Movie Industry Analysis

GROUP 1 PHASE 2 PROJECT

OVERVIEW OF THE PROJECT

❖The project focuses on analysing movie datasets to provide actionable insights for the creation of a new movie studio.

❖Using data visualization and analysis tools, such as Tableau and models

❖The project aims to identify trends and patterns in the movie industry, including top-performing genres, audience preferences and revenue-driving factors.

BUSINESS UNDERSTANDING

❖ The company now sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies.

❖ You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of your company's new movie studio can use to help decide what type of films to create.

DATA UNDERSTANDING

❖ Here will need to understand our data. This involves getting the relevant information from each dataset crucial for our analysis.

❖ We start by loading the various datasets reviewing their various information based on the columns and check which information is necessary for our analysis before beginning the data cleaning.

Movie Performance Data

- **Box Office Data:** Includes revenue by day, week, or year for individual movies. Useful for analysing trends in revenue over time.

DATA UNDERSTANDING Cont……1.

- **Weekend Box Office Data**: Aggregated weekend earnings, helpful for identifying peaks in movie viewership.

- **Cumulative Box Office**: Total earnings over time for each movie.

▶ **2. Movie Metadata**

- **Movie Titles and Release Dates**: Basic information about movies, such as title, genre, release date, and distributor.

- **Genre and Ratings**: Helps in analyzing how genres or ratings affect performance

▶ **3. Home Entertainment Data**

- **DVD/Blu-ray Sales**: Useful for understanding post-theater revenue streams

DATA UNDERSTANDING Cont......1.

- **Weekend Box Office Data**: Aggregated weekend earnings, helpful for identifying peaks in movie viewership.

- **Cumulative Box Office**: Total earnings over time for each movie.

▶ **2. Movie Metadata**

- **Movie Titles and Release Dates**: Basic information about movies, such as title, genre, release date, and distributor.

- **Genre and Ratings**: Helps in analyzing how genres or ratings affect performance

DATA UNDERSTANDING CONT……2

▶ **3. Home Entertainment Data**

• **DVD/Blu-ray Sales**: Useful for understanding post-theatre revenue streams.

• **Digital Sales/Streaming Revenue**: Can show trends in digital distribution.

▶ **4. Market-Level Data**

• **Market Share by Studio**: Helps analyse how different studios perform over time.

• **International Box Office**: Useful for cross-region comparison and global trends.

OBJECTIVES/GOALS

Main Objective

- To analyse movie data and uncover key patterns in revenue, popularity, ratings, and director influence across genres, providing actionable insights for business growth and strategic decision-making.

OBJECTIVES (cont...2)

Specific Objectives

1. Identify the movie genres that generate the highest revenue and analyse the factors contributing to their financial success.

2. Explore audience preferences to uncover the most popular genres and the drivers behind their popularity.

3. Evaluate the ratings of movies across genres to assess trends in critical and audience reception.

4. Examine the role and impact of directors on the success of specific genres.

5. Identify top-performing movies within their respective genres.

6. Highlight the most successful and influential directors based on revenue and popularity.
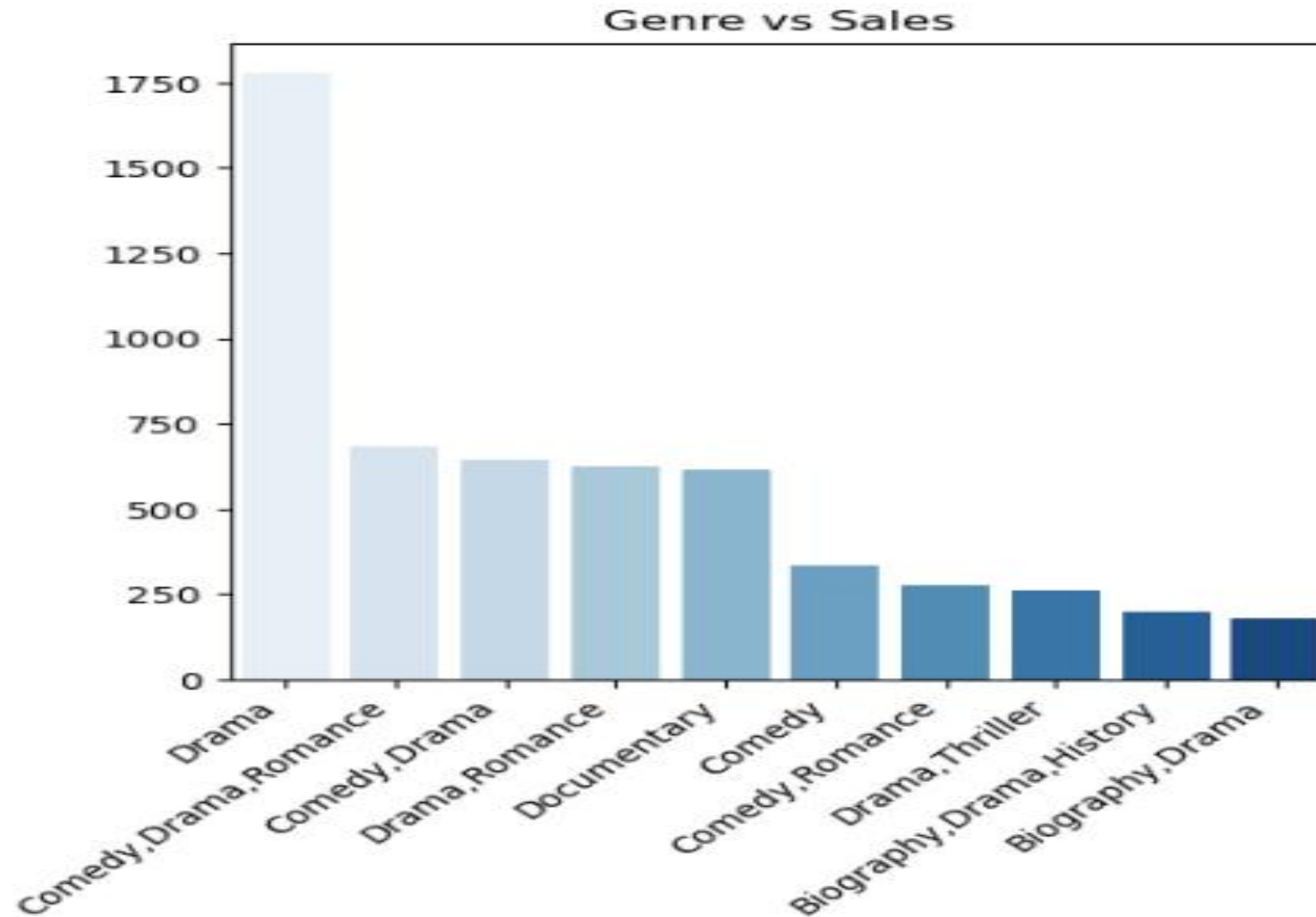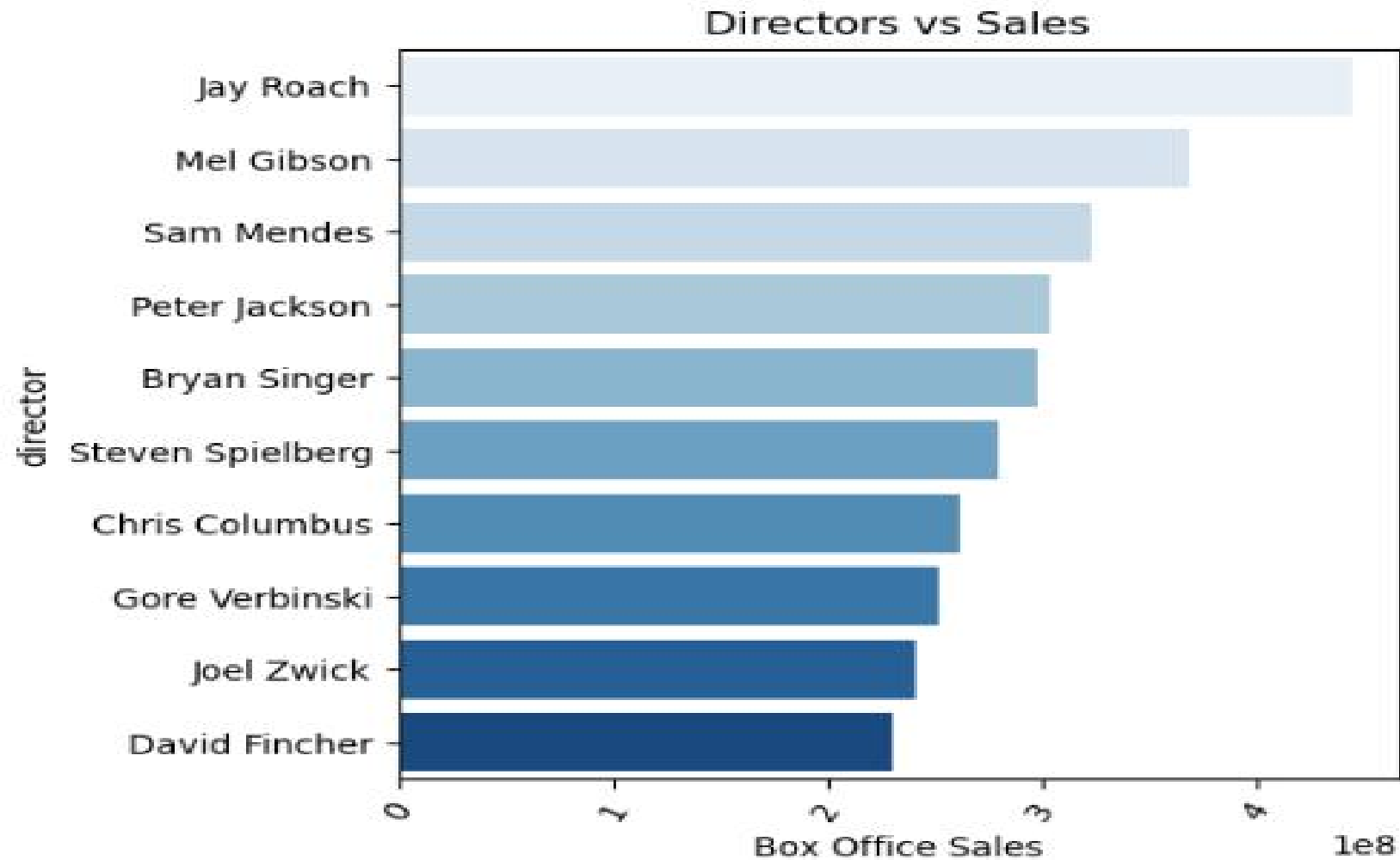
9
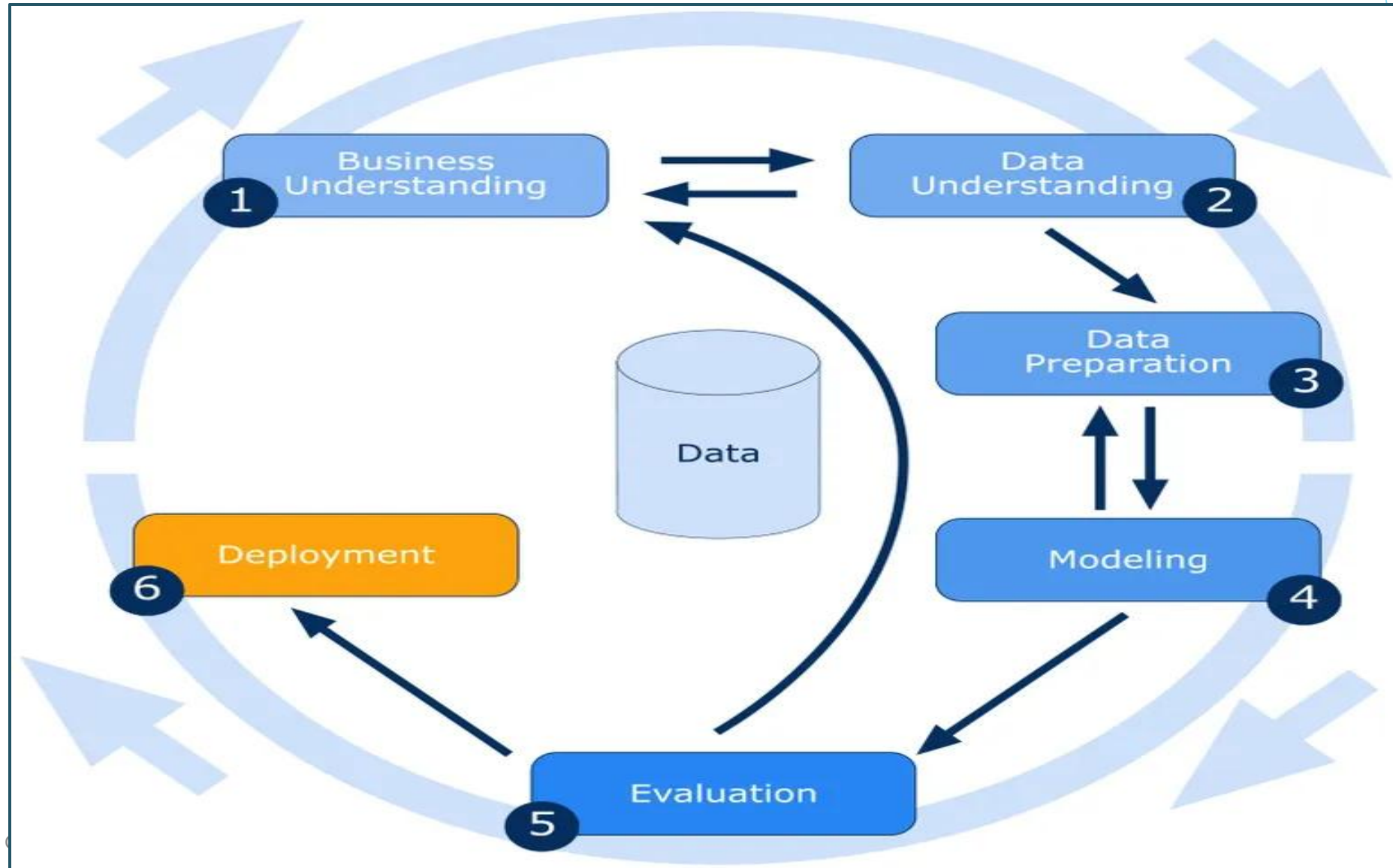
Comparison between genre and rating
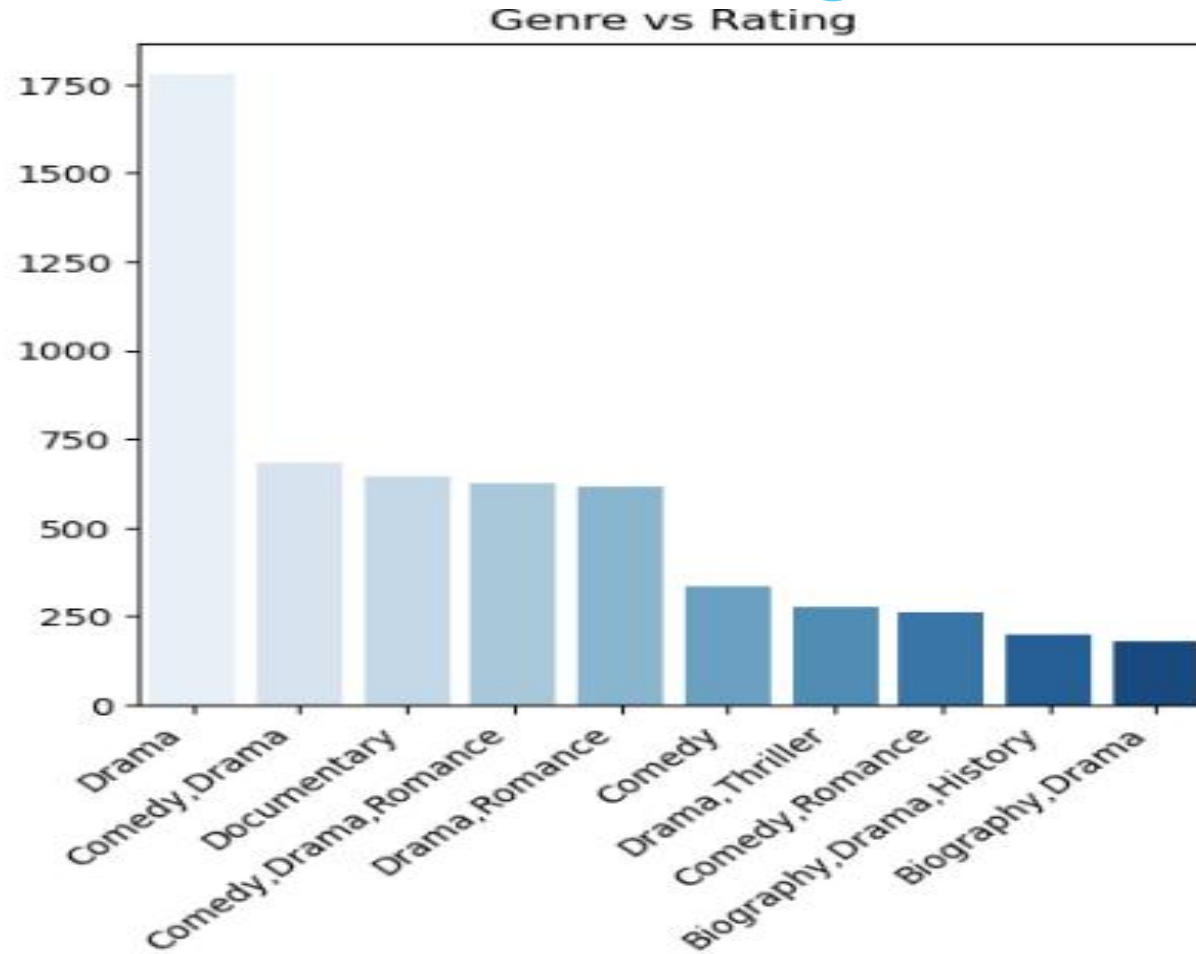
The highly rated genre is drama.
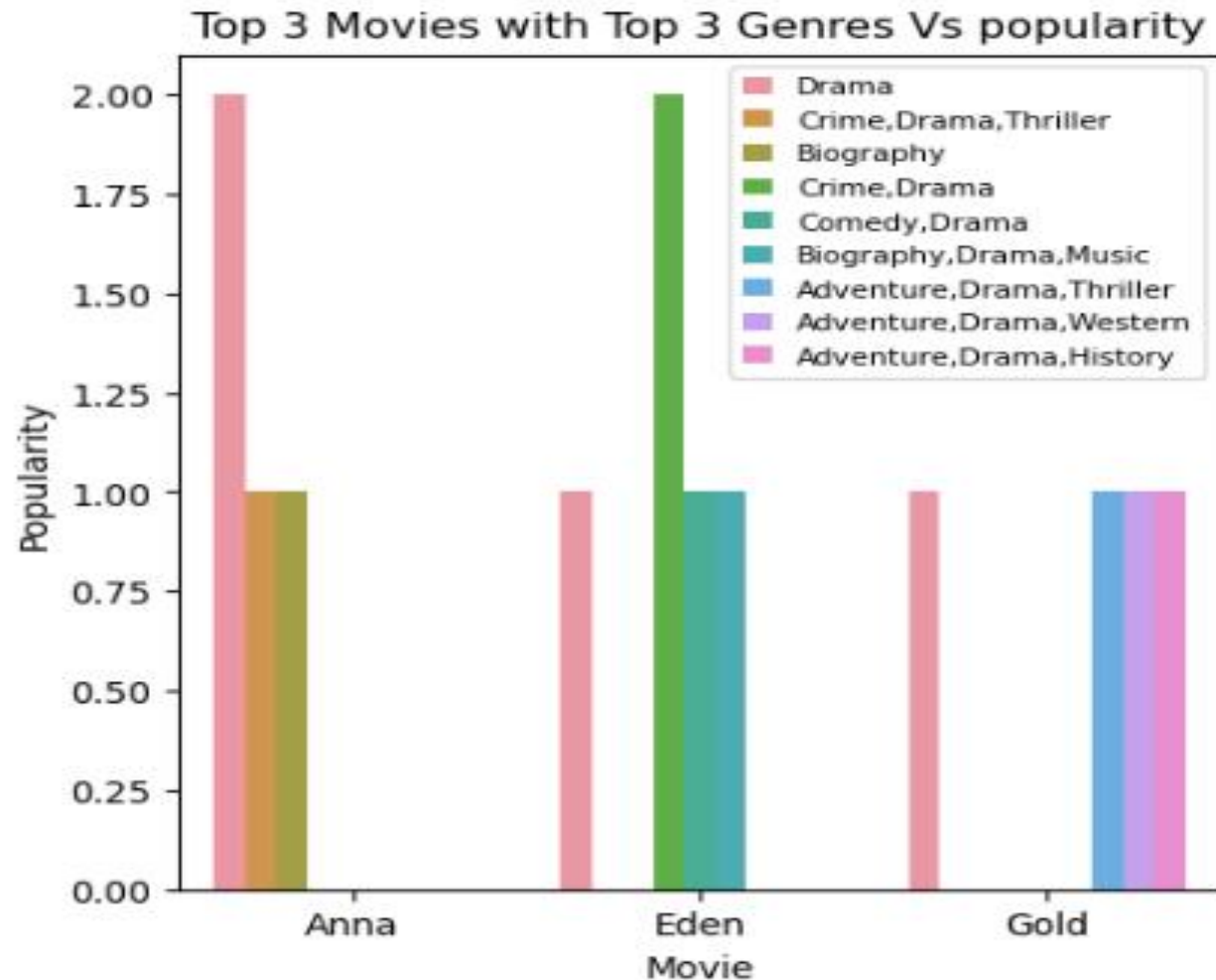
Comparison between genre and sales

Highest sales were from comedy, followed by drama.

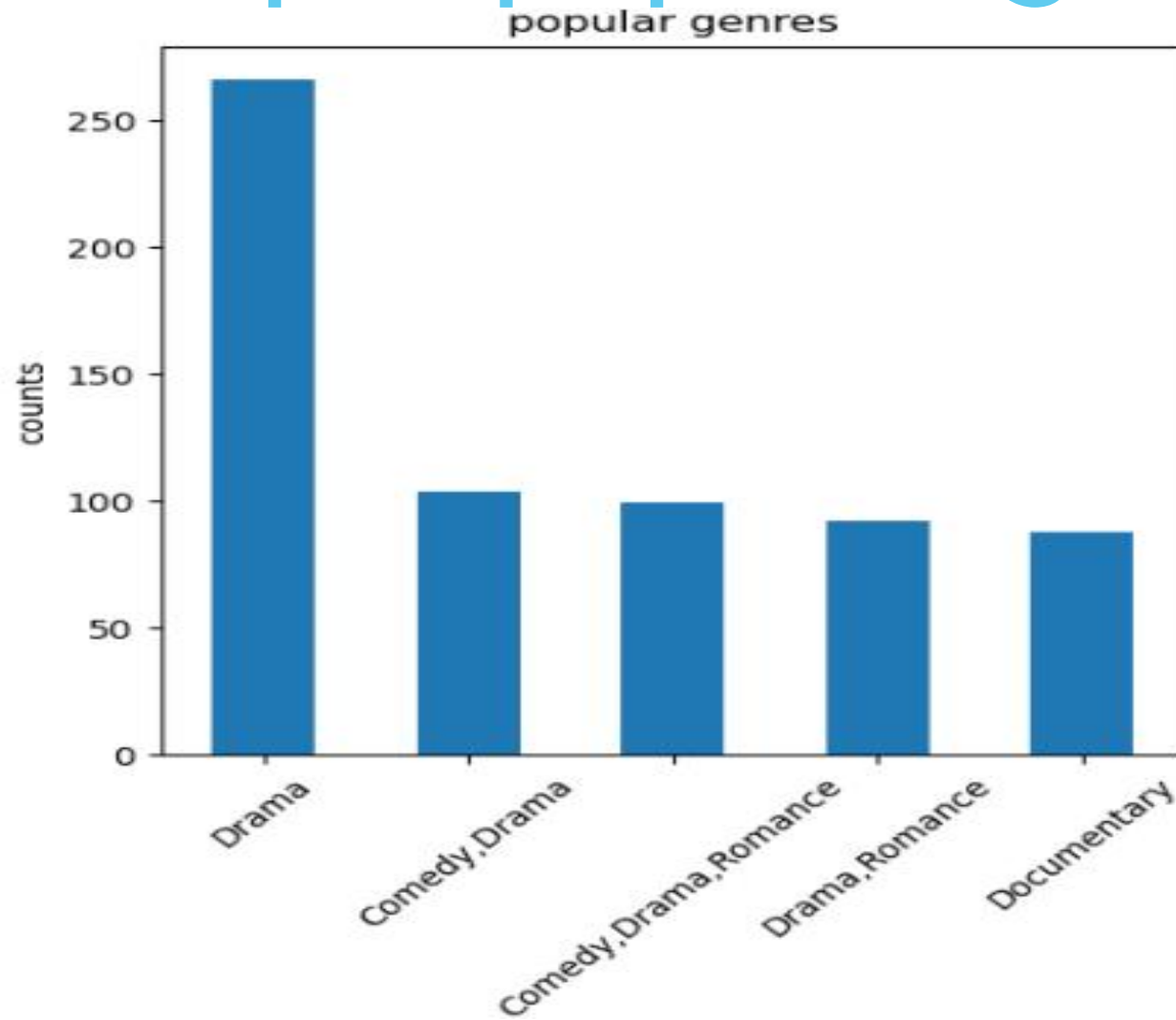# Comparison between directors and sales

Movies and films from Jay roach made highest sales followed by Mel Gibson and Sam Mendes.

Top 3 movies with top 3 genres and popularity

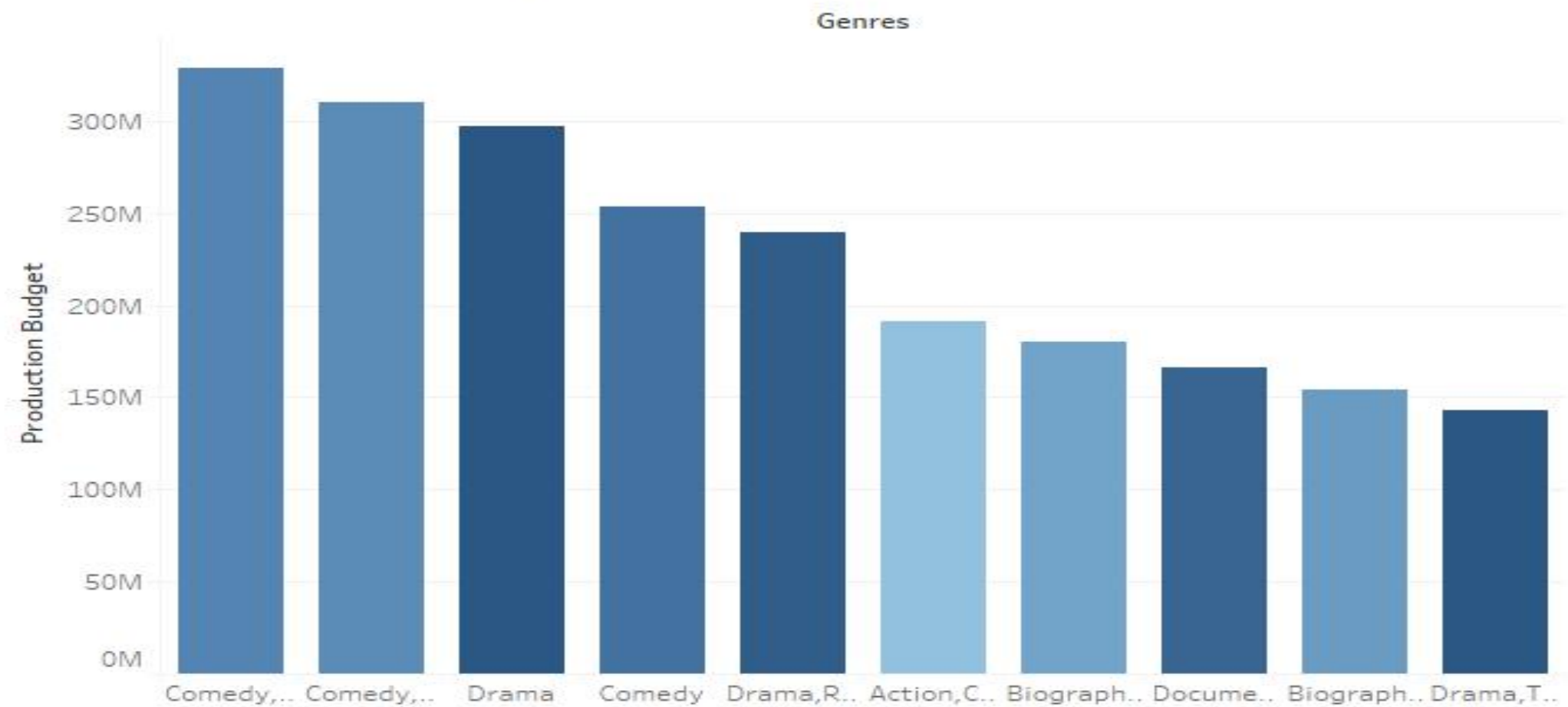Dramas from Anna were the most popular as well as crime, drama, thriller from Eden Movies.

Top 5 popular genres

popular genres

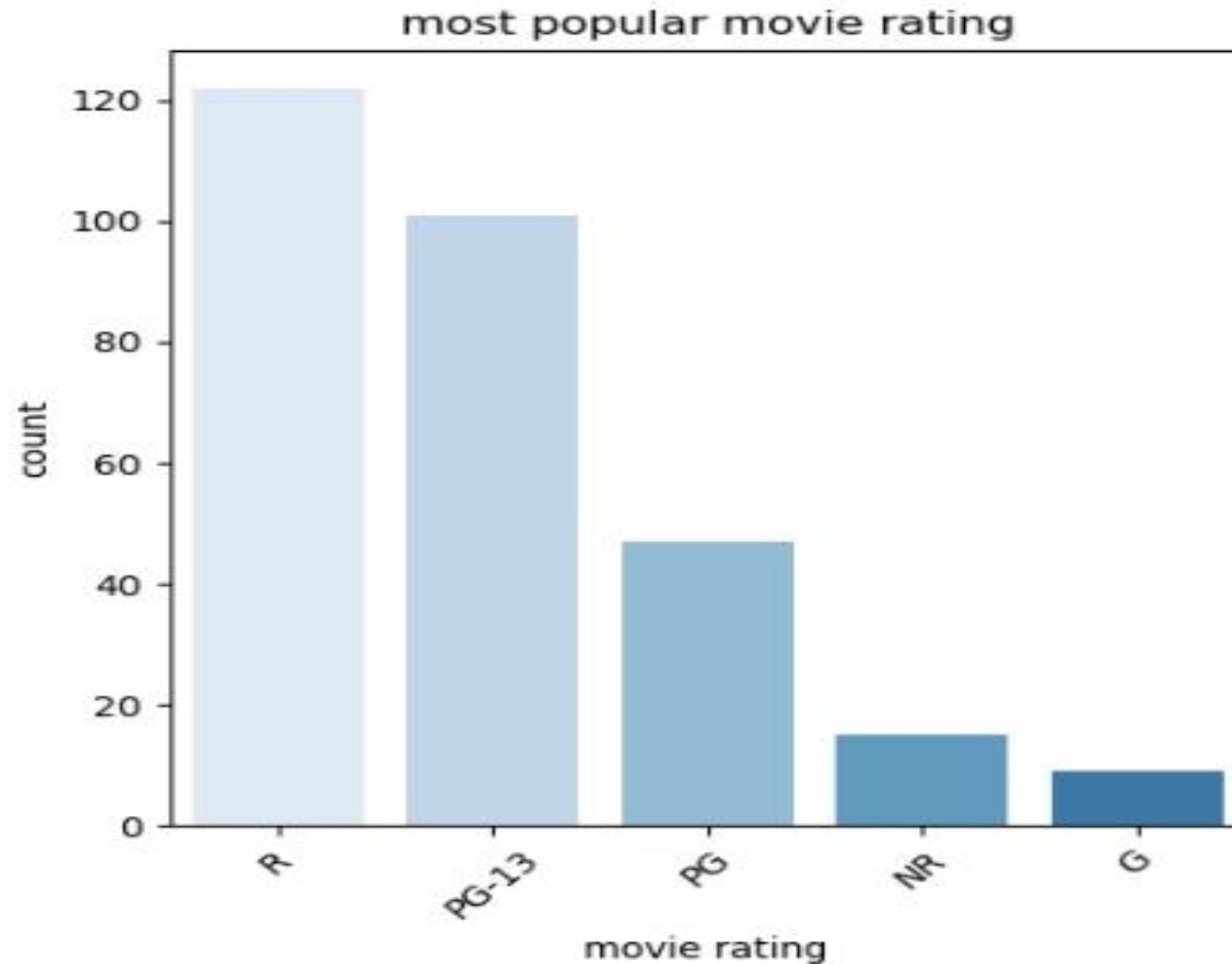Drama was the most popular genre doubling the other categories.

Comparison between genre and production budget

Genre vs Production budget

Comedy ,drama were costly in budget but gave the highest returns on investment.
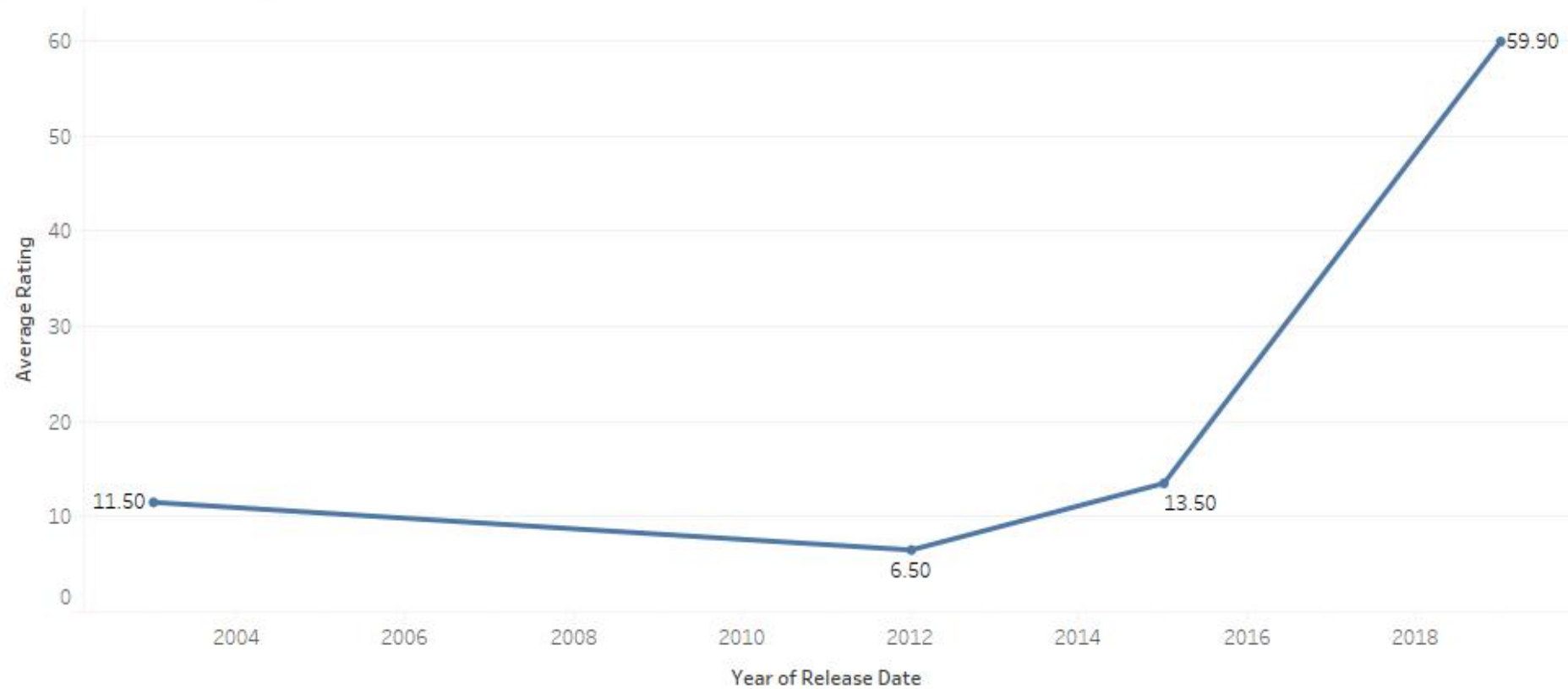
Top 5 most popular movie rating

R,PG-13,PG IN Descending order of popularity.
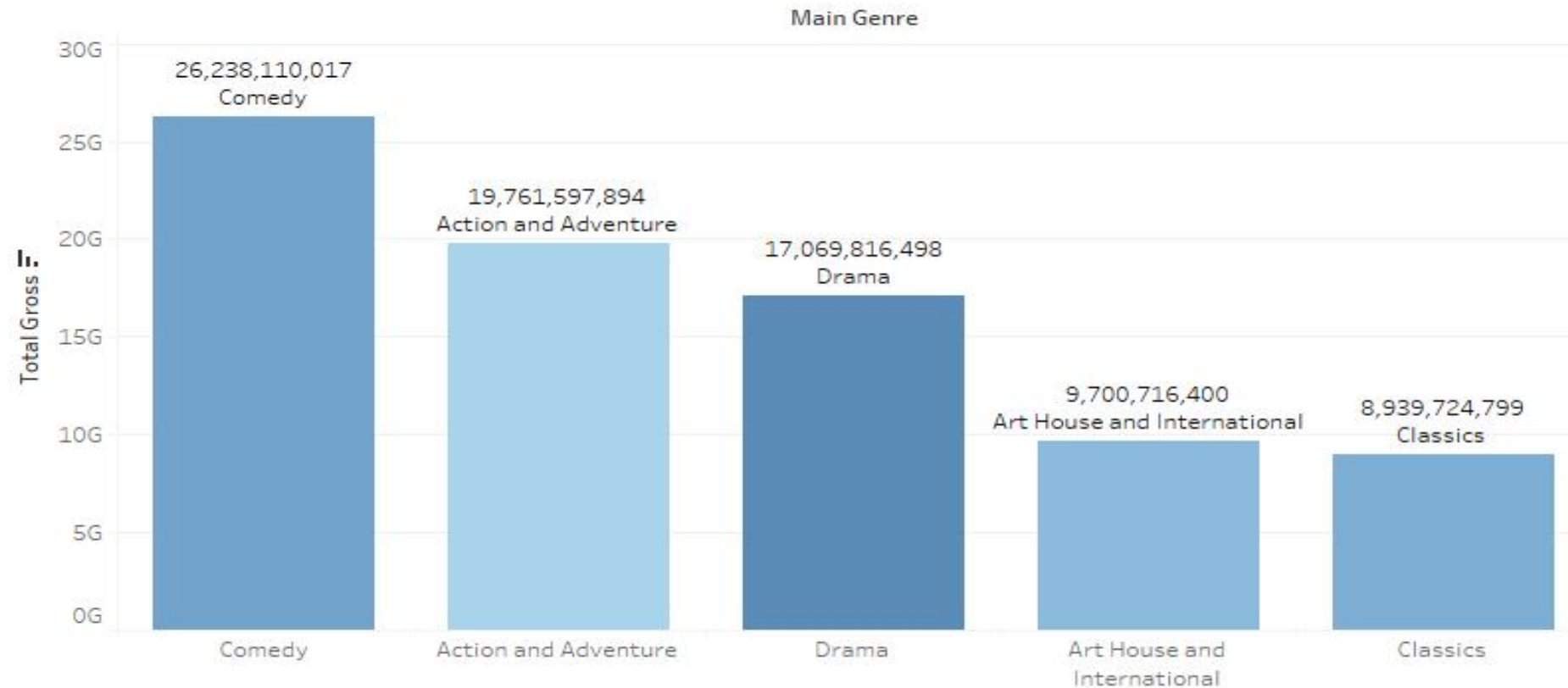
Yearly release timing analysis

Release Timing Analysis

Time series showed an upward trend in terms of Rating worth investing.

Comparison between genres and total gross

Genres vs total gross

Drama, action and adventure, comedy showed an increasing trend in terms of gross profits.

RESULTS/FINDINGS

❖ ROI Leaders: Action and Adventure genres lead in return on investment (ROI), making them highly profitable.

❖ Budget Allocation: Production budgets are highest for Comedy and Drama, reflecting their revenue potential.

❖ Studio Performance: Universal Pictures has the highest average ratings, showcasing its strong audience appeal.

❖ Release Timing: Average ratings peaked in recent years, suggesting growing audience satisfaction with newer movies. ROI Leaders: Action and Adventure genres lead in return on investment (ROI), making them highly profitable.

❖ Budget Allocation: Production budgets are highest for Comedy and Drama, reflecting their revenue potential.

❖ Studio Performance: Universal Pictures has the highest average ratings, showcasing its strong audience appeal.

❖ Release Timing: Average ratings peaked in recent years, suggesting growing audience satisfaction with newer movies.

ACKNOWLEDGEMENT

▶ We would like to acknowledge all the group one members for working tirelessly to contribute to this project:

▶ We would also like to pass our regard tour technical mentor MR.William Okomba having shaped our journey of learning data science.

▶ All our colleagues who have been helping us to debug whenever in need.

HYPOTHESIS TESTING REPORT

1. Testing the nature of the data.
2. Setting confidence interval.
3. Checking on the statistics of the average rating and total gross.
4. Stating the Null and alternative hypothesis.
5. Fail to reject the null hypothesis or not.

Testing the nature of the data

```
merged_movies_final_df.describe()
```

	average_rating	domestic_gross	foreign_gross	year	total_gross
count	1970.000000	1.970000e+03	1.970000e+03	1970.000000	1.970000e+03
mean	6.460812	2.005813e+06	1.474976e+07	2014.114721	1.675557e+07
std	0.999176	3.573622e+06	8.693627e+06	2.366826	9.497521e+06
min	1.600000	1.000000e+02	6.000000e+02	2010.000000	1.080000e+04
25%	5.900000	5.950000e+04	7.000000e+06	2012.000000	1.102500e+07
50%	6.600000	3.280000e+05	1.890000e+07	2014.000000	1.895730e+07
75%	7.100000	2.000000e+06	1.890000e+07	2016.000000	1.970000e+07
max	9.200000	1.710000e+07	5.410000e+07	2018.000000	7.100000e+07

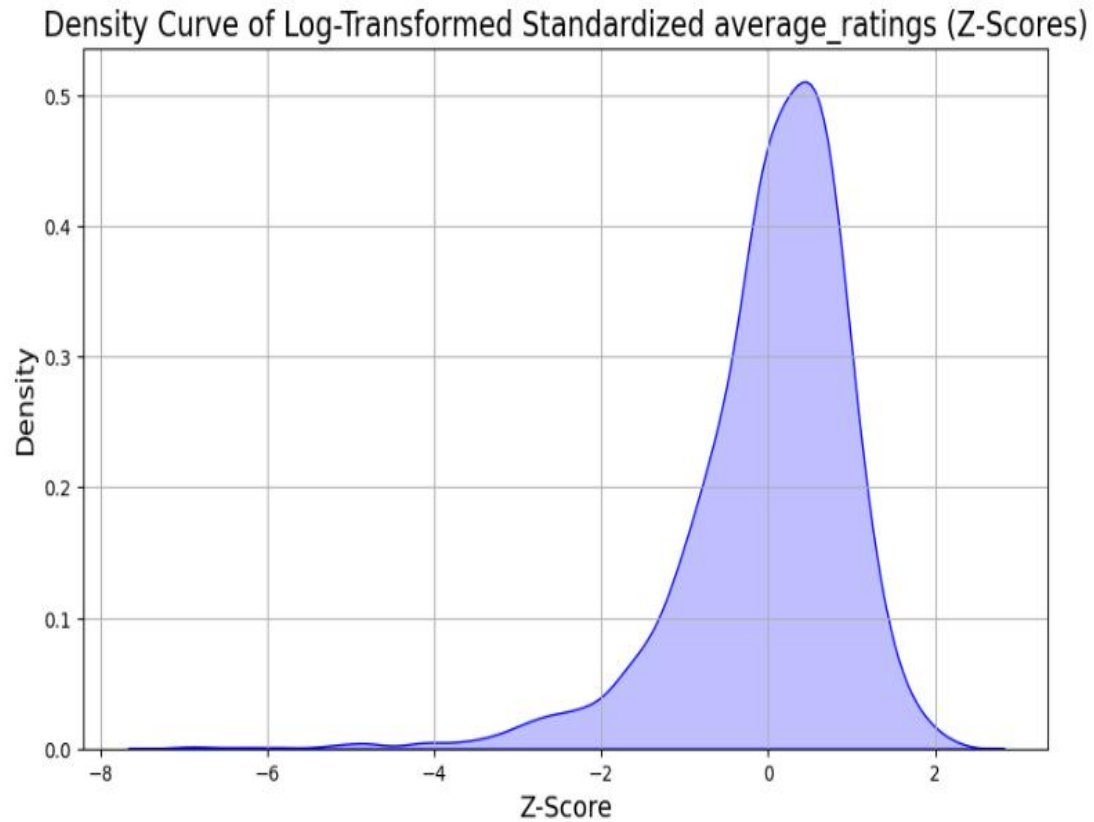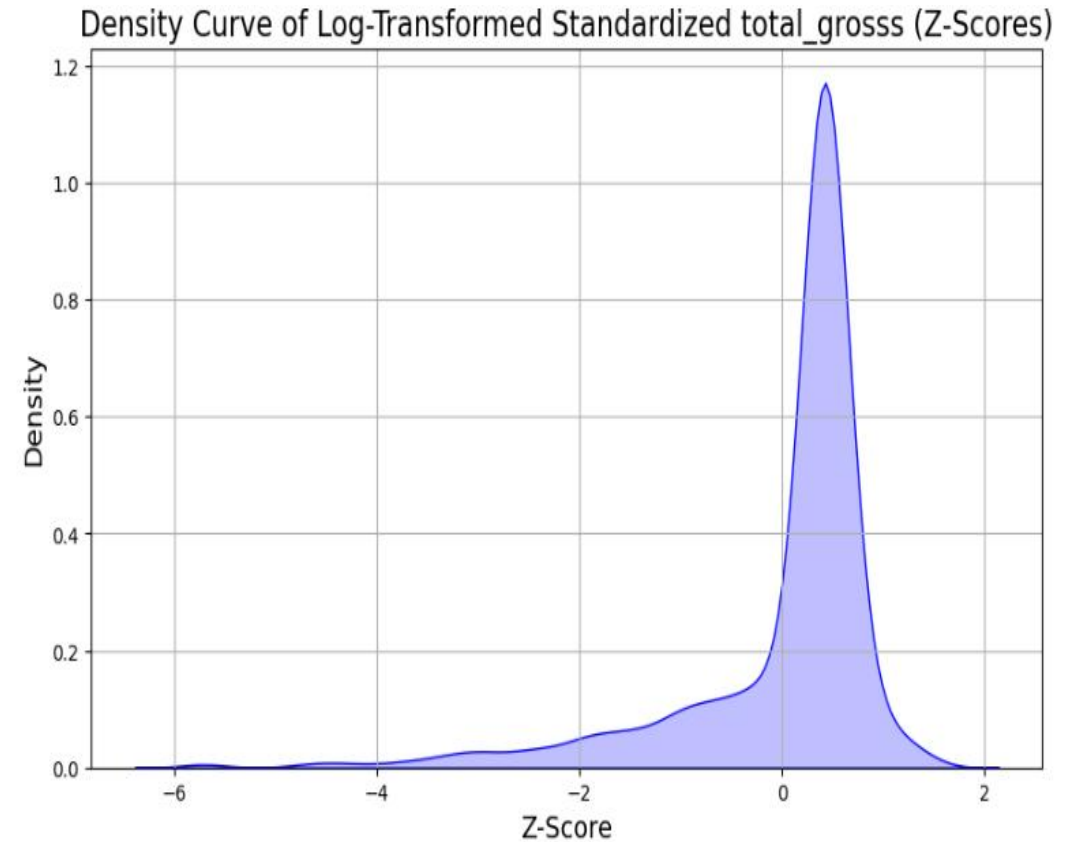# Statistics of the average rating and total gross.

Average rating
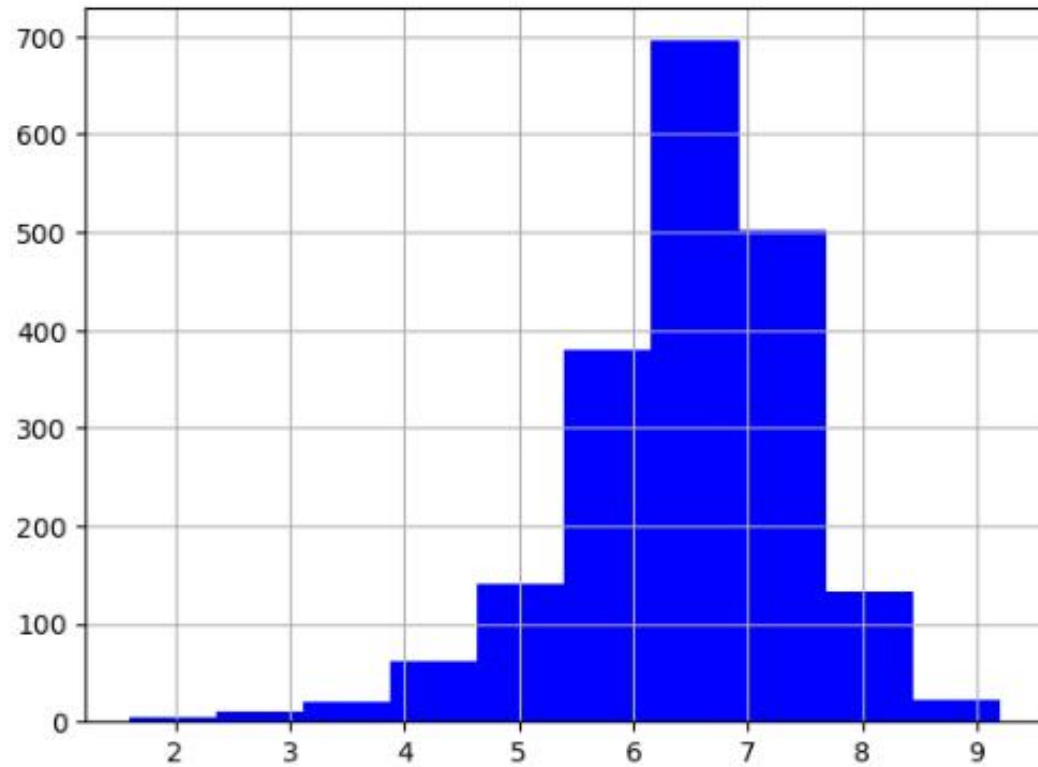
Total gross
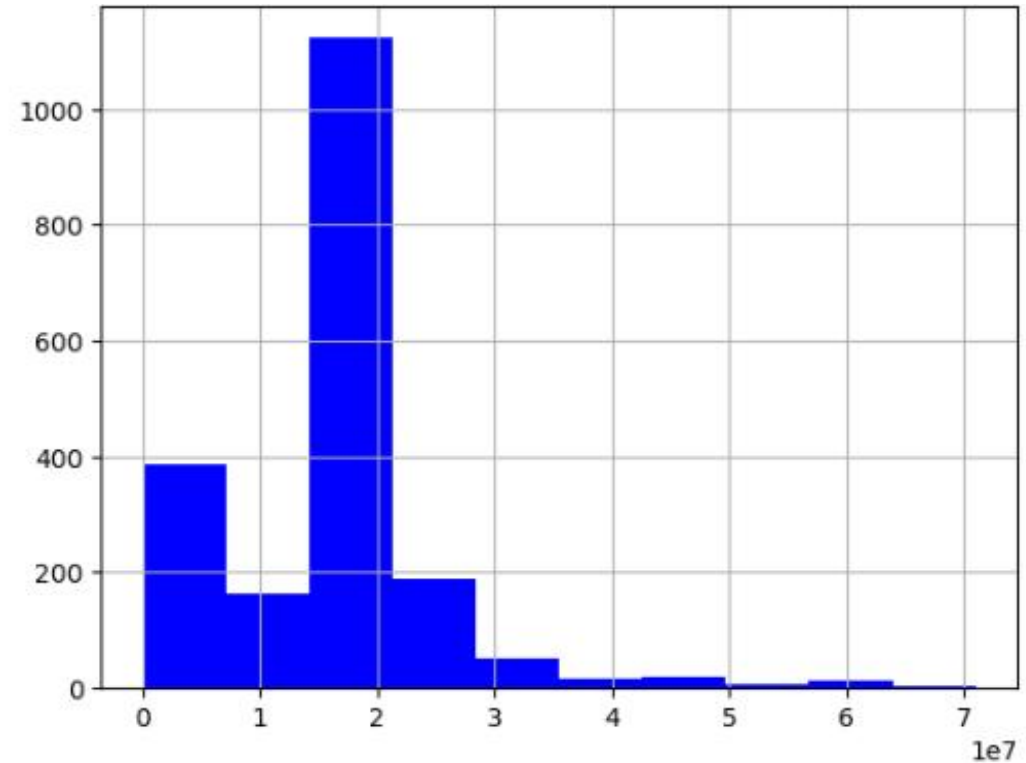
Statistics of the average rating and total gross after log transformation and standardization.

Average rating

Density Curve of Log-Transformed Standardized average_ratings (Z-Scores)

Total gross

Density Curve of Log-Transformed Standardized total_grosss (Z-Scores)

Null and alternative hypothesis.

Average rating

```
print("Null hypothesis\n log_average_rating_mean = ", log_average_rating_mean)
print("Alternative hypothesis\n log_average_rating_mean > ", log_average_rating_mean)
```

```
Null hypothesis
 log_average_rating_mean =  1.99944225935456
Alternative hypothesis
 log_average_rating_mean >  1.99944225935456
```

```
Sample Mean: 1.99944225935456

Z-Statistic: 0.0

P-Value: 0.5

Fail to reject the null hypothesis: There is no significant evidence
 that the mean is greater than  1.99944225935456
```

Total gross

```
print("Null hypothesis\n log_total_gross_mean = ", log_total_gross_mean)
print("Alternative hypothesis\n log_total_gross_mean > ", log_total_gross_mean)
```

```
Null hypothesis
 log_total_gross_mean =  16.269271504462125
Alternative hypothesis
 log_total_gross_mean >  16.269271504462125
```

```
Sample Mean: 16.269271504462125

Z-Statistic: 0.0

P-Value: 0.5

Fail to reject the null hypothesis: There is no significant evidence
 that the mean is greater than  16.26927150446225
```

MODELING REPORT

The logistic regression model's performance is suboptimal, as reflected by the following key observations:

1. High Errors: The MAE (65.66), MSE (8101.98), and RMSE (90.01) indicate significant prediction errors on average.

2. Negative R^2 (-0.378): The negative R^2 suggests the model performs worse than simply predicting the mean of the target variable.

Model Improvements Needed: The model likely requires better feature engineering, data preprocessing, and possibly a different model approach to improve accuracy and fit.

CONCLUSION

- From the analysis, it is clear that the Drama genre enjoys the highest ratings, but its box office success fluctuates over time. While the Adventure, Comedy, and Sci-Fi genres remain the highest-selling genres, this is likely due to their broad appeal and entertainment value.

- Directors like Clint Eastwood and Jay Roach contribute significantly to the sales, with their strong reputations and professional standards. The ratings analysis also highlights the R rating's popularity, reflecting the tastes of adult audiences.

- Overall, the variability in movie sales and ratings over the years is likely driven by factors such as market dynamics, audience preferences, and the quality of direction and production.

RECOMMENDATIONS

1. Prioritize Action and Comedy genres for future productions, as they have proven to be the most lucrative and widely appreciated.

2. Partner with top-rated directors (e.g., Steven Spielberg, Clint Eastwood) and writers (e.g., Woody Allen) to enhance critical and audience reception.

3. Optimize movie runtimes based on audience preferences; longer runtimes for Action and Adventure, and shorter runtimes for genres like Animation.

4. Leverage Universal Pictures for distribution due to their strong track record in generating high ratings.

CONTACT INFORMATION

1. catherine.kiptui@student.moringaschool.com

2. michellekavetza@gmail.com

3. gateromichael@gmail.com

4. noordinoordino470@gmail.com

5. kennethnyangweso99@gmail.com

6. aumakrystel5@gmail.com

7. segomich227@gmail.com