Project Report : Lights, Data, Action

Overview: This comprehensive entertainment landscape analytics project consists of four detailed dashboards analyzing Netflix content, IMDb TV shows, IMDb top 1000 movies, and Oscar awards data. The goal is to provide stakeholders with actionable insights for content strategy, production decisions, and audience targeting.

Problem Statement:

The entertainment industry is vast and complex, with multiple platforms, genres, and metrics influencing content creation, distribution, and audience reception. This project aims to address the following key challenges:

1. Understanding the relationship between critical acclaim (Oscar wins) and commercial success (gross revenue) across different genres.
2. Identifying trends in content production and popularity over time on platforms like Netflix and in the broader film industry.
3. Analyzing the correlation between genre, ratings, and audience engagement across movies and TV shows.
4. Determining the impact of directors and actors on a film's success and recognition.
5. Exploring the distribution of content across different categories and how it relates to viewer preferences.

To tackle these challenges, we've developed a comprehensive analytics suite comprising four interconnected dashboards that provide multi-dimensional insights into the entertainment industry.

Primary Audience

Entertainment Industry Executives

- Studio heads and production companies requiring data-driven insights for content investment decisions
- Streaming platform executives analyzing content performance metrics
- Programming directors evaluating genre success rates and audience preferences

Content Creators

- Film and TV show producers seeking to understand market trends
- Directors analyzing successful genre combinations
- Screenwriters researching popular themes and formats
- Production companies evaluating commercial viability of different genres

Business Analysts

- Market researchers studying entertainment industry patterns
- Data analysts tracking content performance metrics
- Investment analysts evaluating industry trends
- Strategic planners developing content strategies

Academic Researchers

- Media studies scholars analyzing industry trends
- Entertainment industry researchers studying Oscar patterns
- Film historians tracking historical performance data
- Cultural analysts examining viewing preferences

Secondary Audience

Industry Stakeholders

- Investors seeking market intelligence
- Advertising agencies planning media campaigns
- Distribution companies analyzing market potential
- Talent agencies evaluating career opportunities for clients

Media Professionals

- Entertainment journalists requiring industry insights
- Critics analyzing trends in film and television
- Media planners studying audience preferences
- Content acquisition specialists making purchasing decisions

The dashboards are designed to provide comprehensive insights while remaining accessible to both technical and non-technical users, with clear visualizations and intuitive navigation systems.

Data Sources:

Netflix dashboard: https://www.kaggle.com/datasets/shivamb/netflix-shows/data
IMDB Movies Dataset: https://www.kaggle.com/datasets/omarhanyy/imdb-top-1000
Oscar List Dataset: https://www.kaggle.com/datasets/unanimad/the-oscar-award
IMDB TV Shows Datasethttps://www.kaggle.com/datasets/devanshiipatel/imdb-tv-shows

By integrating data from these diverse sources, the project aims to provide a holistic view of the entertainment landscape, enabling data-driven decision-making for content creators, streaming platforms, and industry stakeholders.

IMDb movies dashboard overview:

The dashboard development process involved several key phases to create an interactive visualization of IMDb movie trends.
Data Integration and Preparation:

- Created individual worksheets in Tableau for different visualizations
- Implemented data filtering and cleaning for movie-specific information
- Standardized date formats spanning from 1920 to 2020
- Established connections between different visualization components

Dashboard Components
The final dashboard consists of five main visualizations:

1. A circular information display showing movie details
2. A time series graph showing the number of series released per year
3. A gross earnings trend analysis by year
4. A treemap of top 10 genres by IMDb rating
5. A bar chart displaying top 10 series by vote count

Analysis and Discoveries

Evolution of Top Voted Series:
The "Top 10 Series with Highest number of Votes" visualization shows interesting transitions across different time periods in the dashboard.

Early View (1970):

Initially, the visualization appears empty, indicating limited voting data for movies from this period, which aligns with the pre-internet era when IMDb voting wasn't available.

Mid-Period (1988):

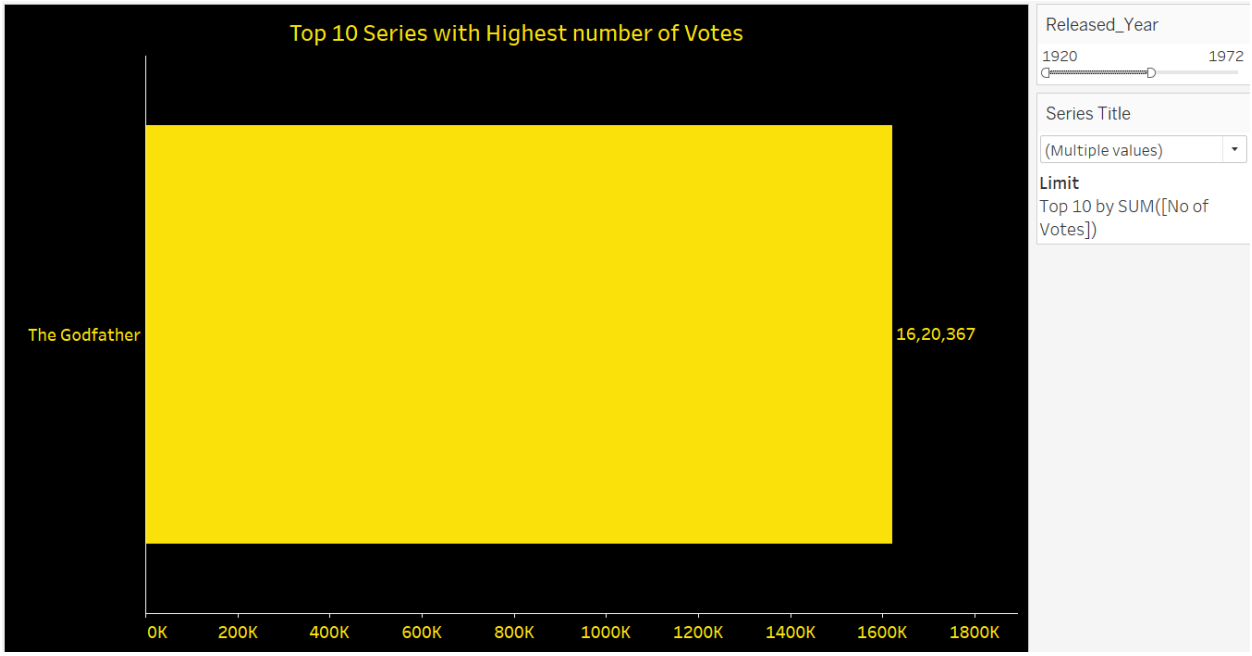
The visualization begins showing data, with "The Godfather" appearing as a significant entry with approximately 1,500K votes, demonstrating the emergence of user engagement on the platform.

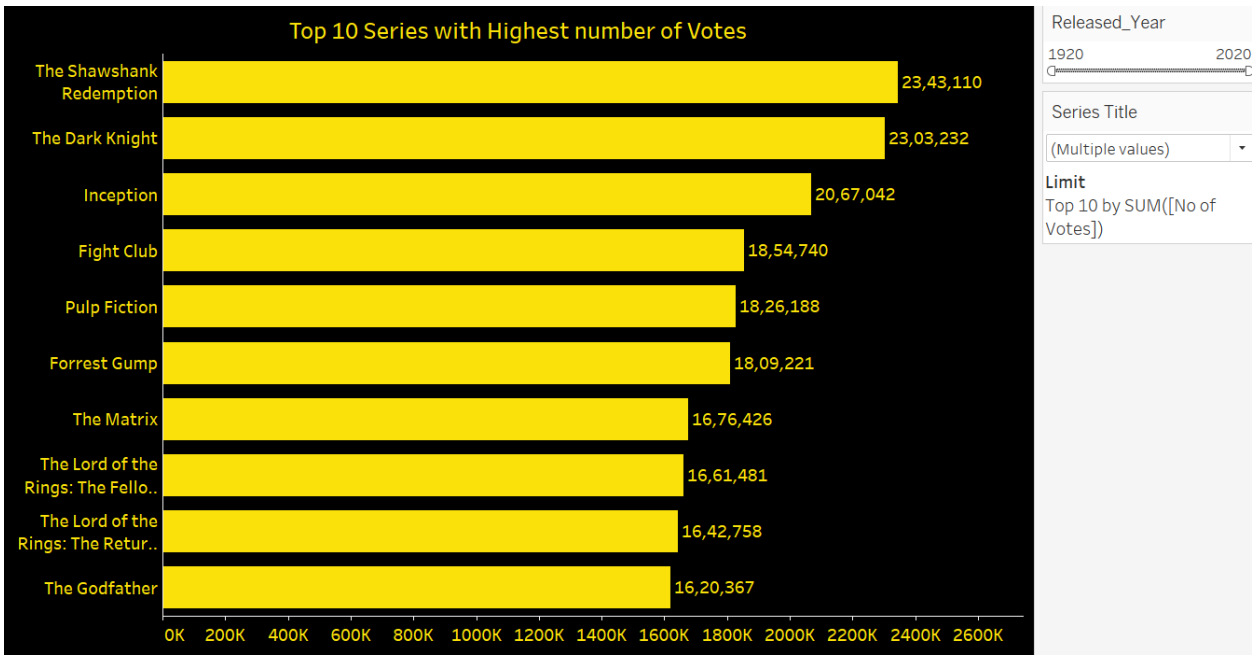
Modern Era (2020):

The visualization shows a dramatic expansion in voter participation:

- "The Dark Knight" leads with 23,03,432 votes
- "Inception" follows with 20,67,042 votes
- "Fight Club" garners 18,54,740 votes

- Classic films like "The Godfather" and "Pulp Fiction" maintain strong voting numbers
- Modern franchises like "The Lord of the Rings" and "The Matrix" show significant voter engagement





This transition reflects both the growing accessibility of online voting platforms and the expanding audience engagement with both contemporary and classic films over time. The data suggests that newer blockbusters tend to accumulate higher vote counts, while certain classic films maintain consistent popularity among viewers.

Implementation Challenges and Solutions:

Technical Challenges

- Creating interactive filters required careful parameter setting to ensure proper data refresh
- Resolved by implementing a dual-filter system at the top (Series selection and Release Year range)
- Addressed visualization space constraints through careful layout optimization

Data Integration Issues

- Multiple genre combinations created complexity in categorization
- Solved by implementing a hierarchical genre classification system
- Managed missing gross earnings data by focusing on complete records for trend analysis

Visualization Improvements

Based on the feedbacks, the dashboard successfully implements:

- Dark theme for better readability
- Clear visual hierarchy with consistent color coding
- Interactive elements allowing users to explore data across different dimensions
- Balanced layout combining both detailed and overview information

Netflix dashboard overview:

The streaming industry has witnessed a massive surge, and Netflix remains a dominant platform with a diverse content library. Our dashboard aims to analyze key trends within the Netflix catalog to provide insights into the distribution of content, trends over time, and factors influencing user engagement. Netflix dashboard development process involved several key phases to create an interactive visualization of movies and TV shows across Netflix platform based on dataset using key fields:

- *Title*: Name of the movie/TV show
- show_id: 8-digit unique identifier (assigned unique identifier for dataset)
- *Type*: Content type – Movie or TV Show
- *Release Year*: Year the content was released

- *Duration*: Runtime for movies and seasons for TV shows
- *Listed_in*: Associated genres for the titles
- *Country*: Country where the movie/show was produced
- *Rating*: TV parental rating of the content
- *Date_added*: When the title was added to Netflix
- Director: Director who created the movie/show
- Cast: Actors involved in the content
- Description: Brief description about the content it consists

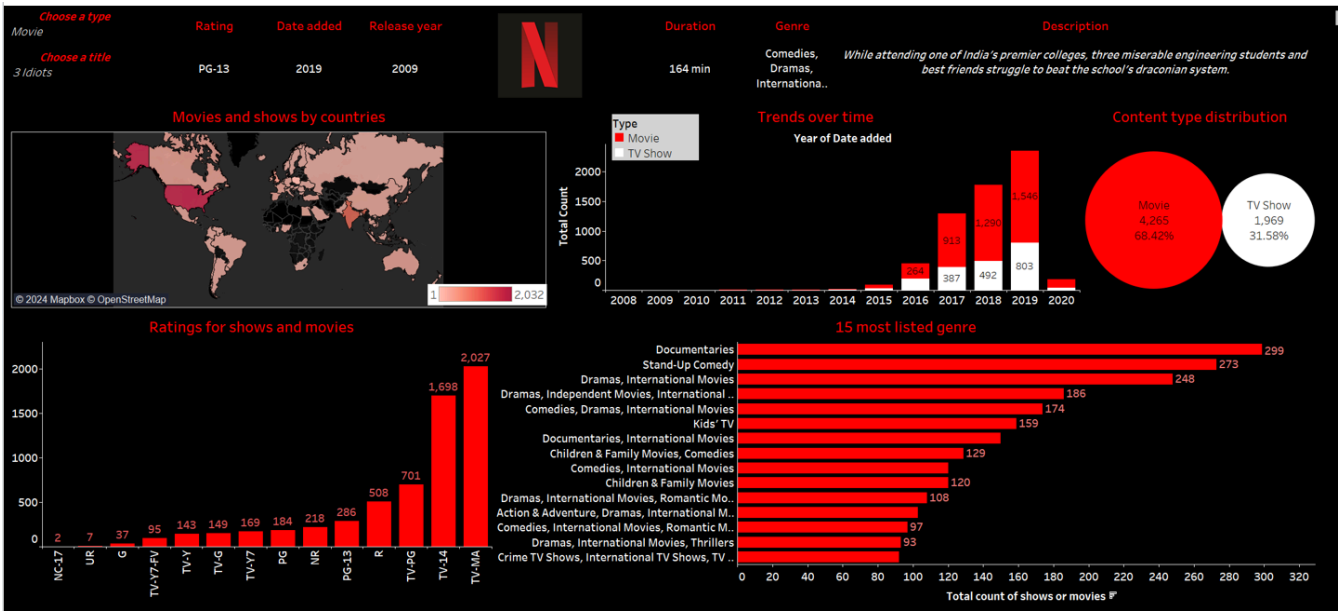
Data Integration and Preparation:

- Assigned 8-digit unique identifier for dataset
- Created individual worksheets in Tableau for different visualizations
- Implemented interactive filters for description, date added, duration, rating, release year and genre to provide a holistic representation of a specific content.
- Integration of the 5 different visualizations on the dashboard alongside an interactive filter which allows a user to show the content type (movie/TV show), the title and view the details associated in a unified view.

Dashboard components:

The main dashboard (custom size: 2200x1100 px) successfully implements:

- Dark theme (Netflix theme) for better readability
- Clear visual hierarchy with consistent color coding
- Interactive filters allowing user to explore titles and content types
- Balanced layout combining both detailed and overview information.

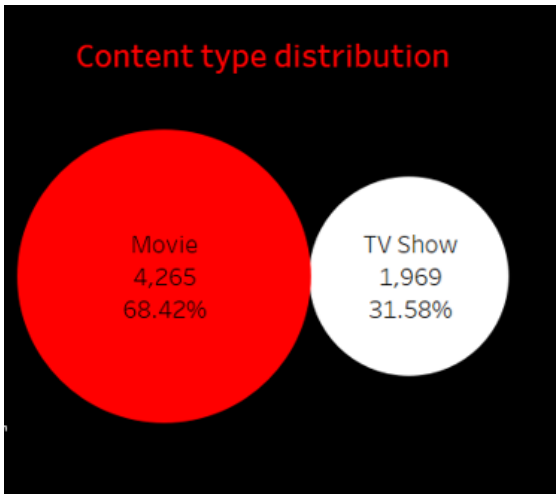


It consists of 5 main and 6 other visualizations. It also consists of 2 filters (single dropdown list) for an easy selection of content type and title. The visualizations are:

6. A packed bubble chart representing the distribution of content between movies and TV shows
7. Content addition of movies and shows yearwise (2008-2020).
8. 15 most listed genres across the Netflix content.
9. A symbol map displaying the content production across countries.
10. Distribution of content across different TV parental ratings.
11. Combination of description, date added, duration, rating, release year and genre associated with an individual content (uses an interactive filter)

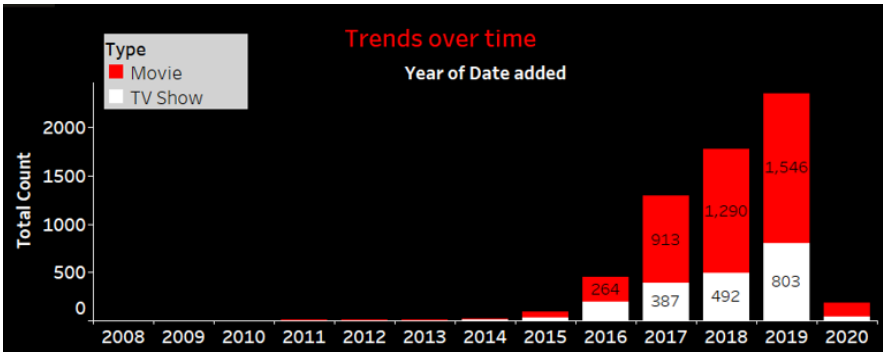
Analysis and Discoveries

Content Type Distribution: A packed bubble chart representing the proportion of movies versus TV shows in Netflix's library.



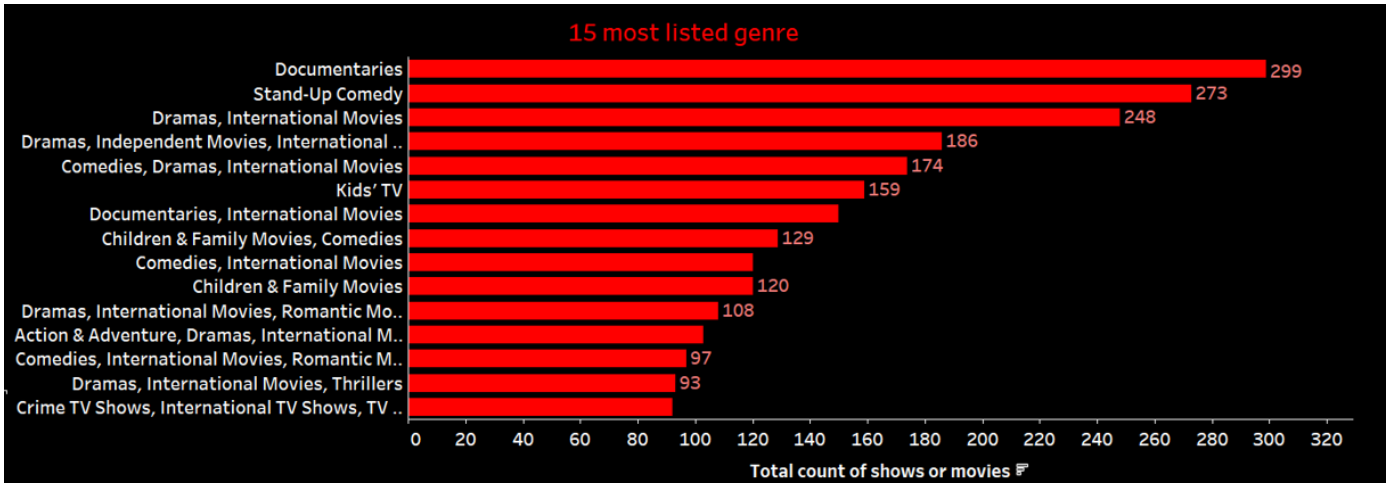
- **Result**: Netflix has a significantly higher number of movies compared to TV shows which shows that Netflix library is primarily dominated by movies which accounts for nearly 70% of its content, whereas TV shows account for 32%.

Trends Over Time: A stacked bar chart displaying the number of titles added per year for the period 2008-2010 (legend and labels for total count included)



- **Result**: A steep increase in content additions was observed after 2015 for both content types- Movies and TV shows, aligning with Netflix's transition to a global streaming platform.

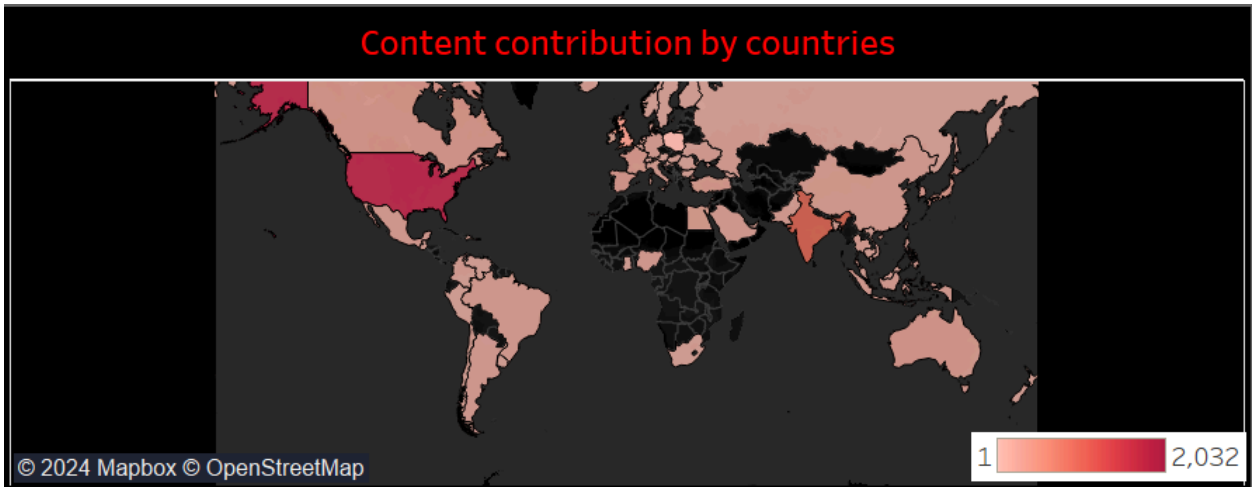
Top 15 most listed genre: A horizontal bar chart visualizing the top 15 genres by genre count across the platform (labels for total count included)



- **Result**: Popular genres include documentaries, stand up-comedy, drama/international movies, reflecting audience preferences for these categories.

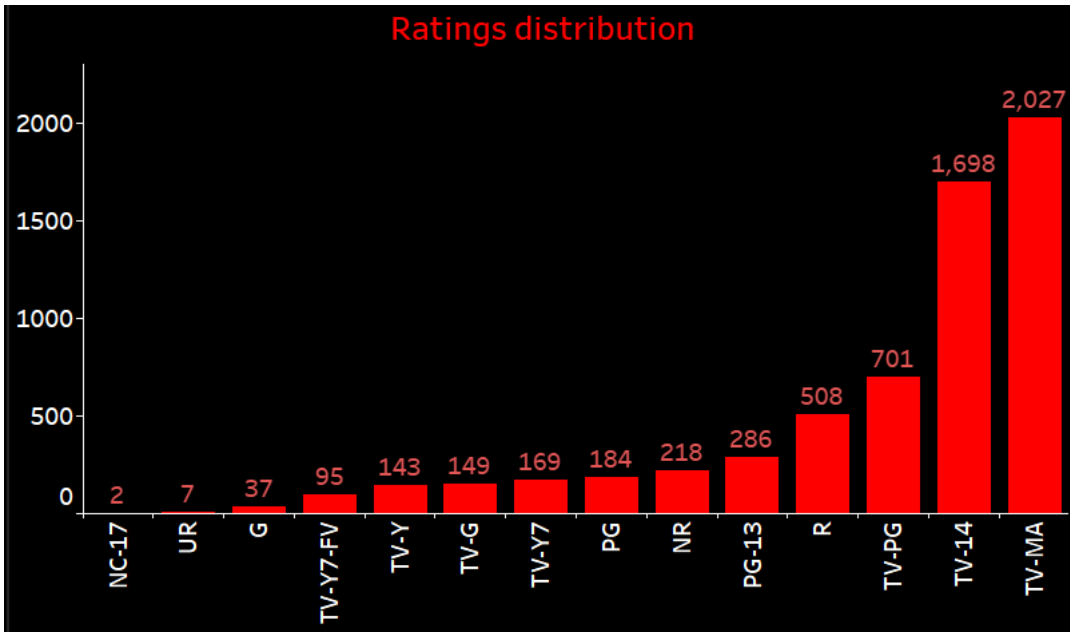
Content contribution by countries: A map showing the countries producing the most content for Netflix. (legend included: darker shade shows greater number of content

produced)



- **Result**: The United States leads in content production, followed by India and the United Kingdom.

Ratings Distribution: A histogram showcasing the spread of movies and shows by ratings given. (labels for total count included)



- **Result**: Most titles have parental ratings of TV-MA , TV-14 and TV-PG, indicating that the content is suitable for mature audiences, parents strongly cautioned and parental guidance suggested respectively.

The key highlights these visualizations are:

- **Dominance of Movies**: Movies from the majority of Netflix's catalog.
- **Content Growth**: There was exponential growth in titles post-2015. However, there's a decline in the number of movies and shows added in 2020, hinting at the global shutdown of movies & TV shows production due to pandemic. This pause in production caused Netflix to put a halt at its year-wise growth in content delivery.
- **Popular Genres**: Documentaries, stand up-comedy and dramas/international movies led the genre mix. It represents the audience preference for these content types and shows diverse, broad appeal among them.
- **Global Trends**: The United States contributed the highest number of titles, but other regions like India are emerging strongly.
- **Content release across years**: Netflix grew its content at an impressive rate every year especially after 2015. In the year 2017, 913 movies were added to its library which shows a ~246% increase over 2016. It also shows that Netflix heavily invested in producing original series and movies, leading to a substantial increase in their unique content library over the past decade.

Implementation Challenges and Solutions

- Challenge: The dataset size occasionally slowed dashboard color rendering. -Solution: Create the data extract for faster data retrieval, and unnecessary fields were excluded.
- Challenge: Aligning the colors for each visualization so it appears consistent on the main dashboard.

  -Solution: Used the same color code for each visualization.

- Challenge: Use of interactive filters and misc.visualizations on the main dashboard so it fits a single view of the dashboard.

  -Solution: Used the different blank layouts for each visualization (rating, description, title, date added, released year) and resized it on the top to make it a banner.

- Challenge: The main dashboard uses two filters, to select content type and title, the interactive filter needed to display only Movie titles when selected as 'movies' and not TV shows.

  -Solution: Kept the option 'Show relevant values only' in the filter customization option and deselected 'Show All values'.

- Challenge: To fit all the visualizations on a single dashboard. Since the dashboard consists of a horizontal bar chart, a packed bubble chart, a stacked bar chart, a world map symbol, a histogram, misc. Visualizations on the top and a Netflix logo, the standard view on Tableau was not enough to view it in a unified view.

-Solution: Increased the dashboard size to custom layout (2200x100 px) which gives a much clearer representation in a unified view.

Feedback and Incorporation

I asked my roommates to review the visualizations and dashboard to know their feedbacks. Based on their suggestions, I added layouts of different visualizations, including filters to choose title and content type. I changed the color codes of legends to match with the dark theme of our workbooks overall.

Additionally, I enhanced the tooltips with more detailed information about movies, directors, and actors.

Movie Oscar dashboard overview:

Data Integration and Preparation

- Imported two datasets: *IMDB Top 1000 Movies* and *Oscar List* into Tableau.
- Cleaned data by ensuring uniform formats for common fields like Movie Title and Year.
- Split the Genre field in the IMDB dataset to handle multiple genres per movie, creating pivot fields for the actors mentioned in the dataset that performed the role in the movies.
- Linked the datasets using Series Title and Film as the key to establish relationships between IMDB ratings, gross revenue, and Oscar table details.

Dashboard Components:

Built the following visualizations to explore and present meaningful insights:

1. **Oscar Wins vs. Gross Revenue by Genre**: Shows the financial success of Oscar-winning movies across genres.
2. **Top Directors with Most Oscar-Winning Movies**: Highlighted directors with exceptional performance in awards.

3. **Top Actors Appearing in Oscar-Winning Movies**: Showcased actors frequently appearing in award-winning films.
4. **Oscar-Winning Movies by Category**: Explored the distribution of awards across categories like Best Picture, Best Director, etc.
5. **Distribution of Oscar-Winning Movies**: Analyzed the prevalence of genres in Oscar-winning films.

Dashboard Creation: Combined all visualizations into a single interactive dashboard, incorporating global filters (Genre, Year, Winner) and action filters for seamless exploration.

The dashboard includes three key interactive filters that enhance user exploration and provide dynamic insights.

Winner Filter: Filters movies based on whether they have won an Oscar or not.

Impact: Updates all charts to display relevant data for Oscar-winning or non-winning movies.

Year Released Filter: Filters movies based on their release year.

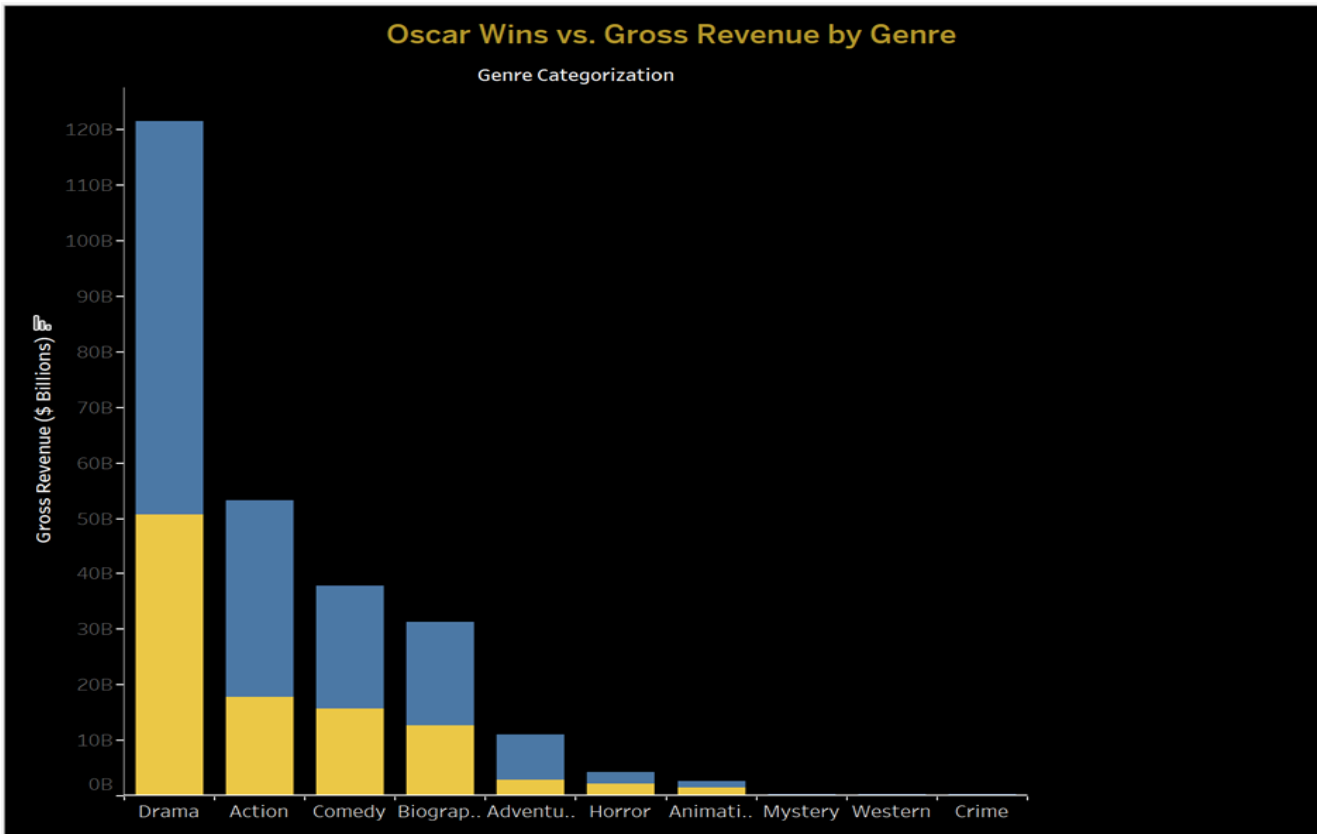
Impact: Updates visualization (ex. gross revenue, top directors/actors) to focus on movies within the selected timeframe.

Genre Categorization Filter: Filters movies based on their genres (e.g., Drama, Action, Comedy, etc.). Multi-select dropdown that enables users to choose one or multiple genres.

Impact: Adjusts charts to show metrics for the selected genres. For example, Revenue distribution, top directors, and actors update based on chosen genres.

Analysis and Discoveries

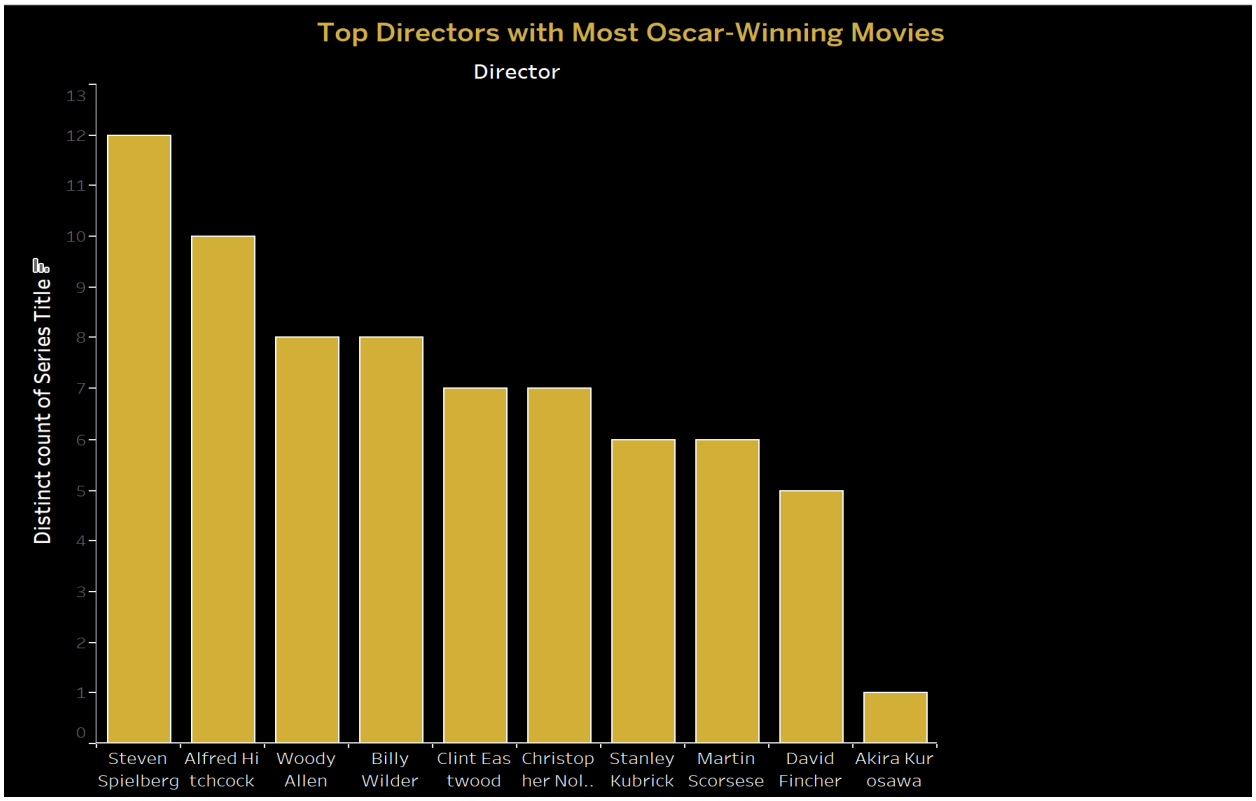
Oscar Wins vs. Gross Revenue by Genre



Key Insights:

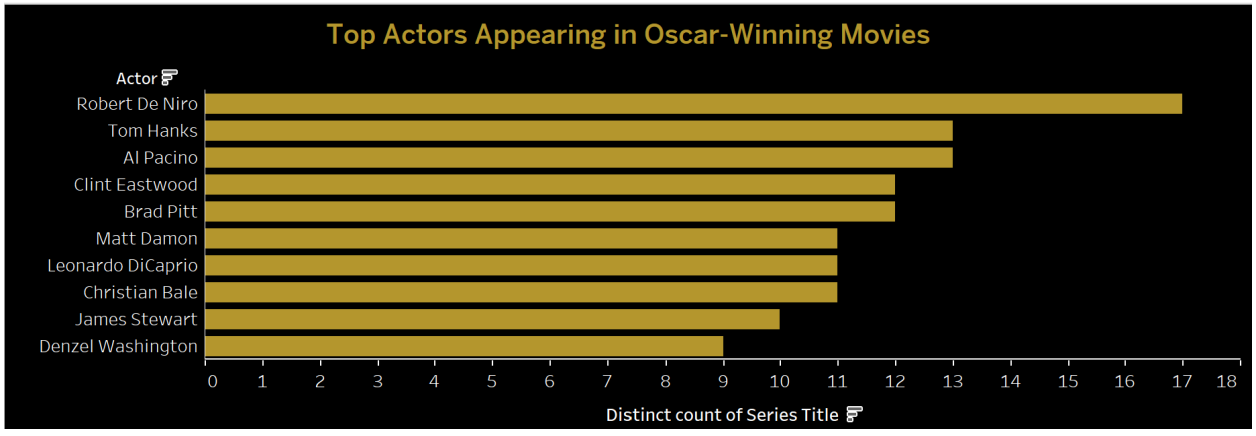
1. **Drama Leads in Revenue and Oscar Wins:**
   Drama stands out as the top-grossing genre and boasts the highest proportion of Oscar-winning movies, highlighting its strong appeal to both audiences and Oscar committees.
2. **Action and Comedy: High Revenue, Low Oscar Recognition:**
   Action and Comedy generate significant gross revenue but their proportion of Oscar-winning movies is relatively low revealing a gap between audience popularity and Oscar committee preferences.
3. **Biography: Critical Success Despite Lower Revenue:**
   The Biography genre, while does not generate much Gross Revenue but still holds a significant share of Oscar wins, which represents alignment with traditional Oscar themes and critical recognition
4. **Lesser Oscar Success in Adventure and Horror:**
   Adventure and Horror genres contribute little to overall gross revenue and have only a few Oscar-winning films, reflecting their limited recognition by the Oscars.
5. **Insights Through Filtering:**
   With the Released Year filter, trends over time can be explored, helping identify shifts in genre popularity and Oscar recognition.

Top Directors with Most Oscar-Winning Movies:



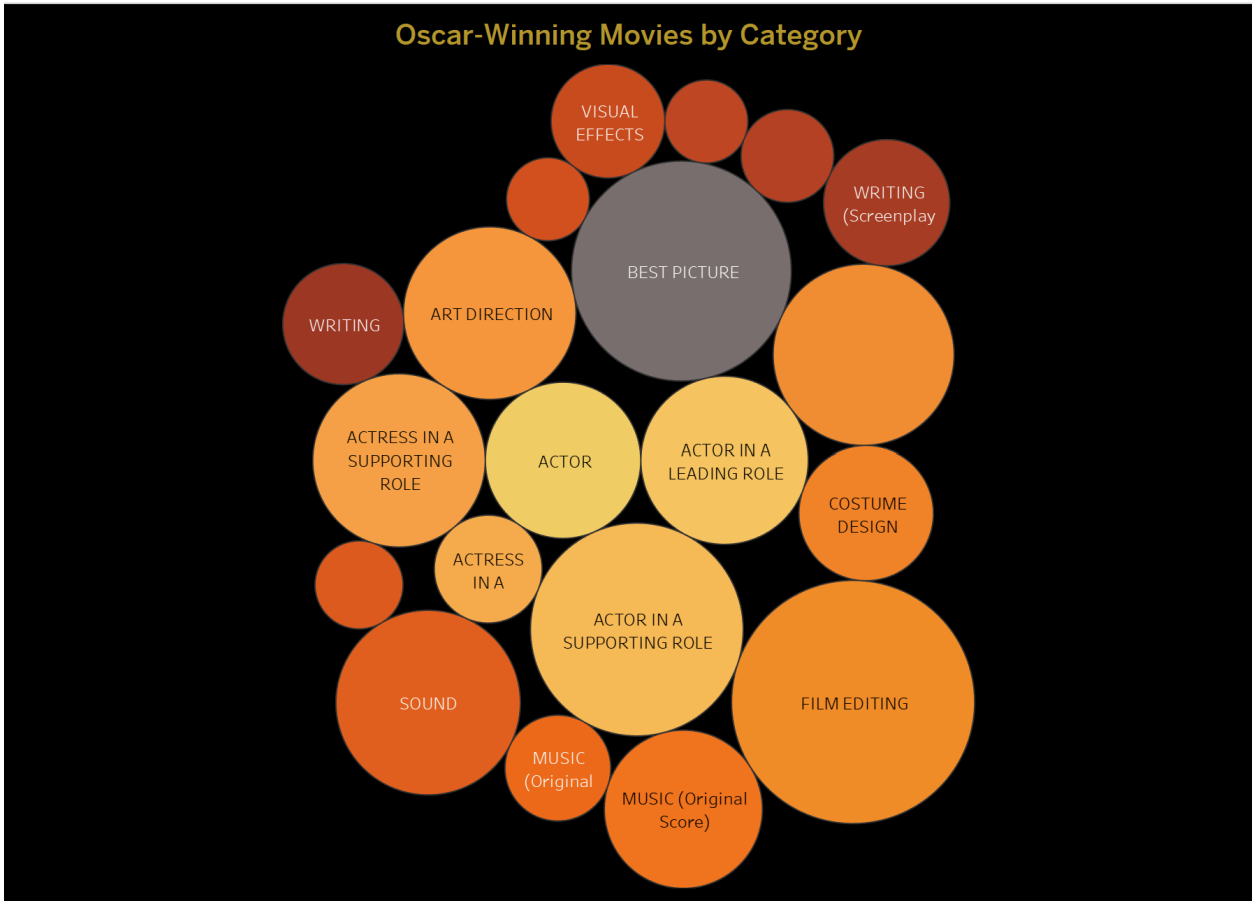
1. Directors like Steven Spielberg and Martin Scorsese have consistently excelled in filmmaking, earning frequent Oscar wins.
2. Identified a pattern of directors excelling in particular genres, such as Scorsese in Drama.

Top Actors Appearing in Oscar-Winning Movies:



1. Actors like Robert De Niro and Tom Hanks frequently starred in Oscar-winning films, showcasing their talent and selection of impactful roles.

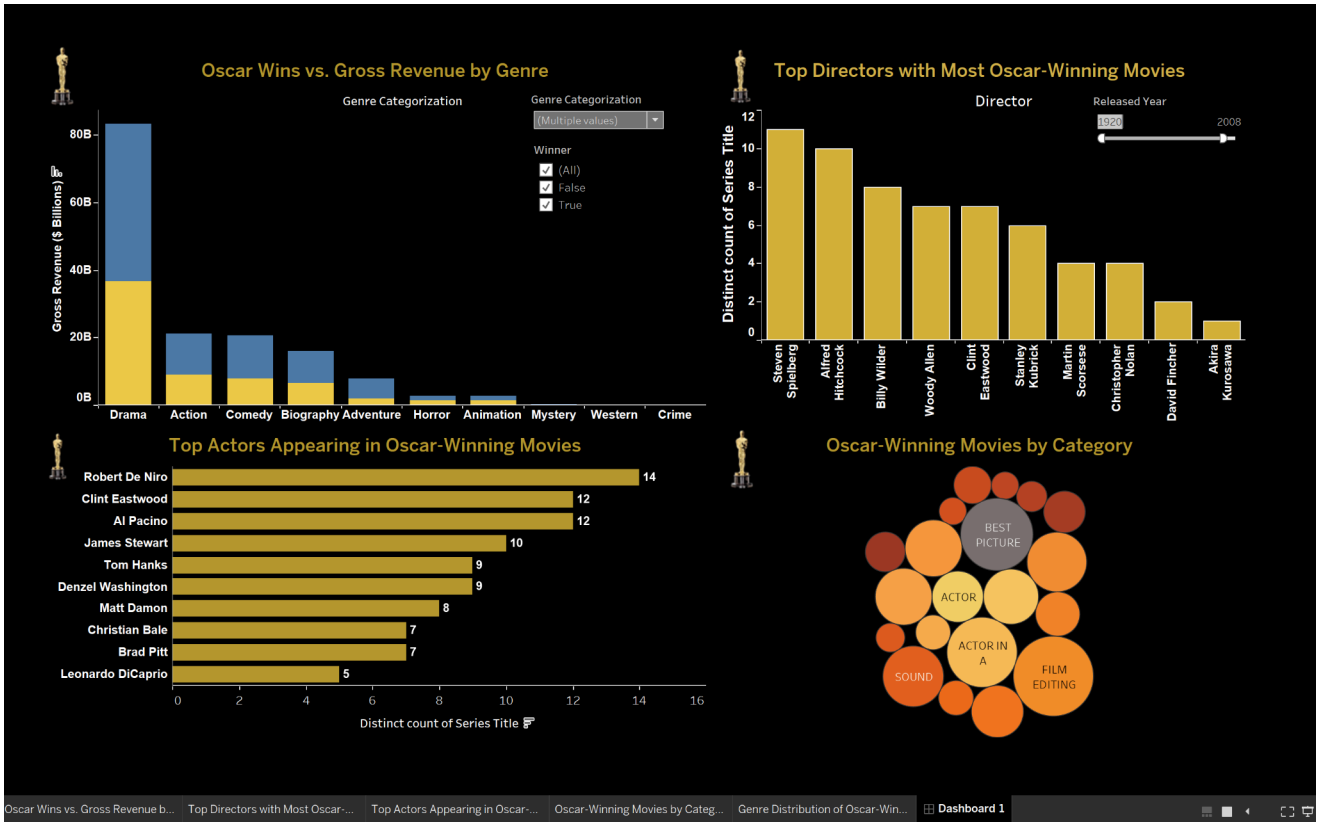
Oscar-Winning Movies by Category:



This visualization highlights the distribution of Oscars across diverse categories, emphasizing the significance of Best Picture, performance awards, and technical achievements while showcasing a balanced evaluation of cinematic excellence.

1. Best Picture is the most prominent category, showcasing its importance as the ultimate recognition of overall movie excellence.
2. Performance awards (e.g. Actor/Actress in Leading and Supporting Roles) dominate alongside technical categories like Film Editing and Sound, highlighting a balance between artistic and technical achievements.

Final Dashboard:



Implementation Challenges and Solutions

Handling Multiple Genres:

- The inclusion of multiple genres for each movie added complexity to the analysis.
- Solution: Split the Genre field into separate rows per movie, enabling granular exploration of genres.

Data Integration Issues:

- Some movies appeared in one dataset but not the other, resulting in incomplete joins. Not all movies were included in the Oscar list dataset, causing issues when merging it with the movies dataset. This led to numerous null values, requiring extensive cleaning for accurate visualization.
- Solution: Conducted a manual review to address mismatches where feasible and removed incomplete entries from key visualizations by generating a cleaned data file for the Oscar_list dataset.

Adjustments from the Original Plan

- Initially planned to analyze all genres equally but shifted focus to the most represented genres (Drama, Action, Adventure) due to data constraints.
- Modified the approach to account for multiple genres per movie, ensuring accurate representation in genre-based analyses.

Feedback and Incorporation:

I asked my friend to review the visualization and share his feedback. Based on his suggestions, I added more interactivity to improve the exploration of genre and award relationships, including filters for Genre, Year, and Oscar Category.

Additionally, I enhanced the tooltips with more detailed information about movies, directors, and actors.

IMDB TV shows Dashboard overview:

The IMDB TV Shows Dashboard we built is an interactive and insightful dashboard designed to analyze trends and patterns in TV show data. We visualized key metrics like genre performance, ratings distribution, episode duration, and show popularity to uncover audience preferences and content success factors. The dashboard highlights correlations, outliers, and insights,giving a comprehensive view of the tv entertainment landscape.

Data Integration and Preparation:

- We identified and downloaded the dataset from IMDB's official dataset portal. The datasets included TV show titles, genres, ratings, votes, episode durations, and other related metadata.
- Missing Fields like ratings and votes contained null entries, which were removed, fields like episode duration was standardized and fields like genres and actors were split for compatibility in visualizations
- Custom bins were created to group ratings and Calculated fields were added to determine average ratings.

Discovery and Analysis

- **Genre Vs Rating** - A heatmap to display the relationship between genres and their average ratings.
  We found that biography emerged as the highest-rated genre, with an average rating of 7.997. Other top-performing genres included Mystery (7.933) and Animation (7.438). Music and Romance, however, were rated lower on average

- **Genre Analysis** - A horizontal bar chart to represent the frequency of each genre.
  Comedy and Drama were the most frequently occurring genres, followed by Animation and Crime. Niche genres like Biography and Short had fewer titles, but their ratings tended to be higher.
- **Rating Distribution** - A histogram to highlight how TV show ratings were distributed.
  Most ratings were concentrated between 6.0 and 8.0, forming a bell-shaped curve. Very few shows achieved exceptional ratings (above 9.0) or fell below 4.0, indicating a general clustering of quality.
- **Episode Duration vs. Rating** - A scatter plot for the relationship between episode duration and ratings.
  Shows with longer episode durations tended to receive higher ratings, suggesting that audiences favor more in-depth storytelling.
- **Top TV Shows by Rating** - A ranked table displayed the highest-rated shows.
  Shows like *The Wire* and *Breaking Bad* achieved near-perfect ratings which shows their broad critical and audience acclaim.

Challenges Faced:

- Filtering required for missing values and inconsistencies in ratings, votes, and episode durations
- Distorted scatter plots, resolved by disabling aggregation and configuring with raw data.
- Selecting appropriate bin sizes for ratings distribution to balance clarity and detail.
- Filtering extreme outliers in episode durations was necessary to ensure clean and interpretable visualizations.

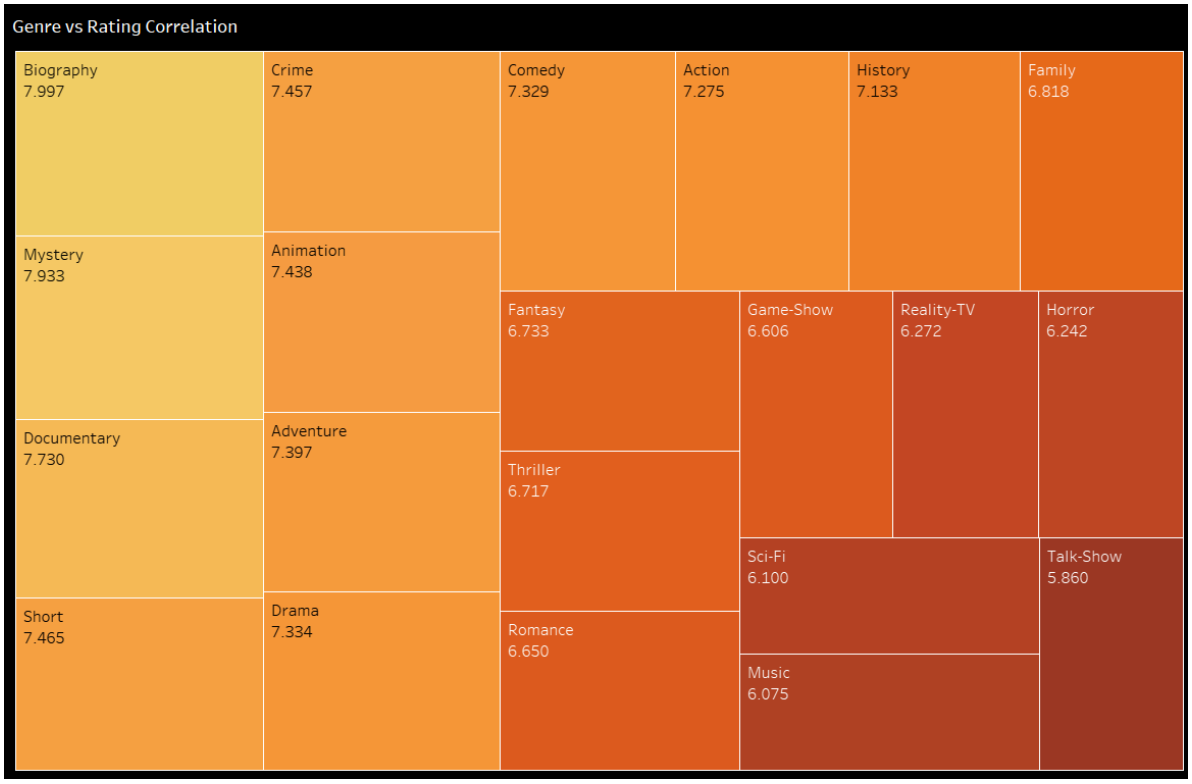
Adjustments made based on feedback:

- We added dropdowns for sub-genres and sliders for rating to make the dashboard more interactive.
- Simplified the dashboard by focusing on the most actionable visualizations, such as the scatter plot and heatmap.

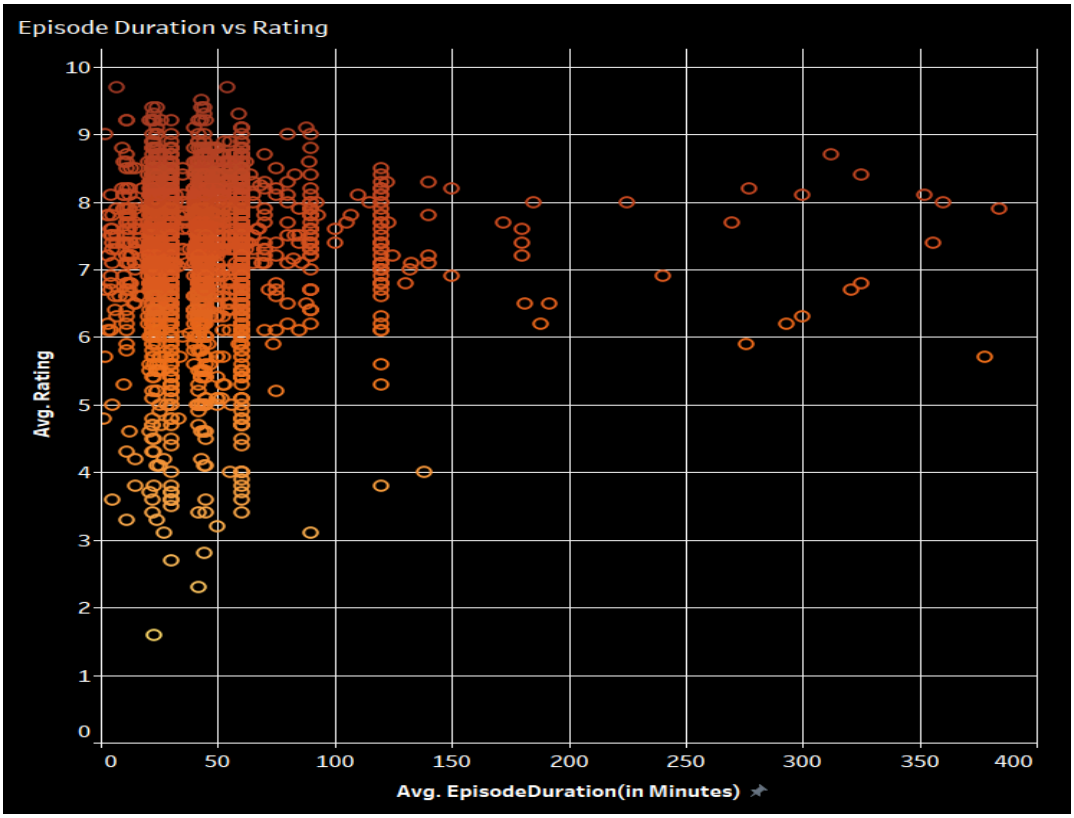
Key Insights:

- The Genre vs. Rating Correlation heatmap revealed disparities in average ratings across genres. Biography is the highest-rated genre followed by Mystery. These genres consistently outperformed others, suggesting that audiences appreciate detailed, real-life stories instead of genres like Music and Romance. This

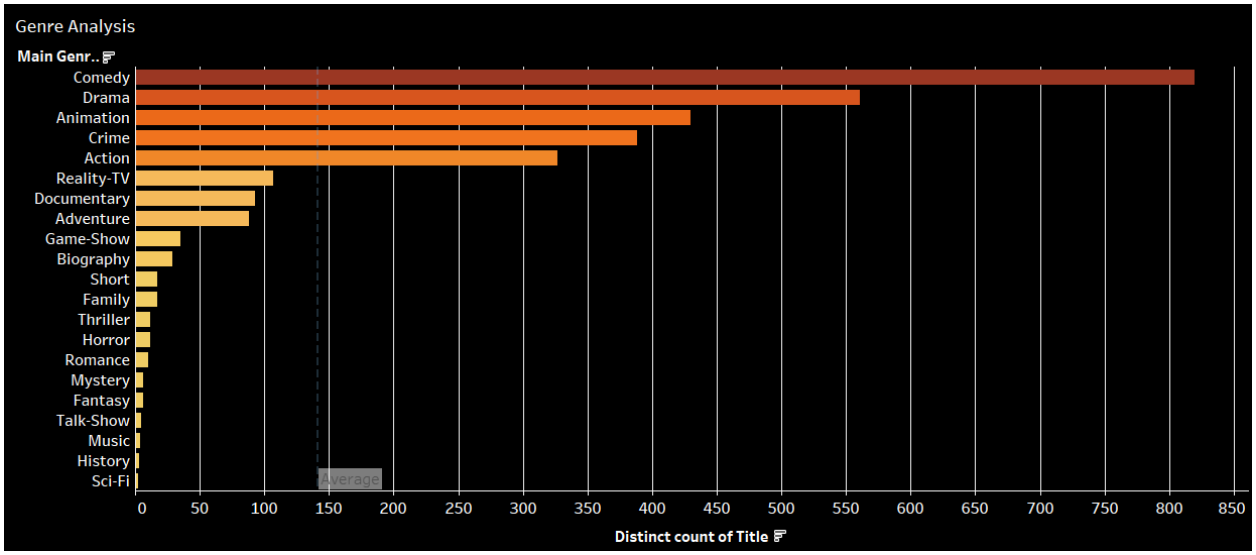
highlights opportunities for content creators to focus on genres that resonate strongly with audiences.



- The Episode Duration vs. Rating scatter plot shows with longer episodes, often above 60 minutes, tended to receive higher ratings, suggesting that audiences appreciate in-depth storytelling and detailed narratives. However, a cluster of shorter episodes also performed well this demonstrated that while longer episodes generally attract higher ratings, creative storytelling can still make shorter content stand out.



- The Genre Analysis bar chart revealed that certain genres, such as Comedy, Drama, and Animation, had the highest number of shows, indicating their popularity among content producers. However, when cross-referenced with average ratings, these high-volume genres did not necessarily achieve the highest ratings.



Conclusion from all dashboards: The dashboard findings underscore the role of platforms and award recognition in shaping viewer habits and the cultural significance of content. This centralized view highlights a dynamic entertainment landscape, emphasizing both critical acclaim and accessibility through streaming.

1. IMDB content & Netflix: The correlation between IMDB top-rated content and their availability on streaming platforms like Netflix reflects the importance of critically acclaimed shows to streaming catalogs. Shows with high IMDB ratings tend to be a key driver for audience retention and platform subscriptions.
2. Oscar Movies & IMDB Movies: Movies listed in the IMDB top 1000 movies often overlap with those awarded Oscars, showcasing a strong connection between critical reception and audience ratings. This validates that films achieving awards also resonate broadly with audiences globally, aligning quality with popularity.
3. Netflix & Oscar Movies: Netflix's growing inclusion of Oscar-winning films in its catalog bridges the gap between streaming convenience and critically acclaimed cinema. Netflix originals receiving Oscar nominations further emphasize the platform's growing influence in producing high-quality, award-worthy content driving user subscriptions.
4. Growth of streaming platform content: The evolution of audience preferences, transitioning from traditional theaters to streaming services.
5. Audience engagement: The importance of quality content, as evidenced by high IMDB ratings and prestigious awards, driving viewership on platforms like Netflix.
6. Popular genre: Across all workbooks, recurring popular genres (Drama, Comedy, and Thriller) appear frequently. This suggests a global preference for narratives that emotionally engage audiences or offer suspenseful entertainment.