

DATA SCIENCE PART TIME- GROUP 15 PHASE 2 PROJECT

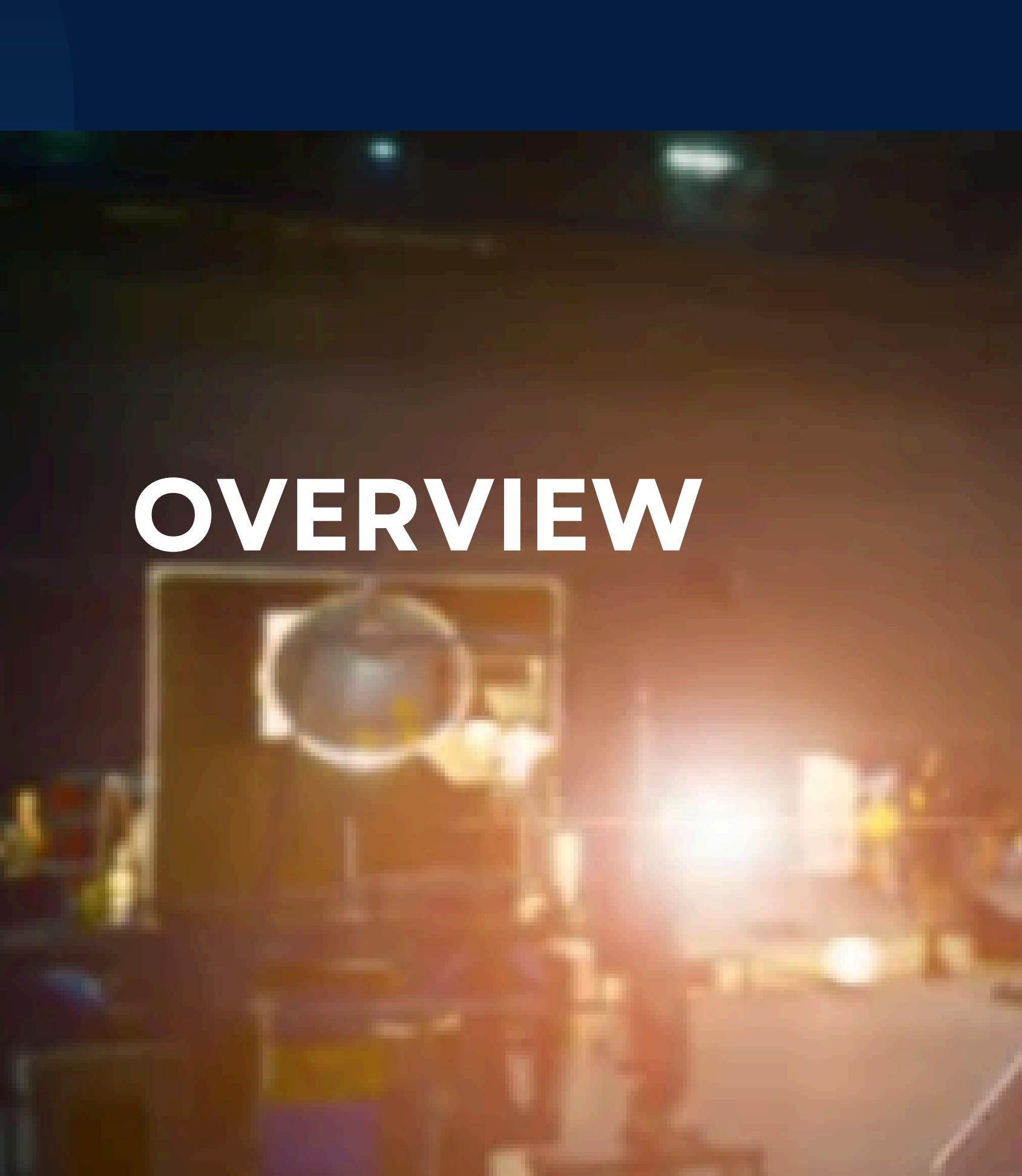
Title: Data-driven insights for a successful movie studio launch

Presented by:

- 1. Diana Aloo**
- 2. Sylvia Wambui**
- 3. Catherine Kaino**
- 4. Vincent Buluma**

14th September 2025

OVERVIEW



01

**Business
Problem &
Objectives**

02

**Data Sources
Understanding**

03

**Data
Preparation**

04

**Analysis &
Findings**

09

Recommendations

10

**Conclusion &
Next Steps**

01 BUSINESS PROBLEM & OBJECTIVES

Problem

Our company plans to enter the movie production industry but lacks experience. This analysis explores historical movie data to identify what makes films successful—financially and critically(reviews)

Objectives

Venturing into the movie industry without prior production experience carries financial risks and therefore the aim is to offer data-backed recommendations to help guide decisions on genre selection, production strategy, and release planning. Specific objectives are to explore the following:

- 1.What factors drive box office success
- 2.Understand how the studios that command high Box Office Revenue achieve success
- 3.Corelation between films ratings and financial perfomance

02 DATA SOURCES UNDERSTANDING

To answer the project objectives, our analysis draws on below reputable film industry datasets which form a comprehensive view of the domain:

1. Box Office Mojo – Provides detailed box office revenue data, both domestic and international.
2. IMDb – Offers extensive metadata on films, including genres, casts, crews, and ratings from a global audience.
3. Rotten Tomatoes – Supplies critic and audience scores, capturing critical reception and sentiment.
4. The Movie Database (TMDb) – Delivers community-driven information such as keywords, genres, release dates, and popularity metrics.
5. The Numbers – Tracks financial performance, including budgets and revenues, allowing deeper profitability insights.

Integrating these sources ensures that data analysis is not only based on financial outcomes but also relatable with audience feedback, critical reviews, and contextual metadata. This holistic approach strengthens the ability to link descriptive insights with actionable business recommendations.

03 DATA PREPARATION

Proper data preparation was crucial. By splitting multi-genre rows and combining key metrics, we ensured the insights reflect real-world trends. The following data preparation approach was taken:

- Cleaned and standardized gross revenue data
- Handled missing values in genres, ratings, and runtime
- Created worldwide gross metric for comparison
- Split multiple genres to analyze individually
- Visualized trends by studio, year, and rating

The following tools and libraries were used;

- Pandas- Used for data manipulation, merging datasets, grouping by genre/year
- Numpy- Used for Numerical operations (e.g. correlation calculation)
- Matplotlib-Used for creating static plots (scatterplots, bar/line charts)
- Seaborn-Used for Statistical data visualization (scatterplots with trendlines)
- Zipfile-Used to Unzip folders
- Sqlite3 used as SQL database engine

04 DATA ANALYSIS & FINDINGS

Each objective in this project guided a specific set of analyses. The steps followed a logical progression from data exploration and visualization. The findings below represent how analytical techniques were used to inform strategic decisions.

1. **Objective 1. Identify Genres that drive high revenue**

a) Analysis Conducted

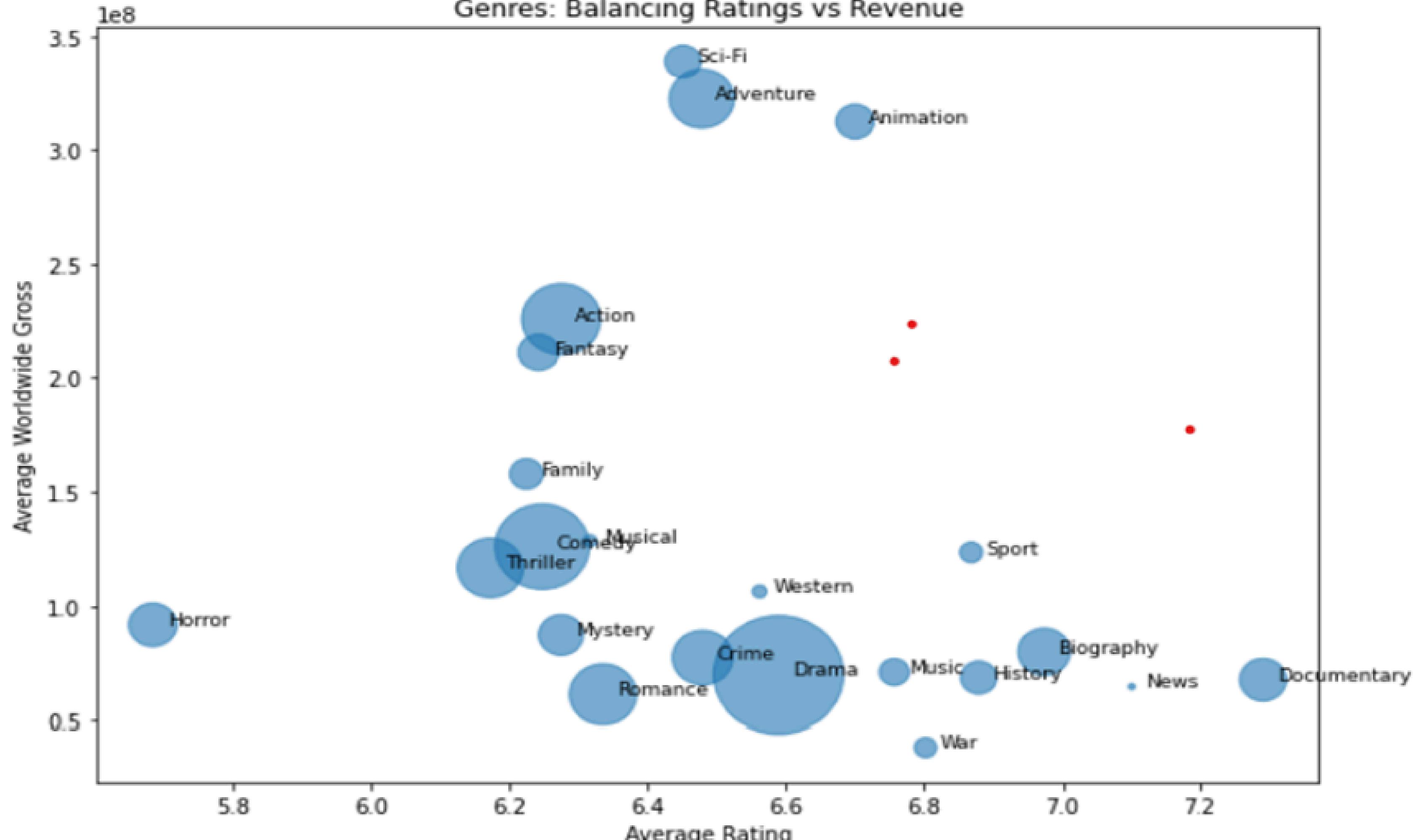
- Grouped movies by genre and computed:
 - Average Worldwide Revenue
 - Count of movies per genre and rating
- Split multi-genre movies into multiple rows to assess each genre independently.

b) Findings

- Sci-Fi, Adventure, and Animation lead in revenue (~\$300M+), though Sci-Fi and Adventure have moderate ratings (~6.4–6.5).
- Animation strikes a strong balance (avg. rating 6.7 and high revenue), making it ideal for broad investments.
- Documentary and Sport achieve high ratings (~7.3 and ~6.9) but lower revenue.

05 DATA ANALYSIS & FINDINGS....CONT'

Genres: Balancing Ratings vs Revenue



06 DATA ANALYSIS & FINDINGS...CONT'

Objective 2. Understand the studios that command high Box Office Revenue

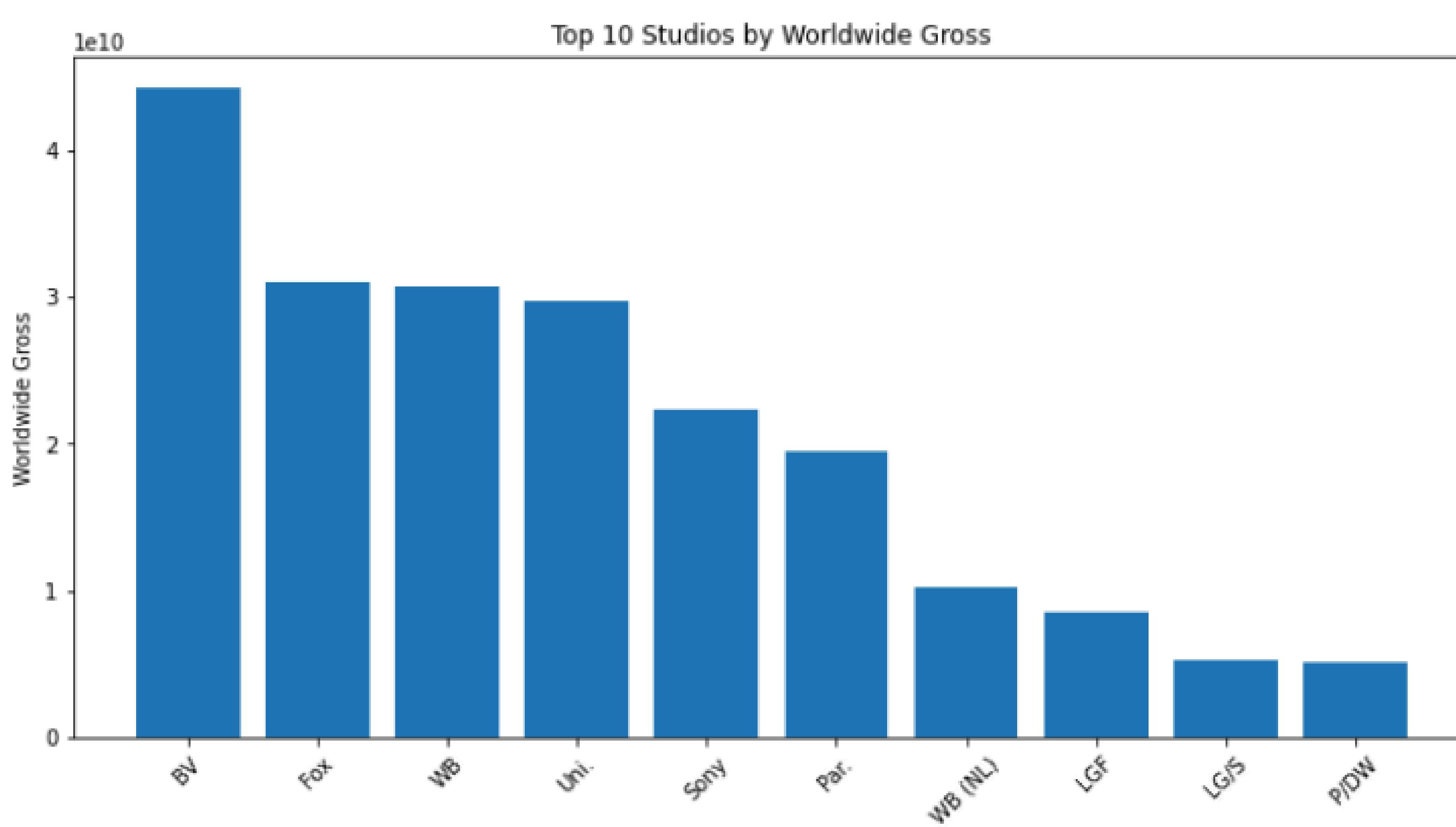
a) Analysis Conducted

- Data Merging:
 - Merged movie basics, ratings, and box office gross datasets.
 - Joined on movie titles and cleaned inconsistencies in naming across datasets
 - Grouped the merged dataset by studio
 - Calculated total and average worldwide gross per studio

b) Findings

- A small group of studios i.e. Disney (BV), Warner Bros (WB), Universal, and Fox – consistently account for the highest worldwide grosses.
- These studios frequently release franchise-backed blockbusters, which significantly boost revenue.
- Studios with long-running IPs (e.g., Marvel, Star Wars, Fast & Furious, Harry Potter) enjoy repeatable, large-scale success.

07 DATA ANALYSIS & FINDINGS.....CONT



08 DATA ANALYSIS & FINDINGS....CONT'

Objective 3. Correlation between film ratings and financial performance

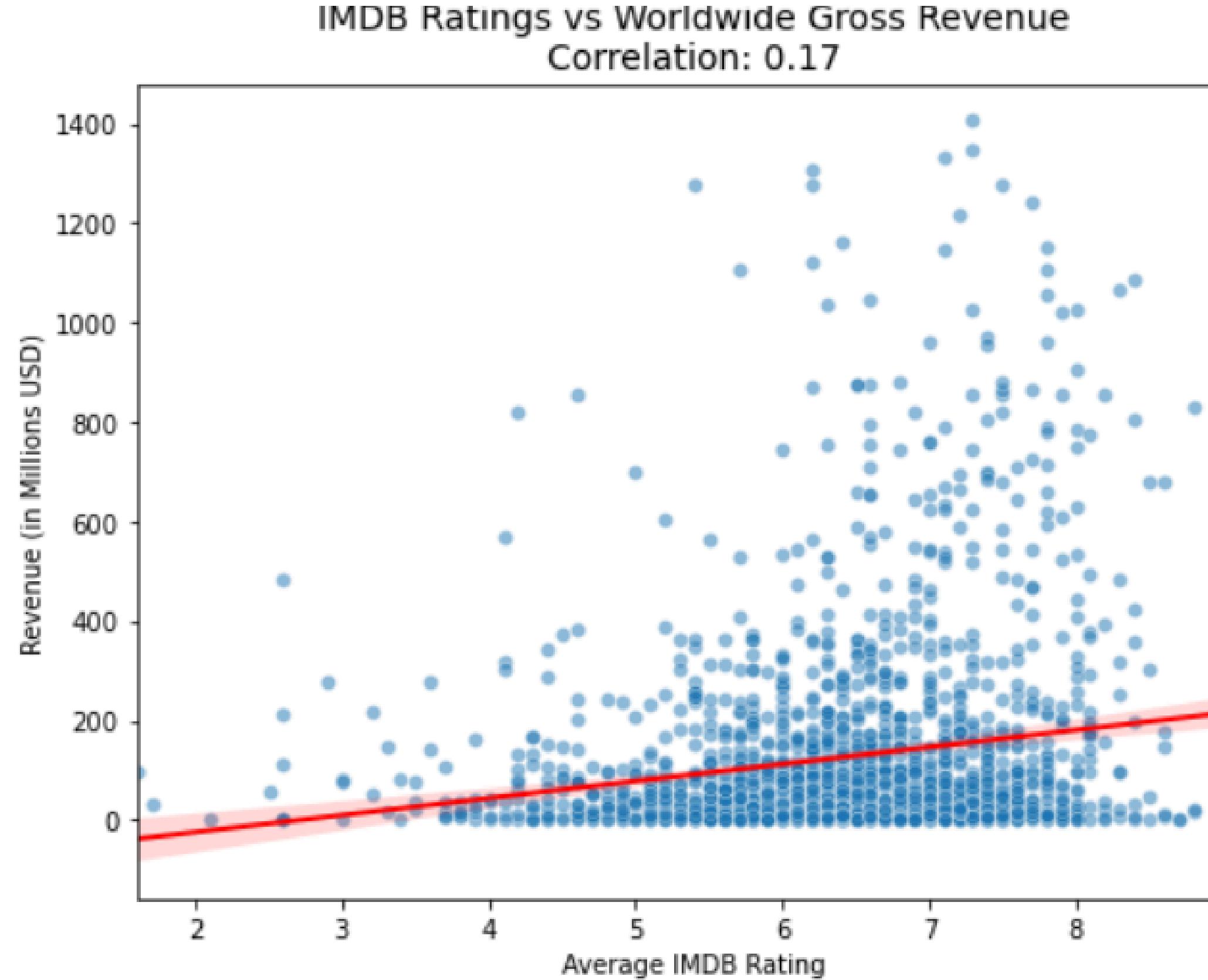
a) Analysis Conducted

- Created scatterplots of individual movies:
 - X-axis: IMDb Ratings
 - Y-axis: Worldwide Revenue
- Overlaid a regression trendline to test correlation

b) Findings

- Very weak correlation between ratings and revenue (correlation coefficient ≈ 0).
- Several blockbuster films earn \$1B+ with ratings around 6-7.
- Conversely, some highly rated films e.g Documentaries (>8) earned modest revenue, showing a weak relationship.

09 DATA ANALYSIS & FINDINGS....CONT'



10 RECOMMENDATIONS

Based on the analysis of IMDB ratings, box office revenues, genres, and studio performance, we provide the following recommendations for balancing **commercial success** and **audience approval**.

1. Genre Strategy

- Maintain a balanced portfolio – invest heavily in blockbuster genres (Sci-Fi, Animation, Adventure) for profitability while supporting Documentary and Sport for awards, critical acclaim, and niche audiences.

2. Studio Strategy

- Top-earning studios focus on Action, Sci-Fi, Animation, and Adventure – genres with global appeal. We recommend that the company should benchmark with best performing studios before starting the business.

3. Ratings vs Revenue Relationship

- Do not rely solely on ratings as a predictor of box office potential.
- Use ratings as a secondary indicator of audience trust and long-term film value (streaming success, awards, cultural impact).

11 CONCLUSIONS & NEXT STEPS

Conclusion

The recommendations provide a balanced approach that ensures both short-term profitability and long-term sustainability, aligning financial growth with audience approval and brand value.

- Commercial Track: Focus on scalable blockbuster genres (Sci-Fi, Animation, Adventure) to maximize box office returns.
- Prestige Track: Support critically acclaimed genres (Documentary, Sport, Indie Drama) to build reputation, win awards, and maintain brand trust.

Next Steps

Based on the recommendation, the following next steps are required to guide strategic roll out of the studio business.

1. Develop a Balanced Genre Portfolio
2. Develop a market entry roadmap and strategy to optimize profitability as per selected portfolio.
3. Budget for the same.

THANK YOU

