The data processing：

About the database design, we have to prepare data to try whether the program can run or not;

The book data are from douban-book, mainly from python crawler,

We select the title, cover-url, author, publishers, date, price star, summary of the top-250

The other data are mainly self-created, except the danmu and the review are from BiliBili, we use python script to create users and other information randomly, and modify the data to make them more Structured.

Finally, we store them into different excels and use python script to read the file to output SQL statement to insert data to the database. The data are mainly for test so there may be some de facto mistakes, for example, the publisher time of press or the birth time and died time of some author. But that's not the point and not affect our database design.