Genre Analysis: Investigate trends in film genres over the years and their correlation with audience ratings and box office performance.

By Edwin Mutendwa

Objectives

Genre Trends Over Time: Explore how the popularity of different film genres has evolved over the years.

Correlation Between Genres, Ratings, and Box Office Performance: Investigate the relationship between film genres, audience ratings, and box office success.

Impact of Genre on Filmmaking Decisions: Discuss how the analysis of genre trends can inform filmmaking decisions.

Data Summary

	runtime_minutes	averagerating	numvotes	domestic_gross
count	2951.000	2951.000	2951.000	2.951000e+03
mean	107.312	6.466	63208.184	3.069066e+07
std	20.044	0.994	126754.906	6.709869e+07
min	3.000	1.600	5.000	1.000000e+02
25%	94.000	5.900	2500.000	1.375000e+05
50%	105.000	6.600	13907.000	2.000000e+06
75%	118.000	7.100	66744.000	3.245000e+07
max	272.000	9.200	1841066.000	7.001000e+08

Missing Values for merged dataset

Percentage of Missing Values in Each Column

No outliers reported

Box Plot of Runtime, Average Rating, Num Votes, and Domestic Gross
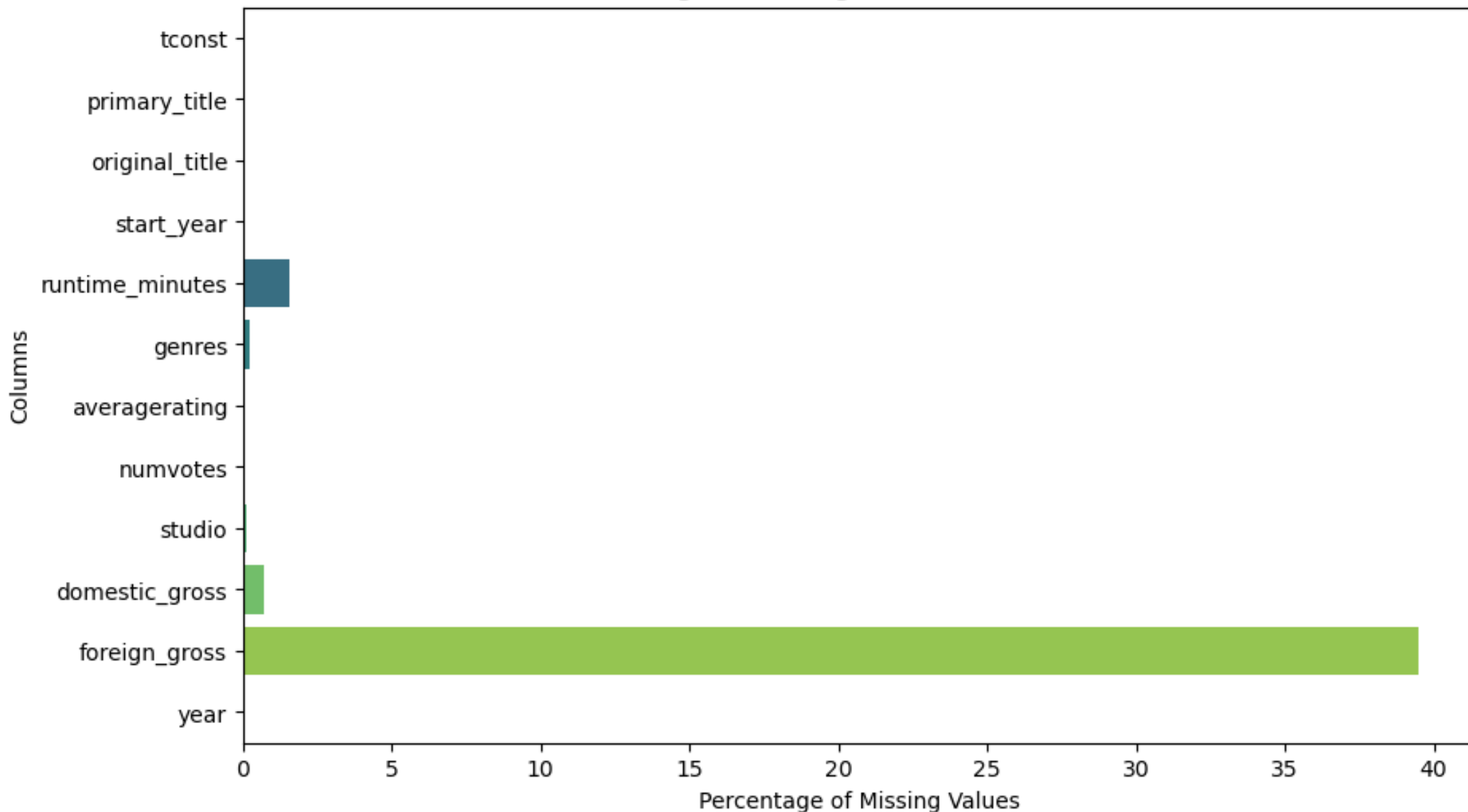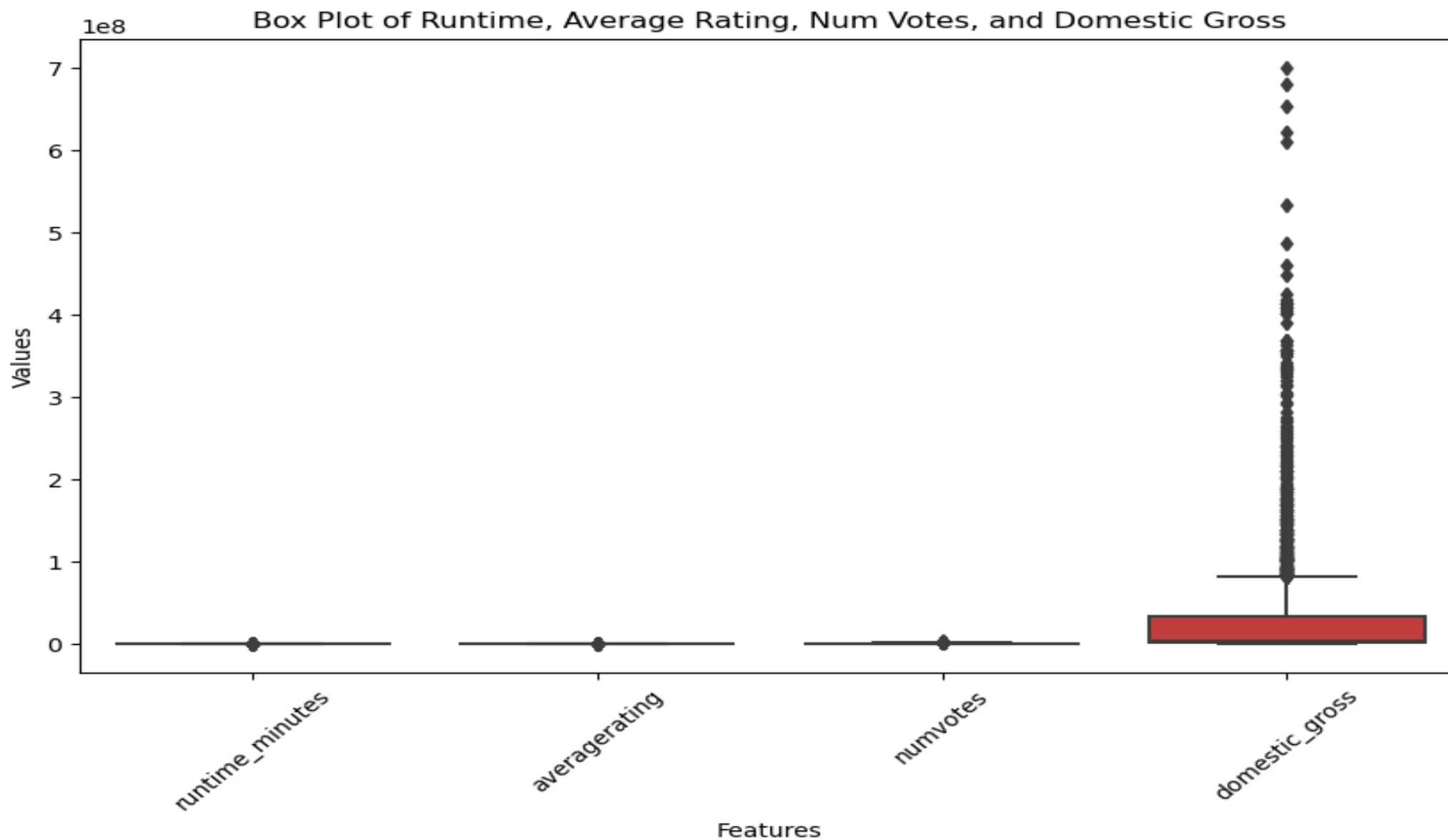
Crucial decisions during data cleaning

Dropped foreign gross column: More than 40 percent of data was missing

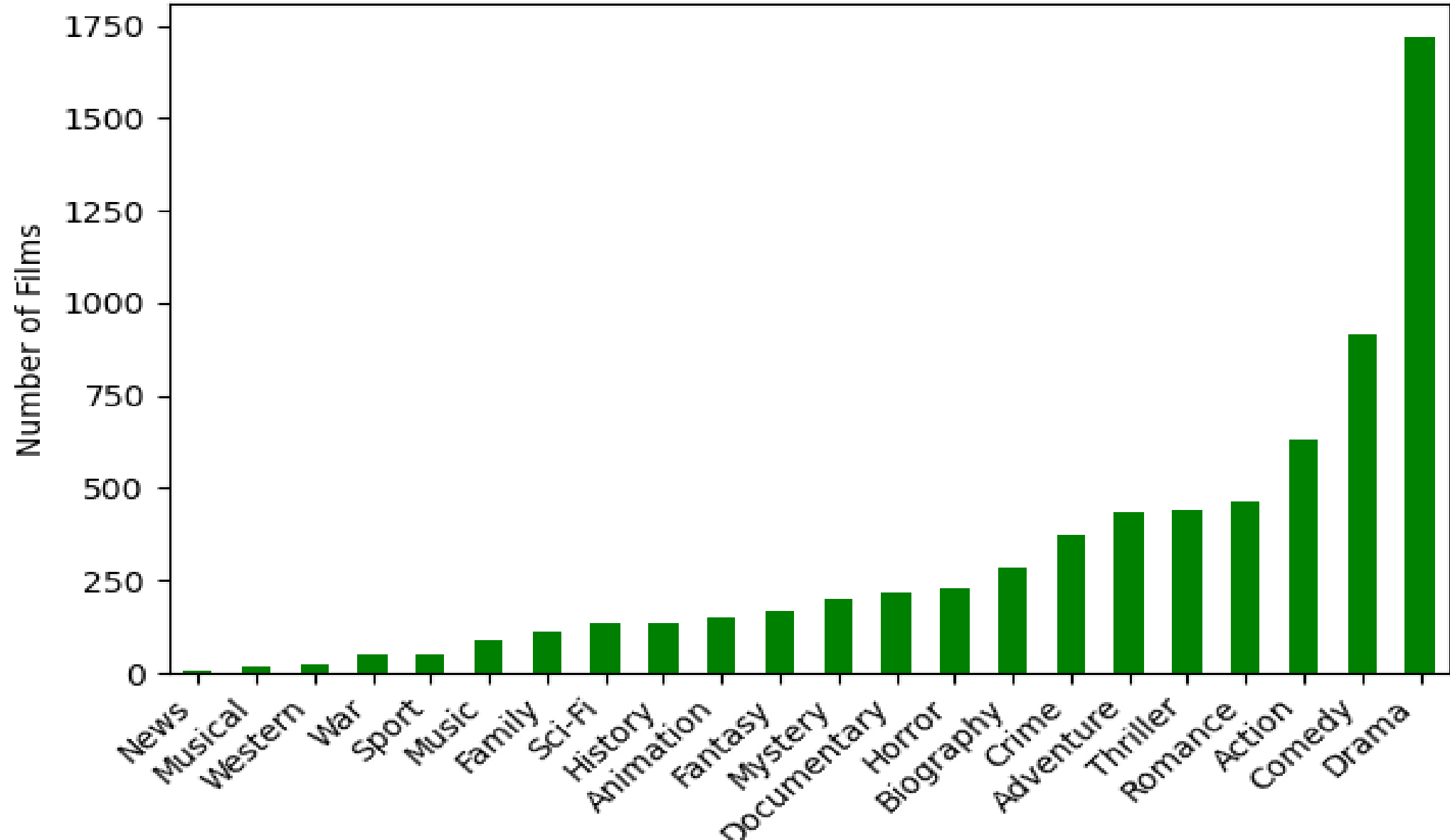Dropped runtime_minute rows with missing data: Less than 0.1 of data was missing

Dropped genre rows with missing data: Less than 0.1 of data was missing

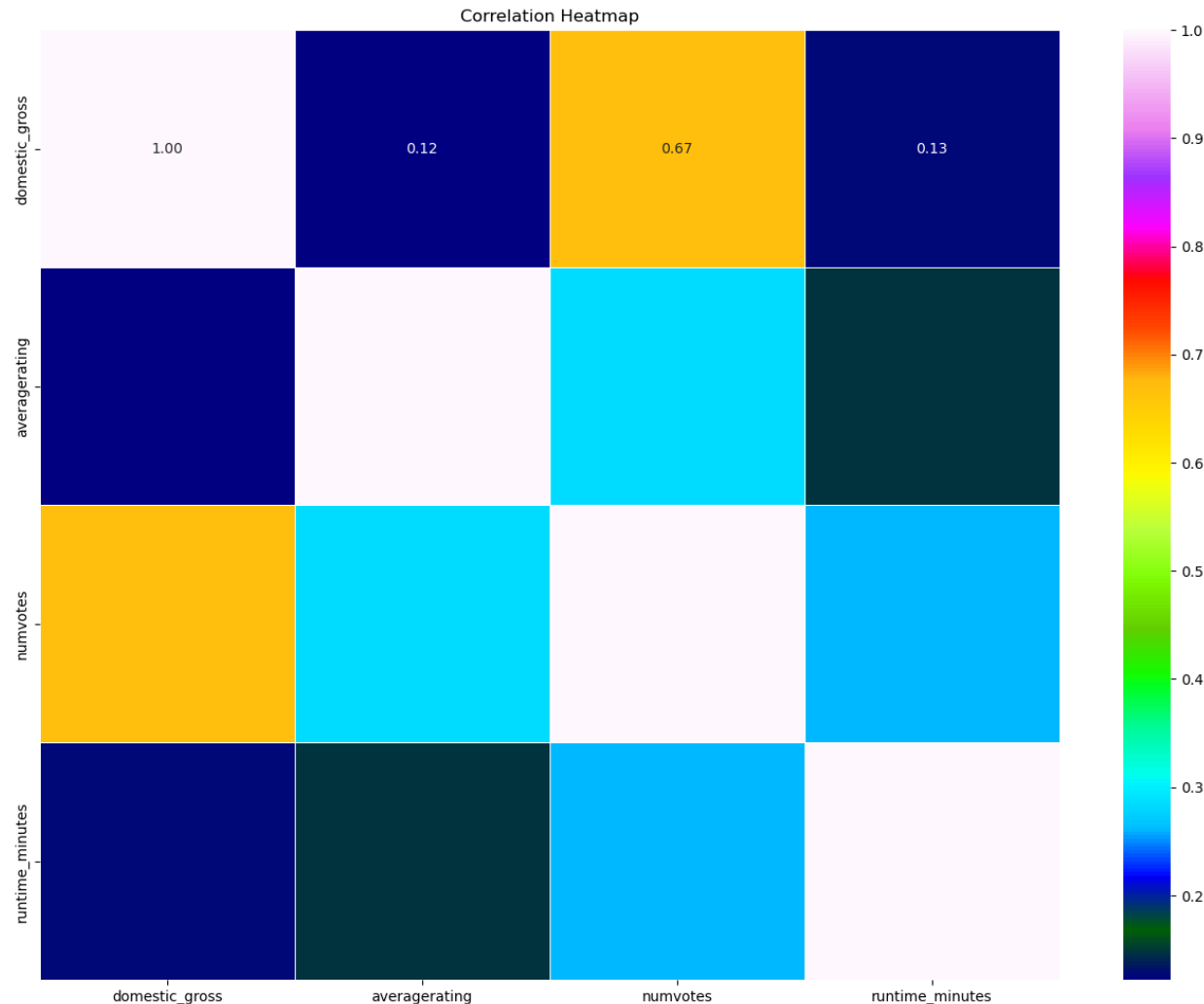Dropped domestic_gross rows with missing data: Less than 0.1 of data was missing

Data Analysis

# Genre Analysis

Distribution of Films Across Different Genres

Correlation between domestic gross, average rating, number of votes, run time minutes


Correlation Heatmap

- The data suggests that films with higher domestic gross tend to attract more votes from the audience, indicating a potential relationship between commercial success and audience engagement.

- While there is a positive correlation between average rating and other metrics, it is relatively weaker compared to the correlation between domestic gross and number of votes.

- Longer runtime does not necessarily guarantee higher ratings or gross earnings, but there is a slight positive correlation between runtime and audience engagement, as reflected in the number of votes.

Movie With highest and lowest average rating

Top 10 Performing Studios by Total Domestic Gross
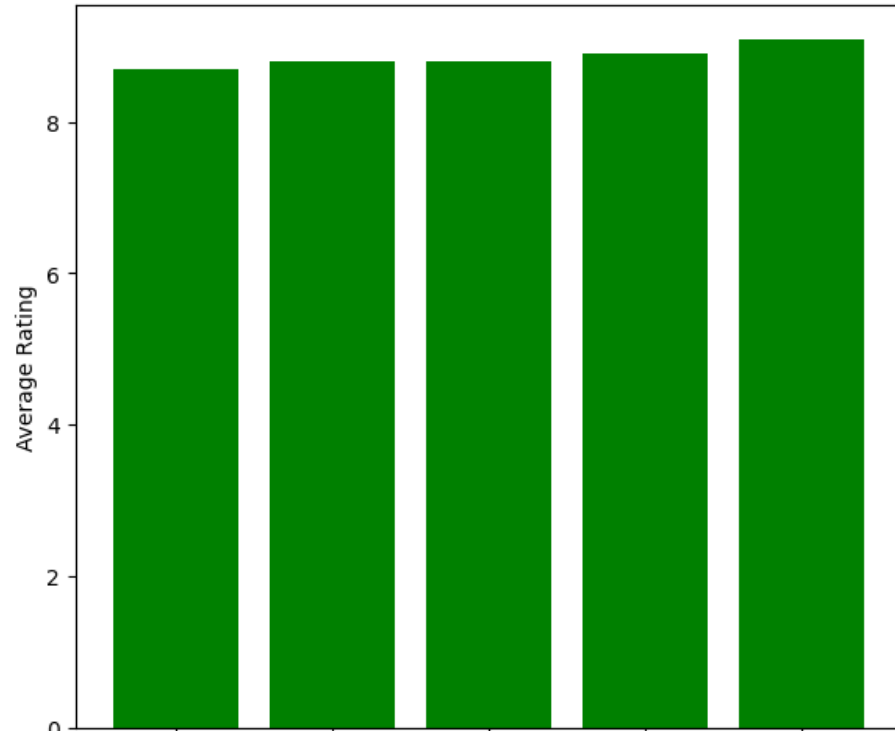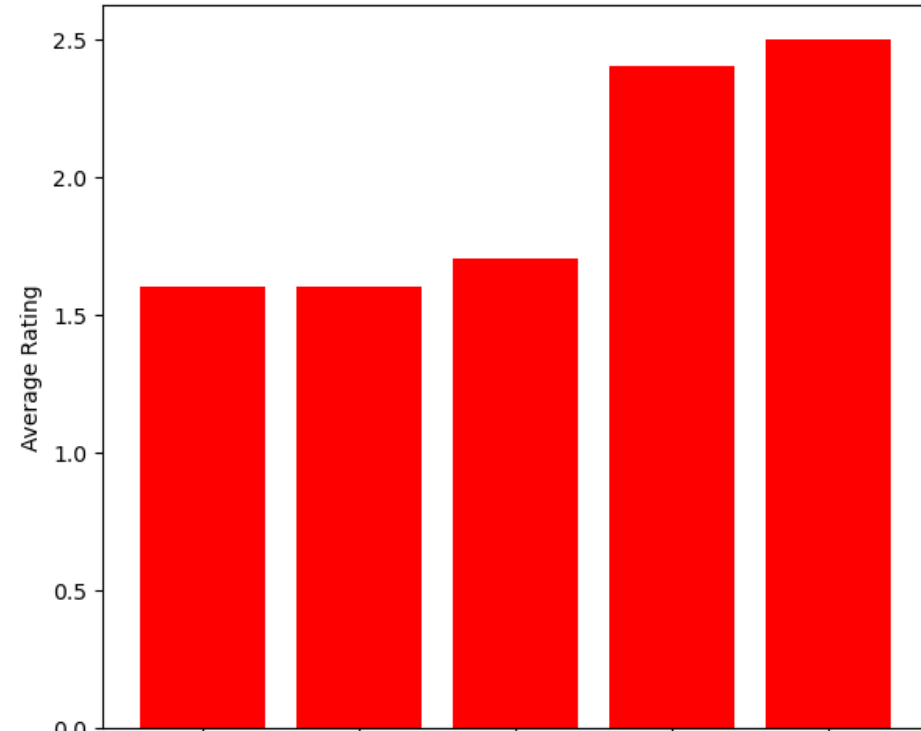
Top Performing Studios by Total Domestic Gross
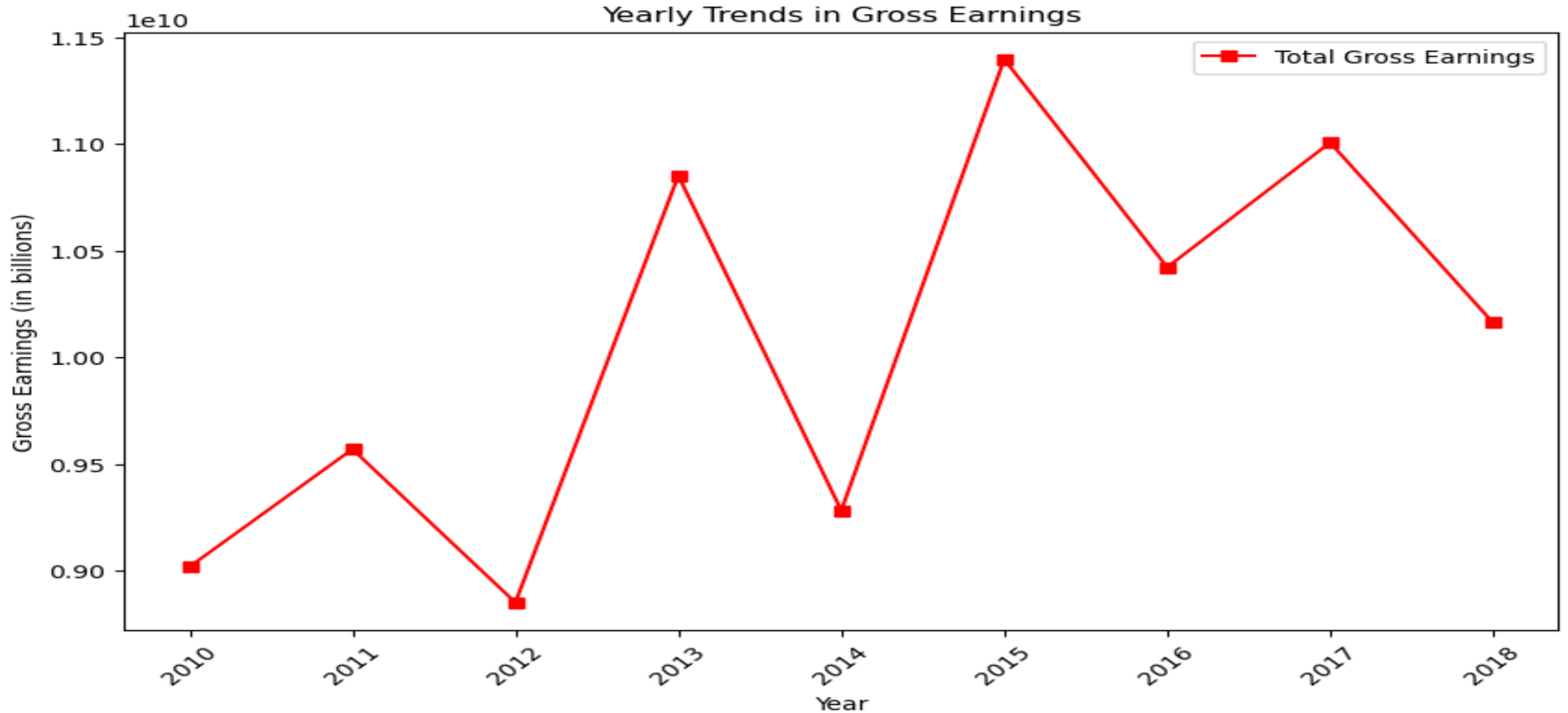
Top 10 Rated Studios

Top Rated Studios by Average Rating
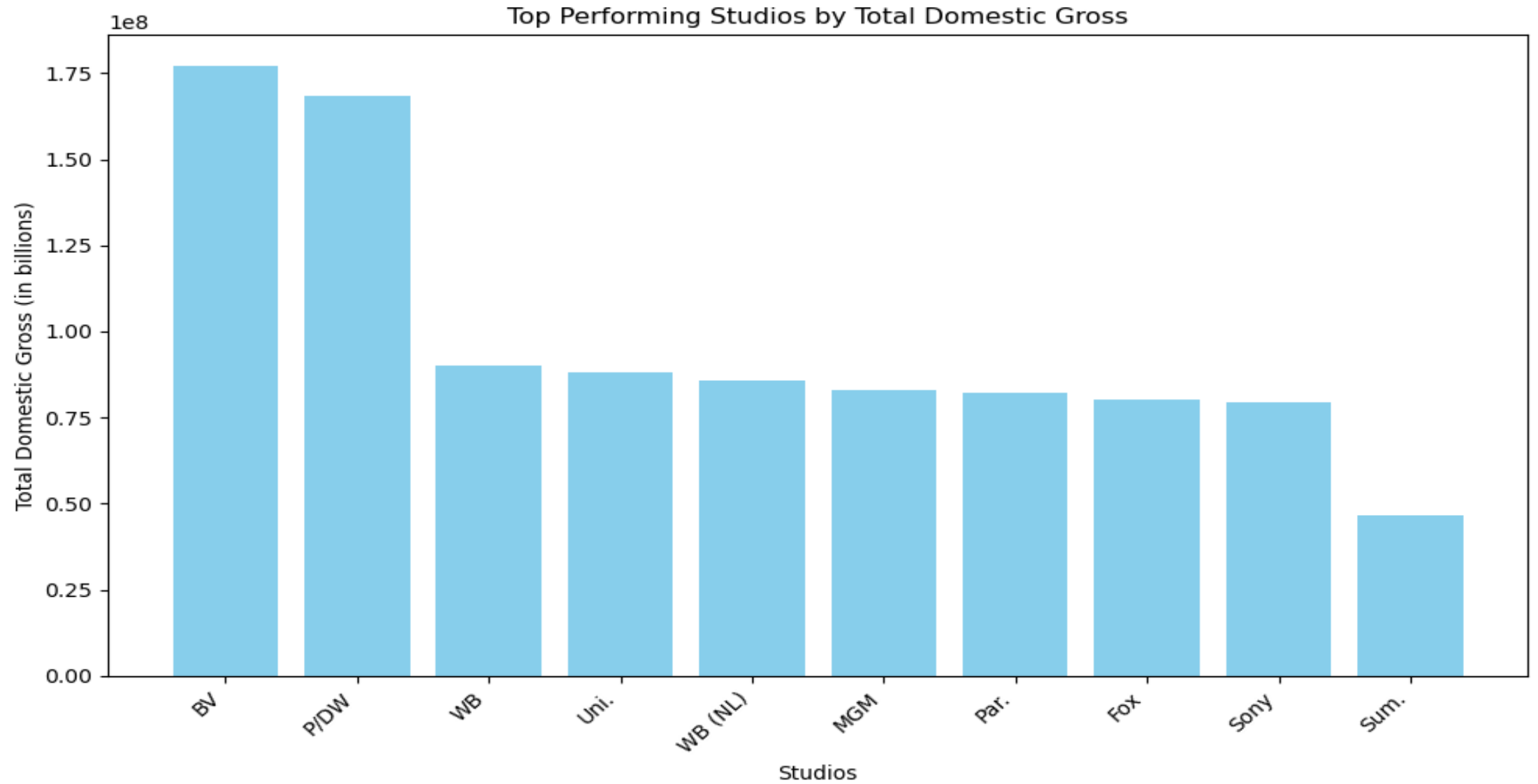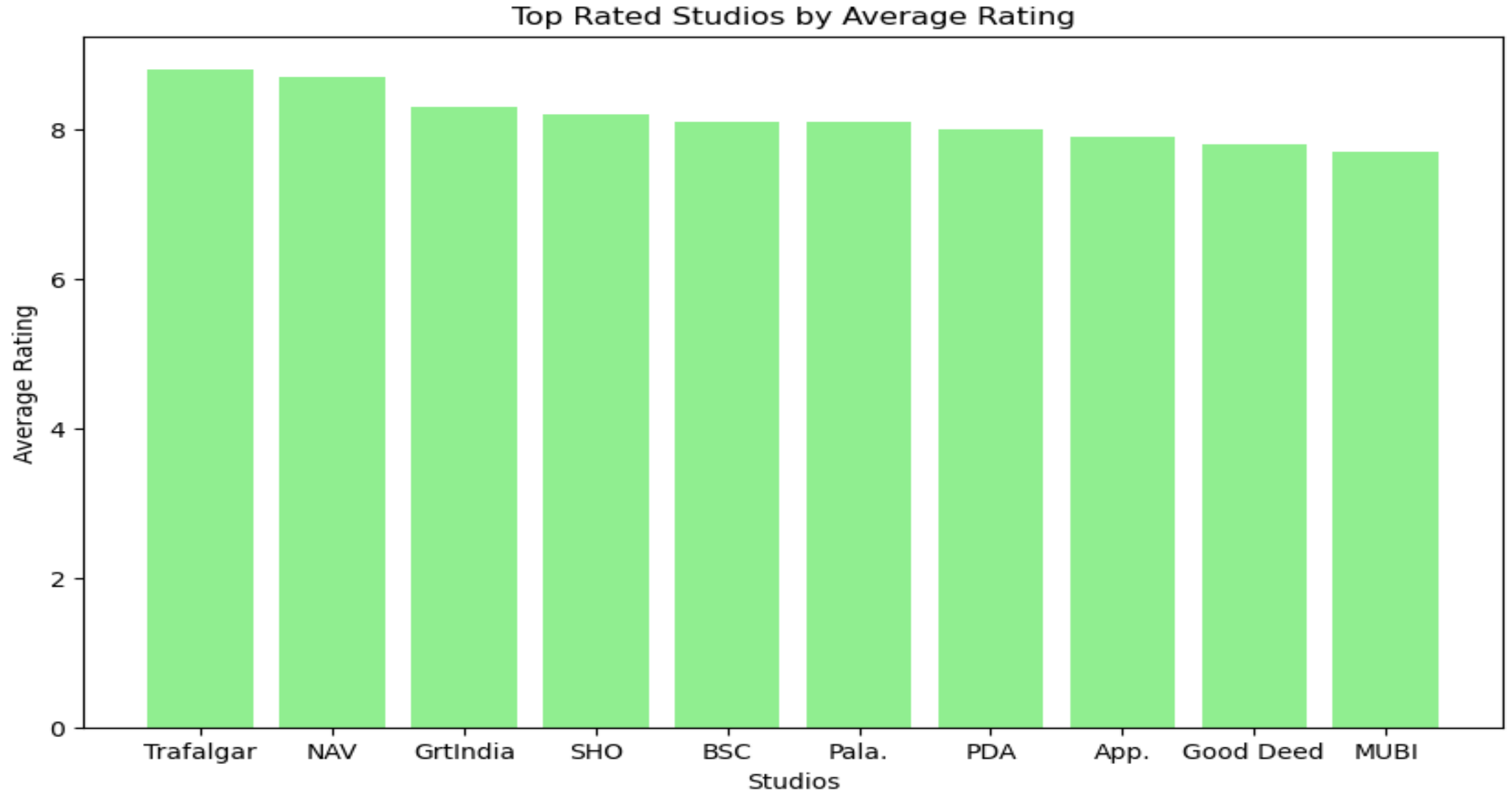
Movie with the highest and lowest domestic gross
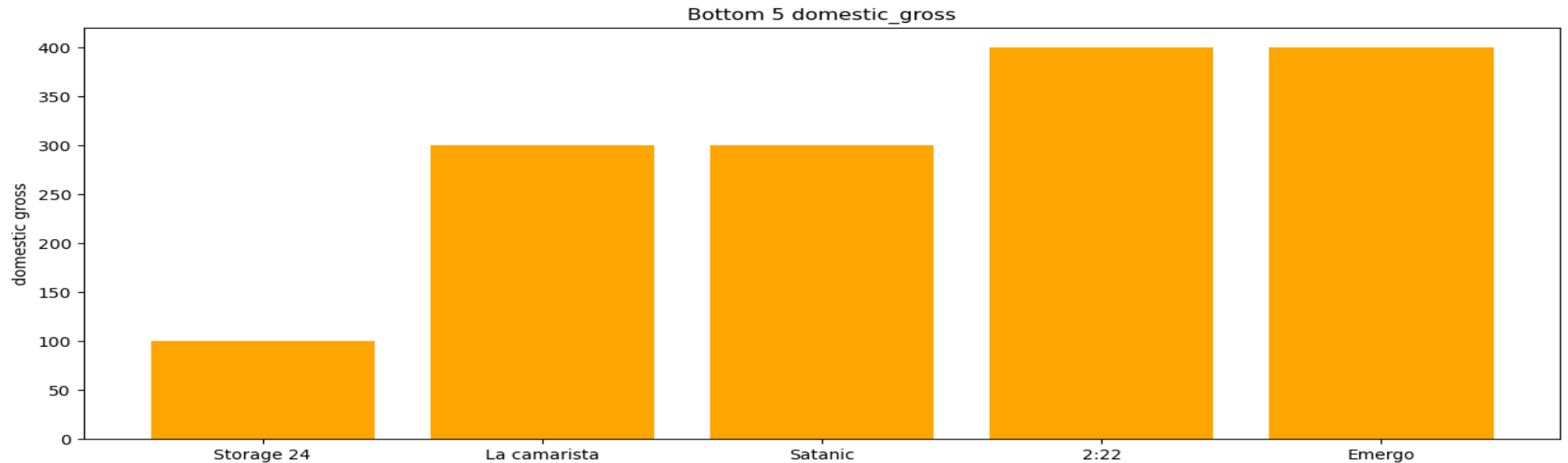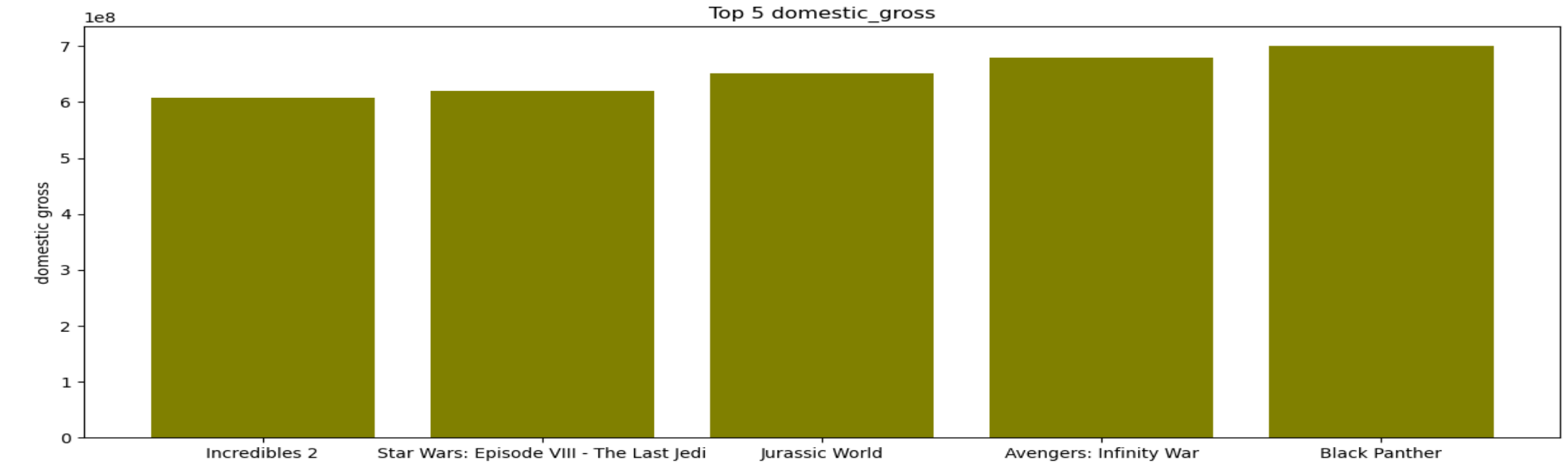
Yearly trends in gross earnings

Regression

Regression between runtime vs number of votes
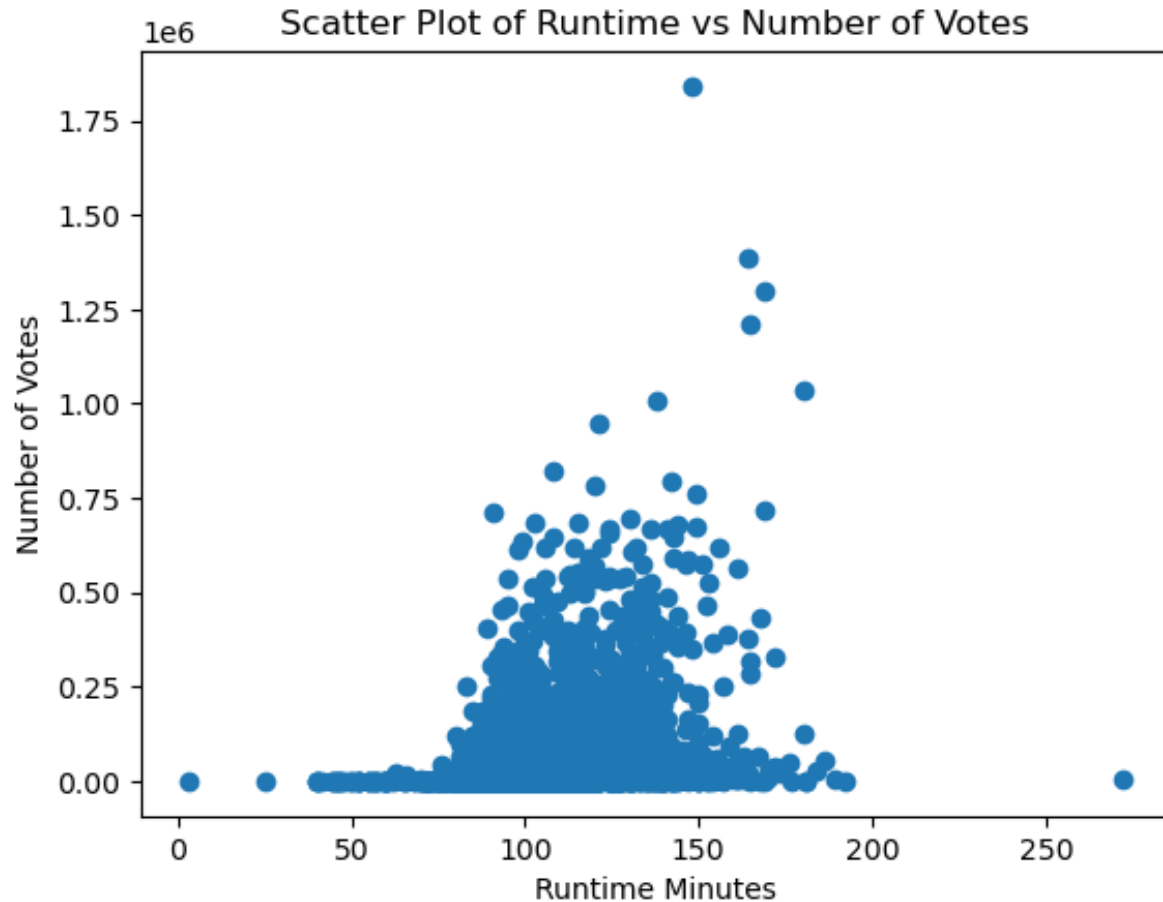
Scatter Plot of Runtime vs Number of Votes

- **Slope (1.80438938e-09)**: The slope of the regression line suggests that for every unit increase in domestic gross earnings, the number of votes increases by approximately 1.80438938e-09.

- **Intercept (6.41029475e+00)**: The intercept represents the predicted number of votes when the domestic gross earnings are zero.

Coefficients of the linear regression model (slope, intercept): [1.80438938e-09 6.41029475e+00]

Regression of gross vs average rating

- **Slope (1.80438938e-09)**: The slope of the regression line suggests that for every unit increase in domestic gross earnings, the average rating increases by approximately 1.80438938e-09

- **Intercept (6.41029475e+00)**: The intercept represents the predicted average rating when the domestic gross earnings are zero.

Scatter Plot of Domestic Gross vs Average Rating
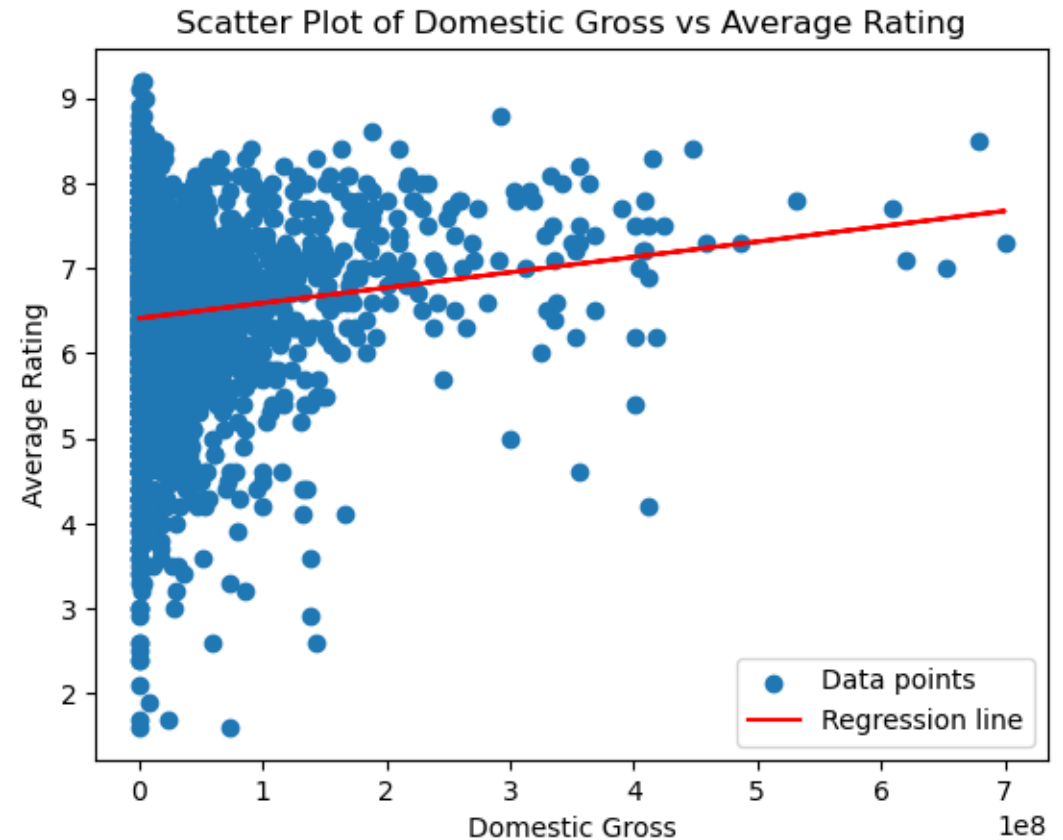
Coefficients of the linear regression model (slope, intercept): [1.8043838e-09 6.41029475e+00]

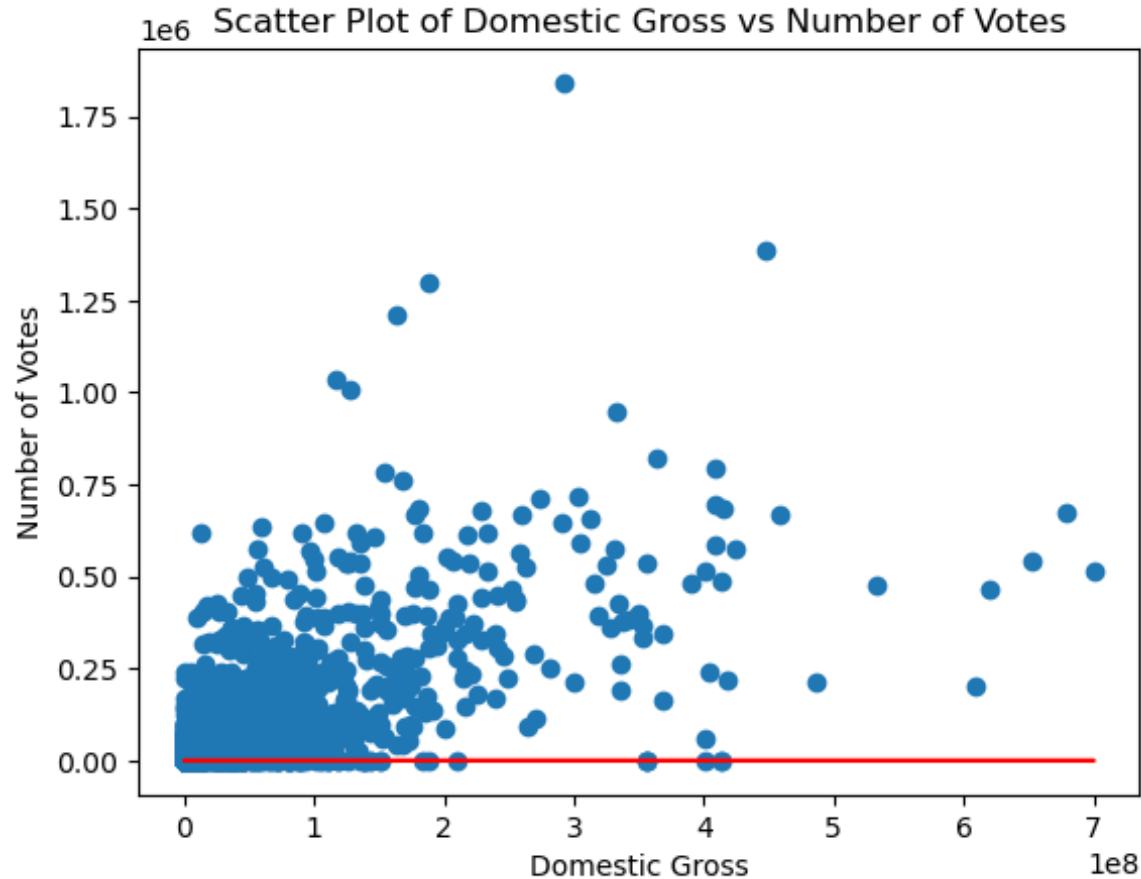# Regression of domestic gross vs number of votes

- The nearly flat slope of the regression line suggests a very weak correlation between domestic gross earnings and the number of votes. This is supported by the small slope coefficient (1.80438938e-09).

- The intercept coefficient (6.41029475e+00) represents the predicted number of votes when the domestic gross earnings are zero, which is approximately 6.41 votes.

Coefficients of the linear regression model (slope, intercept): [1.80438938e-09 6.41029475e+00]

Recommendations

Recommendation and Conclusion

Recommendation	Explanation
Focus on Audience Engagement Strategies	Given the weak correlation between domestic gross earnings and the number of votes received by movies, it's evident that box office success alone does not guarantee high audience engagement. Therefore, businesses in the film industry should prioritize strategies aimed at increasing audience engagement beyond ticket sales.
Diversify Revenue Streams	Instead of relying solely on box office revenue, film businesses should explore diversifying their revenue streams. While box office earnings remain an important metric, the analysis indicates that factors other than domestic gross, such as average rating and runtime, also influence audience behavior.
Invest in Quality Content and Storytelling	Although the analysis suggests a weak correlation between domestic gross earnings and average rating, the importance of producing quality content and compelling storytelling cannot be overstated. Businesses should prioritize investments in developing high-quality, engaging content that resonates with audiences.

Thank you