【项目11】 国产烂片深度揭秘

作业要求

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

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

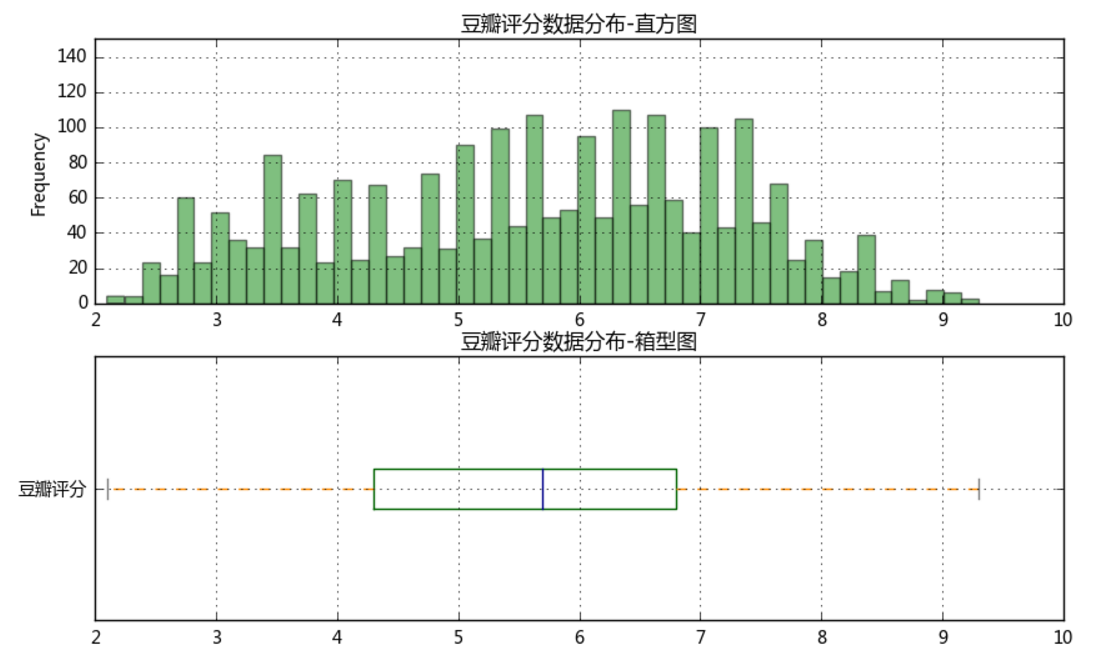
③ 判断“烂片标准” → 这里以上四分位数（该样本中所有数值由小到大排列后第25%的数字）评分为“烂片标准”

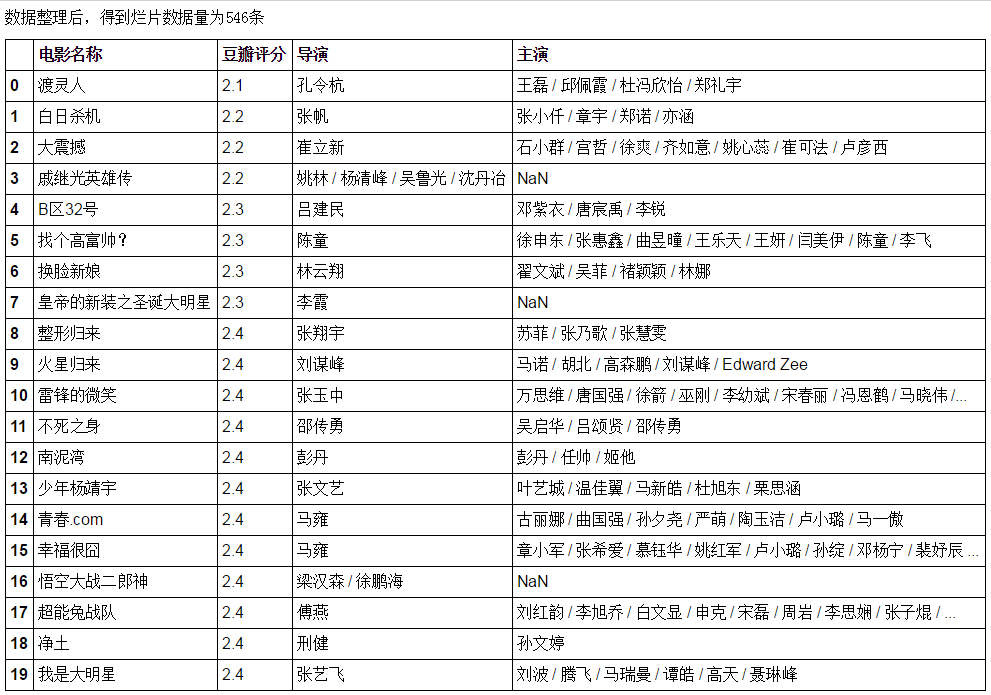
⑤ 筛选出烂片数据，并做排名，找到TOP20

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





2、什么题材的电影烂片最多？

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

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

\*\* 用bokeh制图

\*\* 按照烂片比例做降序排列

提示：

① 删除“类型”字段空值的数据

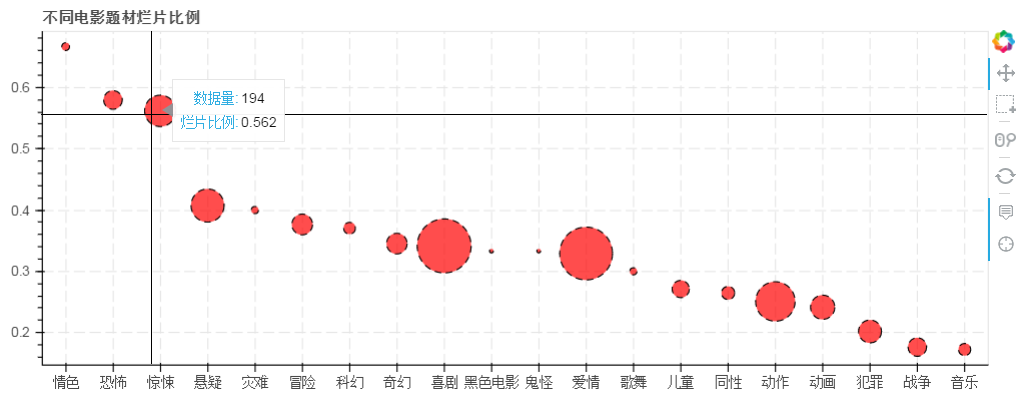
② 由于一个电影“类型”会有多个，这里需要将一个电影每个“类型”都识别出来，在统计某个题材时都需要计算，例如：

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

④ bokeh图设置点大小，这里通过 开方减小数据差距 → size = count\*\*0.5\*系数





3、和什么国家合拍更可能产生烂片？

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

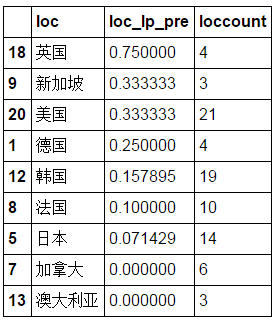
提示：

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



4、卡司数量是否和烂片有关？

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'

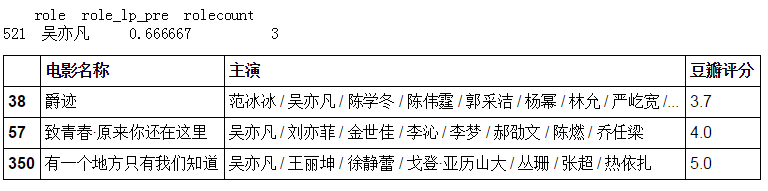
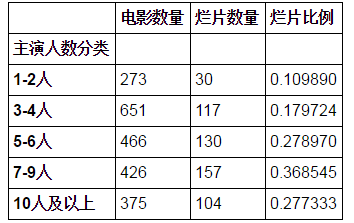
③ 查看烂片比例最高的演员TOP20

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....



5、不同导演每年电影产量情况是如何的？

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

③ 查看不同导演每年的电影产量制作散点图 → 横坐标为年份，纵坐标为每年电影平均分，点大小该年电影数量

\*\* 用bokeh制图

\*\* 横坐标限定为2007-2017年

\*\* 绘制散点图只需要用产出过烂片的导演数据

提示：

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

