



ANALYZING VIEWER VOTING TRENDS IN CLASSIC AND MODERN CINEMA

Nolan Arendt and Robert Ransom

Agenda

- Executive Summary
- Data Source(s)
- Data Processing
- Business Analysis
- Future Research Recommendations
- Conclusions
- Questions

Executive Summary

Elevator Pitch

Our analysis of IMDb ratings will unlock valuable insights into the movie industry. By examining genre, country, duration, and sentiment, we'll identify trends and preferences that shape audience ratings. Demographic variations will reveal gender and geographical biases. Filmmakers and industry professionals will gain actionable insights to enhance the movie-watching experience.

Data Source(s)

Description and Integration

- “IMDb movies.csv” – Structured Internet Movie Database (IMDb) listing of 85,856 titles produced between 1894 and 2020; Title, Production Year, Genre, Duration, Country of Origin, Language, Director, Writer, Production Firm, Actors, Plot Description, Votes (Avg/Num), Budget, Review Count.
- “IMDb ratings.csv” – Structured Internet Movie Database (IMDb) listing of 85,856 titles; voting by title, broken down by Weighted Average, Gender, Age Group, Vote Origin.
- “df” – Merged Pandas Data Frame using ‘imdb_title_id’ as Primary Key, 19,252 titles with 24 variables after cleaning, removal of ‘2019 TV Movies,’ Plot Description Sentiment Analysis.

Data Processing

Project Roadmap



Data Selection

Identifying and selecting the most applicable data to support our business questions.



Data Cleansing

Various techniques to ensure that the data is accurate, complete, and ready for analysis.



Exploratory Data Analysis

Examining and summarizing data to gain actionable insights on the box office dataset.



Data Visualization

Techniques used to identify patterns, trends, and relationships that may not be immediately apparent from the raw data.

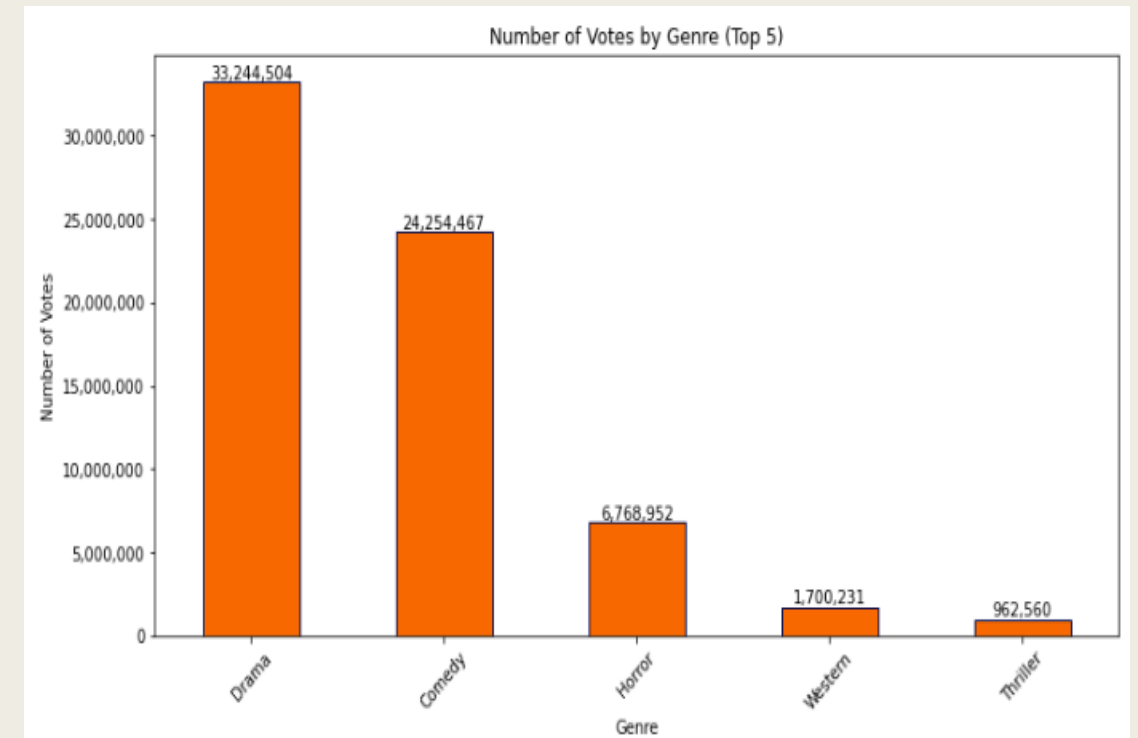
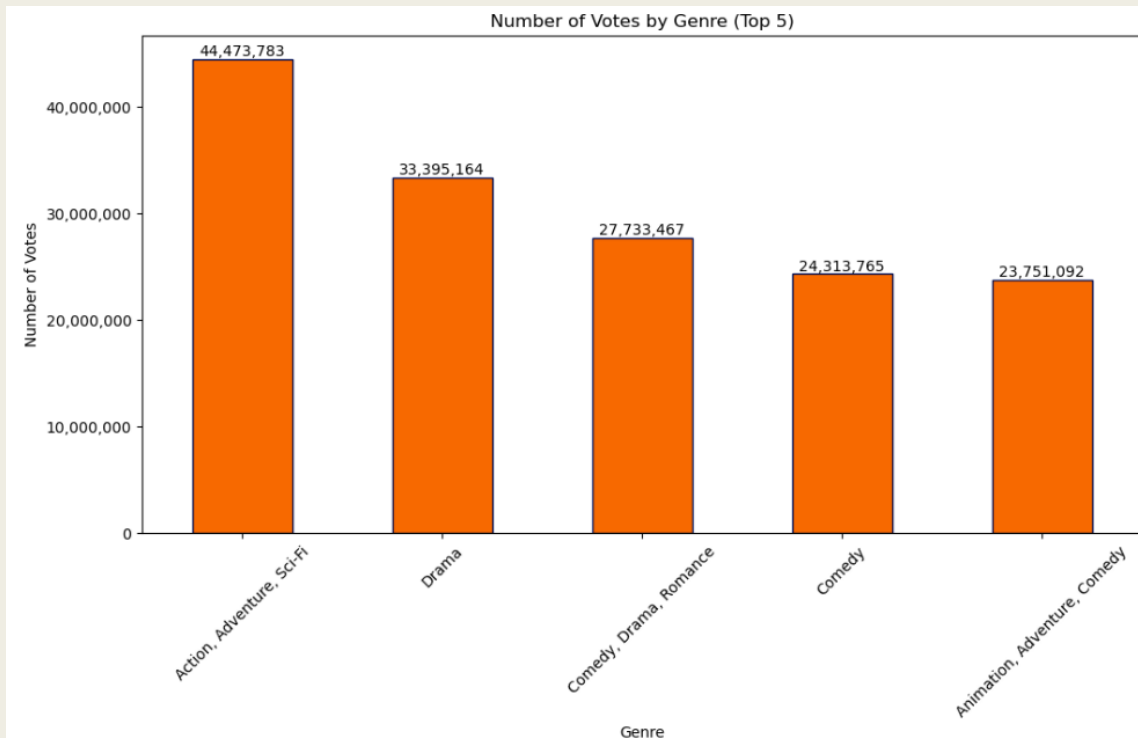
BUSINESS ANALYSIS

Provide analysis of viewer voting trends by genre, plot description, country of origin and identify disparity between gender responses.

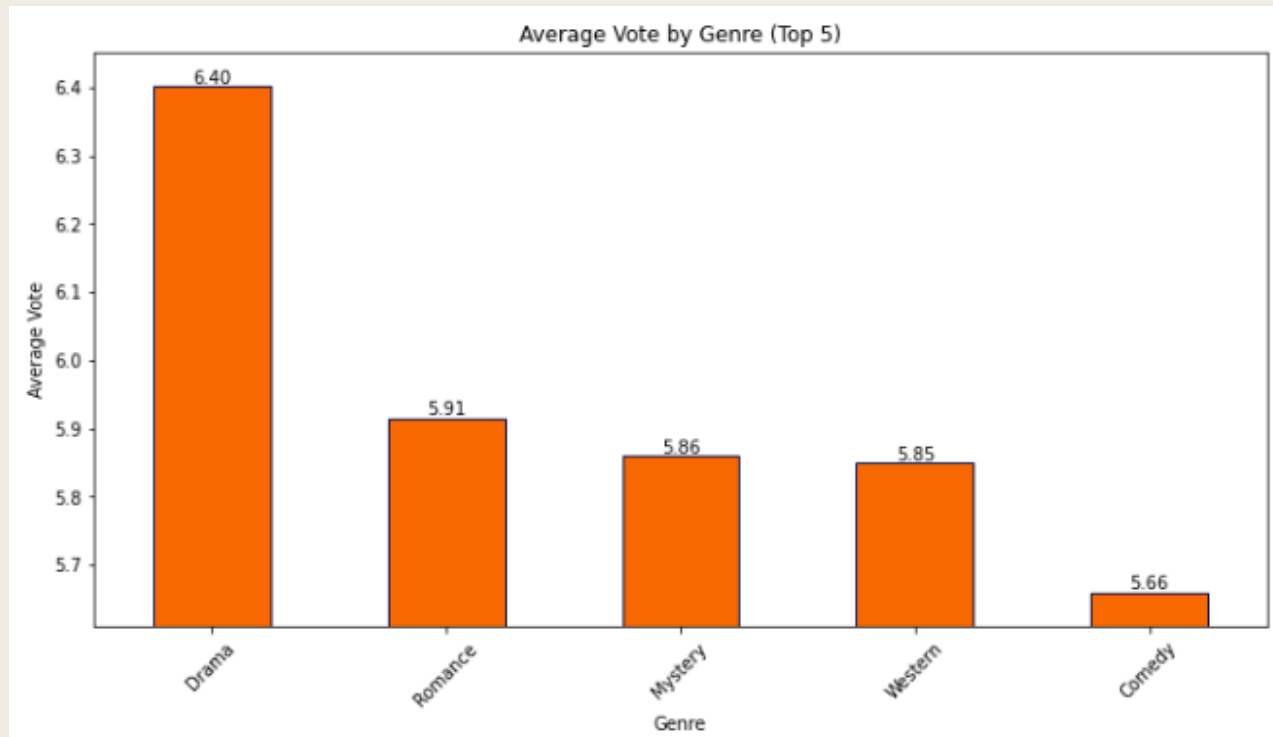
Identify top viewer interaction trends in multi- and single-genre films

Action/Adventure/Sci-Fi by far received the most votes and engagement from viewing audiences. Identifying films with multiple genre labels allows inclusion of sub-genres where single-genre labels may not properly classify media groupings.

Drama, while also receiving the second-highest overall number of ratings, received the highest viewer engagement of single-genre listings. Number of votes by genre may not properly convey audience scores but identifies more responsive viewers for additional surveys or themes for increased interaction.

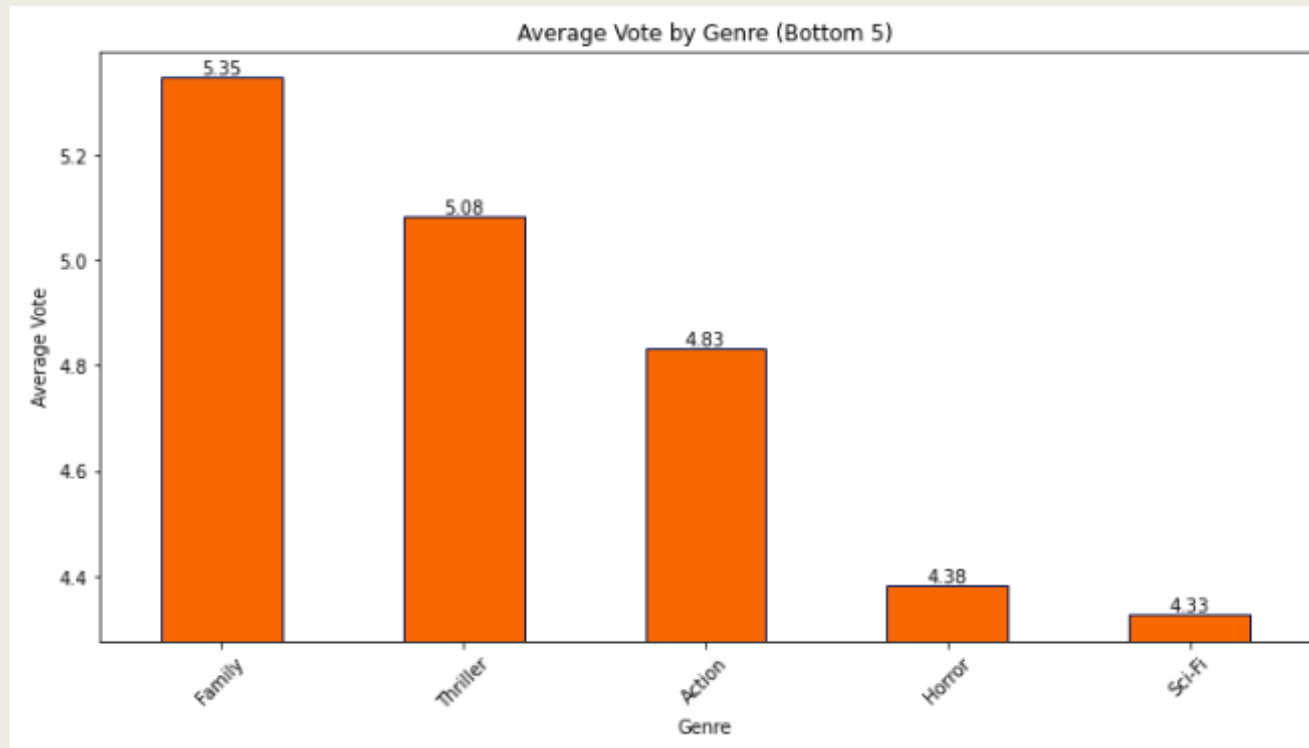


Identify top rated single-genre categories



- Drama (covering a wide range of emotional themes and human experiences) tends to be well received by viewers, distinguished even in the top 5 results for average votes.
- This could be a useful indication of genres to combine for increased viewer interaction and positive reception.

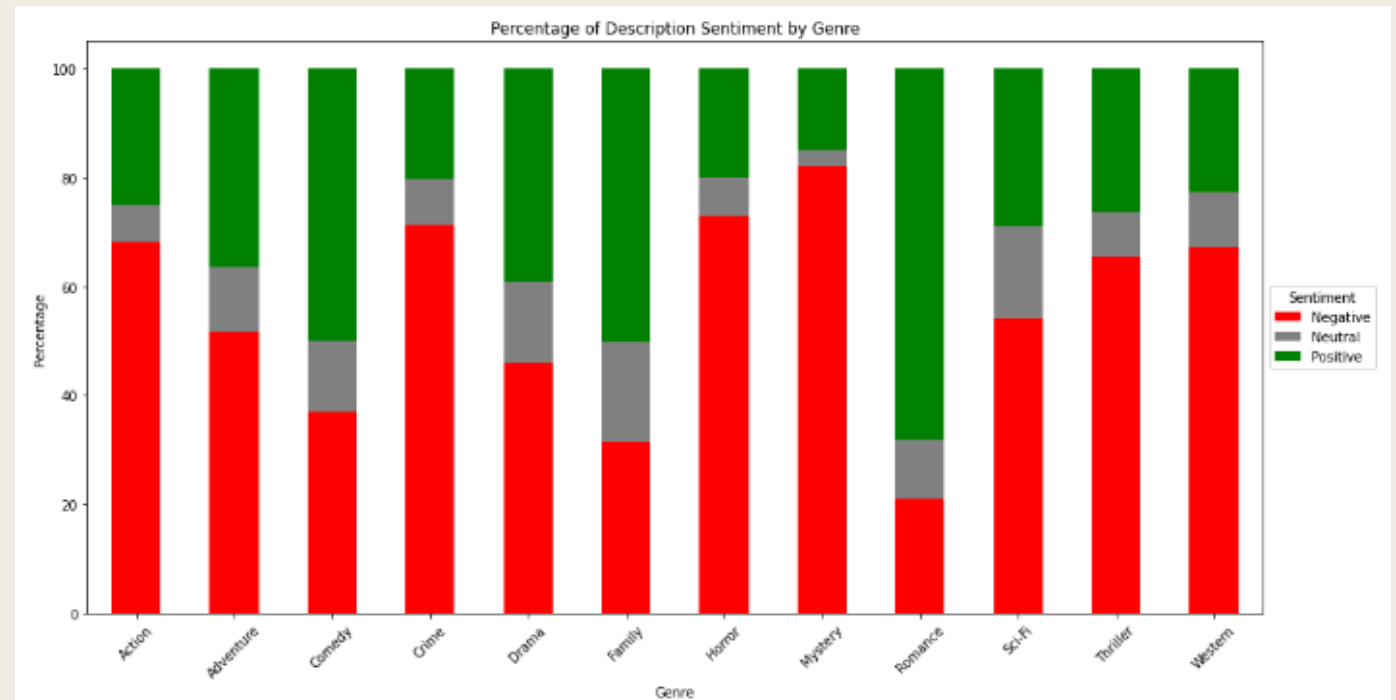
Identify bottom rated single-genre categories



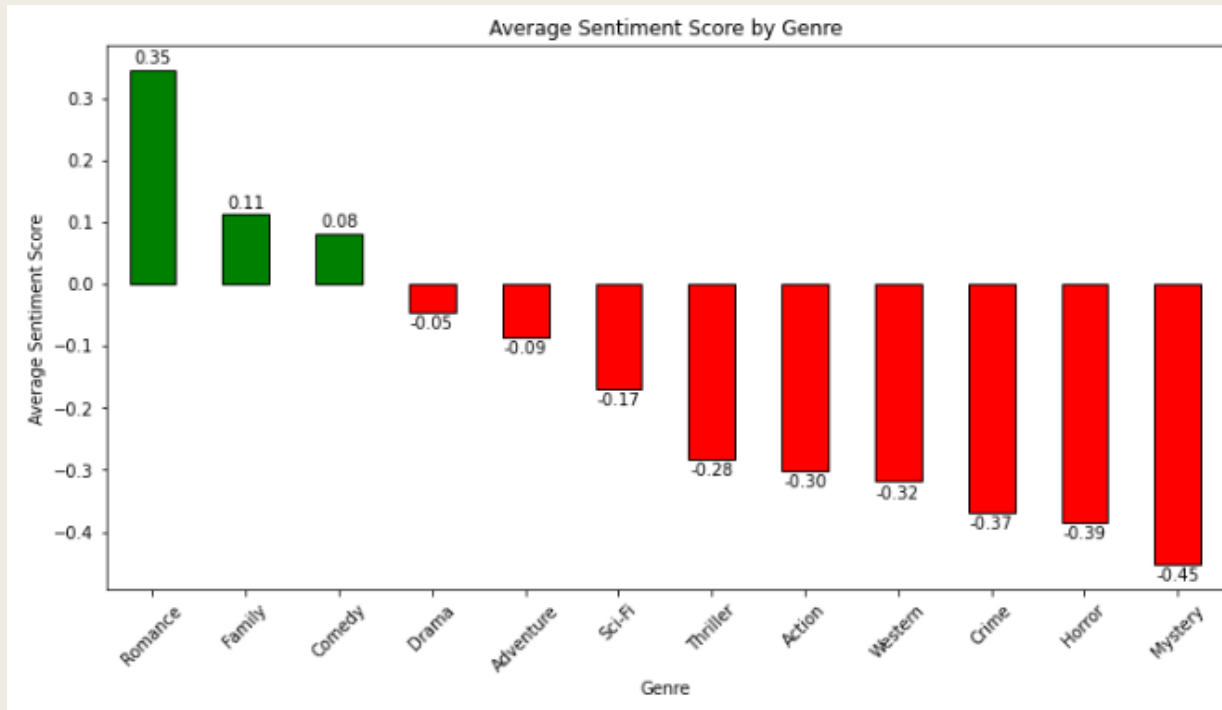
- Surprisingly, single-genre films tend to divert from the voting patterns initially identified in the multi-genre categories. Sci-Fi and Action, while both tied to the most responsive multi-genre category, are found at the bottom of single-genre scale.
- While tied to the most responsive, category, average vote for Sci-Fi places it at the bottom of the scale. Potentially due to the low number of films (257 of 19,252) the Sci-Fi category is more susceptible to outlier votes.

Plot Description Sentiment Analysis

- Correlating the Plot Description Sentiment Analysis with the highest rated genres, preliminary results indicate movies with more positive and emotionally resonant descriptions tend to receive higher than average votes.
- Contrary to this point is placement of “Mystery” with the highest negative description but scored 3rd.

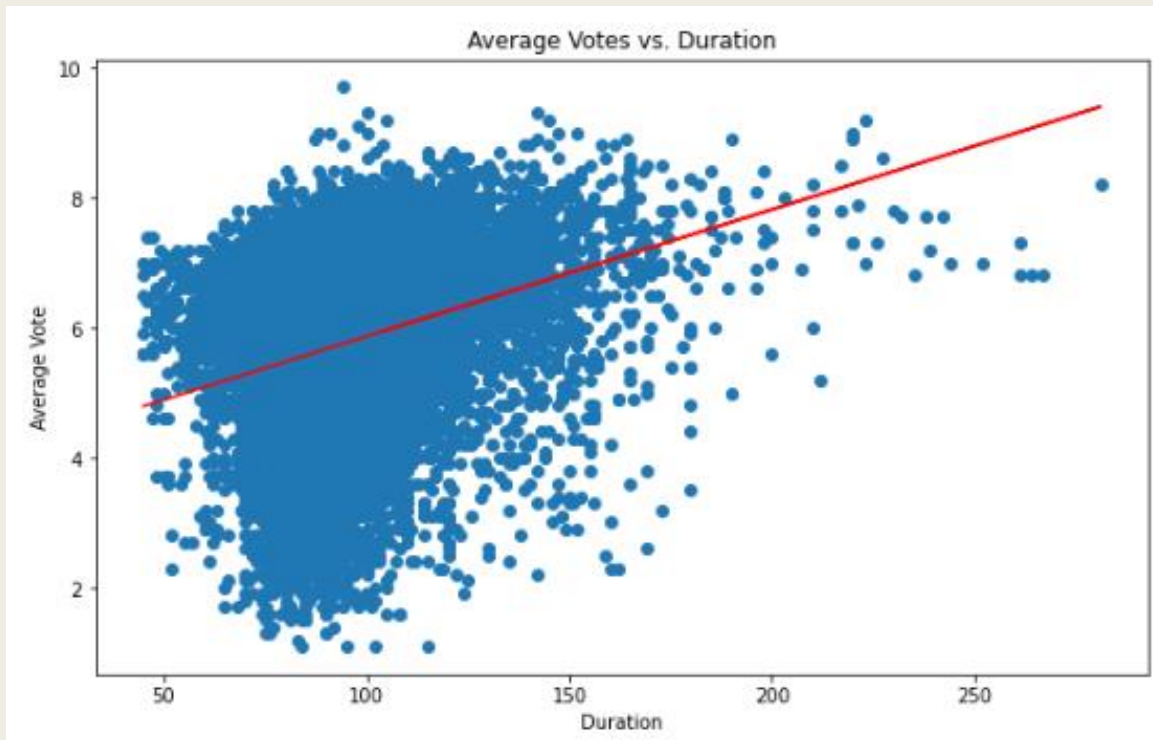


Plot Description Sentiment Score



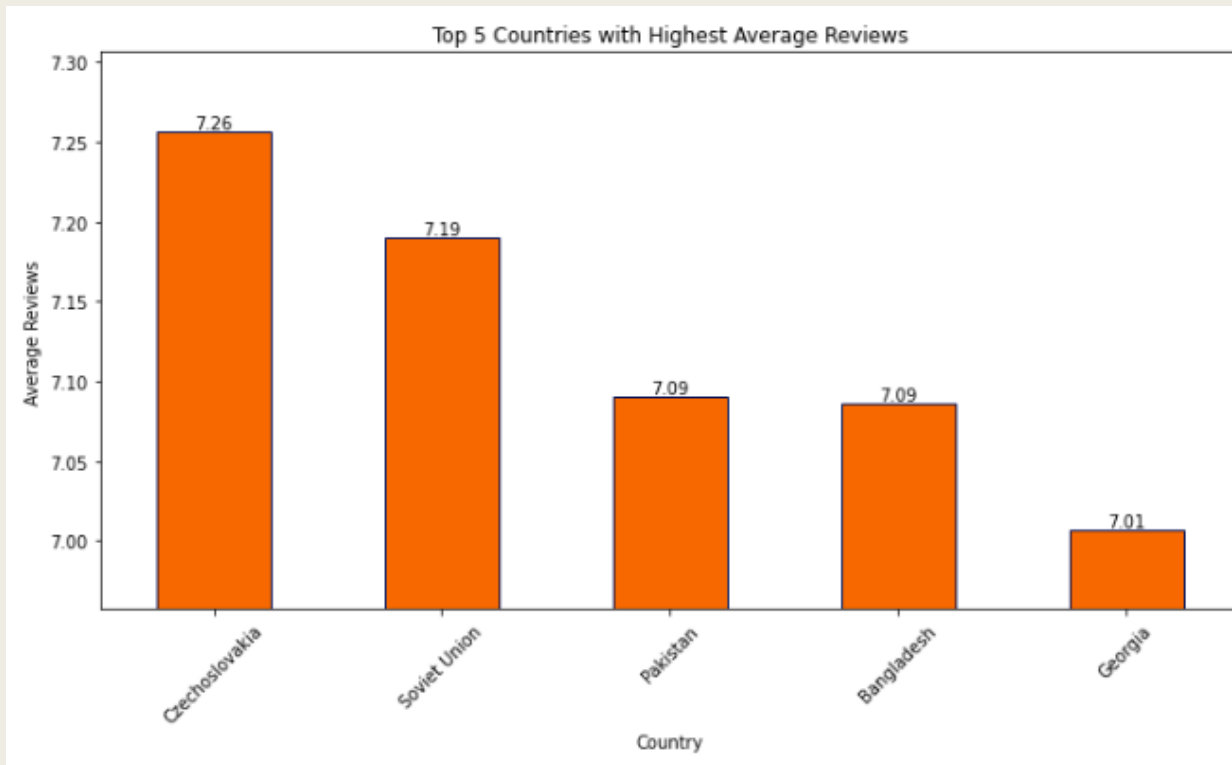
- Additional supporting evidence identifying the average sentiment score, color-coded by Positive/Negative alignment.
- The sentiment score aligns generally with the highest rated genres, but as previously identified, having a positive sentiment score does not guarantee increased voting scores.

Comparison of Avg Vote against Runtime



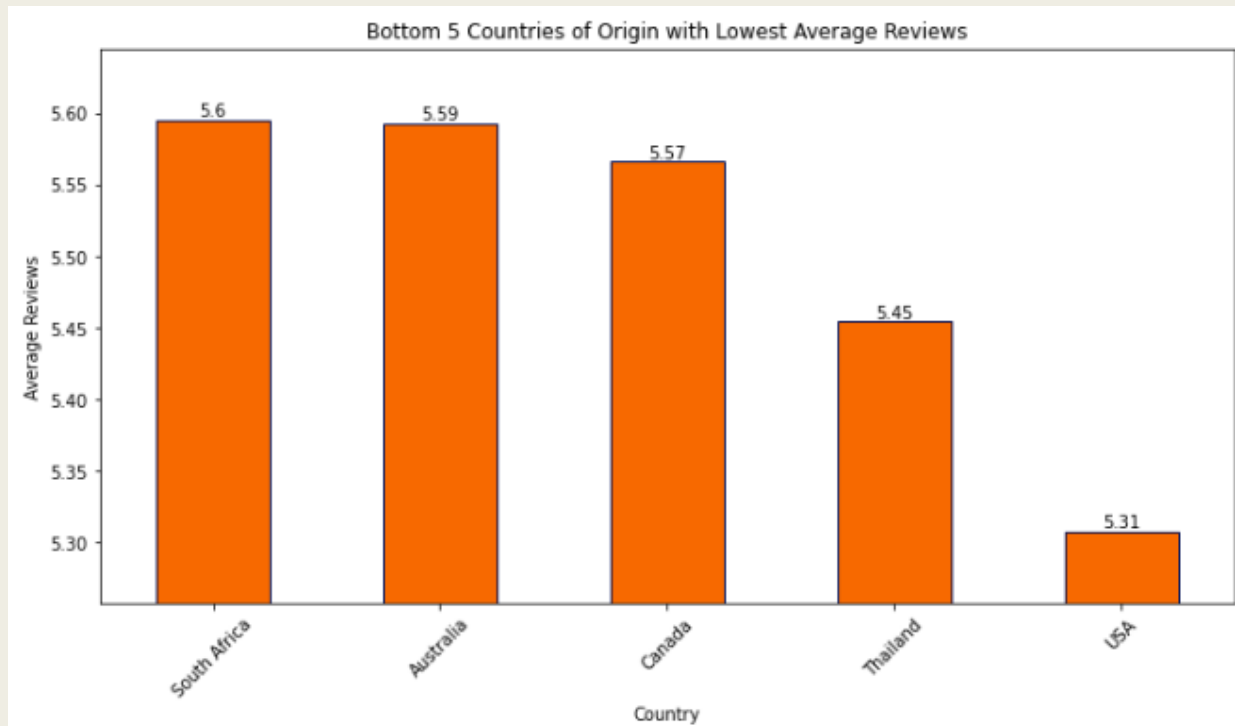
- Linear trend analysis shows a generally upward slope as duration increases.
- Individual genre analysis against runtime demonstrated similar trends in all areas.
- Additional mathematical analysis must be conducted to identify the optimal duration by genre.

Identify top 5 Countries of Origin



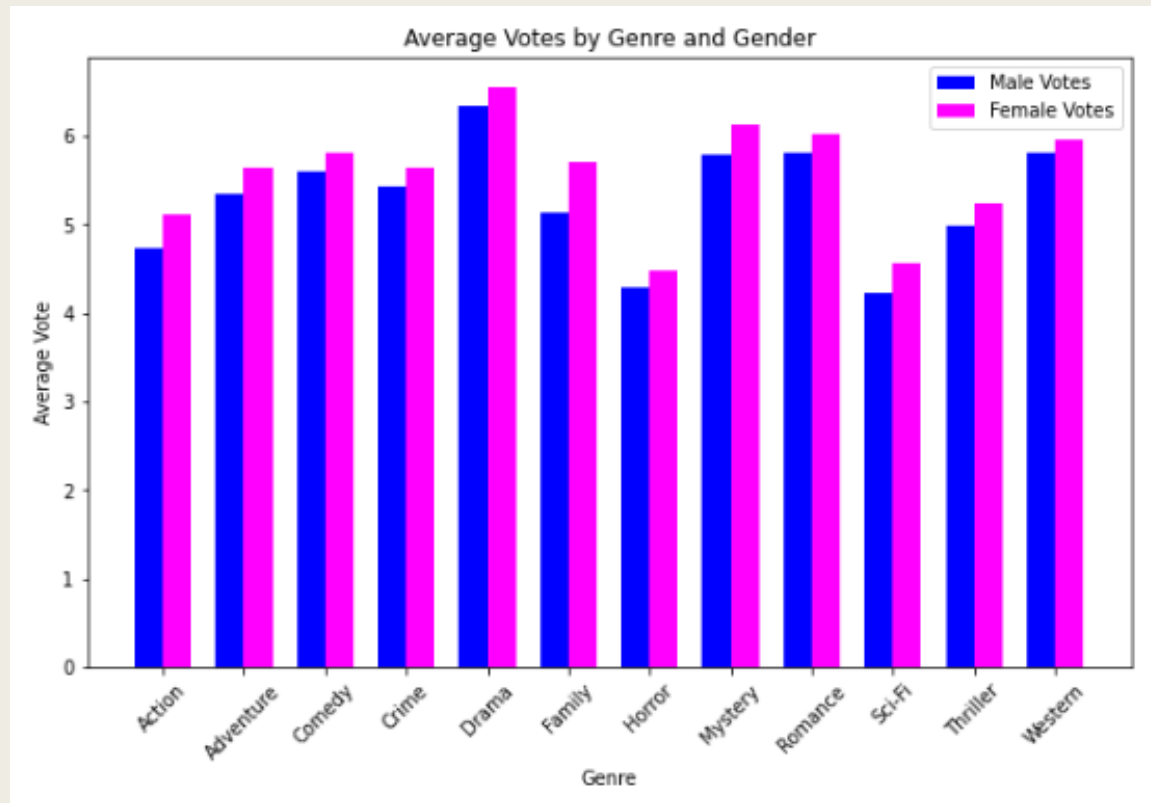
- EurAsian films tend to receive higher average scores than other areas.
- These findings raise questions about the influence of cultural backgrounds, storytelling styles, and regional audience preferences on movie ratings.

Identify bottom 5 Countries of Origin



- The lowest average score belongs to the USA, significantly lower than the surrounding film industries.
- It is important to note that the total number of films from the USA (~7,300 of 19,252, ~38%).
- On a scale of 1-10, the ideal mean would be 5; as the sample set (USA) increases to meet the population size, we should expect results to trend toward the theoretical mean as the number of films increases.

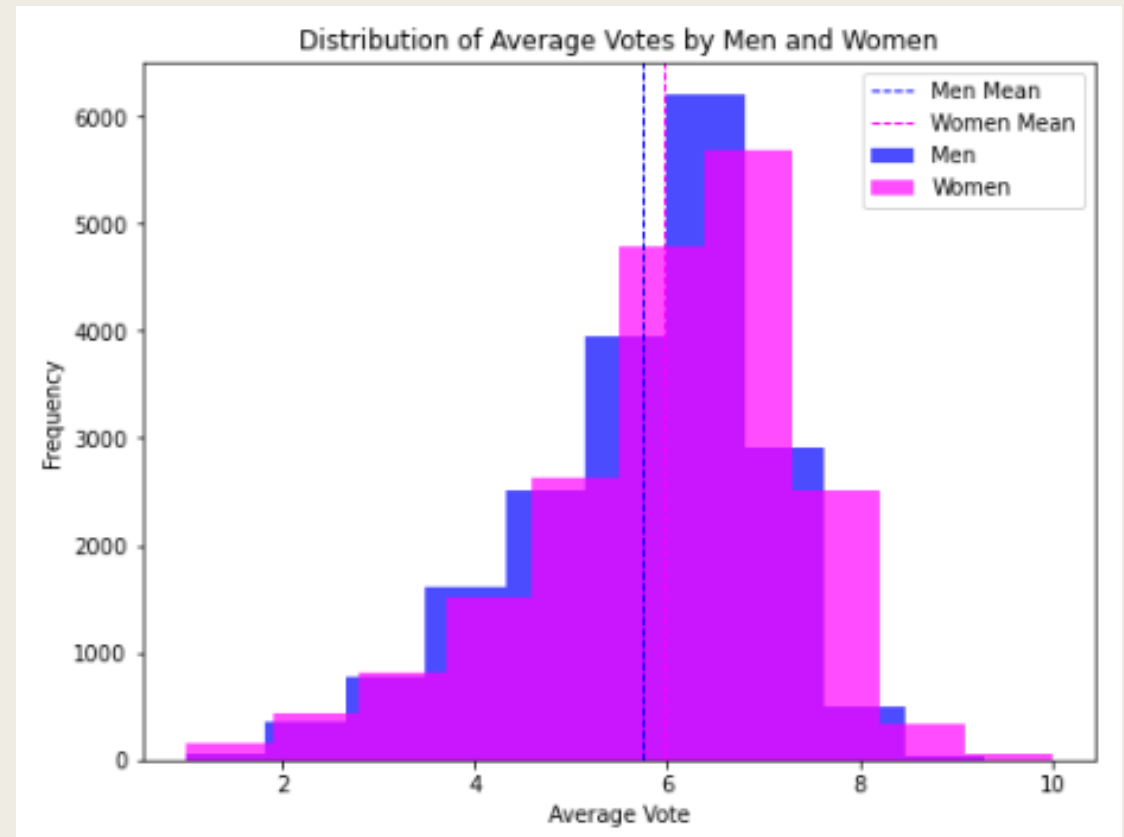
Gender-based voting trends by Genre



- In every instance, Females provide higher average voting scores.
- Save the “Family” genre, the disparity between Male and Female votes maintains a relatively consistent ratio.
- Males tended to return lower voting scores regardless of viewer interaction level (highlighted in later slide).

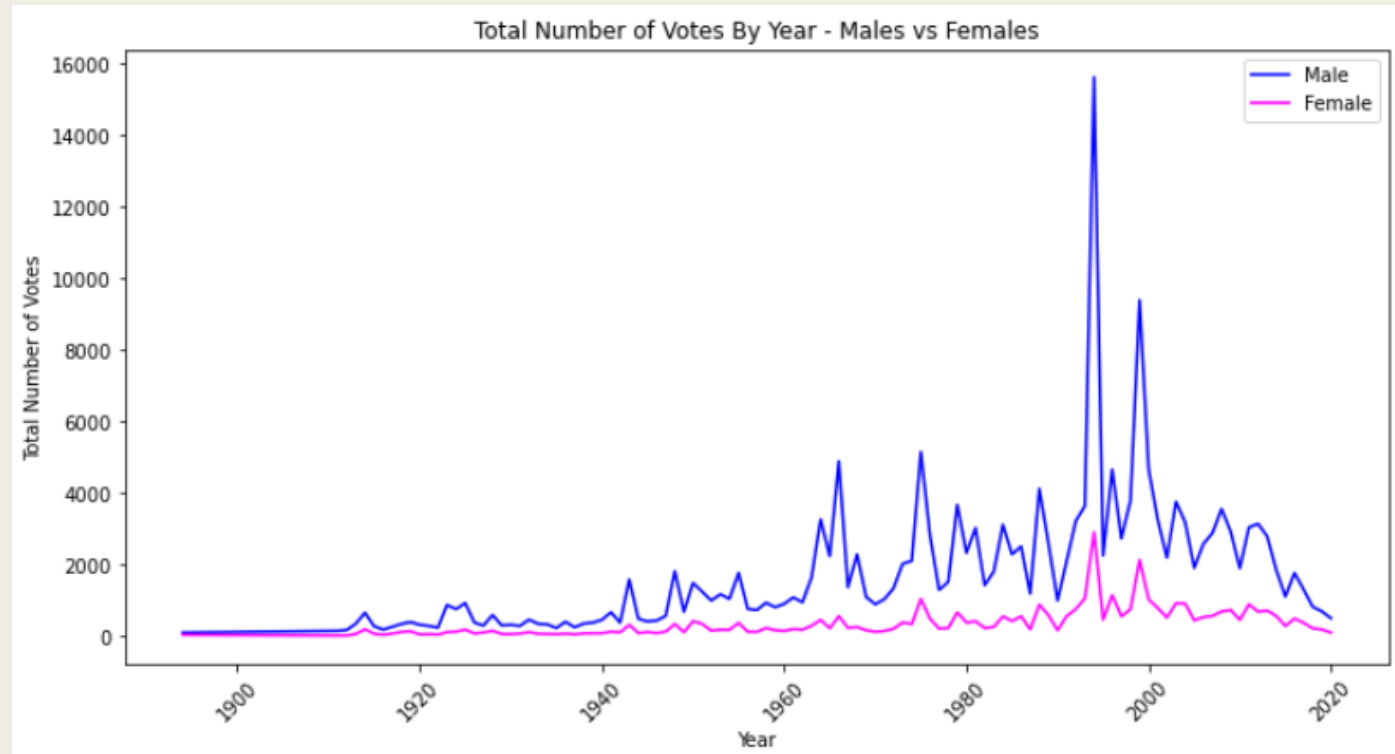
Additional Gender-based voting trends

- Statistical analysis reinforces that Female viewers provided higher voting scores than Males.
- While Males tend to have more interaction and responses, they trend toward lower scores regardless of genre.



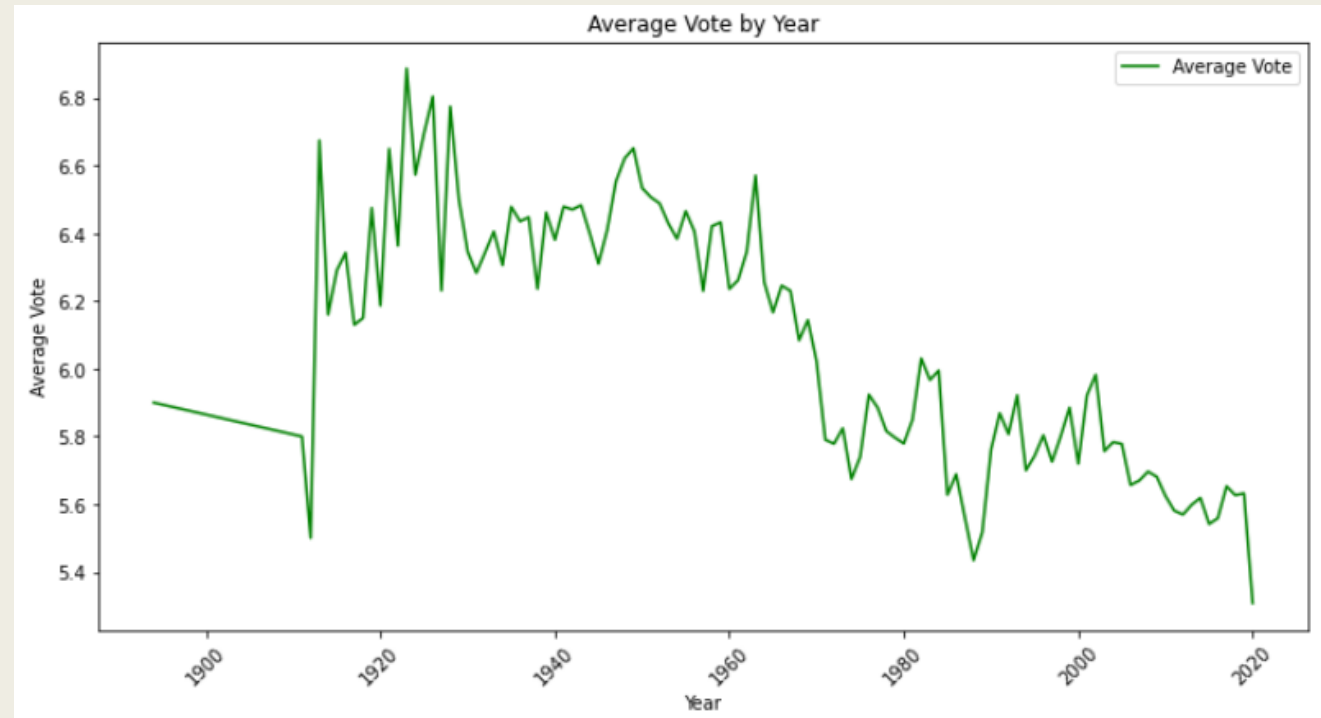
Gender voting trends by Film Production Year

- Despite providing lower average voting scores than Females, Males surpass the total number of votes across the entire duration of film production.
- Particularly starting in the 1960s, with a spike in the late 1990s, total votes increased demonstrating a more interactive viewer base and more active film industry.



Average Score trends by Production Year

- Generally, 1920 to 1960 highlights a golden era in film production.
- Higher scores emphasize either the smaller number of films that are more susceptible to outlying values, or limited input to the film industry due to global activity during those times (WWII, Korea, Vietnam).
- Average score generally decreases as production year increases.



Future Research Recommendations

- Integrate Awards database with ratings to provide correlation of viewer ratings and award success.
- Use integrated Awards database to identify correlation between Director, Production Studio, viewer response, etc.
- Identify optimal film duration compared to average voter score by genre.
- Budget/revenue analysis by genre; return on investment.

Conclusions

- Our analysis highlights the significance of genre in shaping movie ratings and voting patterns. Certain genres, such as Drama and Comedy, attract a larger number of votes and receive higher average ratings, indicating their popularity and appeal among audiences. The lower average votes for Sci-Fi and Horror genres present opportunities for improvement in catering to audience preferences. The gender-based voting disparities emphasize the importance of considering the diverse perspectives and preferences of viewers when creating and marketing movies.
- By examining the significance of genre, runtime, country of origin, and plot sentiment, we offer actionable information for the film industry to better understand audience preferences and create more successful and engaging movies. These insights can lead to improved marketing strategies, content development, and overall audience satisfaction, ultimately driving the success of movie productions.



QUESTIONS?