The Correlation Between Social Media and Consumer Behaviour: A Research on Movie Industry in China

Enjun Long
Hwa Chong International School, 269783 Bukit Timah, Singapore

Keywords: Social Media, Movie Box Office, Consumer Behavior, Movie Industry in China

Abstract: The large domestic coverage of social media makes it significantly influential from a wide perspective. In the recent decade, social media's role has become increasingly important in the commercial world. Over time, social media platforms have transformed into persuasive tools that significantly influence how consumers explore, evaluate, and purchase products. This paper embarks on an exploration of the intricate interplay between social media and the movie box office, employing quantitative research methodologies to examine the impact of social media on the movie box office. In this study, the impact of social media on movie sales performance will be analyzed from two perspectives: The correlation between social media (as alternative evaluation sources) and the total movie box office; and the correlation between social media's engagement indexes (reflected by different social media indexes) and movie box office daily. A significant positive relationship has been found in both two perspectives. The research provides new findings for marketers in the movie industry to better understand the factors influencing movie sales performance.

1 INTRODUCTION

1.1 The Contemporary Social Media Usage and Influence

According to the DataReportal, at the beginning of 2023, there were 1.05 billion internet users in China, accounting for 73.7% of the total domestic population. Among the internet users, 98.1% of them were social media users (Kemp 2023). In addition, a growing trend of social media usage was foreseen in the next few years. The large domestic coverage of social media makes it significantly influential from a wide perspective. In the recent decade, social media's role has become increasingly important in the commercial world. Unlike the traditional way of one-way advertising, two-way communication on social media allows consumers to have peer interactions (Shareef et al. 2019). Social media functions such as like, comment, share, and subscribe enable consumers a much wider communication and stronger connection with each other. Early in 1999, it was proven that an informational society could effectively affect product evaluation in the buying process (Kozinets 1999). Over the years, social media has been advocated to be effective in influencing consumer's perceptions and behavior (Chopra et al. 2021, Nash 2019). With the expansion of social media networks and the maturity of individual opinion dissemination, social media marketing has been taken as a strategic approach by merchants to achieve marketing objectives and boost sales performance (Wibowo et al. 2020).

1.2 The Movie Industry and Social Media

As one of the most popular entertainment options, movies have become a great portion of the media products consumed by people (Kubrak 2020). It provides a two-hour stress-free time and environment for audiences to sit back, and relax, which benefits one's mental and emotional health (Molaie et al. 2010). According to China Movie Database, by 13 November 2023, the movie industry sales performance has topped 50 billion yuan for 2023 (China Film Data Information Network 2023). The industry witnessed a year-on-year growth rate of 52.88% by the first half of 2023, demonstrating a good recovery from the pandemic impact and a vibrant market demand (Chen 2023).

Marketing, specifically social media marketing, has been playing an important role in driving movie sales. Industry studies proved that movie sales performance is significantly affected by the Internet Word of Mouth (IWOM) (Zhu & Ma 2021). Asur and

Long, E.
The Correlation Between Social Media and Consumer Behaviour: A Research on Movie Industry in China.
DOI: 10.5220/0012820000004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 260-266
ISBN: 978-989-758-690-3

Huberman also showed evidence that box-office revenues for movies could be predicted with social media via a specific analysis of Twitter comments (Asur & Huberman 2021). Similarly, the movie box office was proven to be predictive based on the Facebook comments from the research done by Matthias, Michel, Drik, and Asil (Bogaert et al. 2021). Nowadays, as the industry scale grows in China, it becomes a burgeoning interest to unlock the relationship between social media and movie sales performance from different perspectives. However, relevant research in the Chinese market is comparatively little.

1.3 Research Background

Inspired by the increasing impact of social media on consumer purchase decisions and realizing most of the previous research was done based in Europe and America, or on a single social media platform, this paper aims to study the relationship between social media and consumer behavior in the movie industry in China. The consumer behavior model adopted in this paper is the Engel-Blackwell-Kollat Model, of which consumer behavior could be explained in five steps: 1) problem recognition; 2) information search; 3) evaluation of alternatives; 4) purchase; 5) post-purchase evaluation (Engel et al. 1995). Two main perspectives have been explored through the following angles:

The correlation between social media (as alternative evaluation sources) and the total movie box office.

The correlation between social media engagement (reflected by different social media indexes) and movie box office daily.

Through studying the two perspectives, this paper enhances the understanding of the relationship between social media and movie sales performance at the same time brings new insights for the marketer in the Chinese movie industry.

2 METHODOLOGY AND RESEARCH DESIGN

2.1 The Correlation between Social Media (as Alternative Evaluation Sources) and the Total Movie Box Office

From the Engel-Blackwell-Kollat model, it is understood that the purchase decision is affected by the evaluation of alternatives (Nash 2019). Likewise, the movie box office would be affected by the IWOM, peer reviews, the level of popularity, and favourability on social media. This is because the movie, as an experienced product, would also be through the step of evaluation of alternatives in the consumer behavior process. To avoid a bad movie experience, consumers tend to take mass reviews such as movie ratings and audience comments on social media as alternative evaluation factors (Chen et al. 2011). To find out the correlation between the several common-used evaluation factors and movie box office and have a direct visualization of the correlation, Pearson Correlation Coefficient heatmaps are applied. Four alternative evaluation sources have been proposed: 1) professional rating, 2) public rating, 3) social media popularity, and 4) social media favourability.

2.1.1 Data Interpretation

Thirteen movies released in the year 2023 have been chosen for the first perspective. To improve the reliability of the research, four principles were applied when selecting sample movies (Table 1). To accommodate those control variables, the movie box office of the following thirteen movies is chosen in this study: 1) *No more bets,* 2) *Never Say Never,* 3) *Dust to Dust,* 4) *Creation of the Gods l: Kingdom of Storms,* 5) *Just for meeting you,* 6) *Lost in the stars,* 7) *One and only,* 8) *Post Truth,* 9) *Papa,* 10) *Ping Pong: The Triumph,* 11) *Manifesto,* 12) *Too Beautiful to Lie, and* 13) *Heart's Motive.*

Table 1. Sample Selection Principles.

Principles	Significance of controlling
General Releasing Date	Movies released in the following three peak seasons have been excluded from the sample choice: the Chinese New Year period, the first week of May (Labor Day Golden Week), and the first week of October (National Day Golden Week). This is to ensure the movie box office is contributed by a reasonable normal traffic flow instead of seasonal traffic flow.
Similar Cast and Crew	To avoid the celebrity effect that potentially boosts the movie box office.
Series Movies Exclusive	To avoid fans or detractors carrying forward from the previous movies, resulting in a potential increase or decrease in the movie box office.
Same Genre of Movie	Similar to the celebrity effect, choosing movies sharing a similar genre is to avoid high movie box sales driven by high-tech

	animation production. In this sample pool, the movies are all real-life shooting with real actors.

Measurement for social media as evaluation sources is as below.

Professional Rating = 0.5 (Weibo Rating + IMDb Rating)

75% of the Weibo rating is contributed by authority media, industry experts, and top influencers in the movie industry, and thus it has been considered a professional rating source. In the meanwhile, Weibo Rating is retrieved from Weibo. cn and IMDb Rating is retrieved from imdb.com.

Public Rating = 1/3 × (Maoyan Rating + Douban Rating + Taopiaopiao Rating)

Three popular public rating benchmarks are adopted. Maoyan Rating is retrieved from the application Maoyan PRO; Douban Rating is retrieved from m.douban.com; Taopiaopiao Rating is retrieved from the application Taopiaopiao.

Social Media Popularity = 0.68 / (0.68 + 0.87) Accumulated Hottest Topics on Weibo + 0.87 / (0.68+0.87) × Accumulated Hottest Topics on Douyin.

Social Media Popularity is estimated by two sources, the Accumulated Hottest Topics and Accumulated Hottest Topics on Douyin. Both data sources are retrieved from the application Maoyan PRO. Proportion is allocated based on the Social Media Usage Report (Ipsos 2023).

Social Media Favorability = 0.68 / (0.68 + 0.87) × (0.2 × number of like on Weibo + 0.3 × number of comments on Weibo + 0.5 × number of Weibo repost) + 0.87 / (0.68 + 0.87) × number of like on Douyin

Social Media Favourability is estimated by two sources: Weibo engagement in 3 perspectives (retrieved from Weibo.cn.), and several like on Douyin (retrieved from the application Douyin). Proportion is allocated based on the Social Media Usage Report (Ipsos 2023).

Movie box office = Total Movie box office of the movie

Data is retrieved from China Movie Database (* The measurements of popularity and social media favourability are based on movies' official accounts on Weibo and Douyin only.)

2.1.2 Result and Observation

The heatmap is outputted based on the Pearson Correlation Coefficient, which indicates the measurement of the linear association strength between the two variables. According to Sedgwick, a small sample needs a correlation coefficient in larger values so that the linear association could be significant, which is closer to − 1 or 1 (Sedgwick 2012). In terms of the heatmap result obtained, the correlation coefficient value is observed to range from 0.2 to 1. A higher value in the result suggests a stronger positive correlation between the two variables, and the cell would be in a lighter color accordingly. The correlation between social media (as an alternative evaluation source) and the total movie box office in Figure 1.

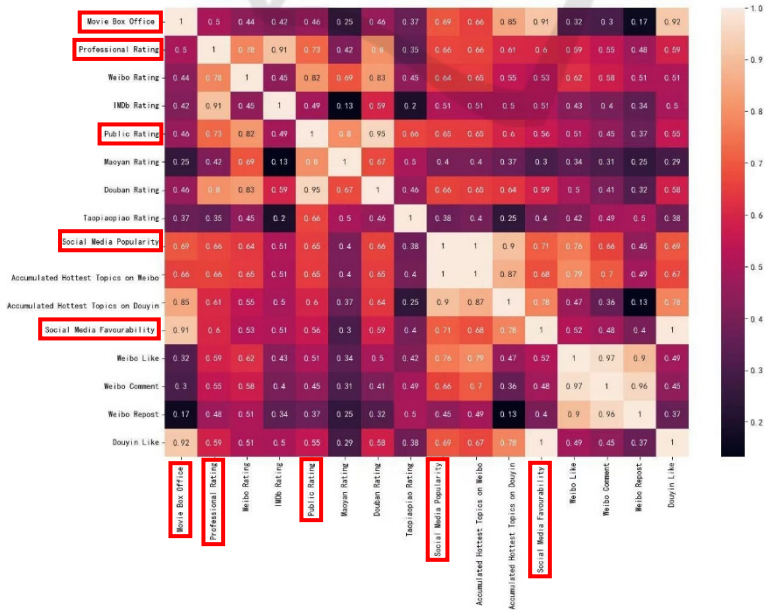

Figure 1: Pearson Correlation Coefficient heatmaps (Picture credit: Original).

Surprisingly, being the most obvious factor when people are looking for reviews on social media, both professional ratings and public ratings don't seem to have a strong correlation with the movie box office. They have only a noticeable relationship of correlation coefficient of 0.5 and 0.46 separately with the movie box office. This indicates when there is an increase in professional rating and public rating there is a tendency for the movie box office to increase while the degree of accuracy would need to be moderated.

Social media popularity has a substantial association with the movie box office than the ratings of 0.69. It indicates that the movie box office is likely to change according to the social media popularity although it is not perfect. The correlation coefficient of social media favourability has the strongest relationship with the movie box office of 0.91 which is close to a perfect correlation. It suggests that when social media popularity increases, the movie box office is very likely to increase similarly in a high degree of confidence. In addition to the main factors being boxed in red, the correlation coefficients are interesting to be found in individual factors. Social media as the factor most related to the movie box office, has a strong positive correlation with the number of likes on Douyin. This implies favourability reflected by Douyin has a strong influence on one's movie consumption. On the other hand, the lowest correlation coefficient of 0.17 is between Weibo repost and the movie box office. All correlation coefficients about Weibo are in the range from 0.17 to 0.66. The influence of Weibo is hence shown to be decreasing.

2.2 The Correlation Between Social Media Engagement (Reflected by Different Social Media Indexes) and Movie Box Office

The data sourse: We Are Social; DataReportal; Meltwater. From the data, the top six social media platforms with the highest share of internet users as of Q3 2023 are WeChat, Douyin, QQ, Baidu Tieba, Xiaohongshu, and Weibo (Fig. 2). Due to the absent of QQ and Xiaohongshu index daily, as well as the temporary maintenance on the Weibo index website, out of the top six social media platforms, only Baidu Index, Douyin Index and Wechat Index have been adopted as independent variables. In addition, Toutiao is a core and significant application of ByteDance China, whose relationship with movie sales is also worth exploring.

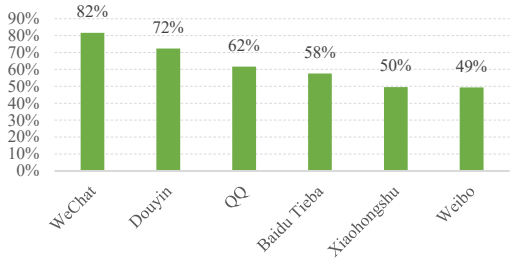

Figure 2: Top 6 social media with the highest share of internet users in China as of Q3 2022 (Original).

To figure out the correlation between social media engagement and movie box office daily, and how the social media platform affects movie sales performance, the Ordinary Least Squares (OLS) Regression model is adopted.

2.2.1 Data Interpretation

Based on the sample in the first prospective research, the following four movies have been chosen for further study in the regression model for movie box office prediction: 1) No more bets, 2) Creation of The Gods I: Kingdom of Storms, 3) Dust to dust", and 4) Papa. The selection of sample movies in the second perspective research has been one step further complied with the principle of the normal release date. The realizing time frame is narrowed from July to September 2023.

Douyin index is extracted from oceanengine.com. And it is calculated based on content score, spread score, search score in three parts.

Baidu index is from index.baidu.com, which calculated based on Baidu's intelligent distribution and recommendation content data, the information index is obtained by weighted summation of social media users' behavior such as reading, comments, sharing, likes, and dislikes.

WeChat index is from the original Tencent WeChat source, reflecting the popularity and importance of keywords in the related content.

Toutiao Keyword Index is extracted from oceanengine.com, and it is calculated based on the sample principle as that of the Douyin index.

Variable measurement:

Dependent variable (Y): Daily movie box office of the four movies.

Independent variable (X1): Daily Douyin Index based on exact movie name.

Independent variable (X2): Daily Baidu Index based on exact movie name.

Independent variable (X3): Daily WeChat Index based on exact movie name.

Independent variable (X4): Daily Toutiao Keyword Index based on exact movie name.

Sample size (n) is 80, which is made up of the first twenty days index of each variable for the four movies.

The four variables multiple linear regression model is estimated as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_4 X_4$$

2.2.2 Hypothesis

H1: Douyin Index has a positive influence on movie box office, $\beta_1 \neq 0$

H2: Baidu Index has a positive influence on movie box office, $\beta_2 \neq 0$

H3: WeChat Index has a positive influence on movie box office, $\beta_3 \neq 0$

H4: Toutiao Keyword Index has a positive influence on movie box office, $\beta_4 \neq 0$

2.2.3 OLS Regression Result and Observation

From the OLS Regression result in 1 (Table 2, 3, 4), the P value of both the Douyin Index and the Baidu Index is less than 0.05, which shows a statistically significant relationship between the two independent variables and the dependent variable of the movie box office. The P value of the Weixin Index and Toutiao Keyword Index is larger than 0.05, indicating both indexes are not statistically significant in predicting movie box office. Therefore, H3 and H4 are rejected.

Table 2: OLS Regression result 1(a).

Dep. Variable:	Movie Box Office	R-squared:	0.835
Model:	OLS	Adj. R-squared:	0.826
Method:	Least Squares	F-statistic:	94.62
No. Observations:	80	Prob (F-statistic):	1.60e-28
Df Residuals:	75	Log-Likelihood:	-1474.7
Df Model:	4	AIC:	2959.
Covariance Type:	nonrobust	BIC:	2971.

Table 3. OLS Regression result 1(b).

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.529e+07	4.14e+06	3.695	0.000	7.05e+06	2.35e+07
Douyin Index	11.1489	2.082	5.355	0.000	7.001	15.296
Baidu Index	7.6118	1.325	5.746	0.000	4.973	10.251
Weixin Index	0.0234	0.042	0.558	0.579	-0.060	0.107
Toutiao Keyword Index	-45.5014	58.701	-0.775	0.441	-162.440	71.437

Table 4. OLS regression result 1(c)

Omnibus:	4.436	Durbin-Watson:	1.431
Prob (Omnibus):	0.109	Jarque-Bera (JB):	4.188
Skew:	0.308	Prob (JB):	0.123
Kurtosis:	3.936	Cond. No.	2.34e+08

Table 5. OLS regression result 2(a).

Dep. Variable:	Movie Box Office	R-squared:	0.833
Model:	OLS	Adj. R-squared:	0.829
Method:	Least Squares	F-statistic:	192.4
No. Observations:	80	Prob (F-statistic):	1.13e-30
Df Residuals:	77	Log-Likelihood:	-1475.1
Df Model:	2	AIC:	2956.
Covariance Type:	nonrobust	BIC:	2963.

Table 6. OLS Regression result 2(b).

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.459e+07	3.96e+06	3.680	0.000	6.69e+06	2.25e+07
Douyin Index	10.2052	1.190	8.576	0.000	7.836	12.575
Baidu Index	8.1049	1.156	7.010	0.000	5.803	10.407

Table 7. OLS Regression result 2(c).

Omnibus:	4.141	Durbin-Watson:	1.405
Prob (Omnibus):	0.126	Jarque-Bera (JB):	3.587
Skew:	0.337	Prob (JB):	0.166
Kurtosis:	3.789	Cond. No.	8.45e+06

With only the Douyin Index and Baidu Index retained for the second round of OLS regression modeling, the result is output as OLS Regression result 2 (Table 5, 6, 7).

P value for both Douyin Index is less than 0.05, indicating a statistically significant relationship between Douyin Index and the Movie Box Office. A coefficient value of 11.1489 suggests every unit increase in Douyin Index will result in an increase in the movie box office by 11.1489 units.

P value for Baidu Index is 0.05, indicating a statistically significant relationship between Baidu Index and the Movie Box Office. The coefficient value of 8.1049 suggests every unit increase in the Baidu Index will result in an increase in the movie box office by 8.1049 units.

R- square value is 0.833 (Adj R- square = 0.829), suggesting 83.3% of the total variation could be measured by the model. Both the R – square value and Adjusted R – square value at 80% + level, reporting the total amount of variation that can be accounted for linear regression model are high.

F-statistic is observed at 192.4. Together with the low P value (0.000), it suggests the model is statistically significant, and the linear regression between Y and X1, and X2 is strong.

Given the coefficient estimators from the OLS Regression result, the two-variable multiple linear regression model is expressed as:

$$Y = 1.459e + 07 + 10.2052\,X1 + 8.1049\,X2$$

Overall, out of the four investigated social media platforms, Douyin and Baidu are the two platforms that appear to be significant in terms of predicting movie sales box. At the same time, that reflect social media has a significant impact on consumer behavior of purchase in the movie market.

3 CONCLUSION

In this paper, the correlation relationship between social media and the movie box was studied from two perspectives. Results did show that social media plays a role in influencing movie sales performance. Specifically, from the two perspectives research results, social media platforms, such as Douyin and WeChat, serving as a general information channel, appear to have a stronger positive impact the movie sales performance. Surprisingly, social media index from specific movie informational platforms, such as Maoyan and IMDb has comparable week correlation with movie sales. The marketer may want to shift their marketing effort to the general social media platform instead to achieve more effective marketing objectives.

In the meanwhile, limitations of the research model are also observed as below. There will be time between information search and purchase. While in the modeling, the factor has not been taken into consideration, mainly because of very limited research on the time buffer between the two steps, especially in experience consumption. Also, considering the convenience of online purchases nowadays, the effect on the time buffer should be minimized. However, its potential influence on consumer behavior remains an area not thoroughly examined in this research.

Due to the researcher's technical expertise and data availability online, this study only considers the impact brought by some social media platforms in China. Other social media platforms such as Youku, IQIYI, Bilibili, as well as news channels, are not in consideration. Thus, this study might not fully capture the diverse impacts different platforms can have on movie box office performances.

The heatmap and multiple regression model only show the correlation between various factors and the movie box office, they do not show the causation caution. The movie box office may be affected due to other external reasons such as political stance and societal zeitgeist which are not accounted for in this study.

Looking forward, it is valuable for researchers to deep dive into the relationship between movie box office performance and the social media impact in a more precise manner. A larger sample size and diversified social media platforms will continue to have a wider outlook.

REFERENCES

S. Kemp, Digital 2023: China. DataReportal - Global Digital Insights, https://datareportal.com/reports/digital-2023-china.

M. A. Shareef, B. Mukerji, Y. K. Dwivedi, N. P. Rana, R. Islam, J Retail Consum Serv 46, 58-69(2019)

R. V. Kozinets, Eur Manag J 3, 252-64(1999)

A. Chopra, V. Avhad, A. S. Jaju, Business Perspectives R 1,77-91(2021)

J. Nash, J Fashion Mark Mana 23, 82-103(2019)

A. Wibowo, S. C. Chen, U. Wiangin, Y. Ma, A. Ruangkanjanases, Sustainability 13, 189(2020)

T. Kubrak, Behav Sci Law 5, 86(2020)

A. Molaie, A. Abedin, M. Heidari, Proc Social Behav Sci 5, 832-837(2010)

China Film Data Information Network. The 2023 film box office surpasses 500 billion yuan. https://www.zgdypw.cn/sc/scfx/202311/20/t20231120_7373236.shtml

Y. Chen, Film Mark Indu Report 45(2023)

R. Zhu, Y. Ma, J Chaohu Univ 23, 71-79(2021)

S. Asur, B. A. Huberman, IEEE 1, 492-499(2021)

M. Bogaert, M. Ballings, D. Van den Poel, A. Aztecan, Decis Support Syst 147, 113517(2021)

J. F. Engel, R. D. Blackwell, P. W. Miniard, J Consum Behav 8(1995)

Y. Chen, S. Fay, Q. Wang, J Interact Mark 25, 85-94(2011)

Ipsos. 2023 Social Media Usage Report. Retrieved from https://www.199it.com/archives/1607624.html.

P. Sedgwick, Pea correlation coefficient 345(2012)