

Chou, Yuntsai; Lin, Wei

Conference Paper

Blockbuster or Flop? Effects of Social Media on the Chinese Film Market

24th Biennial Conference of the International Telecommunications Society (ITS): "New bottles for new wine: digital transformation demands new policies and strategies", Seoul, Korea, 23-26 June, 2024

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Chou, Yuntsai; Lin, Wei (2024) : Blockbuster or Flop? Effects of Social Media on the Chinese Film Market, 24th Biennial Conference of the International Telecommunications Society (ITS): "New bottles for new wine: digital transformation demands new policies and strategies", Seoul, Korea, 23-26 June, 2024, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/302460>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Blockbuster or Flop?

Effects of Social Media on the Chinese Film Market

Yuntsai Chou, Yuan Ze University, ychou@saturn.yzu.edu.tw

Wei Lin, Fujian University of Technology, wlin@fjut.edu.cn

Keywords: digital marketing, electronic word of mouth, box office revenue, random forest model, polynomial regression, Maoyan, Douban

1. The Chinese Market

While the release of most Hollywood films was prolonged in 2020 because of the COVID-19 pandemic, *Ba Bai* (The Eight Hundred) became the highest grossing movie in terms of worldwide box office revenue. In 2021, China also reported the highest box office revenue worldwide, which was approximately 7.3 billion USD (Navarro, 2022). “China’s box office share in the global box office increased to 25.4%, overtaking the US as the largest market (Thomala, 2022).” Because economic development in China has incubated a middle class with high incomes, the demand for entertainment has increased considerably. In the postpandemic era, many expect that “China will gain an even greater share of the global box office” (Thomala, 2022). Amid the fierce competition from streaming services, Hollywood studios rely on international markets, especially China, to recoup box office revenue.

Challenges are increasingly encountered regarding the distribution of foreign films to the Chinese market because China has quotas for foreign films and Chinese moviegoers have distinct tastes. Hollywood studios must identify movies with potential to become blockbusters in China and the individual characteristics of movies that facilitate their successful release in China. International digital advertising accounts for more than 50% of the total budget (Guttmann, 2022), and social media advertising expenditures have increased considerably (Statista Research

Department, 2021). For this reason, studios should understand the effects of social media on the success of blockbusters in China.

Social media are a type of mass media and interpersonal communication tool (Baek, 2017: 19). Mass media involve a one-to-many broadcasting mode, in which messages are generated and spread by public accounts, key opinion leaders, or influencers. In online interpersonal communication, also known as “electronic word of mouth” (eWOM), individuals generate messages and share them from their private accounts to their friend and family networks. Mass media entail low user engagement, whereas interpersonal communication entails high user engagement. eWOM is associated with high trustworthiness because people value the opinions of their friends and family members.

We used a timeline model to differentiate broadcasting from eWOM. Broadcasting is used before and during a film’s release to diffuse messages through official accounts or influencers. eWOM is used after a film’s release to diffuse audience reviews to friend and family networks.

We collected data on movies and social media users from two film-specific platforms, Douban and Maoyan, and extracted budget data from Internet Movie Database (IMDb). Our database comprised a total of 957 movies from 2012 to 2019 (before the pandemic). We postulated that a film’s gross box office revenue would be

affected by social media. We used the number of social media users who rated or commented on a movie as the volume attribute and the rating as the valence attribute.

Film distributors normally allocate resources to marketing to promote big-budget films, and movies with small budgets are promoted through eWOM. In this study, we explored big- and small-budget films by using polynomial regression and random forest (RF) models to determine the effects of marketing and eWOM on box office revenue.

This study contributes to the literature on social media. Most studies on the Chinese film market have analyzed Weibo posts, but this study explored Douban and Maoyan to determine their effects on films' box office performance and to validate the results of previous studies. In addition, this study distinguished between eWOM and broadcasting and empirically tested the validity of these two methods. Studies have yielded inconsistent findings regarding the effectiveness of social media; this study addressed this inconsistency. Moreover, this study evaluated the effects of social media on the box office performance of small- and big-budget movies to determine the effectiveness of digital marketing. Our results can serve as reference for marketing in the Chinese film market.

The rest part of this paper is organized as follows. Section II provides a discussion of the theories of broadcasting and eWOM marketing. In Section III, the history of Chinese cinema is described in detail, and an introduction into Douban and Maoyan is provided. Section IV presents the hypotheses and data sources. In Section V, the polynomial regression tests and empirical results are described. Section VI presents the RF model and the results. Section VII presents the conclusion.

2.Film Discussions on Social Media

2.1 Digital Marketing

Because film production costs have skyrocketed in the past decade, investors and studios have sought to identify an effective formula for making a blockbuster. In addition to the genre, casting, director, and costs, marketing is crucial to a film's success. Distributors usually spend half of production costs on marketing (Follows, 2016). Especially with global gross box office revenue decreasing because of the pandemic, distributors must carefully allocate resources to marketing to maximize its effect (Bogaert et al., 2021).

Global marketing expenditures for blockbusters are rarely less than \$US150 million, and at least 70% is used for TV ad slots (Rainey, 2016). Distributors have begun to use multiple channels to market films (Chen et al., 2019). In the United States, digital marketing expenditures account for only 14% of all marketing expenditures but contribute to approximately 46% of box office revenue (Sweeney, 2018).

Major digital marketing platforms include YouTube, Facebook, and Twitter. Paid Facebook posts account for 4% of movies' media budgets but contribute to approximately 9% of opening box office revenue and 20% of marketing-driven sales. Ding et al. (2017) estimated that a 1% increase in the number of Facebook "likes" in the week prior to the release of a film would cause a 0.2% increase in opening-week box office revenue.

All movie genres are associated with expenditure on marketing channels. The genres that rely the most on TV ads are romance (57% of the total expenditure is spent on TV ads), horror (56%), and comedy (55%). Fantasy and sci-fi films spend the most on outdoor marketing (29% and 28%, respectively). The genre that uses the radio the most is horror. Because online audience tends to be younger, horror, which appeals to younger people, also frequently uses online advertising (Follows, 2019).

2.2 Broadcasting and eWOM

WOM is a form of person-to-person communication without commercial influence (Nieto et al., 2014). WOM messages tend to be more reliable and credible than mass media communication (Brown et al., 2007). Because most consumers can recognize fake positive reviews of products, only trustworthy comments from friends and family members can encourage them to see a movie (Li & Hitt, 2008).

WOM is particularly important for experiential goods, such as those in the tourism and hospitality industries, which are difficult to evaluate before purchase (Baek, 2017: 13 and 14). WOM developed into eWOM because of the Internet. Social media are a platform that can be used to spread eWOM and “accelerate the exchange of information among users and reduce information asymmetries between buyers and sellers, making the selling process more efficient” (Castillo, 2021).

Figure 1 provides a comparison of broadcasting with eWOM. Broadcasting involves the use of distributors’ public accounts for disseminating mass generated content (MGC); social media users only engage in pushed content or content they search for. eWOM is created by users, which is why it is called “user-generated content” (UGC), and it entails users socializing and participating in a community (Castillo, 2021: 113516). eWOM for movies spreads through online reviews, video sites, blogs, and social media pages (Baek, 2017: 15; Kim et al., 2015). Huete-Alcocer (2017) identified three features that distinguish eWOM from other online marketing modes: credibility, privacy (anonymity), and accessibility.

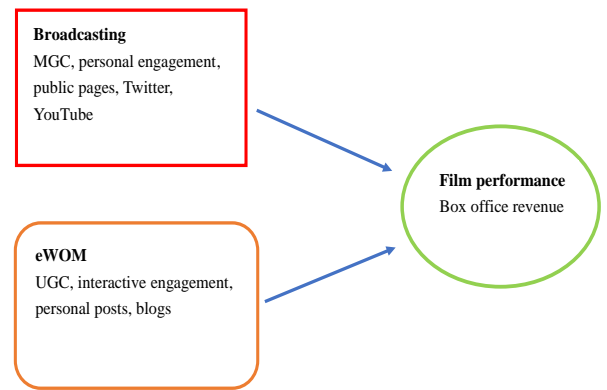


Figure 1

Comparison of Broadcasting and eWOM

Source: (Chen et al., 2020: 1569)

Although many studies have examined the effects of digital marketing on film performance (Ruus & Sharma, 2020), the results regarding communication modes, platforms, and measurement proxies are inconsistent. Gong et al. (2017) revealed that UGC is more effective than MGC in explaining film performance. Castillo et al. (2021) indicated that eWOM affects movie selection, and that interactive engagement mostly occurs during prerelease and the opening weekend. Chen et al. (2020) revealed that MGC has a stronger effect on ticket sales than does UGC.

Liu (2006) identified eWOM as a reliable source of information because it reflects a film’s popularity among moviegoers. Pandya et al. (2018) determined that the effect of preconsumption WOM on ticket sales is stronger than that of postconsumption WOM. Deng (2020) compared users’ reviews of movies with those of critics and determined that users’ reviews were an “average of moviegoers’ opinions, more plausible than respective eulogy” (Deng, 2020: 1246). Thus, (structured) numeric valence in user reviews has a stronger effect on ticket sales than does critics’ (unstructured) textual narratives.

Abidi et al. (2020) predicted a movie’s success by exploring the number of trailer views on YouTube, likes on Facebook, and followers on Twitter. Asur and Huberman (2010) used the

volume of tweets and two measures of sentiment (i.e., polarity and subjectivity) to model box office revenue and observed a strong correlation among them. Rui et al. (2013) determined that the number of tweets and the positive tweet ratio strongly affect box office sales. Baek (2017) noted that Twitter has mass media-like characteristics that have strong effects in the initial stage of a movie's release, and that Yahoo! Movies has interpersonal communication-like characteristics that have strong effects in the late stage of a movie's release (Baek, 2017: 20). Ding et al. observed a positive correlation between the number of prerelease likes on the online movie review site IMDb and the total box office revenue (2017: 79). Ruus and Sharma (2020) compared the predictive power of tweet sentiment analysis, IMDb, and Rotten Tomatoes and discovered that reviews on Twitter are more positive than those on and Rotten Tomatoes.

Oh et al. (2017) revealed that Twitter's power to predict movie sales decreased when Facebook data were included in the model. Bogaert et al. (2021) noted that Twitter has lower source credibility than does Facebook, suggesting that UGC on Facebook predicts movie performance. Ding et al. (2017) examined the effect of the number of likes on Facebook movie pages on box office sales and discovered that a 1% increase in prerelease likes would lead to a 0.2% increase in opening-week box office sales.

Volume and valence are indicators of eWOM (Chen et al., 2020). Studies have demonstrated their effects on box office revenue (Dellarocas et al., 2007: 23–25; Rui et al., 2013: 867). Niraj and Singh (2015) revealed that UGC volume and valence significantly and positively affected box office performance after the effect of critics' reviews was controlled for. Bogaert et al. (2021) studied 218 movies released between 2012 and 2015 and discovered that volume is the most important measure of box office revenue, followed by the combination of volume and valence and valence alone: "Release day performance is more

affected by blog volume and advertising before the release day whereas post-release performance is affected by post-release blog valence and advertising" (Castillo, 2021: 9).

Liu (2006) determined that volume, as measured using the total number of user interactions on Yahoo! Movies, is the most important factor affecting movie sales, and that valence has no significant effect. Duan et al. (2008) demonstrated that the volume of online user reviews, but not their content, significantly affects box office revenue. Lipizzi et al. (2016) analyzed 22 Oscar-winning films in 2014 and the related Twitter posts and concluded that online sentiment has less power than does traffic (i.e., volume) in predicting box office sales during a movie's opening weekend.

Chintagunta et al. (2010) observed that valence, but not volume, is the most influential variable. Bhattacharjee et al. (2017) revealed a significant relationship between the polarity of social media content (i.e., valence) and box office revenue (2017: 365). By analyzing sentiment in users' reviews, Chen et al. (2020) discovered that negative reviews had a stronger effect on box office revenue than did positive ones (2020: 1,567 and 1,568). Apala et al. (2013) demonstrated the significant relationships among sentiment, volume, and box office performance.

In a study on Hollywood cinema, Ruus and Sharma (2020) identified the top three factors predicting box office performance: the number of tweets, budget, and the number of theaters showing the movie (2020: 989). Kim et al. (2018) used lexicon-based sentiment classification and machine learning methods and analyzed 147 Korean films; the results indicated a positive and significant relationship between user sentiment and box office performance. Kang (2021) examined Naver search data for Korea movies in 2018, and the results revealed that the data played a key role in predicting ticket sales (2021: 42).

Liu et al. (2014) researched 57 Chinese movies with 5 million tweets collected from Weibo microblogs and determined that eWOM predicts box office revenue. Qiu and Tang (2018) collected Weibo text and Douban movie reviews for 10 movies in 2017 and demonstrated that they both have strong power for predicting movies' box office performance in China. Tang et al. (2018) compared Western and Chinese movies of the same genre and revealed that both positive and negative sentiments are robust indicators of the popularity of a movie (2018: 80–82).

3. The Chinese Film Market and Social Media

3.1 Douban

The primary social media platforms for video entertainment in China are Douban Movies (Douban) and Maoyan Entertainment (Maoyan). Douban is the Chinese version of IMDb and displays users' reviews and ratings of movies released in China. Douban was created in March 2005 and offers information regarding films, books, and music as well as a space for social interaction in which users rate and share comments. Douban expanded its business to film ticket sales in 2012, covering 81 movie theater chains in 33 major cities (Baidu Encyclopedia, 2012).

Douban attracted its first 1 million users in approximately 2 years, and the number of unique visitors had increased to 8.23 million by 2009 (Sun, 2013). As of 2012, Douban had 62 million users, and the daily number of page views was 120 million, which Douban claimed to be equivalent to 200 million unique monthly visitors (Netease, 2012). Since then, its growth has been moderate compared with that of other Chinese social media platforms because of the elite nature of its users. In 2020, the number of registered accounts was 75 million, and the number of daily page views was approximately 210 million.

Douban launched an app in 2014, and the

number of domestic Android downloads was 5.5 million as of 2022 (Table 1). The number of active monthly users of the app increased from 3 million in 2014 to 7.25 million in 2020 and 12 million in 2021. Douban users are mostly young men living in second-tier cities. Users aged less than 40 years account for 90% of all users, and male users account for 52.9% of all users (qianfan.tech).

Table 1

Number of App Downloads, 2016–2021

PLATFORM	DOUBAN	MAOYAN
2016	20,906,022	22,624,209
2017	31,009,773	10,784,137
2018	48,739,303	22,918,841
2019	34,585,790	22,155,950
2020	47,359,413	18,950,932†
2021	37,239,830	37,909,034

†: The number of downloads decreased because movie theaters were closed during the COVID-19 pandemic.

Source: business analysis platform for mobile products (*qimai.cn*)

3.2 Maoyan

Maoyan was launched as the movie ticket sales platform of Meituan in 2012 and spun off in 2015 (Maoyan, 2022). Maoyan made its initial public offering in Hong Kong in February 2019. Its revenue sources include online ticket sales, film and video royalties, and ads. In 2019, its annual revenue was CN¥4.27 billion, and the net profit was CN¥458.9 million before the pandemic (Table 2). In 2020, its annual revenue decreased to CN¥1.37 billion, and the net profit turned negative CN¥646.3 million (Table 2).

Table 2

Maoyan's Annual Revenue, 2017–2021

unit: CN¥ million

YEAR	REVENUE	NET PROFIT
2017	2,548.0	216.0
2018	3,755.0	(138.4)
2019	4,267.5	458.9
2020	1,365.7	(646.3)
2021	3,323.4	368.5

Source: 2017–2021 Maoyan Entertainment Annual Financial Reports

A survey revealed that more than 90% of Chinese netizens purchase movie tickets online

(China Film Archive [CFA]).¹ Two platforms—Maoyan and Taobao Dianyian (a subdivision of Alibaba Group)—became a duopoly after stiff competition among online ticket sales platforms. In Q4 of 2017, Maoyan had 51.9% market share, and Taobao had 37.7% market share (Maoyan, 2022). Maoyan’s market share increased to 61.3%, but Taobao’s decreased to 0.7% only half a year later (Q2 2018; qianfan.tech). The two platforms combined account for more than 90% market share of online ticket sales. Maoyan has become the leading movie portal, offering real-time box office data from theaters nationwide (Chen et al., 2020).

In 2018, Maoyan’s monthly page views numbered 1.1 billion, the platform had 150 million user ratings, and the number of unique monthly visitors exceeded 150 million (Maoyan, 2019). Maoyan launched an app in 2012, and the Android app has been downloaded 325 million times domestically (qianfan.tech). Unlike Douban users, a small group with high-brow taste, Maoyan users are ordinary moviegoers, which is why Maoyan has more users.

3.3 The Film Market in China

China reformed the film distribution system in 2001, so that private and interprovince cinema chains could be established and Hollywood and Hong Kong films could be imported under the quota control (CFA). In 2010, the Chinese government issued a series of film policies to help the industry thrive.

Private capital and Hong Kong talent were invested in film production to increase the capacity of the industry and improve the quality of its films. Box office revenue increased considerably between 2012 and 2015, and Chinese films accounted for more than 50% of the revenue (HuaJing Info).² This increase in box office revenue in China was mainly due to the increase in movie screens and the moviegoing population in third- and fourth-tier cities (Table 3). In 2021, China surpassed the United States to become the highest grossing film market worldwide, and Chinese films’ share of the domestic box office revenue increased to 84.49% (Figure 2).

¹<http://www.cfa.org.cn>

Table 3

Box Office Revenue in China, 2012–2021

Year	Number of screens	Box office rev. (CNY billion)	Top grossing movie	Box office rev. of top grossing movie (CNY billion)
2013	18,195	21.80	<i>Journey to the West</i>	1.25
2014	23,600	29.60	<i>Transformers: Age of Extinction</i>	1.98
2015	31,600	44.07	<i>Monster Hunt</i>	2.44
2016	41,129	45.40	<i>The Mermaid</i>	3.39
2017	50,076	55.90	<i>Wolf Warrior 2</i>	5.69
2018	60,079	60.98	<i>Operation Red Sea</i>	3.65
2019	69,787	64.27	<i>Ne Zha</i>	5.04
2020	75,581	20.40	<i>The Eight Hundred</i>	3.11
2021	81,317	47.26	<i>The Battle at Lake Changjin</i>	5.70

Source: HuaJing Info (www.huaon.com)

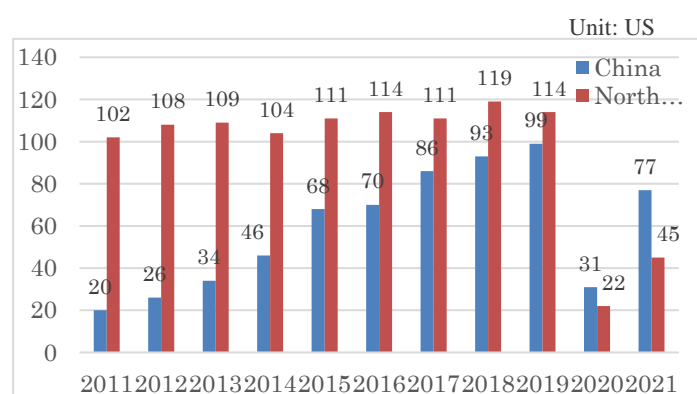


Figure 2

Box Office Revenue in China and North America, 2011–2021

Source: HuaJing Info (www.huaon.com)

4. Modeling the Effects of Social Media

4.1. Proxies

Although many studies have discussed the effects of online ratings and comments on movies’ box office performance, their empirical results have varied depending on the scope of analysis, social media platform, algorithm used for measurement

²<http://www.huaon.com>

and estimation, and film origin. We focused on the Chinese market and explored whether social media ratings and reviews affect films' box office performance. Our first hypothesis was as follows:

H1: The volume and valence of online ratings and comments affect films' box office revenue.

Baek (2017) differentiated broadcasting and eWOM by timeline; MGC is more influential in the initial stage of movie release, whereas UGC, entailing interaction, is more influential in the late stage of the release (Baek, 2017: 20). We used a timeline to visualize the process of broadcasting and eWOM modes (Figure 3). Broadcasting occurs during prerelease, when official accounts and influencers disseminate marketing messages. After a film is released, viewers share comments online, engaging in eWOM.

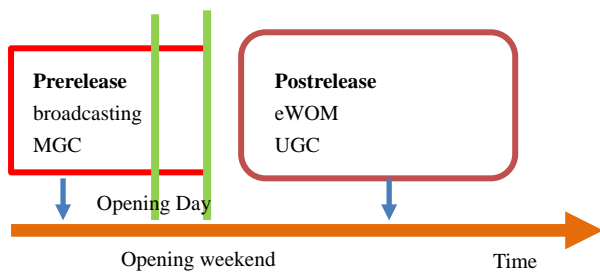


Figure 3
Timeline of Digital Marketing
Source: author

Studies have also demonstrated that eWOM has a lagged effect on box office revenue because information requires time to diffuse (Godes & Mayzlin, 2009: 731; Karniouchina, 2011: 67). Song et al. used a model of the ability of MGC and UGC to predict box office revenue and discovered that it outperformed other models (2019: 191 and 192). Ding et al. provided empirical evidence indicating that the increase in the number of likes on a film's IMDb page in the week prior to its release strongly affects opening-week ticket sales, but this effect weakens over time (2017: 81). These results indicate that eWOM can replace broadcasting after a film's release. We postulated that opening-weekend box office performance is primarily affected by broadcasting, and that eWOM increases box office revenue after the opening weekend because of the lagged effect of social

media (Baek et al., 2017: 18; Figure 3).

Opening-weekend (i.e., Friday through Sunday) box office revenue accounts for approximately one-third of gross box office revenue (Turner, 2015a) and thus serves as a robust indicator of a film's financial success (Oh et al., 2017: 32). We created a dependent variable—opening-weekend box office revenue's contribution to gross box office revenue—to measure the effects of broadcasting and eWOM on box office performance. Strong eWOM should increase box office sales after the opening weekend and decrease opening-weekend box office revenue's contribution to the gross revenue. Weak eWOM should not increase box office revenues after the opening weekend and would result in the high contribution of opening weekend box office revenue to the gross revenue. Thus, the following hypothesis can be proposed:

H2: Strong eWOM decreases opening-weekend box office revenue's contribution to gross revenue.

4.2. Data Collection and Variable Coding

Because Maoyan ticket services have been available since 2012, we collected data on Douban and Maoyan from 2012 to 2019, before the pandemic disrupted theatrical release. A total of 140,502 movies and TV series and 4,428,475 comments were extracted from Douban, and a total of 3,000 movies and TV series were extracted from Maoyan. Our database comprised 957 movies in both Douban and Maoyan. Budget information was collected from IMDb.

Our literature review revealed that volume and valence or their combination are the most cited proxies for eWOM (Asur & Huberman, 2010). For each movie, the number of users who rated or commented on Douban and Maoyan was collected to represent the volume; ratings were used as valence. Ratings were converted from the 5-star system on Douban and Maoyan to a 10-point Likert-type scale. Douban also provides the ratio of positive comments to negative and neutral comments for each movie. Ahmad et al. (2020) used the weighted ratios of positive to negative comments and likes to dislikes to predict movie revenue (2020:102278). We calculated the ratio of

positive to negative comments (p/n) on Douban, which also served as a valence variable (Table 4).

Table 4
Variables

Variable	Value/unit	Description	Source
<i>Box_office</i>	CN ¥10,000	A movie's daily box office revenue in the 2 months after its release	Maoyan
<i>Box_office_p</i>	{1, 0}	Percentage of gross revenue accounted for by opening-weekend revenue (= <i>Box_office</i>)	Generated
<i>Budget</i>	US\$10,000	A movie's production cost	IMDb
<i>Nation</i>	{2, 0}	1 = Hollywood, 2 = other regions, 0 = domestic (China)	Maoyan/generated
<i>Season</i>	{1, 0}	1 = holiday season (Lunar New Year, May 1st holiday, summer, October 1st holiday) 0 = nonholiday	State Council of China
<i>Screen_no</i>	>0	The number of screens showing a given film	Maoyan
<i>Sequel</i>	{1, 0}	1 = sequel, 0 = nonsequel	Maoyan/generated
Volume			
<i>Douban_vote</i>	Frequency	Number of users who rated a given film on Douban	Douban
<i>Maoyan_vote</i>	Frequency	Number of users who rated a given film on Maoyan	Maoyan
<i>Douban_user</i>	Frequency	Number of users who shared comments on a given film over Douban	Douban
Valence			
<i>Douban_score</i>	{10, 1}	Rating for a given film on Douban	Douban
<i>Maoyan_score</i>	{10, 1}	Rating for a given film on Maoyan	Maoyan
<i>Douban_value</i>	>0	p/n value = ratio of positive to negative comments	Douban
Genre			

Drama	{1, 0}	1= drama, 0 = not drama	Douban
Romance	{1, 0}	1= romance, 0 = not romance	Douban
Sci-fi	{1, 0}	1 = sci-fi, 0 = not sci-fi	Douban
Kung-fu	{1, 0}	1= kung-fu, 0 = not kung-fu	Douban
History	{1, 0}	1= period, 0 = not period	Douban
Horror	{1, 0}	1= horror, 0 = not horror	Douban
Adventure	{1, 0}	1= adventure, 0 = not adventure	Douban
Comedy	{1, 0}	1= comedy, 0 = not comedy	Douban
Cartoon	{1, 0}	1= cartoon, 0 = not cartoon	Douban
Children's/family	{1, 0}	1 = children's, 0 = not children's	Douban

source: author

Baek et al. used the screen number, genre, and sequel status to predict box office revenue (2017: 19). Abidi et al. (2020) revealed that genre, Motion Picture Association of America rating, studio, director, and nomination for best picture affect box office performance. Castillo et al. (2021) demonstrated that engagement, country, distributor, and number of screens significantly affect opening-weekend revenue. King created a model comprising genre, nationality, director, distributor, ticket price, and release date to predict box office revenue (Kang, 2021: 42). Deng explored the effects of online reviews on box office performance and controlled for the effects of screen number, marketing cost, budget, release date, and genre (Deng, 2020: 1,248). We used genre, country of origin, screen number, release date, sequel status, and budget as control variables in our models (Table 4). To control for the disturbance from price inflation in the longitudinal data, we set the 2012 consumer price index to 100% to deflate the variables *Box_office* and *budget*.

Film distributors usually allocate resources to marketing resources to promote big-budget (i.e., A-list) films and ensure investment returns (Turner, 2015b). We explored whether social media affect the ticket sales of A-list films differently from other films and whether films with smaller budgets have must rely on eWOM to increase ticket sales.

We conducted a cluster analysis in which films were stratified by budget. We used K-means clustering to minimize the least squares distance from each center of each cluster K and obtained $K = 2$ through iterative tests. We referred to the two groups as the A-list and low-budget films. Figure 4 presents the distinctive distribution of valence scores in the groups. The third hypothesis was as follows:

H3: eWOM decreases opening-weekend revenue's contribution to gross revenue more for low-budget films than for A-list films.

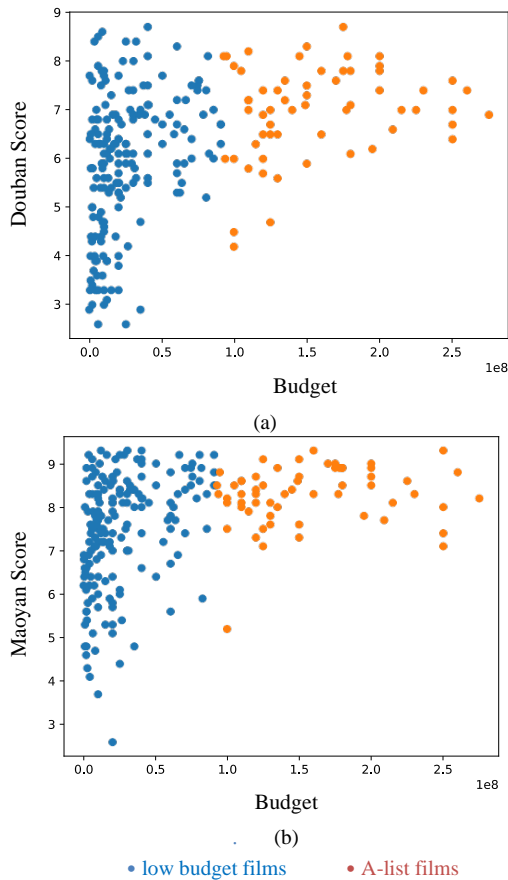


Figure 4
Cluster Analysis of Films Released in China from 2012 to 2019
Source: author

The A-list group included 248 Hollywood and Chinese films released domestically and overseas, and the average budget was US\$58.2 million (Table 5). The low-budget group included 709 Chinese films released domestically, with an average budget of US\$1.7 million. A-list status was

associated with higher volume and valence scores on both Douban and Maoyan than was low-budget status (Table 5).

Table 5
A- List and Low-Budget Films

	Total	A-list	Low budget
Number	957	248	709
Average budget (US\$ million)	15.0	58.2	1.7
Average <i>Douban_vote</i> (user number)	52,071	97,134	36,309
Average <i>Douban_score</i> (rating value)	4.924	6.099	4.513
Average <i>Douban_user</i> (user number)	23,110	38,495	17,729
Average <i>Douban_value</i> (p/n value)	2.76	5.36	1.85
Average <i>Maoyan_vote</i> (user number)	83,591	115,676	72,366
Average <i>Maoyan_score</i> (rating value)	6.825	7.651	6.536

Source: author

4.3. Prediction Models

Four models are often used to test the effects of social media on box office performance: linear regression, polynomial regression, support vector machine, and RF. Ahmad et al. (2020) employed four machine learning algorithms, namely linear regression, support vector machine, a multilayer perception neural network, and RF, to predict movie revenue on the basis of sentiment in YouTube trailers' comments. Huang and Pai (2020) compared least squares support vector regression (LSSVR), the back propagation neural network, generalized regression neural networks, and the multivariate linear regression model and revealed that LSSVR was the best at predicting ticket sales (Huang & Pai, 2020: 8 and 9).

Ma and Wan (2021) evaluated box office records by using unsupervised methods such as association rules and cluster analysis. They also conducted factor analysis, which revealed that nationality and being an adaption of a novel are the

primary factors affecting box office performance. Kang modified the Bass diffusion model by incorporating a marketing-mix variable (i.e., Naver search trend data) to predict ticket sales (2021: 50). Zhang (2022) used a two-stage process to predict box office revenue by first obtaining sentiment values from microblogs by using a naïve Bayes model and by incorporating these values into a Bass model. Castillo (2021) used partial least squares path modeling, a well-known form of structural modeling, to measure latent variables, namely personal engagement and interactive engagement, and to predict opening-weekend box office performance (2021: 113516). Both Oh et al. (2017) and Chen et al. (2020) have conducted stepwise ordinary least square tests to examine eWOM and its relation to box office performance. Kim et al. (2015) applied vector regression to estimate the nonlinear relationships between online reviews and box office performance.

In this study, we conducted regression tests to assess the effects of social media on box office performance. Assuming a nonlinear relationship between *Box_office* (Box_office_p) and social media variables, we estimated the coefficients by using polynomial regression algorithms (Fan, 1996). We also used the RF model to evaluate the explanatory power of social media (Kang, 2021).

5. Effects of Broadcasting and eWOM

5.1 Polynomial Regression Model

Polynomial regression is a special type of [multiple linear regression](#) because it fits a nonlinear relationship between independent and control variables (Freeman, 2009: 26). A linear regression for estimating *Box_office* (Y_i) is given by (F1),

$$Y_i = \alpha + \beta X_i + \gamma SM_i + \delta G_i + \varepsilon_i \quad (F1)$$

where $E(\varepsilon_i) = \mu$; $Var(\varepsilon_i) = \sigma^2$; X_i represents movie-related information, such as screen number, budget, sequel status, nationality, and release date; SM_i represents volume and valence; and G_i denotes a set of discrete genres. Polynomial regression transforms (F1) into a quadratic form, given by (F2), because nonlinear relationships between volume and valence and box office revenue are assumed (i.e., decreasing returns in terms of box office revenue due to volume and

valence).

$$Y_i = \alpha + \beta X_i + \gamma SM_i + \theta SM_i^2 + \delta G_i + \varepsilon_i \quad (F2)$$

In (F2), the least squares algorithm is used to estimate the coefficients in a linear form because of the concern of efficiency, even though nonlinearity is assumed. When the value of a social media attribute is increased from n to $n + 1$, the expected change in box office revenue is $\gamma + 2\theta avg SM_i$, given by (F3).

5.2 Increases in Box Office Revenue Due to Volume

Figures 5 and 6 visualize the correlations between Douban and Maoyan volume and box office revenue in the log form. The correlation coefficients are all positive and significant, which demonstrates the effect of volume on box office revenue when all other factors remain constant.

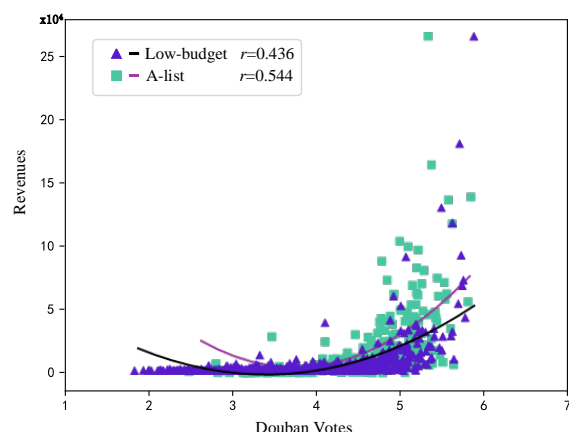


Figure 5

Correlation Between Douban Volume and Box Office Revenue

Source: author

Both Figures 5 and 6 present that the correlation coefficients for the A-list films are higher than those for the low-budget ones. This observation means that the box office revenue of the A-list films can be estimated more effectively by using social media volume than can that of the low-budget films.

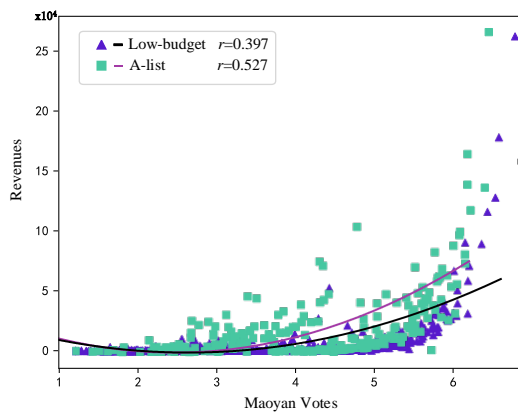


Figure 6

Correlation Between Maoyan Volume and Box Office Revenue

Source: author

Table 6 presents the results of running regressions of social media attributes against box office revenue. Column A presents the coefficient estimates for all 957 films, Columns B and C present those for A-list movies, and Columns D and E present those for the low-budget ones. Columns B and D present the coefficient estimates for the number of Douban users who rated films and the associated rating and Columns C and E present the coefficients for the number of Douban users who commented on films and the associated p/n values. For all estimations, the Maoyan volume is represented by the number of rating users and the valence of their ratings. *Maoyan_vote*'s coefficient is positive (0.05–0.09) and significant at the 0.1% level. This result indicates that Maoyan volume contributes to increases in box office revenue. The coefficients for *Douban_vote* and *Douban_user* are 0.02–0.04 and 0.05–0.09, respectively, and significant at the 0.1% level; both Douban volume attributes increase box office revenue.

	MAOYAN_VOTE	MAOYAN_SCORE	NATION	SEASON	SCREEN_N	SEQUEL	BUDGET	GENRE	ROMANCE	SCI-FI	KUNG-FU	HISTORY	HORROR/THRILLER	COMEDY	CHILDEN'S	CARTOON	ADVENTURE	SAMPLE SIZE	F-VALUE	R-SQUARE VALUE	NATION DUMMY	SEASON DUMMY					
	0.078*** [0.003]	0.050*** [0.005]	0.094*** [0.003]	-511.690* [223.461]	-1,217.973 [807.129]	-379.110 [212.372]	-941.735 [816.346]	-2,749.628* [1,387.661]	-847.318 [1,242.424]	15.002 [8.819]	48.338* [19.658]	0.218 [9.05]	0.026*** [0.002]	0.039*** [0.005]	0.016*** [0.002]	5,908.011*** [1,328.531]	5,915.769** [2,697.5]	3,628.851* [1,429.641]	0.135 [0.172]	0.821*** [0.172]	omitted	DRAMA	-1,735.134* [849.905]	-1,136.897 [1,817.173]	-640.962 [916.194]		
								ROMANCE	-1,405.857 [885.635]	-1,920.475 [2,141.163]	-450.814 [900.501]	SCI-FI	2,970.528 [1,636.825]	-2,399.198 [2,367.401]	105.154 [3,094.151]	KUNG-FU	-2,166.861 [3,499.93]	-14,788.030* [7,340.238]	2,394.239 [3,710.542]	HISTORY	968.370 [2,670.377]	6,294.61 [5,167.427]	2,031.512 [3,039.553]	HORROR/THRILLER	-400.438 [1,130.911]	-998.855 [2,243.034]	134.559 [1,241.743]
								COMEDY	-518.251 [824.121]	258.300 [1,985.981]	-801.501 [837.518]	CHILDEN'S	911.605 [2,260.029]	7,733.617 [12,023.71]	1,003.191 [2,080.725]	CARTOON	-3,027.919* [1,268.433]	-4,987.502 [2,688.087]	-1,067.131 [1,420.131]	ADVENTURE	2,042.356 [1,045.582]	-2,763.212 [2,023.48]	2,195.098 [1,240.235]	SAMPLE SIZE	957	248	709
								F-VALUE	502.26***	80.12***	506.00***	R-SQUARE VALUE	0.904	0.859	0.928	NATION DUMMY	v	v	v	SEASON DUMMY	v	v	v				

Table 6

Effects of Social Media on Box Office

VARIABLE	Revenue		
	(A) BOX OFFICE REVENUE (ALL)	(B) BOX OFFICE REVENUE (A-list)	(D) BOX OFFICE REVENUE (low-budget)
DOUBAN_VOTE	0.030*** [0.005]	0.022* [0.009]	0.039*** [0.006]
DOUBAN_SCORE	232.632 [242.805]	169.139 [651.132]	-75.561 [246.265]

xxx#: the variables are algebraically transformed to the log format; *: 5% significance level; **: 1% significance level; ***: 0.1% significance level; [x]: standard error

Source: author

All coefficients for valence attributes are nonsignificant. As revealed by Liu (2006), Duan et al. (2008), Lipizzi et al. (2016), and Castillo (2021), volume, rather than valence, contributes to box office performance in the Chinese film market (Table 6). Because social media volume refers to the popularity of a film, high volume indicates high

popularity and more ticket sales. Regarding valence, most movies are rated at median values (Table 7), which means that moviegoers must make extra effort to decipher the meaning of ratings. The Maoyan and Douban valence attributes thus do not directly signify the audience’s sentiment.

Table 7
Ratings

Score	Douban	Maoyan
0–1	55	32
1–2	0	0
2–3	88	11
3–4	148	33
4–5	164	55
5–6	176	120
6–7	182	146
7–8	114	243
8–9	30	259
9–10	0	58
Total	957	957

Source: author

Both the Maoyan and Douban volume attributes have significant power for predicting box office revenue (Table 6). Table 8 presents the results calculated using (F3). A-list films’ gross box office revenue increases by CN¥68.6–81.9 for each additional user on Maoyan/Douban that rates the film. Similarly, the gross revenue of the low-budget films increases by CN¥98.1–111.3 for one extra user that rates the movies (Table 8). These results suggest that social media volume affects the ticket sales of the low-budget films more than do that of the A-list films. Therefore, low-budget movies are more dependent on social media to ensure box office success than are A-list films.

Table 8
Effects of One-User Increase on Box Office Revenue

	γ value	θ value	Avg. Maoyan vol.	Avg. Douban vol.	Net revenue growth
Effect of Maoyan ratings on A-list films	0.078	$1.69e^{-8}$	115,676		CN¥81.9
Effect of Douban ratings on A-list films	$-2.892e^{-7}$	$3.533e^{-7}$		97,134	CN¥68.6

Effect of Maoyan ratings on low-budget films	$-8.179e^{-7}$	$6.775e^{-7}$	72,366		CN¥98.1
Effect of Douban ratings on low-budget films	0.111	$4.16e^{-9}$		36,309	CN¥111.3

Source: author

Coefficients for *screen_no* and *sequel* are positive and significant at the 0.1% and 1% levels, respectively (Table 6). An A-list film’s box office revenue increases by CN¥380–390 for each additional screening. Similarly, each additional screening can increase a low-budget film’s box office revenue by CN¥120–160. If an A-list film is a sequel, its box office revenue will increase by CN¥59,157,690–63,570,630. A low-budget film’s box office revenue will increase by CN¥36,288,510–48,293,510 if it is a sequel.

The coefficients for *nation*, *season*, and *budget* against A-list films’ box office revenue are also statistically significant. An A-list film’s box office revenue will increase by CN¥26,768,840–27,496,280 if is a Chinese film, by CN¥483,380–515,150 if it is released during the holiday season, and by CN¥8,210–8,560 for every US\$10,000-increase in its budget. The coefficients for genre are nonsignificant. These results indicate that genre does not affect a movie’s box office revenue when other factors are controlled for.

5.3 Effects of eWOM Due to Social Media Attributes

Figure 7 presents the normalized daily change in A-list and low-budget films’ box office revenue. A-list films’ ticket sales peak during the opening weekend and decrease steadily thereafter. By contrast, the low-budget films experience a second, smaller wave during the second weekend. We performed polynomial regression to examine whether eWOM contributes to box office sales after the opening weekend.

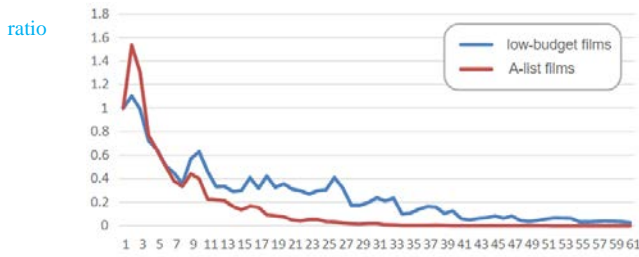


Figure 7

Box Office Revenue (%) by Day After Release

Source: author

5.3.1 Effects of eWOM on A-list Films

Columns F and G in Table 9 present the results of running the volume and valence attributes against A-list films' opening-weekend revenue in relation to gross revenue (*Box_office_p*). The *Douban_vote* and *Douban_user* coefficients are negative and significant at the 1% level. A 1% increase in Douban users rating or commenting on an A-list film causes a 0.06/0.04% decrease in opening-weekend revenue's contribution to gross revenue for the film. The results indicate that because of lagged diffusion, the effects of eWOM become apparent after the opening weekend. Douban valence attributes do not significantly predict opening-weekend revenue's contribution to gross revenue for the A-list films. The results suggest that Douban volume, rather than valence strengthens the effects of eWOM on post-opening weekend revenue's contribution to gross revenue for A-list films.

Table 9

Effects of eWOM Due to Social Media Attributes³

Variable	(F) % OF OPENING- WEEKEND REVENUE (A-list)	(H) % OF OPENING- WEEKEND REVENUE (low-budget)
DOUBAN_VOTE#	-0.058** [0.019]	0.017 [0.011]
DOUBAN_SCORE	-0.001 [0.012]	-0.024*** [0.007]
MAOYAN_VOTE#	0.077*** [0.010]	0.044*** [0.008]
MAOYAN_SCORE	-0.040*** [0.011]	0.022*** [0.005]
NATION	0.037* [0.017]	0.043 [0.024]

³ Because *Box_office_p* is presented as percentages, we algebraically transformed the volume attributes,

SEASON	-0.060 [0.031]	-0.019 [0.023]
SCREEN_NO#	-0.061*** [0.008]	-0.060*** [0.005]
SEQUEL	0.033 [0.033]	-0.048 [0.028]
BUDGET#	-0.002 [0.007]	omitted
GENRE		
DRAMA	0.003 [0.022]	-0.042* [0.018]
ROMANCE	0.001 [0.027]	0.025 [0.017]
SCI-FI	0.007 [0.028]	0.042 [0.060]
KUNG-FU	-0.036 [0.089]	0.000 [0.071]
HISOTRY	-0.015 [0.063]	0.051 [0.059]
HORROR/ THRILLER	-0.069* [0.027]	-0.049* [0.024]
COMEDY	0.014 [0.024]	0.020 [0.016]
CHILDEN'S	-0.070 [0.149]	0.006 [0.040]
CARTOON	-0.023 [0.034]	0.047 [0.030]
ADVENTURE	-0.025 [0.024]	-0.044 [0.024]
SAMPLE SIZE	248	709
F-VALUE	9.29***	11.82***
R-SQUARED VALUE	0.3894	0.2157
NATION DUMMY	v	v
YEAR DUMMY	v	v
SEASON DUMMY	v	v

xxx#: the variables are algebraically transformed to the log format; *: 5% significance level; **: 1% significance level; ***: 0.1% significance level; [x]: standard error

Source: author

In Columns F and G in Table 9, the estimated coefficients for *Maoyan_vote* are approximately 0.08 and are significant at the 0.1% level. A 1% increase in Maoyan volume entails a 0.08% increase in opening-weekend revenue's contribution to gross revenue for the A-list films. Because Maoyan is a ticket presale website and few people can watch a film before its release, the increase in the number of Maoyan users upon a film's release is likely due to broadcasting. Maoyan volume before a film's release thus indicates a

screen_no, and *budget* variables into log form to eliminate bias.

movie's popularity and effectively predicts opening-weekend box office sales.

The estimated coefficients for *Maoyan_score* are approximately -0.04 to -0.03 and significant at the 0.1% level, indicating the power to predict opening-weekend revenue's contribution to gross revenue for A-list films. The results indicate that for every point increase in the films' Maoyan rating, post-opening weekend revenue's contribution to gross revenue increases by 0.03/0.04% for the A-list films. The rating for an A-list film before its release is the result of mass marketing. However, it may reflect public opinion if its value fluctuates considerably after the film's release. Fluctuation in Maoyan valence may be due to moviegoers rating a film online and diffusing information through eWOM.

5.3.2 Effects of eWOM on Low-Budget Movies

Columns H and I in Table 9 present different results for the low-budget movies. Douban valence, rather than volume, produces negative (-0.01 to -0.02) and significant estimates of opening-weekend revenue's contribution to gross revenue for the low-budget films. Post-opening weekend revenue's contribution to gross revenue for the low-budget films increases by 0.01–0.02% for each unit increase in Douban rating or p/n value. Because Douban users appreciate high-brow culture, they usually prefer non-Hollywood films. Their ratings and comments thus serve as a guide for moviegoers seeking non-Hollywood films. When the Douban rating for a low-budget film increases through eWOM, an individual's willingness to watch the film may increase, which would increase post-opening weekend revenue's contribution to gross revenue. Audiences may not watch a film if its rating decreases. On Douban, eWOM results from volume for the A-list films and valence for the low-budget ones.

The results in Columns H and I (Table 9) indicate that the coefficient estimates for *Maoyan_vote* are approximately 0.04–0.05 and significant at the 0.1% level. Opening-weekend revenue's contribution to gross revenue for the low-budget films increases by 0.04% for each percent increase in Maoyan volume. The estimated coefficient for *Maoyan_score* is approximately

0.02 and significant at the 0.1% level. Each 1-point increase in Maoyan rating increases opening-weekend revenue's contribution to gross revenue for the low-budget films by 0.02%. Maoyan volume and valence are positively correlated with opening-weekend revenue's contribution to gross revenue for the low-budget films; this finding indicates that Maoyan does not facilitate eWOM but facilitates broadcasting for low-budget films.

The coefficient estimates of *screen_no#* for both the A-list and low-budget films are -0.06 and significant at the 0.1% level. Each 1% increase in screen number increases post-opening weekend revenue's contribution to gross revenue by 0.06%. The coefficient for the country of origin for the A-list films is significant at the 5% level. The box office sales of Chinese A-list films after the opening weekend are higher than those of Hollywood films because Chinese films benefit from eWOM.

The coefficient for the horror and thriller genre is significant for the A-list films. Because the success of such films depends on their special effects, their quality is difficult to ascertain on the basis of information on websites alone. Therefore, horror and thriller films with well-executed special effects can generate strong eWOM, with increased post-opening weekend revenue's contribution to gross revenue. Drama films do not rely on large production budgets but on the quality of screenplays to succeed. For these films, strong eWOM indicates their popularity; therefore, their post-opening weekend revenue's contribution to gross revenue would be increased.

5.3.3 Comparison of Volume and Valence Attributes

Volume affects a film's gross, opening-weekend, and post-opening weekend revenue (Table 10). High volume means high popularity and strong box office performance. A movie can become a blockbuster if it achieves high volume, irrespective of the valence. A movie may flop if it has low volume, even with positive valence.

Table 10

Comparison of Maoyan and Douban

BOX OFFICE REVENUE	MAOYAN	DOUBAN
--------------------	--------	--------

A-LIST FILMS	Volume (positive)	Volume (positive)
LOW-BUDGET FILMS	Volume (positive)	Volume (positive)
% CHANGE IN OPENING-WEEKEND REVENUE'S CONTRIBUTION TO GROSS REVENUE	MAOYAN	DOUBAN
A-LIST FILMS	Volume (positive) Valence (negative) †	Valence †
LOW-BUDGET FILMS	Volume (positive) Valence (positive)	Valence (negative) †

†: negative effects attributed to eWOM

Source: author

Table 10 provides a summary of the marketing modes of Douban and Maoyan. Maoyan mostly affects movies' box office performance through broadcasting. High Maoyan volume constitutes broadcasting for both the A-list and low-budget films, which guarantees opening-weekend ticket sales. Douban has an artistic community, and evaluations of films by critics on the platform are spread through eWOM. High Douban volume increases post-opening weekend revenue's contribution to gross revenue for the A-list films through eWOM.

Although valence does not affect gross revenue, it affects how revenue is distributed during and after the opening weekend. High Maoyan ratings for the A-list films constitute eWOM because they increase revenue after the opening weekend. The high valence increases low-budget films' opening-weekend box office sales through broadcasting. High Douban valence increases post-opening weekend revenue's contribution to gross revenue for the low-budget films through eWOM (Table 10).

6. Explanatory Power of Social Media Attributes

6.1 RF Model

An RF model consists of several decision trees that provide regression functions, and the final output of RF regression is the average of the outputs of all decision trees. It is an ensemble method that combines all generated decision trees by using a bagging (or bootstrap aggregation) algorithm.⁴ Each bootstrap subset is obtained by randomly selecting n observations from S_n with replacement, where each observation is selected with probability $1/n$. Let $X = \{x_1, x_2, \dots, x_m\}$ represent the input vector containing m features and Y represent the output value. The training set S_n containing n observation can be expressed as follows:

$$S_n = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}, X \in \mathbb{R}^m, Y \in \mathbb{R} \quad (F4)$$

We used an n -fold method to divide the training samples from the test ones before the bootstrap process and to make the experimental results more robust (Figure 8).

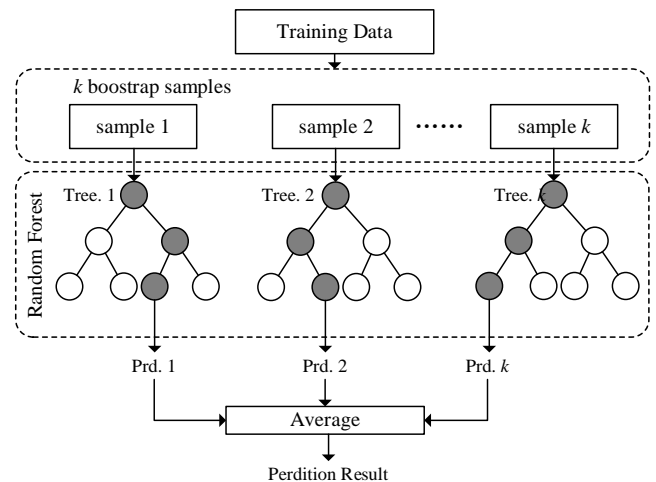


Figure 8

Training of RF Model

Source: author

⁴ Proposed by Breiman (2001), bagging is a technique used with numerous regression methods to reduce the

variance associated with prediction, thereby improving prediction performance.

For the training of each decision tree, a new training set S_n (bootstrap samples) is generated with replacement from the original data. Only n randomly selected samples are used for training, and unselected samples are included as a part of another subset and called out-of-bag (OOB) samples. When a regression tree is constructed with the training samples, the OOB samples are used to evaluate the performance of the regression tree. This built-in cross-validation process enables RF to compute an unbiased estimation of the generalization error without an external data subset (Figure 8). It also increases generalizability and prevents overfitting in RF. To obtain the total learning error, an average of the prediction error of each tree using the OOB sample can be obtained as follows:

$$MSE \approx MSE^{OOB} = n^{-1} \sum_{i=1}^n [\hat{Y}(X_i) - Y_i]^2 \quad (F5)$$

where $\hat{Y}(X_i)$ represents the predicted value of the input sample X_i , Y_i represents the real values, and n represents the total number of OOB samples. The OOB error is an unbiased estimate of the generalization error given that RF produces a limiting value of the generation error.

After training, the prediction function $\hat{h}(X, \Theta_k)$ is obtained, which is constructed over S_n . RF regression involves using the K tree-structured base classifier $h(X, \Theta_k)$, where $K = \{1, \dots, k\}$ and Θ_k represents the parameter and the structure of the k th decision tree. The bagging algorithm determines a set of bootstrap samples $(S_n^{\theta_1}, \dots, S_n^{\theta_k})$ and applies the previous tree decision algorithm to these samples to construct the collection of k prediction trees $\hat{h}(X, S_n^{\theta_1}), \dots, \hat{h}(X, S_n^{\theta_k})$. The ensemble produces k outputs corresponding to each

tree, $\hat{Y}_1 = \hat{h}(X, S_n^{\theta_1})$, $\hat{Y}_2 = \hat{h}(X, S_n^{\theta_2})$, ..., and $\hat{Y}_q = \hat{h}(X, S_n^{\theta_k})$. Aggregation is then performed by averaging the outputs of all trees. Thus, \hat{Y} can be obtained as follows:

$$\hat{Y} = \frac{1}{k} \sum_{l=1}^k \hat{Y}_l = \frac{1}{k} \sum_{l=1}^k \hat{h}(S_n^{\theta_l}) \quad (F6)$$

where \hat{Y}_l represents the output of l th tree.

We measured the importance of volume and valence after training. Zhu et al. (2015) indicated that the importance score for the j th feature can be computed by averaging the difference in the OOB error before and after the permutation over all trees. The features producing large values⁵ for this score are ranked as more important than those that produce small values.

6.2 Importance of Maoyan Volume

We only included the statistically significant variables in Tables 6 and 9 in the RF tests of *Box_office* (Figure 9). The variable *screen_no* generates the estimated value of 0.637, *Maoyan_vote* generates the value of 0.231, *Douban_vote* generates the value of 0.047, *budget* generates the value of 0.041, *Maoyan_score* generates the value of 0.031, and *Douban_score* generates the value of 0.013. The estimated value of Maoyan volume is the greatest among the social media attributes, with a weight of 23% in predicting box office performance. Maoyan volume thus has the strongest predictive power among the attributes. Figure 9 also indicates that the volume attributes are more important ($0.271 = 0.230 + 0.041$) than the valence ones ($0.044 = 0.013 + 0.031$) for determining box office performance. The RF results are consistent with those of the polynomial regression.

⁵ The score was normalized using the standard

deviations of these differences.

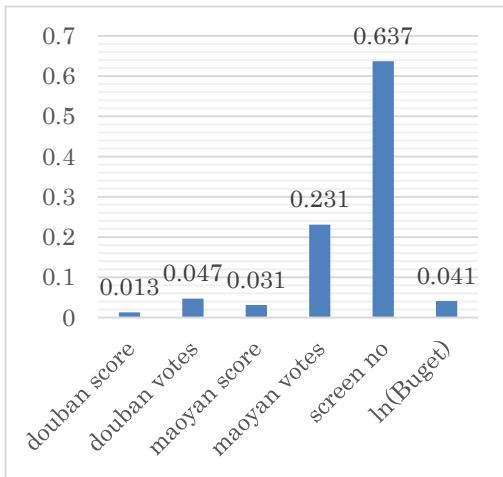
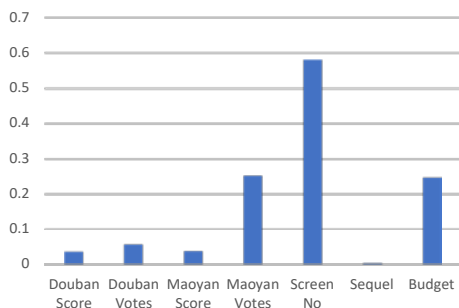


Figure 9

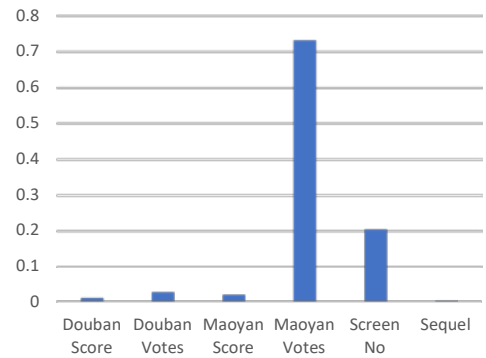
Importance of Social Media Attributes

Source: author

Figure 10 provides a comparison of the RF results for the A-list and low-budget films. *screen_no* has the strongest predictive power for the A-list films, whereas *Maoyan_vote* has the strongest predictive power for the low-budget films (Figure 10). The results indicate that A-list films' box office revenue can increase more when they are screened in a large number of theaters. Low-budget films considerably rely on social media diffusion, especially user volume, to succeed. We also observed that the estimated value of Maoyan volume + valence is greater than that of Douban. Maoyan is thus more capable of increasing movies' gross revenue than Douban.



(a) A-list films



(b) low-budget films

Figure 10

Importance of Social Media Attributes for A-List and Low-Budget Movies

Source: author

6.3 Marketing Modes for Maoyan and Douban

Our empirical results indicate that Maoyan and Douban affect movies' box office performance at different stages and through different digital marketing modes, even if they both increase gross revenue. Maoyan mostly broadcasts information regarding films prior to their release (Figure 11), whereas Douban uses eWOM after a film's release. Because Maoyan is a ticket presale website, its volume indicates the perceived popularity resulting from distributors' marketing before release. Douban has a reputation for being a social space for cultural elites. Douban volume signifies these elites' experience, which can guide audiences. Our results suggest that film distributors should allocate resources to increasing Maoyan volume before release because this is the most important attribute. Distributors can also shift focus to Douban volume after release. To increase ticket sales after the opening weekend, distributors should screen films in more theaters when Douban volume increases.

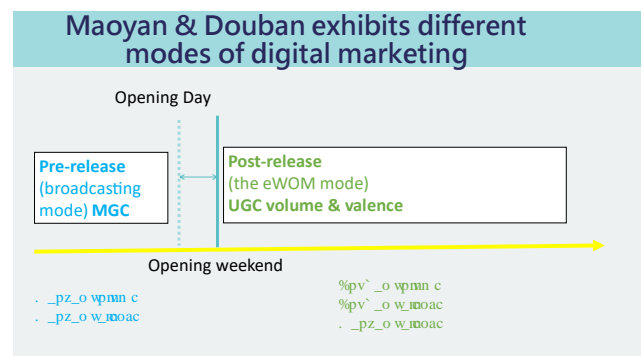


Figure 11

Timeline of Maoyan and Douban Marketing

Source: author

For the A-list films, attention should be directed to Maoyan valence because it increases post-opening weekend revenue's contribution to gross revenue. For the low-budget movies, film distributors should attempt to increase Maoyan ratings before release to increase ticket sales during the opening weekend. Distributors should attempt to increase Douban valence after a film's release because high ratings and p/n values increase box office sales after opening weekends.

Our empirical results partially support H1; that is, only volume affects gross revenue. Douban and Maoyan volume have stronger effects on the low-budget films than on the A-list films. The results also support H2; that is, strong eWOM decreases opening-weekend revenue's contribution to gross revenue. We did not reject the null hypothesis (H3); that is, low-budget films are more influenced by eWOM than are A-list films. Both types of movies are affected by eWOM, but the effects of eWOM on each type of film cannot be compared because they are due to different attributes.

7. Conclusion: Blockbuster or Flop?

Because the Chinese film market has become the second largest market worldwide and the expenditure on digital advertising is more 50% of budgets, investigating Chinese blockbusters and the effects of social media on box office revenue is crucial for film distributors. Our empirical results indicate that both the Douban and Maoyan platforms affect gross revenue through volume, Maoyan increases films' gross revenue more than does Douban, Maoyan affects films' revenue mostly through broadcasting, Douban affects revenue mostly through eWOM, and low-budget films depend more on social media to succeed than do A-list films.

The results can inform marketing strategies for the Chinese film industry. For instance, film distributors should attempt to increase Maoyan

volume prior to the film release and address Douban volume after release. Distributors can increase the number of theaters in which their films are screened to increase ticket sales when Douban volume increases. For A-list films, attention should be directed to Maoyan ratings after release. For low-budget movies, Maoyan ratings should be analyzed before release, and Douban valence should be analyzed after release.

This study contributes to the literature regarding the effects of social media on box office revenue. First, it determined the effectiveness of Douban and Maoyan in improving box office performance in China, which supplements the results of studies on Weibo. Second, it demonstrated that social media volume, rather than valence, significantly affects movies' box office performance. Third, this study demonstrated that Maoyan entails broadcasting, and that Douban entails eWOM. Finally, this study revealed that low-budget movies depend more on social media to succeed than do A-list ones.

The results may only be relevant to 2019; the COVID-19 pandemic affected box office performance. Further research on the post-COVID era should be conducted to examine the effects of social media during different periods. Studies can also examine platforms such as Wechat, TikTok, and Taobao to compare their effectiveness and validate the results of this study.

References

1. Abidi, S. M. R., Xu, Y., Ni, J., Wang, X. & Zhang, W. (2020). Popularity prediction of movies: from statistical modeling to machine learning techniques. *Multimedia Tools and Applications*, 79: 35583–35617.

2. Ahmad, I., Bakar, A., Yaakub (2020). Movie revenue prediction based on purchase intention mining using YouTube trailer reviews. *Information Processing & Management*, 57(5): 102278.
3. Apala, K. R., Jose, M., Motnam, S., Chan, C.-C., Liszka, K. J., & de Gregorio, F. (2013). Prediction of movies box office performance using social media. Paper delivered at the 2013 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.
4. Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. Paper delivered at 2010 *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, <https://doi.org/10.1109/WI-IAT.2010.63>.
5. Awwalu J, B. & Yaakub, MR. (2019) [Hybrid N-gram model using Naïve Bayes for classification of political sentiments on Twitter](#). *Neural Computing and Applications*, 31(12): 9207–9220.
6. Baek, H., S. Oh, S., Yang, H., & Ahn, J. (2017). Electronic word-of-mouth, box office revenue and social media. *Electronic Commerce Research and Applications*, 22 :13-23.
7. Baidu Encyclopedia (2012). The Douban Platform, accessible via http://baike.baidu.com/item/□ □ □ ?fromModule=lemma_search-box#1 on Jun. 30th, 2022.
8. BBC (2015) [EU accuses Google Shopping of search abuse](#). *BBC*. 2015.4.15 website article.
9. Bhattacharjee, B., Sridhar A. & Dutta, A. (2017) Identifying the causal relationship between social media content of a Bollywood movie and its box-office success—a text mining approach. *International Journal of Business Information Systems*, 24(3): 344-368.
10. Boer, R. (2021) *Socialism with Chinese characteristics: A guide for foreigners*. N.Y.: Springer.
11. Bogaert, M., Ballings, M., den Poel, D. A., & Oztekin, A. (2021) Box office sales and social media: A cross-platform comparison of predictive ability and mechanisms. *Decision Support Systems*, <https://doi.org/10.1016/j.dss.2021.11>
12. Breiman, L. (2001). Random Forests. *Machine Learning*, 45: 5–32. <https://doi.org/10.1023/A:1010933404324>
13. Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: conceptualizing the online social network. *Journal of Interactive Marketing*, 21: 2–20. doi: 10.1002/dir.20082.
14. Cao, Y. (2020). Regulating digital platforms in China: Current practice and future developments. *Journal of European Competition Law & Practice*, 11(3-4): 173-180.
15. Castillo, A., Benitez, J., Llorens, J., & Luo, X. (2021). Social media-driven customer engagement and movie performance: Theory and empirical evidence. *Decision Support Systems*, 145, 113516. <https://doi.org/10.1016/j.dss.2021.113516>

16. Chen, J., Kou, G., Peng, Y., Chao, X., Xiao, F. & Alssadi, F. (2020). Effects of marketing messages and consumer engagement on economic performance: evidence from Weibo. *Internet Research*, 30(5): 1565-1581.
17. Chintagunta, P.K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science* 29: 944–957.
18. Dellarocas, C., Chang, X., & Awad, N. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4): 23-25.
19. Deng, T. (2020). Investigating the effects of textual reviews from consumers and critics on movie sales. *Online Information Review*, 44(6): 1245-1265.
20. Ding, C., Cheng, H., Duan, Y., & Jin, Y. (2017). The power of the “like” button: The impact of social media on box office. *Decision Support Systems*, 94:77-84.
21. Duan, W., Gu, B. & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales--An empirical investigation of the movie industry. *Journal of Retailing*, 84(2): 233-242.
22. Fan, J. (1996). Local Polynomial Modelling and its applications: From linear regression to nonlinear regression. *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC. ISBN 978-0-412-98321-4.
23. Follows, S. (2019). [How are movie advertising budgets spent?](#) Accessible via <https://stephenfollows.com/how-are-movie-advertising-budgets-spent/> on Jun. 23rd, 2021
24. Follows, S. (2016). [How movies make money” \\$100m+ Hollywood blockbusters?](#) Accessible via <https://stephenfollows.com/how-movies-make-money-hollywood-blockbusters/> on Jun. 23rd, 2021
25. [Freedman, D.](#) A. (2009). *Statistical Models: Theory and Practice*. Boston: [Cambridge University Press](#).
26. [Godes, D & Mayzlin, D.](#) (2009) Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science*, 28(4): 721-739.
27. Gong, S., Zhang, J., Zhao, P., Jiang, X. (2017). Tweeting as a marketing tool: A field experiment in the TV Industry. *Journal of Marketing Research*, 54: 833–850.
28. Guttman, A. (2022). Advertising industry worldwide - statistics & facts Accessible via https://www.statista.com/topics/990/global-advertising-market/#topicHeader__wrapper, retrieved on Aug. 4th, 2022
29. Huang, Y-T, & Pai, P-F. (2020). Using the least squares support vector regression to forecast movie sales with data from Twitter and movie

databases. *Symmetry*, 12(4): 625.

<https://doi.org/10.3390/sym12040625>

30. Huete-Alcocer, N. (2017) A literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. *Frontier in Psychology*, 8: 1-4.

31. Kang, D. (2021). Box-office forecasting in Korea using search trend data: a modified generalized Bass diffusion model. *Electronic Commerce Research*, 21(1): 41-72.

32. Karniouchina, E. V. (2011). Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, 28(1): 62-74.

33. Kim, S., Park, N., & Park, S. (2013). Exploring the effects of online word of mouth and expert reviews on theatrical movies' box office success. *Journal of Media Economics*, 114 (2): 26–98.

34. Kim, T., Hong, J., & Kang, P. (2015). Box office forecasting using machine learning algorithms based on SNS data. *International Journal of Forecasting*, 31: 364–390.

35. Kim, Y., Kang, M., & Jeong, S. R. (2018). Text mining and sentiment for predicting box office success. *KSII Transactions on Information Systems*, 12(8): 4090-4101

36. Lee, S., Bikash, K. C., & Choeh, J. Y. (2020) Comparing performance of ensemble methods in predicting movie box office revenue. *Heliyon*, 6,

<https://doi.org/10.1016/j.heliyon.2020.e04260>.

37. Li, X. & Hitt, L. M. (2008) Self-selection and information role of online product reviews. *Information Systems Research*, 19: 456–474.

38. Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2016). Combining structure, content and meaning in online social networks: The analysis of public's early reaction in social media to newly launched movies. *Technological Forecasting and Social Change*, 109: 35–49.

39. Liu, T., Ding, X., Chen, Y., Chen, H., & Guo, M. (2014). Predicting movie Box-office revenues by exploiting large-scale social media content. *Multimedia Tools and Applications*, 75(3): 1509-1528.

40. Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70: 74–89.

41. Ma, Y., & Wan, Y. (2021). Data analysis method of intelligent analysis platform for big data of film and television. *Complexity*, 2021.

42. Maoyan (2022). Maoyan Entertainment Annual Financial Report, 2021.

43. Maoyan (2019). Maoyan Entertainment Annual Financial Report, 2018.

44. [Navarro, J. G. \(2022\)](https://www.statista.com/statistics/243180/). Leading box office markets worldwide 2021, by revenue. Accessible via <https://www.statista.com/statistics/243180/>

[leading-box-office-markets-workwide-by-revenue/](#), retrieved on Aug. 4th, 2022.

45. Netease (2012). Douban claims over 100 million monthly active users, close to profit making. Accessible via <https://www.163.com/tech/article/893L4F6N000915BF.html>, retrieved on Aug. 31st, 2022.

46. Nieto, J., Hernández-Maestro, R. M., and Muñoz-Gallego, P. A. (2014). Marketing decisions, customer reviews, and business performance: the use of the Toprural website by Spanish rural lodging establishments. *Tourism Management*, 45, 115–123. doi: 10.1016/j.tourman.2014.03.009

47. Niraj, R. & Singh, J. (2015). Impact of user-generated and professional critics reviews on Bollywood movie success. *Australasian Marketing Journal*, 23(3): 179-187.

48. Oh, C., Roumani, Y., Nwankpa, J., & Hu, H. (2017). Beyond likes and tweets: consumer engagement behavior and movie box office in social media. *Information Management*, 54(1): 25–37.

49. [Oh, S.](#), [Baek, H.](#) and [Ahn, J.](#) (2017). Predictive value of video-sharing behavior: sharing of movie trailers and box-office revenue. *Internet Research*, 27(3): 691 - 708. <https://doi.org/10.1108/IntR-01-2016-0005>

50. Pandya, B., Kala, V., Al-anoud, & Lobo, V. (2018). Sentimental Analysis of Twitter Data to Predict Performance of Movies. *JOSR*

51. Qiu, X. & Tang, T. Y. (2018) Microblog mood predicts the box office

performance. *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*, 129-133.

52. Rainey, J. (2016) The peril of promotion: Pricy TV campaigns, fear of change shackles movie spending, accessible via <https://variety.com/2016/film/features/movie-marketing-advertising-tv-campaigns-1201724468/> on Aug. 3rd, 2021.

53. Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? The effect of tweets on movie sales. *Decision Support Systems*, 55: 863–870.

54. Ruus, R., & Sharma, R. (2020). Predicting movies' box office result – A large scale study across Hollywood and Bollywood. *Complex Networks 2019*, H. Cherifi et al. (eds.), 982-994.

55. [Shen, D. \(2020\). Movie box office prediction via joint actor representation and social media sentiment. Paper delivered at the 34th Conference on Neural Information Processing Systems.](#)

56. Sun, T. (2013). The movie box office market maintained a high growth rate, triggering O2O competition for film theaters. Accessible via <http://zhuanlan.zhihu.com/p/269167775>, retrieved on Aug. 31st, 2022.

57. Song, T., Huang, J., Tan, Y. & Yu, Y. (2019). Using user- and marketer-generated content for box office revenue prediction: differences between microblogging and third-party platforms. *Information Systems Research*, 30(1): 191-203.

58. Statista Research Department (SRD). (2021). Digital ad spend growth 2016-2022, by format. <https://www.statista.com/statistics/260117/growth-of-online-advertising-spending-by-type/>, retrieved on Aug. 4th, 2022.
59. Sweeney, E. (2018). Digital represents 14% of movie ad budgets but drives 46% of revenue, study finds. in *Marketing Dive*, accessible via <https://www.marketingdive.com/news/digital-represents-14-of-movie-ad-budgets-but-drives-46-of-revenue-study/544253/> on Jun. 30th 2021.
60. Tang, T., Winoto, P., Guan, A., & Chen, G. (2018). The “foreign language effect” and movie recommendation: A comparative study of sentiment analysis of movie reviews in Chinese and English. *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, 79-84.
61. Thomala, T. L. (2022). Film industry in China - statistics & facts. accessible via <https://www.statista.com/statistics/1046171/china-box-office-share-in-the-global-box-office/>, retrieved on Aug. 1st, 2022
62. Turner, B. (2015a) How are movie release dates chosen? in *HowStuffWorks*, accessible via [How Are Movie Release Dates Chosen? | HowStuffWorks](#) on Aug. 5th, 2021
63. Turner, B. (2015b) How to they predict the weekend box office before Sunday’s numbers are in? in *HowStuffWorks*, accessible via [How do they predict the weekend box office before Sunday’s numbers are in? | HowStuffWorks](#) on Aug. 5th, 2021
64. Zhang, C., Tian, Y.-X., & Fan, Z.-P. (2022). Forecasting the box offices of movies coming soon using social media analysis: A method based on improved Bass models. *Expert Systems with Applications*, 191: 116241. <https://doi.org/10.1016/j.eswa.2021.116241>.
65. Zhu, R., Zeng, D., & Kosorok, MR. (2015). Reinforcement learning trees. *Journal of American Statistical Association*, 110 (512): 1770-1784. doi: 10.1080/01621459.2015.103699