1. 数据爬虫

爬取豆瓣电影（https://movie.douban.com/）中的电影评论信息，例如《哪吒之魔童降世》的影评



url为<https://movie.douban.com/subject/26794435/reviews?start=20>，其中，subject后面的参数应该是电影的id，start后面的参数是评论的偏移值，应该是20的整数倍。使用requests库对数据进行爬取，并使用pandas库对数据进行存储。



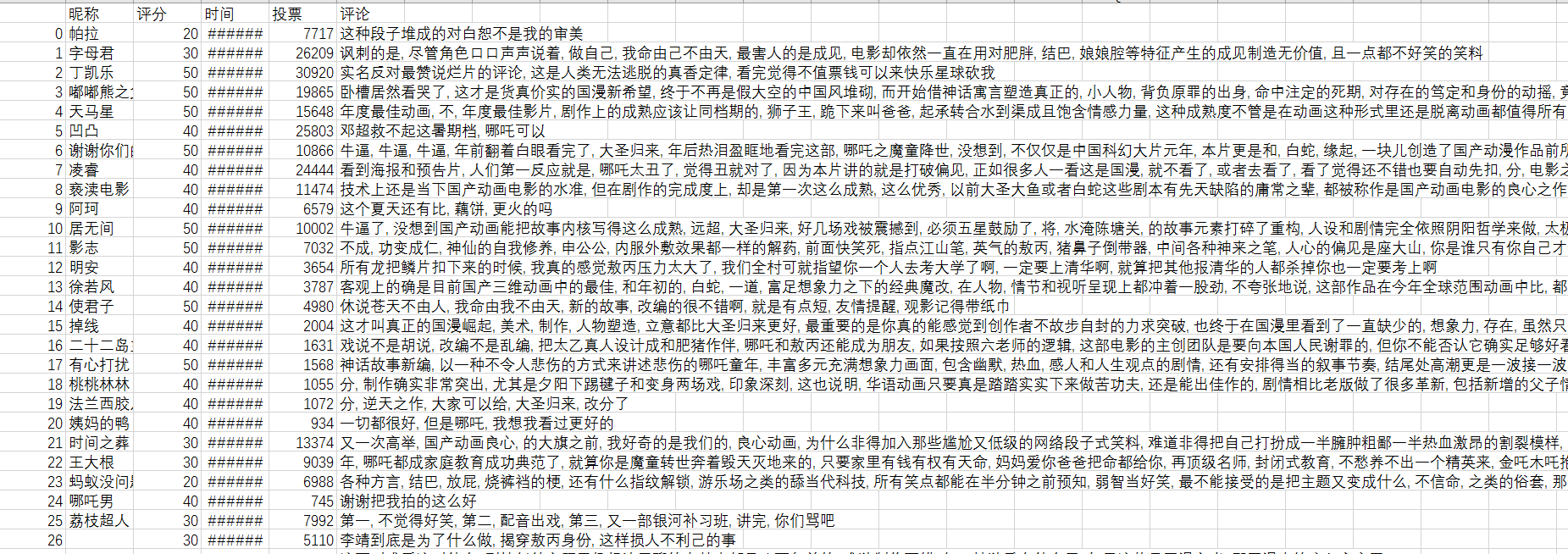
主函数中对url进行规整，并调用get\_url()函数对url进行http请求。



get\_url()函数接受url和函数报表头，请求后返回一个html源码，get\_data()函数对数据进行解析并保存至本地文件。

1. 数据分析

爬取了《哪吒之魔童降世》中的影评，影评效果如下：



1. 从评分上去分析，平均分为31.3分，可以看出，该部电影在豆瓣上的评分只能用中规中矩来形容。
2. 在评论上看，褒贬不一，评分中打了50分的认为这部电影还不错，评价非常高，但是打了低分的用户对这部电影则是满满的吐槽。