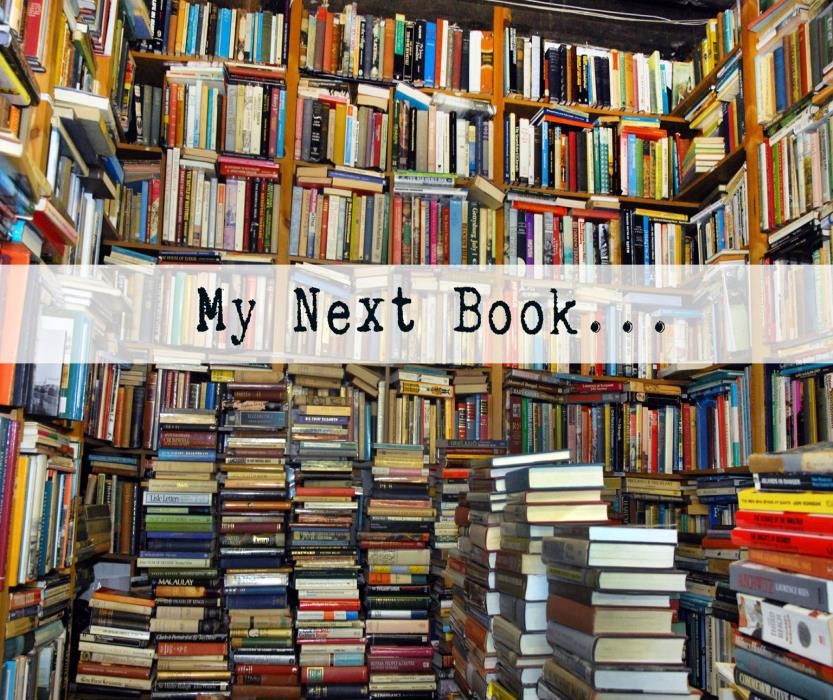
**BOOK RECOMMENDATION SYSTEM**

**Team Members**

**Aditi Jaiswal(219310302)**

**Ankit Kumar(219310325)**

* **Problem statement**
* **Data Summary**
* **Analysis of different datasets**
* **Data Cleaning**
* **Outlier treatment**
* **Imputing missing values**
* **Different Recommendation Model**
* **Challenges**
* **Conclusion**
* **Future Scope**

During the last few decades, with the rise of Youtube, Amazon, Netflix, and many other such web services, recommender systems have become much more important in our lives in terms of providing highly personalized and relevant content.

**The main objective is to create a recommendation system to recommend relevant books to users based on popularity and user interests.**

# Data Summary

The dataset is comprised of three csv files:: User\_df, Books\_df, Ratings\_df

Users\_dataset.

* + User-ID (unique for each user)
  + Location (contains city, state and country separated by commas)
  + Age Shape of Dataset - (278858, 3)

Books\_dataset.

* + ISBN (unique for each book)
  + Book-Title
  + Book-Author
  + Year-Of-Publication
  + Publisher

Ratings\_dataset.

* + User-ID
  + ISBN

● Image-URL-S

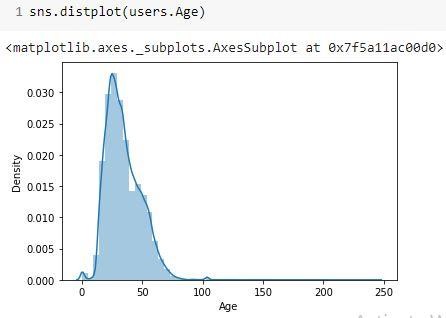
● Image-URL-M

● Image-URL-L

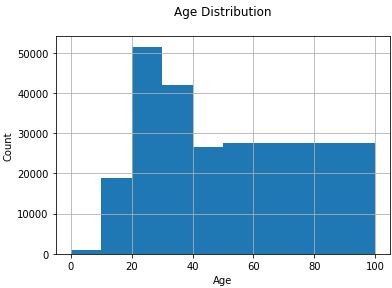
● Shape of Dataset - (271360, 8)

● Book-Rating

● Shape of Dataset - (1149780, 3)

● The Age range given here is from 0 To 250.

● Outliers in the Age column.

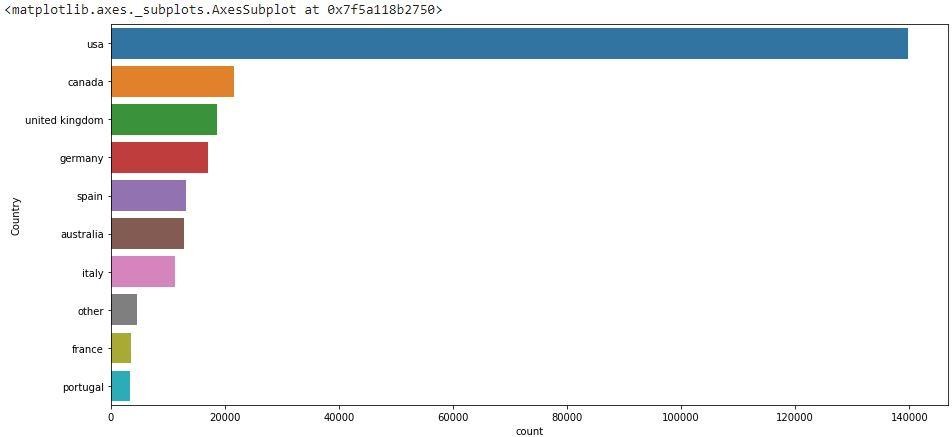


● The Age range distribution is right skewed

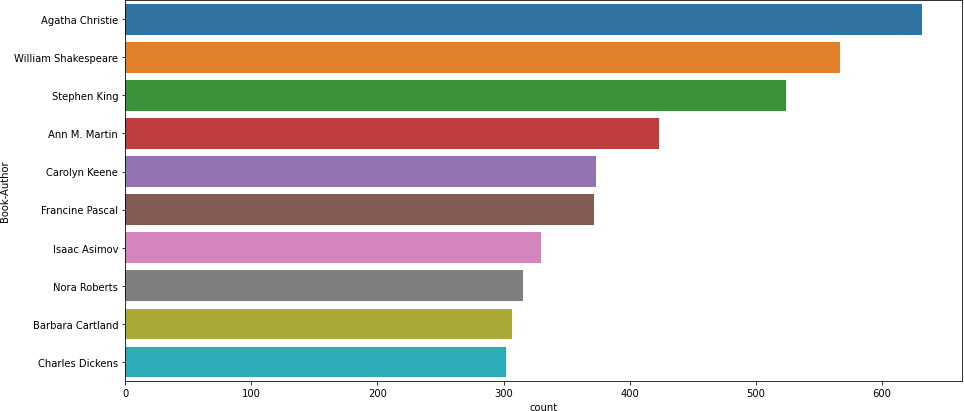
● Most active readers lie in age group 20- 40

● Splitting Location column and analysing country.

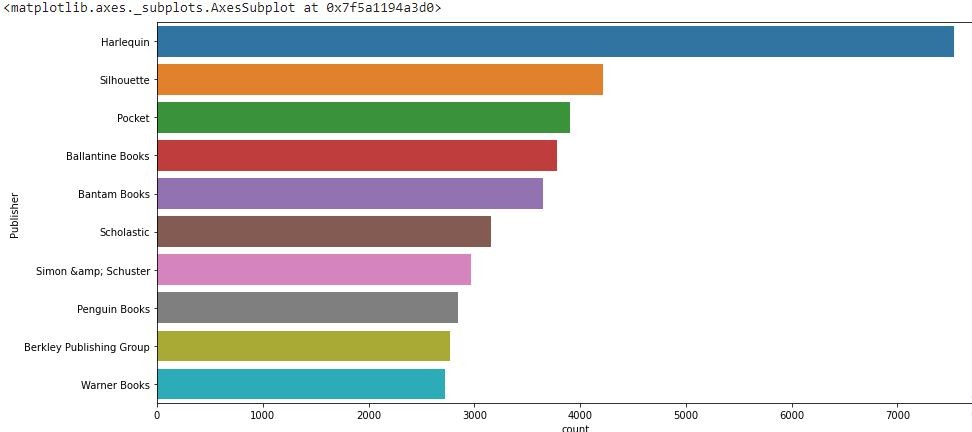
● Most active readers are from USA.



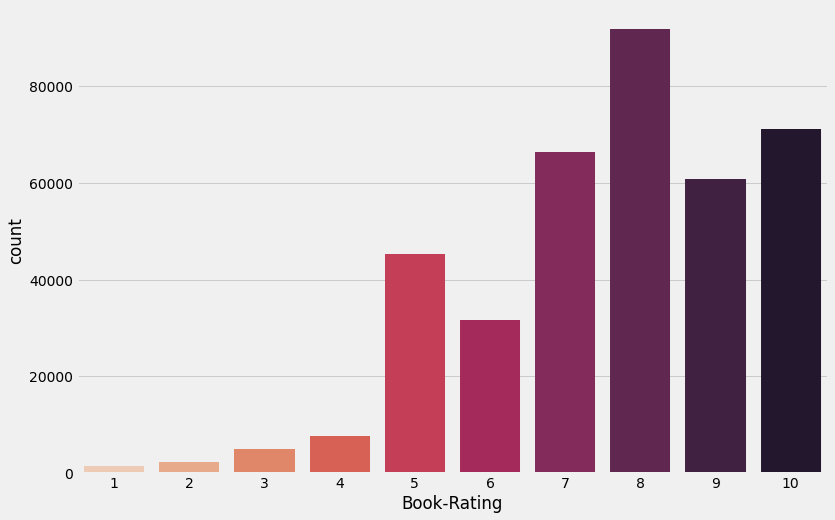
Agatha Christie wrote highest number of books in our given dataset



Harlequin published highest number of books in our given dataset

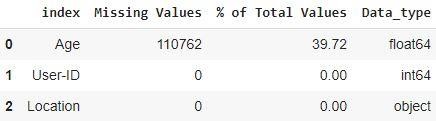


* Higher ratings are more common amongst users
* Rating 8 has been rated the highest number of times

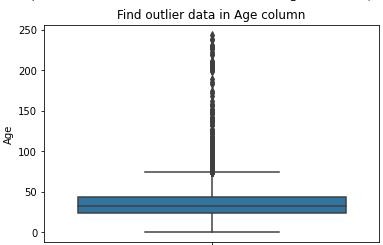


## 1. Null Value Imputation:

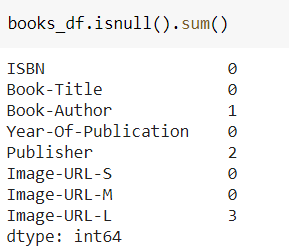
**Age column has 40% missing values**

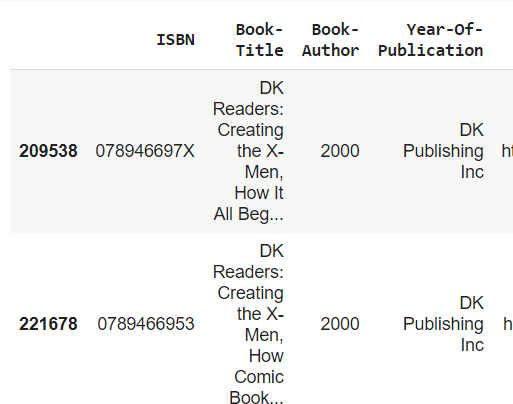


* + **Outliers in Age column**
  + **Age has positive Skewness (right tail) so we can use median to fill Nan values,**



## 1. Null Value Imputation:





### 1.)Popularity Based Recommendation

Book weighted average formula:

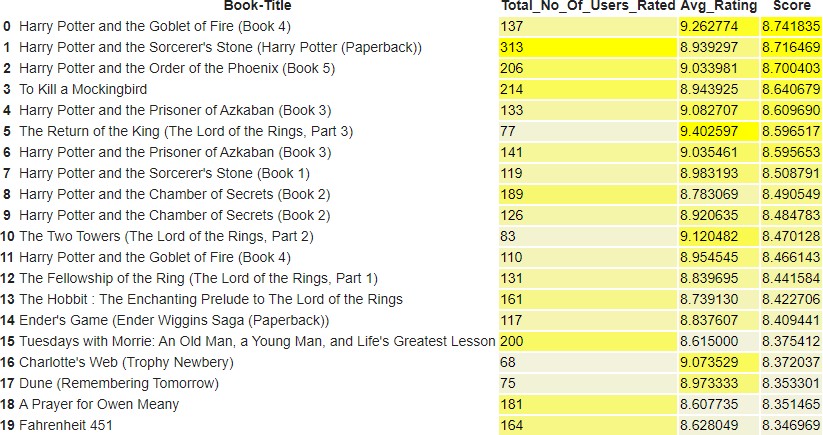
**Weighted Rating(WR)=[vR/(v+m)]+[mC/(v+m)]**

Where,

v is the number of votes for the books;

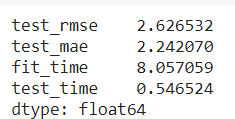
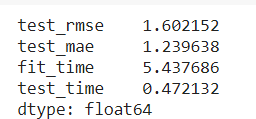
m is the minimum votes required to be listed in the chart; R is the average rating of the book; and

C is the mean vote across the whole report.

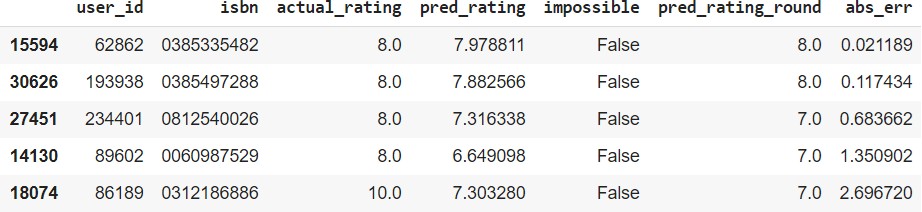


### 2.)Model based collaborative filtering

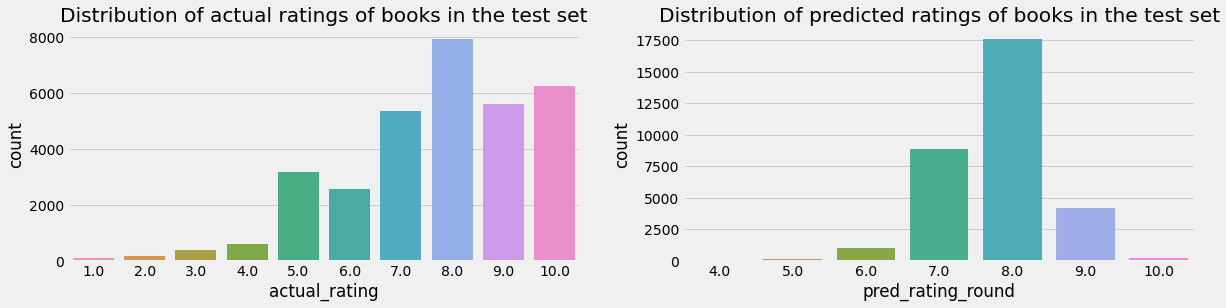
**SVD NMF**



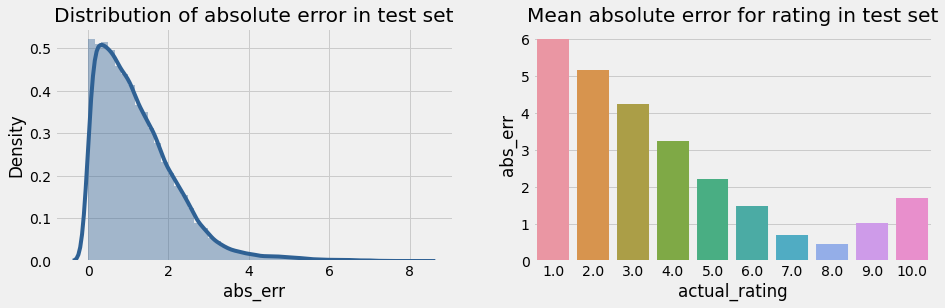
### SVD Model Results



**SVD Model Results**



### SVD Model Results



**User-ID - 193458**

### Test set: predicted top rated books

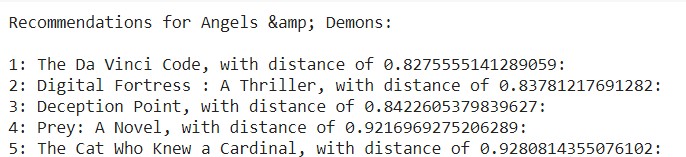


**Test set: actual top rated books**



### 3.)Collaborative Filtering-(Item-Item based)

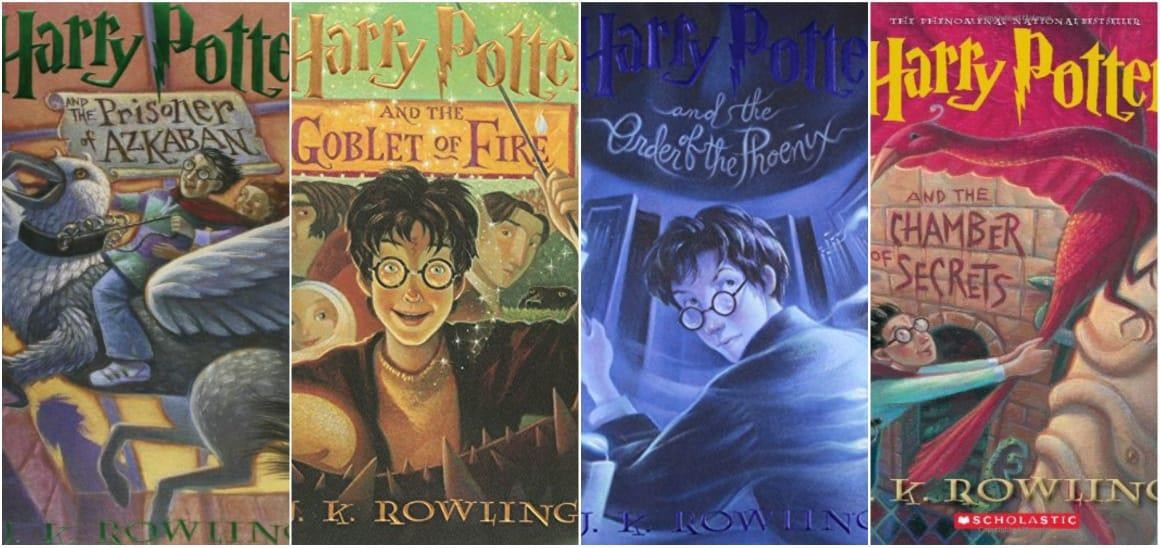
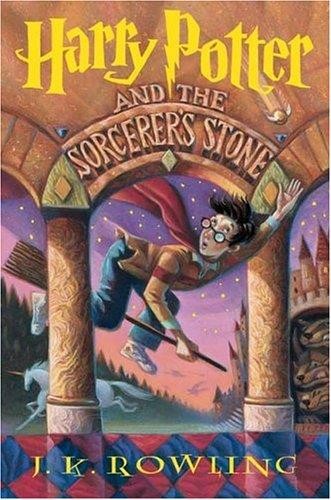
* Cosine Similarity
* Nearest Neighbour



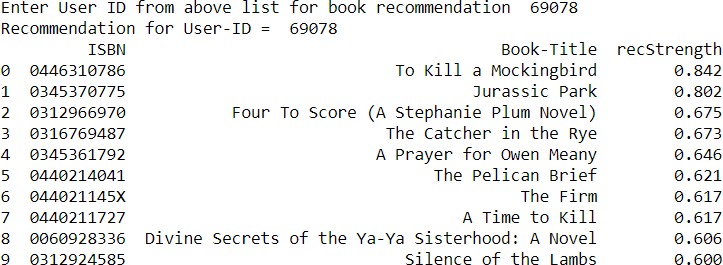
### SVD and Correlation

Recommendations for Harry Potter and the Sorcerer's Stone (Book 1)

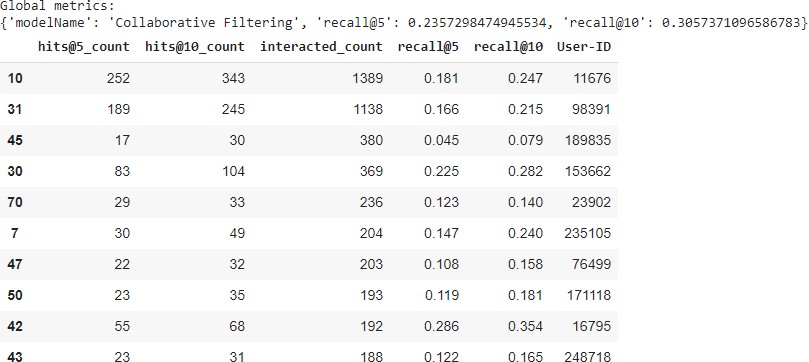
Input Output



### 4.)Collaborative Filtering-(User-Item based)



**Model Results**



# Conclusion

* **In EDA, the Top-10 most rated books were essentially novels. Books like The Lovely Bone and The Secret Life of Bees were very well perceived.**
* **Majority of the readers were of the age bracket 20-35 and most of them came from North American and European countries namely USA, Canada, UK, Germany and Spain.**
* **If we look at the ratings distribution, most of the books have high ratings with maximum books being rated 8. Ratings below 5 are few in number.**
* **Author with the most books was Agatha Christie, William Shakespeare and Stephen King.**
* **For modelling, it was observed that for model based collaborative filtering SVD technique worked way better than NMF with lower Mean Absolute Error (MAE) .**
* A recommendation system helps an organization to create loyal customers. The recommendation system today are very powerful that they can handle the new customer too who has visited the site for the first time. They recommend the products which are currently trending or highly rated and they can also recommend the products which bring maximum profitto the company.

A book recommendation system is a type of recommendation system where we have to recommend similar type of books to the reader based on his interest. The books recommendation system is used by online websites which provide ebooks like google playbooks, open library, good Read’s, etc.

* + **Handling of sparsity was a major challenge as well since the user interactions were not present for the majority of the books.**
  + **Understanding the metric for evaluation was a challenge as well.**
  + **Since the data consisted of text data, data cleaning was a major challenge in features like Location etc..**
  + **Decision making on missing value imputations and outlier treatment was quite**

**challenging as well.**

* + **Given more information regarding the books dataset, namely features like Genre, Description etc, we could implement a content-filtering based recommendation system and compare the results with the existing collaborative-filtering based system.**
  + **We would like to explore various clustering approaches for clustering the users based on Age, Location etc., and then implement voting algorithms to recommend items to the user depending on the cluster into which it belongs.**



**Thank You**