

项目：可视化电影数据

一：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。

在 Excel 表格中将 genres 里的数据拆分成不同的列。删除 homepage, tagline, overview 等不必要的信息。

- 列出你在可视化中会进一步研究的变量。

1, 为了知道电影类型是如何随着时间的推移发生变化的, 我选取研究的变量为: release_year、genres、popularity 和 vote_average。

2, 为了对比 Universal Pictures 和 Paramount Pictures 之间的情况, 我选取研究的变量为: production_companies、release_year、profit (profit = revenue - budget)、popularity 和 vote_average。

3, 为了对比改编电影和原创电影的情况, 我选取的研究变量为: keywords、release_year、profit (profit = revenue - budget)、budget、popularity 和 vote_average。

4, 对比 Steven Spielberg、James Cameron、Michael Bay 三大导演, 选取的变量为: director、profit (profit = revenue - budget)、budget、popularity 和 vote_average。

二：Tableau 可视化和问题回答

- **问题一：**电影类型是如何随着时间推移发生变化的？

https://public.tableau.com/views/p3_9/Q1?:embed=y&:display_count=yes

各个类型的电影发行量随时间的推移基本都呈上涨趋势, 近年来电影发行量比较多的几大类型为 Drama、Comedy、Thriller、Action、Horror; 2010 年以来最受欢迎的六大电影类型为 Drama、Action、Thriller、Adventure、Comedy、Science Fiction; 投票得分最高的六大电影类型为 Drama、Comedy、Thriller、Action、Horror。

- **问题二：**环球影业和派拉蒙影业的电影之间数据指标有什么区别？

https://public.tableau.com/views/p3_9/Q2?:embed=y&:display_count=yes

https://public.tableau.com/views/p3_9/Q2Story?:embed=y&:display_count=yes&publish=yes

派拉蒙影业的电影发行量和总的利润均高于环球影业, 派拉蒙影业取得的利润高峰期基本处于 1985~1995 年间, 派拉蒙影业的电影投票得分基本稍高于环球影业。派拉蒙影业发行量较多的几大电影类型基本符合电影类型的流行趋势。

- **问题三：**和小说改编的电影相比, 基于小说改编的电影表现得怎么样？

https://public.tableau.com/views/p3_9/Q3?:embed=y&:display_count=yes

2010 年以前改编电影每年的发行量基本上没有太大的改变，2010 年以后改编电影的发行量略有增加，与原创电影相比增幅不明显。改编电影的平均预算稍高于原创电影，但是改编电影的平均利润显著高于原创电影。近年来改编电影的受欢迎程度和电影的投票得分也都高于原创电影。

- **问题四：**对比 Steven Spielberg、James Cameron、Michael Bay 三大导演的电影表现如何？

https://public.tableau.com/views/p3_9/Q4?:embed=y&:display_count=yes

选取总利润排名前三的三位电影导演：Steven Spielberg 导演的电影发行量、总预算和总利润均是最高；James Cameron 导演的电影受欢迎程度和电影投票的平均得分均是最高；James Cameron 导演电影发行量和平均利润均较高的电影类型为：Thriller、Action 和 Science Fiction，Steven Spielberg 导演电影发行量和平均利润均较高的电影类型为：Action、Adventure 和 Science Fiction，Michael Bay 导演电影发行量和平均利润均较高的电影类型为：Adventure、Action 和 Science Fiction。