项目：可视化电影数据

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。

将电影风格（genres）、电影制作公司（production_company）、关键字（keywords）进行了清理。将电影风格和电影制作公司数据值都基于"|"拆分扩展成多个数据再融为一列后去除这两项值为空的数据。将关键字列数据根据其中是否包含"based on novel"来判断其是否属于基于小说改编的电影并将判断结果保存为一列新的数据。

- 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。

我希望研究发行总量、收入总额、预算总额、平均评分均值、受欢迎程度均值这 5 个变量。

- 可以使用 Python 中的 Pandas 库来清洗数据。

第二步：Tableau 可视化

我的 Tableau 可视化：https://public.tableau.com/views/Udacity-DA-P3/2?:embed=y&:display_count=yes

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?

电影类型总共有 20 种。从 1960 年到 1985 年左右，各类电影每年的发行总量与总收入变化不大，从 1985 年到 2015 年，各类电影中，Action、Comedy、Drama、Thriller 这四类电影发行量呈大幅增长，Adventure、Crime、Horror、Romance、Science Fiction 这五类电影发行量呈中等幅度增长，其他类型电影发行量呈小幅增长或者没有增长。在收入方面，从 1985 年至 2015 年，Action、Adventure、Comedy、Drama、Science Fiction、Thriller 这几类电影增幅较大，其他类型增幅一般，甚至有的类型的电影收入几乎没有增长。

    - **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?

环球影业的预算、收入比派拉蒙略高，派拉蒙影业电影的平均受欢迎程度略高。环球影业发行的电影总量较派拉蒙影业多，两者的电影平均评分相近，所有电影平均时长相近。

    - **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 1960 年至 2015 年，基于小说改编电影发行总量变化不大、平均受欢迎程度大幅波动小幅上升、平均评分保持稳定震荡、总收入和预算从 1985 年开始大幅波动中等幅度上升，非基于小

说改编电影发行总量大幅上涨、平均受欢迎程度稳定小幅上升、平均评分稳定小幅下降、总收入和预算稳定小幅上升。

- 你提出的另外问题是什么？答案是什么？你是怎么想出这个问题的？

我提出的问题是：我喜欢的几部电影的表现怎么样？有几部电影我非常喜欢我想知道他们的收入、预算、受欢迎程度还有平均评分情况。我喜欢的六部电影中，Interstellar 预算最高、最受欢迎、平均评分最高，而 Inception 收入最高。其他四部电影相较这两部电影来说只能算表现一般。