ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO

TRỰC QUAN HÓA DỮ LIỆU

Đồ án cuối kỳ - 21KHDL

Nhóm 7 - Sinh viên thực hiện:

21127104 - Đoàn Ngọc Mai

21127115 - Trần Thanh Ngân

21127129 - Lê Nguyễn Kiều Oanh

21127229 - Dương Trường Bình

21127616 - Lê Phước Quang Huy

Giảng viên hướng dẫn:

TS. Bùi Tiến Lên

Mục lục

1 Thông tin nhóm và phân công công việc 3

2 Giới thiệu tập dữ liệu 4

2.1 Thông tin cơ bản 4

2.2 Mô tả chi tiết 5

3 Khám phá và phân tích thống kê cho dữ liệu 7

3.1 Kiểm tra trùng lắp dữ liệu 7

3.2 Kiểm tra dữ liệu thiếu 8

3.3 Phân tích thông kê cho các trường dữ liệu 8

3.4 Tương quan giữa các biến 16

4 Tương quan stat model 19

4.1 Giới thiệu 19

4.2 Phương pháp 19

4.3 Kết quả 20

5 Trực quan hóa dữ liệu 24

5.1 Dashboard tổng quan 25

5.2 Movie Dashboard 33

5.3 Genres Dashboard 38

5.4 Actors Dashboard 47

5.5 Directors Dashboard 52

5.6 Company Dashboard 57

6 Dashboard cho người mù màu 64
7 Thành phẩm 65
8 Các tiêu chí đánh giá 68
9 Đề xuất cải tiến 70
Tài liệu tham khảo 71

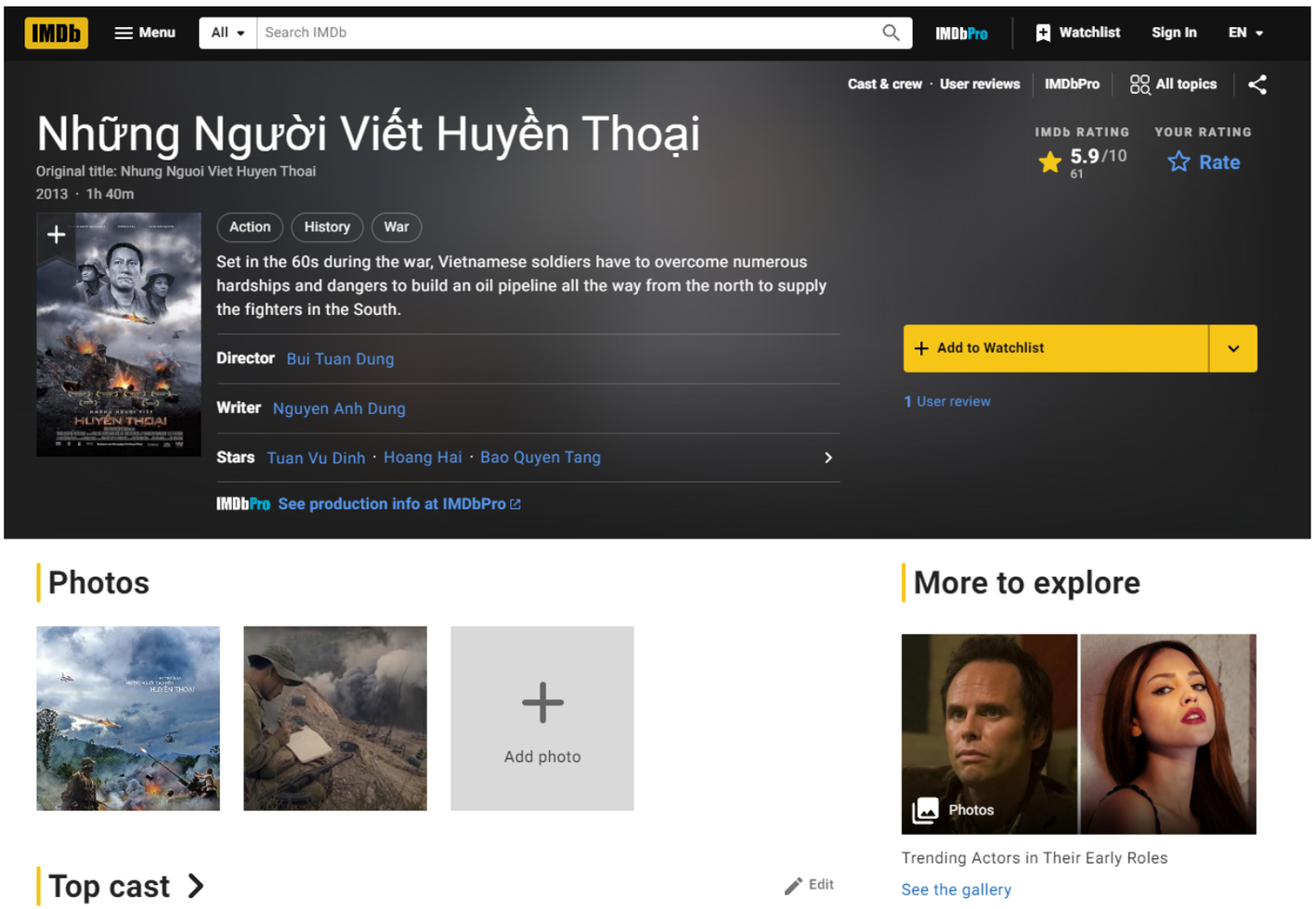
1 Thông tin nhóm và phân công công việc

MSSV	Họ và tên	Công việc được phân công	Mức độ hoàn thành
21127104	Đoàn Ngọc Mai	• Trực quan Actor Dashboard và Director Dashboard. • Thiết kế, trang trí các dashboard phiên bản người mù màu. • Chuẩn bị slide thuyết trình. • Viết báo cáo. • Thuyết trình buổi Seminar.	100%
21127115	Trần Thanh Ngân	• Trực quan Movies Dashboard. • Thiết kế, trang trí các dashboard phiên bản chính. • Thu thập thêm dữ liệu bổ sung. • Viết báo cáo. • Thuyết trình buổi Seminar.	100%
21127129	Lê Nguyễn Kiều Oanh	• Trực quan Genres Dashboard. • Thiết kế, trang trí các dashboard phiên bản người mù màu . • Sắp xếp bố cục cho các dashboard phiên bản chính. • Chuẩn bị slide thuyết trình. • Viết báo cáo. • Thuyết trình buổi Seminar.	100%
21127229	Dương Trường Bình	• Tiền xử lý dữ liệu. • Trực quan Overview Dashboard. • Thống kê và phân tích sự tương quan của các biến. • Viết báo cáo. • Chuẩn bị slide thuyết trình. • Thuyết trình buổi Seminar.	100%
21127616	Lê Phước Quang Huy	• Trực quan Companies Dashboard. • Thu thập dữ liệu từ các trang web. • Hoàn thành phần Stat Model. • Viết báo cáo. • Thuyết trình buổi Seminar.	100%

2 Giới thiệu tập dữ liệu

2.1 Thông tin cơ bản

- Tên: **Movies Shown In Vietnam**
- Nguồn gốc: Nhóm đã thu thập dữ liệu từ nhiều nguồn khác nhau
  - Cào dữ liệu kết hợp từ các website moveek và imdb
  - Sử dụng API của TMDB
- Chủ đề: Thông tin chi tiết của các bộ phim đã từng chiếu rạp tại Việt Nam



Hình 1: Trang web `imdb.com` nơi tác giả đã thu thập dữ liệu

2.2 Mô tả chi tiết

Mỗi dòng trong bộ dữ liệu mô tả các thông tin về các bộ phim (tên, doanh thu, ngân sách thể loại, năm phát hành ...) ngoài ra thông tin như đạo diễn, biên kịch, diễn viên chỉ có ở những phim điện ảnh Việt Nam. Bảng 1 trình bày ý nghĩa của từng cột dữ liệu và hình 2 thể hiện một mẫu nhỏ từ bộ dữ liệu này:

STT	Tên cột	Ý nghĩa
1	Id	Mã định danh cho mỗi bộ phim
2	Title	Tên của bộ phim
3	Original Title	Tên gốc của phim (ngôn ngữ của quốc gia sản xuất phim đó)
4	Original Language	Ngôn ngữ gốc của bộ phim
5	Overview	Tóm tắt và mô tả nội dung của bộ phim
6	Revenue	Doanh thu của bộ phim
7	Budget	Ngân sách sản xuất của bộ phim
8	Runtime	Thời lượng của bộ phim (đơn vị: phút)
9	Release Date	Ngày phát hành của bộ phim
10	Vote Average	Điểm đánh giá trung bình của bộ phim
11	Vote Count	Số lượt đánh giá mà bộ phim nhận được
12	Genres	Thể loại của bộ phim, có thể bao gồm nhiều thể loại khác nhau
13	Production Companies	Các công ty sản xuất bộ phim
14	Production Countries	Các quốc gia tham gia sản xuất bộ phim
15	Spoken Languages	Các ngôn ngữ được sử dụng trong bộ phim
16	Director	Đạo diễn của bộ phim
17	Stars	Các diễn viên chính tham gia trong bộ phim

Bảng 1: Ý nghĩa của các cột trong bộ dữ liệu

Id	Title	riginal Tit	inal Langu	Overview	Popularity	Revenue	Budget	Runtime	elease Dat	ote Averag	vote Coun	Adult	Genres	ction Com	ction Cou	ken Langu	Status	Director	Writer	Stars
291008	Legend is	Huyền Th	vi	An impov	0.703	0	0	90	2009-09-18	7.1	59	FALSE	['Drama']	['Phuoc Sa	['Vietnam	['Tiếng Vi	Released	Luu Huynh	Luu Huynh	['Dustin N
44024	Beauty in	Đẹp Từng	vi	An aspirin	0.6	0	0	109	2009-01-15	5.3	33	FALSE	['Comedy'	[]	['Vietnam	['Tiếng Vi	Released	Ngoc Dang	Vu	['Tang Tha
8681	Taken	Taken	en	While vac	87.205	2.27E+08	25000000	94	2008-02-18	7.391	10753	FALSE	['Action', '	['Wintergr	['France', '	['English',	Released			
16614	Adventur	Adventur	en	In the sun	21.476	17200000	9800000	107	2009-04-03	6.4	2075	FALSE	['Comedy'	['Miramax	['United S	['English']	Released			
44866	Park Benc	Bancs pub	fr	As she do	6.504	0	0	110	2009-07-08	5.3	50	FALSE	['Comedy'	['Why Not	['France']	['Français'	Released			
11322	Public Ene	Public Ene	en	Depressio	37.699	2.14E+08	1E+08	140	2009-07-01	6.679	4254	FALSE	['Crime', 'I	['Universa	['United S	['English']	Released			
14160	Up	Up	en	Carl Fredr	131.776	7.35E+08	1.75E+08	96	2009-05-28	7.957	19483	FALSE	['Animatic	['Pixar']	['United S	['English']	Released			
1072077	Face Off 6	Lật Mặt 6:	vi	A group o	1.153	11538081	0	131	2023-04-28	6.5	224	FALSE	['Action', '	['Ly Hai Pr	['Vietnam	['Tiếng Vi	Released	Ly Hai		['Trung Du
19908	Zombielar	Zombielar	en	Columbus	35.549	1.02E+08	23600000	88	2009-10-07	7.3	11811	FALSE	['Comedy'	['Pariah', '	['United S	['English']	Released			
64937	Passport t	Chuyện Tì	vi		1.196	173828	0	120	2009-02-13	6.5	176	FALSE	['Drama', '	[]	['Vietnam	['Tiếng Vi	Released	Victor Vu	Binh Le	['Binh Min
19913	(500) Days	(500) Days	en	Tom, gree	56.515	60781545	7500000	95	2009-07-17	7.3	9664	FALSE	['Comedy'	['Fox Sear	['United S	['English',	Released			
173380	Burning K	Những Nụ	vi	A group o	1.346	0	240000	98	2010-01-15	6.2	75	FALSE	['Music']	[]	['Vietnam	['Tiếng Vi	Released	Quang Du	Quang Du	['Minh Har
172534	Công Chúa	Công Chúa	vi		0.842			0	2010-02-04			FALSE	[]	[]	['Vietnam	['Tiếng Vi	Released			
11324	Shutter Is	Shutter Is	en	World Wa	94.396	2.95E+08	80000000	138	2010-02-14	8.202	23108	FALSE	['Drama', '	['Paramou	['United S	['English',	Released			
12155	Alice in W	Alice in W	en	Alice, now	83.115	1.03E+09	2E+08	108	2010-03-03	6.632	13485	FALSE	['Family', '	['Walt Dis	['United S	['English']	Released			
22894	Legion	Legion	en	When God	47.781	67900000	26000000	100	2010-01-21	5.8	2479	FALSE	['Horror', '	['Screen G	['United S	['English']	Released			
329851	Let Hoi De	Để Mai Tír	vi	A transger	2.045	165987	1500000	94	2014-01-01	6.6	452	FALSE	['Comedy'	['CJ Entert	['South Ko	['Tiếng Vi	Released	Charlie Ng	Dustin Ng	['Dustin N
1726	Iron Man	Iron Man	en	After bein	101.018	5.85E+08	1.4E+08	126	2008-04-30	7.642	25572	FALSE	['Action', '	['Marvel S	['United S	['English',	Released			
1011477	Untitled K	Untitled K	en		7.781	0	0	0	2024-12-12	0	0	FALSE	['Action', '	['Columbi	['United S	['English']	Planned			
34544	The A-Tea	The A-Tea	en	A group o	39.353	1.77E+08	1.1E+08	117	2010-06-09	6.4	4439	FALSE	['Thriller',	['Phoenix	['United K	['Deutsch'	Released			
37834	Knight an	Knight an	en	A fugitive	41.104	2.62E+08	1.17E+08	109	2010-06-15	6.2	4287	FALSE	['Action', '	['Pink Mac	['United S	['English',	Released			
27205	Inception	Inception	en	Cobb, a sk	149.015	8.26E+08	1.6E+08	148	2010-07-15	8.367	35454	FALSE	['Action', '	['Legenda	['United K	['English',	Released			
27576	Salt	Salt	en	As a CIA o	29.486	2.93E+08	1.1E+08	100	2010-07-21	6.4	5259	FALSE	['Action', '	['Wintergr	['United S	['English',	Released			
299054	Expend4b	Expend4b	en	Armed wi	176.936	58000000	1E+08	103	2023-09-15	6.364	1236	FALSE	['Action', '	['Millenni	['Bulgaria'	['English']	Released			
425724	Inferno	Giao lộ đị	en	Inferno fo	0.623	0	0	100	2010-09-17	6	1	FALSE	['Drama', '	[]	['Vietnam	['Tiếng Vi	Released			
33909	Wall Stree	Wall Stree	en	As the glo	33.094	1.35E+08	70000000	133	2010-09-20	6.1	1447	FALSE	['Drama', '	['Pressma	['United S	['English']	Released			
60084	Floating L	Cánh Đồn	vi	Set in the	1.263	0	800000	105	2010-12-03	7.1	397	FALSE	['Drama']	['Vietnam	['Vietnam	['Tiếng Vi	Released	Phan Quai	Ngu Nguy	['Tang Tha
8848	The Bank	The Bank	en	Terry is a	31.832	64828421	20000000	111	2008-02-28	6.834	2172	FALSE	['Thriller',	['Omnilab	['Australia	['English']	Released			
38055	Megamin	Megamin	en	After Meg	99.147	3.22E+08	1.3E+08	96	2010-10-28	6.976	6551	FALSE	['Animatic	['DreamW	['United S	['English']	Released			
37710	The Touri	The Touri	en	American	29.877	2.79E+08	1E+08	103	2010-12-08	6.183	5236	FALSE	['Action', '	['Columbi	['France', '	['English',	Released			
38745	Gulliver's	Gulliver's	en	Travel wri	25.702	2.37E+08	1.12E+08	85	2010-12-23	5.2	2292	FALSE	['Family', '	['Dune En	['United S	['English']	Released			
40805	The Greer	The Greer	en	Britt Reid,	26.847	2.28E+08	1.2E+08	119	2011-01-12	5.6	3086	FALSE	['Action', '	['Original	['United S	['English',	Released			
405880	Ghost at S	Bóng Ma H	vi	Writer Na	1.093	0	0	83	2011-01-26	3.6	39	FALSE	['Horror', '	[]	['Vietnam	['Tiếng Vi	Released	Le Bao Tru	Le Bao Tru	['Ngoc Die
380265	Battle Of	Cô Dâu Đạ	vi		1.171	0	0	100	2014-01-16	6	97	FALSE	['Comedy'	['Galaxy S	[]	['Tiếng Vi	Released	Victor Vu		['Hoang Ba
38757	Tangled	Tangled	en	When the	127.619	5.92E+08	2.6E+08	100	2010-11-24	7.605	10953	FALSE	['Animatic	['Walt Dis	['United S	['English']	Released			
16710	Meet the	Meet the	en	A single m	8.098	41975388	20000000	100	2008-03-21	6.7	96	FALSE	['Comedy'	['Tyler Per	['United S	['English']	Released			
41630	No Strings	No Strings	en	Emma is a	40.945	1.49E+08	25000000	107	2011-01-21	6.296	4530	FALSE	['Comedy'	['Katalyst	['United S	['English']	Released			

Hình 2: Một mẫu nhỏ trong bộ dữ liệu Movies Shown In Vietnam

3 Khám phá và phân tích thống kê cho dữ liệu

Sau khi thu thập được bộ dữ liệu, nhóm tiến hành khám phá để hiểu rõ bộ dữ liệu, tiền xử lí và thống kê cho các trường dữ liệu trước khi tiến hành trực quan

```
RangeIndex: 2193 entries, 0 to 2192
Data columns (total 17 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Id                    2193 non-null   int64
 1   Title                 2193 non-null   object
 2   Original Title        2193 non-null   object
 3   Original Language     2193 non-null   object
 4   Overview              2083 non-null   object
 5   Revenue               2193 non-null   float64
 6   Budget                2193 non-null   float64
 7   Runtime               2193 non-null   int64
 8   Release Date          2193 non-null   object
 9   Vote Average          2193 non-null   float64
 10  Vote Count            2193 non-null   int64
 11  Genres                2193 non-null   object
 12  Production Companies  2193 non-null   object
 13  Production Countries  2193 non-null   object
 14  Spoken Languages      2193 non-null   object
 15  Director              463 non-null    object
 16  Stars                 456 non-null    object
dtypes: float64(3), int64(3), object(11)
```

Hình 3: Thông tin chi tiết bộ dữ liệu

Đầu tiên, nhóm xác định rằng bộ dữ liệu hiện tại:

- Chứa tổng cộng **2193 dòng** và **17 cột**.
- Trong số đó, có 5 cột bị mất mát dữ liệu đáng kể là `Director` và `Stars`
- Chuyển cột `Release Date`về kiểu dữ liệu datetime

3.1 Kiểm tra trùng lắp dữ liệu

Bộ dữ liệu không có dòng nào bị trùng lắp

Cột	**Số lượng giá trị thiếu**	**Tỉ lệ giá trị thiếu (%)**
Stars	1737	79.2
Director	1730	78.9
Overview	110	5.01

Bảng 2: Các cột bị thiếu dữ liệu ở tất cả các bộ phim

3.2 Kiểm tra dữ liệu thiếu

Vì ngữ cảnh cũng như mục đích của đồ án liên quan đến Việt Nam nên nhóm sẽ tập trung phân tích ở các bộ phim Việt Nam nhiều hơn.

Cột	**Số lượng giá trị thiếu**	**Tỉ lệ giá trị thiếu (%)**
Overview	108	23.5
Stars	12	2.6
Director	4	0.87

Bảng 3: Các cột bị thiếu dữ liệu ở các bộ phim Việt Nam

Dựa vào Bảng 2 và 3, ta có thể thấy:

- Cột `Overview` đối với các phim Việt bị thiếu khoảng 25%, nhưng vì cột này không ảnh hưởng quá nhiều đến giai đoạn trực quan nên ta không xóa.
- Hai cột `Director` và `Stars` ở các bộ phim nước ngoài thiếu rất nhiều, vì nhóm tập trung thu thập cho các bộ phim ở Việt Nam hơn. Do đó, hai cột này ở các phim Việt thiếu rất ít nên ta vẫn giữ lại.

3.3 Phân tích thông kê cho các trường dữ liệu

Ở phần này nhóm sẽ tập trung phân tích cho các bộ phim Việt Nam. Với mỗi cột nhóm sẽ tính các giá trị thông kê của cột (mean, std, min, max...), khoảng tin cậy 95% cho giá trị trung bình và phân phối của dữ liệu.

$$\text{Khoảng tin cậy 95\% của giá trị trung bình} = \bar{x} \pm z_{1-\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$$

Với:

- $\bar{x}$: giá trị trung bình của mẫu
- $z_{1-\frac{\alpha}{2}}$: giá trị z-score tương ứng với $\alpha = 0.05$
- s: độ lệch chuẩn của mẫu
- n: số lượng mẫu

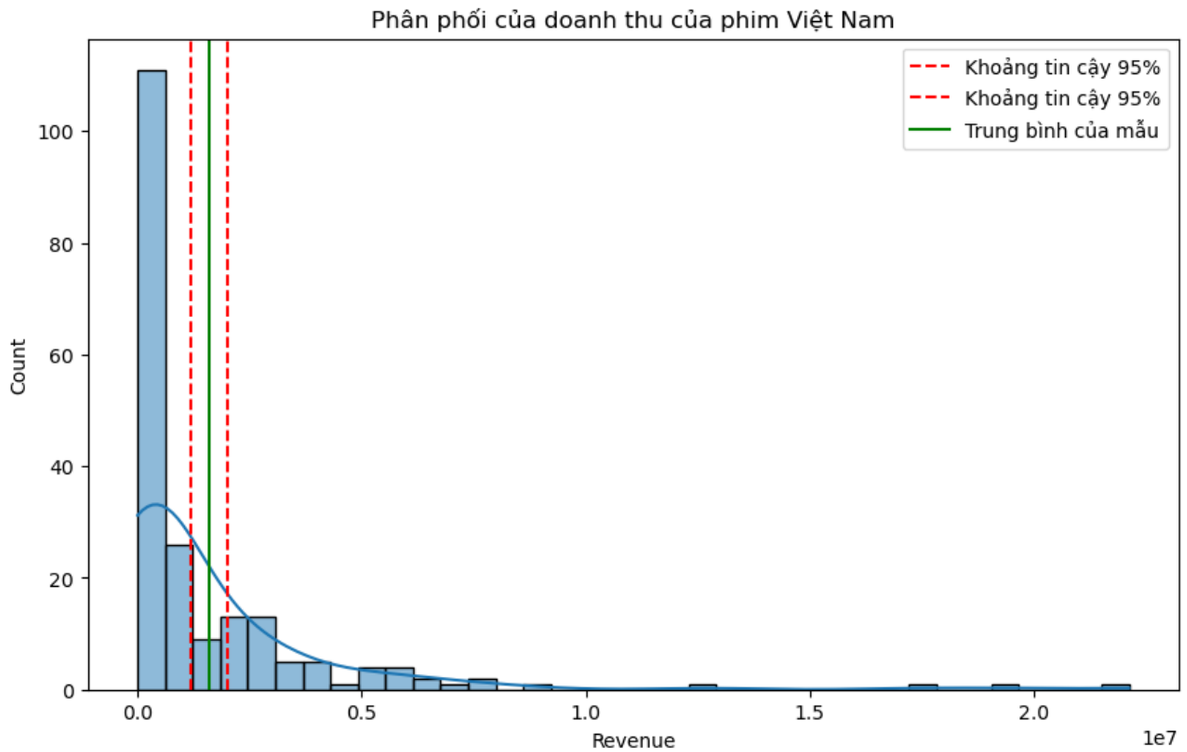
Revenue

Bảng 4: Thống kê và khoảng tin cậy 95%

Thống kê	**Giá trị**
Trung bình (mean)	1,637,583
Độ lệch chuẩn (std)	2,796,444
Giá trị nhỏ nhất (min)	975
Phân vị 25%	124,606
Phân vị 50% (median)	665,224
Phân vị 75%	2,136,495
Giá trị lớn nhất (max)	22,119,910
Khoảng tin cậy 95%	(1,275,393; 1,999,773)

Nhận xét

- Doanh thu của các bộ phim Việt Nam có phân phối lệch phải, khoảng giá trị khá rộng từ khoảng 1000 đô đến hơn 22 triệu đô.
- Phần các bộ phim Việt có doanh thu từ vài trăm nghìn đô đến hơn 2 triệu đô. Tuy nhiên vẫn có các trường hợp ngoại lệ với doanh thu rất thấp hoặc rất cao.
- Doanh thu trung bình của các bộ phim Việt Nam từ 1.27 triệu đô đến 1.99 triệu đô với khoảng tin cậy 95%

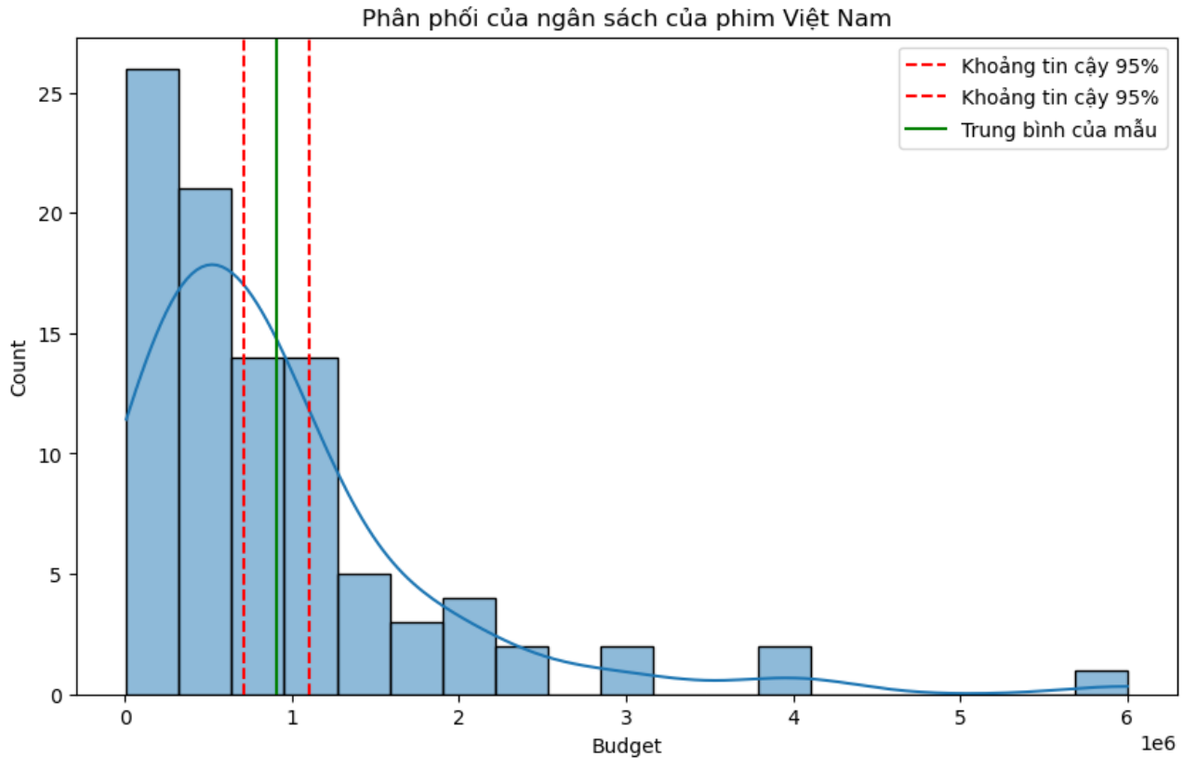


Hình 4: Phân phối của doanh thu phim Việt Nam

Budget

Bảng 5: Thống kê và khoảng tin cậy 95%

Thống kê	**Giá trị**
Trung bình (mean)	906,335
Độ lệch chuẩn (std)	925,063
Giá trị nhỏ nhất (min)	8000
Phân vị 25%	300,000
Phân vị 50% (median)	669,159
Phân vị 75%	1,023,219
Giá trị lớn nhất (max)	6,000,000
Khoảng tin cậy 95%	(725,926; 1,086,744)



Hình 5: Phân phối của ngân sách phim Việt Nam

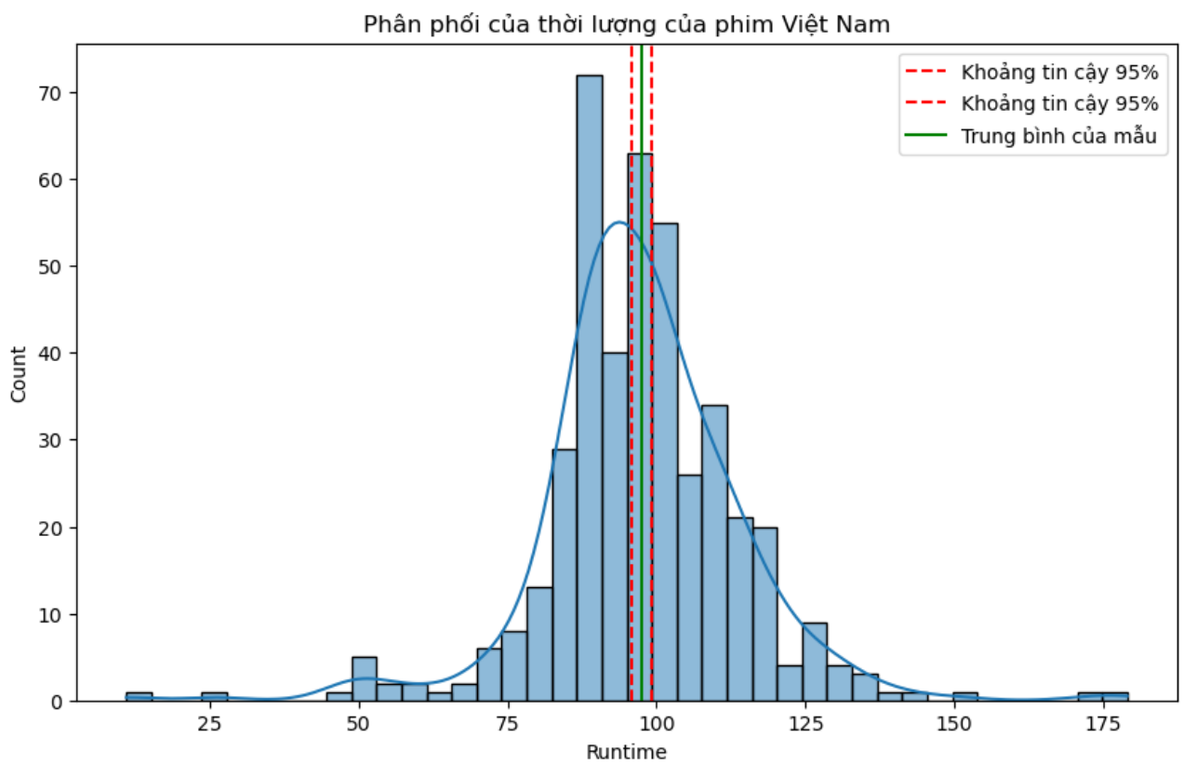
Nhận xét

- Ngân sách của các bộ phim Việt Nam có phân phối lệch phải, khoảng giá trị khá rộng từ khoảng 8000 đô đến 6 triệu đô.
- Phần lớn các bộ phim Việt có ngân sách từ vài trăm nghìn đô đến hơn 1 triệu đô. Tuy nhiên vẫn có các trường hợp ngoại lệ với ngân sách rất thấp hoặc rất cao.
- Ngân sách trung bình của các bộ phim Việt Nam từ 725 nghìn đô đến 1.08 triệu đô với khoảng tin cậy 95%

Runtime

Bảng 6: Thống kê và khoảng tin cậy 95%

Thống kê	**Giá trị**
Trung bình (mean)	97.88
Độ lệch chuẩn (std)	16.04
Giá trị nhỏ nhất (min)	11
Phân vị 25%	90
Phân vị 50% (median)	98
Phân vị 75%	107
Giá trị lớn nhất (max)	179
Khoảng tin cậy 95%	(96.3; 99.4)



Hình 6: Phân phối của thời lượng phim Việt Nam

Nhận xét

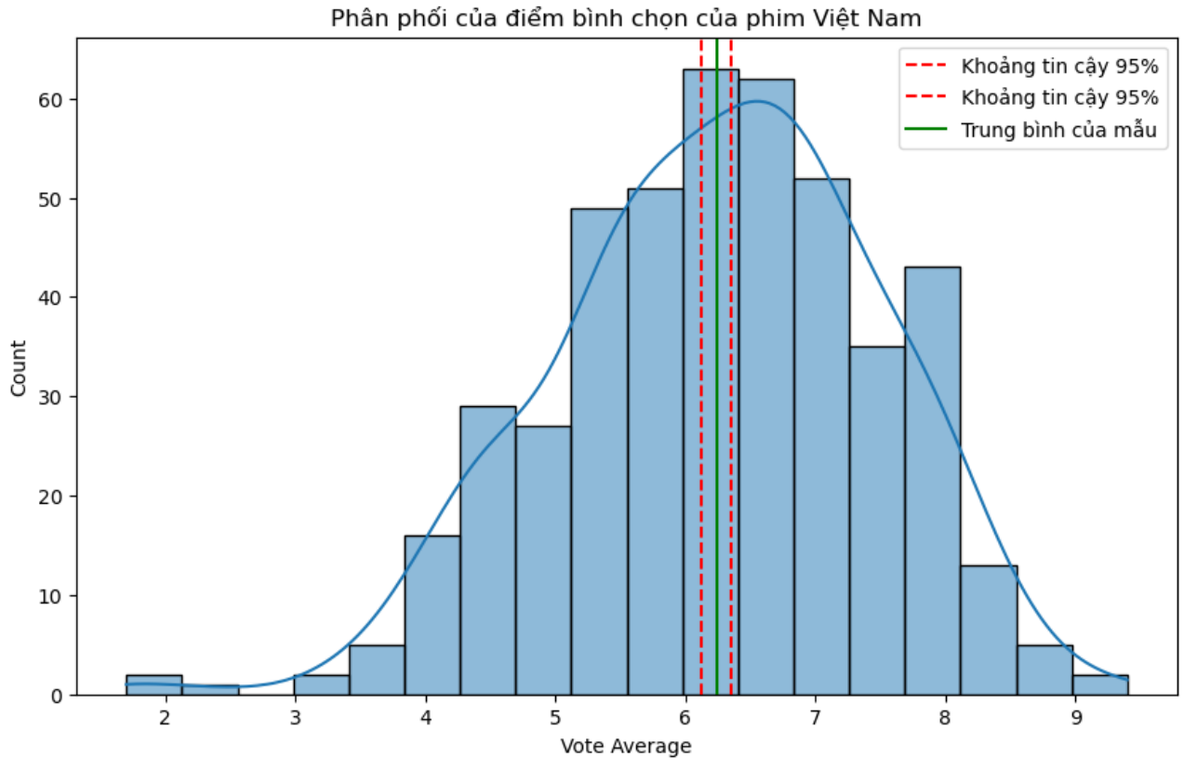
- Thời lượng của các bộ phim Việt Nam có phân phối gần giống với phân phối chuẩn, khoảng giá trị từ 11 đến 179 phút

- Phần lớn các bộ phim Việt có thời lượng từ 75 đến 125 phút. Tuy nhiên vẫn có các trường hợp ngoại lệ với thời lượng rất thấp hoặc rất cao.
- Thời lượng trung bình của các bộ phim Việt Nam từ 96 đến 99 phút với khoảng tin cậy 95%

Vote Average

Bảng 7: Thống kê và khoảng tin cậy 95%

Thống kê	**Giá trị**
Trung bình (mean)	6.15
Độ lệch chuẩn (std)	1.21
Giá trị nhỏ nhất (min)	1.7
Phân vị 25%	5.4
Phân vị 50% (median)	6.2
Phân vị 75%	7.00
Giá trị lớn nhất (max)	9.4
Khoảng tin cậy 95%	(6.03 ; 6.26)



Hình 7: Phân phối của điểm bình chọn phim Việt Nam

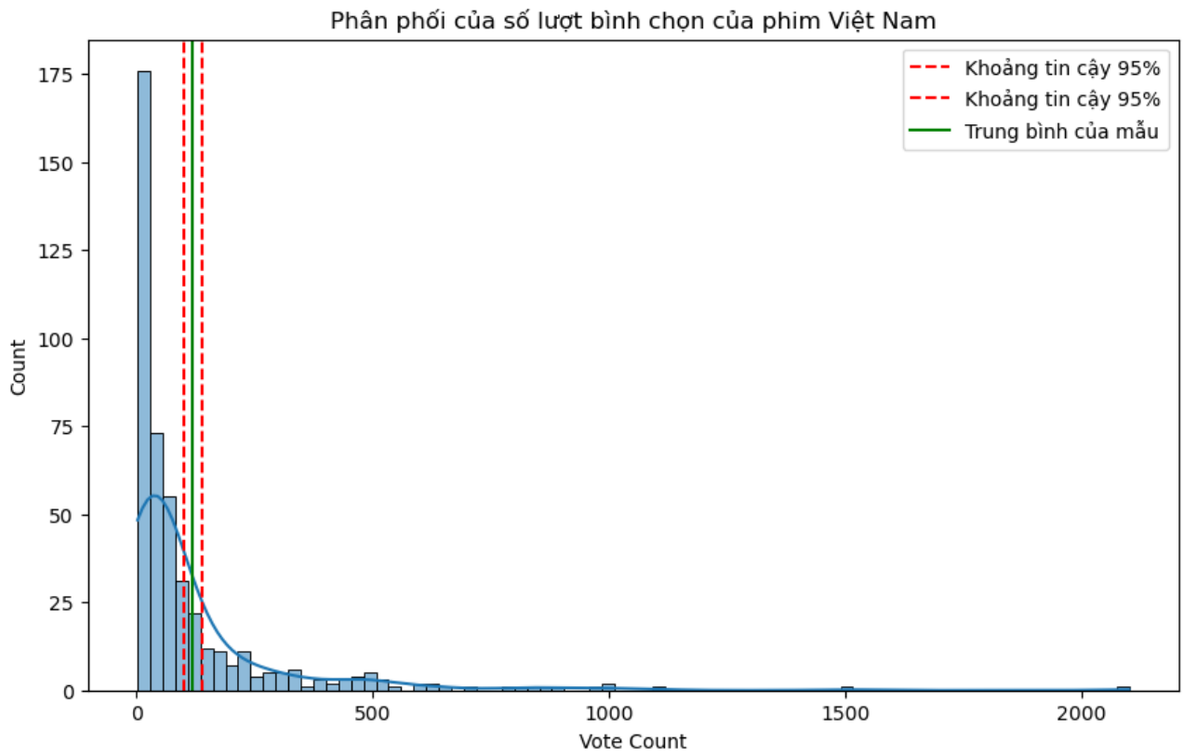
Nhận xét

- Bình chọn trung bình của các bộ phim Việt Nam có phân phối gần giống với phân phối chuẩn, có khoảng giá trị từ 1.7 đến 9.4
- Phần lớn các bộ phim Việt có điểm bình chọn từ 5.5 đến 7.5. Tuy nhiên vẫn có các trường hợp ngoại lệ với điểm bình chọn rất thấp hoặc rất cao.
- Điểm bình chọn trung bình của các bộ phim Việt Nam từ 6.03 đến 6.26 với khoảng tin cậy 95%

Vote Count

Bảng 8: Thống kê và khoảng tin cậy 95%

Thống kê	**Giá trị**
Trung bình (mean)	124.17
Độ lệch chuẩn (std)	211.16
Giá trị nhỏ nhất (min)	2
Phân vị 25%	18.75
Phân vị 50% (median)	49.5
Phân vị 75%	123.25
Giá trị lớn nhất (max)	2100
Khoảng tin cậy 95%	(104.07 ; 144.27)



Hình 8: Phân phối của số lượt bình chọn phim Việt Nam

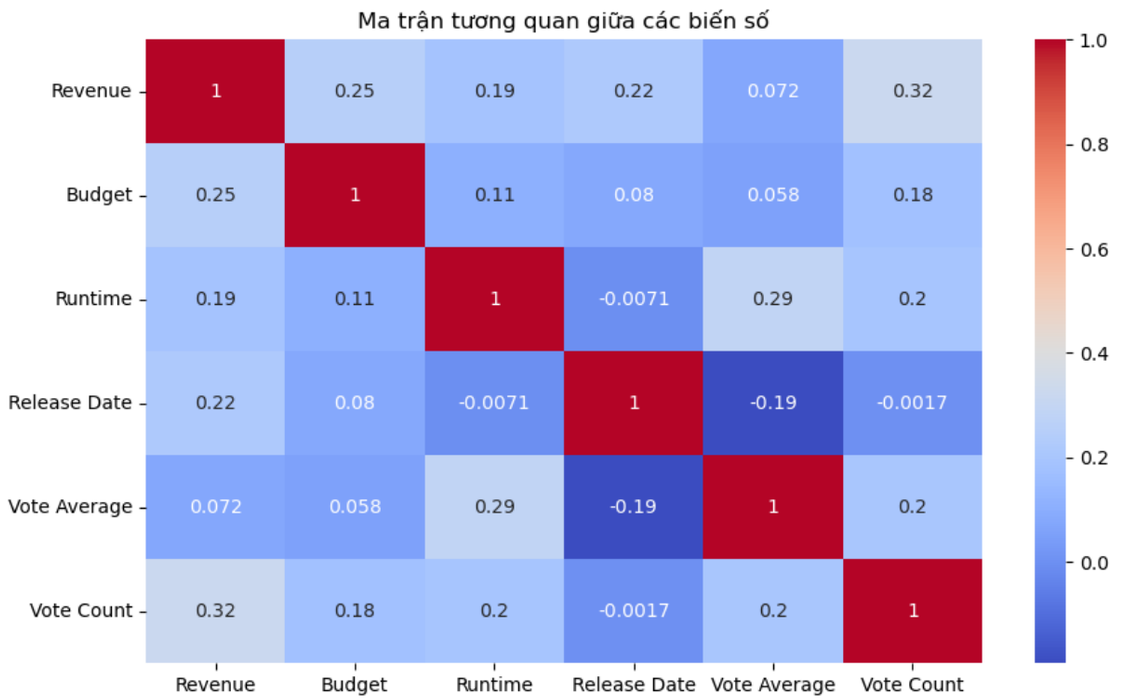
Nhận xét

- Số lượt bình chọn của các bộ phim Việt Nam có phân phối lệch phải, có khoảng giá trị từ 2 đến 2100.

- Phần lớn các bộ phim Việt có số lượt bình chọn từ 10 đến 200. Tuy nhiên vẫn có các trường hợp ngoại lệ với số lượt bình chọn rất thấp hoặc rất cao.
- Số lượt bình chọn trung bình của các bộ phim Việt Nam từ 104 đến 144 với khoảng tin cậy 95%

3.4 Tương quan giữa các biến

Nhóm dùng heatmap để trực quan ma trận hệ số tương quan giữa các biến và dùng scatter plot để trực quan rõ hơn cho từng cặp biến

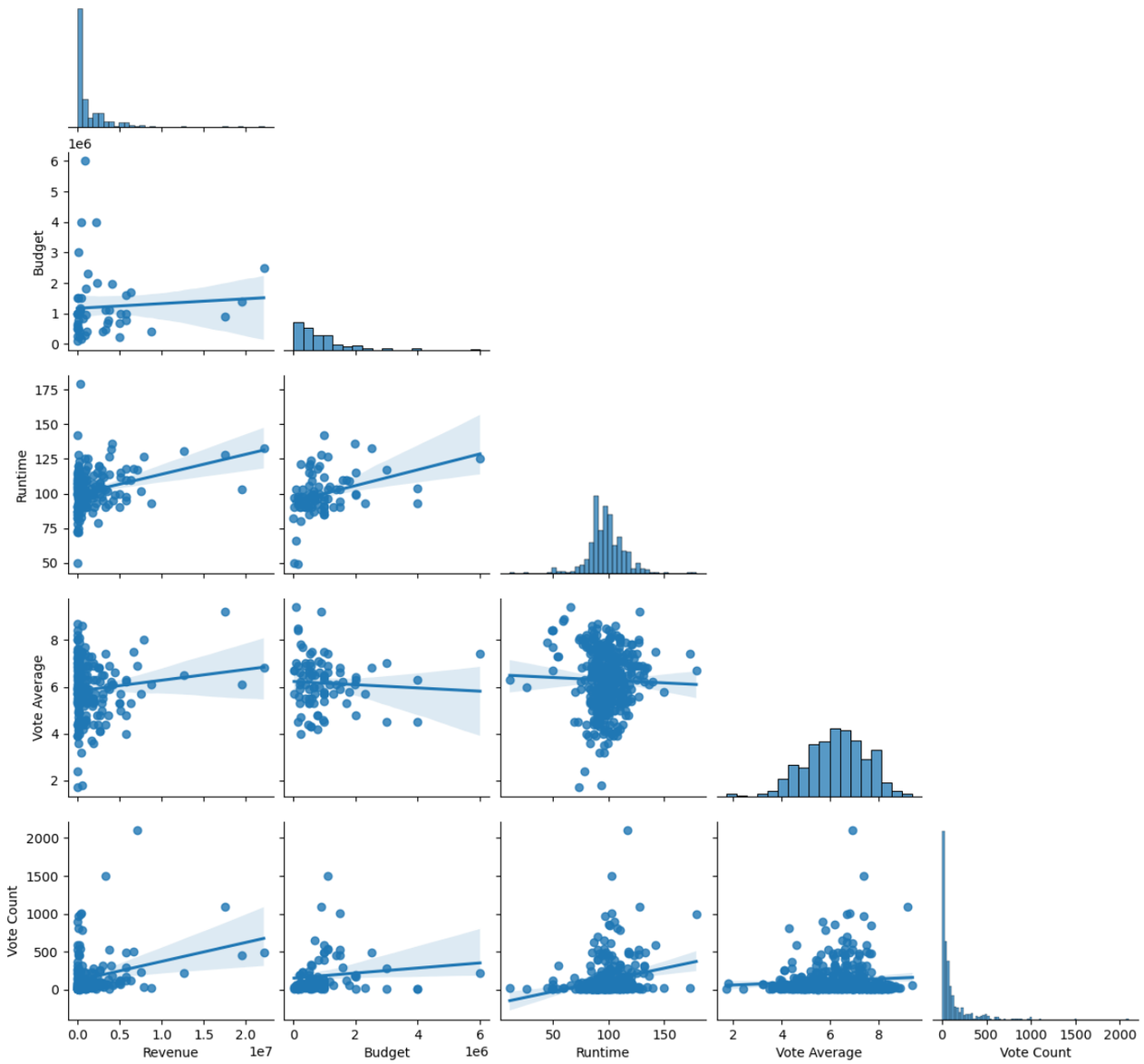


Hình 9: Ma trận hệ số tương quan

Nhận xét

Dựa vào ma trận tương quan (Hình 9) giữa các biến số cũng như biểu đồ phân tán giữa các biến số (Hình 10), ta có thể thấy các biến số không có tương quan nhiều với nhau.

- Đối với ma trận tương quan, các hệ số tương quan đều có độ lớn nhỏ hơn 0.4
- Đối với biểu đồ phân tán và đường thẳng hồi quy, ta cũng không thể thấy được mối tương quan, mối quan hệ rõ ràng giữa các biến số



Hình 10: Biểu đồ phân tán giữa các cặp biến

Chúng ta sẽ phân tích kĩ hơn ở phần tiếp theo

4 Tương quan stat model

4.1 Giới thiệu

Trong báo cáo này, nhóm lần lượt phân tích mối quan hệ giữa các biến mục tiêu là doanh thu (Revenue) và điểm trung bình (Vote Average) với các biến độc lập còn lại. Mục tiêu là xác định liệu có mối quan hệ đáng kể nào giữa doanh thu và [biến độc lập] hay không hoặc điểm trung bình và [biến độc lập], và nếu có, mối quan hệ này mạnh mẽ đến mức nào.

4.2 Phương pháp

Nhóm sử dụng mô hình hồi quy tuyến tính OLS (Ordinary Least Squares) được cung cấp bởi thư viện statsmodels trong Python để ước lượng mối quan hệ. Mô hình được xây dựng như sau:

$$\text{Biến Mục Tiêu} = \beta_0 + \beta_1 \times [\text{Biến Độc Lập}] + \varepsilon \tag{1}$$

$$\text{Trong đó}, (\beta_0) \text{ là hệ số chặn}, (\beta_1) \text{ là hệ số hồi quy cho [biến độc lập], và } (\varepsilon) \text{ là sai số ngẫu nhiên.} \tag{2}$$

4.3 Kết quả

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                Revenue   R-squared (uncentered):                  0.773
Model:                            OLS   Adj. R-squared (uncentered):             0.773
Method:                 Least Squares   F-statistic:                             1880.
Date:                Wed, 08 May 2024   Prob (F-statistic):                       0.00
Time:                        21:41:10   Log-Likelihood:                        -44295.
No. Observations:                2211   AIC:                                 8.860e+04
Df Residuals:                    2207   BIC:                                 8.862e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Vote Average -1.696e+05   1.56e+06     -0.109      0.913   -3.23e+06    2.89e+06
Vote Count    2.538e+04    860.065     29.507      0.000    2.37e+04    2.71e+04
Budget           2.0404      0.065     31.296      0.000       1.913       2.168
Runtime       -1.43e+05   9.72e+04     -1.470      0.142   -3.34e+05    4.77e+04
==============================================================================
Omnibus:                     1869.764   Durbin-Watson:                   2.028
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           114483.019
Skew:                           3.603   Prob(JB):                         0.00
Kurtosis:                      37.508   Cond. No.                     3.83e+07
==============================================================================
```

Hình 11: Mô hình dự đoán Revenue

Tổng Quan Mô Hình: (hình 11)

- **R-squared (uncentered)**: 0.773. Mô hình giải thích được khoảng 77.3% biến thiên của biến phụ thuộc "Revenue"mà không cần một hằng số. Đây là một tỷ lệ khá cao, cho thấy mô hình có khả năng giải thích tốt dữ liệu.
- **Adj. R-squared (uncentered)**: 0.773. Điều chỉnh R bình phương cũng rất cao và gần với R bình phương, cho thấy số lượng biến trong mô hình được chọn một cách hợp lý.
- **F-statistic:** 1880 với **Prob (F-statistic)** là 0.00, chỉ ra rằng mô hình của bạn rất có ý nghĩa thống kê, với khả năng rất thấp là kết quả này xảy ra do ngẫu nhiên.

Phân Tích Các Biến:

- **Vote Average**: Không có ý nghĩa thống kê (p-value = 0.913), cho thấy Vote Average có thể không ảnh hưởng đáng kể đến Revenue. Hệ số âm cũng không phản ánh một mối quan hệ tích cực giữa Vote Average và Revenue.
- **Vote Count**: Có ảnh hưởng tích cực và đáng kể đến doanh thu ($p < 0.001$). Điều này cho thấy số lượng đánh giá càng cao, doanh thu càng tăng.
- **Budget**: Có ảnh hưởng tích cực và rất mạnh mẽ đến doanh thu ($p < 0.001$). Điều này cho thấy ngân sách lớn hơn có thể dẫn đến doanh thu cao hơn.
- **Runtime**: Không có ý nghĩa thống kê do (p-value = 0.142), cho thấy Runtime có thể không là yếu tố quan trọng ảnh hưởng đến Revenue trong mô hình này.

Kết Luận:

- Mô hình cho thấy "Vote Count", và "Budget"là hai yếu tố có ảnh hưởng đáng kể và tích cực đến "Revenue". Điều này có thể hữu ích cho các nhà sản xuất phim khi xem xét đến việc tăng cường chất lượng nội dung, khuyến khích người xem bình chọn, và đầu tư hợp lý vào ngân sách sản xuất để cải thiện doanh thu.
- "Runtime"không chứng minh được mối quan hệ đáng kể với "Revenue"trong mô hình này, điều này có thể gợi ý rằng thời lượng phim không phải là yếu tố quyết định chính cho doanh thu của phim.
- Mô hình có độ tin cậy cao với R bình phương và F-statistic cho thấy mô hình có khả năng giải thích tốt dữ liệu và các biến độc lập được chọn có ý nghĩa.

```
                                 OLS Regression Results
==============================================================================
Dep. Variable:           Vote Average   R-squared (uncentered):              0.826
Model:                            OLS   Adj. R-squared (uncentered):         0.825
Method:                 Least Squares   F-statistic:                         539.3
Date:                Wed, 08 May 2024   Prob (F-statistic):              6.44e-171
Time:                        21:41:10   Log-Likelihood:                    -1071.9
No. Observations:                 458   AIC:                                 2152.
Df Residuals:                     454   BIC:                                 2168.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Vote Count     0.0019      0.001      3.022      0.003       0.001       0.003
Revenue    -6.179e-08      6e-08     -1.031      0.303    -1.8e-07     5.6e-08
Budget      1.464e-07   2.14e-07      0.684      0.494   -2.74e-07    5.67e-07
Runtime        0.0569      0.002     36.627      0.000       0.054       0.060
==============================================================================
Omnibus:                       29.278   Durbin-Watson:                   1.679
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.606
Skew:                           0.504   Prob(JB):                     1.52e-09
Kurtosis:                       4.055   Cond. No.                     3.07e+04
==============================================================================
```

Hình 12: Mô hình dự đoán Vote Average

Tổng Quan Mô Hình: (hình 12)

- **R-squared (uncentered)**: 0.826. Điều này cho thấy mô hình của bạn giải thích được khoảng 82.6% biến thiên của biến phụ thuộc ("Vote Average") mà không cần một hằng số. Đây là một tỷ lệ khá cao, cho thấy mô hình có khả năng giải thích tốt dữ liệu.
- **Adj. R-squared (uncentered)**: 0.825. Điều chỉnh R bình phương cũng rất cao, gần với R bình phương, cho thấy số lượng biến trong mô hình được chọn một cách hợp lý và không có sự thừa thãi đáng kể.
- **F-statistic**: 538.3 với **Prob (F-statistic)** là 6.44e-171, chỉ ra rằng mô hình của bạn rất có ý nghĩa thống kê, với khả năng rất thấp là kết quả này xảy ra do ngẫu nhiên.

Phân Tích Các Biến:

- **Vote Count**: Có mối quan hệ tích cực và có ý nghĩa thống kê với Vote Average (p-value = 0.003). Điều này cho thấy số lượng phiếu bầu là một yếu tố quan trọng, với mỗi đơn vị tăng trong Vote Count, Vote Average dự kiến tăng thêm khoảng 0.0019 đơn vị.
- **Revenue**: Không có ý nghĩa thống kê (p-value = 0.303), cho thấy Revenue có thể không ảnh hưởng đáng kể đến Vote Average trong mô hình này
- **Budget**: Cũng không có ý nghĩa thống kê (p-value = 0.494), cho thấy Budget có thể không là yếu tố quan trọng ảnh hưởng đến Vote Average trong mô hình này.
- **Runtime**: Có mối quan hệ tích cực và có ý nghĩa thống kê với Vote Average (p-value < 0.001). Điều này cho thấy Runtime là một yếu tố quan trọng, với mỗi đơn vị tăng trong Runtime, Vote Average dự kiến tăng thêm khoảng 0.0569 đơn vị.

Kết Luận:

- Mô hình cho thấy "Vote Count"và "Runtime"là hai yếu tố có ảnh hưởng đáng kể và tích cực đến "Vote Average". Điều này có thể hữu ích cho các nhà sản xuất phim khi xem xét đến việc tăng cường chất lượng nội dung và khuyến khích người xem bình chọn để cải thiện đánh giá trung bình của phim.
- "Revenue"và "Budget"không chứng minh được mối quan hệ đáng kể với "Vote Average"trong mô hình này, điều này có thể gợi ý rằng việc đầu tư nhiều hơn vào ngân sách sản xuất hoặc việc thu về doanh thu cao không nhất thiết dẫn đến việc cải thiện đánh giá của phim từ góc độ của người xem.
- Mô hình có độ tin cậy cao với R bình phương và F-statistic cho thấy mô hình có khả năng giải thích tốt dữ liệu và các biến độc lập được chọn có ý nghĩa.

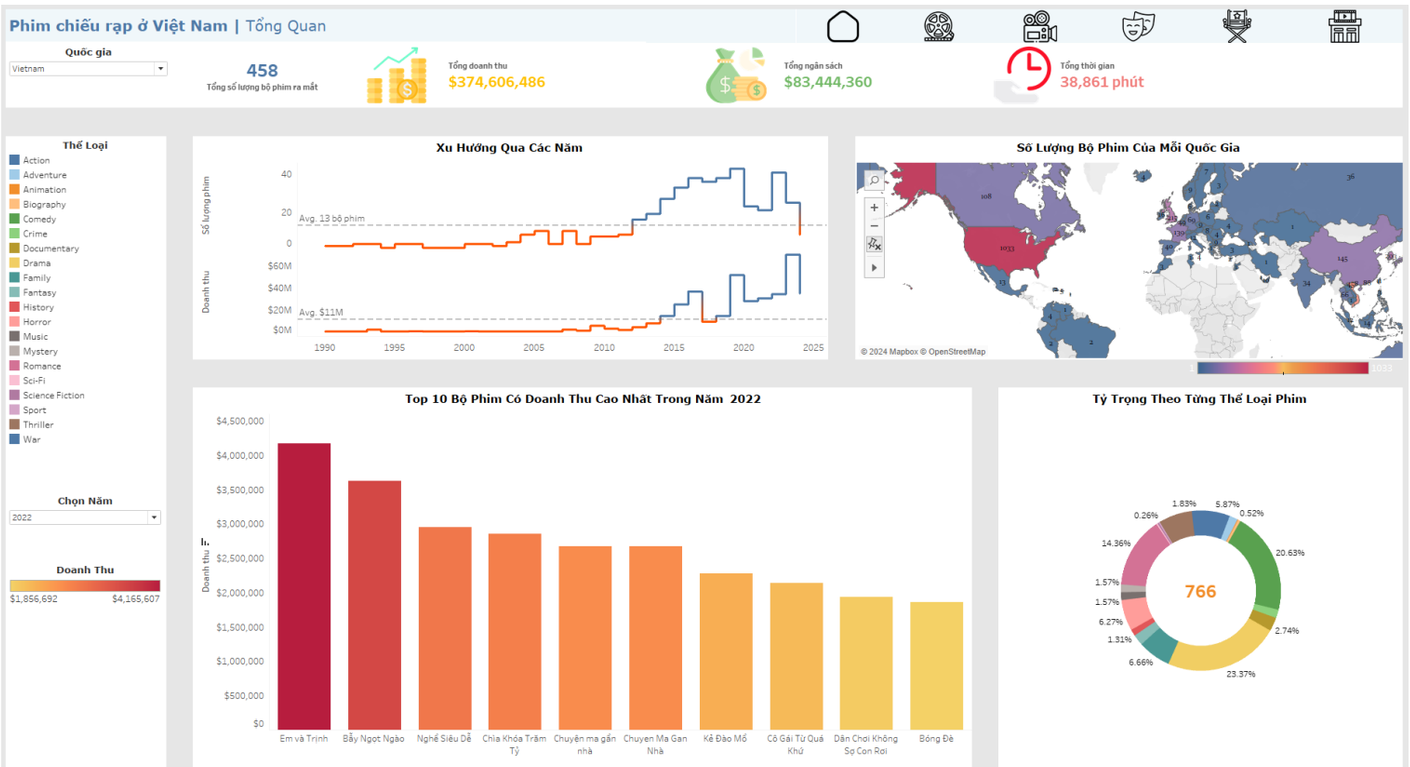
5 Trực quan hóa dữ liệu

Nhóm sẽ sử dụng công cụ Tableau để trực quan cho đồ án này. Sản phẩm cuối cùng mà nhóm muốn thể hiện gồm 6 dashboard chính cùng nhau tạo nên một cái nhìn toàn diện về ngành điện ảnh, trong đó các dashboard sẽ lần lượt đi sâu phân tích từng khía cạnh cụ thể như:

- Tổng quan về bộ dữ liệu
- Các bộ phim
- Diễn viên
- Đạo diễn
- Công ty sản xuất phim
- Thể loại phim

5.1 Dashboard tổng quan

Dashboard đầu tiên giúp người dùng có cái nhìn tổng quan về các bộ phim chiếu rạp ở Việt Nam. Đây là giao diện của dashboard:



Hình 13: Dashboard tổng quan

Bố cục của dashboard gồm:

- Bốn con số đại diện mô tả sơ lược về bộ dữ liệu
  - Tổng số bộ phim
  - Tổng doanh thu trên toàn thế giới của các bộ phim (đô la)
  - Tổng ngân sách của các bộ phim (đô la)
  - Tổng thời lượng của các bộ phim
- Bốn loại biểu đồ khác nhau giúp người dùng hiểu được tổng quan về các bộ phim chiếu rạp
  - **Hai biểu đồ đường**: thể hiện xu hướng theo thời gian qua doanh thu và số lượng phim ra

mắt của các năm. Trục hoành của cả hai biểu đồ đường đều là trục thời gian (năm) và trục tung lần lượt biểu thị số lượng phim và doanh thu. Với một đường thẳng nét đứt của mỗi biểu đồ chính là đường thẳng cho giá trị trung bình, nếu giá trị của năm đó nhỏ hơn giá trị trung bình thì sẽ hiển thị màu đỏ, nếu lớn hơn thì hiển thị màu xanh

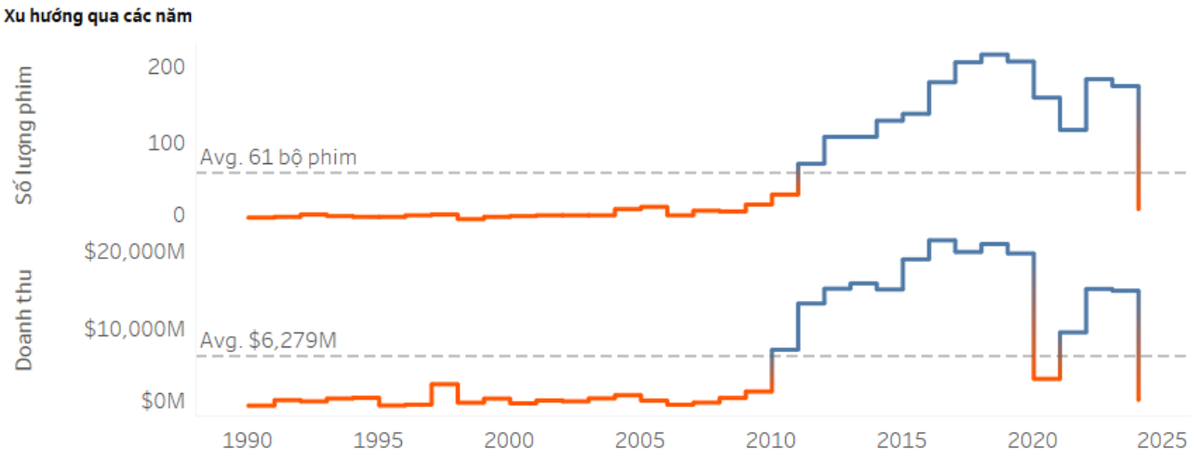
- **Biểu đồ dạng map**: thể hiện số lượng các bộ phim theo các nước sản xuất. Màu của quốc gia có nhiều phim chiếu sẽ càng đậm.
- **Biểu đồ cột** thể hiện 10 phim có doanh thu cao nhất của một đất nước theo năm
- **Biểu đồ donut** thể hiện các thể loại phim phổ biến. Số lượng phim theo mỗi thể loại được chuyển thành tỉ lệ phần trăm để trực quan. Con số ở giữa chính là tổng số thể loại

• Ở bên trái của dashboard có hai ô để người dùng có thể chọn quốc gia và năm của các bộ phim để xem xét và phân tích kỹ hơn với các bộ phim của các đất nước và năm cụ thể.

• Từ dashboard tổng quan ta có thể quan sát và đưa ra các phân tích, nhận xét sau:

- Dataset có tổng cộng **2193** bộ phim đã từng chiếu rạp tại Việt Nam trong đó có **453** bộ phim Việt. Trên toàn thế giới, các bộ phim được đầu tư với tổng ngân sách hơn **67 tỉ đô** và đem lại **226 tỉ đô** doanh thu với tổng thời lượng phim hơn **225 nghìn phút**. Các bộ phim Việt Nam được đầu tư với ngân sách hơn **83 triệu đô** và mang lại tổng doanh thu hơn **374 triệu đô** với thời lượng hơn **38 nghìn phút**. Các con số cho thấy ngành công nghiệp điện ảnh trên thế giới là một ngành rất quan trọng và mang lại khoản lợi nhuận không nhỏ đến các nền kinh tế. Với Việt Nam thì đây là một ngành đầy tiềm năng và vẫn không ngừng phát triển để vươn ra thế giới.
- Hai biểu đồ đường (Hình 15 và 14) l không chỉ thể hiện doanh thu và số lượng phim ra đời ở mỗi năm của thế giới và Việt Nam mà còn phần nào cho thấy tình hình phát triển của ngành điện ảnh của thế giới và ngành điện ảnh nước nhà trong thị trường Việt Nam.
  - ∗ Các bộ phim trên thế giới: giai đoạn những năm 1990 - 2000 chỉ có vài phim ở thế giới

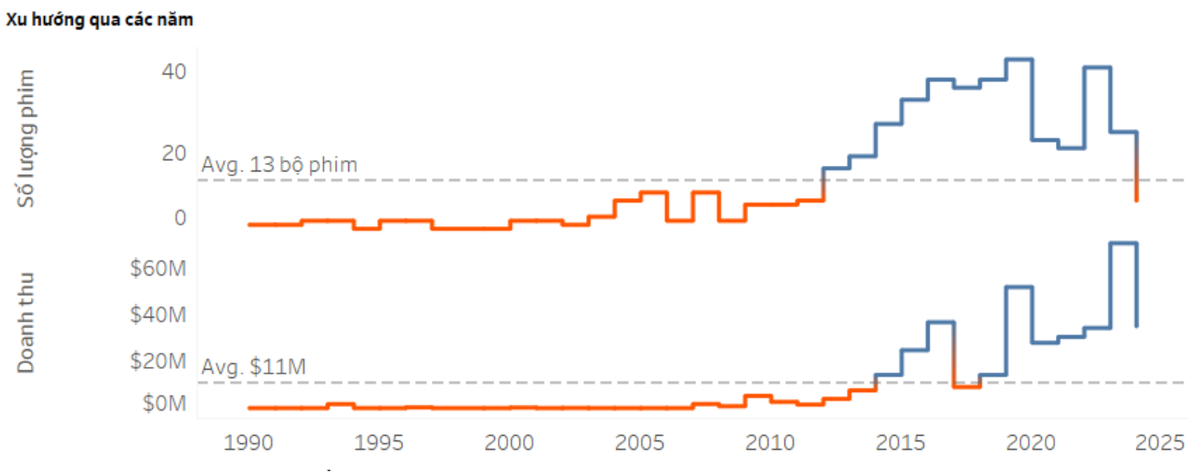
và doanh thu của các bộ này đều dưới mức trung bình, chỉ khoảng 1-2 tỷ đô mỗi năm. Năm 2000-2010 đã có sự gia tăng của các bộ phim nước ngoài chiếu ở Việt Nam nhưng doanh thu thì không tăng đáng kể. Từ 2010 đã có sự nhảy vọt, phát triển mạnh mẽ làn sóng phim nước ngoài chiếu rạp ở Việt Nam, số lượng các bộ phim tăng nhanh gấp nhiều lần qua các năm và doanh thu cũng theo đó mà đột biến. Đỉnh cao là năm 2018 với 215 phim mang về doanh thu toàn cầu hơn 20 tỷ đô. Giai đoạn 2020- 2021 nền điện ảnh thế giới có thể đã chịu một ảnh hưởng không hề nhỏ từ đại dịch Covid cũng như các yếu tố khác nên số lượng các bộ phim đã giảm rõ rệt, kéo theo doanh thu suy giảm theo đó, ít nhất là năm 2021 chỉ khoảng 117 bộ phim và doanh thu năm 2020 chưa đến 3.5 tỷ đô ,thấp nhất trong giai đoạn này. Từ 2021 đến nay, toàn cầu đã dần ổn định hơn, hồi phục dần được tình hình kinh tế cũng như là ngành điện ảnh nên số lượng bộ phim và doanh thu đã tăng trở lại.



Hình 14: Biểu đồ đường thể hiện số lượng và doanh thu của tất cả bộ phim chiếu ở Việt Nam

* Các bộ phim Việt Nam: Giai đoạn 1990 - 2000 ở các bộ phim Việt Nam chỉ có một vài phim và doanh thu cũng dưới mức trung bình chỉ khoảng 2 triệu đô, chủ yếu đều không có số liệu về doanh thu. Từ năm 2000 trở đi nền điện ảnh Việt đã có dấu hiệu phát triển, số lượng các bộ phim tăng rõ, và cũng đã có số liệu về doanh thu, rơi vào khoảng 1-2 triệu đô mỗi năm số. Đặc biệt từ khoảng năm 2011 trở đi có thể nói là sự phát triển bùng nổ, mạnh mẽ của nền điện ảnh nước nhà khi số lượng phim Việt được sản xuất tăng

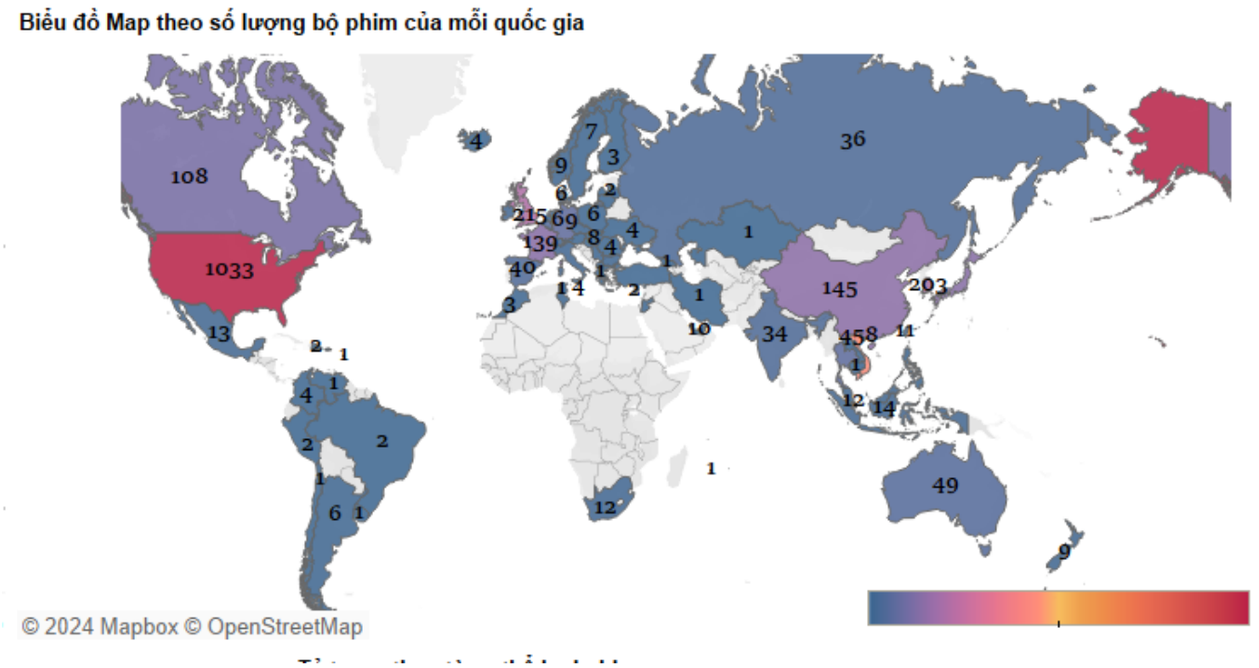
trưởng đột biến. Đồng thời doanh thu cũng tăng trưởng theo gấp nhiều lần, đỉnh cao vào năm 2019 với 43 phim được ra mắt và đạt được doanh thu 52 triệu đô trong năm. Cũng như thế giới, năm 2020 - 2021 đất nước ta cũng như nền điện ảnh cũng bị ảnh hưởng không hề nhỏ, số lượng các bộ phim và doanh thu đem về giảm đáng kể. Từ năm 2022, thị trường phim Việt bắt đầu phục hồi và sôi động trở lại, số lượng và doanh thu các bộ phim tăng trở lại. Đặc biệt năm 2023 đã có tổng doanh thu cao nhất từ trước đến giờ là 70 triệu đô trong năm.



Hình 15: Biểu đồ đường thể hiện số lượng và doanh thu của các bộ phim Việt

– Biểu đồ map (Hình 16) thể hiện nguồn gốc các bộ phim chiếu tại Việt Nam có phân bố không đều nhưng là một bức tranh phản ánh sự đa dạng và đa quốc gia trong nguồn cung phim. Mỹ, với hơn 1000 bộ phim, chiếm gần một nửa tổng số, rõ ràng thể hiện sự thống trị của ngành công nghiệp điện ảnh lớn nhất thế giới. Sự ảnh hưởng của Hollywood không chỉ phản ánh sức mạnh về sản xuất và tiếp thị mà còn phản ánh sự phổ biến của văn hóa Mỹ trên toàn cầu, bao gồm cả Việt Nam.

  Bên cạnh đó, các quốc gia khác như Canada, Anh, Pháp, Trung Quốc, Hàn Quốc và Nhật Bản cũng có sự đóng góp đáng kể, với số lượng phim dao động từ 100 đến 200 bộ mỗi nước. Điều này phản ánh sự đa dạng về nội dung và phong cách của các bộ phim được chiếu tại Việt Nam, cũng như sự phát triển mạnh mẽ của ngành công nghiệp điện ảnh ở những quốc



Hình 16: Biểu đồ map thể hiện số lượng bộ phim của mỗi quốc gia

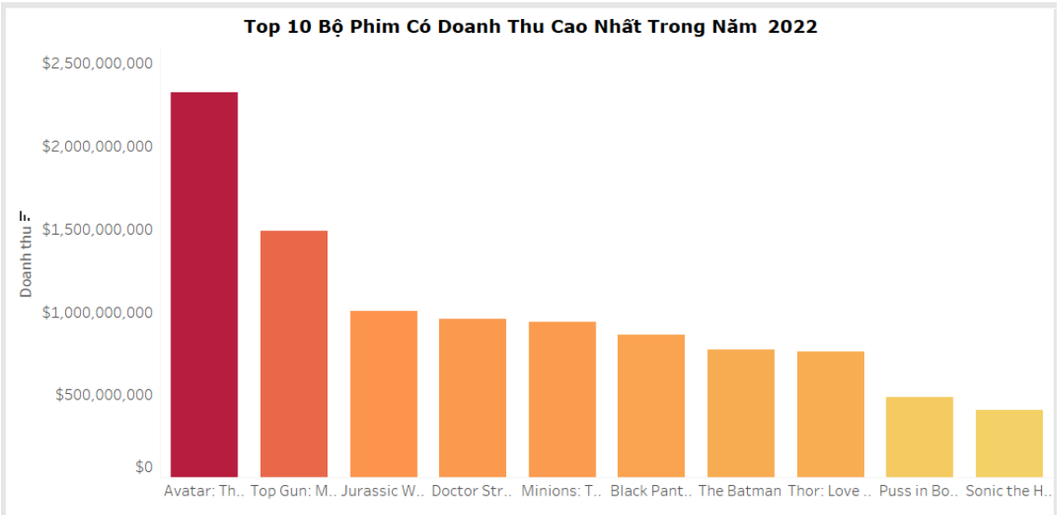
gia này.

Tuy nhiên, không thể phủ nhận sự tiềm năng và nỗ lực của điện ảnh Việt Nam với hơn 450 bộ phim sản xuất trong thời gian này. Mặc dù con số này ít so với các nền điện ảnh lớn, nhưng nó vẫn thể hiện sự phát triển và tiềm năng của ngành công nghiệp điện ảnh nội địa. Sự đa dạng trong nguồn cung phim tại Việt Nam không chỉ làm phong phú lựa chọn cho khán giả mà còn thúc đẩy sự phát triển và cạnh tranh trong ngành điện ảnh quốc gia.

– Biểu đồ cột đứng giúp ta biết được các bộ phim có doanh thu cao nhất qua từng năm:

  * **Các bộ phim nước ngoài**: Trải qua năm 2021, bộ phim "Spider-Man: No Way Home"từ vũ trụ điện ảnh Marvel, với sự tham gia của bộ ba người nhện, đã ghi dấu ấn mạnh mẽ khi chiếm lĩnh doanh số phòng vé toàn cầu gần 2 tỷ đô la, vượt xa tổng doanh thu của hai bộ phim đứng ở vị trí 2 và 3 (No Time To Die và F9). Năm 2022, sự trở lại của bom tấn "Avatar"với phần tiếp theo "Avatar: The Way of Water"đã gây sốc với doanh thu cao nhất lịch sử phòng vé, vượt qua con số 2.3 tỷ đô. Năm 2023 là cuộc đua căng thẳng giữa các bộ phim nổi tiếng nhưng chiến thắng cuối cùng lại thuộc về "The Super

Mario Bros"với doanh thu 1.3 tỷ đô, vượt xa phim đứng hạng hai là "Oppenheimer một bộ phim đạt giải Oscar. Đầu năm 2024, "Dune: Part Two"tiếp tục thống trị là phần tiếp theo của bộ phim "Dune"(2021), được đánh giá cao về mặt nội dung và hình ảnh.

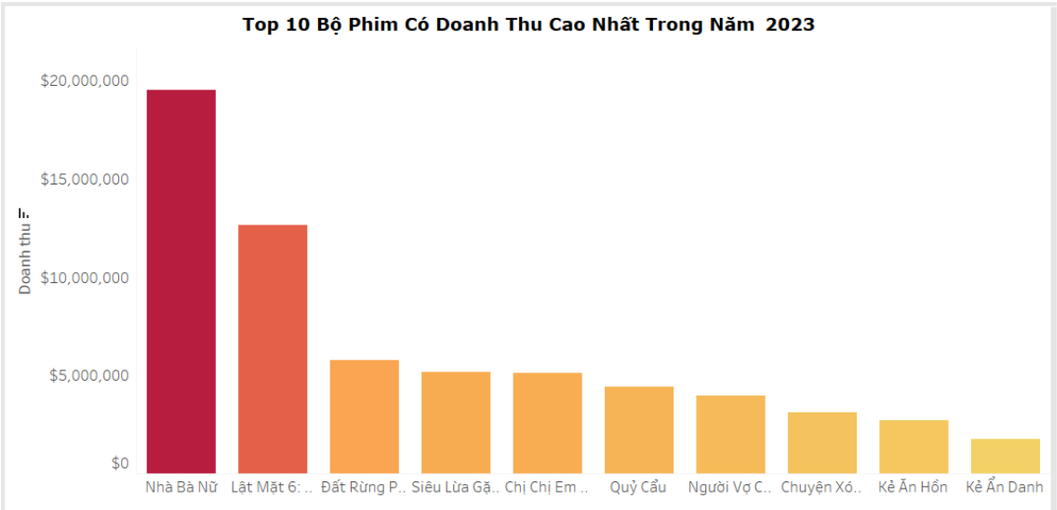


Hình 17: Top 10 bộ phim nước ngoài có doanh thu cao nhất 2022

* **Phim Việt Nam**: Trong năm 2021, bộ phim "Bố Già"của Trấn Thành, dù là lần đầu đảm nhận vai trò đạo diễn, đã thu hút sự chú ý của khán giả trong nước với doanh thu kỷ lục hơn 17 triệu đô, gần ba lần bộ phim "Lật mặt 5: 48h"của đạo diễn Lý Hải. Phim "Em và Trịnh", lấy cảm hứng từ cuộc đời của cố nhạc sĩ Trịnh Công Sơn, đã đạt doanh thu hơn 4 triệu đô. Năm 2023, hai bộ phim của Trấn Thành và Lý Hải một lần nữa đối đầu và chiến thắng thuộc về "Nhà Bà Nữ"với doanh thu hơn 19 triệu đô, trong khi "Lật Mặt 6: Tấm Vé Định Mệnh"thu về hơn 12 triệu đô. Tính đến đầu năm 2024, Trấn Thành lại một lần nữa dẫn đầu với bộ phim "Mai", chiếu vào dịp Tết và mang về doanh thu cao nhất từ trước tới giờ của phim Việt, với hơn 22 triệu đô. Bộ phim "Lật Mặt 7: Một Điều Ước"vẫn đang chiếu rạp và thu về doanh thu gần 8 triệu đô, đứng thứ hai trong bảng xếp hạng

– Biểu đồ donut cho ta thấy tỷ trọng các thể loại phổ biến theo tần suất xuất hiện:

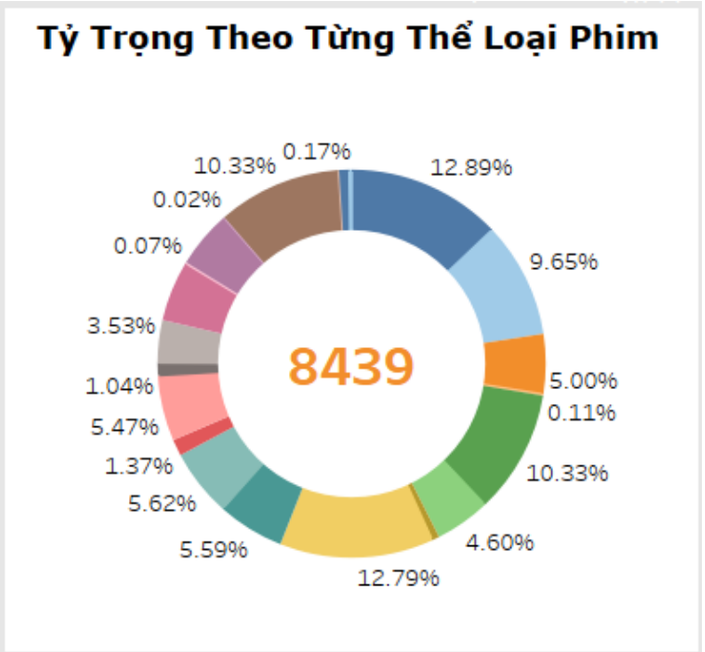
* **Tất cả các bộ phim** : Trong hơn 8400 tần suất của các thể loại, chúng ta nhận thấy có các



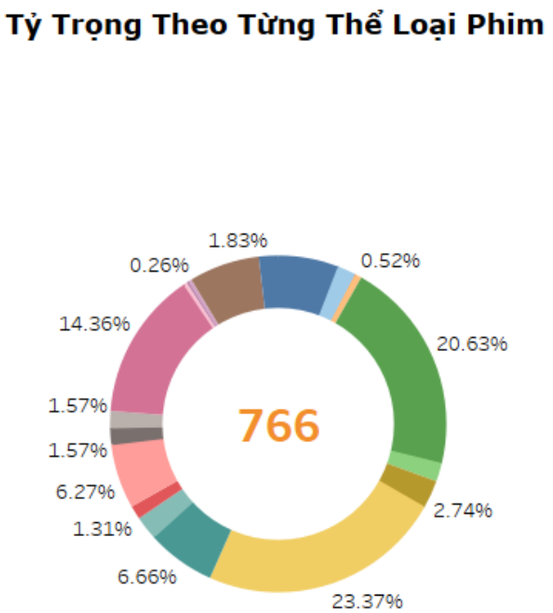
Hình 18: Top 10 bộ phim Việt có doanh thu cao nhất 2023

thể loại nổi bật, chiếm tỷ trọng nhiều hơn hẳn, đứng đầu là Action (12.9%) và Drama (12.8%), theo sau là các thể loại quen thuộc được nhiều khán giả yêu thích Comedy (8.72%), Thriller(10.3%) và Family (5.6%)

* **Các bộ phim Việt Nam**: Khác với thế giới, thể loại phổ biến nhất trong phim Việt là Drama với tỷ trọng 23.3%, gần 1 phần 4 số tần suất xuất hiện. Bên cạnh đó cũng có các thể loại tương đồng với thế giới như Comedy (20.63%), Romance (14,36%), Family (6.6%)



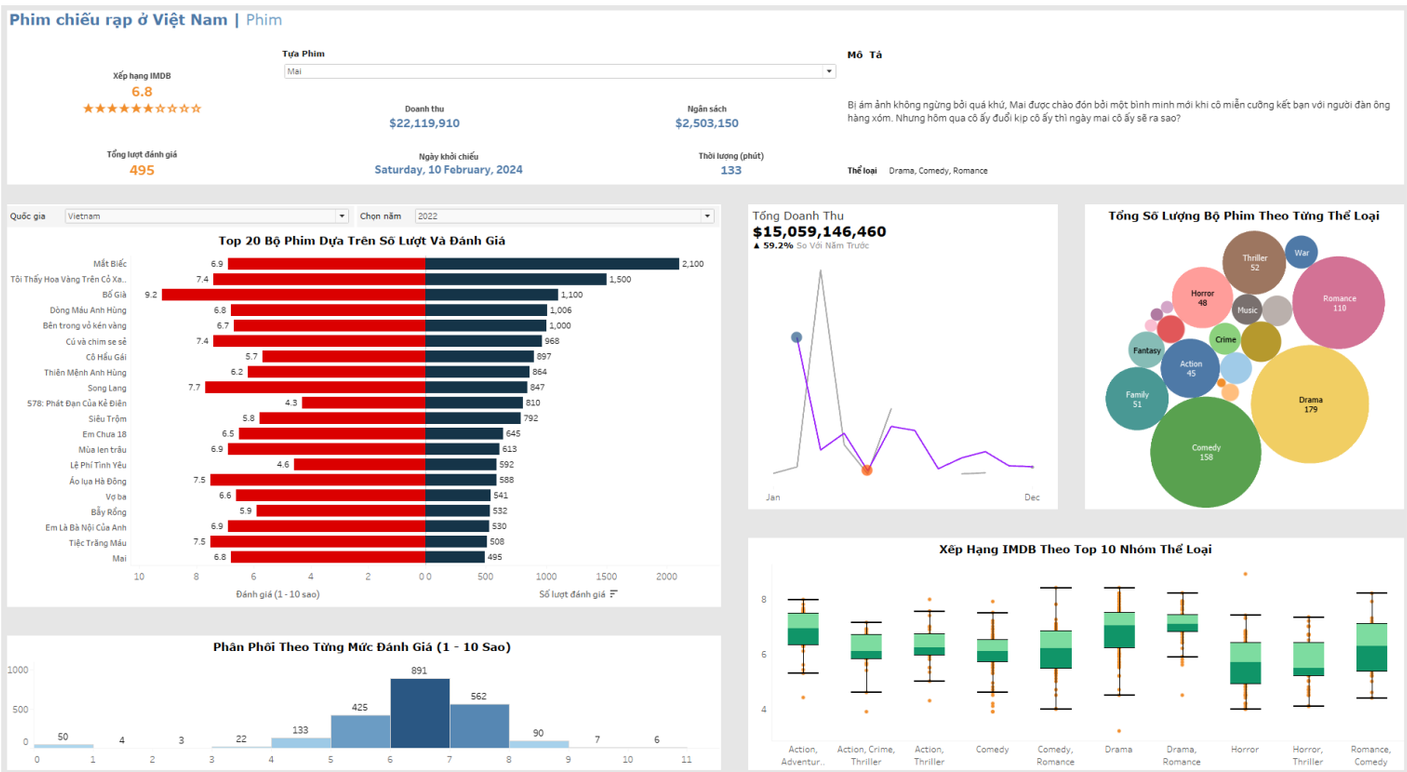
Hình 19: Tỷ trọng các thể loại của tất cả bộ phim



Hình 20: Tỷ trọng thể loại của các bộ phim Việt

5.2 Movie Dashboard

Trong phần này, chúng ta sẽ khám phá và phân tích dữ liệu trực quan về các bộ phim chiếu rạp ở Việt Nam, tập trung chủ yếu vào các yếu tố quan trọng của một bộ phim như doanh thu, ngân sách, điểm đánh giá từ khán giả,... Dashboard này sẽ không chỉ cung cấp cái nhìn tổng quan mà còn hướng tới việc hiểu sâu hơn về các yếu tố ảnh hưởng đến thành công của một bộ phim, cũng như nhận biết một số xu hướng mới nổi bật trong ngành công nghiệp điện ảnh.



Hình 21: Movie Dashboard

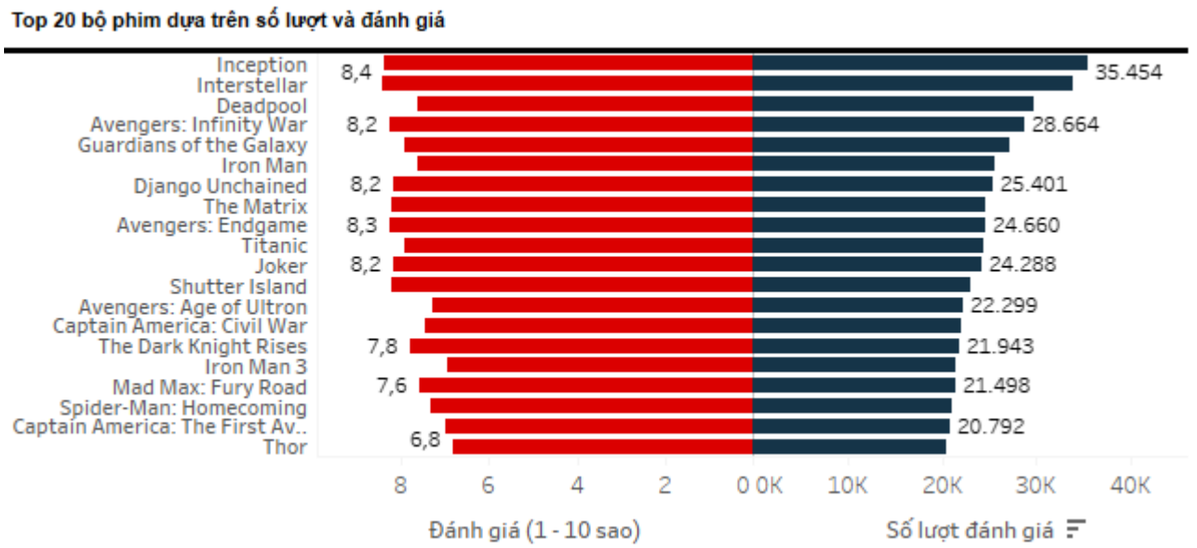
Các thông tin cơ bản của một bộ phim được thể hiện bao gồm:

- **Tựa phim:** Tựa đề của bộ phim không chỉ là cái tên đơn thuần mà còn là điểm khởi đầu để thu hút sự chú ý của khán giả. Một tựa đề đầy ấn tượng và thu hút có thể gợi lên sự tò mò và mong đợi từ phía người xem.
- **Xếp hạng IMDB:** đây là một chỉ số quan trọng để đo lường sự thành công và chất lượng của một bộ phim. Được tính dựa trên đánh giá từ cộng đồng người xem (trên thang đo từ 1 - 10 sao),

điểm số càng cao thường đi kèm với sự đánh giá tích cực và sự ủng hộ mạnh mẽ từ phía khán giả.

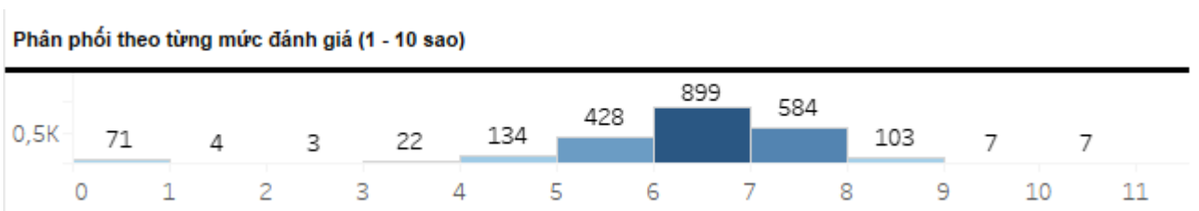
- **Tổng lượt đánh giá:** Số lượng đánh giá cho thấy mức độ phổ biến và sự quan tâm của khán giả đối với bộ phim. Số lượt đánh giá lớn có thể tạo ra một hình ảnh rõ ràng về sự phản ứng và tiếp nhận từ cộng đồng.
- **Doanh thu:** Doanh thu là một chỉ số quan trọng cho thấy sức hút và thành công thương mại của bộ phim. Sự kết hợp giữa doanh thu và ngân sách cung cấp cái nhìn về lợi nhuận và hiệu quả của dự án.
- **Ngân sách:** Ngân sách sản xuất là một yếu tố quyết định quan trọng đối với quyết định về phong cách sản xuất, đội ngũ diễn viên và chất lượng kỹ thuật của bộ phim.
- Bên cạnh đó có thể kể đến các thông tin cũng không kém phần quan trọng khác như: **Ngày khởi chiếu**, **Thời lượng**, **Thể loại** và **Mô tả** của bộ phim.

Để có một cái nhìn tổng quan và chi tiết hơn, thông qua việc sử dụng dữ liệu và các biểu đồ trực quan, mỗi biểu đồ đều mang lại một góc nhìn độc đáo và cung cấp thông tin quan trọng về các yếu tố ảnh hưởng đến sự thành công của các bộ phim. Chúng ta sẽ đi sâu vào từng biểu đồ để phân tích và rút ra những kết luận có ích, từ việc nhận biết các xu hướng đến đánh giá mức độ ảnh hưởng của các yếu tố khác nhau đối với doanh thu và sự phổ biến của các bộ phim.



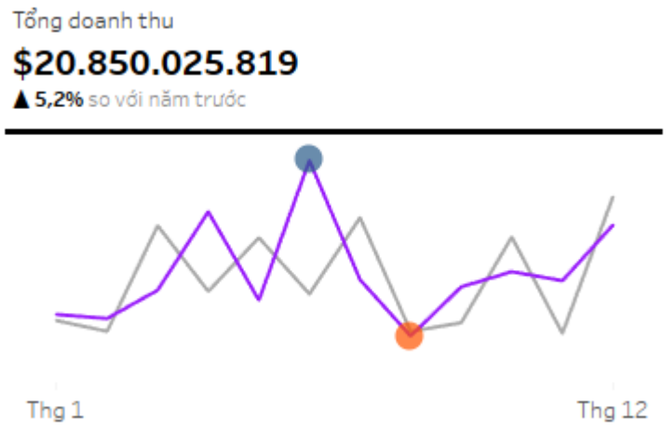
Biểu đồ cột ngang này giúp chúng ta nhìn nhận những bộ phim nào đang được đánh giá cao nhất trên IMDB và thu hút được nhiều sự quan tâm của khán giả dựa trên tổng lượt đánh giá và điểm số trung bình.

Thông qua biểu đồ trên, chúng ta có thể dễ dàng nhận ra các bộ phim nổi bật nhất. Ví dụ, **Inception** là bộ phim đứng đầu với điểm số trung bình cao nhất là **8.4 sao** và thu hút được một lượng lớn lượt đánh giá, lên đến 35,454. Điều này cho thấy rằng Inception không chỉ được công nhận về chất lượng mà còn được đông đảo khán giả đánh giá cao. Mặt khác, bộ phim Thor xếp thứ 20 với điểm số 6.8 sao và 20,792 lượt đánh giá, cũng là một thành công đáng chú ý trong thị trường phim.



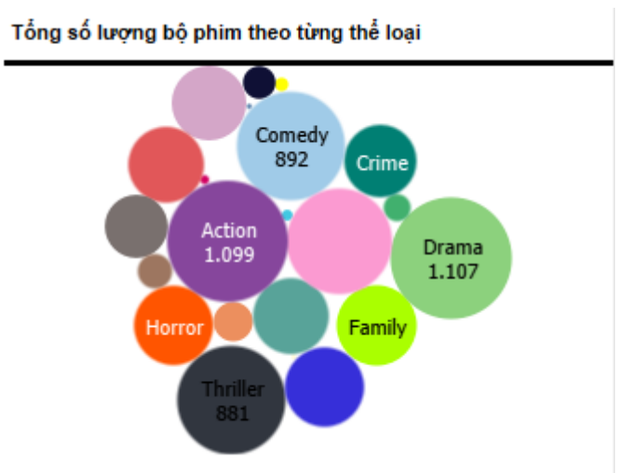
Biểu đồ cột này phản ánh số lượng đánh giá ở từng mức độ xếp hạng khác nhau trên IMDB. Việc này giúp chúng ta hiểu được phản ứng của cộng đồng đối với các bộ phim và phân tích mức độ phổ biến và đánh giá của từng tác phẩm. Chẳng hạn, khung đánh giá từ 6-7 sao phổ biến nhất, điều này cho thấy đa số các bộ phim để lại cho khán giả một mức độ hài lòng tương đối. Trong khi đó, mức

cao nhất - 10 sao, chỉ có 7 bộ phim, và mức thấp nhất là dưới 1 sao, cũng nhận được khá nhiều đánh giá từ khán giả.



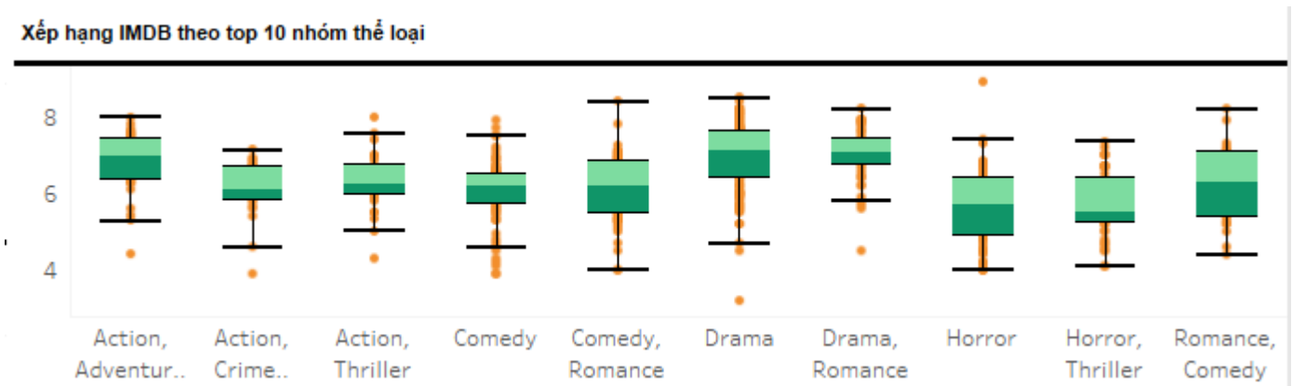
Biểu đồ đường so sánh doanh thu trong năm hiện tại và năm trước đó. Sự biến động của doanh thu qua các tháng trong năm có thể phản ánh sự tăng trưởng, ổn định hay giảm sút so với năm trước. Bên cạnh đó, biểu đồ cũng cung cấp về những điểm nổi bật trong năm, như các tháng có doanh thu cao nhất hoặc thấp nhất. Các yếu tố khác như sự kiện đặc biệt diễn ra vào khoảng thời gian đó cũng có thể ảnh hưởng đến biến động của doanh thu.

Như hình trên, ta có thể thấy doanh thu trong năm **2018** đạt **$20,850,025,819**; tăng **5,2%** so với năm 2017. Trong đó, tháng 6 là tháng đạt doanh thu cao nhất, và tháng 8 là tháng có doanh thu thấp nhất trong năm 2018.



Biểu đồ bong bóng hiển thị sự phân bố của các thể loại phim trong một quốc gia, giúp chúng ta nhận biết được xu hướng và sở thích của khán giả đối với từng thể loại phim cụ thể. Sự hiểu biết sâu sắc về sở thích của người xem có thể giúp các nhà sản xuất định hình chiến lược sản xuất, quyết định xem các bộ phim thuộc các thể loại nào có thị trường tiềm năng cao và có khả năng thu hút đối tượng khán giả lớn nhất. Những thể loại phim nào có kích thước bóng bóng lớn hơn, tức là phổ biến hơn, có thể đối mặt với mức độ cạnh tranh cao hơn.

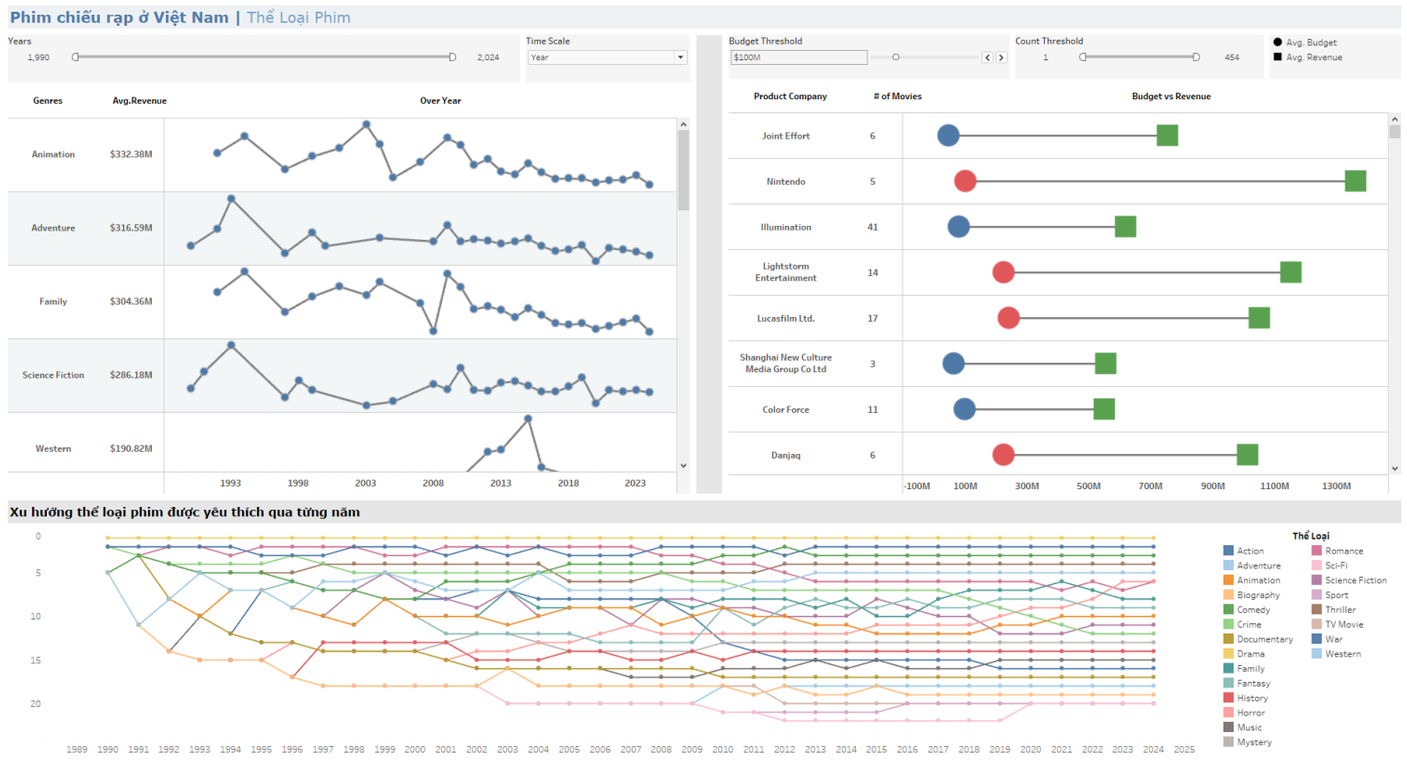
Cụ thể, **Drama** là thể loại được các hãng sản xuất khá ưa chuộng và cho ra khá nhiều bộ phim thuộc thể loại này, lên tới 1107 bộ. Điều này phần nào cho thấy xu hướng của khán giả đa số thích xem thể loại Drama hơn so với những thể loại khác. Đứng thứ 2 và 3, cũng cạnh tranh đáng kể với Drama, lần lượt là Action và Comedy với 1099 và 892 bộ phim được sản xuất.



Biểu đồ boxplot thể hiện sự phân phối của điểm đánh giá của các bộ phim trong từng thể loại, cung cấp thông tin về giá trị trung vị, phạm vi, phân vị, và các ngoại lệ có thể có. Bằng cách so sánh các thể loại phim khác nhau, chúng ta có thể nhận biết sự khác biệt trong đánh giá trung bình giữa chúng. Nếu một boxplot nằm ở vị trí cao hơn so với các boxplot khác, điều này có thể chỉ ra rằng thể loại phim đó thường được đánh giá cao hơn bởi khán giả. Hiểu rõ hơn về sở thích và phản ứng của khán giả đối với từng thể loại phim giúp các nhà sản xuất có thể đề xuất các chiến lược tốt hơn trong tương lai.

5.3 Genres Dashboard

Dashboard này sẽ tập trung phân tích và trực quan sâu hơn về thể loại của các bộ phim chiếu rạp tại Việt Nam (Hình 22).



Hình 22: Genres Dashboard

Bố cục của Dashboard và các thao tác điều chỉnh Dashboard:

- **Biểu đồ đường:** Biểu đồ này trình bày tổng doanh thu trung bình của các loại rạp chiếu phim qua thời gian, cho phép chúng ta theo dõi xu hướng phát triển. Chúng ta có thể điều chỉnh nút để tùy chỉnh hiển thị theo khoảng thời gian phát hành phim (từ năm 1964 đến 2024) và lựa chọn xem theo năm hoặc quý.
  - **Nút lọc thời gian phát hành phim:** từ năm 1964 đến 2024 và thao tác là kéo qua trái hoặc phải.
  - **Nút lọc thang thời gian:** có 2 lựa chọn là Năm hoặc Quý và thao tác là chọn bằng cách nhấp vào lựa chọn.

  - **Đặc biệt là:** thao tác chọn 1 thể loại của biểu đồ này và biểu đồ Dot sẽ thay đổi theo tương tác vừa rồi cho ta biết thêm thông tin chi tiết hơn về thể loại vừa được chọn.

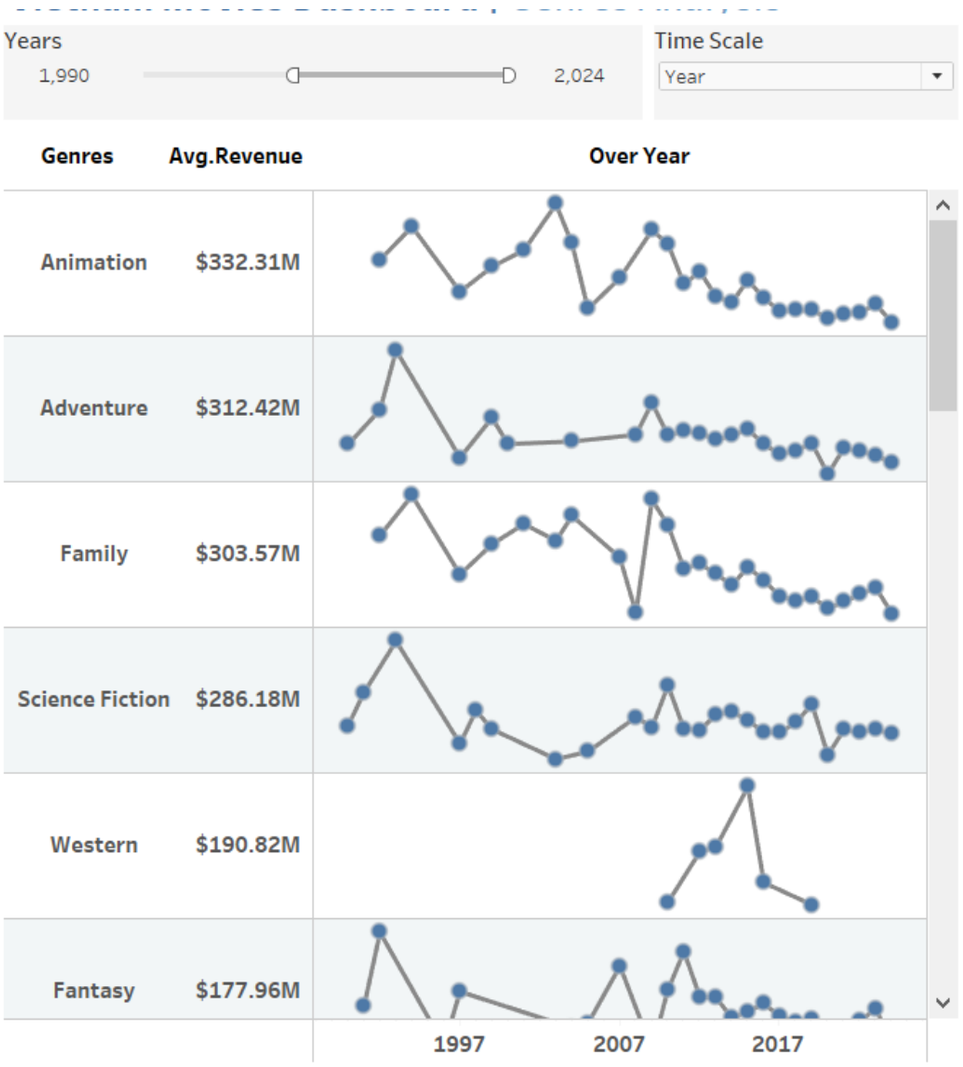
- **Biểu đồ Dot:** Biểu đồ này thể hiện tổng số lượng phim chiếu rạp của mỗi công ty sản xuất, cùng với so sánh giữa kinh phí sản xuất và doanh thu trung bình của từng công ty trên tổng số bộ phim. Chúng ta có thể tùy chỉnh mức kinh phí để so sánh và phân tích hiệu quả đầu tư của các công ty và tìm hiểu được công ty nào đầu tư vượt mức kinh phí. Kinh phí vượt mức và dưới mức được chọn sẽ được thể hiện bằng 2 màu sắc khác nhau và đều có hình tròn. Hình vuông là thể hiện doanh thu trung bình của công ty sản xuất. Thêm vào đó, ta cũng có thể lọc ra các công ty dựa vào số lượng phim sản xuất để tìm ra các công ty sản xuất nhiều phim nhất hoặc ít phim nhất.
  - **Nút điều chỉnh mức kinh phí:** thao tác là kéo sang trái hoặc phải.
  - **Nút lọc số lượng phim:** thao tác là kéo trái hoặc phải hoặc điền số vào ô trống.
  - **Đặc biệt là**, nếu chúng ta chọn 1 thể loại bên biểu đồ đường thì biểu đồ này sẽ thay đổi theo và hiển thị ra những thông tin bên trên nhưng với các công ty sản xuất có sản xuất ra các bộ phim có thể loại được chọn.

- **Biểu đồ Bump:** Biểu đồ này trình bày xếp hạng thể loại của các bộ phim chiếu rạp dựa trên số lượng phim được sản xuất qua từng năm. Điều này giúp chúng ta nhận biết xu hướng sản xuất phim của từng thể loại qua thời gian và đánh giá sự thay đổi về sự ưa chuộng của các thể loại phim trong ngành công nghiệp điện ảnh.

Phân tích và nhận xét chi tiết về Dashboard:

- **Trong biểu đồ đường (Hình 23),** nếu phân tích theo năm, có thể nhận thấy 4 thể loại phim có doanh thu trung bình cao nhất là Animation, Adventure, Family, và Science Fiction. Thể loại Animation đứng đầu với mức doanh thu trung bình là 332.31 triệu đô la Mỹ trong khoảng thời gian từ năm 1990 đến 2024. Đáng chú ý, năm 2003 ghi nhận mức doanh thu cao nhất gần 1 tỷ đô la Mỹ cho thể loại này, tuy nhiên, từ năm 2009 trở đi, doanh thu của phim Animation có xu

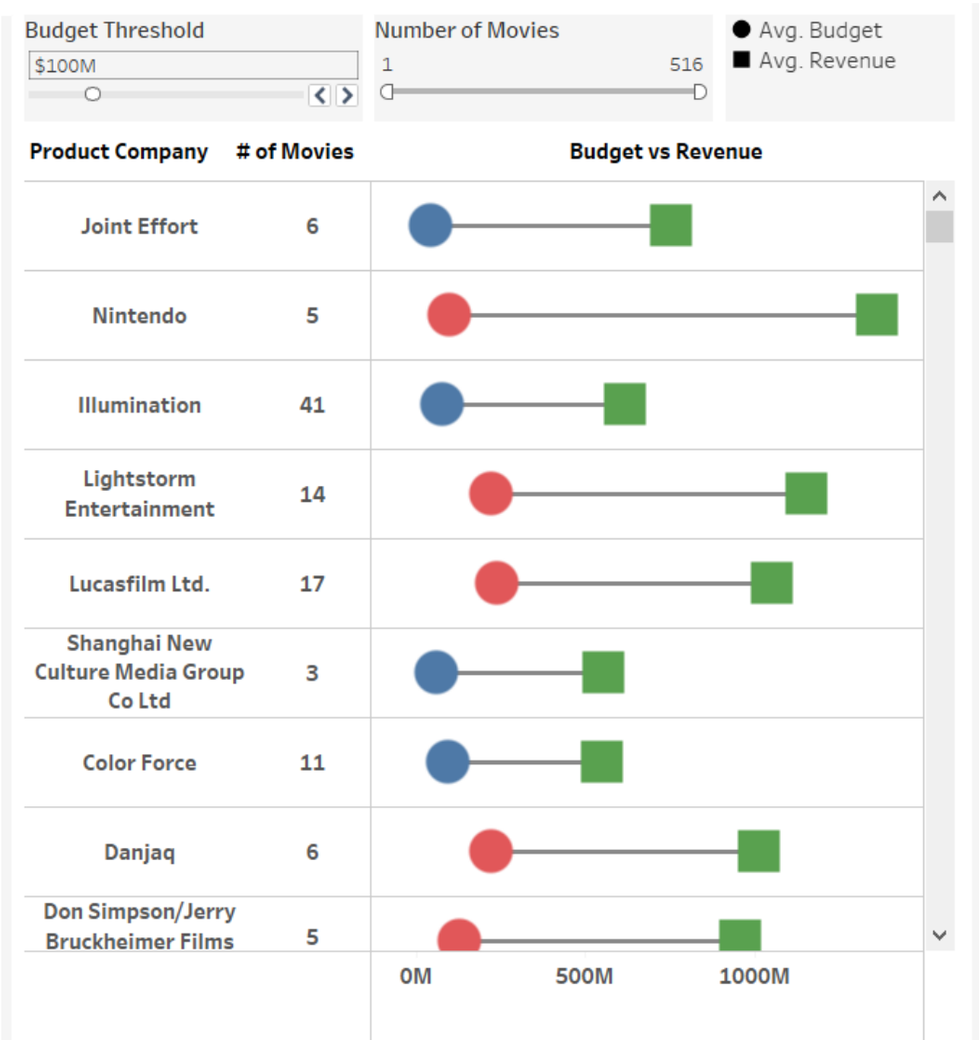
hướng giảm mạnh. Thứ hai là thể loại Adventure, với doanh thu trung bình đạt 312.42 triệu đô la Mỹ. Năm 1993 được biết đến là năm bùng nổ với doanh thu lớn nhất đối với thể loại này, tuy nhiên, sau năm 1993, doanh thu của phim Adventure giảm đáng kể và chỉ có sự phục hồi từ năm 1997 trở đi. Thứ ba là thể loại Family, với doanh thu trung bình đạt 303.57 triệu đô la Mỹ. Thể loại này ghi nhận mức doanh thu cao nhất vào năm 2009, nhưng từ đó đến nay, xu hướng phát triển đã chậm lại và giảm dần. Cuối cùng, thể loại Science Fiction, cũng ghi nhận doanh thu cao nhất vào năm 1993 gần 1 tỷ đô la Mỹ. Tuy nhiên, sau đó, doanh thu giảm mạnh trong khoảng thời gian từ 1993 đến 2003, nhưng bắt đầu có sự phục hồi và phát triển tích cực từ năm 2003 đến nay. Tóm lại, năm 1993 có thể coi là năm bùng nổ của thể loại phim khoa học viễn tưởng và phiêu lưu, với doanh thu của các bộ phim thuộc thể loại này đạt mức cao nhất.



Hình 23: Biểu đồ Đường về doanh thu của các thể loại qua các năm

- **Trong biểu đồ Dot (Hình 24),** có thể quan sát rằng có những công ty sản xuất ít phim nhưng lại đạt được lợi nhuận cao hơn so với các công ty sản xuất nhiều phim. Ví dụ, công ty Nintendo chỉ sản xuất 5 bộ phim trong khoảng thời gian từ 1990 đến 2024, nhưng lợi nhuận thu về lại cao hơn so với công ty Illumination sản xuất đến 41 bộ phim trong cùng khoảng thời gian. Điều này cho thấy sự hiệu quả trong việc đầu tư và sản xuất của công ty Nintendo. Tuy nhiên, cũng có sự đa dạng trong kết quả của các công ty sản xuất phim. Ngoài các công ty thu được lợi nhuận cao, còn xuất hiện các công ty không thu được lợi nhuận và bị lỗ vốn, như công ty CJ Entertainment, Studio and New, vì doanh thu thu được thấp hơn kinh phí sản xuất. Điều này cho thấy rằng,

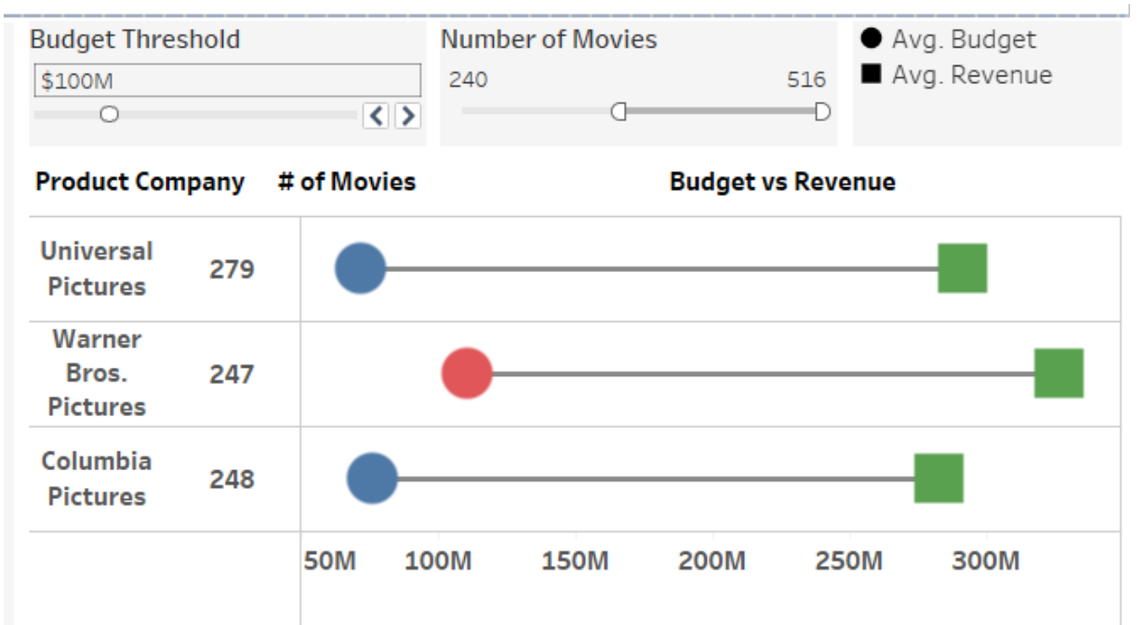
thành công trong ngành công nghiệp điện ảnh không chỉ phụ thuộc vào số lượng phim sản xuất mà còn phải kết hợp với chất lượng sản phẩm và kỹ năng quản lý tài chính.



Hình 24: Biểu đồ Dot về số lượng phim, kinh phí và doanh thu của các công ty sản xuất phim

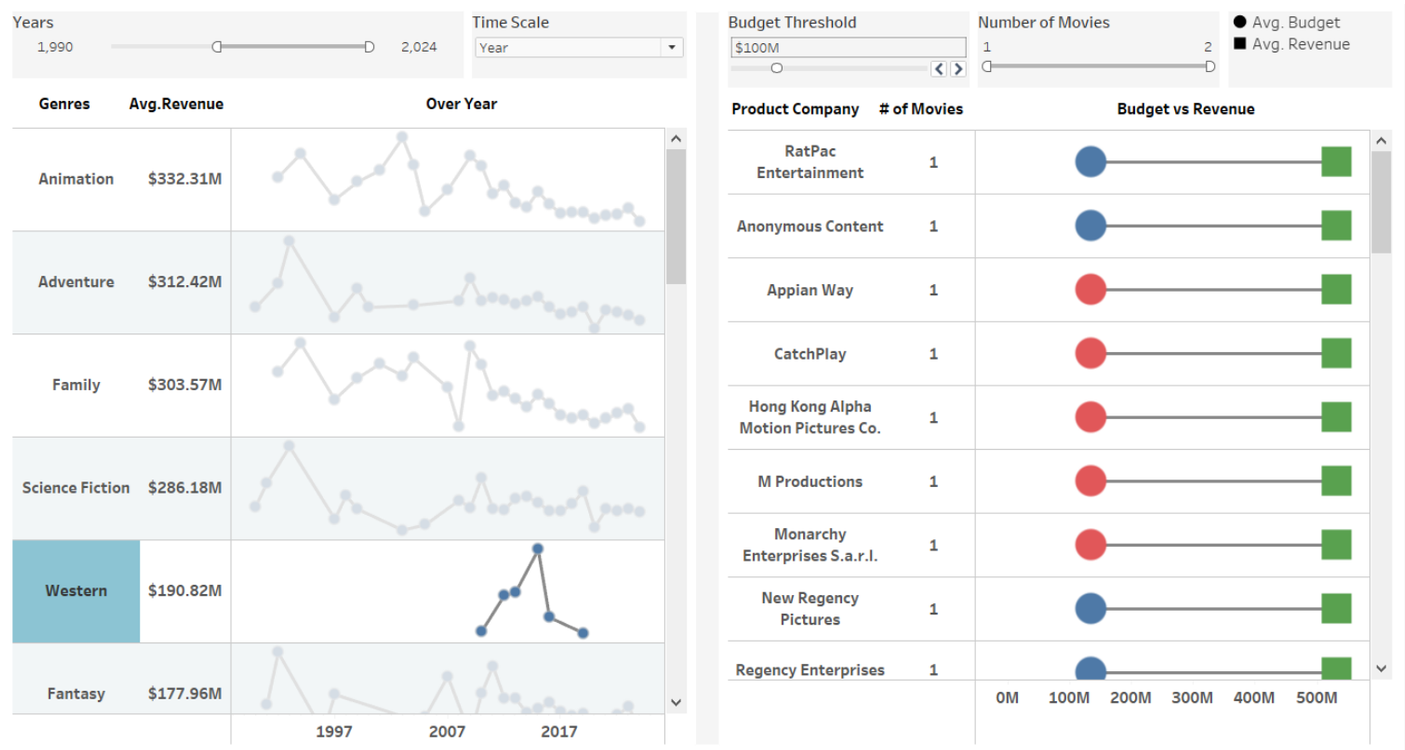
- **Dựa vào biểu đồ Dot (Hình 25),** có thể nhận thấy rằng Universal Pictures là công ty sản xuất nhiều bộ phim chiếu rạp nhất tại Việt Nam, với tổng số 279 bộ phim. Điều này thể hiện sự đóng góp đáng kể của công ty này vào ngành công nghiệp điện ảnh của đất nước. Tiếp theo, Warner Bros Pictures và Columbia Pictures cũng là những đại diện lớn trong ngành, sản xuất hơn 200 bộ phim chiếu rạp và đứng ở vị trí thứ hai và thứ ba. Ngoài các công ty lớn trên, Nintendo cũng là một đối tác đáng chú ý. Dù chỉ sản xuất một số lượng nhỏ các bộ phim, nhưng công ty này đạt được lợi nhuận cao nhất, đặc biệt là từ các bộ phim hoạt hình. Điều này cho thấy sự đa dạng

và sức ảnh hưởng của các công ty trong ngành công nghiệp điện ảnh tại Việt Nam.



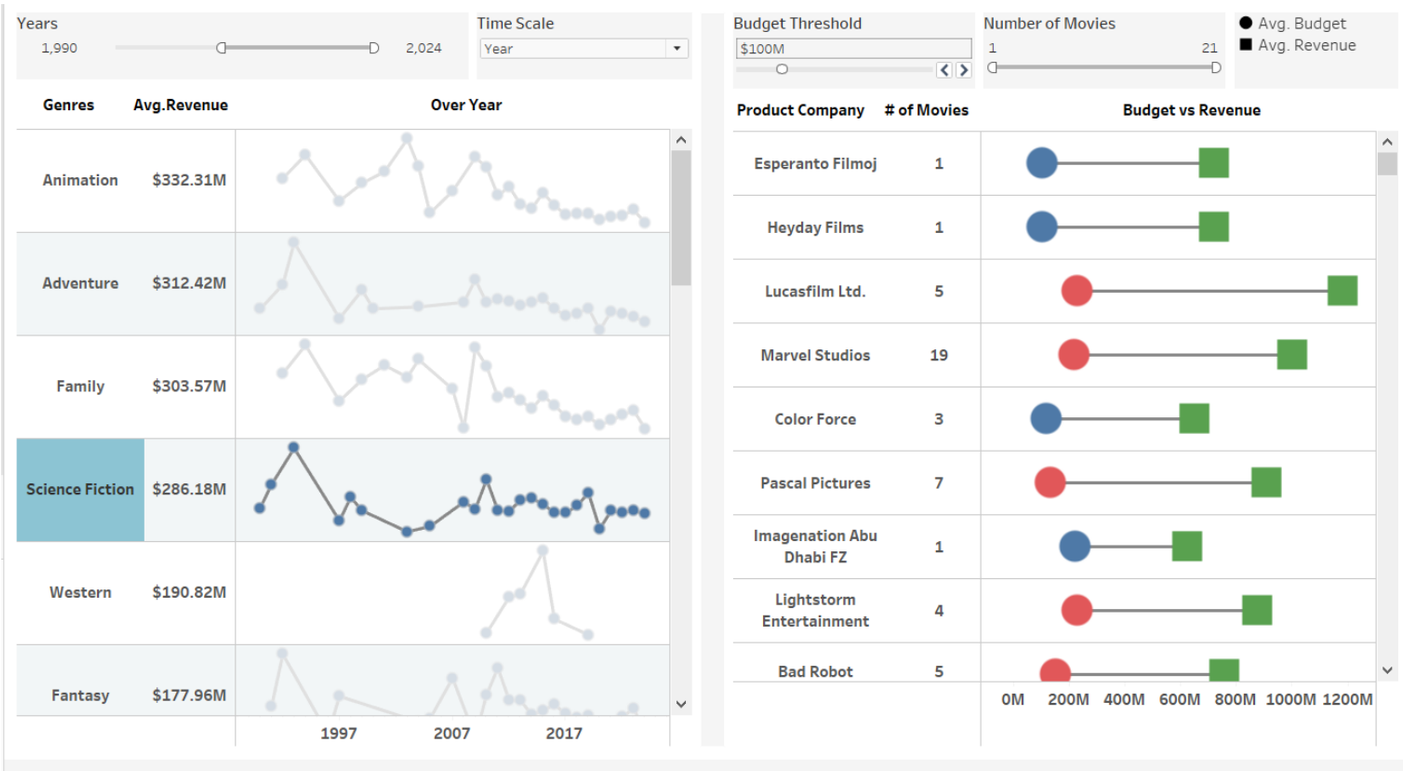
Hình 25: Biểu đồ Dot về số lượng phim, kinh phí và doanh thu của các công ty sản xuất phim

- **Xét biểu đồ Dot, khi chọn thể loại Western (Hình 26),** có thể nhận thấy rằng đa số các công ty sản xuất phim không ưa chuộng thể loại này để đầu tư. Phần lớn các công ty chỉ sản xuất một bộ phim thể loại Western và không có công ty nào sản xuất nhiều hơn một bộ phim trong thể loại này. Điều này cho thấy thể loại phim Western là một lựa chọn đặc biệt và thách thức đối với các nhà sản xuất, có thể là do nó không phổ biến hoặc không thu hút đối với đại đa số khán giả. Tuy nhiên, nhìn vào biểu đồ, ta cũng có thể thấy sự đa dạng trong kết quả lợi nhuận của các công ty sản xuất phim thể loại Western. Có các công ty kiếm được lợi nhuận đáng kể từ thể loại này, cùng với các công ty không đạt được lợi nhuận. Điều này phản ánh rõ ràng sự rủi ro và tiềm năng lợi nhuận khi đầu tư vào thể loại phim này.



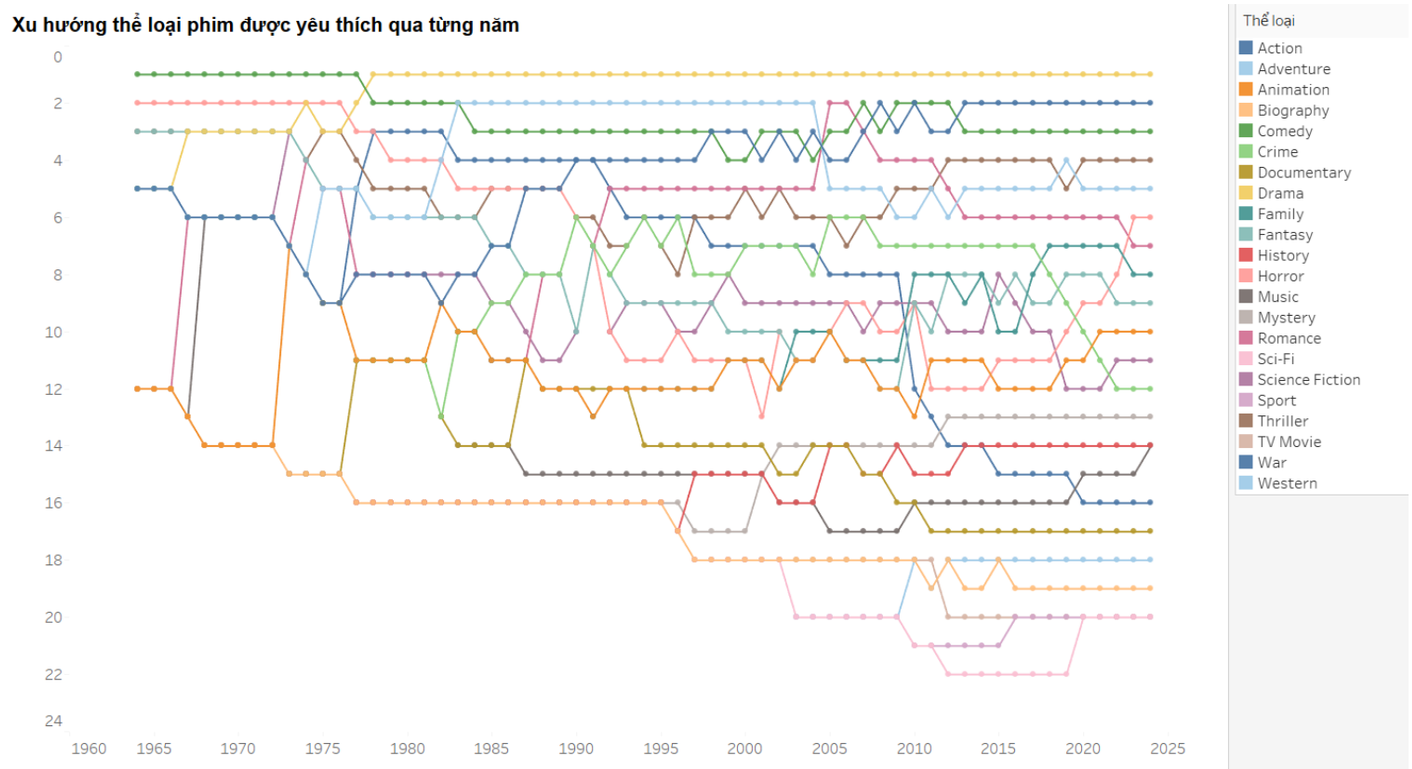
Hình 26: Dashboard khi chọn thể loại Western

- **Xét biểu đồ Dot, khi chọn thể loại Science Fiction (Hình 27),** ta nhận thấy sự phổ biến và đầu tư mạnh mẽ vào thể loại này từ các công ty sản xuất phim. Có một lượng lớn các công ty tham gia sản xuất phim trong thể loại này, điển hình như Marvel Studios, Warner Bros Picture, 20th Century Fox và Columbia Pictures. Các công ty này có sự hiện diện mạnh mẽ trên thị trường và sản xuất một số lượng đáng kể các bộ phim khoa học viễn tưởng. Ví dụ, 20th Century Fox đã sản xuất 21 bộ phim, trong khi Warner Bros Picture, Marvel Studios và Columbia Pictures đã sản xuất 19 bộ phim mỗi công ty. Trong số các công ty này, Lucasfilm Ltd là đơn vị kiếm được lợi nhuận cao nhất từ phim thể loại Science Fiction, với tổng cộng 12 bộ phim và lợi nhuận đáng kể. Điều này cho thấy sự thành công của công ty trong việc tận dụng tiềm năng của thể loại phim này và thu hút khán giả một cách hiệu quả. Tổng quan, biểu đồ Dot cho thấy rằng thể loại Science Fiction là một trong những lựa chọn phổ biến và có tiềm năng lợi nhuận cao trong ngành công nghiệp điện ảnh.



Hình 27: Dashboard khi chọn thể loại Science Fiction

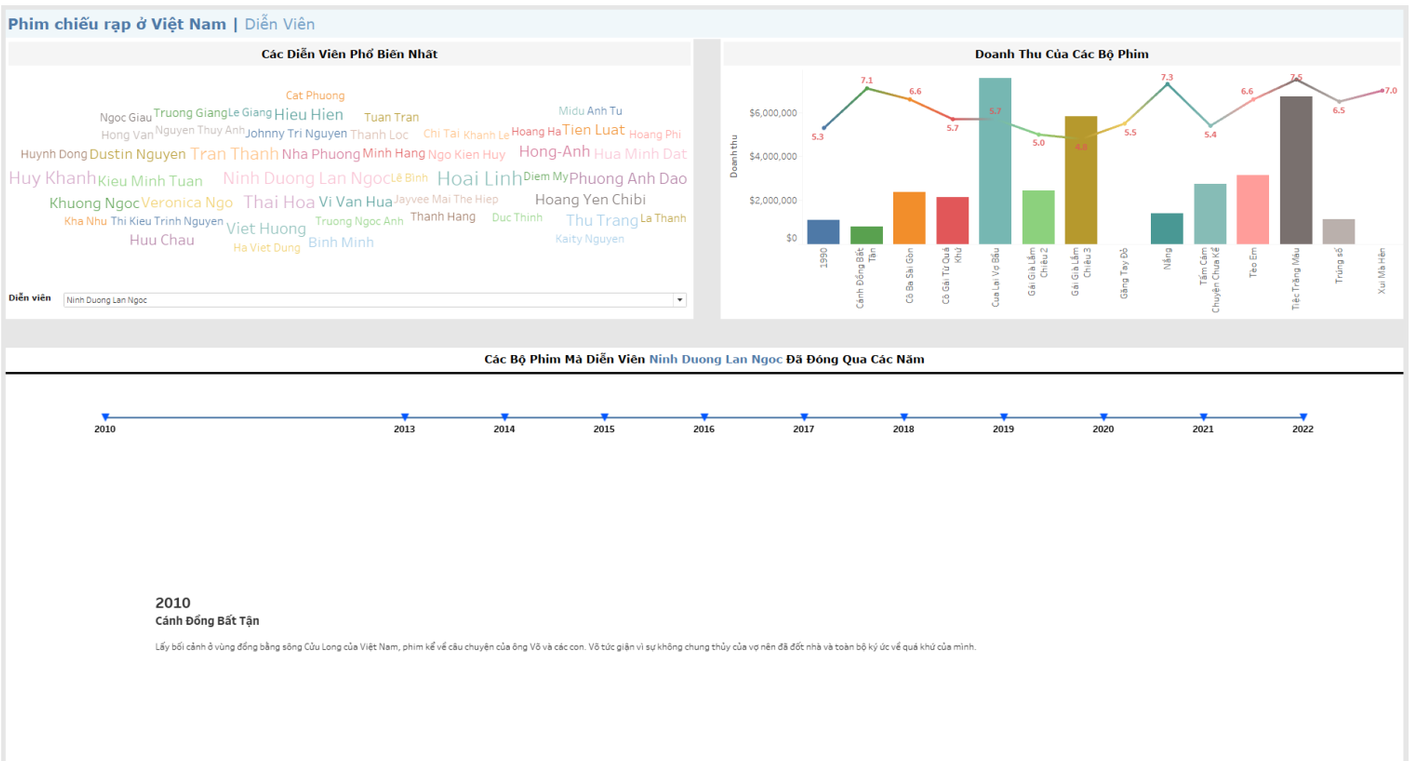
- **Ở biểu đồ Bump (Hình 28),** chúng ta có thể dễ dàng nhận thấy rằng thể loại phim Drama luôn giữ vị trí hàng đầu trong bảng xếp hạng yêu thích, với số lượng phim được sản xuất luôn là nhiều nhất từ năm 1990 đến nay. Điều này cho thấy sức hấp dẫn và sự ổn định của thể loại này trong lòng khán giả qua các thập kỷ. Tuy nhiên, vị trí thứ hai trong bảng xếp hạng thì có sự biến động lớn với sự thay đổi giữa các thể loại như Adventure, Romance, Comedy và Action. Điều này cho thấy sự đa dạng và sự chuyển đổi của sở thích của khán giả qua thời gian, cũng như sự cạnh tranh giữa các thể loại phim để thu hút sự chú ý và lòng tin của khán giả. Tóm lại, biểu đồ Bump cung cấp cái nhìn tổng quan về sự thay đổi và sự ổn định của các thể loại phim hàng đầu qua các năm, giúp cho những quyết định liên quan đến sản xuất và tiếp thị phim trở nên hiệu quả hơn.



Hình 28: Biểu đồ Bump thể hiện xu hướng yêu thích của các thể loại qua từng năm dựa trên số phim được sản xuất

5.4 Actors Dashboard

Dashboard phân tích phim ảnh Việt Nam này được thiết kế để cung cấp cái nhìn chi tiết về sự nghiệp diễn xuất của các diễn viên trong các bộ phim chiếu rạp tại Việt Nam. Thông qua việc sử dụng dữ liệu lịch sử trong nhiều năm, dashboard này không chỉ giúp người dùng theo dõi sự phát triển của từng diễn viên qua từng giai đoạn mà còn cung cấp cái nhìn tổng quan về tình hình tài chính của ngành công nghiệp điện ảnh. (Hình 29).



Hình 29: Actor Dashboard

Bố cục và chức năng của Dashboard phân tích điện ảnh Việt Nam này được thiết kế một cách khoa học và tinh tế, nhằm cung cấp cái nhìn toàn diện về sự nghiệp của các diễn viên trong ngành điện ảnh qua các bộ phim chiếu rạp. Dưới đây là mô tả chi tiết về từng thành phần của dashboard và các thao tác điều chỉnh tương ứng:

- **Biểu đồ Word Cloud:** Biểu đồ này hiển thị sự phân bố của các diễn viên dựa trên số lượng phim chiếu rạp mà họ đã tham gia, giúp người dùng nắm bắt được mức độ hoạt động và phổ

biến của từng diễn viên một cách trực quan.

  - **Chức năng điều chỉnh:**
    * **Nút lọc diễn viên:** Cho phép người dùng chọn ra tên diễn viên từ danh sách để truy cập thông tin chi tiết hơn về sự nghiệp của họ.

- **Biểu đồ kết hợp giữa đường và cột:** Biểu đồ này minh họa mối quan hệ giữa tổng doanh thu và điểm đánh giá trung bình của các bộ phim mà diễn viên được chọn đã tham gia, qua đó phản ánh được sự thành công về mặt thương mại lẫn chất lượng nghệ thuật.
  - **Chức năng điều chỉnh:**
    * **Nút lọc diễn viên:** Tương tự như trong biểu đồ Word Cloud, nút này cho phép người dùng chọn diễn viên để xem xét doanh thu và đánh giá phim một cách chi tiết.

- **Biểu đồ thể hiện dòng thời gian:** Biểu đồ này thể hiện các phim mà diễn viên đã tham gia theo trục thời gian, từ đó giúp người dùng theo dõi sự phát triển và thay đổi trong sự nghiệp của diễn viên qua các năm.
  - **Chức năng điều chỉnh:**
    * **Mô tả phim:** Khi chọn một năm cụ thể trên trục thời gian, người dùng có thể xem tên và mô tả sơ lược về các bộ phim mà diễn viên đã tham gia, cung cấp thông tin chi tiết hơn về nội dung mà diễn viên đã đóng trong từng phim.

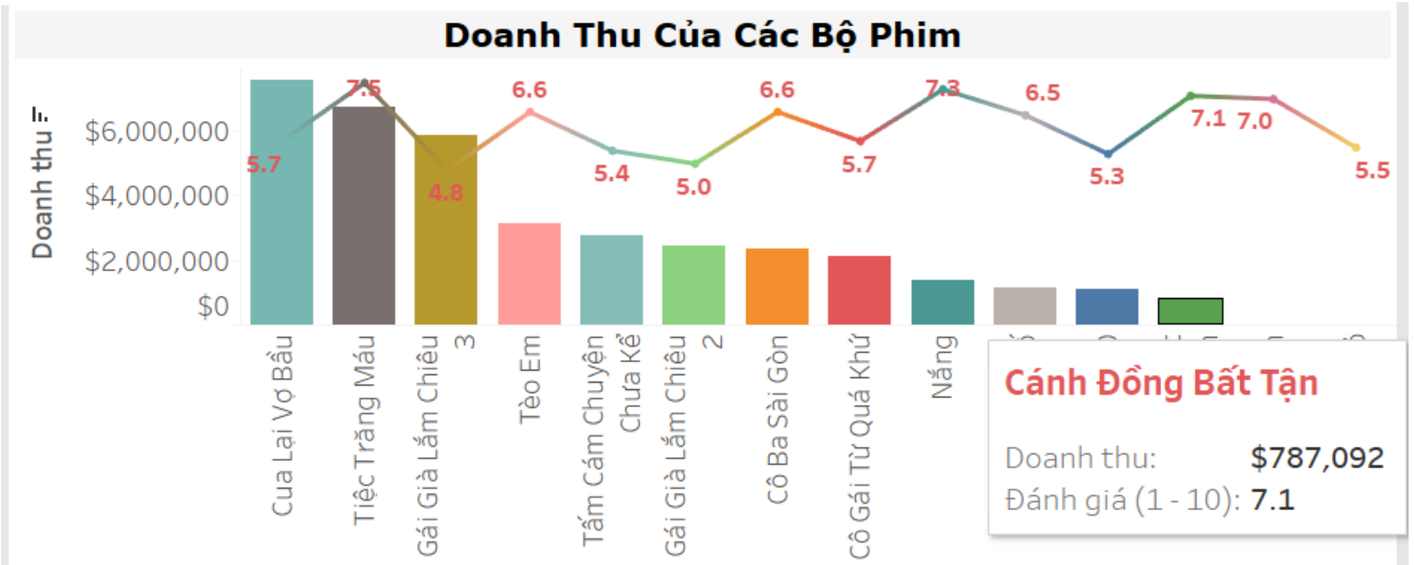
Phân tích và nhận xét chi tiết về Dashboard:

- **Phân tích Biểu Đồ Word Cloud (Hình 30):**

Hình 30: Biểu đồ Word Cloud

- Để phân tích sâu hơn về biểu đồ này, chúng em đã chọn ra 1 diễn viên tiêu biểu là Ninh Dương Lan Ngọc. CCô là một nữ diễn viên được công chúng đón nhận và ủng hộ khi ra mắt tác phẩm mới.
- Cô đã góp mặt trong khoảng 14 phim chiếu rạp, khá nhiều, cho thấy rằng cô hoạt động trong giới nghệ thuật khá năng nổ.

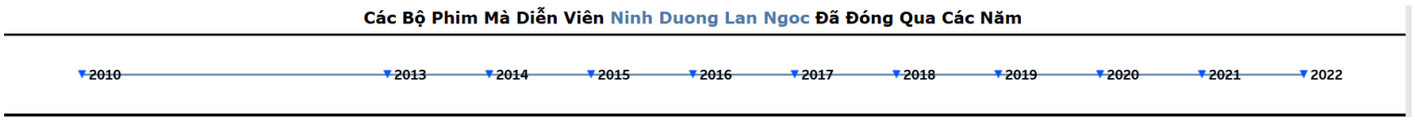
• **Phân tích biểu đồ kết hợp giữa đường và cột (Hình 31):**



Hình 31: Biểu đồ kết hợp đường và cột

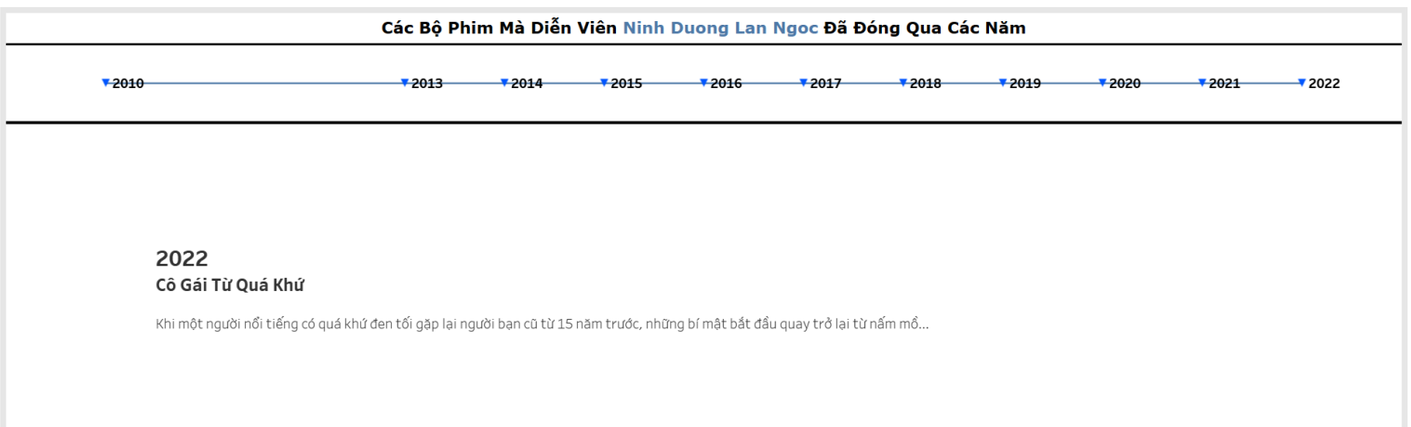
- Ninh Dương Lan Ngọc là một trong những diễn viên hàng đầu của điện ảnh Việt Nam, được đánh giá cao về khả năng diễn xuất và sức ảnh hưởng trong ngành công nghiệp điện ảnh. Với sự xuất hiện trong khoảng 14 bộ phim chiếu rạp, mỗi tác phẩm mà cô tham gia đều đạt được điểm bình chọn trên trung bình, chứng tỏ sự ổn định và đáng tin cậy của cô trong lựa chọn dự án.
- Vai diễn đầu tay của Ninh Dương Lan Ngọc trong "Cánh đồng bất tận"không chỉ giúp cô chứng minh được khả năng diễn xuất của mình mà còn mang về giải thưởng quan trọng như Cánh Diều Vàng [6].
- Ngoài ra, Ninh Dương Lan Ngọc được mệnh danh là "quý cô trăm tỷ"không chỉ phản ánh sức hút của cô đối với khán giả mà còn là sự thể hiện của khả năng thu hút đầu tư từ các nhà sản xuất và đạo diễn. Việc cô góp mặt trong các dự án như "Cua lại vợ bầu"(doanh thu 7 triệu đô) và "Gái già lắm chiêu 3"(5 triệu đô) là minh chứng cho sức ảnh hưởng của cô trong thị trường điện ảnh.
- **Nhận xét:** Tóm lại, Ninh Dương Lan Ngọc không là một diễn viên tài năng, các tác phẩm do cô lựa chọn diễn đều được khán giả phản ứng khá tích cực.

- **Phân tích biểu đồ thể hiện dòng thời gian (Hình 32):**
  - Diễn viên Ninh Dương Lan Ngọc vào khoảng 2013 đến 2022 hoạt động khá năng nổ, hầu như năm nào cũng ra mắt tác phẩm mới. Hơn hết, dù hoạt động khá thường xuyên nhưng mà những tác phẩm của cô vẫn được đánh giá cao và đạt mức doanh thu ổn định. Nhìn chung, sự nghiệp diễn xuất của Ninh Dương Lan Ngọc không chỉ đạt đỉnh cao về mặt nghệ thuật mà còn gặt hái thành công lớn về mặt thương mại, minh chứng bằng những con số ấn tượng về doanh thu.



Hình 32: Biểu đồ dòng thời gian

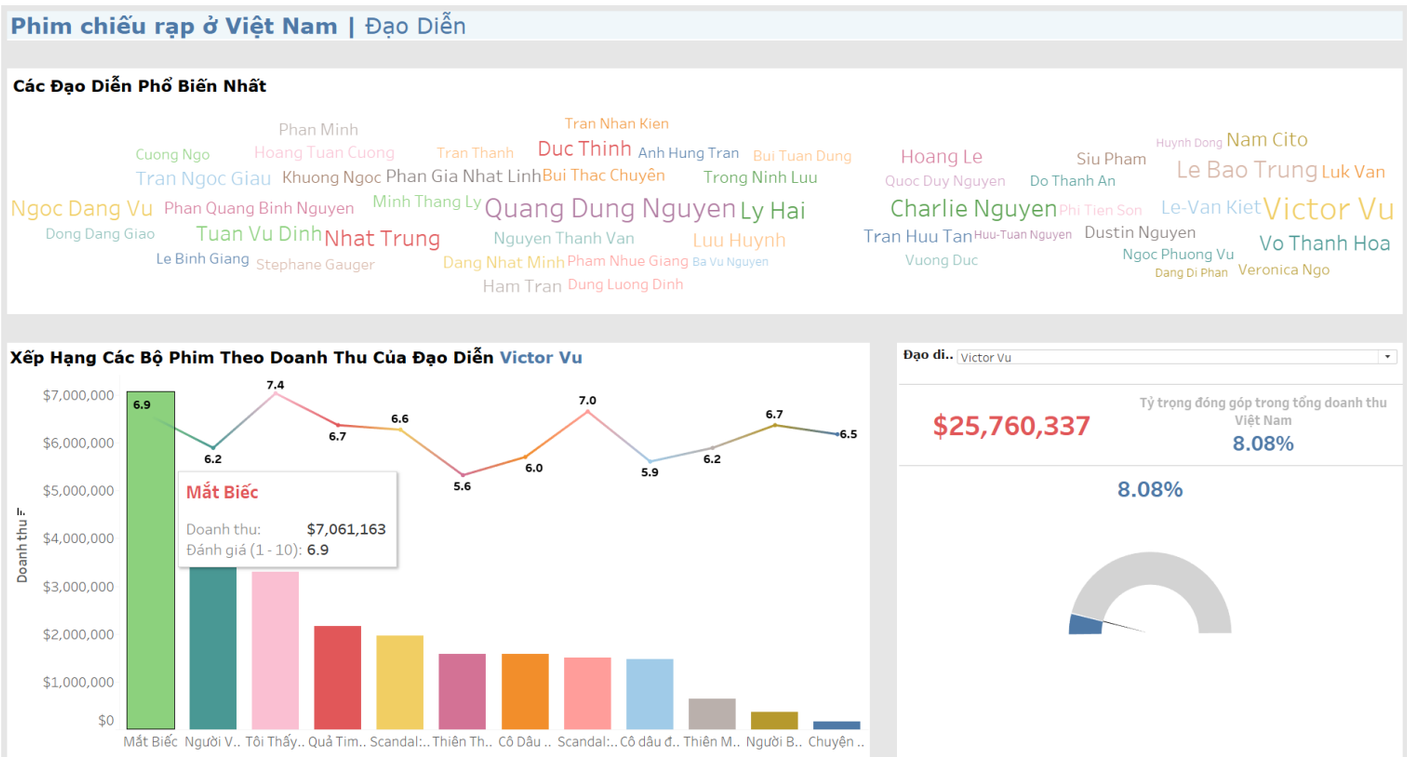
  - **Với sự lựa chọn xem xét năm 2023 trong biểu đồ thời gian (Hình 33),** bảng biểu mô tả bên dưới cho người dùng thông tin sơ bộ về bộ phim mà diễn viên được chọn đã đóng và được công chiếu vào năm đó.



Hình 33: Biểu đồ dòng thời gian với lựa chọn xem xét 2023

5.5 Directors Dashboard

Dashboard này được thiết kế nhằm mục đích phân tích sâu rộng và trực quan hóa các thông tin chi tiết về các đạo diễn, những nhân vật có vai trò then chốt trong quá trình sản xuất các bộ phim chiếu rạp tại Việt Nam. Qua đó, người dùng có thể hiểu rõ hơn về ảnh hưởng và tầm quan trọng của các đạo diễn đối với ngành công nghiệp điện ảnh, từ việc hình thành ý tưởng đến khâu hoàn thiện sản phẩm cuối cùng. Dashboard này sẽ cung cấp các biểu đồ đa dạng, bao gồm các chỉ số về số lượng phim được sản xuất, doanh thu phòng vé, và các chỉ số đánh giá trung bình, giúp người dùng có cái nhìn toàn diện và chi tiết về từng đạo diễn cụ thể (Hình 34).



Hình 34: Directors Dashboard

Bố cục của Dashboard và các thao tác điều chỉnh Dashboard:

- **Biểu đồ Word Cloud:** Biểu đồ được thiết kế để minh họa trực quan cho danh sách các đạo diễn tiêu biểu trong ngành công nghiệp điện ảnh Việt Nam. Kích thước của tên từng đạo diễn trong biểu đồ được xác định dựa trên số lượng phim mà họ đã thực hiện. Biểu đồ này không chỉ

giúp người dùng nhận diện được những cá nhân có ảnh hưởng lớn trong ngành qua số lượng tác phẩm họ đã sản xuất mà còn cung cấp cái nhìn trực quan và dễ hiểu về sự phân bố và tính chất đa dạng của các đạo diễn trong điện ảnh Việt Nam. Biểu đồ Word Cloud là công cụ hữu ích để phân tích và đánh giá năng lực sáng tạo cũng như tầm ảnh hưởng của các đạo diễn trong lĩnh vực điện ảnh.

- **Biểu đồ kết hợp đường và cột:** Biểu đồ này được thiết kế để hiển thị danh sách các bộ phim do từng đạo diễn đã sản xuất, dựa trên lựa chọn của người dùng thông qua bộ lọc. Đặc biệt, biểu đồ sẽ trực quan hóa các bộ phim có doanh thu cao nhất cũng như điểm đánh giá trung bình từ khán giả, giúp người dùng dễ dàng đánh giá mức độ thành công và ảnh hưởng của các đạo diễn trong ngành công nghiệp điện ảnh Việt Nam. Công cụ này không chỉ phản ánh khả năng sáng tạo và đóng góp của các đạo diễn qua từng dự án phim mà còn cung cấp cái nhìn tổng quan về sự tiếp nhận và đánh giá của khán giả đối với sản phẩm điện ảnh. Nhờ đó, biểu đồ này là một phương tiện hiệu quả để phân tích và thể hiện rõ nét những đóng góp quan trọng của các nhà làm phim trong việc phát triển điện ảnh nước nhà.
  - **Nút chọn đạo diễn:** chọn một đạo diễn trong danh sách mà người dùng muốn biết thêm thông tin.
- **Bảng biểu doanh thu và biểu đồ Gauge:** Được thiết kế để trực quan hóa tổng doanh thu từ các bộ phim của từng đạo diễn, dựa trên lựa chọn của người dùng qua bộ lọc. Cụ thể, bảng biểu này không chỉ cung cấp thông tin về tổng doanh thu mà còn phản ánh sự phân bố doanh thu của các đạo diễn trên thị trường điện ảnh Việt Nam. Kết hợp với biểu đồ Gauge, công cụ này giúp người dùng nhận thức rõ ràng hơn về ảnh hưởng và mức độ thành công của từng đạo diễn, qua đó phản ánh khả năng thu hút khán giả và tạo ra doanh thu từ các tác phẩm điện ảnh. Biểu đồ và bảng biểu này là những công cụ hữu hiệu giúp đánh giá và so sánh hiệu quả hoạt động nghệ thuật cũng như kinh tế của các đạo diễn, qua đó nâng cao hiểu biết về bối cảnh và xu hướng phát triển của ngành công nghiệp điện ảnh Việt Nam.
  - **Nút chọn đạo diễn:** chọn một đạo diễn trong danh sách mà người dùng muốn biết thêm

thông tin.

Phân tích và nhận xét chi tiết về Dashboard:

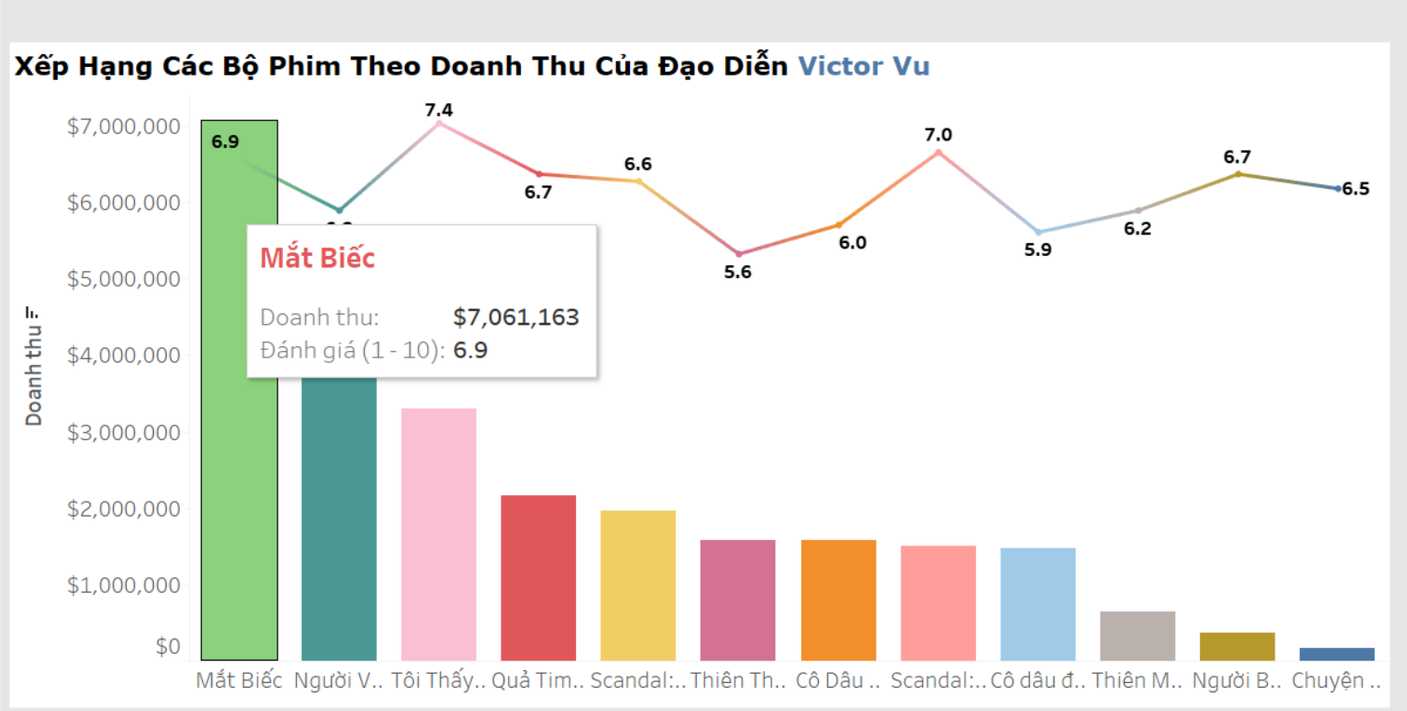
- **Phân tích biểu đồ Word Cloud (Hình 35):**
  - Có thể thấy rõ rằng có sự hiện diện thường xuyên của nhiều đạo diễn danh tiếng như Victor Vũ, Nguyễn Quang Dũng, Lê Hoàng Hoa và Charlie Nguyễn, mỗi người đã cho ra mắt hơn 10 tác phẩm điện ảnh. Bên cạnh những cái tên đã được khẳng định, ngành công nghiệp điện ảnh Việt Nam cũng chứng kiến sự nổi lên của các đạo diễn mới như Trấn Thành, người đã giới thiệu 3 bộ phim trong những năm gần đây.
  - **Nhận xét:** Sự phong phú trong danh sách các đạo diễn từ lâu đời đến mới nổi phản ánh tính đa dạng và sự phát triển không ngừng của ngành công nghiệp điện ảnh Việt Nam. Sự đóng góp từ cả những nhà làm phim có kinh nghiệm lâu năm và những tài năng mới nổi bật đã góp phần làm phong phú thêm diện mạo của điện ảnh trong nước, đồng thời mở rộng ảnh hưởng và sức hấp dẫn của nó trên trường quốc tế.

Hình 35: Biểu đồ Word Cloud

- **Phân tích biểu đồ kết hợp giữa đường và cột (Hình 36):**
  - Để thực hiện một phân tích chi tiết hơn về biểu đồ này, chúng em đã chọn đạo diễn Victor Vũ làm ví dụ. Anh là một nhân vật trong ngành điện ảnh được đánh giá cao và nhận được sự ủng hộ từ công chúng khi sản xuất các tác phẩm mới.
  - Victor Vũ, được biết đến là "Vua phòng vé"ở Việt Nam, là một đạo diễn nổi tiếng với 16 bộ phim đã sản xuất. Xét về điểm đánh giá, tất cả các phim của anh đều nhận được sự yêu mến

từ khán giả, với điểm thấp nhất là 5.6 cho phim "Thiên Thần Hộ Mệnh"và cao nhất là 7.4 cho phim "Tôi Thấy Hoa Vàng Trên Cỏ Xanh".

- Về mặt doanh thu, phim có doanh thu thấp nhất là "Chuyện Tình Xa Xứ"với 174 nghìn đô và cao nhất là "Mắt Biếc"với 7 triệu đô.
- Điểm nổi bật trong sự nghiệp của Victor Vũ là bộ phim "Thiên Mệnh Anh Hùng", lấy cảm hứng từ vụ án Lệ Chi Viên. Phim này không chỉ kiếm được 645 nghìn đô mà còn giành được nhiều giải thưởng danh giá tại Cánh diều vàng năm 2012, bao gồm giải cho Phim điện ảnh xuất sắc nhất và Đạo diễn xuất sắc nhất. [7]
-
- **Nhận xét:**
  * Đánh giá trên biểu đồ cho thấy Victor Vũ là một đạo diễn tài năng, có khả năng tạo ra các bộ phim đạt doanh thu cao và nhận được sự đón nhận nồng nhiệt từ công chúng.
  * Thông qua những phân tích này, chúng ta có thể hiểu rõ hơn về sự phát triển của ngành công nghiệp điện ảnh và sự ảnh hưởng của các đạo diễn tài năng trong việc tạo ra các tác phẩm điện ảnh chất lượng và có sức hấp dẫn lớn.

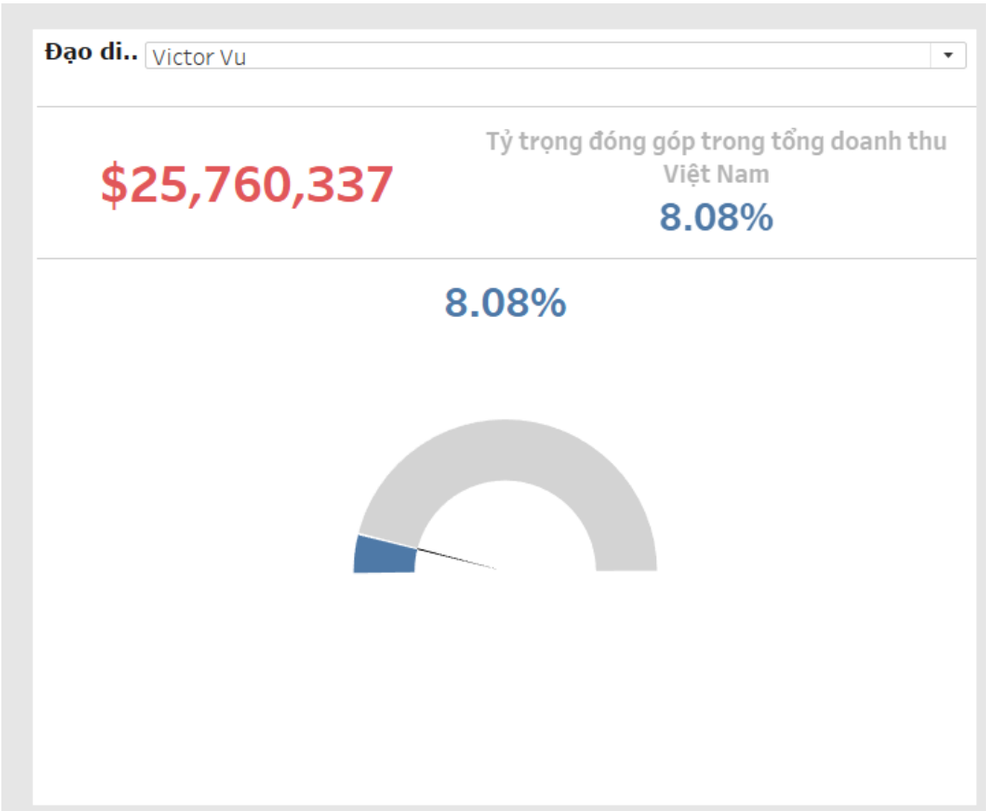


Hình 36: Biểu đồ kết hợp giữa đường và cột

- **Phân tích bảng biểu doanh thu và biểu đồ Gauge (Hình 37):**
  - Các phim của anh đã mang về tổng doanh thu lên tới 25 triệu đô, chiếm 8.08 phần trăm tổng doanh thu phim ở Việt Nam.
  - Biểu đồ Gauge hiển thị tỷ lệ phần trăm này qua một cung nửa vòng, với phần tô màu xanh lam biểu thị tỷ lệ doanh thu màVictor Vũ đóng góp cho điện ảnh Việt Nam. Đây là một công cụ trực quan hữu hiệu giúp nhận diện nhanh chóng mức độ thành công của các đạo diễn qua doanh thu phim, từ đó cung cấp cái nhìn tổng quan và dễ hiểu về tầm ảnh hưởng của họ đối với ngành công nghiệp.
  - **Nhận xét:**
    * Từ dữ liệu thống kê, có thể thấy rằng Victor không chỉ là một nhà làm phim năng nổ mà còn là một trong những nhân tố chủ chốt góp phần vào sự thịnh vượng của điện ảnh Việt Nam. Sự đón nhận nồng nhiệt của công chúng đối với các bộ phim do ông sản xuất

không chỉ thể hiện qua con số doanh thu mà còn qua phản hồi tích cực từ phía người xem.

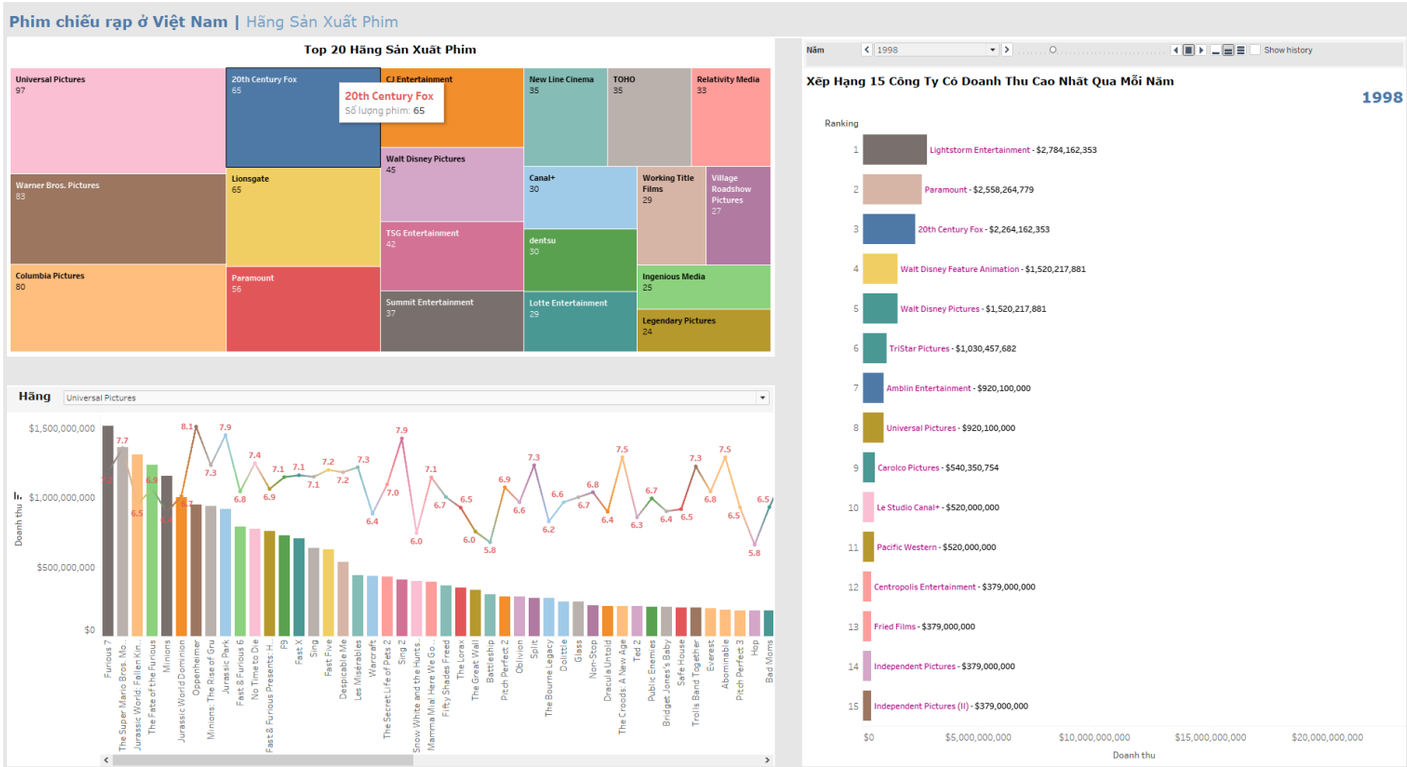
* Nhìn chung, Victor Vũ thực sự đã khẳng định được tài năng và sức ảnh hưởng của mình trong ngành công nghiệp điện ảnh Việt Nam. Các tác phẩm ăn khách đã khiến Victor Vũ trở thành một cái tên đảm bảo doanh thu cho phòng vé điện ảnh Việt Nam. Anh không chỉ tạo ra những bộ phim có doanh thu cao mà còn mang đến những tác phẩm chất lượng, được khán giả và giới chuyên môn đánh giá cao.



Hình 37: Bảng biểu doanh thu chi tiết và biểu đồ Gauge

5.6 Company Dashboard

Dashboard này sẽ tập trung phân tích và trực quan chi tiết hơn cho người dùng về các công ty, họ là chủ dự án phim, kết nối tất cả thành viên, nguồn lực về tài chính – nhân sự, điều phối, quản lý và là người đi giải quyết tất cả mọi vấn đề khó khăn, trở ngại trong một đoàn phim. (Hình 38)



Hình 38: Company Dashboard

Bố cục của Dashboard:

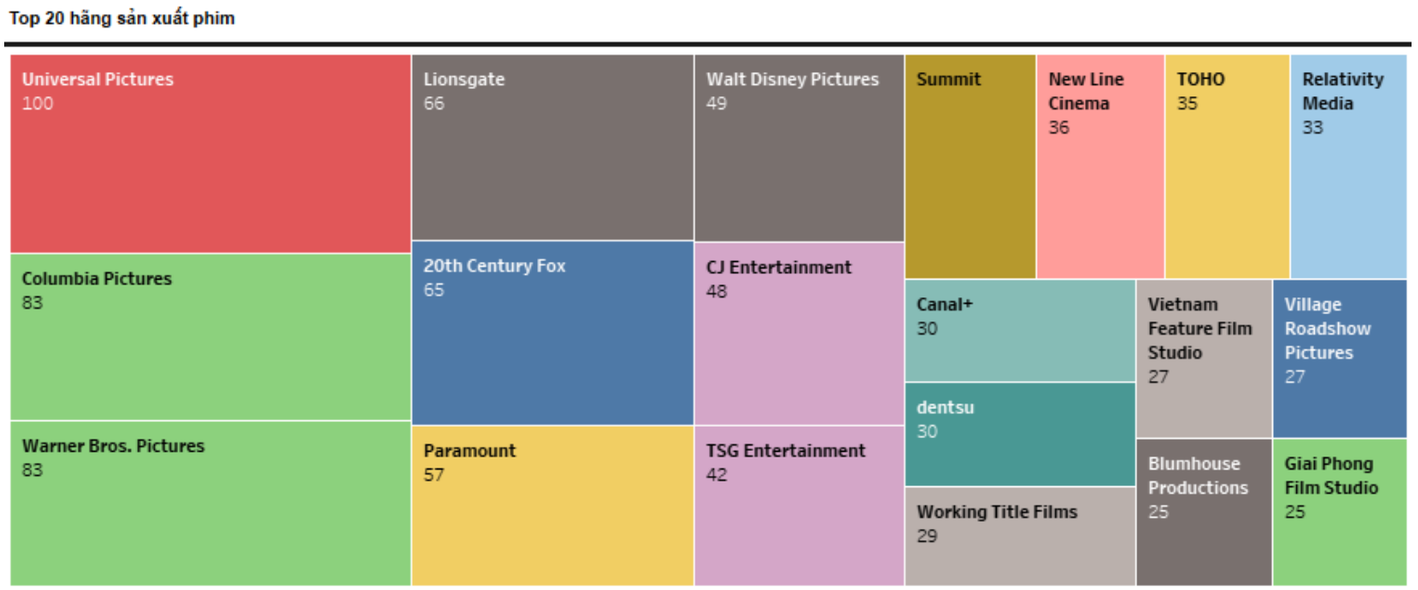
- **Biểu đồ treemap:** hiển thị 20 công ty sản xuất phim có số lượng phim được sản xuất nhiều nhất.
- **Biểu đồ cột và đường:** hiển thị các phim của một công ty sản xuất, trong đó biểu đồ cột cung cấp doanh thu của phim, biểu đồ đường cho biết điểm đánh giá của phim.
  - **Chức năng điều chỉnh:**
    - **Nút lọc công ty:** Cho phép người dùng chọn ra công ty để xem chi tiết thông tin các phim của công ty đó.
- **Biểu đồ cột ngang:** Biểu đồ xếp hạng 15 công ty có doanh thu cao nhất qua từng năm.
  - **Chức năng điều chỉnh:**
    - **Nút lọc năm:** Cho phép người dùng chọn ra năm để xem bảng xếp hạng doanh thu của

năm đó.

  * **Nút hiển thị tăng dần:** Tự động hiển thị 15 phim có doanh thu cao nhất tăng dần qua từng năm đến năm 2024

Phân tích và nhận xét chi tiết về Dashboard

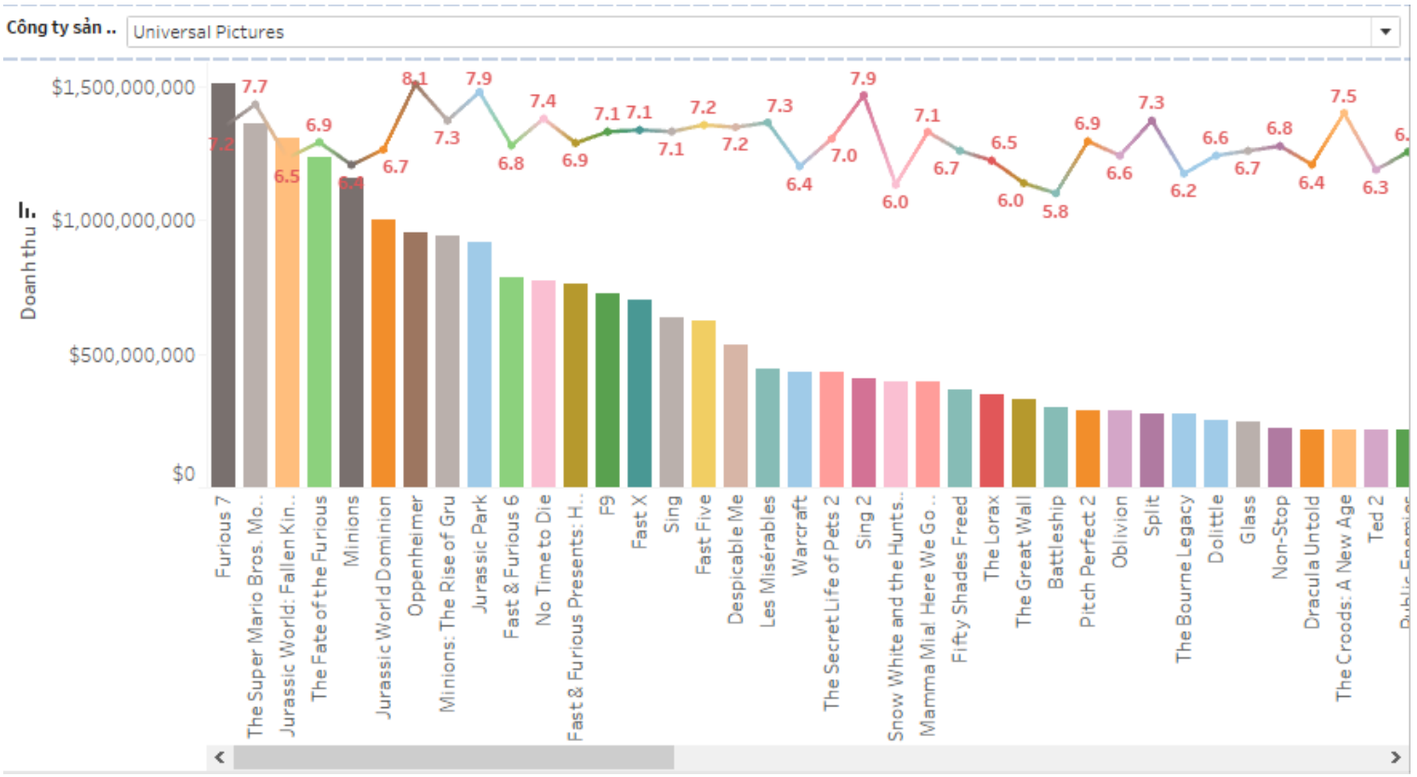
- **Biểu đồ treemap (Hình 39):**
  - Universal Pictures dẫn đầu danh sách với 100 phim đã sản xuất, nhiều nhất trong số các hãng được nêu trên biểu đồ. Đây là một trong những công ty phim lớn nhất và lâu đời nhất ở Mỹ.
  - Columbia Pictures và Warner Bros. Pictures đồng xếp thứ hai với cùng 83 phim.
  - Lionsgate và 20th Century Fox cũng cạnh tranh sát nút với nhau, với Lionsgate sản xuất 66 phim và 20th Century Fox sản xuất 65 phim.
  - Walt Disney Pictures, một hãng phim nổi tiếng khác, có số phim ít hơn so với những hãng dẫn đầu với 49 phim.
  - Có sự xuất hiện của một hãng phim không phải từ Hollywood, đó là CJ Entertainment từ Hàn Quốc với 48 phim.
  - Gái Phòng Film Studio, nằm cuối cùng trong danh sách, cùng với Blumhouse Productions và Village Roadshow Pictures, đều có 25 phim.
  - Biểu đồ cung cấp một cái nhìn tổng quan về ngành công nghiệp sản xuất phim, cho thấy sự đa dạng của các hãng phim từ các khu vực khác nhau của thế giới cũng như số lượng phim họ đã sản xuất.



Hình 39: Biểu đồ treemap

- **Biểu đồ cột và đường (Hình 40):** Biểu đồ có thể xem xét được nhiều công ty và nhóm đã chọn công ty Universal Pictures để phân tích.
  - Phim **Furious 7** có doanh thu cao nhất trong số các phim của công ty này, với khoảng hơn 1,5 tỷ USD và đánh giá là 7.2. Đây là một trong những phim bom tấn về hành động và tốc độ nổi bật trong năm 2015, bộ phim đã gây được tiếng vang vô cùng lớn. Khi nhắc đến những bộ phim về hành động tốc độ người ta thường sẽ nói đến **Furious 7**.
  - Các phim như **Jurassic World**, **Minions**, **The Fate of the Furious** cũng đạt doanh thu rất cao, chúng ngang ngửa hoặc vượt qua mốc 1 tỷ USD.
  - Các phim có doanh thu cao không nhất thiết có điểm đánh giá cao nhất. Ví dụ, **Minions** có doanh thu cao nhưng chỉ có điểm đánh giá là 6.5, trong khi một số phim với doanh thu thấp hơn có điểm đánh giá cao hơn 7.
  - Phim có điểm đánh giá cao nhất trên biểu đồ là 8.1, đây là phim **Oppenheimer** là một bộ phim về quá trình chế tạo bom nguyên tử, nội dung phim khá hàn lâm, phù hợp với những người làm khoa học. Bộ phim đã chiến thắng tới 7 tượng vàng Oscar, do đó không khó hiểu

khi phim đạt điểm 8.1.

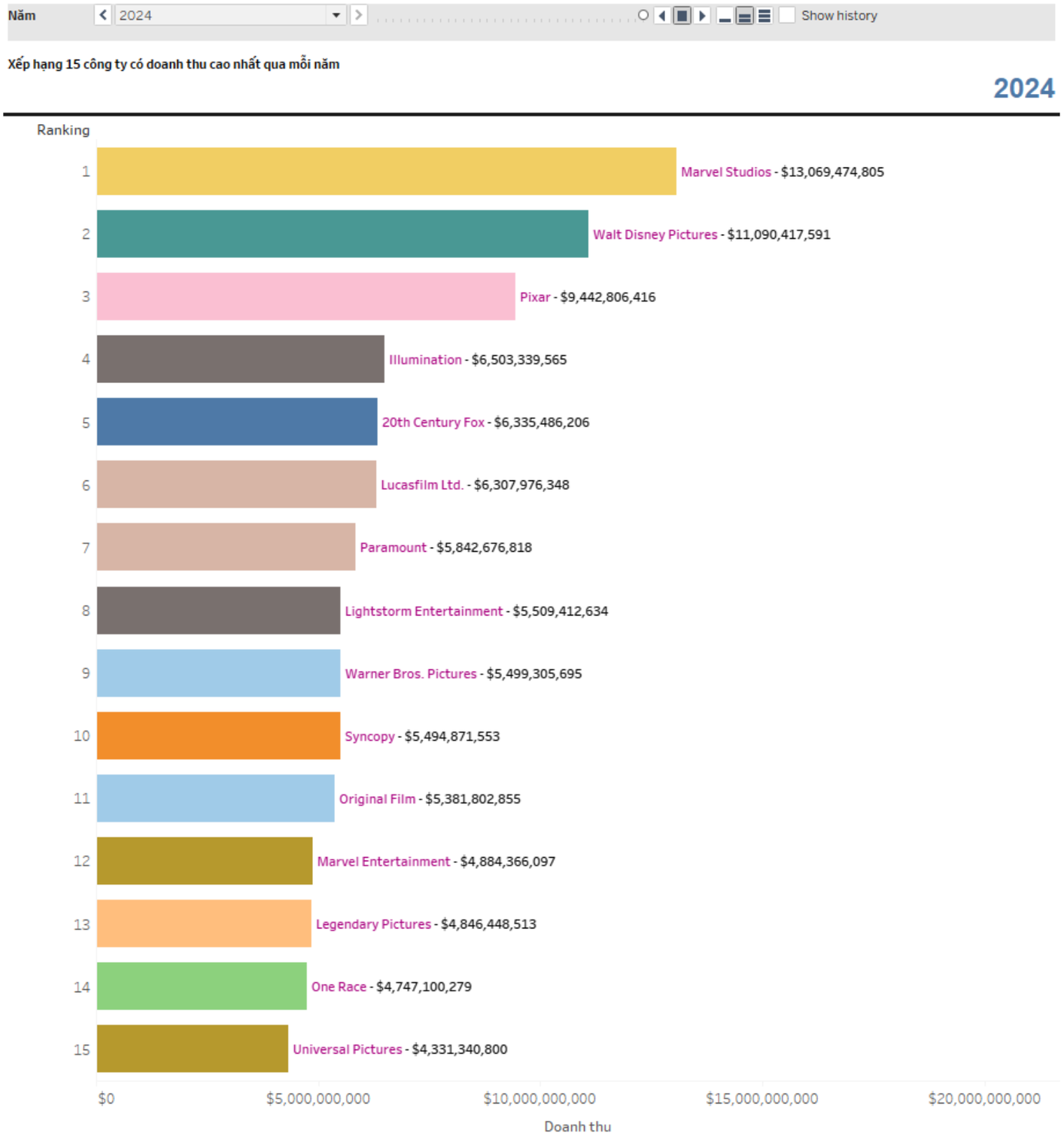


Hình 40: Biểu đồ cột và đường

- **Biểu đồ cột ngang (Hình 41):** Xếp hạng 15 công ty có doanh thu cao nhất qua từng năm, các doanh thu đó đã được cộng dồn, có nghĩa đến năm 2024 thì doanh thu đó là toàn bộ doanh thu từ trước đến nay.
  - Marvel Studios đứng đầu danh sách với doanh thu cao nhất, là hơn 13 tỷ đô. Nếu mọi người đam mê vũ trụ điện ảnh siêu anh hùng thì không thể không biết đến Marvel. Có thể kể đến một số phim nổi tiếng với doanh thu cao như series phim Avengers, Ironman, Thor,... Còn vô số phim nổi tiếng khác trong vũ trụ siêu anh hùng Marvel.
  - Walt Disney Pictures đứng thứ hai, với doanh thu là $11,090,417,591, thấp hơn Marvel Studios khoảng 2 tỷ đô. Hãng phim Disney cũng nổi tiếng với series phim Cướp biển vùng Caribe, Frozen, và hàng loạt phim hoạt hình khác với doanh thu hàng tỷ đô.
  - Cả Marvel Studios và Walt Disney Pictures đều có doanh thu vượt qua mốc 10 tỷ đô la, điều

này cho thấy sức ảnh hưởng lớn của hai công ty này trên thị trường giải trí.

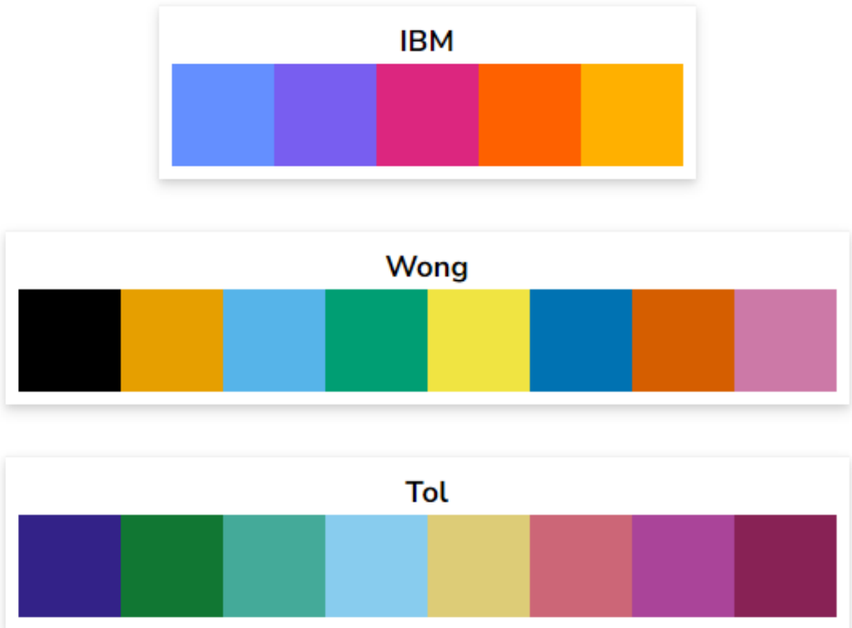
- Pixar, công ty chuyên về phim hoạt hình, đứng ở vị trí thứ ba với doanh thu là $9,442,806,416
- Các công ty trong top 5 (Marvel Studios, Walt Disney Pictures, Pixar) đều có doanh thu trên 9 tỷ đô la, cho thấy sự nổi trội so với những công ty khác trong danh sách.
- Có sự đa dạng trong danh sách về loại hình công ty, bao gồm các studio chuyên về phim hoạt hình như Pixar và Illumination, các studio chuyên về phim siêu anh hùng như Marvel Studios và Warner Bros. Pictures, và các studio với các thể loại phim khác nhau như Universal Pictures và Columbia Pictures.
- Các công ty từ vị trí thứ 4 đến thứ 15 có doanh thu khá gần nhau, nằm trong khoảng từ 4 tỷ đô la đến gần 6.5 tỷ đô la.
- Biểu đồ này cung cấp một cái nhìn tổng quan về ngành công nghiệp giải trí và phản ánh phần nào sự phân phối doanh thu giữa các công ty lớn trong ngành.



Hình 41: Biểu đồ cột ngang

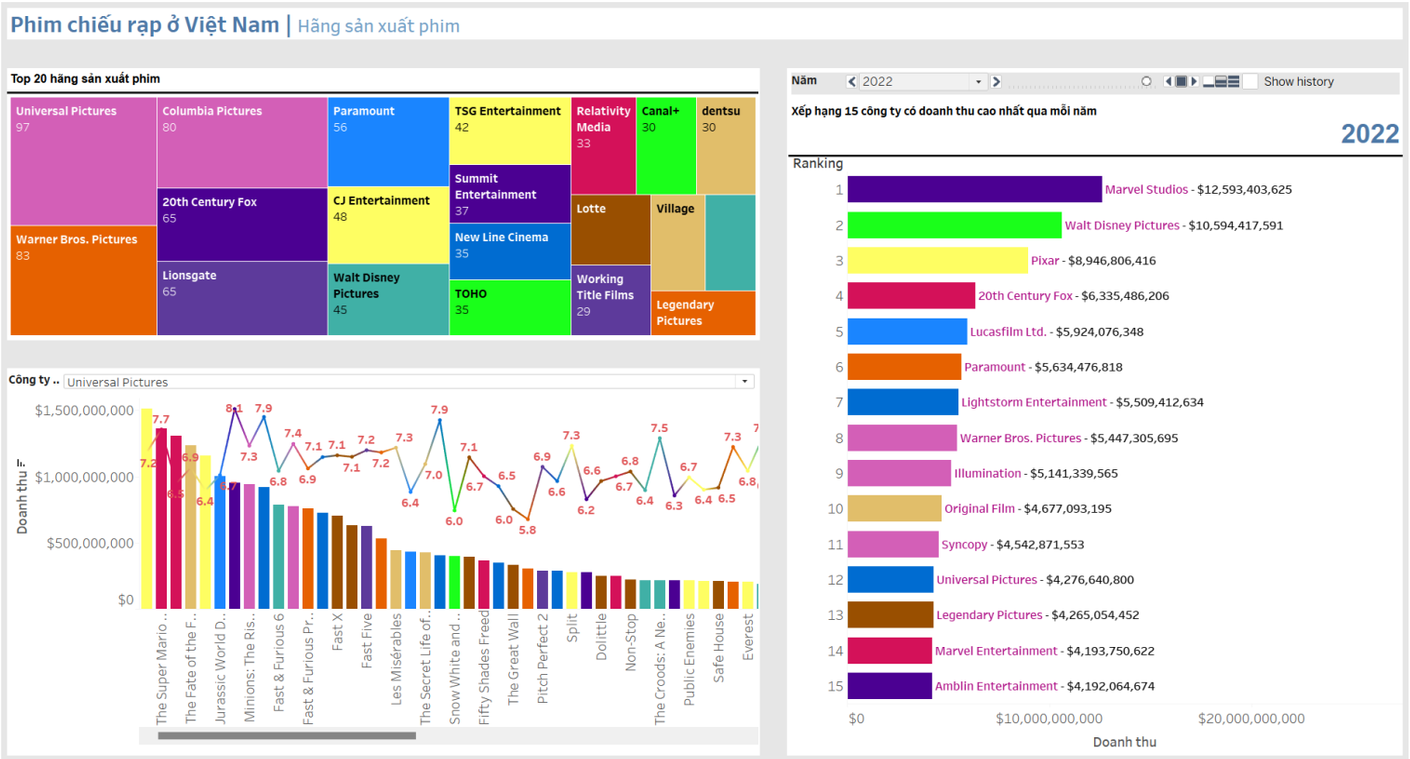
6 Dashboard cho người mù màu

- Nhóm đã tạo thêm các Dashboard tương tự như các Dashboard trên nhưng có thay đổi màu sắc để phù hợp với người mù màu.
- Nhóm sử dụng 3 dải màu từ IBM Design Library , Bang Wong, và Paul Tol để sử dụng cho dashboard. **(hình 42)**



Hình 42: Dải màu sử dụng

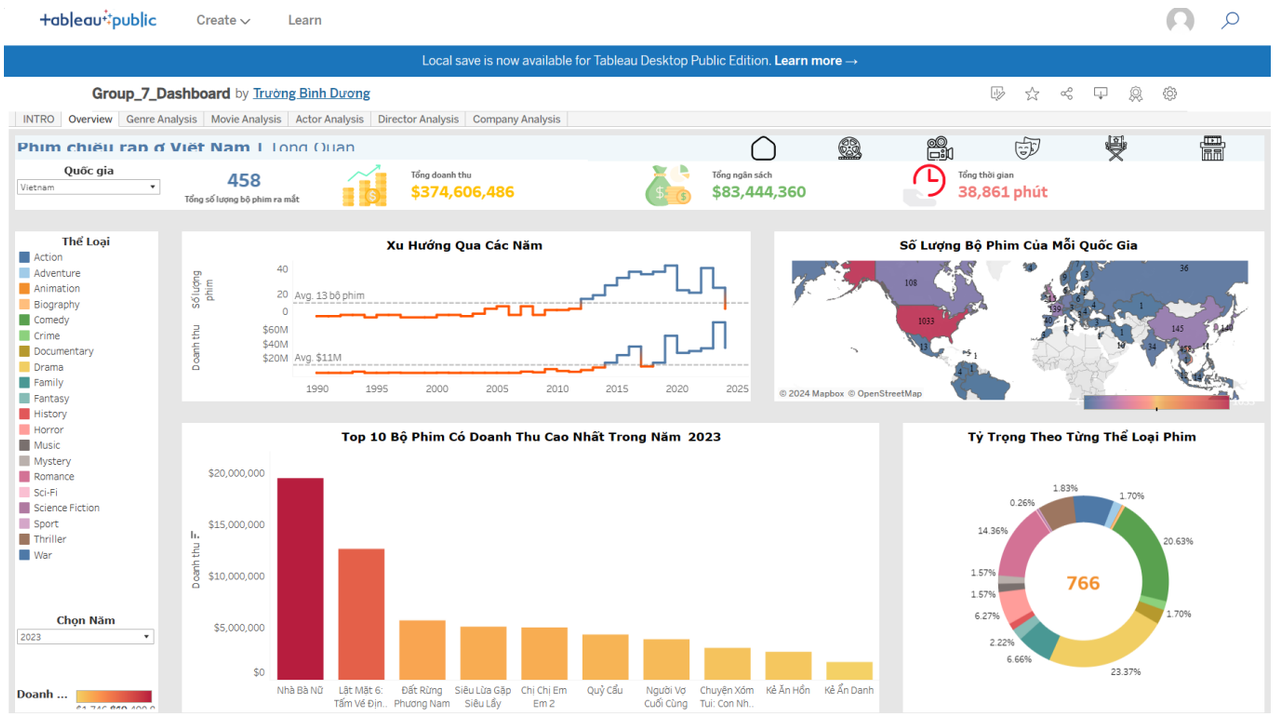
- Và dưới đây là một dashboard ví dụ cho người mù màu trong phiên bản cho người mù màu sau khi sử dụng các dải màu trên **(hình 43)**.



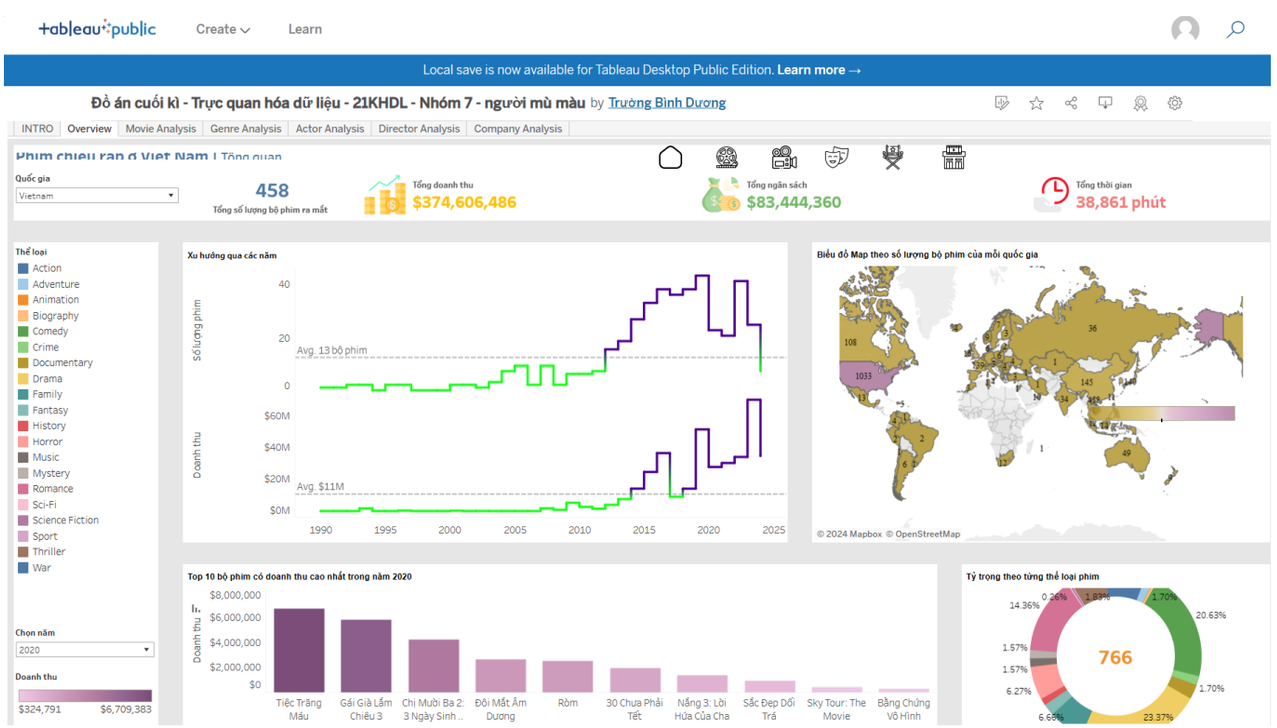
Hình 43: Dashboard Company trong phiên bản người mù màu

7 Thành phẩm

Sau khi hoàn thành, hai sản phẩm nhóm đã được chia sẻ lên Tableau Public để người dùng có thể xem dễ dàng qua link phiên bản chính và phiên bản cho người mù màu. Và đây là hình ảnh sản phẩm sau khi public



Hình 44: Dashboard chính



Hình 45: Dashboard dành cho người mù màu

8 Các tiêu chí đánh giá

STT	Tiêu chí	Mô tả	Đánh giá
1	Kết hợp nguồn dữ liệu đáng tin cậy	• Nhóm crawl dữ liệu từ trang Moveek. • Sử dụng API của TMDB • Hoàn thiện dữ liệu bằng cách crawl từ trang IMDB.	✓
2	Phù hợp với mục đích	Nhóm đã chọn đa dạng các loại biểu đồ phù hợp với mục đích trực quan cho các đối tượng người dùng: • Biểu đồ đường: thể hiện xu hướng, có thể so sánh số lượng theo thời gian •Biểu đồ map thể hiện số lượng phim theo mỗi quốc gia • Biểu đồ donut: thể hiện tỷ trọng giữa các thể loại theo thể loại trò chơi. • Biểu đồ cột: so sánh doanh thu, điểm bình chọn cho mỗi phim. • Biểu đồ bubble: thể hiện top các thể loại phổ biến • Biểu đồ histogram: thể hiện phân phối của điểm bình chọn phim. •Biểu đồ boxplot: so sánh phân bố điểm bình chọn trung bình giữa các thể loại •Biểu đồ đường kết hợp với cột: minh họa mối quan hệ giữa hai thuộc tính và xem xét chất lượng của phim • Biểu đồ bump: thể hiện xu hướng về thể loại qua từng năm • Biểu đồ gauge: thể hiện tỷ trọng doanh thu của đạo diễn so với toàn thị trường • Biểu đồ treemap: trực quan các công ty theo số lượng phim	✓
3	Rõ ràng và dễ hiểu	Biểu đồ và đồ thị được thiết kế một cách rõ ràng và dễ hiểu, với cung cấp chú thích và đơn vị đo lường rõ ràng, giúp người xem có thể nhanh chóng nắm bắt được thông tin.	✓
4	Sự tích hợp và liên kết	Các biểu đồ được tích hợp và liên kết một cách trực quan với các biểu đồ khác, nhằm có cái nhìn toàn diện đối với khu vực đó. Mối quan hệ giữa các biểu đồ được làm rõ qua dashboard	✓

5	Phân tích được sự thay đổi & xu hướng	Trực quan hóa đã thể hiện rõ ràng sự thay đổi số phim được ra mắt và doanh thu theo thời gian và mối quan hệ giữa các biến được làm rõ. Có đưa ra các phân tích, nhận xét và các câu chuyện liên quan về dữ liệu	✓
6	Tương tác và điều hướng	Sự tương tác được tích hợp cho phép người xem khám phá, tùy chỉnh chọn để có thể quan sát doanh thu từng phim theo từng diễn viên, công ty hoặc theo từng năm .	✓
7	Thiết kế hấp dẫn	Dashboard được thiết kế với bố cục hợp lí, cỡ chữ to rõ. Nhóm đã chọn những màu phù hợp để làm dịu mắt và thu hút người xem.	✓

9 Đề xuất cải tiến

Dưới đây là một số đề xuất của nhóm để cải thiện dự án trực quan hóa dữ liệu về phim chiếu rạp ở Việt Nam, đồng thời tạo ra một trải nghiệm phong phú và tương tác hơn cho người dùng, cung cấp cái nhìn toàn diện hơn về thị trường phim điện ảnh, cụ thể:

1. **Tăng cường tích hợp dữ liệu từ nguồn bên ngoài:** Ngoài việc thu thập dữ liệu từ các trang web và diễn đàn, nên cân nhắc tích hợp dữ liệu từ các nguồn khác như dữ liệu từ nền tảng mạng xã hội *(ví dụ: Twitter, Facebook, v.v.)* để hiểu sự phản ứng của cộng đồng đối với các bộ phim chiếu rạp cụ thể. Sự kết hợp này sẽ cung cấp cái nhìn toàn diện hơn về hiệu suất của các bộ phim chiếu rạp ở Việt Nam.

2. **Phân tích sâu hơn về xu hướng và mô hình dự đoán:** Sử dụng các kỹ thuật phân tích dự đoán và mô hình hóa để đề xuất các xu hướng phim trong tương lai. Điều này có thể bao gồm việc áp dụng các mô hình học máy để dự đoán doanh thu phim dựa trên các yếu tố như thời gian phát hành, thể loại phim, hãng phim và đánh giá từ cộng đồng người xem phim.

3. **Tích hợp hệ thống gợi ý và phản hồi từ người dùng:** Xây dựng một hệ thống gợi ý thông minh dựa trên dữ liệu lịch sử và hành vi của người xem. Hệ thống này có thể gợi ý các bộ phim chiếu rạp ở Việt Nam tương tự hoặc phản hồi trực tiếp từ người dùng về trải nghiệm của họ với các bộ phim cụ thể. Điều này sẽ giúp cải thiện tương tác với người dùng và tạo ra một cộng đồng phản hồi tích cực.

4. **Thiết kế luồng dữ liệu tự động cập nhật**: Xây dựng luồng dữ liệu có thể cập nhật và đồng bộ với dashboard để dữ liệu được trực quan luôn chính xác và mới nhất.

Tài liệu tham khảo

[1] Wexler, S., Cotgreave, A., & Shaffer, J. (2017). *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios*. Wiley.

[2] Chuc Nguyen Van, "Data Analysis and Visualization with Tableau", [Youtube Playlist], https://www.youtube.com/watch?v=kCKJQPSmmCQ&list=PL2vTwB9LeGfD1K6nNVESsrg0pz_FlOEoh&pp=iAQB

[3] Oppenheimer đạt 7 giải thưởng oscar (Trích: https://www.aa.com.tr/en/world/oppenheimer-dominates-96th-academy-awards-winning-7-oscars/3161131)

[4] Kì vọng phim điện ảnh Việt Nam đạt 1000 tỷ doanh thu (Trích: https://vietnamnet.vn/mo-ve-phim-viet-doanh-thu-1-000-ty-dong-2275367.html)

[5] Phim "Bố Già"của Trấn Thành mặc dù được doanh thu cao nhưng bị đánh giá thấp ở nước ngoài (Trích: https://plo.vn/bo-gia-cua-tran-thanh-nhan-diem-thap-o-nuoc-ngoai-nha-phe-binh-phim-noi-gi-post668585.html)

[6] Lan Ngọc, Đình Toàn đoạt giải Cánh Diều Vàng (Trích: https://vnexpress.net/lan-ngoc-dinh-toan-doat-giai-canh-dieu-vang-1912145.html)

[7] Phim Victor Vũ đoạt 8 giải Cánh Diều (Trích: https://vnexpress.net/phim-victor-vu-doat-8-giai-canh-dieu-2432777.html)

[8] Danh sách các phim có doanh thu cao nhất trên thế giới (Trích: https://vi.wikipedia.org/wiki/Danh_s%C3%A1ch_phim_c%C3%B3_doanh_thu_cao_nh%E1%BA%A5t)

[9] Danh sách các phim có doanh thu cao nhất Việt Nam (Trích: https://vi.wikipedia.org/wiki/Danh_s%C3%A1ch_phim_%C4%91i%E1%BB%87n_%E1%BA%A3nh_Vi%E1%BB%87t_Nam_c%C3%B3_doanh_thu_cao_nh%E1%BA%A5t)