**Vyakya Task 1**

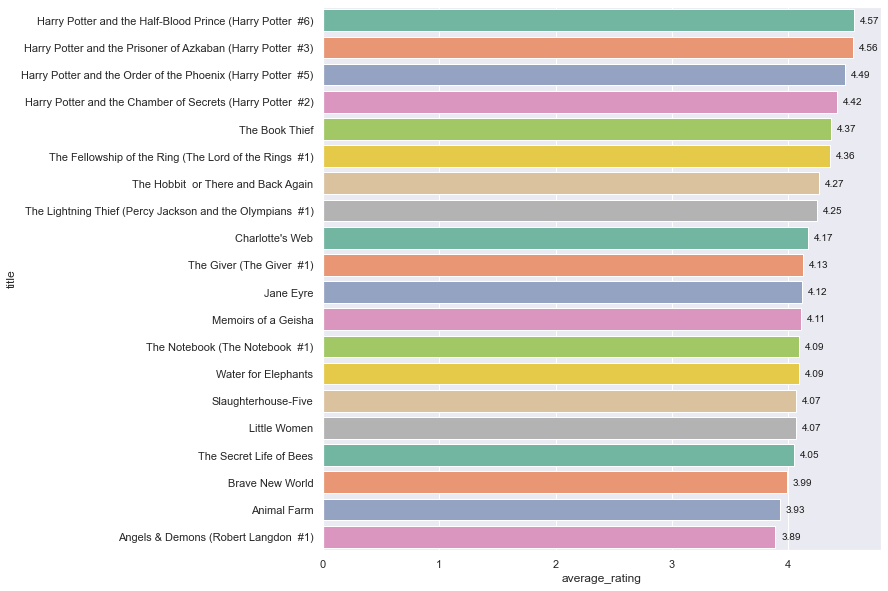
**Book recommendation System**

**Introduction:**

The basic idea behind analyzing the given dataset is to get a fair idea about the relationships between the multiple attributes a book might have, such as the aggregate rating of each book, the trend of the authors over the years and books with numerous languages.

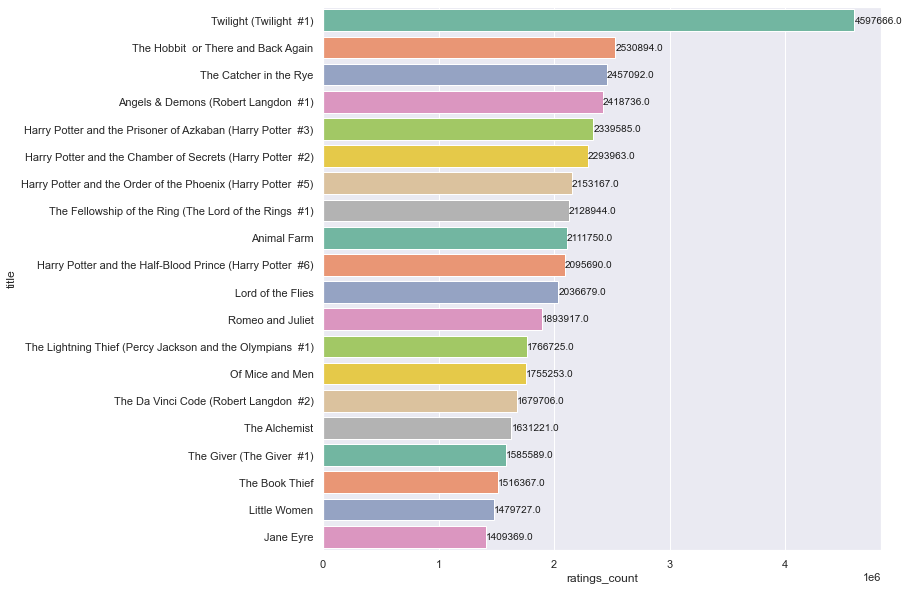
**Exploratory Data Analysis (EDA):**

**Top 20 rated books:**

****

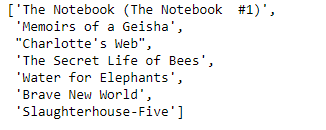
Here from the above graph we can infer that the books from harry potter series have been rated the highest throughout the database.

**Top 20 Top Voted Books**



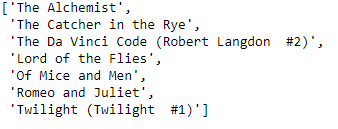
Twilight #1 is the most top voted book with almost double the number of top votes by its successor The hobbit of There and Back Again. We can infer that the harry potter series books dominate their presence here too.

**Top rated but not in top voted**



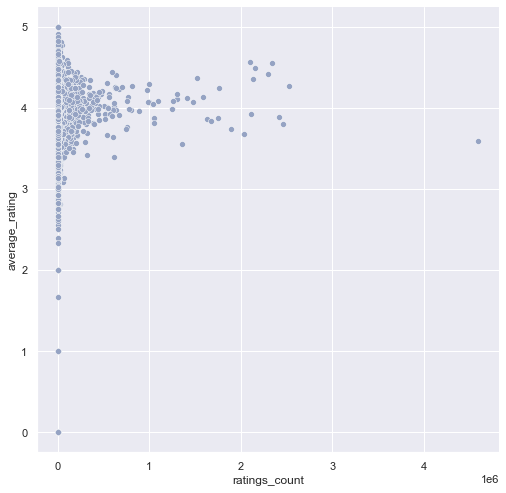
These books are top rated by the readers but not top voted.

**Top voted but not in top rated**

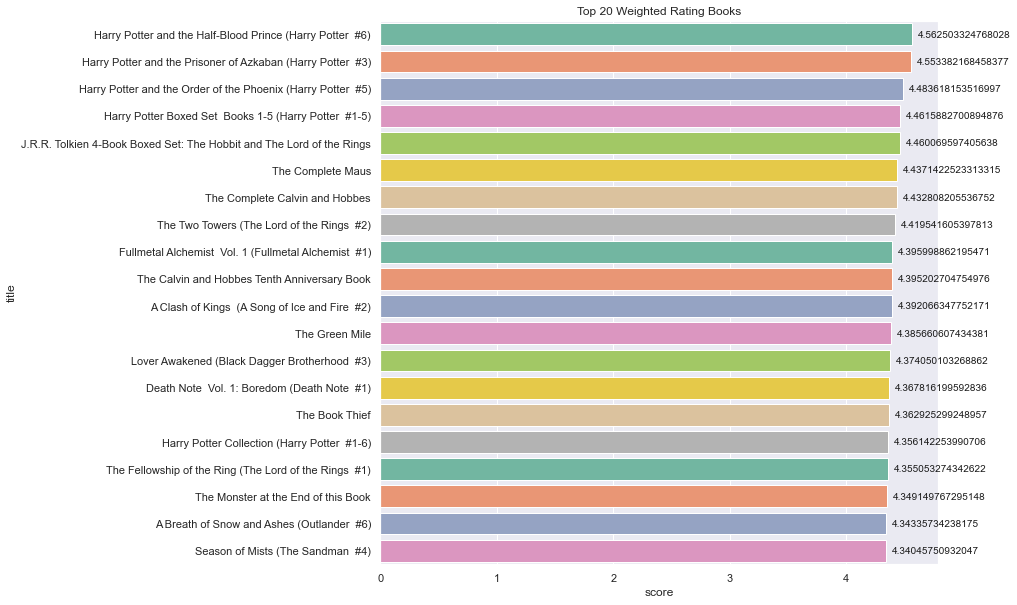


These books are top voted by the readers but not top rated.

**Relationship between rating and vote**

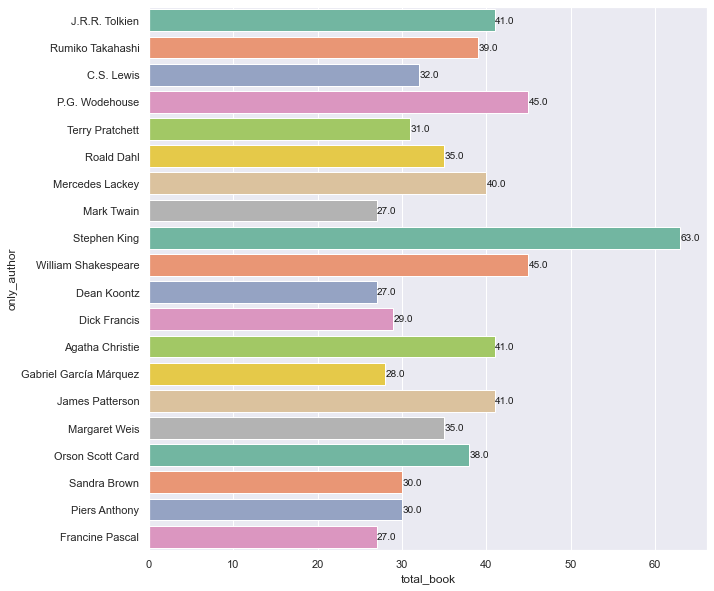


**Top 20 weighted ratings:**



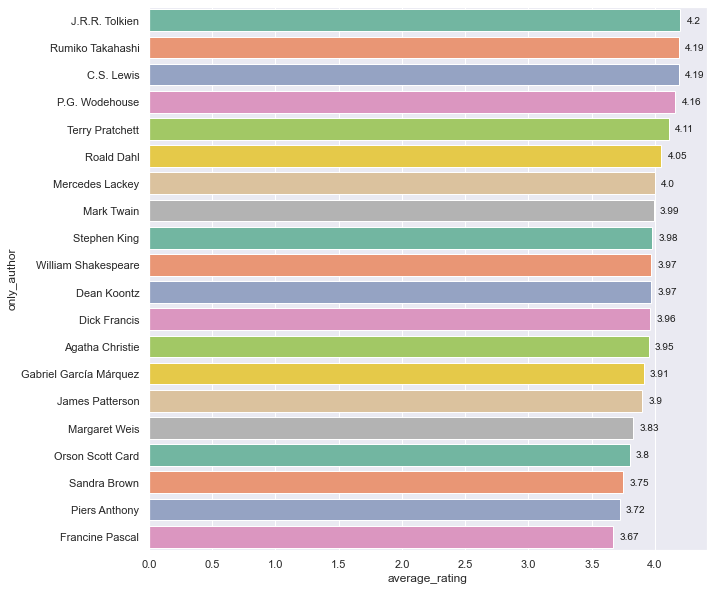
Here from the above graph we can infer that the books from harry potter series have been rated the highest throughout the database.

**Most books of an author**



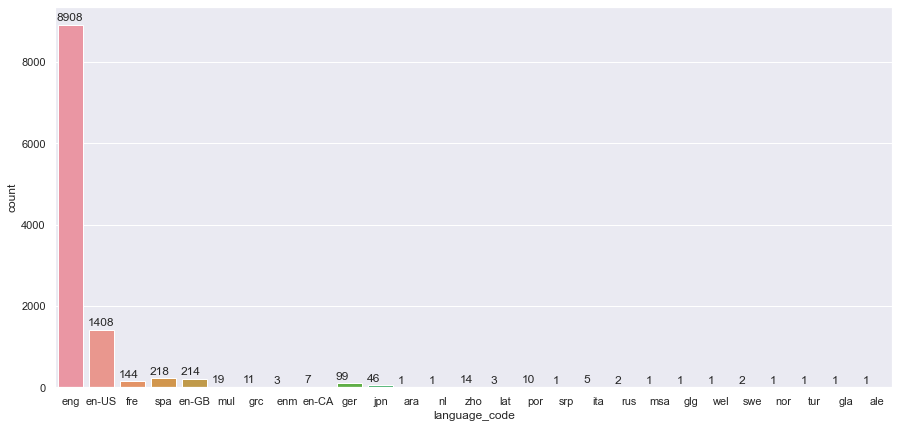
Here from the above graph we can infer that throughout the history Stephen King has wrote 63 titles followed by William Shakespeare with 45 titles.

**Average rating of an author**

****

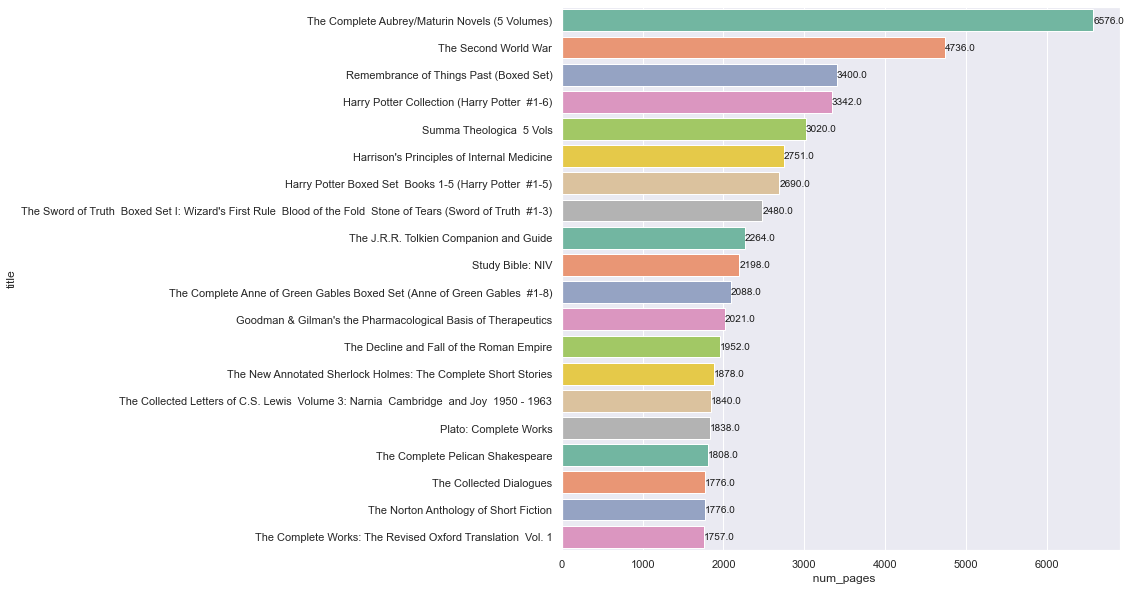
J.R.R.Tolkein is the most rated author with rating 4.2 with close successors Rumiko Takashi and C.S.Lewis with rating of 4.19.

**Language Distribution**

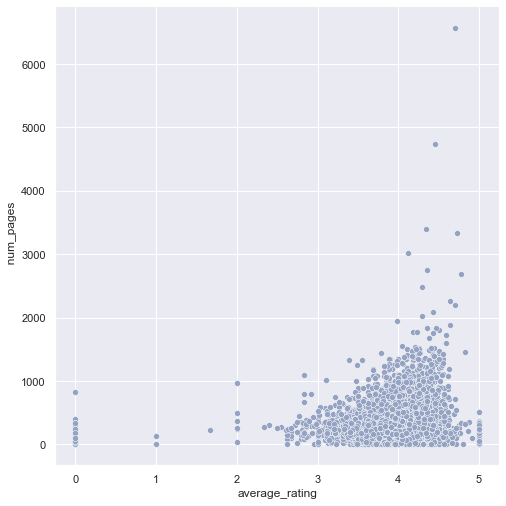
****

English is the most used language in books, There are different versions of English too, like eng-UK, eng-US and eng-CAN. Followed by Spanish

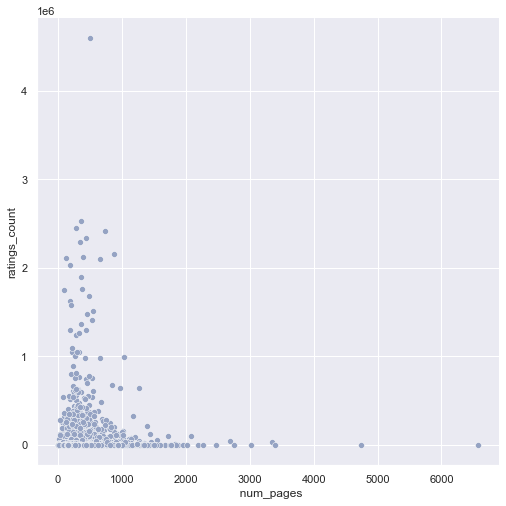
**Top 20 Top Voted Books**

****

**Relationship between rating and pages**

****

**Relationship between pages and ratings\_count**

****

**Recommendation Engine:**

Having seen the clustering, we can infer that there can be some recommendations which can happen with the relation between Average Rating and Ratings Count.

Taking the Ratings\_Distribution (A self created classifying trend), the recommendation system works with the algortihm of K Nearest Neighbors.Based on a book entered by the user, the nearest neighbours to it would be classified as the books which the user might like.KNN is used for both classification and regression problems. In classification problems to predict the label of a instance we first find k closest instances to the given one based on the distance metric and based on the majority voting scheme or weighted majority voting(neighbors which are closer are weighted higher) we predict the labels.

