项目：可视化电影数据

你需要完成每一个部分，当你准备好了，将文件另存为 PDF 文档格式，在教室里提交 。

第一步：清理数据和选择变量

清理缺失的信息，选择在可视化中需要进一步研究的最重要的变量。

列出你在可视化中会进一步研究的变量。你要研究的变量数量应该不超过 8 个。

可以使用 Python 中的 Pandas 库来清洗数据。

本项目使用 python 清洗缺失数据，具体清洗过程见“moves_data_clean.ipynb”文件。

主要研究的变量如下：

1. popularity
2. budget
3. revenue
4. keywords
5. genres
6. release_year
7. production_company
8. original_title

第二步：Tableau 可视化

请确保你遵循了这些提示，并在你的项目中包含了 Tableau 的仪表板（Dashboard）、故事（Stories）模块，以及你的项目评审师会特别检查的可视化类型（小图、散点图、条形图等）。需要特别注意的是，针对每一个问题（Q1-Q4），你都需要创建一个仪表板。另外，你也需要为一个你所选择的问题创建故事模块。

重要提示: 请把你的工作簿上传到 **Tableau Public** 上，并允许你的项目评审师查看你的工作簿。请注意单单把文件简单保存为后缀名为 ".twbx" 不能让所有项目评审师查看文件。如何做到能够允许所有项目评审师查看文件的指示在这里。

第三步：问题

回答下列问题，引用你在线可视化结果去支持你的答案：

问题 1：电影类型是如何随着时代变化而变化的?

在二十世纪 80 年代以前，由于电影数量相对较少，电影类型也相对单一，在这个时期主要以 Drama 为主。

二十世纪 80 年代，Comedy 异军突起，数量迅速增多超过 Drama，并与 80 年代中后期成为发行量第一的电影类型。

二十世纪 90 年代，Thriller、Action 以及 Romance 发行数量逐渐增多，日渐成为电影市场中不可取少的类型。

进入二十一世纪以后，Drama、Comedy、Thriller、Action 都有较快发展，分列电影类型前四名。值得注意的是，Horror 自 2004 年以后飞速发展，截至 2015 年末，以由 2004 年的第 10 名发展为第 4 名，超过 Action 并仅次于 Comedy。

其他类型的电影由于相对小众，发展情况一直不温不火，较为稳定，只是随着电影发行总数的增多而增多。

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?

总体来看，环球影业的电影发行数量高于派拉蒙影业。但从单部影片的平均收益来看，派拉蒙影业要稍高于环球影业。进一步看，派拉蒙影业单部影片的平均预算及平均关注度均要高于环球影业。

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?

从样本整体情况来看，基于小说改编的电影无论是从关注度方面、收益方面、预算收益比（来自预算与收益趋势线的比较）方面，似乎均要优于非小说改编的电影。但从历年的情况来看，选取了几个存在极端情况的年份进行细致分析发现，出现这种表面上非小说改编电影完败的原因是我们采取了平均值方法进行比较，由于非小说改编电影的数量巨大，从而存

在大量低质量的片子拉低了整体水平；反观基于小说改编的电影，一般来说能够用于改编电影的小说均具有一定的名气，一旦上映会天然吸引一定的关注度和票房，且该类型电影数量相对少很多，从而无法有效消除极值影响。

然后，当我使用历年最赚钱、关注度最高的电影形成时间序列进行分析时发现，除个别年份，绝大多数年份当年最赚钱、关注度最高的电影均为非小说改编电影。

综上，和非小说改编的电影相比，基于小说改编的电影预算收益比较高，票房有一定的保证，但是不容易出现收益和关注度极高的电影。

问题 4：电影评分、电影收益与关注度、预算之间的关系。

一般来说，关注度越高，电影评分和电影收益越高。预算越高，电影评分和电影收益越高。在自己平时选择观看哪些电影时，多是先看哪些电影的讨论比较多（关注度），然后看评分高低，从而抉择到底选择观看哪部电影。再个就是如果电影的预算投入比较大的，多数制作较为精良，也能够吸引到很多眼球。

Tableau Public 链接：https://public.tableau.com/profile/.37707676 - !/vizhome/P3_23/Q1