# MOVIE
## STUDIO PROJECT

# Overview

- We will be presenting our findings and insights on the current box office trends and their implications for our company's new movie studio.

- As a team, we have been tasked with exploring the types of films that are currently performing the best at the box office.

- We'll then translate these findings into actionable insights for the head of the movie studio.

- Our focus will be on making recommendations for the type of movies that will be the most profitable for our new movie division

# Data Understanding

We used the following datasets in our data analysis:

- **Rotten Tomatoes Review dataset - To sort genres by critic review ratings**

- **Rotten Tomatoes Movie Info Dataset**

- **IMDB Movie database - To analyze best performing by genres**

- **The Numbers Movie Budget Dataset - To calculate ROI**

# Project
## Objectives

- Recommending the most popular genres.
- Recommend the most popular genre combinations.
- Recommend the best month to release movies.
- Draw a correlation between domestic and worldwide revenue.
- Suggest the best director by critic rating.

# Project Scope

- Analyze the frequency of genre combinations, their average ratings, and their success metrics.
- Analyze average production budgets and movie runtimes across different MPAA ratings and genres.
- Assess the performance of movies across domestic and international markets.

## Assumptions

- We assumed that our datasets were an accurate representation of the movie industry.
- The genres are specified consistently and accurately.
- The ratings accurately reflect the audience perception.

# Data Analysis

**Tools used**

- **Python - This was the primary language for data manipulation, analysis and visualizations**

  **Libraries**

  - **Numpy - for mathematical computations.**
  - **Pandas - for filtering, grouping, summarizing.**
  - **Matplotlib - for creating static visualization.**
  - **Seaborn - for advanced visualizations.**

- **Jupyter Notebook - It's an interactive computing environment for writing code.**
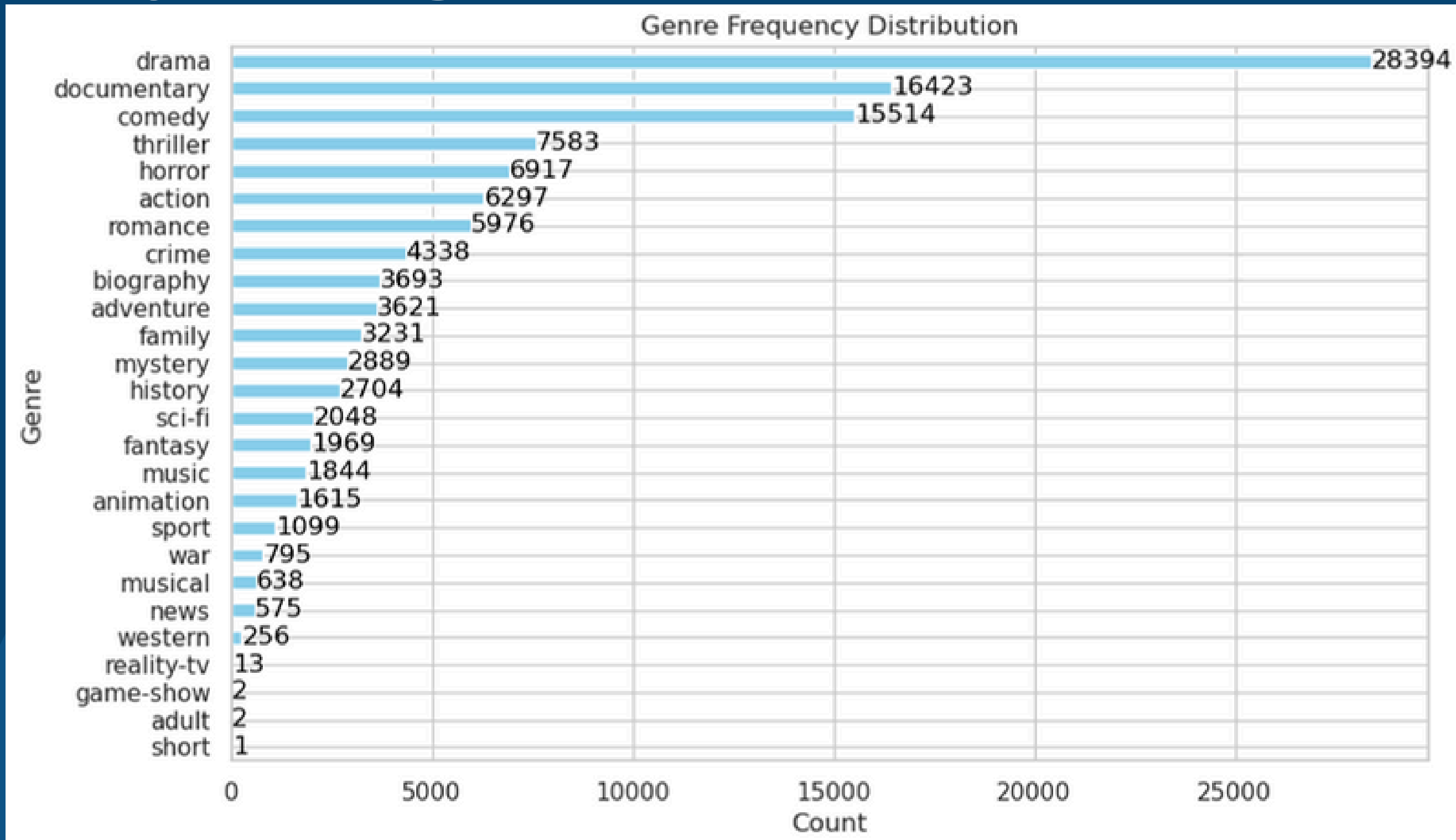
- **Tableau - to create interactive visualizations**

# Methodology

1. Accessing data - we imported appropriate libraries and loaded the data using pandas.
2. Data preparation - Identified and rectified inconsistencies by handling missing values through dropping and imputing missing values, and standardizing data formats and units, and removing duplicates.
3. Descriptive Analysis - Computed basic statistics including mean, standard deviation and the inter-quartile ranges to understand the distribution of and central tendencies of key variables.
4. Visualizations - Created visual representations of data to identify patterns and trends. We used histograms, box plots, bar charts, pie charts, heat maps, scatter plots and line graphs.

# Key Insights - Genre Popularity



Genre Frequency Distribution

| Genre | Count |
|---|---|
| drama | 28394 |
| documentary | 16423 |
| comedy | 15514 |
| thriller | 7583 |
| horror | 6917 |
| action | 6297 |
| romance | 5976 |
| crime | 4338 |
| biography | 3693 |
| adventure | 3621 |
| family | 3231 |
| mystery | 2889 |
| history | 2704 |
| sci-fi | 2048 |
| fantasy | 1969 |
| music | 1844 |
| animation | 1615 |
| sport | 1099 |
| war | 795 |
| musical | 638 |
| news | 575 |
| western | 256 |
| reality-tv | 13 |
| game-show | 2 |
| adult | 2 |
| short | 1 |

The 'Drama' category is the most popular among viewers with a significantly high number of occurences meaning 'Drama' is more appealing and engaging to a higher audience.

Genre Pairwise Co-occurrence
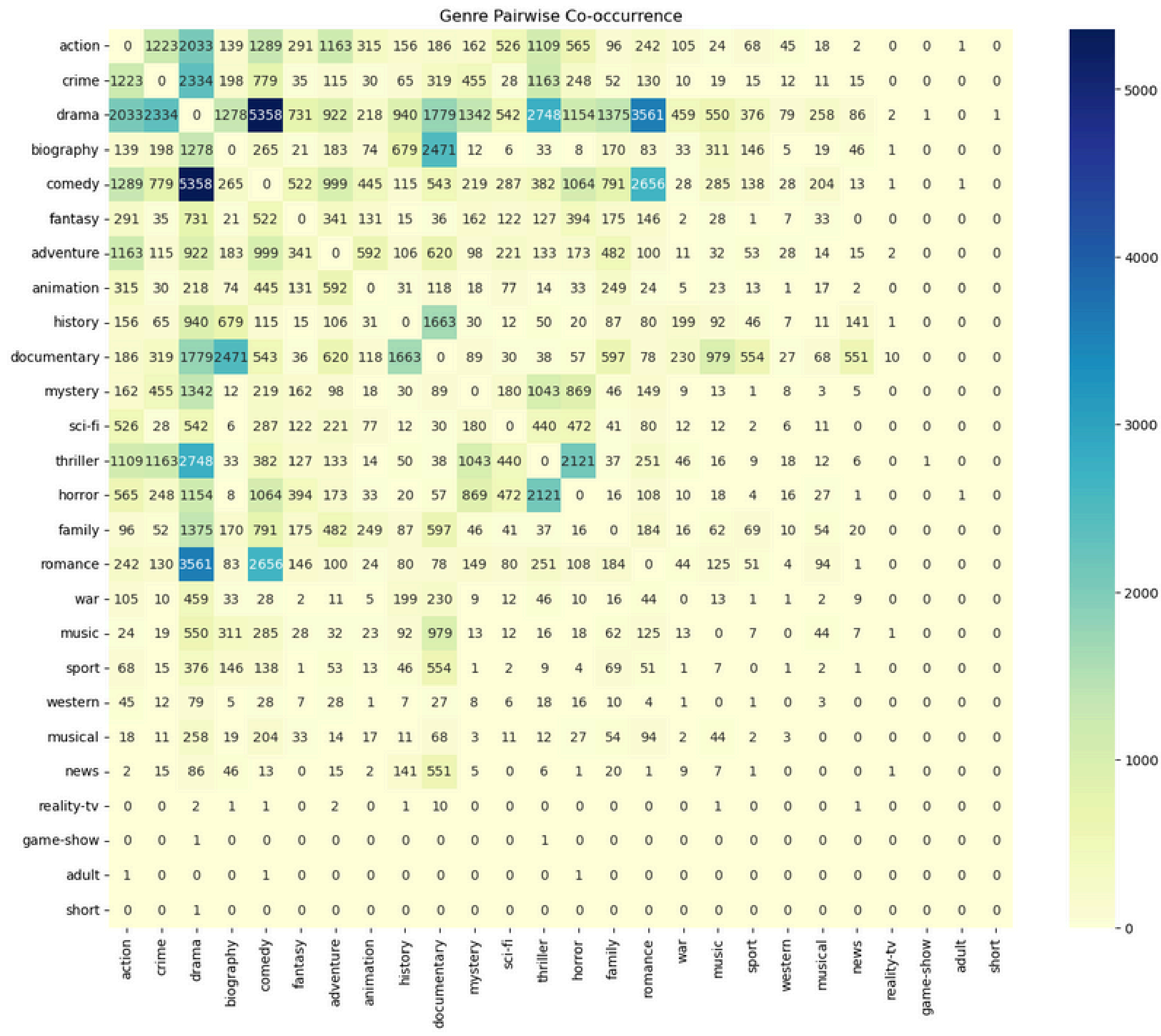
# Key Insights - Genre Co-occurence

- **The combination for Comedy & Drama, has the highest occurrence of 5358 and a rating of 6.742 indicating a successful formula/combination.**

- **This shows that movies that blend elements of both comedy and drama are more appealing to a larger audience.**

# Key Insights - Budget v. Revenue



**Production Budget vs. Worldwide Gross**
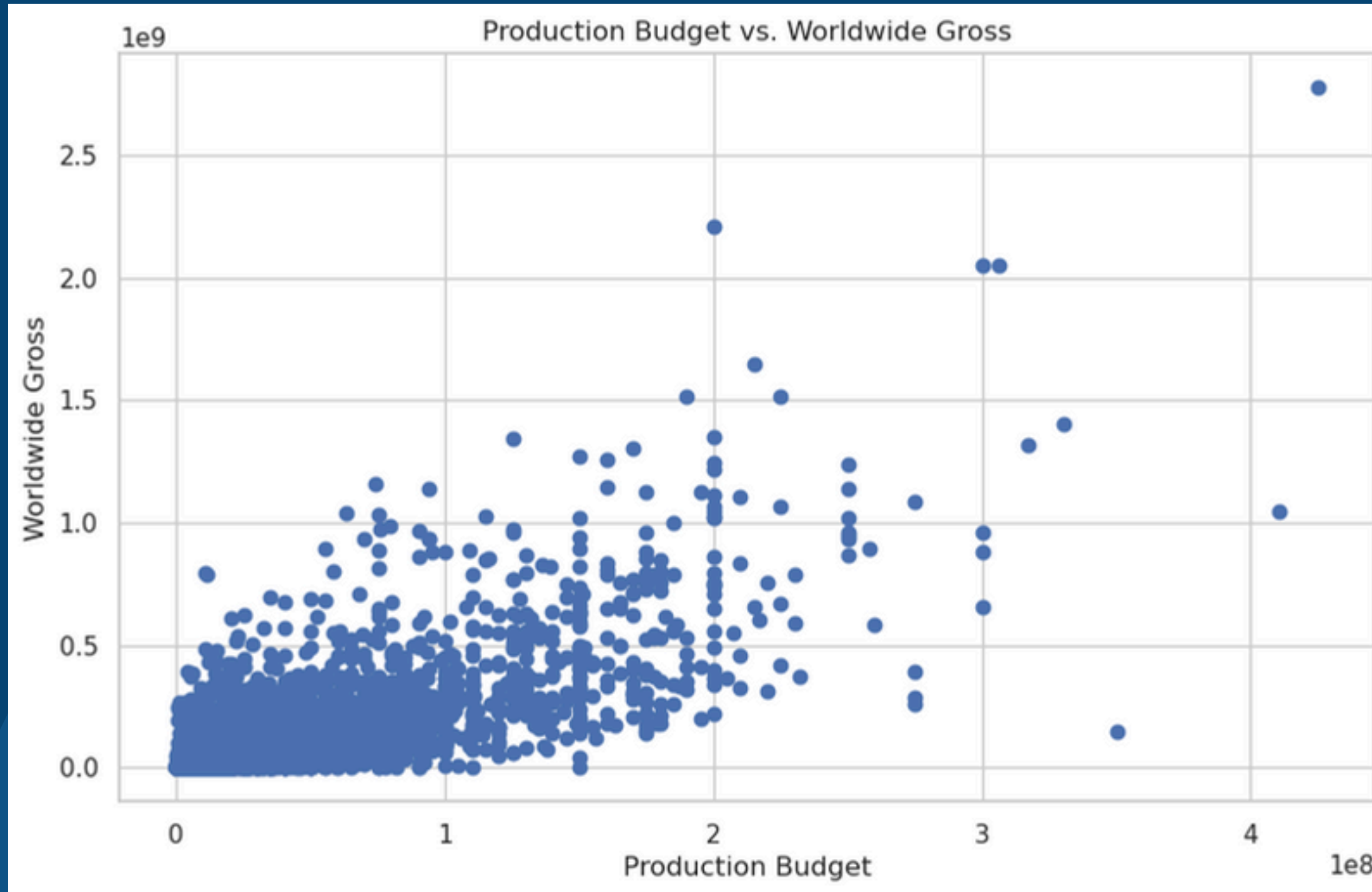
There is a strong positive correlation between production budget and worldwide gross revenue of 0.748, implying that high budgets have more impact on global earnings.

# Key Insights - Average Profit by Month



Average Profit by Month of the Year
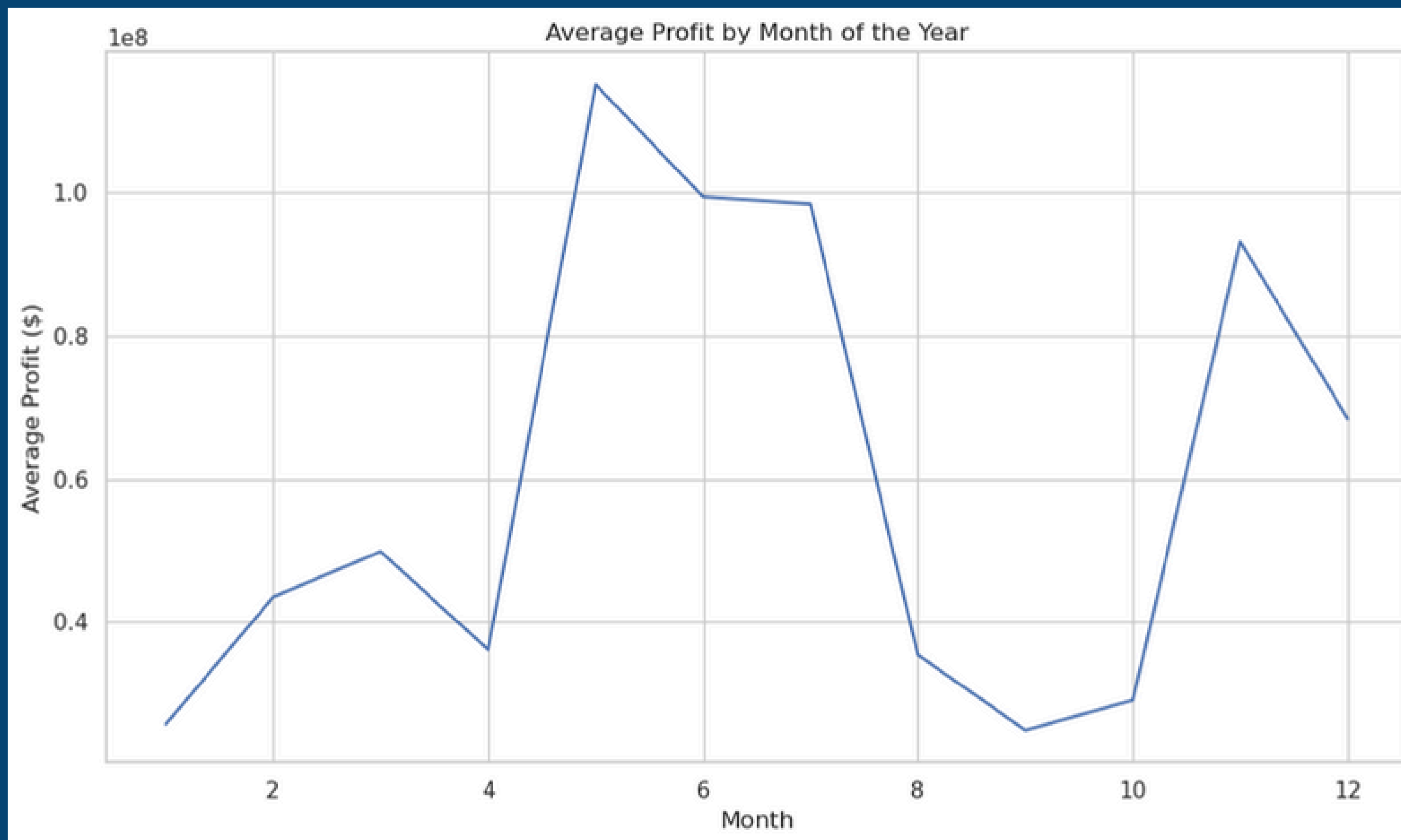
- It is evident that May, June, and July are the most profitable months for movies. May has an average profit of 11.5 million followed closely by June with 10 million and July with 9.8 million.

- Therefore, we should consider releasing movies during one of these three months to bolster profitability.

# Key Insights - Directors by Critic Rating



Top 20 Directors by Mean Standardized Rating
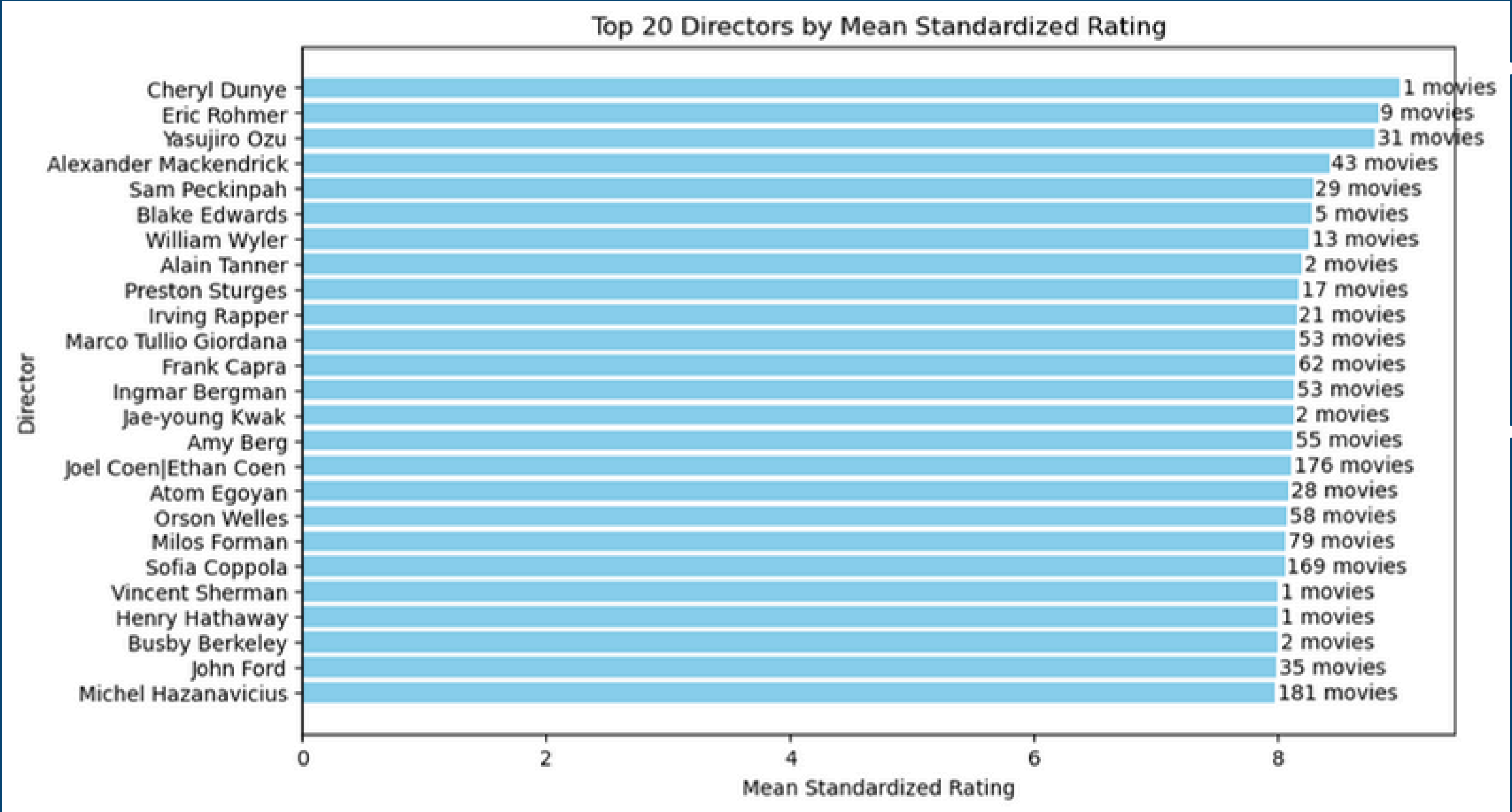
- The Coen brothers have directed 176 movies with an average rating of 8.105. This indicates that they are highly respected film makers with a strong track record of producing quality films.

- The next recommended director is Sofia Coppola whose movies have an average rating of 8.052129 having directed 169 movies. Her high average rating is exceptionally high which suggests that her movies are consistently well received by critics and audiences alike.

# Team Members

- **Michael Achoki** ✉ 
- **Esther Cheruiyot** ✉ 
- **Floyed Muchiri** ✉ 
- **Patricia Ngari** ✉ 
- **Kelvin Murithi** ✉