

基于某国际影视传媒投资制作公司的票房数据分析及投资项目选择

项目描述：该项目通过构建数据仓库与可视化分析体系，辅助企业科学决策影视投资项目。

项目职责：数据分析、可视化、建模

工具：SQL, Python, Kettle, SQL Server, Power BI

时间：2022.12-2023.01

一、项目背景与研究目标

1. 背景：电影市场高速发展，投资决策风险高；依赖数据分析提升决策科学性。

2. 目标：

- 分析历史票房数据，识别影片要素（类型、演职员、发行区域等）对表现的影响；
- 基于多维数据建模提出优化策略，提升投资回报率；
- 为影视公司提供数据驱动的定量分析参考。

二、数据来源与处理流程

1. 数据来源：

- 阿里云开放数据平台：影片基础信息
- Kaggle 电影数据集：3400 部影片日票房数据

2. ETL 流程：

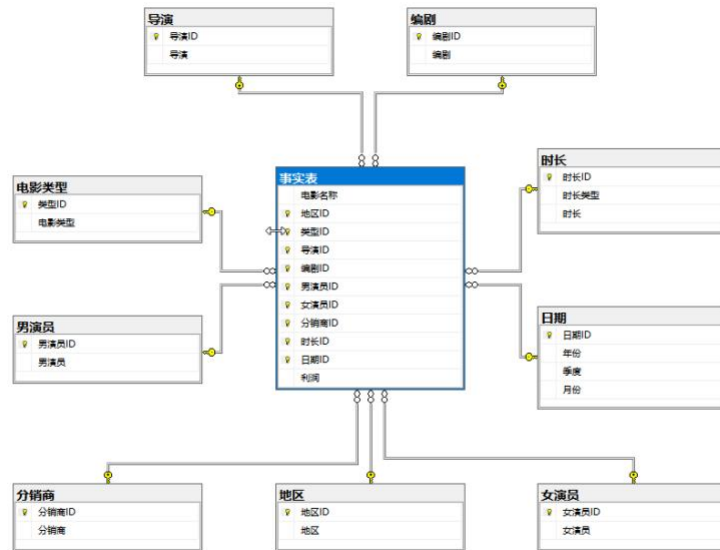
- （1）抽取：SQL 导出数据；
- （2）清洗：Python 标准化影片标题等字段；
- （3）加工：聚合为月维度数据，处理缺失值；
- （4）加载：使用 Kettle 创建事实表，加载进 SQL Server。

P1	电影名称	地区	时长	电影类型	编剧	导演	男演员	女演员	分销商	成本	放映年份	放映季度	放映月份	利润
1	9334 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	2	4	0
2	9335 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	2	5	0
3	9336 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	2	6	0
4	9337 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	3	7	0
5	9338 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	3	8	0
6	9339 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	3	9	0
7	9340 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	4	10	11083
8	9341 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	4	11	33293
9	9342 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2007	4	12	0
10	9343 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	1	1	0
11	9344 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	1	2	0
12	9345 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	1	3	0
13	9346 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	2	4	0
14	9347 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	2	5	0
15	9348 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	2	6	0
16	9349 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	3	7	0
17	9350 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	3	8	0
18	9351 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	3	9	0
19	9352 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	4	10	0
20	9353 The Nightmare Before Christmas	印度尼西亚	76	爱情	斯坦利·库布里克	詹姆斯·卡梅隆	周星驰	Kirsten Dunst	Manhattan Pictures	100000000	2008	4	11	0

数据库中的原始数据



Kettle 流程图



数据库雪花模型

- (5) 数仓建模：构建数据库的雪花模型，并部署数据集市。
- (6) 分析：使用 powerBI 对数据进行可视化，并提供投资建议。

该项目使用 **powerBI** 设计指标面板，支持多维动态筛选与趋势分析。重点聚焦于不同影片维度与票房波动的关系，为后续策略建议提供数据支持。

电影类型

票房表现(按
时长/时长类型)

所有

时长/时长类型

所有

票房表现(按 日期/年份 和 电影类型/电影类型)

电影类型: 爱情 ● 动作 ● 犯罪 ● 惊悚 ● 卡通 ● 科幻 ● 恐怖 ● 魔幻 ● 喜剧 ● 战争

时长: ...
● 一般...
● 特殊...
● 长片
● 超长...
● 短片
● Unk...

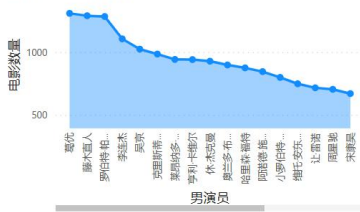
日期/年份

编剧/编剧	导演/导演	地区/大洲	地区/地区	电影类型/电影类型	分销商/分销商	男演员/男演员	女演员/女演员	日期_季度	日期_年份	日期_月份	时长/时长类型	票房表现
Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	0.00
坂元裕二	阿米尔汗	非洲	埃及	恐怖	Capelight Pictures	亨利·卡维尔	Scarlett Johansson	3	2005	7	一般时长	24,319,260.00
坂元裕二	阿米尔汗	非洲	埃及	恐怖	Capelight Pictures	亨利·卡维尔	Scarlett Johansson	3	2005	8	一般时长	11,371,589.00
坂元裕二	阿米尔汗	非洲	埃及	恐怖	Capelight Pictures	亨利·卡维尔	Scarlett Johansson	3	2005	9	一般时长	128,064.00
坂元裕二	阿米尔汗	欧洲	俄罗斯	爱情	Cohen Media Group	克里斯蒂安·贝尔	Anna Popplewell	4	2019	10	一般时长	1,011,305.00
坂元裕二	阿米尔汗	欧洲	芬兰	科幻	OTL Releasing	让·雷诺	Keira Knightley	2	2001	4	一般时长	13,194,345.00
坂元裕二	阿米尔汗	欧洲	芬兰	科幻	OTL Releasing	让·雷诺	Keira Knightley	2	2001	5	一般时长	1,076,670.00
坂元裕二	阿米尔汗	欧洲	芬兰	恐怖	Mirror/LD Entertainment	刘德华	Zoe Saldana	1	2020	1	一般时长	502,205.00
坂元裕二	阿米尔汗	欧洲	芬兰	恐怖	Mirror/LD Entertainment	刘德华	Zoe Saldana	4	2019	11	一般时长	810,079.00
坂元裕二	阿米尔汗	欧洲	芬兰	恐怖	Mirror/LD Entertainment	刘德华	Zoe Saldana	4	2019	12	一般时长	9,823,800.00
坂元裕二	阿米尔汗	欧洲	孟加拉国	卡通	Codeblack Entertainment	刘德华	Carey Mulligan	1	2018	3	一般时长	5,167,410.00
坂元裕二	阿米尔汗	欧洲	孟加拉国	卡通	Codeblack Entertainment	刘德华	Carey Mulligan	2	2018	4	一般时长	22,037,389.00
坂元裕二	阿米尔汗	欧洲	孟加拉国	卡通	Codeblack Entertainment	刘德华	Carey Mulligan	2	2018	5	一般时长	4,060,648.00
坂元裕二	阿米尔汗	欧洲	孟加拉国	卡通	Codeblack Entertainment	刘德华	Carey Mulligan	2	2018	6	一般时长	587,483.00
坂元裕二	阿米尔汗	欧洲	孟加拉国	卡通	Codeblack Entertainment	刘德华	Carey Mulligan	3	2018	7	一般时长	39,586.00
总计												200,037,940,262.00

[illegible]

（3）演员票房表现

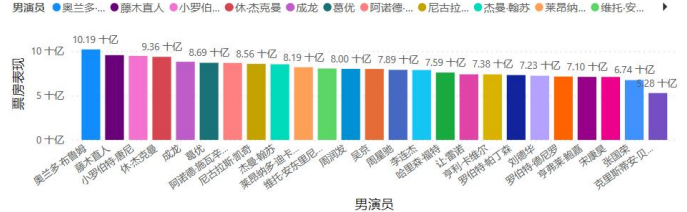
电影数量(按 男演员)



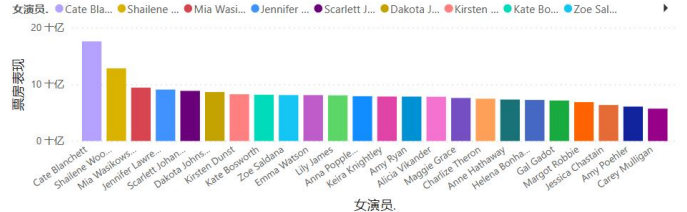
电影数量(按 女演员)



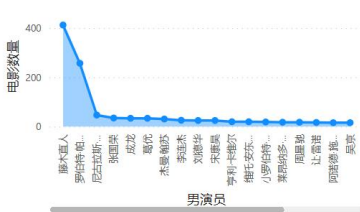
演员的票房表现



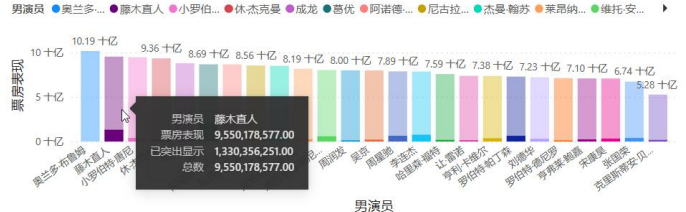
演员的票房表现



电影数量(按 男演员)



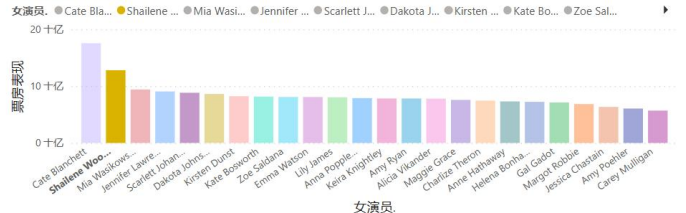
演员的票房表现



电影数量(按 女演员)



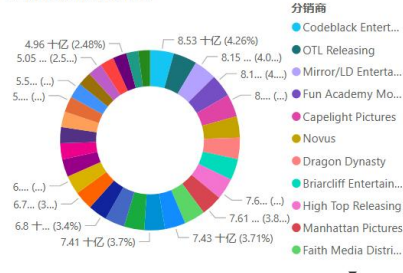
演员的票房表现



(4) 各地区的电影票房情况和最佳分销商

- 大洲、地区
- ☐ 北美洲
 - ☐ 大洋州
 - ☐ 非洲
 - ☐ 南美洲
 - ☐ 欧洲
 - ☐ 亚洲

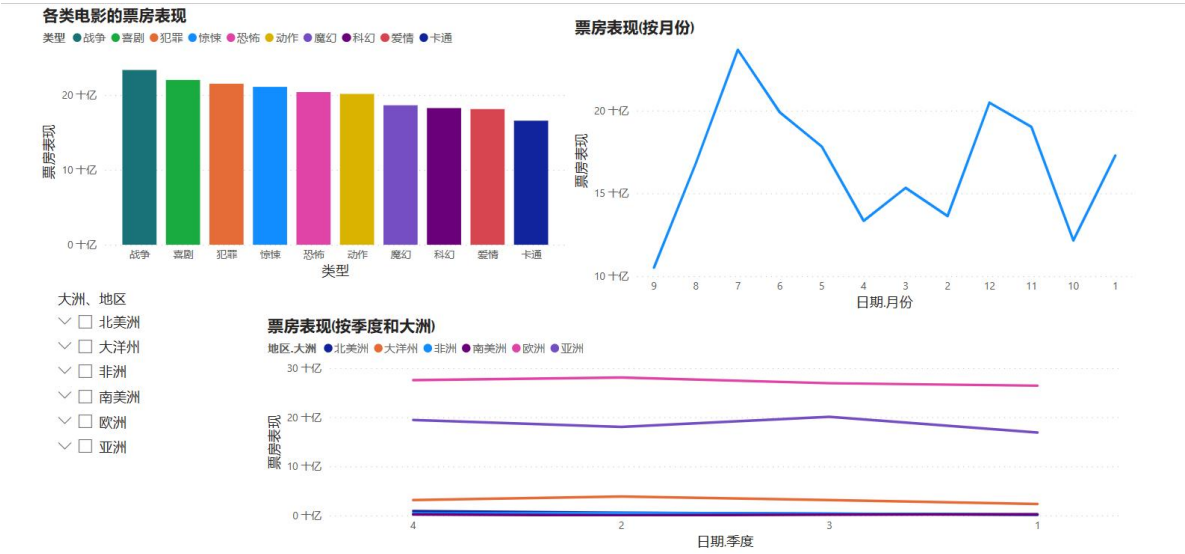
票房表现(按 分销商)



票房表现(按 地区)

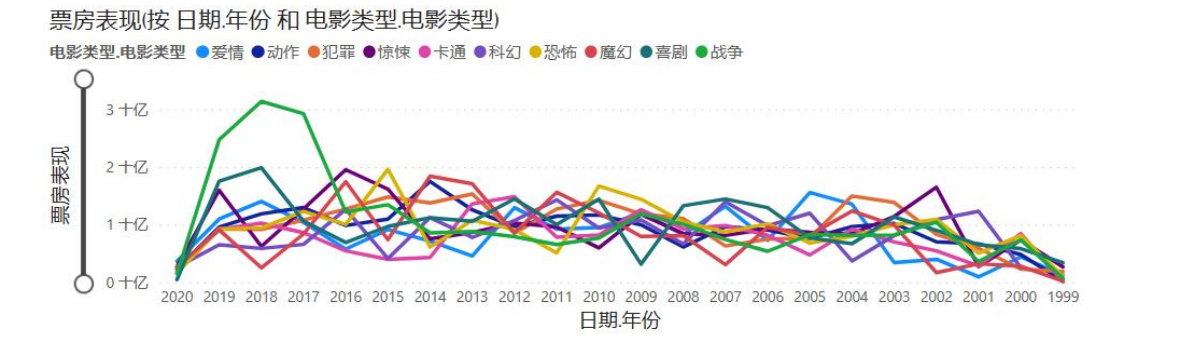


(5) 不同地区的上映时间选择



三、分析成果示意

1. 战争题材影片近年票房涨幅最大，显示出强劲的票房潜力，可作为重点关注的题材方向。



2. 欧洲电影市场比较成熟，分布密集，可能存在比较激烈的竞争；亚洲市场发行渠道，特别是东南部地区票房表现突出，且存在开发空间，建议重点布局。

- 问题设置与指标体系的构建仍有优化空间；
- 可视化面板部分功能尚不完善，少数细节没有优化；
- 未纳入社交热度、评分等外部因素；
- 后续可结合票价、评分等数据进行预测建模。