

Lights Camera, Data: Analyzing Movie Trends

Group 4 Project



INTRODUCTION.

Problem statement

- Data serves as a foundation for identifying factors that contribute to a movie's success in terms of box office performance and audience reception.
- Analyzing metrics such as genre, popularity, ratings, and release timing, we aim to uncover actionable insights that align with our business goal of creating commercially successful and audience-appealing films.
- This understanding will guide the strategic decisions for our new movie studio, ensuring data-driven approaches to film production.

INTRODUCTION.

Objectives

- To identify high-performing genres with the best ratings to optimize budgets for maximum ROI.
- To align movie production with audience preferences by focusing on popular genres and attributes.
- To plan releases during peak seasons for optimal audience reach and profitability.
- To leverage movie length and quality metrics to determine ideal formats for various genres.

Data Overview

IMDb database

Key Features:

- Movie details: Titles, cast, crew, genres, runtime.
- Ratings: Audience and critic scores (e.g., IMDb Score).
- Revenue & Budget: Estimated earnings and production costs.
- Popularity: Reviews, votes, and trends.

TMDb dataset

Key Features:

- Titles, release dates, genres, runtimes.
- Production companies.
- Popularity: User engagement and trends.
- Comprehensive genres and thematic tags.

Movie info

Key Features:

- Movie title, release date, genre.
- Director and producers.
- Language and country of origin.
- Runtime and ratings.

Data Acquisition Procedure

- **Data Sources**

- SQLite Database: Extracted data from im.db, using movie_basics and movie_ratings tables.
- CSV File: Supplemented data with tmdb.movies.csv for additional information.

- **SQL Query**

- Performed an inner join on movie_basics and movie_ratings using id.
- Extracted fields:
 - id, title, original_title, start_year, runtime_minutes, genres, averagerating, numvotes.

- **Data Enrichment**

- Conducted a left join with tmdb.movies.csv on id to add:
- release_date.

- **Output Dataset**

- Final columns:
- id, title, original_title, start_year, runtime_minutes, genres, averagerating, numvotes, release_date.

- **Data Validation**

- Verified column mappings during joins.
- Checked for missing values and inconsistencies.

Data Analysis

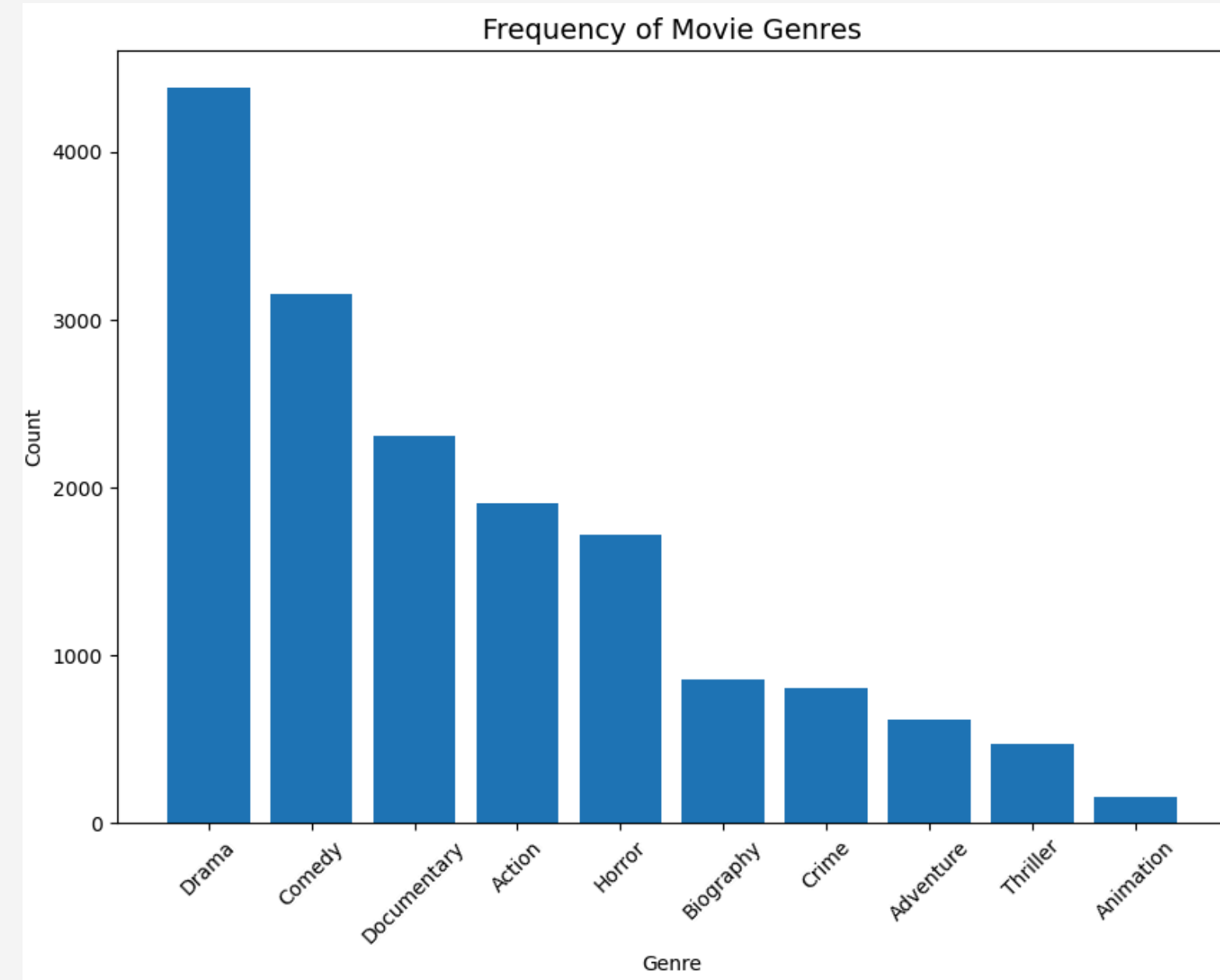
Fact Finding



Genre analysis

- **Popular Genres**

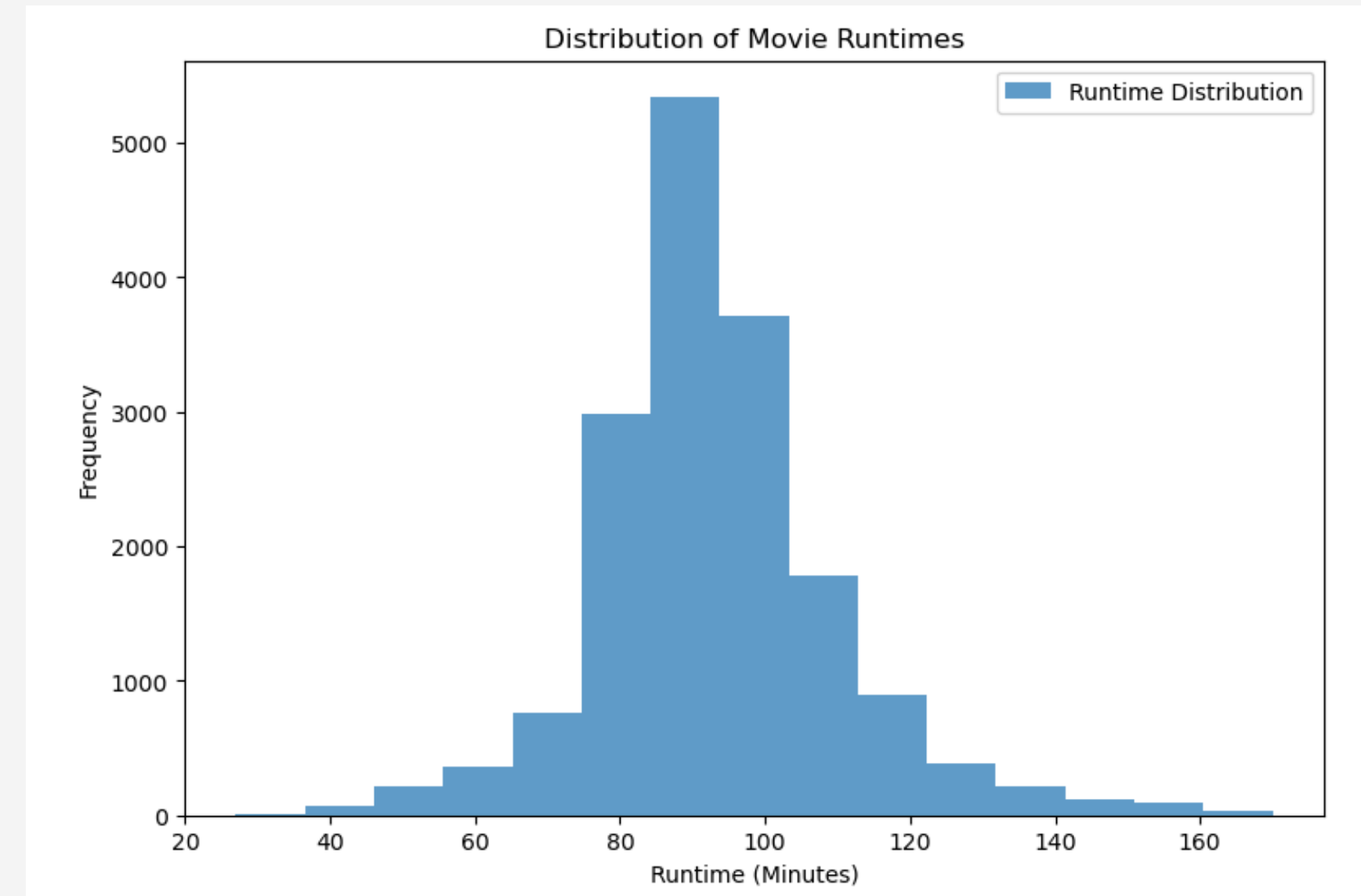
- Drama is the most frequent genre, significantly higher than all others.
- Documentary and Comedy & Action follow, with fewer movies than Drama.
- Horror , Biography, Crime and Adventure also appear among the top 10 genres.
- Animation and Thriller are the least common genres.



Runtime Analysis

- **Runtime Distribution**

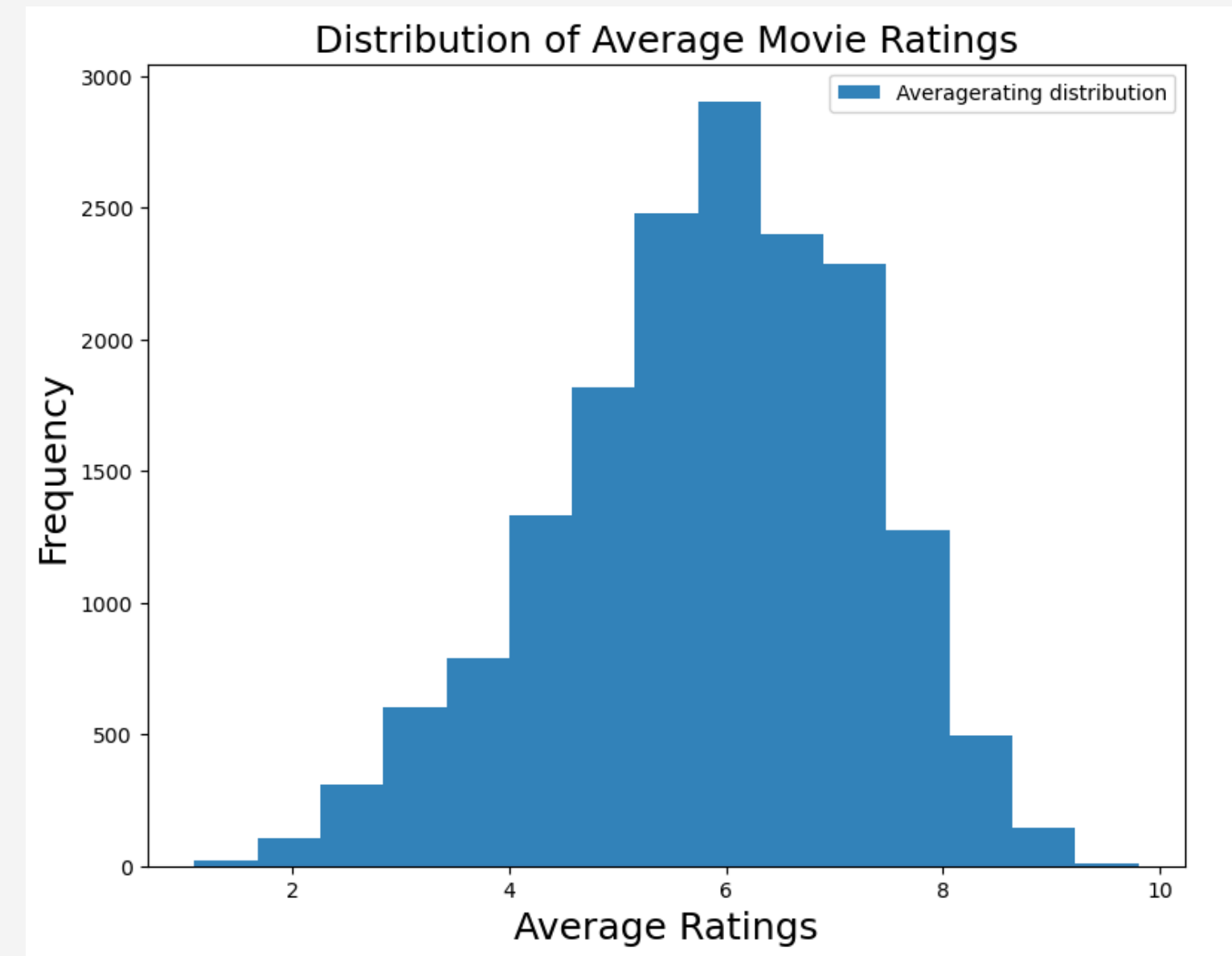
- Runtime is Clustered around 90 minutes, with a secondary cluster near 100 minutes.
- Majority of movies fall within the 80 to 110-minute range.
- The distribution is right skewed with a longer tail extending to longer runtimes.
- Movies with very long runtimes are relatively less frequent.



Movie Ratings Distribution

- **Runtime Distribution**

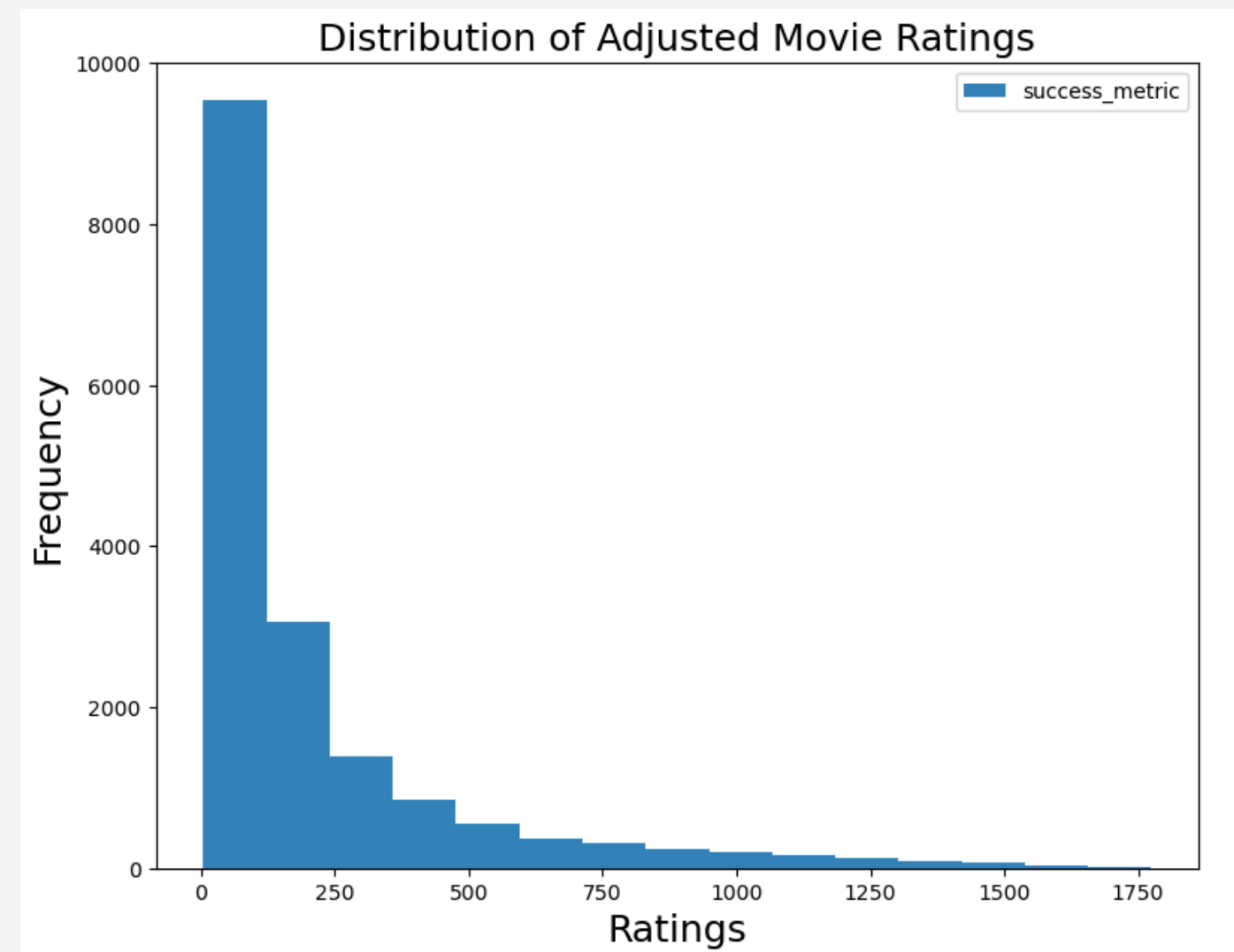
- Peak Frequency: The highest frequency of movies is within the 6-7 rating range.
- Typical Range: Most movies have average ratings between 5 and 8.
- The distribution is somewhat bell-shaped but slightly skewed to the left with a longer tail toward lower ratings.
- Low Frequency Extremes: There are relatively few movies with average ratings below 2 and above 9..



Adjusted Movie Ratings

- **Measuring Quality and Reach**

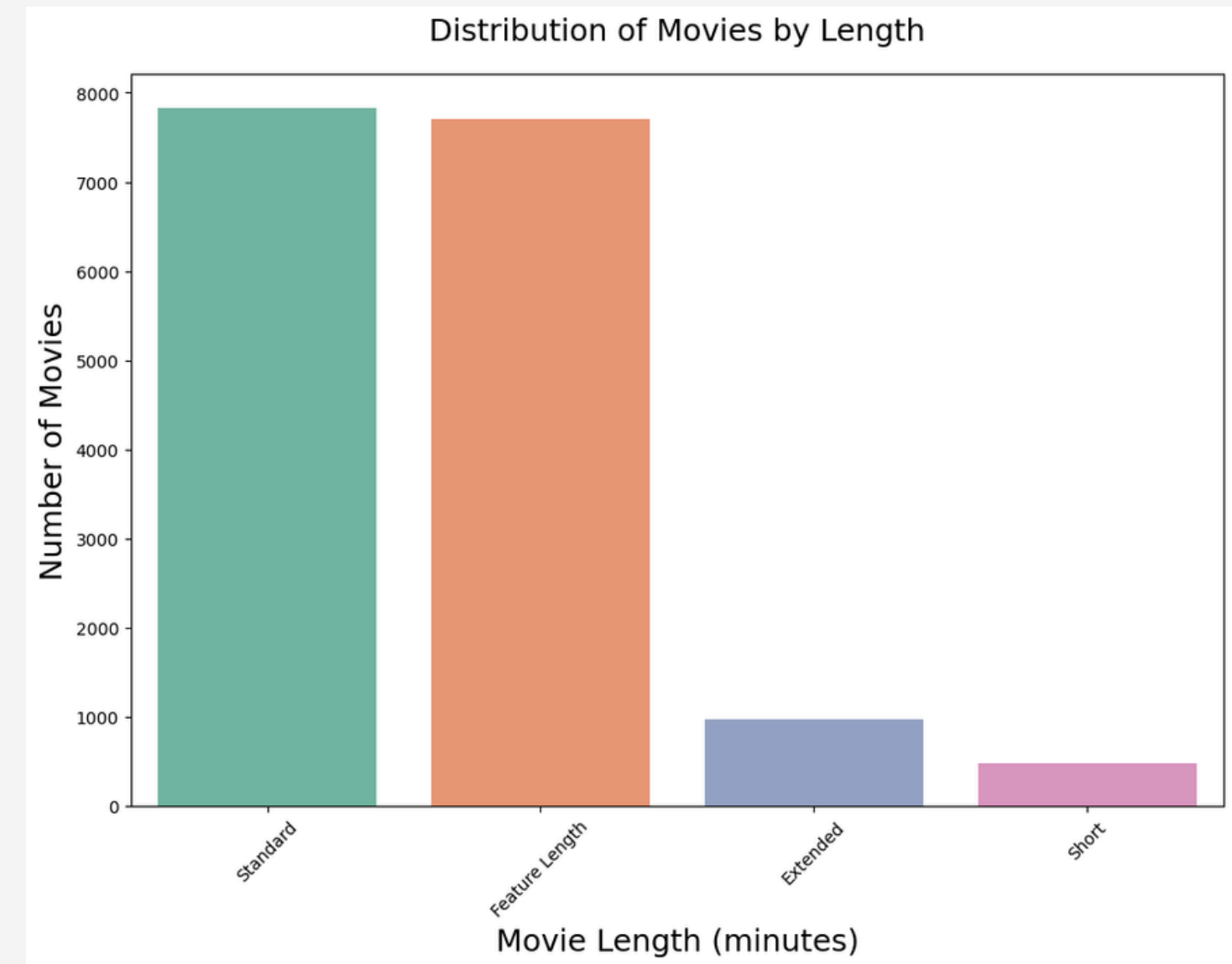
- The histogram plot peaks at the lower end, indicating most movies have low success metrics (low ratings, fewer votes, or both).
- The distribution is right-skewed, with a small number of movies having exceptionally high success metrics
- The highest density is within the 0-250 range, meaning most movies fall in this category with moderate ratings or limited popularity.
- A long tail showing a few movies achieve extremely high success, often blockbusters with high ratings
- There are no distinct peaks beyond the concentration of lower-success movies, indicating no clear tiers of success in the data.



Distribution of Movie Lengths

- **Classifying Movie Lengths**

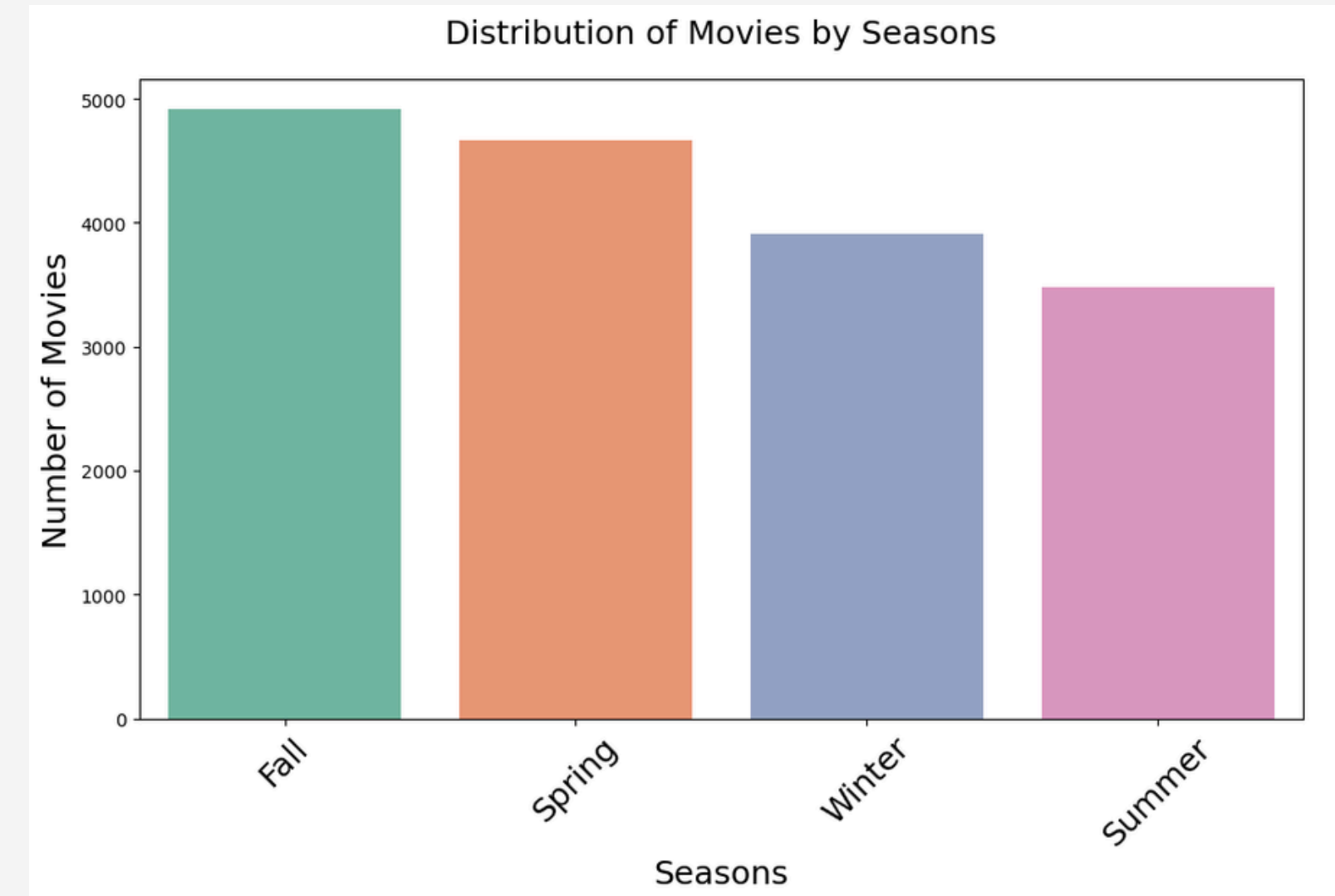
- **Dominant Lengths:** The "Standard" and "Feature Length" categories have the highest number of movies.
- **Similar Frequency:** The counts for "Standard" and "Feature Length" movies are relatively similar.
- **Low Extended:** The "Extended" category has lower number of movies compared to "Standard" and "Feature Length."
- **Lowest Short:** The "Short" category has the lowest count of movies.



Movie Releases by Season

- **Classifying Movie Lengths**

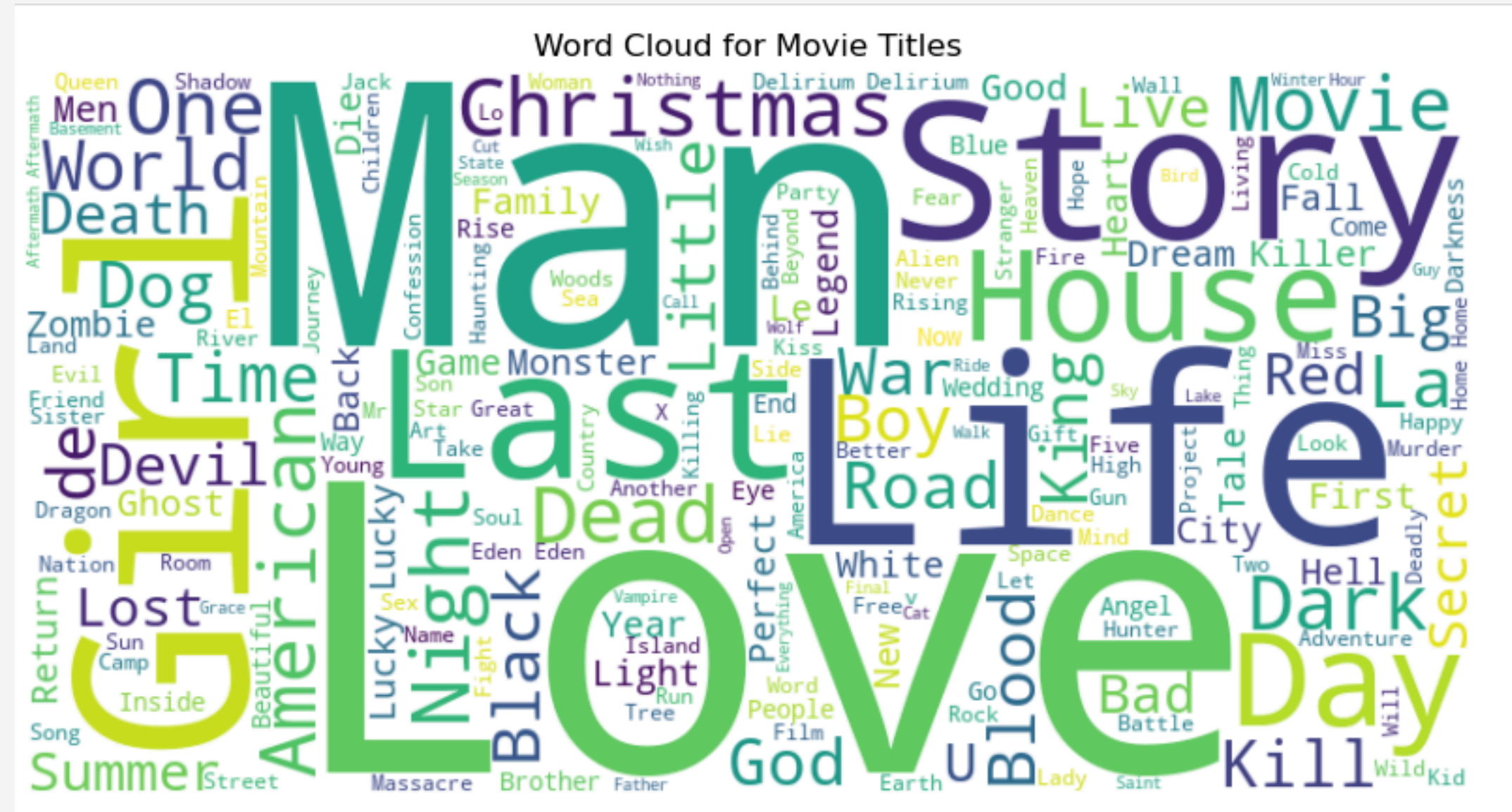
- **Fall Dominance:** "Fall" has the highest count of movie releases.
- **Spring Second:** "Spring" has the second highest amount of movie releases.
- **Lower Winter and Summer:** "Winter" and "Summer" have notably lower counts compared to "Fall" and "Spring."
- **Summer Least:** "Summer" has the lowest count of movie releases..



Movie Title Themes

- **Title Themes**

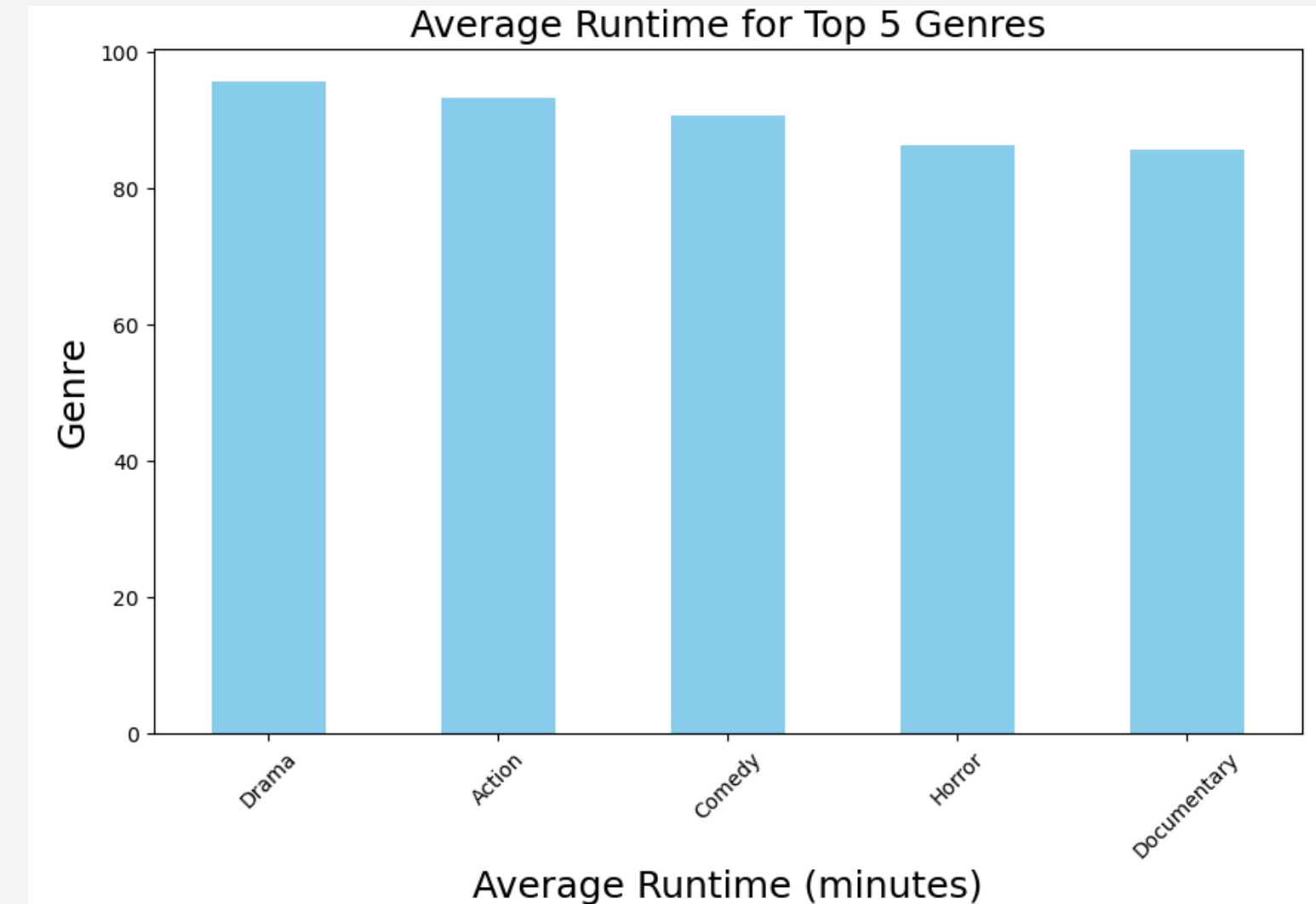
- **Central Themes:** "Man," "Life," "Girl", and "Love"—suggest recurring central themes.
 - **Common Title Elements:** Words like "Story", "Last," and "Day" are also prominent and used in movie titles.
 - **Action & Conflict:** Words like "Kill," "Dead," and "Blood" indicate action or conflict.
 - **Genre Indicators:** Some words appear that can hint to the genres of movies, such as "Devil," "Zombie", "Ghost"
 - **Common Nouns:** Words like "House", "World," "Night", "Time," and "Family", indicate locations or concepts that are widely used in movie titles.
- Word Cloud for Movie Titles



Average Movie Runtime by Genre

- **Genre Average Runtimes**

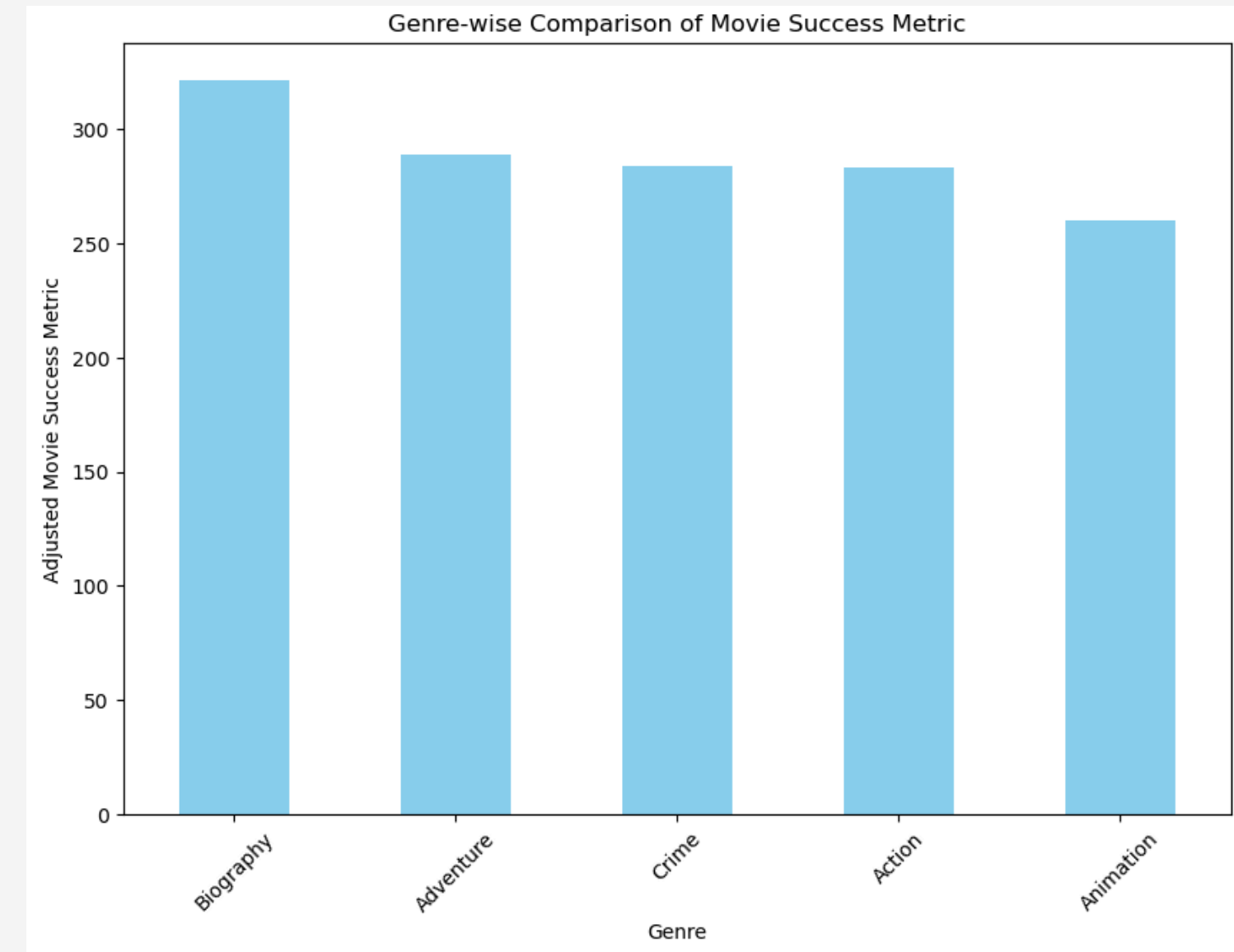
- Drama Highest: The "Drama" genre has the highest average runtime.
- Action Second: The "Action" genre has the second-highest average runtime.
- Comedy Middle: The "Comedy" genre has an average runtime in the middle.
- Horror Lower: The "Horror" genre has a lower average runtime.
- Runtime Range: Average runtimes are within a close range, from around 86 to 96 minutes.



Average Movie Runtime by Genre

- **Genre Average Runtimes**

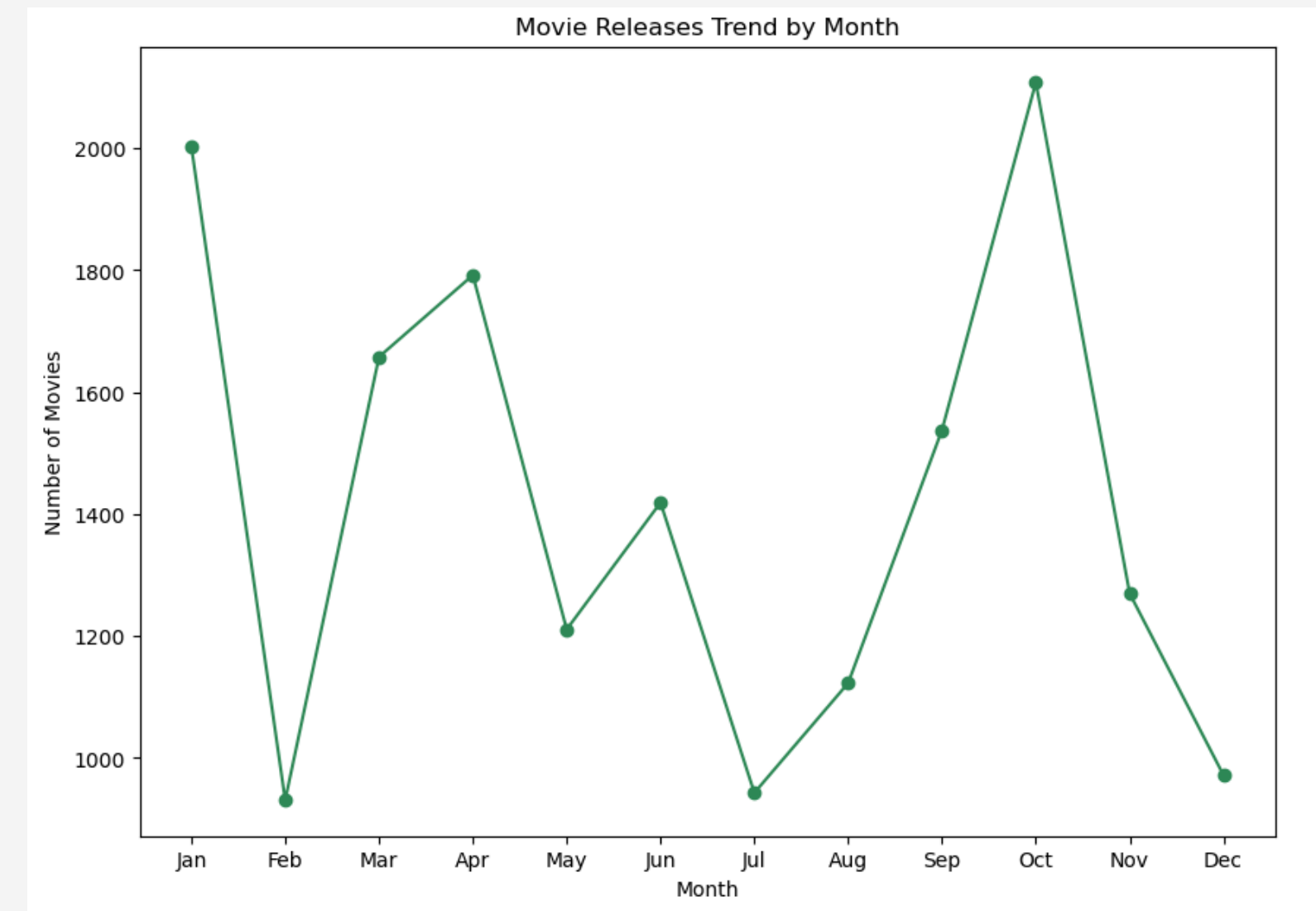
- **Biography:** Movies in this genre have the highest success metric, performing well in ratings and audience engagement.
- **Adventure, Crime, Action:** These show similar success metrics, suggesting they are popular but not as dominant as Biography.
- **Animation:** Among the top genres, Animation has the lowest success metric but still ranks well compared to others.
- **Insights:**
 - Biography movies resonate more with audiences, balancing ratings and voting volume.
 - Action and Adventure appeal broadly, driving consistent popularity



Movie Releases by Month

- **Releases Monthly**

- October Peak: There is a clear peak in movie releases during October, the highest point.
- February Dip: February shows the lowest number of releases, a noticeable dip.
- Releases fluctuate throughout the year, with an upward trend from July to October, followed by a sharp decline.
- Early Year Releases: January starts strong but declines quickly.
- Mid-Year Dip: Lower release count in July, followed by an upward trend.
- Late-Year Drop: Post-October peak, releases decline in November and December



Movie Ratings Across Genres - ANOVA Analysis

- **Key Insights**

- Null Hypothesis (H_0): All genres have the same average ratings.
- Alternative Hypothesis (H_1): At least one genre has a significantly different average rating.

- **Results**

- F-statistic: 39.83 (Indicates variability between genre ratings)
- P-value: Extremely low ($2.33e-159$) → Statistically Significant

- **Conclusion**

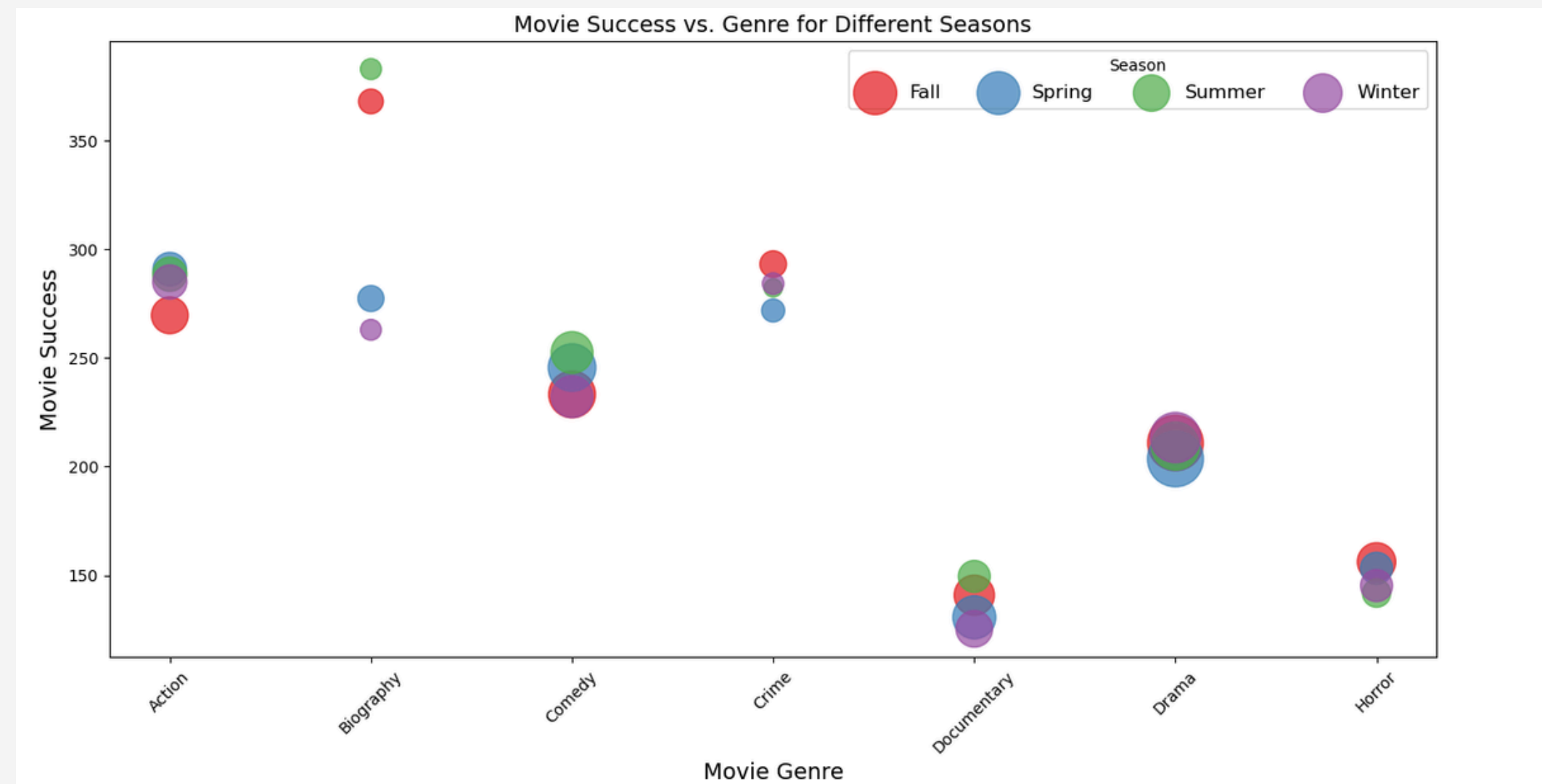
- There is a significant difference in ratings across different genres, meaning some genres perform notably better or worse in terms of ratings.

Significant Difference in Movie Runtimes Across Genres.

- **Key Insights:**
 - Null Hypothesis (H_0): The average runtimes are the same for all genres.
 - Alternative Hypothesis (H_1): At least one genre has a significantly different average runtime.
- **Results:**
 - F-statistic: 62.31 (Indicates variability between genre ratings)
 - P-value: Extremely low ($4.92e-254$) → Statistically Significant
- **Conclusion:**
 - There is a significant difference in movie runtimes between genres. Some genres have notably longer or shorter average runtimes.

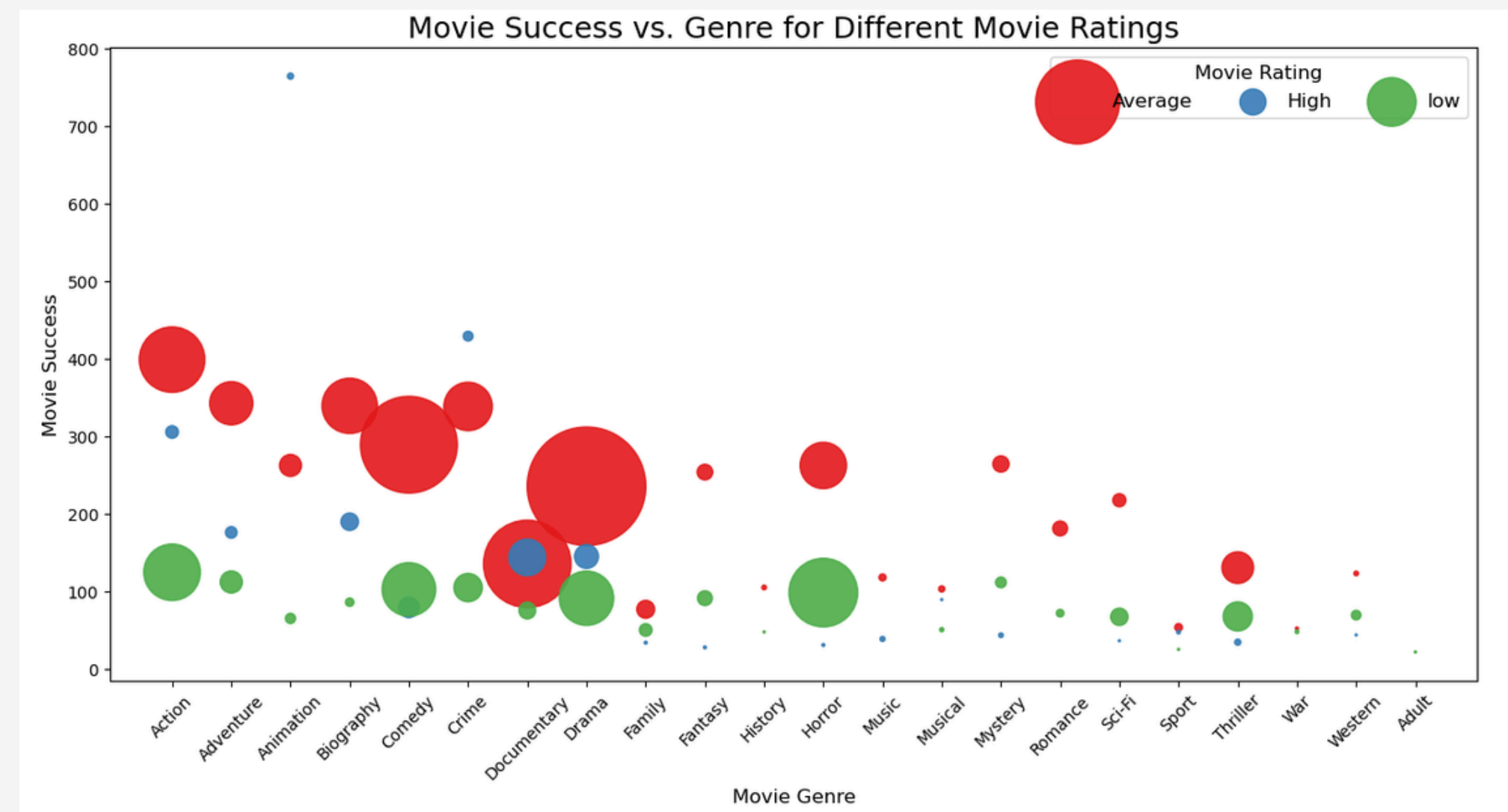
Seasonal Trends in Movie Success Across Genres

- Biography genre peaks highest in success, particularly during summer releases, with fall also performing well.
- Action films show consistent high success (250-300) across all seasons, indicating year-round appeal and seasonal immunity.
- Drama genre maintains moderate success (~200), with slightly better performance in winter, aligning with audience preferences for emotional depth.
- Horror films show the lowest success (~150) with minimal seasonal variation, highlighting a niche but limited audience reach.
- Crime films perform best in fall, suggesting strategic release timing enhances connection with intense, darker themes.



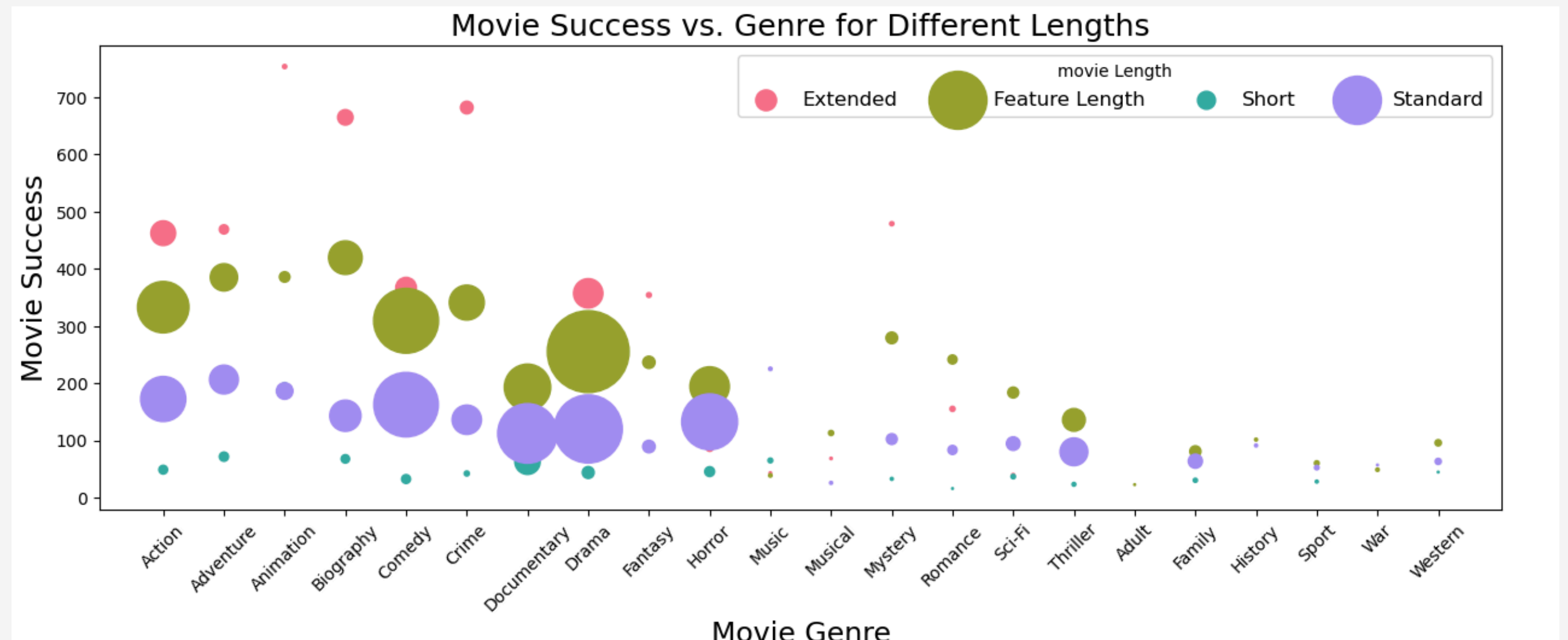
Movie Success Across Genres and Rating Categories

- Documentaries, comedy: High success in "Average" ratings; consistent performance despite moderate ratings.
- Action & Adventure: Strong performance with "Average" ratings; popularity achieved without exceptional ratings.
- Most successful movies are in the "Average" category; critical acclaim does not guarantee commercial success.
- Horror, Music, Romance: Smaller bubbles but maintain moderate success levels.
- Low-Rated Movies: Consistently lower success; clear gap between "Average" and "Low" rated films.
- Comedy & Drama: Large bubbles in "Average"; reliable popularity but less frequent exceptional success.



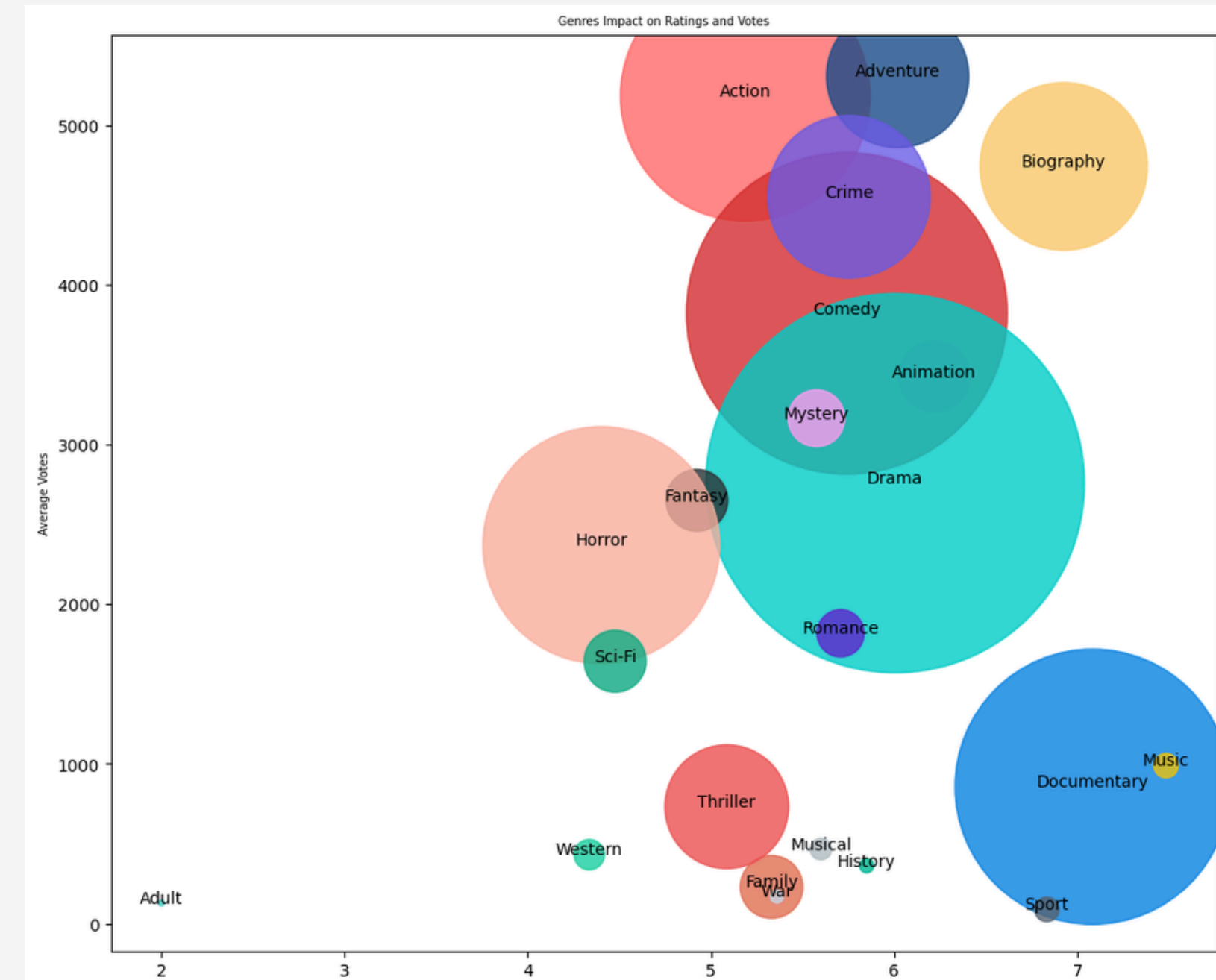
Movie Success by Length and Genre

- Extended-length movies outperform others, especially in Biography, and Action, showing audience preference for longer content.
- Documentaries thrive in the feature-length format with high success metrics, making it the ideal format for this genre.
- Length hierarchy: Extended > Feature > Standard > Short, suggesting longer formats generally correlate with higher success.
- Mainstream genres (Action, Adventure, Comedy, Drama) perform well across multiple lengths
- Niche genres (Horror, Western, War) show consistently lower success metrics.
- Short films consistently underperform across genres, highlighting their limited commercial appeal.



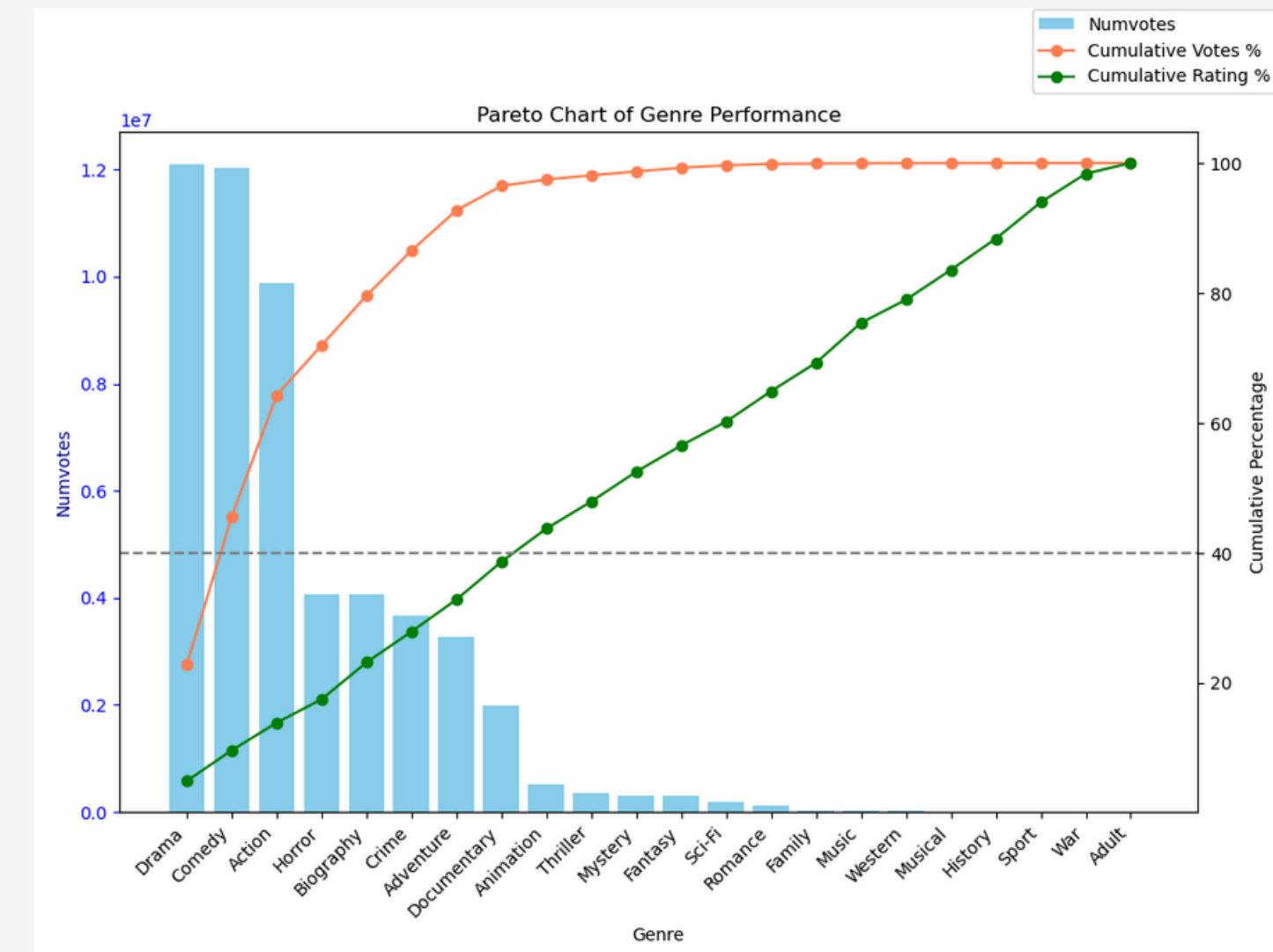
Audience Votes and Ratings Across Genres

- Action, Adventure, and Crime genres have high ratings (~5-6) and large vote counts (~4000-5000), showcasing mass appeal and quality content.
- Documentaries achieve the highest ratings (~7) but have low vote counts (~1000), indicating niche audience appeal despite high quality.
- Drama features strong ratings (~5-6), substantial votes (~3000), and large bubble sizes, reflecting its popularity and consistent reception.
- Genres like Documentary, History, and Music excel in ratings, while Horror and Sci-Fi attract higher votes. Few genres like Action, Adventure, and Drama succeed in both metrics.



Pareto Analysis: Genre Contribution to Audience Engagement

- Drama, Comedy, and Action account for approximately 60-70% of total votes, making them priority genres for studios.
- The first 7-8 genres contribute to roughly 80% of all votes, confirming that a few genres dominate audience engagement.
- A steep rise in the percentage of the cumulative vote for leading genres contrasts with a long tail of low-contribution genres, emphasizing the gap between popular and niche categories.
- The cumulative rating percentage increases more gradually than votes, indicating that high audience engagement doesn't always align with higher ratings.
- Sport, War, and Adult genres contribute minimally to votes and ratings, suggesting limited commercial viability



Key Findings

High-Performing Genres

- Drama, Comedy, and Action are the most engaging genres, contributing 60-70% of audience votes.
- Biography is the most successful genre with high ratings and significant engagement.
- Documentary achieves high ratings but appeals to a niche audience with low vote counts.

Success Metrics

- High-rated movies do not guarantee mass appeal (e.g., Documentaries with high ratings but low popularity).
- Most commercially successful films fall in the "Average" ratings category.

Seasonal Trends

- Fall dominates with the highest movie releases and success metrics, particularly in October.
- February and mid-year (July) show dips in releases and engagement.

Runtime Preferences

- Most successful movies have runtimes between 80–110 minutes.
- Extended-length movies perform better in genres like Biography, Animation, and Action.

Engagement Analysis

- Genres like Action, Adventure, and Crime lead with high votes and solid ratings, showing broad appeal.
- Short films and niche genres like Horror and War consistently underperform.

Recommendations

Focus on Popular and Successful Genres

- Prioritize Drama, Comedy, and Action for mainstream appeal and financial returns.
- Allocate resources to Biography films to balance critical acclaim and audience satisfaction.
- Reduce investments in low-performing genres like Horror unless targeting niche markets.

Strategic Release Planning

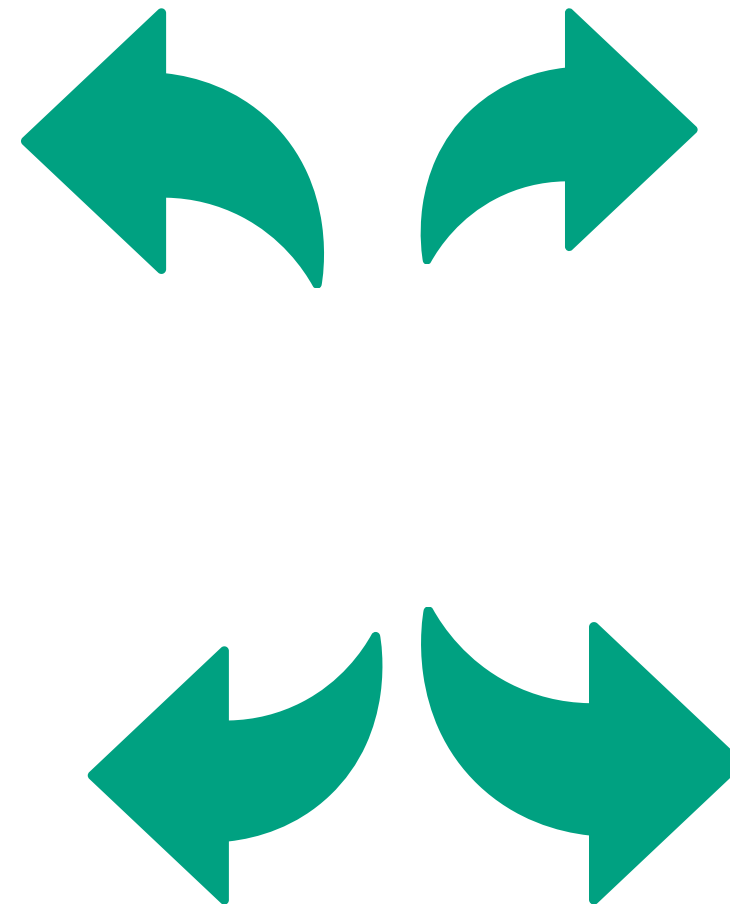
- Release high-budget films in the Fall, especially in October, to leverage audience engagement peaks.
- Avoid releases during February and mid-year dips for blockbuster productions.

Target Optimal Runtime Formats

- Focus on movies with runtimes between 80-110 minutes for mainstream genres.
- Produce longer films for Animation, Biography, and Action, which benefit from extended runtimes.

Diversify Formats for Niche Markets

- Consider documentaries in feature-length formats to maximize quality and niche engagement.
- Leverage high-rating genres like Documentary and History for targeted marketing.



Conclusion

- Data-driven decision-making is essential for successful movie production.
- High-performing genres like Drama, Comedy, and Action should be prioritized for maximum audience engagement and ROI.
- Aligning movie runtimes with audience expectations enhances overall success.
- Strategic release timing during peak seasons, such as Fall, optimizes visibility and profitability.
- While high ratings are valuable, commercial success often depends on balancing quality with broad audience appeal.
- This approach allows the studio to effectively target diverse audience segments and maintain a competitive edge in the market.

Contributors

Do you have any questions?

Feel free to reach out!



Western



Amase



Ryan



Abdirahaman



Linnet