

网易云课堂 - 数据分析课程（高薪全能班）

# 毕业项目： 国产烂片深度揭秘

版本: V 1.0



网易云课程 x 微专业

## 项目介绍

### 1. 考核基本要求：

- 使用 Jupyter Notebook 完成毕业项目开发
- 提交 Jupyter Notebook 文件
- 尽量在 10 个小时以内完成毕业项目

### 2. 考核要求：

#### 2.1 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

##### 2.1.1 要求：

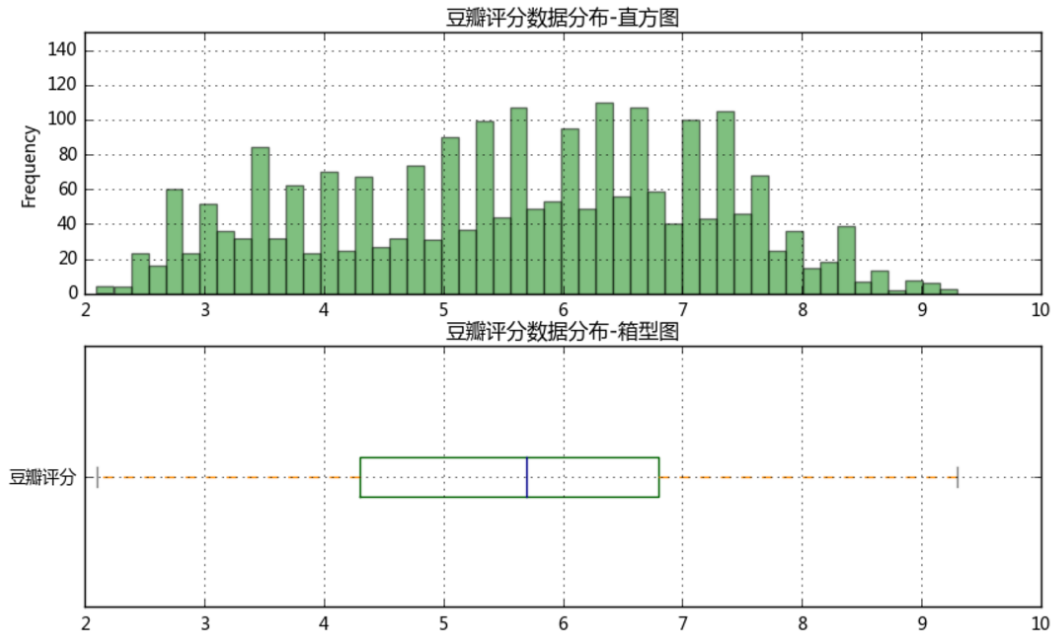
- ① 读取数据 “moviedata.xlsx”
- ② 查看“豆瓣评分”数据分布，绘制直方图、箱型图
- ③ 判断“烂片标准” → 这里以上四分位数（该样本中所有数值由小到大排列后第 25%的数字）评分为“烂片标准”
- ④ 筛选出烂片数据，并做排名，找到 TOP20

##### 2.1.2 注意：

- 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

- 图示：



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 / 邱佩霞 / 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 / 沈丹泊	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 / 马一微
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 / 裴好辰 ...
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 / 李思娴 / 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

## 2.2 什么题材的电影烂片最多？

### 2.2.1 要求：

- 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

### 2.2.2 注意：

- 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

	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

## 2.3 和什么国家合拍更可能产生烂片？

### 2.3.1 要求：

- ① 按照“制片国家/地区”字段分类，筛选不同电影的制片地
- ② 整理数据，按照“题材”汇总，查看不同题材的烂片比例

### 2.3.2 注意：

- ① 删除“制片国家/地区”字段空值的数据
- ② 删除“制片国家/地区”中不包括“中国大陆”的数据
- ③ 制片地删除“中国大陆”、“中国”、“台湾”、“香港”等噪音数据
- ④ 筛选合作电影大于等于 3 部以上的国家

	loc	loc_lp_pre	loccount
18	英国	0.750000	4
9	新加坡	0.333333	3
20	美国	0.333333	21
1	德国	0.250000	4
12	韩国	0.157895	19
8	法国	0.100000	10
5	日本	0.071429	14
7	加拿大	0.000000	6
13	澳大利亚	0.000000	3

## 2.4 卡司数量是否和烂片有关？

### 2.4.1 要求：

- 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 烂片比例最高的演员 TOP20

### 2.4.2 注意：

- 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 这里可以按照参演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

```
role  role_lp_pre  rolecount
521  吴亦凡      0.666667      3
```

	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

## 2.5 不同导演每年电影产量情况是如何的？

### 2.5.1 要求：

- ① 通过“上映日期”筛选出每个电影的上映年份
- ② 查看不同导演的烂片比例、这里去除掉拍过 10 次电影以下的导演

### 2.5.2 注意：

- ① 注意要删除“上映日期”中的空格字符
- ② 绘制图表时，分开建立数据绘制

	dir	dir_lp_pre	dircount
1	王晶	0.210526	19
4	徐克	0.090909	11
3	周伟	0.071429	14
2	邓衍成	0.031250	32
0	海涛	0.000000	12
5	胡明凯	0.000000	16

	电影名称	导演	豆瓣评分	上映日期	year
7	我心雀跃	刘紫微	6.7	2017-04(中国大陆)/2016-06-14(上海国际电影节)	2017
19	龙门劫案	王勇	4.0	2016年	2016
20	东北小佛爷	谢云鹏	6.5	2016-6-16	2016
21	一个成人电影工作者的自白	贾小熊	7.6	2016-6-15	2016
23	宠灵实验室之狸奴艾莉	郝昭赫	3.7	2016-4-1	2016