ALY 6010 – Probability and Introductory Statistics
Project Report

Analysis of Movies/TV Shows in Netflix and Amazon Prime Video

Source: *Popular Movies and TV shows Amazon Prime, Netflix*. (2021, January 29). Kaggle.
https://www.kaggle.com/datasets/jyotmakadiya/popular-movies-and-tv-shows-amazon-prime-netflix/code

Devi Somalinga Bhuvanesh
NUID – 001428225

Final Project
Analysis of Movies/TV Shows in Netflix and Amazon Prime Video

Introduction

The entertainment industry has made its own space in the online medium. Movies, Songs, and TV shows of different genres and languages can now be viewed by individuals anywhere across the world through the online platforms. Of them, Netflix which started on 29th August 1997 and Amazon Prime Video on 7th September 2006 are among the most popular online platforms where the old as well as recent movies and TV shows are present. The chosen dataset from Kaggle will help to analyse the trend of movies and TV shows, their genres, and ratings present in Netflix and Amazon Prime Video.

Data Cleaning

The given dataset consists of information with **24,665 rows** of Movies and TV shows which are present in Netflix and Amazon Video Prime with **10 different attributes.** A detailed data cleaning process was carried out in Python.

Duplication of entries were found in following instances which was cleaned:

- Few movies/TV shows belong to multiple genres. Such movies/TV shows were repeated in several rows for each genre which resulted in duplication of movie name, IMDb ratings, Rotten Tomatoes, and related fields for the same movie/TV show. Therefore, its title and related fields were merged into a single one with genre as string delimited by "," using 'group by' function. This helps to uniquely count the movies/TV shows and its related fields present in the database.
- Similarly, the same movie/TV shows were listed in Netflix and Amazon prime video streaming platform in different entries. In order to avoid duplication, it was combined using the 'group by' function in Python.

In order to bring in uniformity, data errors were found in following situations:

- The IMDb ratings and Rotten Tomatoes for each movie/TV show was represented as a decimal value between 1-10 and 1-100 respectively. Yet, IMDb had few rows mentioned as "d;}" and Rotten tomatoes had few rows mentioned as "na". Here, substituting the respective data error with "0" could mislead interpretation. Therefore, "NaN" (Not a Number) was used to replace a blank cell using the 'replace' function.
- The first column in the original database is just a row index which is unnamed and not required for the analysis. In Python, an automatic row index value starting from "0" is generated. Hence, to avoid duplication and confusion, it was removed using the 'drop' function.

Further, by grouping the values of Rotten Tomatoes as per Tomatometer guidelines and IMDb, **two new variables were also included, i.e., Category of Rotten Tomatoes and Category of IMDb which is further explained in the table below.**

Table 1: Data type and definition of attributes

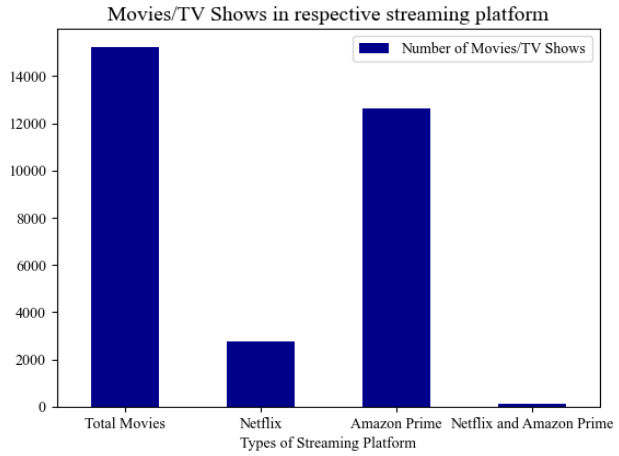
Attributes	Data Type	Definition
ID	Object	Unique ID for each row
Title	Object	Name of the movie/TV show
Year	Integer	Year of movie release in the industry
Rating	Object	Age Restriction or Targeted audience for each movie/TV show
IMDb	Object	IMDb Rating out of 10.0 which has audience score
Rotten Tomatoes	Object	Rotten Tomatoes percentage (out of 100) which has critic and audience score
Genre	Object	Genre of a movie/ TV show - 13 different genres are present
Netflix	Integer	1 if on Netflix else 0
Amazon Prime Video	Integer	1 if on Amazon Prime Videos else 0
Category of Rotten Tomatoes	Object	Fresh >=60.0, Rotten <60.0, No Tomatometer - Blank or NaN
Category of IMDb	Object	Good >= 7, Average 3-7, Bad <3, No ratings - Blank or NaN

Therefore, the cleaned database consists of **15,238 rows and 11 attributes** as mentioned above**.**

Data Analysis

A comparative study is carried out to analyse the data present in overall data with each of its subset databases which was extracted separately based on the streaming platforms, Netflix, Amazon Prime Video, and common movies/TV shows present in both streaming platforms.

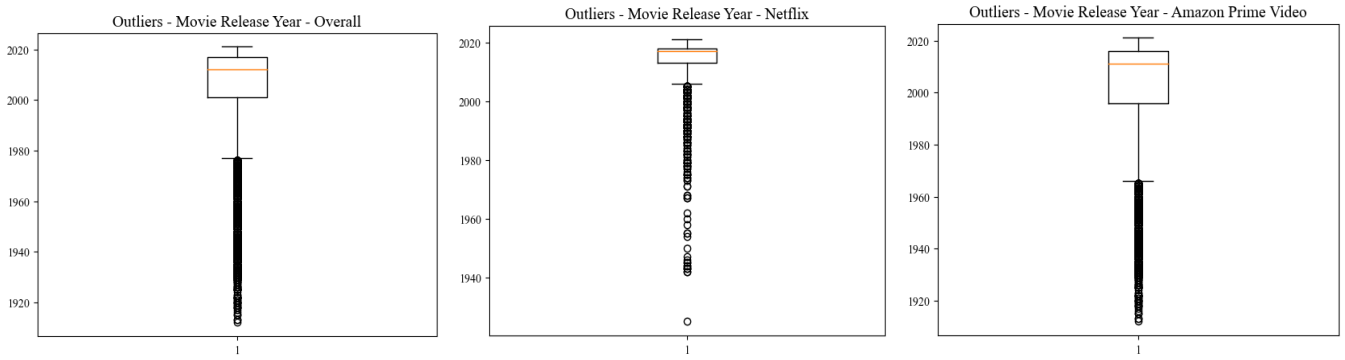
Movies/TV Shows in respective streaming platforms



The database for analysis consists of 15,238 unique movies/TV shows. Of them, 18% (2765) movies/TV shows are present in Netflix, and 80.75% (12,306) are present in Amazon Prime Video. This indicates that the sample size of Netflix is comparatively lesser than that of movies/TV shows present in Amazon Prime Video. In addition, less than 1%, i.e., 0.88% (135) movies/TV shows are commonly present in both the streaming platforms.

Outlier analysis of Movies/TV Shows release year

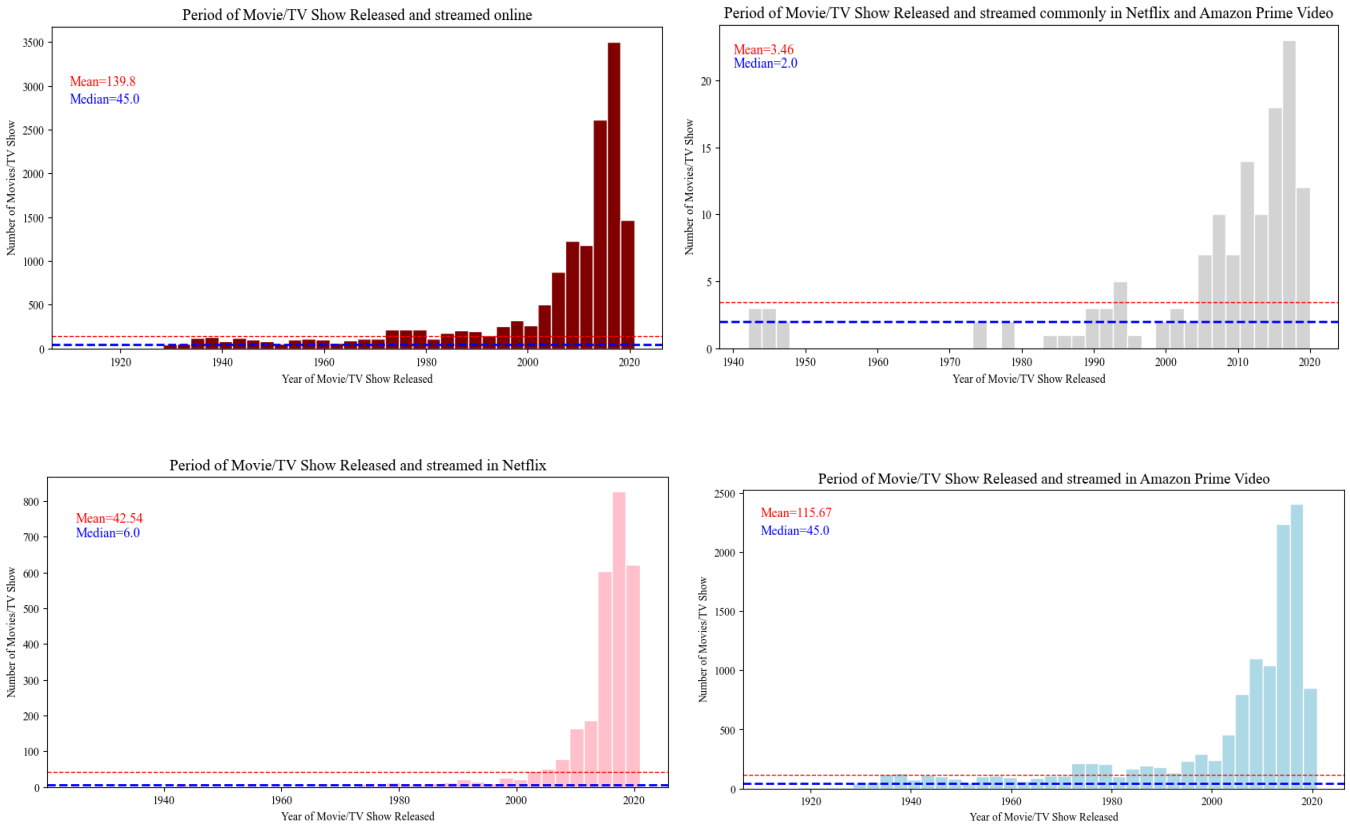
To determine whether there is any huge variation in the number of movies/TV shows selected from each year period, an outlier analysis was carried out. In the overall database, it was observed that 2.65% (404), i.e., movies/TV shows below 1980 are found as outliers. Here, 2001 is 25th quartile and 2017 as 75th quartile with 2012 as median. While looking at each of the streaming platforms, Netflix consists of 2.35% (65) movies/TV shows below the 2002 year period and 1.49% (183) movies/TV shows in Amazon Prime Video below the 1970 year period were found to be Outliers.



In order to compare between old and recent movies present in both platforms in terms of its ratings, genre, and age restrictions, the outliers are retained for the further analysis to see the trend.

Period of Movie/TV show released and streamed in online platforms

The movies/TV shows released from 1912 to 2021 in the industry are present in the database for analysis which has the average shows in 2003 period. The major variation can be observed in Netflix with movies/TV shows starting from 1925 are present with an average number of movies released in 2013.



With the advent of online streaming platforms in the late 20th century, people started preferring to watch movies/TV shows from home rather than going to theatres (*Engadget is part of the Yahoo family of brands, 2016).* Therefore, movies/TV shows released before and after the 2000 year period present in each of the streaming platforms are analysed.

In the overall database, less than a quarter (24%) movies/TV shows released before 2000 and the remaining 76% are from the recent post 2000 year period. Similarly, among each streaming platform, 6% (162) and 28% (3473) movies/shows before the 2000 year period are present in Netflix and Amazon Prime Video respectively while the remaining movies are from post 2000 year. In addition, 29 and 106 movies/TV shows released before and after 2000 respectively are present commonly in both the streaming platforms.

Therefore, it is evident that even though the entertainment stream in online platforms is showing an increasing trend on the selection of movies/shows released over the years, with more shows. Netflix platform consists of more movies released in recent times (94%) when compared to Amazon Prime Video (72%).

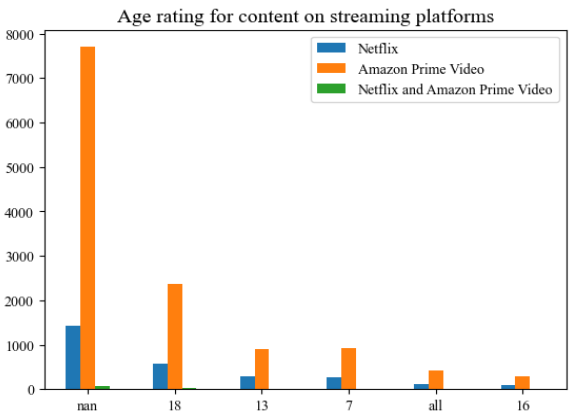
IMDb ratings and Rotten Tomatoes of movies/TV shows - Year wise

In order to analyse further on the quality of movies released based on each year period, the IMDb ratings and Rotten Tomatoes are compared.

The movies/TV shows released over different years are present in Netflix and Amazon prime video. While comparing the IMDb and Rotten tomatoes ratings, the movies released before the 2000 year period are having a majority of good IMDb scores when compared to Rotten Tomatoes. On the other hand, movies/shows released after 2000 years seem to have a similar range of IMDb and Rotten tomatoes ratings. Over the years, an increasing trend of movies/TV shows with more IMDb scores below the mean value of 58.9 can be observed while in Rotten Tomatoes, more scores above its average of 43.77.

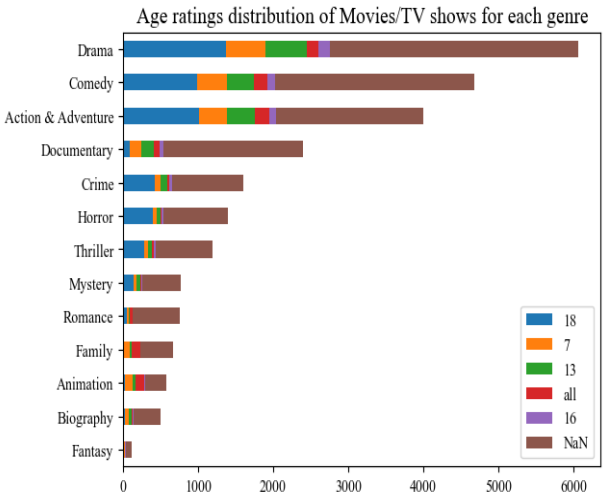
Age rating for content on streaming platforms

Each of the movies/TV shows offer a film rating system which provides parents with the information needed to determine if a film is appropriate for their children or not. Overall, 19% (2927) of the movies/TV shows in Netflix and Amazon Prime are for adults, i.e., age group above 18 years old followed by movies/shows for 13+ and 7+ years old with 7.7% (1175). Around 4% (544) of the content belong to all the categories of age group and 2% (365) of the movies are restricted for 16+ years.



On the contrary, it can also be noted that more than half of the movies, i.e, 59% (9054), do not have any film rating mentioned for respective content in the film/show which could mislead adults/children from watching movies/TV shows which is not under their category. This needs to be looked upon by each streaming platform, i.e., 52% and 61% in Netflix and Amazon prime video respectively. There are 78 common movies/shows streamed in both the platforms which do not have any rating.

Age rating distribution of movies/TV shows for each genre



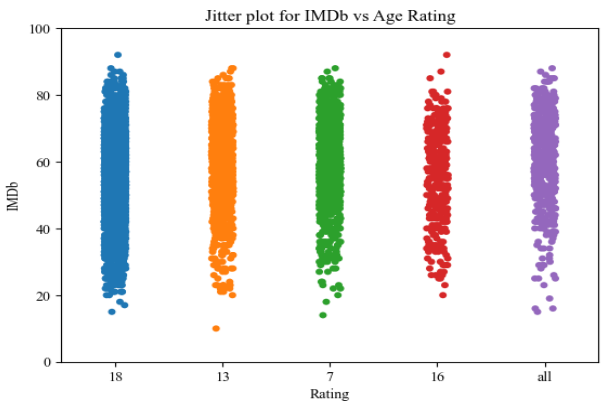
To assess the type of movies/TV shows present under each age rating, a genre wise analysis was carried out.

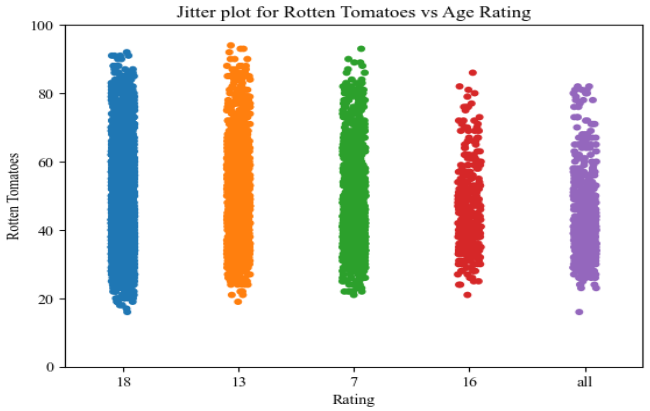
13 genres are present in the database where the maximum number of movies/shows are present in the Drama category with 25% followed by Comedy 19% and Action & Adventure 16%. The least number of movies/TV shows are found in Fantasy with 0.43% followed by Biography 2% , and Animation 2.31% genres.

Except for the "All" category, the Drama genre consists of the highest number of movies/shows in all other age ratings. Among Drama genres, the highest number of movies/ TV shows are present under the 7+ year category. 19% of the Action & Adventure genre represent the highest genre in the "All" category. On the other hand, the Fantasy genre has the least number of movies/shows in all the categories. In addition, Romance and Horror genres are also among the least under 16+ years and "All" categories respectively.

IMDb ratings and Rotten Tomatoes of movies/TV shows based on Age Rating of films

To further assess the quality of movies present in each of the age ratings, the IMDb ratings and Rotten Tomatoes are analysed.





For films rated under the 18+ year category, it has almost equal IMDb and Rotten Tomatoes ratings in the 20-85 range with few movies with higher Rotten Tomatoes more than 80 for the 18+ year category. Apart from 18+years, movies/TV shows under other age groups such as 13+, 7+, 16+, and all years categories have a similar pattern where the movies/shows have a majority of good rating in IMDb and low rating in Rotten Tomatoes. Therefore, movies/shows have different ratings and views in IMDb and Rotten Tomatoes except under the 18+ years category.

Comparison of IMDb ratings and Rotten Tomatoes in Netflix Versus Amazon Prime Video

The comparison of ratings in IMDb ratings and Rotten Tomatoes of movies/TV shows present in the database will help to identify which platform has films with good ratings.

Table 2: Key statistical details of IMDb score in each streaming platform

	Overall (n=15238)	Netflix (n=2765)	Amazon Prime Video (n=12608)	Netflix and Amazon Prime Video (n=135)
Count	14902	2728	12306	132
Mean	5.89	6.23	5.82	6.30
Median	6.00	6.30	6.00	6.40

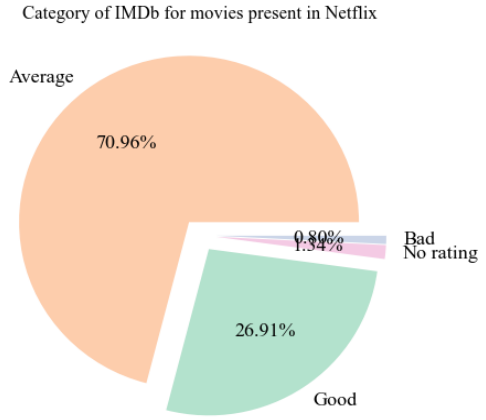
Table 3: Key statistical details of Rotten Tomatoes score in each streaming platform

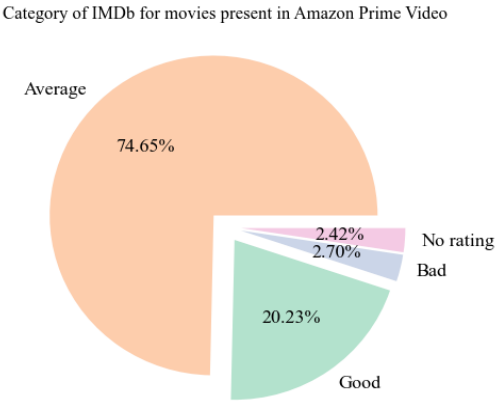
	Overall (n=15238)	Netflix (n=2765)	Amazon Prime Video (n=12608)	Netflix and Amazon Prime Video (n=135)
Count	15150	2758	12525	133
Mean	43.77	51.58	42.11	49.90
Median	40.00	49.00	39.00	48.00

Through the statistical analysis (refer Figure 1.1 in Appendix), the following key inferences can be derived:

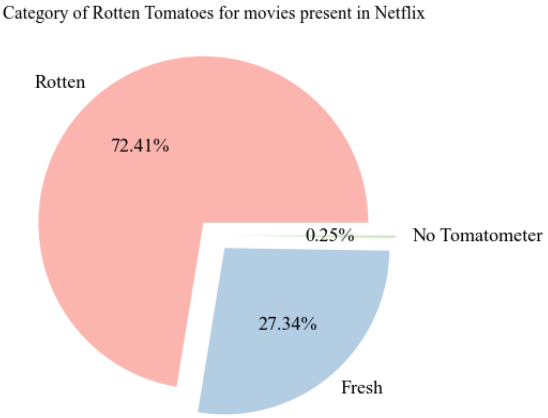
- It can be found that not all movies/TV shows are provided with IMDb and Rotten Tomatoes scores. However, Rotten tomatoes (99%) are comparatively provided to more movies/TV shows than IMDb score (98%).
- The mean of Rotten Tomatoes is greater than its median in the overall database as well as in sub-database of Netflix, Amazon Prime Video and common movies/shows in both platforms which indicates that it is positively skewed. On the other hand, IMDb in all platforms is negatively skewed as the mean is lesser than median.
- IMDb rating, which is the ratings given by the user, is rated out of 10 and it ranges from 1.3 to 9.8 with an average rating of 5.89. No movies/TV shows have full IMDb ratings.
- Similarly, Rotten Tomatoes which is rated out of 100 from both critics and audience view. In the current cleaned database, the Rotten Tomatoes rating ranges from 10 to 94 with an average rating of 43.77. No movie/show has a 95 or more score.

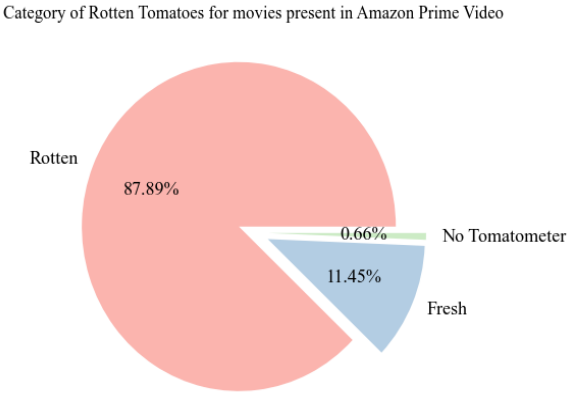
IMDb - Netflix vs Amazon Prime Video





Rotten Tomatoes - Netflix vs Amazon Prime Video





Netflix has more good IMDb ratings (26.91%) and more fresh Tomatoes (27.34%) when compared to Amazon Prime video (20.23% good IMDb ratings and 11.45% fresh tomatoes). IMDb ratings in both the streams are negatively skewed while Rotten Tomatoes are positively skewed (Refer Figure 1.2 and Figure 1.3 in Appendix).

Genre Wise comparison of IMDb and Rotten Tomatoes

To further understand which genre of movies/TV shows has good IMDb ratings and Rotten tomatoes, a genre wise comparison was carried out. Through the analysis, it is evident that Drama, Animation, and Biography genres have majorly good IMDb and Rotten Tomatoes ratings while Fantasy, Horror, Thriller, and Mystery have majorly bad ratings. Overall, few movies/TV shows in Documentary, Family, Fantasy, Biography, and Animation genres did not have any Tomatometer ratings as well as IMDb ratings. (Refer Figure 1.4 and Figure 1.5)

Table 4: Genre Wise Comparison of IMDb vs Rotten Tomatoes

Genre Wise Comparison	IMDb	Rotten Tomatoes
Good IMDb/ Fresh Tomatoes The movies/TV in these genres are quite good comparatively with other genres.	• Documentary (50.79%) • Biography (48.07%) • Family (28.12%) • Animation (27.59%) • Drama (20.62%)	• Drama (21.18%) • Action and Adventure (16.69%) • Animation (15.11%) • Biography (15.01%) • Comedy (14.73%)
Bad IMDb/ Bad Rotten Tomatoes The movies/TV shows in these genres are not good enough when compared to other genres.	• Horror (8.98%) • Fantasy (4.67%) • Action & Adventure (4.28%) • Thriller (4.12%) • Mystery (2.61%)	• Fantasy (99.07%) • Romance (97.73%) • Horror (97.56%) • Thriller (97.39%) • Mystery (97.25%)
No IMDb rating/ No Tomatometer	• Documentary (8.89%) • Animation (4.22%) • Family (3.16%) • Fantasy (2.80%) • Biography (2.64%)	• Documentary (2.71%) • Family (1.20%) • Fantasy (0.93%) • Biography (0.61%) • Animation (0.53%)

Hypothesis testing on movie quality in both streaming platform

In order to understand which streaming platform has quality movies and TV shows with good ratings, a hypothesis testing was carried out to analyse IMDb and Rotten Tomatoes for movies present in Netflix and Amazon Prime Video. Two separate hypothesis testing was conducted to check the same.

Selection of Samples - Overall, the current database has 2765 movies/TV shows in Netflix and 12,306 movies/TV shows in Amazon Prime Video. To conduct the Hypothesis testing and to ensure that there is no bias, 2000 movies/TV shows in both the streaming platforms were randomly selected. Therefore, a random sampling method is adopted for the hypothesis testing.

Nature of sample - Each movie/TV show chosen through random sampling method is independent in nature.

Type of Distribution - As per Central limit theorem, the sampling distribution of the mean approaches a normal distribution as the size of the sample increases, regardless of the shape of the original population distribution. In this study, each streaming platform has 2000 sample sizes each which is large enough to state that it is a normal distribution.

Type of Statistic - Since the two independent samples are greater than 30, a Z-statistic is employed in the study to test the hypothesis.

Even though 2000 sample sizes were randomly chosen from each platform, few movies/TV shows in both the platforms did not have any IMDb ratings or Rotten Tomatoes mentioned. Hence, there is a variation in "n" value in Netflix and Amazon Prime Video.

Hypothesis 1 - Comparison of IMDb in both streaming platforms		
Null Hypothesis H0 = The average IMDb rating of movies in Amazon Prime Video is equal to or higher than the average IMDb rating of movies in Netflix.		
Alternate Hypothesis Ha = The average IMDb rating of movies in Netflix is higher than the average IMDb rating of movies in Amazon Prime Video.		
Type of Tail - Based on the alternate hypothesis, it is evident that it is a Right tail		
	Netflix	**Amazon Prime Video**
Sample size n	1976	1950
Sample Mean x	6.21	5.83
Sample Standard Deviation s	1.11	1.34
Z Statistic (n>30)	0.38	
P Value	0.35	
Significance Level	0.05	
Result	Fail to reject the null hypothesis. We do not have sufficient evidence to reject that the average IMDb rating of movies in Amazon Prime Video is equal to or higher than the average IMDb rating of movies in Netflix.	

Hypothesis 2 - Comparison of Rotten Tomatoes in both streaming platforms		
Null Hypothesis H0 = The average Rotten Tomatoes rating of movies in Amazon Prime Video is equal to or higher than the average rating of movies in Netflix.		
Alternate Hypothesis Ha = The average Rotten Tomatoes rating of movies in Netflix is higher than the average rating of movies in Amazon Prime Video.		
Type of Tail - Based on the alternate hypothesis, it is evident that it is a Right tail		
	Netflix	**Amazon Prime Video**
Sample size n	1995	1988
Sample Mean x	51.27	42.09
Sample Standard Deviation s	13.89	12.63
Z Statistic (n>30)	9.18	
P Value	0.00	
Significance Level	0.05	

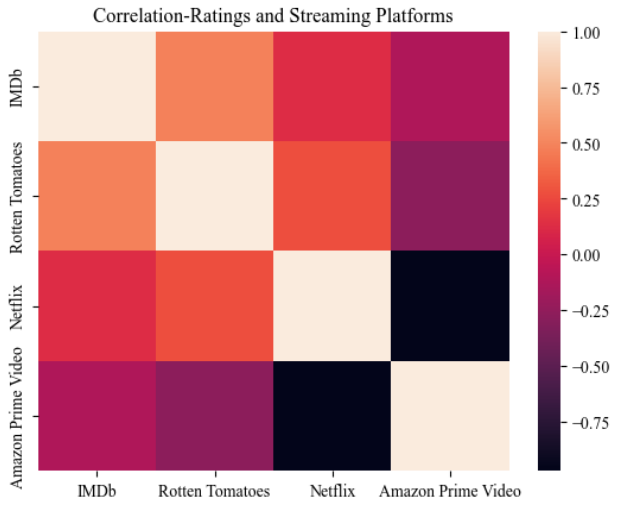
Result	Reject the null hypothesis. Therefore, the average Rotten Tomatoes rating of movies in Netflix is higher than the average rating of movies in Amazon Prime Video.

Through the hypothesis testing, at 95% significance level, it is evident that we do not have sufficient evidence to reject the claim that the average IMDb rating of movies in Amazon Prime Video is equal to or higher than the average IMDb ratings of movies in Netflix. However, the average Rotten Tomatoes rating of movies in Netflix is higher than the average rating of movies in Amazon Prime Video.

In summary, even though the movies/TV shows in Netflix have good Rotten Tomatoes ratings when compared to Amazon Prime video, it is not so while comparing IMDb ratings.

Correlation between streaming platforms and ratings

To determine whether the rating of a movie/TV show influences a streaming service's decision to buy the movie or vice versa, a correlation analysis was carried out.



A moderate positive correlation was observed between IMDb and Rotten Tomatoes (0.48) which could be due to the difference in rating methodology followed in each stream. However, it's important to consider other factors such as the demographic of the raters and the number of ratings provided when comparing the two platforms.

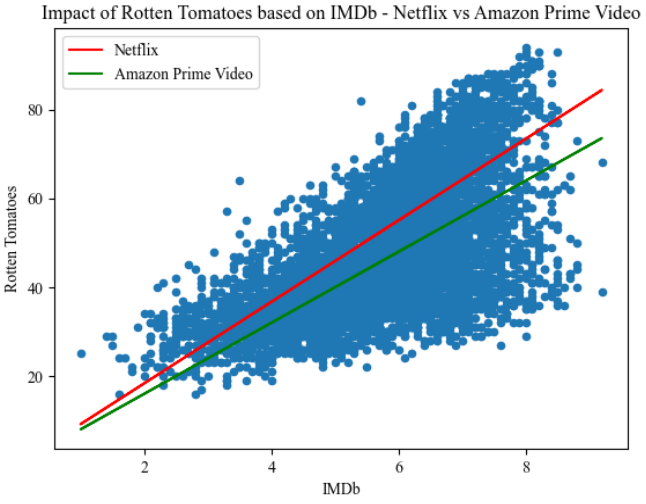
In contrast, a strong negative correlation of -0.97 was observed between the Netflix and Amazon Prime Video variables. This indicates that generally, when one streaming platform buys streaming rights for a movie, the other platform cannot stream the same.

A weak positive correlation of 0.12 between IMDb ratings and Netflix platform was observed while a weak negative correlation of -0.11 between IMDb and Amazon Prime Video. This indicates that either the streaming platforms' decision to buy a movie/TV show is not influenced by IMDb rating or, IMDb ratings are not influenced by streaming platforms streaming the movie.

Linear Regression on Ratings, Streaming Platforms, and Genre

Since similar factors like storyline, cinematography, etc could influence the ratings of movies/TV shows, we wanted to perform linear regression between IMDb and Rotten Tomatoes ratings. IMDb ratings consist only of audience ratings while Rotten Tomatoes consists of both audience and critics ratings. Hence, Rotten tomatoes is considered as a dependent variable and it is assessed with the independent variable, IMDb ratings. At 96.1% and 94.5% R^2 value, the linear model fits well to determine the impact of IMDb ratings on Rotten Tomatoes in Netflix and Amazon Prime Video respectively. Also, through hypothesis testing, at

95% confidence level, there is a linear relationship between IMDb and Rotten Tomatoes in both the streaming platforms. With increase in IMDb ratings, it can also be observed that Netflix has movies/TV shows with higher increase in Rotten Tomatoes when compared to Amazon Prime Video (regression line has higher slope for Netflix compared to Amazon Prime Video). (Refer Figure 1.6, Figure 1.7, and Figure 1.8 from Appendix section)



Similarly, a linear regression was carried out to predict IMDb and Rotten Tomatoes ratings for each of the 13 genres of movies/TV shows from Netflix and Amazon Prime Video**.** Dummy encoding method was used to encode categorical genre values into numerical variables that can be used for regression. At 95.6% and 92.5% R^2 value, the linear model fits to determine the impact of IMDb ratings and Rotten Tomatoes on each of the 13 genres in both the streaming platforms, respectively. Through the hypothesis testing it is evident that at 95% confidence level, there is a linear correlation between IMDb ratings with all the genres except Action & Adventure, Crime, Fantasy, and Mystery. Similarly, while comparing Rotten Tomatoes, all genres except Animation, Comedy, and Fantasy have a linear correlation. (Refer Figure 1.9 from Appendix section)

By predicting the values through a linear regression model, it can be observed that a slight variation was observed in actual IMDb ratings with predicted IMDb ratings. On the other hand, a huge variation was observed in actual and predicted ratings of Rotten Tomatoes. (Refer Figure 1.10 from Appendix section)

Summary

To understand the nature of 15,238 movies/TV shows released during 1912 to 2021 present in Netflix (18%) and Amazon prime video (80.75%), a comparative analysis was carried out with 11 attributes. Less than 1% (135) of movies/TV shows are present commonly on both the streaming platforms. Netflix consists of more movies released in recent times, i.e., after the 2000 year period (95%).

Even though the majority of the movies/TV shows are concentrated towards adults (19%) and Drama genre (25%), more than half (59%) of the film rating is not available which makes it difficult to determine if a film is appropriate for children or not. The least number of movies was found in Fantasy with 0.43% followed by Biography 2% , and Animation 2.31% genres. Also, the major difference in IMDb and Rotten Tomatoes ratings can be found in 4 genres, i.e., Documentary, Family, Fantasy and Romance.

Netflix has more good IMDb ratings (26.91%) and more fresh Tomatoes (27.34%) when compared to Amazon Prime video (20.23% good IMDb ratings and 11.45% fresh tomatoes). IMDb ratings in both the streams are negatively skewed while Rotten Tomatoes are positively skewed. Drama, Animation, and

Biography genres have majorly good IMDb and fresh Tomato ratings while Fantasy, Horror, Thriller, and Mystery have majorly bad ratings. Overall, few movies/TV shows in Documentary, Family, Fantasy, Biography, and Animation genres did not have any Tomatometer ratings as well as IMDb ratings.

With 95% confidence, the hypothesis testing proved that even though the movies/TV shows in Netflix have good fresh Tomato ratings when compared to Amazon Prime video, it is not so while comparing IMDb ratings. Through strong negative correlation of -0.97 between Netflix and Amazon Prime Video variables, we can conclude that in general, when one streaming platform buys streaming rights for a movie, the other platform cannot stream the same.

Through correlation and linear regression, it was evident that there is a linear relationship between IMDb and Rotten Tomatoes across the streaming platforms. On the other hand, no linear relationship exists between IMDb ratings with Action & Adventure, Crime, Fantasy, and Mystery genres as well as Rotten Tomatoes with Animation, Comedy, and Fantasy.

References

Popular Movies and TV shows Amazon Prime, Netflix. (2021, January 29). Kaggle. https://www.kaggle.com/datasets/jyotmakadiya/popular-movies-and-tv-shows-amazon-prime-netflix/code

pandas.DataFrame.copy — pandas 1.5.3 documentation. (n.d.). https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.copy.html

ferdio. (2021, June 21). *Jitter Plot | Data Viz Project*. Data Viz Project. https://datavizproject.com/data-type/jitter-plot/

seaborn.kdeplot — seaborn 0.12.2 documentation. (n.d.). https://seaborn.pydata.org/generated/seaborn.kdeplot.html

Matplotlib y axis values are not ordered. (n.d.). Stack Overflow. https://stackoverflow.com/questions/48062499/matplotlib-y-axis-values-are-not-ordered

matplotlib.pyplot.yticks — Matplotlib 3.7.1 documentation. (n.d.). https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.yticks.html

3.5.1 - Scatterplot with Groups | STAT 200. (n.d.). PennState: Statistics Online Courses. https://online.stat.psu.edu/stat200/lesson/3/3.5/3.5.1#:~:text=A%20scatterplot%20with%20groups%20can,x%2D%20and%20y%2Daxes.

Carvalho, T. (2022, August 15). Jitter plots with Python's Seaborn - Towards Data Science. Medium. https://towardsdatascience.com/jitter-plots-with-pythons-seaborn-62188bf511b8

GeeksforGeeks. (2020b, November 26). Density Plots with Pandas in Python. https://www.geeksforgeeks.org/density-plots-with-pandas-in-python/

Figure title with several colors in matplotlib. (n.d.). Stack Overflow. https://stackoverflow.com/questions/9350171/figure-title-with-several-colors-in-matplotlib

Different colors for points and line in Seaborn regplot. (n.d.). Stack Overflow. https://stackoverflow.com/questions/48145924/different-colors-for-points-and-line-in-seaborn-regplot

seaborn.FacetGrid.add_legend — seaborn 0.12.2 documentation. (n.d.). https://seaborn.pydata.org/generated/seaborn.FacetGrid.add_legend.html

How to add a legend on Seaborn facetgrid bar plot using Matplotlib. (n.d.). https://www.tutorialspoint.com/how-to-add-a-legend-on-seaborn-facetgrid-bar-plot-using-matplotlib

How to add a title to Seaborn Facet Plot. (n.d.). Stack Overflow. https://stackoverflow.com/questions/29813694/how-to-add-a-title-to-seaborn-facet-plot

Density Plot Python Pandas. (n.d.). Stack Overflow. https://stackoverflow.com/questions/59696944/density-plot-python-pandas

How do you create a legend for kde plot in Seaborn? (n.d.). Stack Overflow. https://stackoverflow.com/questions/64014746/how-do-you-create-a-legend-for-kde-plot-in-seaborn

How to color different seaborn kdeplots in one figure? (n.d.). Stack Overflow. https://stackoverflow.com/questions/68096271/how-to-color-different-seaborn-kdeplots-in-one-figure

GeeksforGeeks. (2023, January 13). Plot a Vertical line in Matplotlib. https://www.geeksforgeeks.org/plot-a-vertical-line-in-matplotlib/

GeeksforGeeks. (2021a, March 4). Bar Plot in Matplotlib. https://www.geeksforgeeks.org/bar-plot-in-matplotlib/

pandas.Series.plot.bar — pandas 1.5.3 documentation. (n.d.). https://pandas.pydata.org/docs/reference/api/pandas.Series.plot.bar.html

matplotlib.pyplot.subplots — Matplotlib 3.1.2 documentation. (n.d.). https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.subplots.html

Extract values in Pandas value_counts(). (n.d.). Stack Overflow.
https://stackoverflow.com/questions/35523635/extract-values-in-pandas-value-counts

pandas.Index.value_counts — pandas 0.17.0 documentation. (n.d.).
https://pandas.pydata.org/pandas-docs/version/0.17.0/generated/pandas.Index.value_counts.html

pandas.Series.str.lower — pandas 1.5.3 documentation. (n.d.). https://pandas.pydata.org/docs/reference/api/pandas.Series.str.lower.html

Plotting a stacked Bar Chart. (n.d.). Stack Overflow. https://stackoverflow.com/questions/23415500/plotting-a-stacked-bar-chart

How to merge pandas value_counts() to dataframe or use it to subset a dataframe. (n.d.). Stack Overflow.
https://stackoverflow.com/questions/35809098/how-to-merge-pandas-value-counts-to-dataframe-or-use-it-to-subset-a-dataframe

How to rename a pandas Series? (n.d.). Stack Overflow. https://stackoverflow.com/questions/18553503/how-to-rename-a-pandas-series

pandas.Series.rename — pandas 1.5.3 documentation. (n.d.). https://pandas.pydata.org/docs/reference/api/pandas.Series.rename.html

pandas.Series.value_counts — pandas 1.5.3 documentation. (n.d.).
https://pandas.pydata.org/docs/reference/api/pandas.Series.value_counts.html

pandas.Series.plot — pandas 1.5.3 documentation. (n.d.). https://pandas.pydata.org/docs/reference/api/pandas.Series.plot.html

How do I order my x-axis on pandas bar plot? (n.d.). Stack Overflow.
https://stackoverflow.com/questions/70306401/how-do-i-order-my-x-axis-on-pandas-bar-plot

Sorting the order of bars in pandas/matplotlib bar plots. (n.d.). Stack Overflow.
https://stackoverflow.com/questions/22635110/sorting-the-order-of-bars-in-pandas-matplotlib-bar-plots

How to plot multiple dataframes in subplots. (n.d.). Stack Overflow.
https://stackoverflow.com/questions/22483588/how-to-plot-multiple-dataframes-in-subplots

How to add a title to Seaborn Facet Plot. (n.d.). Stack Overflow.
https://stackoverflow.com/questions/29813694/how-to-add-a-title-to-seaborn-facet-plot

freeCodeCamp.org. (2017, April 10). *Whose ratings should you trust? IMDB, Rotten Tomatoes, Metacritic, or Fandango?*
https://www.freecodecamp.org/news/whose-reviews-should-you-trust-imdb-rotten-tomatoes-metacritic-or-fandango-7d1010c6cf19/

GeeksforGeeks. (2022, January 14). *Multi plot grid in Seaborn*. https://www.geeksforgeeks.org/multi-plot-grid-in-seaborn/

Pie charts — Matplotlib 3.7.1 documentation. (n.d.). https://matplotlib.org/stable/gallery/pie_and_polar_charts/pie_features.html

Seperate title for each subplot in a for loop in Python. (n.d.). Stack Overflow.
https://stackoverflow.com/questions/65007756/seperate-title-for-each-subplot-in-a-for-loop-in-python

Matplotlib - selecting colors within qualitative color map. (n.d.). Stack Overflow.
https://stackoverflow.com/questions/59149279/matplotlib-selecting-colors-within-qualitative-color-map

GeeksforGeeks. (2023, March 21). *Python map function*. https://www.geeksforgeeks.org/python-map-function/

How to get the index and values of series in Pandas. (n.d.).
https://www.tutorialspoint.com/how-to-get-the-index-and-values-of-series-in-pandas

Engadget is part of the Yahoo family of brands. (2016, October 11).
https://www.engadget.com/2016-10-11-how-the-internet-change-the-way-we-watch-movies.html

Appendix

Figure 1.1. Descriptive Statistics

Overall Database

	Year	IMDb	Rotten Tomatoes	Netflix	Amazon Prime Video
count	15238.000000	14902.000000	15150.000000	15238.000000	15238.000000
mean	2003.378134	5.891907	43.771551	0.181454	0.827405
std	21.203389	1.307138	13.368272	0.385407	0.377909
min	1912.000000	0.000000	10.000000	0.000000	0.000000
25%	2001.000000	5.100000	34.000000	0.000000	1.000000
50%	2012.000000	6.000000	40.000000	0.000000	1.000000
75%	2017.000000	6.800000	51.000000	0.000000	1.000000
max	2021.000000	9.800000	94.000000	1.000000	1.000000

Common Movies/TV Shows in Netflix and Amazon Prime

	Year	IMDb	Rotten Tomatoes	Netflix	Amazon Prime Video
count	135.00000	132.000000	133.000000	135.0	135.0
mean	2005.40000	6.308333	49.909774	1.0	1.0
std	18.51107	1.050121	12.478138	0.0	0.0
min	1942.00000	3.300000	28.000000	1.0	1.0
25%	2004.50000	5.600000	41.000000	1.0	1.0
50%	2012.00000	6.400000	48.000000	1.0	1.0
75%	2017.00000	7.100000	57.000000	1.0	1.0
max	2020.00000	8.500000	93.000000	1.0	1.0

Netflix

	Year	IMDb	Rotten Tomatoes	Netflix	Amazon Prime Video
count	2765.000000	2728.000000	2758.000000	2765.0	2765.000000
mean	2013.811573	6.231745	51.583756	1.0	0.048825
std	9.307126	1.111270	14.033797	0.0	0.215540
min	1925.000000	1.500000	10.000000	1.0	0.000000
25%	2013.000000	5.500000	41.000000	1.0	0.000000
50%	2017.000000	6.300000	49.000000	1.0	0.000000
75%	2018.000000	7.000000	61.000000	1.0	0.000000
max	2021.000000	9.000000	93.000000	1.0	1.000000

Amazon Prime Video

	Year	IMDb	Rotten Tomatoes	Netflix	Amazon Prime Video
count	12608.000000	12306.000000	12525.000000	12608.000000	12608.0
mean	2001.111675	5.821039	42.116487	0.010707	1.0
std	22.339557	1.333556	12.596599	0.102926	0.0
min	1912.000000	0.000000	10.000000	0.000000	1.0
25%	1996.000000	5.000000	34.000000	0.000000	1.0
50%	2011.000000	6.000000	39.000000	0.000000	1.0
75%	2016.000000	6.800000	49.000000	0.000000	1.0
max	2021.000000	9.800000	94.000000	1.000000	1.0

Figure 1.2. Rotten Tomatoes vs IMDb - Overall, Netflix, and Amazon Prime Video

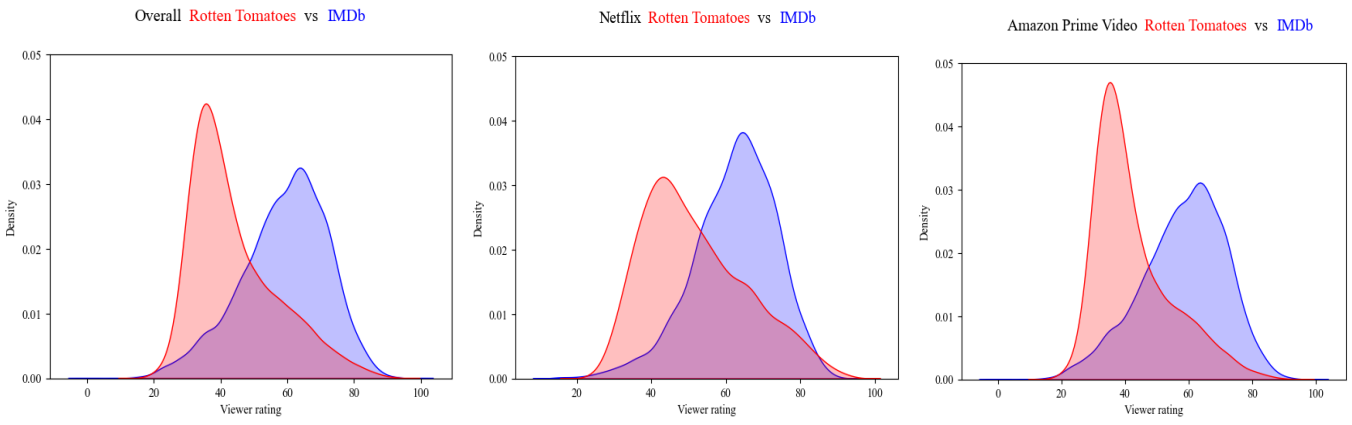


Figure 1.3. Rotten Tomatoes vs IMDb - Genre wise

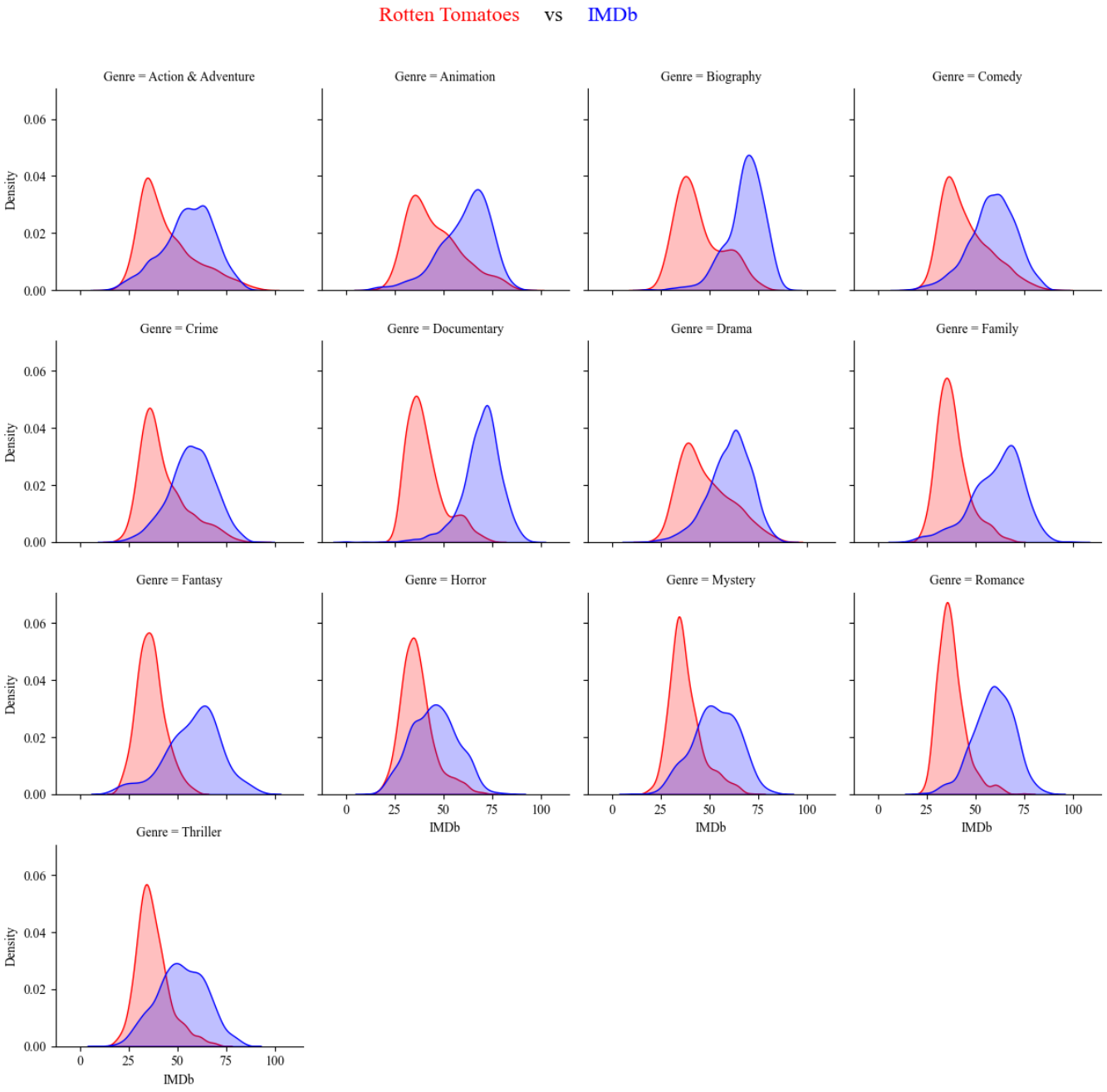


Figure 1.4. Genre Wise Comparison of Rotten Tomatoes Ratings

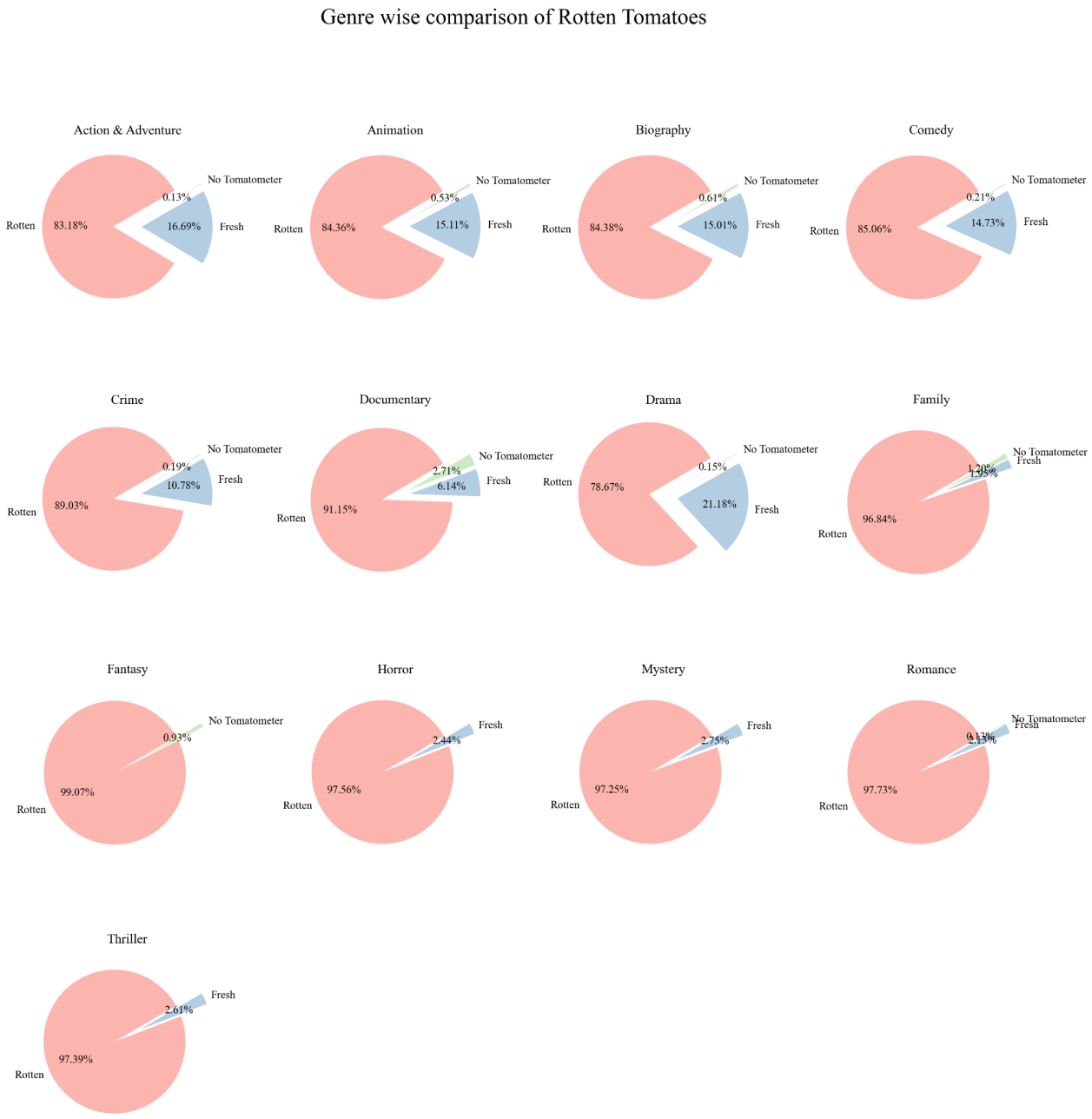


Figure 1.5. Genre Wise Comparison of IMDb Ratings

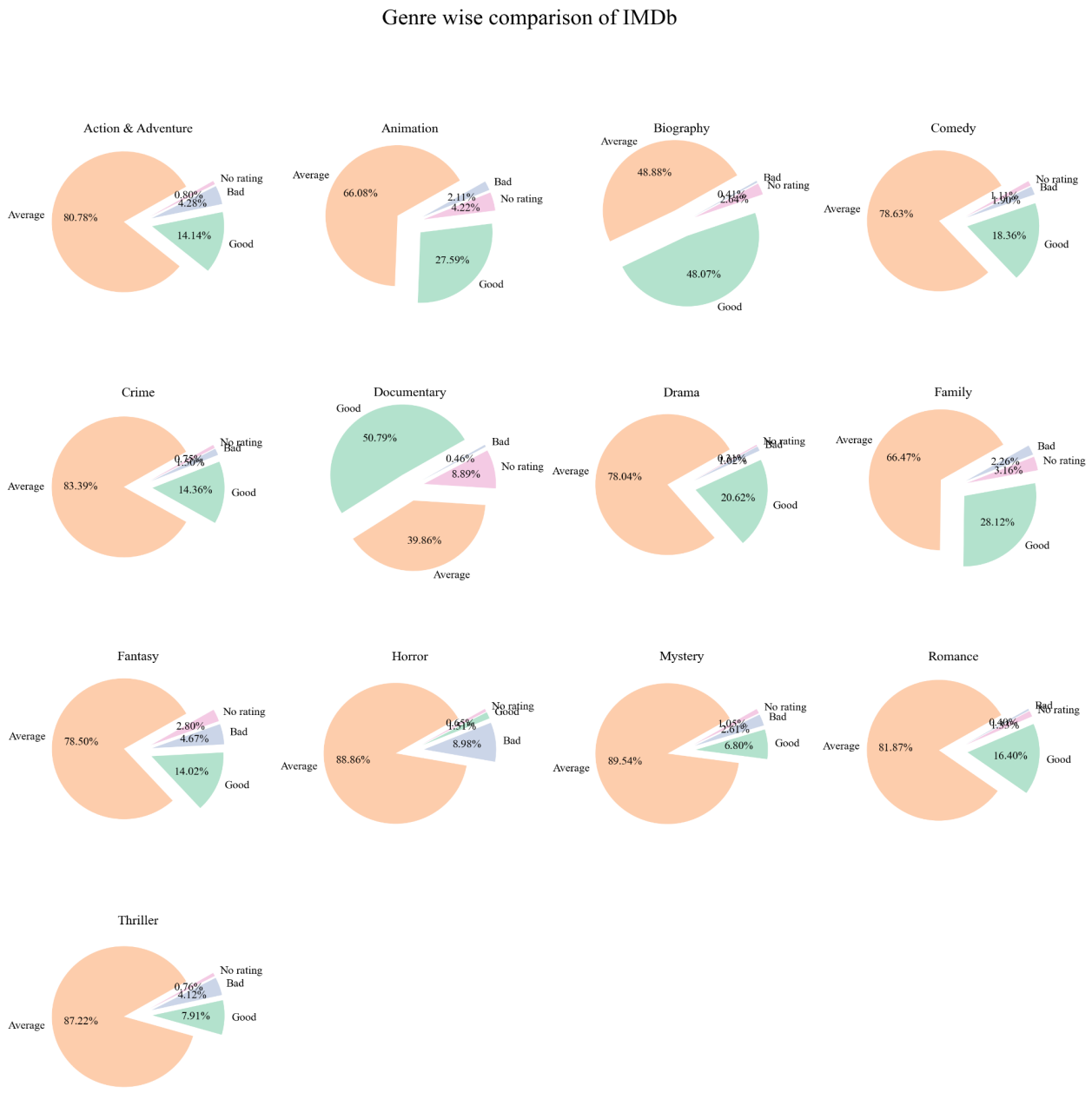
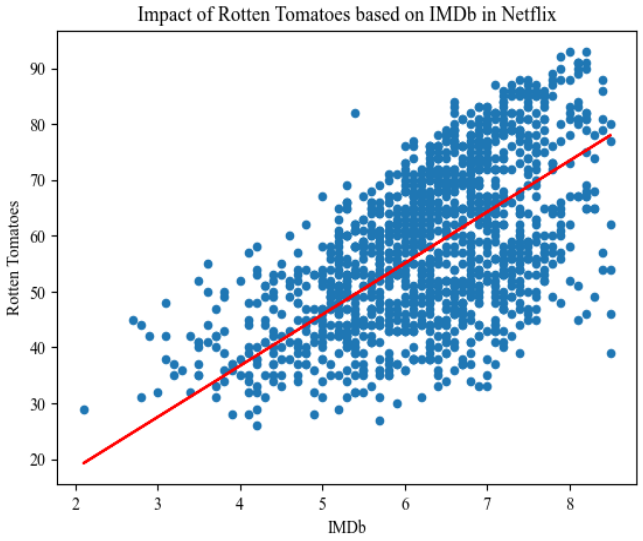


Figure 1.6. Linear Regression of IMDb and Rotten Tomatoes for Netflix and Amazon Prime Video



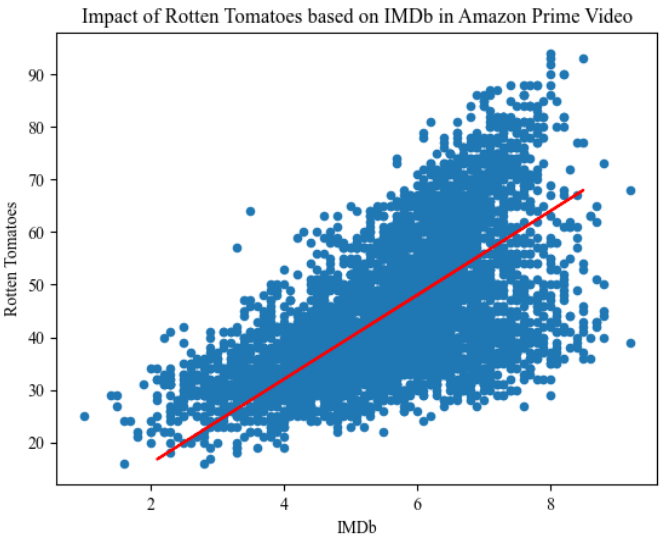


Figure 1.7. Hypothesis of IMDb and Rotten Tomatoes for Netflix and Amazon Prime Video

OLS Regression Results

Dep. Variable:	Rotten Tomatoes	**R-squared (uncentered):**	0.961
Model:	OLS	**Adj. R-squared (uncentered):**	0.961
Method:	Least Squares	**F-statistic:**	3.303e+04
Date:	Mon, 03 Apr 2023	**Prob (F-statistic):**	0.00
Time:	16:39:35	**Log-Likelihood:**	-5154.9
No. Observations:	1329	**AIC:**	1.031e+04
Df Residuals:	1328	**BIC:**	1.032e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
IMDb	9.1749	0.050	181.749	0.000	9.076	9.274

Omnibus:	79.374	**Durbin-Watson:**	1.095
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	47.070
Skew:	-0.319	**Prob(JB):**	6.01e-11
Kurtosis:	2.334	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS Regression Results

Dep. Variable:	Rotten Tomatoes	**R-squared (uncentered):**	0.945
Model:	OLS	**Adj. R-squared (uncentered):**	0.945
Method:	Least Squares	**F-statistic:**	8.367e+04
Date:	Mon, 03 Apr 2023	**Prob (F-statistic):**	0.00
Time:	15:25:55	**Log-Likelihood:**	-18726.
No. Observations:	4884	**AIC:**	3.745e+04
Df Residuals:	4883	**BIC:**	3.746e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
IMDb	7.9965	0.028	289.257	0.000	7.942	8.051

Omnibus:	25.508	**Durbin-Watson:**	1.046
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	24.307
Skew:	-0.145	**Prob(JB):**	5.27e-06
Kurtosis:	2.812	**Cond. No.**	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 1.8. Prediction of Rotten Tomatoes in Netflix and Amazon Prime Video based on IMDb

	IMDb	Actual value of Rotten Tomatoes	Netflix	Amazon Prime Video
46	7.0	73.0	64.224139	55.975431
50	5.7	53.0	52.296799	45.579994
159	7.2	62.0	66.059114	57.574729
264	5.6	46.0	51.379311	44.780345
588	8.0	93.0	73.399016	63.971921

Figure 1.9. Linear Regression of IMDb and Rotten Tomatoes for each Genre

Dep. Variable:	IMDb	**R-squared (uncentered):**	0.957
Model:	OLS	**Adj. R-squared (uncentered):**	0.956
Method:	Least Squares	**F-statistic:**	9009.
Date:	Mon, 03 Apr 2023	**Prob (F-statistic):**	0.00
Time:	01:25:38	**Log-Likelihood:**	-10025.
No. Observations:	6156	**AIC:**	2.008e+04
Df Residuals:	6141	**BIC:**	2.018e+04
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Action & Adventure	-0.0612	0.039	-1.581	0.114	-0.137	0.015
Animation	0.4715	0.080	5.919	0.000	0.315	0.628
Biography	0.5971	0.112	5.352	0.000	0.378	0.816
Comedy	0.1805	0.038	4.729	0.000	0.106	0.255
Crime	-0.0860	0.053	-1.627	0.104	-0.190	0.018
Documentary	1.5249	0.066	23.020	0.000	1.395	1.655
Drama	0.6273	0.037	16.800	0.000	0.554	0.701
Family	0.2220	0.086	2.571	0.010	0.053	0.391
Fantasy	0.1243	0.236	0.527	0.598	-0.338	0.586
Horror	-0.7957	0.062	-12.793	0.000	-0.918	-0.674
Mystery	-0.0188	0.091	-0.206	0.837	-0.198	0.160
Romance	0.2343	0.112	2.086	0.037	0.014	0.454
Thriller	-0.2465	0.073	-3.396	0.001	-0.389	-0.104
Amazon Prime Video	5.1951	0.043	119.701	0.000	5.110	5.280
Netflix	5.6093	0.050	111.500	0.000	5.511	5.708

Omnibus:	1082.040	**Durbin-Watson:**	1.899
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3117.242
Skew:	-0.926	**Prob(JB):**	0.00
Kurtosis:	5.953	**Cond. No.**	16.6

Dep. Variable:	Rotten Tomatoes	**R-squared (uncentered):**	0.925
Model:	OLS	**Adj. R-squared (uncentered):**	0.925
Method:	Least Squares	**F-statistic:**	5063.
Date:	Mon, 03 Apr 2023	**Prob (F-statistic):**	0.00
Time:	01:25:36	**Log-Likelihood:**	-24884.
No. Observations:	6156	**AIC:**	4.980e+04
Df Residuals:	6141	**BIC:**	4.990e+04
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Action & Adventure	1.6370	0.433	3.783	0.000	0.789	2.485
Animation	-0.1052	0.890	-0.118	0.906	-1.850	1.640
Biography	3.3274	1.247	2.668	0.008	0.883	5.772
Comedy	0.3567	0.426	0.836	0.403	-0.479	1.193
Crime	-4.6153	0.591	-7.815	0.000	-5.773	-3.458
Documentary	-4.1070	0.740	-5.547	0.000	-5.558	-2.656
Drama	5.1982	0.417	12.455	0.000	4.380	6.016
Family	-7.6762	0.965	-7.954	0.000	-9.568	-5.784
Fantasy	-4.3494	2.635	-1.651	0.099	-9.515	0.816
Horror	-7.1512	0.695	-10.287	0.000	-8.514	-5.788
Mystery	-2.2306	1.021	-2.186	0.029	-4.231	-0.230
Romance	-6.5840	1.255	-5.245	0.000	-9.045	-4.123
Thriller	-6.6958	0.811	-8.253	0.000	-8.286	-5.105
Amazon Prime Video	44.9224	0.485	92.611	0.000	43.972	45.873
Netflix	53.3171	0.562	94.827	0.000	52.215	54.419

Omnibus:	161.251	**Durbin-Watson:**	1.431
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	393.500
Skew:	0.021	**Prob(JB):**	3.57e-86
Kurtosis:	4.238	**Cond. No.**	16.6

Figure 1.10. Prediction of IMDb and Rotten Tomatoes for each type of movie

	Actual IMDb	Predicted IMDb
0	6.2	5.133895
1	5.7	5.133895
2	6.1	5.133895
3	4.9	5.133895
4	6.1	5.133895
...	...	...
15171	6.1	6.236594
15182	5.8	6.236594
15186	5.7	6.236594
15194	7.3	6.236594
15232	5.3	6.236594

6156 rows × 2 columns

	Actual Rotten Tomatoes	Predicted Rotten Tomatoes
0	81.0	46.559435
1	74.0	46.559435
2	72.0	46.559435
3	50.0	46.559435
4	79.0	46.559435
...	...	...
15171	38.0	58.515270
15182	39.0	58.515270
15186	27.0	58.515270
15194	42.0	58.515270
15232	34.0	58.515270

6156 rows × 2 columns