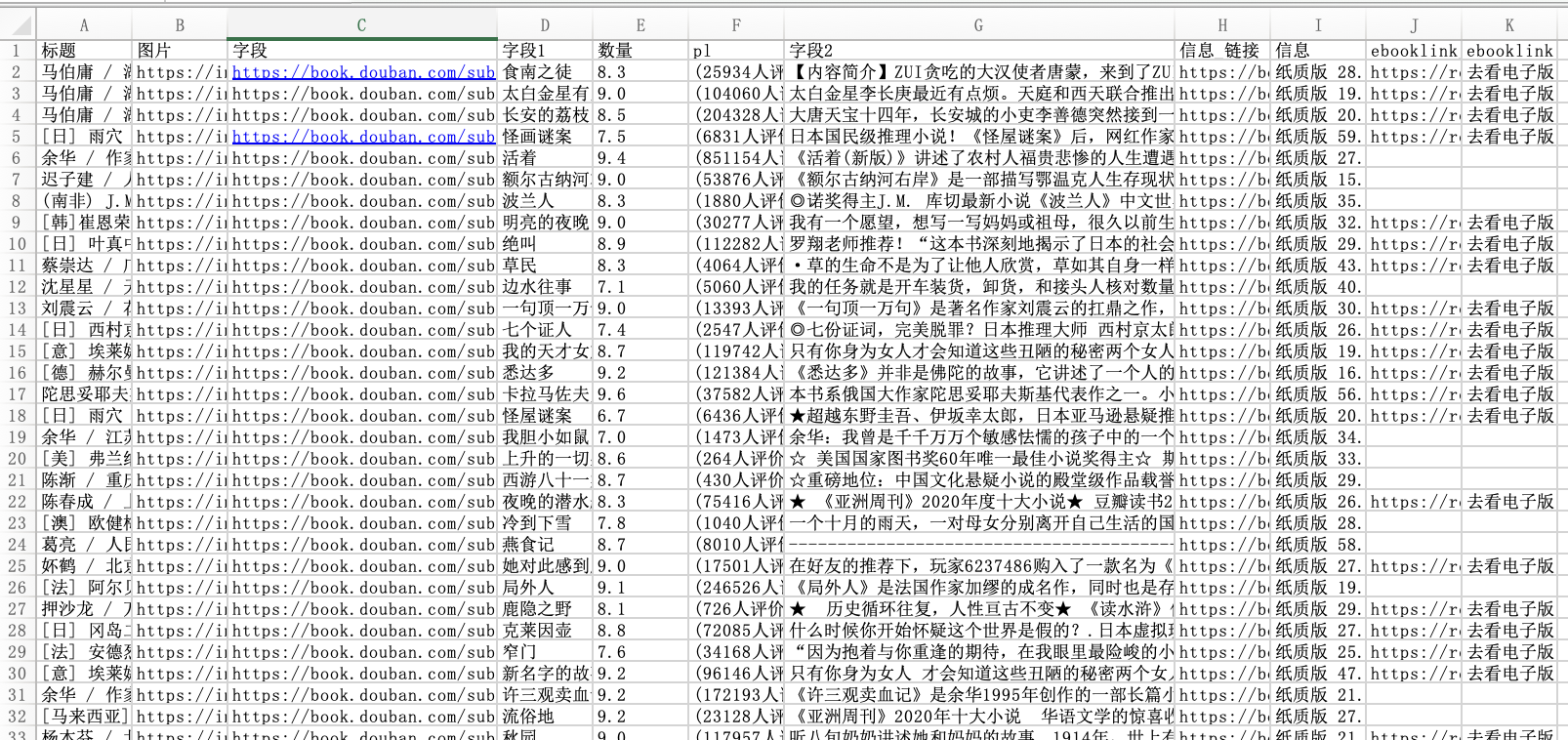
1. 豆瓣书评数据采集报告

1.进入页面后，首先设置循环翻页，循环次数设置为50次，每页都提取如图中的列表数据，摘取每一本书的信息。



完成后得到如下图所示的1000行图书的数据：



2.我们把C列中每一个书的链接复制，粘贴到八爪鱼中，使其循环网址以爬取每本书的短评：

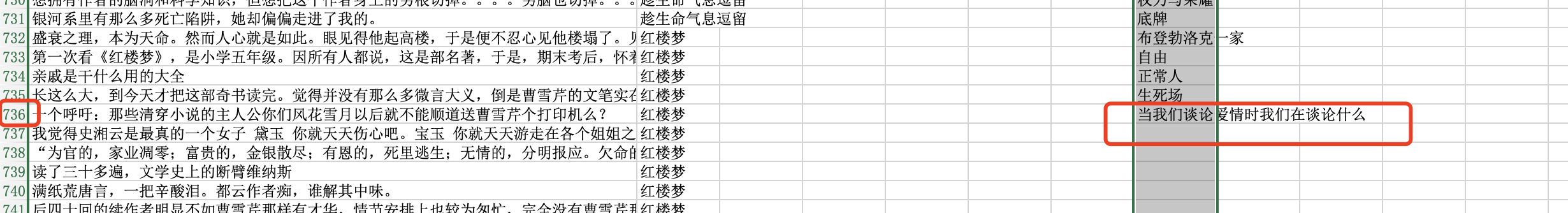


’点击元素‘对应的就是‘更多短评’， 点进去之后提取前十条的列表数据

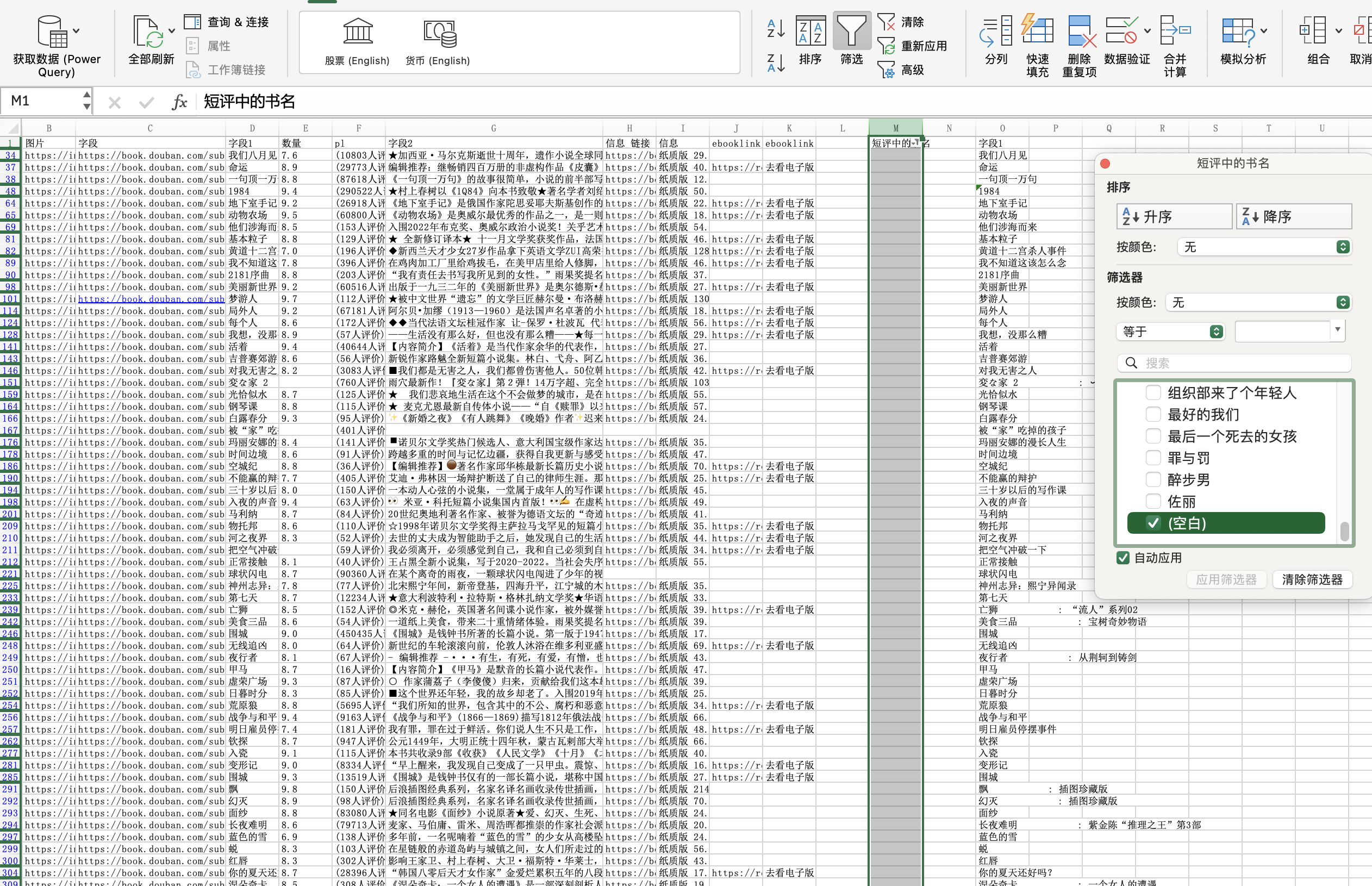


3.得到如下图的短评数据，每本书对应10条，共计**8150**条记录：

4.随后我们发现短评对应的书的数量只有735个，低于我们采集的1000本书：



我们使用excel中的公式=IFERROR(INDEX(M:M, MATCH(D2, D:D, 0)), "")将短评表中出现的书名与书信息表中出现的书名进行对应，然后将没有对应的书的链接筛选出来：



将这些没有成功爬到短评的书的链接再次拷贝到八爪鱼中进行短评数据提取，做法与步骤2相同：



最终将这些短评合并到原先的短评表中，操作完成。