好莱坞影片-数据可视化分析

**一、课程设计题目背景**

看电影是大家日常娱乐休闲的热门选择之一，看一场精彩的电影无疑是一种享受。然后电影制片商如何才能制作出既迎合大众口味又能获得高票房的电影呢？一部电影包含大量信息，其中不乏一些非常重要的信息，通过对电影数据进行分析，也许可以从中发现一些有用的观点和结论。

**二、数据集**

数据集介绍：2007-2011好莱坞和美国主要电影，每一部都有大量的数据：包括他们的预算、评分、票房、类型、利润和获奖情况等，具体如下：

|  |  |
| --- | --- |
| 变量 | 变量说明 |
| Film | 电影名称 |
| Major\_Studio | 制片公司 |
| Rotten\_Tomatoes | 烂番茄评分 |
| Audience\_score | 观众评分 |
| Genre | 影片类型 |
| Domestic\_Gross | 国内总票房($m：百万美元) |
| Foreign\_Gross | 海外总票房($m：百万美元) |
| Worldwide\_Gross | 全球总票房($m：百万美元） |
| Budget | 预算($m：百万美元） |
| Profitability | 收益率(%) |
| Oscar | 奥斯卡金像奖获奖情况 |
| Release\_time | 上映时间 |

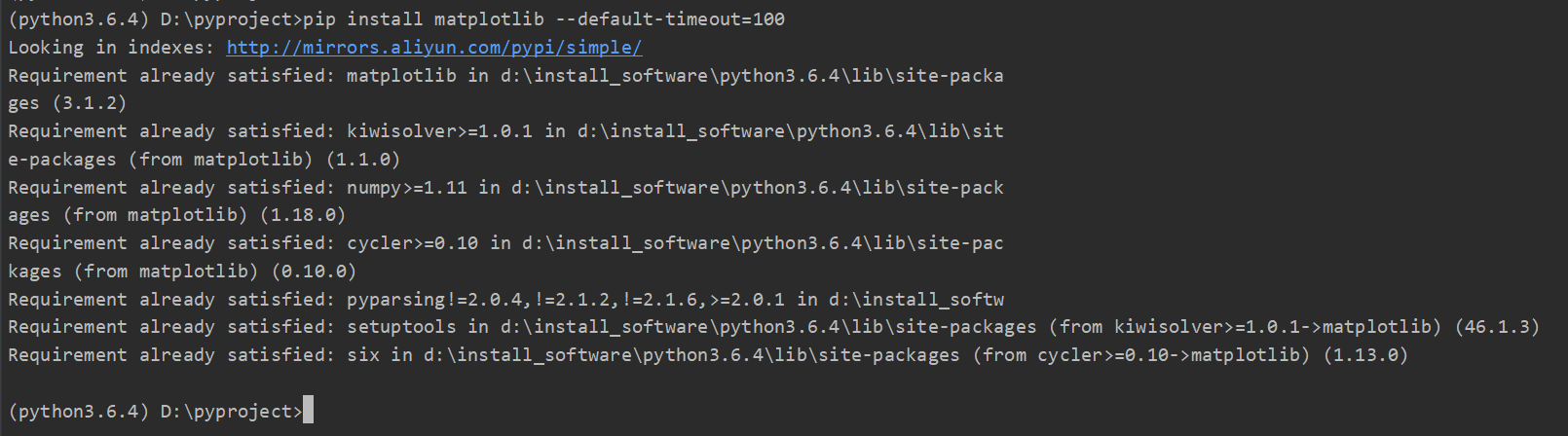
**三、课程设计要求**

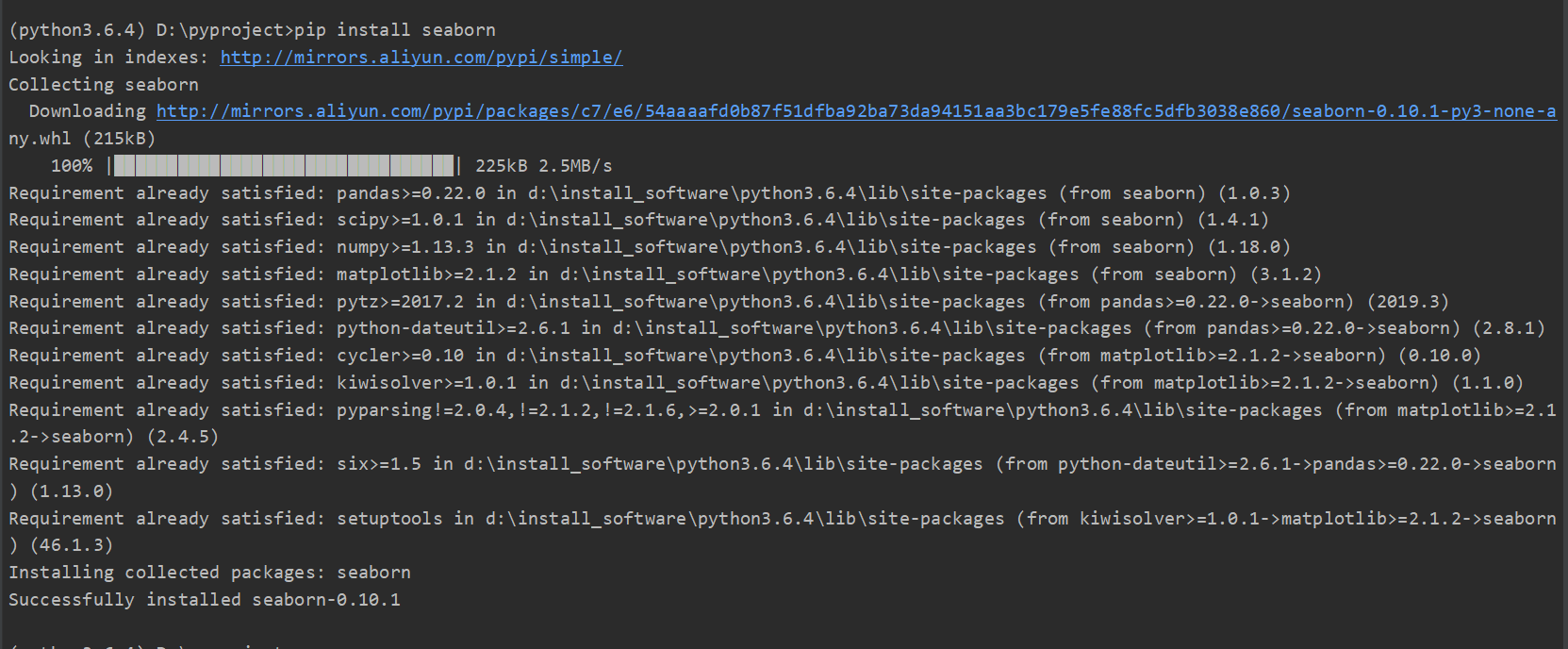
经过python数据可视化技术的学习，急需要通过一些简单的课程设计实例来将所学知识和用法融会贯通。课程设计要求利用matplotlib、seabron等数据可视化工具对好莱坞影片数据进行可视化分析。

1. 引入相关包和导入数据（10分）

1.安装

pip install matplotlib --default-timeout=100

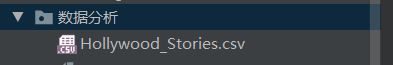


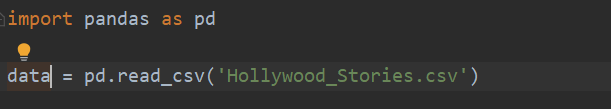


2.python引入包



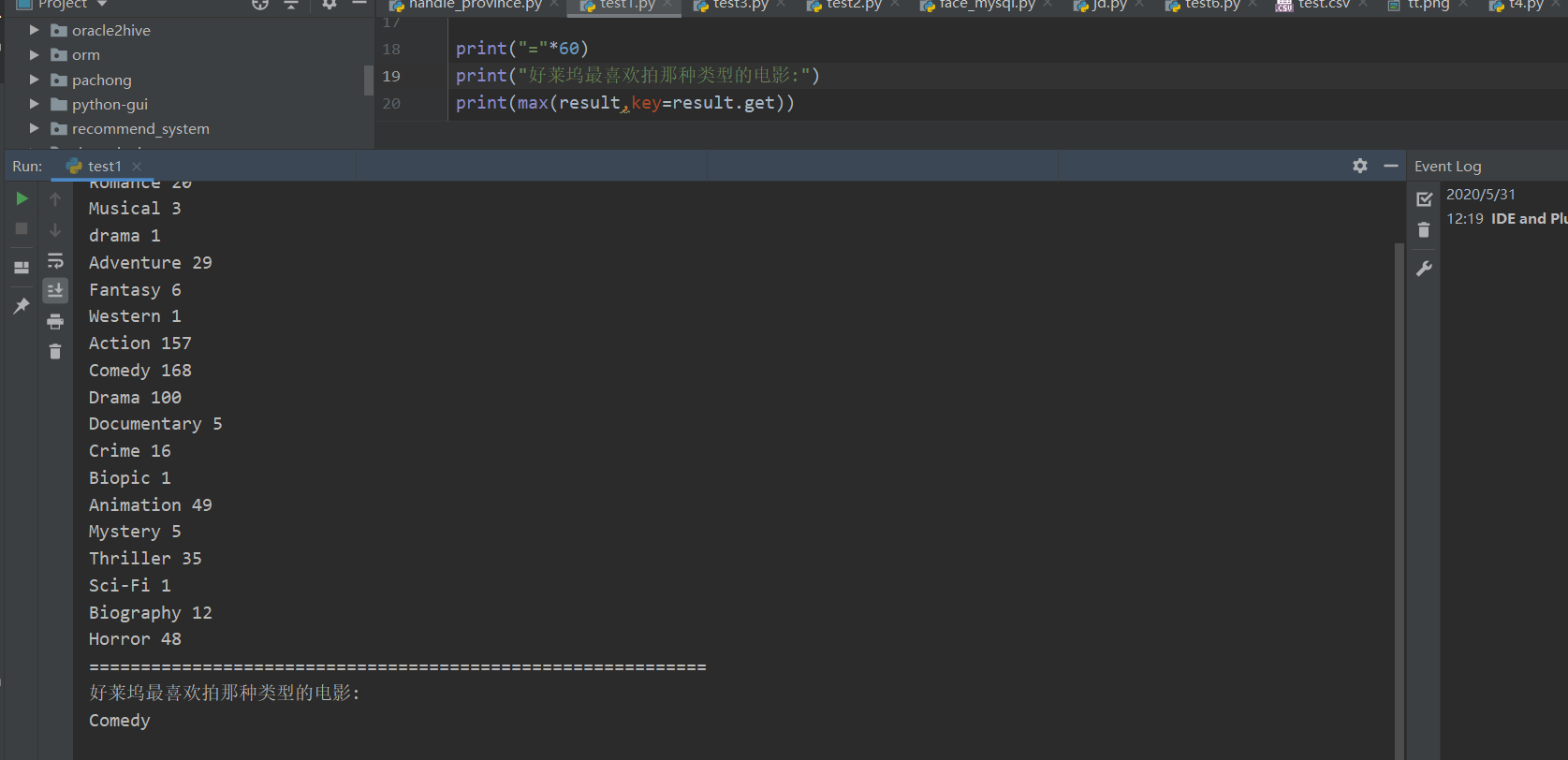
3.导入数据



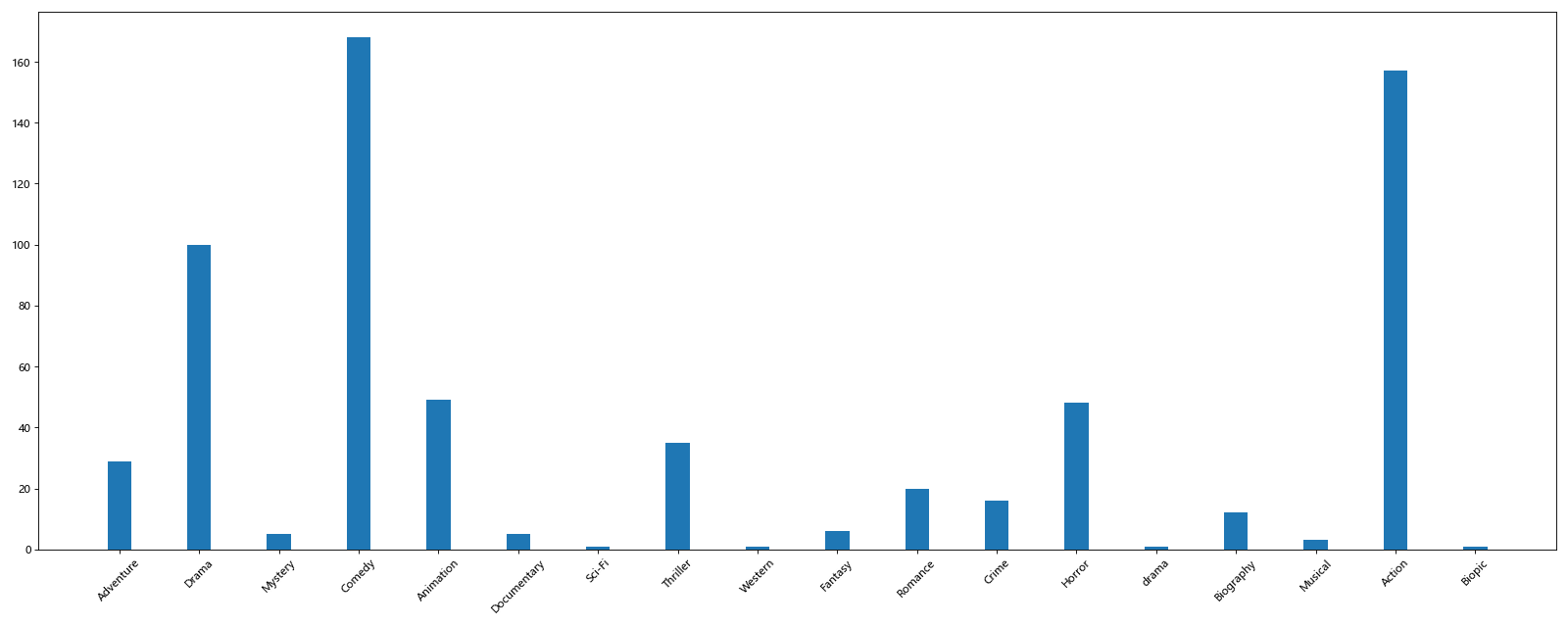


（2）具体实现可视化内容（要求：实现代码并形成可视化效果图）

1. 根据数据分析好莱坞最喜欢拍那种类型的电影？如：可采用组成成分可视化方法完成影片类型分析（10分）

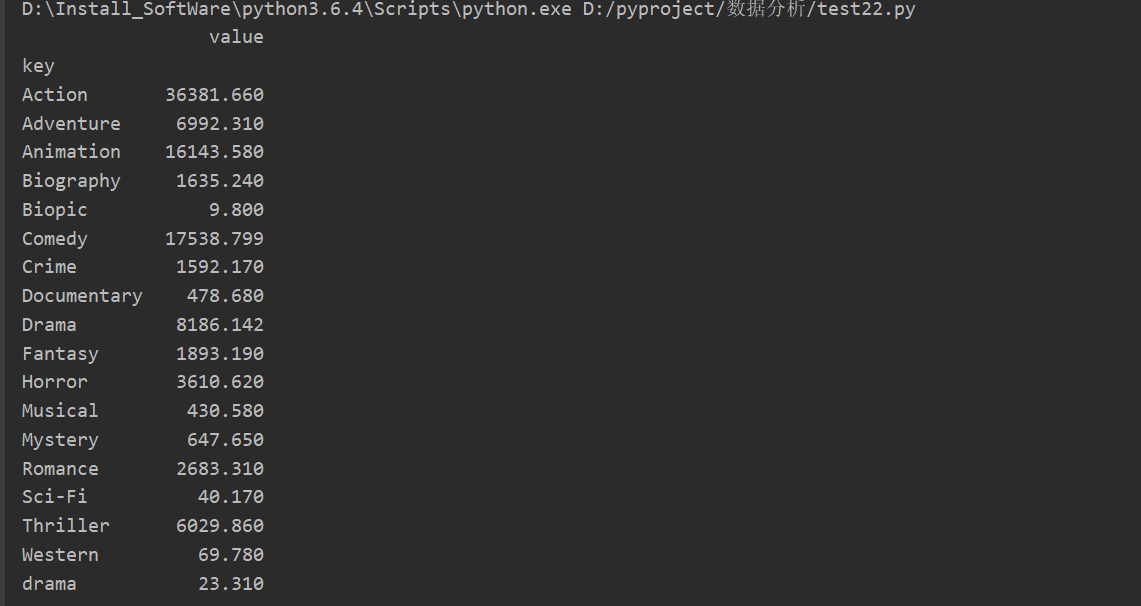


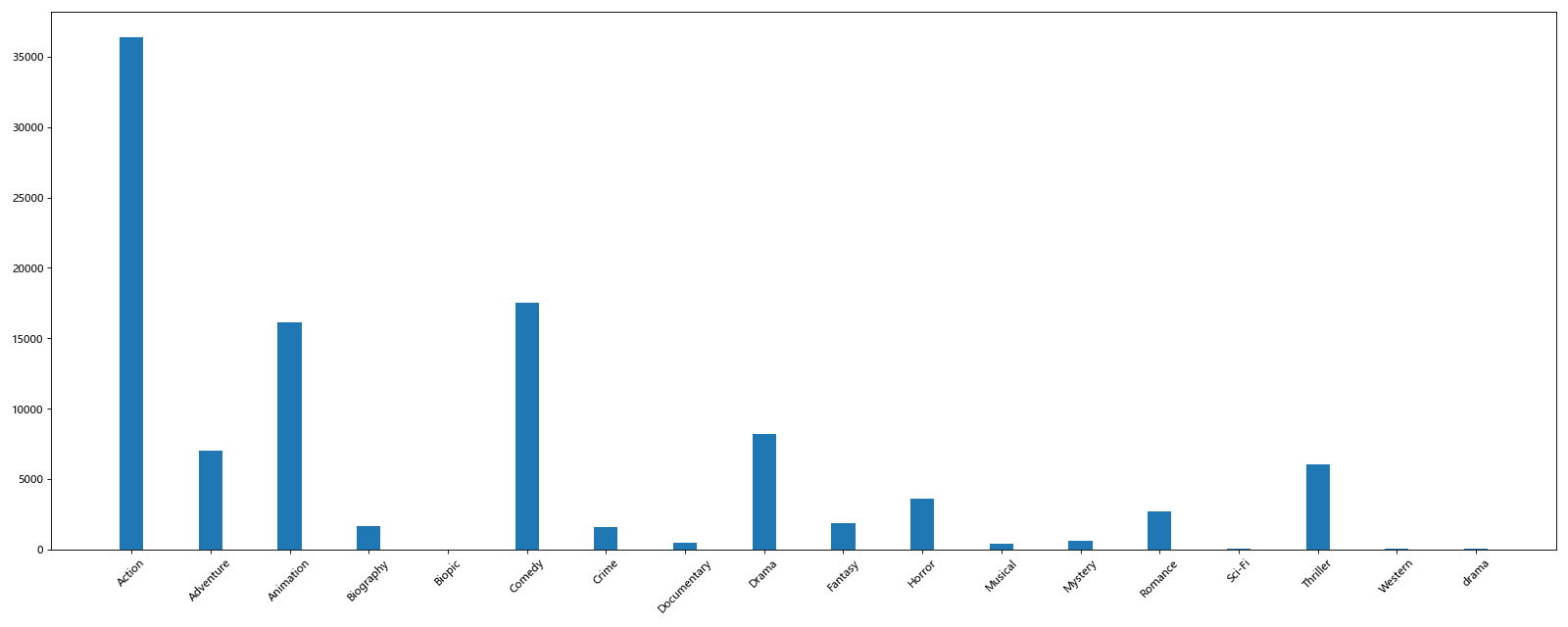
最喜欢Comedy 有168次



1. 根据数据可视化分析好莱坞哪种类型的电影票房最高？如：可采用排序图可视化展示电影票房最高的影片类型（10分）







Action全球票房最高

1. 根据数据可视化分析哪个影片发行公司的平均收益率最高？如可采用排序可视化方法展示影片发行公司的平均收益率（10分）

20th Century Fox 3425.42

Aardman Animations 197.97

Buena Vista 679.04

CBS 703.61

CBS Films 169.61

Columbia 839.58

Crest 370.19

Disney 14496.27

DreamWorks 796.87

DreamWorks Animation 442.02

DreamWorks Pictures 1646.26

Focus 932.30

Fox 22858.78

Happy Madison 426.82

Happy Madison Productions 212.88

Highlight Communications 493.70

Independent 1367519.39

Independent\* 175.23

Legendary Pictures 836.33

Liberty Starz 398.69

Lionsgate 6399.92

MGM 1374.05

Mediaplex 99.00

Miramax 397.33

Miramax Films 126.18

Morgan Creek Productions 77.00

New Line 359.98

New Line Cinema 394.72

Overture 130.75

Paramount 23887.74

Pixar 275.93

Regency Enterprises 318.48

Relativity 646.87

Relativity 229.00

Relativity Media 4994.08

Reliance Entertainment 54.22

Sony 24142.56

Sony Classics 347.87

Sony Pictures Animation 511.05

Spyglass Entertainment 948.92

Summit 6731.55

Summit Entertainment 415.57

The Weinstein Company 3449.64

Universal 14082.67

Vertigo Entertainment 234.48

Village Roadshow Pictures 78.89

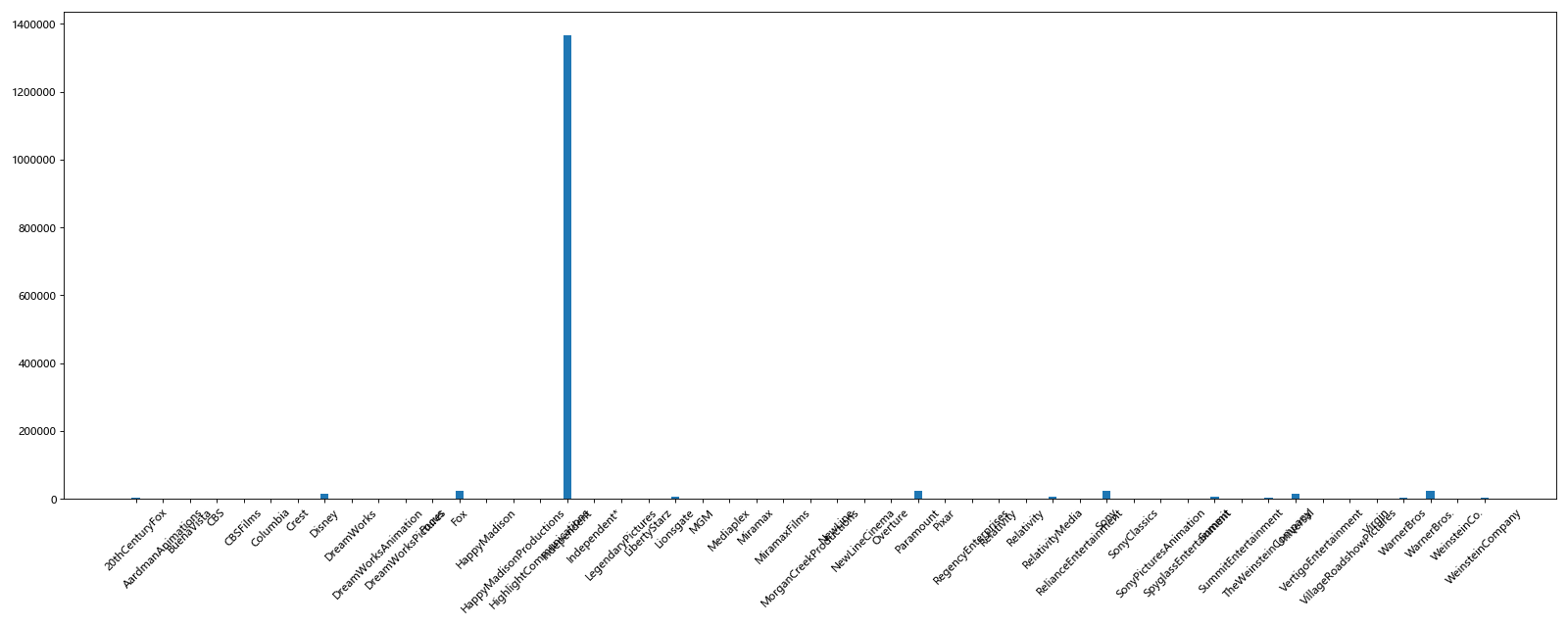
Virgin 599.44

Warner Bros 4153.86

Warner Bros. 23439.72

Weinstein Co. 1438.68

Weinstein Company 2347.56



Independent 最高

1. 根据数据可视化分析预算对票房的影响？（10分）
2. 根据数据可视化分析观众评分跟票房的关系？（5分）
3. 根据数据可视化分析烂番茄评分跟票房的关系？（5分）

7、根据数据可视化分析好莱坞影业预算、票房逐年发展趋势？（10分）

**四、结果分析及总结（30分）**

（1）根据上述可视化分析，阐述通过好莱坞影片数据发现的数据内涵？（15分）

1．通过通过好莱坞影片数据发了不同因素之间相互制约的关系

2. 通过数据可视化能够非常直观的体现出数据的内在规律，将纷繁芜杂的数据通过各式各样的图像来展示，生动形象的展示了事物的内在规律及价值

（2）浅谈个人对数据可视化设计认识，不少于300字。（15分）

数据可视化就是用图表的形式展示数据的信息；让展示的东西有视觉冲突，让人能出图表中一目了然看出你想要表达的信息；这里很关键一点就是，你想要展示数据的什么信息，很多人为了盲目的追求可视化的一些酷炫效果，得到的东西却不知道表达的是什么，那么这个可视化是失败的；当然可视化也要简洁，简单明了才是可视化的目标所在，毕竟做数据可视化是帮助人们去理解数据，认识数据，从多个侧面去观察数据，并告知告知别人你想表达的数据中的信息。

数据可视化就是了简单明了的了解数据，不过你用一个全新的角度去诠释你的数据，让读者读懂了你的信息又能从更高的角度去解读这些信息；第二点信息要充实，充实不代表越多越好，因为一个图表展示的信息过多，会让读者成为一种累赘，很难读懂你的想法，充实的信息就表明你的图表能够能到他们想要的信息又不造成信息过多的情况，所以我们就要考虑业务应用场景了，什么样的场景需要用那些信息结合那些是多余的需要去判断，第三点高效，也就是我们所说的简单明了，一眼就能知道数据中的信息，所以我们在可视化过程中一定要把一些也主题不相关的因素排除掉，过多不相关的因素存在只会增加读者的阅读时间和难度；第四点美感，图片的布局，色彩，形状等一系列的组成部分都是相当和谐的