

Research on the Influence of Data Analysis on Movie Box Office under the Big Data Environment

Xudong Zhou

Yinxing Hospitality Management College of CUIT, Chengdu 611743, Sichuan, China

21402282@QQ.COM

Abstract. With the advent of the information age, the concept of “big data” has gradually infiltrated into various industries, the analysis and mining of mass data can bring unlimited business opportunities to the industry, and the movie industry is no exception. This paper collects various data of 548 domestic movies in MaoYan Movie, Douban Movie, Entgroup Consulting through the octopus collector, selects factors that may affect the movie box office, and uses SPSS 19.0 software platform to conduct data analysis influencing factors of movie box office. After research, it has been clear that different movie types, release periods and sequels, IP adaptation, and distributing and releasing corporation have significant influence on the movie box office, which reduces investment risks for investors and studio in the movie industry, and provides references for formulating precise marketing production strategies and increase box office income.

Keywords: movie box office, influencing factor, SPSS.

1. Introduction

The rapid development of the movie industry has brought huge benefits to the country, as a part of the cultural industry; it occupies an extremely important position. Under the growing information environment, the concept of “big data” generated from the Internet industry has become more and more frequent in the public life, in order to obtain better box office income; the domestic movie industry has gradually introduced big data technology, in recent years, many movies are inseparable from the data analysis behind the success of the box office, it can be said that under the background of the information age, the use of data analysis is a successful prerequisite to achieve precision marketing.

In the past decade, the Chinese movie industry has developed rapidly, and the box office growth momentum has been rapid, from 2008 to 2017, the box office income of Chinese movies had shown a linear growth trend, the specific data is shown in Fig.1:

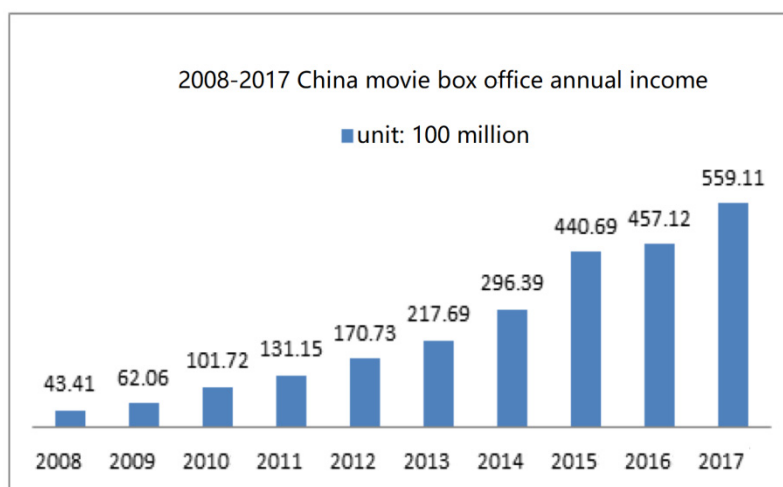


Fig. 1 2008-2017 China movie box office annual income

2. Analysis of Influencing Factors of Movie and Construction of Indicators

This paper takes the China mainland movie market as the key point of data analysis, uses the Octopus collector to collect data, and then uses SPSS to analyze the influencing factor of the movie box office based on data.

2.1 Data Fetching

2.1.1 Data Acquisition

The data acquisition software is used to collect relevant data from MaoYan Movie, Entgroup Consulting, Douban movie, etc., the data acquisition process is shown in Fig.2:

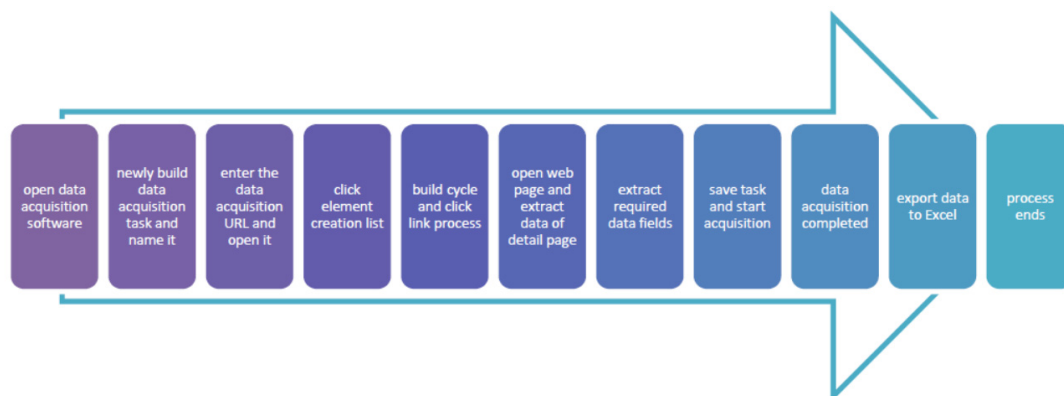


Fig. 2 flow diagram of data acquisition

2.2 Variable Design of data Analysis

The dependent variable of data fetching setting is movie box office, the independent variables include: movie types, IP adaptation, sequel, network score, movie schedule, number of followers, distributing and releasing corporation, specific variable settings, as shown in Table 1:

Table. 1 variable definition and data description

variable type	variable	variable value	data sources
dependent variable	movie box office	0-99(100 million)	EntGroup, MaoYan Movie
independent variable	movie types	Drama movie=1, action movie=2, romance movie=3, comedy movie=4, fantasy movie=5, suspense movie=6	EntGroup, box office ranking list
	IP adaptation	yes=1, no=0	Baidu baike
	sequel	yes=1, no=0	Baidu baike
	movie schedule	New Year movie season =1, summer movie season=2, others=0	EntGroup, Douban movie
	network score	1-10 point	Douban movie,Mtime
	number of followers	0.1-40 (ten thousand people)	Douban movie,Mtime, MaoYan Movie
	distributing and releasing corporation	strong marketing ability=1, weak=0	EntGroup Consulting

3. Data Analysis

3.1 Data Preprocessing

The Octopus collector is used to obtain related data of 1000 top-ranking movies publicly released in the mainland from Douban Movies, Maoyan Movies, and EntGroup, including imported movies and domestic movies, Excel is used to remove imported movies and invalid data, and finally 548 valid data is obtained. The data is imported into SPSS and variables are processed. As shown in Table.2:

Table. 2 variable processing diagram

	name	type	width	decimal	tag	value	loss	array	alignment	measure standard	role
1	name	character string	86	0		no	no	4	left	Name (N)	input
2	box office	numerical value (N)	12	2		no	no	5	right	Measure (S)	input
3	network score	numerical value (N)	3	1		no	no	3	right	Name (N)	input
4	number of followers	numerical value (N)	12	2		no	no	7	right	Measure (S)	input
5	movie type	numerical value (N)	6	0		(1 plot)	no	6	right	Name (N)	input
6	show schedule	numerical value (N)	12	0		(0 other)	no	4	right	Name (N)	input
7	sequel	numerical value (N)	12	0		(0 other)	no	5	right	Name (N)	input
8	IP adaptation	numerical value (N)	12	0		(0 no)	no	5	right	Name (N)	input
9	distributing and releasing corporation	numerical value (N)	12	0		(0 no)	no	6	right	Name (N)	input

3.2 Data Analysis

3.2.1 Descriptive Statistics

In order to understand the basic characteristics of the analytical data, the variables are described and counted, respectively, the analysis results are shown in Fig.3:

	N	minimal value	maximal value	sum	mean value	standard deviation
box office	548	.30	56.79	1400.49	2.5556	4.39007
network score	548	2.5	9.2	3177.3	5.798	1.3451
number of followers	548	.10	41.21	2891.23	5.2760	6.12342
movie types	548	1	6	1484	2.71	1.479
show schedule	548	0	2	531	.97	.805
sequel	548	0	1	67	.12	.328
IP adaptation	548	0	1	166	.30	.460
distributing and releasing corporation	548	0	1	219	.40	.490
valid N (list state)	548					

Fig. 3 descriptive statistics

The maximal value of box office of 548 movies is 5.679 billion Yuan, the minimal value is 30 million yuan, and sum is 1400.49; mean value is 256 million Yuan. The maximal value of network score is 9.2, the minimal value is 2.5, and the mean value is 5.8, mean value number of followers is 412,100, the minimal value is only 1000, mean value is 52,800, the sum of the sequel movies is 67, the number of IP adaptations is 166, and the number of movies released by the top ten movie distributing and releasing corporation is 219.

3.2.2 Variance Analysis

In order to study whether different movie types, different show schedules, sequels, etc. have significant differences on movie box office; each independent variable is brought into SPSS for single factor analysis.

4. Movie Types

To analyze whether there are significant differences in the box office among different movie types, and analyze the variance of the movie type, first, assume that the null hypothesis is tenable, and there are significant differences among the movie box of different movie types. The result of the homogeneity test of variance is shown; the associated probability is 0, it is less than significance level 0.05, so the population variance of each group is not equal. Because of the heterogeneity of the variance, the Tamhane's T2 method is used to test, the result of Tamhane's multiple comparison is shown, it can be seen that the associated probability between the drama movie and the comedy movie is 0.007, which is less than the significance level, namely there are significant differences between the two groups, there are significant differences among box offices of different movie types.

5. Show Schedule

To analyze whether there are significant differences in the box office among different show schedules, variance analysis is carried out on the show schedule, first, assume that the null hypothesis is tenable, and there are significant differences among movie box that are shown in different schedules. The results of the homogeneity test of variance are shown; the associated probability is 0.002, it is less than the significance level 0.05, so the population variance of each group is equal. Due to the heterogeneity of variance, Tamhane's T2 method is used to test, the results of Tamhane's multiple comparison are shown, it can be seen that the associated probability between the New Year movie season and other schedules is 0.018, which is less than the significance level, there are significant differences among the New Year movie season and other movie boxes in other schedules.

6. IP Adaptation, Sequel Movie and Distributing and Releasing Corporation

The IP adaptation, the single-factor variance analysis is carried out for sequel movie and distributing and releasing corporation, the associated probabilities of analysis results are all 0, which was less than the significance level, and it shows that the IP adaptation, sequel and marketing ability of the release company had a significant impact on the box office.

dependent variable: box office

source	III type quadratic sum	df	mean square	F	Sig.
correcting model	1207.741 ^a	7	172.534	9.981	.000
intercept	3371.684	1	3371.684	195.053	.000
IP adaptation	.001	1	.001	.000	.994
sequel	189.177	1	189.177	10.944	.001
distributing and releasing corporation	32.342	1	32.342	1.871	.172
IP adaptation* sequel	197.548	1	197.548	11.428	.001
IP adaptation* distributing and releasing corporation	31.993	1	31.993	1.851	.174
sequel* distributing and releasing corporation	.091	1	.091	.005	.942
IP adaptation * sequel * distributing and releasing corporation	12.910	1	12.910	.747	.388
error	9334.428	540	17.286		
total	14121.296	548			
corrected total	10542.168	547			

a.R square = .115 (adjust R square = 1.03)

Fig. 4 inter-group effect test

In addition, the three factors of IP adaptation, sequel film and distributing and releasing corporation are jointly investigated, and multi-factor variance analysis are carried out, as shown in Fig.4, the results of the inter-group effect test are shown, the associated probability of the sequel is

0.001, it is smaller than the IP adaptation (0.994) and the distributing and releasing corporation (0.172), namely the influence of the sequel on the box office is greater than the IP adaptation and distributing and releasing corporation. The associated probability of the combination of IP adaptation and sequel is 0.001, namely the interaction between the two has a significant influence on the box office of the movie, it shows that IP adaptation and sequel are combined, which has a positive effect on the box office.

7. Result Discussions and Analysis

7.1 The Movie Types Produce Significant Influence on the Box Office

Through data analysis on the factors affecting movie box office, it is found that the movie types has a significant influence on the box office. As shown in Fig.5, it can be seen that the audience prefers comedy and action movies. According to the data, in all the films released last year, the proportion of action and comedy-type movie box office accounts for more than 20%, the two occupies more than 50% of the total.

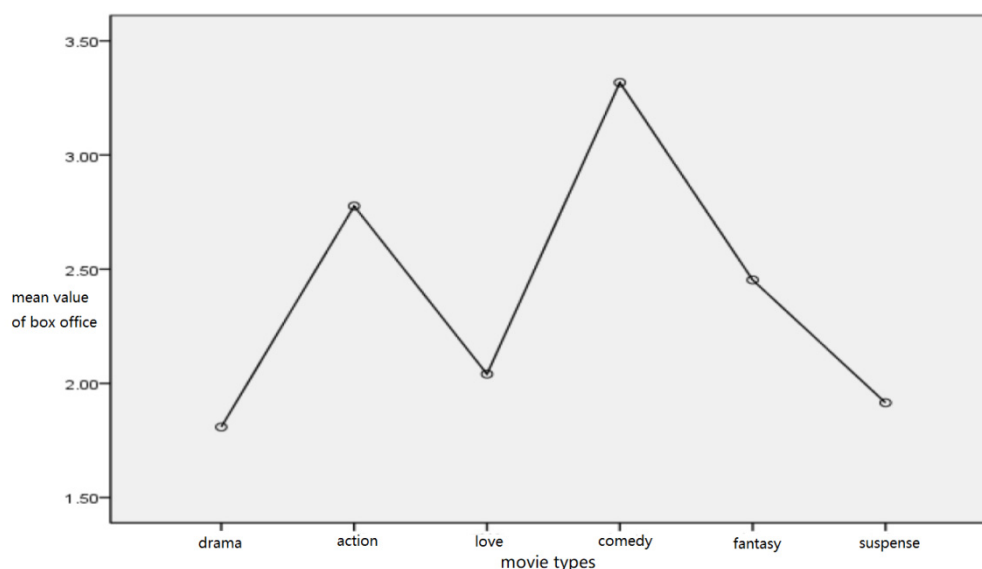


Fig. 5 box office-movie type mean value diagram

According to the regression analysis, it was found that love and drama films had a negative impact on the movie box office. This reflects taste changes of audience, in the domestic movies released in 2017, the number of love movies is the largest, but the box office accounted for only 5%. By consulting relevant data, it can be found that in the hundreds of millions of movies in recent years, the number of love movie is indeed many, but the movies with good reputation and good quality are very few, and the long-term accumulated negative reviews has also affected the audience's choice. The drama movies in this paper include history, war, western movies, etc., movie audiences are few, and investment is basically lower, most are low-cost movies, these films lack certain marketing method in the Internet age, which makes audiences lack understanding of these movies.

7.2 The Show Schedule has an Impact on the Movie Box Office

Although the show schedule is not introduced into the equation in the regression analysis, the reason may be that in the multiple linear regression analysis, the influence of other independent variables is greater, so the show schedule with less impact on the box office is eliminated. However, according to variance analysis of show schedule, the New Year movie season in the show schedule has a significant impact on the movie box office. As shown in Fig.6, it can also be seen that the average box office of the New Year movie season is the highest, much higher than other schedules. This aspect also shows that the audience is more inclined to go to the cinema to watch movies on New Year's Day, Spring Festival and other festivals.

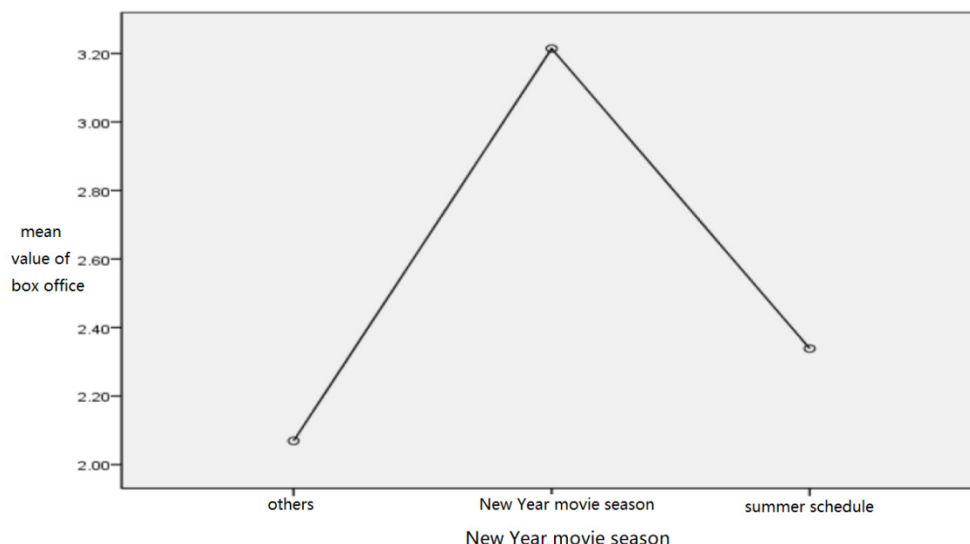


Fig. 6 box office and show schedule mean value diagram

7.3 IP Adaptation, Sequel, and Distributing and Releasing Corporation have a Positive Influence on Movie Box Office

In the analysis of variance, IP adaptation, sequel, distributing and releasing corporation and movie box office have significant differences, and the combination of sequel + IP adaptation has a more significant impact on the box office. The regression coefficients in the regression equation are IP adaptation (0.823), sequel (1.584), and distributing and releasing corporation (0.974), all shows that they have a positive influence on the box office.

The popular IPs to movies is adapted to movies, which is very hot in recent years, these popular IPs have a strong fan base, the audience has expectations for stories that they are familiar with and interested in, and become a potential driving force for box office growth, in addition to the marketing of the studio, they lay a solid foundation for box office income. Generally speaking, the sequel will be shot after the previous movie has achieved a relatively good box office score, so the audience expects a higher sequel and thus has a positive impact on the box office.

According to the actual situation, powerful distributing and releasing corporations usually have more and wider cinema resources, which provides basic guarantee for the sale of the film. The ability of the distributing and releasing corporation often determines the release time of the movie, which has a great impact on the movie box office.

7.4 The Number of Followers has a Positive Impact on the Movie Box Office

According to the relevant analysis, there is a significant positive correlation between the number of followers and the movie box office. The coefficient of number of followers in the regression equation is 0.336, which has a positive impact on the movie box office. As shown in Fig.7, it can be seen that there is a basically positive correlation between the number of followers and the box office. The number of followers represents a part of the people who are interested in the movie; these people have the willingness to go to the movies, which has a positive impact on the box office. On the other hand, it is also the embodiment of movie marketing. The Internet is more and more developed, and the development of new media is also very rapid, the studio can increase the attention of the film in various ways, develop more potential viewing groups, and increase the box office income.

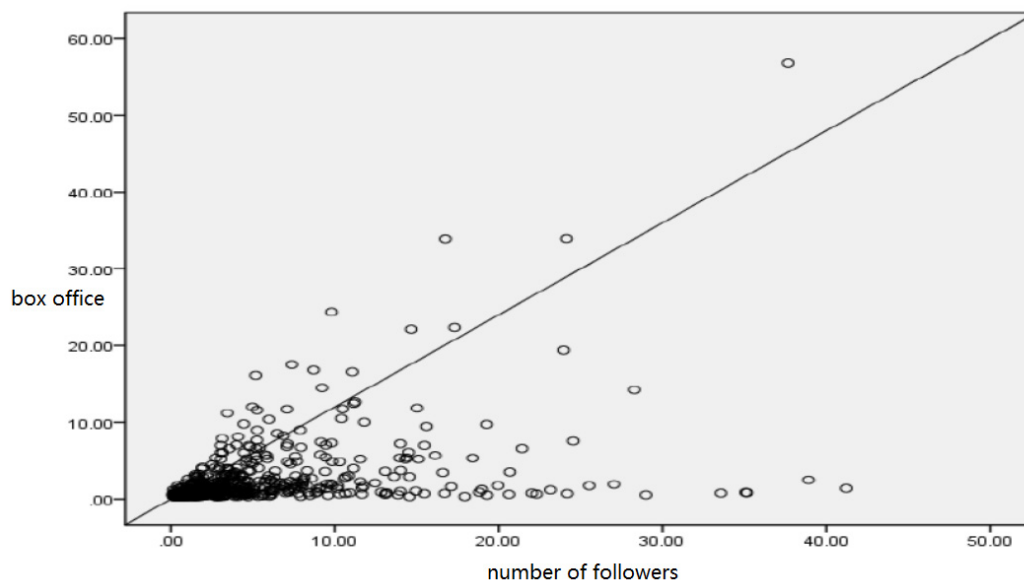


Fig. 7 box office-the number of followers scatter diagram

7.5 Network Score Negatively Influence Movie Box Office

The coefficient of the network score in the regression equation is -0.362 , which has a negative effect on the box office of the movie to a certain extent, as shown in Fig.8, which means that the higher the score, the better the box office. This shows that although the studio can improve the attention of the film through various marketing methods and increase the box office income, the promotion cannot bring high evaluation to the film, in recent years, the movie "best game no one played" is everywhere, and the studio focuses on the promotion and ignores the quality and lead to this phenomenon.

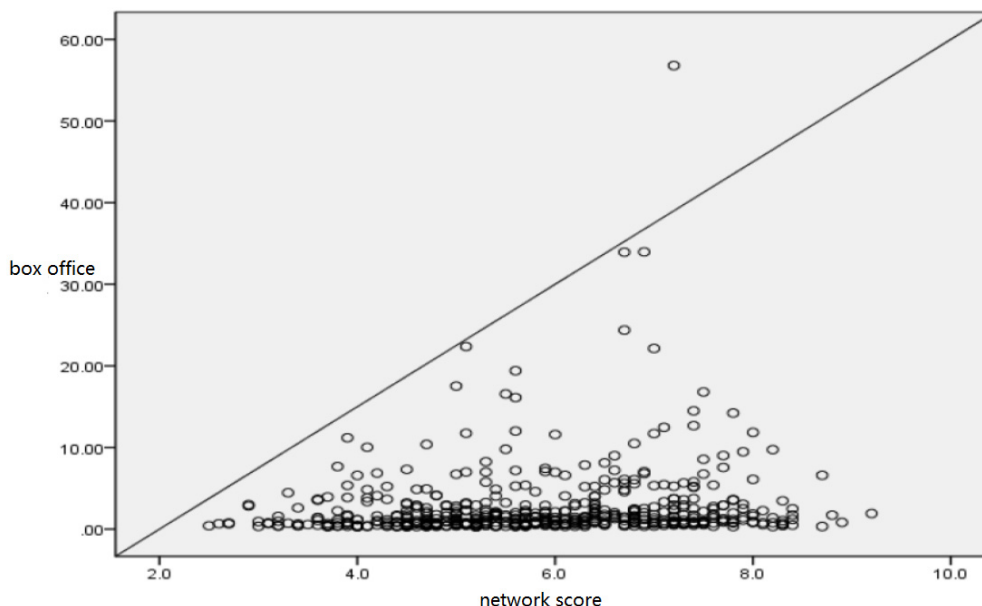


Fig. 8 box office-network score scatter diagram

8. Conclusion

Although the movie market is developing rapidly, the annual box office is rising year by year, and the box office record is constantly being refreshed, however, the relevant data shows that the film market has always been in a minority profit, part is flat, and there is still a part of the loss, which is not optimistic for the movie market. However, the development of technology and the Internet has

affected people's lives, especially with the arrival of big data, which has brought new impetus to the movie market and brought the vitality to weak market.

According to the variance analysis, different film types, show schedule and sequels, IP adaptation, and distributing and releasing corporation all have significant effects on the movie box office; according to relevant analysis, the number of followers and the movie box office present significant positive correlation; according to the results of multiple regression analysis, distributing and releasing corporation, the number of followers, IP adaptation have a positive effect on the box office, and love and drama movies have a negative influence on the box office, and the influence is greater, the network score negatively affects the movie box office .

This paper proposes the following suggestions in combination with the results of data analysis, it is necessary to consider the movie box office income with data analysis results, the following points must be paid attention to: attach importance to the choice of theme, enrich the film type; correctly locate the film schedule, choose a favorable schedule help the growth of the box office, people prefer to go to the cinema on important holidays such as New Year's Day and Spring Festival, the studio should seize the important schedule, the favorable schedule will contribute to achieving better box office results under the low cost of the movie; the higher the attention of the movie, the more people know, the more people who are interested in the film, thus promoting the increase of the box office; creating high-quality sequels and adaptation movies; pay attention to the issuer's publicity ability, the release can be said to be the "last mile" before the movies enter the market, a good release strategy will greatly help the box office.

The use of big data in the Internet era can make the work more successful, the movie industry will increasingly integrate the concept of big data in the future, data analysis technology is used to formulate a more scientific movie shooting strategy for the studio, and obtain a higher box office income.

Acknowledgements

Project fund: scientific research project of sichuan county economic research center in 2018, No: xy2018040.

References

- [1]. Pei Pei, Jiang Yinlong, Analysis of Influencing Factors of Domestic and Foreign Movie Box Office [J]. Co-Operative Economy & Science. 2014, (02).
- [2]. Guo Xinru, Wang Wei, Zhang Nan. Research on the Box Office's Influence Factors of Small and Medium Budget Movies of China-Based on the Empirical Analysis of 100 Movies [J]. Cultural Industry Research. 2015, (03): 33-40.
- [3]. Lu Wenjing, Feng Xiao. The Confirmatory Analysis of the Influence Factors of Movie Box Office in the Era of Big Data [J]. Journal of Communication University of China Science and Technology. 2017, (02): 41-46.
- [4]. EntGroup Consulting. 2017 Domestic Film Release Market Inventory [R]. Beijing: EntGroup Consulting, 2018.