
DATA ANALYTICS AND VISUALIZATION USING TABLEAU UTILITARIAN FOR IMDB MOVIES & SHOWS

Raparathi Bhoomika¹, Koormachalam Amulya², Dara Hannah Saejal³, Dr. Diana Moses⁴

^{1,2,3}Student, Department of Artificial Intelligence & Data Science, Methodist College of Engineering and Technology, Abids, Hyderabad, India.

⁴Associate Professor, Department of Computer Science & Engineering, Methodist College of Engineering and Technology, Abids, Hyderabad, India.

ABSTRACT

In the rapidly evolving digital entertainment landscape, understanding the dynamics of movie and show production, quality, and popularity is crucial for content creators, distributors, and audiences. This paper presents a comprehensive visual analysis of an IMDb dataset, aiming to uncover trends and patterns that illuminate the evolution of the entertainment industry. Utilizing the powerful capabilities of Tableau Desktop, we analyze various facets of the dataset, including IMDb scores, release years, genres, production countries, age certifications, runtimes, and viewer engagement through IMDb votes.

The analysis is organized into thematic dashboards, each focusing on different aspects of the data. The "Trends Over Time" dashboard examines how the quality and quantity of content have changed over the years, revealing significant trends in production activity and audience preferences. The "Genre Analysis" dashboard delves into different genres' popularity and average ratings, highlighting viewer tastes and industry focus shifts. The "Country and Production Insights" dashboard identifies leading countries in content production and their respective quality ratings, providing a global perspective on the entertainment landscape. The "Ratings and Popularity" dashboard explores the relationship between IMDb scores and viewer engagement, shedding light on factors that drive content popularity.

Key findings from the visual analysis reveal that certain countries, such as the United States, United Kingdom, and India, are prolific producers of movies and shows, with countries like Japan and South Korea consistently receiving high ratings for their content. Genre analysis indicates a persistent preference for Drama and Thriller genres, while temporal trends highlight the impact of streaming platforms on genre popularity and production volumes. Additionally, the correlation between IMDb scores and votes underscores the importance of quality in driving viewer engagement and popularity.

By leveraging visual analytics, this study provides valuable insights into the intricate dynamics of the entertainment industry. The findings offer actionable information for stakeholders, including content creators, distributors, and platforms, aiming to navigate and thrive in the competitive digital entertainment landscape. This paper demonstrates the efficacy of Tableau Desktop in transforming raw data into meaningful visual narratives, enhancing our understanding of what drives success in movies and shows today.

1. INTRODUCTION

In today's data-driven world, extracting actionable insights from vast datasets is critical across various industries. The entertainment industry, particularly film, is no exception. With the proliferation of streaming platforms and the continuous production of new content, understanding audience preferences, trends, and factors influencing movie success has become paramount for filmmakers, studios, and distributors.

This paper explores the application of data analysis and visualization techniques using Tableau, a powerful data analytics and visualization tool, to delve into the vast dataset provided by IMDb (Internet Movie Database). IMDb, one of the most comprehensive sources of film-related information, offers a treasure trove of data on movies and shows, including title, type, release year, age certification, runtime, genres, production, seasons, IMDb ID, IMDb score and IMDb votes. Tableau can connect to files, and relational and Big Data sources to acquire and process data. The software allows data blending and real-time collaboration, making it unique. Data analysis is speedy with Tableau and the visualizations created are in visual form with dashboards and worksheets. A Tableau dashboard allows for multiple visualizations to be seen within a single view. It's often used to show only the most important data and is sometimes personalized. It works by connecting to data stored in various places. It can pull data from any source imaginable. From simple Excel sheets to PDFs, complex databases like Oracle, and the next level of cloud, such as Amazon web services, Microsoft Azure SQL database, and Google Cloud SQL, can be extracted by Tableau. Therefore, we introduce Tableau and present the procedure of using Tableau for the interactive visualization and analysis of IMDb Movies & Shows data to encourage its widespread use. Tableau is a new-age data analytics and visualization tool that provides flexibility and ease of use with a smooth experience for the users.

Our approach focuses on a practical perspective, aiming to extract actionable insights that can inform decision-making processes in the film industry. By harnessing the capabilities of Tableau, we strive to uncover patterns, correlations, and trends within the IMDb dataset, shedding light on various aspects of movie production, distribution, and reception.

2. DATASET DESCRIPTION

The IMDb dataset used in this analysis encompasses various fields crucial for understanding the characteristics and performance of movies and shows taken from Kaggle. The key attributes include:

- **Title** - Name of the movie/show.
- **Type** - It is a Movie or a Show.
- **Release Year** - The year in which that movie was released.
- **IMDB Score** - Rating of the movie/show at the IMDb site.
- **IMDB ID** - ID of the movie/show at the IMDb site.
- **Runtime** - Total runtime of the movie/show.
- **Genres** - Genre of the movie.
- **Production Countries** - The country in which the movie/show is produced.
- **Seasons** - Number of seasons of the show.
- **IMDB Votes** - Votes of the movie at the IMDb site.
- **Age Certification** - Age certificate earned by that movie/show.

These fields collectively provide a rich tapestry of information that allows for nuanced exploration into various aspects of content creation, distribution, and reception within the entertainment industry.

3. INTRODUCTION TO TABLEAU

Tableau is a powerful tool used for data analysis and visualization. It allows the creation of amazing and interactive visualization and that too without coding. Tableau is very famous as it can take in data and produce the required data visualization output in a very short time. It can elevate your data into insights that can be used to drive your action in the future.

Tableau is a great data visualization and business intelligence application that can be used to report and analyze massive amounts of data. Salesforce purchased Tableau in June 2019 an American firm founded in 2003. It enables users to build various charts, graphs, maps, dashboards, and stories for visualizing and analyzing data to aid in business choices. Tableau offers several unique and fascinating features that make it one of the most popular Business Intelligence (BI) applications.

3.1 Tableau Features

- Tableau supports powerful data discovery and exploration that enables users to answer important questions in seconds
- No prior programming knowledge is needed; users without relevant experience can start immediately with creating visualizations using Tableau
- It can connect to several data sources that other BI tools do not support. Tableau enables users to create reports by joining and blending different datasets
- Tableau Server supports a centralized location to manage all published data sources within an organization

3.2 Values in Tableau

There are two types of values in the tableau:

- **Dimensions:** Discrete Values (which can not change concerning time) in nature called Dimension in tableau. Example: city name, product name, country name.
- **Measures:** Continuous Values (which can change concerning time) in nature called Measures in Tableau. Example: profit, sales, discount, population.

3.3 Advantages of Tableau

- **Quick calculation-** All the calculations on the tableau are done by the backend, so it is relatively faster than any other tool.
- **Interactive dashboards-** Tableau dashboards are very interactive and easy to draw.
- **No manual calculation-** All the calculations are done by the tableau only. There is no manual calculation, but in some specific cases, we used calculated fields for calculation.
- **A large amount of data-** Tableau can handle a large amount of data. Different types of visualization can be created with a large amount of data without impacting the performance of the dashboards.

4. DATA ANALYSIS

The analysis of the IMDb dataset involves creating a series of visualizations using Tableau Desktop to uncover trends and patterns in the data. Below, we provide detailed descriptions of each visualization, including the storytelling aspect, dimensions, measures, and the insights derived from them.

4.1 VISUALIZATIONS

4.1.1 Visualization-1

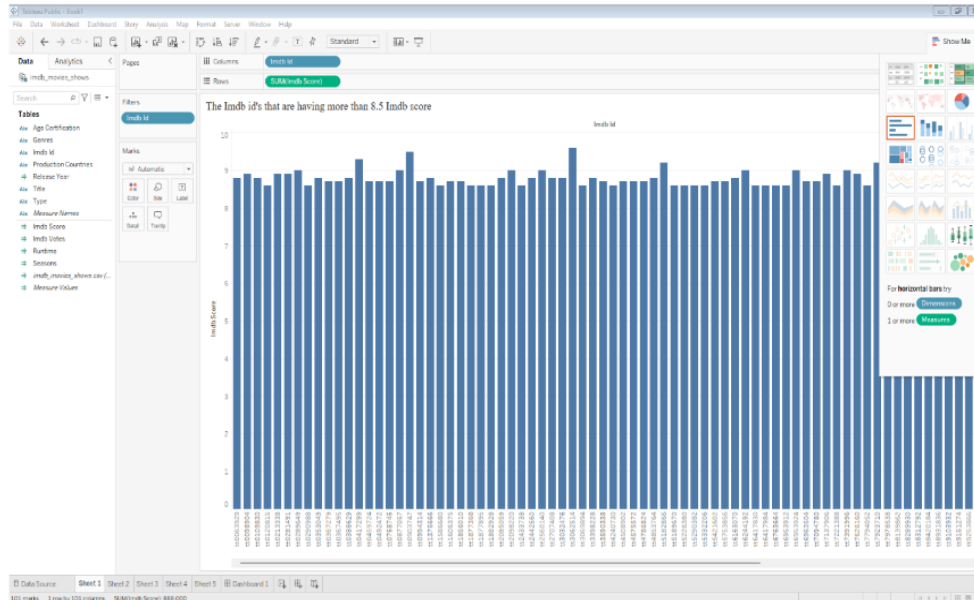


Fig.1: The IMDB IDs with scores above 8.5

- **Visualization:** Bar Chart
- **Storytelling:** This bar chart visualization focuses on the IMDb dataset, showcasing only those titles that have achieved an IMDb score higher than 8.5. By filtering out lower-rated titles, we can easily identify the most acclaimed movies and shows, which can be particularly useful for viewers seeking high-quality content and for industry stakeholders aiming to understand the characteristics of highly-rated productions.
- **Dimensions/columns:** IMDb ID
- **Measures/rows:** SUM (IMDb Score)
- **Filters:** IMDb Score > 8.5
- **Insights:** The chart easily identifies highly-rated movies and shows. Understand genres and styles that receive high ratings. Determine high standards expected by audiences. Highlight titles with significant viewer engagement.

4.1.2 Visualization-2

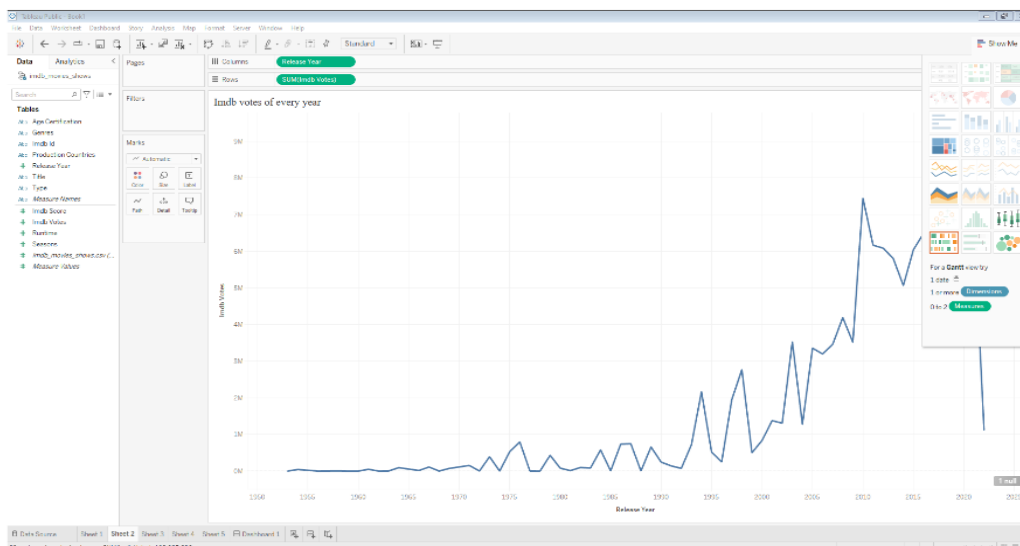


Fig.2: IMDB votes per year.

- **Visualization:** Line Chart
- **Storytelling:** This line chart tracks the total number of IMDb votes each year, providing insights into viewer engagement and the popularity of movies and shows over time.
- **Dimensions:** Release Year
- **Measures:** Sum of IMDb Votes
- **Filters:** None
- **Insights:** It identifies years with spikes or drops in total IMDb votes, indicating periods of high or low viewer engagement. Pinpoint specific years where movies and shows were particularly popular, showcasing major releases or industry trends.

4.1.3 Visualization-3

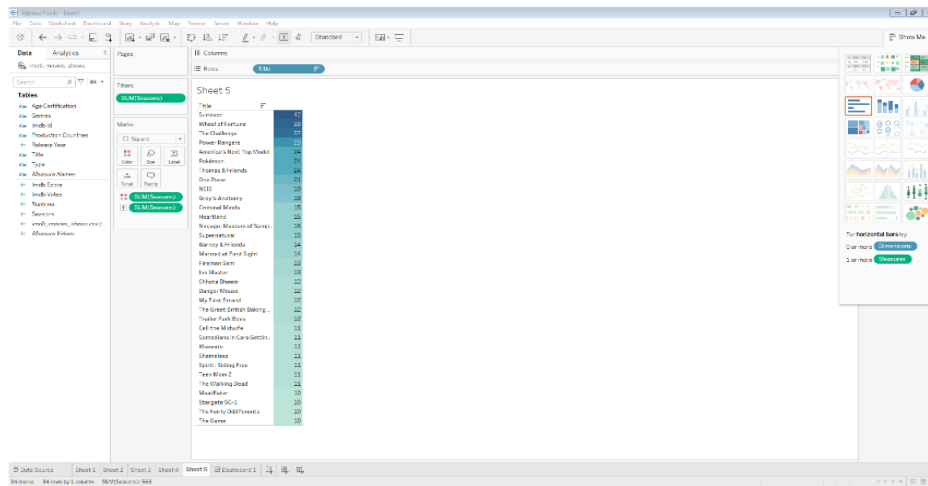


Fig.3: Number of Seasons for Long-Running TV Shows

- **Visualization:** Tree Map
- **Storytelling:** This treemap displays TV shows with 10 or more seasons, sorted by the number of seasons. It highlights the most enduring and long-running TV shows.
- **Dimensions:** Title
- **Measures:** Seasons
- **Filters:** Seasons \geq 10
- **Insights:** It identifies which TV shows have the most seasons, showcasing their longevity. Long-running shows often indicate strong viewer loyalty and consistent quality. Networks and producers can analyze the traits of successful long-running shows to inform future content development.

4.1.3 Visualization-4

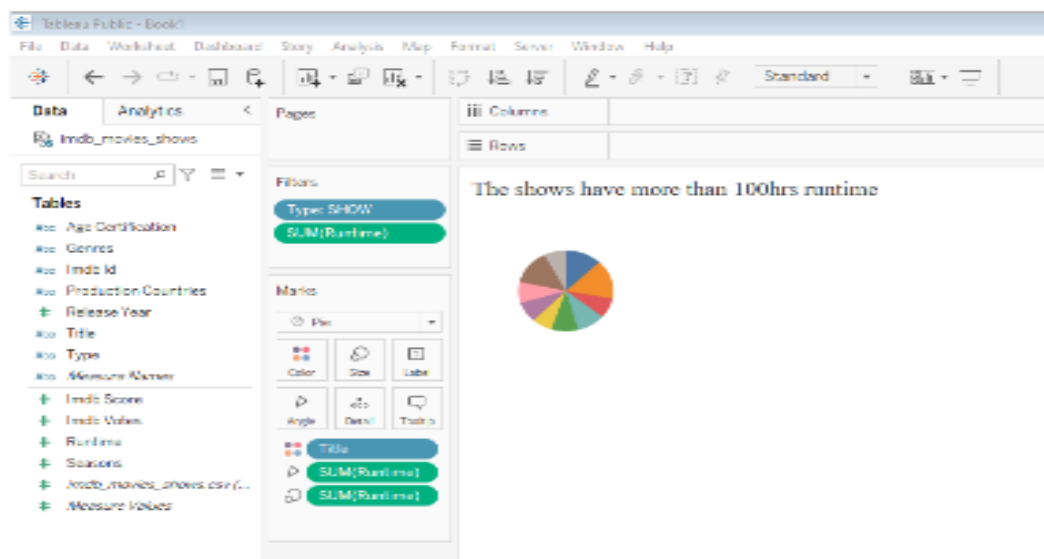


Fig.4: Total Runtime of TV Shows (Runtime > 100 hours)

- **Visualization:** Pie Chart
- **Storytelling:** This pie chart illustrates the distribution of total runtime among TV shows with a runtime exceeding 100 hours. It provides a visual breakdown of how much content falls into different categories based on runtime.
- **Dimensions:** Type
- **Measures:** Sum of Runtime
- **Filters:** Runtime > 100 hours, Type = Show
- **Insights:** It Visualizes the proportion of total runtime contributed by different types of TV shows that exceed 100 hours. Identify if certain types of shows (e.g., Drama, Comedy) dominate the total runtime in the dataset. Understand the genre or type preferences among viewers for longer-format TV shows, which can inform content strategy and viewer engagement.

4.1.5 Visualization-5

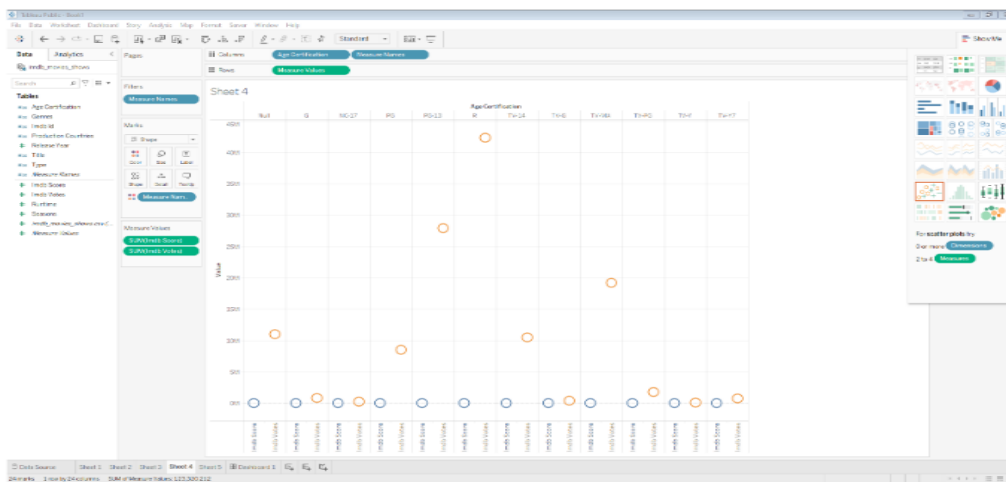


Fig.5: Scatter Plot with Age Certification, IMDb Score, and IMDb Votes

- **Visualization:** Scatter Plot
- **Storytelling:** This scatter plot visualizes the relationship between age certification (like PG-13, R, etc.), IMDb scores, and IMDb votes. Each point represents a movie or show, with the x-axis showing the age certification, the y-axis representing the sum of IMDb scores, and the size of the points representing the sum of IMDb votes.
- **Dimensions:** Age Certification
- **Measures:** Sum of IMDb Votes (Filtered as Rows), Sum of IMDb Scores (Size of Points)
- **Filters:** Age Certification (Selected specific age certifications)
- **Insights:** It identifies if there's a correlation between higher IMDb scores and more IMDb votes across different age certifications. Understand how age certifications influence viewer ratings and engagement. It can inform content creators and distributors about the relationship between age certifications, viewer reception, and popularity, helping in strategic decision-making.

4.1.6 Visualization-6

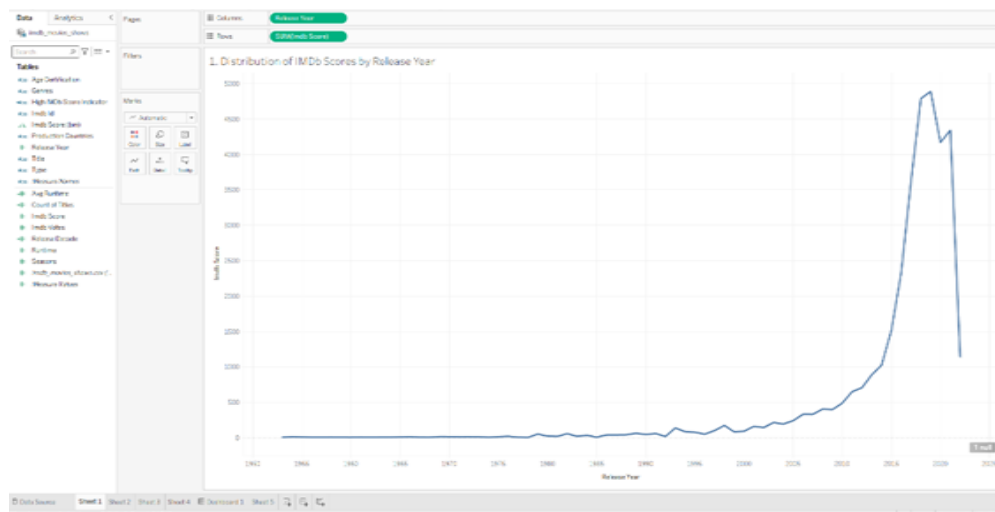


Fig.6: Distribution of IMDb Scores by Release Year

- **Visualization:** Line Chart
- **Storytelling:** Examines the distribution of IMDb scores over different release years, highlighting trends in quality.
- **Dimensions:** Release Year
- **Measures:** IMDb Score
- **Filters:** None
- **Insights:** Identifies periods of high-quality or low-quality production, providing insights into industry trends over time.

4.1.7 Visualization-7

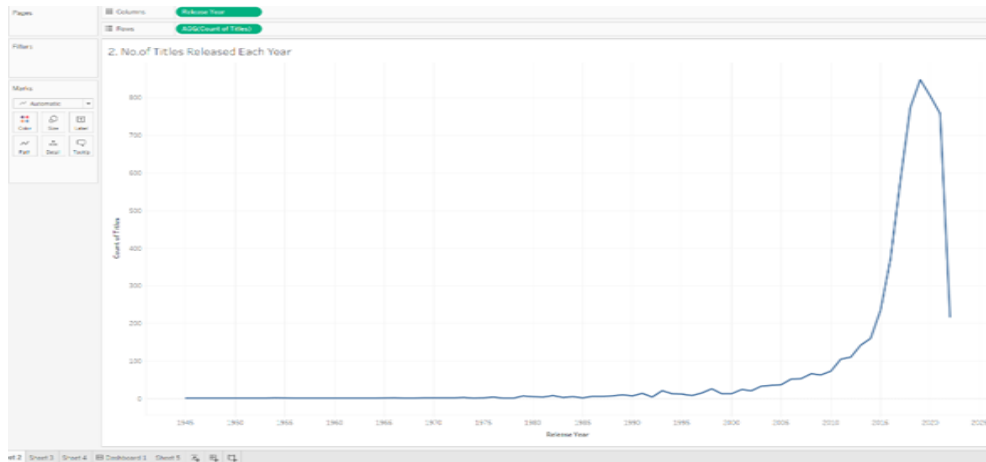


Fig.7: Number of Titles Released Each Year

- **Visualization:** Line Chart
- **Storytelling:** Tracks the volume of movies and shows produced annually.
- **Dimensions:** Release Year
- **Measures:** Count of Titles
- **Filters:** None
- **Insights:** Reveals trends in production activity, indicating industry growth or decline periods.

4.1.8 Visualization-8

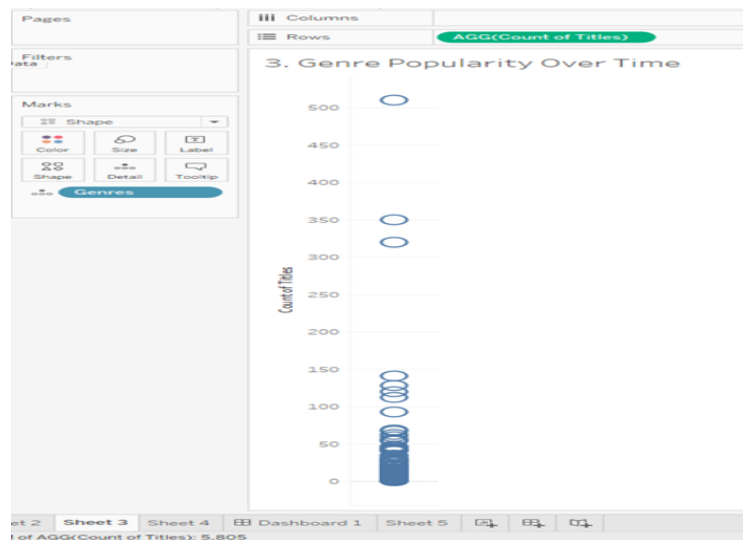


Fig.8: Genre Popularity Over Time

- **Visualization:** Circle View
- **Storytelling:** Displays the changing popularity of different genres over the years.
- **Dimensions:** Release Year, Genre
- **Measures:** Count of Titles
- **Filters:** None
- **Insights:** Shows shifts in audience preferences and industry focus on various genres.

4.1.9 Visualization-9

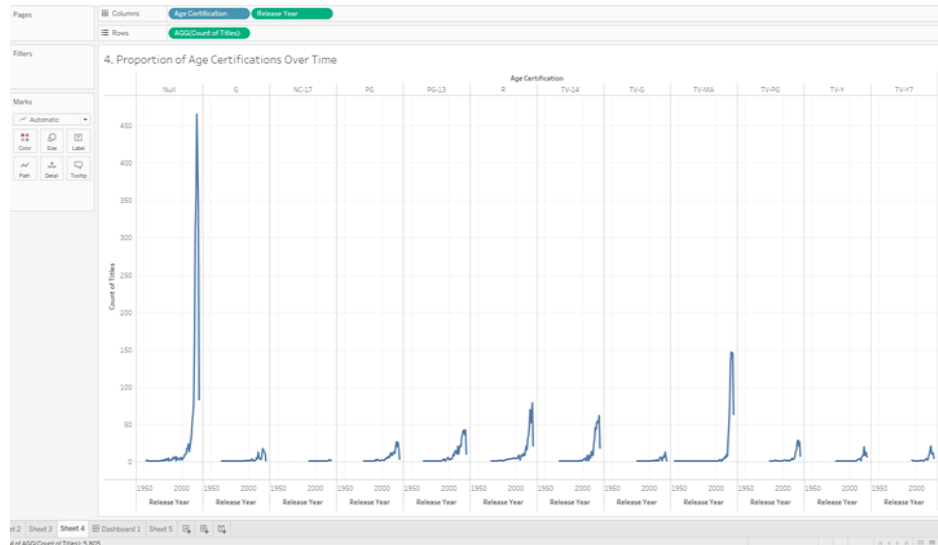


Fig.9: The Proportion of Age Certifications Over Time

- **Visualization:** Stacked Area Chart
- **Storytelling:** This stacked area chart visualizes the proportion of different age certifications (such as G, PG, PG-13, R) for movies and shows over time. It shows how the distribution of content suitable for different age groups has changed through the years.
- **Dimensions:** Release Year, Age Certification
- **Measures:** Count of Titles
- **Filters:** None
- **Insights:** Used in trend analysis, to observe how the proportion of titles in different age certification categories has shifted over time. For instance, you might see an increase in PG-13-rated titles and a decrease in G-rated titles.

4.1.10 Visualization-10

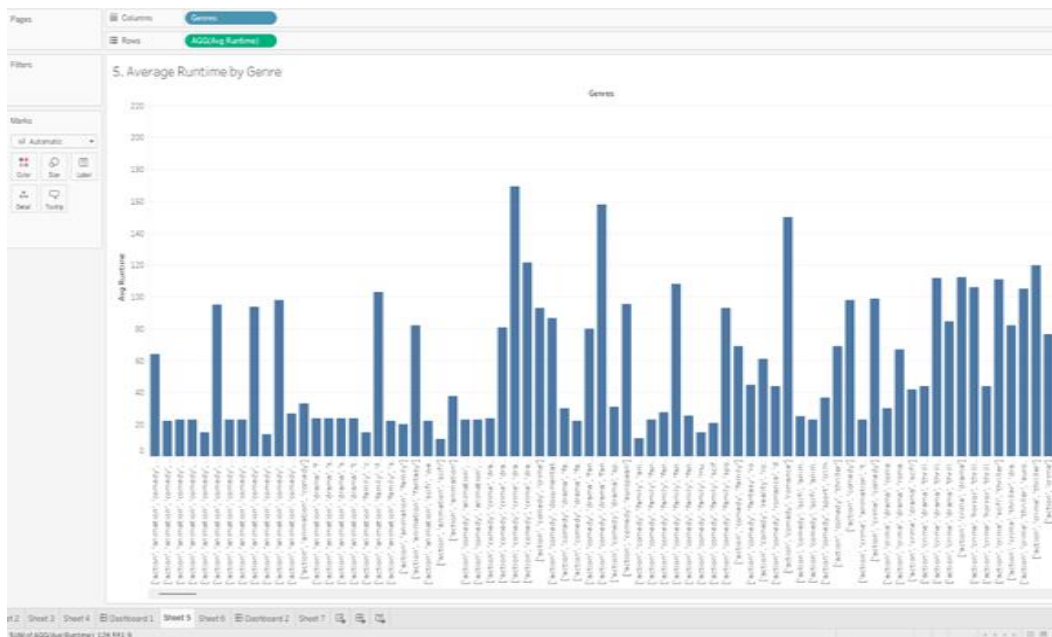


Fig.10: Average Runtime by Genre

- **Visualization:** Bar Chart
- **Storytelling:** Shows the average runtime of movies and shows across different genres.
- **Dimensions:** Genre
- **Measures:** Average Runtime
- **Filters:** None
- **Insights:** Highlights production characteristics of different genres, such as longer runtimes for dramas.

4.1.11 Visualization-11

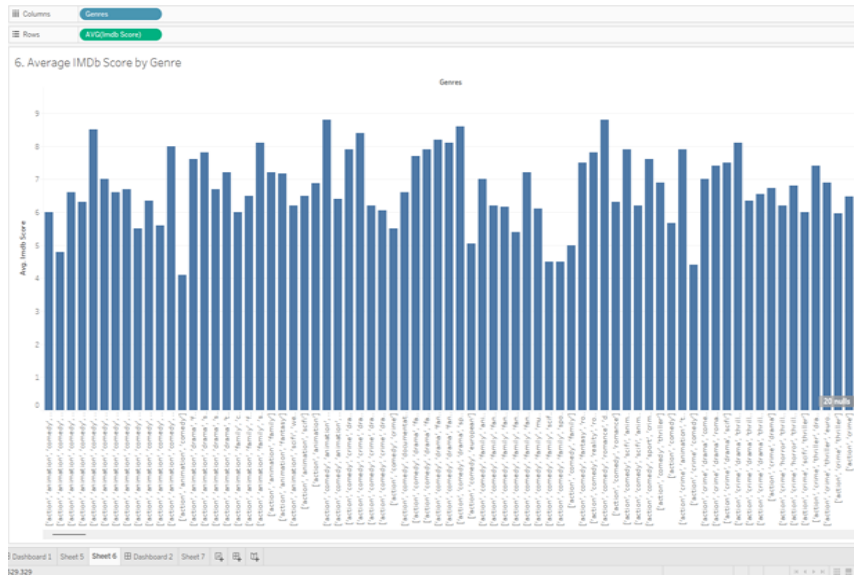


Fig.11: Average IMDb Score by Genre

- **Visualization:** Bar Chart
- **Storytelling:** This bar chart illustrates the average IMDb score for different genres of movies and shows, providing insights into which genres tend to receive higher or lower ratings from viewers.
- **Dimensions:** Genre
- **Measures:** Average IMDb Score
- **Filters:** None
- **Insights:** By analyzing the average IMDb score by genre, stakeholders can gain valuable insights into the quality and viewer perception of different types of content, aiding in strategic decision-making for content creation and distribution.

4.1.12 Visualization-12

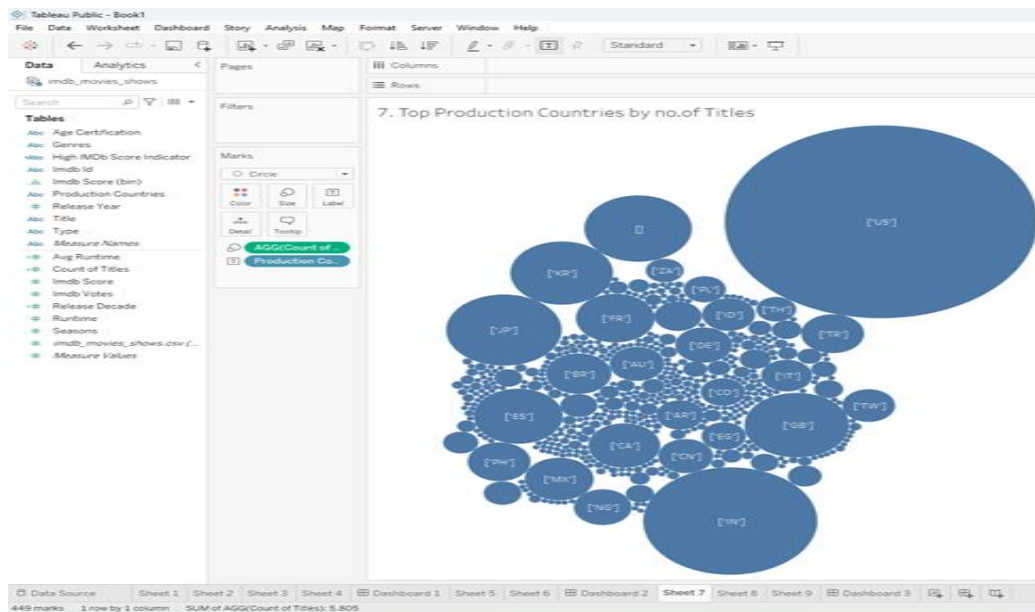


Fig.12: Top Production Countries by Number of Titles

- **Visualization:** Bubble Chart
- **Storytelling:** Identifies leading countries in movie and show production.
- **Dimensions:** Production Country
- **Measures:** Count of Titles
- **Filters:** None
- **Insights:** Highlights dominance of certain countries in the entertainment industry.

4.1.13 Visualization-13

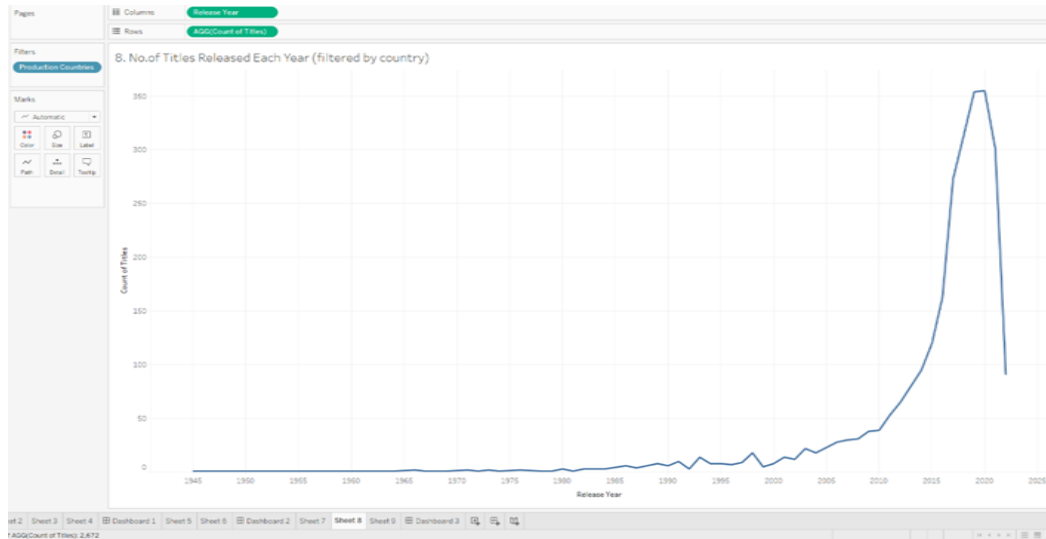


Fig.13: Number of Titles Released Each Year (Filtered by Country)

- **Visualization:** Line Chart
- **Storytelling:** This line chart visualizes the number of movies and shows released each year, filtered by specific countries. It highlights trends in production volume over time for different countries.
- **Dimensions:** Release Year
- **Measures:** Count of Titles
- **Filters:** Country (to focus on specific countries of interest)
- **Insights:** By analyzing the number of titles released each year filtered by country, stakeholders can gain valuable insights into the development and dynamics of the global entertainment industry, informing strategic decisions for content production, investment, and distribution.

4.1.14 Visualization-14

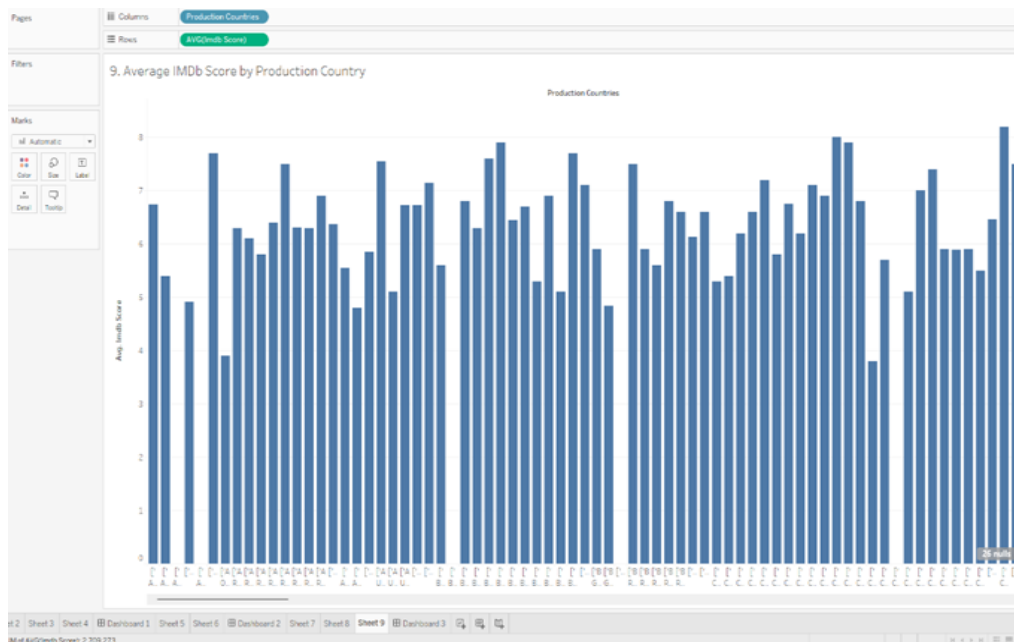


Fig.14: Average IMDb Score by Production Country

- **Visualization:** Bar Chart
- **Storytelling:** Shows the average IMDb score of titles produced in different countries.
- **Dimensions:** Production Country
- **Measures:** Average IMDb Score
- **Filters:** None
- **Insights:** Indicates the quality of content produced by different countries.

4.1.15 Visualization-15

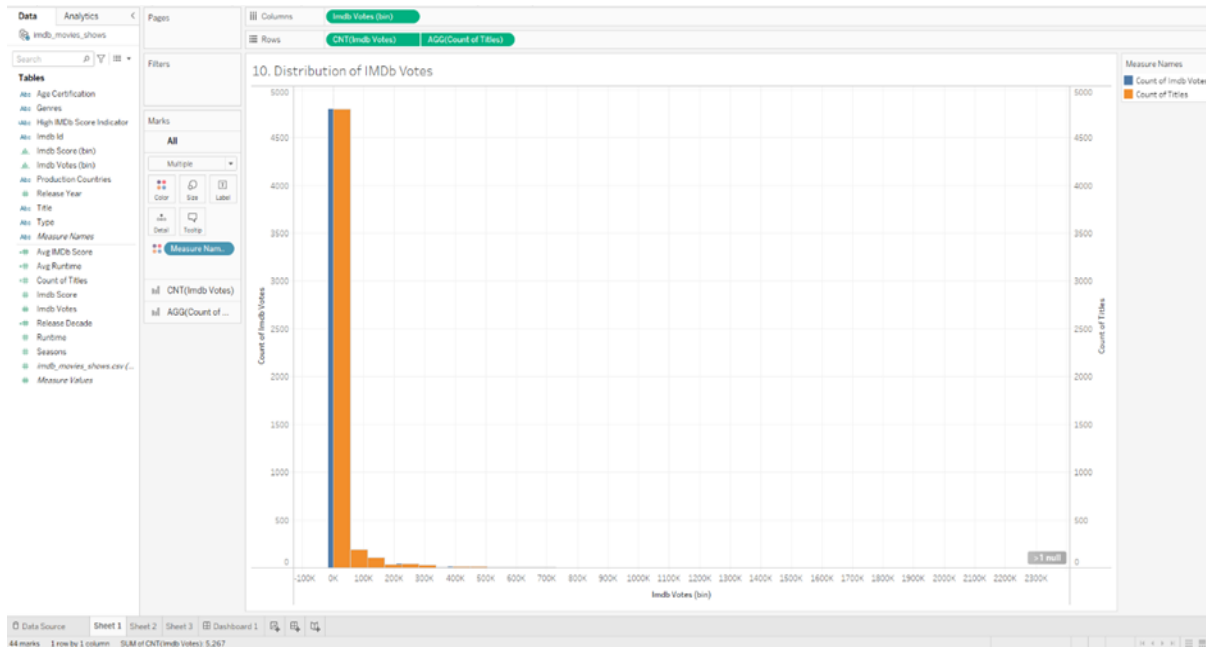


Fig.15: Distribution of IMDb Votes

- **Visualization:** Histogram
- **Storytelling:** Displays the distribution of IMDb votes received by movies and shows.
- **Dimensions:** IMDb Votes (Binned)
- **Measures:** Count of Titles
- **Filters:** None
- **Insights:** Reveals viewer engagement levels and popularity distribution.

4.1.16 Visualization-16

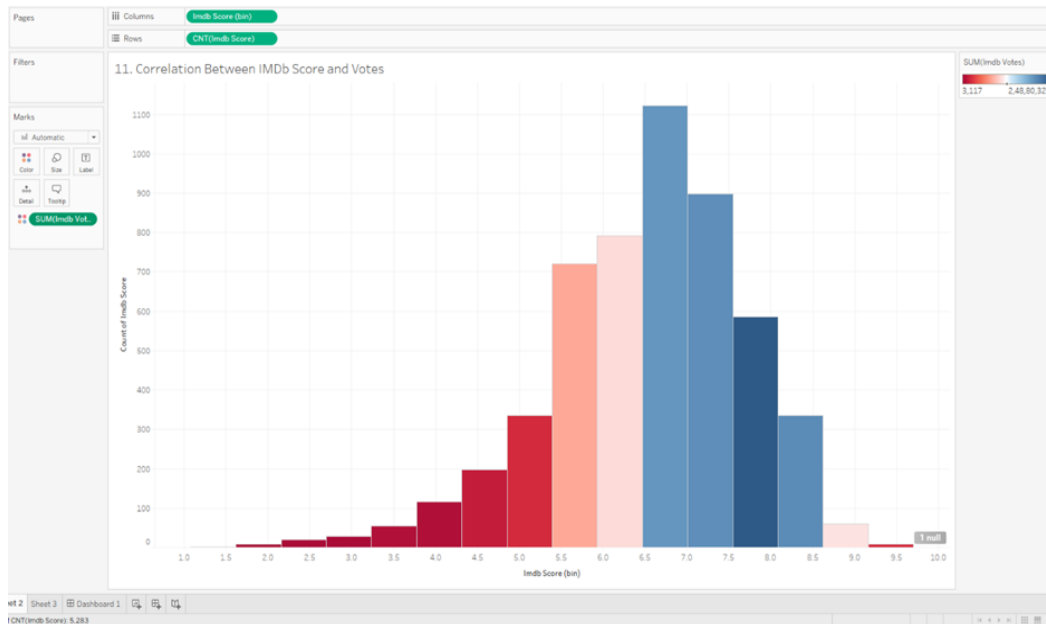


Fig.16: Correlation Between IMDb Score and Votes

- **Visualization:** Histogram
- **Storytelling:** Explores the relationship between IMDb scores and the number of votes received.
- **Dimensions:** IMDb Score, IMDb Votes
- **Measures:** IMDb Score, IMDb Votes
- **Filters:** None
- **Insights:** Highlights whether high-rated titles also receive more votes, indicating strong engagement.

4.1.17 Visualization-17

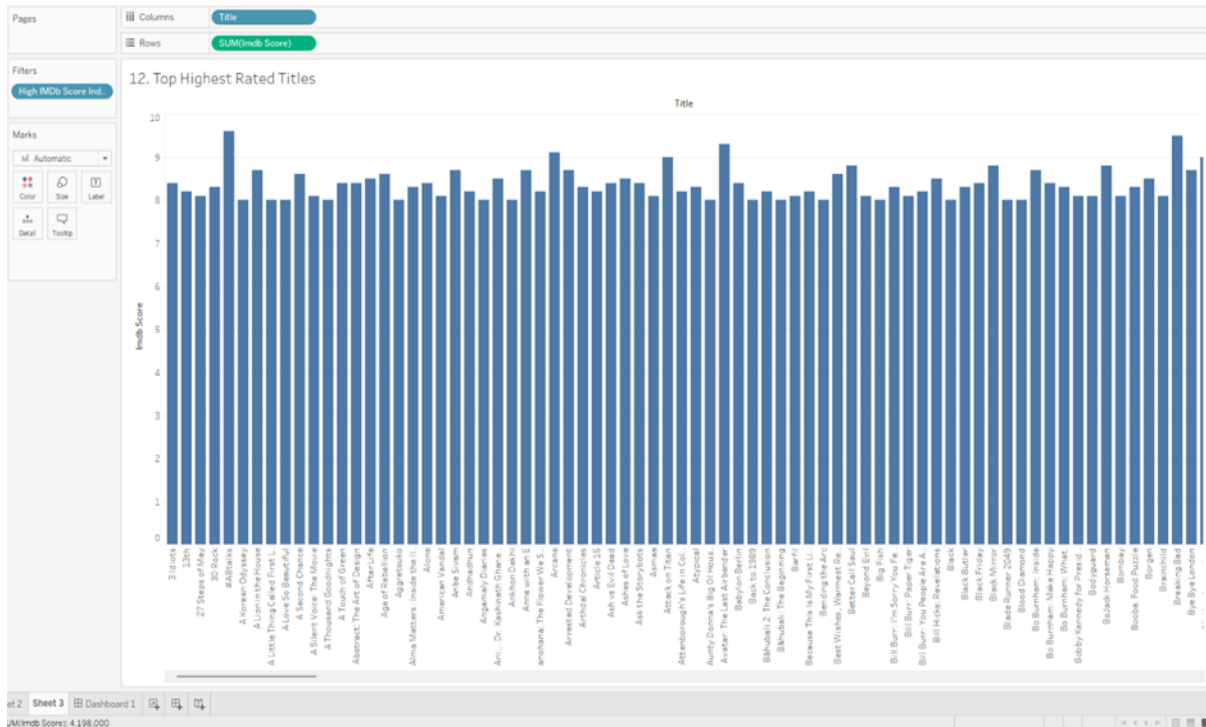


Fig.17: Top Highest-Rated Titles

- **Visualization:** Bar Chart
- **Storytelling:** Highlights the titles with the highest IMDb scores.
- **Dimensions:** Title
- **Measures:** IMDb Score
- **Filters:** IMDb Score > 8.5
- **Insights:** Showcases the top performers in the dataset, highlighting critically acclaimed and audience-loved titles.

4.1.18 Visualization-18

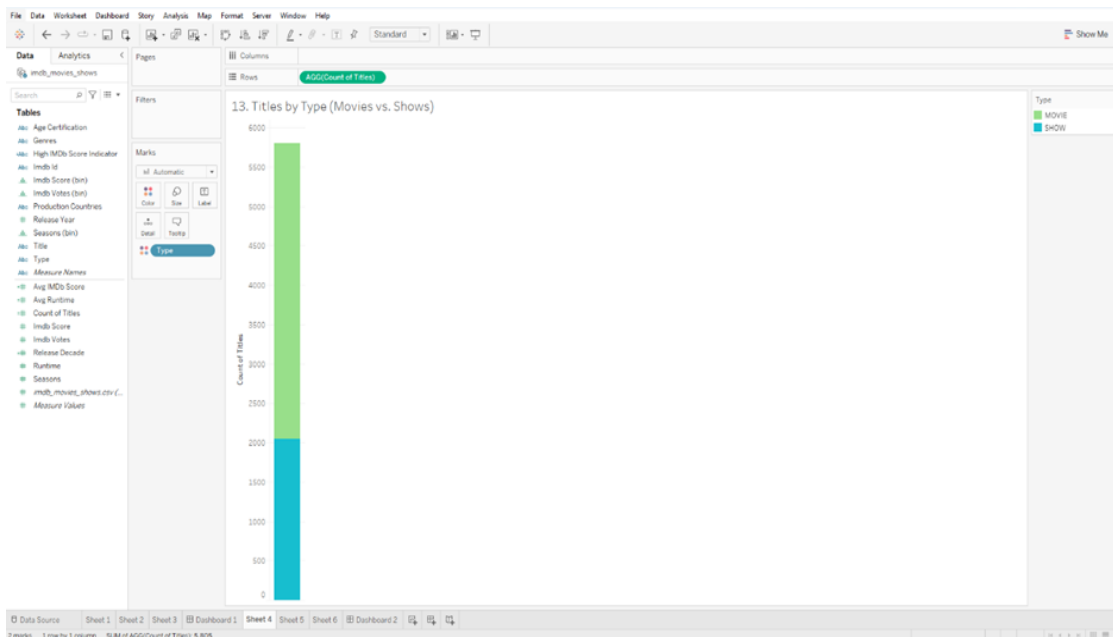


Fig.18: Titles by Type (Movies vs. Shows)

- **Visualization:** Stacked Bar Chart
- **Storytelling:** This stacked bar chart visualizes the number of titles categorized as movies and shows, highlighting the distribution of content types over different years.
- **Dimensions:** Type (Movies, Shows), Release Year

- **Measures:** Count of Titles
- **Filters:** None
- **Insights:** By analyzing the number of titles by type using a stacked bar chart, stakeholders can gain a comprehensive understanding of the evolving landscape of the entertainment industry, informing strategic decisions for content production and market positioning.

4.1.19 Visualization-19

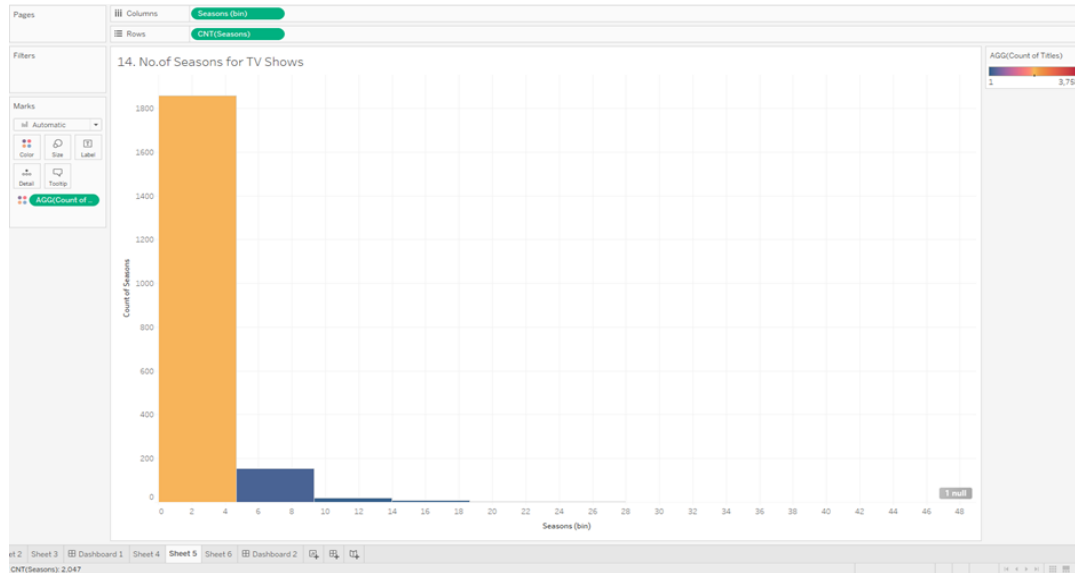


Fig.19: Number of Seasons for TV Shows

- **Visualization:** Histogram
- **Storytelling:** Displays TV shows with 10 or more seasons, sorted by the number of seasons.
- **Dimensions:** Title
- **Measures:** Seasons
- **Filters:** Seasons \geq 10
- **Insights:** Identifies the most enduring and long-running TV shows, indicating strong viewer loyalty and consistent quality.

4.1.20 Visualization-20

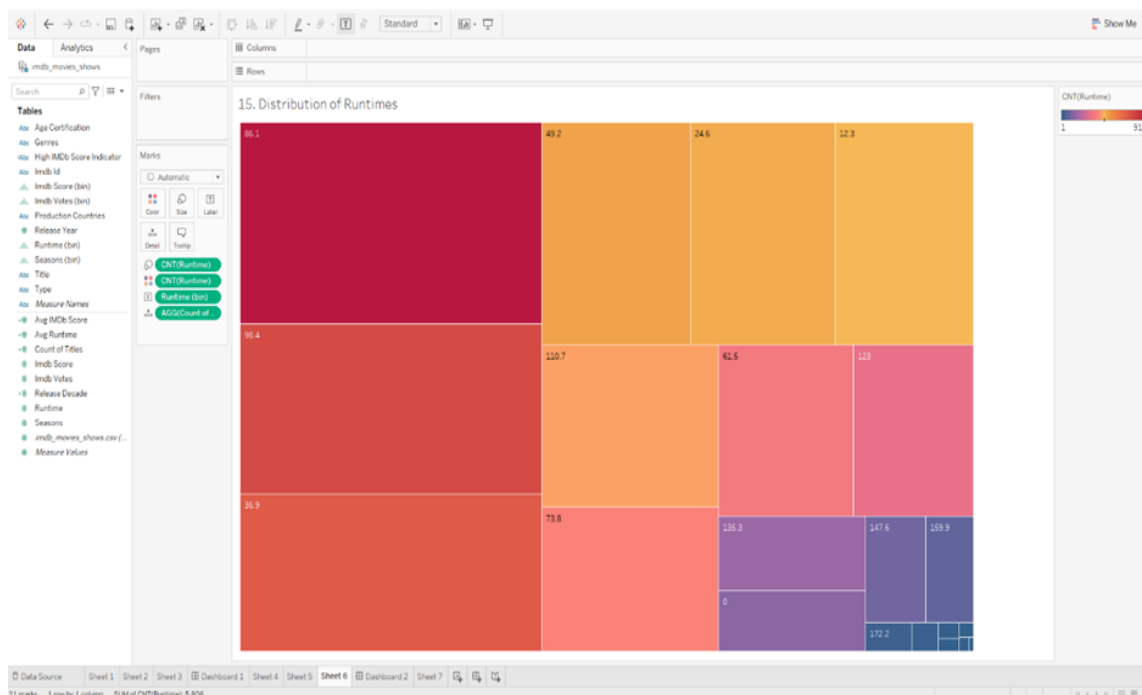


Fig.20: Distribution of Runtimes

- **Visualization:** Treemap
- **Storytelling:** This treemap visualizes the distribution of movie and show runtimes, providing insights into the range and frequency of different runtime durations.
- **Dimensions:** Runtime (Binned into ranges), Title
- **Measures:** Count of Titles
- **Filters:** None
- **Insights:** By analyzing the distribution of runtimes using a treemap, stakeholders can gain valuable insights into the lengths of movies and shows, informing strategic decisions for content creation, acquisition, and scheduling to meet audience preferences and industry standards.

4.1.21 Visualization-21

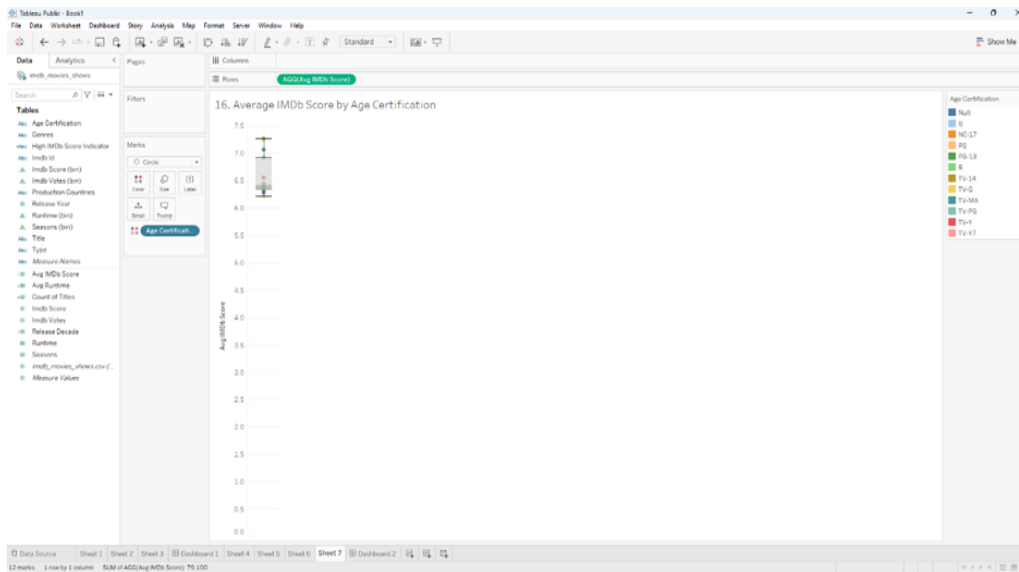


Fig.21: Average IMDb Score by Age Certification

- **Visualization:** Box and Whisker Plot
- **Storytelling:** This box and whisker plot visualize the distribution of IMDb scores for different age certifications, showing the spread and central tendency of ratings for each certification category.
- **Dimensions:** Age Certification
- **Measures:** IMDb Score
- **Filters:** None
- **Insights:** By analyzing the average IMDb score by age certification using a box and whisker plot, stakeholders can gain a deeper understanding of the distribution and central tendency of ratings for different age groups, informing strategic decisions on content targeting, creation, and quality assurance.

4.1.22 Visualization-22

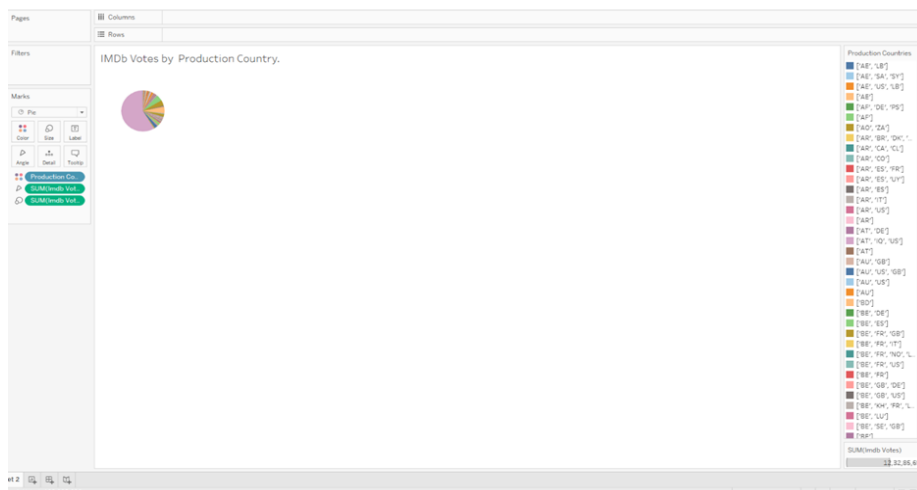


Fig.22: IMDb Votes by Production Country

- **Visualization:** Pie Chart
- **Storytelling:** This pie chart visualizes the distribution of IMDb votes across different production countries. It provides insights into which countries' movies and shows receive the most audience engagement in terms of votes.
- **Dimensions:** Production Country
- **Measures:** Sum of IMDb Votes
- **Filters:** None
- **Insights:** By analyzing IMDb votes by production country using a pie chart, stakeholders can gain valuable insights into the global distribution of audience engagement, informing strategic decisions for content production, acquisition, and market positioning.

4.1.23 Visualization-23

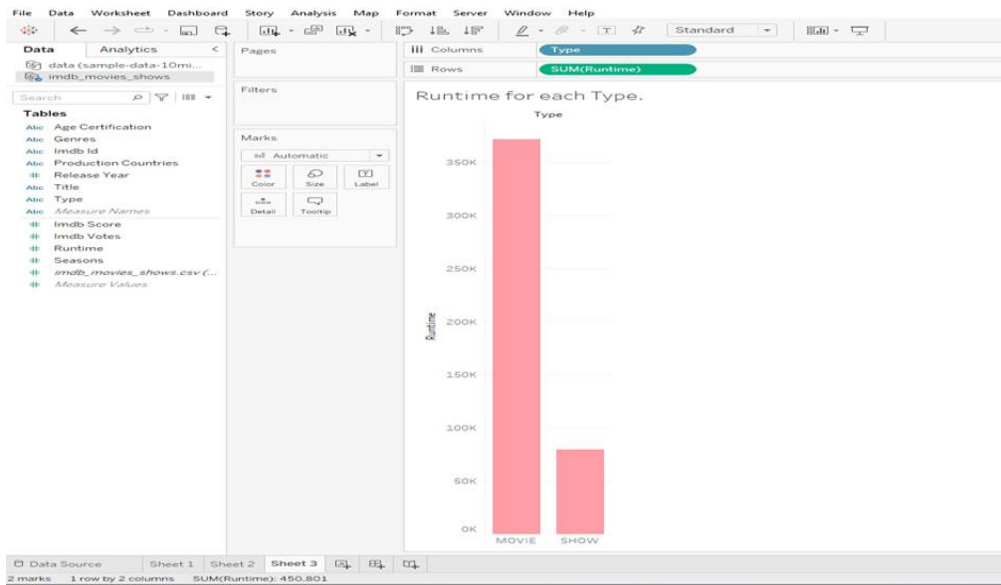


Fig.23: Runtime for Each Type

- **Visualization:** Bar Chart
- **Storytelling:** This bar chart visualizes the average runtime for movies and shows, providing insights into typical content lengths for each type.
- **Dimensions:** Type (Movies, Shows)
- **Measures:** Average Runtime
- **Filters:** None
- **Insights:** By analyzing the average runtime for each type using a bar chart, stakeholders can gain valuable insights into industry standards and audience preferences, informing strategic decisions for content creation, acquisition, and scheduling.

4.1.24 Visualization-24

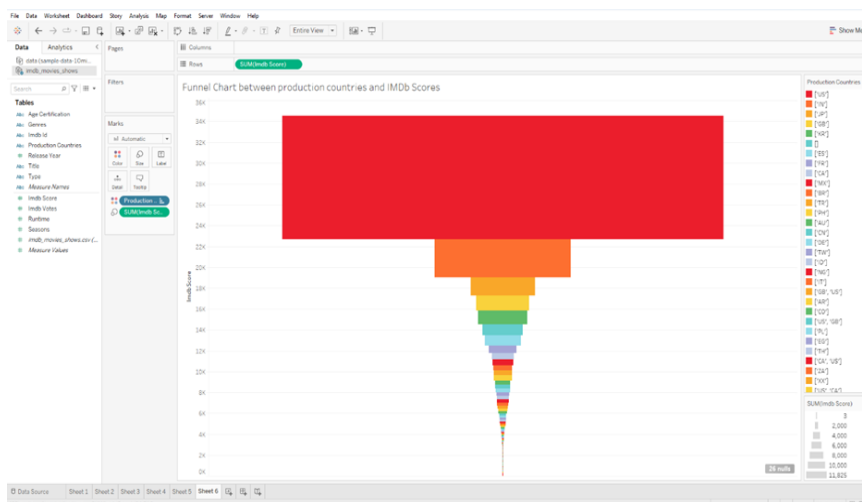


Fig.24: Funnel Chart between Production Countries and IMDb Scores

- **Visualization:** Funnel Chart
- **Storytelling:** This funnel chart visualizes the average IMDb scores across different production countries, illustrating the comparative performance and viewer ratings of content produced in various countries.
- **Dimensions:** Production Country
- **Measures:** Average IMDb Score
- **Filters:** None
- **Insights:** By analyzing the average IMDb scores across production countries using a funnel chart, stakeholders can gain valuable insights into the global distribution of content quality, informing strategic decisions for content creation, acquisition, and international partnerships.

4.2 DASHBOARDS

4.2.1 Dashboard-1

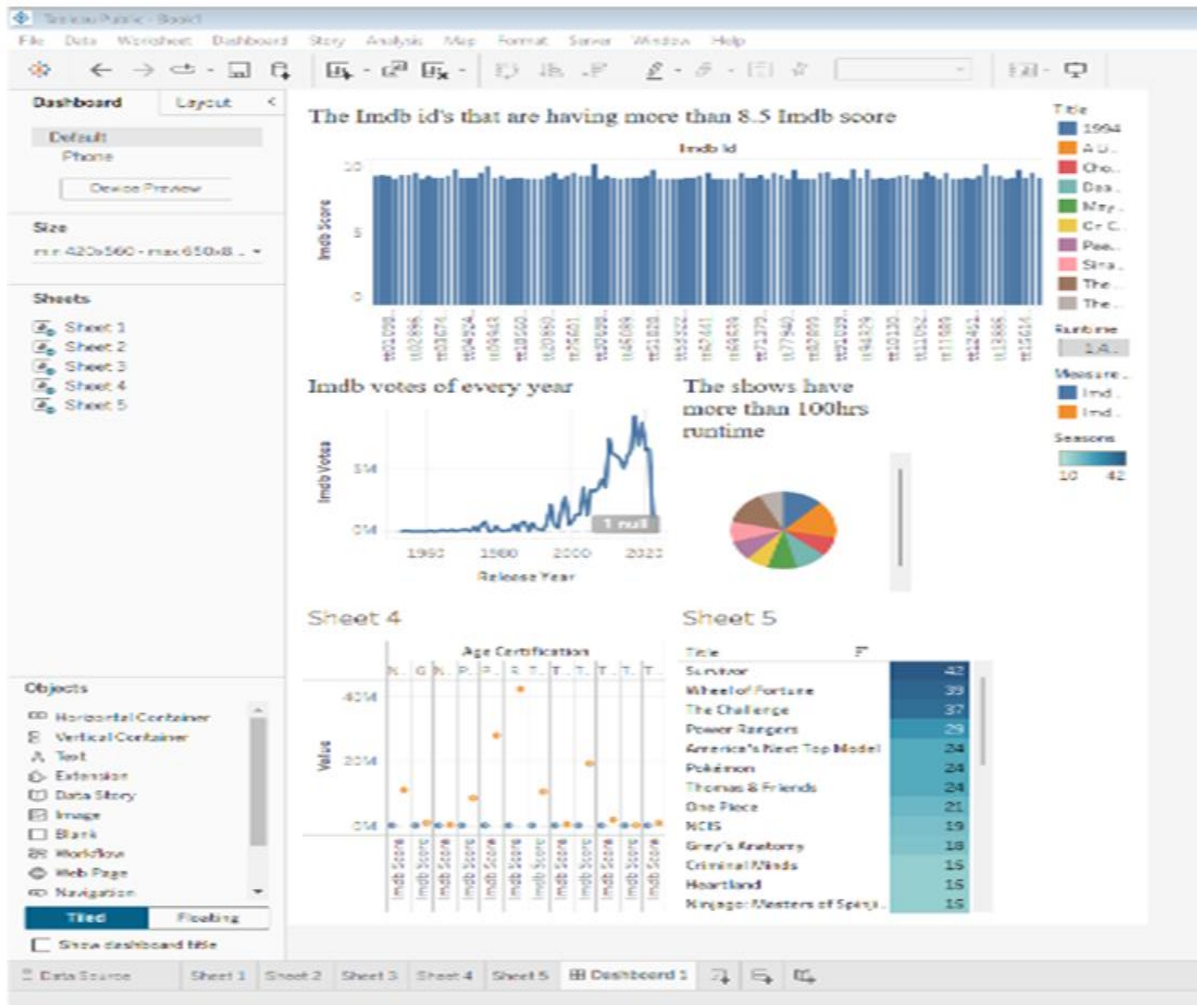


Fig.25: High-Quality Content and Viewer Engagement

This dashboard focuses on identifying high-quality content and analyzing viewer engagement patterns. By highlighting top-rated titles, trends in viewer votes, characteristics of long-running TV shows, and the relationship between age certifications and content performance, the dashboard offers valuable insights for content creators, producers, and marketers. The analysis helps to understand what drives high ratings and viewer engagement, informing strategic decisions in content creation and distribution.

It contains the sheets:

1. IMDb IDs with Scores Above 8.5
2. IMDb Votes Per Year
3. Number of Seasons for Long-Running TV Shows
4. Total Runtime of TV Shows (Runtime > 100 hours)
5. Scatter Plot with Age Certification, IMDb Score, and IMDb Votes

4.2.2 Dashboard-2

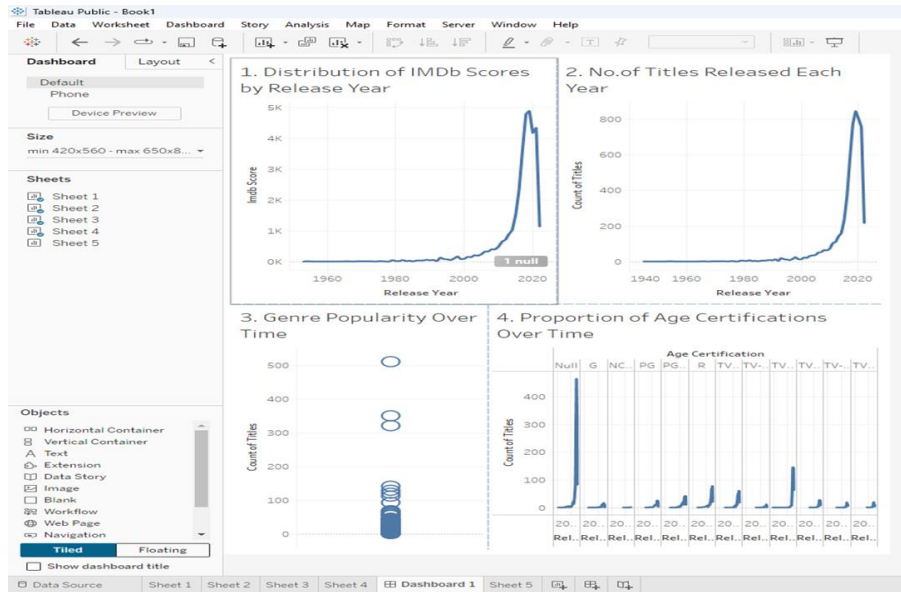


Fig.26: Trends Over Time

This dashboard focuses on the historical trends and shifts in the entertainment industry, offering a comprehensive view of how content quality, production volume, genre preferences, and age certifications have evolved. By examining the distribution of IMDb scores, the number of titles released each year, the popularity of genres, and the proportion of age certifications, stakeholders can gain valuable insights into industry trends and audience preferences. This analysis helps content creators, producers, and marketers to understand past trends, predict future shifts, and make informed decisions in content strategy and development.

It contains the sheets:

1. Distribution of IMDb Scores by Release Year
2. Number of Titles Released Each Year
3. Genre Popularity Over Time
4. Proportion of Age Certifications Over Time

4.2.3 Dashboard-3

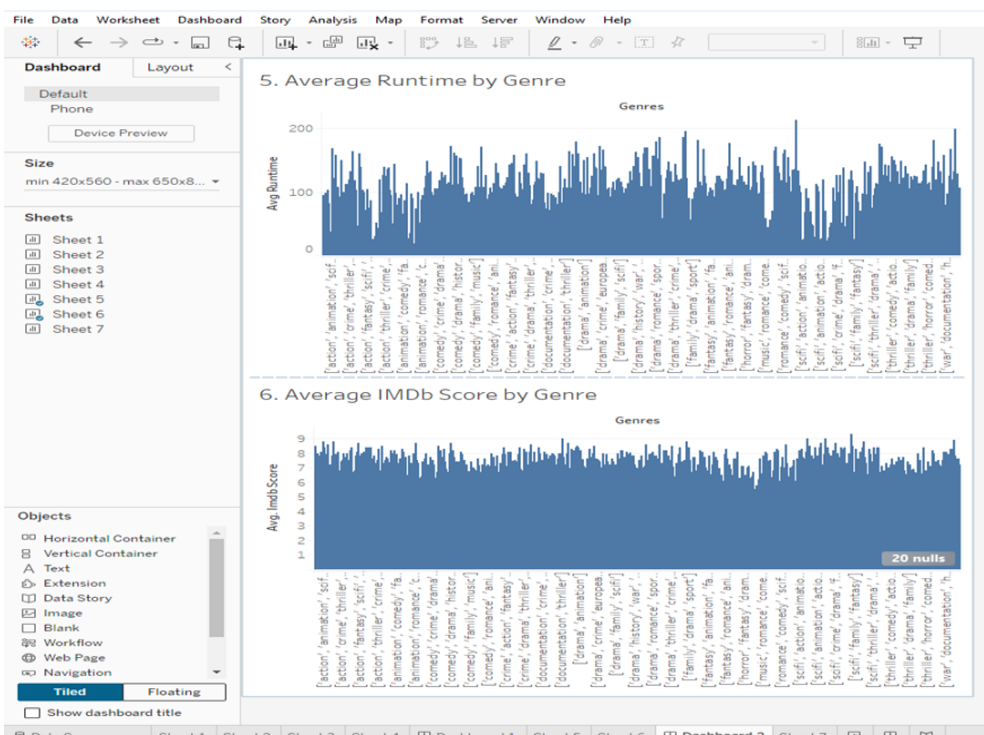


Fig.27: Genre Analysis

This dashboard focuses on analyzing the quality and duration of content across different genres. Visualizing the average runtimes and IMDb scores for various genres provides valuable insights into how different types of content are structured and perceived by audiences. Stakeholders can use this information to understand genre-specific trends, make informed decisions about content creation and acquisition, and tailor their strategies to meet audience preferences. This analysis is crucial for identifying successful genres and optimizing content to enhance viewer satisfaction and engagement.

It contains the sheets:

1. Average Runtime by Genre
2. Average IMDb Score by Genre

4.2.4 Dashboard-4

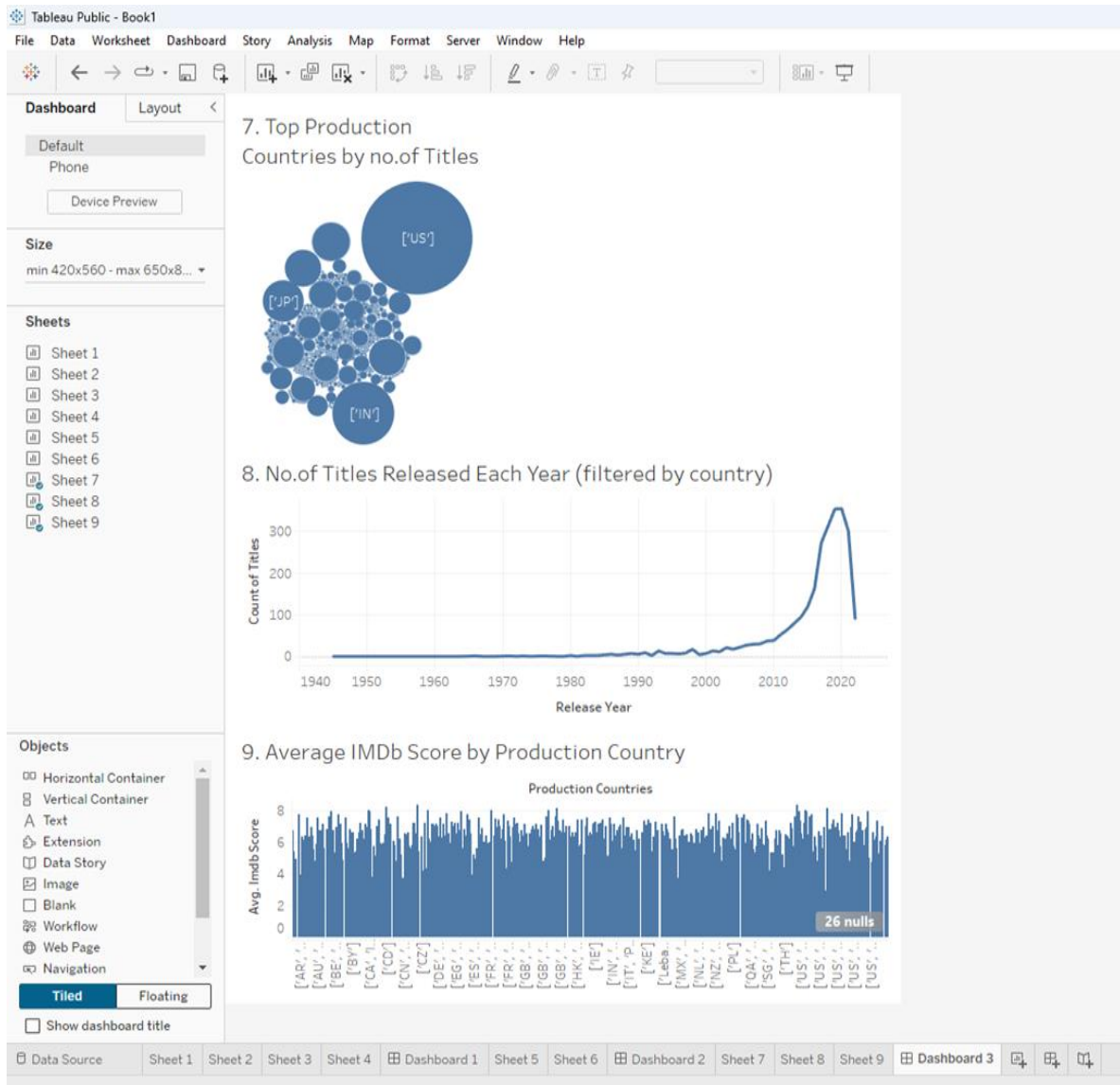


Fig.28: Country and Production Insights

This dashboard focuses on analyzing the global landscape of content production, offering a detailed view of the leading production countries, their release patterns over the years, and the average IMDb scores for their titles.

By examining the number of titles produced, the annual release trends, and the average IMDb scores, stakeholders can gain valuable insights into the activity and performance of different countries in the entertainment industry. This analysis helps content creators, producers, and marketers to understand global trends, identify key players, and make informed decisions in content strategy and international collaborations.

It contains the sheets:

1. Top Production Countries by Number of Titles
2. Number of Titles Released Each Year (Filtered by Country)
3. Average IMDb Score by Production Country

4.2.5 Dashboard-5

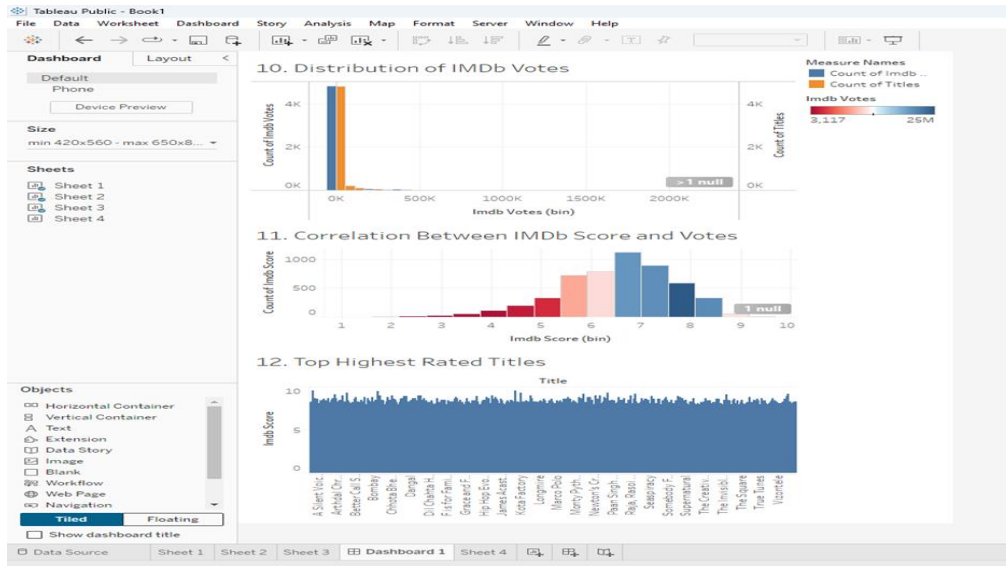


Fig.29: Ratings and Popularity

This dashboard focuses on analyzing audience engagement and identifying top-rated titles on IMDb. By examining the distribution of votes, the correlation between IMDb scores and votes, and highlighting the highest-rated titles, stakeholders can gain valuable insights into what drives viewer engagement and which titles are considered the best by audiences. This analysis helps content creators, producers, and marketers understand the factors contributing to high ratings and engagement, guiding strategic decisions in content creation, marketing, and curation.

It contains the sheets:

1. Distribution of IMDb Votes
2. Correlation Between IMDb Score and Votes
3. Top Highest Rated Titles

4.2.6 Dashboard-6

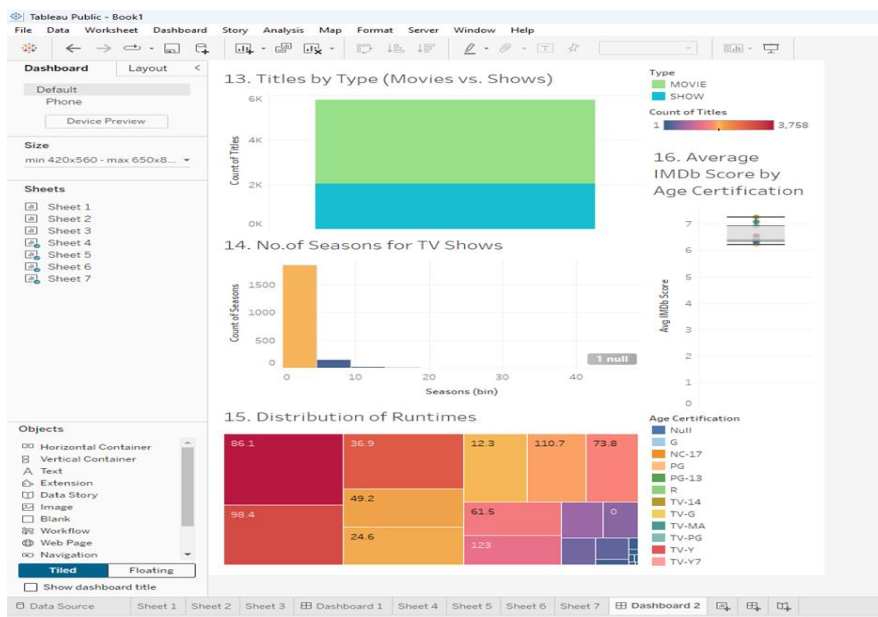


Fig.30: Title Characteristics

This dashboard focuses on analyzing the types, characteristics, and audience ratings of different content. By examining the proportion of content types (movies vs. shows), the number of seasons for TV shows, the distribution of runtimes, and the average IMDb scores by age certification, stakeholders can gain valuable insights into industry trends, production norms, and audience preferences. This analysis helps content creators, producers, and marketers understand the structure, duration, and reception of popular content, informing strategic decisions in content development, scheduling, and targeting specific audience groups.

It contains the sheets:

1. Titles by Type (Movies vs. Shows)
2. Number of Seasons for TV Shows
3. Distribution of Runtimes
4. Average IMDb Score by Age Certification

4.2.7 Dashboard-7

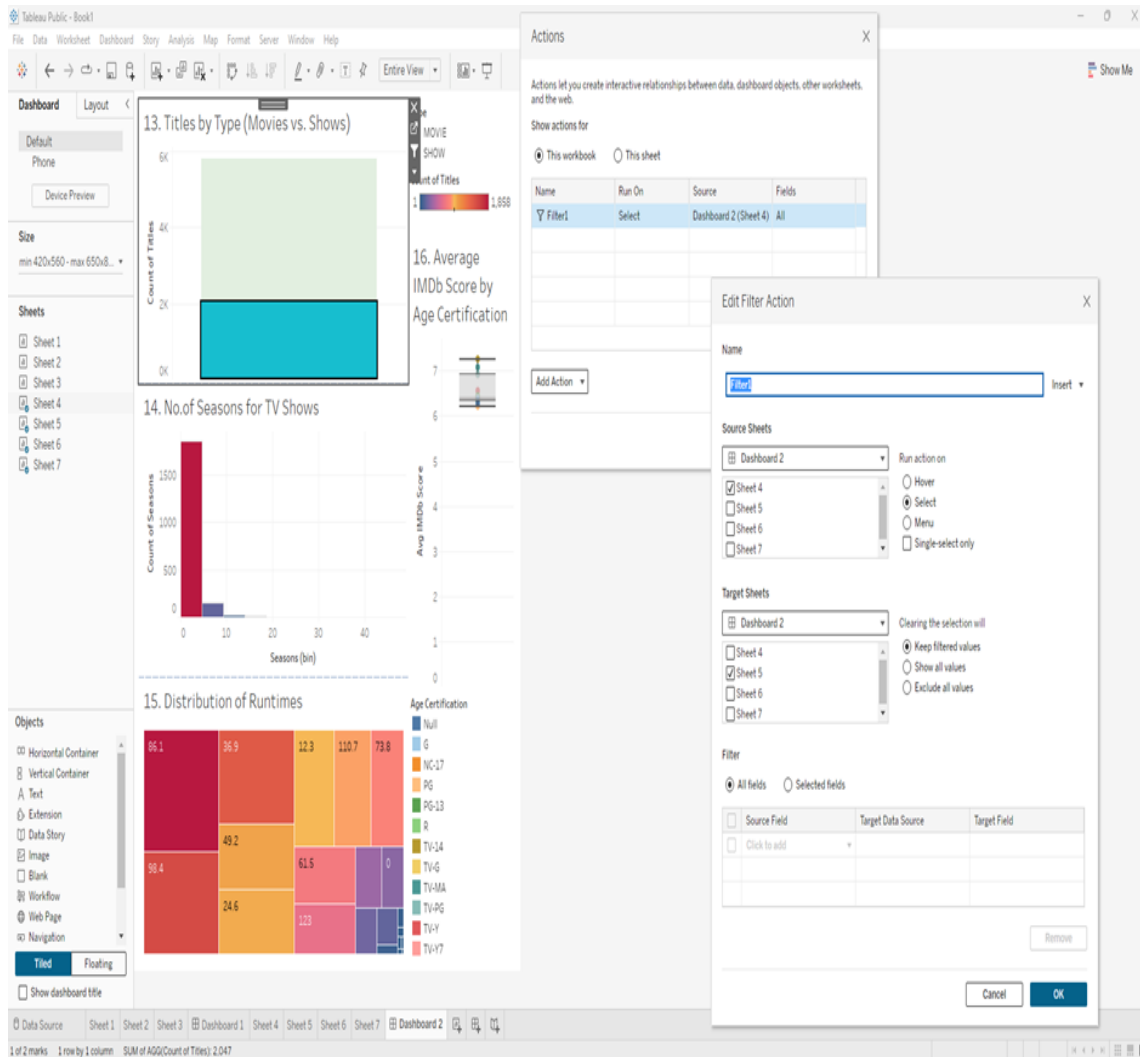


Fig.31: Filter Action

This interactive dashboard focuses on analyzing the types, characteristics, and audience ratings of different content with an added layer of interactivity through filter actions. By examining the proportion of content types (movies vs. shows), the number of seasons for TV shows, the distribution of runtimes, and the average IMDb scores by age certification, and allowing users to filter the number of seasons based on content type, stakeholders can gain valuable insights into industry trends, production norms, and audience preferences. This analysis helps content creators, producers, and marketers understand the structure, duration, and reception of popular content, and make informed strategic decisions in content development, scheduling, and targeting specific audience groups.

It contains the sheets:

1. Titles by Type (Movies vs. Shows)
2. Number of Seasons for TV Shows
3. Distribution of Runtimes
4. Average IMDb Score by Age Certification

In Filter Action,

Source Sheet: Titles by Type (Movies vs. Shows)

Target Sheet: Number of Seasons for TV Shows

4.2.8 Dashboard-8

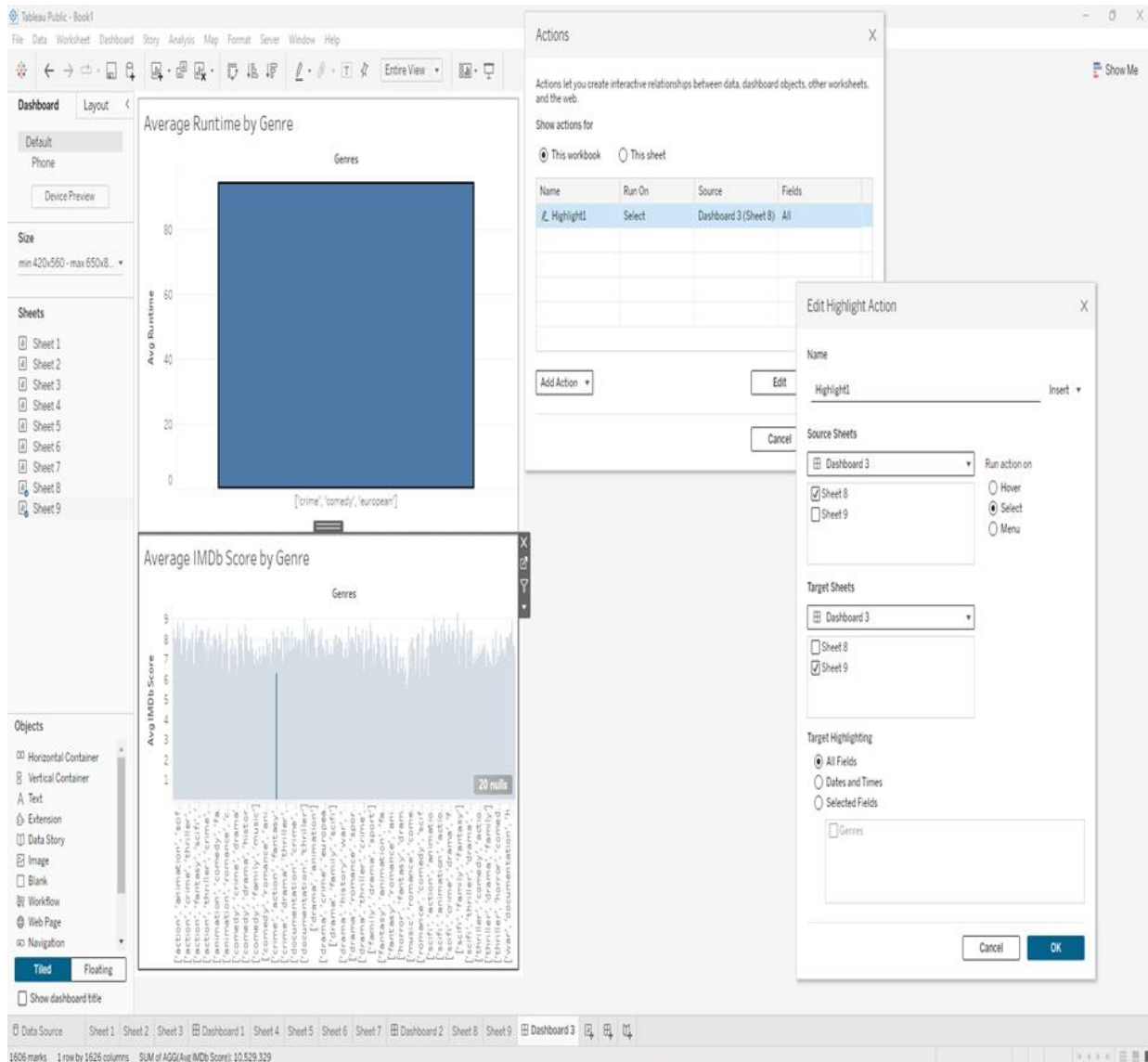


Fig.32: Highlight Action

This dashboard focuses on analyzing the characteristics and performance of different genres in the entertainment industry. By examining the average runtime and IMDb scores for various genres and utilizing a highlight action for interactive exploration, stakeholders can gain valuable insights into industry standards, audience expectations, and the perceived quality of content. This analysis helps content creators, producers, and marketers understand which genres perform well in terms of viewer ratings and engagement, informing strategic decisions in content creation, acquisition, and marketing.

It contains the sheets:

1. Average Runtime by Genre
2. Average IMDb Score by Genre

In Highlight Action,

Source Sheet: Average Runtime by Genre

Target Sheet: Average IMDb Score by Genre

5. RESULTS AND DISCUSSIONS

5.1 Summary of the Dataset

The dataset provides a detailed overview of IMDb movies and TV shows, encompassing various attributes such as title, type (movie or show), release year, age certification, runtime, genres, production countries, seasons, IMDb ID, IMDb score, and IMDb votes. This extensive dataset enables a thorough entertainment industry analysis, capturing trends in content production, viewer engagement, and critical reception.

5.2 Key Observations

- Trends in IMDb Scores:** IMDb scores vary significantly across different years, genres, and age certifications, reflecting shifts in production quality, audience preferences, and industry standards.
- Viewer Engagement:** IMDb votes show fluctuating patterns over the years, with peaks indicating periods of high viewer interaction, often coinciding with the release of popular titles.
- Content Production Dynamics:** The number of titles released each year has shown both growth and contraction phases, influenced by market conditions and industry trends. Certain countries dominate content production, highlighting regional industry leaders.
- Genre Popularity:** Genre preferences have evolved over time, with certain genres experiencing cyclical popularity. This indicates changing audience tastes and the impact of cultural and societal trends on content consumption.
- Content-Length and Seasons:** There are significant variations in the runtime of movies and shows, with long-running TV series (ten or more seasons) indicating strong viewer retention and loyalty.

5.3 Discussions and Insights

- Evolving Standards and Preferences:** Over time, the variation in IMDb scores suggests that production quality and audience expectations are continually evolving. Higher scores in recent years may indicate improvements in production techniques and storytelling.
- Influence of Major Releases:** Peaks in IMDb votes are often associated with releasing highly anticipated titles, suggesting that certain movies and shows significantly boost viewer engagement. These insights can guide marketing strategies and content release schedules.
- Regional Production Strengths:** Analysis of production countries reveals that certain regions consistently produce high volumes of content. This highlights the importance of understanding regional strengths and opportunities for international collaborations and market expansion.
- Adapting to Genre Trends:** The cyclical nature of genre popularity underscores the need for content creators to stay attuned to audience preferences and cultural trends. This can inform genre-specific content development and acquisition strategies.
- Audience Engagement and Longevity:** Long-running TV shows with multiple seasons demonstrate strong audience retention, indicating the value of developing content that can sustain viewer interest over time. This is crucial for platforms aiming to build a loyal user base.

5.4 Overall Insights

The comprehensive analysis using Tableau has provided actionable insights into the entertainment industry's dynamics. By visualizing complex datasets, we have been able to uncover key trends and patterns that inform strategic decision-making.

Tableau's interactive features have enhanced our ability to explore data in depth, facilitating a nuanced understanding of content performance, audience preferences, and market opportunities. These insights are instrumental for content creators, distributors, and marketers aiming to optimize their strategies and stay competitive in the ever-evolving entertainment landscape.

6. CONCLUSION

Utilizing Tableau to create and analyze the extensive visualizations and dashboards across the IMDb dataset has provided profound insights into the intricate workings of the entertainment industry. From exploring IMDb scores and viewer engagement trends to dissecting genre preferences and production dynamics, Tableau's robust visualization capabilities have allowed us to uncover actionable intelligence.

The interactive nature of Tableau has been pivotal, enabling us to dynamically interact with data, uncover hidden patterns, and gain a deeper understanding of audience behaviours and market trends. This agility has empowered stakeholders to make informed decisions, optimize content strategies, and capitalize on emerging opportunities.

Furthermore, Tableau's ability to seamlessly integrate and visualize complex datasets has streamlined the analysis process, offering clarity and precision in interpreting industry trends.

By presenting data in intuitive and visually compelling formats, Tableau has facilitated communication and alignment across teams, fostering collaborative insights and strategic alignment. Ultimately, the insights derived from Tableau visualizations have not only enhanced our understanding of audience preferences and content performance but have also equipped us with the tools to innovate and drive sustainable growth in the dynamic landscape of the entertainment sector.

7. REFERENCES

- [1] Zhang Y, Zhang L. Movie recommendation algorithm based on sentiment analysis and LDA. *Procedia Computer Science* 2022; 199: 871–878. doi: 10.1016/j.procs.2022.01.109
- [2] *Environment and Social Psychology* (2023) Volume 8 Issue 3. Analyze IMDb movies by sentiment and topic analysis, doi: 10.54517/esp.v8i3.1958.
- [3] Bhuvaneshwari P, Rao AN, Robinson YH, Thippeswamy MN. Sentiment analysis for user reviews using Bi-LSTM self-attention-based CNN model. *Multimedia Tools and Applications* 2022; 81(9): 12405–12419. doi: 10.1007/s11042-022-12410-4.
- [4] Topal K, Ozsoyoglu G. Movie review analysis: Emotion analysis of IMDb movie reviews. In: *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*; 18–21 August 2016; San Francisco, CA, USA. pp. 1170–1176.
- [5] Shaukat Z, Zulfiqar AA, Xiao C, et al. Sentiment analysis on IMDB using lexicon and neural networks. *SN Applied Sciences* 2020; 2(2): 1–10. doi: 10.1007/s42452-019-1926-x.
- [6] Sharma, Rasagy, and Venkatesh Rajamanickam. "Using interactive data visualization to explore non-linear movie narratives." *Parsons Journal for Information Mapping* (2013).
- [7] Burghardt, Manuel, Michael Kao, and Niels-Oliver Walkowski. "Scalable MovieBarcodes—an exploratory interface for the analysis of movies." *IEEE VIS Workshop on Visualization for the Digital Humanities*. Vol. 2. 2018.
- [8] Jorge, Ana, Nuno Gil, and Teresa Chambel. "Time for a New Look at the Movies through Visualization." *Proceedings of Artech* (2012): 269-278.
- [9] Kang, Hyeonsu, and Jennifer He. "Chromatography. A Quantified Visualization of Movies." (2005).
- [10] Magni, Giovanni, et al. "BORDERS. Visual analysis of Cinema's inner dynamics and evolutions. A case study based on the Internet Movie Database." *A Matter of Design. Making Society Through Science and Technology*. 2014. 647-665.