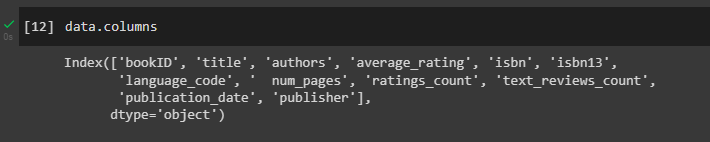
**EDA Report - Niyas Hameed R F**

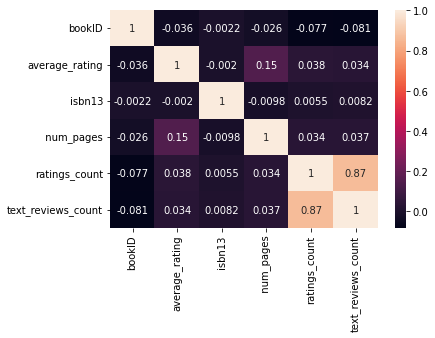
With the given books.csv dataset, I was able to identify a small number of patterns and identified certain measures that might be helpful for a recommendation system.

Starting with the given dataset features,

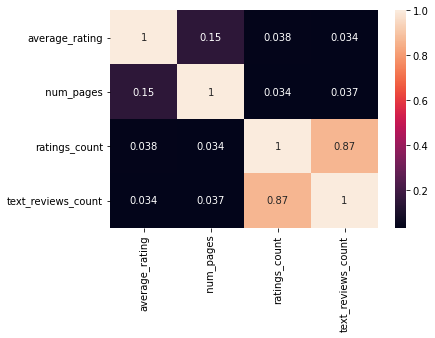


The columns given here were more or less, the summary of the books. Yet I observed that this is missing a key feature, which might be Genre. Books can generally be classified with more info given about the genre, which would be more appropriate for recommendation systems.

I mapped the correlation between the columns in the sample dataset.



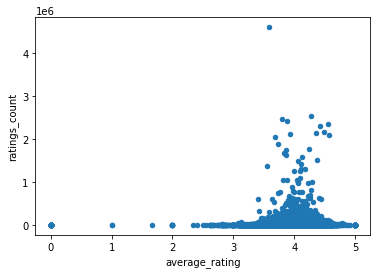
As a result, I removed all the unnecessary columns which do not provide an effect as much as the other.



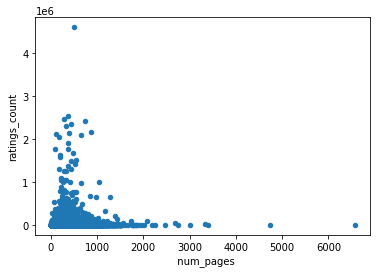
The new alteration contains the

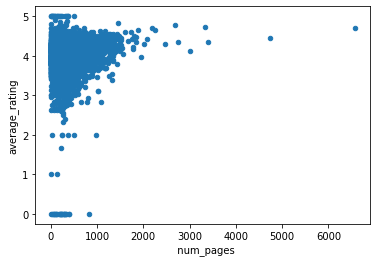
* Average rating
* Number of Pages
* Ratings count
* Text reviews count (more or less the same as ratings\_count)

Now, I had to map and visualize the data.



This shows us that, the average rating is between 3-5 mark, but it is mostly 4. Users and critics tend to give “Not the best, but very good” ratings often.





Observing the number of ratings given for books with number of pages, we observe that the mark is very similar to the number of books with pages between 0 - 1000, but a mean of 336.6 pages for every book was observed.

Other than this, I plot various scatter plots according to the attributes, and there’s not much useful or decision making information given from these plots.

