

# Behind the success of movies

## 1. Introduction

Everyone loves a good movie!

Movie industry is a multi-billion-dollar business. Lots of movies are being released in every year. All of these movies have different features, budget, cast crew, etc. but one thing is in common - all want to be successful. Success of a movie depends on various factors of past and present. Identifying the right factors can predict the profitability, rating and award winning of a movie that we define as the most important criteria for movie's success. Some of the factors in predicting movie success are budget, actors, director, IMDB vote count, social fan following and likes, etc. Movie's success prediction plays an indispensable job in the film industry, some researchers over the years predicted success of a movie by its profitability or IMDB rating, In this project, we will test various factors and find a movie's success model that consider both in addition to award winning of a movie. The success can not be predicted based on one specific factor, what makes it harder to analyze.

Our main goal in this project is to find out what are the most significant factors and connections that influence our success indicators by analysis and manipulations of our data, what causes the difference between the ups and downs of the success indicators, and try to explain the reasons for this in a scientific and research way.

## 2. Data overview

The data for the project included two datasets that was taken from Kaggle and one dataset contains movie's released month that scrapped from IMDB site.

**Important features:** Director, Actor, Genres, Release month, Facebook likes, gross and imdb score.

**Calculated important features:** ROI, Oscar winner(binary), succeed Director and Actor rating/gross history.

**Less important features/ useless:** Title, content rating, Aspect Ratio, Color Company, Duration, Language, Filming Location, number of critic reviews, number of voted users.

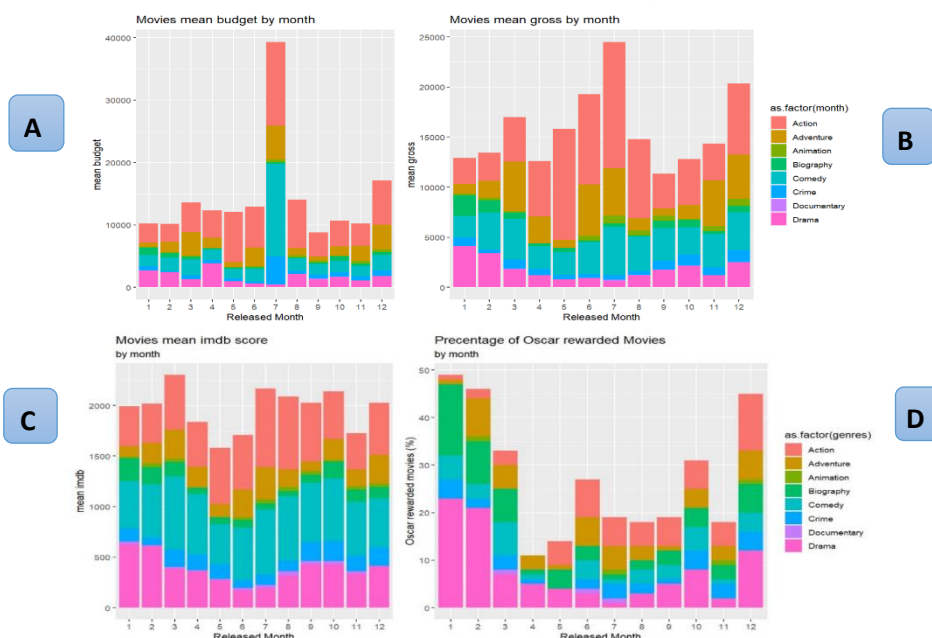
We cleaned our data from unnecessary features, Na's and anomalies. we left with 23 variables And 3885 movies.

## 3. Methods and results

After a research and analysis of the data we found that the most relevant features for our success parameters are: release month, genre and average director and actor rating.

In the following section we answer the most relevant question that came up from our features and explain our insights.

### 1. What are the best months to release a movie and why?



By the success factors we defined for successful movie we found that even the release month of a movie depends on the movie's goals.

- As we can see in chart A, represent the budget by month, the summer season, used to be from June to late July, is the time for big budget breaking records movies, such as, The Dark Knight Series, Star Wars Series, etc. In addition, chart B, represent gross by month, shows that the most profitable month are also the summer months.

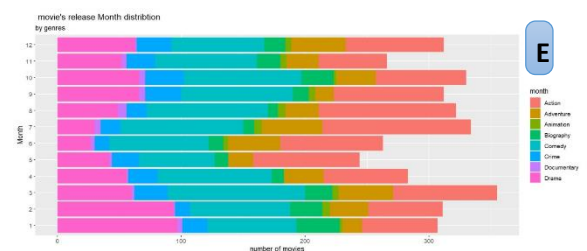
When the distributors are looking for pure box office revenue and the quickest return on the initial investment, then the summer months are the way to go. An explanation we found to this subject briefly says, "Summer movies are marketed as events for a reason. Kids are out of school, many parents are taking vacation time, etc. There is more money to be made because there are more butts to fill the seats". (attached link in appendix section).

- As we can see in chart D, December through February are the key months for distributors looking to market their films for the upcoming award season. The reason these two-three months are the best for awards contenders is simple... it's just a couple months before ballots, Oscar and more awards go out, when the films are fresh in the public's and especially the voters' minds. It's the 'kiss of death' to release a hopeful awards contender in March or April. People won't remember them.

These insights shows that big blockbusters in the summer and award contenders in the winter, but the charts also show some more depth. After the analysis we found that there is a non-random connection between movie's release month to it's genre, that leads us to the next insights.

## 2. How movie's genre affects it's success factors?

- If we look deeper into chart B, we can spot that Action & Adventure films have taken the top spot at the box office, chart E shows that about half of the summer movies are Action & Adventure.



So why there are still other genres taking part in the film industry?

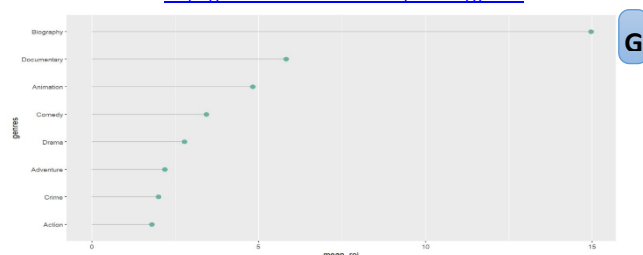
fun fact: For every Action film made in North America since 1995, five Drama films were made. As far as genre goes, Drama and Comedy make up the majority of the production pie (chart F), but here's the rub , Action & Adventure make up most of the profit pie.

Market Share for Each Genre 1995-2022

Rank	Genre	Movies	Total Box Office	Tickets	Share
1	Adventure	1,132	\$65,407,204,383	9,314,074,616	27.01%
2	Action	1,135	\$51,937,626,071	7,195,741,303	21.45%
3	Drama	5,543	\$35,686,156,829	5,429,219,473	14.74%
4	Comedy	2,445	\$33,757,187,802	5,164,650,305	13.94%
5	Thriller/Suspense	1,232	\$19,893,431,315	2,898,981,781	8.21%
6	Horror	744	\$13,684,039,487	1,954,993,306	5.65%
7	Romantic Comedy	642	\$10,747,901,795	1,727,932,197	4.44%
8	Musical	204	\$4,336,246,864	573,155,343	1.79%
9	Documentary	2,449	\$2,523,420,699	376,555,994	1.04%

Table from: <https://www.the-numbers.com/market/genres>

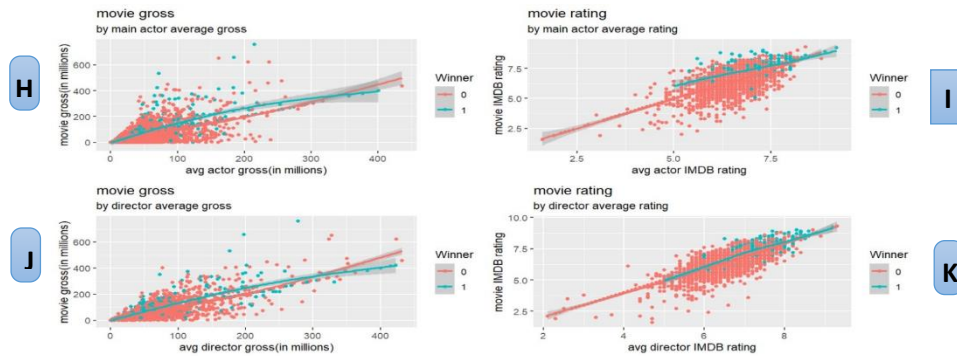
- The reason is return on investment . That is a financial ratio used to calculate the benefit an investor will receive in relation to their investment cost. In chart G we can see clearly that even though Action & Adventure have higher profit than all other genres, the ROI is far from best, and that is why production companies still making other genres more than Action & Adventure.



- Last insight from months and genres factors is the majority of Oscar awarded movies. In chart D we can see that most of them are Drama & Biography movies, especially on the months before award season.

From both months and genres factors we found connection to profitability and rewards, but still no connection to movie's IMDB rating. For that, we investigated the contribution of the directors and main actor rating.

## 3. How director and main actor rating affect the movie's success parameters?



When thinking about influential roles in production, the first that springs to mind is of course the director. The clue is in the name – they direct the film, meaning they have a lot of control over the outcome. In this project we calculated the average rating of every director and actor in our data until the last movie they made, by that, we could get a little sense about the quality of them in any given time and see how it affects the success features.

As we can see in charts H-K there is a connection between main actor and director rating to all three parameters of success. The blue point represent the Oscar rewarded movies are in the right-upper part of each graph, what means as high gross/rating when higher director/actor rating. Like Oscar awarded factor, movie's gross and rating that getting higher rate while director/main actor rating rise.

After we investigated all the features and found the important features that influence the success factors the most, we built a machine learning algorithm to assess our assumptions and maybe predict the next movie success.

#### 4. Model

We used a logistic regression model to classify if the movie succeeded in all 3 parameters, our model uses all the important features that we found as relevant. because the Imbalanced data we use random oversampling to enrich the amount of succeed observation. This method works with minority class. It replicates the observations from minority class to balance the data. Although, the training accuracy of such data set will be high, but the accuracy on unseen data will be worse – the model help us to assess that our success factors are not random.



#### 4. Limitations and Future work

Our approach in this project to movie's success factors has some limitations, the main thing is the lack of observation. We only have 191 observations for success in all three factors, mostly because Oscar winner title. It was harder to analyze and train the data for reliable result. If we could invest more time on this project, we would explore other movie titles and nominees, to expand our observation for success.

If we had additional three month, we would like to explore more relevant data from the social media and audience review, what could have been helpful to our opinion, because of the rising impact of the social media charts on our life.

## 5. Appendix

- ✚ Genres: <https://screencraft.org/blog/this-movie-genre-box-office-king/>
- ✚ <https://www.ukessays.com/essays/film-studies/director-influence-film-success-3738-7716.php>
- ✚ <https://screencraft.org/blog/this-movie-genre-box-office-king/>

## 6. Interesting insights not scientifically based

- ✚ The most common (participated in the most movies together) Director Actor due is Clint Eastwood and Clint Eastwood with 8 joint movies together.
- ✚ The most successful due that is In the top 10 movies in all categories (number of Oscar winners, avg gross, avg imdb) is Peter Jackson, Christopher Lee.

director_name <chr>	main_actor <chr>	join_movies <dbl>	g_mean <dbl>	imdb_mean <dbl>	winner_sum <dbl>
Peter Jackson	Christopher Lee	2	327158238	8.75	2

- ✚ Time Isn't Money: Does Runtime Affect a Film's Box Office?
  - films lasting 136 minutes or more tend to have much larger worldwide grosses than the films of shorter lengths