

# Journal of Mixed Methods Research

## **Encapsulating Abstract Consequences of the Pandemic on the Film Industry through Neural Networks via Mixed Methods Analysis.**

Journal:	<i>Journal of Mixed Methods Research</i>
Manuscript ID	Draft
Manuscript Type:	Methodological/Theoretical Discussion
Manuscript Keywords:	Mixed methods, Deep learning, Big Data, Movie, Supervised learning

SCHOLARONE™  
Manuscripts

**Encapsulating Abstract Consequences of the Pandemic on the Film Industry  
through Neural Networks via Mixed Methods Analysis.**

**Abstract**

The field of mixed methods can benefit from utilizing various Deep Learning and Machine Learning informed techniques to encapsulate abstract concepts into quantifiable data. This article shows how the ‘ordinalization’ of different aspects of movies through Neural Networks can help investigate the cause of the revenue crash of film studios with excellent track records after the Covid-19 pandemic. Mixed methodology is enforced by considering the Video, Audio, Script, and other metadata pertaining to a movie. Apart from the aforementioned holistic case study, each aspect of the movie is also individually evaluated to highlight the settings of a successful movie. These examinations revealed subconscious biases that most successful movies exploit to generate high revenue.

**Keywords**–Mixed methods; Deep learning; Big Data; Movie; Supervised learning.

# Introduction

The film industry is massive, generating billions of dollars in revenue annually[1]. The industry has always been known for its ability to captivate audiences through a combination of compelling storytelling, stunning visuals, and engaging marketing. However, the COVID-19 pandemic has significantly impacted the industry, particularly the revenue generated from theatrical releases.

Despite the reopening of theatres in 2021, many successful film franchises and studios failed to produce any good films, which suggests that the pandemic's impact on the industry goes beyond just the closure of movie theatres. This raises questions about the underlying factors that contribute to a film's success, particularly in the post-pandemic era.

Therefore, this research seeks to investigate the impact of the pandemic on the film industry and its various aspects, including the production, distribution, and reception of films. The research questions that guide this study are:

- How to quantify the pandemic's impact on revenue?
- How do filmmakers design the film to exploit unexplained human characteristics?
- What footprint do good films leave?

These research questions will be addressed through a mixed-methods approach, utilizing quantitative and qualitative data sources to comprehensively understand the film industry's current state and future prospects.

This article begins with Literature Review, which discusses all the previous work around the specific algorithms discussed in the methodology section. The Data Collection section describes how the large dataset was created from different parts of the internet. The Methodology

describes all the algorithms used to generate results. In Discussions, chosen results that have significant meaning regarding the study are showcased.

## Related Work

For this study, research gaps have multiple meanings due to the enormous scope of the thesis. Thus, multiple research profiles are studied to identify current advancements and where the actual work needs to be done.

## Mixed Methods Analysis

In mixed methods analysis, both quantitative and qualitative data are given equal weight and are used to complement and validate each other. For example, qualitative data might provide a more in-depth understanding of a phenomenon's context or subjective experience. In contrast, quantitative data might be used to measure the prevalence or frequency of certain aspects of that phenomenon.

The process of mixed methods analysis involves integrating and synthesizing the results of both types of data to draw conclusions and make recommendations. This process can be complex and requires careful attention to issues such as data integration, analysis and interpretation, and triangulation of findings.

1  
2  
3 Creswell, J.W. (2003). Research design: qualitative, quantitative, and mixed methods approach.

4 This book provides an overview of different research designs, including qualitative, quantitative,  
5  
6 and mixed methods approaches, and discusses the strengths and weaknesses of each approach.  
7

8  
9  
10 This book covered the entire history of MMA up to 2003 when traditional statistics was used in  
11  
12 this domain. This research misses the functionality that an ML model can provide.  
13

14 Onwuegbuzie, A.J., & Leech, N.L. (2005). On becoming a pragmatic researcher: the importance  
15  
16 of combining quantitative and qualitative research methodologies. This article argues that  
17

18 combining qualitative and quantitative research methodologies can provide a more  
19  
20 comprehensive understanding of research problems. This aspect has been missing in the realm of  
21  
22 films as most of the research involved user reviews which fundamentally assume unbiased  
23  
24 reviews and ratings.  
25  
26

27  
28 Greene, J.C., & Caracelli, V.J. (2003). Making paradigmatic sense of mixed methods practice.  
29

30 This article discusses the theoretical and philosophical foundations of mixed-methods research  
31  
32 and provides examples of mixed-methods research in practice. This research highlights how data  
33  
34 science is not just using statistical models to summarize the data but using the author's  
35  
36 perspective as a critical perspective in evaluation.  
37  
38

39  
40 Creswell, J.W., & Plano Clark, V.L. (2007). Designing and conducting mixed methods research.  
41

42 This book provides a detailed guide to designing and conducting mixed-methods research,  
43  
44 including practical data collection, analysis, and reporting advice. This work showcases how data  
45  
46 collection, which is rarely elaborated, defines the entire trajectory of the project.  
47  
48

49 Morse, J.M. (1991). Approaches to qualitative-quantitative methodological triangulation. This  
50  
51 article discusses different approaches to combining qualitative and quantitative research methods  
52  
53  
54  
55  
56  
57  
58  
59  
60

and provides methodological triangulation examples. Triangulation is extremely important to avoid misunderstanding of analyzed data.

## Video Translation

Translating a video into data for statistical analysis depends on the nature of the video and the type of analysis you want to perform. For this study, Color Coding is a technique we propose that can be significantly useful in encapsulating the psychology of films.

The article "Complementary Color Wavelet: A Novel Tool for the Color Image/Video Analysis and Processing" presents a new method for color image and video analysis and processing. The authors propose using complementary color wavelet transforms, which can decompose color signals into different scales and orientations, similar to other wavelet transforms. The main advantage of this method is that it preserves more color information than other wavelet transforms, making it suitable for various color-based image and video processing applications. The authors provide several examples of how this method can be applied to tasks such as image denoising, image segmentation, and video compression. The results of the experiments show that the proposed method outperforms other existing methods in terms of color preservation and overall performance. The paper concludes that complementary color wavelet transforms can be valuable for color image and video analysis and processing.

Apart from creating a Color Coding technique, Image Segmentation has been used to identify a number of faces using OpenCV's Face Recognition library.

The paper "Robust Face Detection and Recognition using Image Processing and OpenCV" presents a method for detecting and recognizing faces using image processing techniques and the OpenCV library. The authors propose a four-stage approach, which includes face detection, face

alignment, feature extraction, and face recognition. The face detection stage uses the Haar cascade classifier to detect faces in an image or video stream. The face alignment stage aligns the detected face using the Affine transformation technique. The feature extraction stage uses the Local Binary Patterns (LBP) algorithm to extract features from the aligned face image. Finally, the face recognition stage uses the k-nearest neighbor algorithm to match the extracted features with a database of known faces.

## Audio Translation

In this study, Audio is translated to its corresponding notes to utilize the expansive libraries in music note analysis.

The paper titled "Converting MIDI notes to ABC notes in pencil code" by Heines (2016) presents a technique to convert MIDI (Musical Instrument Digital Interface) notes to ABC (A Better Chance) notes in Pencil Code. The author describes the process of converting the MIDI notes into a textual format that Pencil Code can read. The conversion process involves mapping the MIDI notes to their corresponding ABC notation and writing a program in Pencil Code to automate the conversion process.

## Script Sub-Setting

Scripts have been analyzed to understand the emotional timelines throughout the duration of the film. In the study, the script is converted to a sub-set of definitive words associated with emotions.

In their paper titled "Evaluating keyphrase extraction algorithms for finding similar news articles using lexical similarity calculation and semantic relatedness measurement by word embedding," Sarwar, Noor, and Miah (2022) evaluate several keyphrase extraction algorithms for finding similar news articles. The authors use both lexical similarity calculation and semantic relatedness measurement by word embedding to compare the effectiveness of various algorithms. They use two similarity measures to evaluate the algorithms' performance: lexical similarity and semantic relatedness. They use the Jaccard similarity coefficient for lexical similarity, and for semantic relatedness, they use the cosine similarity of word embeddings. The results show that combining the TextRank algorithm with semantic relatedness measurement using word embeddings outperforms other algorithms in finding similar news articles. The authors also provide a detailed analysis of the performance of each algorithm and the impact of various parameters on the performance.

Bhamare and Prabhu (2021) propose a supervised scheme for aspect extraction in sentiment analysis using a hybrid feature set of word dependency relations and lemmas. The authors note that aspect extraction is an essential step in sentiment analysis, as it helps to identify the specific aspects of a product or service that customers are expressing opinions about.

The authors use a dataset of hotel reviews to evaluate their proposed scheme. They first preprocess the data by tokenizing the text, removing stop words, and performing dependency parsing to identify the syntactic relationships between words. They then use a combination of dependency relations and lemmas as features for their model.

The authors compare the performance of their model with several baseline models and state-of-the-art models for aspect extraction. The results show that their proposed scheme outperforms other precision, recall, and F1 score models.



# Metadata Analysis

Metadata analysis refers to the process of analyzing and interpreting metadata, which are data that describe other data. Metadata can provide valuable insights into the characteristics and properties of the data they describe, such as the author, date of creation, location, format, and subject matter. Metadata analysis can be applied in various domains, including digital libraries, social media, and scientific research.

In scientific research, metadata analysis enhances the reproducibility and transparency of research results. Metadata can be used to document the research process, including the data sources, methods, and software used to analyze the data. Metadata analysis can also help to identify errors and inconsistencies in research data and to facilitate data sharing and reuse.

Long (2016) describes the process of building and searching the ECHO Early Cinema Credits Database using metadata analysis. The database contains information about over 34,000 films produced between 1895 and 1915, including the credits for each film, such as the director, cast, and crew.

The author explains how metadata analysis was used to extract information from film archives and other sources, such as trade journals and newspapers. The metadata was then encoded using a standardized format based on the Dublin Core metadata schema. The database was designed to allow for complex searches and queries based on various criteria, such as the name of a specific actor or the genre of a film. Overall, the paper demonstrates the potential of metadata analysis for excavating film history and for creating new tools and resources for film scholars and enthusiasts.

# Movie Analysis

Manual Movie analysis gives a comprehensive view of what factors a typical critic looks at to evaluate a film. This will help narrow down results and focus on what matters, not on lucky consistencies.

Gorinski (2018) presents an automatic approach to movie analysis and summarization based on natural language processing and machine learning techniques. The proposed method involves several stages: scene segmentation, shot boundary detection, keyframe extraction, and text summarization.

The author explains how the method can be used to generate a summary of a movie, which captures the main events, characters, and themes of the film. The summary is generated in the form of a textual description, which can be used for various purposes, such as movie reviews, recommendations, and marketing.

The author evaluates the method's effectiveness on a dataset of movie summaries using various metrics such as recall, precision, and F1-score. The results show that the method achieves high summarization accuracy and readability performance compared to other state-of-the-art methods.

Overall, the paper demonstrates the potential of automatic movie analysis and summarization for improving the efficiency and accessibility of movie-related information. The proposed method represents a promising direction for future research in the field of natural language processing and multimedia analysis.

# Data Collection

Each film's contribution is adhering copyright policy enforced by the movie. The movies are collected from respective distribution authorities in view of their copyrights policy. The AFX is extracted via sound intensity fluctuations. Reference [10] is a personal blog hosting subset (stills) of 250 films that are considered aesthetic and iconic. These are used to train the subsetting algorithm to reduce data footprint. Posters are collected through [3] data API. The score of the films is collected via authorized distributors of the films. The scripts of the movies are scraped from [9]. [2] is used to establish a corpus of the movies. This data is readily available as a large database. This, however, does not fulfill the data requirements. [3] API is yet again used to access other metadata related to the films, like reviews, ratings, etc. The movies' financial data was gathered by scraping [4]. [5] data was used to collect the search performance of chosen films. [6] data and [8] data were collected to count the number of instances of film instances available on their respective social platforms. All scrapings to collect data adhere to the scraping policy in their websites `/robots.txt` directory.

## Data Analysis Techniques

### Video to Snippets

A neural network is being used to reduce video data by eliminating non-essential frames. It takes 3 continuous frames of the film and returns a binary value that determines if the center frame

must be saved or not. This neural network is trained using [richards] data of 250 films. Only 10% of the total frames are reserved at the end of downsizing. The architecture of the neural network is described as follows:

<b>ALGORITHM 1: Video Subsetting Neural Network</b>
<i>model = Sequential_Network()</i>
<i>// Left Frame</i>
<i>model.add(Convolutional(64, (3, 3), activation='relu', input_shape=(1920, 1080, 3), disjoint=true))</i>
<i>model.add(MaxPooling((2, 2)))</i>
<i>// Current Fram</i>
<i>model.add(Convolutional(64, (3, 3), activation='relu', input_shape=input, disjoint=true))</i>
<i>model.add(MaxPooling((2, 2)))</i>
<i>// Right Frame</i>
<i>model.add(Convolutional(64, (3, 3), activation='relu', input_shape=input, disjoint=true))</i>
<i>model.add(MaxPooling((2, 2)))</i>
<i>// Dense Layers, Fully Connected</i>
<i>model.flatten()</i>
<i>model.add(Dense(128, activation='relu'))</i>
<i>model.add(Dense(1, activation='sigmoid'))</i>
<i>model.compile(optimizer='adam', loss='binary_crossentropy')</i>

## Fourier Transform

Music comprises of notes aka frequencies which must be extracted using Fourier Transforms. Identifying the spikes in transformed frequencies will give the note played at a particular time.

This can later be compared with adjacent notes to find closest scale of music being played. These scales can be analysed by associating them with performance of the film, or peculiarity of notes.

<b>ALGORITHM 2:</b> Audio Fourier Transform using Cooley Tukey Optimization		
<i>routine FFT(data):</i>		
	$n = \text{data.length}$	
	<i>if</i> $n == 2$ :	
		$\text{return } [\text{data}[0] + \text{data}[1], \text{data}[0] - \text{data}[1]]$
	$\text{even} = \text{FFT}(\text{data}[:,2])$	
	$\text{odd} = \text{FFT}(\text{data}[1:,2])$	
	$\text{ret} = []$	
	<i>for</i> $k$ <i>in</i> $\text{range}(\text{floor}(n/2))$ :	
		$\text{factor} = \text{exponential}(-2j * \pi * (k/n))$
		$\text{ret}[k] = \text{even}[k] + \text{factor} * \text{odd}[k]$
		$\text{ret}[k + \text{floor}(n/2)] = \text{even}[k] - \text{factor} * \text{odd}[k]$
	<i>return</i> $x$	

## White Noise Detection

Detecting white noise can be achieved by rapidly monitoring x and y axes of the fourier transformed audio for specific frequencies. This can be done using the following algorithm.

**ALGORITHM 3:** White Noise Detection

*audio = #file\_read*

*cutoff\_frequency = 1000*

*freq, psd = audio.welch() // Computing Power Spectral Density*

*high\_psd = mean(psd[freq > cutoff\_frequency])*

*total\_psd = mean(psd)*

*Ratio = high\_psd / total\_psd*

*threshold = 0.9*

*if ratio > threshold:*

*#mark*

*else:*

*#de-mark*

# Ordinalisation Neural Network

This neural network’s primary purpose is to overfit the network to understand underlying weightage of disjoint input nodes. The Mixed Methods aspects can be quantified here by having 4 distinct aspects of films. Video snapshots using CNN, Audio after FFT using LSTM, Script words using LSTM, and Reviews using LSTM are connected to 64, 32, 16, 1 Deep Neural Network in disjoint manner.

**ALGORITHM 4:** Ordinalisation Neural Network

```
model = Sequential_Network()
```

```
model.add(VGG16([32, 64], activation='relu', input=(1920, 1080, 3)), disjoint=true)
```

```
model.add(LSTM(activation='relu', input=(n, 1)), disjoint=true)
```

```
model.add(LSTM(activation='relu'), disjoint=true)
```

```
model.add(LSTM(activation='relu'), disjoint=true)
```

```
model.flatten()
```

```
model.add(Dense([64, 32, 16], activation='relu'))
```

```
model.add(Dense(1, activation='sigmoid'))
```

```
model.compile(optimizer='adam', loss='binary_crossentropy')
```

## Other Algorithms

- Genre Informed Word Rank Subsetting from Script: The pretrained google GoEmotions dataset has been used to determine if a word is appropriate for the genre of the film.
- Color Grade Extraction: The color grade is extracted into one of four categories: Blue, Green, Orange, and Violet.
- Poster Face Detection: Using Open CV Haarcascade presets, posters can be analyzed to find number of faces in the poster.
- Content Overlap Measurement: This is implemented using frame hashing to find quantity of content overlap of videos.
- Film Review Emotion Classifier: This uses GoEmotions classifier to correlate reviews with emotion.

- AFX Intensity Mapper: The Audio Effects intensity can help map impactful times of a film. This is found by calculating mean and median of all audio intensity levels. The top 20% loud impacts with sinusoidal peak are marked as high AFX intense scenes.

## Results & Discussion

The following findings were observed to have a significant impact on the study's objective after performing the mixed methods analysis methods discussed earlier. Each methodology resulted in multiple quantitative outputs, which were filtered using qualitative analysis.

### Money is blue, not green.

The visual analysis of films has revealed something very peculiar regarding the color grading of the movie. Typical high-production quality film undergoes color grading to have a personalized tone for the film. When this tone revolves around *blues*, the film has a high chance of earning big revenue. This can be linked to humans associating blue with coziness. Other significant color grades like green, red, and violet also formed clusters w.r.t performance of the film.

Insert Figure 1 here.

Figure 1: Correlation of Revenue with color palette's base color.



Films that use unnatural tones in their color palette also tend to rack higher than other blues. Films Avatar, Avatar 2, and Blade Runner 2049 have proven this theory. The Avatar franchise has niche blues and racked the top 5 all-time grossing. Blade Runner 2049 used extreme oranges and, unfortunately, gave a loss. So the uniqueness of the color has some level of influence on performance. Nevertheless, they tend to imply higher production quality meaning a higher budget.

## Reviews work at extremes.

Film reviews and ratings have been the basis of the audience's ability to qualify for a film before watching it. However, film reviews and ratings only seem to show good or bad films significantly.

Insert Figure 2 here.

Figure 2: Good Film ratings, Bad Film Ratings, Avg Film Ratings.

Moreover, even the decisive ratings do not indicate a significant correlation to revenue. This inconsistency in ratings is due to the time factor. Re-releases sometimes boost the ratings of the films. Moreover, award shows tend to bloat the actual ratings of the film.

# Ears are easiest to exploit.

Converting movie scores to notes has uncovered something very bizarre. Any musical piece employs many chord progressions, and with regard to films, these chords are prolonged over a long time to produce what are known as “swells.” These swells are typically soothing to ears, but when swells are used in a unique way called *Lydian* mode, where a semitone moves up the fourth note of any scale, the score tends to be associated with a successful film, or in rare cases, composed by a successful composer. All movies analyzed have some form of Lydian mode, and successful films tend not to overuse it and use it only in the film's climax.

For instance, the famous composer James Newton Howard sparingly uses the Lydian notes in the movie *Unbreakable*. Also, he plants smaller pieces of the notes at the beginning of the film. This not only induces some form of familiarity, but when the actual notes occur, there is some form of revelation to the unheard notes, and the audience feels awe. It is important to note that this film had shallow marketing and budget, yet it amassed immense revenue.

# A signature is Essential.

It is difficult to quantify the uniqueness of any movie’s soundtrack. Franchises intentionally share large parts of the score to evoke familiarity, and notes often are the same. The only changing factor is the instrument and synthesizers. Thus any movie innovating with their synthesizer has amassed greater revenue.

## C Major for the win.

Films that incorporate famous artists to write pop songs for the film have shown a trend of developing the music in C major and G major scales. This is common practice for any pop artist. However, the peculiar correlation is when this song is used. The movie franchise James Bond uses all these pop songs initially, while all other movies use them during credits. This can be associated with them leaving the theatre in a brighter mood and has definitely led to reviewing bias.

## Piracy does not kill the box office.

Contrary to popular misconception, piracy of films has no correlation with the revenue a film generates. However, interesting insight has been into what is more often pirated. Typically piracy is significantly higher in TV Shows than compared to movies.

## Concept of *Basal* audience.

Movie revenue for franchises has shown a peculiar pattern. Every decade, a franchise seems to have a base audience who are willing to buy the tickets, but this basal audience seems to fade away quickly at the end of the decade. An even more interesting pattern is found when this same franchise takes a longer break in between producing a new film. The basal audience reemerges to

make the film more successful. This trend is seen in the X-Men franchise in the 2000s and 2010s. Marvel from 2010 to 2020.

## Experience matters a lot.

The experience of *cinema* is very crucial. We can estimate how often impactful audio effects are used by analyzing how the film’s non-score and non-vocal audio, aka AFX, is distributed in frequencies. This can roughly tell us how well the movie is able to capture the audience’s attention. This factor has been used less frequently but, when used, has shown great box office success. For Instance, the movies Interstellar and Top Gun: Maverick are the highest-grossing films of their release years; their high-quality AFX seems to be a contributing factor that other films fail to produce.

## Reddit is more powerful than Twitter.

Social media is typically targeted to increase hype among the audience. This is done by releasing special teasers and creating *#Hashtags* that try to embed the film’s name in the audience’s mind. Reddit seems to be more efficacious in amassing an audience. Almost all films with subreddits made more money than those using other platforms.

## Factoring *Word of mouth*.

It is very difficult to factor word of mouth into something that has a global impact. But Google trends correlate to word of mouth very well. Take Avengers: Endgame, for instance; this has garnered more searches before and after but more after the release. And it has gained a lot of money. The film Arrival is not a franchise like Endgame, yet word of mouth gave this film exponentially higher traction after its release and gave it a profitable season. Whereas the film Babylon had very high hype yet after its release instantly died. This probably translates to people hyping up a movie before its release causing high expectations, which clearly the film wouldn't be able to deliver. That is the curse of the subjective opinion of trailers and posters. Whereas films that quietly promoted the movie did very well.

## A 5\$ Poster for the rescue.

It is shocking how significantly impactful a simple movie poster is. Not only do posters give the audience a glimpse into the film, but they also attract the common audience and coerce them to watch the film competing in the theatre. A good poster does not need to have layers of subliminal information; it must only need to have a collage of ALL starlist cast members of the film. Fewer faces do not indicate poor performance, but posters with more than 5 faces have excellent box office revenue. There are few exceptions, but this theory weirdly holds regardless of the film's genre, script, or release date.

# China controls big Hollywood studios.

It is no doubt that Chinese revenue is more than domestic. If a film gets approved to be released in China, it is with no doubt that the film will generate profit. But the restrictions in China are also very severe. And Chinese Film Board requires films to adhere to their country’s socialistic ideas. This, in turn, causes massive creative differences and restricts the freedom of Hollywood.

# Misleading/Controversial trailers amass interest.

The number of movie trailer comments is correlated to the amount of confusion the trailer creates. This controversy is necessary to keep the audience hooked until the film gets released. However, this interest has no significant impact on revenue due to the number of comments being extremely small as compared to the basal audience.

# More trailers, the merrier.

There is a unique correlation between the number of trailers to the revenue it generates. Films that are marketed with multiple trailers generate significant revenue, **but** only if the trailers overlap more than 80% of the video content.

## White Noise.

Highly grossing films have comparatively more white noise. And it can be said that films with long sequences with only white noise or binaural frequencies soothe the audience into feeling 'home' like the blue color palette theory.

## Cause of Covid-19 Collapse

- During the pandemic, revenue generated from OTT platforms skyrocketed.
- This caused the film production studios to launch their own counterpart.
- During this time, audiences became familiar with various genres that weren't typically launched at theatres.
- The audience was also recognizing more unpopular actors.
- After the relaxation of the curfew, the theatres only released films that were already queued before the lockdown, and these were old films that didn't interest the audience as much as the OTT variety.
- This collapse was due to poor timing of film schedules.
- But now, 2 years after the first relaxation, the OTT crew got jobs in big film studios and actively produced highly capable films.

Insert Figure 3 here.

Figure 3: Similarity between successful post-pandemic theatre films and OTT films during the pandemic.

These graph nodes are from a script that quantifies a film’s medium of quality and rates between 1 to 100. We can use these same neural network weights to understand how audience taste has changed before and after the pandemic.

Insert Figure 4 here.

Figure 4: Before and After pandemic nodes.

We can see how audience trusts in reviews has decreased significantly. This is probably owing to them discovering many underrated films during the pandemic.

- This shows how Hollywood is adapting to this new palette of the audience.
- The impact of familiarity of the audience with previously unpopular actors has also played a role in boosting a film’s success.
- Arguably, the quality of films has increased with natural diversity in the cast, and this is a silver lining to new and healing Hollywood.



# Conclusion

The effectiveness of the methods proposed can be estimated from the results. Incorporating the broad area of Artificial Intelligence into Mixed Methods Analysis can significantly help find meaning from abstract ideas. This will help in bridging the gap between qualitative and quantitative analyses. The use of overfitting brought a quantitative basis for understanding the effects of the pandemic. Machine Learning techniques made this process easier by filtering the data and reducing human effort.

Looking forward, there are several areas where future work could enhance the multi-modal approach to film analysis. One important avenue for future research is the development of robust AI tools that can analyze the actual contents of frames. While some progress has been made in this area, current technology is still limited. There is a significant need for more advanced tools that can extract meaningful insights from visual data.

Another area of interest is the analysis of actual conversations and the creation of dialog networks. By understanding the patterns and structures of dialog in successful films, we can gain insights into how to create more engaging and memorable characters and stories.

Additionally, the influence of film critique and the types of movies that command fanart are important areas for future research. By analyzing the reception of films by critics and fans, we can better understand what drives success in the industry and how to appeal to different audiences.

The impact of YouTube hype and other channels that discuss and promote films is also an area that warrants further exploration. By analyzing the reach and impact of these channels, we can gain insights into how to market and promote films to different audiences effectively.

Finally, the impact of awards such as the Oscars and Bafta on the success of films is an important area for future research. By analyzing the patterns and trends in award-winning films, we can gain insights into what factors contribute to success in the industry.

Overall, our project provides much scope for creativity and innovation in the field of film analysis. As technology advances and new data sources become available, future work has great potential to enhance our understanding of what makes films successful and how to create engaging and impactful stories.

## References

[1] "Global Movie Production & Distribution Industry: Industry Market Research Report". IBISWorld. August 2018. Retrieved 22 January 2019.

[2] INTERNET MOVIE DATABASE, <https://www.imdb.com/>, Last Accessed, October 22, 2023.

[3] THE MOVIE DATABASE, <https://www.themoviedb.org/>, Last Accessed, October 22, 2023.

[4] BOX OFFICE MOJO, <https://www.boxofficemojo.com/>, Last Accessed, October 22, 2023.

[5] GOOGLE TRENDS, <https://trends.google.com/>, Last Accessed, October 22, 2023.

[6] TWITTER, <https://www.twitter.com/>, Last Accessed, October 22, 2023.

- [7] SPOTIFY, <https://www.spotify.com/>, Last Accessed, October 22, 2023.
- [8] REDDIT, <https://www.reddit.com/>, Last Accessed, October 22, 2023.
- [9] Open Source, THE INTERNET MOVIE SCRIPT DATABASE, <https://imsdb.com/>, Last Accessed, October 22, 2023.
- [10] Evan E. Richards, BLOG, <https://www.evanerichards.com/>, Last Accessed, October 22, 2023.
- [11] Creswell, J. W. (2003). Research design: Qualitative, quantitative, and mixed methods approaches (2nd ed.). Sage Publications.
- [12] Plano Clark, V. L., & Creswell, J. W. (2010). Understanding research: A consumer's guide. Pearson.
- [13] Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, 8(5), 375-387.
- [14] Greene, J. C., & Caracelli, V. J. (2003). Making paradigmatic sense of mixed methods practice. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 91-110). Sage Publications.
- [15] Collins, K. M. T., & O'Cathain, A. (2009). Introduction: Ten points about mixed methods research to be considered by the novice researcher. *International Journal of Multiple Research Approaches*, 3(1), 2-7.
- [16] Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- [17] Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Sage Publications.

[18] Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(2), 120-123.

[19] Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97-113.

[20] Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Sage Publications.

[21] Chen, Y., Li, D., & Zhang, J.Q. (2019). Complementary Color Wavelet: A Novel Tool for the Color Image/Video Analysis and Processing. *IEEE Transactions on Circuits and Systems for Video Technology*, 29, 12-27.

[22] Bhat, A., Jha, R., & Kedia, V. (2022). Robust Face Detection and Recognition using Image Processing and OpenCV. 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), 1273-1278.

[23] Heines, J.M. (2016). Converting MIDI notes to ABC notes in pencil code. *ACM Inroads*, 7, 84 - 84.

[24] Sarwar, T.B., Noor, N.K., & Miah, M.S. (2022). Evaluating keyphrase extraction algorithms for finding similar news articles using lexical similarity calculation and semantic relatedness measurement by word embedding. *PeerJ Computer Science*, 8.

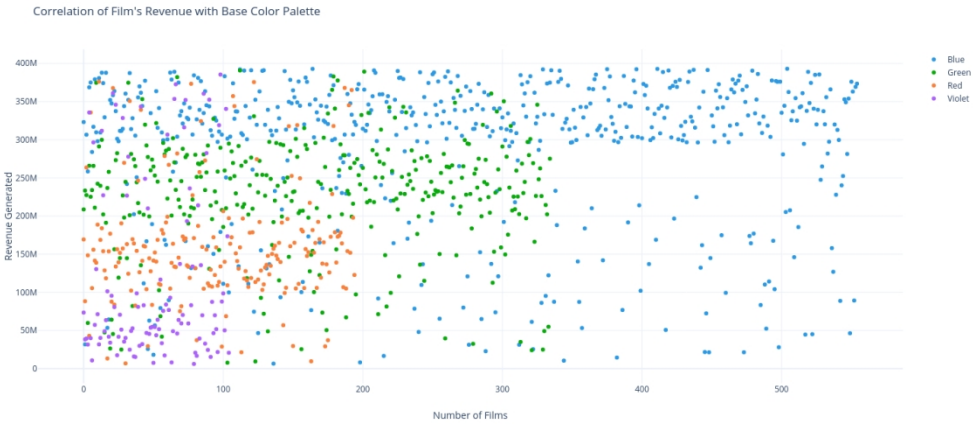
[25] Bhamare, B.R., & Prabhu, J. (2021). A supervised scheme for aspect extraction in sentiment analysis using the hybrid feature set of word dependency relations and lemmas. *PeerJ Computer Science*, 7.

[26] Long, D. (2016). *Excavating Film History with Metadata Analysis: Building and Searching the ECHO Early Cinema Credits Database*.

[27] Gorinski, P.J. (2018). *Automatic movie analysis and summarisation*.

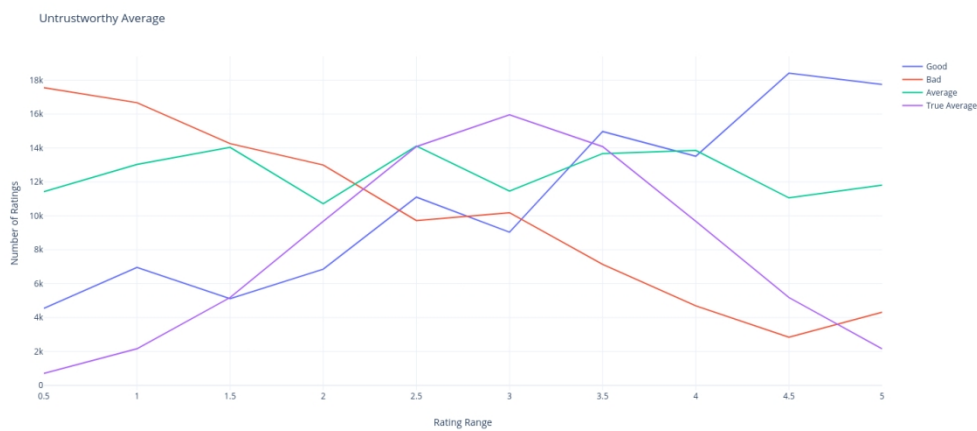
[28] Kantheti, Bhargav, (2023, March 22), LARGE MOVIE DATABASE,  
<https://www.kaggle.com/datasets/bharxhav/large-movie-database>, Last Accessed, March 22,  
2023.

For Peer Review



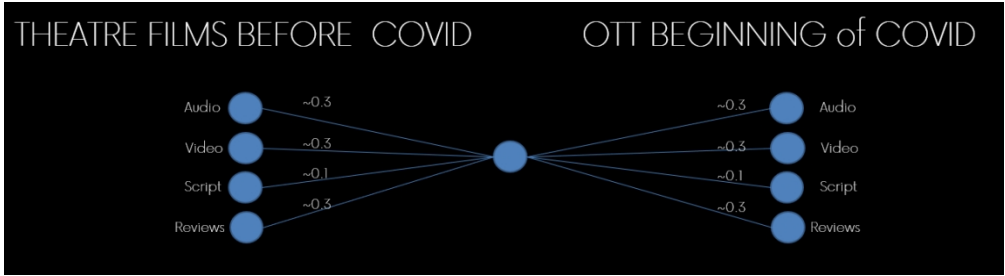
Correlation of Revenue with color palette's base color.

514x235mm (72 x 72 DPI)



Good Film ratings, Bad Film Ratings, Avg Film Ratings.

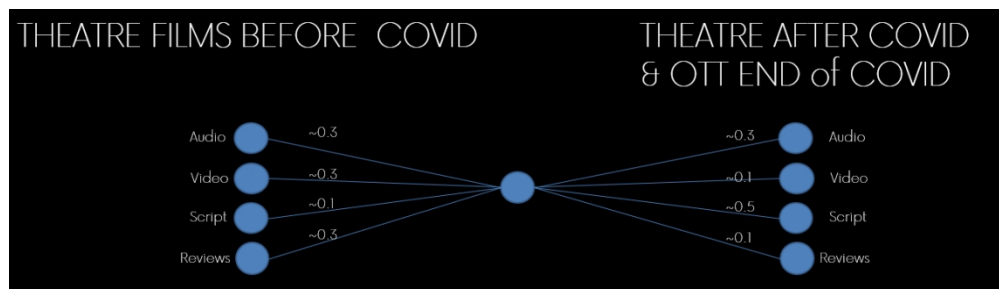
514x235mm (72 x 72 DPI)



Similarity between successful post-pandemic theatre films and OTT films during the pandemic.

868x237mm (38 x 38 DPI)





Before and After pandemic nodes.

862x247mm (38 x 38 DPI)