

Web Scrapping: Metacritic

Introduction

The film industry is the major entertainment in the world with increasing number of movies released each year across a variety of genres. In this project, the horror and comedy genres were selected because they are opposite to each other so making it easier to observe the differences between them. This is because both genres differ significantly in terms of tone, themes and critical reception by audiences. Understanding the different between the genre can provide a valuable insight to the audience preference, production trend and cultural impact.

The Metacritic website was chosen in this web scrapping analysis. Metacritic is review website that covers from games, movies, music album and tv shows where the user can give a scores that contribute to the Metascore. Movie was chosen due to relevance to public interest and trends because analyzing movies helps understand cultural shifts, preferences and genre evolution over time. In this project, movies web scrapping was done with attributes such as year of release, genre, streaming services and release type are available for selection. All years, streaming services and release types was selected while only comedy and horror genres were chosen for comparison. This analysis includes both quantitative metrics for metascore distributions and qualitative assessments for sentiment analysis from the movie descriptions.

The objectives of this report are:

1. To compare the temporal release patterns of horror and comedy films by year, month, and day.
2. To evaluate the distribution of metascores, highlighting differences in critical reception between the two genres.
3. To forecast the distribution of metascores and movies produce for next 5 years.
4. To explore textual trends using word frequency, TF-IDF analysis, and sentiment scoring from movie plot descriptions.
5. To derive insights on audience and critical preferences, especially how metascore trends vary over time and across weekdays or weekends.

By combining web scraping, text mining and statistical analysis, the overview of how the horror and comedy films performs in the entertainment industry.

Methodology

This project combines web scrapping, data preprocessing, text mining, exploratory data analysis (EDA) and time series forecasting to compare between both genres. Here the flow of the analysis.

First, data was scraped from Metacritic website to extract movie attributes such as title, release date, description and metascore. There are 128 pages of horror movies and 338 for comedy movies was scrapping on the website. The scraped data was cleaned and split into day, month and years for relevant analysis later on.

Then, data integration and cleaning was done to easier the comparison. New genre column was created to facilitate comparison. Missing data inside the metascore was imputed with average value based on their genre.

Next, text preprocessing and analysis where it was conducted by tokenization text into words, removal of stop words and elimination of numeric tokens. Lemmatization or stemming is not applied in order to preserve the richness in analysis especially since description of the movie is very short. There are several type of text analysis were performed which are word frequency, word clouds, Term Frequency-Inverse Document Frequency (TF-IDF) and sentiment analysis.

Furthermore, exploratory data analysis (EDA) was done to produce a release patterns, metascore distribution, missing data analysis and temporal metascore trends. Lastly, forecasting was done to built two time series model which are total movie count per genre by year and average metascore per genre by year.

Analysis and Discussion

The attributes in this dataset or web scrapping are title, date, description and metascore. Title is title of the movie and date is date movie release. Movie dates are split into day, month and

year. Figure 1 and Figure 2 shows some data extract of the first 3 pages from web scrapping for horror and comedy movies.

	title	date	description	metascore	day	month	year
1	Garden of Eden	Jun 20, 2025	The wealthy and devout Eden family lives by the word of Go...	NA	20	June	2025
2	Alma and the Wolf	Jun 20, 2025	After a violent animal attack, paranoia spreads through Spir...	NA	20	June	2025
3	Eye for an Eye	Jun 20, 2025	When she becomes a bystander to an unforgivable act of vi...	NA	20	June	2025
4	Found Footage: The Making of the Patterson Project	Jun 20, 2025	A crew of documentarians chronicle an amateur filmmaker's...	NA	20	June	2025
5	28 Years Later	Jun 20, 2025	It's been almost three decades since the rage virus escaped ...	76	20	June	2025
6	Soul Reaper	Jun 17, 2025	Nightmares of violent deaths torment Respati—until he real...	NA	17	June	2025
7	House of Abraham	Jun 13, 2025	Looking for an end to her suffering, a woman checks into th...	NA	13	June	2025
8	Gator Lake	Jun 13, 2025	Lurking in Florida's deadliest lake, a bloodthirsty alligator go...	NA	13	June	2025
9	Best Wishes to All	Jun 13, 2025	A young woman's visit to her grandparents' home leads to t...	70	13	June	2025
10	The Other	Jun 13, 2025	After years of infertility, Robin and Daniel adopt Kathelia, a ...	NA	13	June	2025
11	Bark	Jun 13, 2025	Nolan Bentley wakes up in the middle of a remote forest, tie...	NA	13	June	2025
12	Bleeding	Jun 10, 2025	In a world where vampire blood is harvested as a drug, two ...	NA	10	June	2025
13	Resurrection Road	Jun 6, 2025	An elite squad of six soldiers, led by ex-slave Barabbas, is se...	NA	6	June	2025
14	The Ritual	Jun 6, 2025	When the Church sanctions the exorcism of Emma Schmidt, ...	33	6	June	2025
15	Dangerous Animals	Jun 6, 2025	When Zephyr, a rebellious surfer, is abducted by a shark-ob...	65	6	June	2025
16	I Don't Understand You	Jun 6, 2025	Dom (Nick Kroll) and Cole (Andrew Rannells), a couple on th...	57	6	June	2025
17	Bogieville	Jun 3, 2025	A young couple on the run come across a derelict trailer par...	NA	3	June	2025
18	Bring Her Back	May 30, 2025	A brother and sister uncover a terrifying ritual at the seclude...	75	30	May	2025
19	Pretty Boy	May 27, 2025	Suffering from blindness after a botched laser surgery, Holly...	NA	27	May	2025
20	A Hard Place	May 27, 2025	After a botched job leaves a trail of bodies behind, a gang o...	NA	27	May	2025

Figure 1 Data from web scrapping for first 3 pages for horror movies

	title	date	description	metascore	day	month	year
1	Semi-Soeter	Jun 20, 2025	Power couple Jaci and JP find themselves in a bumpy predic...	NA	20	June	2025
2	Shanghai Blues	Jun 20, 2025	In 1937 after The Second Sino-Japanese War breaks out a s...	70	20	June	2025
3	Noah's Arc: The Movie	Jun 20, 2025	Fans and new audiences will reunite with the beloved chose...	NA	20	June	2025
4	HI-FIVE	Jun 20, 2025	In the dead of night, a body donated for organ transplants ...	NA	20	June	2025
5	Don't Tell Larry	Jun 20, 2025	After lying to her new eccentric co-worker Larry about a co...	NA	20	June	2025
6	The Queen of My Dreams	Jun 20, 2025	Against the backdrop of a shared obsession with Bollywood...	64	20	June	2025
7	Found Footage: The Making of the Patterson Project	Jun 20, 2025	A crew of documentarians chronicle an amateur filmmaker's...	NA	20	June	2025
8	Everything's Going to Be Great	Jun 20, 2025	There's no business like show business — for Buddy and Ma...	54	20	June	2025
9	Elio	Jun 20, 2025	A space fanatic with an active imagination and a huge alien ...	66	20	June	2025
10	Bride Hard	Jun 20, 2025	Sam (Rebel Wilson) is a world-class secret agent... and a pre...	23	20	June	2025
11	KPop Demon Hunters	Jun 20, 2025	When they aren't selling out stadiums, Kpop superstars Ru...	75	20	June	2025
12	Maxxie LaWox: Drag Super-shero	Jun 17, 2025	A shy barista's life transforms when a magic wig turns him i...	NA	17	June	2025
13	Band on the Run	Jun 17, 2025	At the height of the Detroit Garage Rock movement in 1999...	NA	17	June	2025
14	Materialists	Jun 13, 2025	A young, ambitious New York City matchmaker finds herself...	71	13	June	2025
15	How to Train Your Dragon	Jun 13, 2025	On the rugged isle of Berk, where Vikings and dragons have...	61	13	June	2025
16	Gator Lake	Jun 13, 2025	Lurking in Florida's deadliest lake, a bloodthirsty alligator go...	NA	13	June	2025
17	Deep Cover	Jun 12, 2025	Three improv actors are asked to go undercover by the poli...	67	12	June	2025
18	Cheers to Life (Viva a Vida)	Jun 11, 2025	Jéssica (Thati Lopes) is a simple young woman, dedicated to...	NA	11	June	2025
19	Oxy Morons	Jun 10, 2025	Wayne hates his logistics job. So when an underworld chieft...	NA	10	June	2025
20	Serpentine Pink	Jun 10, 2025	Set in the desert-weird wilds of Joshua Tree, California, Serp...	NA	10	June	2025

Figure 2 Data from web scrapping for first 3 pages for comedy movies.

First three pages of the website, there are 72 observations for each attribute for both genres. However, there are 40 missing values for comedy and 45 missing values for horror at Metascore. Then the missing value ratio for both genres are high which at 0.556 for comedy and 0.625 for horror which is majority of the data for first three pages. This is due to these

movies recently release. he first three pages consist entirely of newly released films, with the most recent release dated 20 June 2025 for both genres. The last movie on page three was released on 11 April 2025 for comedy and 14 March 2025 for horror. The high number of missing Metascore values is likely due to the fact that many audiences have not yet watched or reviewed the films.

Then, boxplot and t-test were done to see the outlier and significant difference between both genres. Figure 3 shows Metascore boxplot for first three pages.

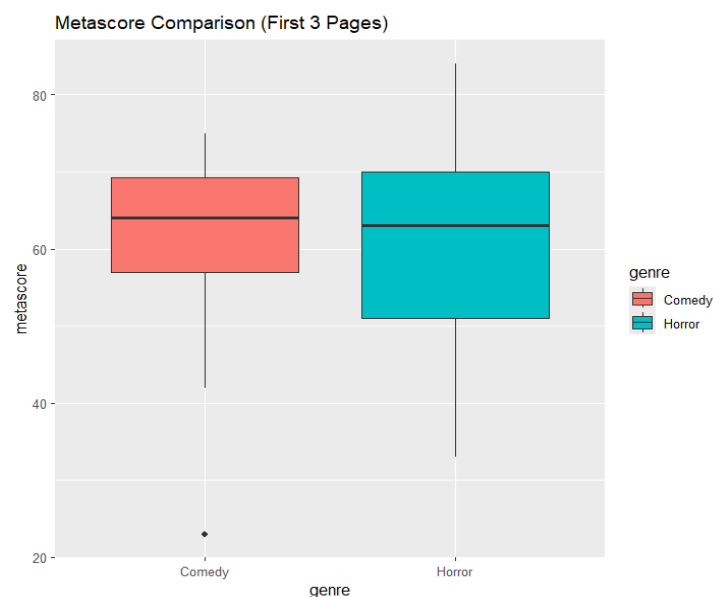


Figure 3 Metascore boxplot for first three pages.

The first three pages doesn't show any outlier for both genres and the average Metascore for Comedy is 62 and Horror is 60. From the t-test, p-value is 0.6757 which shows there are no significant differences between both genres Metascore. This shows the user in the Metacritic tend to rate Comedy and Horror movies similarly for the newest movies. It indicating no clear preference or bias in Metascore ratings between the two genres in the movie listed in the first three pages. Figure 4 below shows top 10 words for both genre in the descriptions.

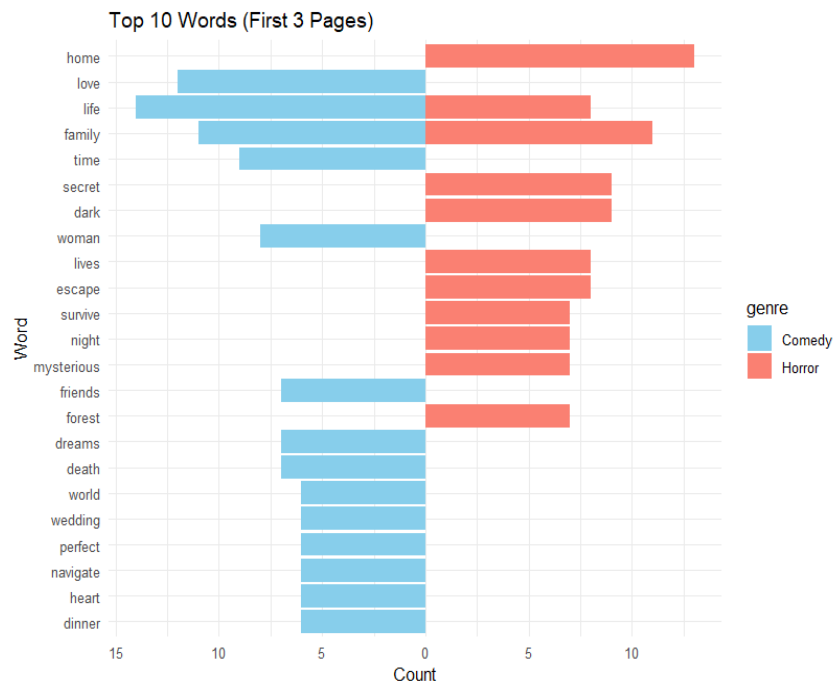
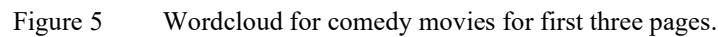


Figure 4 Top 10 words for both genres in descriptions.

The top 10 words for each genre in descriptions reflects their distinct themes. In Comedy, the frequency words was found are life (14 times), love (12 times) and family(11 times) which suggesting stories focus on relationship, personal experiences and emotional warmth. The words like dreams add a inspirational tone while wedding and dinner words show a social events which usually used as humorous setting. In contrast, Horror is dominated by home (13 times), dark (11 times), secret (9 times) and escape (8 times) shows the emotion such as fear, mystery and survival. Term like forest and night shows the isolation and suspense. Even though both genres share words like family and life their context is different like comedy uses them light-heartedly while horror present them in lens of danger and tensions. Figures 5 and Figure 6 shows the wordcloud for comedy and horror movies.



6

baby, wedding, couple and musician which suggest a relaxed and slow-pace movie. Figure 7 below shows sentiment analysis for both genres.

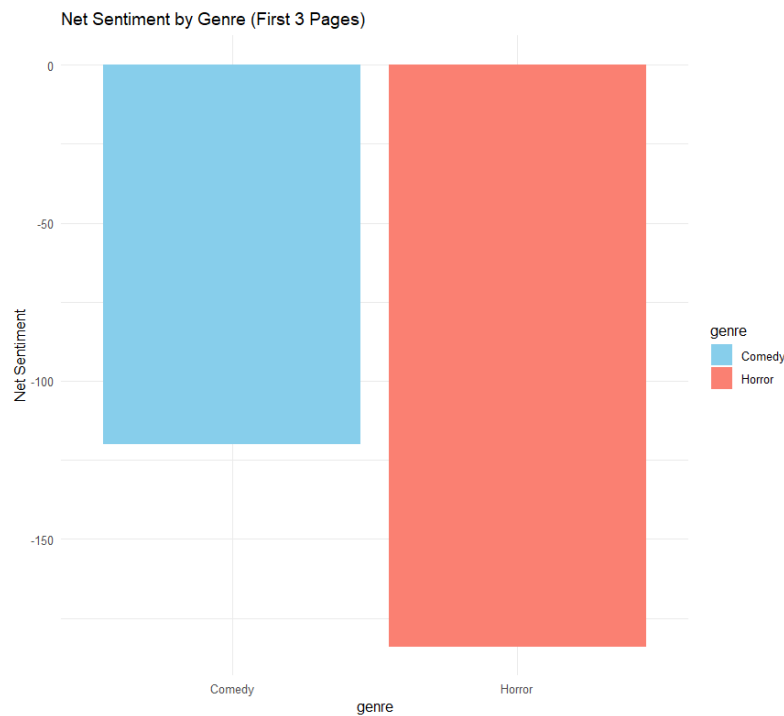


Figure 7 Sentiment analysis by genres.

The sentiment analysis for first three pages reveals that both Comedy and Horror genres exhibit an overall negative tone where Comedy has -120 net sentiment while Horro has -184 net sentiment. This indicates that horror description are understandably more negative in tone while comedy is expected to be more positive. This could be due to dramatic or conflict-driven plot element commonly found in descriptions to attract the audience. Figure 8 below shows top TF-IDF words for both genres.

Web Scrapping

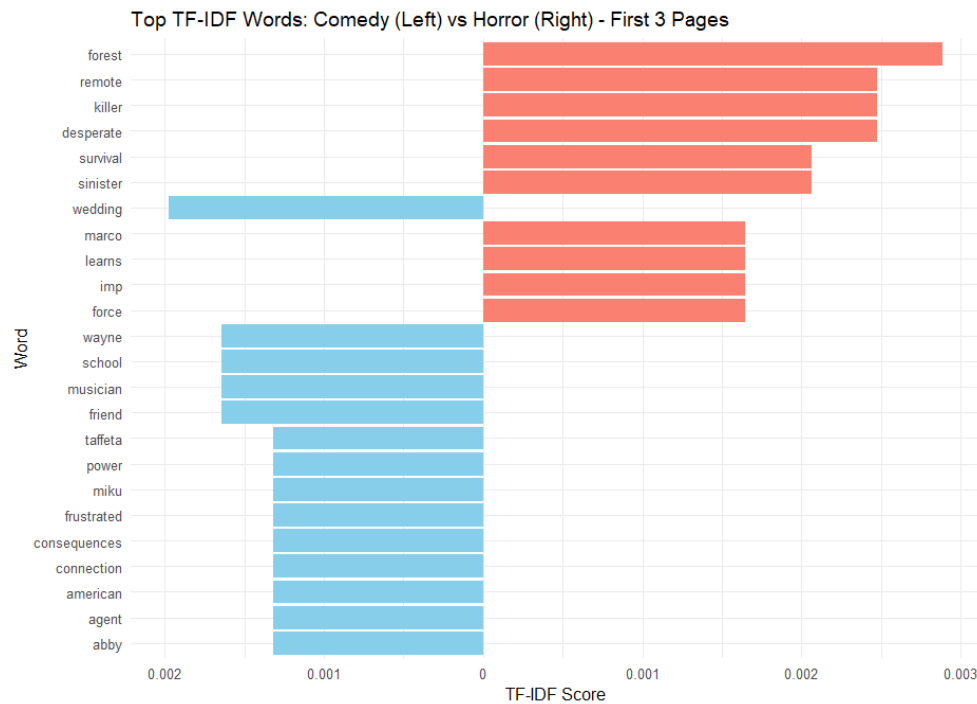


Figure 8 Top TF-IDF words for both genres.

TF-IDF is used to identify words are most important or unique for specific group. It combine how often words appear and how unique it is across the genre. In Horror, top TF-IDF shows words like forest, desperate and killer which highlighting fear-based theme. Meanwhile, Comedy emphasize words like wedding, friends and school which reflect social everyday life themes. This can be used to define the tone or narrative content of each genre rathe than common words that appear across both genres. Figure 9 below shows movie release by month and day for first three pages.

Web Scrapping

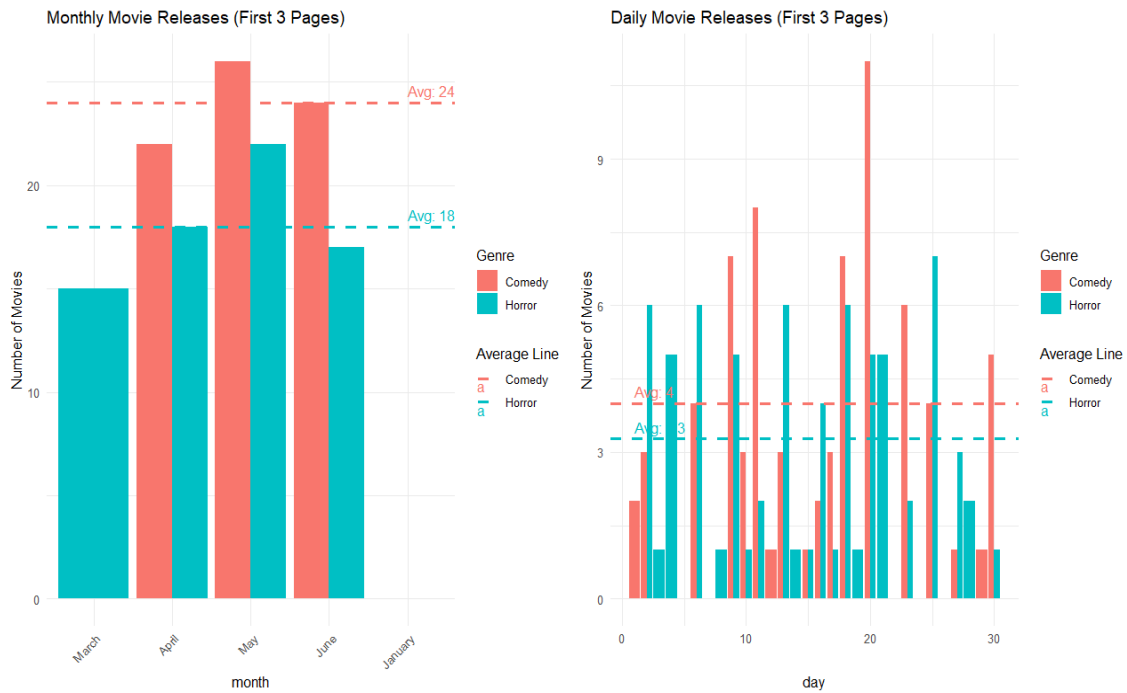


Figure 9 Movie release by month and day for first 3 pages

The total movie release for first three pages are 72 but the average horror movies release for 4 month (March to June) are 18 movies per month while the average comedy movie release for 3 months (April to June) are 24 movies per month. This is show on March there are no any movie with Comedy genre was produced and released and the highest movie release is on May for Comedy. Then, the average comedy release per day are 4 while horror movies is 3.27 and the highest movie release are on 20th day.

Next is about the web scrapping for all the pages in the Metacritic website for movies with genre Comedy and Horror. There are 8088 Comedy movies in 337 pages and 3048 Horror movies in 127 pages throughout year 1910 until 2025. Then, there are 1715 missing values for Comedy and 948 missing values for Horror which represent 21.2% and 31.1% of missing values for Metascore which is not that hight when comparing with first three pages. Figure 10 below shows percentage of missing Metascore by year and genre.

Web Scrapping

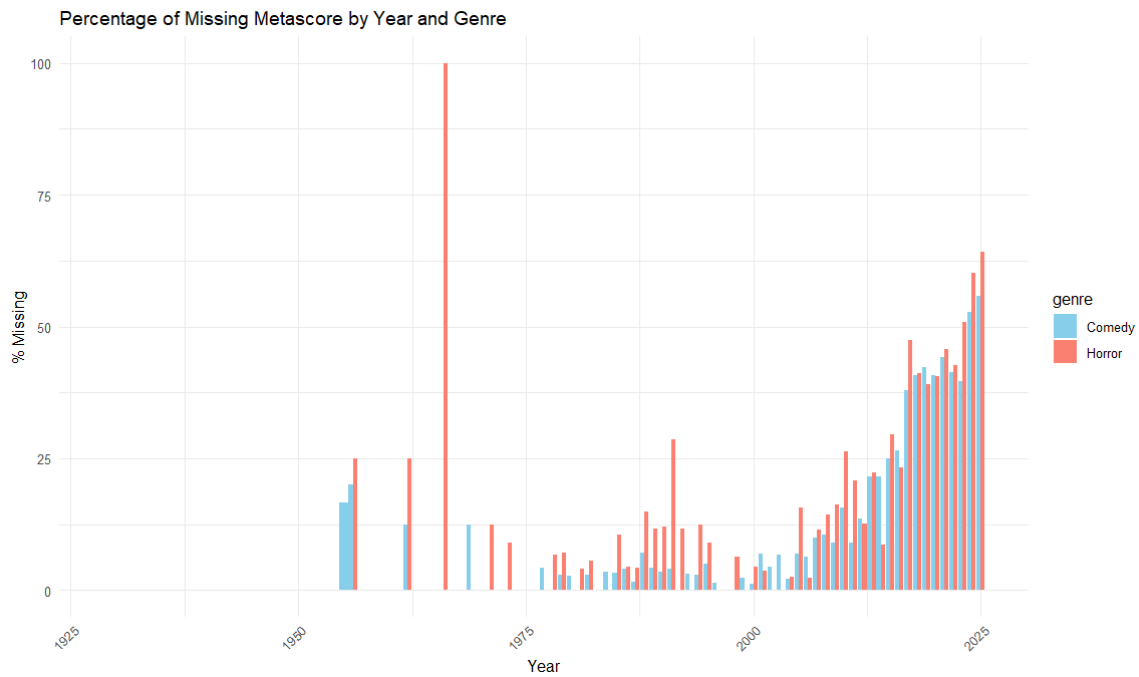


Figure 10 Percentage of missing Metascore by year and genre.

From the figure, number of missing Metascore values keep increasing by year due the rising number of movies releases each year. This is also because not everyone can keep-up new movies all the time due to time constraints such as work commitments. In 1966, only one Horror movies was released which is “The Hands of Fate” and no Metascore was recorded resulting in a 100% missing rate. This suggest that no one reviewed the film possibly because it failed to attract viewers due to an unappealing title, description or poster. Then, the missing values is imputed with average of the Metascore due to numerical rating and using round number to make it consistent. Figure 9 below shows Metascore comparison for both genre.

Web Scrapping

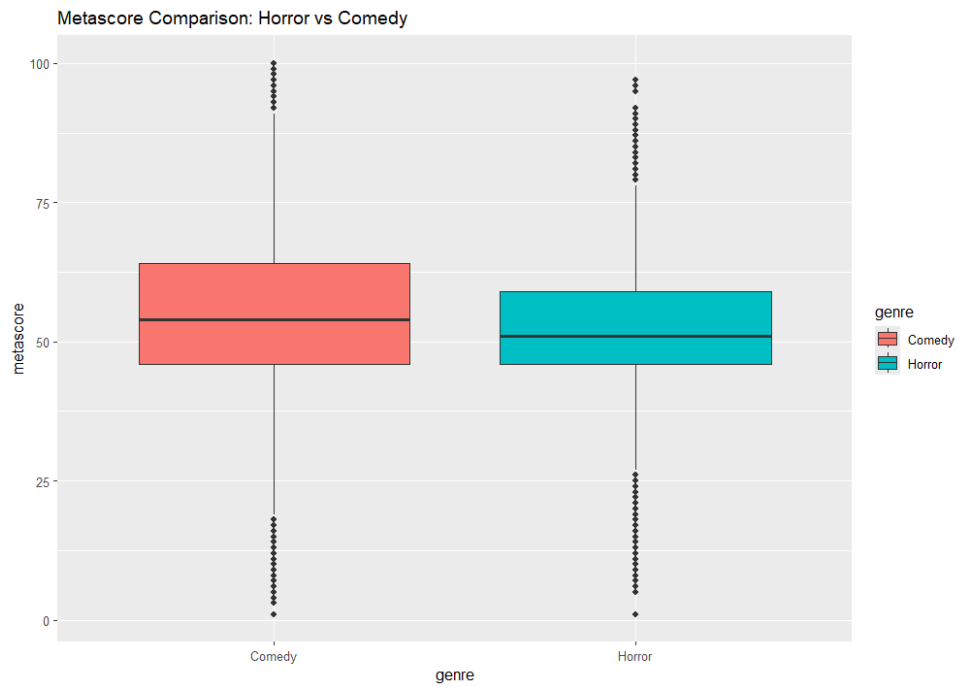


Figure 11 Metascore comparison both genre

There are quite a few outliers can be seen from both genre which is 195 outliers for Comedy movies and 268 outliers for Horror movies. From the t-test, p-value is below than 0.05 so the difference between average Metascore for both genre are significant where Comedy at rating 54 and Horror at 51. Figure 12 below shows top 10 words for both genres in all pages.

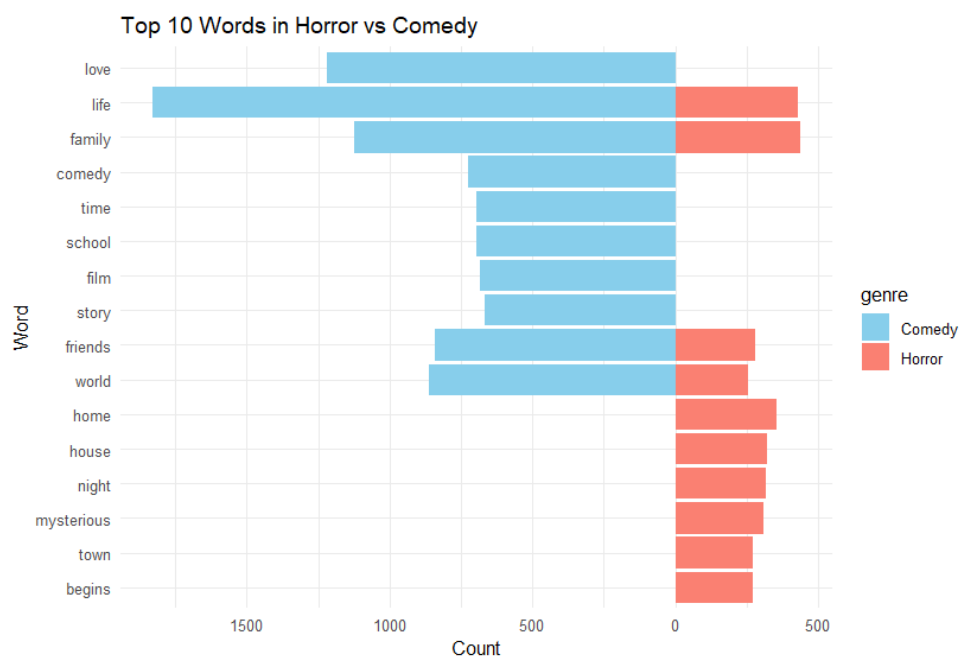


Figure 12 Top 10 words for all pages in both genre

In the full dataset, Comedy focuses on themes like life (1829 times), love (1218 times), and family (1125 times) while Horror emphasizes mysterious (308 times), home (354 times) and night (317 times) which reinforcing the earlier pattern of emotional vs suspenseful storytelling. Figure 13 and Figure 14 below shows wordcloud for both genre in all pages

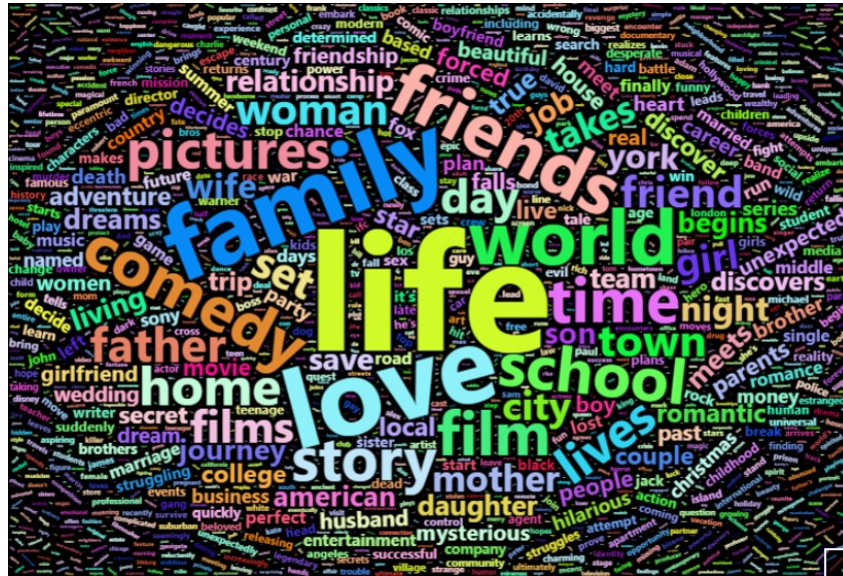


Figure 13 Wordcloud for comedy movies for all pages.

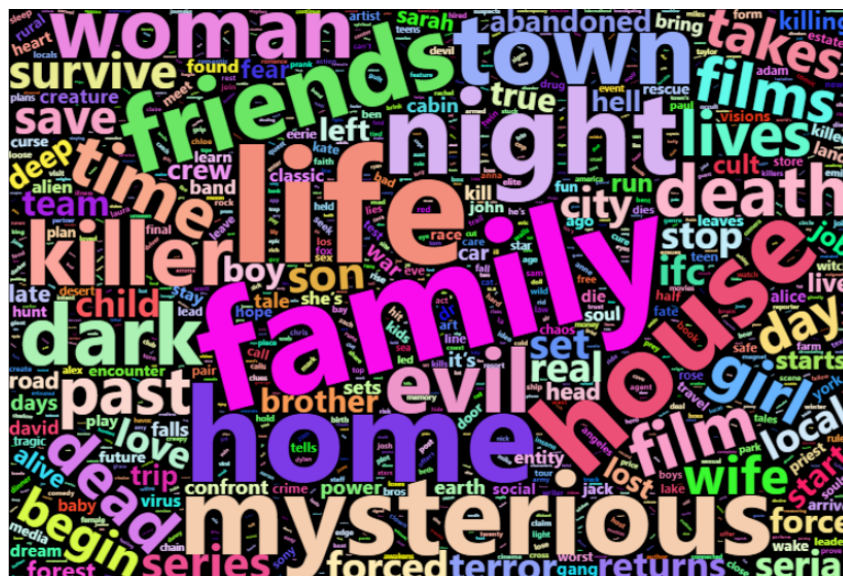


Figure 14 Wordcloud for horror movies for all pages.

As previously mentioned, words from Comedy movies tend to be heartwarming while words from Horror movies very suspenseful. From the wordcloud, can observe that Comedy movies have higher number of words compared to Horror movies which reflects the greater number of Comedy movies produce. Comedy movies is a highly versatile genre where it can

Web Scrapping

combined with others to create horror-comedy, action-comedy, romance-comedy and more. This versatility contributes to the higher number of Comedy movies available. Figure 15 below shows net sentiment by genre for all pages.

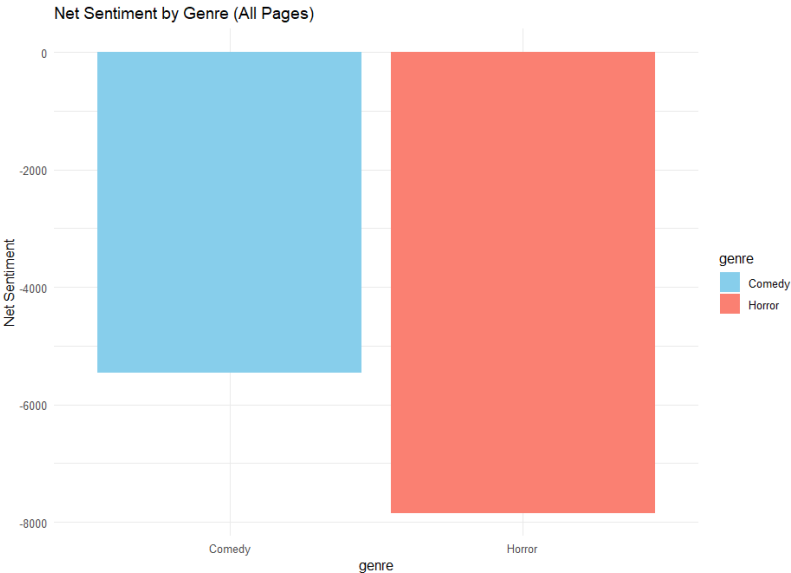


Figure 15 Net sentiment by genres for all pages.

Same as above where both of the net sentiment is negative which at -5467 for Comedy and -7848 for Horror. Horror have higher negative sentiment due to its genre while Comedy expected to have positive sentiment. Figure 16 below show top TF-IDF words for both genres.

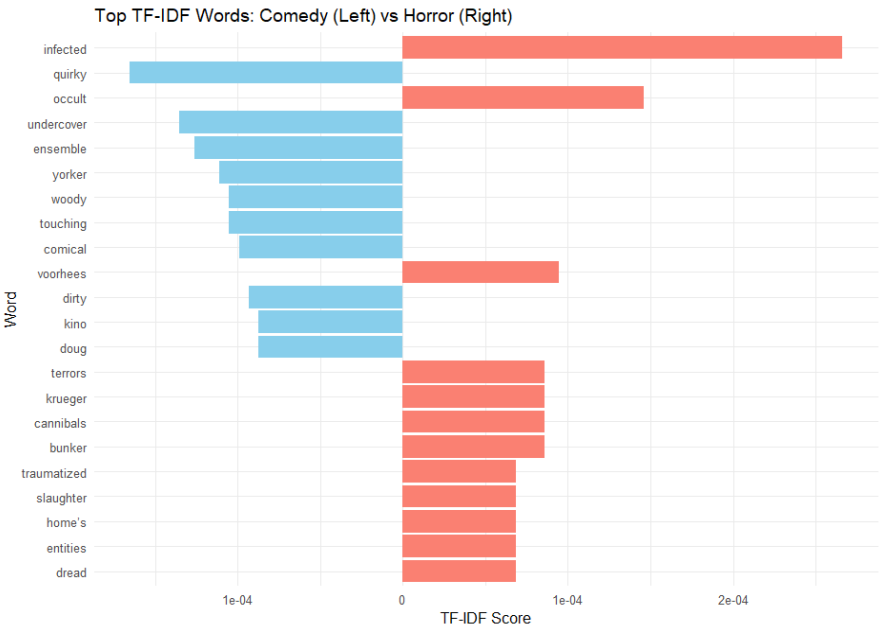


Figure 16 Top TF-IDF words for both genres in all pages.

Web Scrapping

In Comedy words like quirky, undercover, ensemble and touching appear frequently which emphasizing light-hearted, humororus and emotionally warm themes. In contrast, Horror shows a strong presence of disturbing terms such as infected, occult, voorhees and slaughter which suggest darket, violent and supernatural elements. This analysis allowing to better understand the linguistic signature of each genre's movies descriptions. Figure 17 below shows the number of movies release by year, month and day.



Figure 17 Movie release by year, month and day.

From 1910 to 2025, the number of movies release for both genres are keep increasing with the average movie release per year are 83 and 37 for Comedy and Horror movies. By month also shows the Comedy movies has the highest average movie release per month which is at 674 and 254 for Horror movies. Then, the average movie release by day which is 31 days for Comedy are 261 and Horror are 98. Figure 18 below shows the average Metascore value for both genres by month and based on weekend and weekday release date.

Web Scrapping

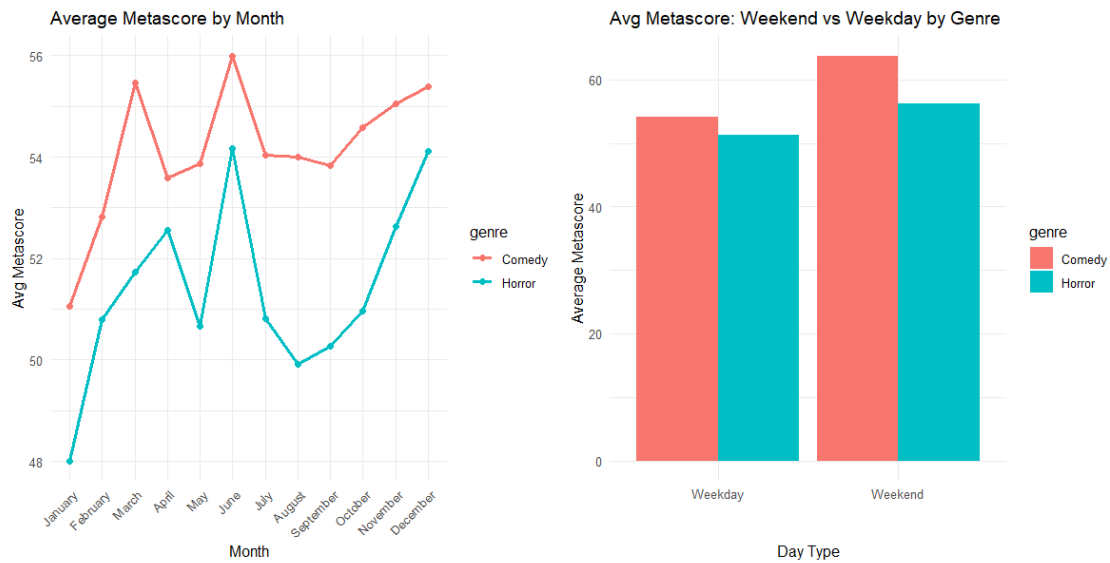


Figure 18 Average Metascore by month and weekday or weekend

From figure above shows Comedy movies has the highest average rating per month across all month which indicating the reviewers generally favors This also suggest Comedy movies may have better commercial potential. Additionally, movies released on weekend tends to have higher Metascore compared to those related to weekdays for both genres. Based on this observation, producer can strategically schedule release date on weekends to potentially achieve higher ratings to attract the potential customers later. Figure 19 below shows the forecasting next 5 years for number of movies released and average Metascore for both genres.

Web Scrapping

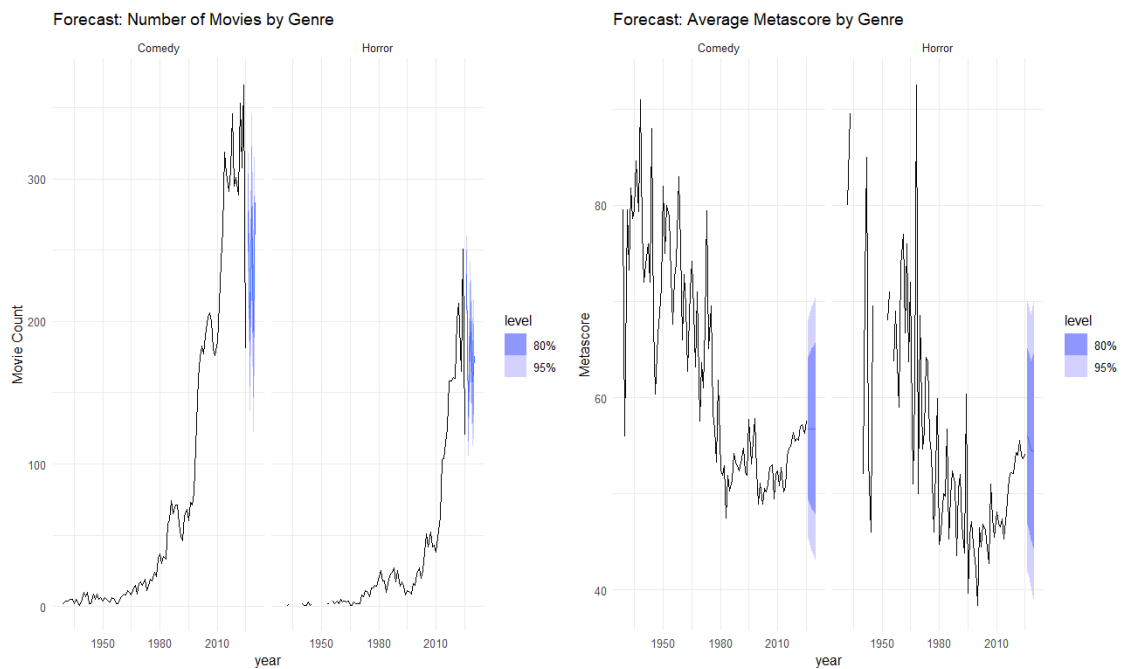


Figure 19 Forecasting for next 5 years for number of movies released and average Metascore for both genres.

From the forecasting above, the number of movie release for both genres will decreasing over the year while Metascore value will increase for next 5 year for both genres.

Conclusion

In conclusion, the analysis for both genres which is Comedy and Horror movies from Metacritic reveals for first 3 pages and full pages show the distinct patterns. Overall, Comedy has the higher movie release and also higher Metascore compared to Horror. This both genres have relatively consistent release trends throughout the year though Comedy has higher volume. Text analysis for Comedy shows the positive and social words while Horror favoured darker and suspense related words. The TF-IDF show the genre distinction between the both by highlighting the uniques keywords. However, sentiment analysis showed both genres leaned toward negative sentiment with Horror significantly more negative. Then the data show releasing the movie on weekend lead to higher Metascore. Lastly, these insights are valuable for understanding genre trends, audience expectations and content focus in movies.