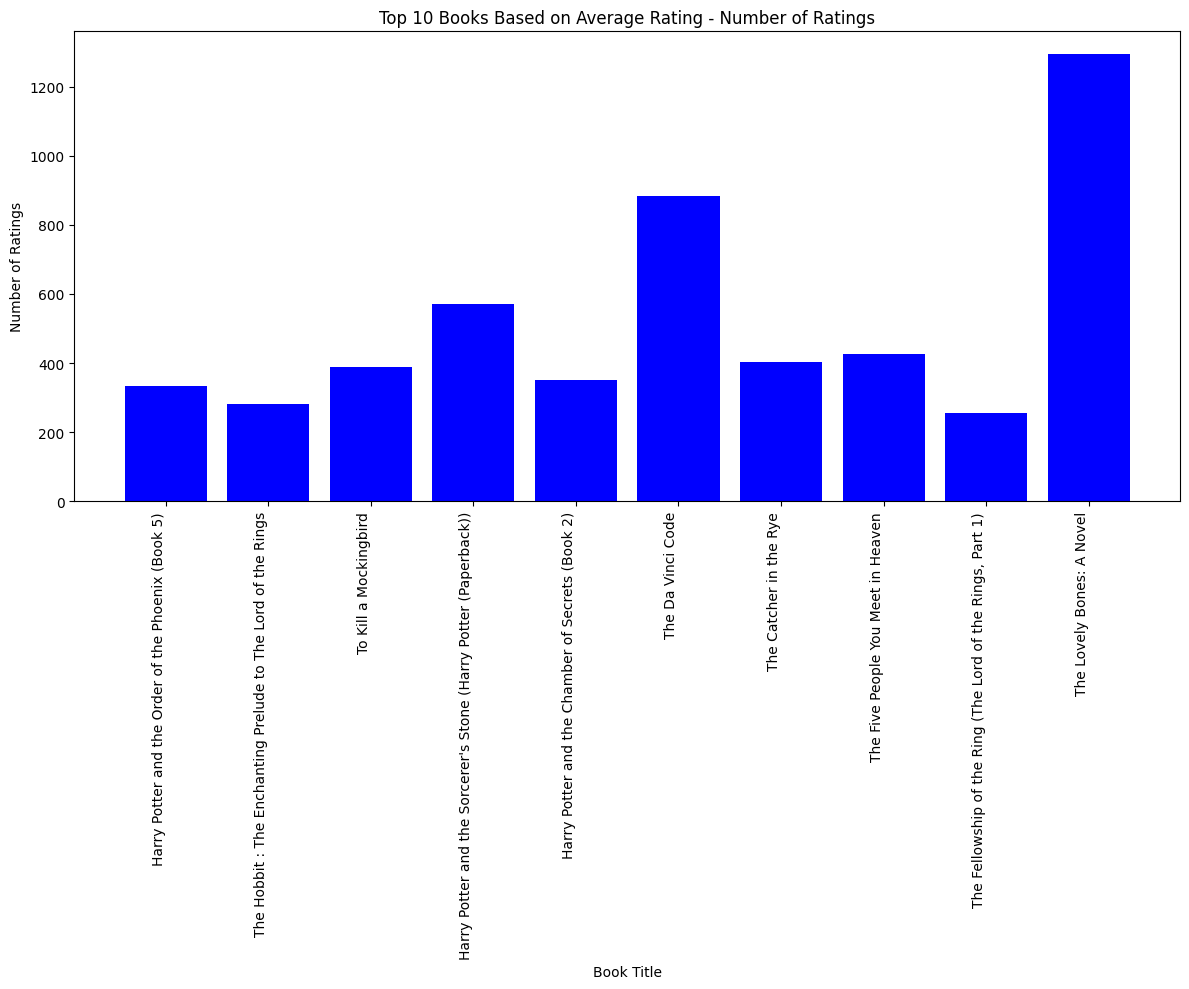
Out of the three CSVs, only two CSV files **Books.csv** and **Ratings.csv** are needed. As the recommender system suggests similar books to the provided book based on user interactions. The user interactions with books can be found in **Ratings.csv** and **Books.csv** has the information related to each book.

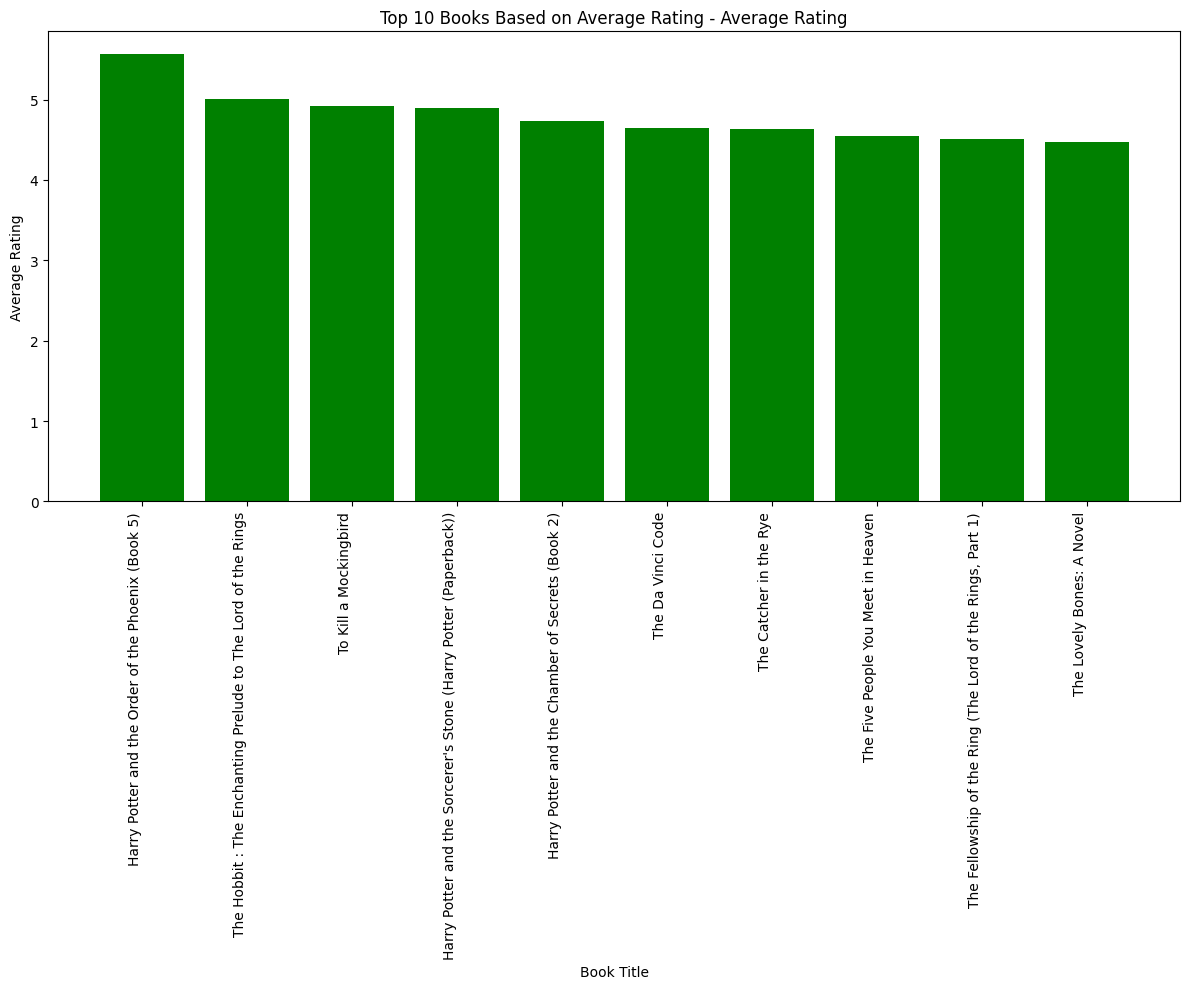
In the dataset, there are duplicate books. This is because a same book is published by different publishers at different time. Thus a same book is present with different **ISBN** numbers. These duplicates are removed by filtering data based on **Book-Title** not **ISBN**.

Initially, there are a total of **271360** books which after removing duplicates and removing a null case for Book-Author column, remains to be **242134** in number. There are a few other null values as well in columns like Publisher and Image-URL-L but they are not used in training thus not handled. There are about **1149780** interactions or ratings available. The total number of users are **278858**.

The rating distribution initially in the dataset is as follows:  


The dataset shows many interactions with 0 ratings, suggesting that users might not like several books. To focus on recommending well-received books, we only consider those with at least 250 ratings. This leaves us with 99 popular books, improving the chances of suggesting enjoyable reads to users. The top 10 popular books with average ratings and number of ratings are displayed below:

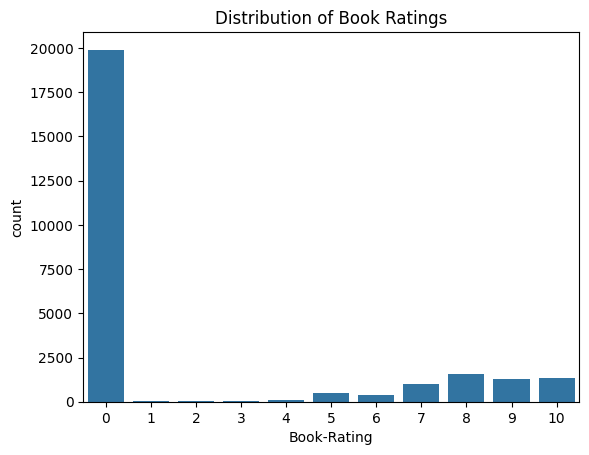




To create a recommender system using item-item collaborative filtering, we start by building an interaction matrix. In this matrix, the rows represent book titles, and the columns represent users. The system relies on finding similarities between items based on user ratings.

To enhance the learning process, we've refined our approach through multiple experiments. Now, our interaction matrix only includes users who have rated more than 200 books and books that have received more than 50 ratings. After applying these filters, we are left with 342 books, 26,162 interactions, and 665 users.

This refined dataset now presents a clearer picture of the rating distribution:



**Approach:**

Two kind of recommender systems are used.

* **Popularity based recommender system:** It suggests top 50 books based on average ratings. Books which have received more than 250 ratings are first filtered, then are sorted based on their average ratings.
* **Item-item collaborative filtering based recommender system:** It is based on the similarity between books calculated using the rating users have given to items. The similarity is found using cosine similarity between books. Top 5 books based on similarity are suggest for the given book. The interaction matrix used only includes users who have rated more than 200 books and books that have received more than 50 ratings.

**Results:**

Top 10 popular books are as follows:

1. **Harry Potter and the Prisoner of Azkaban (Book 3)**Author: J. K. Rowling  
   Number of Ratings: 428  
   Average Rating: 5.85
2. **Harry Potter and the Goblet of Fire (Book 4)**  
   Author: J. K. Rowling  
   Number of Ratings: 387  
   Average Rating: 5.82
3. **Harry Potter and the Sorcerer's Stone (Book 1)**  
   Author: J. K. Rowling  
   Number of Ratings: 278  
   Average Rating: 5.74
4. **Harry Potter and the Order of the Phoenix (Book 5)**Author: J. K. Rowling  
   Number of Ratings: 347  
   Average Rating: 5.50
5. **Harry Potter and the Chamber of Secrets (Book 2)**Author: J. K. Rowling  
   Number of Ratings: 556  
   Average Rating: 5.18
6. **The Hobbit: The Enchanting Prelude to The Lord of the Rings**  
   Author: J.R.R. Tolkien  
   Number of Ratings: 281  
   Average Rating: 5.01
7. **The Fellowship of the Ring (The Lord of the Rings, Part 1)**Author: J.R.R. Tolkien  
   Number of Ratings: 368  
   Average Rating: 4.95
8. **Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))**Author: J. K. Rowling  
   Number of Ratings: 575  
   Average Rating: 4.90
9. **The Two Towers (The Lord of the Rings, Part 2)**Author: J.R.R. Tolkien  
   Number of Ratings: 260  
   Average Rating: 4.88
10. **To Kill a Mockingbird**Author: Harper Lee  
    Number of Ratings: 510  
    Average Rating: 4.70

Top 5 item-item Collaborative filtering based recommendations for **Book = “1984”**

1. **Book Title: Animal Farm**  
   Author: George Orwell  
   Number of ratings received: 233  
   Average rating: 4.274678111587983
2. **Book Title: The Handmaid's Tale**Author: Margaret Atwood  
   Number of ratings received: 311  
   Average rating: 3.3987138263665595
3. **Book Title: Brave New World**Author: Aldous Huxley  
   Number of ratings received: 226  
   Average rating: 4.331858407079646
4. **Book Title: The Vampire Lestat (Vampire Chronicles, Book II)**Author: ANNE RICE  
   Number of ratings received: 301  
   Average rating: 3.777408637873754
5. **Book Title: The Hours : A Novel**Author: Michael Cunningham  
   Number of ratings received: 216  
   Average rating: 3.44

Top 5 item-item Collaborative filtering based recommendations for **Book = “To Kill a Mockingbird”**

1. **Book Title: The Catcher in the Rye**  
   Author: J.D. Salinger  
   Number of ratings received: 449  
   Average rating: 4.5456570155902005
2. **Book Title: Five Quarters of the Orange**  
   Author: Joanne Harris  
   Number of ratings received: 207  
   Average rating: 3.57487922705314
3. **Book Title: Drowning Ruth**  
   Author: Christina Schwarz  
   Number of ratings received: 170  
   Average rating: 2.8117647058823527
4. **Book Title: The Bean Trees**Author: Barbara Kingsolver  
   Number of ratings received: 389  
   Average rating: 4.087403598971722
5. **Book Title: The Color of Water: A Black Man's Tribute to His White Mother**Author: James McBride  
   Number of ratings received: 223  
   Average rating: 4.1076233183856505