

Presented by
Yim Sie Lam, Sheron
3036389359

Are Film Awards Losing Their Authorities?

Project Presentation
COMP2501 Introduction to Data Science and Engineering



I am going to cover the content below...

1

Problems to be solved

Identifying research questions and the respective importance



2

Data Collection

Finding and processing data from Oscars and HKFA



3

Data Analysis

Implications from the processed data



4

Conclusion

Evaluate the problems and limitations



Problem

Data Collection



Analysis

Conclusion



PROBLEMS

Problem



Data Collection



Analysis

Conclusion

Recent Trends and Controversies on Filming Awards



Decline in interest in annual film awards

Globally...

...

- Oscars **TV ratings** have dropped sharply
 - From **43M** in 2014 to **10.4M** viewers in 2021

Locally.....

Official HKFA clips received low **social media engagement**

- Got **50–200K** views on YouTube in 2020s
- Peak = **1M** from 2000 to early 2010

76%

Decline in Oscars
TV Viewers

95%

Decline in HKFA
TV Viewers



Recent Trends and Controversies on Filming Awards

'Anora' wins best picture at the 2025 Oscars

MARCH 2, 2025 · 11:07 PM ET

Film Books Music Art & design TV & radio Stage Classical Games

Oscars 2017

This article is more than 8 years old

Wrong Best Picture
Winner announced?

Anatomy of an Oscars fiasco: how La La Land was mistakenly announced as best picture

How could the wrong film have been named as the winner of the Oscars' most important award? Here are the pictures that tell the story

Oscars 2017: full list of winners



Problem to be solved and related research questions

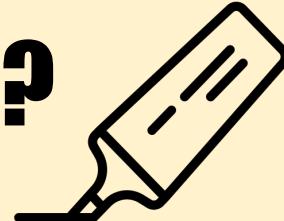


Research Questions

1. Are award-winning films always blockbusters?



2. How has the judging criteria of the judges of the awards changed over time?



Are film awards losing their authorities?



DATA COLLECTION

Problem

Data Collection



Analysis

Conclusion

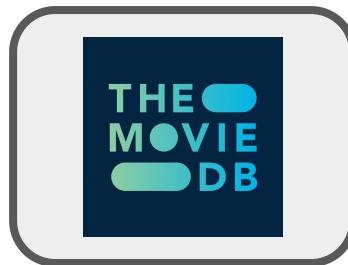
Finding and processing suitable data

Oscars-related data

Available Datasets



Useful Website



Filter targeted Data



the_oscar_award.csv

- Nominees and awardees of oscars from 1927–2025

The Movie Database (TMDB)

- Box offices of films
- Box office reflects public interest and film preference.



Analysis

Problem

Data Collection

Conclusion

Data Collection

Data manipulation on dataset from kaggle

Raw data in .csv:

year_film	year_ceremony	ceremony	category	canon_category	name	film	winner
1927	1928		1 ACTOR	ACTOR IN A LEADING	Richard Barthelmess	The Noose	FALSE
1927	1928		1 ACTOR	ACTOR IN A LEADING	Richard Barthelmess	The Patent Leather Kid	FALSE
1927	1928		1 ACTOR	ACTOR IN A LEADING	Emil Jannings	The Last Command	TRUE
1927	1928		1 ACTOR	ACTOR IN A LEADING	Emil Jannings	The Way of All Flesh	TRUE
1927	1928		1 ACTRESS	ACTRESS IN A LEADIN	Louise Dresser	A Ship Comes In	FALSE
1927	1928		1 ACTRESS	ACTRESS IN A LEADIN	Janet Gaynor	7th Heaven	TRUE
1927	1928		1 ACTRESS	ACTRESS IN A LEADIN	Janet Gaynor	Street Angel	TRUE
1927	1928		1 ACTRESS	ACTRESS IN A LEADIN	Janet Gaynor	Sunrise	TRUE
1927	1928		1 ACTRESS	ACTRESS IN A LEADIN	Gloria Swanson	Sadie Thompson	FALSE
1927	1928		1 ART DIRECTION	ART DIRECTION	Rochus Gliese	Sunrise	FALSE
1927	1928		1 ART DIRECTION	ART DIRECTION	William Cameron Me	The Dove	TRUE
1927	1928		1 ART DIRECTION	ART DIRECTION	William Cameron Me	Tempest	TRUE
1927	1928		1 ART DIRECTION	ART DIRECTION	Harry Oliver	7th Heaven	FALSE
1927	1928		1 CINEMATOGRAPHY	CINEMATOGRAPHY	George Barnes	The Devil Dancer	FALSE
1927	1928		1 CINEMATOGRAPHY	CINEMATOGRAPHY	George Barnes	The Magic Flame	FALSE
1927	1928		1 CINEMATOGRAPHY	CINEMATOGRAPHY	George Barnes	Sadie Thompson	FALSE
1927	1928		1 CINEMATOGRAPHY	CINEMATOGRAPHY	Charles Rosher/Karl	Sunrise	TRUE
1927	1928		1 DIRECTING (Comedy	DIRECTING (Comedy	Lewis Milestone	Two Arabian Knights	TRUE
1927	1928		1 DIRECTING (Comedy	DIRECTING (Comedy	Ted Wilde	Speedy	FALSE
1927	1928		1 DIRECTING (Dramati	DIRECTING (Dramati	Frank Borzage	7th Heaven	TRUE
1927	1928		1 DIRECTING (Dramati	DIRECTING (Dramati	Herbert Brenon	Sorrell and Son	FALSE
1927	1928		1 DIRECTING (Dramati	DIRECTING (Dramati	King Vidor	The Crowd	FALSE
1927	1928		1 ENGINEERING EFFEC	VISUAL EFFECTS	Ralph Hammeras		FALSE
1927	1928		1 ENGINEERING EFFEC	VISUAL EFFECTS	Roy Pomeroy	Wings	TRUE



Data manipulation on dataset from kaggle

```
library(tidyverse)
oscar_data <- read_csv("the_oscar_award.csv")

# Filter for Best Picture winners only
best_picture_winners <- oscar_data |>
  filter(category == "BEST PICTURE" & winner == TRUE) |>
  distinct(year_ceremony, film) |>
  arrange(year_ceremony)
```

- 1. Read the dataset by `read_csv()` [by “tidyverse”]**
- 2. Filtered all the Best Picture winners**



	year_ceremony	film
1	1963	Lawrence of Arabia
2	1964	Tom Jones
3	1965	My Fair Lady
4	1966	The Sound of Music
5	1967	A Man for All Seasons
6	1968	In the Heat of the Night
7	1969	Oliver!
8	1970	Midnight Cowboy
9	1971	Patton
10	1972	The French Connection
11	1973	The Godfather
12	1974	The Sting
13	1975	The Godfather Part II
14	1976	One Flew over the Cuckoo's Nest
15	1977	Rocky
16	1978	Annie Hall

Data extraction from TMDB



Original page:

Anora (2024)
10/18/2024 (US) • Drama, Comedy, Romance • 2h 19m

User Score 70% What's your Vibe? 0

Love is a hustle.

Overview

A young sex worker from Brooklyn gets her chance at a Cinderella story when she meets and impulsively marries the son of an oligarch. Once the news reaches Russia, her fairytale is threatened as his parents set out to get the marriage annulled.

Sean Baker
Director, Writer

AWARDS 3 Wins | 4 Nominations →

Top Billed Cast

Mikey Madison Ani	Mark Rydelshteyn Ivan	Yura Borisov Igor	Karren Karagullian Toros	Vache Tovmasyan Gernik	Luna Sofia Miranda Lulu	Lindsey Normington Diamond
----------------------	--------------------------	----------------------	-----------------------------	---------------------------	----------------------------	-------------------------------

Status Released
Original Language English
Budget \$6,000,000.00
Revenue \$56,286,295.00

Target

- Extracting Box office from TMDB
- Making a table include
 - I. Year of ceremony
 - II. Film name
 - III. Box office



Data extraction from TMDB by RStudio



Rationale

1. Get the box office by searching for the closest film name
2. Use free API key from TMDB to access box office
3. Extract the movie detail and box office as the table “best_picture_box_office”

```

library(httr)
library(jsonlite)

get_box_office <- function(film_title, year, api_key) {
  search_response <- GET(
    "https://api.themoviedb.org/3/search/movie",
    query = list(
      api_key = api_key,
      query = film_title,
      year = year,
      include_adult = FALSE
    )
  )

  if (http_error(search_response)) {
    message(paste("Search failed for:", film_title, year))
    return(tibble(Film = film_title, Year = year, BoxOffice = NA, Status = "Search Failed"))
  }

  search_content <- fromJSON(rawToChar(search_response$content))

  if (length(search_content$results) == 0) {
    return(tibble(Film = film_title, Year = year, BoxOffice = NA, Status = "No Match Found"))
  }

  # Get the most likely match (highest popularity)
  movie_id <- search_content$results %>%
    arrange(desc(popularity)) %>%
    pull(id) %>%
    first()

  details_response <- GET(
    paste0("https://api.themoviedb.org/3/movie/", movie_id),
    query = list(api_key = api_key)
  )

  details_content <- fromJSON(rawToChar(details_response$content))

  return(tibble(
    Film = film_title,
    Year = year,
    BoxOffice = ifelse(!is.null(details_content$revenue), details_content$revenue, NA),
    Status = "Success"
  ))
}

tmdb_key <- "ea0d7ba2a2f3f20d09e7196f104c4a2e"

best_picture_box_office <- best_picture_winners %>%
  mutate(map2_df(film, year_ceremony, ~get_box_office(.x, .y, tmdb_key), .progress = TRUE))

```



Data extraction from TMDB by RStudio



!! INFLATION !!

Note: 1 Billion in 1990s >>> 1 Billion Now



Underestimation of past winners



Data Collection

Data extraction from TMDB by RStudio

!! INFLATION !!

Note: 1 Billion in 1990s >>> 1 Billion Now



Underestimation of past winners



Incorporate Box Office Mojo's Inflation
Calculator [Movie-specific!!]



```
mojo_adjustments <- tribble(  
  ~Year, ~adjustment_factor,  
  2025, 1.00,  
  2024, 1.00,  
  2023, 1.00,  
  2022, 1.02,  
  2021, 1.04,  
  2020, 1.06,  
  2019, 1.08,  
  2018, 1.10,  
  2017, 1.12,  
  2016, 1.15,  
  
  # Apply adjustments  
  adjusted_data <- best_picture_box_office %>%  
    left_join(mojo_adjustments, by = "Year") %>%  
    mutate(  
      adjusted_gross = BoxOffice * adjustment_factor  
    ) %>%  
    select(Film, year_ceremony, Nominal_BoxOffice = BoxOffice,  
          Adjusted_BoxOffice = adjusted_gross, adjustment_factor, status)
```

1995, 2.10,
1994, 2.17,
1993, 2.24,
1992, 2.32,
1991, 2.40,
1990, 2.48,
1989, 2.57,
1988, 2.66,
1987, 2.76,
1986, 2.86,
1985, 2.97,
1984, 3.08,
1983, 3.20,
1982, 3.33,
1981, 3.46,
1980, 3.60,
1979, 3.75,
1978, 3.91,
1977, 4.08,
1976, 4.26,
1975, 4.45,
1974, 4.65,
1973, 4.87,
1972, 5.10,
1971, 5.35,
1970, 5.61,
1969, 5.89,
1968, 6.19,
1967, 6.51,
1966, 6.85,
1965, 7.21,
1964, 7.60,
1963, 8.01

*Try to simulate the impact of inflation on box office



Finding and processing suitable data

HKFA-related data

List Available



Manual Search



Web scraping



Wikipedia

- Best Film awardees list

HK Box office of the winners

- Collect together by excel

RStudio

- Scrape data from Wikipedia and excel



Data Collection

Web scraping by RStudio

Winners of Best Film Over the Years [\[edit\]](#)

Year	Best Film	Original title	Director	Country
1982	Father and Son	父子情	Allen Fong	Hong Kong
1983	Boat People	投奔怒海	Ann Hui	Hong Kong
1984	Ah Ying	半邊人	Allen Fong	Hong Kong
1985	Homecoming	似水流年	Ho Yim	Hong Kong / China
1986	Police Story	警察故事	Jackie Chan	Hong Kong
1987	A Better Tomorrow	英雄本色	John Woo	Hong Kong
1988	An Autumn's Tale	秋天的童話	Mabel Cheung	Hong Kong
1989	Rouge	胭脂扣	Stanley Kwan	Hong Kong
1990	Beyond the Sunset	飛越黃昏	Jacob Cheung	Hong Kong
1991	Days of Being Wild	阿飛正傳	Wong Kar-wai	Hong Kong
1992	To Be Number One	跛豪	Poon Man Kit	Hong Kong
1993	Cageman	籠民	Jacob Cheung	Hong Kong
1994	C'est la vie, mon cheri	新不了情	Tung-Shing Yee	Hong Kong
1995	Chungking Express	重慶森林	Wong Kar-wai	Hong Kong
1996	Summer Snow	女人四十	Ann Hui	Hong Kong
1997	Comrades: Almost a Love Story	甜蜜蜜	Peter Chan	Hong Kong

```
library(rvest)
library(tidyverse)

# Scrape Best Film winners from wikipedia
url <- "https://en.wikipedia.org/wiki/Hong_Kong_Film_Awards"
hkfa_winners <- read_html(url) %>%
  html_nodes("table.wikitable") %>%
  html_table() %>%
  .[[1]] %>%
  # First table contains Best Film winners
  select(Year = 1, Film = 2, ChineseName = 3) %>%
  mutate(Year = as.numeric(str_extract(Year, "\\\d{4}")))
```



Perform web scraping [by “rvest”]

List in Wikipedia W



Problem

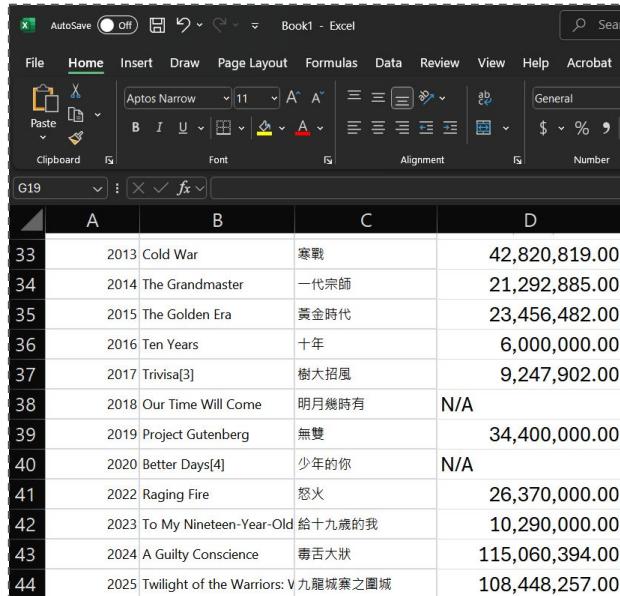
Data Collection

Analysis

Conclusion

Data Collection

Searching HK box office manually online



	A	B	C	D
33	2013	Cold War	寒戰	42,820,819.00
34	2014	The Grandmaster	一代宗師	21,292,885.00
35	2015	The Golden Era	黃金時代	23,456,482.00
36	2016	Ten Years	十年	6,000,000.00
37	2017	Trivisa[3]	樹大招風	9,247,902.00
38	2018	Our Time Will Come	明月幾時有	N/A
39	2019	Project Gutenberg	無雙	34,400,000.00
40	2020	Better Days[4]	少年的你	N/A
41	2022	Raging Fire	怒火	26,370,000.00
42	2023	To My Nineteen-Year-Old	給十九歲的我	10,290,000.00
43	2024	A Guilty Conscience	毒舌大狀	115,060,394.00
44	2025	Twilight of the Warriors: V 九龍城寨之圍城		108,448,257.00



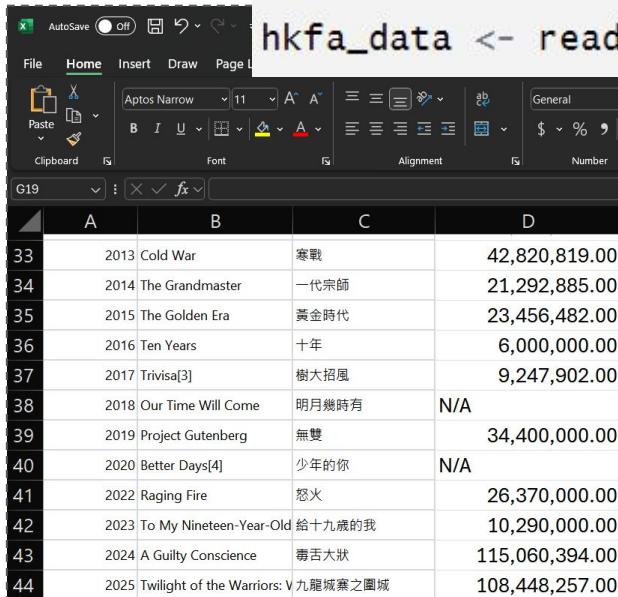
Data collected in Excel

*N/A: Box office data is combined with Mainland → Cannot estimate HK box office separately



Data Collection

Searching HK box office manually online



	A	B	C	D
33	2013	Cold War	寒戰	42,820,819.00
34	2014	The Grandmaster	一代宗師	21,292,885.00
35	2015	The Golden Era	黃金時代	23,456,482.00
36	2016	Ten Years	十年	6,000,000.00
37	2017	Trivisa[3]	樹大招風	9,247,902.00
38	2018	Our Time Will Come	明月幾時有	N/A
39	2019	Project Gutenberg	無雙	34,400,000.00
40	2020	Better Days[4]	少年的你	N/A
41	2022	Raging Fire	怒火	26,370,000.00
42	2023	To My Nineteen-Year-Old	給十九歲的我	10,290,000.00
43	2024	A Guilty Conscience	毒舌大狀	115,060,394.00
44	2025	Twilight of the Warriors: V 九龍城寨之圍城		108,448,257.00

Copy the excel and paste into R by read.table()

Data collected in Excel

*N/A: Box office data is combined with Mainland → Cannot estimate HK box office separately



Data Collection

Searching HK box office manually online

	A	B	C	D
33	2013	Cold War	寒戰	42,820,819.00
34	2014	The Grandmaster	一代宗師	21,292,885.00
35	2015	The Golden Era	黃金時代	23,456,482.00
36	2016	Ten Years	十年	6,000,000.00
37	2017	Trivisa[3]	樹大招風	9,247,902.00
38	2018	Our Time Will Come	明月幾時有	N/A
39	2019	Project Gutenberg	無雙	34,400,000.00
40	2020	Better Days[4]	少年的你	N/A
41	2022	Raging Fire	怒火	26,370,000.00
42	2023	To My Nineteen-Year-Old	給十九歲的我	10,290,000.00
43	2024	A Guilty Conscience	毒舌大狀	115,060,394.00
44	2025	Twilight of the Warriors: V 九龍城寨之圍城		108,448,257.00

→ Data collected in Excel

```
hkfa_data <- read.table(file = "clipboard", sep = "\t", header = TRUE)
```

Copy the excel and paste into R by read.table()

Year	Film	ChineseName	BoxOffice_HKD
1982	Father and Son	父子情	1,483,317.00
1983	Boat People	浪奔怒海	15,475,087.00
1984	Ah Ying	半邊人	1,915,243.00
1985	Homecoming	似水流年	4,556,416.00
1986	Police Story	警察故事	26,626,760.00
1987	A Better Tomorrow	英雄本色	34,161,324.00
1988	An Autumn's Tale	秋天的童話	25,546,552.00
1989	Rouge	胭脂扣	17,476,414.00
1990	Beyond the Sunset	飛越黃昏	4,350,582.00
1991	Days of Being Wild	阿飛正傳	9,751,942.00
1992	To Be Number One	跛豪	38,703,363.00
1993	Cageman	龍民	1,767,657.00
1994	C'est la vie, mon cheri	新不了情	1,149,673.00
1995	Chungking Express	重慶森林	7,678,549.00
1996	Summer Snow	女人四十	14,020,825.00
1997	Comrades: Almost a Love Story	甜蜜蜜	15,557,580.00
1998	Made in Hong Kong	香港製造	1,917,330.00
1999	Beast Cops	野獸刑警	8,317,750.00

Final processed data saved
in “hkfa_data”

→ Fine adjustment according
to inflation

*N/A: Box office data is combined with Mainland → Cannot estimate HK box office separately



Data Collection

Processed data (Box office)



Oscars

	Year	Film	BoxOffice_Millions	Status
45	2007	The Departed	425.54	Success
46	2008	No Country for Old Men	243.71	Success
47	2009	Slumdog Millionaire	522.19	Success
48	2010	The Hurt Locker	66.01	Success
49	2011	The King's Speech	542.62	Success
50	2012	The Artist	169.46	Success
51	2013	Argo	288.05	Success
52	2014	12 Years a Slave	226.27	Success
53	2015	Birdman or (The Unexpected Virtue of Ignorance)	121.79	Success
54	2016	Spotlight	113.49	Success
55	2017	Moonlight	72.85	Success
56	2018	The Shape of Water	214.83	Success
57	2019	Green Book	347.49	Success
58	2020	Parasite	273.05	Success
59	2021	Nomadland	41.04	Success
60	2022	CODA	139.50	Success
61	2023	Everything Everywhere All at Once	139.20	Success
62	2024	Oppenheimer	952.00	Success
63	2025	Anora	56.29	Success



Problem

Data Collection



HKFA



	Year	Film	ChineseName	BoxOffice_Millions_HKD
25	2006	Election	黑社會	23.85
26	2007	After This Our Exile	父子	8.45
27	2008	The Warlords	投名狀	39.65
28	2009	Ip Man	葉問	35.30
29	2010	Bodyguards and Assassins	十月圍城	22.28
30	2011	Gallants	打擂台	1.49
31	2012	A Simple Life	桃姐	35.40
32	2013	Cold War	寒戰	53.10
33	2014	The Grandmaster	一代宗師	25.76
34	2015	The Golden Era	黃金時代	27.68
35	2016	Ten Years	十年	6.90
36	2017	Trivisa	樹大招風	10.36
37	2018	Our Time Will Come	明月幾時有	NA
38	2019	Project Gutenberg	無雙	37.15
39	2020	Better Days	少年的你	NA
40	2022	Raging Fire	怒火	26.90
41	2023	To My Nineteen-Year-Old Self	給十九歲的我	10.29
42	2024	A Guilty Conscience	毒舌大狀	115.06
43	2025	Twilight of the Warriors: Walled In	九龍城寨之圍城	108.45

Analysis

Conclusion



DATA ANALYSIS

Problem

Data Collection



Analysis



Conclusion



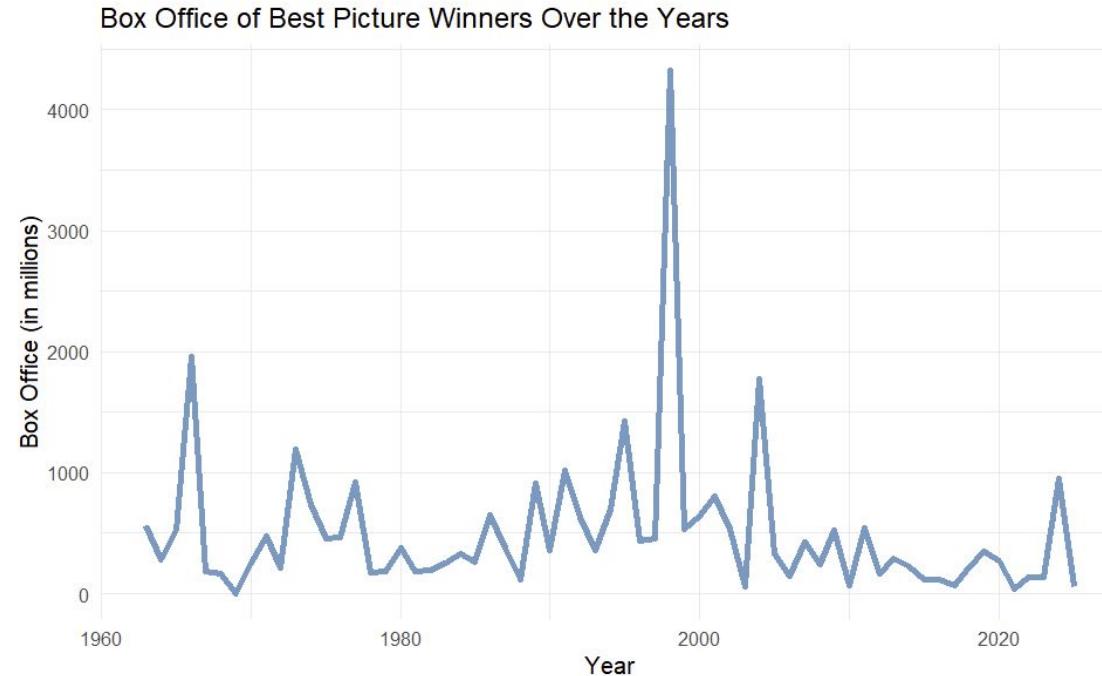
Using ggplot2 to visualize the processed data

```
library(ggplot2)
ggplot(final_results, aes(x = Year, y = BoxOffice_Millions)) +
  geom_line(linewidth = 1.5, color = "steelblue", alpha = 0.8) +
  labs(title = "Box Office of Best Picture Winners Over the Years",
       x = "Year",
       y = "Box office (in millions)") +
  theme_minimal()
```

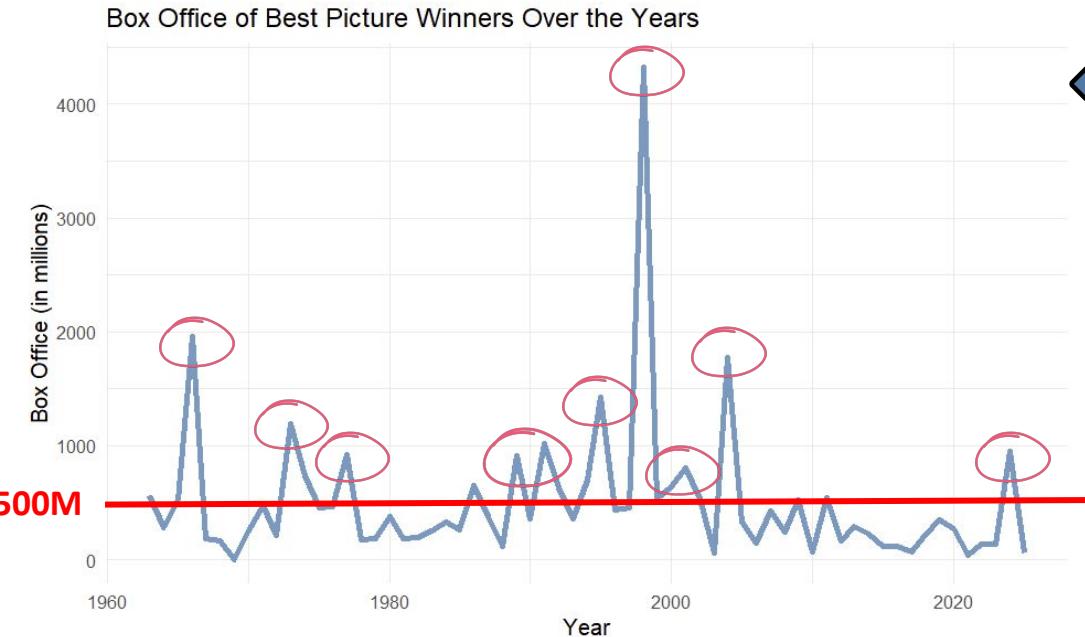
To show correlation of box office and award-winning over years



Using ggplot2 to visualize the processed data



Indications from Oscars' data



Implications

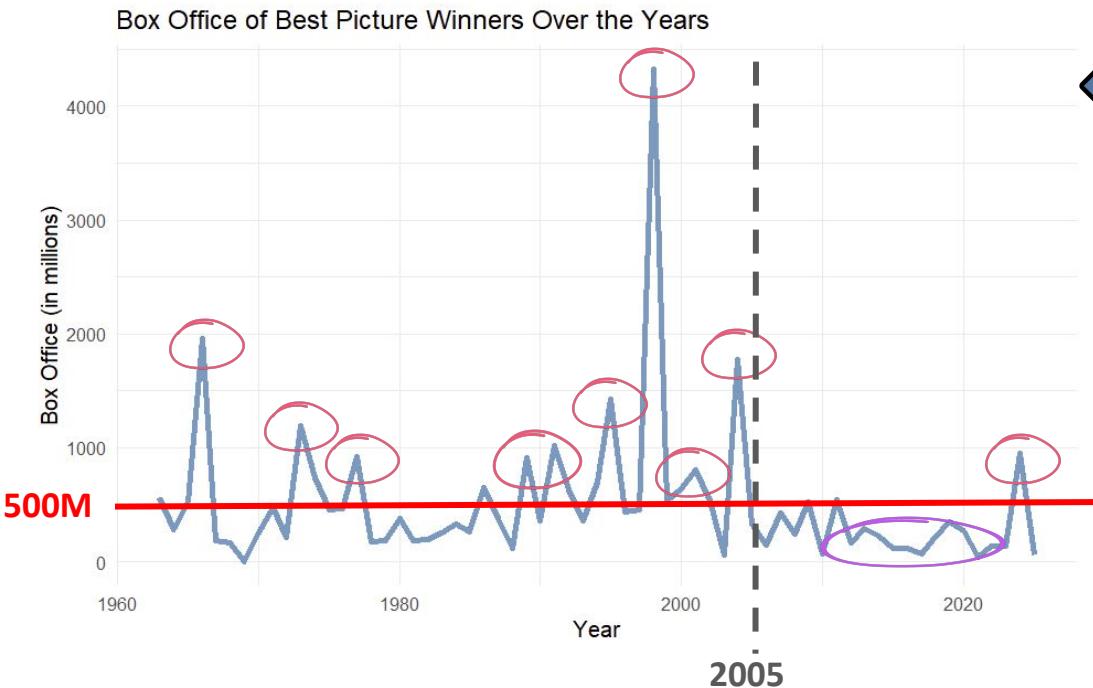
1. Not all award-winning movies achieve high box office revenue

```
average_value <- mean(final_results$BoxOffice_Millions)
average_value
[1] 506.8714
```

- Box office fluctuated
- Average box office = ~500M USD
- ~10 out of 63 films > Average significantly [16%]



Indications from Oscars' data



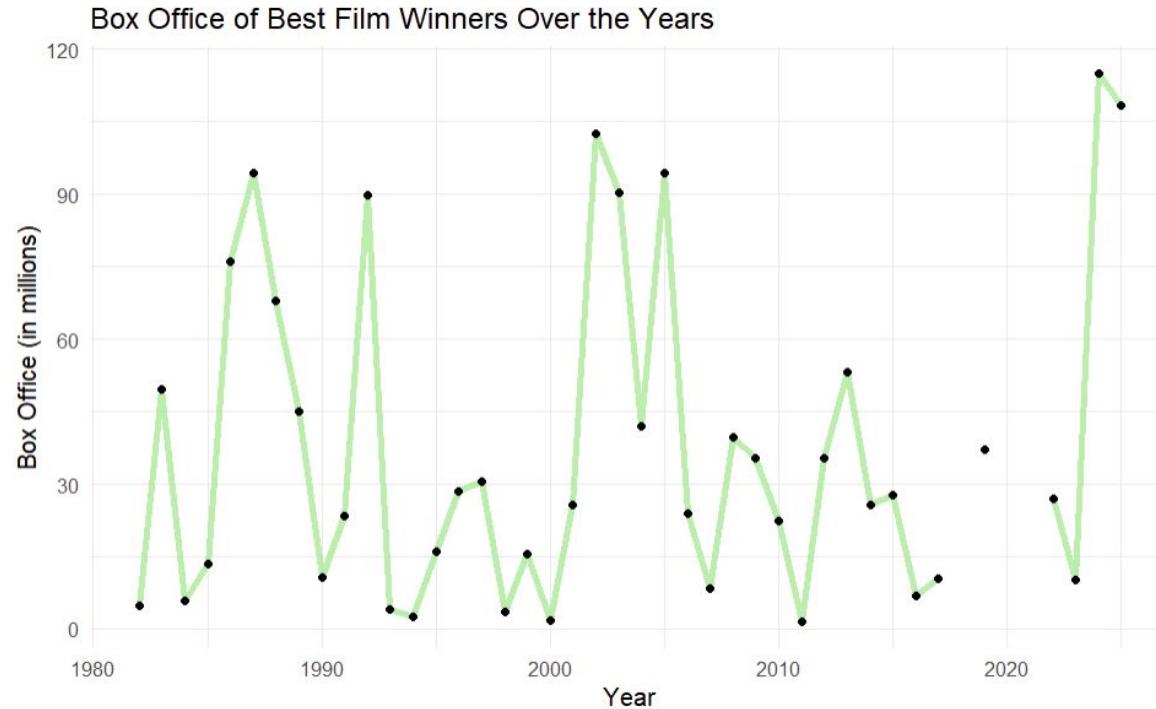
Implications

2. High box office films rarely win Best Picture in the recent years

- Only 1 film in the recent 20 years is way higher than average
- 2012–2022: Best Picture winners had lowest box office earnings.



Using ggplot2 to visualize the processed data



Analysis



Conclusion

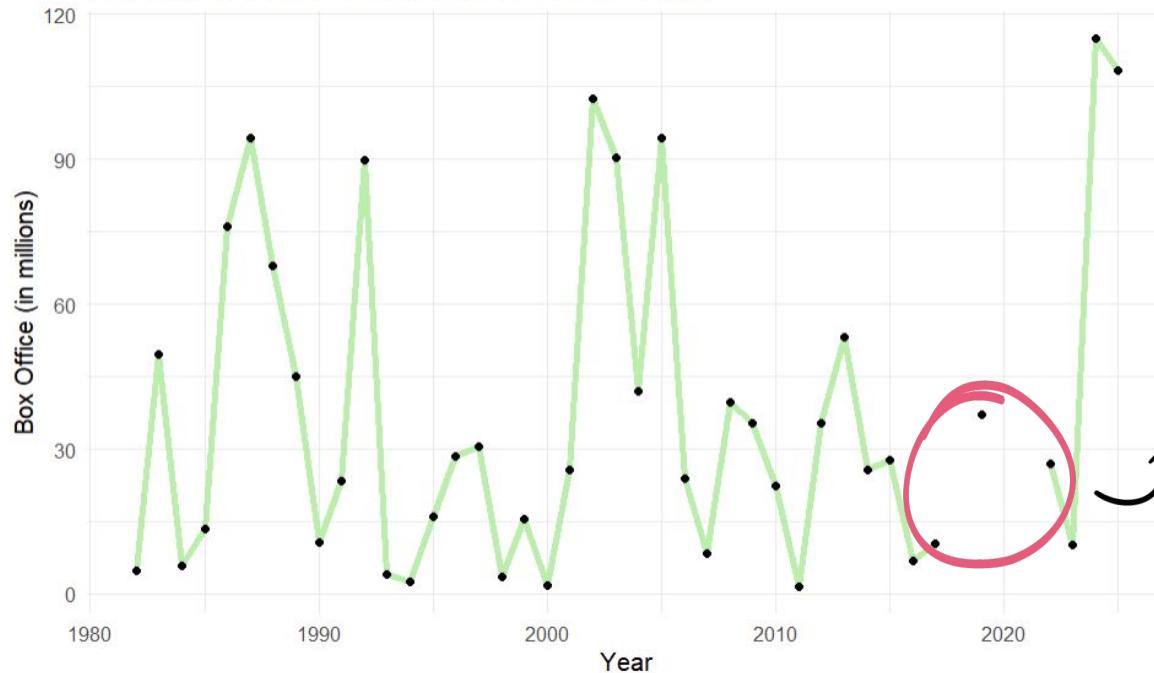
Problem

Data Collection

Using ggplot2 to visualize the processed data



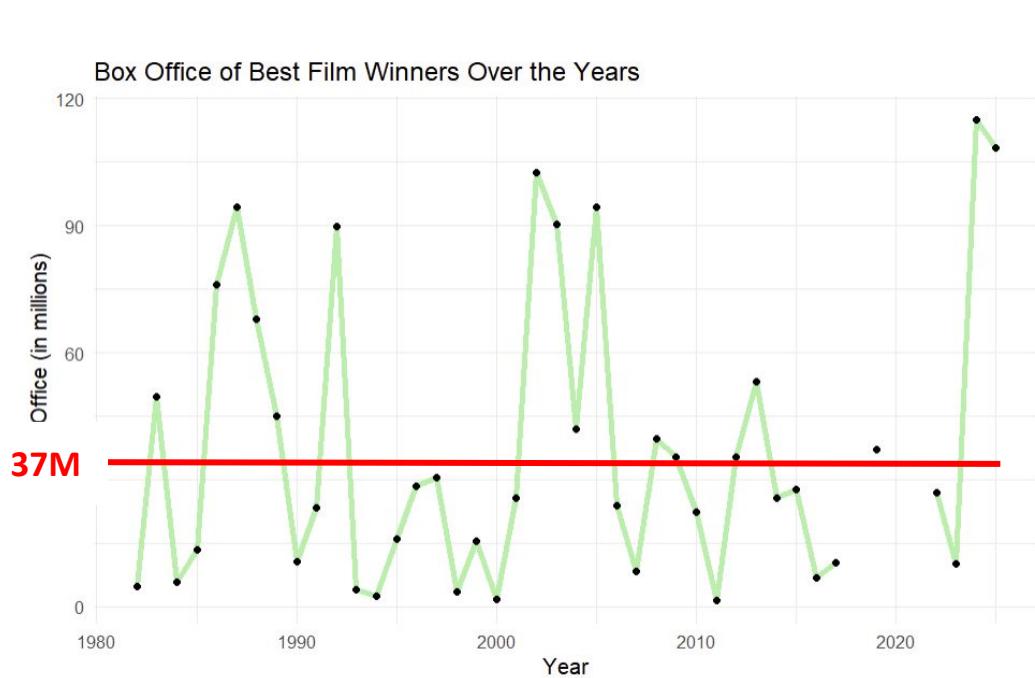
Box Office of Best Film Winners Over the Years



Missing data



- HK box office combined with the mainland box office
- Cannot separate HK data only
- Unfair to use mainland figures.



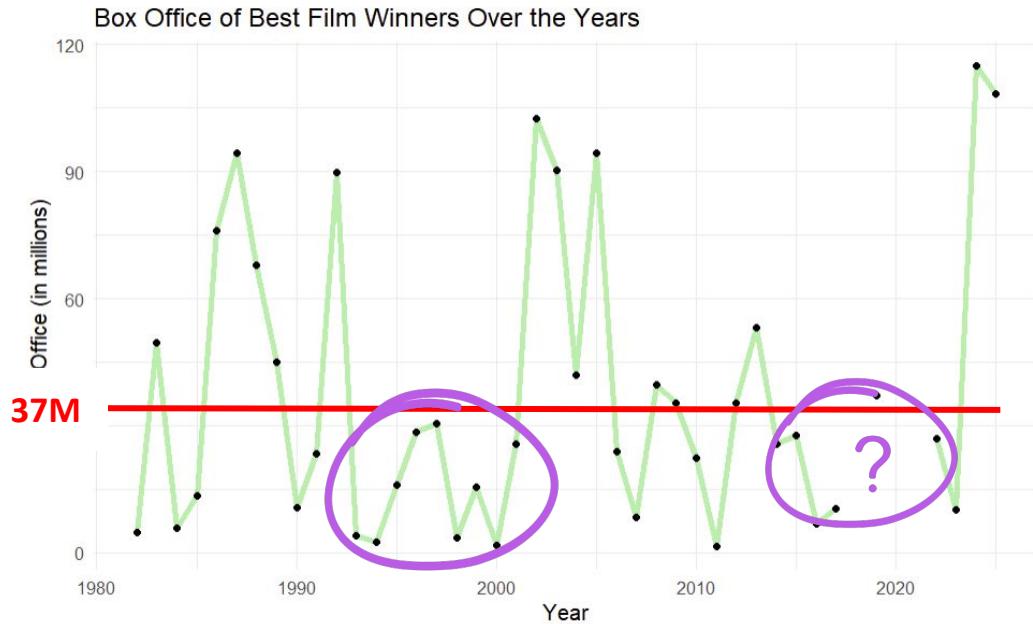
Implications

1. Not all award-winning movies achieve high box office revenue.

```
hk_average_value <- mean(hk_final_results$Boxoffice_Millions_HKD, na.rm = TRUE)
hk_average_value
[1] 37.23415
```

- Box office fluctuated
- Average box office = ~37M HKD
- ~14 out of 41 films > Average significantly [34%]





Implications

2. High box office films have a moderate chance of winning Best Film

- 8 films in the recent 20 years is higher than average
- 1992–2000: Best Film winners had lower box office earnings.
- Uncertain if 2014–2022 Best Film winners had lower box office earning



CONCLUSION

Problem

Data Collection



Analysis

Conclusion



What can we draw from the above?



Research Questions

1. Are award-winning films always **blockbusters**?



2. How has the **judging criteria** of the judges of the awards changed over time?



What can we draw from the above?



Research Questions

1. Are award-winning films always **blockbusters**?



NO for both Oscars and HKFA

- Fluctuated greatly
- Less than half above average → not equally distributed
- Public preference related but **not controlling the awards**

What can we draw from the above?



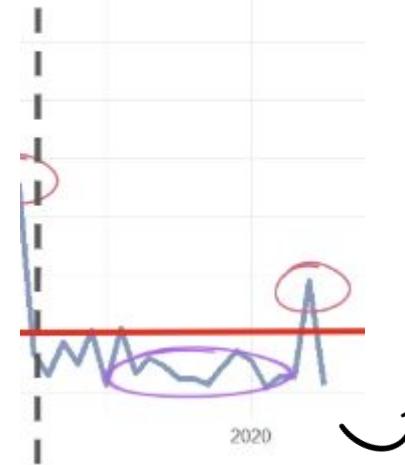
Research Questions

2. How has the **judging criteria** of the judges of the awards changed over time?



Oscars

- Shift to films with lower box office
- **Reduced public engagement**
- **Disconnected??**



- Only 2024's winner > Average
- 2025's winner: Anora
- One of the lowest in the past 20 years



What can we draw from the above?



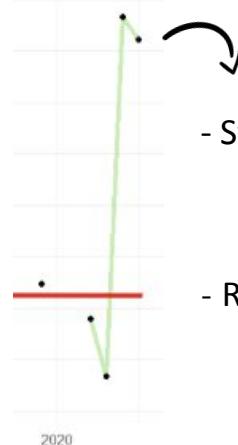
Research Questions

2. How has the **judging criteria** of the judges of the awards changed over time?



HKFA

- More matched with local audiences
- **Rise in public engagement from 2024**



- Stark rise starting from 2024
 - A Guilty Conscience (毒舌大狀) 115M!!
 - Twilight of the Warriors: Walled In (九龍城寨之圍城)
- Revitalization of HK films?





Are film awards losing their authorities?



	Oscars	HKFA (with 2 missing data)
% of Best Film winners \geq average box office	16%	33%
No. of Best Film winners \geq average box office in the past 20 years	3	8
Public engagement with awards	⬇️	⬆️



Are film awards losing their authorities?

Oscars	HKFA
<ul style="list-style-type: none">- High box office does not equal higher award chances- Public preference is less influential- Awards less representative of true public interest	<ul style="list-style-type: none">- 34% doubled from 16%- Greater focus on public preference- Better representation of the public (HK audiences)

Limitations



Missing data

Analysis focused solely on Best Picture winners

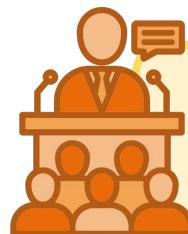


Box office doesn't 100% reflect public interest



Conclusion

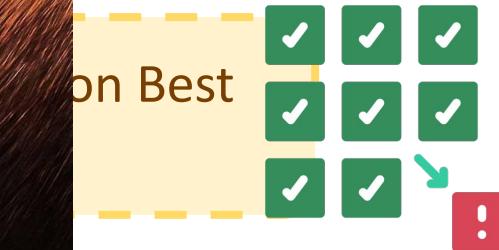
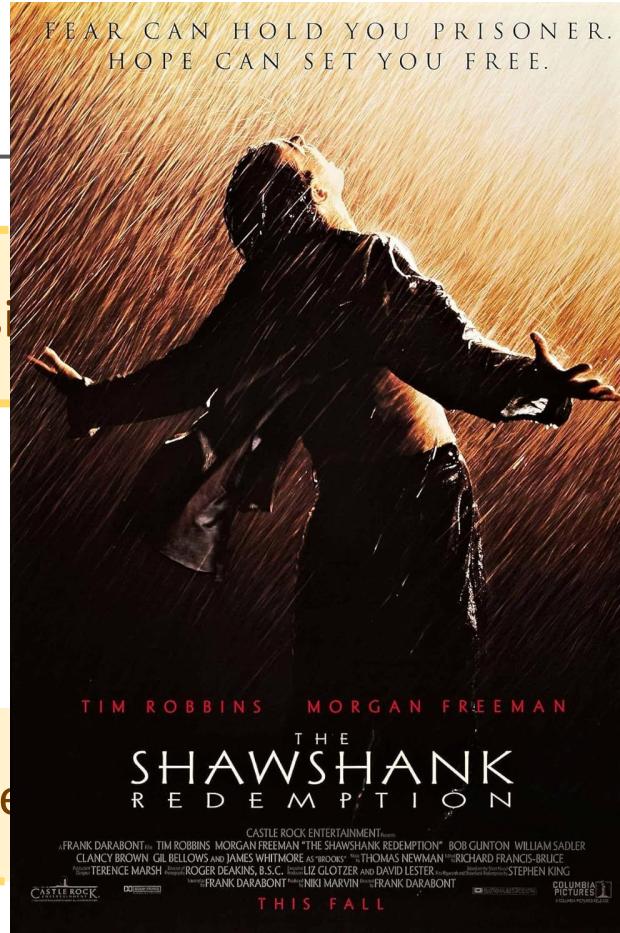
Limitations



Box office



Missing



Public interest



Analysis

Problem

Data Collection

Conclusion





THANK YOU

-- The people's will is the true victory --