



DSF-PT08: Group 1

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



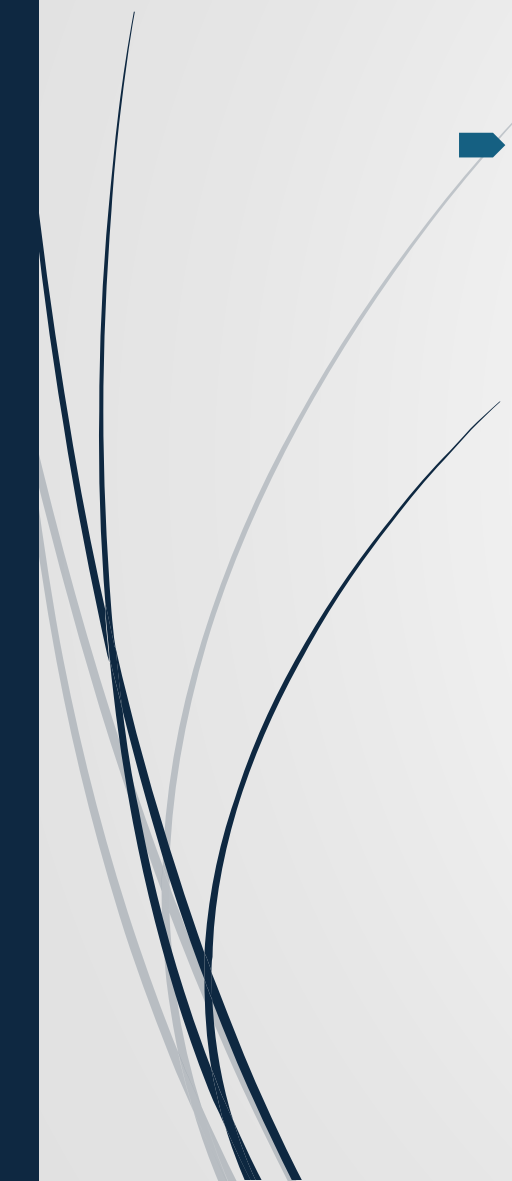
Agenda



- 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. Establish trends in audience preferences and box office performance over recent years.
3. Evaluate production costs associated with different genres.
4. Understand the relationship between production cost and revenue streams of movies, runtime, rating
5. Evaluate if movie rating affect production costs.
6. Provide actionable insights for the new movie studio.



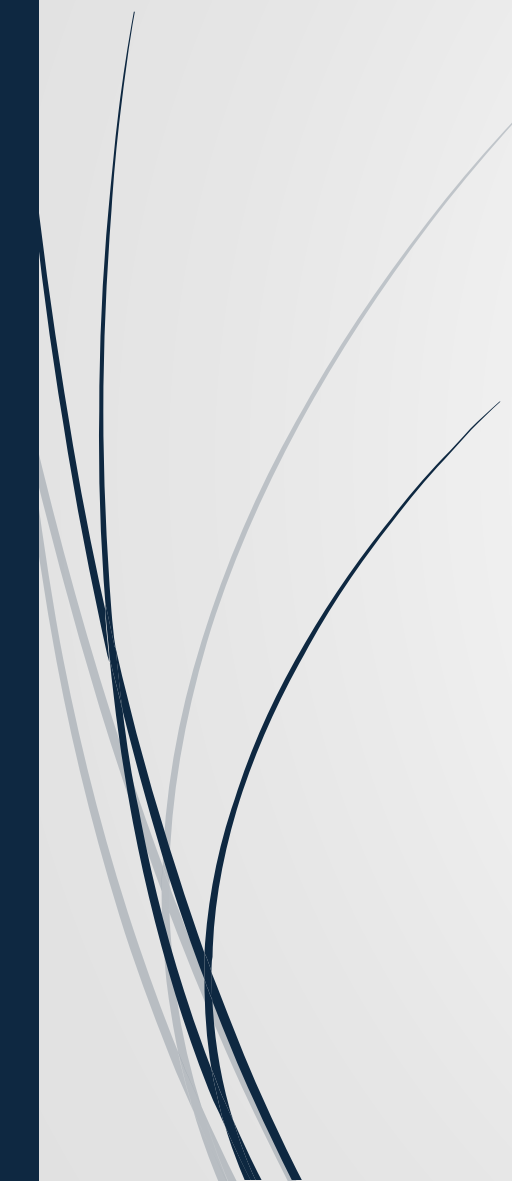
Key questions to answer



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 (Return on Investment) of a movie?
3. What is the relationship between foreign gross and domestic gross?



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



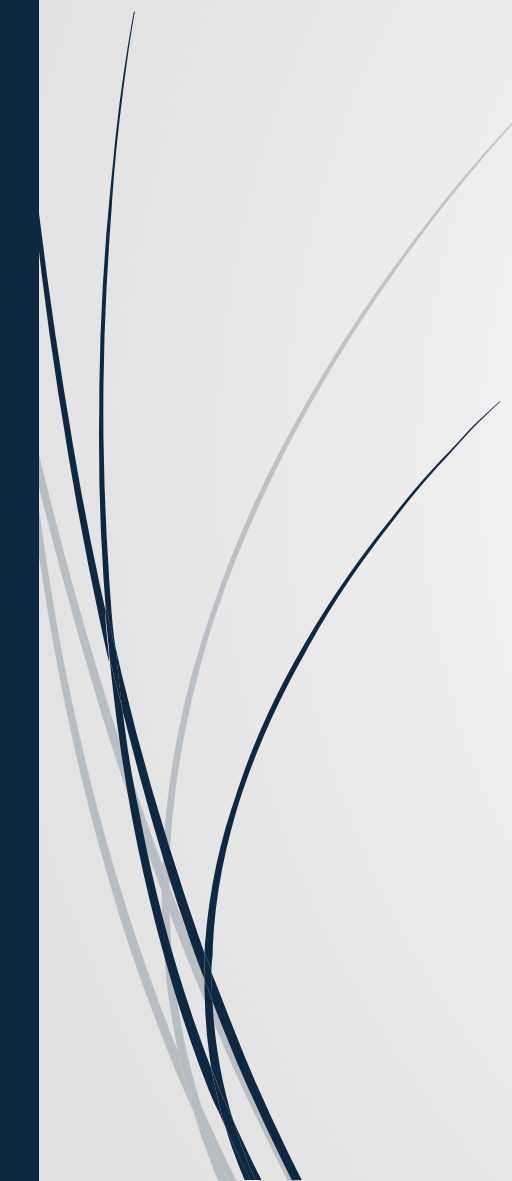
Data Understanding



- For this study, the datasets namely *Movie Gross*, *Movie Budget*, *Movie Info* and *imdb* have 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.

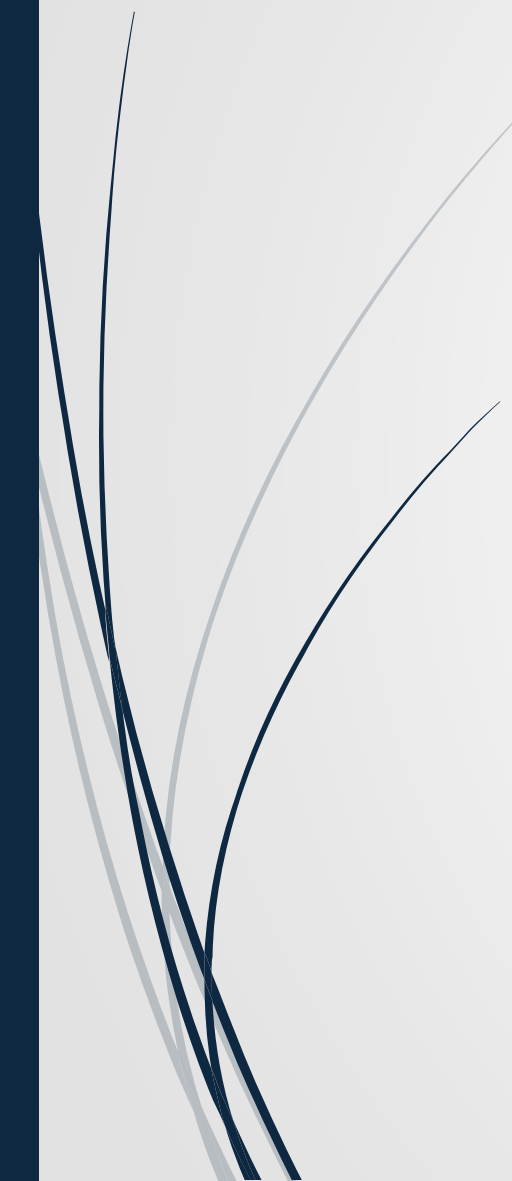


Data Preparation

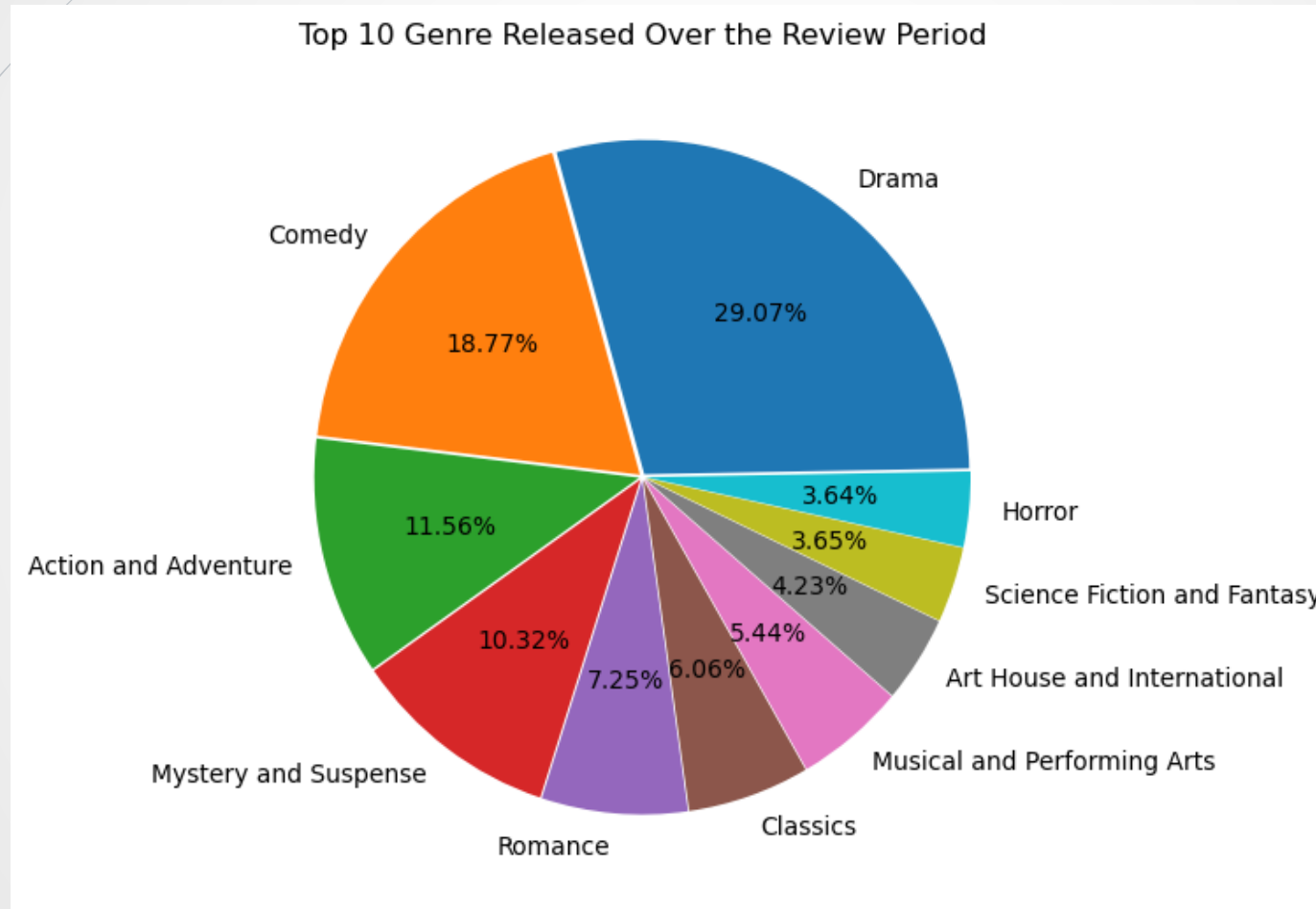
- The analysis involved understanding data spread, identifying inconsistencies, and missing values.
 - Key statistics determined dataset completeness and reliability.
 - Libraries and movie datasets were loaded on Jupyter Notebook and missing values were fixed.
 - Feature engineering was performed, merging 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.
- 

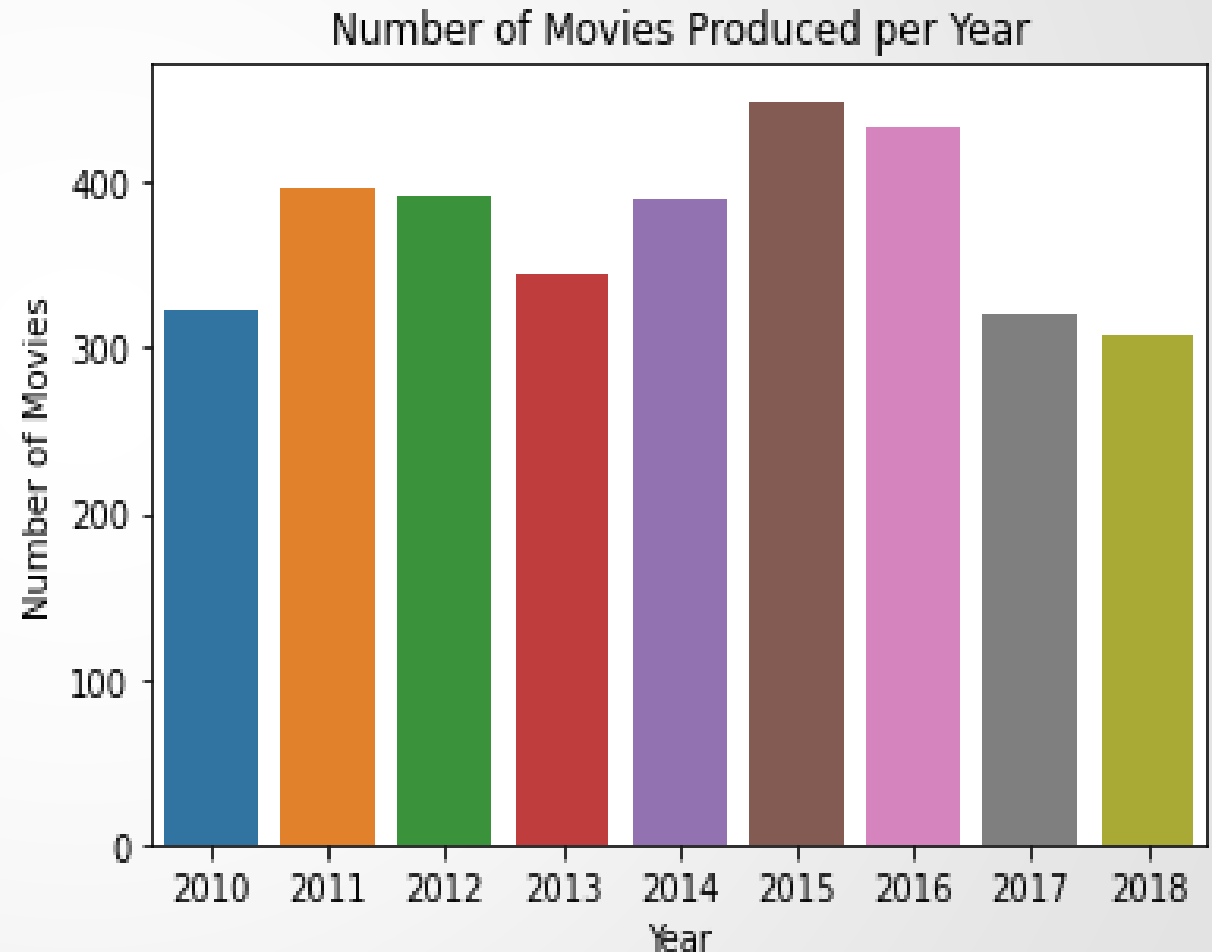
Top 10 genres within the review period



Drama was the most released genre during the review period followed by Comedy and Action & Adventure genres, respectively.

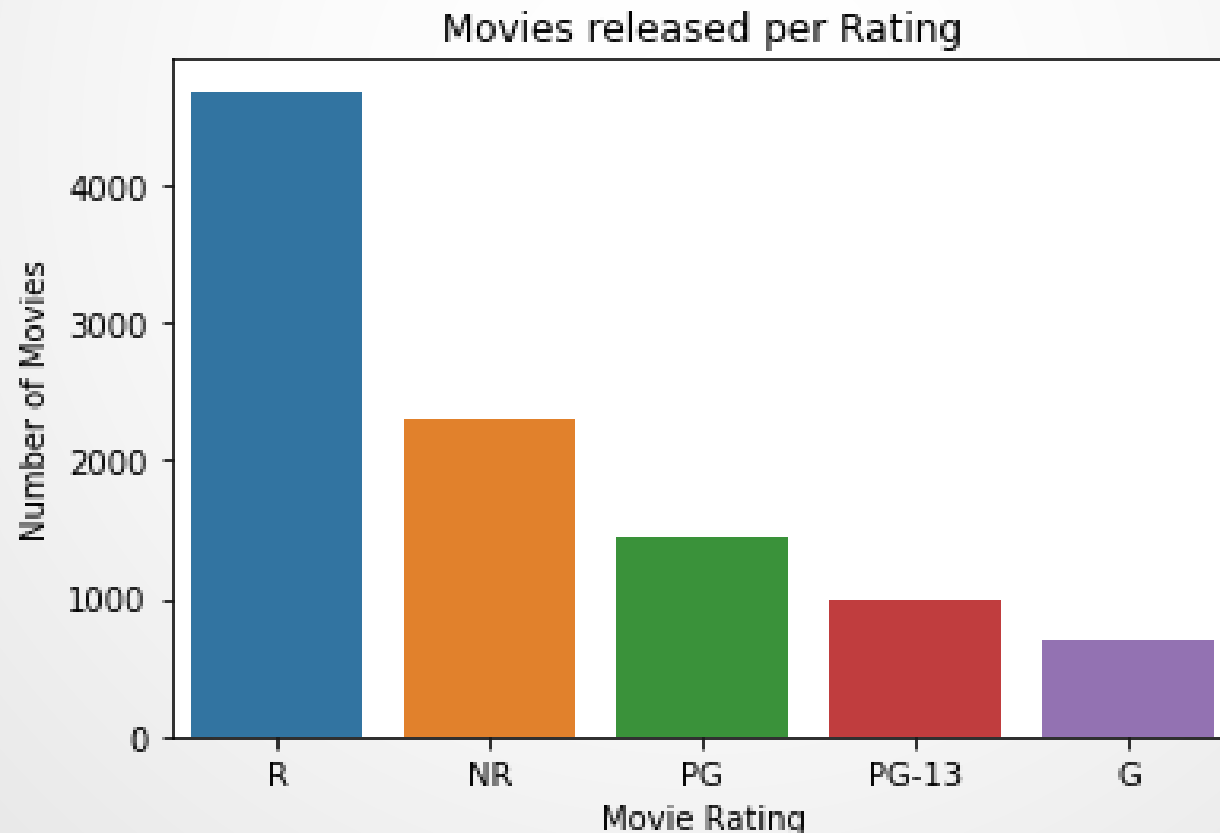
Production Numbers per Year

- The number of movies produced over the review years is cyclical with the peak periods during 2011 and 2015. the troughs in 2010, 2013 and 2018.
- In conclusion, peaks and troughs of the business cycle are collectively known as turning points, which provide valuable information about the state of the economy.

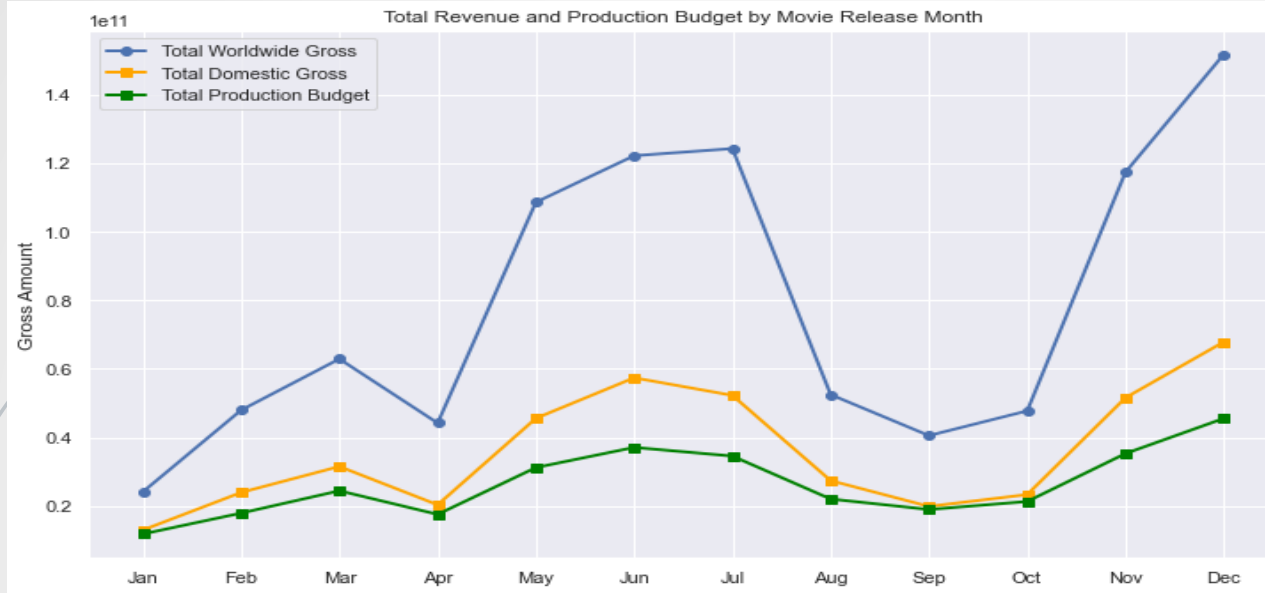


Movies released as per Ratings

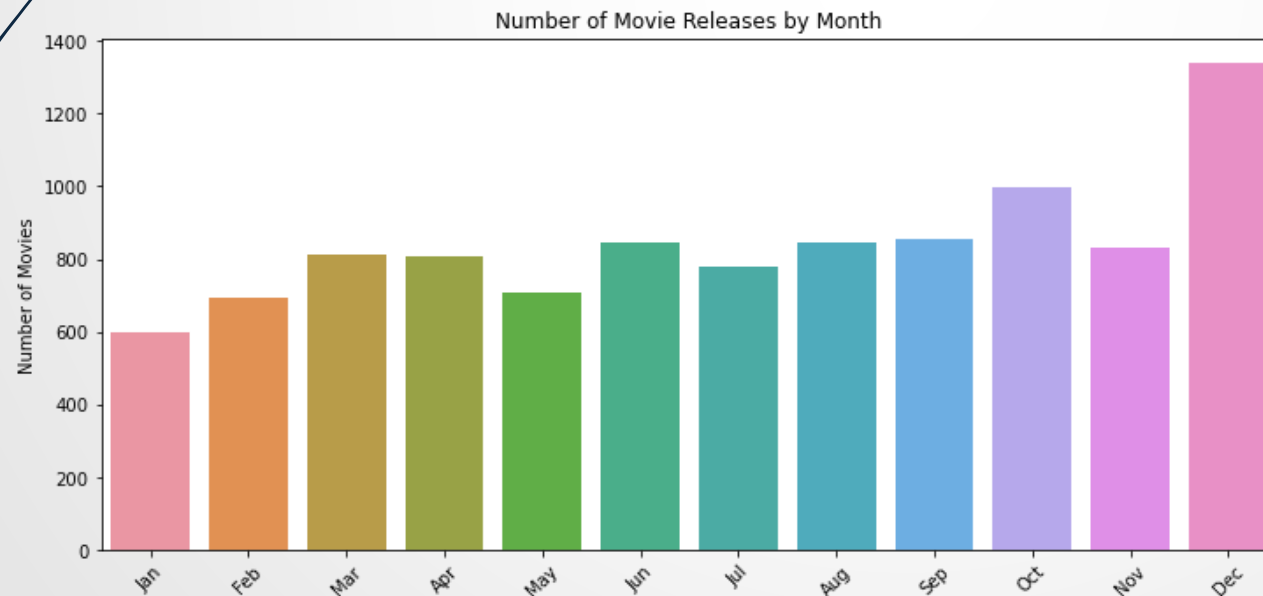
According to the data reviewed, majority of the films fall into one of two categories: restricted (R) or nonrestricted (NR). The general guidance (G) is the lowest and the bulk of the films made are intended for adults.



Production Budget and Revenue per Release Month

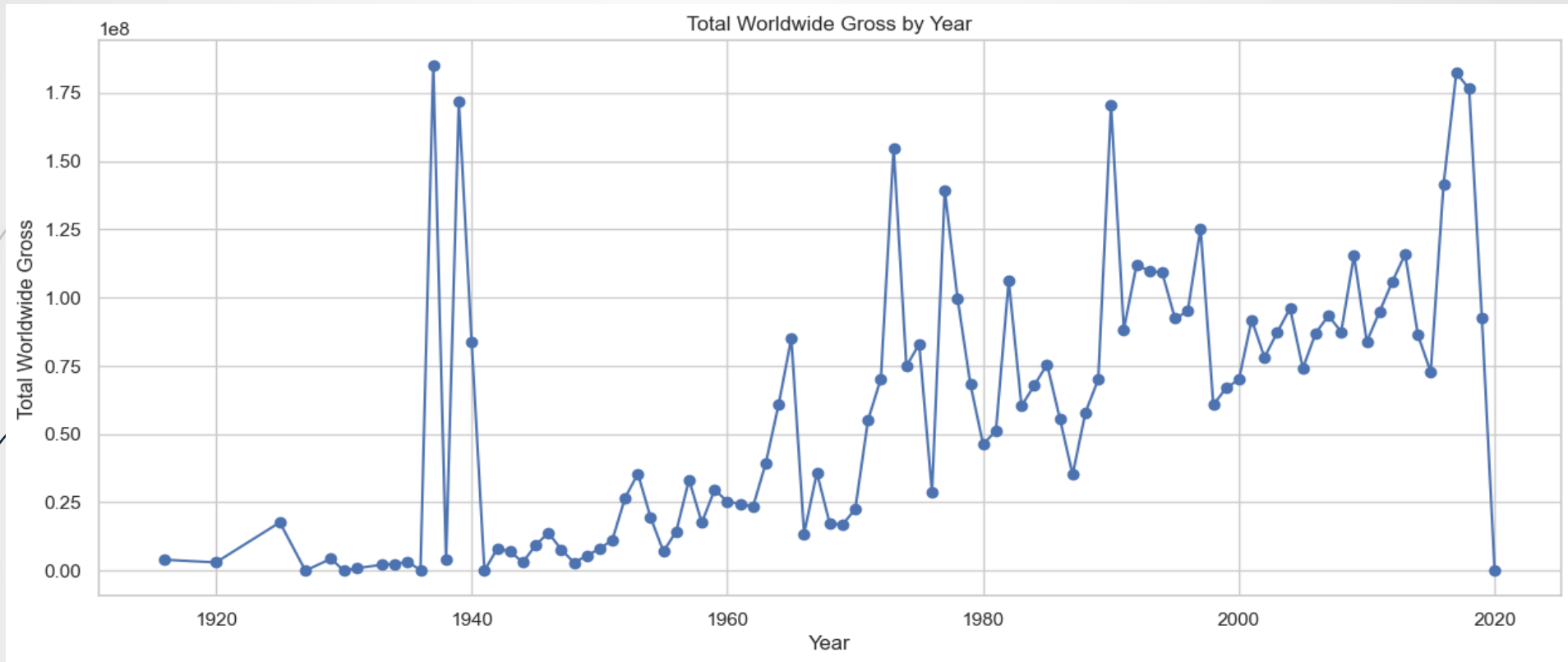


- June, July and December consistently show the highest worldwide gross and domestic gross.
- The more movies produced the more the domestic and worldwide gross (same cycle)



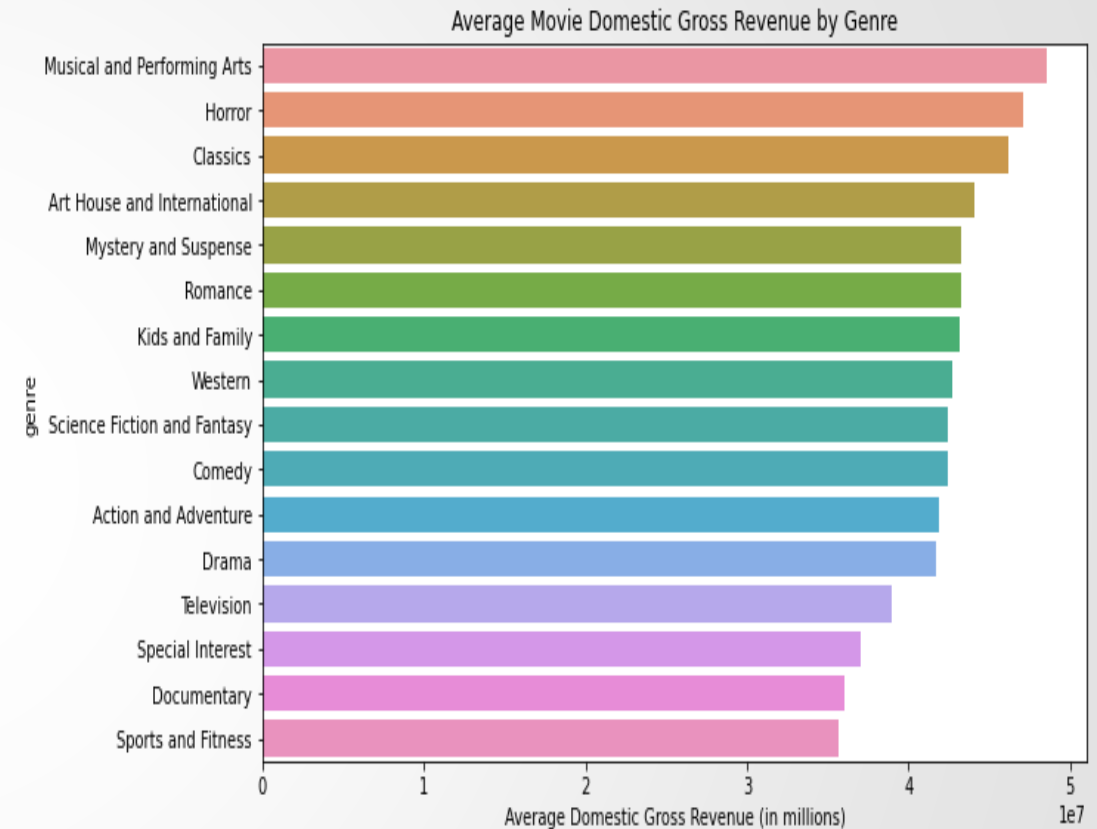
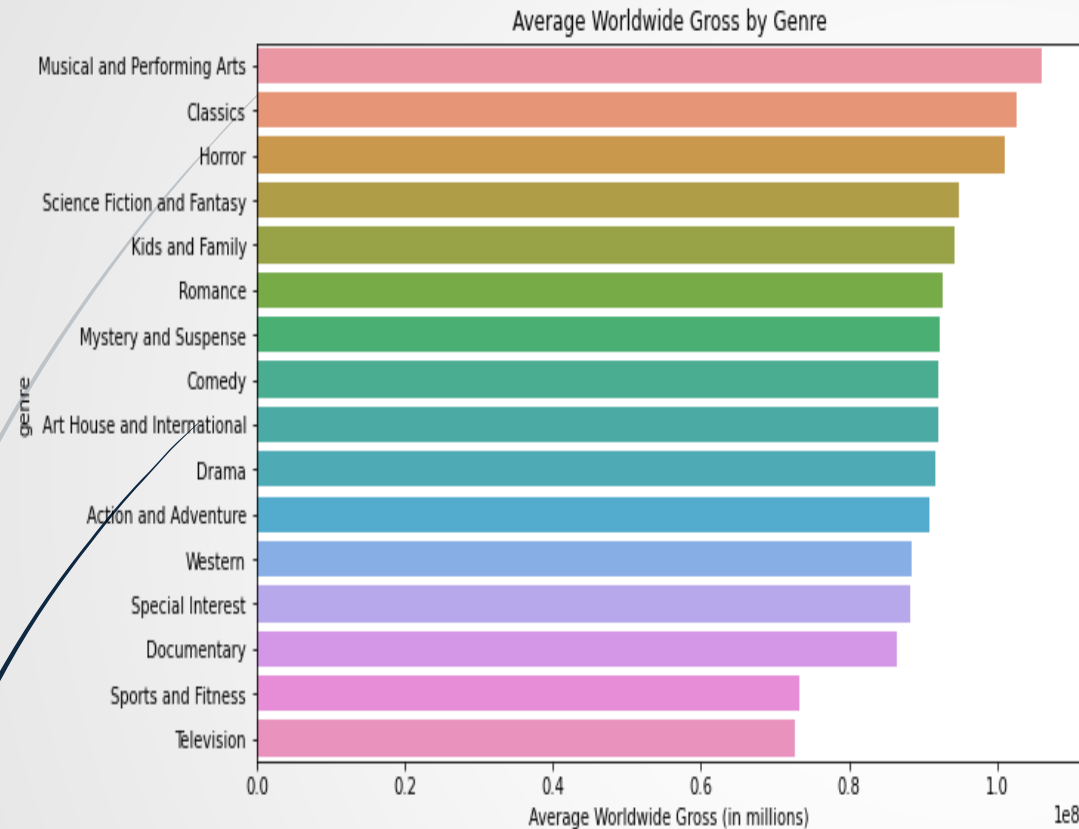
- The data shows more movies are released during the month of December and the revenue shoots up.

Worldwide gross performance during the review period



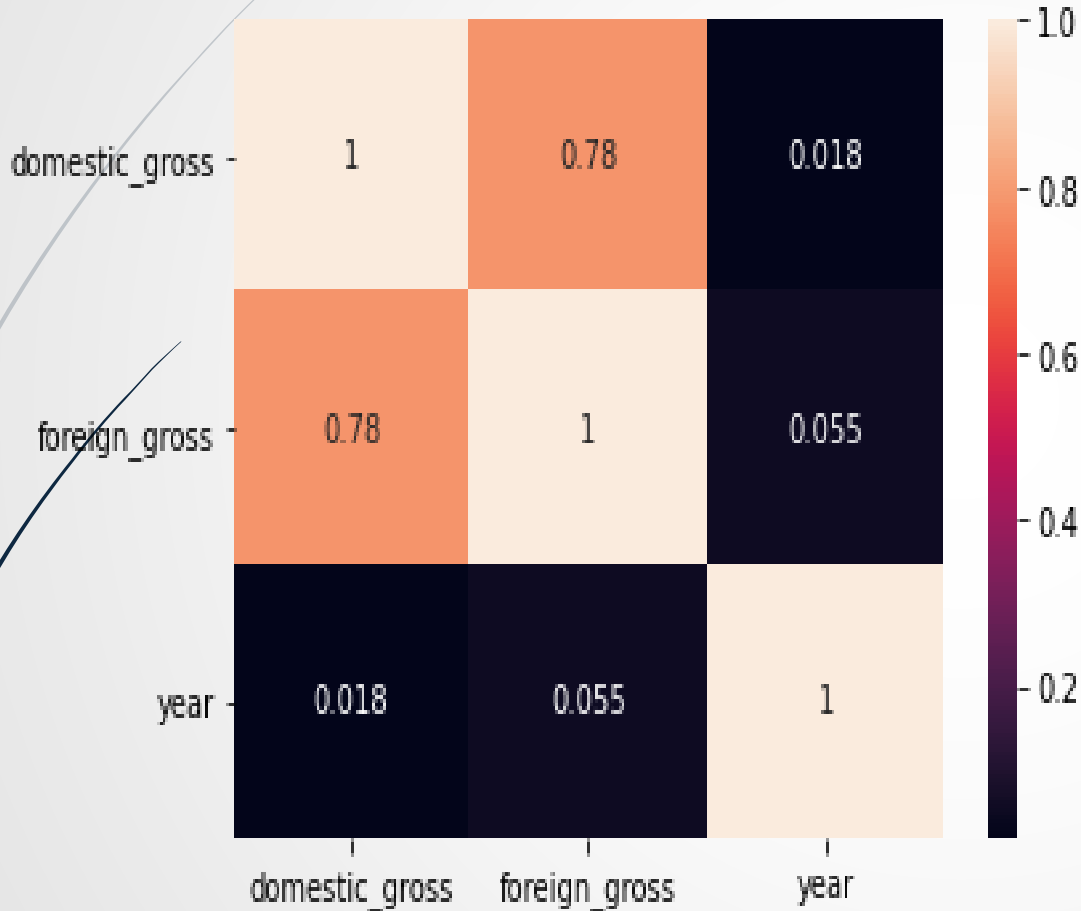
- The worldwide gross increases over the years.
- A plunge during the year of 2020: An event might have happened which needs further research

Genre performance by revenue



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 – Revenue



➤ Strong Relationship Between Domestic and Foreign Earnings

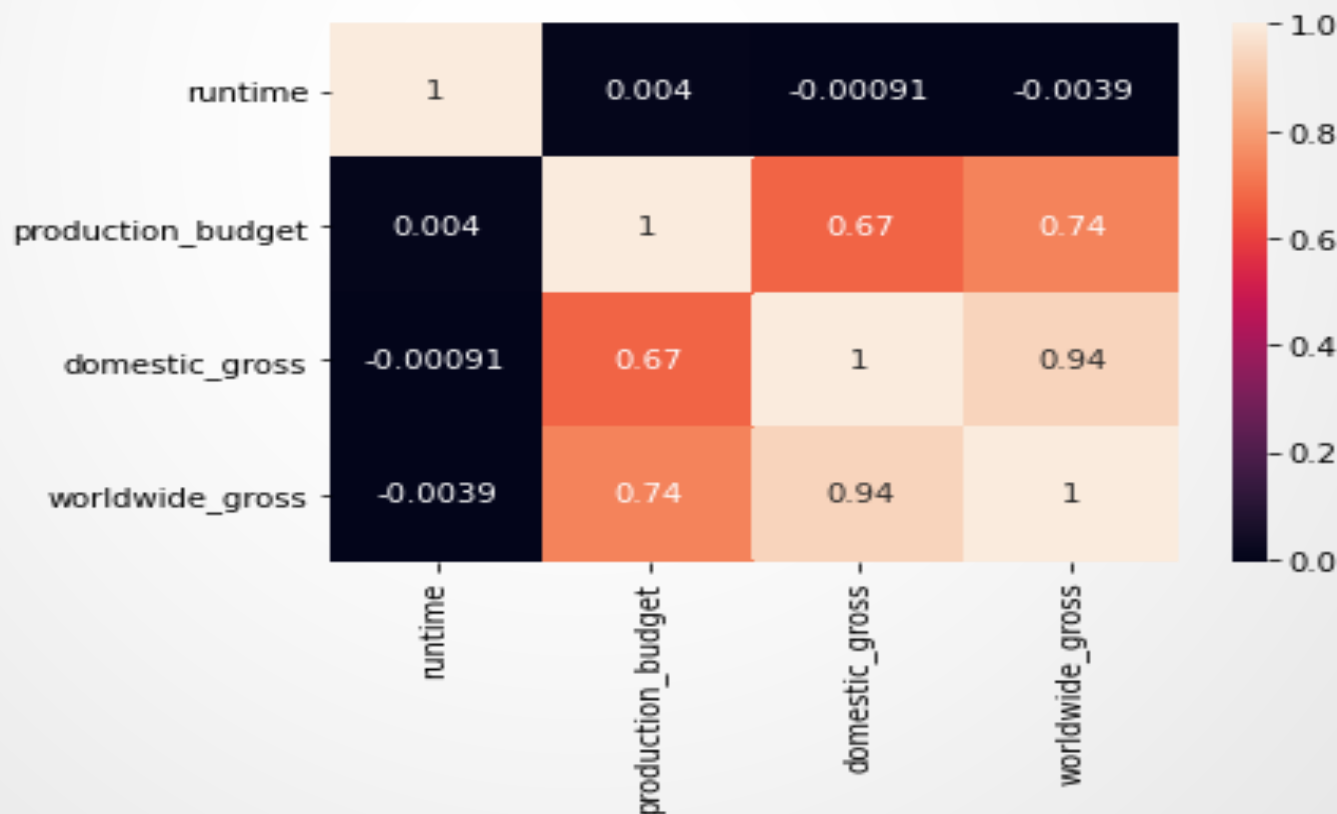
- Domestically successful movies generally may perform well internationally, but the year of release doesn't significantly impact their earnings.

➤ Conclusion

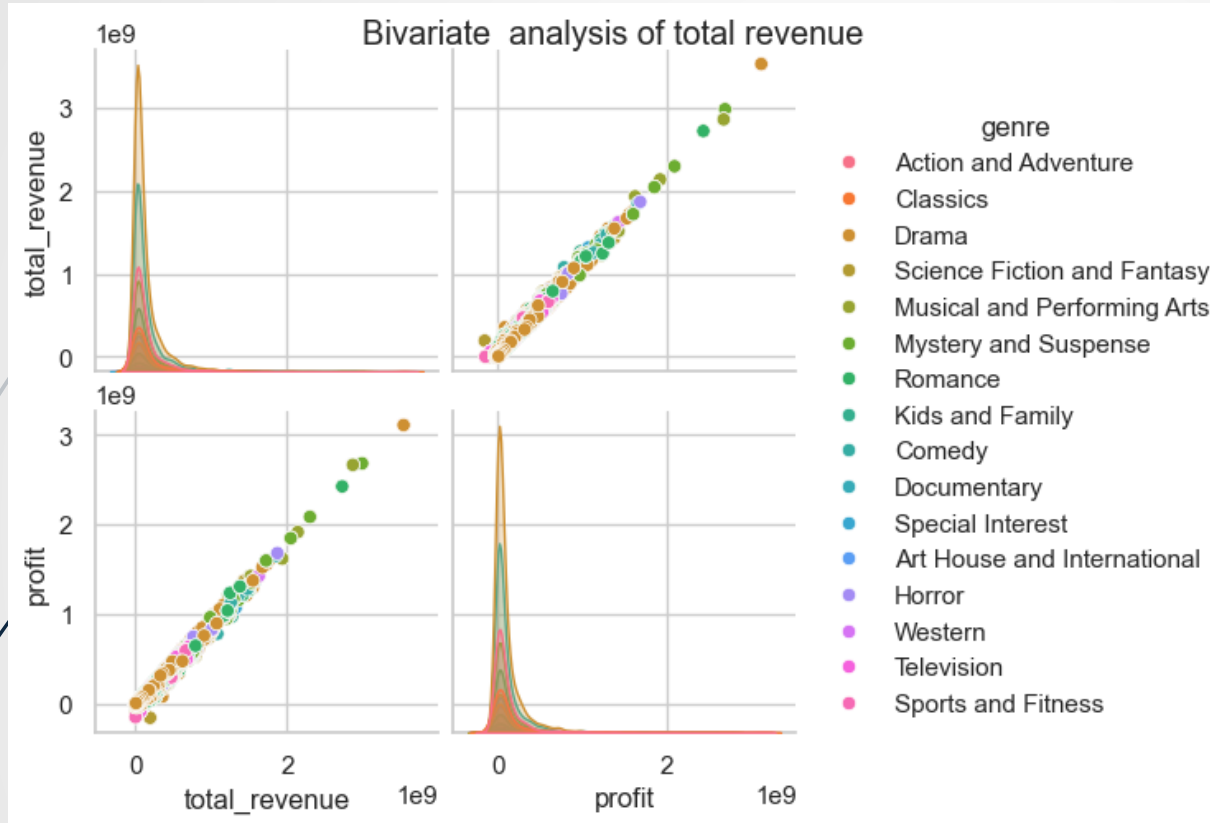
- The strong correlation between domestic and international movie earnings is crucial for predicting its global success

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



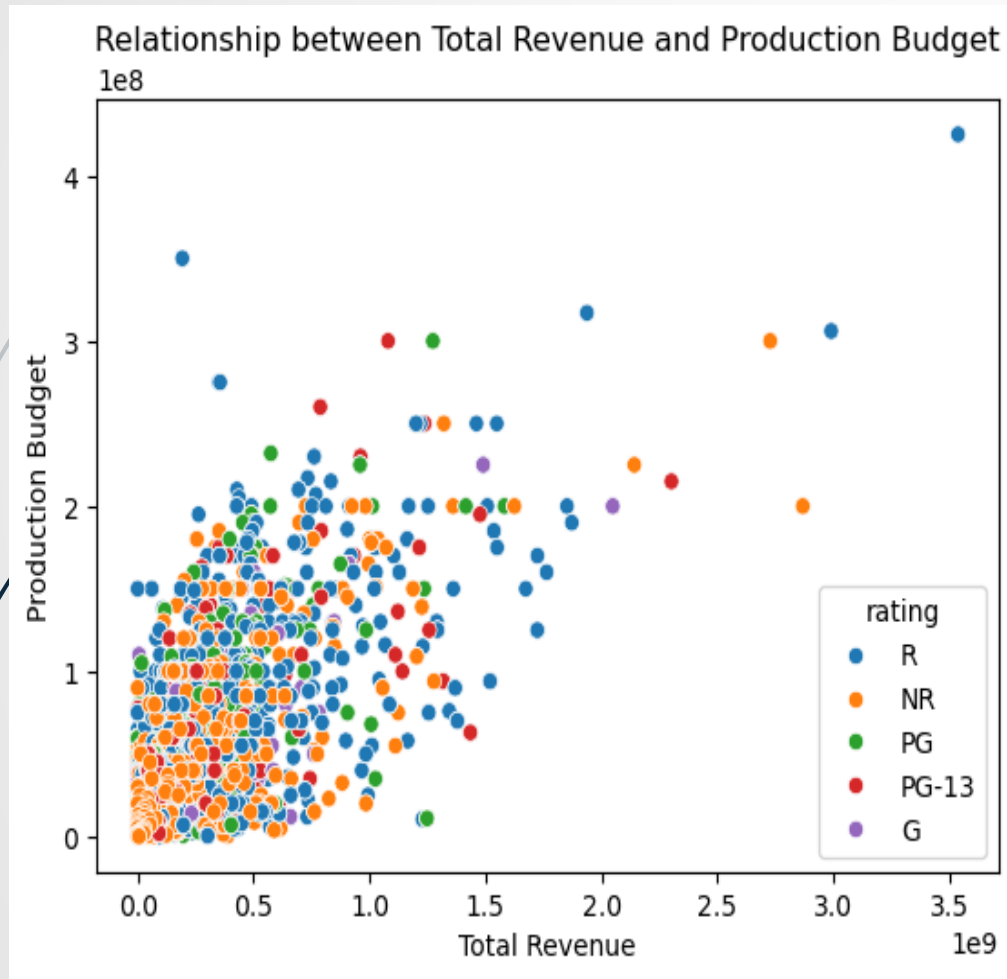
Revenue trends per genre



- **Revenue and Profit Relationship:**
As revenue increases, profit tends to rise proportionally

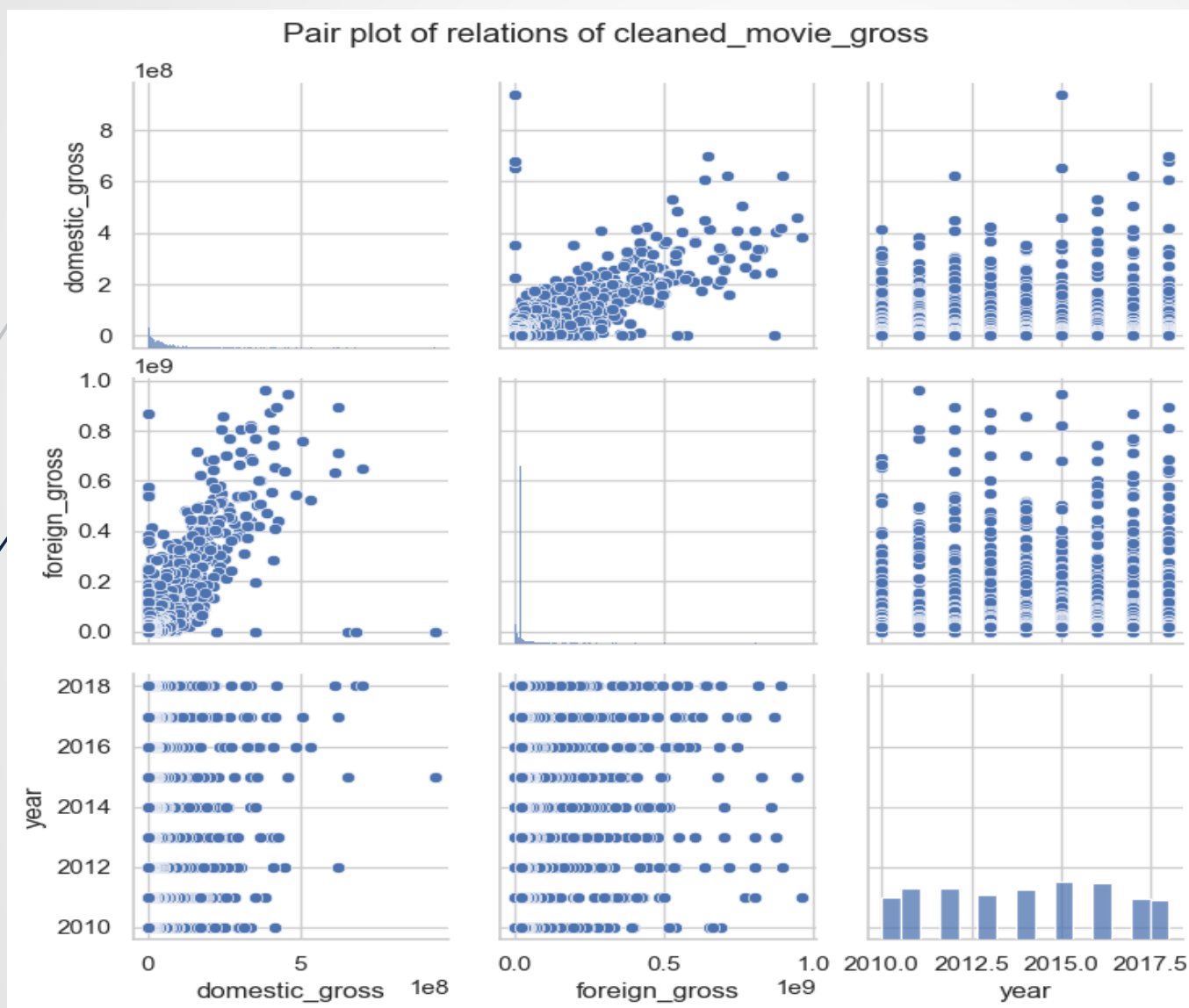
Genre-Based Trends: Certain genres, like Drama, Science Fiction, and Fantasy, often reach higher revenue and profit levels. In contrast, genres like Documentary and Special Interest usually perform on the lower end financially. Action, Adventure, Science Fiction, and Fantasy show high potential for greater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

Production Budget vs Total Revenue




- 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

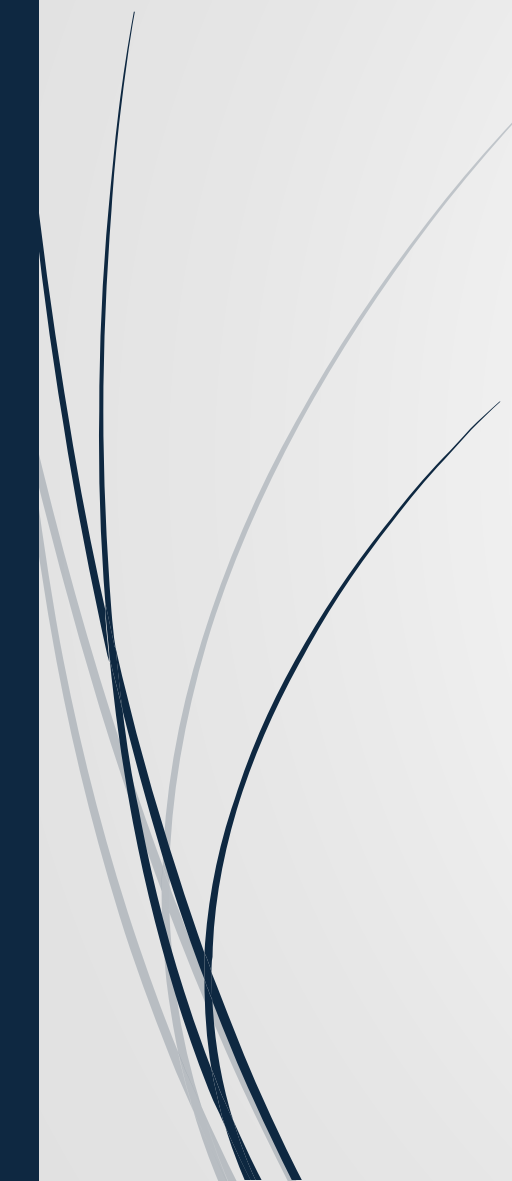
Relationships for the budget data



- ➡ There is a positive linear relationship between domestic gross and foreign gross.
- ➡ An increase in domestic gross may lead to a corresponding increase in 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.

```
# predict worldwide_gross at domestic_gross of $3,516,297.93
domestic_gross=3516297.93
worldwide_gross_est=2.3684*domestic_gross+-8.328e+06
worldwide_gross_est
```

Using Multilinear 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 revenue **domestically** and **worldwide**. 44.9% of the budget differences are influenced by other factors that weren't included in this analysis.
- Revenue provides a preview 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 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.
- Months with Lower Revenue: utilize lower budget or experimental films in months like January and September.

Recommendations cont.

- Increase marketing efforts leading up to and during January and September to maximize on overall revenue.
- Release High-Potential Films in June, July, and December as it shows strong performance in revenue.
- Diversify the film portfolio by including more investments in lower-budget projects to enhance overall profitability and mitigate financial risk.

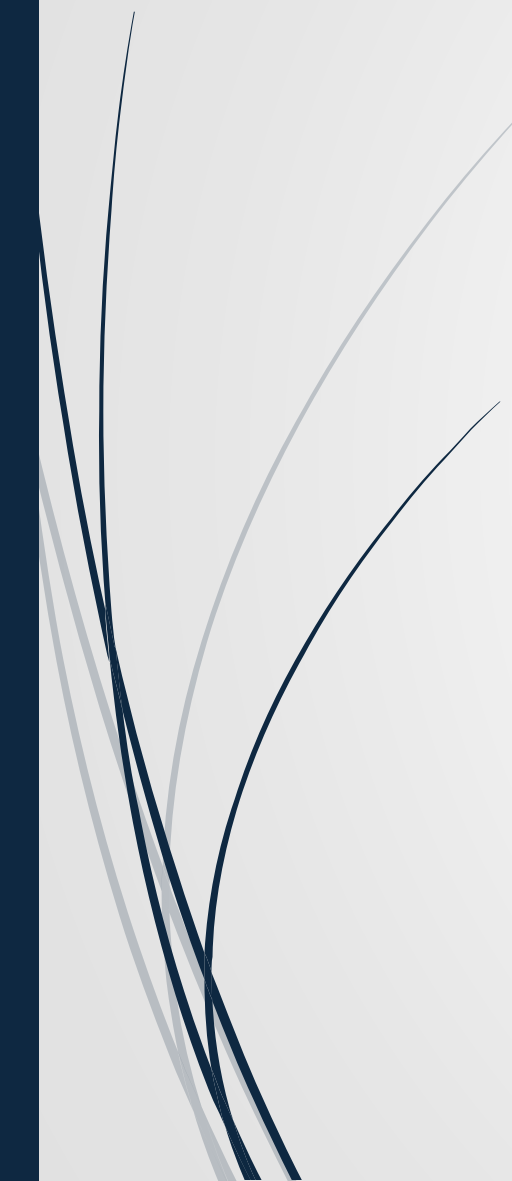


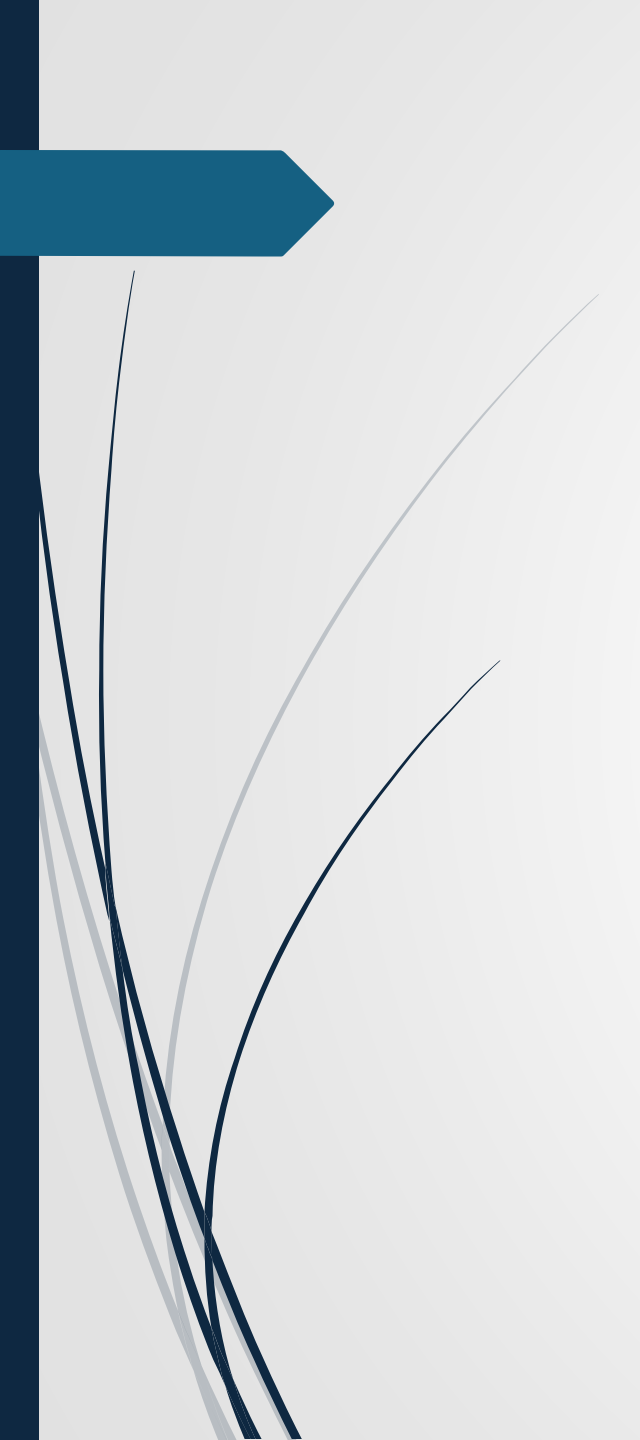
Recommendations cont.

- 
- Develop tailored marketing campaigns for lower-budget films, utilizing cost-effective promotional strategies to enhance revenue potential
 - Consider producing shorter films for wider audience appeal while also investing in select, high-quality longer films that can captivate audiences and drive engagement.
 - Allocation of more resources to produce a mix of shorter, high-rating-friendly films and select longer, high-quality projects that will meet the expectations of more targeted audiences



Next Steps

- Further research to find other factors that will influence the production costs. This will help in prediction on production cost.
 - Research to be done on an event may have happened in the year 2020 as there was a plunge on gross revenue
 - Given the higher production budget indicated more gross revenue, the company needs to diversify its funding to sustain the production budget
- 



Q&A



Contact Info



- <https://www.linkedin.com/in/gilbert-kipkirui-cheruiyot-82387425>
- <https://www.linkedin.com/in/daniel-muigai-358707292/>
- <https://www.linkedin.com/in/millicent-cheptoi-068235221/>
- <https://www.linkedin.com/in/angela-mwanzia-2b993b309>
- <https://www.linkedin.com/in/edwin-george-181b33171/>
- <http://www.linkedin.com/in/reagan-adajo-5a1803103>
- <https://www.linkedin.com/in/james-wachira-a7771644/>
- <https://www.linkedin.com/in/mercy-ayub-774a877b/>



Thank You 😊