What Makes a Movie Successful?

Miguel Pinto (up201905004), Ruben P. Pombo (up 202302830), Bárbara Ramos (up200405638)

***Abstract*— This project sought to extract insights from a dataset comprising over 5000 movies, employing diverse visualization techniques to understand the factors influencing movie success. The hypothesis, supported by several visualizations, posits that film success is intricately linked to thematic elements, particularly those depicting alternative fantastical and animated worlds. This is supported by observations in top-performing genres, descriptions of successful films, and seasonal trends. Additionally, we propose that a director's creative control, especially when closely connected to the writer, plays a crucial role in a film's success, highlighting the impact of cohesiveness on the team's performance.**

I. Introduction

The movie industry is filled with risks, where a single unsuccessful movie can have profound financial consequences for production companies. Despite substantial capital investments, predicting a movie's success remains highly uncertain, creating a demand for reliable forecasting methods during the critical investment phase (Packard et al., 2016). Success in the movie industry can take various forms, including critical acclaim, box office success, or cultural impact. The unique nature of each film adds complexity to success forecasting, particularly in the early stages of production (Marshal et al., 2013).

The preproduction phase, involving decisions on storyline, director, writer selection, and casting, plays a pivotal role in shaping a movie's potential success (Wei and Yang, 2022) and therefore, this study aims to contribute to a deeper understanding of the factors influencing a movie's success, offering valuable insights for producers, financiers, and directors to optimize their preproduction decision-making processes. This project will involve the pre-processing of a dataset containing multiple pieces of information about various movies in the IMDb database and, by obtaining visualizations, the aim is to discover patterns and information that might be relevant to major movies producers companies when making a successful movie.

II. First Steps

A. Exploration of the dataset

The IMDb dataset contains detailed movie information with 5487 entries and 24 variables.

B. Preprocessing

In this phase, we tackled data issues like duplicates, missing values, and illogical release years. New features, including a publication month tracker, were introduced, and a profit column was calculated based on the given description: ***Profit=Worldwide Gross Income - Budget***.

Financial metrics were converted to USD using an exchange dataset and scaled for millions, for ease of representation. Moreover, we simplified the country column to feature only the first country and expanded the dataset by creating new rows for unique combinations in the genre, director, and writer columns.

C. Success function

The success function considers financial, social, and global success with specific variables. Economic success is gauged by profit, focusing on the median of highly profitable movies. For audience rating (avg_vote) and social success, a minimum average vote threshold was set, excluding poorly rated movies and also movies with a reduced number of votes. This resulted in a subset of 1490 movies for further analysis. To predict success and after obtaining the top movies dataset, a new metric (success_rating_0_10) was also created based on the same variables used in the success function, in order to facilitate some of our analyses and visualizations. This metric was created by giving different weights to the variables that we consider to be related to success according to the formula:

Success_rating_0_10 = (0.5 * profit) + (0.25 * avg_vote) + (0.25 * votes)

Success values were then normalized to a 0-10 scale for a structured rating system, streamlining analysis and visualizations.

III. Studying the Variables Influence

Each variable was studied to unveil its potential significance and correlation with overall movie success in the dataset.

A. Publication month

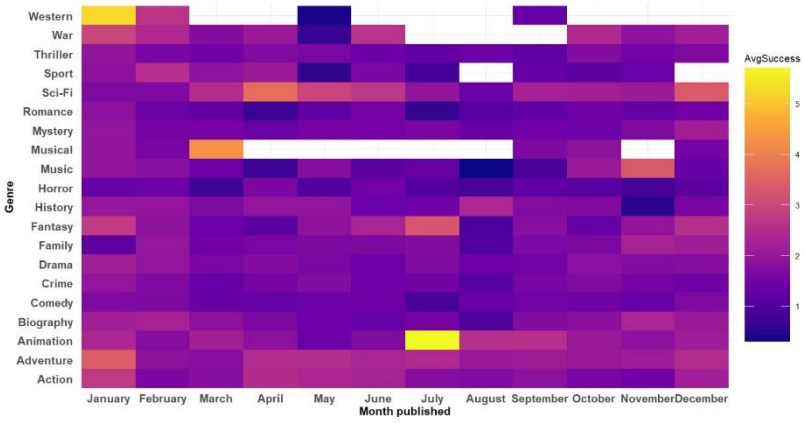

Figure 1 - Genres average success by month

In Figure 1, it's evident that movies falling under the Animation and Fantasy genres released in July tend to achieve a considerable high success. Additionally, the Sci-Fi genre consistently secures the highest success ratings in April and December. The very reasonable explanation could be attributed to demographic preferences as children and young audiences, who predominantly favor animated and fantasy movies (e.g. Disney, Pixar) often indulge in movie-watching during breaks and vacation periods, potentially driving higher success rates during this summer month (Grodal, 2017).

B. Duration

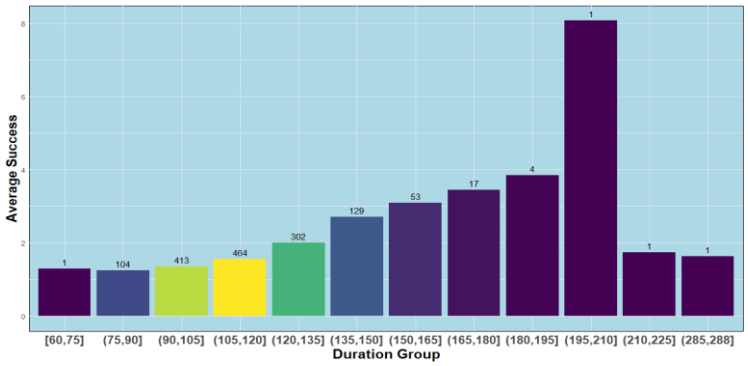

Figure 2 - Average success of different movie duration intervals. The number of movies for each duration is represented above each bar.

We categorized movie durations into 15 minute ranges, allowing each movie to fall within a specific range. Notably, Figure 2 indicates that the most successful ranges spanned from about 140 to 190 minutes, but it could not be selected a smaller range because of the uncertainty and a relative lack of data for longer intervals.

C. Genre

1) Genre choice

In order to focus our analysis on the most successful genres, we decided to create a plot (Figure 3) of all the genres metrics in the dataset and ended up choosing Adventure, Sci-Fi, and Animation. From the top 5 we decided to eliminate Western and War as they had few representatives in the dataset (4 and 18 films, respectively), and we considered that their analysis would be inconclusive and not so accurate. This was also supported by the fact in the last decade these genres have been in decline in terms of movie releases (article) which can also be seen in Figure 4.

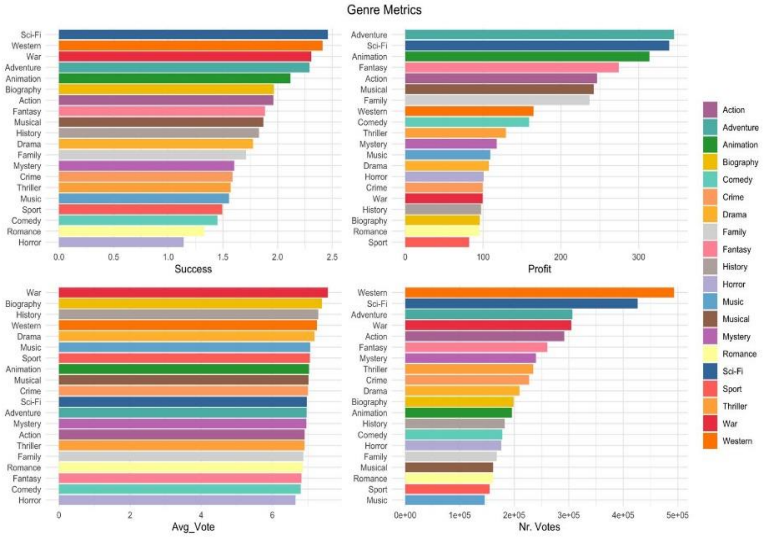

Figure 3 - Genres different success metrics (profit, avg_vote, votes and sucess_rating)

On the other hand, Adventure draw inspiration from basic mammalian behaviors like fighting, bonding, and exploration for resources, while Animation and Sci-Fi explore alternative reality scenarios. Genre preferences and viewer choices are significantly influenced by biological factors, particularly age and gender, serving as crucial determinants. Age, in particular, shapes genre preferences along two dimensions: the balance between fantasy and realism and the level of arousal prompted by the genre (Grodal, 2017).

In fact, as seen in Figure 4 the genres we focused on have seen a significant increase over the years in the number of films released, which indicates that there has been a stronger investment in these genres as the years go by and, for this reason, studying them can provide us with very important insights into what can lead a production company to make the best possible profits from the production of a film. The abrupt decrease in the number of films released in 2020 could be due to the fact that it was the year of the COVID pandemic or simply the stoppage of data collection for the dataset on that date.

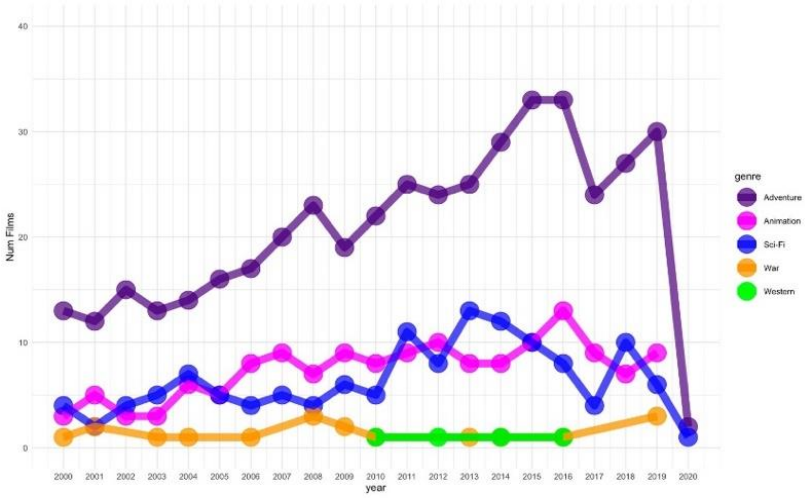

Figure 4 - Top Genres movie releases throughout the years

2) Top genres analysis

a) Adventure

(1) Actors

The data from Figure 5 indicates that while John Oliver (green line) has high profits, his average number of votes is low, suggesting that Mark Hamill is a more consistent choice for this genre, as he excels in all metrics. Additionally, the overlapping lines for Chiwetel Ejiofor and John Oliver suggest they could be an effective actor combination for this type of genre, as evidenced by their overlapping lines indicating they participated in the same movie which had a great success.

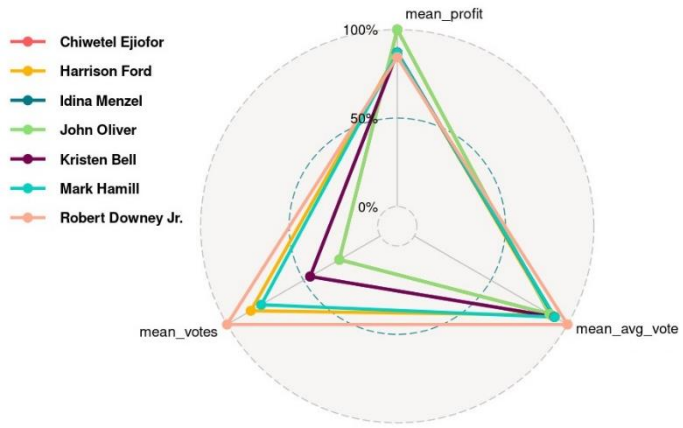

Figure 5 - Top actors of the 'Adventure' genre

(2) Directors

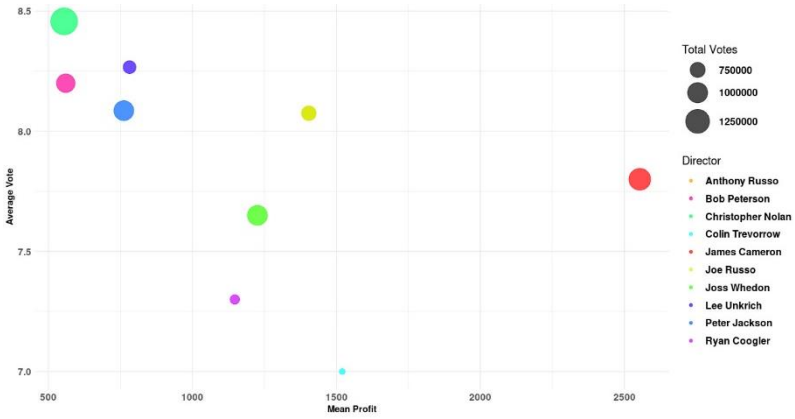

Figure 6 - Top Directors from the 'Adventure' genre

Analyzing the director's bubble graph (Figure 6), we can see that for this specific genre, James Cameron (red bubble) clearly stands out however Joe Russo (yellow bubble) would also be a good choice as he presents a medium profit and a good average vote.

(3) Writers

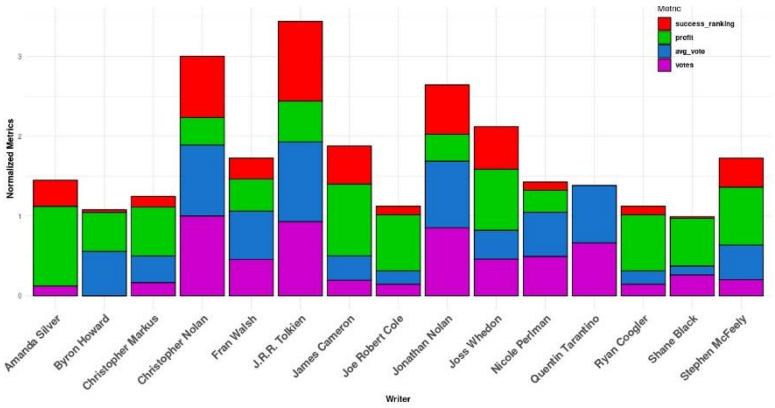

Figure 7 - Top Writers for the 'Adventure' genre

In the top writer's bar plot (Figure 7) we can see JRR Tolkien stands out followed close behind by the Nolan brothers – Cristopher and Jonathan. Collaborating with family members not only adds cohesiveness but also elevates creative vision, fostering efficiency—an integral combination for crafting impactful cinematic journeys. (Kozlowski and Ilgen, 2006).

b) Sci-Fi

(1) Actors

As it's seen in Figure 8, Mark Hamill's metrics are notably impressive in this genre, and Chris Hemsworth's similar performance suggests he could be an equally interesting choice as an actor for a Sci-Fi movie. The overlap between Chadwick Boseman and Michael B. Jordan lines indicates their collaboration in a movie that achieved strong metrics. This evidence suggests that this duo should be seriously considered for roles in Sci-Fi movies.

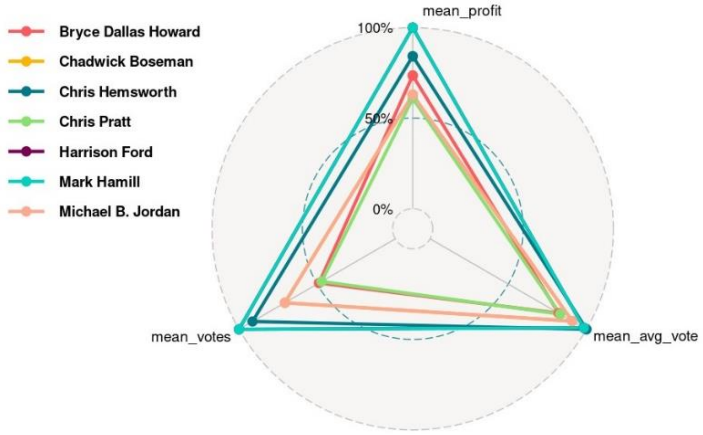

Figure 8 - Top Actors for the 'Sci-Fi' genre

(2) Directors

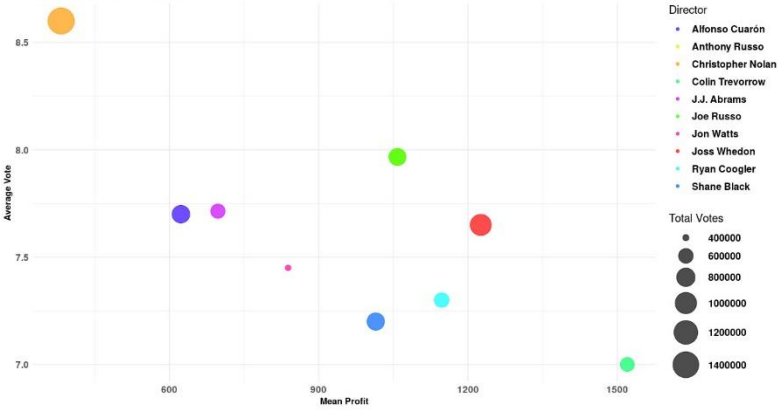

Figure 9 - Top Directors for the 'Sci-Fi' genre

Joe Russo and Anthony Russo, being brothers, share similar metrics in the Sci-Fi genre, suggesting their effective collaboration as directors. Christopher Nolan stands out with high average votes and a large number of votes, yet his mean profit for Sci-Fi movies is comparatively low. Conversely, Colin Trevorrow shows a strong profit margin, but his average vote count and number of votes fall short. A more balanced option appears to be Joss Whedon, who demonstrates solid performance across profit, average votes, and number of votes.

(3) Writers

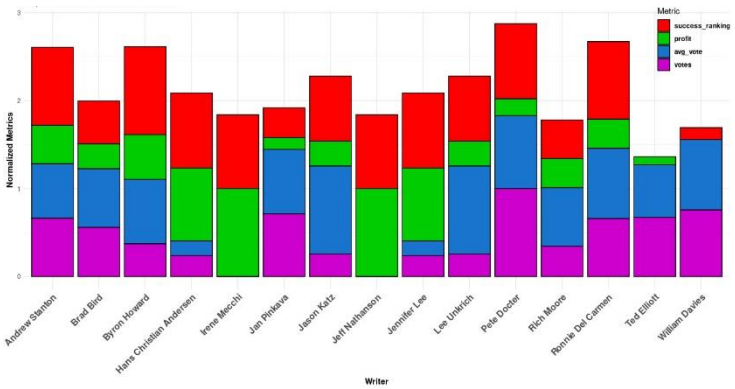

Figure 10 - Top Writers for the 'Sci-Fi' genre

Christopher Nolan notably excels as a writer in the Sci-Fi genre, outshining others, and his brother ranks as the third best writer in this category. Joss Whedon also emerges as a commendable choice, demonstrating high metrics across all evaluated fields.

c) Animation

(1) Actors

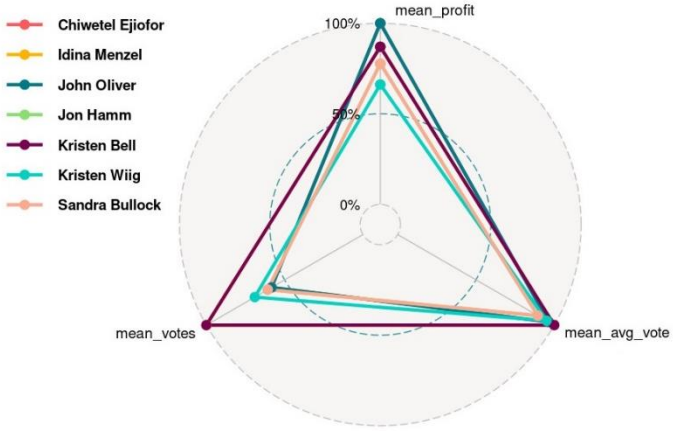

Figure 11 - Top Actors in the 'Animation' genre

While John Oliver's animated movies show amazing profits, they fall short in other metrics. Interestingly, John Oliver and Chiwetel Ejiofor share the same metrics in this genre, evidencing their joint participation in a successful movie, making them a duo worth exploring in future Animation movies.

Kristen Bell stands out as the optimal choice for Animation movies, outstanding in all metrics. Additionally, both Kristen Bell and Idina Menzel demonstrate matching metrics, indicating their potential as a formidable pairing in this genre.

(2) Directors

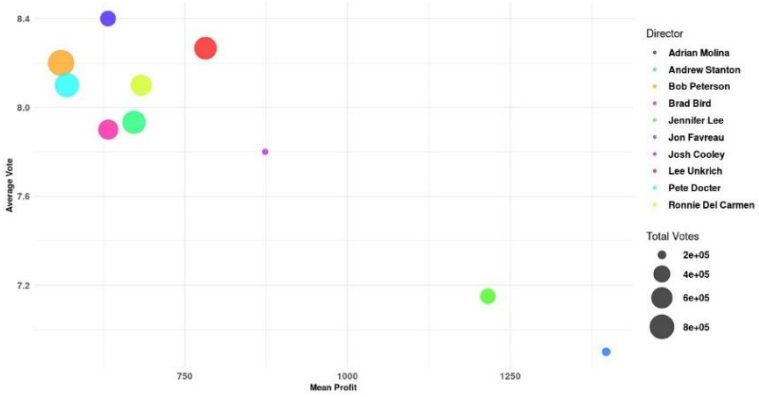

Figure 12 - Top Directors in the 'Animation' genre

The directors in the Animation genre predominantly cluster together, indicating a pattern of similar average votes and number of votes, although profits tend to be lower. Selecting a director with both a lower number of votes and average vote, as shown by two distinct circles on the right side of the graph, may not be advisable. Given this analysis, Lee Unkrich emerges as a commendable choice, boasting a high number of votes, a robust average vote, and moderate profit.

(3) Writers

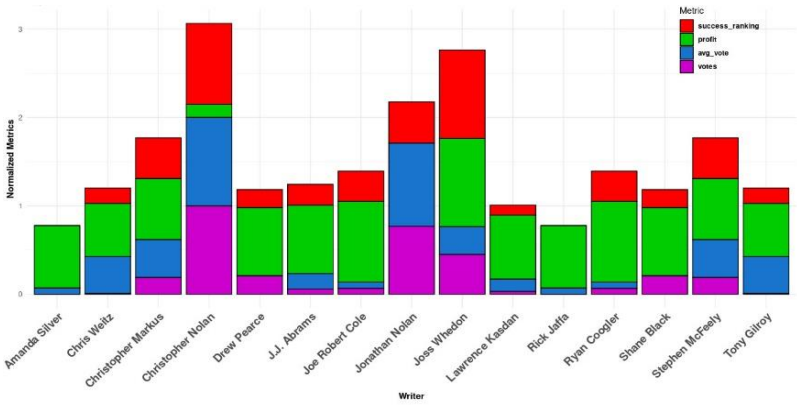

Figure 13 - Top Writers in the 'Animation' genre

The majority of writers in the Animation genre display similar metrics, making the choice challenging. Yet, Peter Docter distinctly stands out from the rest. Additionally, three writers — Ronnie Del Carmen, Byron Howard, and Andrew Stanton — exhibit identical metrics, indicating they are also promising choices in this field.

D. *Combinations*

To explore intricate dynamics shaping the financial and social success of a film, our investigation centered on the interplay between directors and writers, as well as on the dynamics among principal and secondary actors. Our analytical focus was anchored on the success rating.

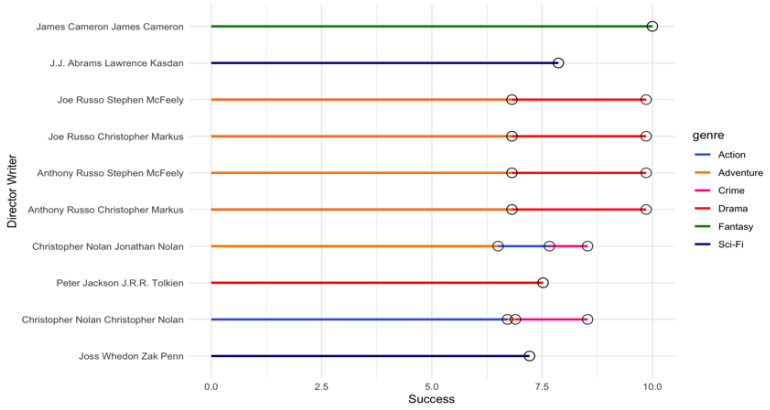

Figure 14 – 10 best combinations of director and writer based on the success rating, across all the dataset genres.

James Cameron, serving as both director and writer, excels with the highest success score. The Russo brothers, joined by writers McFeely and Markus, also secure impressive scores. Christopher Nolan consistently achieves high success across genres, whether working independently or in collaboration with his brother Jonathan as the writer. This evidence highlights the pivotal role of a director's creative control in a movie's success. The director assuming the role of the writer or collaborating with a familial connection enhances cohesiveness enhances creative vision, fostering efficiency—crucial elements for crafting successful cinematic experiences.

Figure 15 - 10 best combinations of main and secondary actor based on the success rating.

The most meaningful dynamics among the actor pairs are observed in the collaborations of Sam Worthington with Zoe Saldana in one dynamic, and Robert Downey Jr. with Chris Evans in another (Figure 15). Additionally, the noteworthy pairing of Christian Bale with Heath Ledger and Noel Appleby with Ali Astin. Essentially, Figure 14 underlines the pivotal influence that actor performances and dynamics can exert on the success of a film.

E. Language

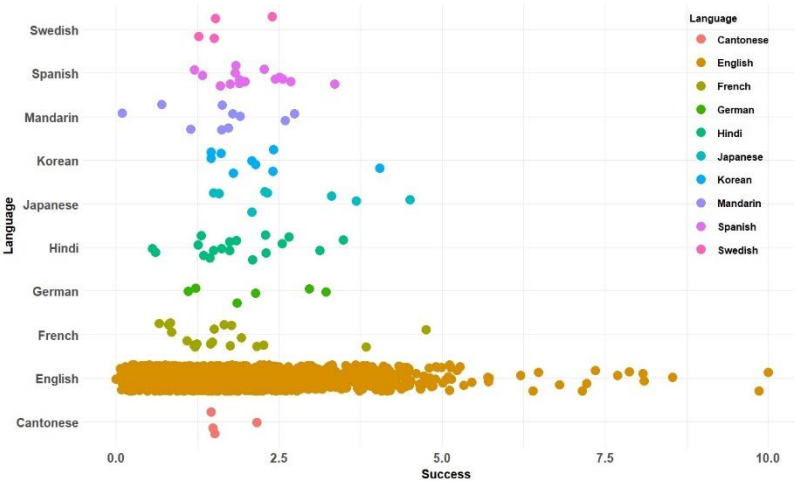

Figure 16 - Scatter plot of the success rating across different main languages

The abundance of English-language films doesn't inherently assure higher success, as most movies, regardless of language, hover between 0 and 5 on the success scale. Yet, films exceeding a success value of 5 are exclusively in English. This suggests that opting for English as the primary language tends to correlate with greater success, as illustrated by Figure 16.

F. Description

1) Most frequent words

Figure 17 highlights the words in the dataset's movie descriptions most strongly linked to success. The prominence of 'world' in the top spot aligns with our hypothesis that movies associated with an imaginative, created world (Sci-Fi) tend to exhibit a positive performance.

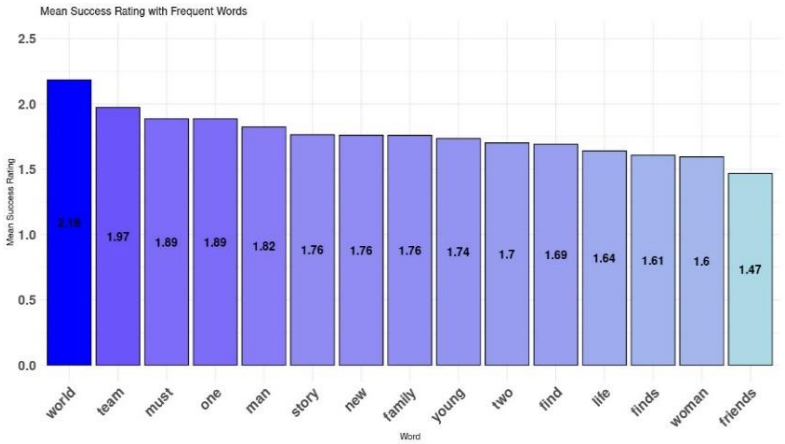

Figure 17 - Success rating of the most frequent words in the 'Description' column

2) Sentiment Analysis

By analyzing the sentiment of the description column, we can conclude that descriptions with a more negative or even neutral sentiment are generally more successful than descriptions with a more positive sentiment. This, coupled with the previous section on the most frequent words, could provide very valuable clues for developing a possible basic plot for a movie that revolves around a story in an imaginary world (usually associated with Sci-Fi movies) that ends up having a more negative outcome or a certain moment in the movie that might be sadder and have a greater impact on the movie's audience.

IV. Conclusion

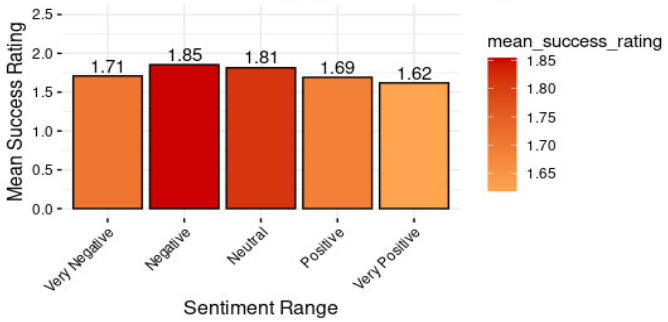

Figure 18 - Sentiment analysis of the 'Description' column based on the success rating.

Genre preferences and viewer choices are significantly influenced by biological factors, particularly age and gender, serving as crucial determinants. Age, in particular, shapes genre preferences along two dimensions: the balance between fantasy and realism and the level of arousal prompted by the genre.

Our hypothesis links film success to story lines of alternative fantastical worlds, observed in top genres, descriptions, and performances.

Seasonal genre trends indicate a nuanced blend of themes and temporal factors, aligning with demographic-based film consumption patterns.

Additionally, our findings underscore the significant impact of a director's creative control. Directing and writing, or collaborating with family, contribute can bring cohesiveness,

fostering increased creativity and efficiency for successful cinematic experiences.

So, here's a summary of the key points to take away from our visualizations, which could be useful information for a film production company in the future:

- In terms of the movie release month, it would be better to choose a time when younger populations are on school breaks (e.g. July for Animated movies in the case of children) as these months are usually more likely to be successful, as Figure 1 shows.
- In terms of preferred genres, it would be interesting to opt for genres that are getting more and more exposure and have shown the best results over the years, such as Adventure, Sci-Fi and Animation.
- In the case of the Adventure genre, Mark Hamill would be an excellent choice as lead actor, Chiwetel Ejiofor and John Oliver as a possible duo, James Cameron as director and JRR Tolkien as writer.
- In the case of the Sci-Fi genre, Mark Hamill and Chris Hemsworth would be good choices as lead actors, Chadwick Boseman and Michael B. Jordan as a good possible duo, Joss Whedon or the Russo brothers as directors, Cristopher Nolan as writer, who could be accompanied by his brother Jonathan given his good metrics, or Joss Whedon also as writer.
- In the case of the Animation genre, Kristen Bell stands out as an actress alongside Idina Menzel because they share the same metrics. John Oliver and Chiwetel Ejiofor would also be a great choice for the lead duo and in terms of director and writer, a solid choice would be Lee Unkrich and Peter Docter, respectively.
- James Cameron seems to be an excellent choice when he has the double job of writer and director. Cristopher Nolan is also an excellent choice, either alone or with his brother Jonathan writing the movie script.
- The language chosen for the film is suggested to be English, as it is the language with the greatest global reach and is generally more successful than others.
- Opting for a sadder description or plot of the movie, with more sentimental impact on the audience, along with setting it in a fantastical alternative world, indicates good results in achieving greater success.

V. REFERENCES

[1] Chen, W.-K. *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] Grodal, T., 2017. *How film genres are a product of biology, evolution and culture—an embodied approach*. Palgrave Commun 3, 17079;

[3] Marshall, P.,Dockendorff, M., Ibáñez, S., 2013. *A forecasting system for movie attendance,* Journal of Business Research, 66(10), pp.1800-1806;

[4] Packard, G., Aribarg, A., Eliashberg , J., Foutz, N. Z. 2016. *The role of network embeddedness in film success*, International Journal of Research in Marketing,33 (2), pp. 328-342

[5] Poor, H. *An Introduction to Signal Detection and Estimation*. New York: Smith, B. “An approach to graphs of linear forms (Unpublished work style),” Springer-Verlag, 1985,4.

[6] Wei,L.,Yang , Y. , 2022. *An empirical investigation of director selection in movie preproduction: A two-sided matching approach*, International Journal of Research in Marketing, 3(2), pp.888-906

[7] Young, G.O. “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.