

中国海洋大学

## 可视化技术期末项目报告

----可视分析项目实战作品

小组编号：第 1 组

成员姓名和专业年级：曾祥辉(21 化学)、邱雨晨(20 工商管理)、梁

汉林(20 网络空间安全)

指导老师：解翠

提交时间：2023 年 06 月 24 日

# 1.项目简介

## ①简明摘要：

该可视化项目旨在为电影投资商提供交互式的数据分析平台，帮助他们更好地了解电影市场趋势和关键因素，从而做出明智的投资决策。该系统的目标用户是电影投资商和相关决策者，包括电影制片公司高层、投资基金经理等。通过可视化展示全球电影市场数据和趋势，系统解决了投资与反馈关系、电影类型影响、演员导演影响以及剪辑时间选择等关键问题。系统提供了全面的市场数据和决策支持，帮助投资商减少风险、选择潜力项目，提高投资成功率。通过系统的应用，投资商可以获得有价值的见解和指导，增强对市场的理解，从而做出更明智的投资决策。

## ②整体界面：

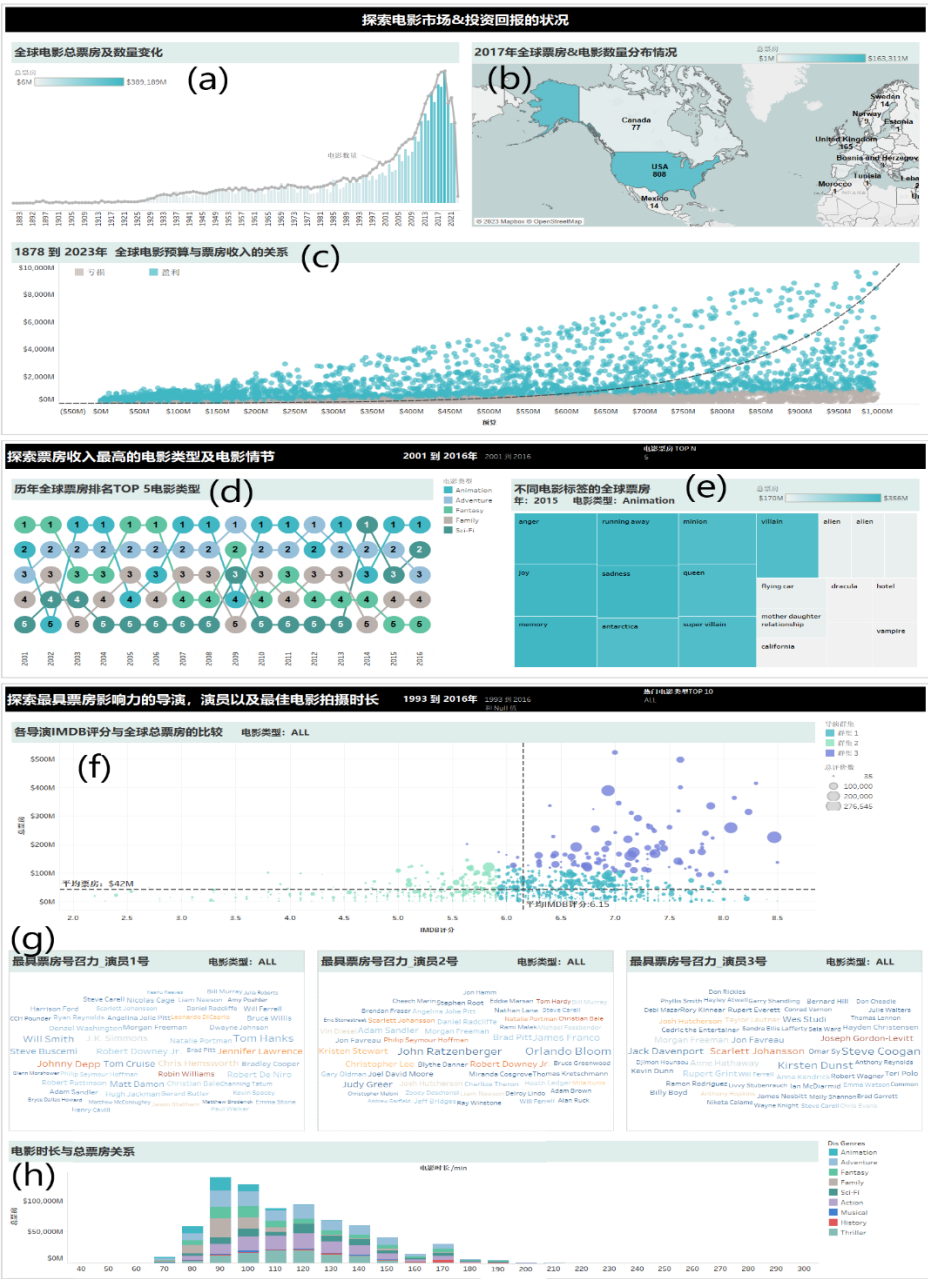


图 1. 电影投资洞察: 赋予电影行业明智的投资决策。(a) 电影的全球总票房及数量变化视图展示了随时间推移的电影票房和电影数量的变化情况。(b) 全球票房与电影数量分布情况视图通过地图展示了电影票房在不同国家的分布情况。颜色深浅表示票房高低, 投资商可以比较各国电影市场的潜力和发展情况。(c) 电影预算与票房收入的关系视图通过散点图展示了电影预算与票房收入之间的关系。数据点代表每部电影, 位置反映了预算和票房情况。趋势线标注了盈利和亏损电影的趋势, 帮助投资商理解预算与票房之间的关联性。(d) 历年全球票房排名 TOP N 电影类型视图以点图和线图展示了不同年份中总票房排名前 N 的电影类型, 并根据选择的电影类型进行高亮显示。这个视图让投资商了解每年最受欢迎的电影类型, 为他们的投资决策提供参考。(e) 不同电影标签的全球票房视图通过热力图展示了不同电影情节的全球票房情况。不同的电影情节在热力图中用颜色深浅和面积大小表示票房的高低。这个视图帮助投资商了解各种电影情节的受欢迎程度和票房表现。(f) 各导演的 IMDB 评分与全球总票房的比较视图展示了各导演的平均 IMDB 评分和平均票房数, 并将导演数据点按名称分布在散点图中。通过参考线划分导演为三个集群, 投资商可以了解哪位导演具有最高的投资价值。(g) 最具票房号召力的演员视图展示了对票房具有较大影响力的演员。演员名称的大小和颜色的深浅反映了他们的票房收入和粉丝数量。这些视图帮助投资商了解演员对电影票房的贡献程度。(h) 电影时长与总票房的关系视图通过直方图展示高票房电影的主要时长范围。电影时长进行分桶, 并将票房作为 Y 轴进行堆叠。

## 2. 数据来源及数据处理

### 一、数据预处理

#### (1) 数据来源:

##### ① 第一部分数据:

##### 1.1 数据介绍:

数据来自 IMDB (Internet Movie Database) 数据库。IMDB 是一个广泛使用的在线电影和电视节目数据库, 提供了关于电影、电视节目、演员、导演等方面的信息。该数据库包含了丰富的电影相关数据, 如电影的基本信息、演职员名单、票房收入、评论和评分等。

在提供的数据中, 每条数据包含了一部电影的多个属性, 这些属性描述了电影的不同方面, 如导演、演员、时长、票房等。这些属性可以用于分析和研究电影产业的趋势、电影类型的受欢迎程度、不同导演或演员的影响力以及电影特征与票房之间的关联等问题。

##### 1.2 数据下载地址:

<https://www.kaggle.com/datasets/carolzhangdc/imdb-5000-movie-dataset>

##### 1.3 属性类型:

数据集包含了多个属性, 如电影的颜色 (Color)、导演名字 (director\_name)、评论家评论数量 (num\_critic\_for\_reviews)、电影时长 (duration)、导演在 Facebook 上的点赞数 (director\_facebook\_likes)、演员 3 在 Facebook 上的点赞数 (actor\_3\_facebook\_likes)、演员 2 的名字 (actor\_2\_name)、演员 1 在 Facebook 上的点赞数 (actor\_1\_facebook\_likes)、票房收入 (gross)、电影类型 (genres)、演员 1 的名字 (actor\_1\_name)、电影标题 (movie\_title)、用户投票数量 (num\_voted\_users)、演职员总的 Facebook 点赞数 (cast\_total\_facebook\_likes)、演员 3 的名字 (actor\_3\_name)、海报中的人脸数量 (facenumber\_in\_poster)、剧情关键词 (plot\_keywords)、电影的 IMDB 链接 (movie\_imdb\_link)、用户评论数量 (num\_user\_for\_reviews)、语言 (language)、国家 (country)、内容评级 (content\_rating)、电影预算 (budget)、电影上映年份 (title\_year)、演员 2 在 Facebook 上的点赞数 (actor\_2\_facebook\_likes)、IMDB 评分 (imdb\_score)、宽高

比 (aspect\_ratio) 和电影在 Facebook 上的点赞数 (movie\_facebook\_likes) 等属性。

#### **1.4 数据集规模:**

数据集的规模/基数是 5043 行, 表示数据集中包含了 5043 部电影的信息。每部电影具有 28 个属性。

#### **1.5 语义特征:**

数据集包含了关于电影的各种信息, 如导演、演员、评论、票房、时长、类型等。这些信息可以用于分析和探索电影行业的趋势、演员和导演的影响力、电影类型的受欢迎程度以及电影时长与票房之间的关系等。

### **②第二部分数据:**

#### **2.1 数据介绍:**

数据是来自 TMDB (The Movie Database) 数据库。TMDB 是一个综合性的电影和电视节目数据库, 提供了有关电影、电视节目、演员和制作公司等方面的信息。数据案例中的每条数据包含了电影的多个属性, 如电影的标题、类型、原始语言、简介、受欢迎程度、制作公司、上映日期、预算、票房收入、时长、状态、标语、平均评分、投票数、演职员、关键词、海报路径、背景图片路径和推荐等信息。通过使用 TMDB 数据库, 可以进行各种电影相关的分析, 例如根据类型、语言或制作公司对电影进行分类和筛选、评估电影的受欢迎程度、分析预算和票房之间的关系、研究电影时长对受欢迎程度的影响、探索不同类型电影的特点、利用推荐数据发现相似的电影等。

#### **2.2 数据下载地址:**

<https://www.kaggle.com/datasets/akshaypawar7/millions-of-movies>

#### **2.3 属性类型:**

数据集包含了多个属性, 共有 13 列。具体属性包括电影的 ID (id)、电影的标题 (title)、电影的流行度 (popularity)、电影的投票平均数 (vote\_average)、电影的投票数量 (vote\_count)、电影的原语言 (original\_language)、电影的发布日期 (release\_date)、电影的预算 (budget)、电影的收入 (revenue)、电影的类别 (genres)、电影的主页 (homepage)、电影的标记 (keywords) 和电影的标记数量 (tagline) 等。

#### **2.4 数据集规模:**

数据集共包含 720,117 行数据, 表示数据集中包含了 720,117 部电影的信息。

#### **2.5 语义特征:**

TMDB 数据集包含了关于电影的多种信息, 如电影的标题、流行度、投票数、预算、收入、类别等。这些信息可以用于分析和探索电影的受欢迎程度、评分分布、电影预算与收入的关系、不同类别电影的分布情况等。此外, 数据集还包含了电影的发布日期、原语言、主页链接和标记等信息, 可以用于更深入的分析 and 挖掘。

### **(2)数据处理:**

#### **①第一部分数据处理:**

##### **1.读取数据:**

```
import pandas as pd
import numpy as np
movies = pd.read_csv('original_data.csv', encoding='utf-8')
```

movies

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore	1000.0	760505847.0
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	309404152.0
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	200074175.0
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale	27000.0	448130642.0
4	NaN	Doug Walker	NaN	NaN	131.0	NaN	Rob Walker	131.0	NaN
...	...	...	...	...	...	...	...	...	...
5038	Color	Scott Smith	1.0	87.0	2.0	318.0	Daphne Zuniga	637.0	NaN
5039	Color	NaN	43.0	43.0	NaN	319.0	Valorie Curry	841.0	NaN
5040	Color	Benjamin Roberts	13.0	76.0	0.0	0.0	Maxwell Moody	0.0	NaN
5041	Color	Daniel Hsia	14.0	100.0	0.0	489.0	Daniel Henney	946.0	10443.0
5042	Color	Jon Gunn	43.0	90.0	16.0	16.0	Brian Herzlinger	86.0	85222.0

5043 rows × 28 columns

## 2.直接删除有缺失值的行:

```
movies = movies.dropna(how='any')
```

movies

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore	1000.0	760505847.0
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	309404152.0
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	200074175.0
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale	27000.0	448130642.0
5	Color	Andrew Stanton	462.0	132.0	475.0	530.0	Samantha Morton	640.0	73058679.0
...	...	...	...	...	...	...	...	...	...
5026	Color	Olivier Assayas	81.0	110.0	107.0	45.0	Béatrice Dalle	576.0	136007.0
5027	Color	Jafar Panahi	64.0	90.0	397.0	0.0	Nargess Mamizadeh	5.0	673780.0
5033	Color	Shane Carruth	143.0	77.0	291.0	8.0	David Sullivan	291.0	424760.0
5035	Color	Robert Rodriguez	56.0	81.0	0.0	6.0	Peter Marquardt	121.0	2040920.0
5042	Color	Jon Gunn	43.0	90.0	16.0	16.0	Brian Herzlinger	86.0	85222.0

## 3.删除电影名重复的行:

```
movies.drop_duplicates(subset=["movie_title"],keep='first',inplace=True)
movies
```

C:\Users\zxh20\AppData\Local\Temp\ipykernel\_19660\656663475.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
movies.drop_duplicates(subset=["movie_title"],keep='first',inplace=True)
```

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore	1000.0	760505847.0
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	309404152.0
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	200074175.0
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale	27000.0	448130642.0
5	Color	Andrew Stanton	462.0	132.0	475.0	530.0	Samantha Morton	640.0	73058679.0
...	...	...	...	...	...	...	...	...	...
5026	Color	Olivier Assayas	81.0	110.0	107.0	45.0	Béatrice Dalle	576.0	136007.0
5027	Color	Jafar Panahi	64.0	90.0	397.0	0.0	Nargess Mamizadeh	5.0	673780.0
5033	Color	Shane Carruth	143.0	77.0	291.0	8.0	David Sullivan	291.0	424760.0
5035	Color	Robert Rodriguez	56.0	81.0	0.0	6.0	Peter Marquardt	121.0	2040920.0
5042	Color	Jon Gunn	43.0	90.0	16.0	16.0	Brian Herzlinger	86.0	85222.0

4.保存数据：

```
In [26]: movies.to_csv("processed_data.csv", index=False)
```

②第二部分数据处理：

1.读取数据并合并：

```
In [1]: import pandas as pd

# 读取contentDataPrime.csv文件
prime_data = pd.read_csv('contentDataPrime.csv')

# 读取contentDataRegion.csv文件
region_data = pd.read_csv('contentDataRegion.csv')

# 使用dataId作为合并键，将两个数据集合并
merged_data = pd.merge(prime_data, region_data, on='dataId')

# 打印合并后的数据
merged_data
```

Out[1]:

	dataId	contentType	title	length	releaseYear	endYear	votes	rating	gross	certificate	description	region
0	102795	movie	Ratha Kanneer	154	1954	-1	349	8.5	-1	NaN	The story revolves around Mohanasundaram, a re...	India
1	102796	movie	All Star Comedy Jam	89	-1	-1	349	7.5	-1	Not Rated	The Original King of Comedy and Shaquille O'Ne...	United States
2	102797	movie	Faces of Death: Fact or Fiction?	54	-1	-1	349	2.9	-1	Unrated	Add a Plotin	United States
3	102798	movie	Don't Look	71	2018	-1	349	2.8	-1	NaN	In the tradition of classic horror movies, 'Do...	United States
4	102799	movie	The Escape from Auschwitz	80	2020	-1	349	1.7	-1	TV-MA	After capturing Kazimierz Piechowski at the Hu...	United Kingdom
...	...	...	...	...	...	...	...	...	...	...	...	...
127931	497	tvSeries	The Man in the High Castle	60	2015	2019	103828	7.9	-1	TV-MA	In a dystopian America dominated by Nazi Germa...	Canada
127932	498	tvSeries	Justified	44	2010	2015	103427	8.6	-1	TV-MA	U.S. Marshal Raylan Givens is reassigned from ...	United States
127933	499	tvSeries	Misfits	60	2009	2013	103121	8.2	-1	TV-MA	A group of young offenders doing community ser...	United Kingdom

在这里，我们以 dataId 为依据合并这些数据，此时这些数据还没有 revenue 和 budget 字段。

2.合并数据：

另一份数据有 revenue 和 budget 字段，但是没有 region 字段，所以我们以 title 为依据合并这两部分字段。

```
movies_data = pd.read_csv('movies.csv')
```

movies\_data

	id	title	genres	original_language	overview	popularity	production_companies	release_date	budget	revenue	runtime
0	385687	Fast X	Action-Crime-Thriller	en	Over many missions and against impossible odds...	6682.100	Universal Pictures-Original Film-One Race-Perf...	2023-05-17	340000000.0	6.867000e+08	142.0
1	758323	The Pope's Exorcist	Horror-Mystery-Thriller	en	Father Gabriele Amorth: Chief Exorcist of the V...	5953.227	Screen Gems-2.0 Entertainment-Jesus & Mary-Wor...	2023-04-05	180000000.0	6.567582e+07	103.0
2	640146	Ant-Man and the Wasp: Quantumania	Action-Adventure-Science Fiction	en	Super-Hero partners Scott Lang and Hope van Dy...	4425.387	Marvel Studios-Kevin Feige Productions	2023-02-15	200000000.0	4.757662e+08	125.0

```
# 使用"title"作为合并键，将两个数据集合并
merged_data_movies = pd.merge(merged_data, movies_data, on='title')
```

3.选择我们需要的列：

```
: # 选择需要的列
merged_data_movies = merged_data_movies[['title', 'releaseYear', 'region', 'budget', 'revenue']]
```

```
: merged_data_movies
```

```
:
:
```

	title	releaseYear	region	budget	revenue
0	Ratha Kanneer	1954	India	0.0	0.0
1	All Star Comedy Jam	-1	United States	0.0	0.0
2	Faces of Death: Fact or Fiction?	-1	United States	0.0	0.0
3	Don't Look	2018	United States	0.0	0.0
4	Don't Look	2018	United States	0.0	0.0
...	...	...	...	...	...
195118	Misfits	2009	United Kingdom	0.0	0.0
195119	I, Robot	2004	United States	120000000.0	347234916.0
195120	I, Robot	2004	United States	0.0	0.0
195121	I, Robot	2004	Germany	120000000.0	347234916.0
195122	I, Robot	2004	Germany	0.0	0.0

#### 4.去除重复数据:

```
: # 根据"title"列删除重复的记录
merged_data_movies = merged_data_movies.drop_duplicates(subset=['title'])

# 打印去重后的数据
print(merged_data_movies)
```

```

      title  releaseYear  region \
0      Ratha Kanneer      1954    India
3      Don't Look      2018  United States
6  The Escape from Auschwitz      2020  United Kingdom
11 Super Stooges vs the Wonder Women      1974      Italy
14  It's a Mad, Mad, Mad World      1987    Hong Kong
...
195112  Under the Dome      2013  United States
195113  Unbelievable      2019  United States
195114  Titans      2018  United States
195115  Misfits      2009  United Kingdom
195119  I, Robot      2004  United States
```

#### 5.去除异常值和缺失值的数据:

```
1: # 根据条件筛选删除记录
merged_data_movies = merged_data_movies[merged_data_movies['releaseYear'] != -1]
merged_data_movies = merged_data_movies[(merged_data_movies['revenue'] != 0.0) & (merged_data_movies['budget'] != 0.0)]
merged_data_movies
```

```
1:
:
```

	title	releaseYear	region	budget	revenue
208	First Love, Last Rites	1997	United States	300000.0	40542.0
2303	Behind the Burly Q	2010	United States	250000.0	23859.0
2394	Aagathan	2010	India	822000.0	1200000.0
2396	Nasty Old People	2009	Sweden	15000.0	5300.0
2543	Black Button	2007	Australia	200.0	210.0
...	...	...	...	...	...
194989	Sex and the City	1998	United States	65000000.0	415252786.0
195011	Supergirl	2015	United States	35000000.0	14296438.0
195029	Revenge	2011	United States	2900000.0	692079.0
195077	The Fall	2013	United Kingdom	30000000.0	3669465.0
195119	I, Robot	2004	United States	120000000.0	347234916.0

#### 6.关键字替换

```
merged_data_movies['region'] = merged_data_movies['region'].replace('United States', 'USA')
```

这样是为了后期实现地图的时候更容易找到对应的经纬度。

## 7.删除 region 的异常值:

观察发现 region(地区)有些记录对应的是”n”，需要进行删除，否则后期无法正确识别到相应的地区：

```
: # 删除“region”列值为“n”的数据
merged_data_movies = merged_data_movies.drop(merged_data_movies[merged_data_movies['region'] == 'n'].index)
```

## 8.保存文件:

保存文件到磁盘中：

```
merged_data_movies.to_csv('merged_data_movies.csv', index=False)
```

## 二、Tableau 当中的数据处理

在进行分析之前，我们需要确认数据源中各字段的定义，并对数据的存储格式进行清洗。在这次分析中，我们主要清洗的字段是电影的类型（genres）和电影情节关键字（plot\_keywords）。由于不同电影可能具有多个类型和关键字，它们通过竖线符号“|”进行区分。为了方便后续的分析，我们对 genres 和 plot\_keywords 这两个字段进行了拆分，并重新汇总了数据。

对于电影的类型字段（genres），我们将每个电影的类型拆分为独立的数据，以便更好地进行分类和统计。这样，我们可以了解每种类型电影的数量和分布情况，为后续的票房分析和投资决策提供基础。

而对于电影情节关键字字段（plot\_keywords），同样进行了拆分操作。通过将每个电影的关键字分离出来，我们可以更好地理解不同关键字与电影票房之间的关联。这将帮助我们发现在电影市场中受欢迎的情节元素，并对投资决策产生积极影响。

通过对 genres 和 plot\_keywords 字段的拆分和重新汇总，我们可以更好地利用这些数据进行后续的分析工作。这将为电影投资商提供更准确的信息，帮助他们了解电影类型的趋势和受欢迎程度，以及各种情节关键字对电影票房的影响。这样，投资商可以更有针对性地制定投资策略，提高电影项目的成功概率。

# 3.目标和分析任务

### 【目标&分析】

根据电影投资商的需求，本次可视化分析项目的目标是为他们提供关于电影市场趋势和关键因素的深入洞察，以帮助他们做出明智的投资决策。以下是针对电影投资的分析任务和对应的分析子任务：

### ①分析任务 1：投资与反馈关系分析

(1)分析子任务 1：展示全球近几年电影市场的变化情况。

(2)分析子任务 2：研究预算与票房之间的关系，探索投资与反馈的直接关系。

(3)分析子任务 3：提供数据洞察，帮助投资商了解他们的投资在市场中的表现情况。

### ②分析任务 2：电影类型与票房关系分析

(1)分析子任务 1：比较不同电影类型在票房上的表现，利用条形图或热力图进行可视化展示。



(2)分析子任务 2: 发现和分析各种电影类型之间的差异, 帮助投资商选择最有潜力的电影类型进行投资。

(3)分析子任务 3: 提供趋势分析, 让投资商了解不同类型电影的市场需求和趋势。

③分析任务 3: 演员和导演的知名度影响分析

(1)分析子任务 1: 比较电影演员和导演的知名度与电影票房之间的关系, 利用散点图或箱型图进行可视化展示。

(2)分析子任务 2: 揭示知名演员和导演对吸引观众和提高票房的重要性, 为投资商提供决策依据。

(3)分析子任务 3: 让投资商了解知名演员和导演的市场价值和潜在影响。

④分析任务 4: 电影剪辑时间对票房的影响分析

(1)分析子任务 1: 通过比较不同剪辑时间的电影票房表现, 找出观众偏好的电影时长。

(2)分析子任务 2: 通过分析观众对电影时长的偏好, 帮助投资商选择最适合的剪辑时间, 提高观众吸引力和票房收益。

(3)分析子任务 3: 提供剪辑时间与票房的关联度分析, 让投资商了解影响票房的关键时间因素。

4.可视分析系统实现技术介绍

(1)探索电影市场与投资回报的状况:

第一个主题旨在帮助投资商了解电影市场的情况以及投资的回报状况。在这个主题中, 仪表板提供了三个关键视图: 「电影的全球总票房及数量变化」、「全球票房与电影数量分布情况」和「电影预算与票房收入的关系」。

1.1 可视化设计与实现

整体效果如图 2 所示:

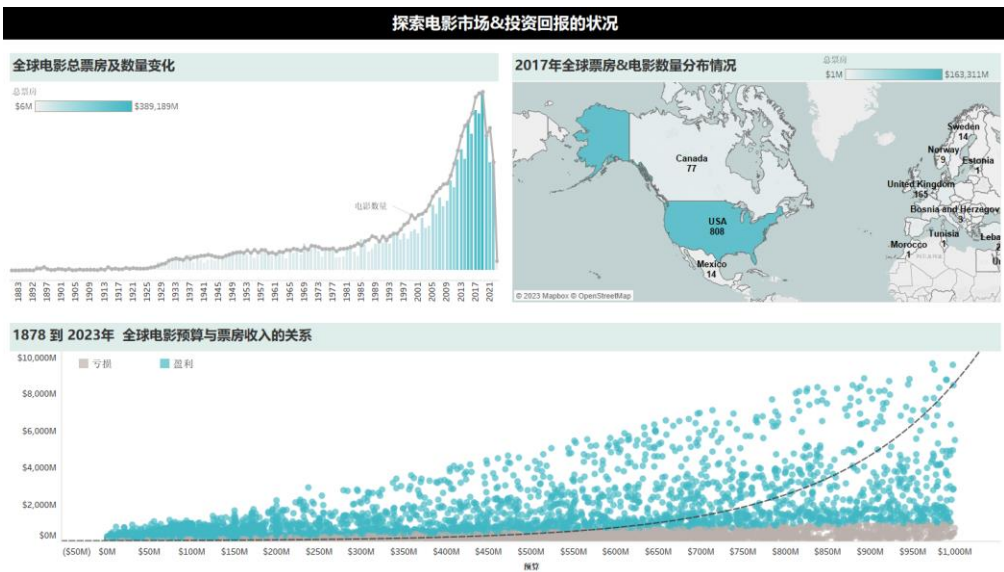


图 2 探索电影市场与投资回报的状况

在「电影的全球总票房及数量变化」视图(图 3)中, 使用柱状图和线图展示了

随着时间的推移，电影票房和电影数量的变化情况。柱状图呈现了每年的总票房，而线图则显示了每年的电影数量。通过颜色的深浅，柱状图反映了票房的高低。这个视图让投资商能够直观地了解全球电影市场的趋势和规模。

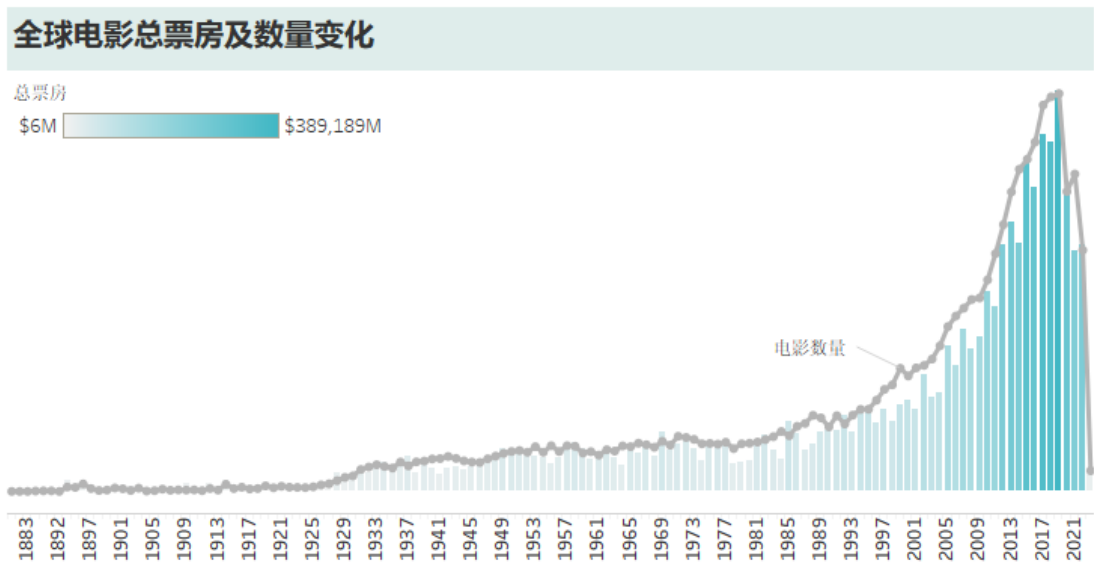


图 3 电影的总票房及数量变化(折线图和柱状图)

「全球票房与电影数量分布情况」视图(图 4)通过地图可视化展示了电影票房在全球范围内的分布情况。各个国家的票房通过填充颜色进行比较，红色越深表示票房越高。这个视图帮助投资商了解不同国家的电影市场潜力和发展情况。

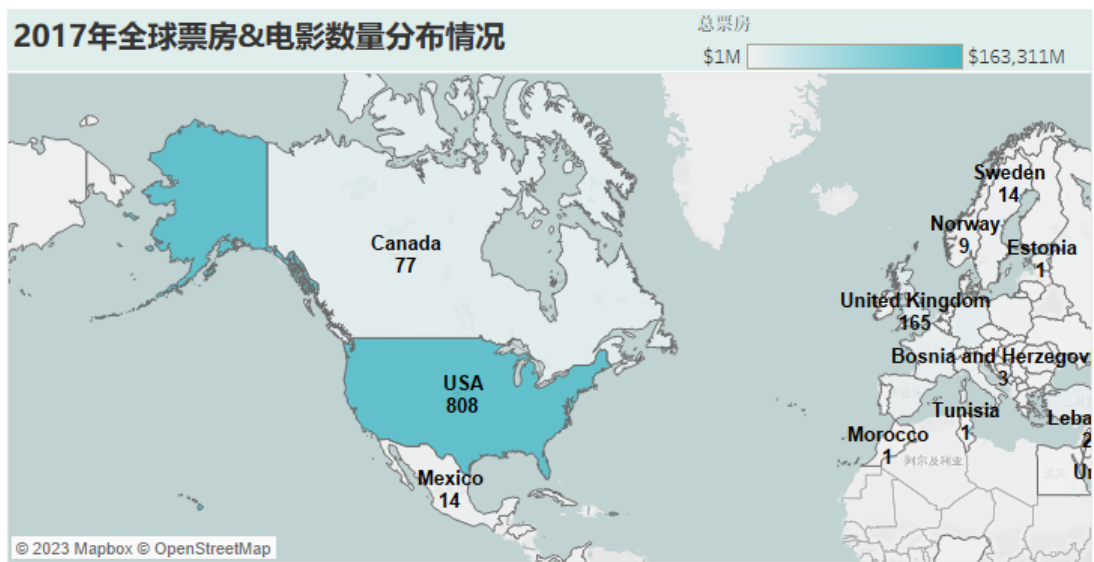


图 4 电影总票房及数量变化(地图)

「电影预算与票房收入的关系」视图(图 5)则通过散点图展示了电影预算与票房收入之间的关系。每个数据点代表一部电影，投资商可以通过数据点的位置了解该电影的预算和票房情况。视图还标注了盈利和亏损电影的趋势线，帮助投资商判断预算与票房之间的关联性。

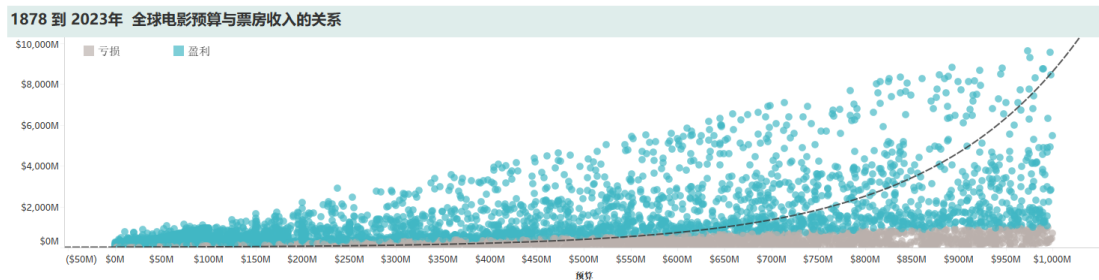


图 5 电影总票房及数量变化(散点图)

## 1.2 交互设计与实现

为增加互动性，仪表板提供了一些操作功能。在「电影的全球总票房及数量变化」视图中，将鼠标悬停在柱状图上，可以根据年份筛选展示「全球票房与电影数量分布情况」和「电影预算与票房收入的关系」的数据(如图 6)。同时，在「全球票房与电影数量分布情况」视图中，鼠标悬停在不同国家的区域上时，工具提示会显示该国家票房前五的电影名称和票房收入。

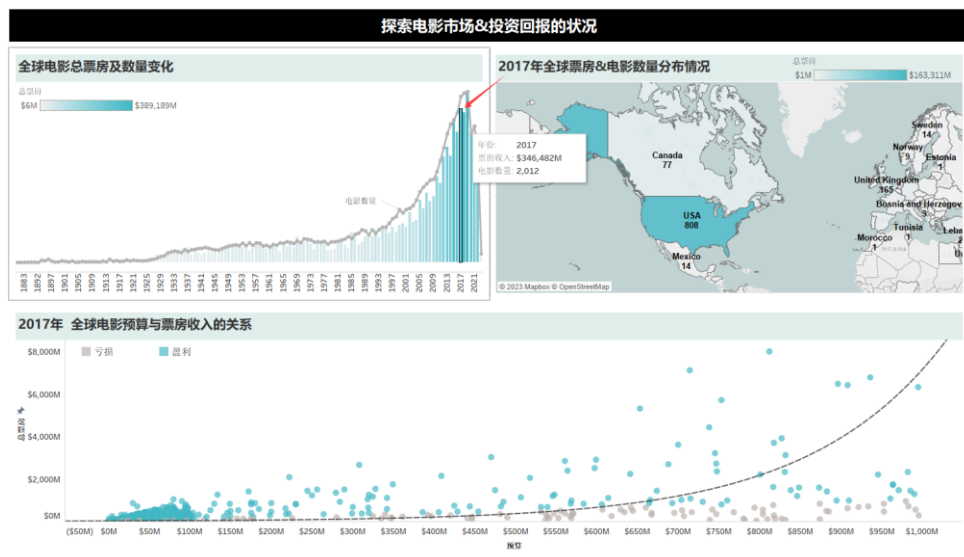


图 6 根据年份进行筛选

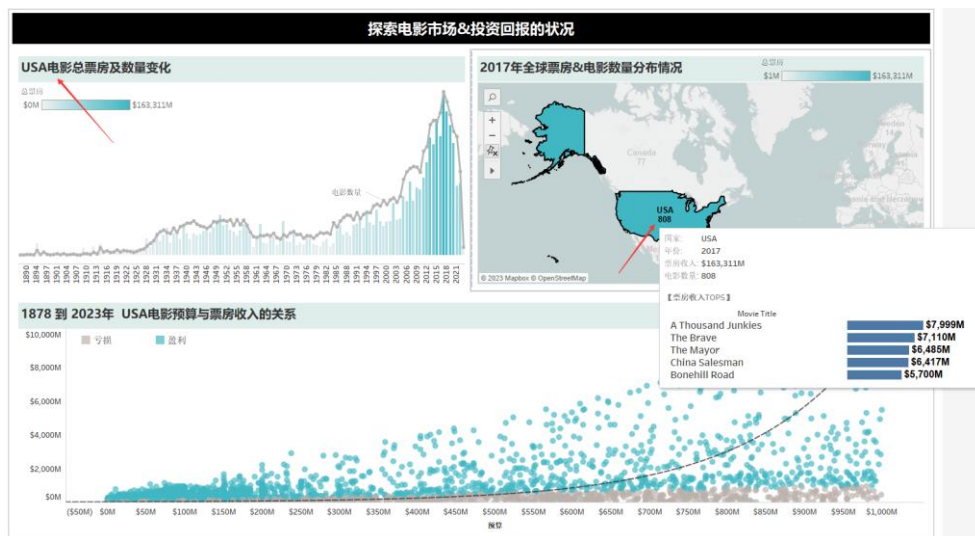


图 7 根据区域进行筛选

自然地，对于图中的散点，柱状条自然也设计了相应的提示条(如图 8、9)，

方便用户知晓相应图形对应的具体数据：

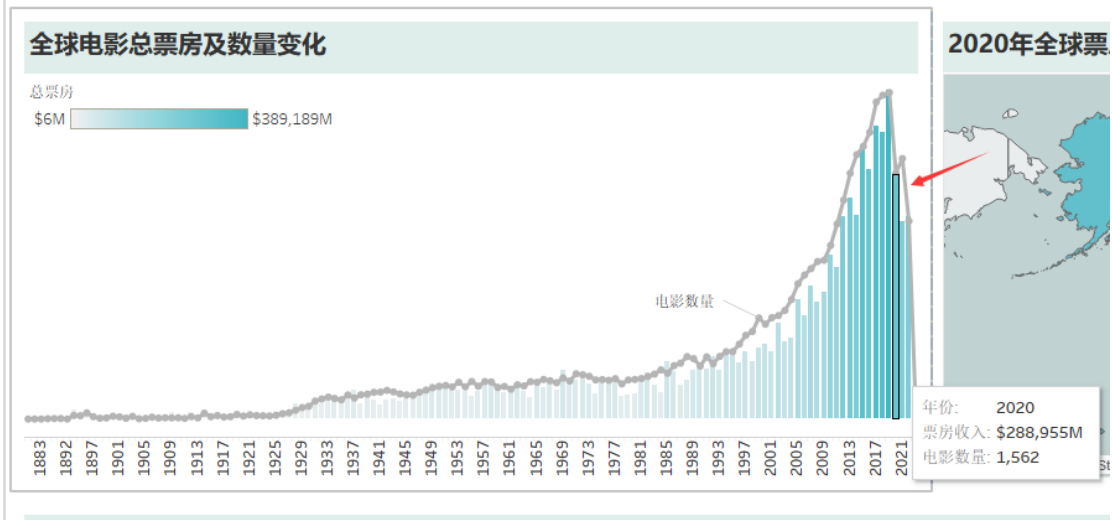


图 8 柱状图提示条

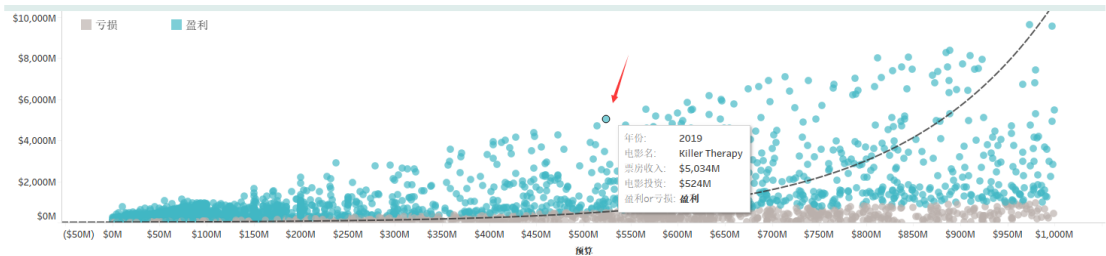


图 9 散点图提示条

### 1.3 案例分析

假设投资商使用上述可视化分析系统来探索电影市场与投资回报的状况。他们想了解全球电影市场的趋势、不同国家的票房分布情况以及电影预算与票房收入的关系。

#### ①探索电影的全球总票房及数量变化：

投资商打开仪表盘，首先查看「电影的全球总票房及数量变化」视图。他们注意到柱状图显示了每年的总票房，线图显示了每年的电影数量。通过鼠标悬停在柱状图上，他们选择了特定年份，如 2019 年。随着年份的选择，其他两个视图也相应更新。他们注意到 2019 年的总票房较高，同时电影数量也比较多。这表明投资回报可能较好，显示效果如图 10：

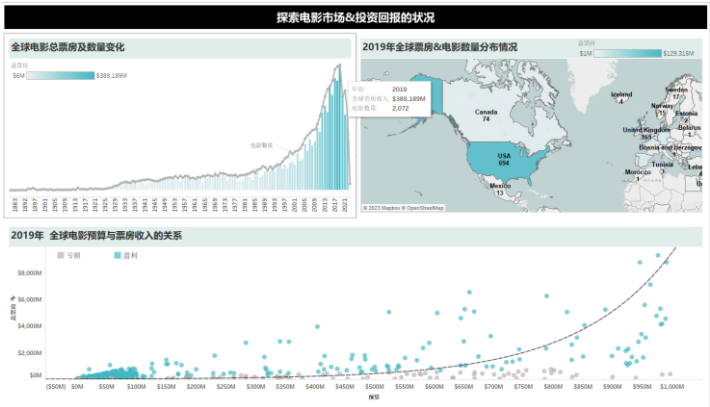


图 10 探索电影的全球总票房及数量变化



## ②全球票房与电影数量分布情况：

投资商进一步查看「全球票房与电影数量分布情况」视图。他们注意到地图上不同国家的填充颜色反映了票房的高低，颜色越深表示票房越高。他们鼠标点击在美国的区域上，工具提示显示美国票房前五的电影名称和票房收入。其余两个视图也自动切换显示美国相关的信息。投资商看到美国电影市场的潜力很大，并且有一些电影在美国取得了很高的票房收入，显示效果如图 11：

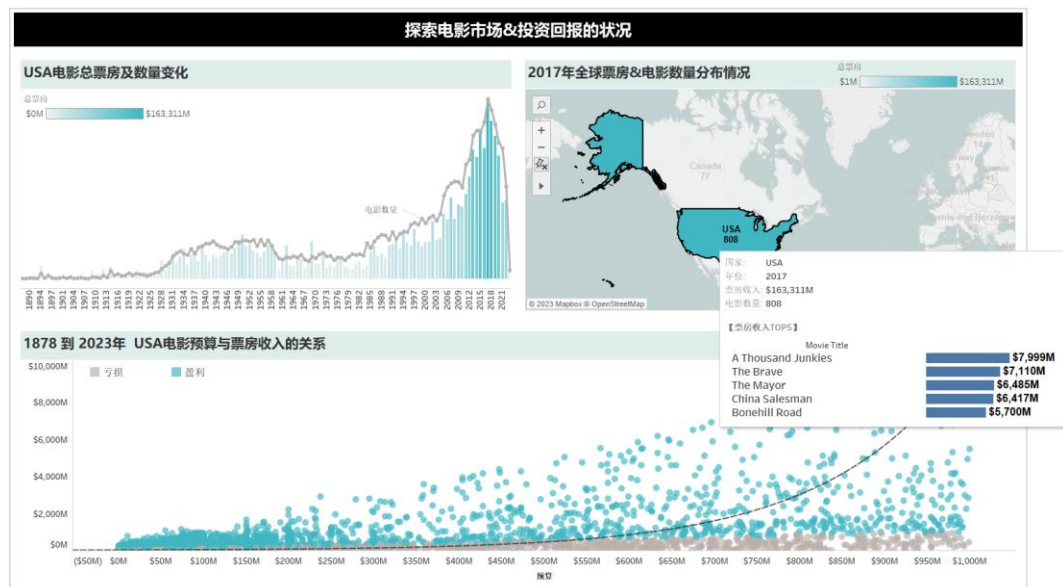


图 11 全球票房与电影数量分布情况

## ③电影预算与票房收入的关系：

投资商切换到「电影预算与票房收入的关系」视图。他们注意到散点图显示了电影的预算和票房收入，并且有盈利和亏损电影的趋势线。他们观察了一些数据点，发现一些低预算的电影取得了较高的票房收入，而一些高预算的电影却没有获得预期的回报。这让他们意识到预算并不是唯一决定票房的因素，其他因素如电影类型、演员和导演的知名度等也会影响投资回报，显示效果如图 12：

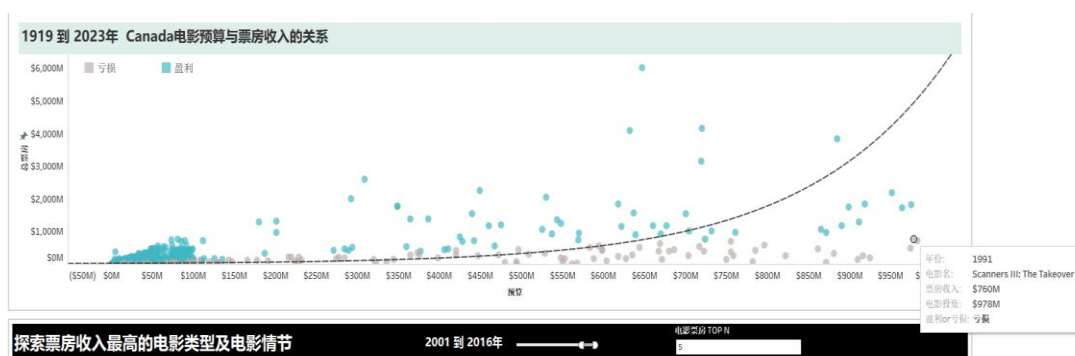


图 12 电影预算与票房收入的关系

通过不断地切换视图、筛选数据和观察趋势，投资商在可视分析系统中进行了迭代的探索过程。他们获得了以下答案和洞察：

- 1.2019 年是一个票房较高的年份，可能是一个投资的好时机。
- 2.美国电影市场具有潜力，某些电影在美国取得了显著的票房收入。
- 3.预算并非是决定票房的唯一因素，其他因素也有一定的作用。

## (2)探索票房收入最高的电影类型及电影情节：

第二个主题旨在帮助投资商了解票房收入最高的电影类型和电影情节。在这个主题中，仪表板提供了两个关键视图：「历年全球票房排名 TOP N 电影类型」和「不同电影标签的全球票房」。

### 2.1 可视化设计与实现

整体效果如图 13 所示：

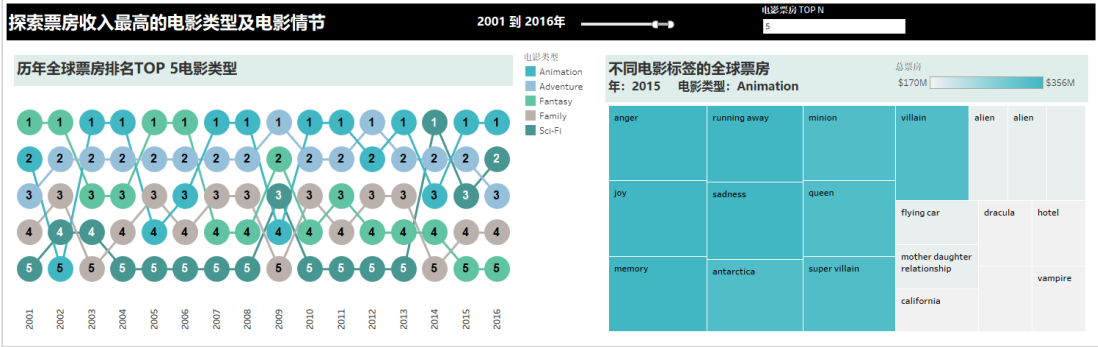


图 13 探索票房收入最高的电影类型及电影情节

「历年全球票房排名 TOP N 电影类型」视图(如图 14)以点图和线图的形式展示了不同年份中总票房排名前 N 的电影类型，并且根据选择的电影类型进行高亮显示。投资商可以通过这个视图了解各个年份最受欢迎的电影类型，并据此作出投资决策。

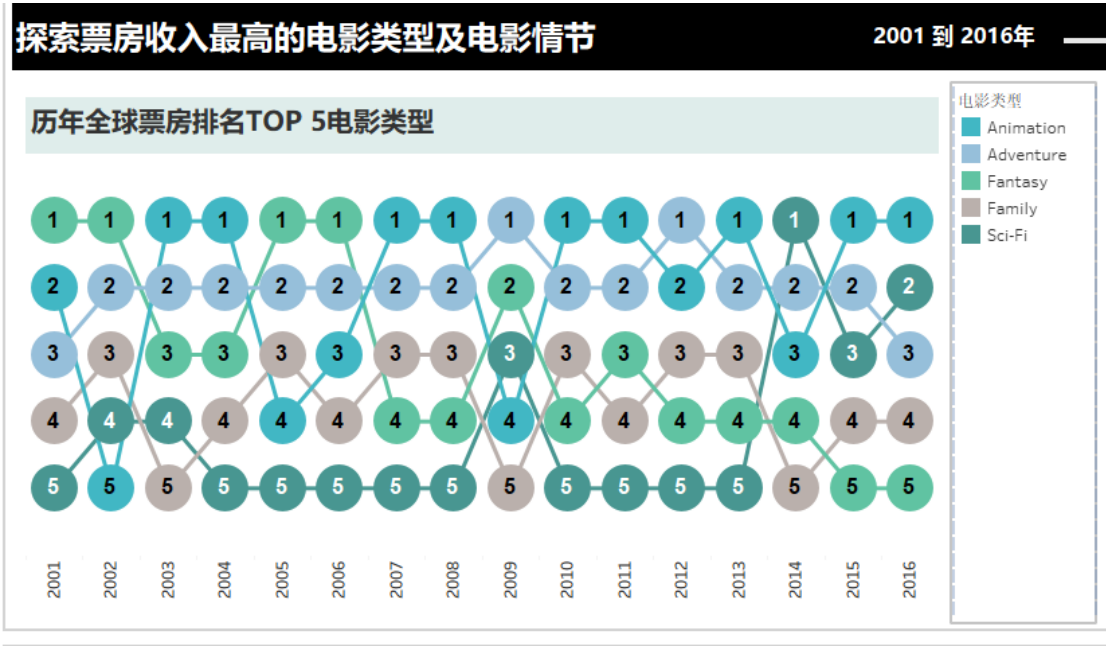


图 14 历年全球票房排名 TOP N 电影类型

「不同电影标签的全球票房」视图(如图 15)则通过热力图展示了不同电影情节的全球票房情况。热力图中，颜色的深浅和面积的大小反映了不同电影情节的票房高低。这个视图帮助投资商了解各种电影情节的受欢迎程度和票房表现。

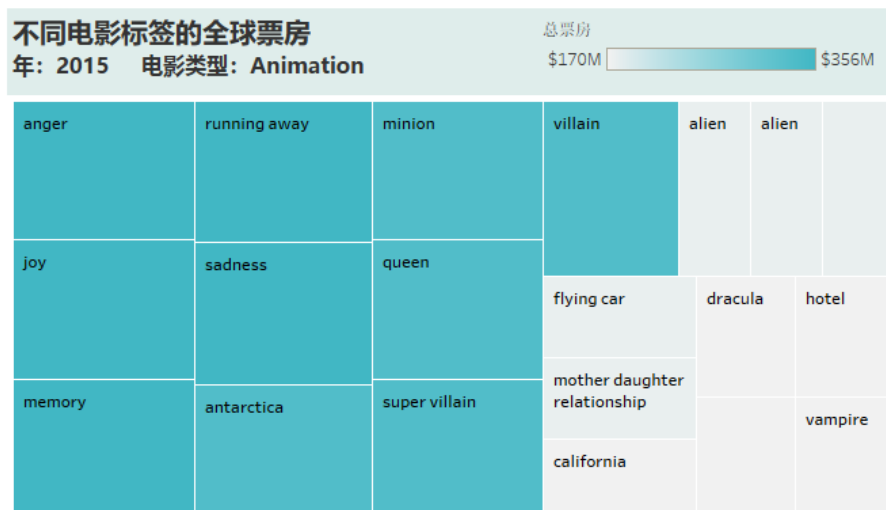


图 15 不同电影标签的全球票房

## 2.2 交互设计与实现

为增加交互性，仪表板提供了一些操作功能。仪表板提供了年份的筛选器和电影票房排名的参数设置(如图 16、17)，通过修改这些操作，可以调整视图的呈现效果。

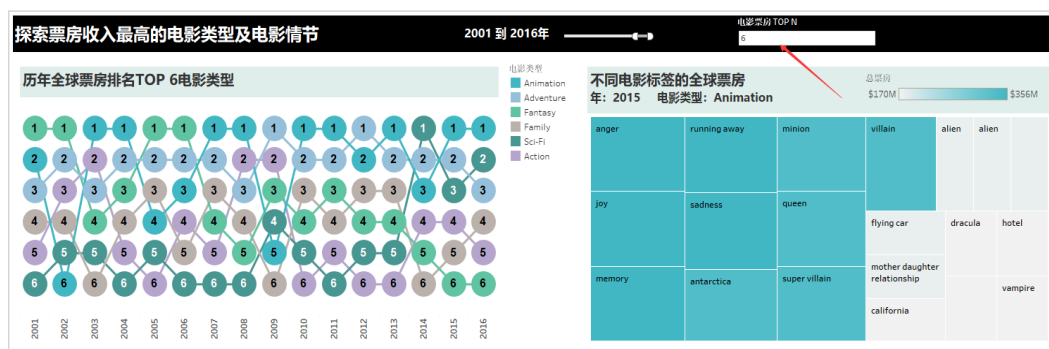


图 16 电影票房排名的筛选

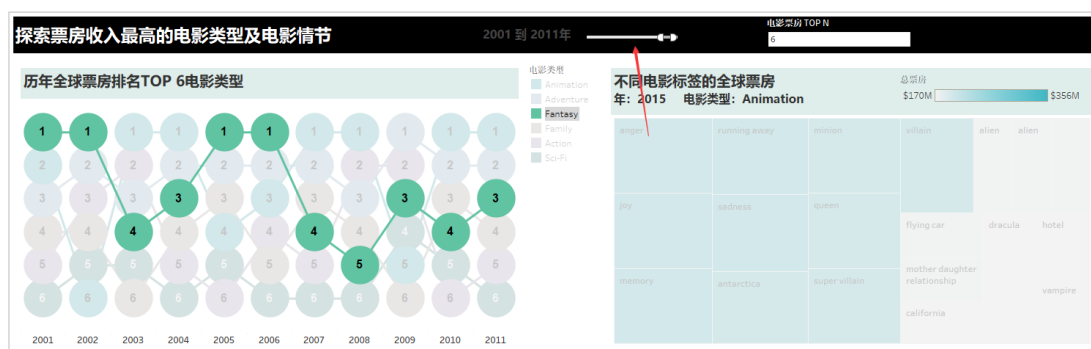


图 17 电影年份的筛选

同时，在「历年全球票房排名 TOP N 电影类型」视图中，选择特定的数据点，会对「不同电影标签的全球票房」视图中的年份和电影类型进行筛选(如图 18)，展示相应年份和电影类型下不同电影情节的票房收入比较。

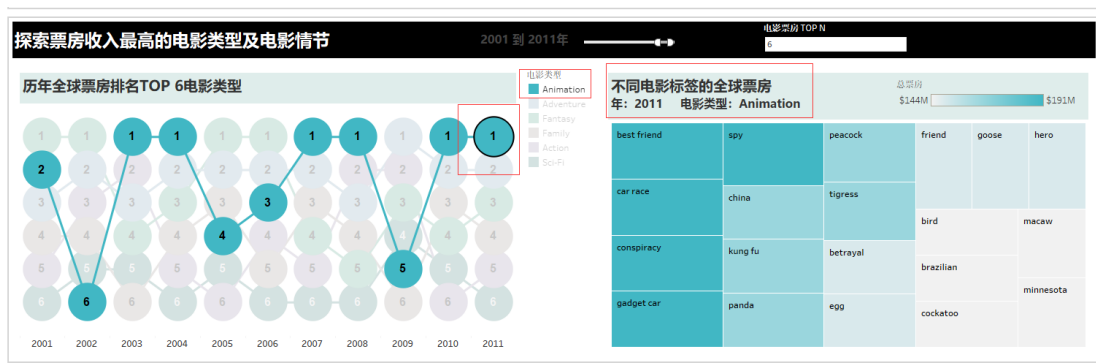


图 18 选择特定的数据点

自然地, 对于图中的点图, 矩形块自然也设计了相应的提示条(如图 19、20), 方便用户知晓相应图形对应的具体数据:

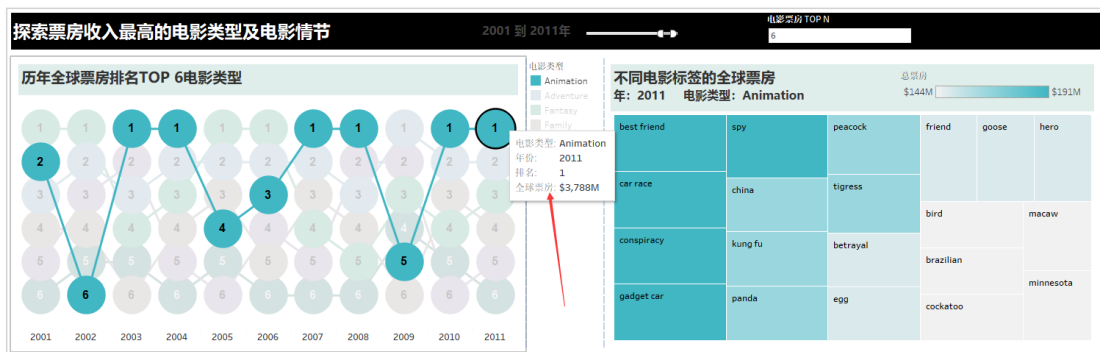


图 19 点图提示条

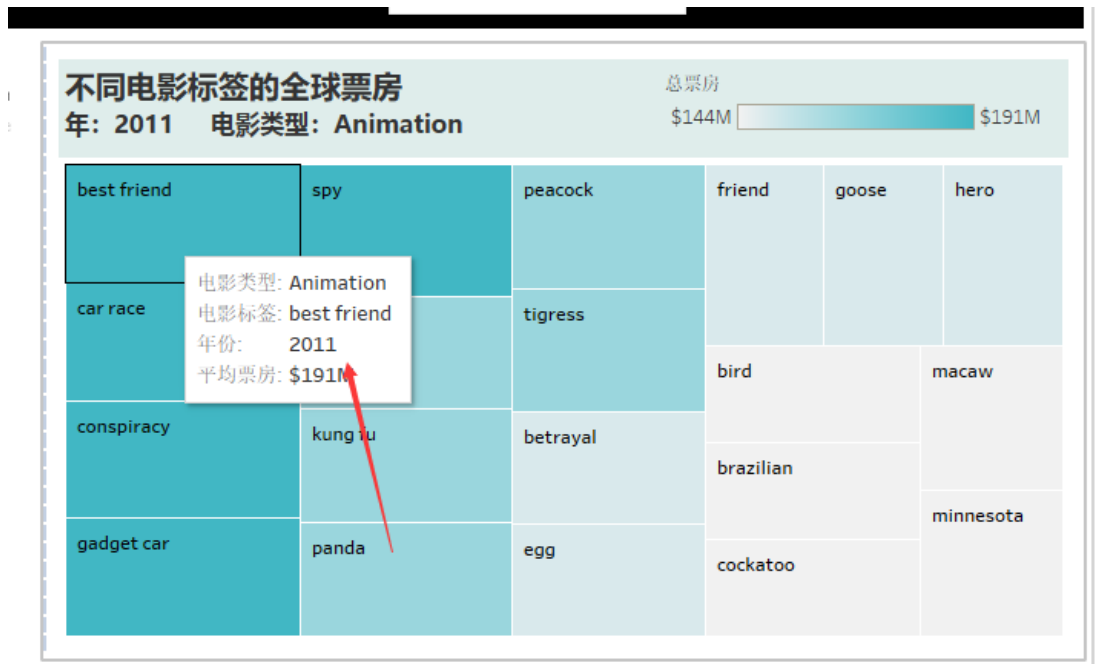


图 20 矩形图提示条

## 2.3 案例分析

假设投资商使用上述可视化分析系统来探索票房收入最高的电影类型和电影情节。他们想了解各个年份中票房排名前 N 的电影类型, 并了解不同电影标签的票房表现。



### ①历年全球票房排名 TOP N 电影类型：

投资商打开仪表盘，首先查看「历年全球票房排名 TOP N 电影类型」视图。他们注意到点图和线图展示了不同年份中票房排名前 N 的电影类型，并且根据选择的电影类型进行高亮显示。他们使用电影票房排名的筛选器，设置为显示前 10 名电影类型，并选择特定年份，如 2016 年。根据图中的高亮显示，他们得知 2016 年动画、冒险和科幻是票房最高的电影类型。这为他们提供了指导，可以关注这些类型的电影投资，如图 21 所示：

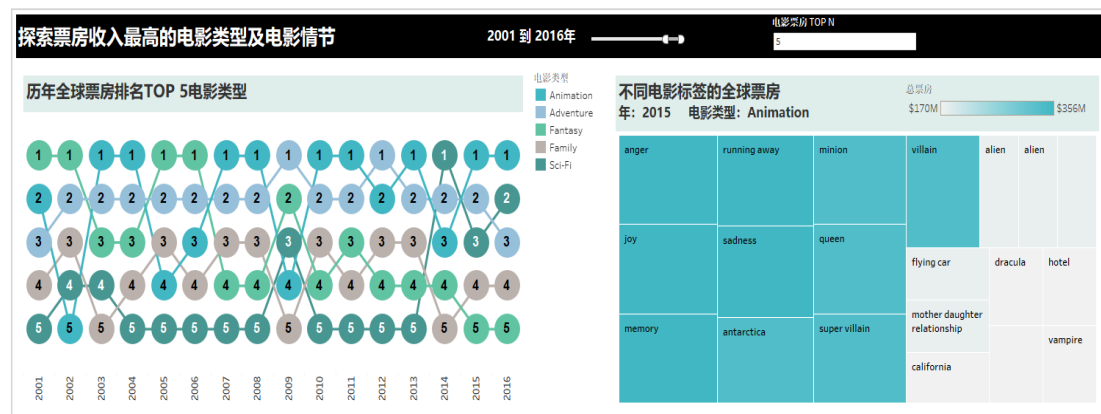


图 21 历年全球票房排名 TOP N 电影类型

### ②不同电影标签的全球票房：

投资商切换到「不同电影标签的全球票房」视图。他们注意到热力图展示了不同电影情节的全球票房情况。通过电影年份的筛选，他们选择了特定年份，如 2016 年。然后，他们在「历年全球票房排名 TOP N 电影类型」视图中选择了动作类型，以便更深入地了解该类型下不同电影情节的票房表现。在「不同电影标签的全球票房」视图中，他们观察到动画类型中，动物控制、混血动物和宠物是票房最高的电影情节。这些洞察为他们确定了在动画类型中投资的具体电影情节提供了指导，如图 22 所示：

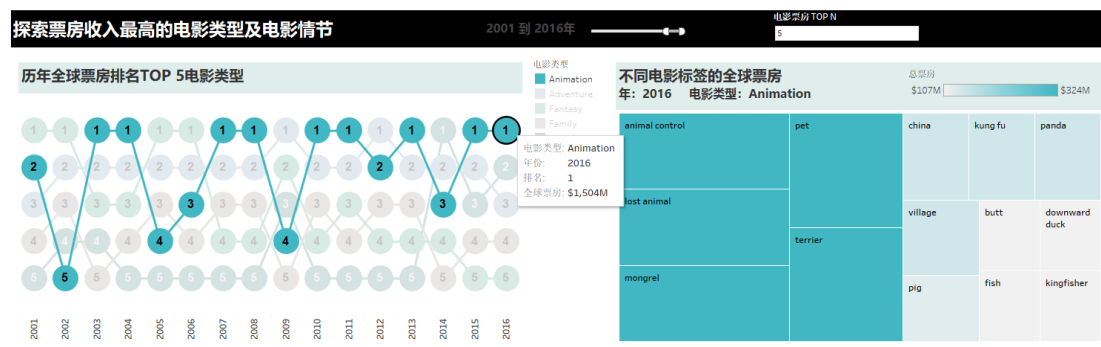


图 22 不同电影标签的全球票房

通过交互和多视图协同联动，投资商能够根据年份、电影类型和电影情节等维度深入探索票房收入最高的电影类型和电影情节。他们获得了以下答案和洞察：

1. 2016 年的票房最高的电影类型是动画、冒险和科幻。

2. 在动画类型中，动物控制、混血动物和宠物是票房最高的电影情节。

(3) 探索最具票房影响力的导演、演员以及最佳电影拍摄时长：

第三个主题包含了五个关键视图：「各导演的 IMDB 评分与全球总票房的比较」、「最具票房号召力的演员 1 号」、「最具票房号召力的演员 2 号」、「最具票房

号召力的演员 3 号」和「电影时长与总票房的关系」。

### 3.1 可视化设计与实现

整体效果如图 23 所示：

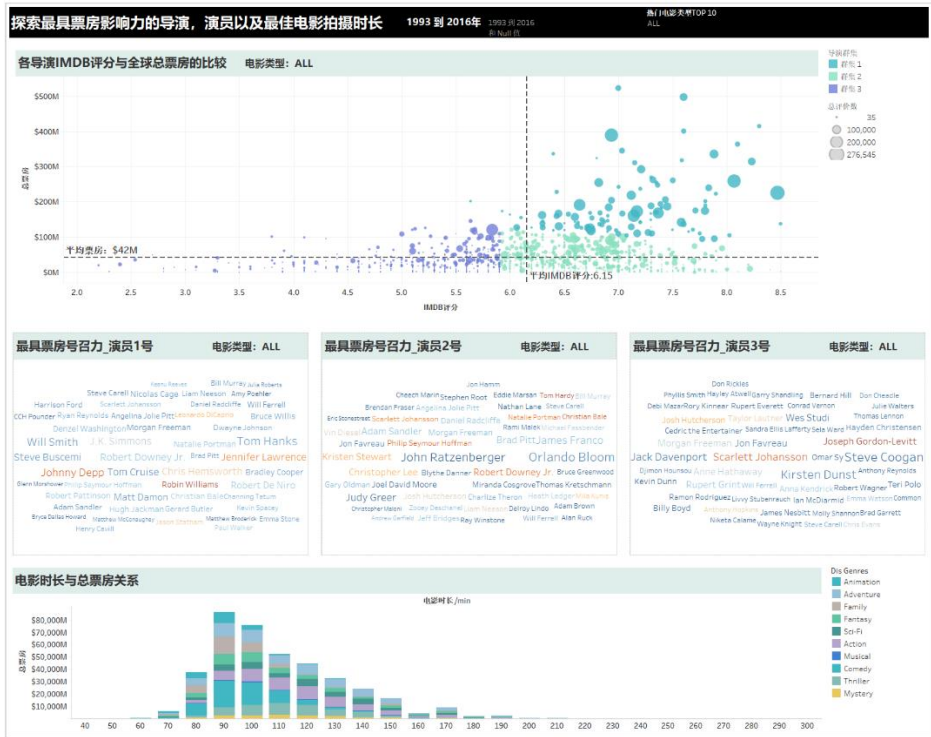


图 23 探索票房收入最高的电影类型及电影情节

「各导演的 IMDB 评分与全球总票房的比较」视图(如图 24)通过散点图展示了各导演的平均 IMDB 评分和平均票房数，并根据导演名称将数据点分布在图中。视图还根据评分和票房数的平均值设置了参考线，将导演划分为三个集群。通过这个视图，投资商可以了解哪位导演具有最高的投资价值。

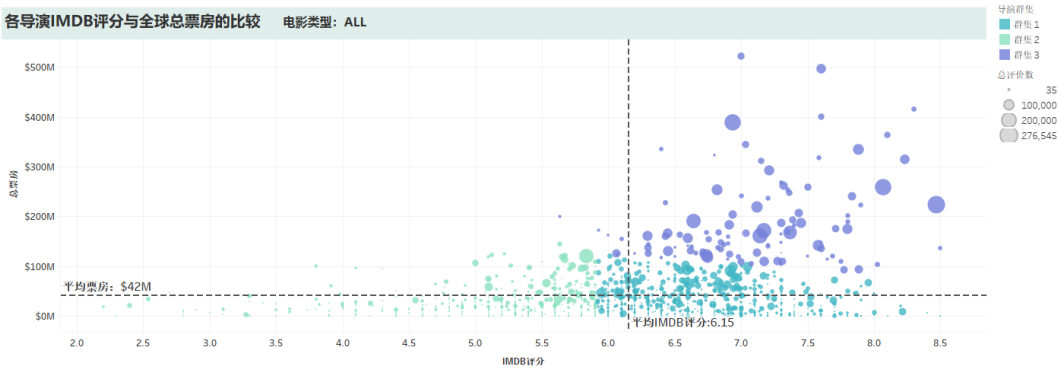


图 24 各导演的 IMDB 评分与全球总票房的比较

「最具票房号召力的演员 1 号」、「最具票房号召力的演员 2 号」和「最具票房号召力的演员 3 号」视图(如图 25)，通过词云图的形式分别展示了哪些演员对票房具有较大影响力。通过演员名称的大小和颜色的深浅，展示了演员的票房收入和粉丝数量。这些视图帮助投资商了解演员对电影票房的贡献程度，以便做出相应决策。



图 25 最具票房号召力的演员

「电影时长与总票房的关系」视图(如图 26)主要通过直方图展示高票房电影的主要时长范围。将电影时长进行分桶，并将电影票房作为 Y 轴进行堆叠，可以清楚地看出高票房电影时长主要集中在 90-110 分钟的范围。

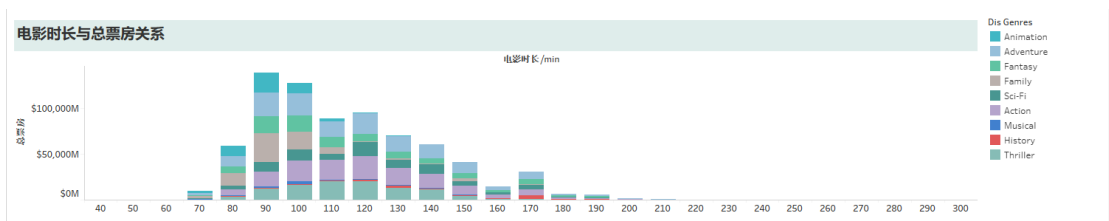


图 26 电影时长与总票房的关系

### 3.2 交互设计与实现

这部分的交互操作主要包括时间筛选器和热门电影类型 TOP N 参数设置(如图 27、28)。通过修改这些条件，投资商可以调整视图呈现的数据。在「各导演的 IMDB 评分与全球总票房的比较」视图中，选择特定导演的数据点(如图 29)，可以查看该导演拍摄过的影片中，最具票房号召力的演员。



图 27 通过时间筛选



图 28 通过电影类型筛选

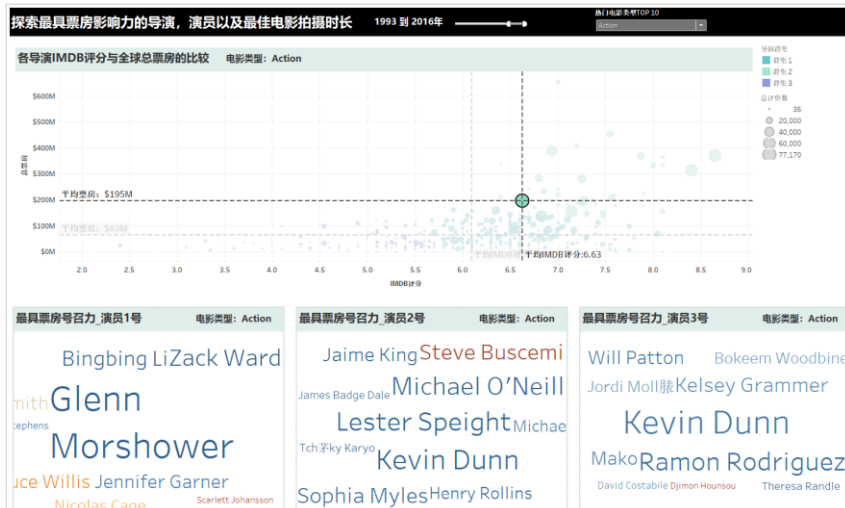


图 29 通过导演筛选

自然地, 对于图中的散点, 词块自然也设计了相应的提示条(如图 30、31), 方便用户知晓相应图形对应的具体数据:

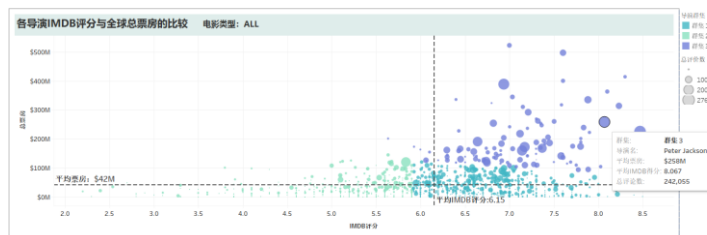


图 30 散点提示条



图 31 词块提示条



3.3 案例分析

假设投资商使用上述可视化分析系统来探索最具票房影响力的导演、演员以及最佳电影拍摄时长。他们想了解各导演的 IMDB 评分和总票房的比较，最具票房号召力的演员，以及电影时长与总票房之间的关系。

①各导演的 IMDB 评分与全球总票房的比较：

投资商打开仪表盘，首先查看「各导演的 IMDB 评分与全球总票房的比较」视图。他们注意到散点图展示了各导演的平均 IMDB 评分和平均票房数，并根据导演名称将数据点分布在图中。视图还通过参考线将导演划分为三个集群。他们使用时间筛选器，选择特定年份范围，如 2010 年至 2016 年，以查看该时期的导演表现。通过观察散点图，他们发现某些导演的平均 IMDB 评分和平均票房数都较高，如导演 A 和导演 B。这为他们确定具有最高投资价值的导演提供了指导，如图 32 所示：

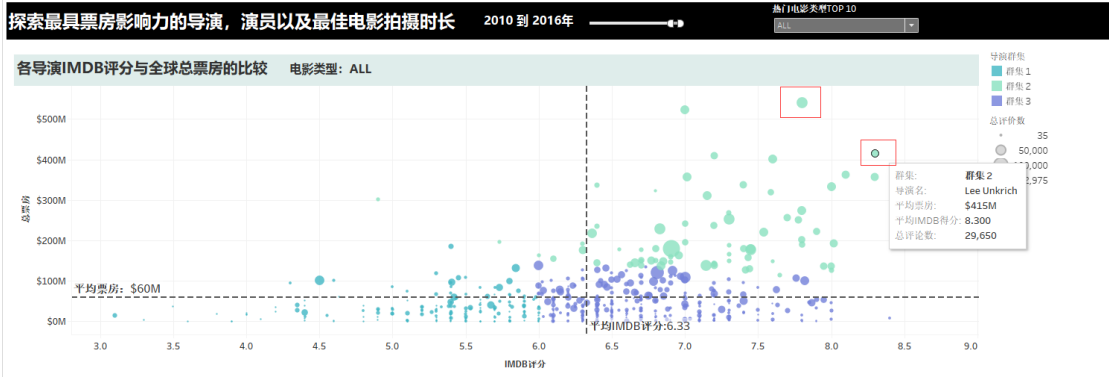


图 32 各导演的 IMDB 评分与全球总票房的比较

②最具票房号召力的演员：

投资商切换到「最具票房号召力的演员 1 号」、「最具票房号召力的演员 2 号」和「最具票房号召力的演员 3 号」视图。他们注意到词云图展示了对票房具有较大影响力的演员。通过演员名称的大小和颜色的深浅，他们可以了解演员的票房收入和粉丝数量。他们使用热门电影类型的参数设置，选择显示前 10 个电影类型，以查看这些类型下的最具票房号召力的演员。观察词云图，他们发现**演员 X**、**演员 Y** 和**演员 Z** 是票房上最具影响力的演员。这些视图帮助他们了解演员对电影票房的贡献程度，从而做出相应决策，如图 33 所示：



图 33 各导演的 IMDB 评分与全球总票房的比较

③电影时长与总票房的关系：

投资商切换到「电影时长与总票房的关系」视图。他们注意到直方图展示了高票房电影的主要时长范围。通过时间筛选器，投资商在「电影时长与总票房的关系」视图中使用时间筛选器，选择特定年份范围，如 2010 年至 2016 年，以查看该时期的电影时长和总票房之间的关系。观察直方图，他们发现高票房电影的主要时

长集中在 90-110 分钟的范围内。这为他们提供了关于最佳电影拍摄时长的信息，如图 34 所示：

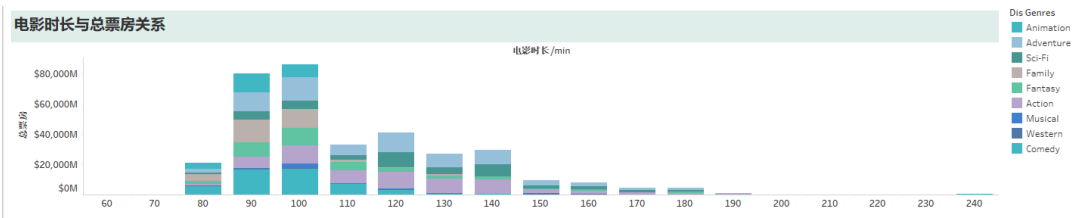


图 34 电影时长与总票房的关系

通过整个探索分析过程，投资商利用交互式的可视化分析系统获得了以下答案：

- 1.各导演的 IMDB 评分与全球总票房的比较：他们发现某些导演在平均 IMDB 评分和平均票房数方面表现出色，这些导演可能具有最高的投资价值。
- 2.最具票房号召力的演员：他们确定了对票房具有较大影响力的演员，这些演员可能在特定电影类型下表现出色，对电影票房贡献较大。
- 3.电影时长与总票房的关系：他们了解到高票房电影的主要时长集中在 90-110 分钟的范围内，这为他们提供了关于最佳电影拍摄时长的指导。

## 5. 项目总结与反思

(1) 描述您的项目是如何从最初的项目提案到最终形成较完善的系统定稿的。

从最初的项目提案到最终形成较完善的系统定稿，项目经历了以下主要步骤：

①**项目提案和需求分析：**在项目提案阶段，确定了开发一个电影可视分析系统的目标和初步需求。进行了对电影市场和投资回报相关数据的调研，明确了系统需要提供的功能和可视化视图。

②**原型设计和验证：**根据项目需求，进行可视化系统的原型设计，包括确定系统的整体布局、各个视图的样式和交互操作等。通过原型的设计和验证，可以更好地理解和评估系统的功能和使用体验，及时调整和优化设计方案。

③**数据收集和处理：**在开始开发系统之前，需要收集电影市场相关的数据，并进行清洗和处理。这涉及到获取数据，并进行数据清洗、转换和合并等操作，以便后续的可视化展示和分析。

④**确定分析主题：**小组通过对处理后数据的梳理，罗列几个分析的方向。例如：电影市场的发展方向、最受欢迎的电影类型、观众最喜欢的导演和演员、导演和演员是否对票房有一定的影响等等。基于以上几个问题，最终确认了本次分析的主题，面向对象是电影投资商，从投资商的角度对数据进行分析。

⑤**系统开发和实现：**根据需求和设计，进行可视化系统的开发和实现。这包括 Tableau 仪表盘的开发、可视化组件的编码实现、数据交互和图表的调整及优化等。同时，需要进行可视化系统的调试，确保系统的稳定性和正确性。

(2) 这期间你的可视化目标和分析任务是否有一些改变及如何改变的？

这次可视化的目标是为电影投资商提供有价值的市场信息，帮助他们更好地了解市场趋势和发掘更具潜力的投资项目。在这个过程中，可视化目标和分析任务有一些改变。最初的设想是通过可视化分析电影市场的趋势、票房与投资回报的关系等方面。然而，在数据分析过程中，发现对**电影类型**的分析存在一些困难，因为**电影的类型**字段通常是以多个标签的形式呈现，而不是单一的分类，很难画

出适合的图表。这给可视化分析带来了一定的挑战。

为了解决这个问题，小组决定对电影类型字段进行清洗和拆分，将多个类型标签分开，以便更好地进行分析和可视化呈现。这个改变的目的是为了确​​保分析结果的准确性和可解释性，并且使得投资商能够更好地理解电影类型对票房的影响，最终实现了对电影类型的分析。

通过对电影类型字段的清洗和拆分，小组成功地实现了对电影类型的可视化分析。以历年的票房数据为基础，分别分析了票房数据最高的电影类型、导演和演员等方面。结合使用 Tableau 进行可视化分析，并且提供了关于历年全球票房排名前十的电影类型的数据和趋势。在数据分析过程中，本次项目在数据梳理阶段消耗的时间较多，尤其是对电影类型和标签字段的清洗。

### **(3) 您的可视化实现技术方面发生了怎样的变化？**

**①数据清洗:**为了进行数据可视化分析，首先需要进行数据模型设计和数据清洗，以便在分析过程中获得准确和可靠的数据。在这个过程中，需要的技术包括数据清理和数据转换，使数据可以更好地适用于可视化分析。

**②数据处理和可视化技术:**使用了 Tableau 进行数据可视化分析，应用了 Tableau 相关分析和计算功能，以便可以实现丰富的数据展示方式并更好地表达数据分析结果。同时也需要从更高的角度考虑表格设计原则，包括颜色、字号、排版和美学等方面。

**③数据展示交互技术:**由于可视化分析旨在更好地帮助投资商了解数据信息，因此此次数据可视化分析项目要求对数据的呈现方式更具有可操作性、互动性和实用性。这将需要使用技术来实现图形交互，例如下拉筛选、点击筛选和参数设置进行交互等，通过这些技术方式，投资商可以根据自己的需求更深入地探究数据分析结果。

### **(4) 您是否有任何想要实现但最终无法解决的问题？如果是这样的话，那么您采用了什么变通方法，或者您放弃了最初的什么想法吗？**

**可视化展示效果:**在面对电影类型种类较多导致可视化展示效果不理想的情况下，我采取了一些变通方法来改善情况。我尝试增加交互性体验，通过提供给用户一定区间的电影类型来限制展示的范围，例如只展示票房排名前 10 的电影类型。这样可以减少展示的数据量，并使图表更加清晰和易于理解。此外，我还采用了 TOP N 的交互操作，让用户可以根据自己的需求选择要查看的前 N 个电影类型，以提高可视化展示的个性化和实用性。

### **(5) 如果您要从头开始设计和开发该项目（或任何其他交互式可视化），您会做些什么改变？**

将可视化与实际决策紧密结合。在开发数据可视化分析项目时，除了关注数据可视化技术及数据可视化交互设计外，应重视可视化内容如何结合实际业务走向。即使分析和可视化技术配有优秀的可视化交互，也需要开发出能够直接转化为实际决策的内容，而不只停留在给用户带来了一定的信息。具体来说，我们可以考虑以下改变：

**①强调需求分析:**在项目开始阶段，我们会更加重视用户需求的调研和分析。了解用户的背景、目标和需求，确保项目的设计和开发能够满足用户的实际需求，在设计过程中将用户体验放在首位。

**②数据质量和准确性:**我们会更加注重数据的质量和准确性，确保所使用的数据是可靠的和具有代表性的。这包括数据清洗、预处理和验证，以确保数据的准确性和一致性。同时，我会更加关注数据的来源和采集方法，确保数据的可信



度和可用性。

**③增加实时性和互动性：**为了使可视化项目更加实用和有吸引力，我们会考虑引入实时数据和互动性的功能。这可以通过与数据源的实时连接、交互式操作和动态更新等方式来实现。这样可以让用户更加直观地了解当前的市场情况和趋势，并根据实时数据做出更准确的决策。

**④强化可解释性和故事性：**除了提供数据和图表，我们会注重将可视化结果与故事性结合起来，以更好地传达数据背后的信息和洞察。通过设计引人入胜的数据故事和有意义的可视化呈现，可以帮助用户更好地理解数据，并将其应用到实际的决策过程中。

**(6) 通过该项目你提升最大的方面是哪些？哪些是你觉着最难的部分？**

在这个项目中，我们最大的提升是在以下方面：

**①数据分析技能：**通过对电影市场数据的分析和处理，我们提升了我们的数据分析技能。我们学会了如何清洗和处理复杂的数据集，以及如何应用统计和可视化方法来揭示数据中的模式和趋势。我们也学会了如何使用工具如 Tableau 进行数据可视化，以及如何将分析结果以简洁而有效的方式展示给其他人。

**②交互式可视化技能：**通过项目的开发，我们增强了我的交互式可视化技能。我们学会了设计和实现交互式功能，例如过滤器、动态图表和可视化控件，以提供更丰富和灵活的数据探索和分析体验。这使我能够将复杂的数据集转化为用户友好的可视化工具，帮助用户更好地理解 and 利用数据。

在这个项目中，最难的部分：

整个项目最具挑战性的部分是交互图表设计。设计交互式图表需要考虑多个方面，包括用户体验、数据呈现方式、交互功能和界面设计等，我们需要考虑以下问题：

**①思考用户的需求和使用场景：**确定应该提供哪些交互功能和如何呈现数据。这需要对用户的行为和期望有深入的理解，并结合数据的特点和分析目标进行设计。例如，在电影市场分析项目中，用户可能希望通过选择特定的年份或电影类型来查看相关的票房数据，因此需要设计相应的交互控件和过滤器。

**②考虑以清晰和易懂的方式呈现数据：**交互图表应该能够有效地传达信息，帮助用户理解数据的含义和趋势。这涉及到选择适当的图表类型、设计直观的视觉元素和标注，并考虑如何组织和布局图表元素以支持数据的比较和分析。

**③如何设计良好的交互界面：**界面应该简洁明了，易于导航和操作，同时保持足够的灵活性和可扩展性。我们需要考虑如何平衡界面的美观性和功能性，以确保用户能够轻松地与图表进行交互，并获得他们需要的信息。

在应对这些挑战的过程中，我们小组锐意进取，不断学习和借鉴优秀可视化设计案例，并与老师交流，以确保交互图表设计效果。我们不断改进和优化交互图表的设计，提升数据呈现的效果。

**(7) 如果要你对课程提出一些建设性的意见，你会给出哪 3 条最可行、且最能显著提升理论和实践教学质量和效果的可实施的建议？你课程学习的个人感悟和体会是？**

**1.曾祥辉：**

**(1)建议：**

- 1.调整作业量：可以适度减少课程作业的数量，以减轻学生的负担。这样可以给我们更多的时间和空间来深入理解和掌握课程内容，并进行更多的实践和探索。
- 2.添加代码和算法讲解：课堂上理论知识太多，希望可以在课堂上增加一些代码



和算法的讲解，帮助我们更好地理解和应用数据可视化的相关技术和方法。

3.增加互动：课堂上有效互动不多，如果可以多点互动环节可以加强我们对于知识的理解。

**(2)体会：**

通过学习这门课程，我深刻认识到数据可视化重要性和应用价值。我学会了使用各种可视化工具和技术，将复杂的数据转化为直观的图表和图形，从而更好地理解和分析数据。在实践过程中，我也发现了数据可视化中的挑战和难点，特别是在交互和图表设计方面。然而，通过不断学习和实践，我逐渐提升了自己的数据分析和可视化能力。

同时，我也意识到数据可视化需要综合运用多个学科领域的知识，这门课程为我提供了一个全面了解和学习的平台，培养了我团队合作学习中的能力。我通过与同学的合作项目和讨论，学会了团队协作和沟通，提高了自己解决问题的能力。

**2.邱雨晨**

**(1)建议：**

课程作业量稍大，可以略微减少一些作业量；课堂上可以讲一些代码和算法；实验可以全部用 d3 来做；

**(2)体会：**

我深刻认识到可视化在数据分析和沟通中的重要性。通过合适的图表、图形和可视化工具，我们能够更好地理解和解释数据，发现隐藏在数据背后的模式和趋势。这对于决策制定、问题解决和对数据进行更深入的分析至关重要。课程还鼓励了团队合作和沟通的重要性。在课程项目中，我有机会与其他同学合作，共同解决数据可视化的挑战。这锻炼了我的团队合作能力和沟通技巧，让我意识到在实际工作中与他人协作的重要性。

**3.梁汉林**

**(1)建议：**

1.增加实验数量，本次课程除了期末的大作业，只布置了 6 个实验，只这 6 个实验就占了 20 分之多，个人觉得可以适当增加平时的实验数量。

2.增加实验难度，实验任务主要都是前端相关的工作，网上能搜索到很多教程，使得实验难度大大下降了，可以适当增加实验的难度。

3.增加平时作业，本学期一共只有 3 次平时作业，尤其是读英文论文，然后严格按照论文格式写一篇论文的的作业很有意思，建议多多增加。

**(2)体会：**

课程内容十分丰富，不仅能学到与可视化相关的技术，还了解了人眼的一些特性，并且实验和作业都不太占用课后时间，希望能布置更多任务让学生充实自我。

## 6. 任务分工与团队评估

在本项目中，团队成员分工如下：

**1.邱雨晨：**

(1)负责电影数据的收集和处理。从相关网站上，获取电影相关的数据，并确保数

据的准确性和完整性。

(2)进行数据清洗和转换。清洗电影数据，处理缺失值、异常值和重复值，以保证数据的质量和可用性。

(3)数据整合和准备。将不同数据源的电影数据整合到一起，并对其进行必要的处理和格式转换，以便后续的数据分析和可视化。

## 2.梁汉林:

(1)负责数据分析的框架和思路。制定可视化分析的方法和策略，确定关键指标和分析维度，以便深入探索电影数据并发现有意义的信息。

(2)进行数据探索和分析。使用适当的统计分析方法对电影数据进行探索性分析，发现数据中的趋势、模式和关联性。

(3)提供可视化分析的案例。将分析结果转化为有意义的见解，并与团队和利益相关者分享，提供对业务决策的指导和支持。

## 3.曾祥辉:

(1)负责图表制作和仪表板的制作。使用 Tableau 工具创建各种图表和可视化元素，以展示电影数据的洞察和发现。

(2)设计和布局仪表板。根据项目要求，设计仪表板的结构、导航和交互功能，以提供友好的交互体验。

(3)优化和调整图表和仪表板。确保图表的美观和可读性，进行必要的调整和优化。

(4)完成项目报告、宣讲 ppt 和 readme 文档的撰写与制作。

在整个项目中，小组所有成员都参与到了可视化分析任务的提出当中来，并一起确定了最终可视化系统的界面图。此外，邱雨晨负责数据的收集和处理，确保电影数据的可靠性和准备就绪；梁汉林提供可视化分析的框架和思路，进行深入探索和分析；曾祥辉负责图表制作和仪表板的制作，以优雅和直观的方式展示分析结果。整个小组通过各自的工作，相互协作，为项目的成功提供了重要的贡献。

# 7. 主要参考资料

在项目开发过程中，我们参考了以下主要资料：

## 1.书籍

(1)王国平：《Tableau 数据可视化从入门到精通》

(2)喜乐君：《数据可视化分析：Tableau 原理与实践》

## 2.博客和教程:

(1)Tableau 社区论坛 (<https://public.tableau.com/app/discover>)

(2)Tableau 官方教学视频(<https://www.tableau.com/learn/training>)

## 3.开源项目和示例代码:

(1)Tableau 社区优秀作品：Movie Waves(<https://public.tableau.com/app/profile/alexandervar/viz/MovieWaves/MovieWaves>)

(2)Tableau 社区优秀作品: Dis-Order Podcast(<https://public.tableau.com/app/profile/klg27/viz/Dis-OrderPodcast/MainDashboard>)

(3)Tableau 社区优秀作品: Movies in Turkey(<https://public.tableau.com/app/profile/azuzu/viz/MoviesinTurkey/MoviesinTurkey>)