Book Recommendation System

The Book-Crossing dataset comprises three files:

- Users

Contains the users. Note that user IDs (User-ID) have been anonymized and map to integers. Demographic data is provided (Location, Age) if available. Otherwise, these fields contain NULL values.

- Books

Books are identified by their respective ISBNs. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (Book-Title, Book-Author, Year-Of-Publication, Publisher), obtained from Amazon Web Services. Note that in the case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (Image-URL-S, Image-URL-M, Image-URL-L), i.e., small, medium, and large.

- Ratings

Contains the book rating information. Ratings (Book-Rating) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.

Steps followed:

1. Data Preparation (Data cleaning and Feature Engineering)
2. Exploratory Data Analysis
3. Got top ten books as per rating
4. Collaborative filtering
    a. Memory-based approach- Cosine similarity
    b. Model-based approach- Singular value decomposition (SVD)
5. Model evaluation metrics
    RMSE score
6. Getting top recommendations of books and ratings

Conclusion:

- Most of the users are between 30 and 40 and prefer more books and we can also view the same between 20 and 30.
- As per ratings "Selected Poems" has been rated most followed by "Little Women". The counterplot shows users have rated 0 the most, which means they haven't rated books at all.
- The top 10 book recommendations as per ratings with top "The Lovely Bones: A Novel" with 707 book ratings. But this is not based on some recommendation system. They are top 10 books as per ratings.
- As we perform by cosine similarity in the recommendation system it gives a 7.94 RMSE score and SVD improved score it to 1.63 RSME score by Singular Value Decomposition model (SVD).
- The model-based approach was best to signify and at last, we got the top 10 recommended books and ratings respectively.