

Goodreads Books (Recommender Systems)

Description

Context

The primary reason for creating this dataset is the requirement of a good clean dataset of books. Being a bookie myself (see what I did there?) I had searched for datasets on books in kaggle itself - and I found out that while most of the datasets had a good amount of books listed, there were either a) major columns missing or b) grossly unclean data. I mean, you can't determine how good a book is just from a few text reviews, come on! What I needed were numbers, solid integers and floats that say how many people liked the book or hated it, how much did they like it, and stuff like that. Even the [good dataset](#) that I found was well-cleaned, it had a number of interlinked files, which increased the hassle. This prompted me to use the Goodreads API to get a well-cleaned dataset, with the promising features only (minus the redundant ones), and the result is the dataset you're at now.

Acknowledgements

This data was entirely scraped via the [Goodreads API](#), so kudos to them for providing such a simple interface to scrape their database.

Inspiration

The reason behind creating this dataset is pretty straightforward, I'm listing the books for all book-lovers out there, irrespective of the language and publication and all of that. So go ahead and use it to your liking, find out what book you should be reading next (there are very few free content recommendation systems that suggest books last I checked), what are the details of every book you have read, create a word cloud from the books you want to read - all possible approaches to exploring this dataset are welcome.