

2023 年第二届全国大学生数据统计与分析竞赛

题目B：电影评分的大数据分析

大数据无处不在，在运营一个成功的商业的过程中，大数据起到的非常重要的作用。同样通过观众对电影的评分的分析，可以在一定程度上给电影行业启发。

网络影评是伴随着电子媒介和数字技术诞生的，呈现出鲜明的数据化倾向，比如电影评分、打星、榜单等。网络评分因其便捷、高效，成为大部分观众评价电影的首选方式，也是电影口碑最直观的体现，评分高低直接影响到大众的观影取向和消费选择。与文字、视频相比，大数据搜集、统计的评分信息可以让我们对观众的观影体验，审美趣味、情感态度一目了然。视频网站、电影门户网站等的后台会根据观众的评分、打星和消费行为来推断用户的审美趣味和消费需求，从而为客户提供个性化推荐和引导。

电影的评分更多反映的是大众审美，并非绝对的评判标准，不应该奉为最终裁决，主要功能是用来参考，规避较差电影的同时，发现好的电影，尤其是在规避较差电影方面有着非常显著的效果。

附件 1. 电影评分.csv 中包括了豆瓣影评排名前 250 名电影的名称、导演、电影类型、国家、上映年份、评分和评论人数。附件 2. 电影票房.csv 中包括了 10605 部电影的票房排名、电影名称、上映时间、总票房(元)、平均票价和平均场次，其中，部分数据有缺失和异常。

请你们团队使用附件中的相关数据，进行数据统计与分析，回答下列问题：

问题一：请分析附件 1 中最受欢迎的电影类型是什么？排名前 250 名电影中出现次数最多的导演前 10 名是谁？出现次数最多的国家前 5 名是哪些国家？

问题二：请分析附件 1 中排名前 250 名电影的上映年份主要集中在哪几年？排名前 250 名电影的评分与评论人数、国家、导演和电影类型是否有关系？

问题三：请你们收集相关数据，分析附件 2 中电影票房较高的电影主要是什么类型的电影？并给出这些电影的上映时间、总票房(元)、平均票价和平均场次的相关统计图表。

问题四：某导演拟定于 2024 年春节档推出一部电影，请从数据分析的角度给导演一个提案，例如考虑：电影的类型、题材、上映时间、票价等方面。