

项目：可视化电影数据

第一步：清理数据和选择变量

我利用 Python 编程进行数据的梳理，源码为文件夹中的 test.py。

我选择以下 9 个变量进行数据分析：

- 1.genres: 风格列表
- 2.production_companies: 制作公司列表
- 3.release_year: 发行年份
- 4.revenue: 收入（美元）
- 5.vote_count: 评分次数
- 6.director: 导演列表
- 7.runtime: 电影时长
- 8.budget: 预算（美元）
- 9.popularity: 在 Movie Database 上的相对页面查看次数

第二步：Tableau 可视化

Tableau 链接：

Q1:

https://public.tableau.com/profile/vayne#!/vizhome/Q1_170/Q1?publish=yes

Q2:

https://public.tableau.com/profile/vayne#!/vizhome/Q2_159/Q2?publish=yes

Q3:

https://public.tableau.com/profile/vayne#!/vizhome/Q3_153/Q3?publish=yes

Q4:

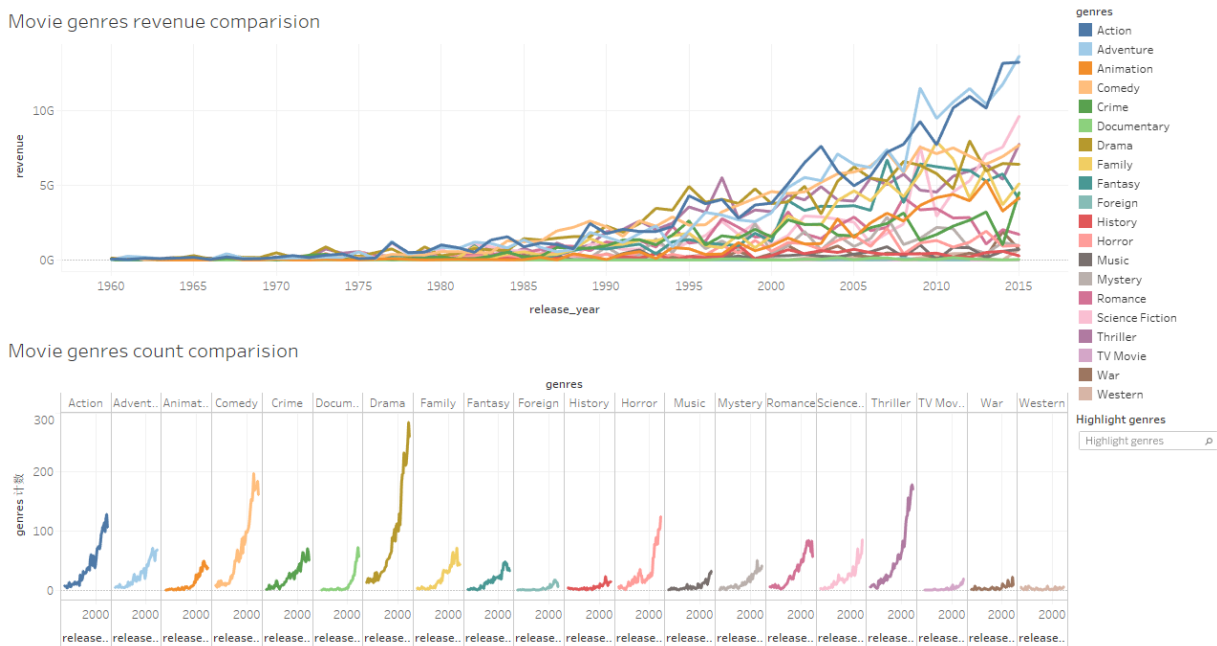
https://public.tableau.com/profile/vayne#!/vizhome/Q4_128/Q4?publish=yes

Story1:

https://public.tableau.com/profile/vayne#!/vizhome/Story1_173/Story1?publish=yes

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？

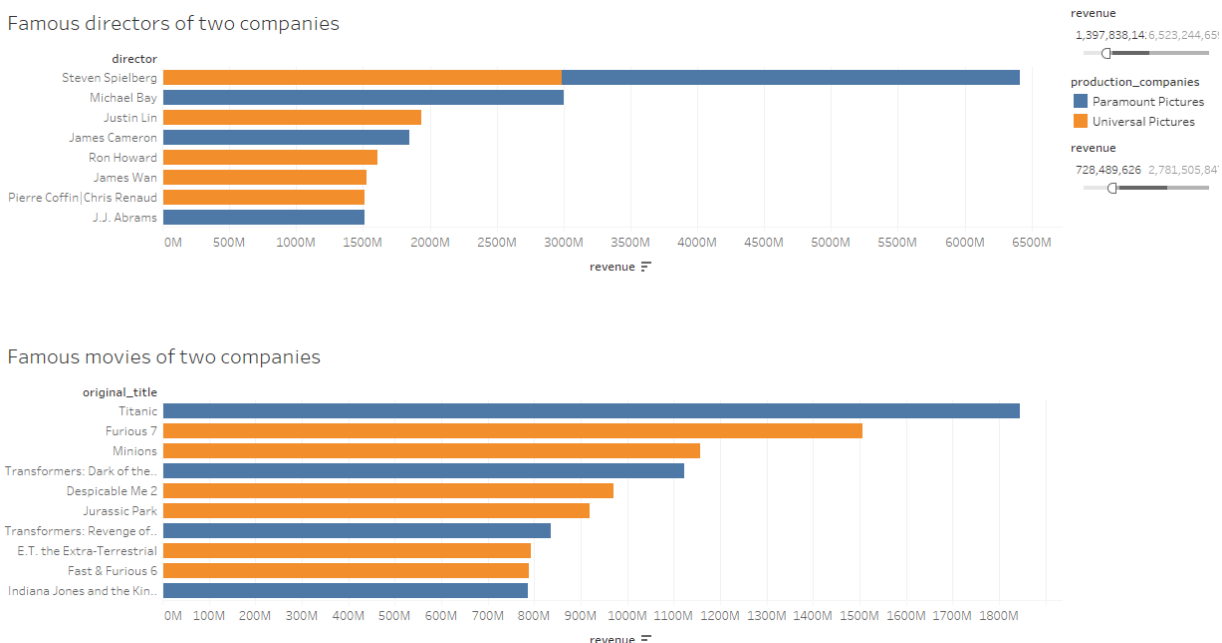


图一

解答 1:

我从两个方面来观察，第一个是随着时间变化，将每年不同电影类型的收入进行比较；第二个是随着时间变化，将每年不同电影类型的数量进行比较。从上半部分可以看出 **Adventure** 类型有利于提升电影票房，从下半部分可以看出，**Drama** 类型的电影数量越来越多。

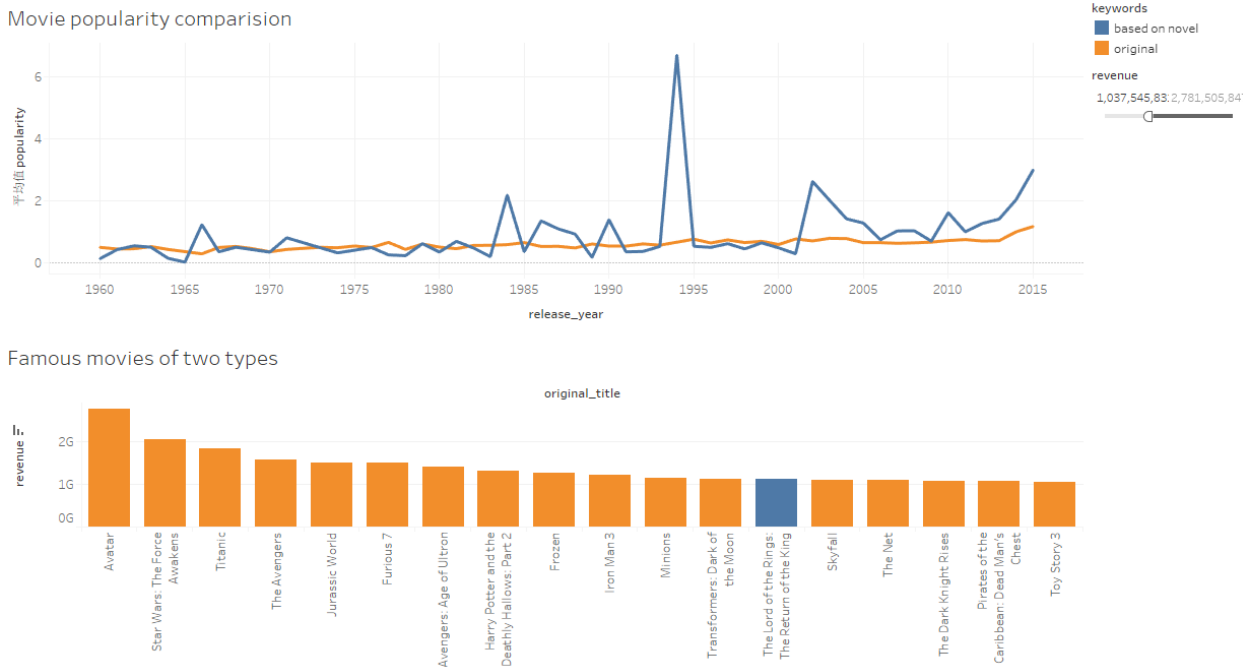
问题 2: 环球影业和派拉蒙影业的电影之间数据指标有什么区别？



图二

我从两个方面来观察，第一个观察两家公司对应的著名导演；第二个观察两家公司的著名电影作品。从上半部分可以看出 **Steven Spielberg** 同时是环球影业和派拉蒙影业当红导演，**Michael Bay** 是派拉蒙影业专属当红导演，**Justin Lin** 是环球影业专属当红导演。从下半部分可以看出，环球影业的代表作为 **Furious 7**，派拉蒙影业代表作为 **Titanic**。

问题 3： 和非小说改编的电影(原创)相比，基于小说改编的电影(改编)表现得怎么样？

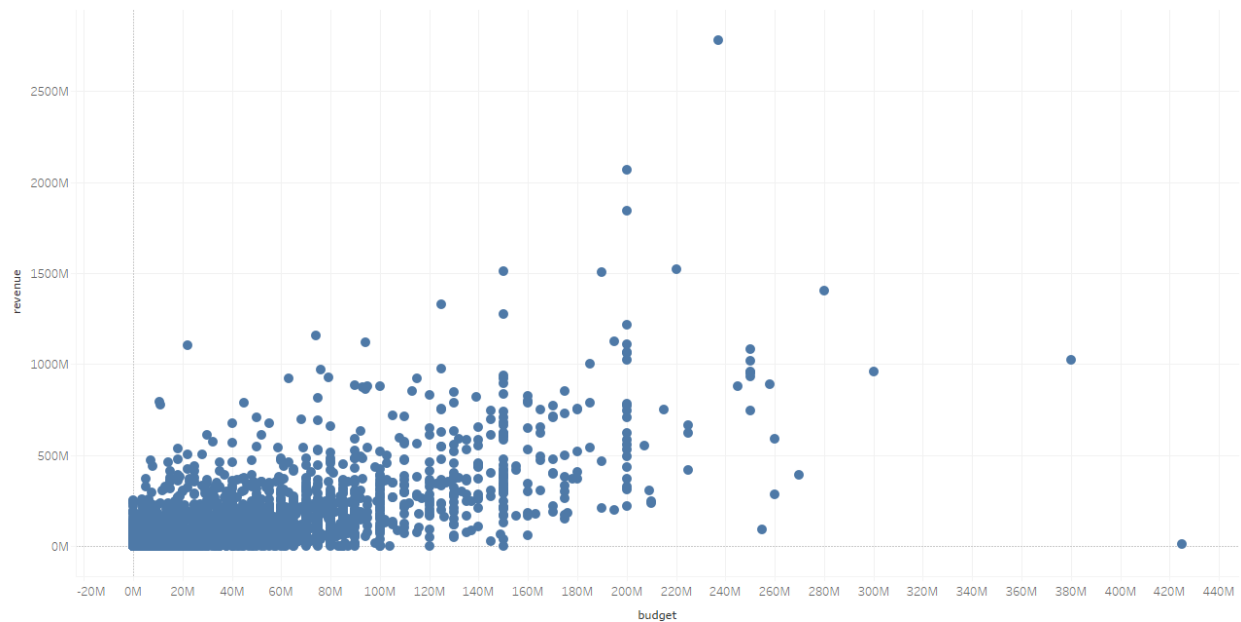


图三

我从两个方面来观察，第一个是比较这改编电影与原创电影的指标随时间的变化趋势；第二个是将改编电影与原创电影的代表作品。从上半部分可以看出改编电影票房通常比原创电影更受欢迎，从下半部分可以看出，改编电影的代表作品为 **The Lord of the Rings: The Return of the King**，原创电影的代表作品为 **Avatar**。

问题 4： 预算和收入的关系是怎么样的？

The relationship between budget and revenue

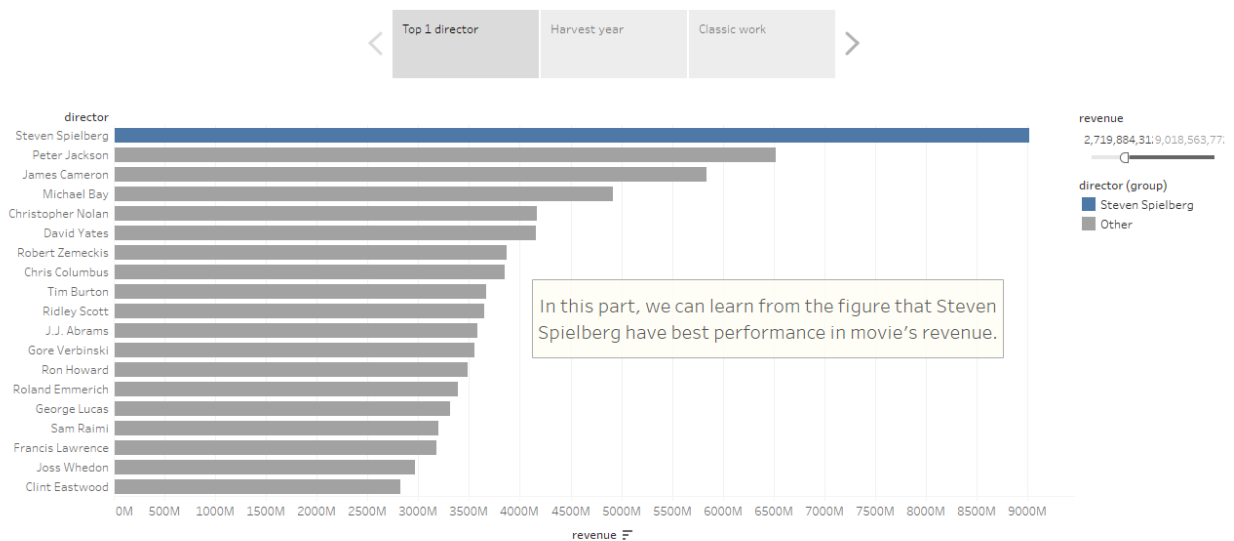


图四

我了解预算和收入的关系，因此制作了对应的散点图。从图四中可以看出，通常来说预算越高，对应的收入也越高，但是仍然会存在一些血本无归或远超预期的情况。

故事 1：故事 1 是对导演情况深入探索。我了解票房最好的导演所制作电影情况。首先是找出票房最高的导演，再了解哪一年是他的丰收之年，最后是了解哪一部是他生涯票房最好的电影。

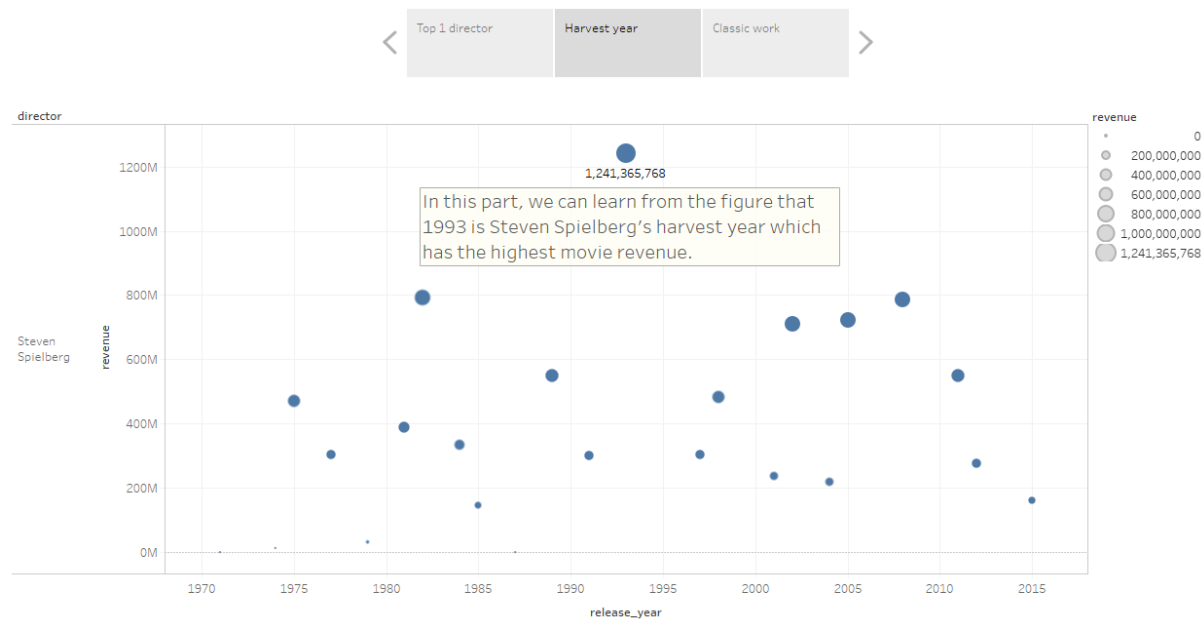
Story 1



图五

从图五中可以看出， Steven Spielberg 是票房总收入最高的导演。

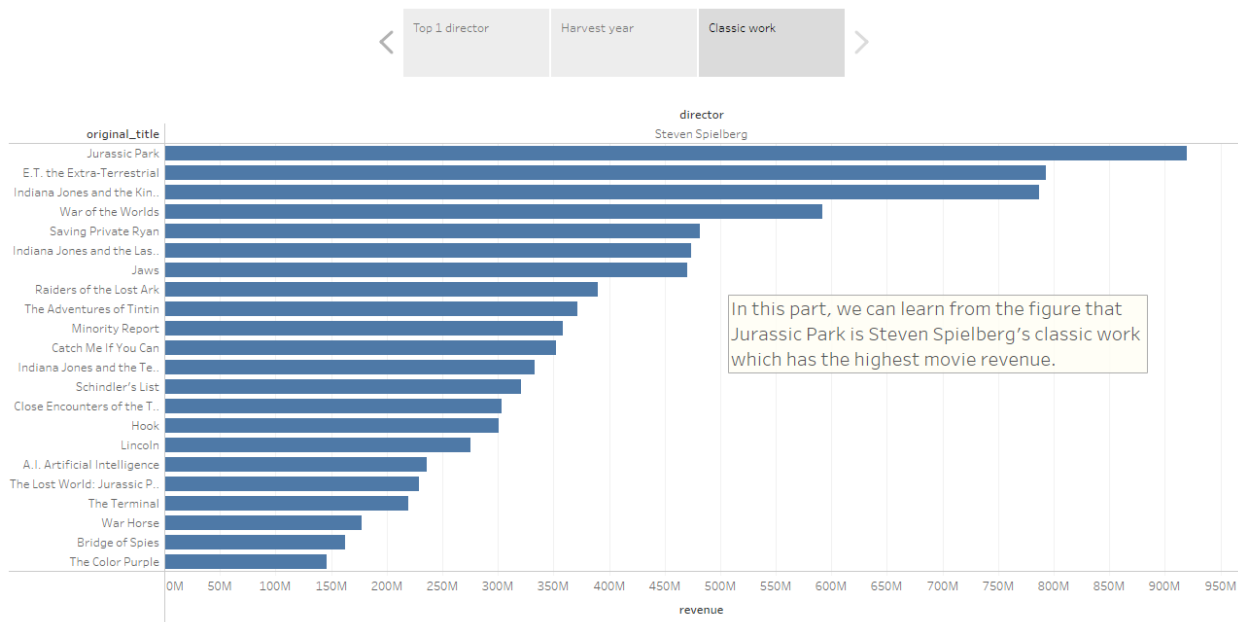
Story 1



图六

从图六中可以看出，1993 年是 Steven Spielberg 的奇迹年，这一年他的电影票房是他人生中最高的。

Story 1



图七

从图七中可以看出，Jurassic Park 是 Steven Spielberg 的扛鼎之作，是票房总收入最高电影。