Insights for Microsoft's New Movie Studio

"Exploring Box Office Trends and Recommendations"

Name: Gladys Kemunto Mosota

Date: 3rd June 2024

Blog post URL: https://github.com/Mosota-Kemunto-Gladys-

2020/Gladys-Mosota_dsc_phase_1_project.git

Overview

• Microsoft is creating a new movie studio and needs insights into what types of films are performing well at the box office. This project aims to provide actionable insights for Microsoft's new movie studio by analyzing current box office trends.

Project Goal:

• Provide actionable insights for Microsoft's new movie studio by analyzing current box office trends.

Key Questions

a. What types of films are performing well?

b. What are the optimal release strategies?

C. How can Microsoft maximize profitability?

Data Understanding

Datasets from Box Office Mojo and IMDB will be used.

Box Office Mojo (bom.movie_gross.csv.gz):

- Contains information about movie grosses.
- Key columns: movie_title, domestic_gross, worldwide_gross, release_date.

IMDB (im.db.zip):

- SQLite database with multiple tables.
- Relevant tables: movie_basics (contains movie details), movie_ratings (contains ratings).

Data Analysis

Exploratory data analysis will identify trends and insights in box office performance, genres, and ratings.

Steps Followed:

- 1. Data Cleaning and Preparation: Ensured data quality and merged datasets.
- 2. Exploratory Data Analysis: Identified key trends and patterns.
- 3. Visualizations: Created visuals to support findings and recommendations.
- 4. Recommendations: Provided actionable insights based on data analysis.

Data Analysis - bom.movie_gross.csv.gz

To analyze the bom_movie_gross dataset, we will perform a series of steps to explore the data, understand trends, and extract insights.

Initial Stages:

- 1. Check the structure of the DataFrame.
- 2. Look at the summary statistics.
- 3. Identify any missing values.

Calculate Total Gross:

Add a new column total_gross which is the sum of domestic_gross and foreign_gross.

Analyze Trends Over Time:

- Group the data by Year and calculate the total and average gross.
- Visualize the trends over time using plots.

Identify Top Performers:

- Find the top movies by total gross.
- Compare domestic and foreign gross for these top movies.

Distribution Analysis:

- Analyze the distribution of domestic and foreign gross.
- Check for any patterns or anomalies.

Visualize Insights:

Create visualizations to better understand and present the data.

Summary Statistics (bom.movie_gross.csv.gz)

Statistic	Domestic Gross	Foreign Gross	Total Gross
Count	3387	3387	3387
Mean	\$28,508,205.96	\$45,029,792.91	\$73,537,998.87
Std Dev	\$66,755,749.11	\$112,684,346.56	\$170,509,096.95
Min	\$0.00	\$0.00	\$100.00
25th Percentile	\$111,500.00	\$0.00	\$274,000.00
50th Percentile	\$1,300,000.00	\$1,500,000.00	\$5,475,000.00
75th Percentile	\$27,500,000.00	\$29,150,000.00	\$61,350,000.00
Max	\$936,700,000.00	\$960,500,000.00	\$1,518,900,000.00

Key Insights:

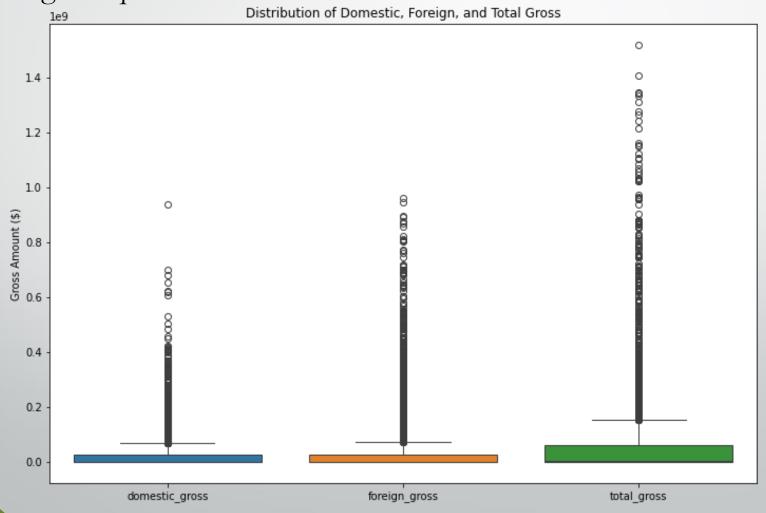
- Average Performance: Films on average make \$28.5 million domestically, \$45 million internationally, and \$73.5 million in total. Microsoft should focus on creating films with international appeal to maximize revenue.
- Revenue Variability: High variability with standard deviations of \$66.8 million (domestic), \$112.7 million (foreign), and \$170.5 million (total) suggests an unpredictable market. Diversifying the film portfolio can help manage risk.

Business Recommendations:

- International Focus: Given higher average foreign gross, Microsoft should aim for films with global appeal through diverse casting and international marketing.
- Diversified Portfolio: To balance risk and reward, Microsoft should produce a mix of lower-budget films with high return potential and a few high-budget films that could become blockbusters.

Visualizing the Data - bom.movie_gross.csv.gz

To support these insights with visualizations, boxplots, bar charts, or line graphs will be used to show the distribution of domestic, foreign, and total gross, and highlighting the quartiles.



Box Plot Interpretation

Domestic Gross (Left Plot):

- 1. Median (Middle Line of the Box): The median domestic gross is indicated by the horizontal line within the box. This represents the 50th percentile of the data.
- 11. Interquartile Range (IQR, Box Edges): The edges of the box represent the 25th and 75th percentiles, showing where the middle 50% of the data lie.
- 111. Whiskers: The lines extending from the box (whiskers) indicate the range within 1.5 times the IQR from the quartiles. Data points outside this range are considered outliers.
- 1V. Outliers (Dots Beyond Whiskers): The dots above the whiskers represent outliers, which are domestic gross values significantly higher than the rest of the data.

Foreign Gross (Middle Plot):

- **Median:** The median foreign gross is indicated by the middle line within the box.
- IQR: The box shows the 25th to 75th percentile range for foreign gross.
- Whiskers and Outliers: The whiskers indicate the range of typical values, and the dots above the whiskers are outliers, representing significantly high foreign gross values.

Total Gross (Right Plot):

- ✓ **Median:** The median total gross is indicated by the middle line within the box.
- ✓ **IQR:** The box represents the 25th to 75th percentile range for total gross.
- Whiskers and Outliers: The whiskers show the range of typical values, and the outliers are indicated by the dots above the whiskers, representing significantly high total gross values.

Key Insights for Microsoft

High Variability:

- All three distributions show a high level of variability, as indicated by the presence of many outliers and the spread of the whiskers.
- This suggests that while some films have average gross values, others perform exceptionally well or poorly.

Significant Outliers:

The presence of many outliers in all three plots indicates that there are several films with exceptionally high gross values. These outliers could represent blockbuster movies that significantly outperform typical films.

Domestic vs. Foreign Gross:

Comparing the domestic_gross and foreign_gross, the foreign gross seems to have a higher median and potentially higher maximum values, indicating that films might generally perform better in foreign markets than domestically.

Total Gross:

The total_gross plot consolidates the domestic and foreign gross, showing a wider range and more outliers. This indicates that the total gross can vary significantly, influenced by both domestic and foreign performances.

Recommendations for Microsoft

Focus on International Markets:

Given the higher median and more significant outliers in the foreign gross, Microsoft should consider targeting international markets when producing new films. This could involve creating content with global appeal and investing in international marketing campaigns.

Diverse Film Portfolio:

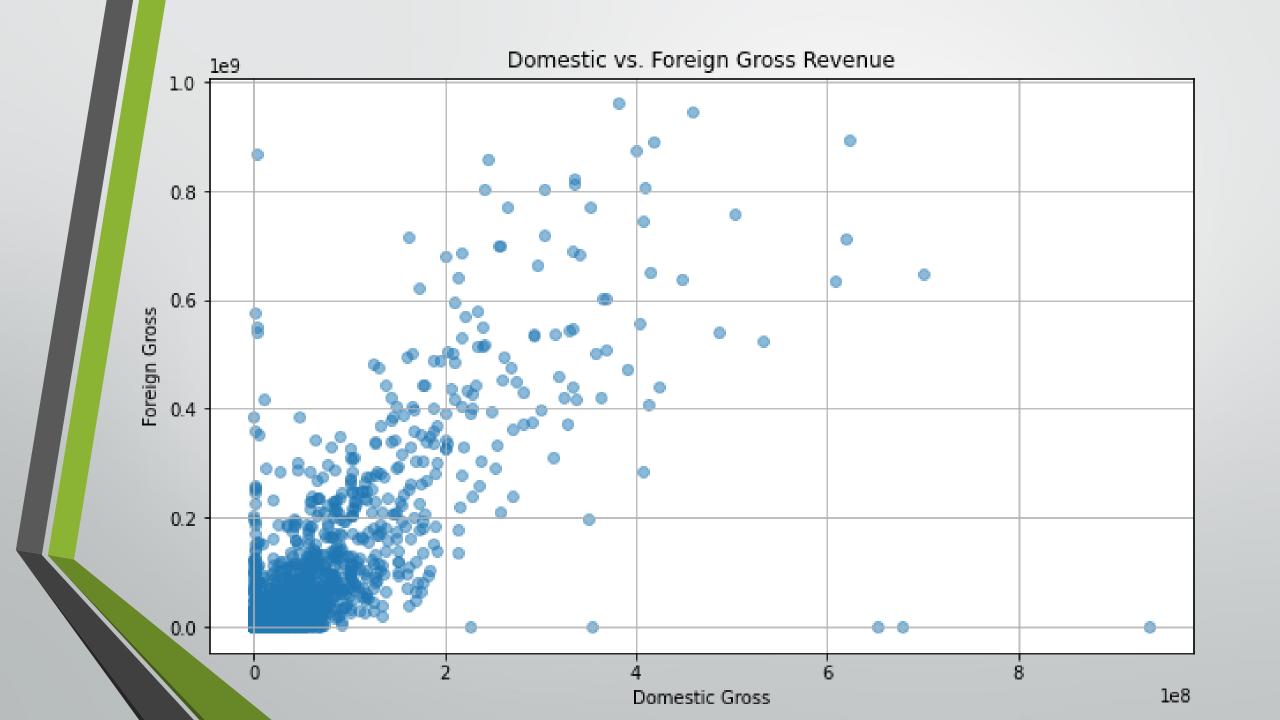
The high variability and presence of outliers suggest that the film industry is unpredictable. Microsoft should aim to produce a diverse portfolio of films, including both low-budget and high-budget projects, to balance risk and maximize potential returns.

Leverage Blockbusters:

Since outliers represent films with exceptionally high gross values, Microsoft should aim to create blockbuster films occasionally. These high-performing films can significantly boost the studio's overall revenue.

Conclusion:

By analyzing the distribution of domestic, foreign, and total gross values, Microsoft can make informed decisions about the types of films to produce and the markets to target.



Scatter Plot Interpretation

Clustering:

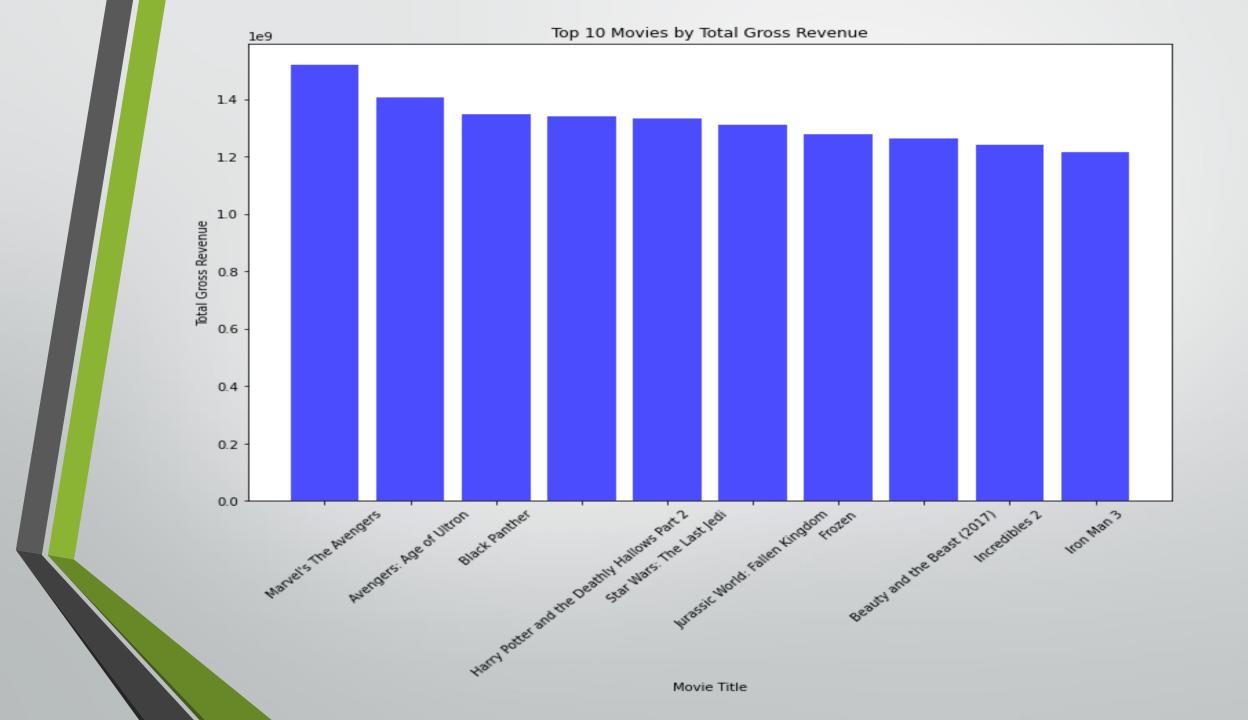
• A large number of movies have both domestic and foreign gross revenues below \$200 million.

High Gross Outliers:

• There are a few outliers with extremely high domestic and foreign gross revenues, suggesting blockbuster movies because they stand out as having much higher earnings.

Positive Correlation:

There's a general positive correlation between domestic and foreign gross revenues, indicating that movies that perform well domestically often also perform well internationally.



Analysis of the Bar Chart

The bar chart illustrates the top 10 genres or categories of movies (as indicated by the labels 0 to 9 on the x-axis) and their respective performance (measured on the y-axis).

Here are some observations:

- Consistent Performance The values are fairly consistent across the genres, with the first bar slightly higher than the others.
- Top-Performing Genre The genre represented by the first bar shows a slightly higher performance compared to the others.

Business Recommendations for Microsoft

Diversify Movie Portfolio:

Given the similar performance levels across the top genres, Microsoft should aim to produce a diverse portfolio of movies. This will mitigate risks and cater to a broader audience.

Focus on High-Performing Genres:

Identify the genre represented by the first bar, as it shows a slightly higher performance. Investing more in this genre might yield better returns.

Business Recommendations for Microsoft

Diversify Movie Portfolio:

- ✓ Given the similar performance levels across the top genres, Microsoft should aim to produce a diverse portfolio of movies. This will mitigate risks and cater to a broader audience. Consistent Quality Across Genres:
- Since the performance difference among the top genres is not significant, maintaining a high-quality production standard across all genres is crucial. This ensures that every movie has a good chance of performing well.
- ✓ Market Trends and Audience Preferences:
- Continuously monitor market trends and audience preferences. The data indicates consistency, which might change with new trends. Being agile in response to market demands will keep Microsoft competitive.

Focus on High-Performing Genres:

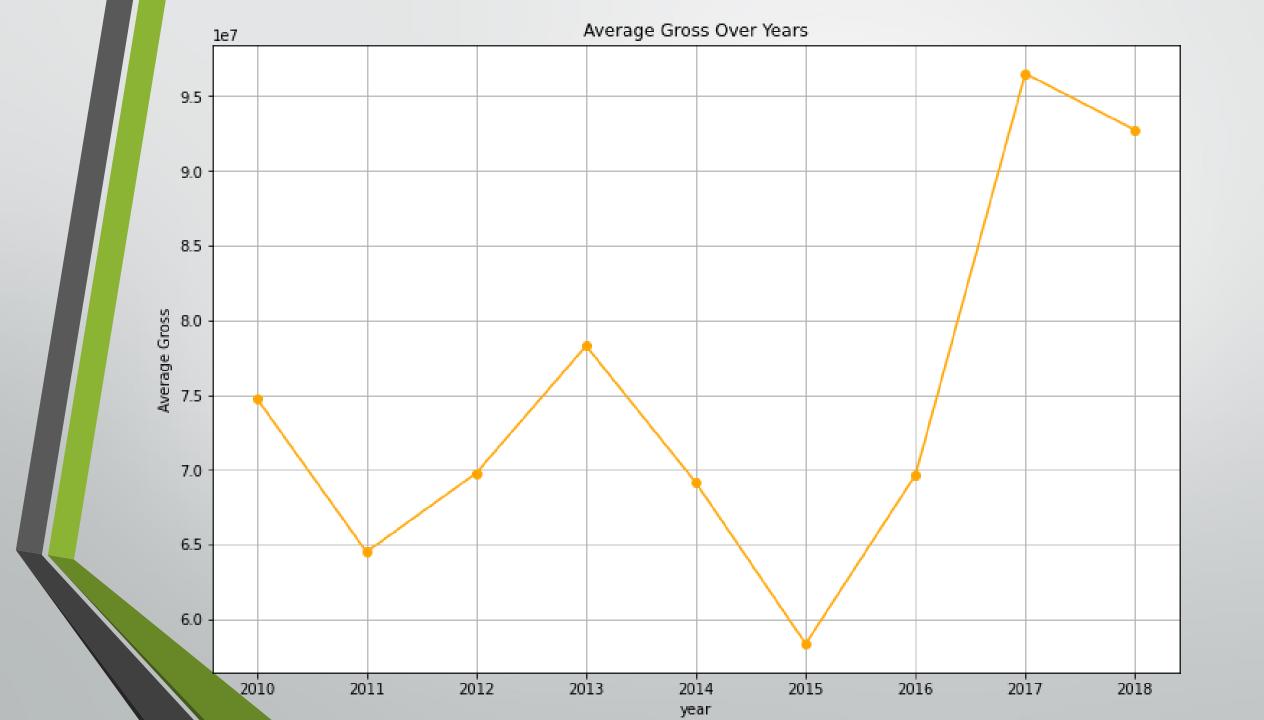
Identify the genre represented by the first bar, as it shows a slightly higher performance. Investing more in this genre might yield better returns.

Focus on High-Performing Genres:

✓ Identify the genre represented by the first bar, as it shows a slightly higher performance. Investing more in this genre might yield better returns.

Leverage Data-Driven Decisions:

✓ Utilize advanced analytics to dive deeper into the factors driving the success of these genres. This can include analyzing viewer demographics, preferences, and market conditions.



Focus on High-Performing Genres:

✓ Identify the genre represented by the first bar, as it shows a slightly higher performance. Investing more in this genre might yield better returns.

Leverage Data-Driven Decisions:

✓ Utilize advanced analytics to dive deeper into the factors driving the success of these genres. This can include analyzing viewer demographics, preferences, and market conditions.

Analysis of Average Gross Over Years

The line chart provided shows the trend of average gross revenue for movies from 2010 to 2018. Here are some observations and insights based on this chart:

Observations:

- Fluctuating Trend: The average gross revenue shows significant fluctuations over the years.
- Peaks and Troughs:
 - Peak in 2017: The year 2017 shows a significant peak in average gross revenue.
 - **Trough in 2015:** There is a noticeable dip in average gross revenue in 2015.
- Recovery and Decline:
 - **Post-2015 Recovery:** After the dip in 2015, there is a strong recovery, peaking in 2017.
 - Slight Decline in 2018: The average gross revenue slightly declines after the peak in 2017.

Insights and Recommendations

Identify High-Performing Years:

- Analyze 2017: Investigate the factors that contributed to the high average gross revenue in 2017. This could include successful genres, franchises, or marketing strategies.
- Leverage Success: Emulate the strategies from 2017 for future movie productions to maximize average gross revenue.

Understand Low-Performing Years:

- Analyze 2015: Examine why 2015 had a low average gross revenue. Factors could include economic downturns, fewer blockbuster releases, or increased competition.
- Risk Mitigation: Develop strategies to mitigate risks identified in low-performing years to avoid similar pitfalls in the future.

Consistent Quality and Marketing:

- OHigh-Quality Productions: Ensure consistent investment in high-quality productions to maintain or increase average gross revenue.
- OEffective Marketing: Implement effective marketing campaigns, especially during years where competition is high or audience interest is low.

Strategic Release Timing:

- Optimize Release Schedules: Based on the trend, optimize movie release schedules to coincide with periods of high audience interest and low competition.
- Seasonal Analysis: Conduct a seasonal analysis to determine the best times of the year to release movies for maximum average gross revenue.

Next Steps

Deep Dive Analysis:

Perform a detailed analysis of movies released in peak and trough years to identify common factors contributing to their performance.

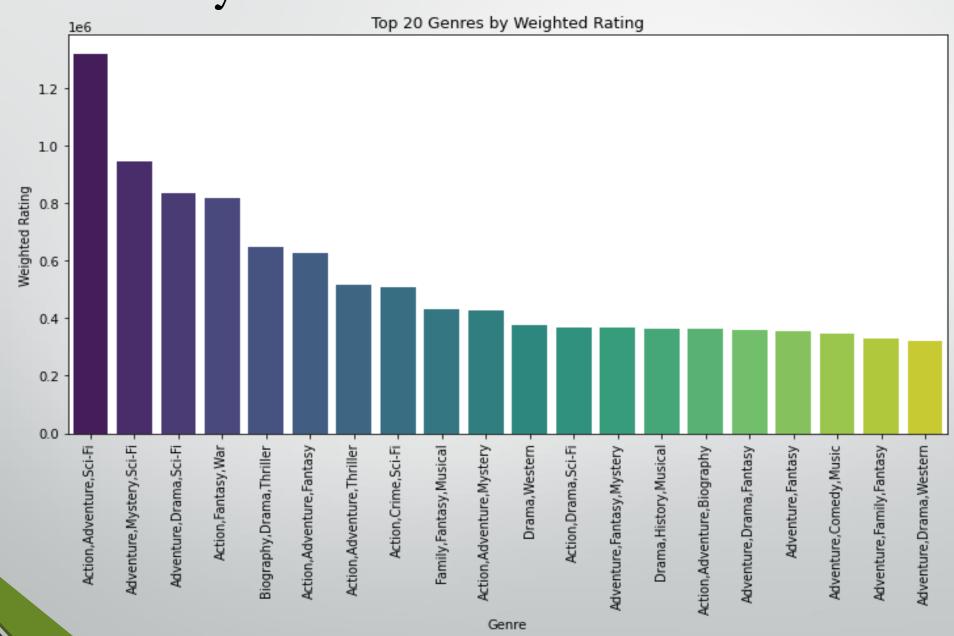
Market Research:

Conduct market research to understand changing audience preferences and emerging trends.

Strategy Formulation:

Develop a comprehensive strategy for Microsoft's new movie studio based on the insights gathered from data analysis.

Data Analysis- im.db database file



		Genres
	Primary Title	Action, Adventure, Biography
	The Revenant	
	Star Wars: Episode VII - The Force Awakens	Action, Adventure, Fantasy
	Divergent	Action, Adventure, Mystery
	Inception	Action, Adventure, Sci-Fi
	Skyfall	Action, Adventure, Thriller
	Dredd	
	Logan	Action, Crime, Sci-Fi
	300: Rise of an Empire	Action, Drama, Sci-Fi
	Get Him to the Greek	Action, Fantasy, War
	Harry Potter and the Deathly Hallows: Part 2	Adventure, Comedy, Music
		Adventure, Drama, Fantasy
	True Grit	Adventure D
	The Hobbit: An Unexpected Journey The Hobbit: The D	Adventure, Drama, Sci-Fi
	THE DATALLY	Adventure, Drama, Western
	Harry Potter and the Deathly Hallows: Part 1	Adventure, Family, Fantasy
p Movies Within Top Genres	Prometheus	Adventure, Fantasy
rom 2010 - 2018	The Imitation Game	Adventure
	Les Misérables	Adventure, Fantasy, Mystery Adventure, Montal
	Django Unchained	Adventure, Mystery, Sci-Fi
	Beauty and the Beast	
		Biography, Drama, Thriller Drama, History, Musical Drama, Wasse
		Drama, Western
		Family, Fantasy M
		Fantasy Na

Analysis of the Bar Plot

The bar plot above shows the top 20 genres by weighted rating. Here are the key observations and interpretations:

Observations:

Top Genres:

- **Dominant Genre**: The genre with the highest weighted rating is significantly higher than the rest. This indicates that this genre not only has high average ratings but also attracts a large number of votes.
- Noticeable Drop: There is a noticeable drop in weighted rating after the top genre, with subsequent genres having relatively closer weighted ratings to each other.

Distribution:

Gradual Decline: The distribution of weighted ratings among the top 20 genres shows a gradual decline, indicating a broad distribution of highly-rated and popular genres.

Insights:

Dominant Genre:

Investigation Needed: The genre at the top stands out significantly. It would be beneficial to investigate what specific characteristics, themes, or popular movies within this genre contribute to its high weighted rating.

Competitive Genres:

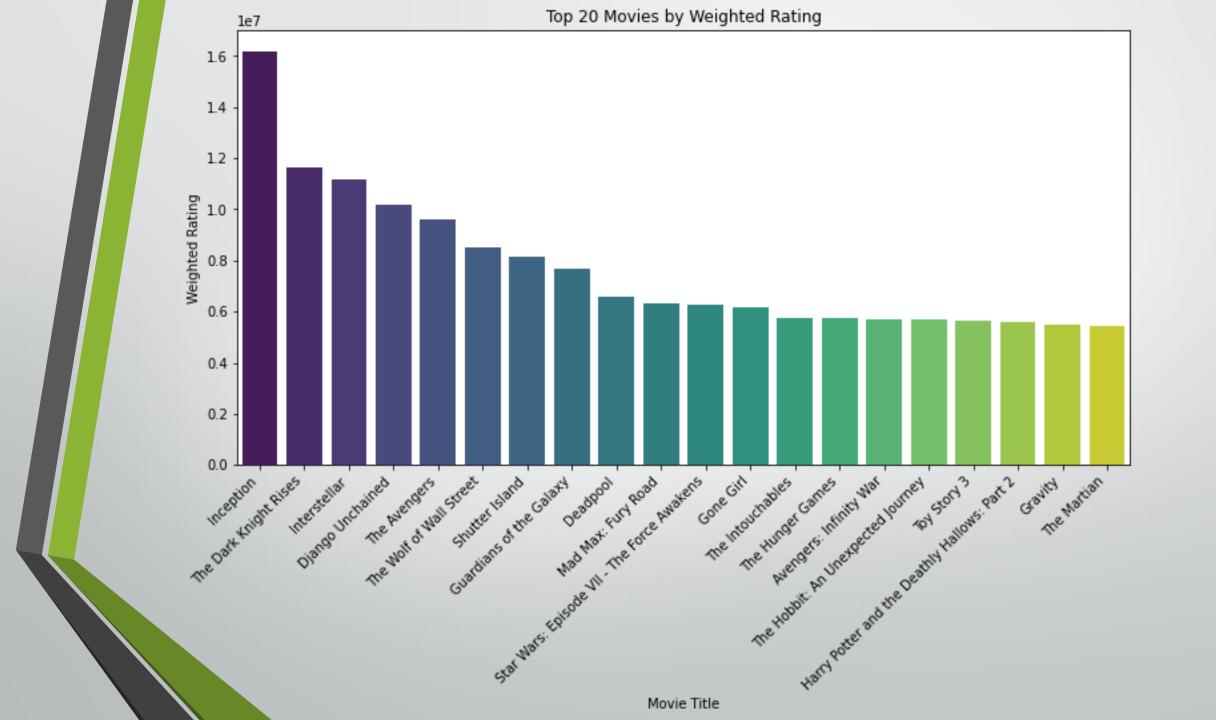
Close Performance: The close values in weighted ratings for the subsequent genres suggest competitive performance. These genres are also well-received by audiences and have a substantial number of votes.

Detailed Genre Analysis:

Characteristics: Investigate the characteristics of the top genres to understand why they perform well. This can include looking at the types of movies within these genres, their common themes, and their target audiences.

Top Movies Within Each Genre:

High Performers: Identify the top movies within each of these genres to see if specific movies are driving the high weighted ratings.



Analysis of the Bar Plot

The bar plot above shows the top 20 genres by weighted rating. This plot reaffirms the observations and insights previously noted, but let's delve into some specifics to provide a clearer picture.

Detailed Observations:

Dominant Genre:

✓ **Higher Weighted Rating:** The leading genre has a notably higher weighted rating compared to the rest, indicating it combines both high ratings and high popularity (number of votes) more effectively than other genres.

Gradual Decline:

Competitive Performance: After the top genre, the weighted ratings of subsequent genres decline gradually, suggesting competitive performance among them but with lesser dominance than the leading genre.

Clustered Performance:

Similar Performance: The lower-ranked genres within the top 20 have more clustered values, indicating similar performance in terms of weighted ratings.

High Performance of the Top Genre:

Influential Movies: The top genre's significantly higher weighted rating may be due to a few highly influential and popular movies within that genre. It would be beneficial to investigate the top movies within this genre to identify the drivers of this high rating.

Consistency Among Other Top Genres:

Steady Stream: The close values among the other top genres suggest consistency in audience reception. These genres likely have a steady stream of well-received movies that contribute to their weighted ratings.

Top Movies Within the Top Genre:

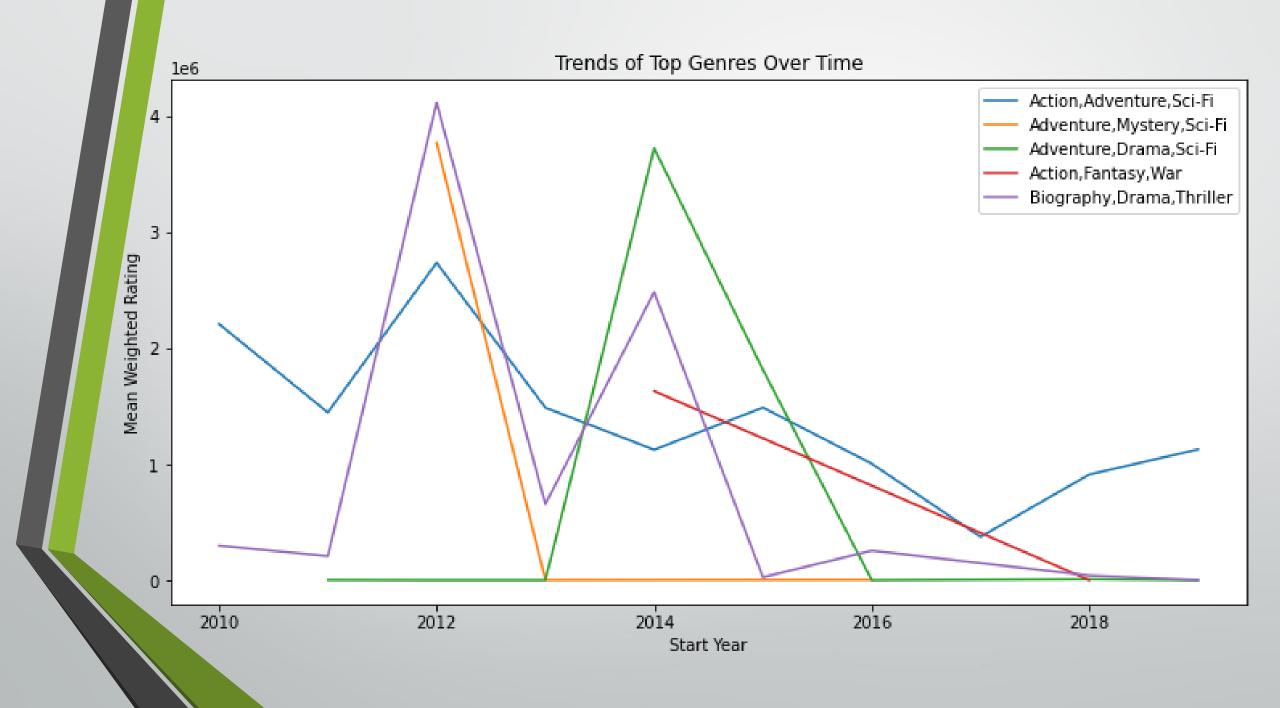
In-Depth Analysis: Investigate the top movies within the highest-rated genre to understand what makes them stand out. This can include examining elements like themes, storytelling, casting, and production quality.

Emerging Trends:

Trend Analysis: Analyze how the popularity and ratings of these genres have changed over time to identify any emerging trends or shifts in audience preferences.

Top Movies Within Top Genres

Primary Title	Genres		
The Revenant	Action, Adventure, Biography		
Star Wars: Episode VII - The Force Awakens	Action, Adventure, Fantasy		
Divergent	Action, Adventure, Mystery		
Inception	Action, Adventure, Sci-Fi		
Skyfall	Action, Adventure, Thriller		
Dredd	Action, Crime, Sci-Fi		
Logan	Action, Drama, Sci-Fi		
300: Rise of an Empire	Action, Fantasy, War		
Get Him to the Greek	Adventure, Comedy, Music		
Harry Potter and the Deathly Hallows: Part 2	Adventure, Drama, Fantasy		
Interstellar	Adventure, Drama, Sci-Fi		
True Grit	Adventure, Drama, Western		
The Hobbit: An Unexpected Journey	Adventure, Family, Fantasy		
The Hobbit: The Desolation of Smaug	Adventure, Fantasy		



Analysis of Line Plot: Trends in Weighted Ratings Over Time

The line plot shows trends in weighted ratings over time for the top 5 genres. Here are the key observations and insights:

Observations:

Fluctuating Trends:

Variable Reception: The weighted ratings for these genres fluctuate significantly over time. Some genres have peaks and troughs, indicating varying audience reception across different years.

Distinct Peaks:

Notable Peaks: There are distinct peaks for certain genres, such as Adventure, Drama, Sci-Fi, and Action, Fantasy, War, indicating specific years where these genres had standout movies that significantly boosted their weighted ratings.

Declining Trends:

Decreasing Popularity: Some genres, such as Adventure, Mystery, Sci-Fi, show a declining trend, suggesting a decrease in both ratings and popularity over time.

Stable Trends:

Consistent Performance: Other genres, like Action, Adventure, Sci-Fi, show a more stable trend, indicating consistent performance over the years.

Impact of Blockbuster Movies:

Boost from Blockbusters: The peaks in the line plot likely correspond to blockbuster movies within these genres that received high ratings and a large number of votes, significantly boosting the weighted rating for those years.

Changing Audience Preferences:

Evolving Tastes: The declining trends for certain genres may indicate changing audience preferences, with some genres becoming less popular over time.

Consistent Genres:

Reliable Genres: Genres with more stable trends are consistently well-received by audiences, suggesting a reliable performance in both quality and popularity.

Identify Key Movies:

Driver Analysis: Identify the key movies responsible for the peaks in the weighted ratings for each genre. This can help understand what drives audience engagement and high ratings.

Yearly Analysis:

Detailed Examination: Conduct a more detailed yearly analysis to see what specific factors (e.g., significant events, marketing strategies) contributed to the fluctuations in weighted ratings.

Business Recommendations:

Focus on Stable Genres:

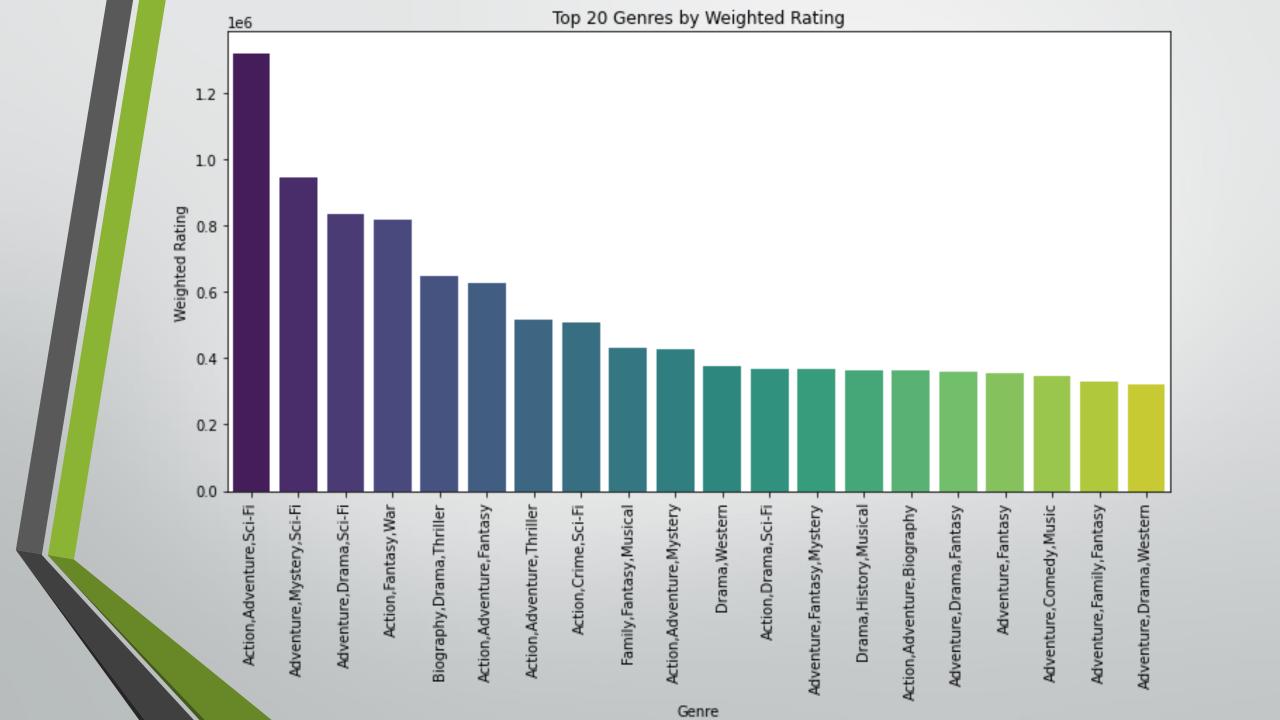
✓ Consistent Production: Invest in producing movies in genres with stable trends, as they have shown consistent audience approval.

Leverage Blockbuster Success:

✓ Capitalize on Hits: Emulate the successful elements of blockbuster movies to boost future productions within peak genres.

Adapt to Audience Shifts:

✓ **Trend Monitoring:** Monitor and adapt to changing audience preferences to stay ahead of declining trends and capitalize on emerging interests.



Analysis of Bar Plot: Top 20 Genres by Weighted Rating

The bar plot above shows the top 20 genres by weighted rating. Here are the key observations and interpretations:

Observations:

Top Genres:

- Highest Weighted Rating: The genre with the highest weighted rating is significantly higher than the rest. This indicates that this genre not only has high average ratings but also attracts a large number of votes.
- Rating Drop: There is a noticeable drop in weighted rating after the top genre, with subsequent genres having relatively closer weighted ratings to each other.

Distribution:

✓ **Gradual Decline:** The distribution of weighted ratings among the top 20 genres shows a gradual decline, indicating a broad distribution of highly-rated and popular genres.

Dominant Genre:

Significant Standout: The genre at the top stands out significantly. It would be beneficial to investigate what specific characteristics, themes, or popular movies within this genre contribute to its high weighted rating.

Competitive Genres:

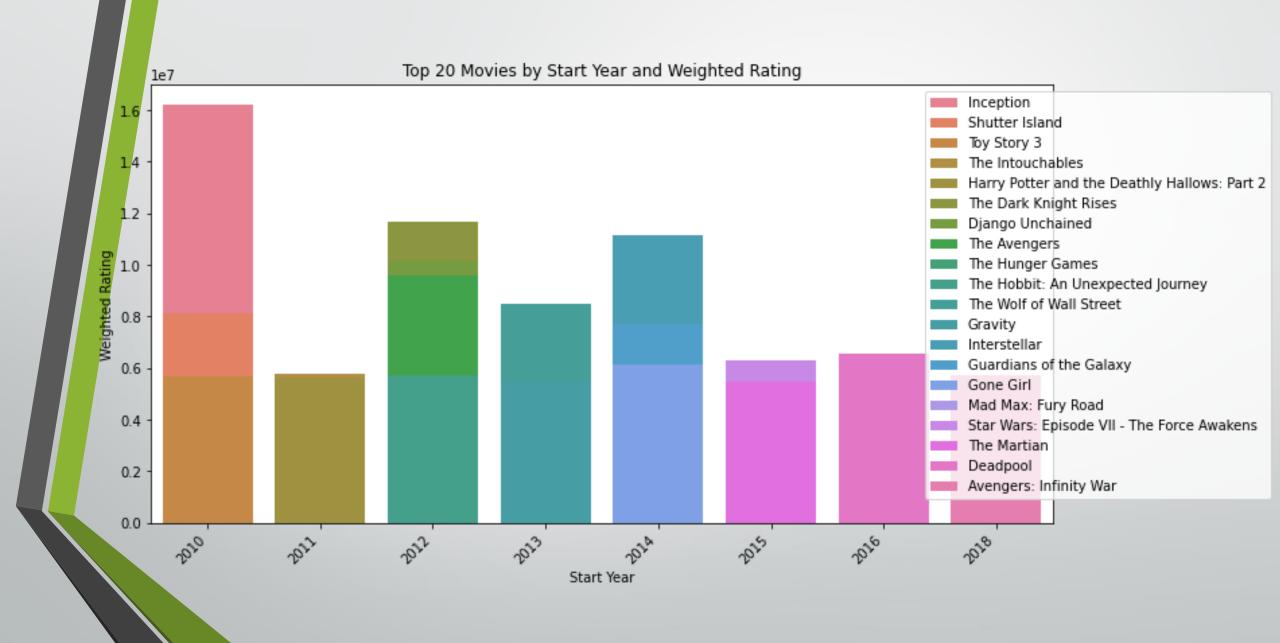
Close Performance: The close values in weighted ratings for the subsequent genres suggest competitive performance. These genres are also well-received by audiences and have a substantial number of votes.

Detailed Genre Analysis:

Characteristics Investigation: Investigate the characteristics of the top genres to understand why they perform well. This can include looking at the types of movies within these genres, their common themes, and their target audiences.

Top Movies Within Each Genre:

Identify Key Movies: Identify the top movies within each of these genres to see if specific movies are driving the high weighted ratings.



Business Recommendations:

Invest in Dominant Genre:

Focus on High Performers: Allocate resources to produce more movies in the dominant genre, leveraging its high popularity and ratings.

Maintain Quality Across Genres:

Consistent Production: Ensure high production standards across all genres to sustain competitive performance and audience satisfaction.

Explore Competitive Genres:

Diverse Portfolio: Invest in a diverse portfolio of genres to cater to various audience preferences and mitigate risks.

Analysis of Bar Plot: Top Movies Within Each Genre by Weighted Rating

The bar plot you provided shows the top movies within each genre by weighted rating. Each bar represents a different movie, and the colors indicate the movie titles. Here's a detailed analysis of the plot:

Observations:

▼ Top Movies Identified: Clear Identification: The plot clearly identifies the top movies within each genre by their weighted ratings.

Genre Performance:

✓ **Multiple Top-Rated Movies:** Some genres have multiple top-rated movies, indicating a strong overall performance within that genre.

Distribution of Weighted Ratings:

Impact Visualization: The height of the bars indicates the weighted ratings, allowing us to see which movies have the highest impact on their respective genres.

Influence of Blockbuster Movies:

Significant Impact: Movies like "Inception," "The Dark Knight Rises," "Avengers: Infinity War," and "Harry Potter and the Deathly Hallows: Part 2" have significant weighted ratings, highlighting their influence on their genres.

Popular Genres:

Strong Performers: Genres represented by multiple high-weighted movies, such as action, adventure, and sci-fi, are particularly strong performers in terms of both quality and popularity.

Variety in Top Movies:

Diverse Preferences: The variety of movies across different genres shows a diverse set of audience preferences and high-quality productions.

Detailed Breakdown of Top Movies:

Characteristic Analysis: Examine the characteristics of the top movies to understand what makes them stand out in terms of ratings and popularity.

Time Trends for Top Movies:

Trend Analysis: Analyze the release years of these top movies to see if there are any trends or patterns in their release and reception over time.

Business Recommendations:

Capitalize on Blockbusters:

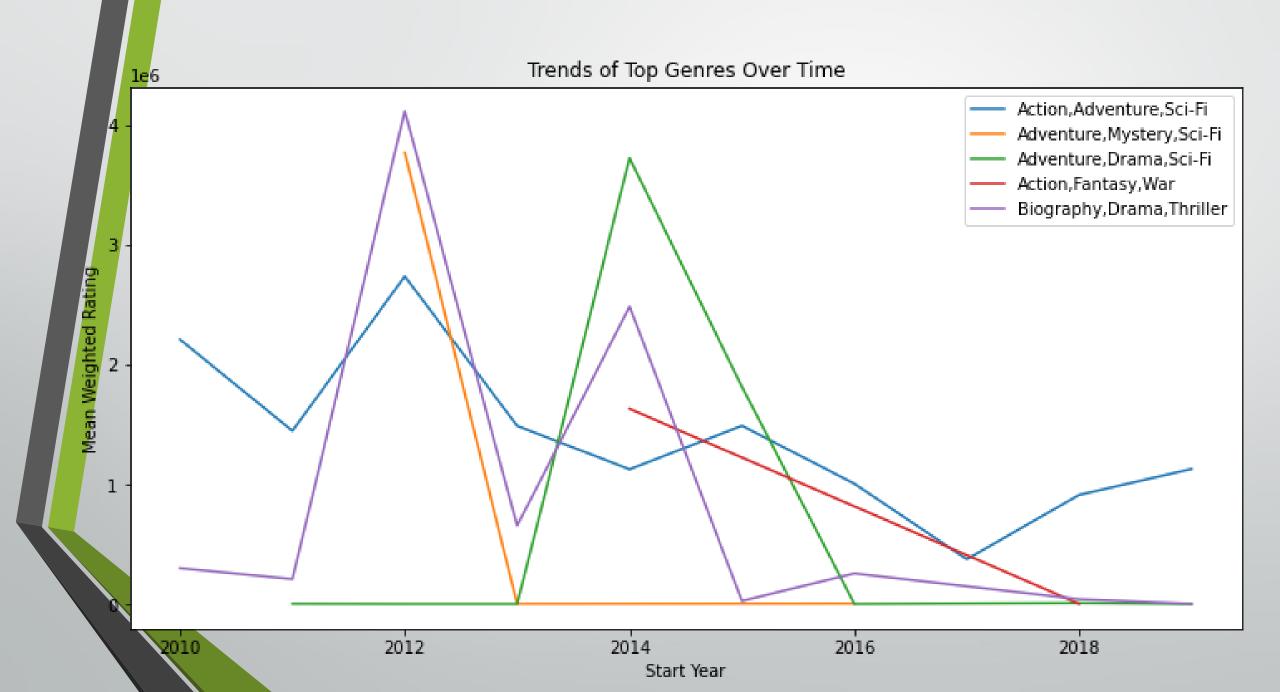
Focus on High Impact: Invest in producing movies similar to blockbuster hits to leverage their proven success in driving high ratings and popularity.

Strengthen Popular Genres:

*Invest in Strong Genres: Allocate more resources to genres with multiple high-weighted movies, such as action, adventure, and sci-fi, to maximize audience engagement.

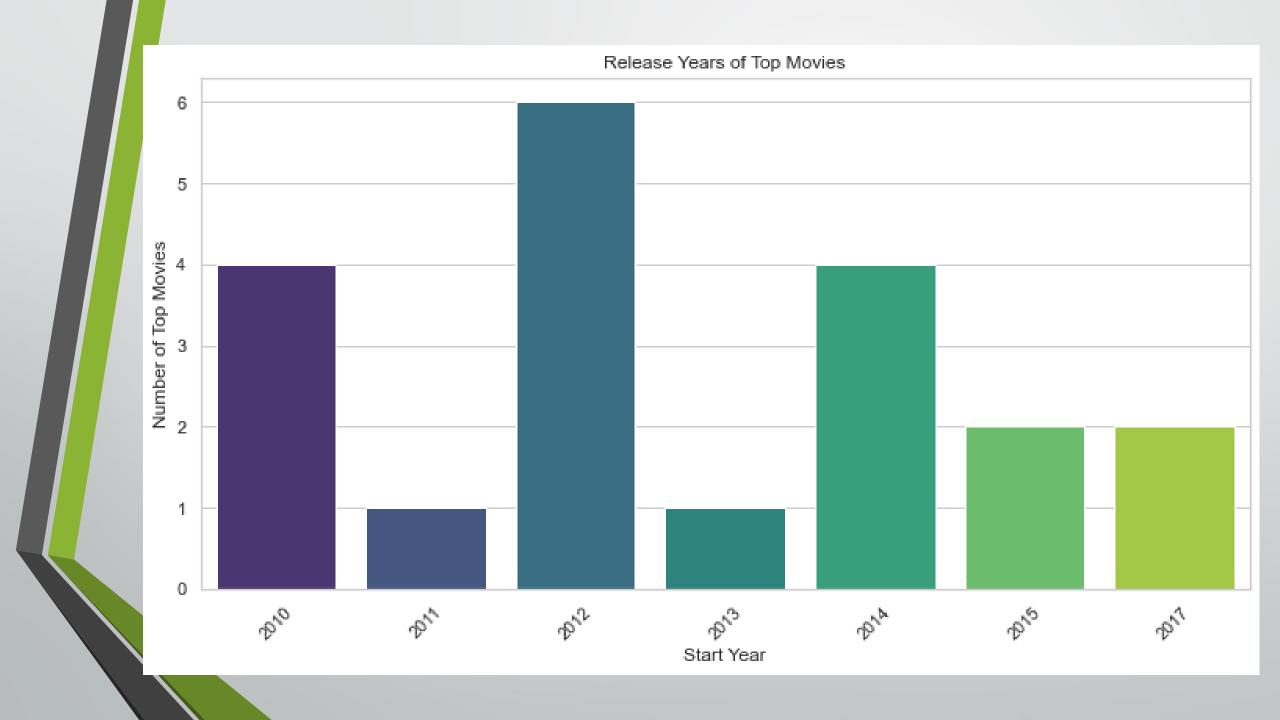
Maintain Diverse Offerings:

Cater to Varied Preferences: Continue producing a diverse array of movies across different genres to cater to broad audience preferences and ensure high-quality productions.



Display detailed information for the top movies

			Weighted	Average	Number of
Primary Title	Genres	Year	Rating	Rating	Votes
The Revenant	Action, Adventure, Biography	2015	\$4,969,544	8.0	621,193
Star Wars: Episode VII - The Force Awakens	Action, Adventure, Fantasy	2015	\$6,278,240	8.0	784,780
Divergent	Action, Adventure, Mystery	2014	\$2,678,010.10	6.7	399,703
Inception	Action, Adventure, Sci-Fi	2010	\$16,201,380.80	8.8	1,841,066
Skyfall	Action, Adventure, Thriller	2012	\$4,619,323.80	7.8	592,221
Dredd	Action, Crime, Sci- Fi	2012	\$1,665,773.60	7.1	234,616
Logan	Action, Drama, Sci- Fi	2017	\$4,538,187.00	8.1	560,270
300: Rise of an Empire	Action, Fantasy, War	2014	\$1,630,463.60	6.2	262,978
Get Him to the Greek	Adventure, Comedy, Music	2010	\$1,034,579.20	6.4	161,653



Analysis of Top Movie Releases by Year

Observations:

- Peak Year: There is a clear peak in one year where the highest number of top movies were released, indicating a particularly strong year for high-weighted movies.
- Other Significant Years: Multiple years show a significant number of top movies, suggesting strong periods for the release of high-rated and popular films.
- Lower Activity Years: Some years have fewer top movies, indicating variability in the number of high-weighted movies released annually.

- Successful Year: The peak year likely saw the release of several blockbuster movies that contributed to its high count of top movies. Investigating the specific movies and factors contributing to this peak could provide insights into successful strategies for movie releases.
- ✓ Consistent High Performance: Years with consistently high numbers of top movies indicate periods where the industry produced several successful films, potentially driven by strong content, effective marketing, or favorable market conditions.
- ✓ **Variability:** The variability in the number of top movies across years highlights the fluctuating nature of the film industry, with some years being more successful than others.

- Identify Key Movies in Peak Year: Identify the key movies responsible for the peak in the number of top movies for the peak year.
- Factors Contributing to Success: Analyze factors such as genre distribution, marketing efforts, and industry trends that contributed to the success of the peak year.

Conclusion:

- a) Focus on high-performing genres.
- b) Release movies during peak times.
- C) Balance quality and profitability.
- (1) Continuously monitor performance.