

电影票房数据可视化系统

- **信息获取**

- 实现方式和用到的库等

- beautifulsoup,xpath

- js逆向

- 字体转换 (TTFONT, ddddocr)

- 爬取的网站和内容

- 共爬取了猫眼, 豆瓣, BOX_OFFICE

- 豆瓣TOP250中英文名称, 评分, 评价人数, 影片信息, 简要影评

- 猫眼: 正在上映的电影名称, 票房, 上座率

- BOX-OFFICE: 2016~2022全球票房TOP200 (共1400条数据), 最赚钱的TOP50电影公司

- **数据库存储**

- 利用sqlite3

- 进行了数据库连接, 数据存储, 数据读取

- 共建立了3个数据库, 豆瓣数据库, 电影公司数据库, 全球票房2016~2022数据库

- **数据清洗**

- 利用pandas,numpy

- 对读取到的进行排序, 统计

- **可视化展示**

- 一共采用了两种可视化 (分别主要利用pyecharts和echarts)

- 利用pyecharts可视化, 共生成10个html文件

7个柱状图，对应2016~2022各年TOP15的电影。鼠标放在其上可以显示具体信息。对大于十亿的电影用红色突出，大于7亿美元的用粉色显示，剩下的用蓝色显示

1个饼图，显示2016~2022年全球电影票房Top100的年代分布。鼠标放在其上可以显示具体信息。可以隐藏某个年份显示剩余的。

1个环形图，显示Top15电影制作公司(Brands)，鼠标放在其上可以显示具体信息。可以隐藏某个电影公司，显示剩余的。

1个折线图，票房收入年份折线图(单位:\$)

1个矩形树状图，显示2016~2022Top电影矩形树状图(单位:\$)

- 在一个网页，利用echarts，table等进行可视化

利用javascript,html5等语言设计一个网页，优雅地展示信息

精心设计页面，精选页面下拉动态效果，标题有颜色渐变动态效果

主要组成有：

2021~2022 全球票房**TOP10** 柱状图，在同一页可选择切换

2021~2022 全球票房**TOP10** 柱状图

Top15电影制作公司 (Brands) 单位：万\$

2016~2022年全球TOP200总票房收入年份折线图(单位：\$)

近期热门电影票房及其相关数据可视化（柱状图和饼图）

豆瓣评分TOP250，利用表格实现，名称有超链接，点击可跳转到豆瓣中关于影片的详细内容

尾声：介绍项目内容，心得等，有海报欣赏

点击从头浏览，四字可以返回首页

网页结尾：标注信息来源网站，并添加超链接

所有的图表信息均来自数据库后端输入