Microsoft Movie Analysis

By Bryn Bowlden

Project overview

This project employs exploratory data analysis to derive insights for Microsoft Corporation's initiative to establish a new movie studio. Drawing from datasets sourced from IMDB and The Numbers, recommendations are provided regarding the genres that Microsoft should consider exploring. The metrics utilized to gauge the success and popularity of each genre include total net profit, ROI and average rating and average run times.

Business Problem

Microsoft is venturing into the film industry with the establishment of a new movie studio, seeking to leverage exploratory data analysis to navigate the complex landscape of successful films. With limited prior expertise in the field, they aim to explore historical and current trends to glean actionable insights. Through this analysis, Microsoft seeks to inform their decision-making process and determine the most promising film genres to pursue.

Questions Addressed

The questions addressed by this analysis include:

•What is the average runtime a movie should aim to be within?

•What are the most popular movie genres to produce, and which have the greatest average ratings?

•What genres have the highest net profit or ROI?

Provided Data Overview

The data sets used in this analysis can be seen below with a brief description for each in the associated subpoints:

• <u>imdb.title.basics:</u>

- Contains movie title details, including runtime duration and genre

• <u>imdb.title.ratings:</u>

- Contains IMDb's vote and rating data for movie titles

• <u>tn.movie_budgets.csv.gz:</u>

- Contains movie title production costs as well as their domestic and worldwide gross
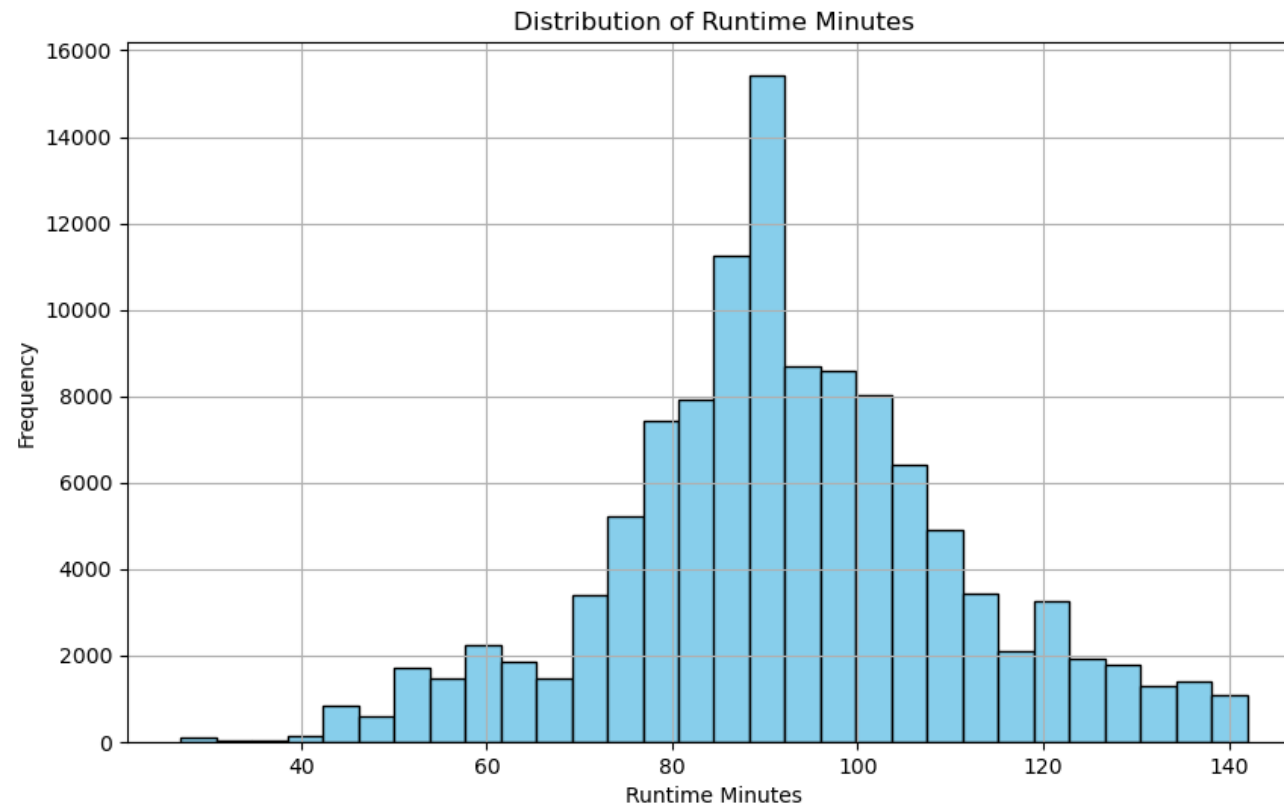
How the Data Will be Used

There is a significant discrepancy between the amount of data available in each data set. Therefore, utilization of the data will depend on its size and content diversity.

Whenever feasible, the analysis will prioritize the utilization of the largest datasets to address the specified problem. However, for inquiries such as identifying the genres with the highest net profit or ROI, due to the data available, smaller yet comprehensive datasets will be employed to ensure precise analysis.

Average runtime

The average runtime of a movie didn't seem to have a direct correlation to the rating a movies received but the finding show that the majority of movies between 80 to 100 minutes in duration with 85 minutes being the average.
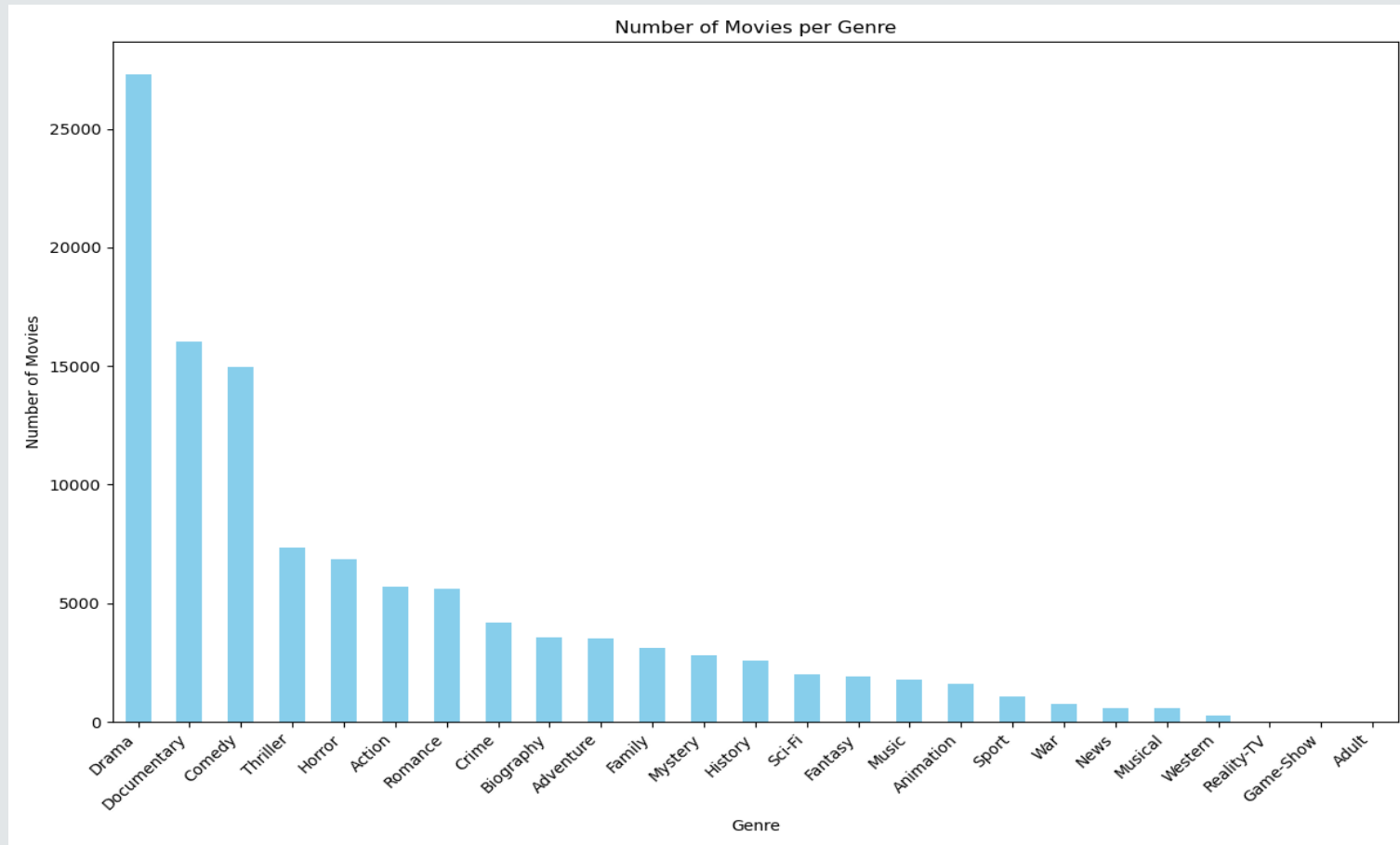
Movies per Genre

When analyzing the number of movies per genre, we can see a the distribution of genres being produced. This could indicate a genres popularity.
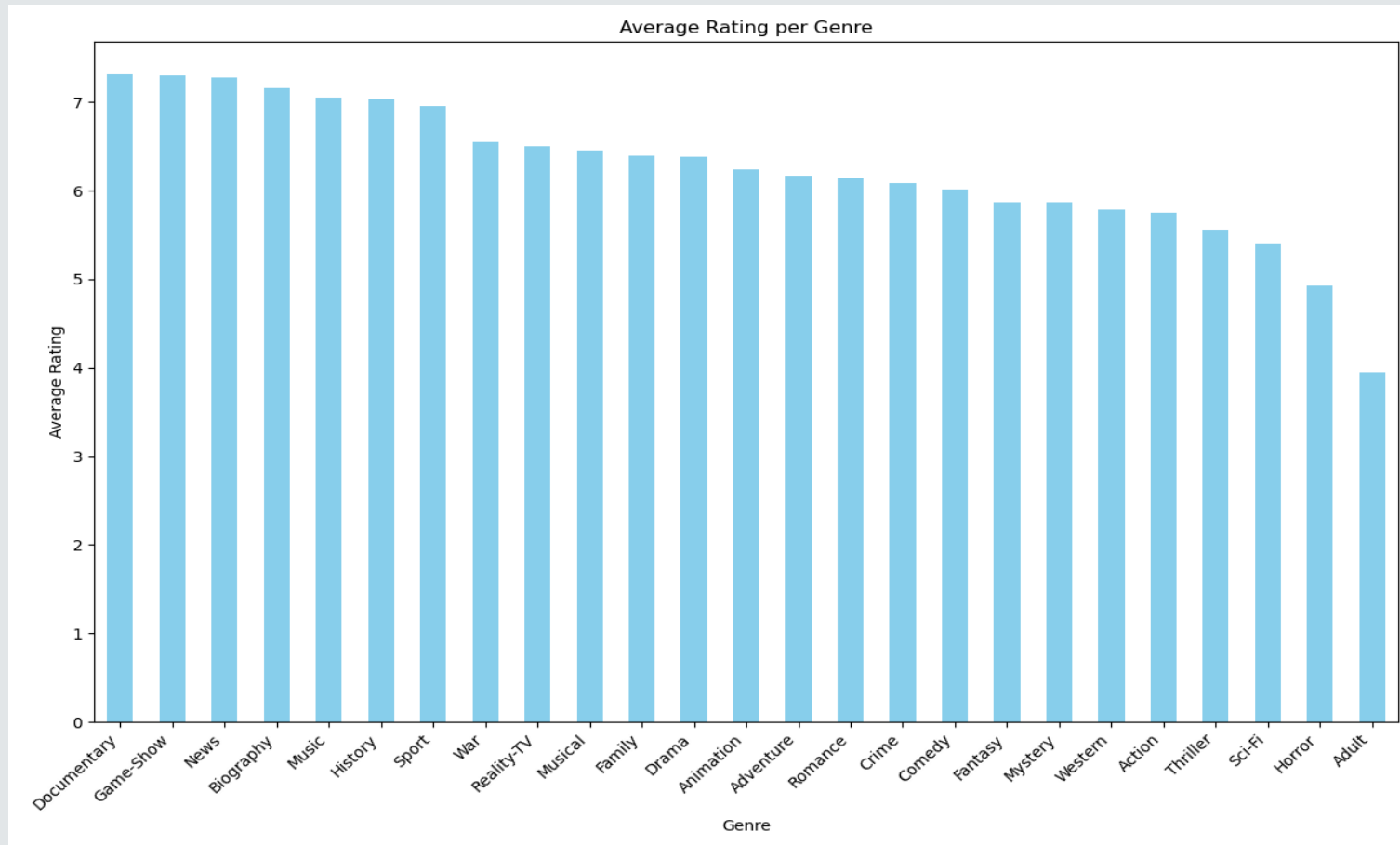The top 10 most produced genres in descending order are:

- •Drama
- •Documentary
- •Comedy
- •Thriller
- •Horror
- •Action
- •Romance
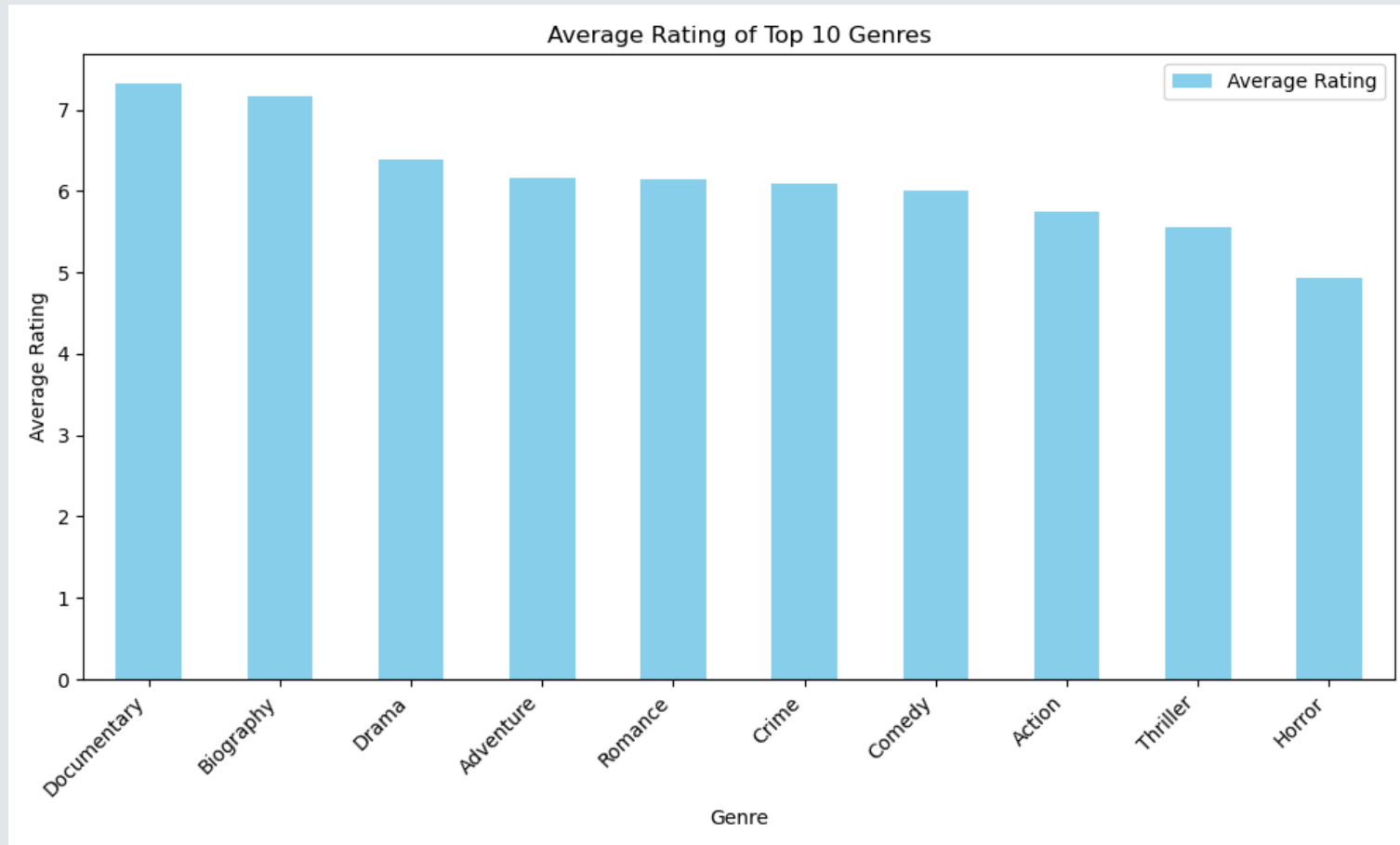- •Crime
- •Biography
- •Adventure

Average Rating per Genre

The difference in average ratings between genres is relatively small, yet several less popular genres exhibit notably high average ratings.
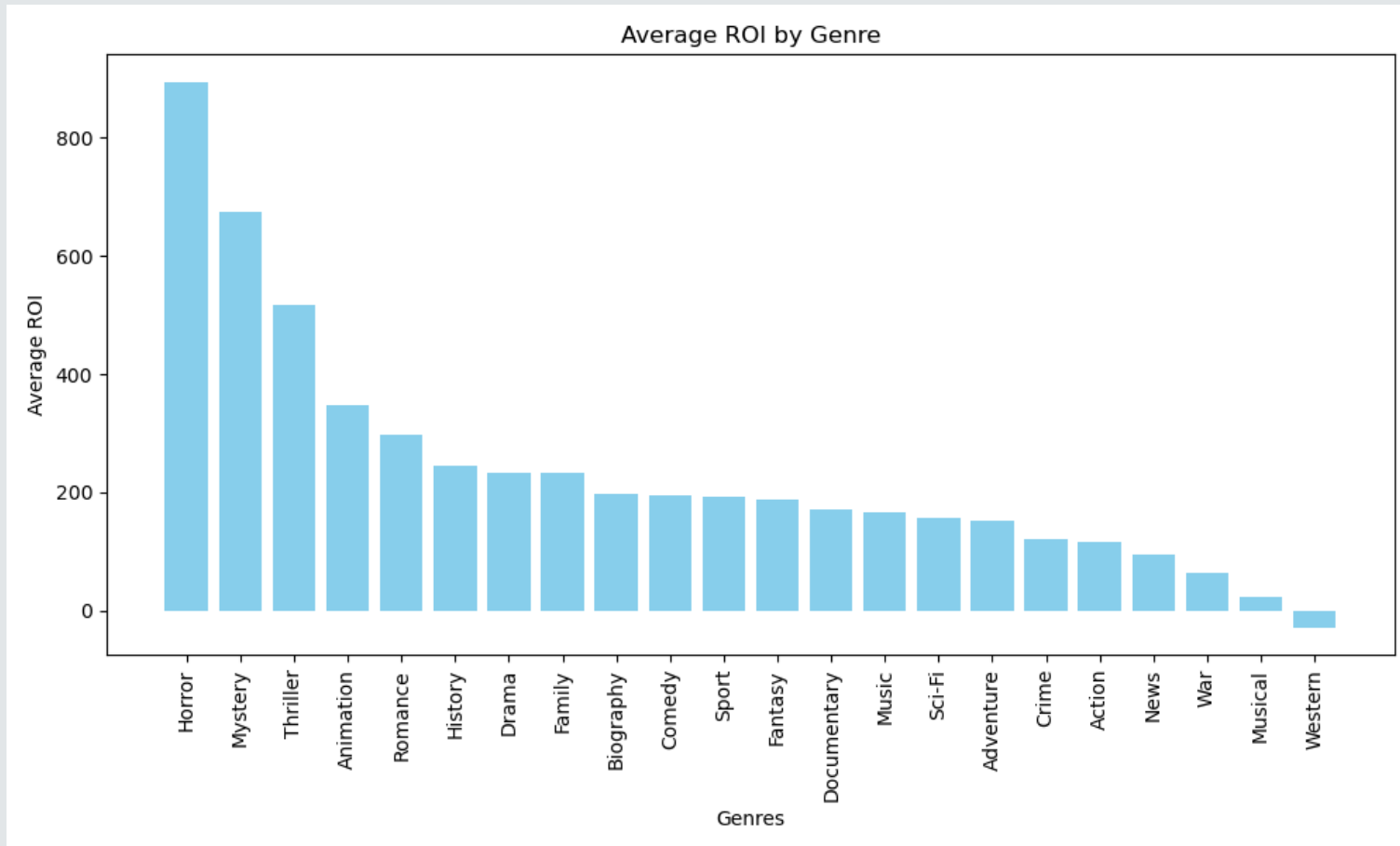
Average Rating per Genre – Top 10

Among the top 10 genres, there remains a relatively narrow variance in ratings across genres. The top three genres are documentary, biography, and drama, while the bottom three are action, thriller, and horror.
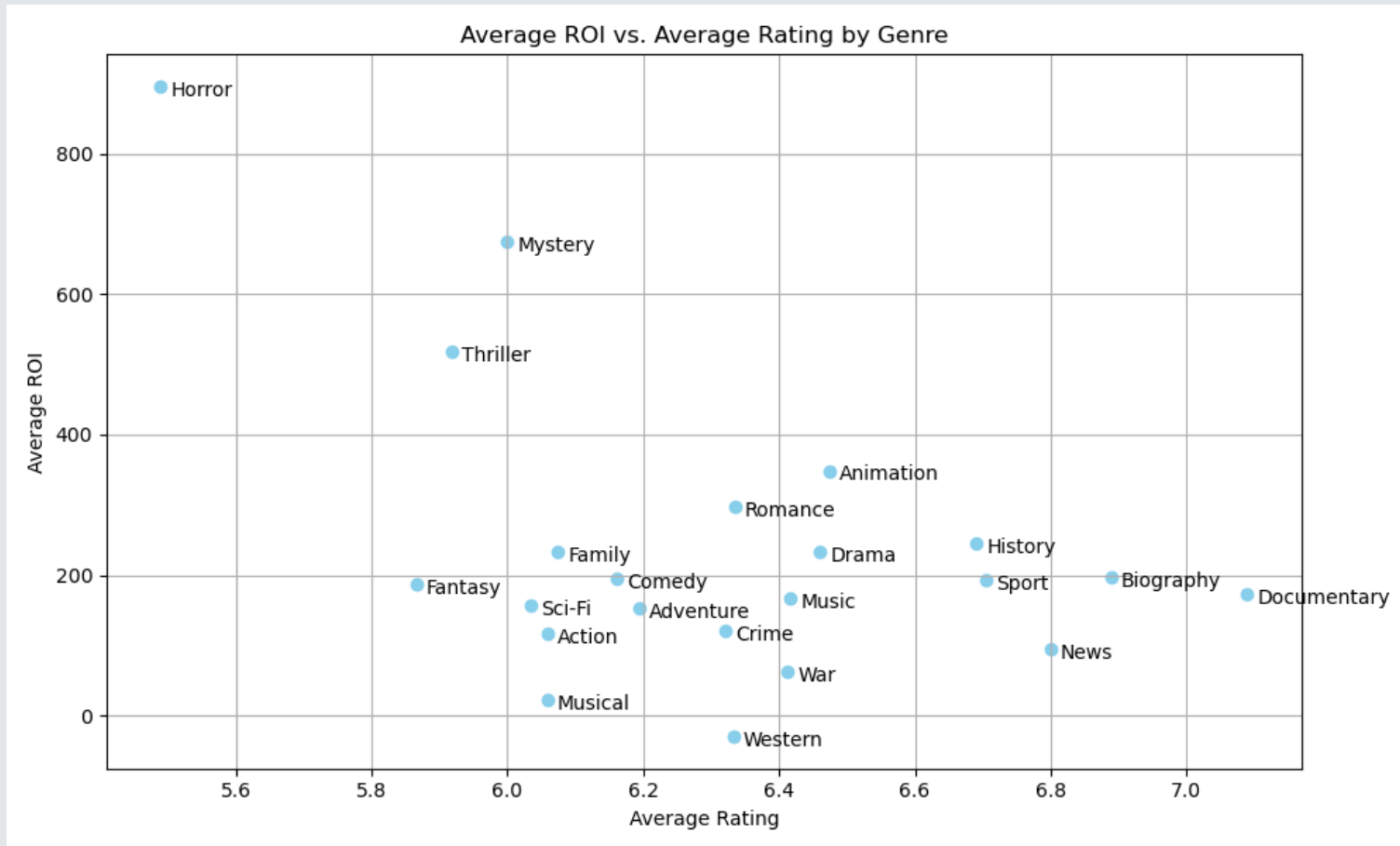

Average Rating of Top 10 Genres

Return on Investment

By computing the net profit of movies within the provided dataset, I derived the ROI for each movie. This information was then utilized to determine the average ROI for each genre. It's evident that the top three genres with the highest ROI are horror, mystery, and thriller, while the lowest are war, musical, and western.

Average ROI vs Average Genre Rating

We observe a negative correlation between ROI and the average rating of genres. This trend is likely attributed to niche genres, which, despite having a smaller but dedicated audience, tend to yield higher average ratings while garnering lower returned on investments.

Conclusions

Based on the thorough analysis of the dataset, several noteworthy findings emerge:

•**Runtime Influence on Rating:** The data suggests that the length of a movie does not necessarily dictate its rating. Despite this, the majority of films fall within the 80 to 100-minute range, with an average runtime of approximately 85 minutes.

•**Genre Popularity Insight:** Examination of movie counts across different genres sheds light on their popularity within the industry and among audiences.

•**Dominant Genres:** The top 10 most prolific genres, ranked in descending order, include Drama, Documentary, Comedy, Thriller, Horror, Action, Romance, Crime, Biography, and Adventure.

•**Rating Variance Among Genres:** While the average ratings across genres exhibit minimal differences, certain less mainstream genres exhibit surprisingly high average ratings.

•**Genre-Specific Ratings:** Focusing on the top 10 genres allows for a more nuanced understanding of their respective ratings. Documentary, Biography, and Drama emerge as top performers, whereas Action, Thriller, and Horror genres tend to receive lower average ratings.

•**Financial Performance Analysis:** By computing the net profit and ROI for each film, valuable insights into the financial viability of different genres are revealed. Horror, Mystery, and Thriller genres demonstrate the highest ROI, while War, Musical, and Western genres yield the lowest returns.

•**ROI vs. Audience Reception:** A negative correlation is observed between ROI and the average rating of genres. Niche genres, although catering to smaller but dedicated audiences, tend to achieve higher average ratings while yielding lower returns on investment.

In summary, while factors like runtime and genre popularity provide valuable insights, understanding the financial dynamics, particularly ROI, is crucial for decision-making in the film industry. This comprehensive analysis aids filmmakers and investors in making informed choices regarding genre selection and resource allocation, balancing creative aspirations with financial considerations.

Final Recommendations

Based on the insights gleaned from the data and the considerations outlined above, the recommendations can be categorized into two main approaches: focusing on return on investment and prioritizing popularity and audience reception.

These recommendations are as follows:

• **For Return on Investment:** If the primary consideration is maximizing return on investment, the data suggests that producing a horror movie would be the most lucrative option.

• **For Popularity and Audience Reception:** Alternatively, if the emphasis is on popularity and appealing to a wide audience, opting for a drama movie would be a prudent and reliable choice, given its consistent appeal and broad audience acceptance.

Both of the aforementioned genres should aim to align with the average runtime duration of 85 minutes, which appears to be the standard duration among movies in the dataset. This duration could help maintain audience engagement while ensuring efficient storytelling within the typical attention span.

Thank you!