**Python大作业结果分析**

**班级：电信科20-1班 姓名：洪祎麟 学号：2020217793**

在本次Python大作业中，我主要是针对四个问题进行了分析与研究。分别是：

1. 动画片在近年来的发展现状。
2. 了解动画片具体讲述的主题，统计动画片在当季播出之后的观看情况
3. 根据一部看过的动画片，分析与该动画片最相似的动画片。
4. 根据用户所输入的动画片，搜索出对应动画片的相关信息

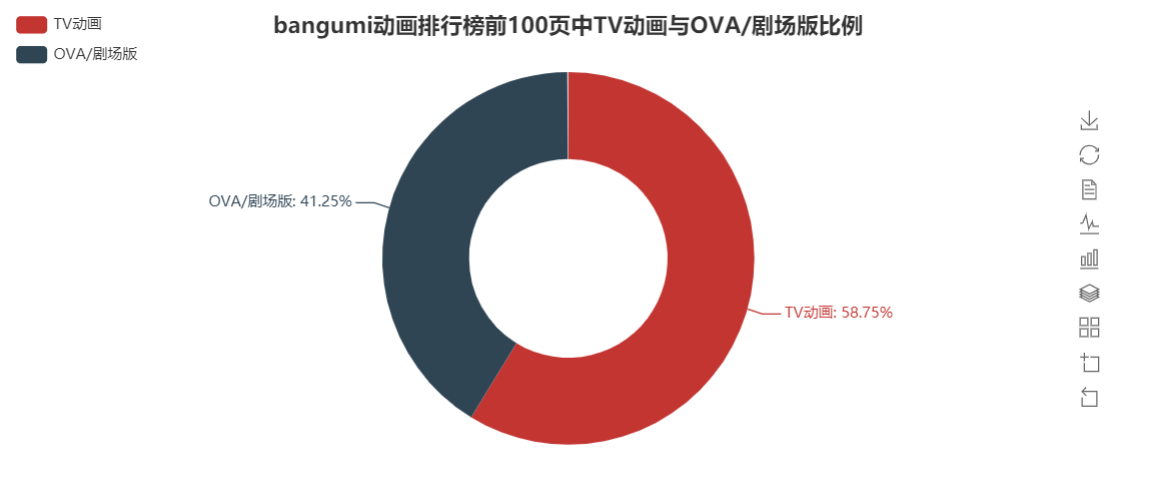
下面分别对四个问题所获得的数据进行分析。

1. 动画片近年来的发展现状

针对动画片的发展现状问题，我在Bangumi中爬取了动画片历史排行榜的前100页，2400条数据。对于数据进行清洗，将部分数据导演为空/评分为空/日文名为空的情况进行去除。对于为空的时候，除了评分均使用空替代，评分以均分替代，最终提取出的结果保存到AnimeRanking.xlsx中：

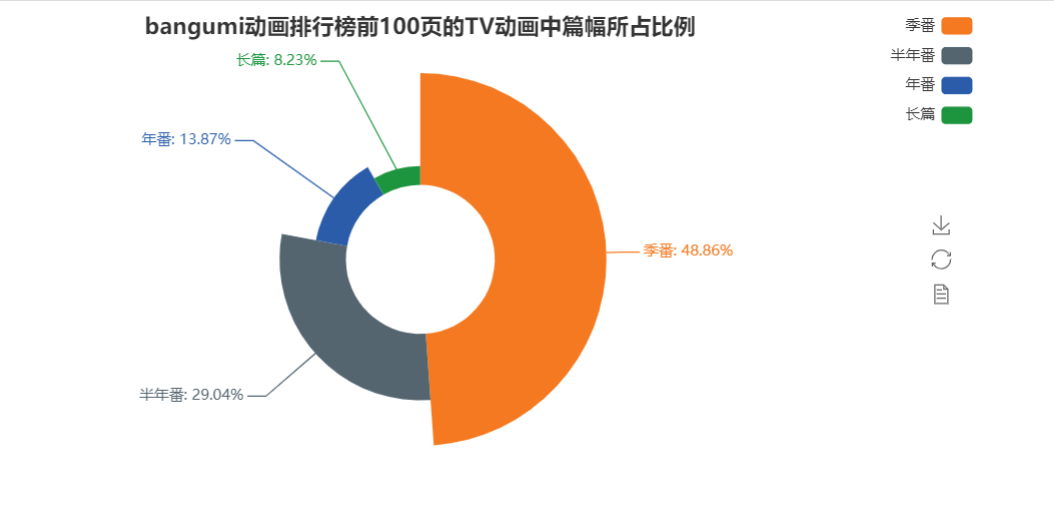


首先：在这个数据中，以动画的集数为判断标准，可以筛选出TV动画和剧场版/OVA动画，具体的比例如下图：



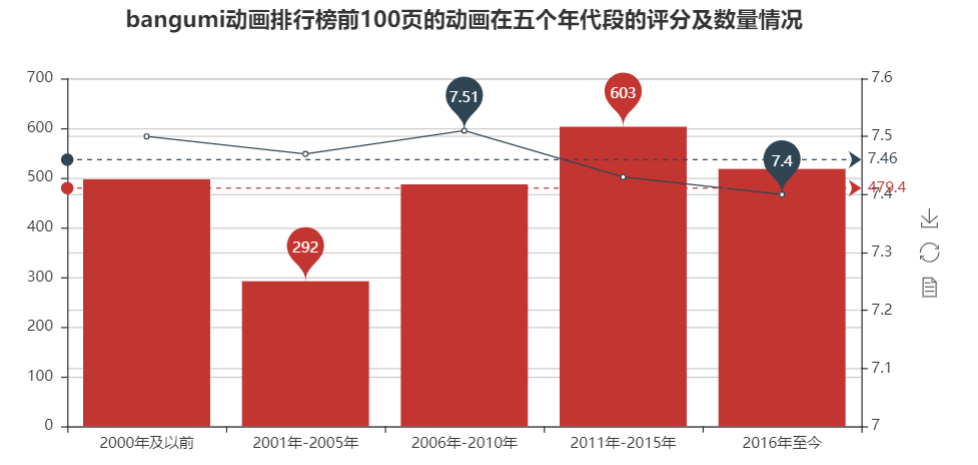
可以看出，前100页的动画中TV动画略多于剧场版动画，说明相对来说还是TV动画更受欢迎一些，每年的TV动画也比剧场版动画稍多一些。

其次，对于动画片的篇幅，所得到的具体结果如下图：



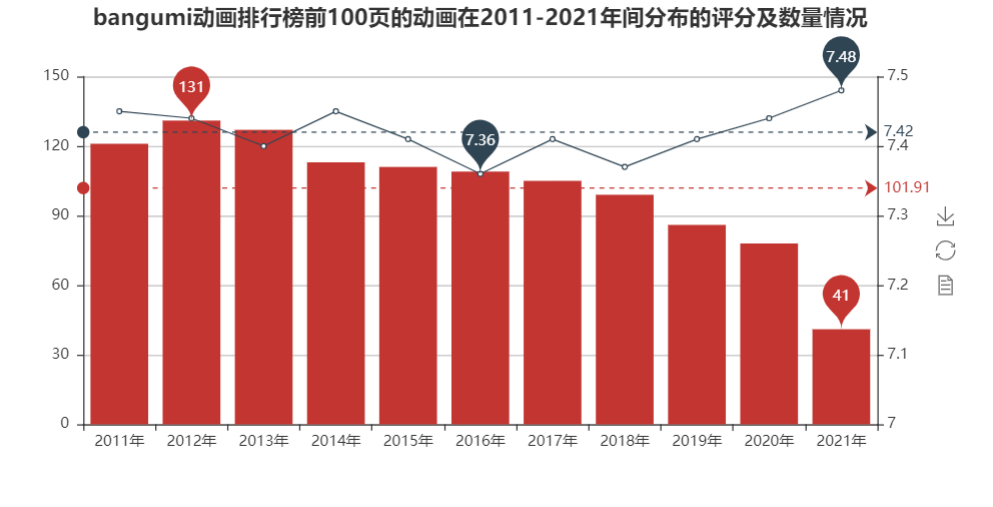
从图中可以看出，在Bangumi排行榜的前100页中，季番的数量相对还是最多的，大约占了半数左右，半年番和年番相对来说少一些。

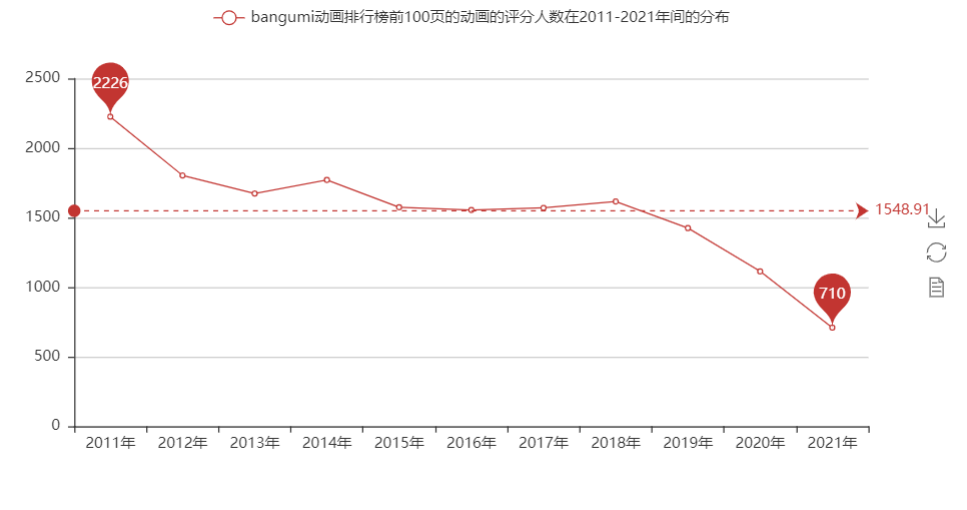
接着，我对前100页动画中的动画片出现的年代进行分析，分析具体哪个时间段的动画片质量较高，这里我选取了动画片的评分和出现年代进行分析，分析的结果如下图：



根据统计结果可以看出，动画片的平均评分确实在渐渐走低，而动画片出现的数量有所增加。说明动画片近年来渐渐走向走量的方式，而相对来说动画片的质量有所下降，总体的均分也有所下跌。但总的来说分数变化幅度不大，较为稳定。

最后，我选取2011-2021这些年的动画片进行分别分析，分析所观看动画片的人数和动画片的均分数量情况，具体的情况如下图：





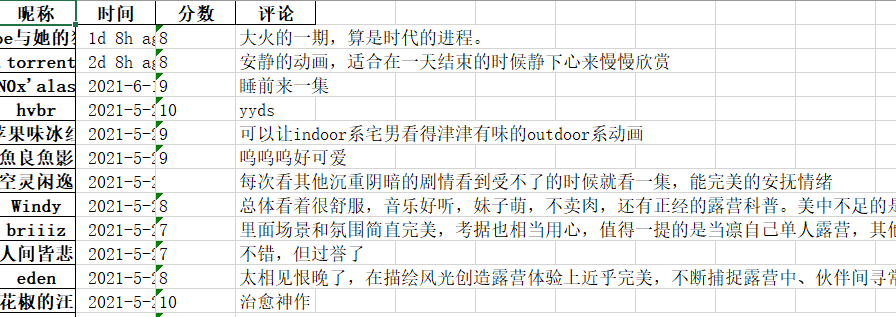
由于2021年动画片只有两季，评分人数不好分析，但而可以看过观看动画片的人数确实有所下降。观众们更倾向于观看之前的老番，最主要就是2011年的番剧，观看的人数最多。在评分方面，近11年出现在前100页动画片数量大体相同。从18年开始，这几年的评分有走高趋势，可以看出动画片的质量在这几年慢慢上升。

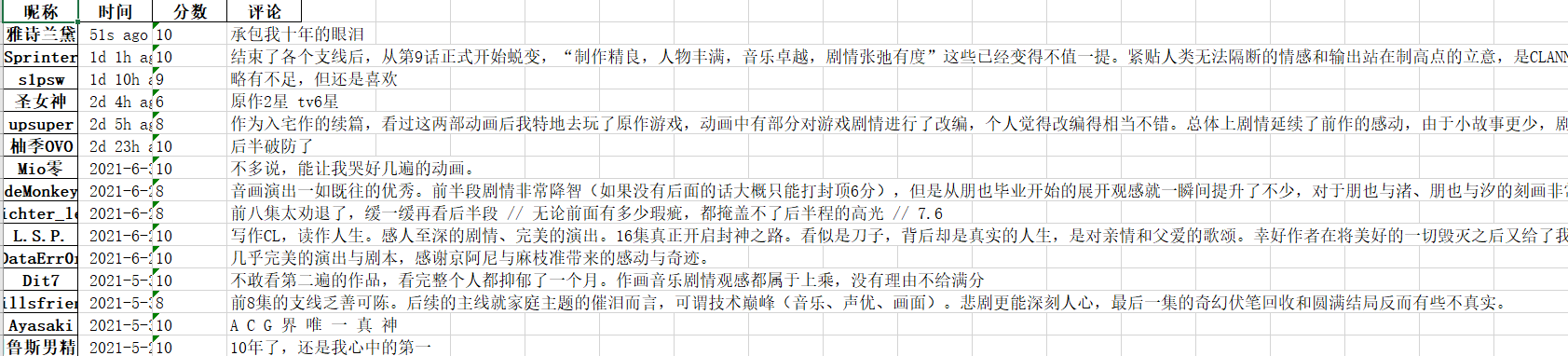
总的来说，这些年动画片的发展较为平稳，这几年的发展稳中有升。

1. 动画片的评论及观看情况分析

在本部分，我主要选取了两部动画片进行分析：《Clannad After Story》和《摇曳露营》，通过对这两部动画片的评论区进行重点分析。

首先，我先提取了《Clannad After Story》和《摇曳露营》的评论区数据。CLAS的评论区选取了145页，摇曳露营的评论区选取了70页，提取出的评论区对评分和特殊字符进行清洗处理，得到了的两个excel表格即为筛选出来的评论数据。提取出来的评论数据如图：





针对每位用户的具体评论信息，我分别制作了两部动画片评论的词云图。两部动画片的评论区词云图分别如下：

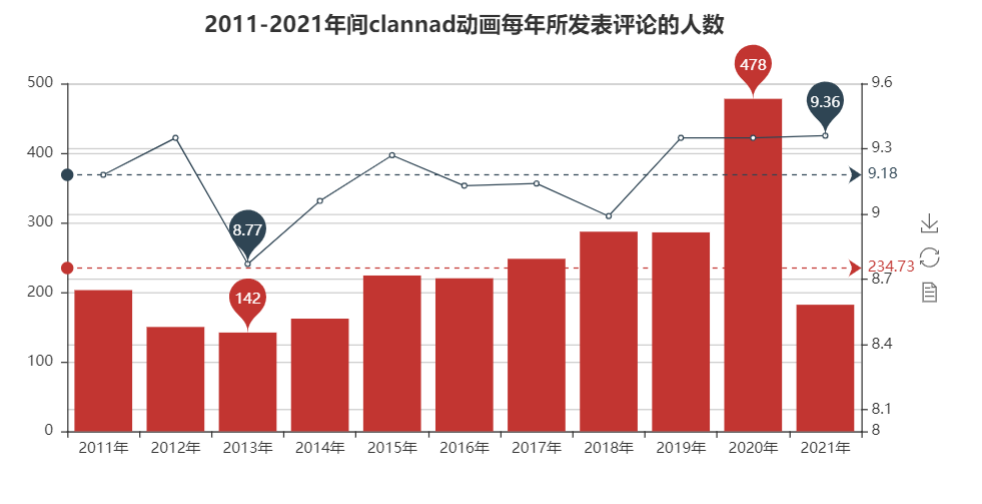


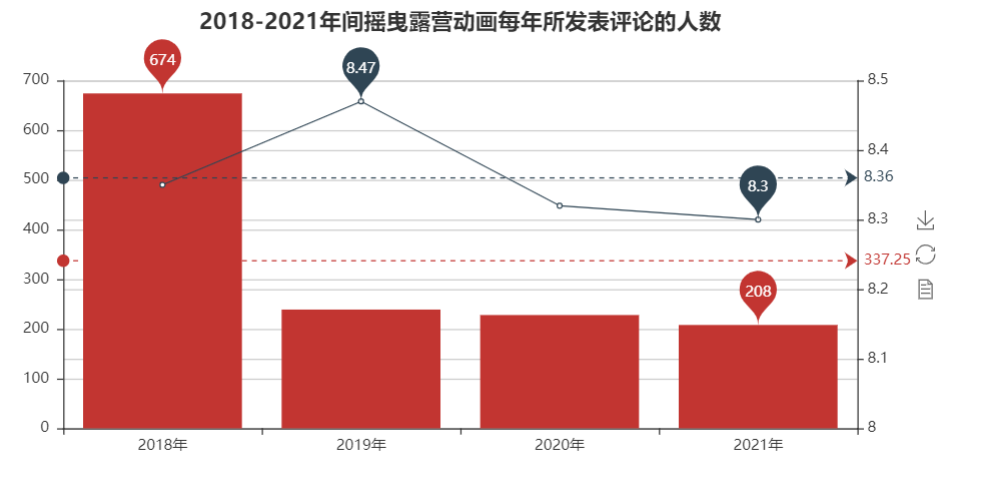


通过词云图可以清楚的看出，对于《CLAS》许多人的评论主要集中在“人生”，“剧情”，“催泪”，“感动”，“亲情”上，说明这部作品可能是一部催泪作品，部分剧情非常令人感动，可能是关于一位动画人物的亲情故事，刻画了一个人物完整的人生，让很多人想到了自己。

而对于《摇曳露营》，主要出现的词语有“露营”，“舒服”，“音乐”，“感觉”，“轻松”等，说明这部动画主要是一部悠闲的动画，因为和氛围非常令人舒服，让人感到放松，主要讲述的可能是主角们前去露营的故事。

最后，我对两部动画在播出时间所发表的评论人数和平均分进行对比，生成的图表如下：



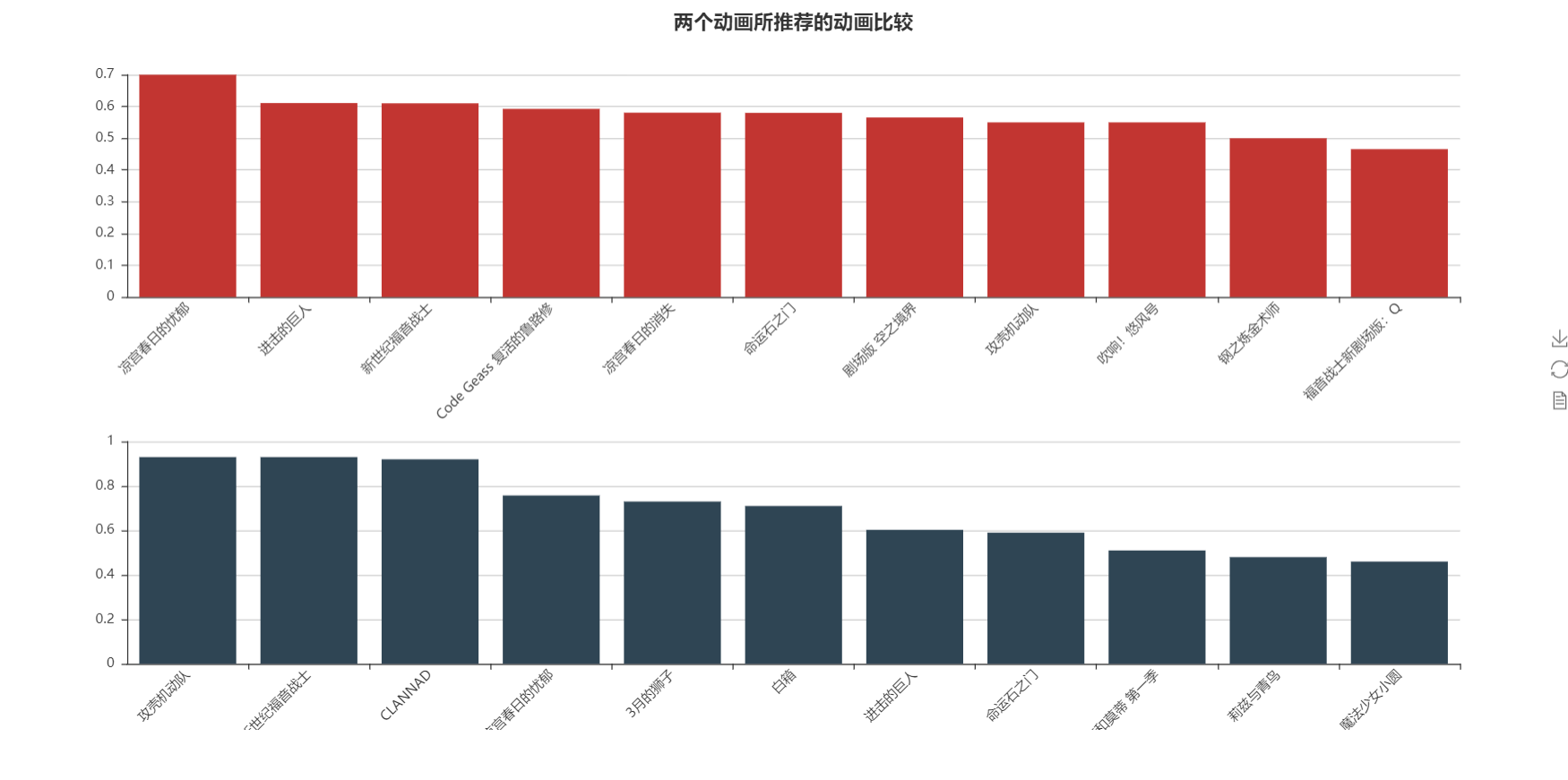


根据生成的图表可以看出，动画在刚播出的一年观看的人数非常多，在后续会保持一个比较稳定的水平。整体动画的评分会发生一些小的波动，但大体维持在稳定的水平。对于clannad可能是由于2020年的疫情原因，更多人在居家隔离时看的，因此发表评论的人数明显增加。

1. 动画片的相似推荐

在这一部分中，我也主要以《CLANNAD After Story》和《摇曳露营》这两部动画为例，进行动画片的相似性推荐。

针对相似性推荐，我主要选择了两个指标，动画的标签和动画的评论者所喜欢看的动画。根据这两个指标我设置了1:1权重，最终统计出的两部动画推荐结果如图：（上为CLAS，下为摇曳露营）



根据结果可以大致的看出适合用户看的其他动画有哪些。针对部分动画（如进击的巨人等），在两部动画的推荐榜中都有出现，说明这部动画的受众相对较广。在两部动画的推荐也可以明显看出，比较符合词云图中所主要出现的数据，大体上都符合响应的类型。

针对10个数据可能并不明显，因此，也可以根据excel文件输出的数据反映这一情况（前一部分），具体的表格数据见下图（上为CLAS，下为摇曳露营）：





通过excel中表格的数据，可以发现更多与之类型相似的动画片出现在其中，进一步验证了推荐的准确性。

1. 动画片的信息查询

第四部分主要是一个动画片查询模块，可以在这里输入想要查询的动画片信息，在多选框选择后，即可跳转到对应bangumi网站的对应动画片。这里主要是进行了pyqt多窗口的控制，和爬虫，根据输入的关键词进入对应网站进行搜索，返回对应信息。本部分的截图主要如下：



