



Prepared by group 3

# *Group Project*

Movie Data Analysis



# *Group Members*

**Lorena Terah**  
Leader

**Edgar Owuor**  
member

**Sharon Kipruto**  
member

**Stephen Musyoka**  
member

**Salma Mwende**  
member

**Dawa Jarso**  
member



# *Introduction*



---

This project involves conducting an in-depth analysis aimed at generating actionable insights for stakeholders in the movie business.

The analysis will explore various aspects of the movie industry, including financial performance, audience reception, and market trends. Our goal is to equip stakeholders with a comprehensive understanding of key metrics and trends, enabling them to make informed decisions that drive the company's growth and success in the competitive movie market.



## Problem objectives

1. Identify types of films that are performing the best.
2. Explore and analyze current box office trends.
3. Identify factors that contribute to the success of a movie.
4. Translate findings into actionable insights.



# **DATA QUESTION TO BE ANSWERED**

## **1. Top-Performing Films:**

- *Characteristics: Traits contributing to high box office gross.*
- *Financial Success: Movies with the greatest financial success.*

## **2. Popular and Successful Genres:**

- *Trends: Popular and successful genres over time.*
- *Audience Preferences: Content that resonates with target audiences based on ratings and reviews.*

## **3. Impact of Budget on Revenue:**

- *Budget vs. Revenue: How movie budgets affect box office revenue.*
- *Profitability: Can smaller budget films be profitable? Identifying an optimal budget range.*



# DATA SOURCES

## 1. Box Office Mojo

- Tracks box office revenue data
- Analyzes financial performance of films

## 2. IMDb (Internet Movie Database)

- Provides info on movies, TV shows, and entertainment industry
- Includes cast/crew, plot summaries, reviews, and ratings

## 3. Rotten Tomatoes

- Aggregates critic and audience reviews for movies and TV shows
- Measures overall reception via percentage scores

## 4. The Movie DB (TMDb)

- User-driven database for movies, TV shows, and celebrities
- Offers cast/crew info, plot details, ratings, and reviews

## 5. The Numbers

- Provides detailed box office and financial data
- Includes budgets, home video sales, and industry statistics



# Data Exploration and Cleaning Process

Familiarizing with the Datasets:

- Explored various datasets and analyzed their columns.

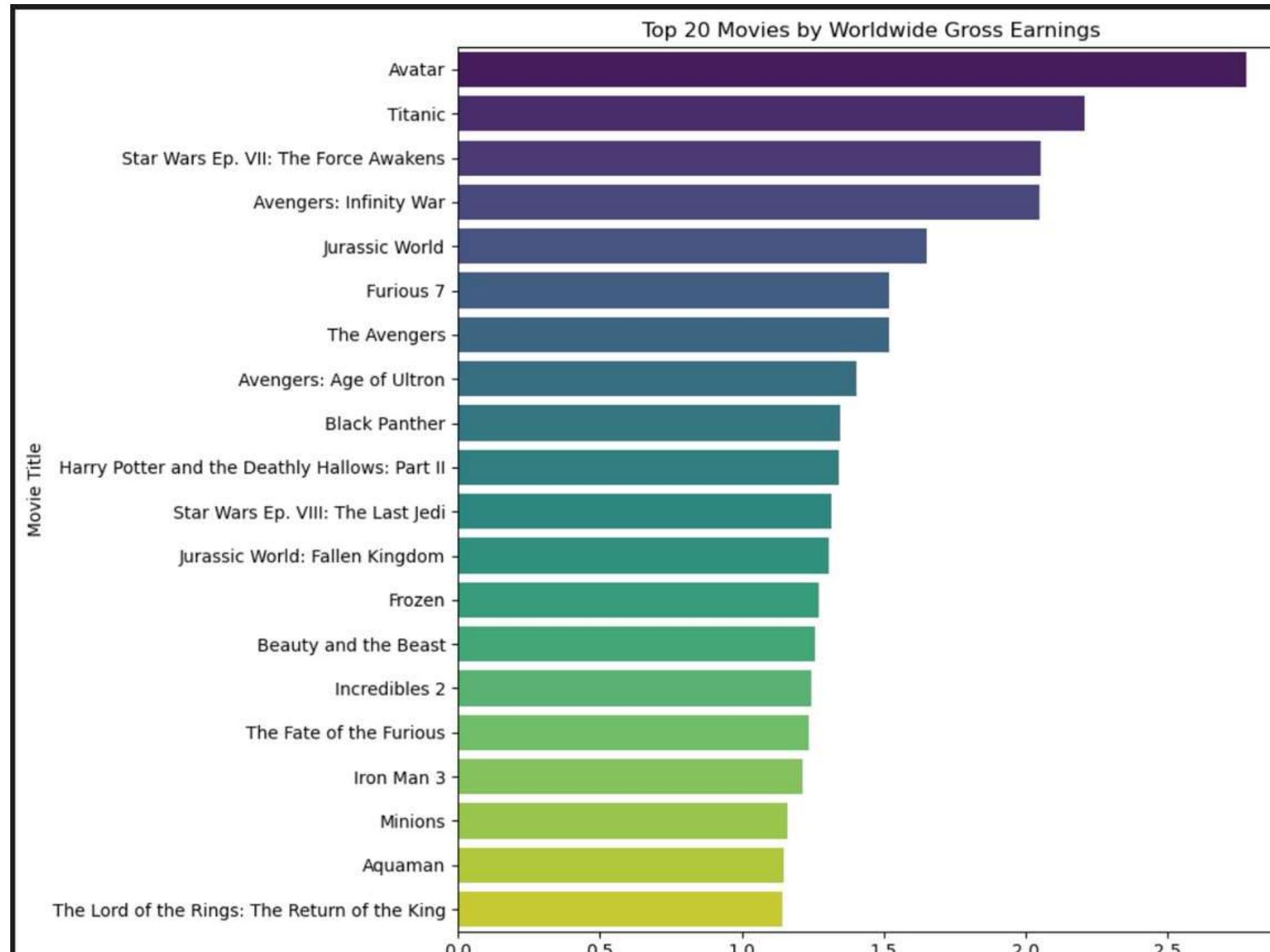
## Data Cleaning Methods

- Handling Missing Values:
  - Identified and managed missing data through imputation or removal.
- Data Transformation:
  - Converted data types and standardized formats for consistency.
- Removing Duplicates:
  - Checked for and removed duplicate records to ensure data uniqueness.
- Outlier Detection:
  - Identified and handled outliers using techniques such as z-score and IQR method.



# Data Visualization Results

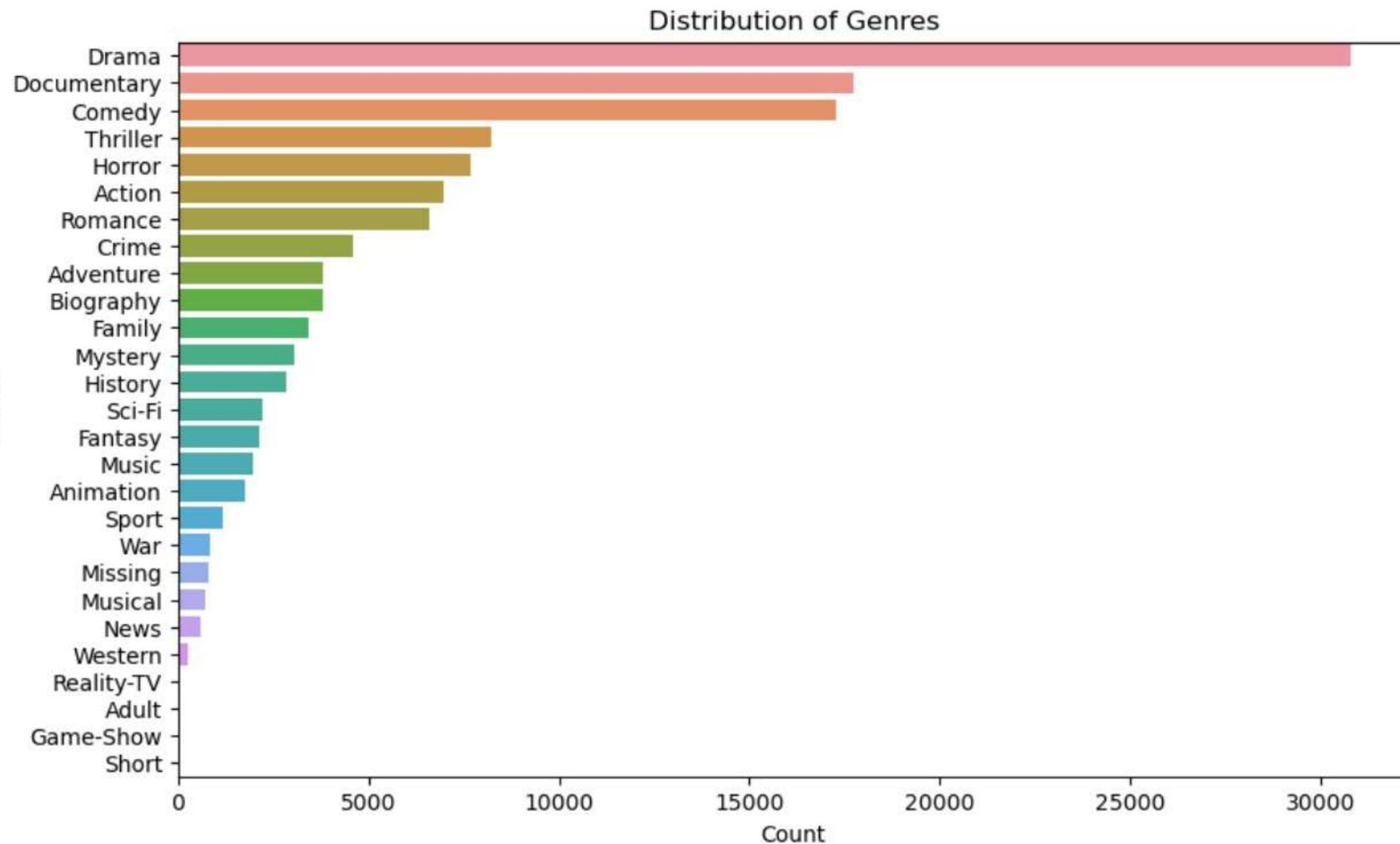
## Top 20 movies by worldwide gross earnings



- The top three movies by revenue are Avatar with approximately \$2.8 Billion followed by Titanic at \$2.2 Billion and Star Wars with \$2.0 Billion

# Data Visualization Results

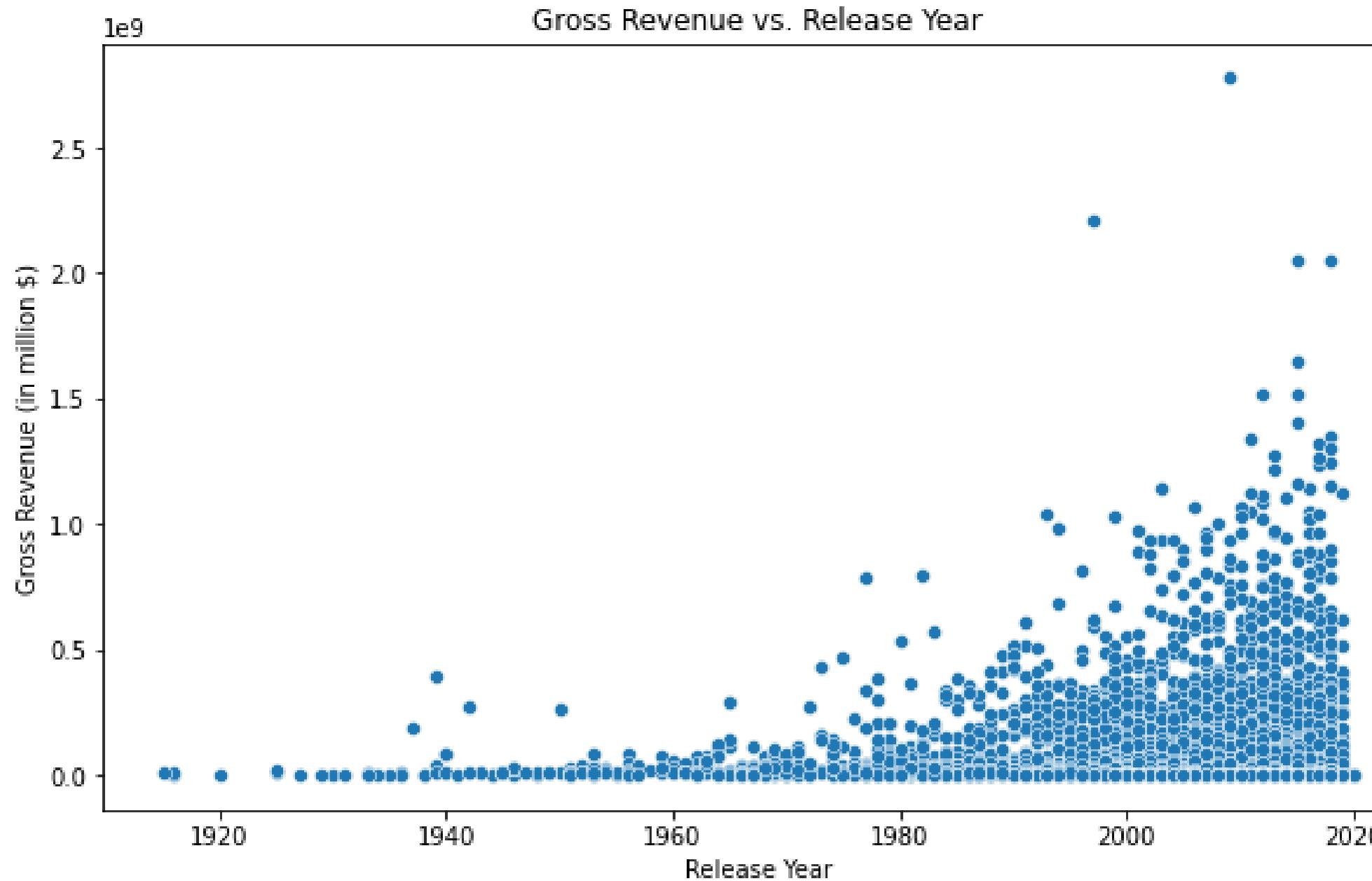
## Distribution of Genres



- The genres with the most distributed is Drama followed by Documentary and Comedy

# Data Visualization Results

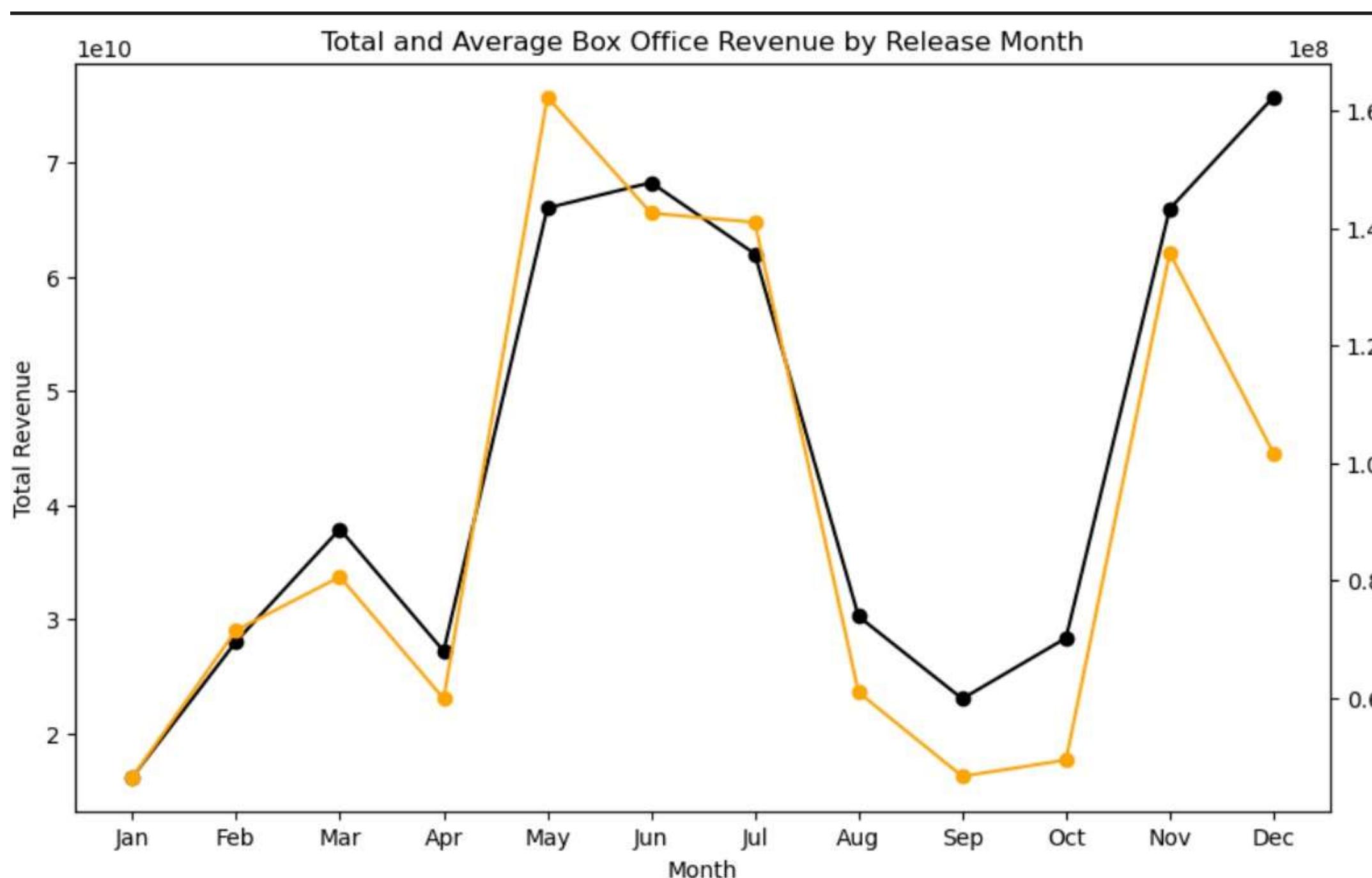
## Gross Revenue vs Release year



The plot above shows that the gross revenue from movies has generally increased from 1920 to 2020. This suggests that movies released in later years tend to earn more revenue.

# Data Visualization Results

## Total revenue release by Release month



**The ideal months to release movies**

**are:**

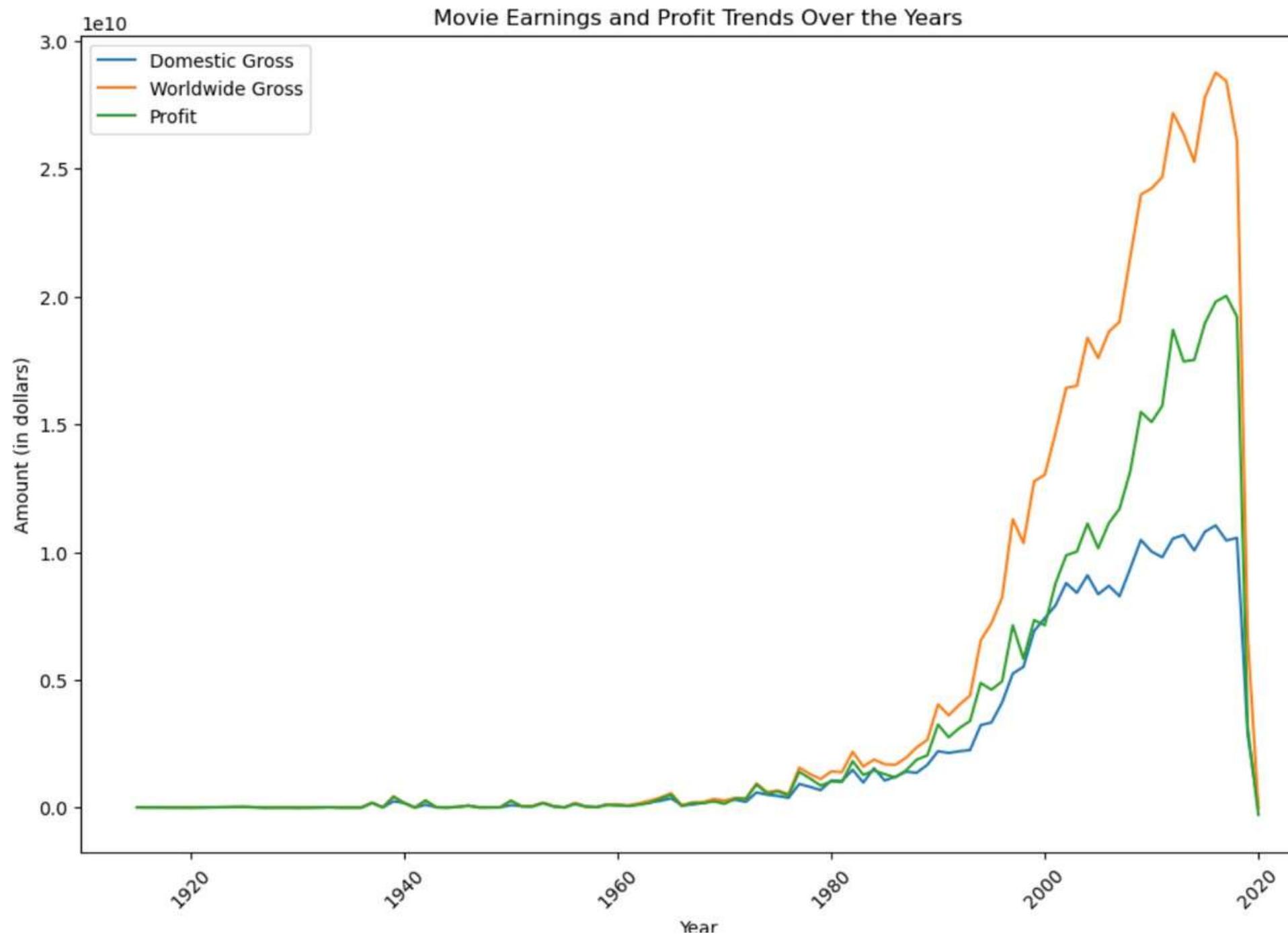
**May** has the highest average worldwide gross earnings,  
**June** and then **July**

\* These months could be considered the best time to release a movie for maximum revenue as they generate the highest worldwide gross revenue on average.

**December** has the highest total revenue

# Data Visualization Results

## Movie earnings and profit Trends over the years



Over time, we notice  
`domestic\_gross`, `worldwide\_gross`  
and `profit` have been generally  
increasing.

# **Recommendations**

1. Focus on **Action, Adventure, Animation, Comedy, Drama, Fantasy, Horror, Romance, Sci-Fi, and Thriller** genres, since they are the most profitable
2. **Prioritize Action** movies, as they are the most popular, since the top 5 popular movies are from the action genre
3. Target mid-range budgets for optimal ROI and prioritize international appeal in story and marketing.
4. Schedule movie releases in **May, June, July, November, or December**, since movies released in these months shown greater revenue
5. Tailor campaigns for domestic and international audiences separately and leverage high ratings when available for genres like Documentary or Drama.



# *Thank you*

