

Audience Preference Analysis

Trends in Movies
from 2014 to 2024

Reporter: HE Shuqing



CONTENTS

This project examines movie popularity trends from 2014 to 2024 by analyzing IMDb data and audience interests in movie themes through Reddit comments

Introduction

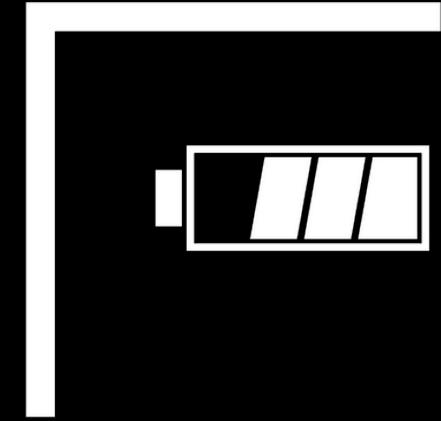
- *Project Objectives*
- *Research Questions*

Changes in Film Popularity Trends

- *Data Collection*
- *Data Cleaning*
- *Data Analysis*
- *Summary*

Changes in audience attention in film themes

- *Data Collection*
- *Data Cleaning & Text Enhancement*
- *Topic Modeling (LDA)*
- *Topic Interpretation & Matrix Construction*
- *Data Analysis*
- *Summary*



INTRODUCTION

Project Objectives

Movie Market



Focus on overall trends in the movie market

Audience Interest



Focus on changes in topics of interest among audiences

Research Questions

1. Changes in Film Popularity Trends:

What significant changes can be identified through analyzing **IMDb data** on box office revenue, ratings, and genre evolution?

2. Changes in Audience Attention to Film Themes:

What topics have audiences focused on each year regarding the film market, based on **Reddit comment texts**?



CHANGES IN FILM + POPULARITY TRENDS

Data Collection

CSV file loaded successfully.
Processing movies from the year 2014...
Found 400 movies from the year 2014.
Processing movie 1/400: X-Men: Days of Future Past
Processing movie 2/400: If I Stay
Processing movie 3/400: Leviathan
Processing movie 4/400: The Purge: Anarchy
Processing movie 5/400: Adult World
Processing movie 6/400: Batman: Assault on Arkham
Processing movie 7/400: Get on Up
Processing movie 8/400: Doraemon: New Nobita's Great Demon – Peko and the Exploration Party of Five
Processing movie 9/400: Godzilla
Processing movie 10/400: Gun Woman
Processing movie 11/400: The Town that Dreaded Sundown
Processing movie 12/400: Dolphin Tale 2
Processing movie 13/400: What We Did on Our Holiday
Processing movie 14/400: Transcendence
Processing movie 15/400: Zombie Fight Club
Processing movie 16/400: Kill the Messenger
Processing movie 17/400: The Salvation
Processing movie 18/400: Birdman or (The Unexpected Virtue of Ignorance)
Processing movie 19/400: The Captive
Processing movie 20/400: Alexander and the Terrible, Horrible, No Good, Very Bad Day
Processing movie 21/400: Beyond the Reach
Processing movie 22/400: Whiplash
Processing movie 23/400: About Last Night
Processing movie 24/400: The Boxtrolls
Processing movie 25/400: Before I Go to Sleep
Processing movie 26/400: That Awkward Moment
Processing movie 27/400: Spring

Overview:

- **Data source:** TMDb API and Cinemagoer
- **Valid date:** 2014 to 2024
- **Sample Size:** 4,400 films (400 films per year)

	Title	Year	IMDb ID	Director	Rating	Votes	Box Office	Budget	Genres	Actors	Languages	Countries
0	X-Men: Days of Future Past	2014.0	1877832	Bryan Singer	7.9	765945.0	\$748,121,534.28 Feb 2015	\$200,000,000 (estimated)	Action, Adventure, Sci-Fi, Thriller	Hugh Jackman, James McAvoy, Michael Fassbender...	English, Vietnamese, French	United Kingdom, United States, Canada
1	If I Stay	2014.0	1355630	R.J. Cutler	6.7	129700.0	\$78,874,843	\$11,000,000 (estimated)	Drama, Fantasy, Music, Romance, Sci-Fi	Chloë Grace Moretz, Mireille Enos, Jamie Black...	English	United States
2	Leviathan	1989.0	97737	George P. Cosmatos	5.8	25227.0	NaN	\$25,000,000 (estimated)	Adventure, Horror, Mystery, Sci-Fi, Thriller	Peter Weller, Richard Crenna, Amanda Pays, Dan...	English	Italy, United States
3	The Purge: Anarchy	2014.0	2975578	James DeMonaco	6.4	167191.0	\$110,602,999	\$9,000,000 (estimated)	Action, Horror, Sci-Fi, Thriller	Frank Grillo, Carmen Ejogo, Zach Gilford, Kiel...	English	United States, France
4	Adult World	2013.0	1067765	Scott Coffey	6.1	18975.0	NaN	\$4,000,000 (estimated)	Comedy, Drama	Emma Roberts, Summer Shelton, Chris Riggie, Sha...	English	United States
...
4383	The First Omen	2024.0	5672290	Arkasha Stevenson	6.5	62703.0	\$50,760,266	\$30,000,000 (estimated)	Horror	Nell Tiger Free, Ralph Ineson, Sonia Braga, Ta...	English, Italian	United States, Italy, Serbia, Canada
4384	Someone Like You	2024.0	22334258	Tyler Russell	5.8	1962.0	NaN	NaN	Romance	Sarah Fisher, Jake Allyn, Lynn Collins, Robyn ...	English	United States
4385	Monster Summer	2024.0	3954936	David Henrie	5.7	3359.0	NaN	NaN	Adventure, Fantasy, Mystery	Mel Gibson, Mason Thames, Julian Lerner, Abby ...	English	United States
4386	Juror #2	2024.0	27403986	Clint Eastwood	7.0	87604.0	NaN	\$30,000,000 (estimated)	Crime, Drama, Mystery, Thriller	Nicholas Hoult, Zoey Deutch, Megan Mieduch, To...	English	United States
4387	Wolfs	2024.0	14257582	Jon Watts	6.5	76575.0	NaN	NaN	Crime, Thriller	Amy Ryan, Austin Abrams, George Clooney, Brad ...	English	United Kingdom, United States

4388 rows × 12 columns

Methods:

- **TMDb API:** Obtain movie titles and IDs
- **IMDb Data:**
 - Matched titles from TMDb with IMDb using Cinemagoer
 - Gathered data on *IMDb ratings, votes, production budgets, box office revenue, genre combinations*

Data Cleaning

Steps :

I. Data Filtering and Initial Processing:

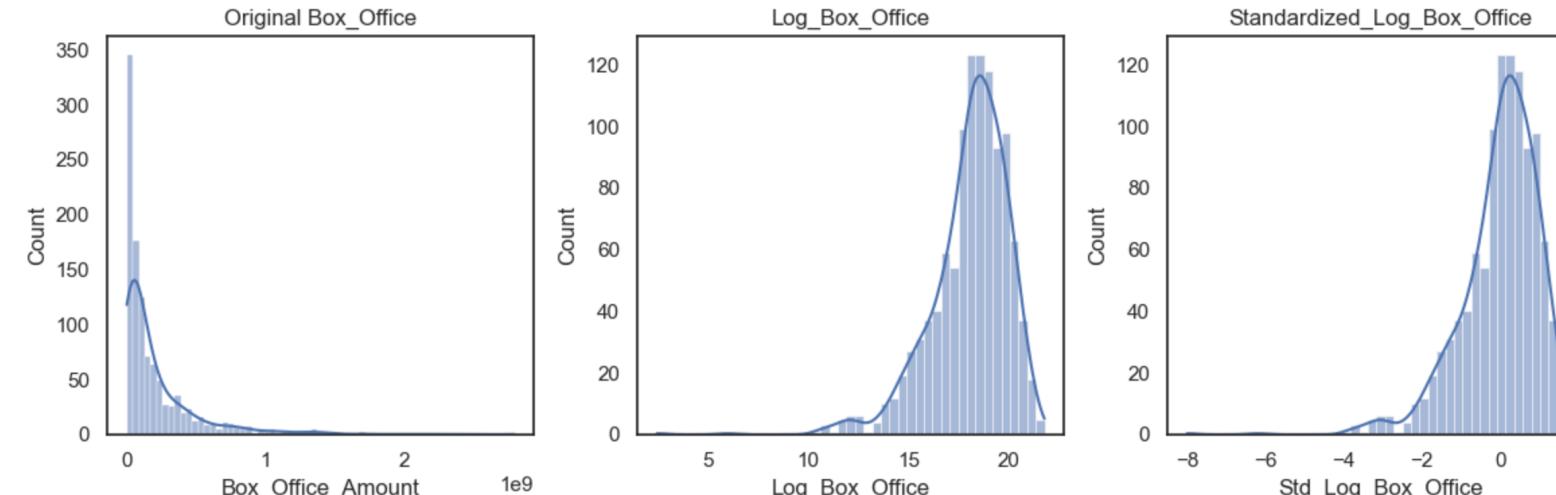
- Limited dataset to films from **2014 to 2024**
- Extracted and converted Budget and Box Office into new columns: **Budget_Amount** and **Box_Office_Amount**

2. Handling Missing Values:

- Filled missing Budget_Amount values with the median by year
- Removed other remaining missing values

3. Data Standardization and Optimization:

- Applied logarithmic transformation to box office data to address right skewness, standardizing values for normal distribution



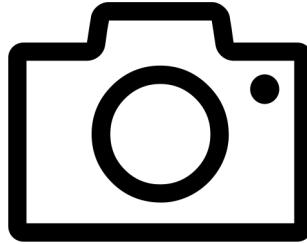
Overview:

Resulted in **1,095** valid movie records for analysis.

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 1095 entries, (2014, 0) to (2024, 4383)
Data columns (total 11 columns):
 #   Column           Non-Null Count   Dtype  
--- 
 0   Title            1095 non-null    object 
 1   Year             1095 non-null    int64  
 2   Director         1095 non-null    object 
 3   Rating           1095 non-null    float64
 4   Votes            1095 non-null    int64  
 5   Genres           1095 non-null    object 
 6   Actors           1095 non-null    object 
 7   Languages        1095 non-null    object 
 8   Countries        1095 non-null    object 
 9   Budget_Amount    1095 non-null    int64  
 10  Box_Office_Amount 1095 non-null    int64  
dtypes: float64(1), int64(4), object(6)
```

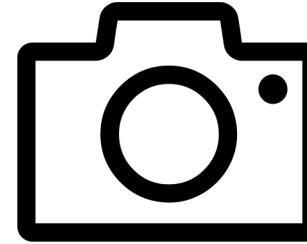
Data Analysis

3 Key Angles to Decode Decade-Long Trends



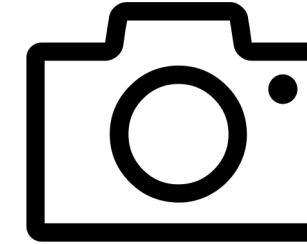
Box Office

How did the box office change before and after the pandemic?



Ratings

How have movie ratings changed over the past decade?



Genres

Which genres became more popular?

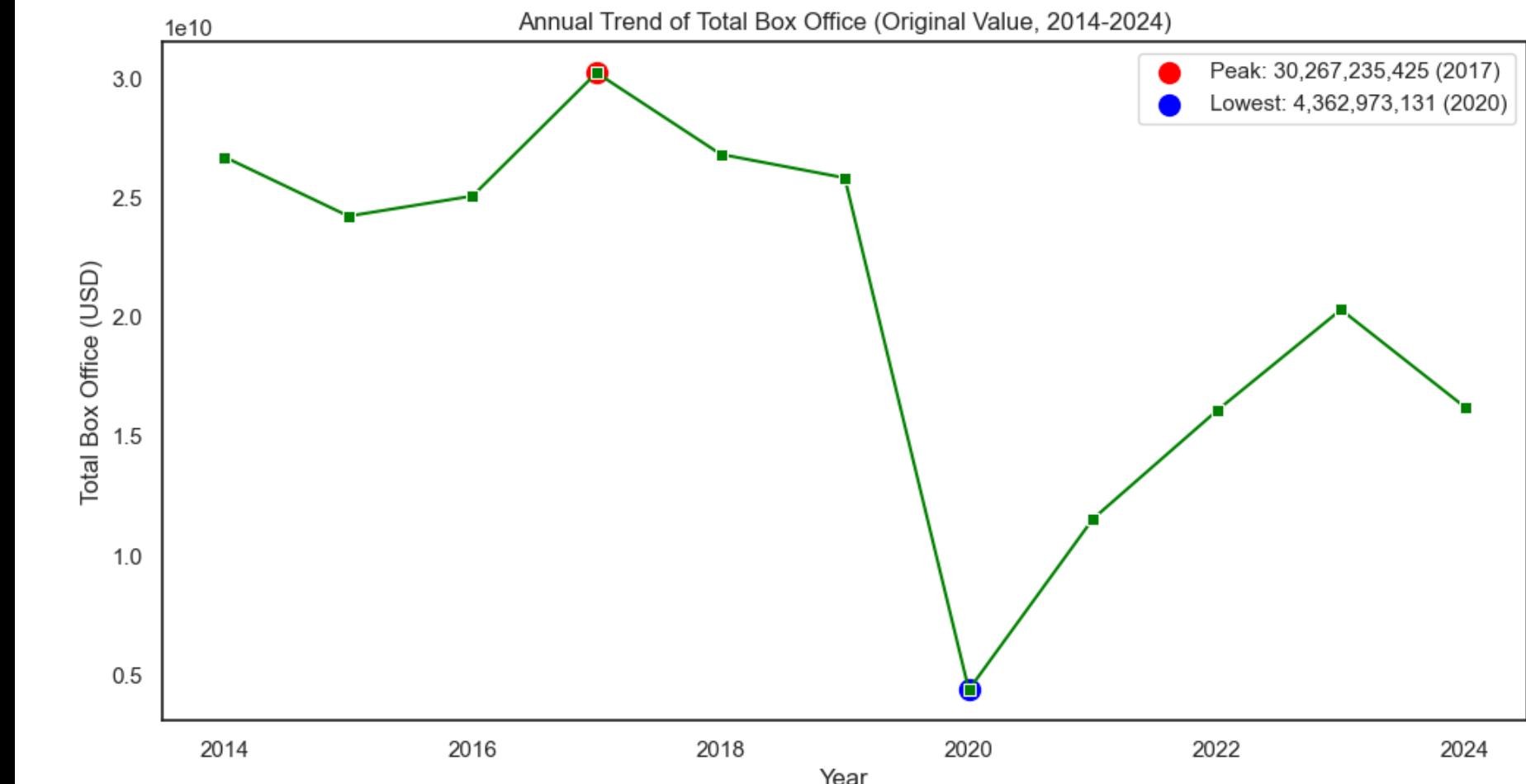
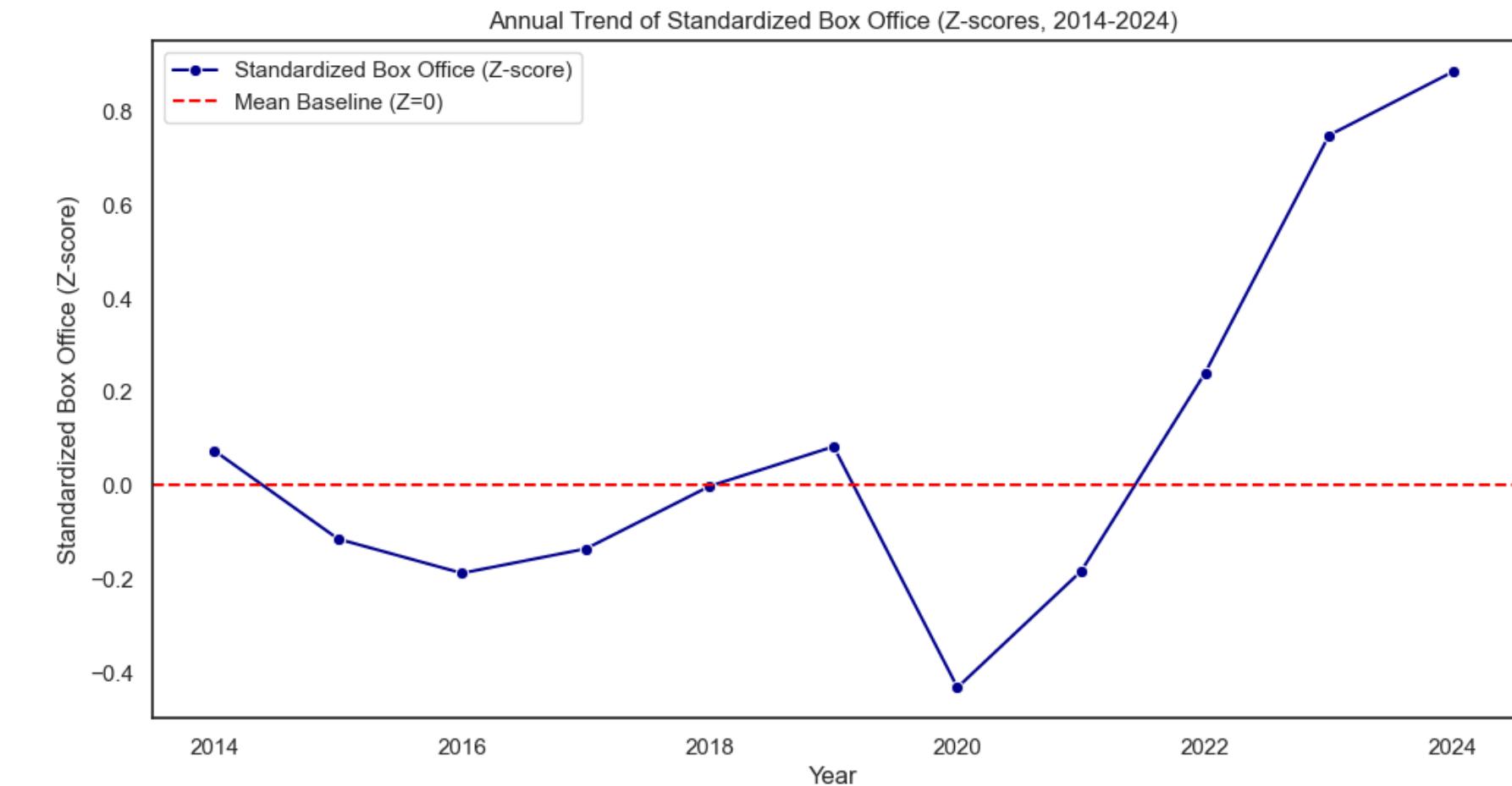
Box Office Trends

“Blockbusters Got Richer, Industry Got Poorer”

- Individual film profitability ↑ [2023-24 Z-scores >0.7]
- Total industry output ↓ [2023 = 67% of 2017 production capacity]

Metric	2020 Trough	2024 Peak
Standardized Z	-0.43	+0.89
Total Revenue	\$4.36B (17% of 2019)	\$20.33B(2023) 2017 peak = 67% recovery

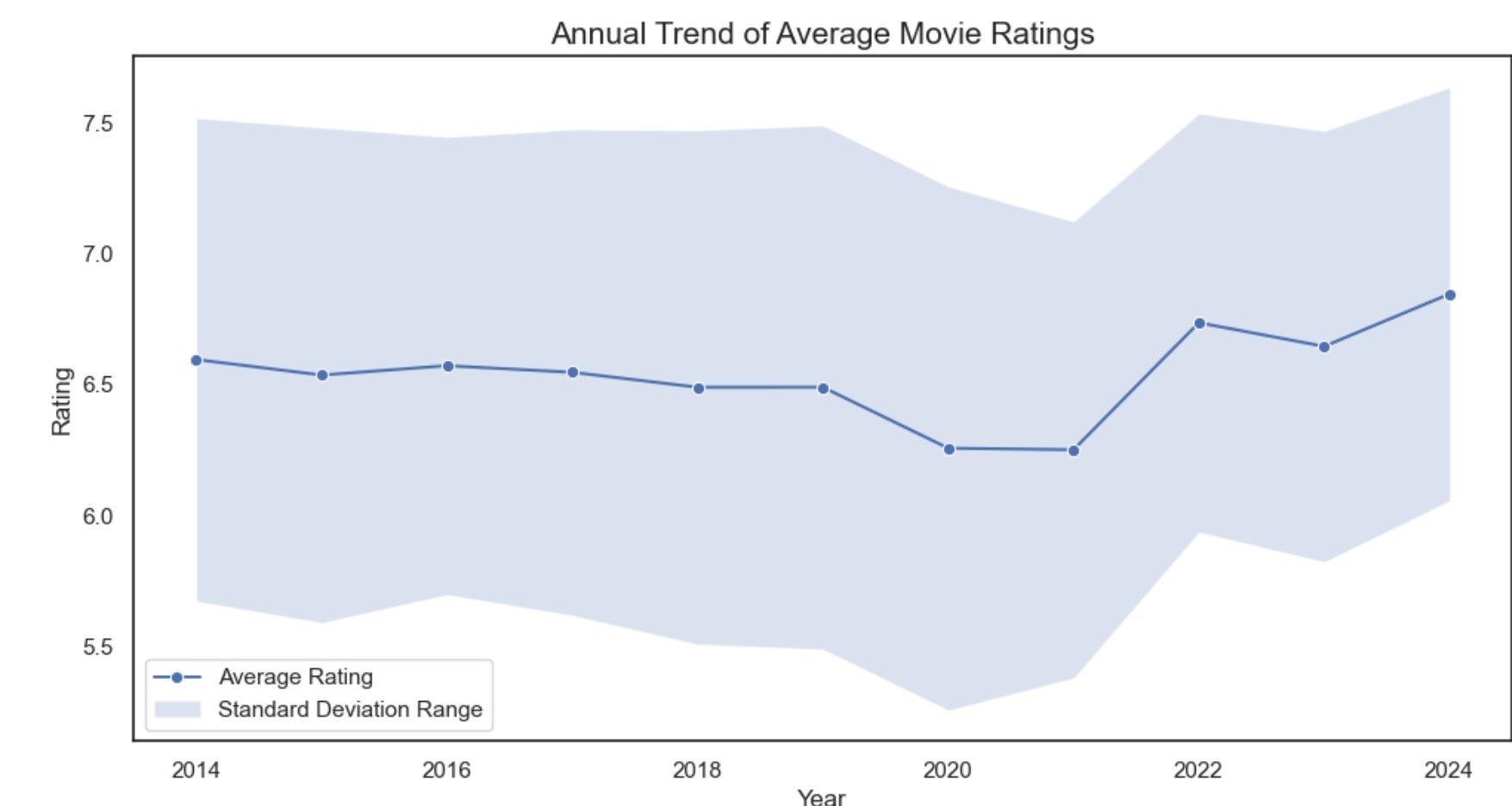
Year	Total Box Office (USD)	Change Rate (YoY)
2017	30,267,235,425	+16.8% (vs 2016)
2020	4,362,973,131	-83.1% (vs 2019)
2023	20,325,215,368	+26.3% (vs 2022)



Movie Rating Trends

“Stable Quality, Unchanged Standards”

- 10-year avg. change: +0.012/year *[$p=0.495$, NS]*



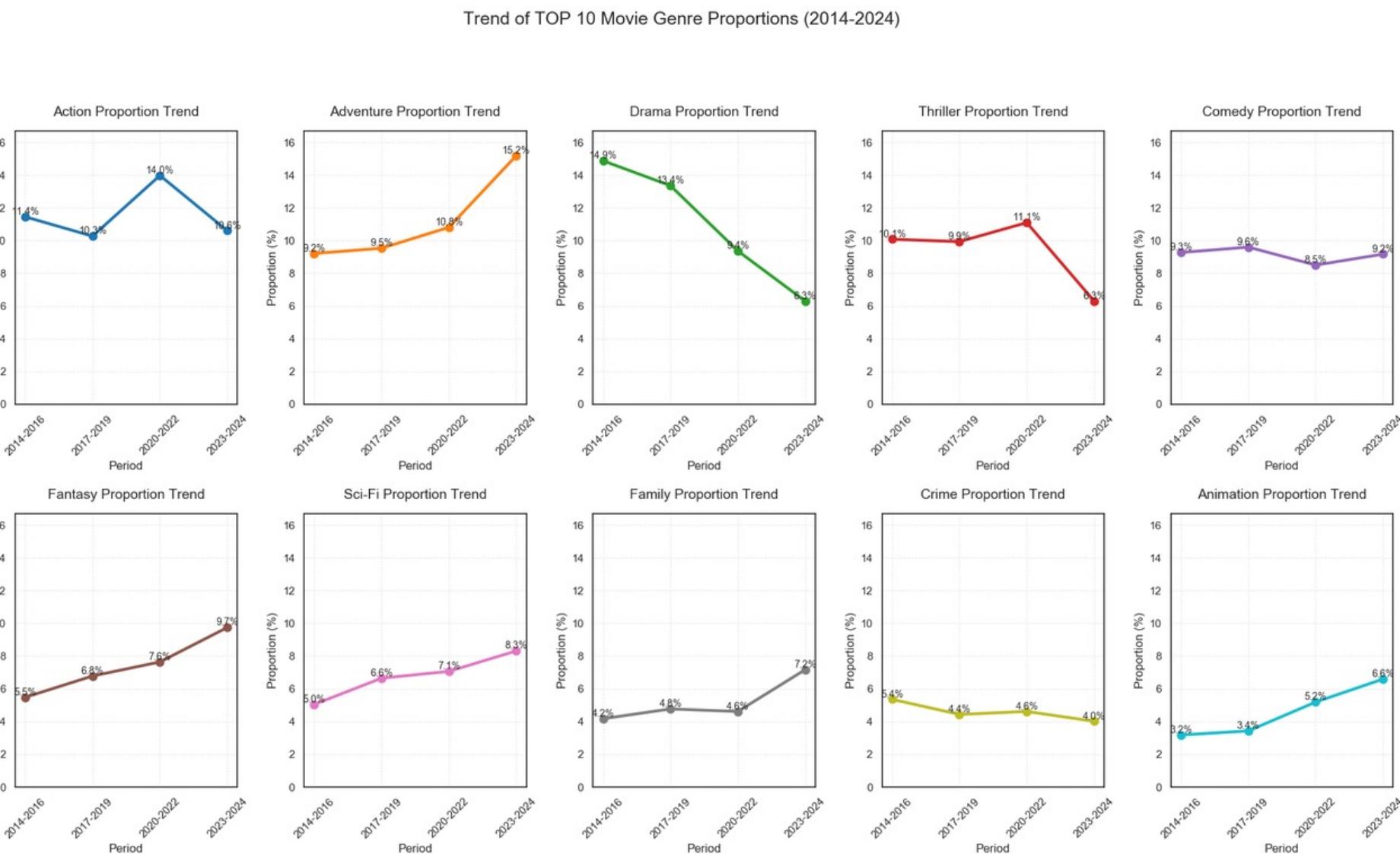
Movie Genres Evolution

“Market Share”

- Animation **▲106%** (Disney+ shadow effect)
- Action spiked to 14% in 2020 (escape reality demand)

Genres	Action	Adventure	Drama	Thriller	Comedy	Fantasy	Sci-Fi	Family	Crime	Animation
Period										
2014-2016	11.4%	9.2%	14.9%	10.1%	9.3%	5.5%	5.0%	4.2%	5.4%	3.2%
2017-2019	10.3%	9.5%	13.4%	9.9%	9.6%	6.8%	6.6%	4.8%	4.4%	3.4%
2020-2022	14.0%	10.8%	9.4%	11.1%	8.5%	7.6%	7.1%	4.6%	4.6%	5.2%
2023-2024	10.6%	15.2%	6.3%	6.3%	9.2%	9.7%	8.3%	7.2%	4.0%	6.6%

Trend of TOP 10 Movie Genre Proportions (2014-2024)

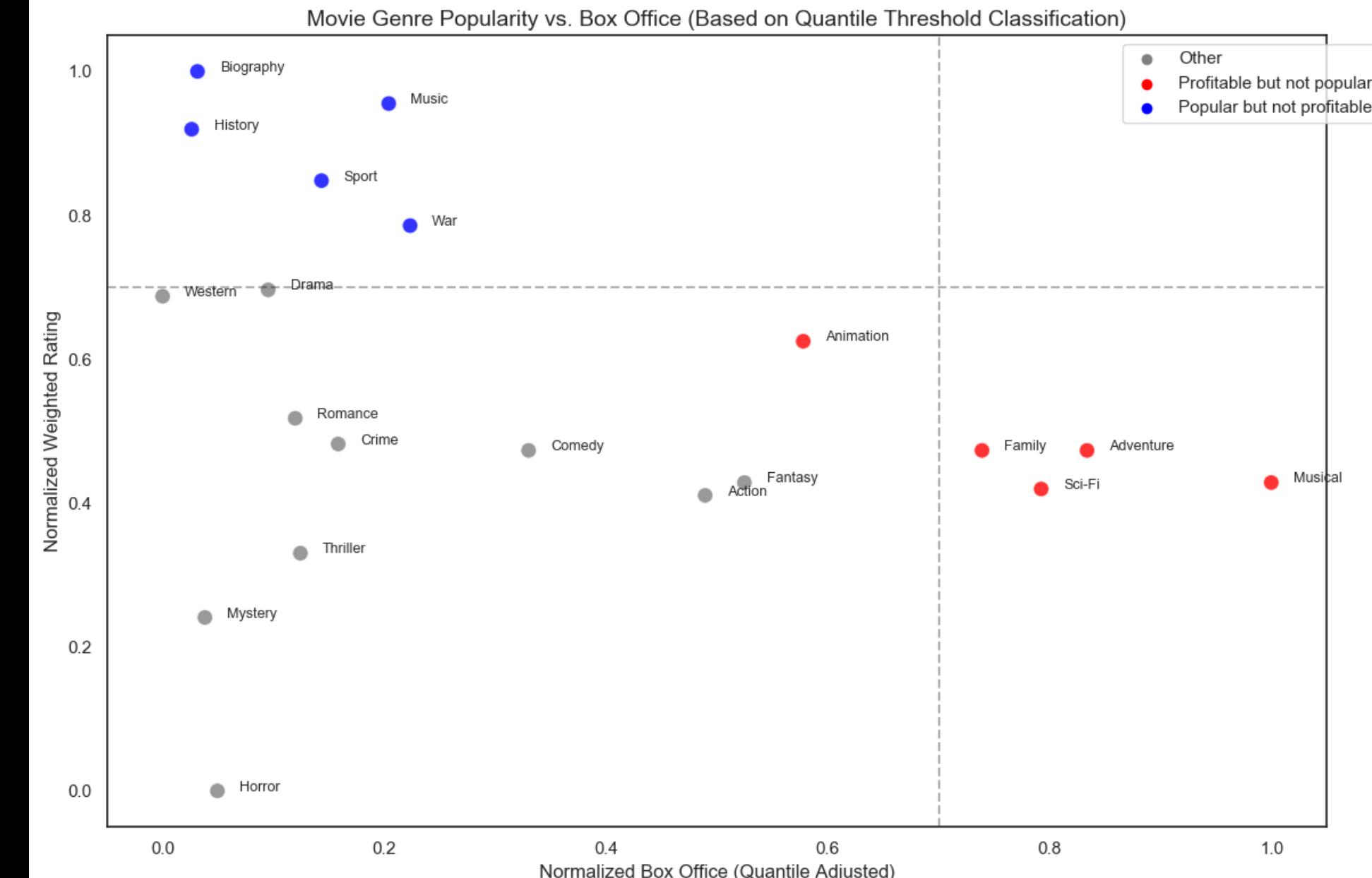


Movie Genres Evolution

“Art vs Commerce”

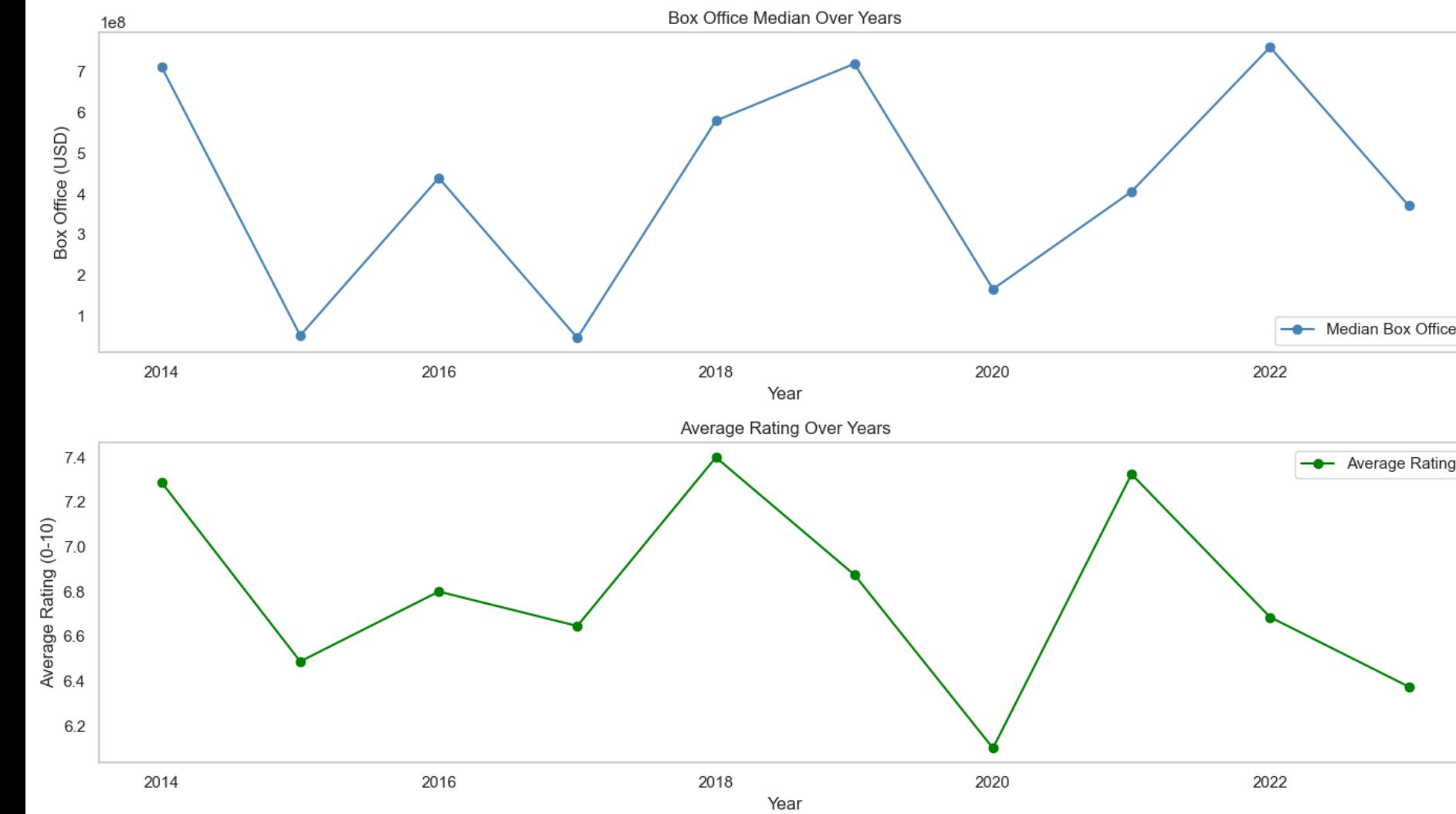
[Popular but Not Profitable] $\leftarrow \bullet \rightarrow$ [Profitable but Not Popular]
 Biographical/Musical [>7.0] Family/Adventure [$6.4-6.55$]

Genre	Rating	Median Box Office
Biography	7.14	\$50.37 million
Music	7.09	\$86.17 million
Family	6.55	\$197 million
Adventure	6.55	\$217 million



Superhero Fatigue

Era	Median Box Office	Avg Rating
Peak (2014-2019)	\$710M	7.40
2023	\$372M ▼51%	6.38 ▼12.5%



Summary

1. Winner Takes All:

Fewer films now earn more money.

2. Streaming Effect:

Platforms like Disney+ directly fueled a surge in animation and family films.

3. Pandemic Psychology:

During lockdowns, audiences craved action films. ("emergency" consumption nature)

4. Superhero Fatigue:

A dual crisis--box office drop and rating decline.



● REC

CHANGES IN AUDIENCE

+

ATTENTION IN FILM THEMES

Data Collection

	Movie Title	Year	Post Title	Post Author	Post Time	Post Score	Comment	Comment Author	Comment Time	Comment Score
0	X-Men: Days of Future Past	2014	X-Men: Days of Future Past Trailer 2	Looperinator	2014-03-24 12:56:19	3165.0	I was excited to see some colossus love	PillsberryDboy4	2014-03-24 13:47:42	600.0
1	X-Men: Days of Future Past	2014	X-Men: Days of Future Past Trailer 2	Looperinator	2014-03-24 12:56:19	3165.0	"You'll need to go into the past, to fix every...	notshawnvaughn	2014-03-24 15:54:08	506.0
2	X-Men: Days of Future Past	2014	X-Men: Days of Future Past Trailer 2	Looperinator	2014-03-24 12:56:19	3165.0	Anna Paquin is still being listed as a cast me...	annachie-gordon	2014-03-24 13:58:09	114.0
3	X-Men: Days of Future Past	2014	X-Men: Days of Future Past Trailer 2	Looperinator	2014-03-24 12:56:19	3165.0	Was that Storm getting killed by a sentinel? I...	Moussekateer	2014-03-24 13:13:31	1095.0
4	X-Men: Days of Future Past	2014	X-Men: Days of Future Past Trailer 2	Looperinator	2014-03-24 12:56:19	3165.0	Superhero movies seem to really love football ...	NaN	2014-03-24 13:15:35	300.0
...
357554	The First Omen	2024	'The Omen' Prequel 'The First Omen' In the Wor...	MarvelsGrantMan136	2022-05-10 22:05:44	29.0	No I'm sorry you're too late, this isn't 2014 ...	Mission-Two1325	2022-05-11 06:29:25	2.0
357555	The First Omen	2024	'The Omen' Prequel 'The First Omen' In the Wor...	MarvelsGrantMan136	2022-05-10 22:05:44	29.0	Disney owns Fox now, so it will be "Omen The M...	whitehavenmansions	2022-05-11 15:06:06	1.0
357556	The First Omen	2024	What would a future sequel of the first omen l...	trover2345325	2024-04-14 15:37:06	0.0	I don't know , but I hope Russel Crowe is goin...	ferrarinobrakes	2024-04-14 19:22:47	2.0
357557	The First Omen	2024	The First Omen Official Trailer 20th Centu...	naaz0412	2024-01-03 19:41:57	6.0	Great trailer. Had my attention for sure.	Bubba_Blade	2024-01-03 23:15:46	7.0
357558	The First Omen	2024	The First Omen Official Trailer 20th Centu...	naaz0412	2024-01-03 19:41:57	6.0	Is she typecast now? First was the servant an...	messengers1	2024-01-04 07:43:02	3.0

357559 rows x 10 columns

Overview:

- Data source:** Reddit comment data
- Valid date:** 2014 to 2024
- Sample Size:** Approximately 300,000 comments (50 posts per valid movie)

Methods:

- PRAW Library:** Ensures legal access to Reddit data
- Post Traversing:** Searched the "movies" subreddit using movie titles.
- Post Filtering:**
 - Were created within the defined time range.
 - Contained the movie title in the title.
- Comment Extraction:**
 - High Interaction: 40 comments if comments ≥ 100 .
 - Medium Interaction: 20 comments if comments ≥ 50 .
 - Low Interaction: 10 comments if comments < 50 .

Data Cleaning & Text Enhancement

Steps :

I. Data Cleaning :

- Filtered comments from **2014 to 2024** and made sure they are no earlier than the year the film was released
- Removed rows with missing data in key fields
- Retained the highest-scoring comment per author in each movie
- Created a new column 'Comment Year' from Comment Time
- Defined a function to remove HTML tags, URLs, and empty comments

Data filtering summary:
Original records: 262005
Filtered records: 222792
Records removed: 39213

Year validation:
Minimum release year: 2014
Maximum comment year: 2024
Comments before movie release: 0 records

Initial Quality Metrics:
contains_url 0.0%
html_tags 0.0%
non_ascii 0.6%
excessive_spaces 10.0%
dtype: object

Quality Metrics After Processing:
contains_url 0.0%
html_tags 0.0%
non_ascii 0.6%
excessive_spaces 0.0%
dtype: object

Overview:

Resulted in **22,2520** valid movie comment records

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 222520 entries, 0 to 222519
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   Movie Title      222520 non-null   object 
 1   Year              222520 non-null   int64  
 2   Post Title        222520 non-null   object 
 3   Post Author       222520 non-null   object 
 4   Post Time         222520 non-null   object 
 5   Post Score        222520 non-null   float64
 6   Comment           222520 non-null   object 
 7   Comment Author    222520 non-null   object 
 8   Comment Time      222520 non-null   object 
 9   Comment Score     222520 non-null   float64
 10  Comment Year      222520 non-null   int64  
dtypes: float64(2), int64(2), object(7)
memory usage: 18.7+ MB
```

Data Cleaning & Text Enhancement

Steps :

2. Text Enhancement :

- **NLP Pipeline Setup:**
 - spaCy's English model for core NLP tasks
 - NLTK for lemmatization and stopword removal
- **Cleaning Enhancement:**
 - Removed noise like HTML, URLs and emojis
 - Standardized text by expanding contractions ("can't" → "cannot")
- **Domain Protection:**
 - Franchise names (MCU → Marvel Cinematic Universe)
 - Technical terms like "CGI" and "frame rate"
- **Linguistic Optimization:**
 - Meaningful word types (nouns, verbs, adjectives)
 - Removed fillers like articles and pronouns

== Processed Samples ==

Original: All these shaky camera movies set in a house run together in my mind. I cant even remember what The Conjuring was about. I dont even know if I've seen Conjuring 2 or not. These shaky horror movies with the black and white scenes in bedrooms with people sleeping in bed for ten minutes are all the same. And then a door swings open. Oh, big deal, I'm so fucking afraid.

Processed: shaky camera movies set house run together mind even remember conjuring even see conjure shaky horror movies black white scenes bedrooms people sleep bed minutes door swing open big deal fuck afraid

Overview:

Resulted in **22,2520** valid movie comment records

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 222520 entries, 0 to 222519
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
---  -- 
 0   Movie Title      222520 non-null   object 
 1   Year              222520 non-null   int64  
 2   Post Title        222520 non-null   object 
 3   Post Author       222520 non-null   object 
 4   Post Time         222520 non-null   object 
 5   Post Score        222520 non-null   float64
 6   Comment           222520 non-null   object 
 7   Comment Author    222520 non-null   object 
 8   Comment Time      222520 non-null   object 
 9   Comment Score     222520 non-null   float64
 10  Comment Year      222520 non-null   int64  
dtypes: float64(2), int64(2), object(7)
memory usage: 18.7+ MB
```

Topic Modeling with LDA

Steps :

- **Longitudinal LDA Approach:**
 - Annual text segmentation → tokenized docs
 - Gensim Dictionary for document-term matrices
- **Model Training:**
 - 5-10 potential topics per year
 - Coherence scoring (c_v: 0.5-0.7)
- **Topic Analysis:**
 - Top 10 representative terms per topic
 - Document-topic probability distributions
 - Dominant topic assignments for each comment

```
== 2022 (Optimal Topics: 8, Coherence: 0.585) ==
Topic 1: poster, michael, friend, george, mirabel, christmas, minions, listen, bay, wear
Topic 2: marvel_cinematic_universe, cinematic_universe, office, box, rock, universe, batman, likely, adam, eye
Topic 3: strange, star, fast, dr, YEAR_2022, suicide, kingdom, squad, posters, dwayne
Topic 4: try, people, end, fuck, use, take, family, life, point, kill
Topic 5: imax, worse, months, animate, avatar, 3d_technology, okay, odd, musical, thor
Topic 6: james, human, sonic, bruno, onion, art, wild, lee, chris, door
Topic 7: history, elvis, picture, light, oscar, king, win, feature, choose, ignore
Topic 8: really, watch, good, movies, love, make, get, go, time, great
```

```
== 2023 (Optimal Topics: 9, Coherence: 0.587) ==
Topic 1: black, school, roles, cry, soundtrack, complete, jack, week, appear, multiple
Topic 2: napoleon, bourne, star_wars, jason, beat, wars, beast, roll, film, sub
Topic 3: marvel_cinematic_universe, cinematic_universe, recommend, paul, review, thanks, rudd, creed, highly, ho
Topic 4: imax, certain, fly, bigger, credits, audiences, glass, split, planet, dog
Topic 5: john, wick, wes, anderson, marketing, holy, muppets, plan, reboot, mission
Topic 6: voice, tom, david, office, dc_extended_universe, box, opening, max, cruise, lmao
Topic 7: go, people, get, want, even, take, way, end, only, make
Topic 8: look, wait, book, poster, godzilla, forward, michael, video, mario, oppenheimer
Topic 9: watch, really, good, movies, love, great, see, much, very, feel
```

```
== 2024 (Optimal Topics: 7, Coherence: 0.574) ==
Topic 1: effects, computer_generated_imagery, team, robert, david, cars, type, gun, visual_effects, oppenheimer
Topic 2: watch, movies, go, make, time, get, want, see, come, work
Topic 3: music, john, game, story, shoot, script, base, mention, play, world
Topic 4: themes, burton, jason, statham, anxiety, robot, mandalorian, fiction, tim, culture
Topic 5: love, really, good, great, much, people, look, feel, very, too
Topic 6: king, train, ryan, sell, road, lion, pay, joy, market, gosling
Topic 7: godzilla, poster, star_wars, mr, dune, shin, max, ass, amy, smith
```

Topic Interpretation & Matrix Construction

Steps :

I. Topic Interpretation :

- **Keyword Analysis:**
 - 2019 Topic with key words ("marvel", "endgame", "ironman") → *Avengers: Endgame*
- **Contextual Grounding:**
 - Topic with key words ("vfx, cgi") → *Film Technologic*
- **Three-tier taxonomy:**
 - Macro Categories (e.g., Film_Tech, Franchise)
 - Specific Labels (e.g., CGI, Superhero)
 - Franchise Series (e.g., DC, Marvel)

	Year	Topic#	Keywords	Specific Context	Series	Label	Category
0	2014	Topic 0	3d_technology, imax	3D/IMAX technology discussion	-	3D_IMAX	Film_Tech
1	2014	Topic 1	apes, planet	Dawn of the Planet of the Apes	Planet of the Apes	Sci-Fi	Franchise
2	2014	Topic 2	really, good	General positive reviews	-	Positive	Audience_Response
3	2014	Topic 3	action, fight	Action movie elements	-	Action	Genre
4	2014	Topic 4	sony, transformers	Transformers 4	Transformers	Sci-Fi	Franchise
...
82	2024	Topic 2	music, game	Video game adaptations	-	Game Adaptation	Genre
83	2024	Topic 3	robot, ai	AI-themed films	-	Sci-Fi	Genre
84	2024	Topic 4	love, feel	Emotional connections	-	Positive	Audience_Response
85	2024	Topic 5	gosling, lion	The Fall Guy	-	Comedy	Genre
86	2024	Topic 6	godzilla, dune	Godzilla x Kong 2	MonsterVerse	Kaiju	Franchise

87 rows × 7 columns

Topic Interpretation & Matrix Construction

Steps :

2. Matrix Construction :

- Data Merging:**
 - Joined LDA outputs (topic distribution) with raw comments using comment_id*
- Document-term frequency matrix (LDA):**
 - Rows = Comments*
 - Columns = Topics (e.g., "2020_Topic8")*
 - Values = Probability scores [0–1]*
- Human-Machine Mapping:**
 - Linked machine-generated topics (e.g., "2020_Topic8") to human labels (e.g., "Superhero")*
- Dimensionality Reduction:**
 - Merged topics with the same human label and calculated their combined probabilities*

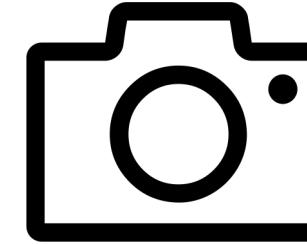
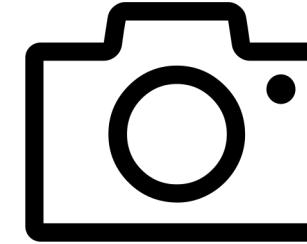
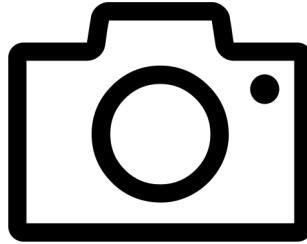
comment_id	2014_Topic0	2014_Topic1	2014_Topic2	2014_Topic3	2014_Topic4	2014_Topic5	2014_Topic6	2014_Topic7	2014_Topic8	...	2023_Topic8
5	0.015880	0.023429	0.381898	0.271167	0.166802	0.029295	0.030989	0.039260	0.022075	...	0.0
24	0.000000	0.000000	0.570520	0.365656	0.000000	0.000000	0.000000	0.011750	0.000000	...	0.0
58	0.012435	0.018347	0.507401	0.217727	0.024250	0.022941	0.024267	0.030980	0.126612	...	0.0
60	0.015849	0.023383	0.381149	0.133607	0.168137	0.029238	0.168173	0.039183	0.022032	...	0.0
101	0.000000	0.013864	0.700770	0.174051	0.018325	0.017595	0.018338	0.023232	0.013063	...	0.0
...
222477	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0
222483	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0
222503	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0
222511	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0
222519	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0

document-term frequency matrix (LDA)												Script Development	Superhero
comment_id	Marketing	Social Diversity	3D_IMAX	Mixed	IP Management	Family Themes	Political Themes	Positive	Neutral	...	Fantasy	Script Development	Superhero
5	0.030989	0.0	0.015880	0.0	0.0	0.0	0.0	0.381898	0.0	...	0.019205	0.0	0.0
24	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.570520	0.0	...	0.000000	0.0	0.0
58	0.024267	0.0	0.012435	0.0	0.0	0.0	0.0	0.507401	0.0	...	0.015040	0.0	0.0
60	0.168173	0.0	0.015849	0.0	0.0	0.0	0.0	0.381149	0.0	...	0.019250	0.0	0.0
101	0.018338	0.0	0.000000	0.0	0.0	0.0	0.0	0.700770	0.0	...	0.011365	0.0	0.0
...
222477	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.293647	0.0	...	0.000000	0.0	0.0
222483	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.241635	0.0	...	0.000000	0.0	0.0
222503	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.272317	0.0	...	0.000000	0.0	0.0
222511	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.310642	0.0	...	0.000000	0.0	0.0
222519	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.209369	0.0	...	0.000000	0.0	0.0

document-term frequency matrix (Merged data)

Data Analysis

Audience Preference Evolution Through Reddit Discussions



Data

222,520 Reddit movie
comments

Technique

Latent Dirichlet
Allocation (LDA) topic
modeling

Timeframe

2014-2024

Movie Topic Categories Shifts

2014-2017→

Genre debates [30%] & Industry [25%] → Traditional cinema culture

2017-2019 →

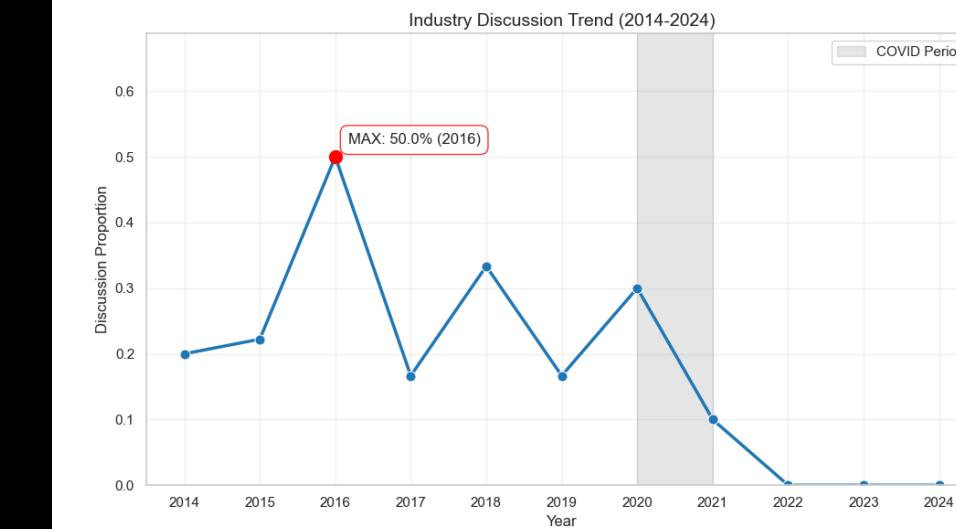
Franchise hype [50%] → Marvel/Star Wars dominance

2020-2021 →

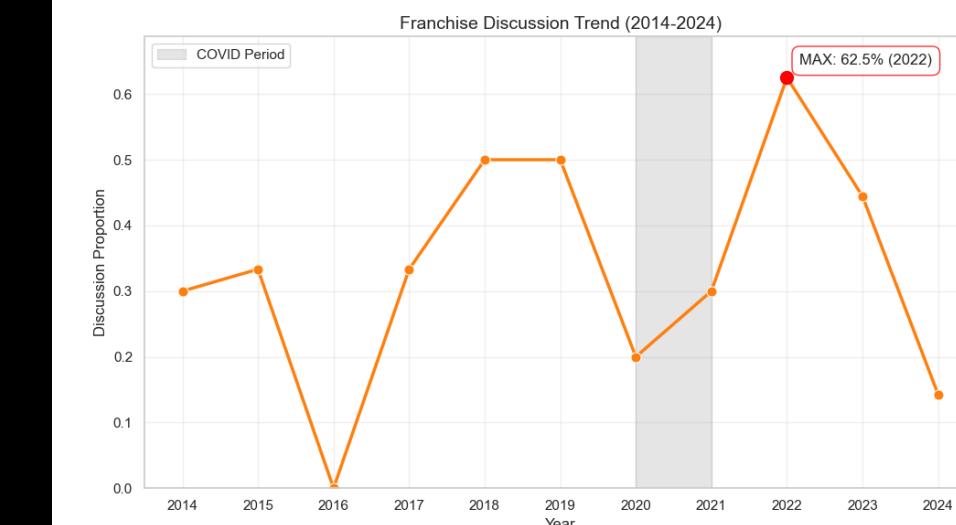
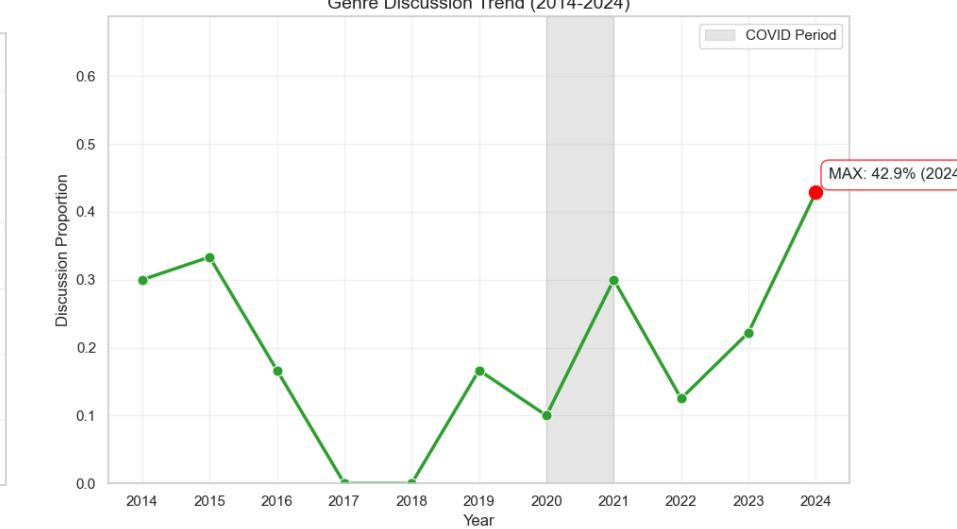
Pandemic → Social context and family viewing

2022-2024 →

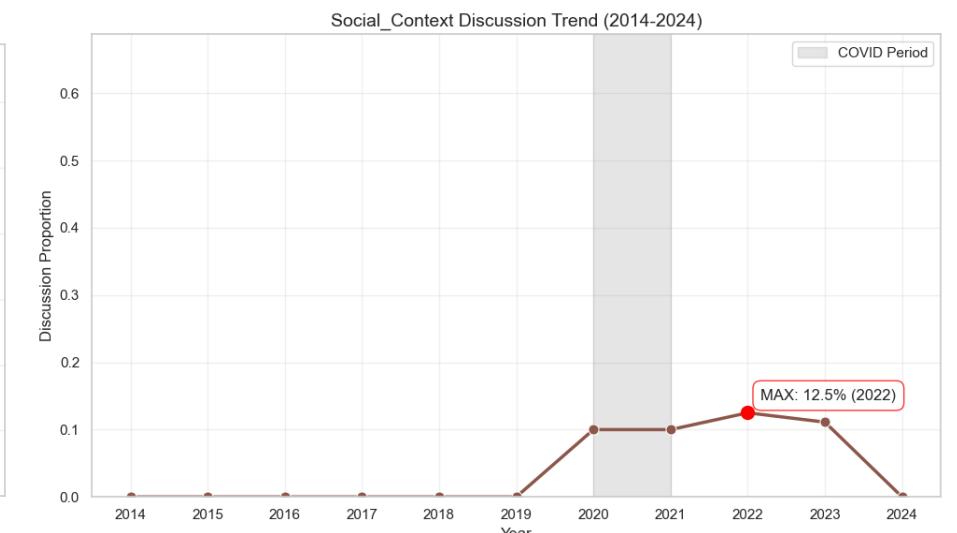
Avatar 2 [2022], Barbie[2023] → AI tech + Social issues



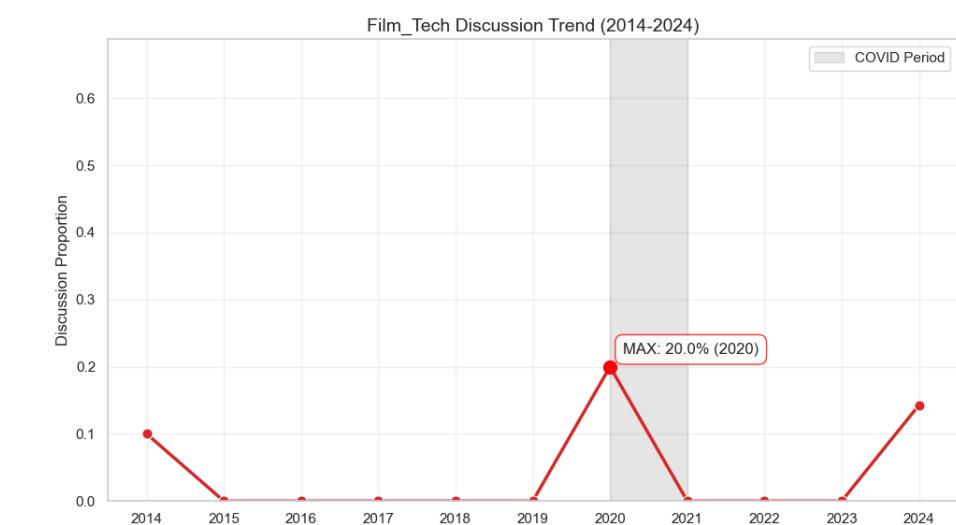
2014-2017_Dominance of Genre and Industry



2017-2019_Dominance of Franchise



2020-2021_New theme



2022-2024_Social issues + AI tech

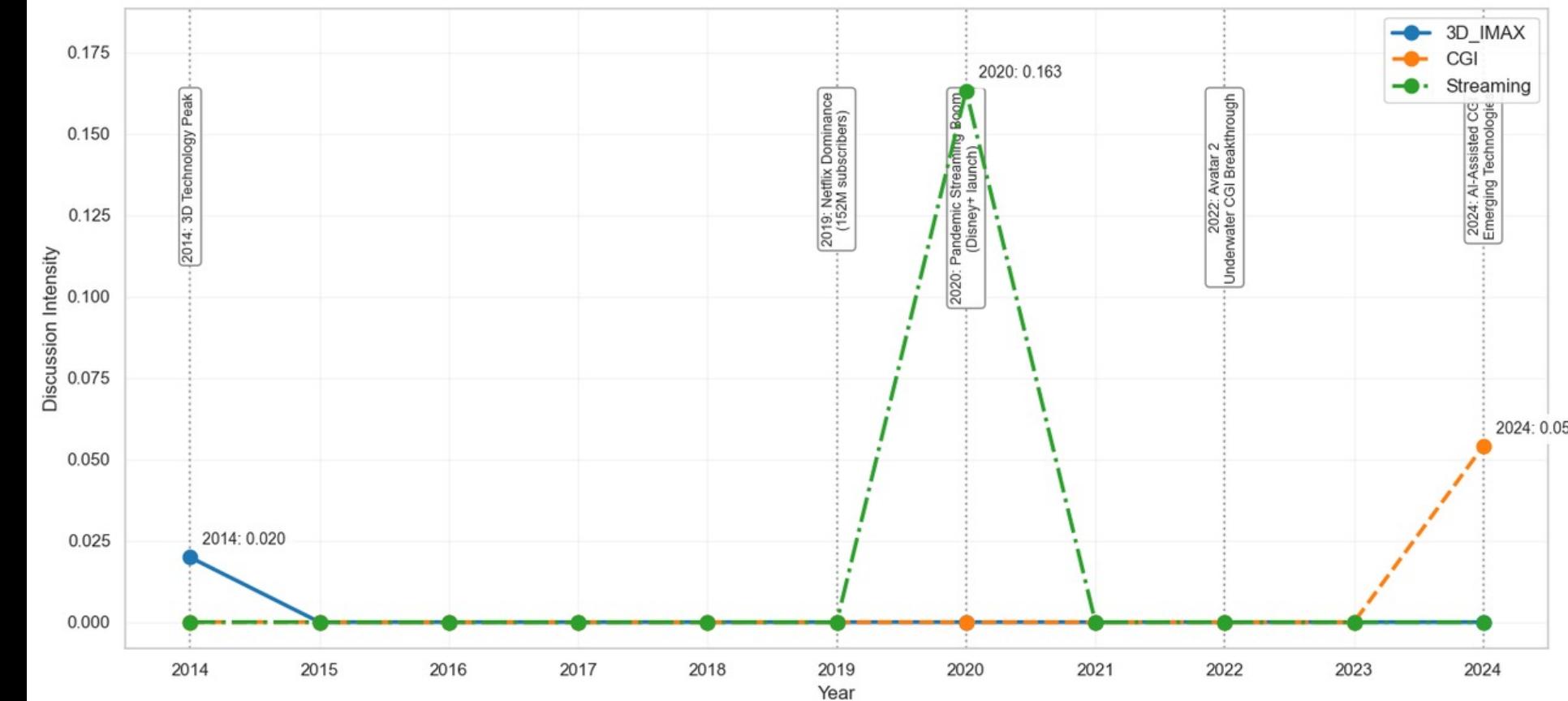
Technology Discussion Evolution

2014: 3D/IMAX [0.0199 intensity]

2020: Streaming [0.1633 intensity]

2024: AI/CGI [Avatar 2 driven]

Film Technology Discussion Trends with Key Milestones (2014-2024)



Technology Discussion Evolution

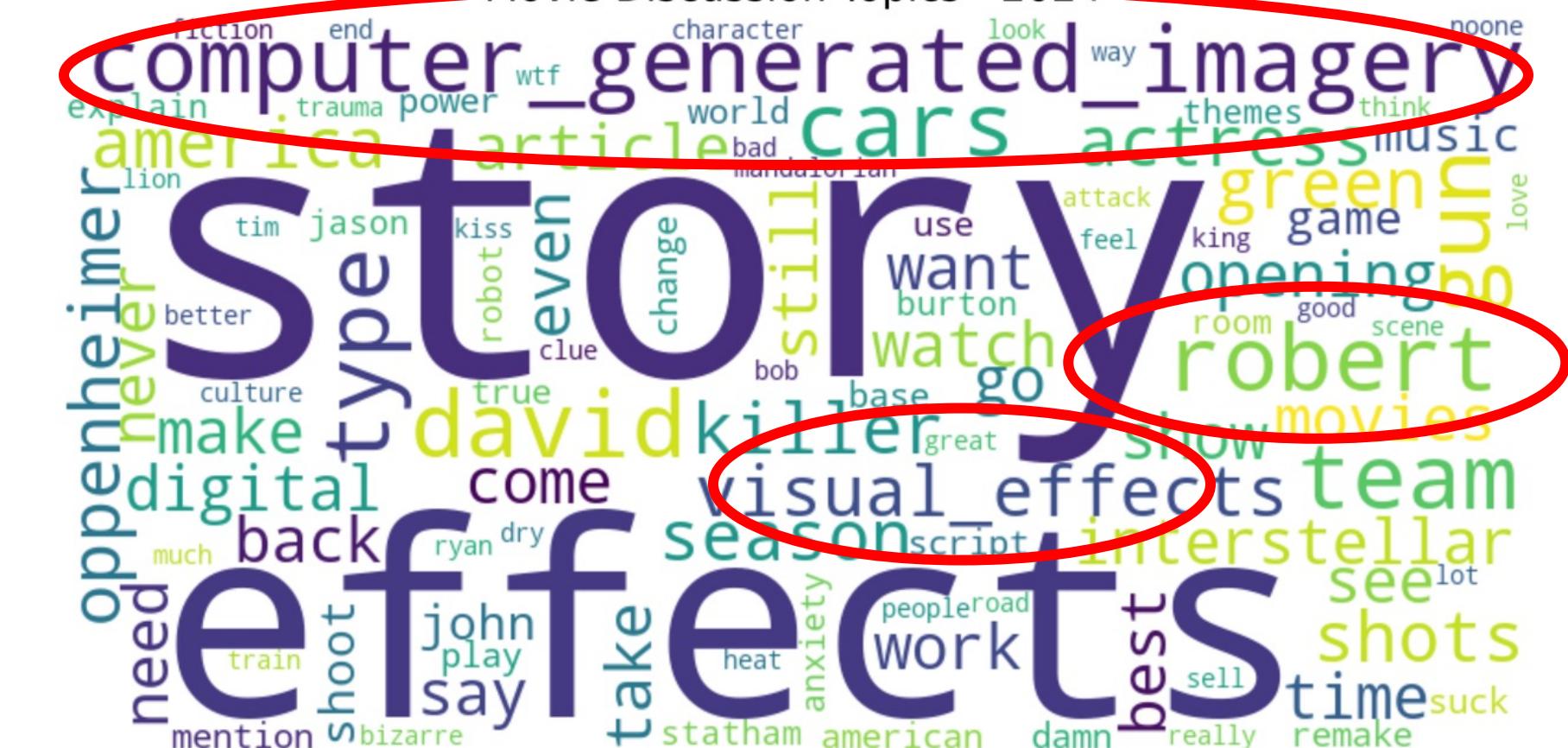
2014: *3d_technology, imax,*

2024: CGI, robert

Movie Discussion Topics - 2014



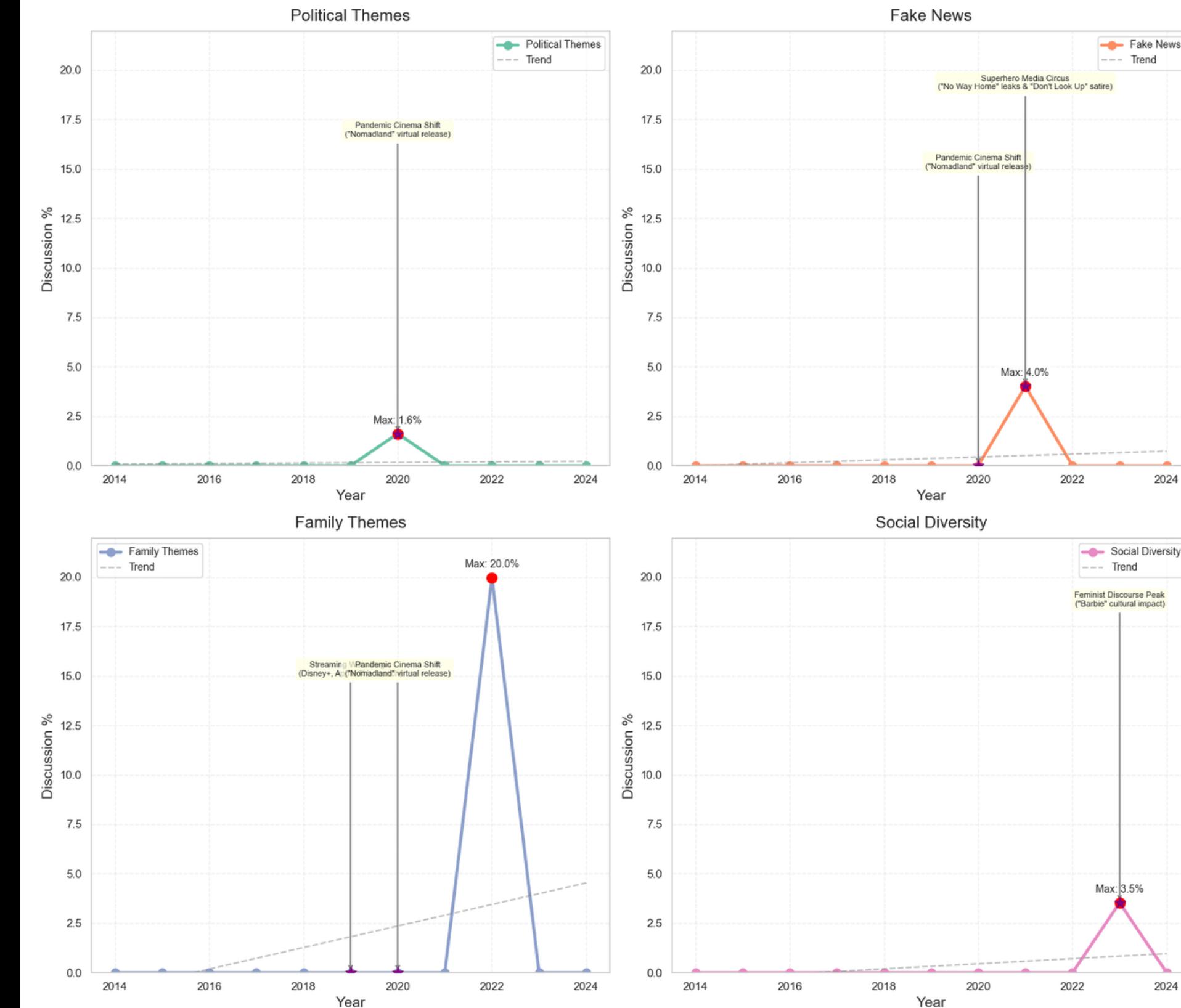
Movie Discussion Topics 2024



Social Issue Discussions Evolution

Event-Driven Spikes:

- 2020: *Nomadland* (political themes - 1.6%)
 - 2021: *Don't Look Up* (fake news - 4%)
 - 2022: *Post-Pandemic & Lockdowns* (Family Themes- 4%)
 - 2023: *Barbie* (diversity - 3.5%)



Social Issue Discussions Evolution

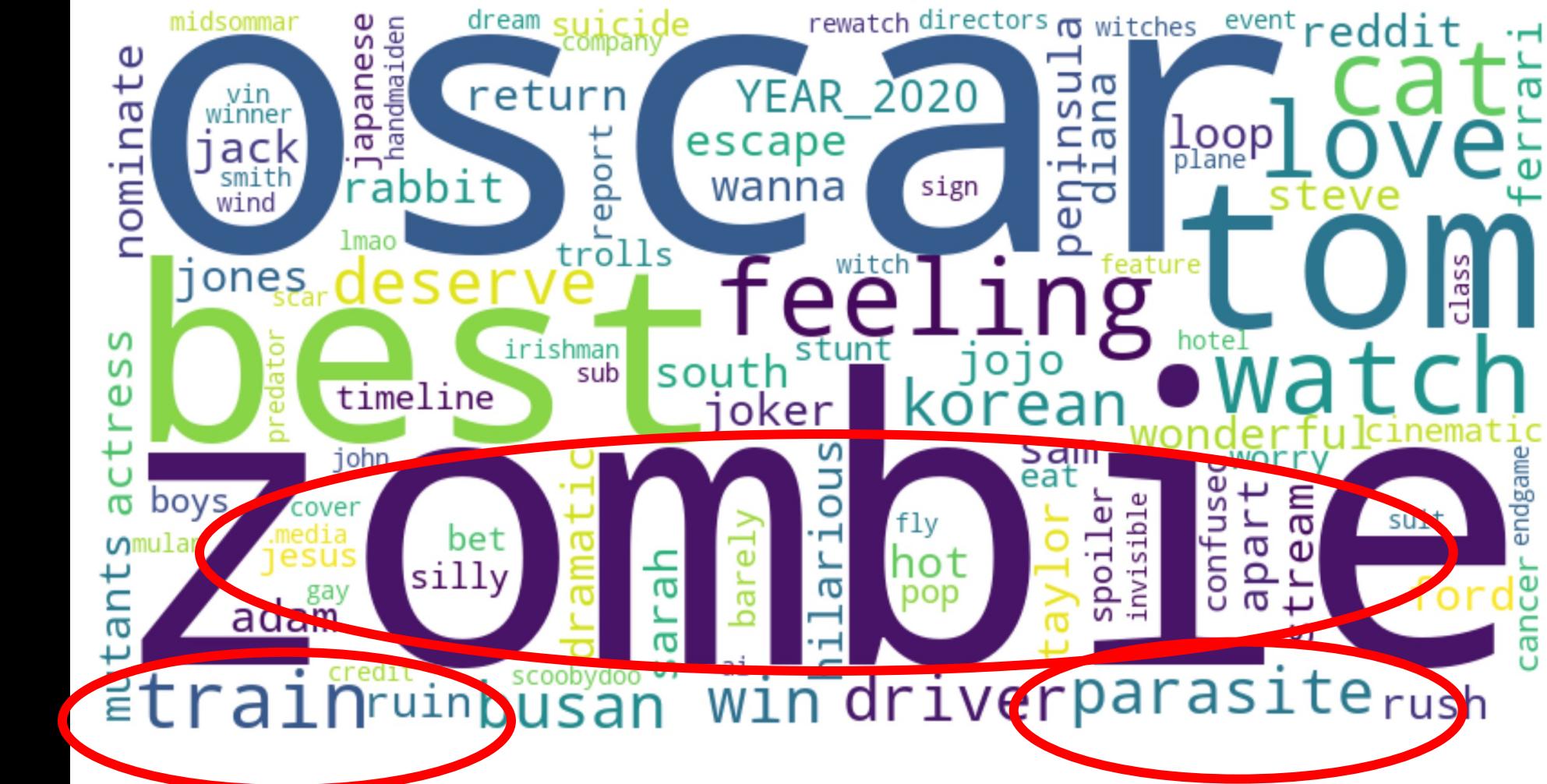
2020 Audience psychological projections related to real-world crises

Zombie

Busan Train (2016)

Parasite (2019)

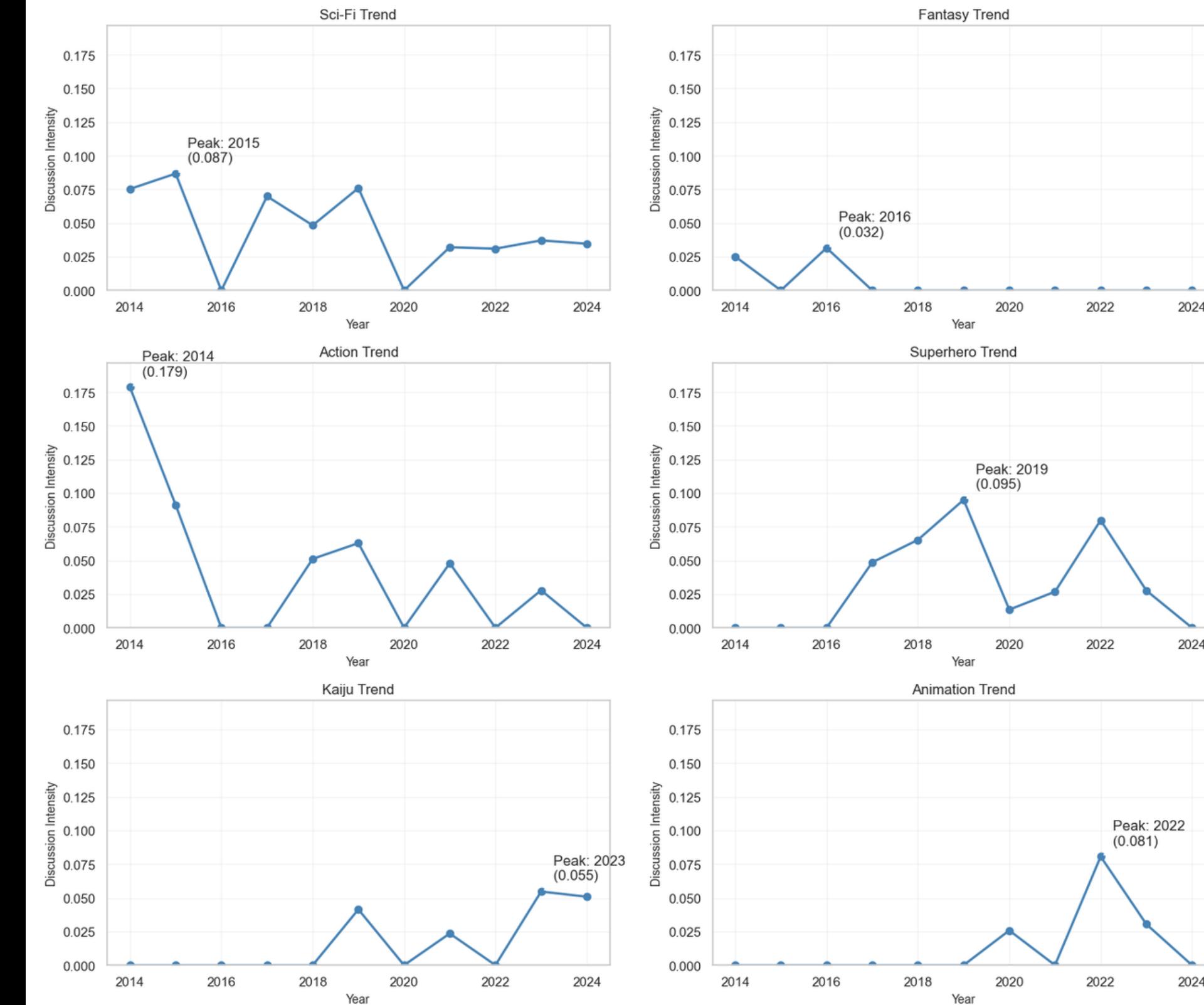
Movie Discussion Topics - 2020



Franchise Movie Discussions Evolution

“Genre Dominance Over Time”

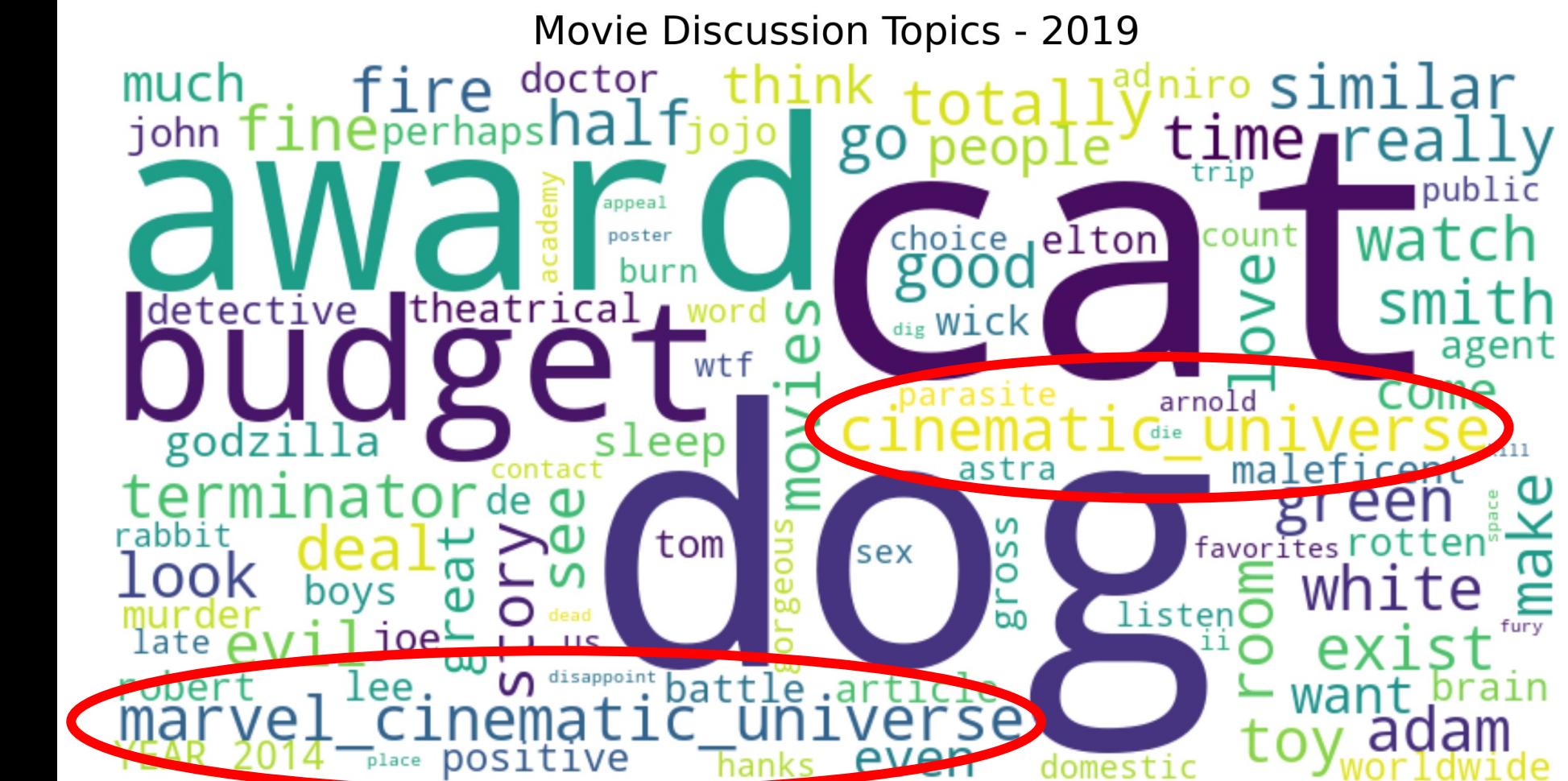
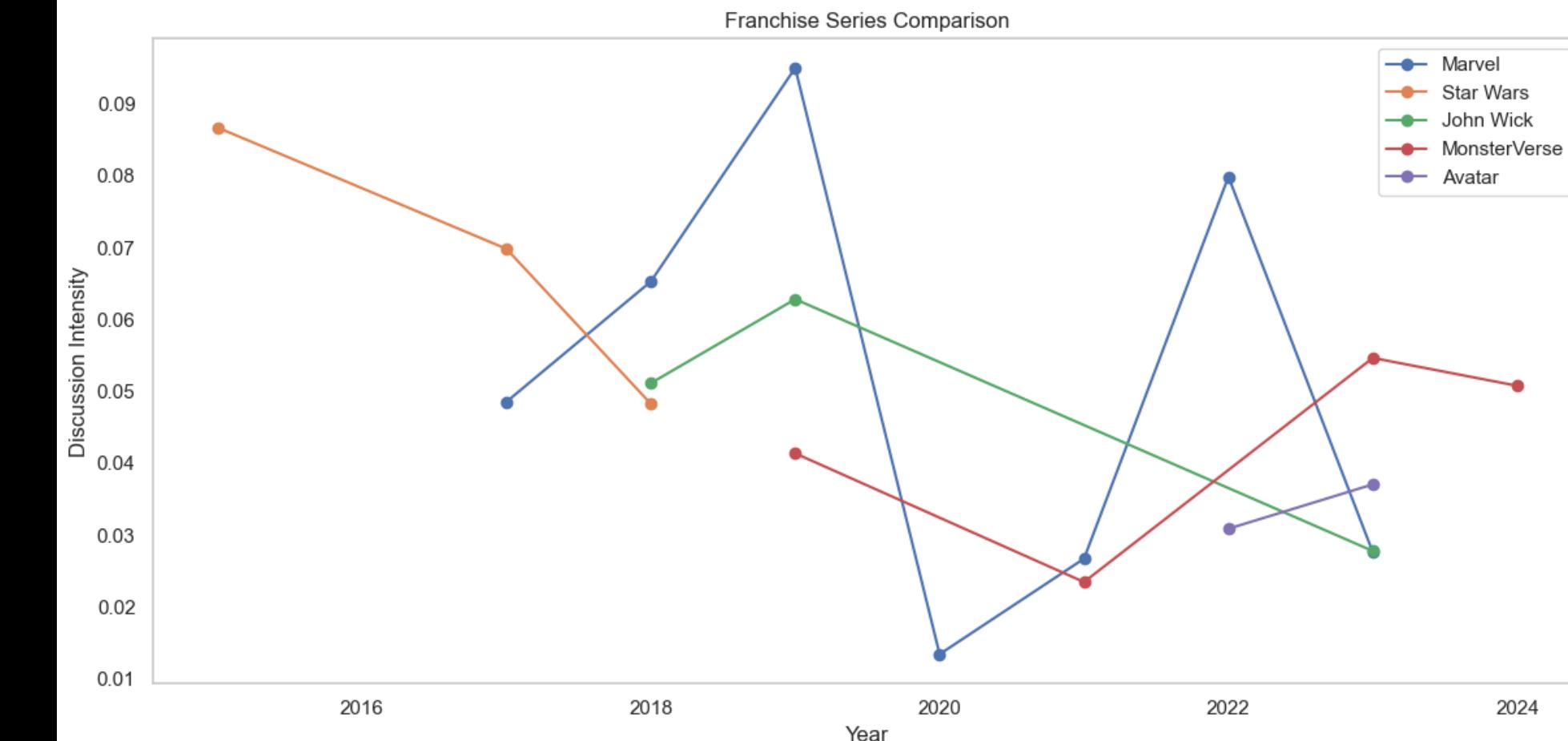
- 2014-2016: Action films peak (17.85%)
 - Key films: *Transformers 4, Mad Max: Fury Road*
- 2017-2019: Superhero era ($\Delta 4.86\% \rightarrow 9.50\%$)
 - Peak: *Avengers: Endgame* (2019)
- 2020-2022: Animation surges (8.07%)
 - Pandemic effect: *Minions: The Rise of Gru*
- 2023-2024: No clear leader → Market diversification



Franchise Movie Discussions Evolution

“Franchise Case Studies”

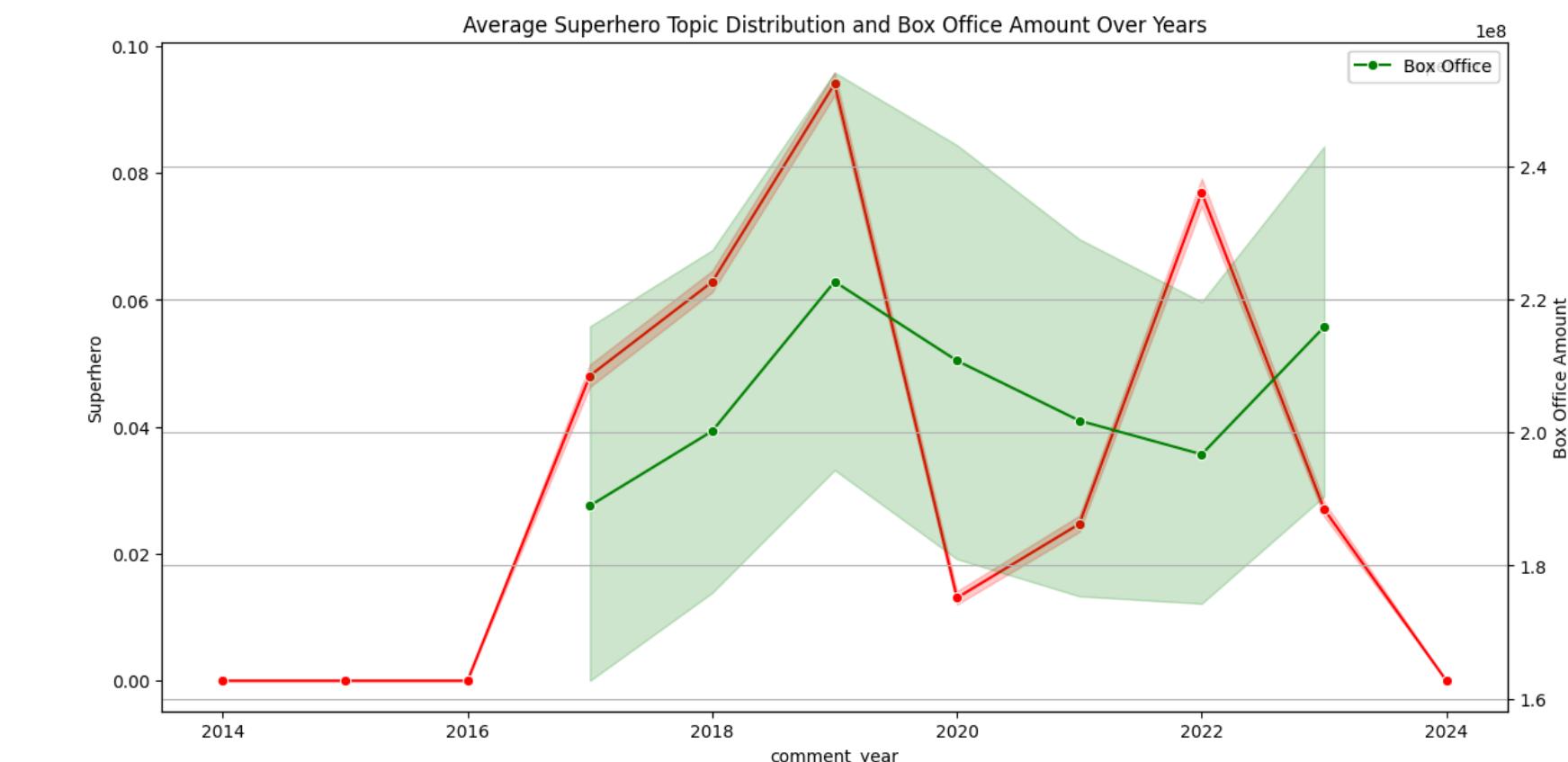
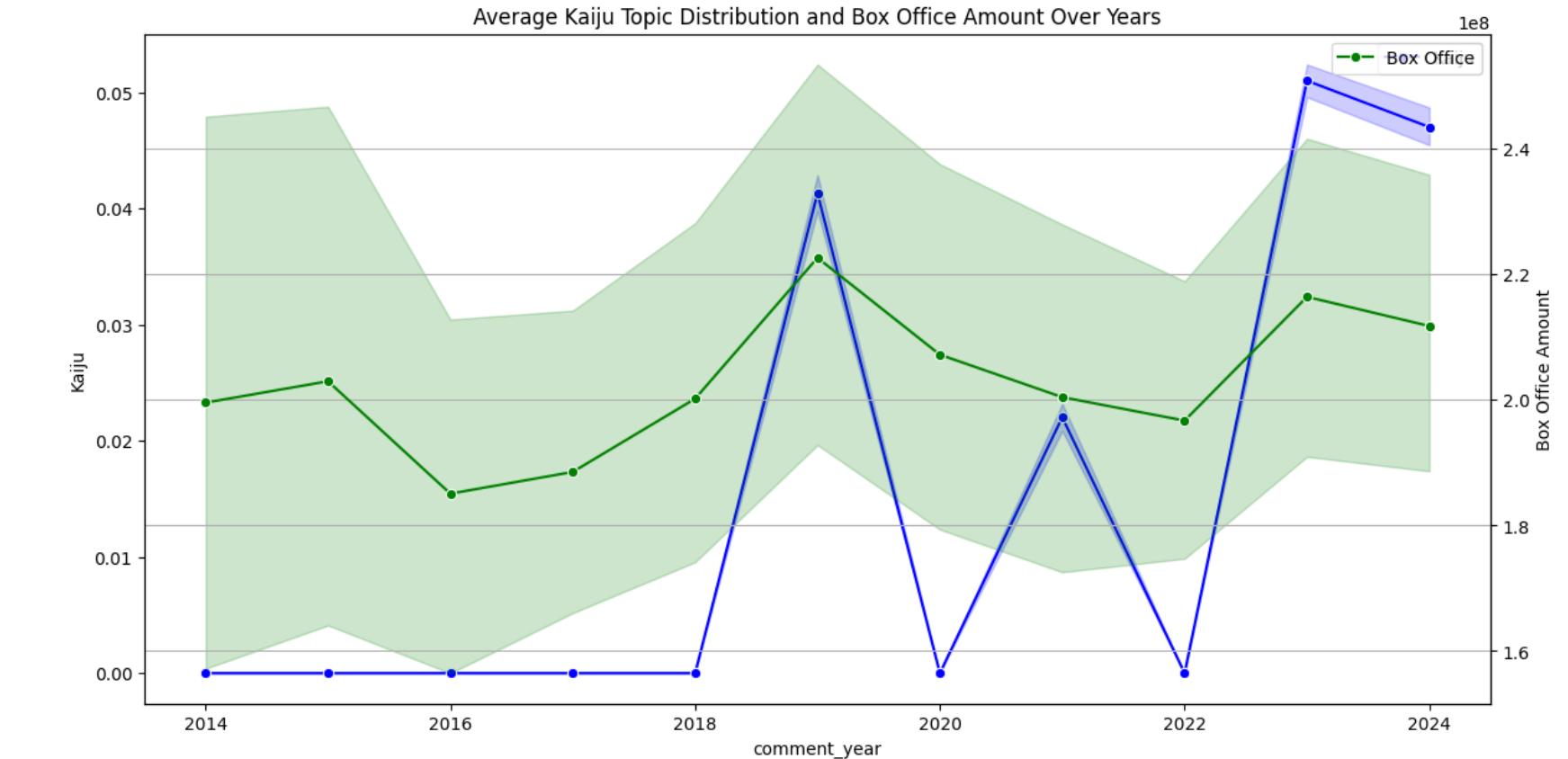
- *Marvel*:
 - 2019 peak → 2020 drop → 2022 rebound → 2023 struggles
 - *Marvel_cinematic_universe*
- *MonsterVerse*:
 - Long lifecycle with accelerating trend
- *Avatar*:
 - 2022 resurgence (*Avatar 2* tech innovation)



Box Office & Audience Discussion

“Emotion drives tickets. Analysis drives clicks”

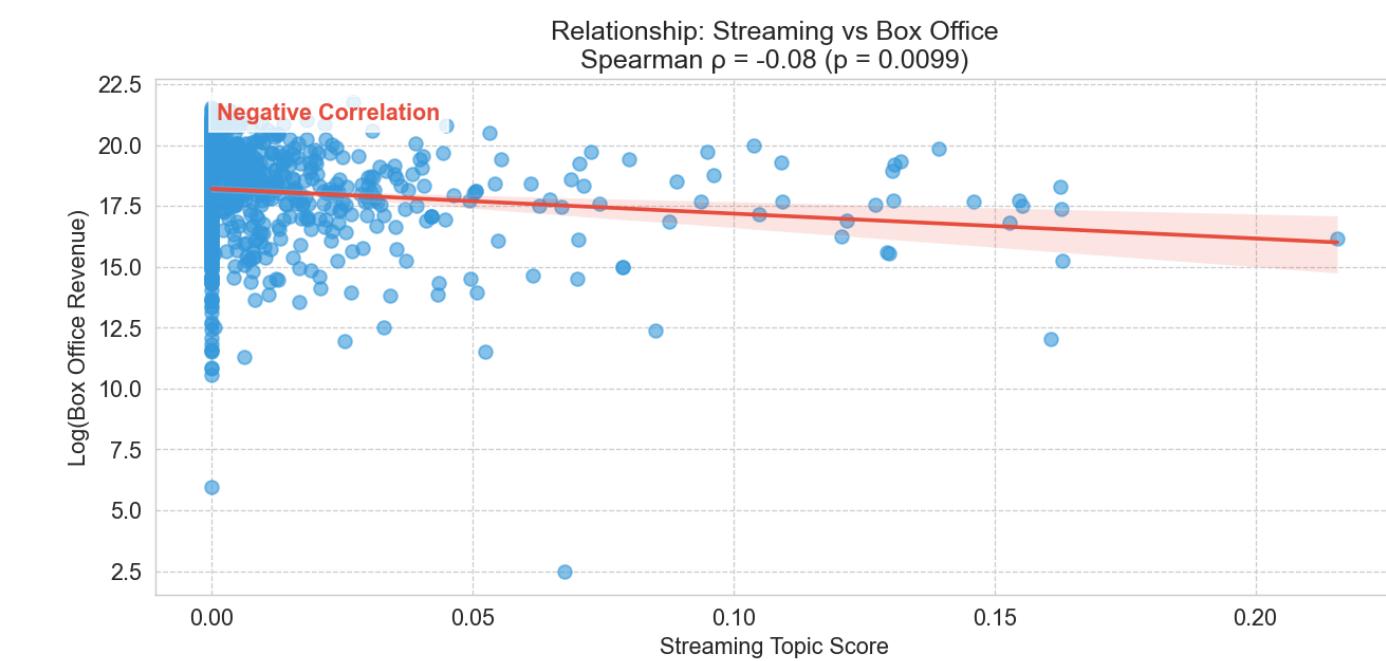
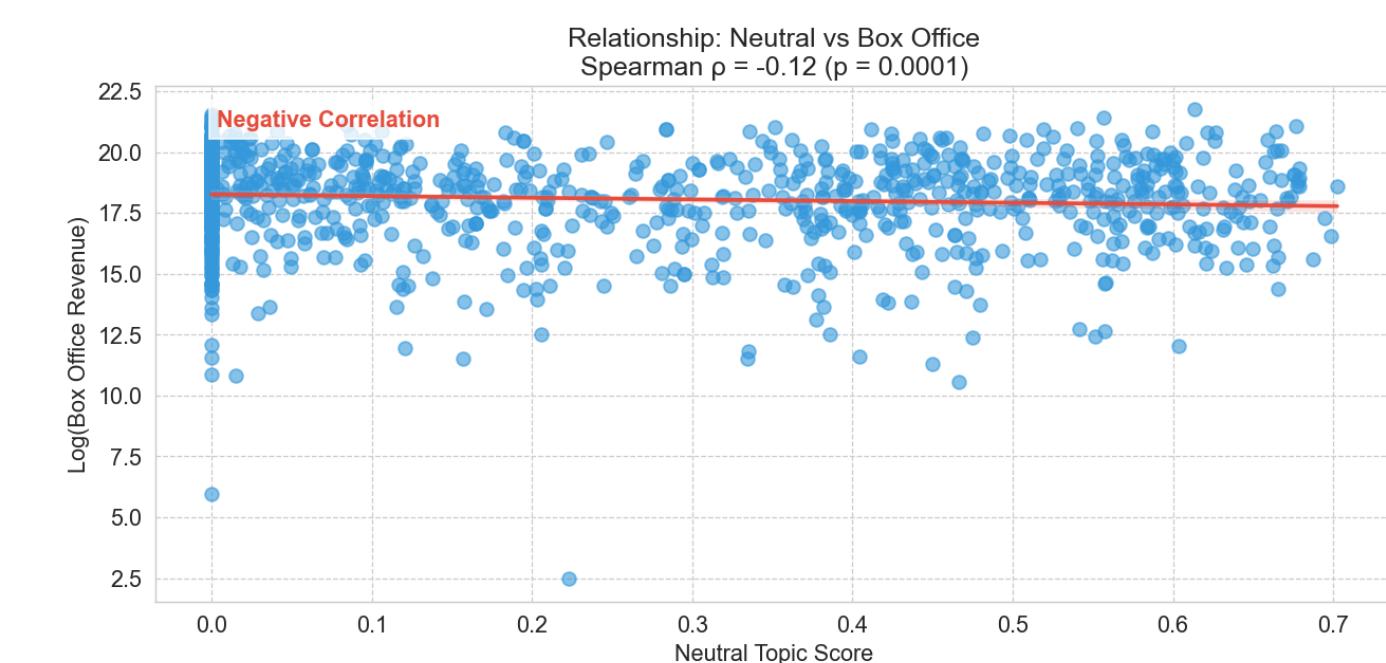
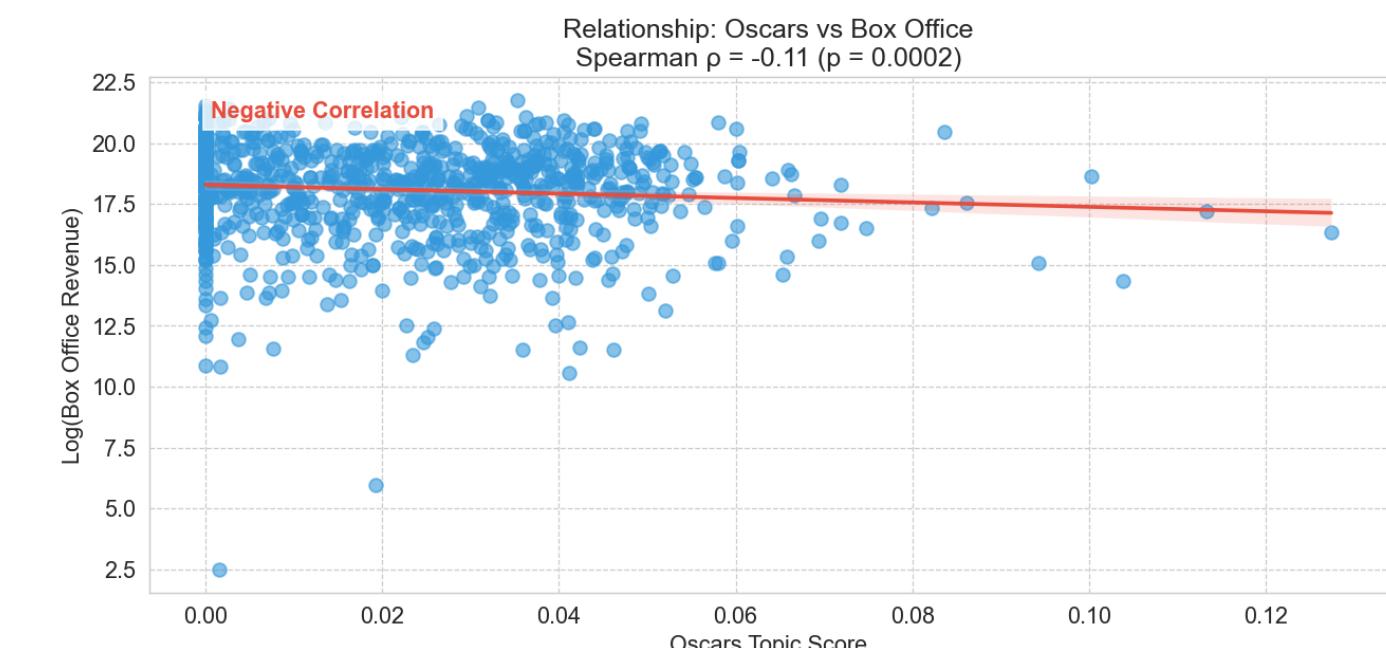
- *Do's & Don'ts:*
 - *Boosters: Kaiju films (Godzilla), practical effects hype*
 - *Saturated: Superhero chatter ≠ revenue*
 - *Killers: Oscar talk, neutral reviews, streaming mentions*



Box Office & Audience Discussion

“Emotion drives tickets. Analysis drives clicks”

- *Do's & Don'ts:*
 - *Boosters: Kaiju films (Godzilla), practical effects hype*
 - *Saturated: Superhero chatter ≠ revenue*
 - *Killers: Oscar talk, neutral reviews, streaming mentions*



Summary

1. Franchise Fatigue:

Audiences are craving fresh stories beyond recycled universes

2. Tech Shift:

IMAX→Streaming→AI production

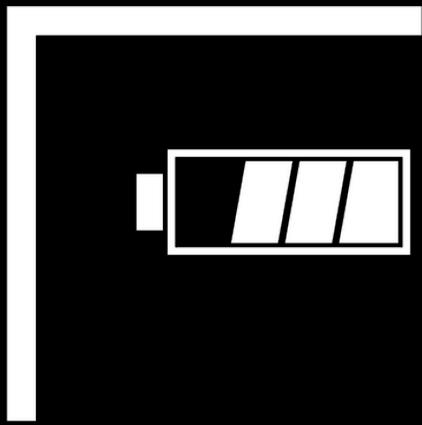
3. Social Lens:

A battleground for societal debates (2020's political themes to 2023's social diversity)

4. Box Office Truths:

Emotional > rational discussions

innovative tech + social relevance + raw emotional power + no franchise burnout



THANK YOU!

