

## 项目：可视化电影数据

### 第一步：清理数据和选择变量

- 清理过程详细见data\_cleaning.html
- 使用的语言和库：Python / Pandas
- 清理后得到cleaned\_movies.csv文件
- 可视化过程中研究的变量：popularity / keywords / release\_year / budget\_ad / revenue\_ad / genre

### 第二步：Tableau 可视化

- Q1: [https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3\\_138/Q1](https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3_138/Q1)
- Q2: [https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3\\_138/Q2](https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3_138/Q2)
- Q3: [https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3\\_138/Q3](https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3_138/Q3)
- Q4: [https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3\\_138/Q4](https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3_138/Q4)
- Q4 Story: [https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3\\_138/Q4Story](https://public.tableau.com/profile/kexin.lin#!/vizhome/Project3_138/Q4Story)

### 第三步：问题

- 问题1：电影类型是如何随着时代变化而变化的？

从所有类型电影的总体占比来看，大部分的电影类型虽然百分比上下波动较大，但总体趋于稳定，对比1960年的总体占比和2015年的总体占比区别不大。然而也有特殊，比如纪录片的总体占比在二十一世纪后开始增长，而西部风格电影的总体占比在二十世纪八十年代之后锐减，从1960年的7.69%到了2015年的0.43%。

- 问题2：环球影业和派拉蒙影业的电影之前数据指标有什么区别？

从Universal Pictures和Paramount Pictures投资的不同电影风格的数量来看，两个公司的风格偏好总体来看比较相似，只存在很小的差异。从Universal Pictures和Paramount Pictures各自的预算/收入对比图来看，两个公司在1970年以后基本都处于盈利状态。对比两个公司各自的预算数目，他们都选择了在1980年左右开始增加电影预算。不同的是，Paramount Pictures在电影的资金投入方面相对Universal Pictures更加保守。对比在1985年及以前两个公司各自的收入数目，Universal Pictures的年票房收入在20世纪90年代后相比Paramount Pictures表现更好，一是因为没有任何一年的总票房收入低于五十亿，二是因为年总票房收入高于一百亿的时候也比Paramount Pictures更多。

- 问题3：和非小说改编的电影相比，基于小说改编的电影表现得怎么样？

从改编电影和原创电影的总体占比可以看出，改编电影在1960年至1980年之间中有几年的占比超过了5%，然而在1980年之后改编电影的占比在绝大多数时候不超过

5%，其余皆为原创电影。虽然在数量方面改编电影大大低于原创电影，但是在票房方面改编电影在某些年份里的平均票房大大超过了原创电影，从中可以得到猜测，改编电影相比原创电影似乎更有票房大爆的可能性。从不同风格电影的改编和原创数量来看，剧情片无论是改编还是原创的数量都是最高的，而电视电影和音乐片从小说改编来的数量最低。

- 问题4：科幻片、爱情片、战争片、恐怖片 and 纪录片这五种电影类型中，哪一种的票房回报率最高，哪一种最低？

首先通过受欢迎度来看，科幻片在以上五种电影类型中无论是平均受欢迎度还是总受欢迎度都排在第一位，而纪录片无论是平均受欢迎度还是总受欢迎度都排在五种类型片的最后。接下来查看票房收入和电影预算的散点图，发现根据趋势线的预测，每投入一美金给科幻片能够获得\$3.04左右的票房回报，然而每投入一美金给记录片只会得到\$1.25左右的票房回报。最后分析五种类型电影的平均票房收入和资金投入。可以看到科幻片的平均票房收入和资金投入的差异在五种类型片中是最大的，而纪录片是最小的，这说明科幻片可能会带来非常高的票房回报，而纪录片的票房回报很可能只能回本。