DSF-PT08: Group1

By: Gilbert Cheruiyot, Millicent Cheptoi, Edwin George, Mercy Ayub, James Muthee, Daniel Muigai, Angela Mwanzia, Reagan Adajo
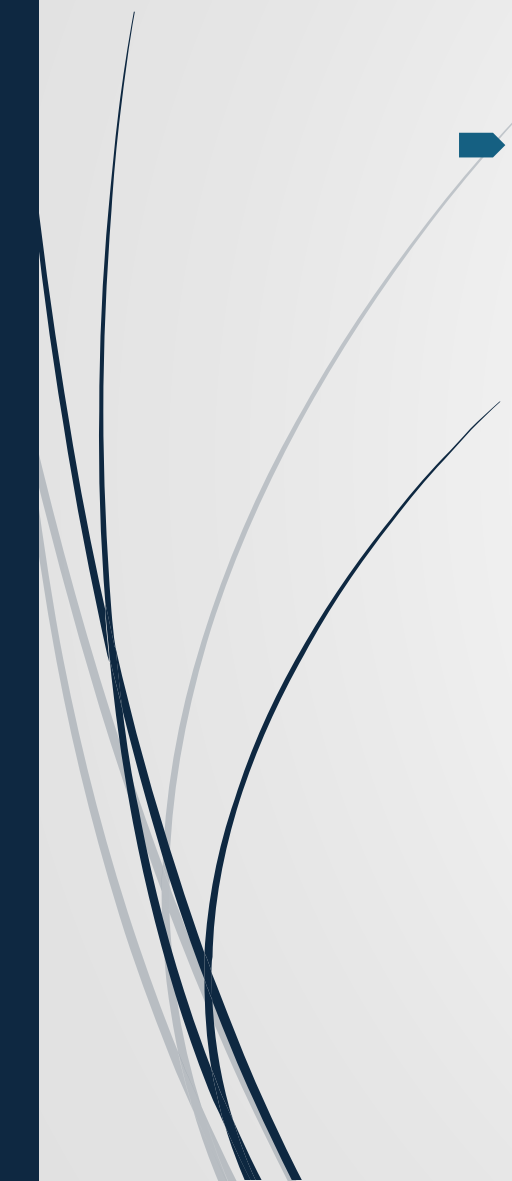
Agenda

- Overview
- Business Problem Definition
- Business Understanding
- Data Understanding
- Data Analysis
- Recommendations
- Next Steps
- Q&A (Thank You)

Business Problem

- Our company wants to invest in a new movie studio, but they don't know anything about creating movies. A research is to be done to explore what types of films are currently doing the best at the box office. And translate the findings into actionable insights to help decide what type of films to create.

Objectives

1. Determine current performing best film genres at the box office.

2. Look at trends in audience preferences and box office performance over recent years.

3. Understand <mark>Audience Demographics</mark> by identifying the demographics of audiences for top-performing genres.

4. Evaluate production costs associated with different genres.

5. Understand the relationship between production cost and revenue streams of movies

6. Evaluate if movie rating affect production costs.

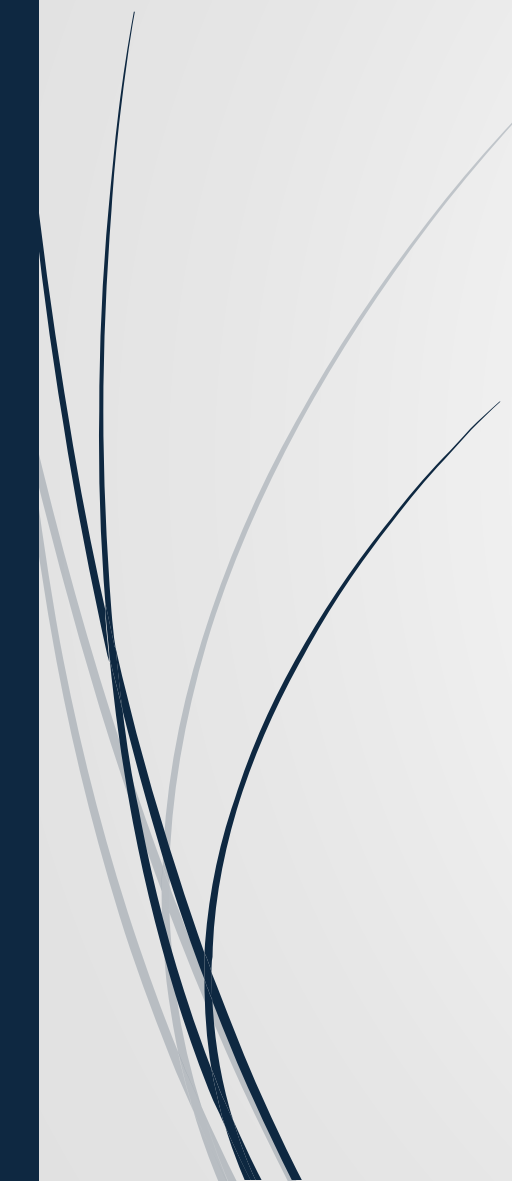7. Provide actionable insights for the new movie studio.

Key questions to answer

1. What are the trends in movie release dates and what impact do they have on revenue?

2. Is there a relationship between production budget and profitability and ROI of a movie? i.e. Does a higher production budget automatically result to a higher profitability and vice versa.

Steps to achieve objectives

1. *Research Box Office Data*: Gather genre-specific box office revenue data.

2. *Analyze Historical Performance*: Identify trends in genre performance.

3. *Evaluate Production Costs*: Examine typical budgets for each genre.

4. *Compile Findings*: Summarize top genres, trends, audience demographics, and costs.

5. *Provide Recommendations*: Suggest genres for the new studio to consider based on the analysis
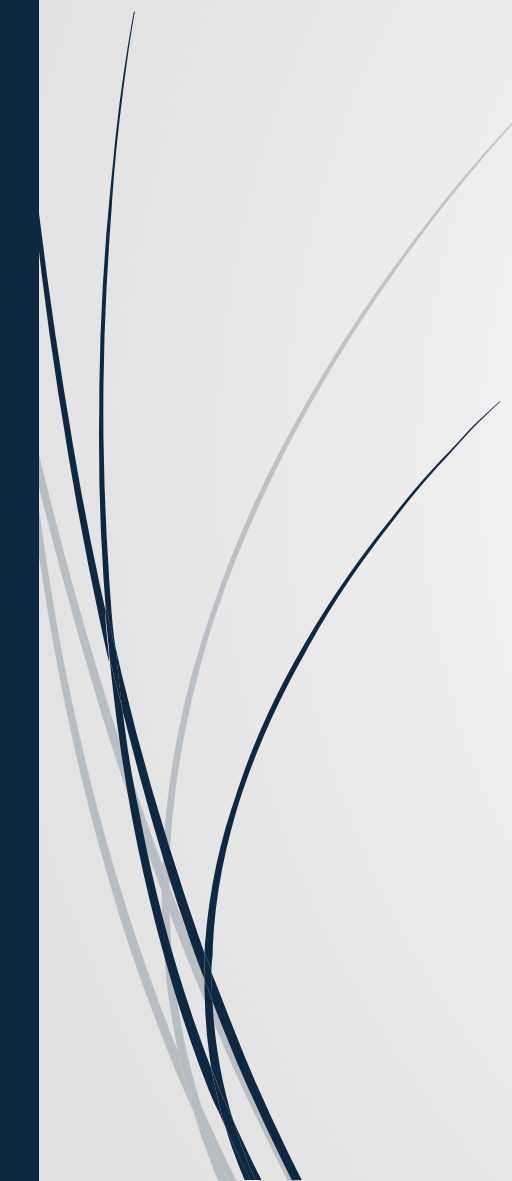
Data Understanding

- For this study, the datasets namely Movie Gross, Movie Budget and Movie Info has been used to arrive at actionable insights for the movie studio project of the company.

- Data:

  - **movie_gross**: Contains data on movie gross earnings

  - **movie_budget**: Contains information about movie budgets

  - **movie_info**: Additional movie metadata

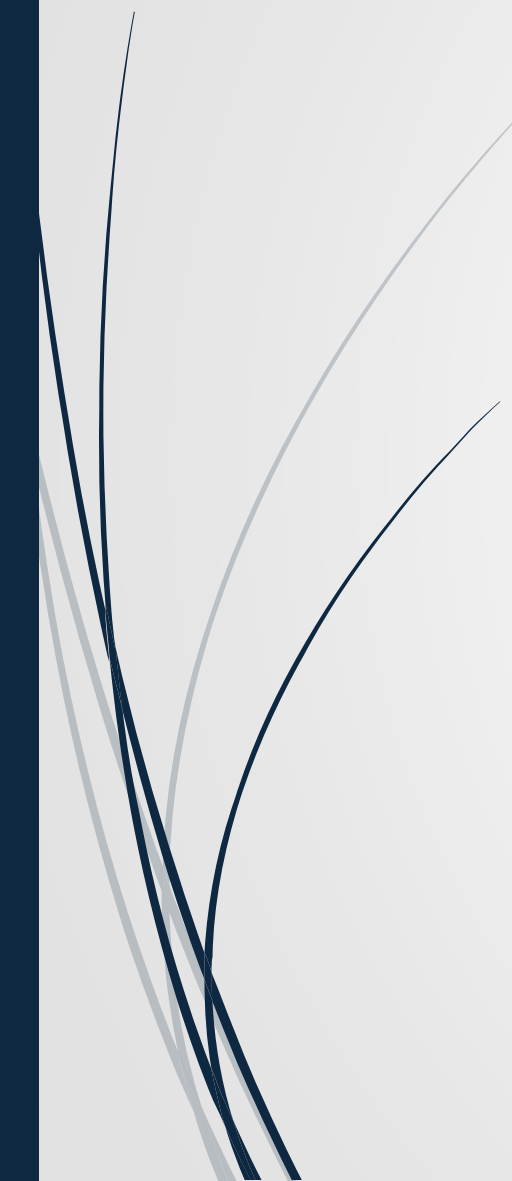 - **im.db**: SQL database containing additional tables for analysis.

Data Preparation and Cleaning

- The analysis involved understanding data spread, identifying inconsistencies, and missing values.

- Key statistics determined dataset completeness and reliability.

- Libraries, movie datasets, and Jupyter Notebook were loaded, and missing values were fixed.

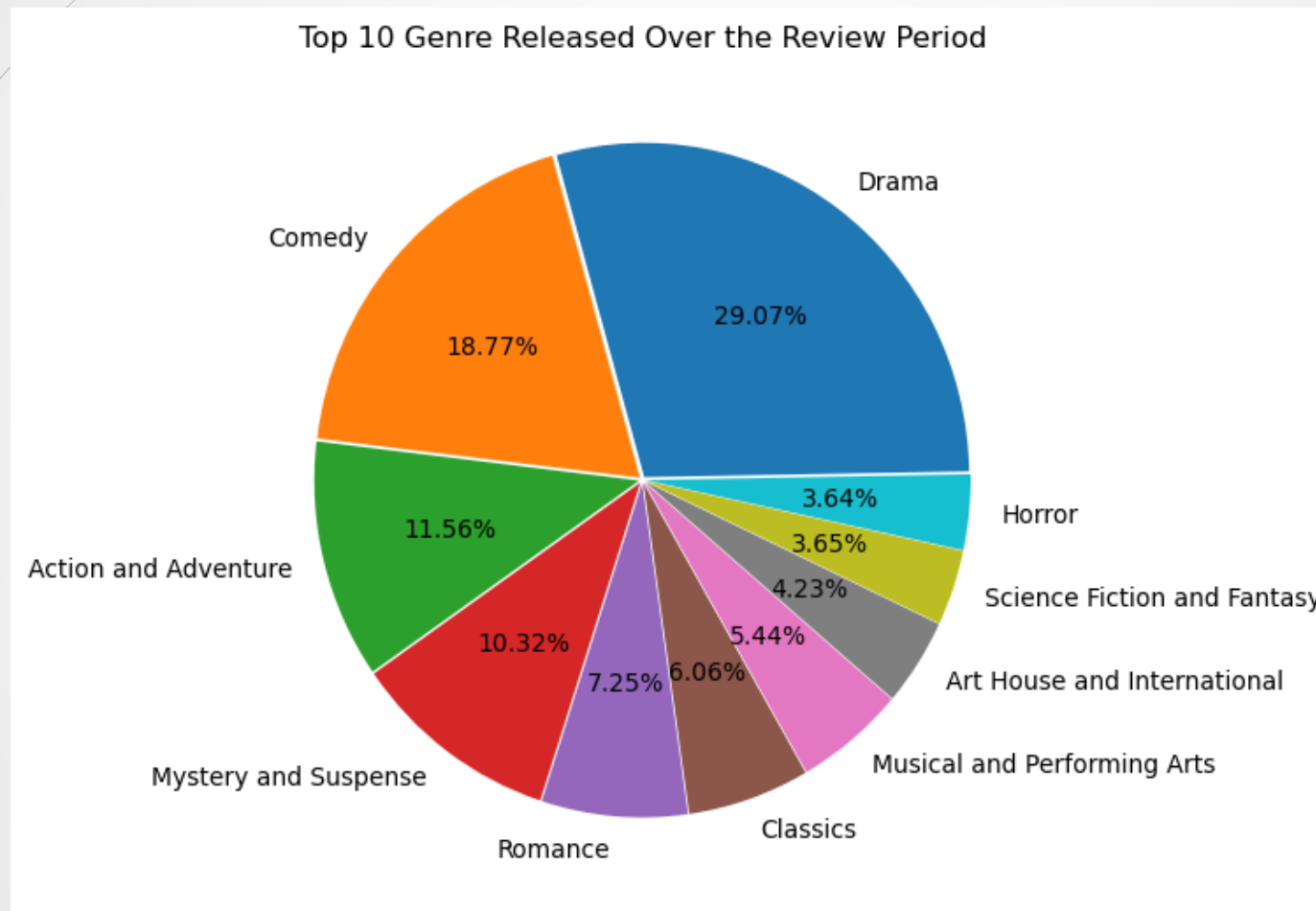- Feature engineering was performed, merging important data frames.

Exploratory Data Analysis

- Summary statistics, correlation analysis, and standard deviation were used in an analysis to understand the distribution of movie information, budgets, and revenues across different genres.

- Trend analysis revealed seasonal patterns, while genre analysis revealed common and rare trends.

- Correlation analysis revealed relationships between variables, and heatmaps visualized movie performance by release date.

The top 10 genres

Top 10 Genre Released Over the Review Period

- Drama — 29.07%
- Comedy — 18.77%
- Action and Adventure — 11.56%
- Mystery and Suspense — 10.32%
- Romance — 7.25%
- Classics — 6.06%
- Musical and Performing Arts — 5.44%
- Art House and International — 4.23%
- Science Fiction and Fantasy — 3.65%
- Horror — 3.64%

The data indicates the top 10 most released genres with Drama being the most released genre during the review period.
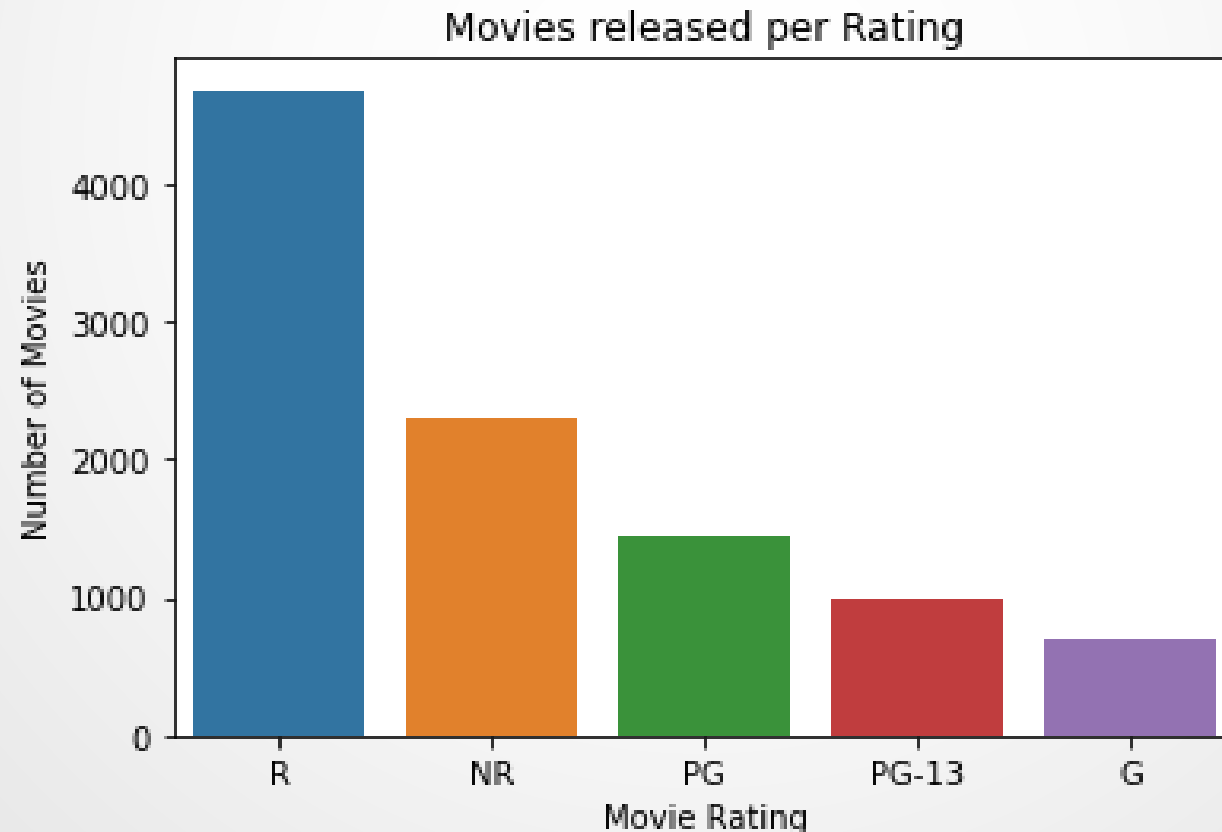
Production Numbers per Year

- ddd
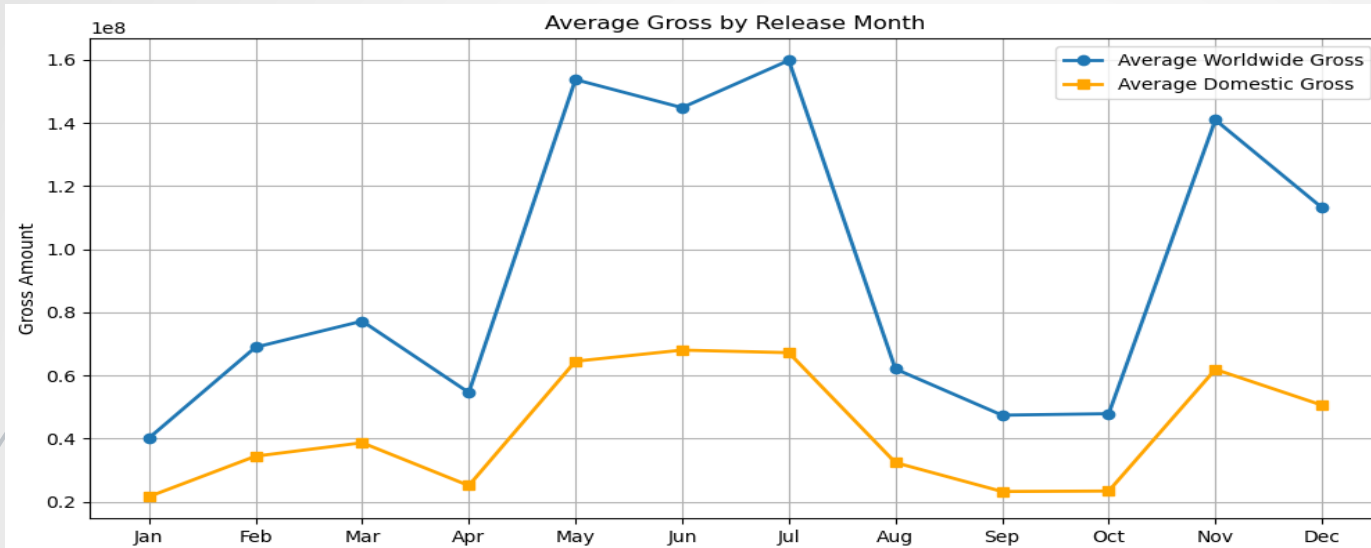
Number of Movies Produced per Year
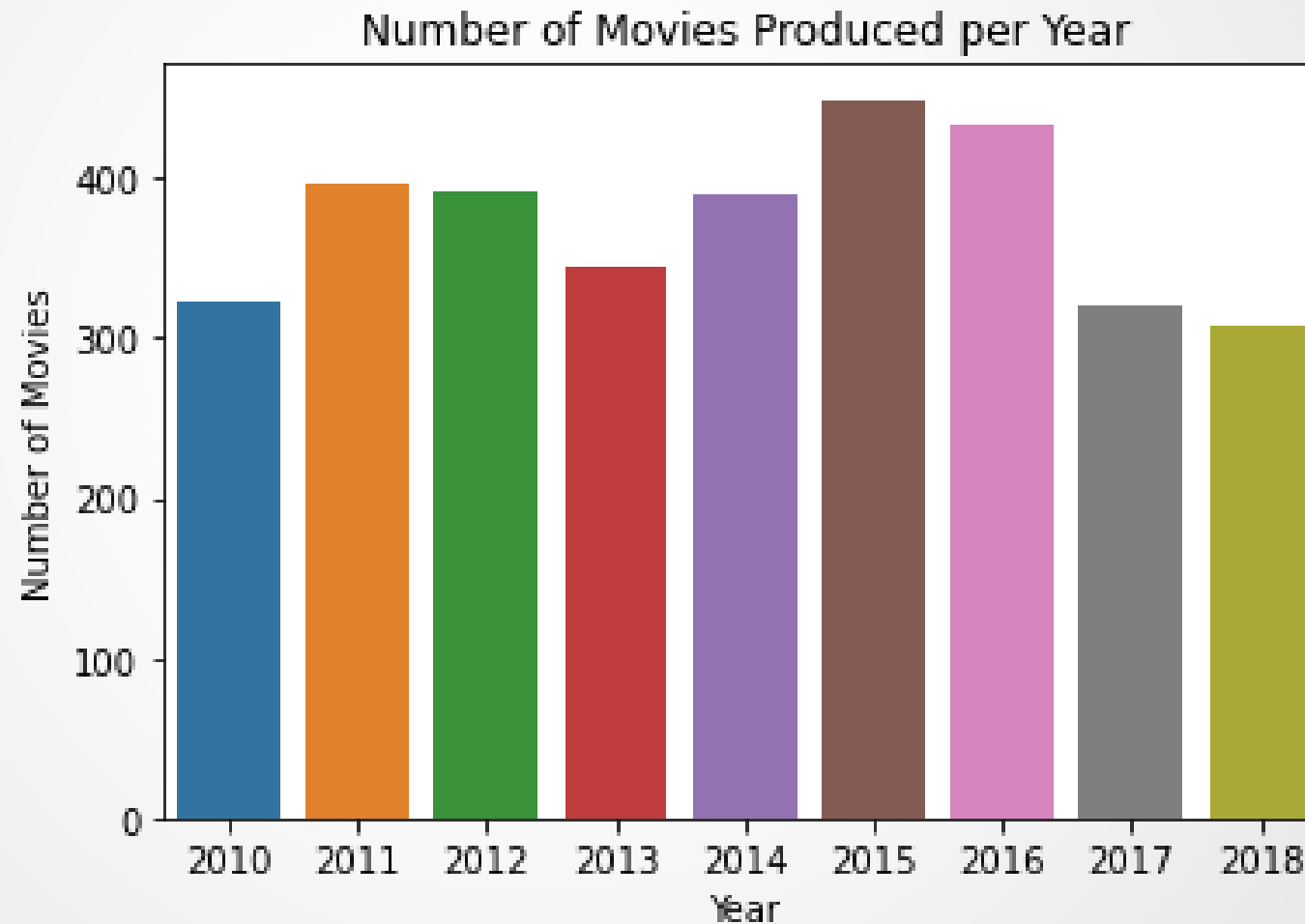
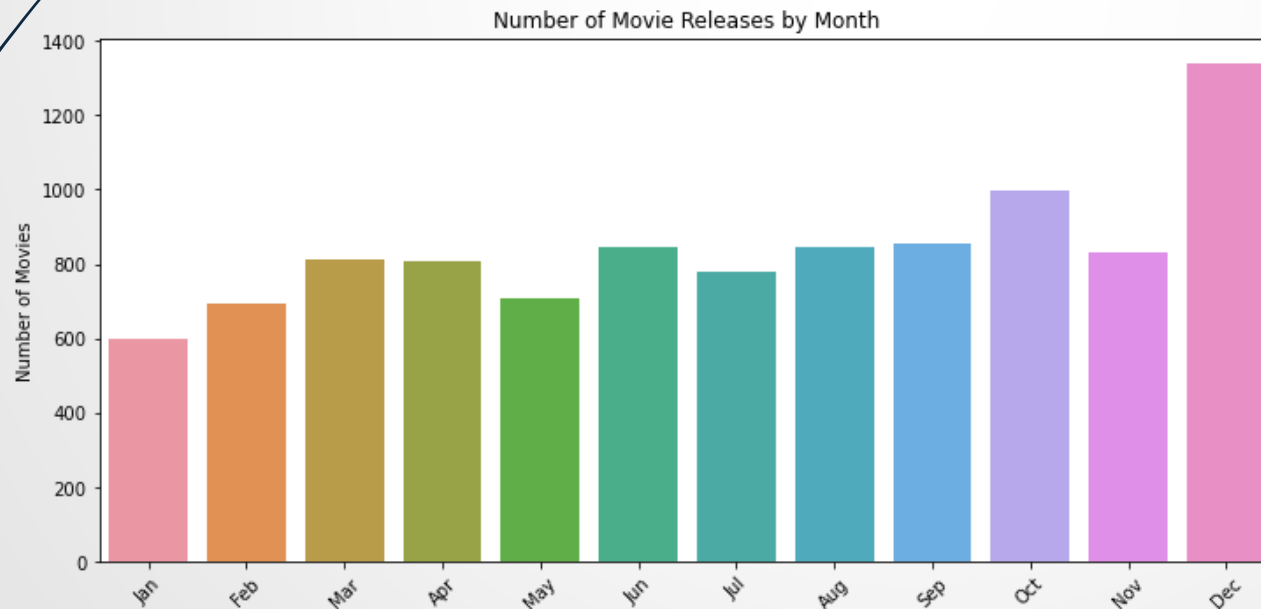# Movies released as per Ratings

Majority of the films fall into one of two categories: restricted (R) or nonrestricted (NR).

According to the research, majority of the films made are intended for adults.

Revenue per Release Month

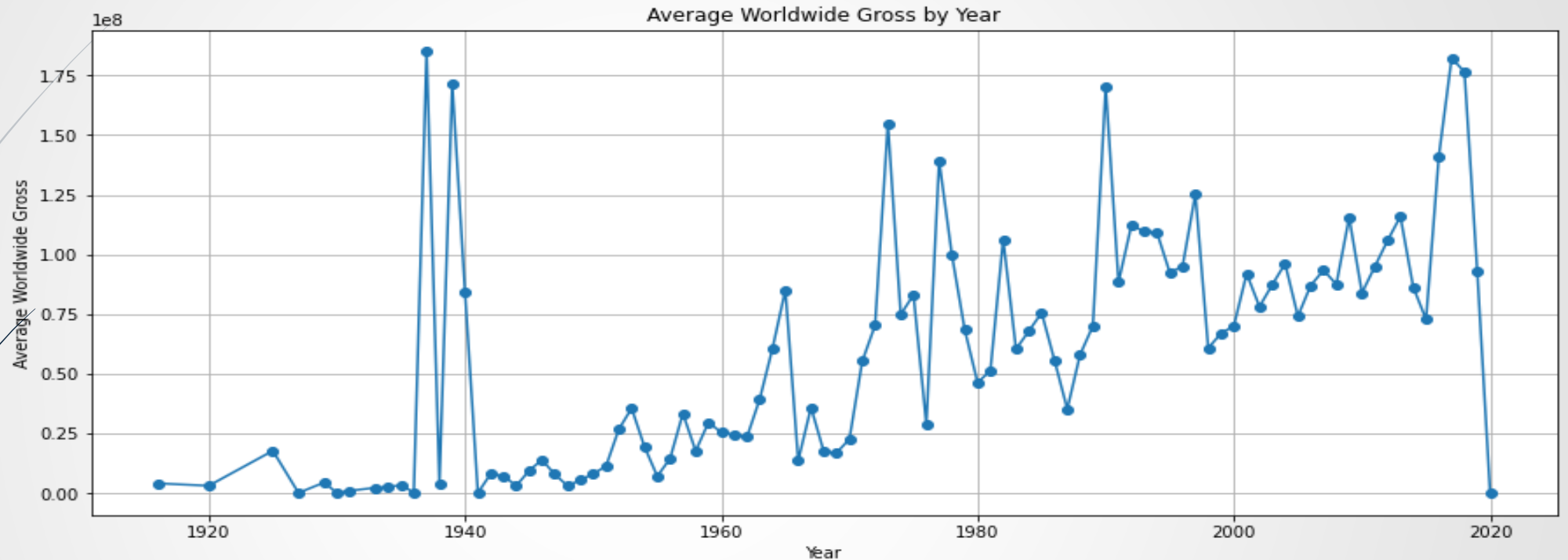Months between May-July consistently show the highest average worldwide and domestic gross.

The data shows more movies are released during the month of December

Worldwide gross performance during the review period

Average Worldwide Gross by Year

- The worldwide gross increases over the years.

- A plunge during the year of 2020: An event might have happened which needs further research

Genre performance by revenue

Average Worldwide Gross by Genre

Average Movie Domestic Gross Revenue by Genre
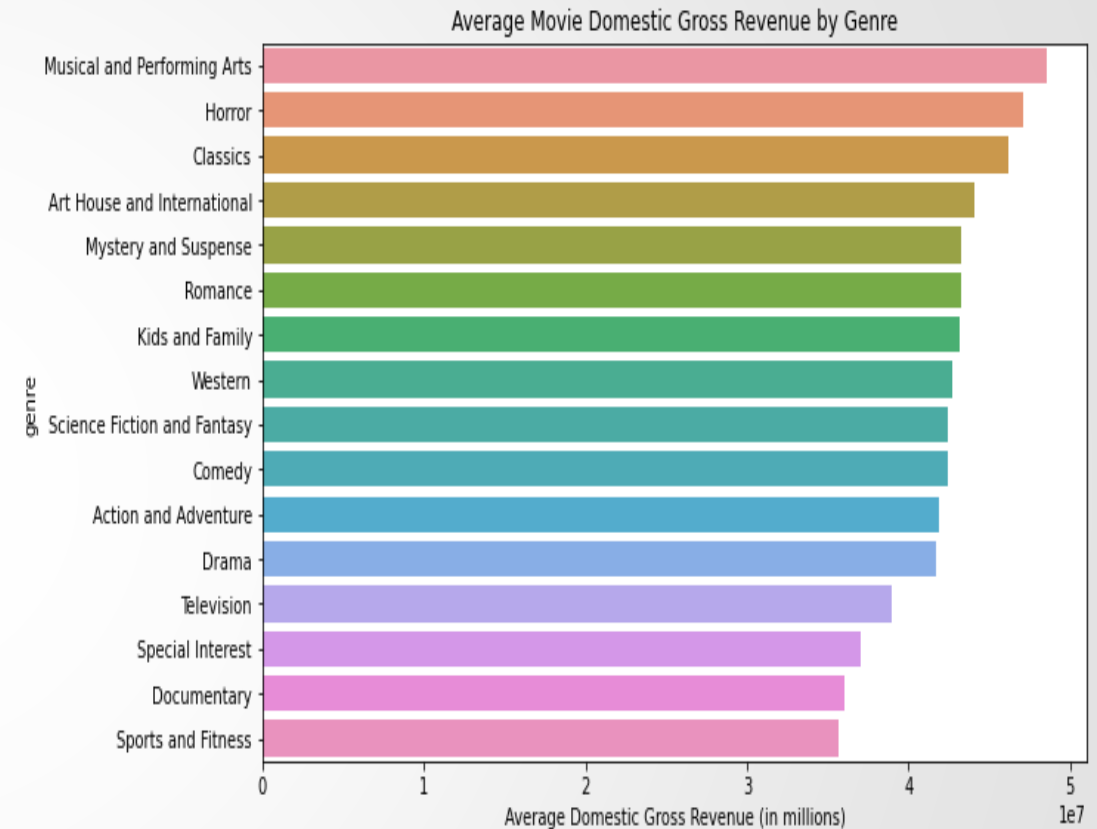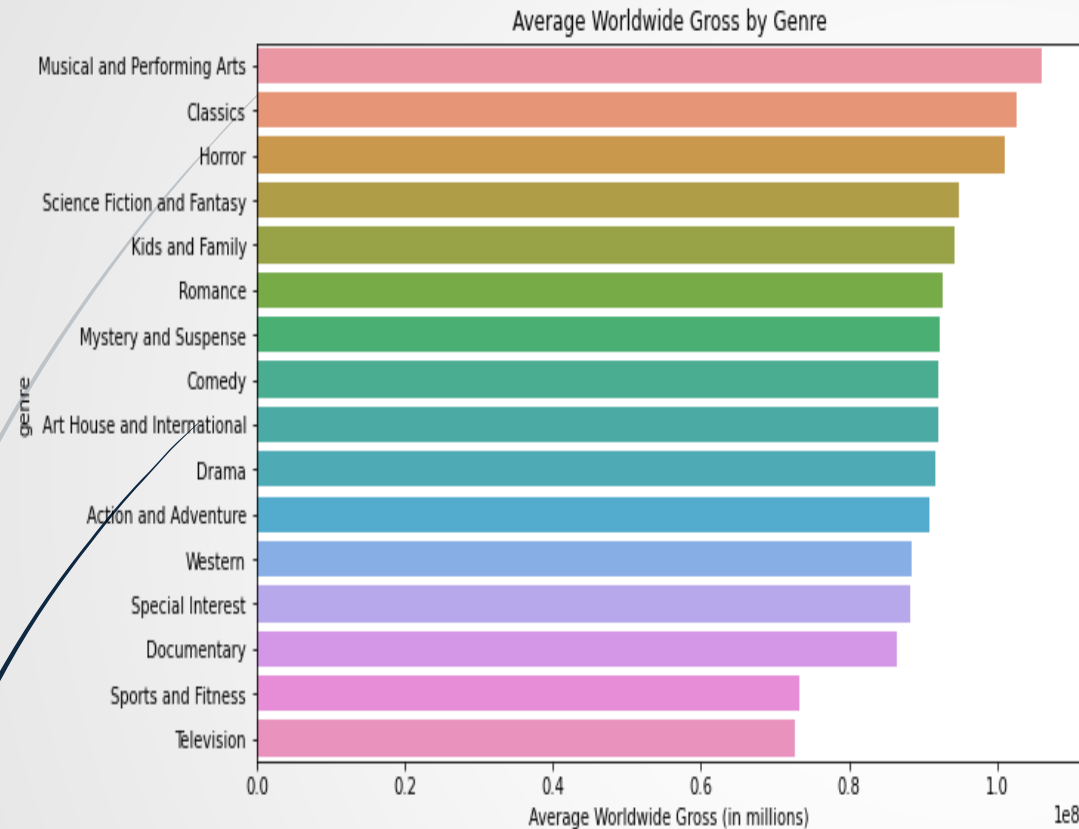
Musical and Performing Arts consistently outperform other genres in domestic and worldwide gross revenue, while Horror, Science Fiction, and Fantasy show strong performance.

Statistical Analysis – Movie gross
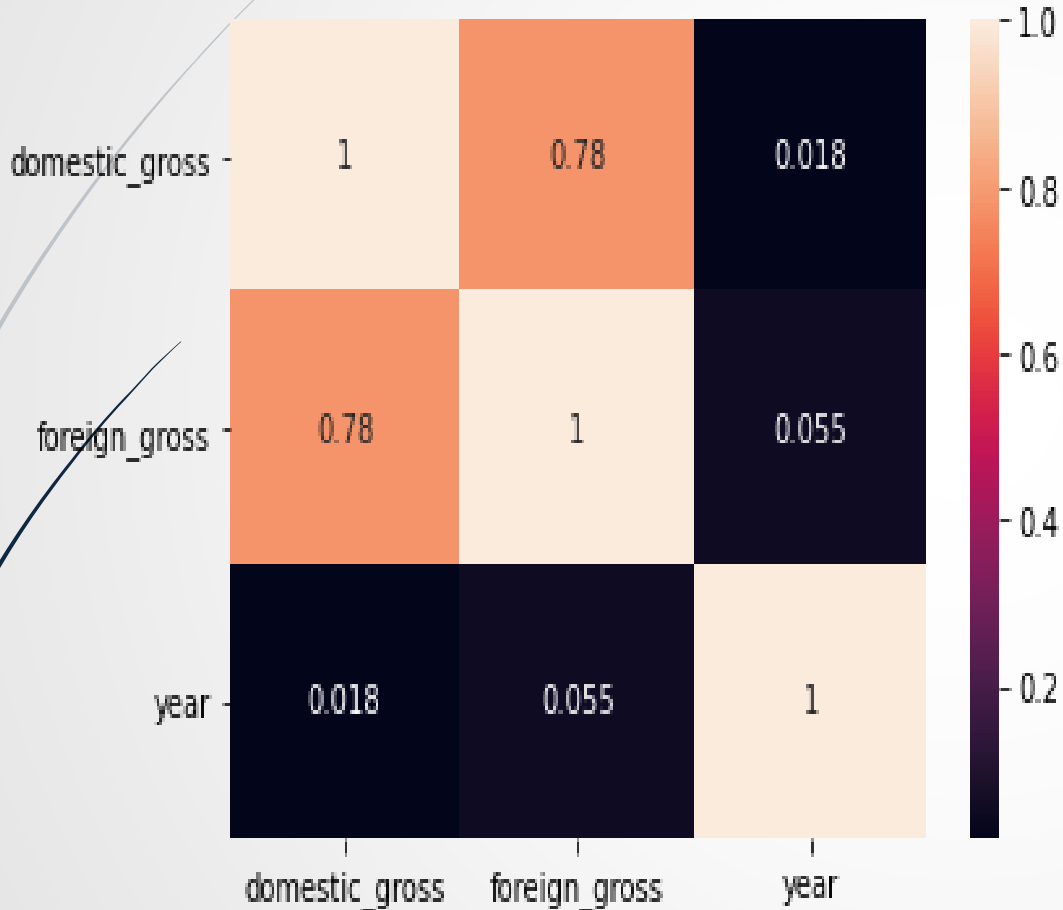
- ► Strong Relationship Between Domestic and Foreign Earnings
  - • Domestically successful movies generally may perform well internationally, but the year of release doesn't significantly impact their earnings.
- ► Conclusion
  - • The strong correlation between domestic and international movie earnings is crucial for predicting its global success

production budget vs domestic gross and worldwide gross

➥ The relationship between the production budget with both domestic gross and worldwide gross indicates that a higher production budget is likely to boost both domestic and worldwide gross revenues for movies
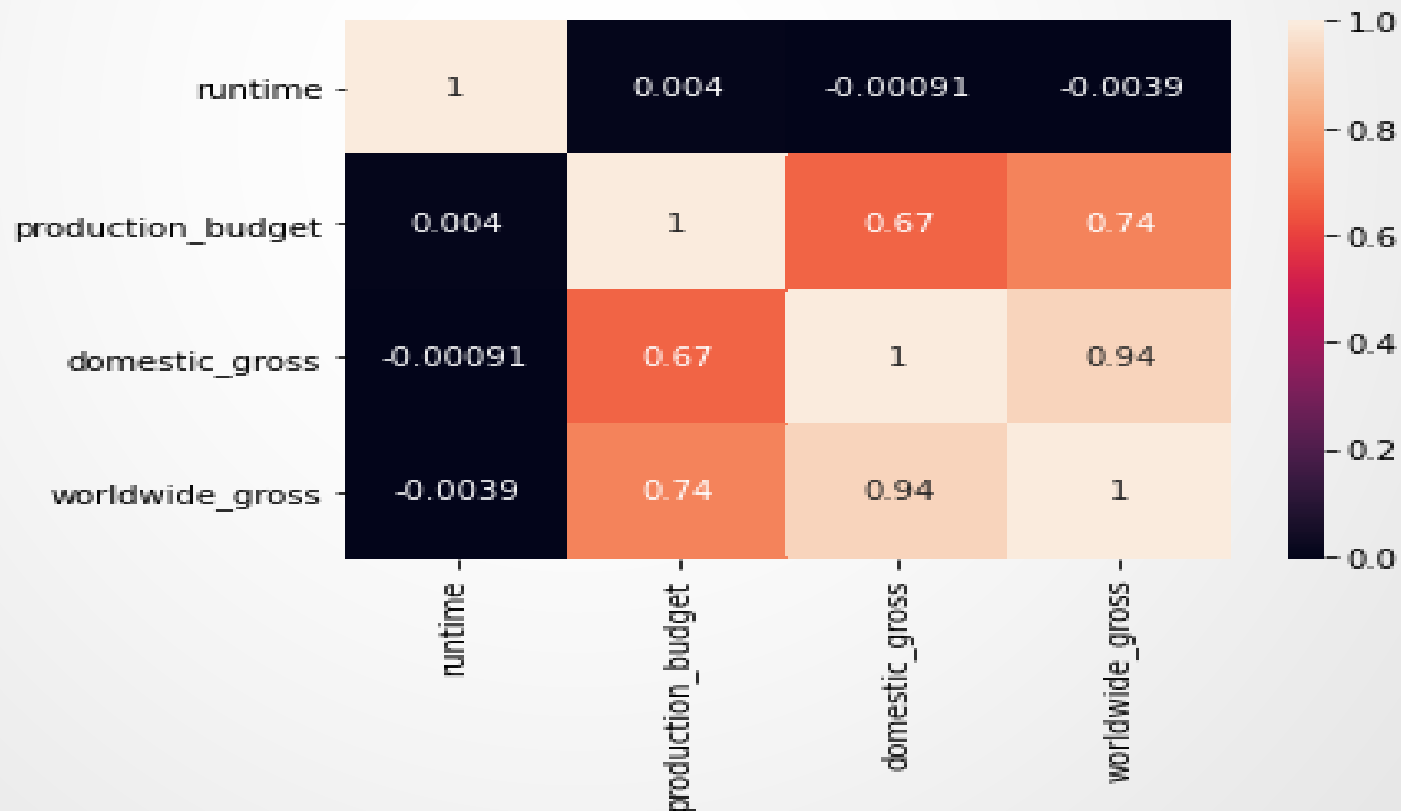
Revenue trends per genre
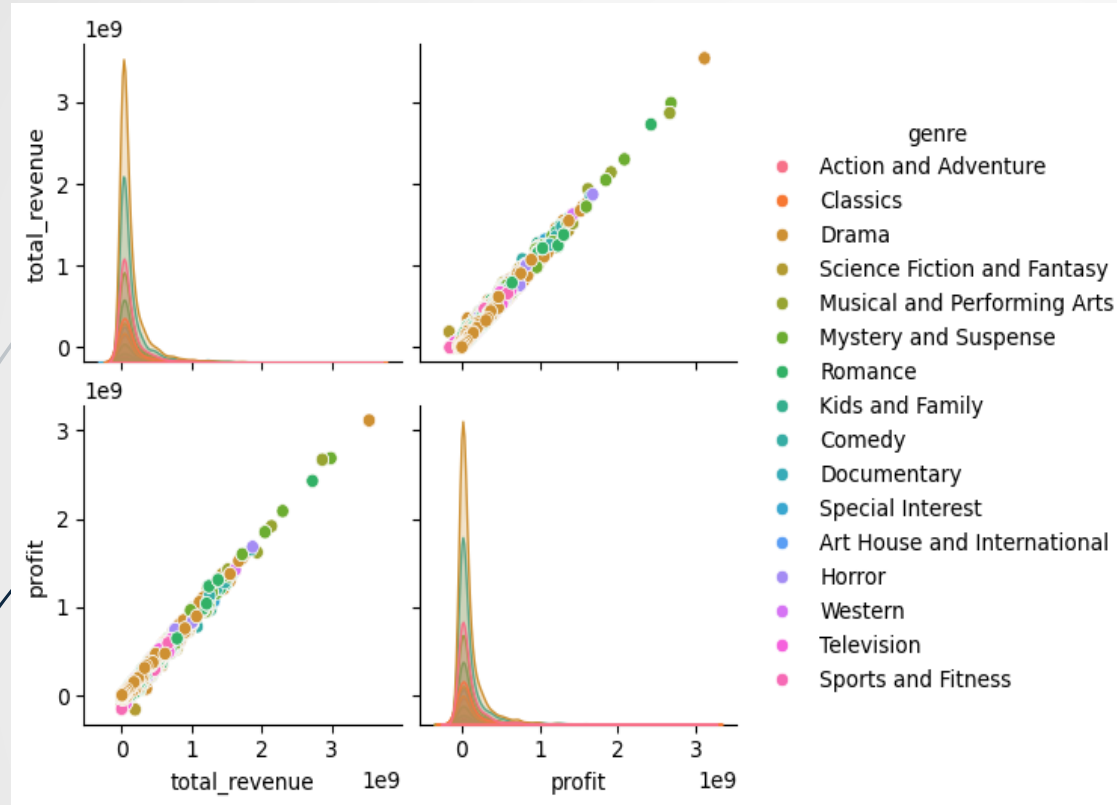
Revenue and Profit Relationship: As revenue increases, profit tends to rise proportionally

Genre-Based Trends: Certain genres, like Drama, Science Fiction, and Fantasy, often reach higher revenue and profit levels. In contrast, genres like Documentary and Special Interest usually perform on the lower end financially. Action, Adventure, Science Fiction, and Fantasy show high potential for greater profitability and revenue.

Production Budget vs Total Revenue

Relationship between Total Revenue and Production Budget
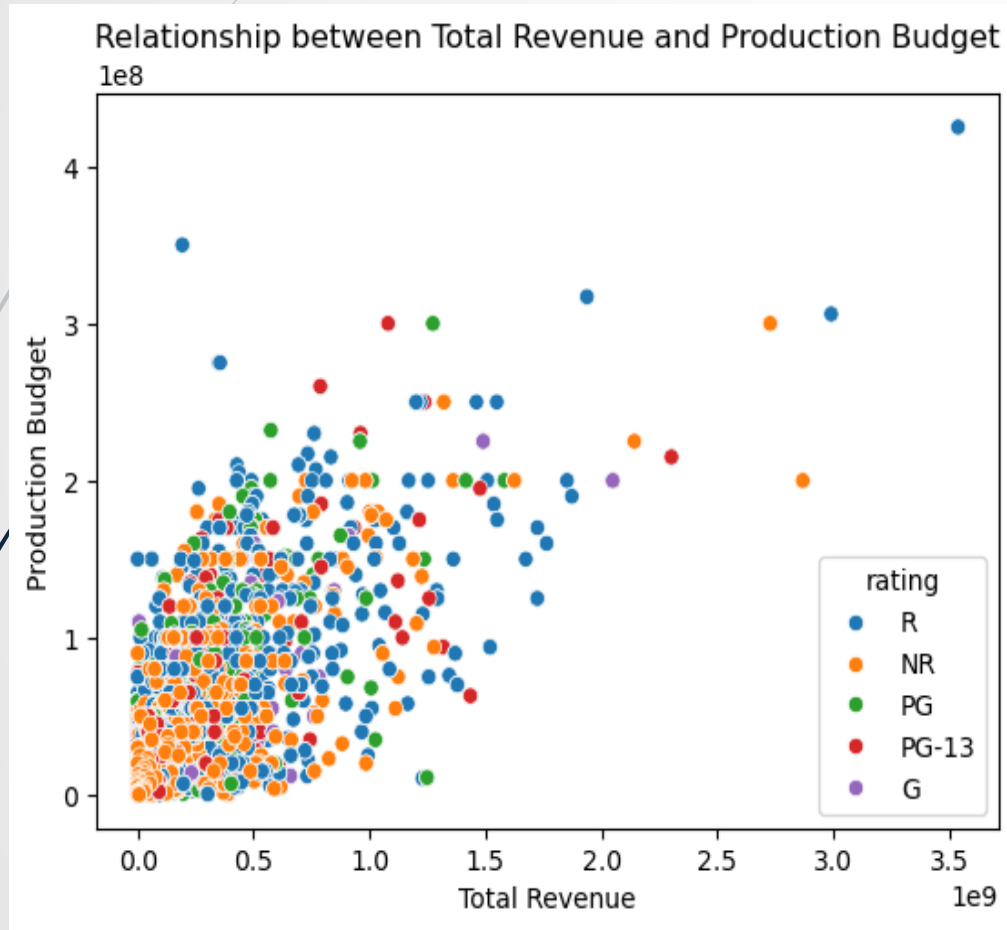
- Movies with higher production budgets tend to earn more revenue, although most films remain within a lower budget and revenue range

- Outliers with high budgets and revenue are exceptional cases, likely representing blockbuster films.

- The plot shows that ratings have little correlation with production budgets or revenue

Relationships for the budget data
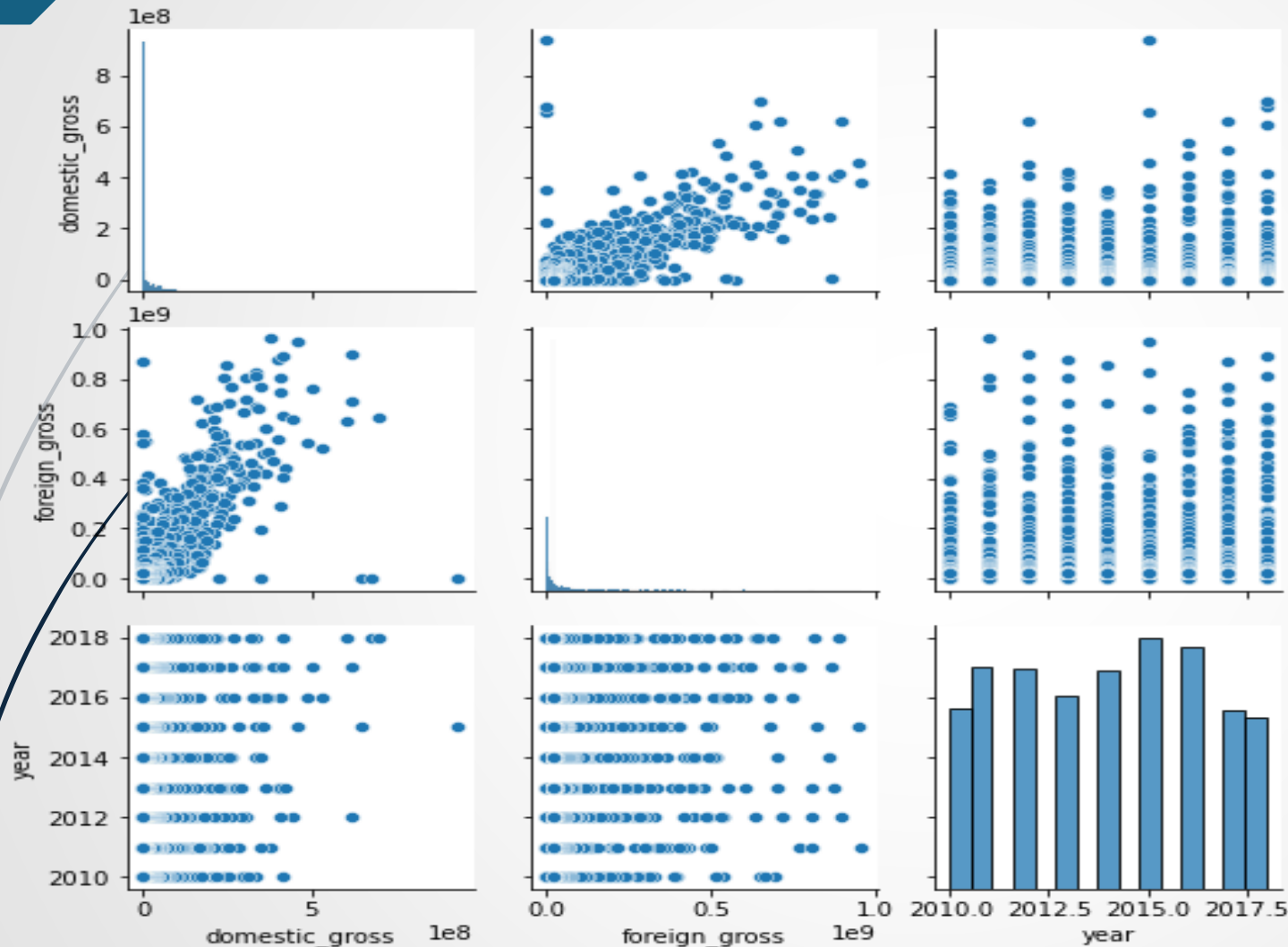
- There is a positive linear relationship between domestic gross and foreign gross.

- An increase in domestic gross indicates in an increase foreign gross

Data Modelling – Simple Linear Regression

- The Ordinary Least Squares(OLS) model was used to determine the relationships between different variables: domestic_gross, production_budget and worldwide_gross, foreign_gross.

- The model was used to predict various variables on the data.

Rainbow test – Domestic gross to predict worldwide gross

- Using this model, an R-squared of 88.2% shows that domestic revenue is a very reliable indicator for predicting worldwide revenue.

OLS Regression Results

Dep. Variable:	worldwide_gross	R-squared:	0.882
Model:	OLS	Adj. R-squared:	0.882
Method:	Least Squares	F-statistic:	7.581e+04
Date:	Thu, 14 Nov 2024	Prob (F-statistic):	0.00
Time:	02:05:22	Log-Likelihood:	-1.9571e+05
No. Observations:	10111	AIC:	3.914e+05
Df Residuals:	10109	BIC:	3.914e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-8.328e+06	7.16e+05	-11.637	0.000	-9.73e+06	-6.93e+06
domestic_gross	2.3684	0.009	275.342	0.000	2.352	2.385

Omnibus:	7952.448	Durbin-Watson:		1.794
Prob(Omnibus):	0.000	Jarque-Bera (JB):		596457.368
Skew:	3.214	Prob(JB):		0.00
Kurtosis:	40.074	Cond. No.		9.71e+07

- 88.2% of the variability in worldwide gross revenue can be explained by domestic gross revenue.

- According to the model, if the company wants to start earning worldwide gross revenue, it should aim to generate at least $3.5 million in domestic gross revenue.

Using <mark>Multi</mark>linear regression - production cost

```
                           OLS Regression Results
===============================================================================
Dep. Variable:        production_budget   R-squared:                    0.551
Model:                            OLS     Adj. R-squared:               0.551
Method:                 Least Squares     F-statistic:                  6193.
Date:                Thu, 14 Nov 2024     Prob (F-statistic):           0.00
Time:                        02:05:23     Log-Likelihood:          -1.8760e+05
No. Observations:               10111     AIC:                      3.752e+05
Df Residuals:                   10108     BIC:                      3.752e+05
Df Model:                           2
Covariance Type:            nonrobust
===============================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
const            1.648e+07   3.23e+05     50.985      0.000    1.58e+07    1.71e+07
domestic_gross     -0.1087      0.011     -9.661      0.000      -0.131      -0.087
worldwide_gross     0.2102      0.004     47.104      0.000       0.201       0.219
===============================================================================
Omnibus:                     3417.810   Durbin-Watson:                1.234
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         36708.047
Skew:                           1.309   Prob(JB):                      0.00
Kurtosis:                      11.960   Cond. No.                  2.56e+08
===============================================================================
```

➤ An R-squared of 0.551 means that 55.1% of the **production budget** differences can be predicted based on a movie's earnings **domestically** and **worldwide**. 44.9% of the budget differences are influenced by other factors that weren't included in this analysis.

➤ Earnings provide a glimpse into a movie's budget, but they don't reveal the full picture; other factors significantly influence production costs

Recommendations

- To maximize revenue, the movie studio should focus on releasing their films in the months with the highest revenue, such as December, June, and July.

- They should prioritize releasing high-budget films with a high return on investment (ROI) to ensure their production budgets are not wasted.

- They should invest in marketing and promotional efforts to reach a wider audience and generate more revenue.

- They should prioritize releasing their films in the top performing genres, such as Musical, Performing Arts, Horror, Science Fiction, and Fantasy. This will help maximize their box office revenue and attract more viewers.

Next Steps

- Thank You ☺