

项目四 国产烂片深度揭秘 —— 电影豆瓣评分数据挖掘

考题01 - 电影评分整体情况如何？

作业要求：读取数据，以“豆瓣评分”为标准，看看电影评分分布，及烂片情况

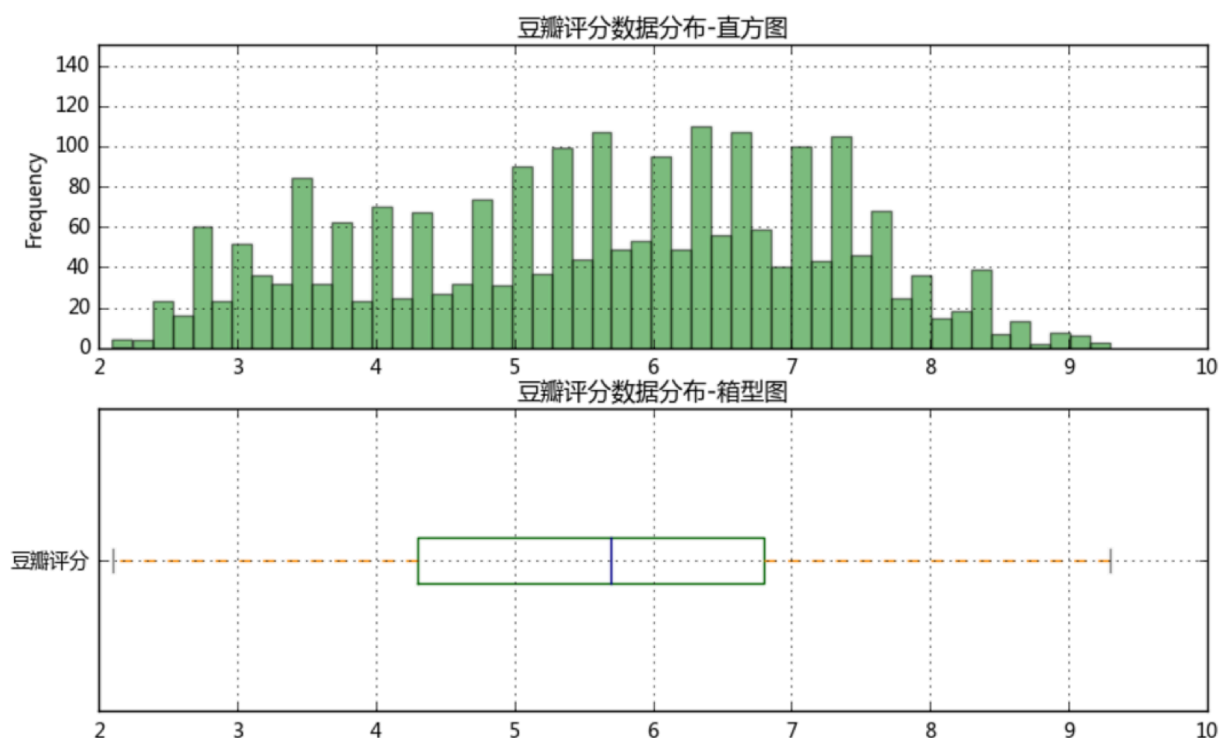
要求：

- ① 读取数据“moviedata.xlsx”
- ② 查看“豆瓣评分”数据分布，绘制直方图、箱型图
- ③ 判断“烂片标准” → 这里以上四分位数（该样本中所有数值由小到大排列后第25%的数字）评分为“烂片标准”
- ⑤ 筛选出烂片数据，并做排名，找到TOP20

提示：

- ① 读取数据之后去除缺失值

烂片评价标准：4.3分，整理后烂片数据大概546条



数据整理后，得到烂片数据量为546条

	电影名称	豆瓣评分	导演	主演
0	渡灵人	2.1	孔令杭	王磊 / 邱佩霖 / 杜冯欣怡 / 郑礼宇
1	白日杀机	2.2	张帆	张小仟 / 章宇 / 郑诺 / 亦涵
2	大震撼	2.2	崔立新	石小群 / 宫哲 / 徐爽 / 齐如意 / 姚心蕊 / 崔可法 / 卢彦西
3	威继光英雄传	2.2	姚林 / 杨清峰 / 吴鲁光 / 沈丹治	NaN
4	B区32号	2.3	吕建民	邓紫衣 / 唐宸禹 / 李锐
5	找个高富帅?	2.3	陈童	徐申东 / 张惠鑫 / 曲昱瞳 / 王乐天 / 王妍 / 闫美伊 / 陈童 / 李飞
6	换脸新娘	2.3	林云翔	翟文斌 / 吴菲 / 褚颖颖 / 林娜
7	皇帝的新装之圣诞大明星	2.3	李霞	NaN
8	整形归来	2.4	张翔宇	苏菲 / 张乃歌 / 张慧雯
9	火星归来	2.4	刘谋峰	马诺 / 胡北 / 高森鹏 / 刘谋峰 / Edward Zee
10	雷锋的微笑	2.4	张玉中	万思维 / 唐国强 / 徐箭 / 巫刚 / 李幼斌 / 宋春丽 / 冯恩鹤 / 马晓伟 / ...
11	不死之身	2.4	邵传勇	吴启华 / 吕颂贤 / 邵传勇
12	南泥湾	2.4	彭丹	彭丹 / 任帅 / 姬他
13	少年杨靖宇	2.4	张文艺	叶艺城 / 温佳翼 / 马新皓 / 杜旭东 / 栗思涵
14	青春.com	2.4	马雍	古丽娜 / 曲国强 / 孙夕尧 / 严萌 / 陶玉洁 / 卢小璐 / 马一傲
15	幸福很囧	2.4	马雍	章小军 / 张希爱 / 慕钰华 / 姚红军 / 卢小璐 / 孙旋 / 邓杨宁 / 裴妤辰 ...
16	悟空大战二郎神	2.4	梁汉森 / 徐鹏海	NaN
17	超能兔战队	2.4	傅燕	刘红韵 / 李旭乔 / 白文显 / 申克 / 宋磊 / 周岩 / 李思媚 / 张子焜 / ...
18	净土	2.4	刑健	孙文婷
19	我是大明星	2.4	张艺飞	刘波 / 腾飞 / 马瑞曼 / 谭皓 / 高天 / 聂琳峰

考题02 - 什么题材的电影烂片最多？

- ① 按照“类型”字段分类，筛选不同电影属于什么题材
- ② 整理数据，按照“题材”汇总，查看不同题材的烂片比例，并选取TOP20
- ③ 将得到的题材烂片比例TOP20制作散点图 → 横坐标为“题材”类型，纵坐标为烂片比例，点大小为样本数量

提示：

- ① 删除“类型”字段空值的数据
- ② 由于一个电影“类型”会有多个，这里需要将一个电影每个“类型”都识别出来，在统计某个题材时都需要计算，例如：

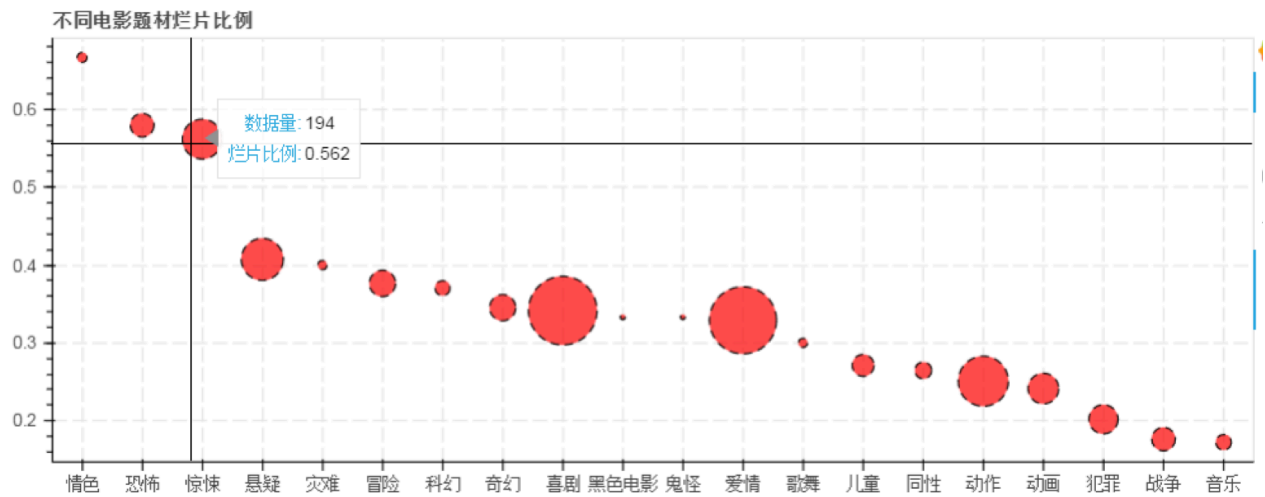
如果一个电影的类型为：“喜剧/爱情”，则在计算“喜剧”、“爱情”题材的烂片比例时，都需要将该电影算上

- ③ 注意类型字段中，要删除空格字符

结果参考：

	type_lp_pre	typecount	typename
22	0.666667	12	情色

32	0.579710	69	恐怖
29	0.561856	194	惊悚
28	0.407407	216	悬疑
14	0.400000	10	灾难
18	0.376471	85	冒险
16	0.370370	27	科幻
34	0.345238	84	奇幻
33	0.341463	574	喜剧
4	0.333333	3	黑色电影
8	0.333333	3	鬼怪
31	0.329091	550	爱情
27	0.300000	10	歌舞
2	0.271186	59	儿童
26	0.264706	34	同性
0	0.250814	307	动作
15	0.241379	116	动画
25	0.201923	104	犯罪
24	0.176471	68	战争
17	0.172414	29	音乐



考题03 - 卡司数量是否决定电影成为烂片？

① 计算每部电影的主演人数

② 按照主演人数分类，并统计烂片率

** 分类: '1-2人','3-4人','5-6人','7-9人','10以上'

③ 查看烂片比例最高的演员TOP20

提示：

① 通过“主演”字段内做分列来计算主演人数

② 需要分别统计不同主演人数的电影数量及烂片数量，再计算烂片比例

③ 这里可以按照明星再查看一下他们的烂片率，比如：吴亦凡、杨幂、黄晓明、甄子丹、刘亦菲、范冰冰....

	电影数量	烂片数量	烂片比例
主演人数分类			
1-2人	273	30	0.109890
3-4人	651	117	0.179724
5-6人	466	130	0.278970
7-9人	426	157	0.368545
10人及以上	375	104	0.277333

	电影名称	主演	豆瓣评分
38	爵迹	范冰冰 / 吴亦凡 / 陈学冬 / 陈伟霆 / 郭采洁 / 杨幂 / 林允 / 严屹宽 / ...	3.7
57	致青春·原来你还在这里	吴亦凡 / 刘亦菲 / 金世佳 / 李沁 / 李梦 / 郝劭文 / 陈燃 / 乔任梁	4.0
350	有一个地方只有我们知道	吴亦凡 / 王丽坤 / 徐静蕾 / 戈登·亚历山大 / 丛珊 / 张超 / 热依扎	5.0