

# Research on 10000 IMDB Movie

Yifan Zhang

## Abstract

This report explains the statistics on the box office and genre of 10000 IMDB films. At the same time, the author analyzed a large number of data to find the correlation and studied the trend of the number of movies from 1902 to 2021, the correlation between the box office and movie scores of movies in different languages, and the reasons for this phenomenon.

**Keywords:** IMDB films, film volume trends, movie ratings, movie popularity

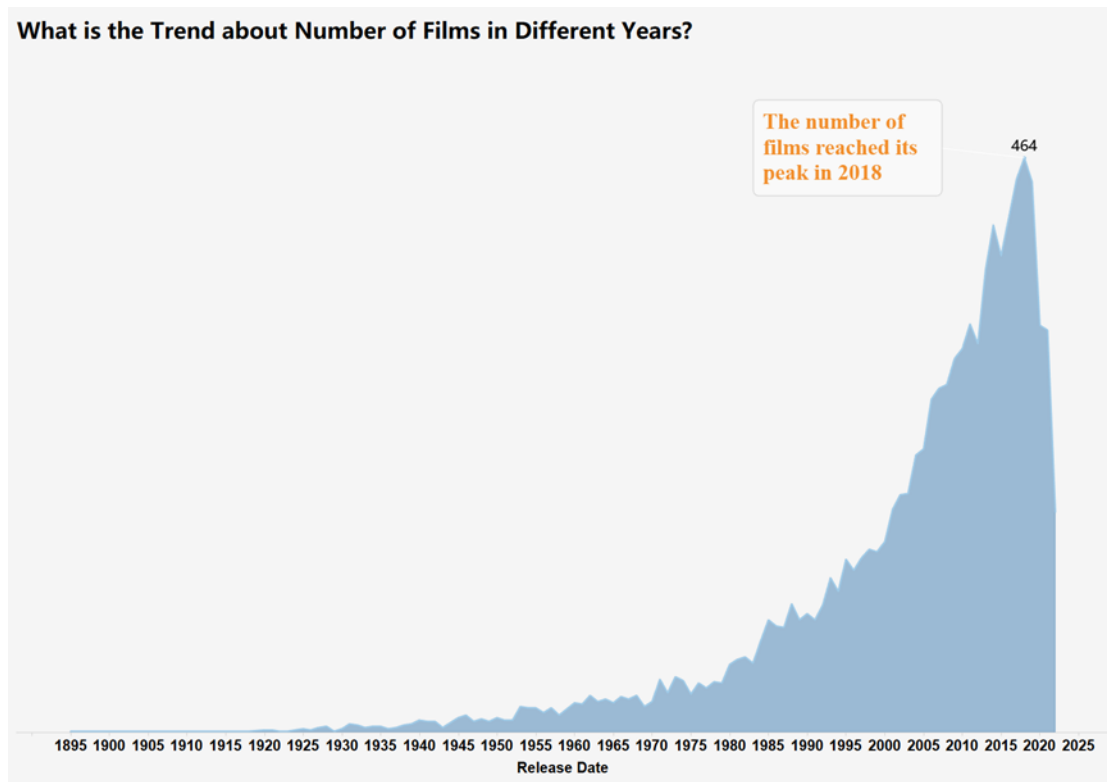
## 1.0 Introduction

Through the dataset of about 10000 IMDB movies from Kaggle, the author has a strong research interest in it. The vast amount of data shows a complex correlation between the categories, popularity, ratings, and other information of these 10,000 IMDB movies. Therefore, the author

wants to make clear the trend of these data, the reasons behind the correlation, and the implications by visualizing the data.

## 2.0 Body

### 2.1 Changing Trend of the Number of Films



This table shows the number of films produced per year recorded on IMDB from 1895 to 2022. This shows that IMDB recorded only one film annually from 1895 to 1918. The number of films produced each year began to slowly increase, and by 1983, the number of films had increased to 56. After this, the number of films began to

grow rapidly, and by 2018, the number of films produced reached a peak of 464. After 2019, the number of films began to decline rapidly again, and by 2022, the number of films produced this year will be only 178, which is 62% less than in 2018.

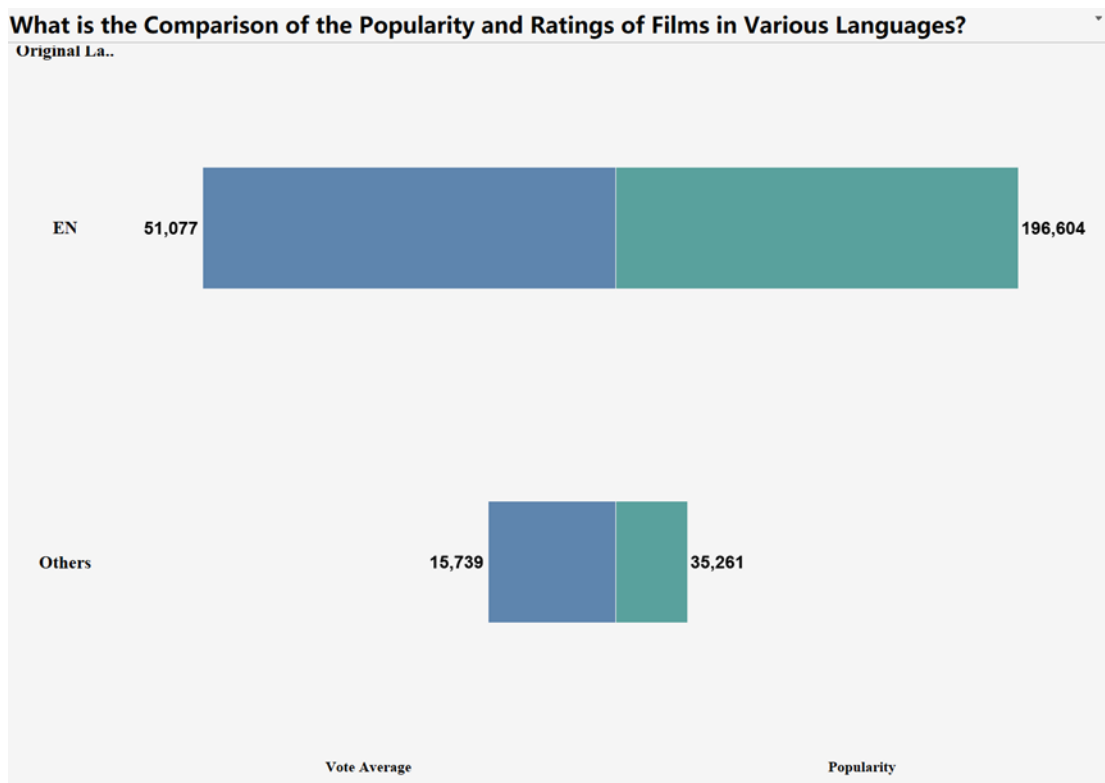
The authors of this report believe that the change in the

number of films reflects the development of the film industry over the past century, as well as technological, economic, cultural, and even social factors at different times. The increasing number of movies in the past century also shows the development of world culture, technology, economy, and even the happiness index. From 360 doc.com, at the end of the 1980s, the global network gradually formed, culture, science and technology, and society developed rapidly, the world economy also tended to internationalization, the rise of multinational companies, and the openness of countries to the film industry gradually increased. Hence, the world

film trade developed rapidly.

The most important reason for the sharp drop in the number of films after 2019 is the COVID-19 pandemic. The epidemic's impact is many, including health, social productivity, and economic problems. Therefore, it has indirectly led to a certain degree of shrinkage of the film industry, and the number of films produced yearly has fallen off a cliff.

## 2.2 Compare the Relationship Between Movie Popularity and Movie Rating



This graph shows the relationship between English and non-English movies' popularity and movie ratings. This shows that the popularity of English films is more closely related to film ratings than non-English films.

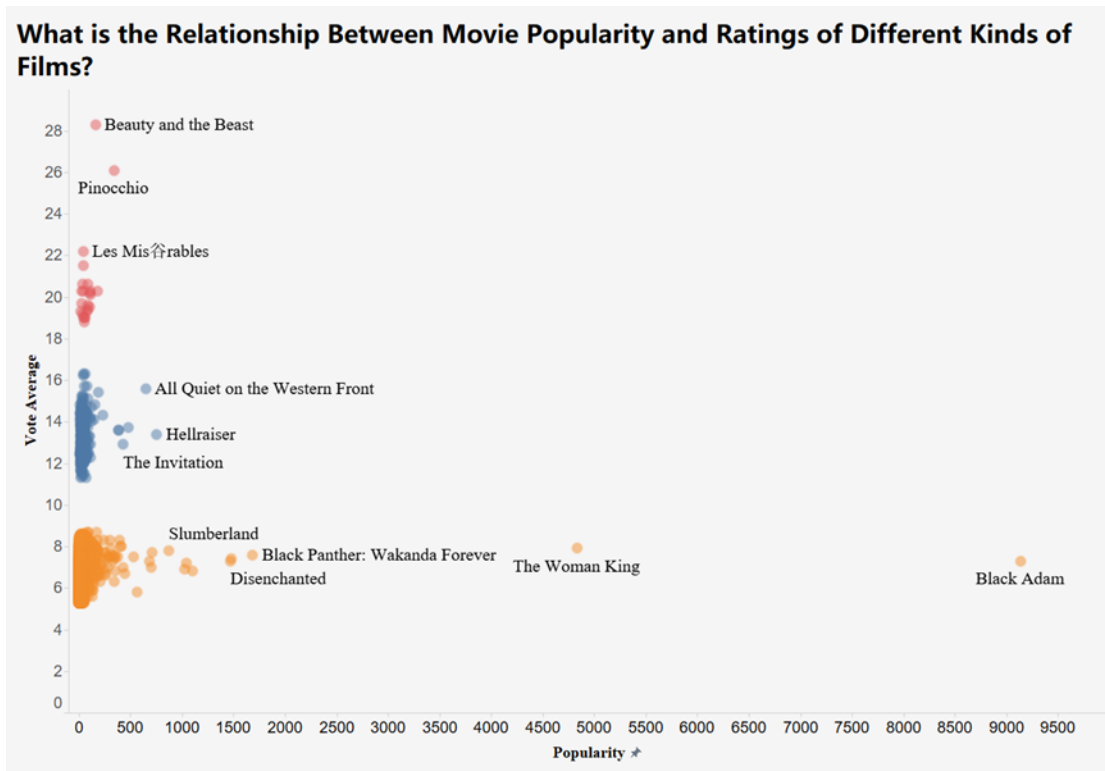
From Alexa, the author learned that IMDB is an American company, and visits from the United States are the most, accounting for 28.9%; India is second, accounting for 6.9%, and a total of about 50% of visits are from countries where the mother tongue or second language is English. China accounts for 3 percent. This indicates that the majority of IMDB users are English speakers. A large number of user participation determines IMDB scores. And people who speak English tend to watch movies in English. Therefore, the author believes that this is the

reason why the popularity of English films is more closely related to film ratings.

Therefore, to a certain extent, the data of non-English films can more objectively reflect the popularity of films and the general relationship with film ratings. As can be seen from the graph, there is a certain correlation between movie popularity and movie rating, but not a decisive correlation. Many factors exist, so I have researched this issue and worked out the following report content.

## 2.3 The Relationship Between the Popularity of Different Kinds of Movies and Movie

Ratings



The author divides the data in the graph into three categories and turns these three categories of data into red, blue, and yellow.

The yellow data shows that although the movie ratings are generally low, there are the most popular movies. This shows that some films can attract audiences through other factors, break through the rating limits, and win their love. By searching for information on such films, the author found that it may be because the actors' fame increased the number of people watching them.

From the blue data, it can be seen that the ratings of the movie and the audience's love are the same. This shows these films have broad appeal and can gain audience recognition and affection.

As can be seen from the red data, the movie has a high rating, but the audience's love is not high. This may be because these films are too artsy or specialized to resonate with the general audience.

From IMDB, there is generally a certain correlation between movie ratings and movie popularity, but the correlation is not very strong. The popularity of a film is related to many factors, such as the type of film, the release date, the distribution company, the publicity, and the cast. The movie score reflects the quality of the movie more, which does not necessarily determine the number of

viewers of the movie but is often related to the prediction of the movie's box office and many complex factors such as emotion analysis and logic model. Therefore, the correlation between IMDB movie ratings and movie popularity is not very strong.

**3.0 Conclusion**

This report makes and analyzes visual data based on the TMDb 10000 Movie Dataset. This paper first expounds on the trend of changes in the number of films recorded by IMDB from 1895 to 2022 and the background and reasons behind it, as well as the important reasons for the rapid development of science, technology, and culture in the late 1980s and the impact of the epidemic in 2019. Therefore, the author believes that the trend in the number of films can also reflect society's technological, economic, and cultural development to a certain extent. Then, this paper compares the relationship between IMDB English movies and non-English movies' popularity and movie ratings and finds that English movies are more correlated due to the distribution of IMDB users. Therefore, the author believes that IMDB can more objectively explain the general relationship between movie rating and movie popularity for non-English language movies and also shows that the relationship between movie rating and movie popularity

is certain but not strong. Finally, it summarizes the correlation between the popularity of different types of movies and movie scores and the complicated reasons behind it, such as the release date the amount of publicity.

### 4.0 Reference

Kaggle.(2023). TMDb 10000 Movies Dataset

. [Online] Available at: <TMDb 10000 Movies Dataset | Kaggle> [Accessed 20 September 2023].

360 doc.com.(2016). Current Situation and Trend of World Film

Development

. [Online] Available at: < [http://www.360doc.com/content/16/0120/20/28093736\\_529387124.shtml](http://www.360doc.com/content/16/0120/20/28093736_529387124.shtml) > [Accessed 20 September 2023].

Douban.(2023). Data about IMDB and Douban Films, China, Japan, Korea and India films

. [Online] Available at: < [https://www.douban.com/note/703905592/?\\_i=5482220XeR0iKt,5487363XeR0iKt](https://www.douban.com/note/703905592/?_i=5482220XeR0iKt,5487363XeR0iKt) > [Accessed 20 September 2023].