

# IMDb Movies



Cassidy Bell, Gorgina Kareem, Jason Johnson, Erika Pino

# Our Questions & Thesis

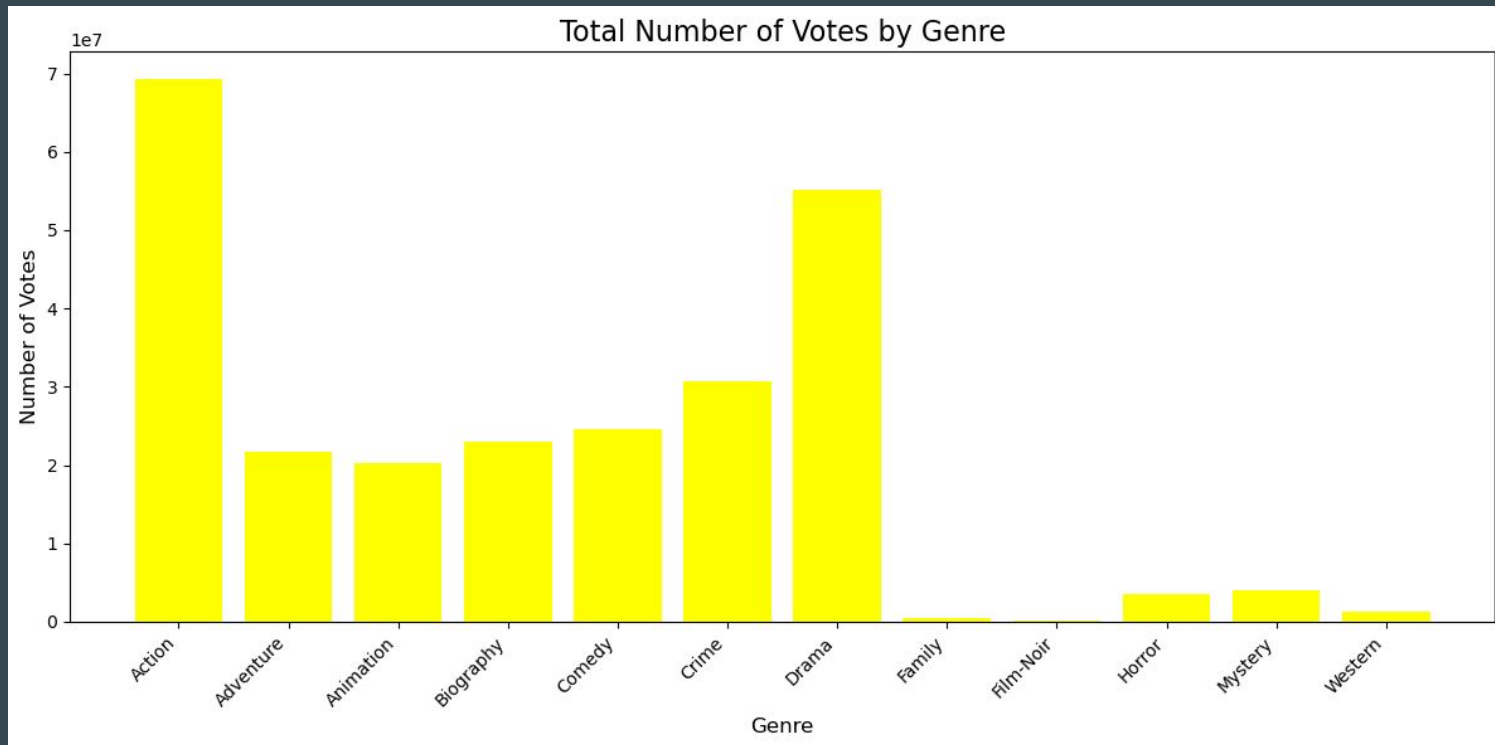
- How would the movie runtime affect the rating of the movie itself?
- Does the number of votes affect the IMDb rating?
- What movie genre has the most number of votes?
- Which movie rating certificate has the highest average Metascore - in comparison to IMDb review scores?

We aim to uncover factors that possibly contribute, or correlate, to a movie's success and how these elements align with audience and critical reception.

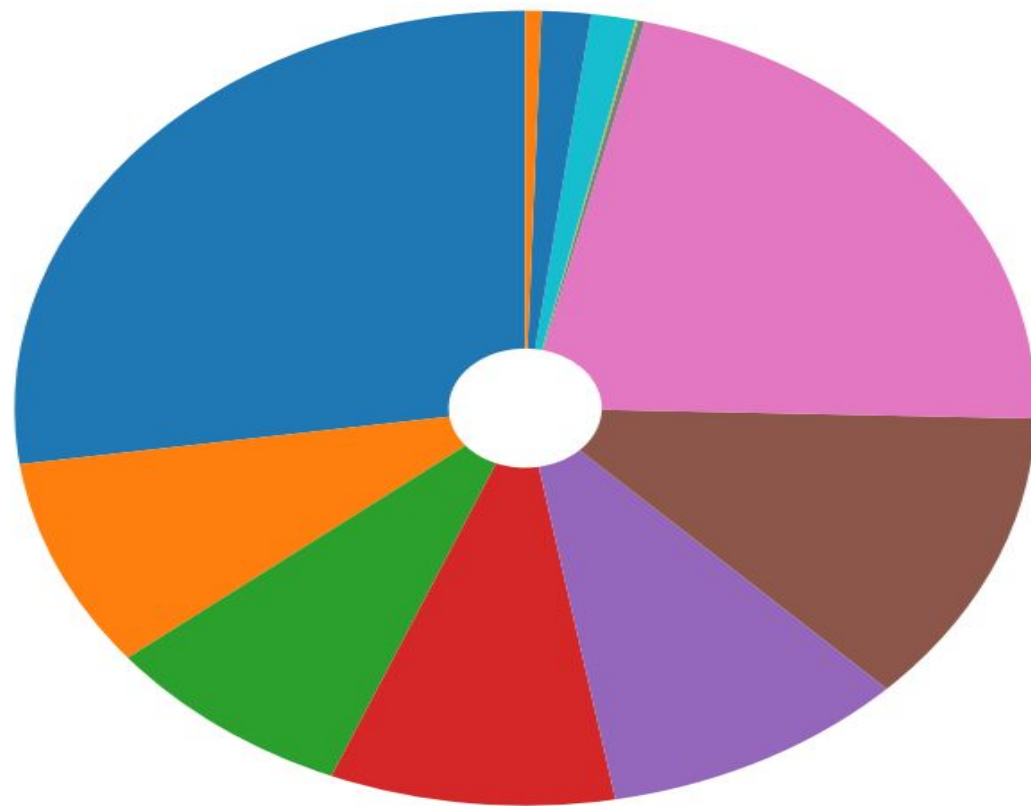
# Data Cleaning Steps

1. Gathered information on Data
2. Checked for missing values & non-missing values
3. Count missing values per column
4. Dropped some columns (Poster Link, Overview, Star 2, Star 3, Star 4)
5. Clean run time & gross columns (removed commas & converted to integers), genre (extract only the 1st genre listed)
6. Combine Certification Ratings from obscure ratings 50+ years ago with today's ratings- ended with only 5 values
7. Renamed the column names to cleaner options

What movie genre has the most number of votes?



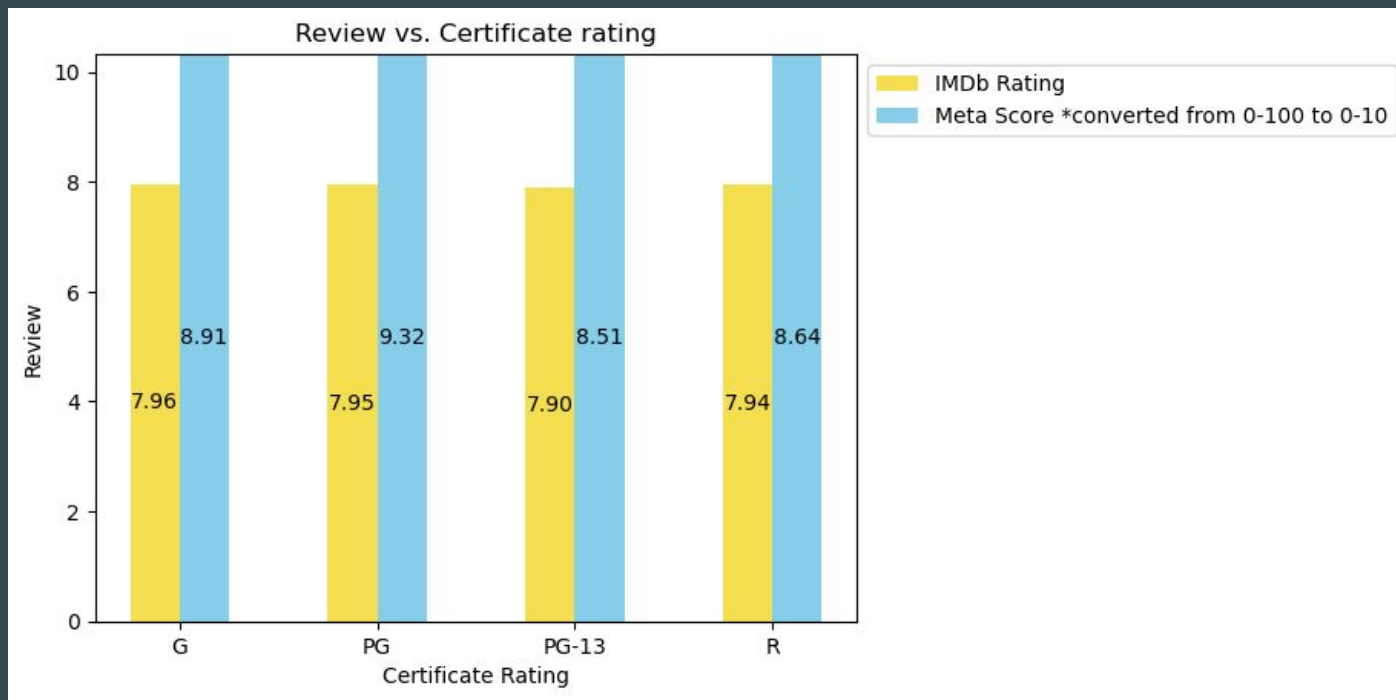
Distribution of Votes by Genre



Genres

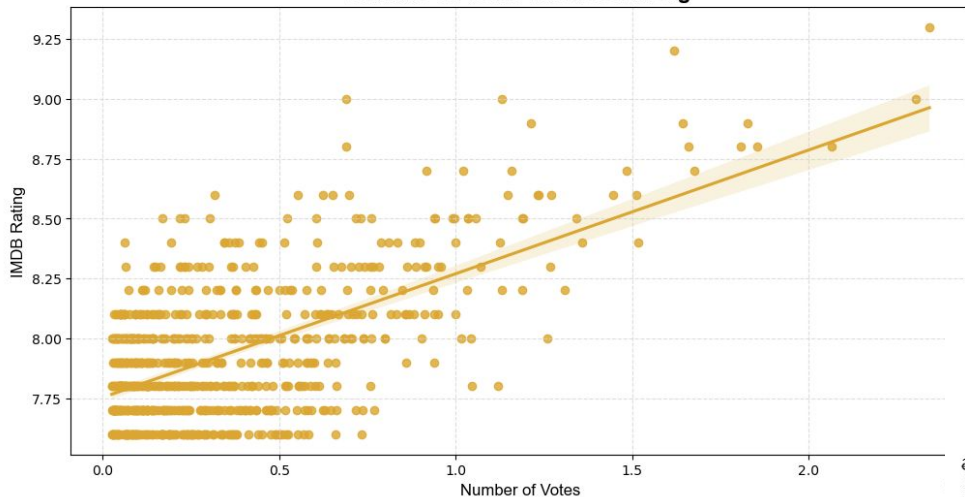
- Action: 27.3%
- Adventure: 8.6%
- Animation: 8.0%
- Biography: 9.0%
- Comedy: 9.7%
- Crime: 12.1%
- Drama: 21.7%
- Family: 0.2%
- Film-Noir: 0.1%
- Horror: 1.4%
- Mystery: 1.6%
- Western: 0.5%

Which movie rating certificate has the highest average Metascore - in comparison to IMDb review scores?

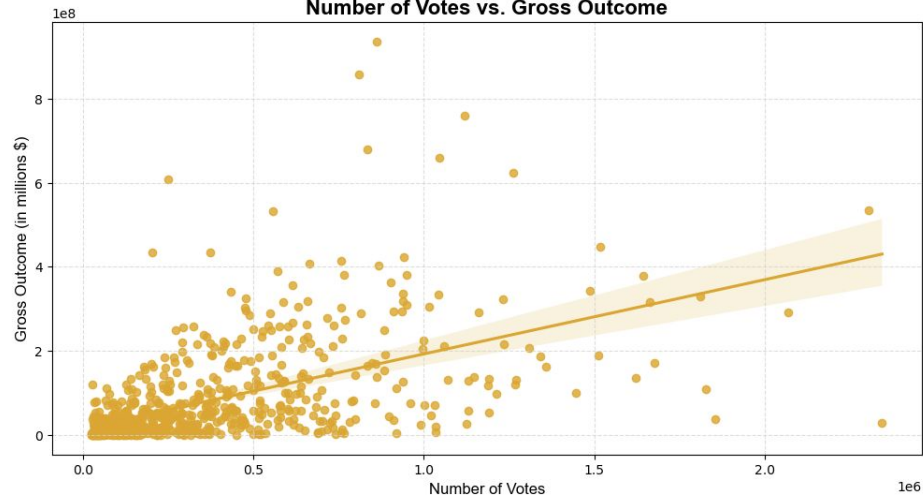


Does the number of votes affect the IMDb rating or gross outcome more?

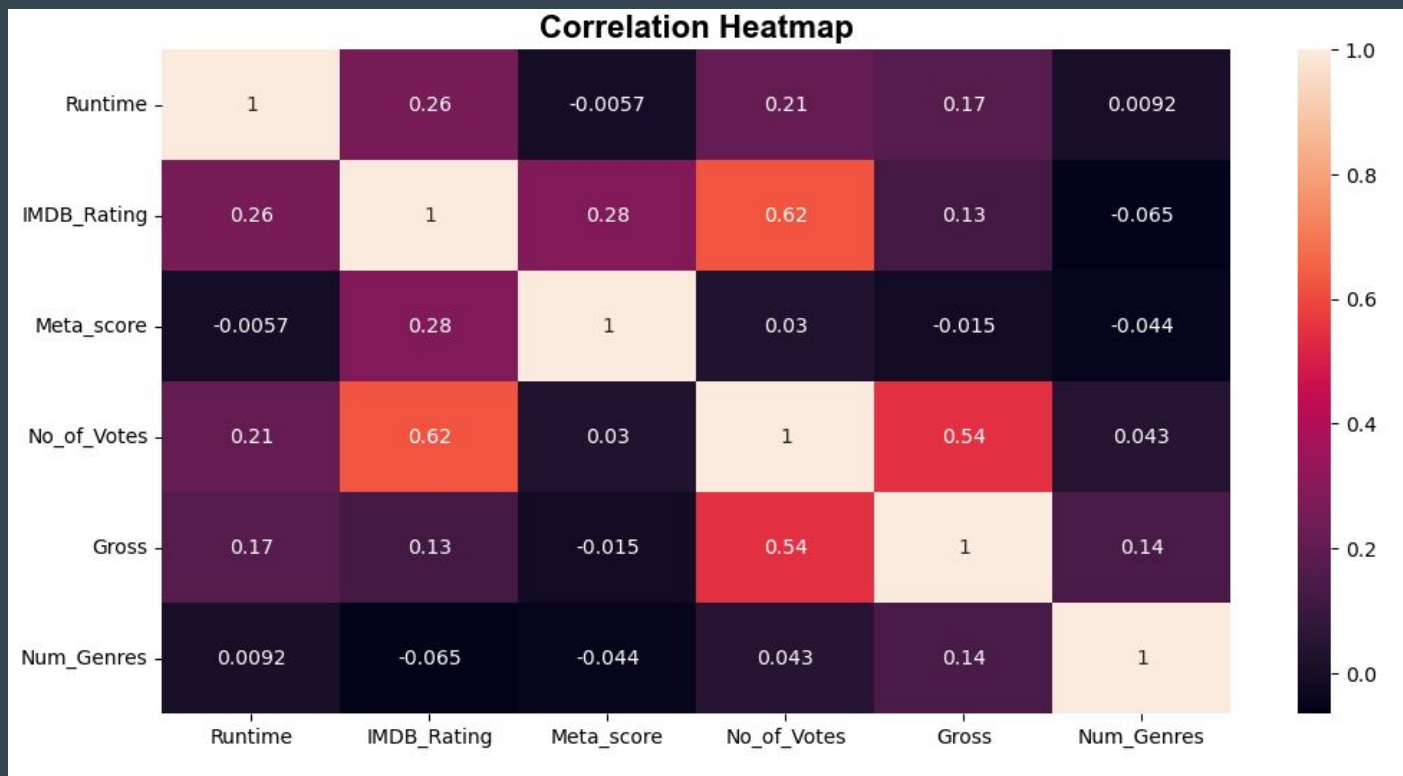
Number of Votes vs. IMDb Rating



Number of Votes vs. Gross Outcome

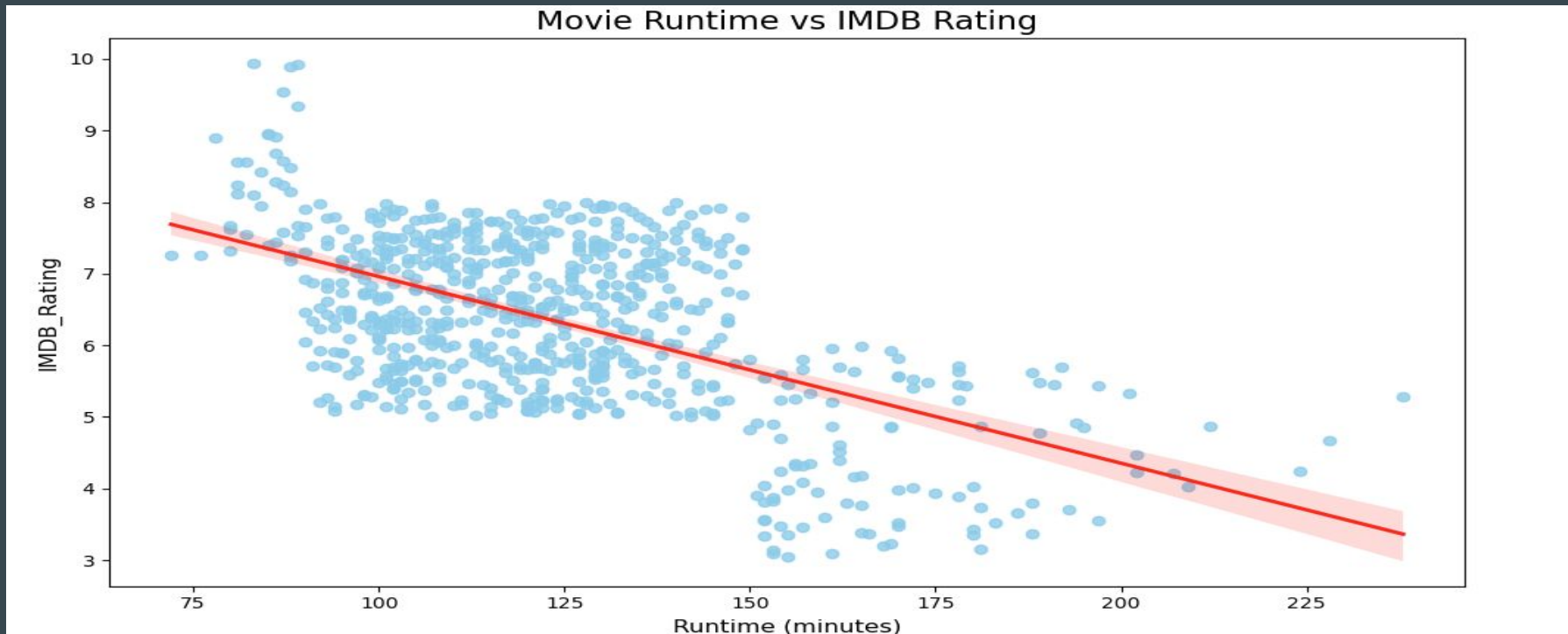


# Correlation

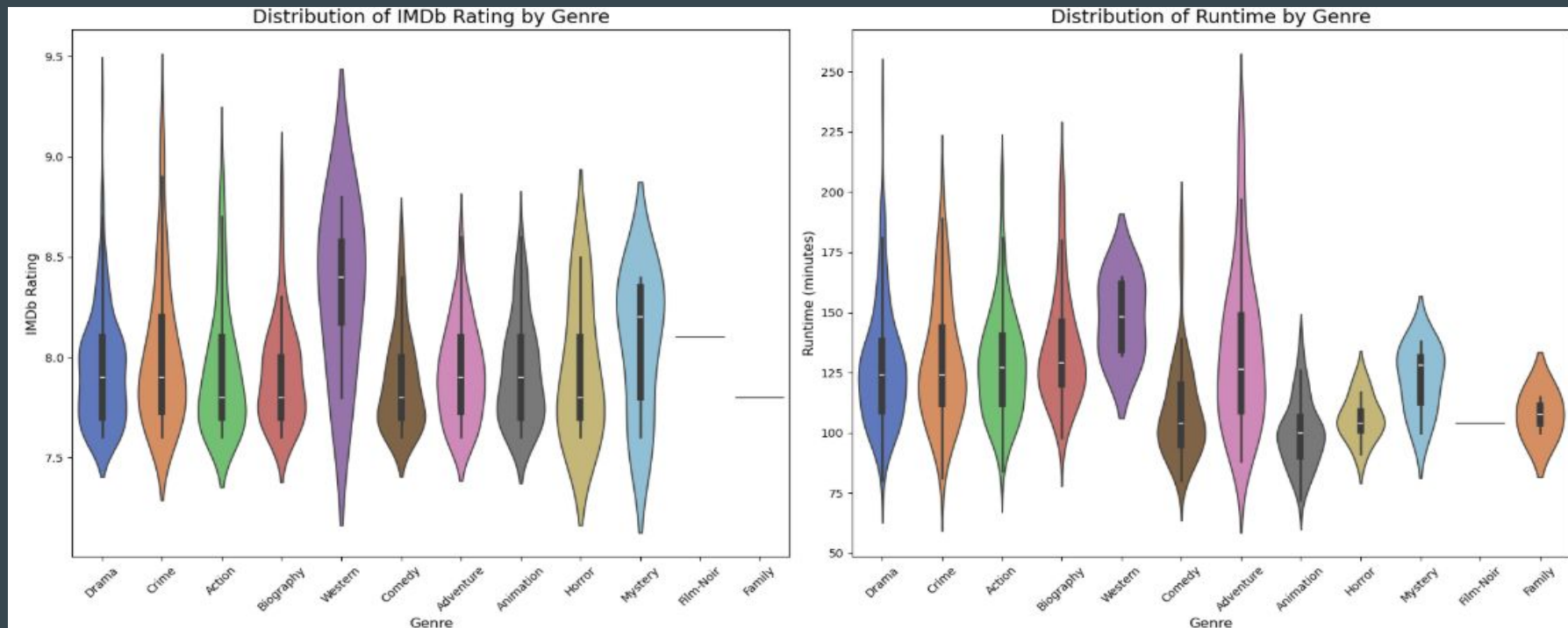




How would the movie runtime affect the rating of the movie itself?



# Violin Plots for IMDb Rating and Runtime by Genre



## Question Answer:

In regards to IMDb ratings and runtime, there are some factors that need to be considered. In this Analysis, the movies that have the shorter runtime of under 150 minutes have a higher IMDb movie rating. Whereas, the movies that have the longer runtime of 150 minutes and beyond have resulted in lower IMDb movie ratings.

# Bias & Limitations

- Cleaning Certification Ratings Limitation
  - No clear definition as to what the outdated movie ratings would be equivalent to in today's movie ratings so we had to use our best judgement to combine them.
- Genre Limitation
  - Most movies fall under multiple genres, so when we cleaned the genre column, we only had it show the first genre listed, which could make some of our data inaccurate.
- Age Bias
  - IMDb respondents are not a true representation of all movie watchers as younger children and older adults are less likely to post a review on the movies they watch.

# Future Work & Conclusion

In summary, our exploration of the IMDb dataset revealed intriguing correlations between movie attributes and their reception by both audiences and critics. By investigating factors such as runtime, genre, vote counts, and certifications, we identified patterns that highlight how specific attributes correlate with higher IMDb ratings and Metascores. Action, Drama, and Crime emerged as dominant genres in popularity, while child-friendly certifications surprisingly garnered the highest average ratings across both metrics.

Future: Relationship between Genre and Certificate with respect to IMDb rating and Metascore.

# Works Cited

<https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows>

<https://chatgpt.com/>

<https://help.imdb.com/article/imdb/track-movies-tv/ratings-faq/G67Y87TFYYP6TWAV?showReportContentLink=false#>