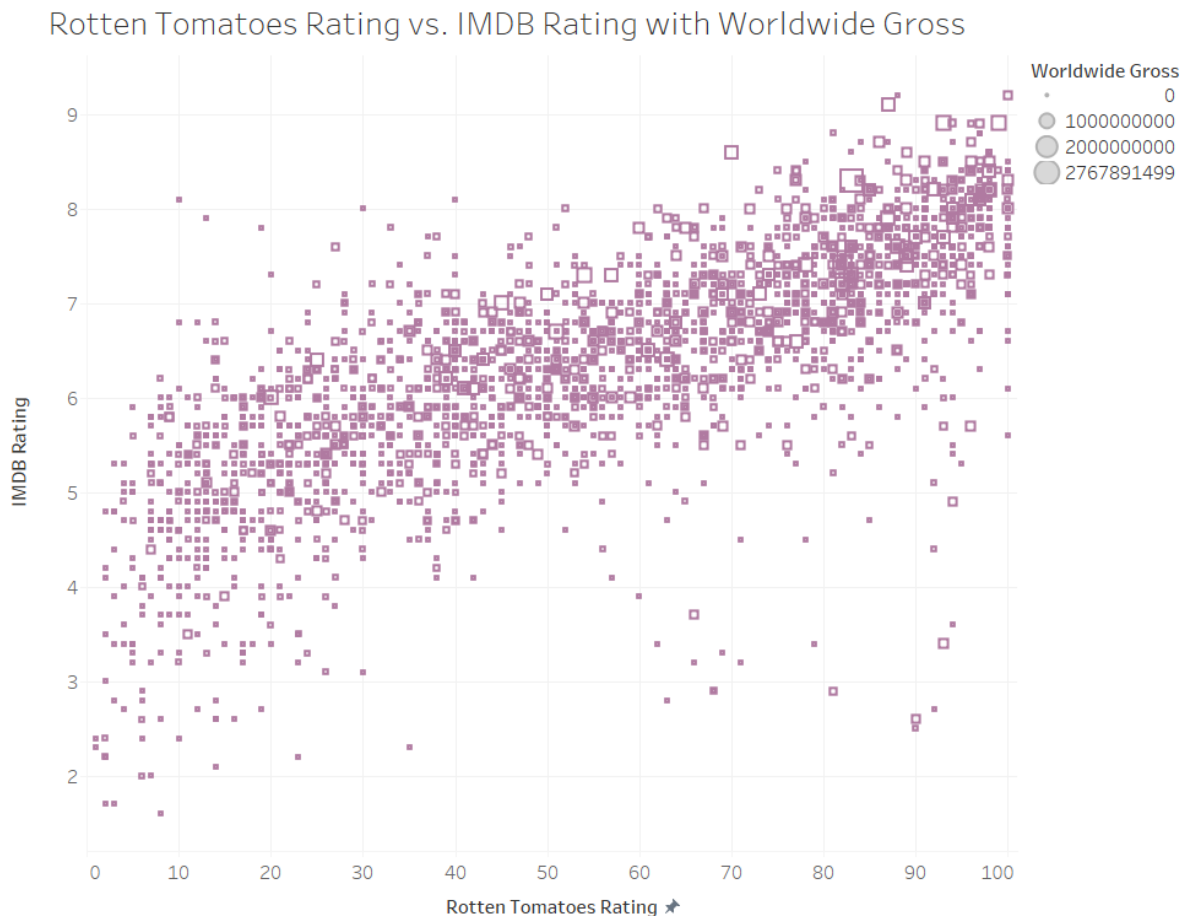


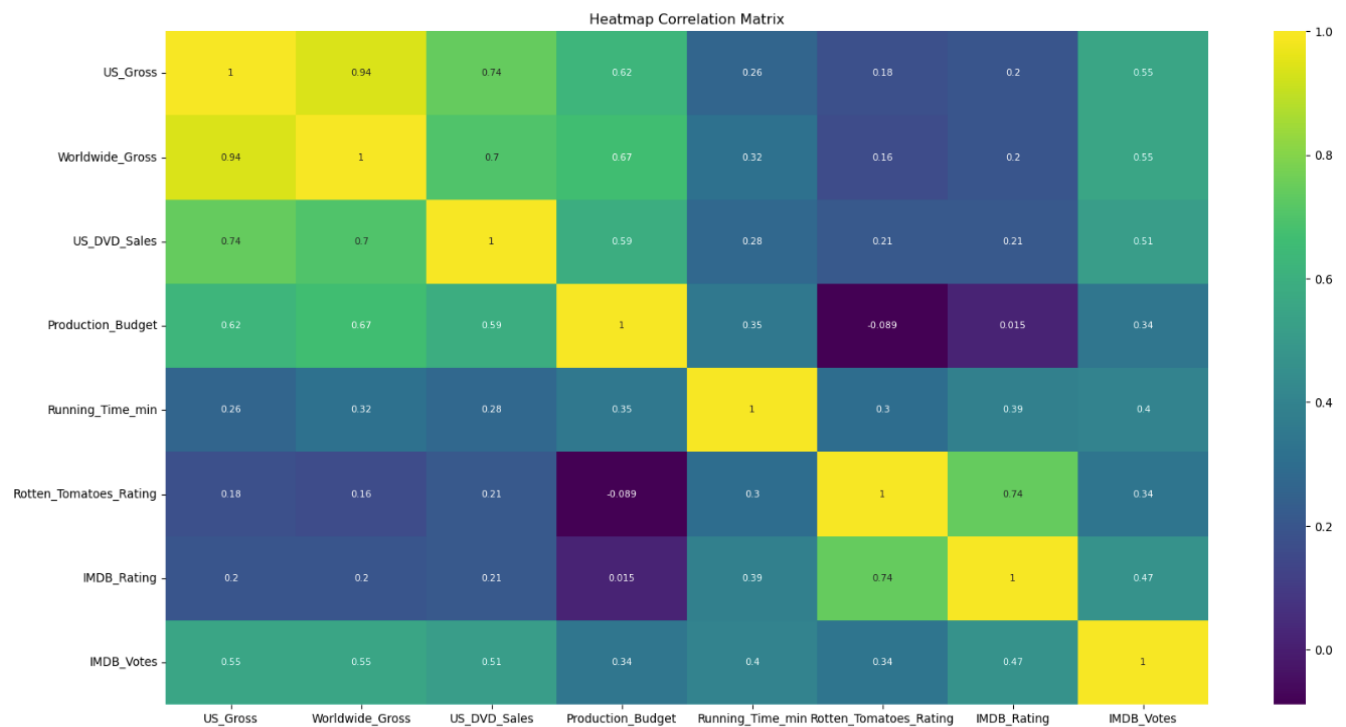
Visual Analytics Group Work



The purpose is to visualise a correlation between the Rotten Tomatoes rating and the IMDb rating and how this influences the worldwide gross of each movie.

On this scatterplot, the RT rating is on the x-axis while the IMDb rating is on the y-axis. This enables us to see the correlation between them. Another variable is the worldwide gross per movie. This is represented by size because we can better distinguish smaller differences between them than with another design such as colour. Finally, the points are empty squares to symbolise a movie screen. Some limitations are that some points overlap each other but this is because two movies may have identical ratings.

On this graphic we can see the linear correlation between the two ratings as most of the points form a straight increasing line. This means that the two ratings agree with each other, so a higher RT rating gives a higher IMDb rating. RT is based on the percentage of positive reviews and IMDb (Internet Movie Database) is a vote out of 10. Additionally, when we compare the size of the points, we can tell that the larger ones, signifying a greater worldwide gross, have higher ratings. This would mean that better movies make more money.



HEATMAP using CORRELATION MATRIX

Purpose:

The purpose of using a heatmap correlation matrix is to identify relationships, patterns between several variables in our dataset. The colour gradients make it easy to spot the areas of high correlation. This visualization helps us to identify which variables are most significant for our analysis. The values range from -1 to 1; 1 being the strongest positive correlation and -1 being the strongest negative correlation and 0 being no correlation between the variables at all.

Justification:

Position & Length: Data points are arranged in a grid, and it allows for a clear, structured comparison between different categories. Unlike other plots, the length in a heatmap refers to the size of the grid cells and they are all uniform. This makes it consistent for an unbiased visual.

Colour: The use of colour gradient represents the intensity of a value. The changes in the colour gradient make it easier to spot the stronger correlations, lighter colours indicate a stronger correlation between the variables and darker colour indicate a weaker correlation.

Size: As explained in 'Length', the size for each grid is fixed, making each cell has the same dimensions. The absence of varying sizes makes it easier to focus purely on colour differences, making it clean and simple.

Design trade-offs: The limitation of this heatmap matrix is that it could only be used for numerical values therefore could lead to an incomplete understanding of the data. If there are datatypes other than float or integers, they could not be included, or these must be converted into numerical data if possible and then could be included in this plot. The other

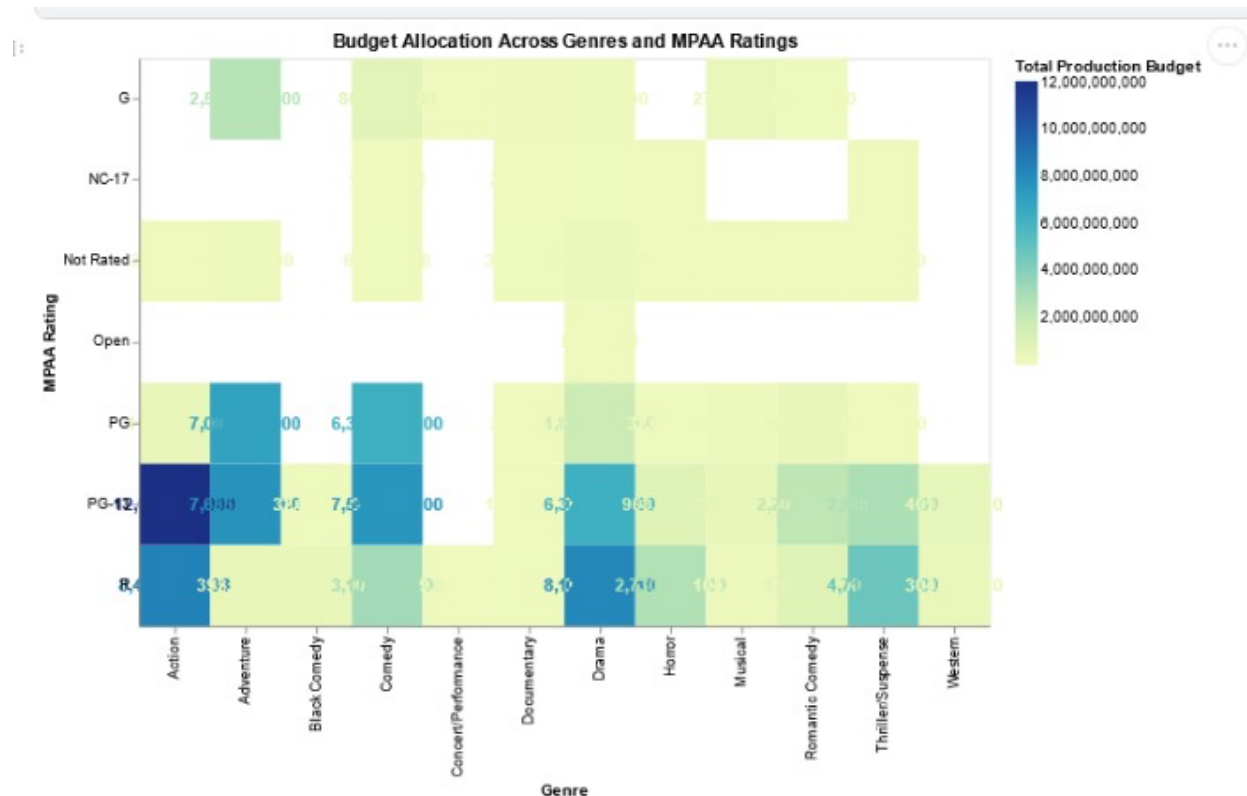
limitation is that correlations do not consider how data is distributed, so if there are outliers, the accuracy of the correlation will be affected.

Interpretation:

When two variables have a strong positive correlation, an increase in one variable is accompanied by an increase in the other, and a decrease in one lead to a decrease in the other. This means the relationship between the variables always moves in the same direction. If we set a minimum of %60 threshold between variables, the following variables have significant positive correlations.

- 1- US_Gross & Worldwide_Gross – 0.94 = As US Gross sale increases, naturally the Worldwide Gross sale increases too. They have a very strong positive correlation.
- 2- US_Gross & US_DVD_Sales – 0.74 = Movies with higher gross sales tend to have a higher US DVD sale. (There are many missing values in US_DVD_Sales so it's important to approach this analysis with caution.)
- 3- US_Gross & Production_Budget – 0.62 = As the Production Budget of the movie increases, US Gross sales tend to increase as well.
- 4- Worldwide_Gross & US_DVD_Sales – 0.70 = Movies with higher worldwide gross sales tend to have a higher US DVD sale. There are many missing values in US_DVD_Sales so it's important to approach this analysis with caution.)
- 5- Worldwide_Gross & Production_Budget – 0.67 = Movies with higher production budgets typically achieve higher Worldwide Gross sales.
- 6- IMDB_Rating & Rotten_Tomatoes_Rating – 0.74 = There is a strong positive relationship, movies with higher IMDB Rating tend to have a higher Rotten Tomatoes Rating. (There are many missing values in Rotten_Tomatoes_Rating so it's important to approach this analysis with caution.)

When two variables have a strong negative correlation, an increase in one variable corresponds to a decrease in the other, and a decrease in one lead to an increase in the other. This means the relationship between the variables always moves in opposite directions. There are no strong negative correlations on this matrix.



Purpose: illustrate how much financial investment is directed toward various genres under different content ratings.

Justification: A treemap is ideal for this data because it efficiently conveys both the categorical nature of the genres and MPAA ratings, as well as the quantitative variable (production budget). The size and color of each rectangle represent the total production budget, allowing for a clear visual comparison across multiple categories. By organizing the data in a grid-like structure, it is easy to see the proportion of budgets allocated to each genre and rating. The use of hue as the color encoding is effective for this quantitative variable, providing a gradient that represents the varying magnitudes of budgets. The tooltip adds interactivity and enhances the detail in interpreting specific budget values.

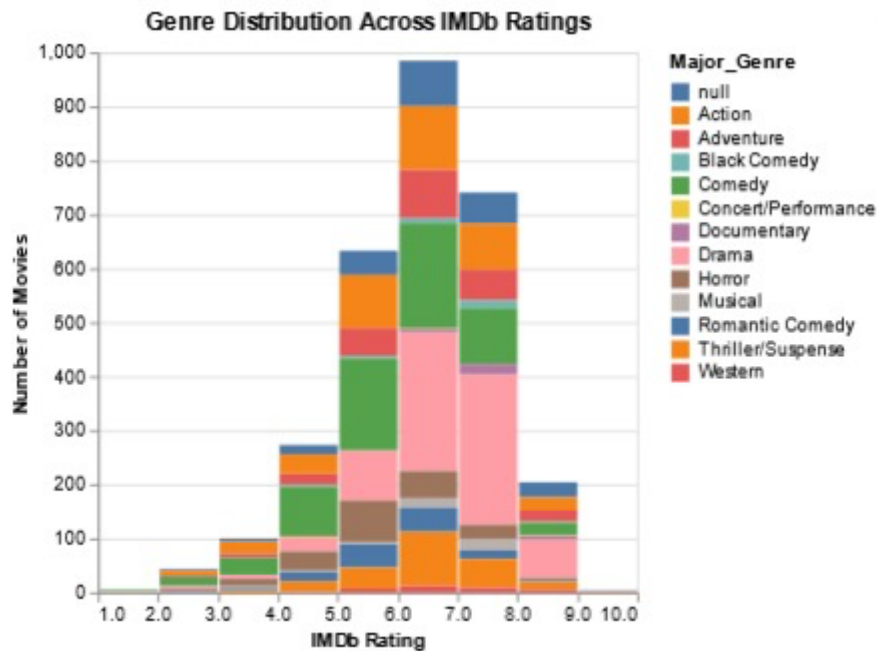
Interpret : Action and Drama genres dominate the budget allocation, particularly within the PG-13 and R ratings, indicating a strong focus on content with broad audience appeal and mature themes. The high budgets allocated to these genres suggest that they are considered lucrative and popular choices for producers. Also receiving moderate funding across a spread of MPAA ratings are the Comedy and Adventure genres, reflecting a steady demand for lighter or family-friendly entertainment. This agrees with the versatility of these genres in drawing diverse audiences, although they are generally not budgeted as high as Action or Drama films.

The PG-13 and R ratings are the most frequent ratings for big-budget films across all genres, which testifies to a tendency for producers to invest large sums of money in productions oriented toward a wide but mature audience. This implies a marketing focus on film products, proving accessible both to teenagers and adults who form great divides in the cinematic audience.

Genres like Horror and Black Comedy tend to get relatively smaller budget shares and show more scattering across the rating groups. This might be an indication of the niche appeal of

these genres, targeting more specialized audiences, and may not deserve huge financial investments due to their divisive nature.

Less popular genres like Westerns and Musicals get the least budgetary resources for each rating, suggesting they really are the least prioritized in terms of financing. Such infrequency could signal their niche appeal and narrow market demand, which, in turn, indicate smaller capital investments by production companies.

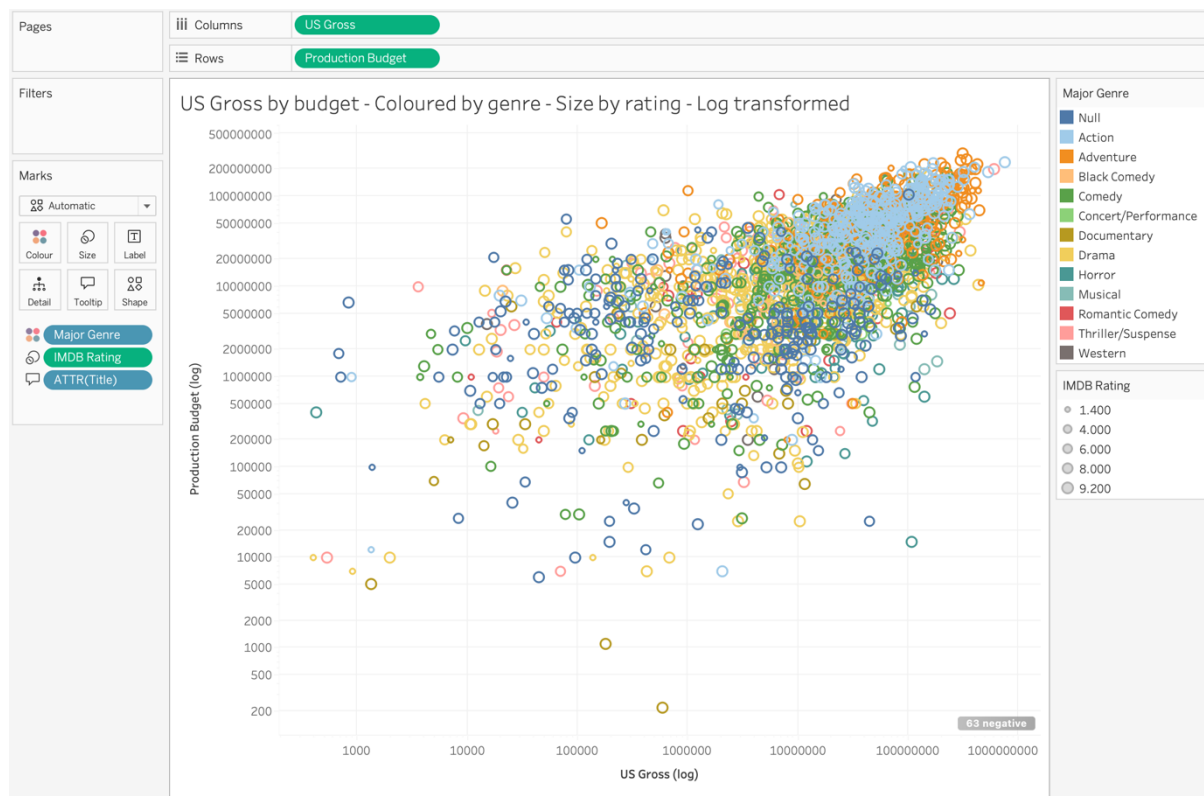


Purpose: To compare the distribution of IMDb ratings across different movie genres, showcasing how the frequency of movies per genre varies within each IMDb rating range.

Justification: The number of movies per IMDb rating is the key variable in this analysis, so it is represented using the aligned length of stacked bars, making it easy to compare quantities. IMDb rating is displayed along the x-axis as a continuous variable, which helps in identifying rating trends across genres. Color (hue) is used to differentiate between genres, which is appropriate for this nominal variable with multiple distinct categories, enabling clear genre comparison within each rating group.

Interpret:

Most films, regardless of genre, are concentrated between the 5.0 and 7.5 IMDb ratings; this means that on average, most films are rated good. Indeed, drama is the most frequent genre for all rating ranges, but leading in mid-range from 5.0 to 8.0 shows its mass market appeal and yet oscillates in reception. Genres like Adventure, Action, and Thriller/Suspense do well in the higher ratings, 8.0+; this indicates that when these films do well, they are rated a little more favorably. Conversely, Horror and Black Comedy are mostly on the lower ratings due to their highly polarizing genres or too specific niche appeal. Western and Musical genres are much rarer throughout the dataset, reflecting their more niche production and average reception since they are not in the majority of any rating category.



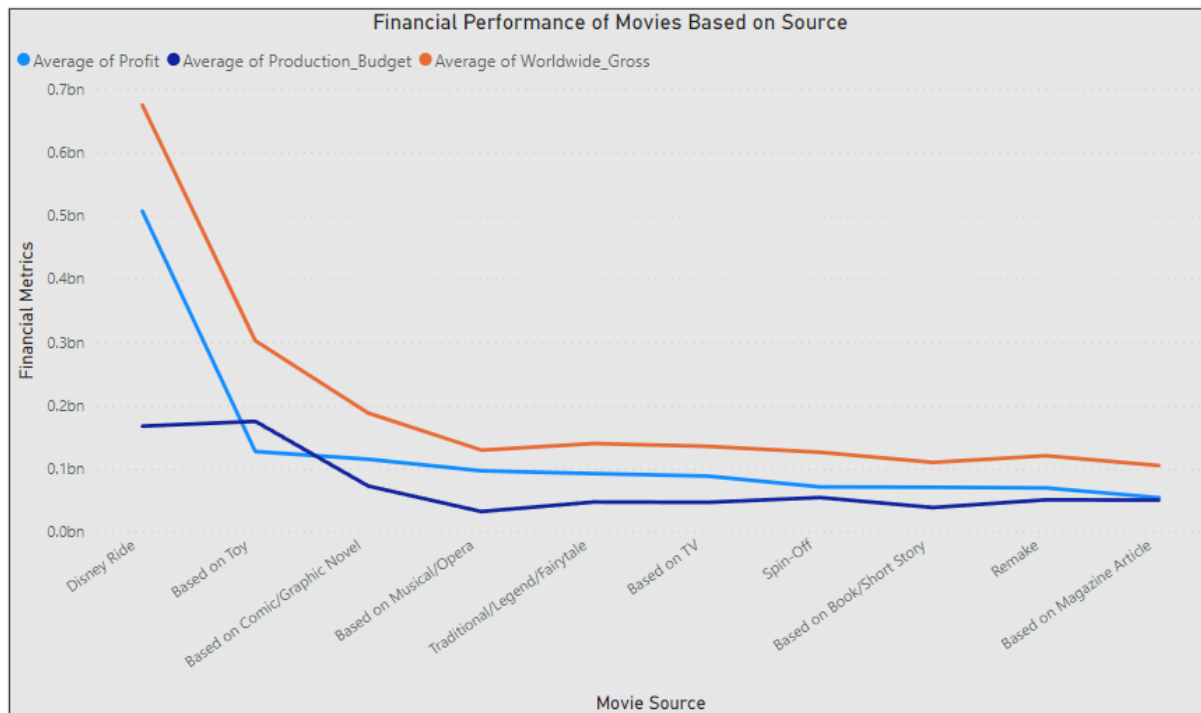
The purpose of this scatter plot is to explore the relationship between production budget and US gross revenue of the film across various genres. A logarithmic transformation has been applied to both axes to better visualise the distribution of the dataset. The scatter plot uses colour to represent genre and sizing of the marks represents the IMDb rating. The aim of this analysis is the observe whether budget and genre correlate with a movie's financial success in the US.

The visual effectively utilises position, scale, colour and size to represent 4 important aspects of the movies dataset. Displaying US gross revenue on the x-axis and production budget on the y-axis using a logarithmic scale is effective for financial data; by condensing the wide range of values across the dataset, the visualisation makes it easier to detect patterns and trends across genres with diverse budgets. Through colouring the marks, it allows clear visual distinction between the genres, allowing readers to see clustering patterns ain budget and revenue. Through sizing by IMDb rating, with larger circles indicating higher ratings, we have added a nuanced layer of information that allows the viewers to see relationships between audience reception and financial success of the movie.

While the visuals experience design trade-off, like overlapping data marks due to high density, particularly among genres with similar budgets and revenue ranges. However, the use of logarithmic scaling, genre colour coding, and size differentiation provide sufficient visual cues to identify general trends and patterns within the dataset, even in the more crowded areas of the visual.

The scatter plot uncovers several insights about genre-based patterns in budget, revenue, and IMDb ratings. Action, adventure, and comedy films, represented in green, light blue, and orange show a broad spectrum in both production budget and gross revenue. Many of

the high grossing films fall within these genres, particularly those with high production budgets, suggesting their significant financial appeal within the wider film industry. However, drama and documentary films, shown as yellow and brown, tend to have smaller budgets and lower earnings. However, it is worth noting that there are outliers in the drama genre that illustrate substantial revenue can be achieved even with a modest budget. Highly rated films, based on their IMDb ratings and shown as larger circles in the scatter plot are distributed relatively evenly between genres, but more frequently around higher budget movies. This implies a trend that higher-budget movies gain improved audience ratings, though it cannot be inferred that there is a straightforward correlation between budget size and IMDb rating. The logarithmic transformation applied to the axes highlight the positive association between budget and gross revenue, yet shows instances when lower-budget films achieve impressive box office numbers, particularly in the horror genre. The findings suggest that many genres can thrive financially even with limited production budgets, highlighting individual's unique preferences and contrasting audience demands.



Purpose:

This line graph shows the relationship between the "source" of movies and three key financial indicators: profit, production budget, and worldwide gross. The aim is to analyse how different movie sources affect these financial metrics, providing insights into which source materials are most profitable and where higher production investments yield the most returns.

Justification:

Axes: The x-axis represents different movie sources, categorizing films based on their origin or inspiration, such as toys, comics, or novels. The y-axis shows financial metrics in billions, specifically the average profit, average production budget, and average worldwide gross, allowing for a comparison of how movies from each source performed financially.

Lines: The graph features three distinct lines:

- **Orange:** Represents the average worldwide gross, showing the global earnings of films by source.
- **Blue:** Depicts the average profit, showing how much profit films from each source generated.
- **Light blue:** Shows the average production budget, indicating how much was invested in producing these films.

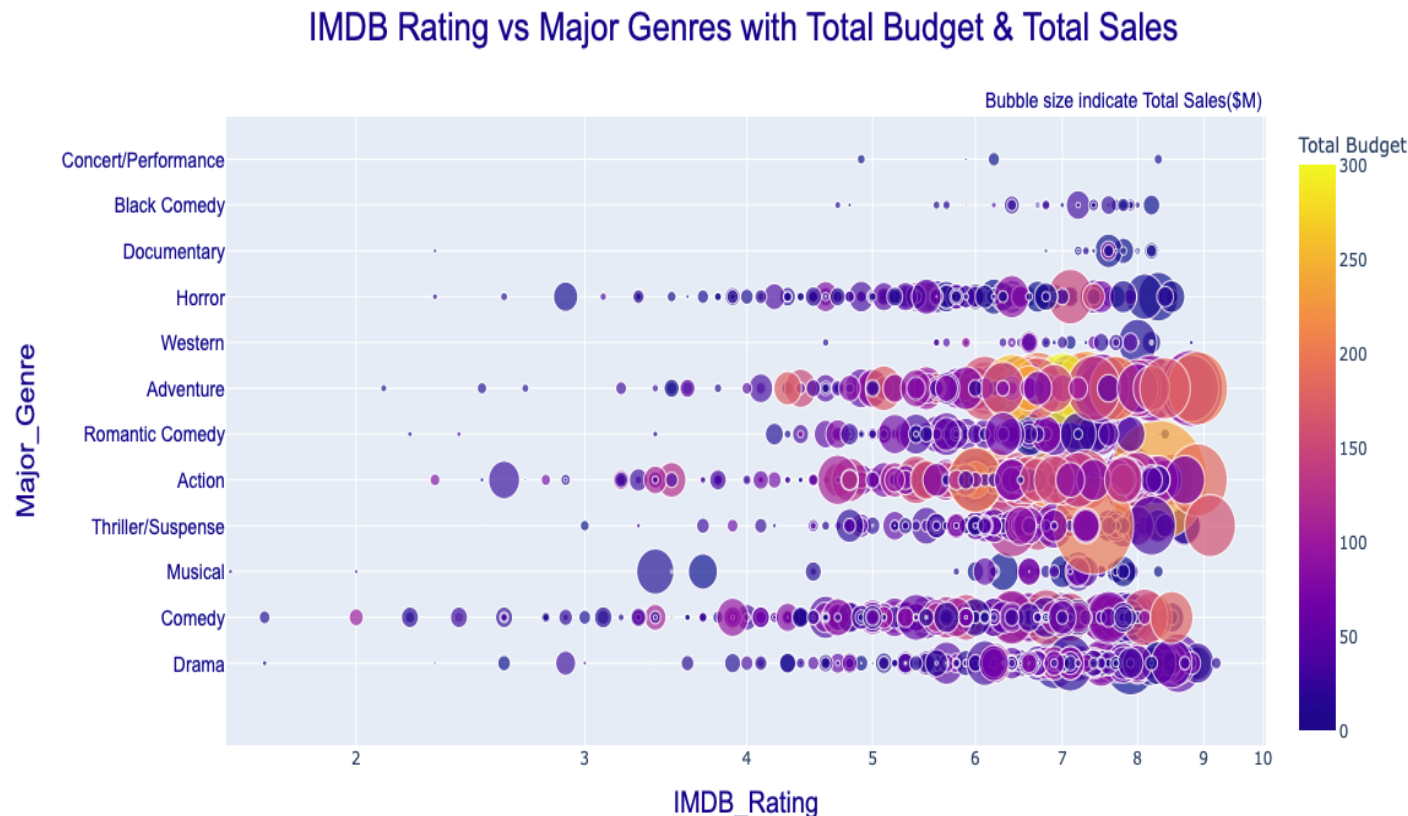
These lines help compare sources in terms of investment versus returns.

Colour Scheme: The choice of orange, blue, and light blue helps distinguish between the different metrics clearly. The colours ensure readability and facilitate quick visual comparisons across movie sources.

Interpretation:

The graph reveals significant trends in the financial performance of films based on their source material:

1. Disney Ride: This source has the highest production budgets and worldwide gross, resulting in higher profits compared to other categories.
2. Based on Toys: Films based on toys show a sharp decline in profit, production budgets, and worldwide gross, indicating overall less financial success.
3. Based on Comic/Graphic Novels: This category demonstrates moderate investment in production and a steady worldwide gross, but yields higher profits compared to many other categories. This suggests that films based on comics tend to perform well relative to their budgets.
4. Based on TV/Traditional Legends/Spin-Offs: These sources exhibit lower production budgets and profits, indicating that they are typically less profitable.
5. Remakes and Magazine Articles: These categories display similar trends, showing slightly lower financial performance than the top categories, although they maintain a moderate return on investment.



Purpose:

Plotly's interactive bubble scatter plot is a very good tool for visualizing multi-dimensional data, for the relationships between movie genres and IMDB ratings, with worldwide gross sales and production budget. We can observe which movies from different genres financially did very well even though they did not always have the highest IMDB ratings or vice versa. This plot is a good starting point in analyses.

Justification:

Position & Length: IMDB Rating is placed on X-axis and shows how audience and critical reviews vary across different movies. Movies with higher ratings can easily be identified and can be compared to those with lower ratings across genres. Major Genre is placed on Y-axis and allows a clear comparison of how different genres relate to the IMDB Rating and their overall financial performance.

Colour: The colour of each bubble represents the Production Budget, providing the level of financial investment in each movie. Movies with higher production budgets are indicated with lighter colours, allowing users to quickly differentiate between low-budget and high-budget films. This highlights potential trends, such as whether higher-budget movies tend to achieve higher IMDB ratings or specific genres with higher budgets tend to achieve lower IMDB Ratings.

Size: The bubble size represents the financial performance of a movie for worldwide total sales, with larger bubbles are identified as higher sales at the world box office. This

makes it easy to spot the most successful movies financially even though they are not highly rated. Hovering over any bubble provides detailed information, such as the movie title, total sales, and budget, offering further context.

Design trade-offs: The plot's interactivity improves the user experience by allowing detailed exploration of data points without cluttering the view. However, including many variables can make the plot more challenging to understand at first, especially for viewers unfamiliar with this type of visualization. Also it makes it harder to differentiate movies when movies have the same IMDB Rating.

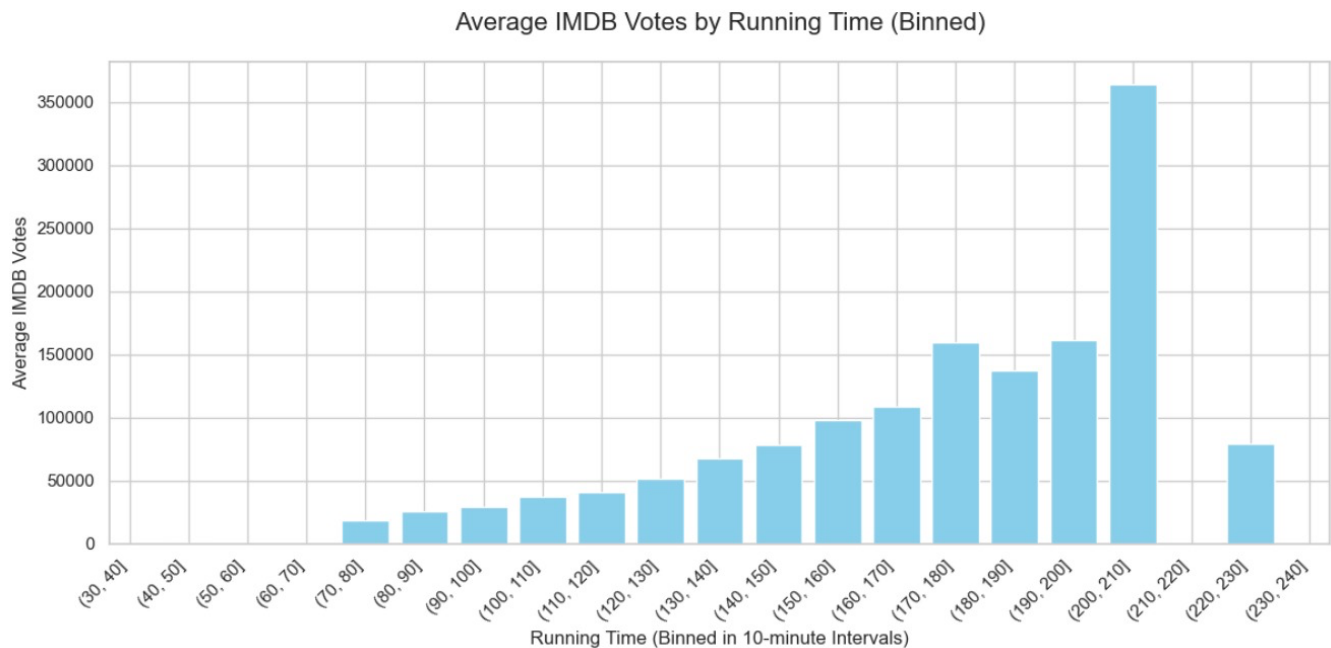
Interpret:

Many of the movies have an IMDB Rating of 4 and above.

Larger bubbles signify movies with higher sales, especially in three main genres, Action, Adventure and Thriller. Some of these high sales genres also have the highest budgets shown with lighter colours even though they are not always rated close to 10. 'Titanic' is a very good example of being a very profitable Thriller movie despite achieving only 7.4 rating. Whereas another Thriller movie 'Inception' has a similar budget with a decent profit but not as high as Titanic, achieves a rating of 9.1. In Adventure category, 'Avatar' is the most profitable movie of all with a rating of 8.3. Another similar example is Adventure movie 'Spiderman-3' with one of the highest budgets with a high profit but again a lower rating of 6.4

Looking at this plot, we can safely say a high IMDB rating do not always produce a profitable movie.

There are not enough movies for Concert, Black Comedy or Documentary categories when compared to the rest of the categories.



Purpose:

This bar graph reveals “audience engagement” across movie runtimes, measured by “average IMDB votes”. Standard runtimes (70–130 mins) draw steady interest, while extended lengths (130–200 mins) show rising engagement, likely due to richer storytelling. A striking peak at 200–210 mins depict high appeal for epic films, which resonate deeply with audiences. Insights from this graph can guide filmmakers on optimal runtimes to maximize reach, highlighting how runtime directly shapes a movie’s impact and viewer engagement.

Justification:

Position & Length:

In the bar graph, the “x-axis” represents “running time” binned into “10-minute intervals”, indicating the length of the movies. The “y-axis” shows the “average IMDB votes”, reflecting the level of viewer engagement for movies within each runtime category. This visualization helps illustrate how different movie lengths correlate with audience interest as measured by IMDB ratings.

Colour: The bars in the graph are coloured **sky blue**, providing a bright and visually appealing contrast against the white background. This colour choice enhances readability, making it easier to distinguish between different runtime intervals. We have kept the colour consistent since we are only working on 2 parameters that is running time and IMDB votes.

Size:

The varying heights of the bars in the graph indicate fluctuations in **average IMDB votes** across different **runtime intervals**. Taller bars represent higher viewer engagement, while shorter bars indicate lower average votes, revealing audience preferences for specific movie lengths. This variance not only highlights trends in viewer interest but also emphasizes the correlation between runtime and audience votes.

Design trade-offs:

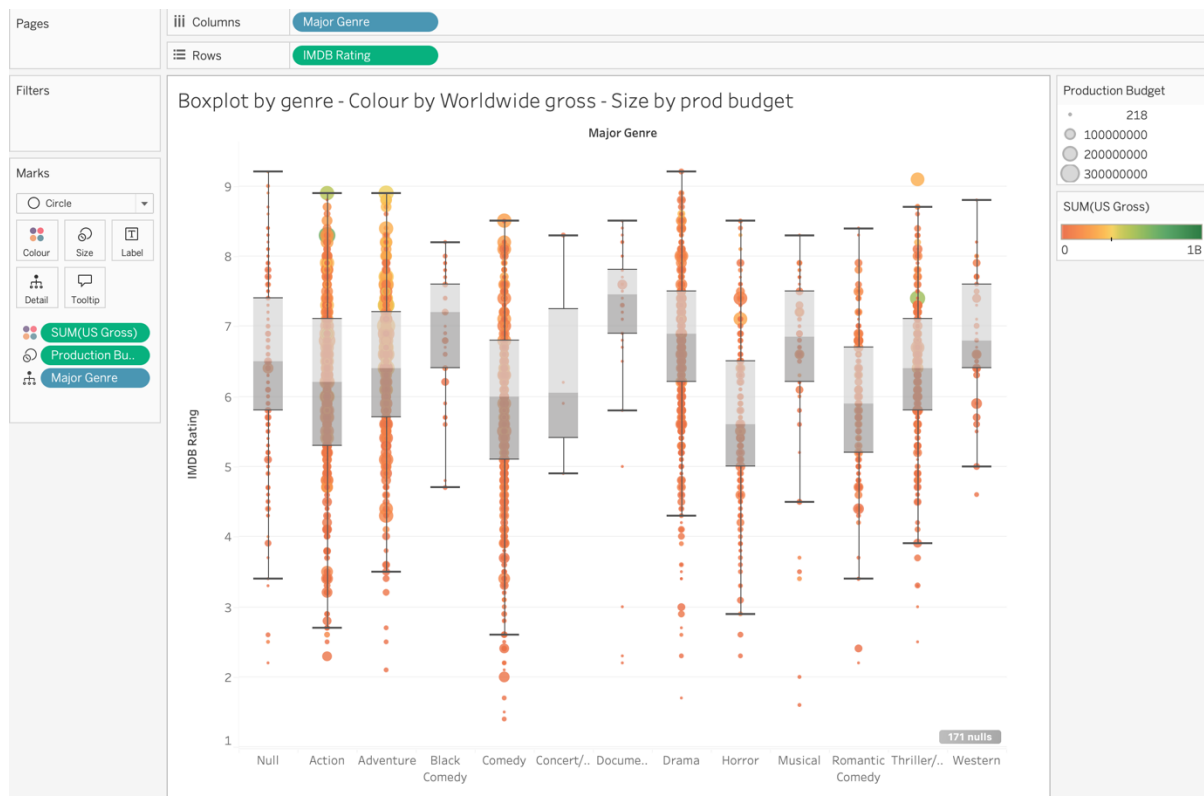
The bar graph's cohesive use of sky blue enhances visual harmony, making it easy for viewers to focus on the data. Its larger size improves readability and allows for detailed labels, ensuring that the information is accessible and engaging.

Interpret:

The graph reveals significant trends in audience engagement based on film runtimes:

1. "Short Films (70-130 minutes)": With less than 50,000 votes, these films may struggle for visibility and appeal, indicating that audiences prefer more substantial narratives.
2. "Moderate-Length Films (130-160 minutes)": Receiving 50,000 to 100,000 votes, this range balances viewer commitment with narrative depth, attracting a larger audience.
3. "Long Films (160-200 minutes)": The spike to over 100,000 votes suggests that epic or critically acclaimed films resonate strongly, likely due to ambitious storytelling.
4. "Blockbusters (200-210 minutes)": A significant peak with over 350,000 votes indicates a strong preference for blockbuster films, reflecting broad marketing and viewer engagement.
5. "Extended Runtimes (220-230 minutes)": Again seeing 50,000 to 100,000 votes, this suggests interest from dedicated fans but lower overall popularity compared to blockbuster films.

These insights highlight how runtime influences viewer engagement, informing production and marketing strategies.



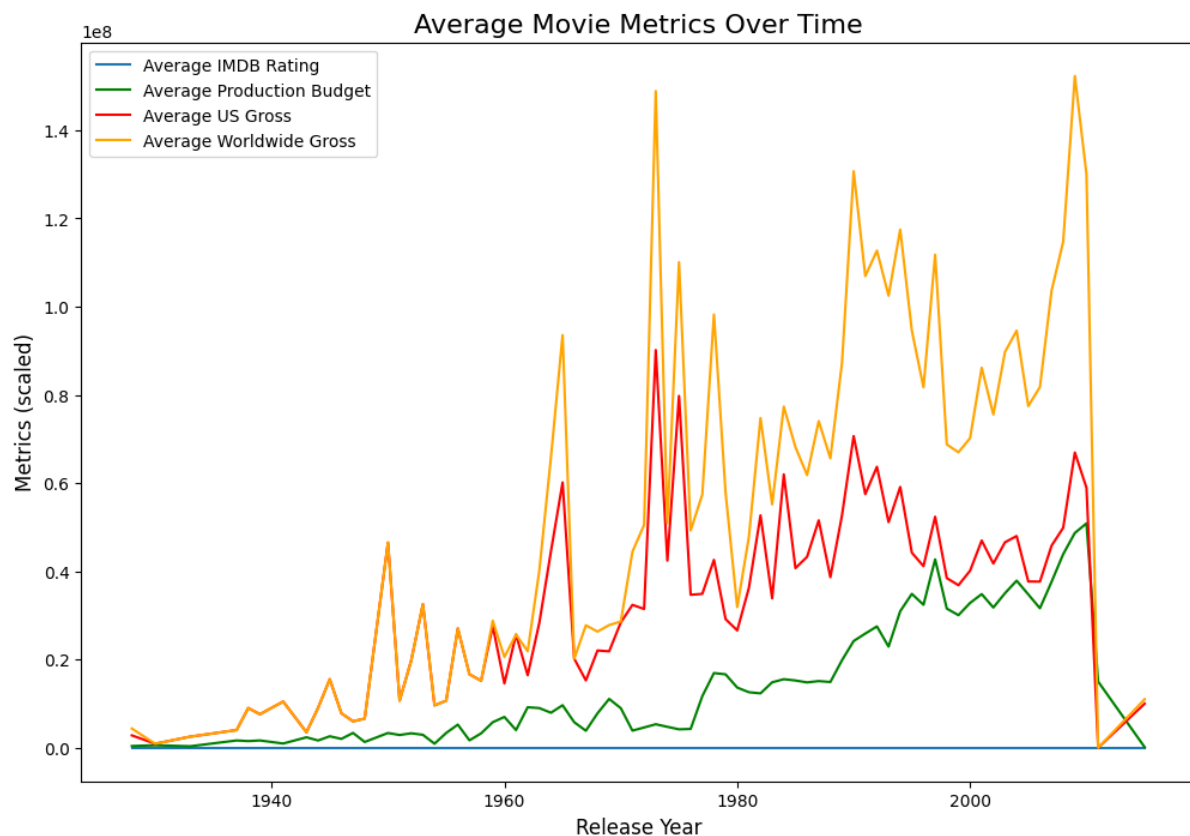
The purpose of this boxplot is to understand the distribution of IMDb ratings across movie genres, while incorporating financial metrics by superimposing through colour and sizing. The colour of each mark behind the boxplot indicated US gross, using a red-amber-green scale with red representing low gross and green indicating higher, each mark is also sized based on the production budget of the movie. The design tries to reveal relationships between a movie's genre, audience rating, and financial performance. The aim of this analysis then is to offer insights into how the genre and budget of a movie may impact the commercial reception and its financial success.

The choice of visualisation can be well justified for the purpose of this analysis. With genre on the x-axis, and IMDb rating on the y-axis, the visualisation forms a boxplot that displays the spread and central tendency of ratings within each major genre. The arrangement easily shows the comparison of median rating, interquartile ranges, and any outliers across the genres. With colour being used to show US gross, this provides an additional dimension to the visual, highlighting the financial performance across the genres, so that viewers can easily assess the revenue-based success of the films. Furthermore, production budget is represented by the size of the marks, with larger marks indicating larger budget, this allows a further layer of financial information that views can easily digest and gain further insight to budgetary trends across genres.

There are numerous trade-offs within the design, particularly as the plot can become clustered due to the high number of data points, and using sizing techniques within the data marks, which can obscure smaller-budget films or those with moderate ratings. The US gross variable also experiences negative skewness, which then visualise many of the marks as a red, which struggles to truly show the distribution of this across some of the genres. On

the whole, the use of colour and size of marks does improve the ability to see more patterns within the data, despite the density of the data.

The analysis successfully visualised several trends within the dataset. Genres like action, adventure and comedy display a wide range of IMDb ratings with many outliers, yet their median ratings fall between 6 and 7. Black comedy, documentaries and musicals have comparatively higher median ratings, displaying a strong audience appreciation within these genres, even with generally lower budgets, shown through the smaller marks behind the box plot. Films with higher budgets are scattered across all genres, however genres with higher US gross (green and yellow marks) are more prevalent in adventure, action and comedy, which shows that they perform well financially, even though their ratings vary. In contrast, the horror genre shows a much lower average rating and smaller production budgets, with very few instances of high-grossing films, highlighting the fact that while the genre does not attract high ratings or revenue, they do not cost as much to produce.



Purpose:

This line plot aims to Visualize changes in key movie metrics – Average IMDB Rating, Production Budget, US Gross, and Worldwide Gross – over time, specifically from the earliest recorded release years up to 2024. By tracking these metrics, the plot highlights long-time trends in the movie industry, such as the impact of increasing budgets on revenue and any shifts in audience rating over decades, this visualization provides insights into how the financial and audience reception aspects of movies have evolved historically.

Justification:

- Each metric is represented by a distinct coloured line, with IMDB Rating (blue), Production Budget (green), US Gross (red), and Worldwide Gross (orange). The use of colour for each metric provides clear differentiation, allowing viewers to distinguish and track each trend individually.
- The y-axis represents the scaled values of the metrics, making it easy to observe the relative trends and patterns. Positioning is effective for comparisons across metrics because each change is plotted consistently over time on the x-axis (Release Year).
- Given the different magnitudes of the metrics, the y-axis is scaled to facilitate comparisons without overwhelming the plot. This scaled axis allows for a balanced view that brings attention to relative changes rather than absolute values.
- Giving multiple metrics in one graph could be overcrowded, but by using different colours and carefully scaling, the plot came up clearly without overwhelming the viewer.
- By limiting the data to years up to 2024, we ensure the analysis reflects observed data rather than any speculative future data.

Interpretation:

- The plot represents that the Worldwide Gross and US Gross have both experienced significant growth over the years, with the Worldwide Gross consistently higher, which shows the increasing global appeal and market for films. Peaks in the 1980s and later years may suggest periods of blockbuster hits with high revenue.
- The green line for the Average Production Budget represents gradual growth over time, keeping in line with increases in both US and Worldwide Gross. This correlation represents that higher budgets can contribute to higher revenue, even though trends indicate it is not a direct one-to-one relationship.
- IMDB Rating remains flat compared to the financial metrics, which represent the audience rating that has been stable over the decades. This shows that during production budgets and revenues have soared, audience appreciation, as measured by average IMDB ratings, has not significantly increased.