

SQL

The coronavirus took the entire world by surprise, changing everyone's daily routine. City dwellers no longer spent their free time outside, going to cafes and malls; more people were at home, reading books. That attracted the attention of startups that rushed to develop new apps for book lovers.

Description of the data

books

Contains data on books:

- book_id

You've been given a database of one of the services competing in this market. It contains data on books, publishers, authors, and customer ratings and reviews of books. This information will be used to generate a value proposition for a new product.

- author_id
- title
- num_pages — number of pages
- publication_date
- publisher_id

authors

Contains data on authors:

- author_id
- author

publishers

Contains data on publishers:

- publisher_id
- publisher

ratings

Contains data on user ratings:

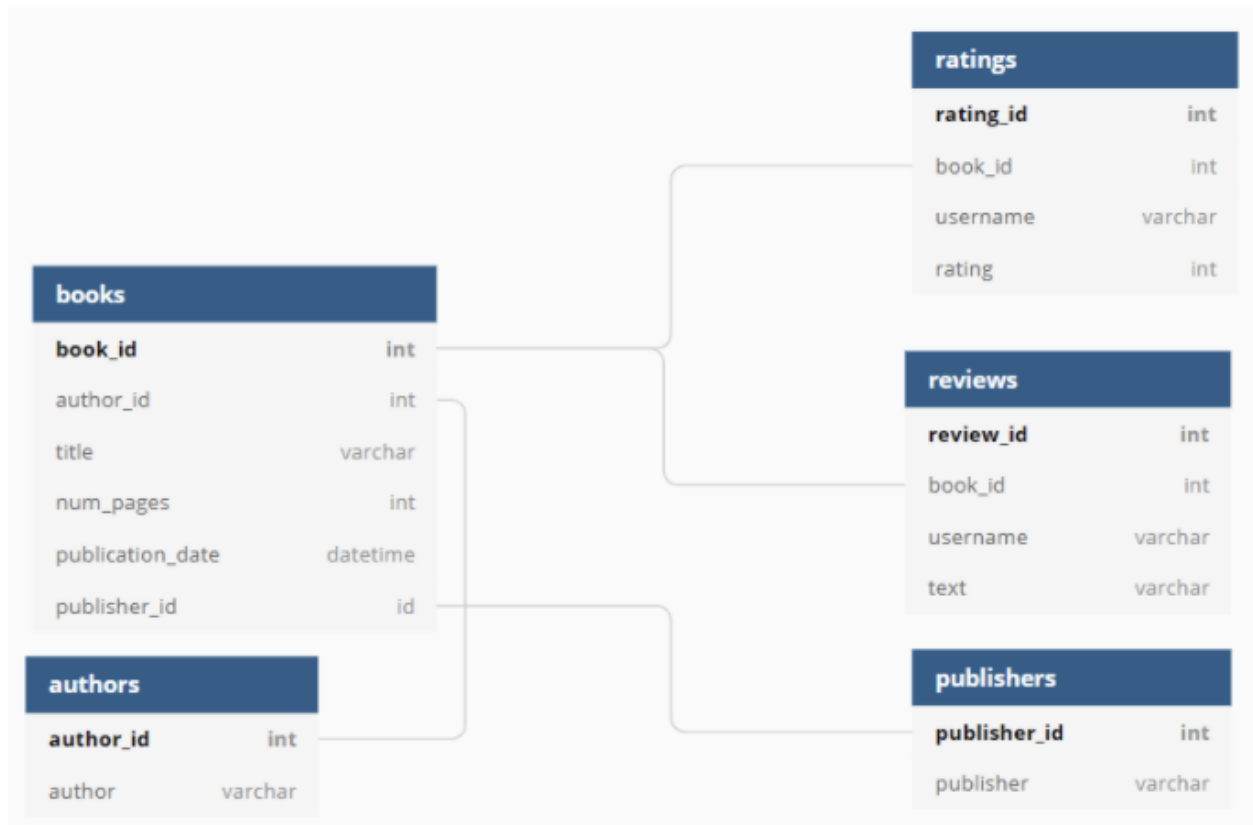
- rating_id
- book_id
- username — the name of the user who rated the book
- rating

reviews

Contains data on customer reviews:

- review_id
- book_id
- username — the name of the user who reviewed the book
- text — the text of the review

Data diagram



Task

- Find the number of books released after January 1, 2000.
- Find the number of user reviews and the average rating for each book.
- Identify the publisher that has released the greatest number of books with more than 50 pages (this will help you exclude brochures and similar publications from your analysis).
- Identify the author with the highest average book rating (look only at books with at least 50 ratings).
- Find the average number of text reviews among users who rated more than 50 books.

Instructions for completing the task

- Describe the goals of the study.
- Study the tables (print the first rows).
- Make an SQL query for each of the tasks.
 - Output the results of each query in Copy link the Notebook.
- Describe your conclusions for each of the tasks.

Access to the data base

Follow the instructions on connecting to the data base.

Notes

- Don't forget functions! They can make your life and running queries considerably easier.
- Your results must be obtained with SQL. Use pandas only to print and store query results.