Phase 1 Project:

MICROSOFT MOVIE STUDIO: DATA-DRIVEN DECISIONS FOR SUCCESS

AUTHORS: FREDRICK MUNYAO KYEKI

Summary

The project aims to provide data-driven insights to support **Microsoft's** new movie studio in establishing a successful presence in the movie industry. The business problem revolves around Microsoft's lack of expertise in film-making and the need to make informed decisions for profitable movie production. I analyzed diverse movie datasets from IMDb, Box Office Mojo, and The Numbers to answer key data questions related to movie success.

The results suggest focusing on producing **Action, Adventure, Animation, Comedy, Documentary, Drama, Fantasy, Sci-fi, Thriller** movies with moderate budgets between $**170,001,400 - $255,001,400** , releasing them during peak movie-going seasons **August, December, July, June, May, November**, and creating audience-centric content that aligns with user preferences.

Outline

- **Business problem**
- **Data**
- **Methods**
- **Results**
- **Conclusions**

Business problem

Microsoft is entering the movie industry and wants to establish a successful movie studio. However, they lack expertise in the movie-making domain and need to make informed decisions to create profitable films.

The business's pain points include:

- uncertainty about the types of movies that perform well at the box office,
- the preferences of the target audience,
- and how to effectively compete in a crowded market.

Data

Data Description: The data used for this project comes from multiple movie-related datasets from various sources, including:

1. **IMDb (Internet Movie Database)**:
   - **IMDB Basics**:
   - **IMDb Ratings**:
2. **Box Office Mojo**:
3. **The Numbers**:

Data

Data Representation:

The datasets represent a diverse sample of movies, including various genres, budgets, and performances. Each movie record contains attributes such as movie title, genre(s), budget, box office gross, user ratings, and release date.

- **Target Variable**: The target variable for this project is the **"worldwide gross"** of movies. Box office gross represents the total revenue generated by the movie in theaters and serves as an indicator of movie financial success.

Properties of Variables of interest:

- **Movie Name** - Categorical variable which is a textual label or name of the movie.
- **Genre**: Categorical variable representing the type or category of the movie (e.g., Action, Drama, Comedy).
- **Budget**: Continuous variable representing the production cost or budget of the movie.
- **worldwide Gross**: Continuous variable representing the total revenue generated by the movie at the box office.
- **User Ratings**: Continuous variable representing the average ratings or scores given by users for the movie.
- **Release Date**: Temporal variable indicating the date when the movie was released in theaters.

Data

Data questions:

The following are the **data questions** answered in this analysis:

1. What types of films are currently performing best at the box office (based on box office gross)?
- What are the characteristics of top-performing movies based on box office gross?
- Which movies have been the most successful financially?

2. Which movie genres have been the most popular and successful over time?
- What are the trends in genre preferences?
- What are the preferences of the target audience based on user ratings and reviews?
- What type of content resonates well with the audience?

3. How does the movie budget impact box office revenue, and can smaller budget films be profitable?
- How does the movie budget affect box office revenue?
- Can smaller budget films be profitable, and is there an optimal budget range?

4. Are there seasonal trends in movie performance, and when is the best time to release a movie?
- Are there seasonal patterns in movie performance?
- When is the best time to release a movie for maximum revenue?

Methods

1. Data Preparation:

- **Data Loading**:
  - Load the required datasets into the analysis environment.
- **Data Cleaning**:
  - Handle Missing Values:
  - Drop Irrelevant Variables:
  - Merge Data:
- **Feature Engineering:**
  - Calculate Profit: subtracting the budget from the box office gross
  - Calculate foreign gross: from the difference between worldwide gross and domestic gross
- **Handling Outliers:**
  - Analyze and Address Outliers: Identify outliers in numeric variables like budget and box office gross. Outliers may affect our analysis, and we need to decide whether to remove or transform them based on their impact on the results.

Results

1. What types of films are currently performing best at the box office (based on box office gross)?

Top 5 best performing genres in terms of gross are:

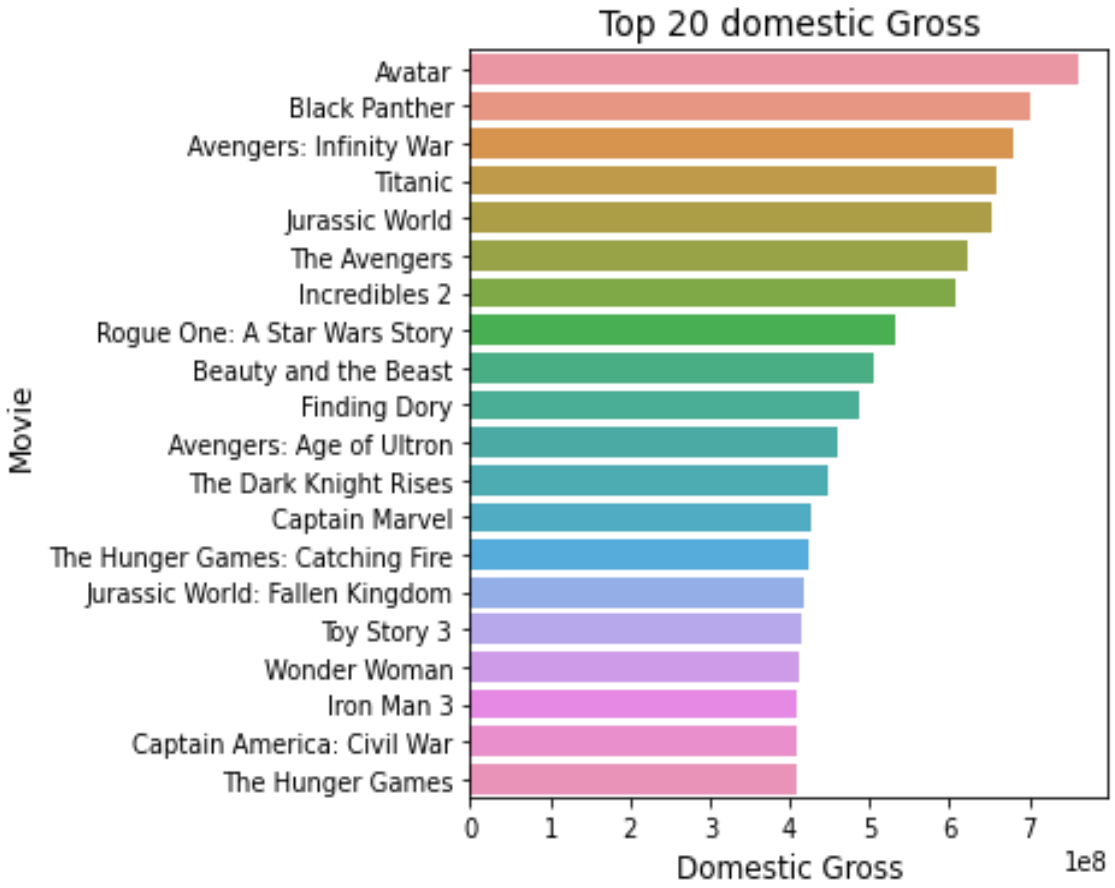
Worldwide gross

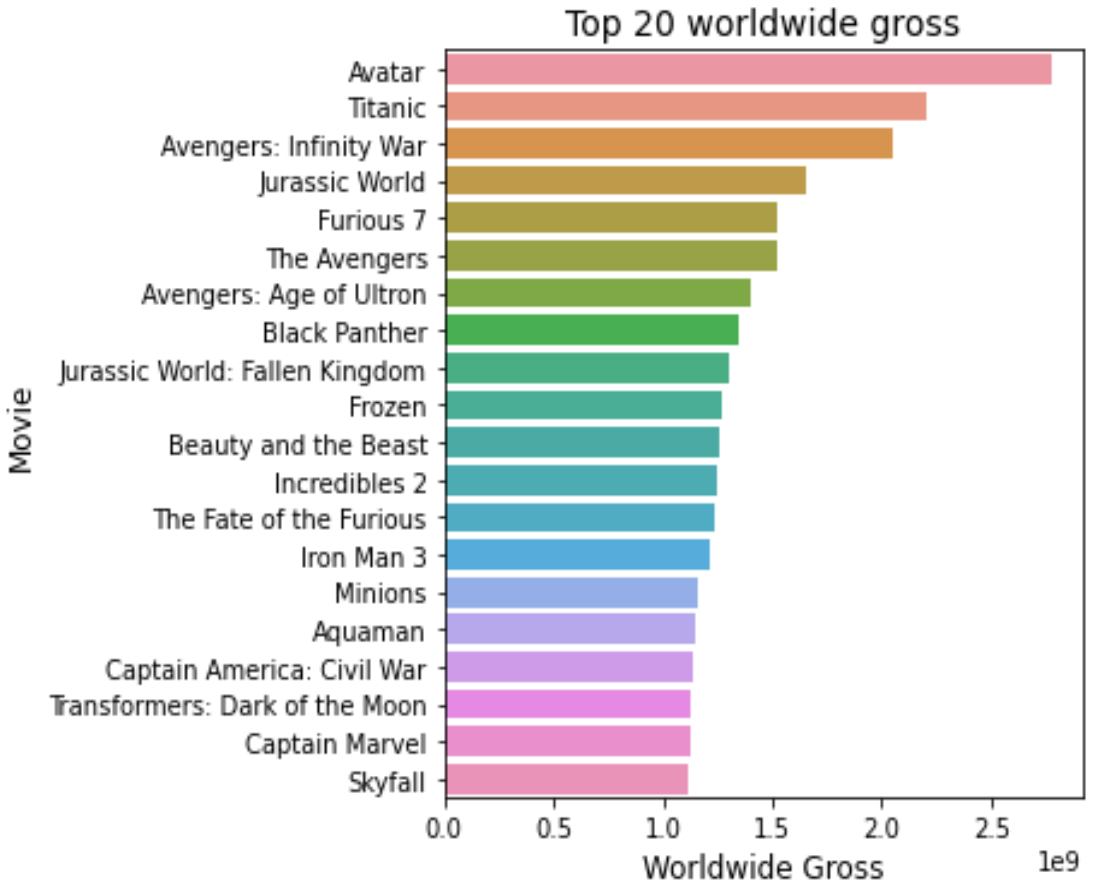
- Adventure
- Action
- Scifi
- Thriller
- Drama

Domestic gross

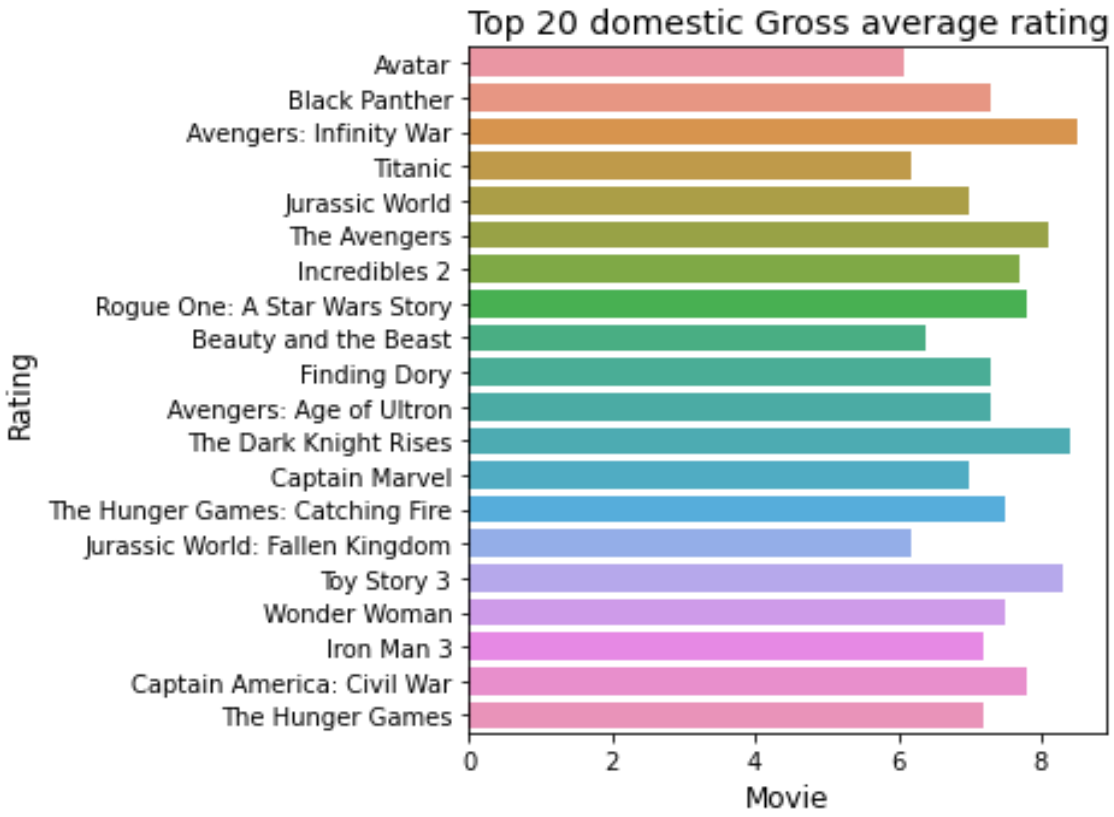
- Adventure
- Action
- Scifi
- Animation
- Fantasy

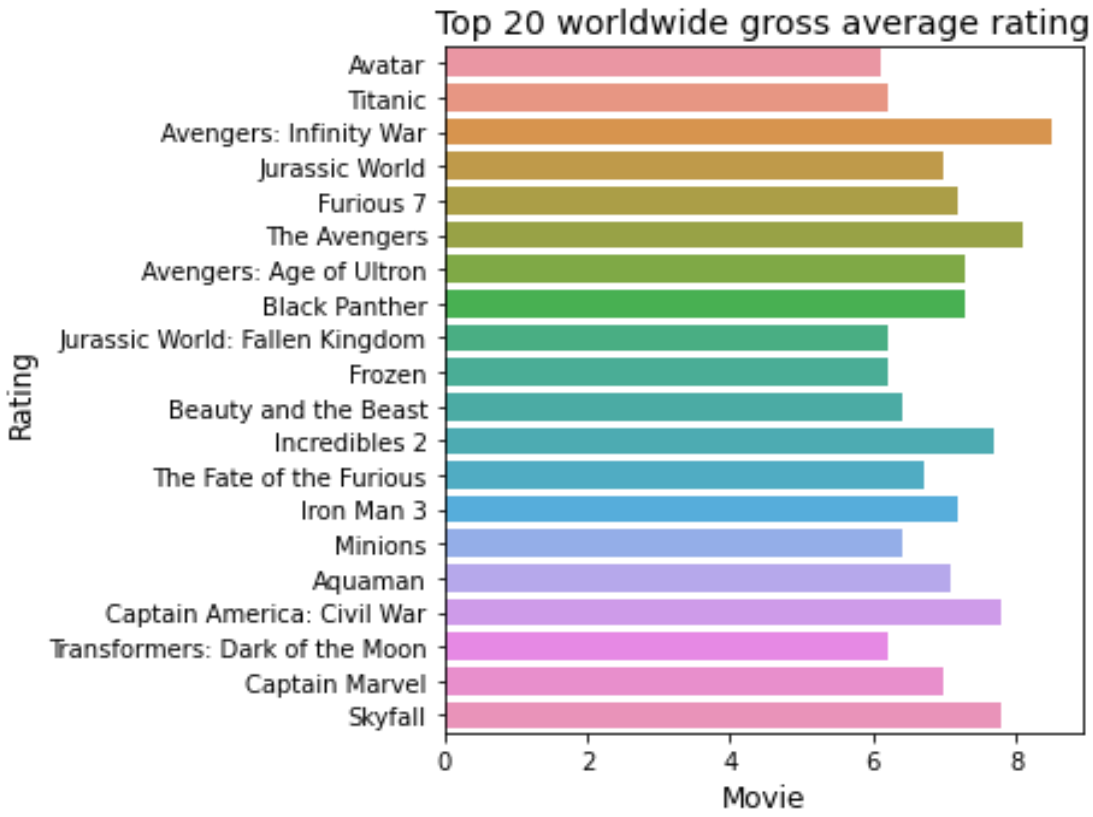
Top 20 movies by domestic gross and worldwide gross



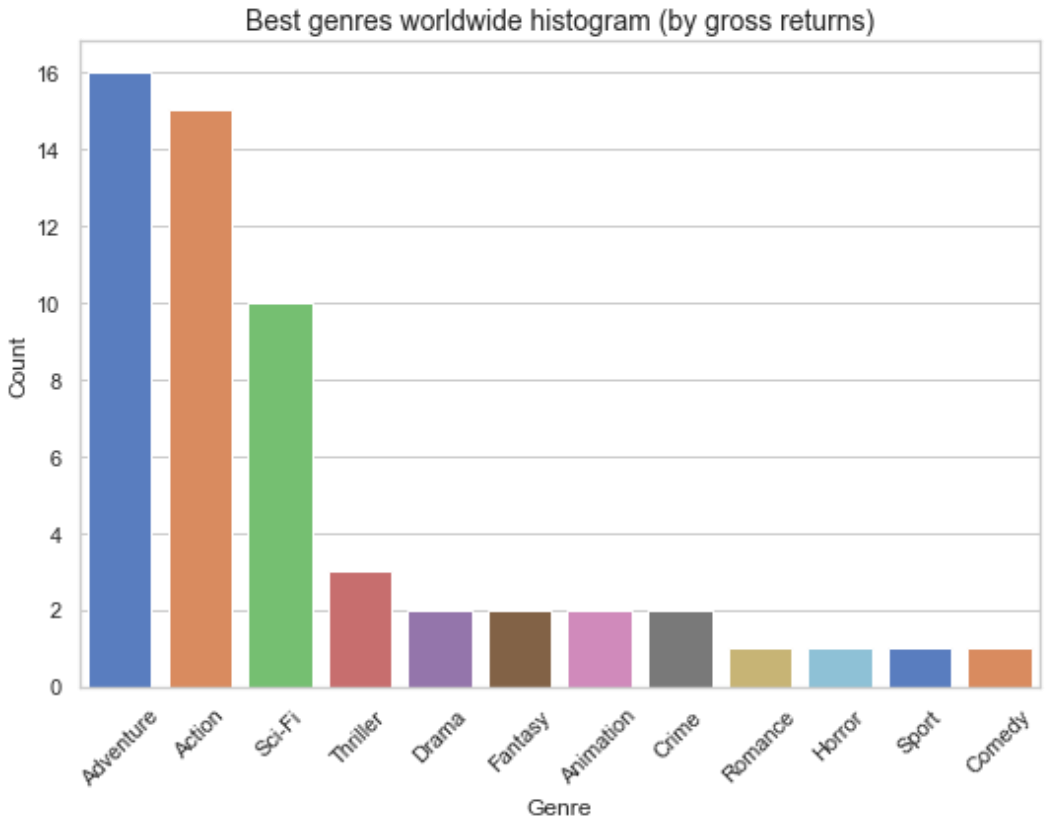


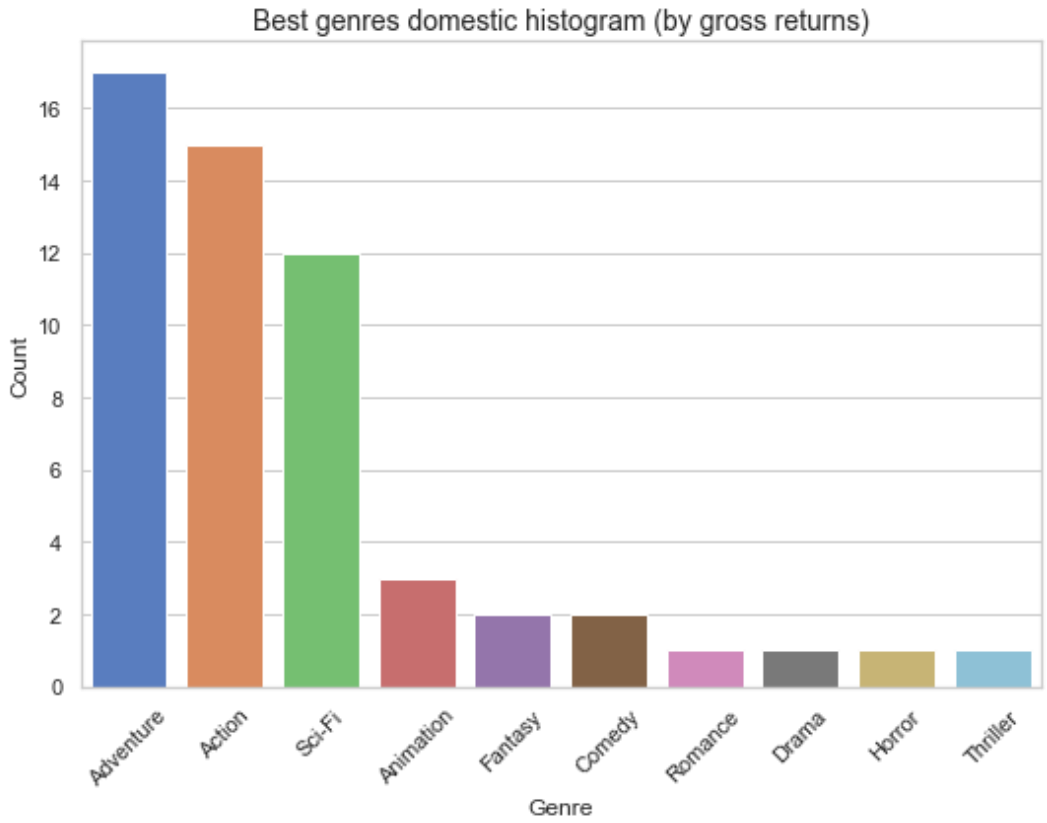
Top 20 movies by domestic gross and worldwide gross (Average Ratings)





Genre Distribution in the top 20 best grossing movies



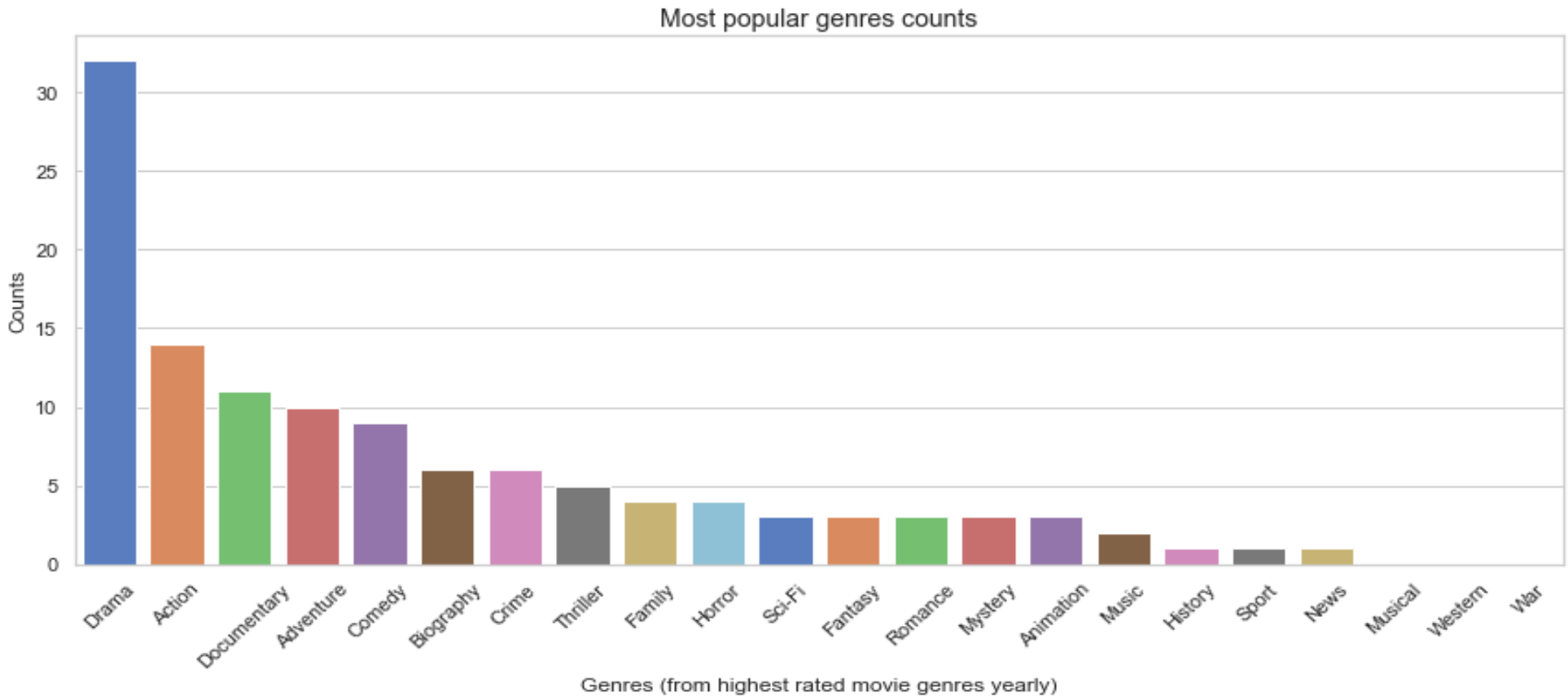


Results

2. Which movie genres have been the most popular and successful over time?

- The top 5 most popular (highly rated yearly / most loved) genres are
  1. Drama
  2. Action
  3. Documentary
  4. Adventure
  5. Comedy

Most popular genres (by highest yearly rating)

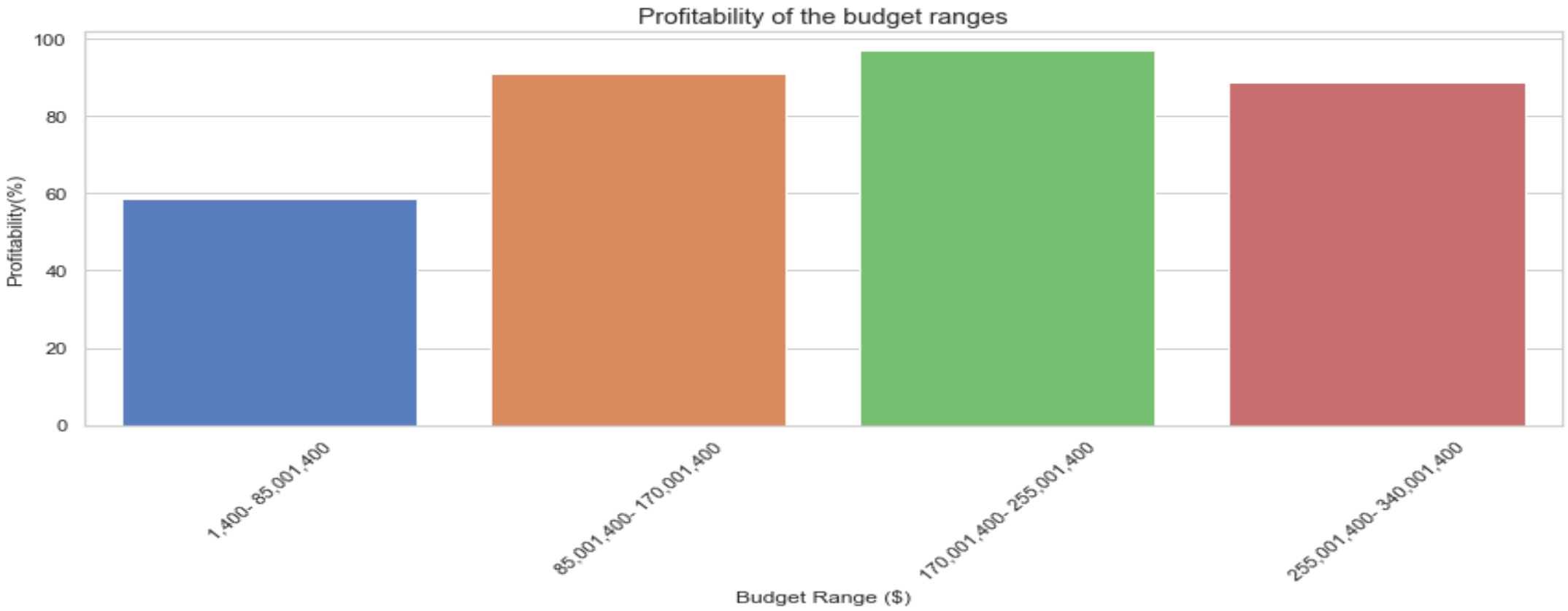


Results

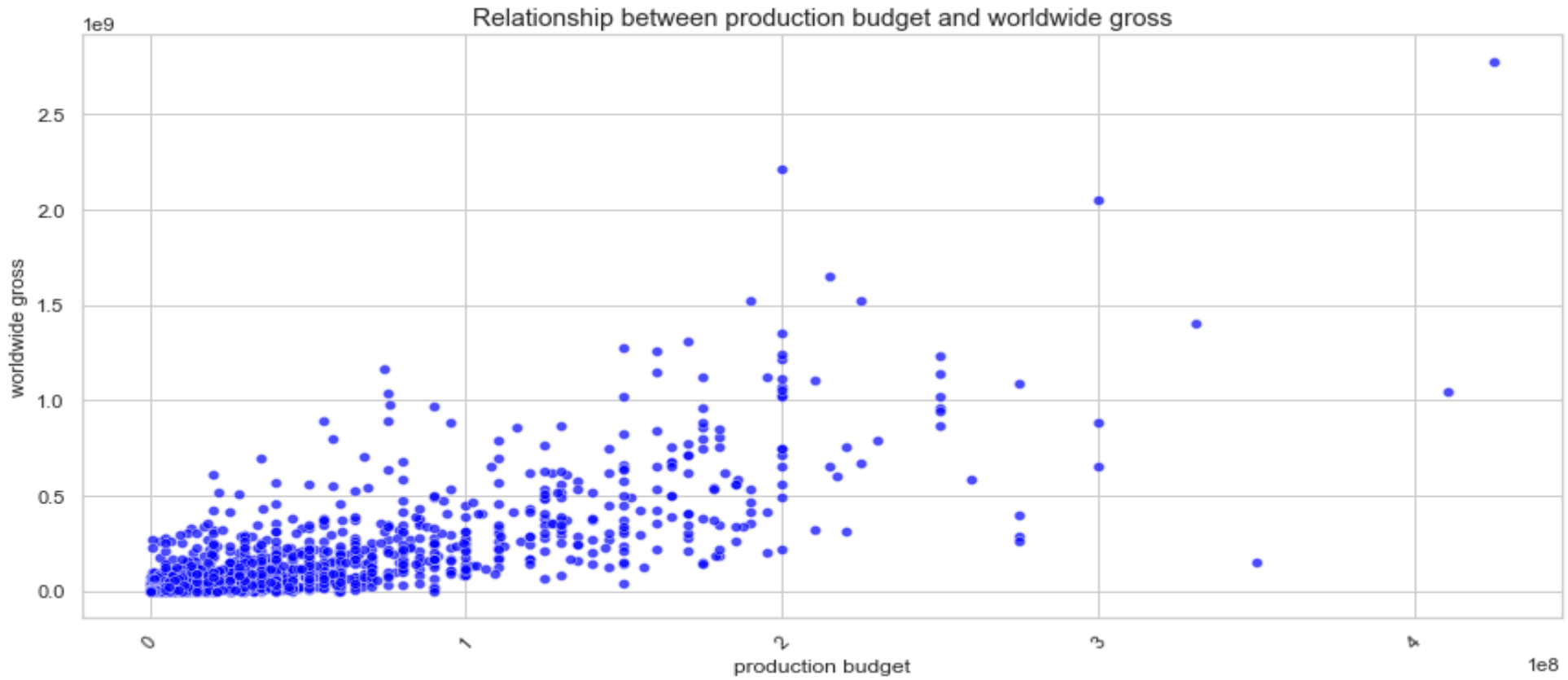
3. How does the movie budget impact box office revenue, and can smaller budget films be profitable?

- A correlation coefficient of approximately **0.78** between `production_budget` and `worldwide_gross` indicates a **strong positive linear correlation**.
- This means that there is a significant and positive relationship between the movie's production budget and its worldwide box office revenue.
- The optimal budget range is: **$170,001,400 - $255,001,400**
- The **correlation coefficient** of approximately **0.66** between `profitability` and `production budget` indicates a **moderately positive** linear correlation.
- This means that there is a positive relationship between a movie's production budget and its profitability.

Profitability index of various budget ranges

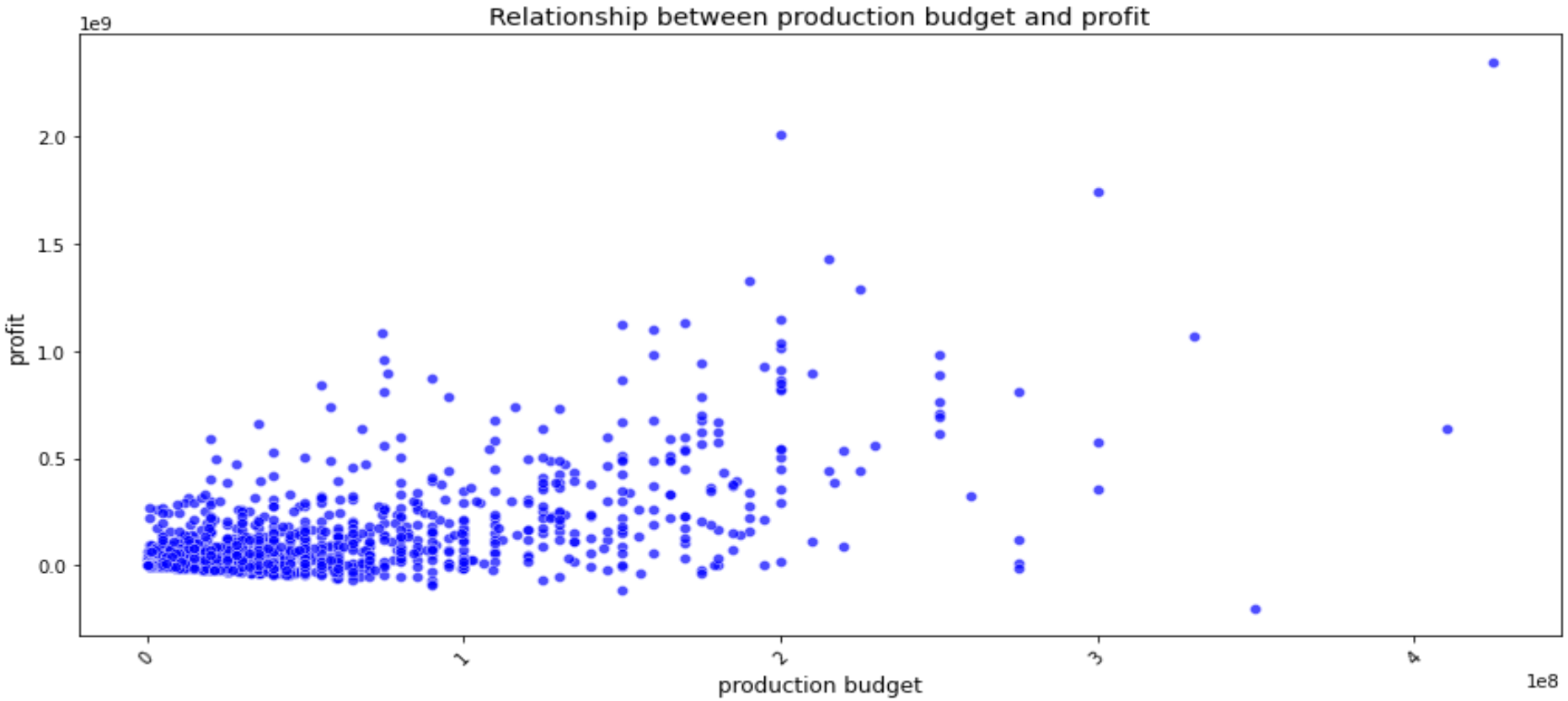


Production budget vs Worldwide Gross relationship



Correlation Coefficient: 0.78

Production budget vs Profit relationship



Correlation Coefficient: 0.66

Results

4. Are there seasonal trends in movie performance, and when is the best time to release a movie?

Months with Higher Average Worldwide Gross:

- **June (Jun)** has the highest average worldwide gross among all months, followed closely by **May (May)** and **December (Dec)**.
- These months could be considered the best time to release a movie for maximum revenue, as they tend to generate higher box office gross on average.

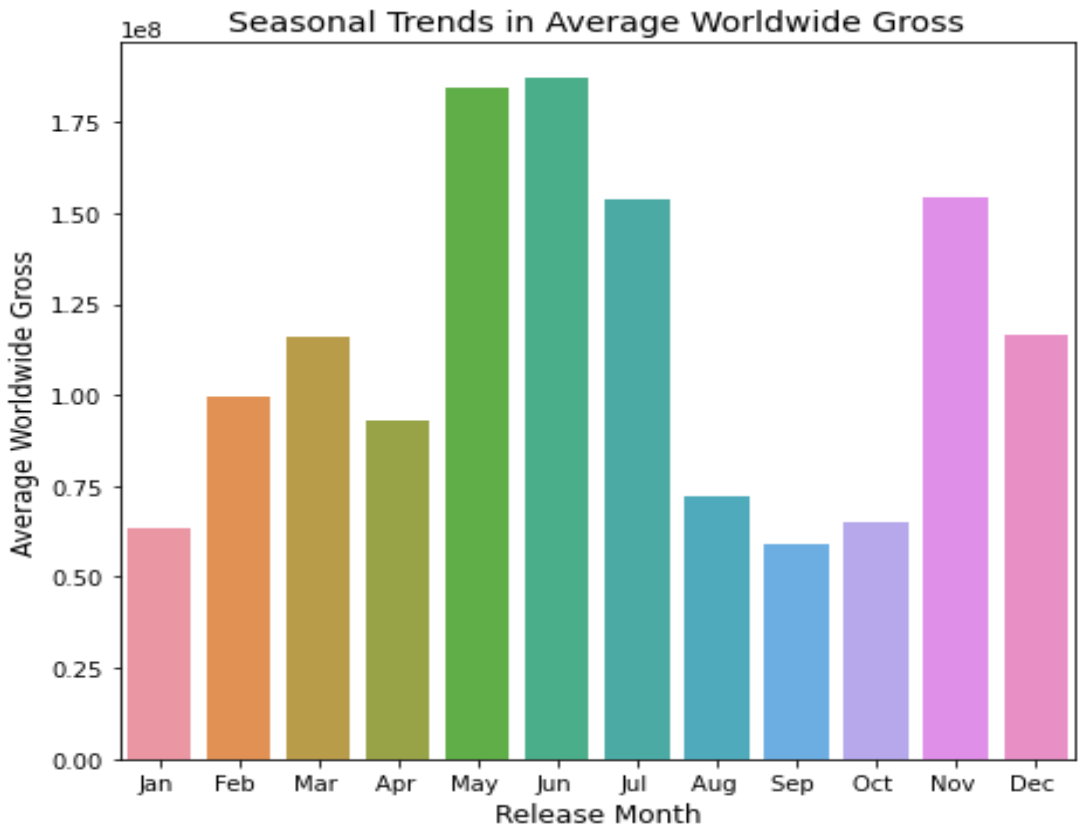
Months with Higher Average Rating:

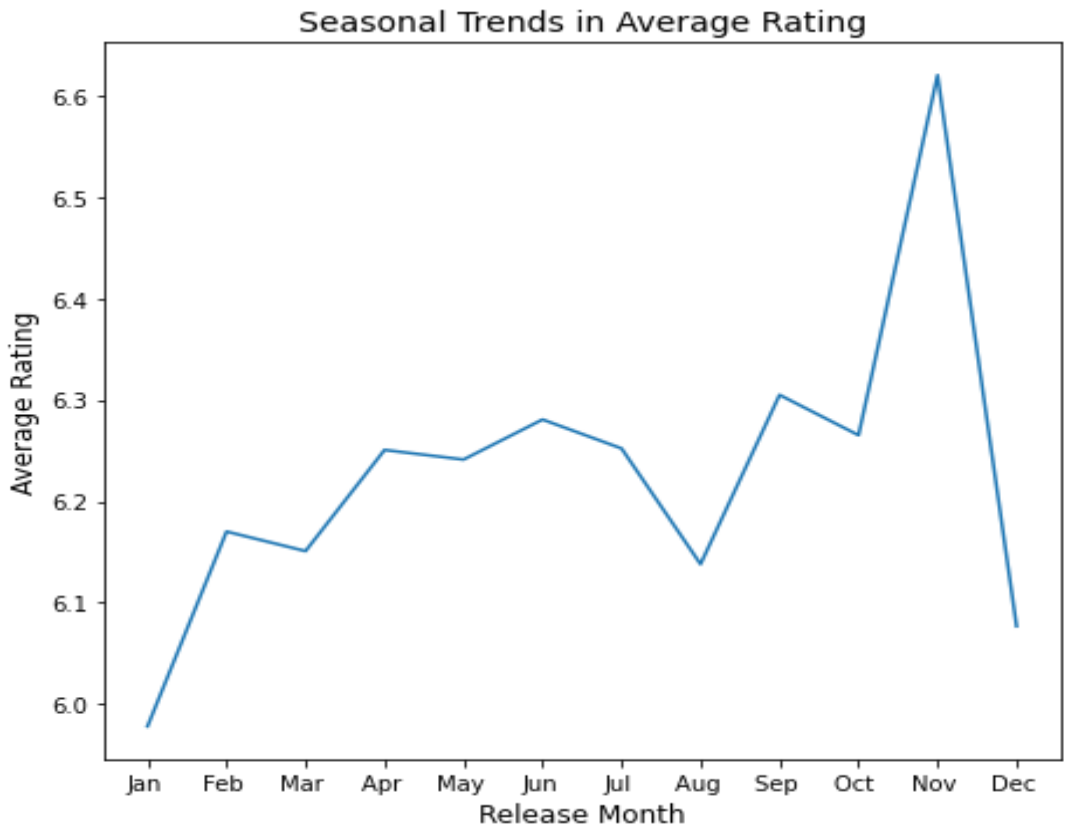
- **November (Nov)** has the highest average rating among all months, followed by **June (Jun)** and **May (May)**.
- **November** seems to be a favorable month for releasing movies that receive higher audience ratings on average.

Seasonal Patterns:

- Summer months ***(June, July, and August)*** show *strong performance* in terms of both average worldwide gross and average rating.
- This could be due to the summer vacation period and increased movie going during this time.
- **December**, being the holiday season, also exhibits high average worldwide gross and average rating, making it a potential lucrative period for movie releases.

Seasonal (monthly) trends in gross and rating





Variable correlations

Based on the correlation coefficients you provided, we can interpret the following relationships between the variables:

Budget and **Domestic Gross**: The correlation coefficient is **0.25**. There is a **weak positive correlation** between the movie's budget and its domestic gross. This suggests that, on average, movies with higher budgets tend to have slightly higher domestic grosses, but the relationship is not very strong.

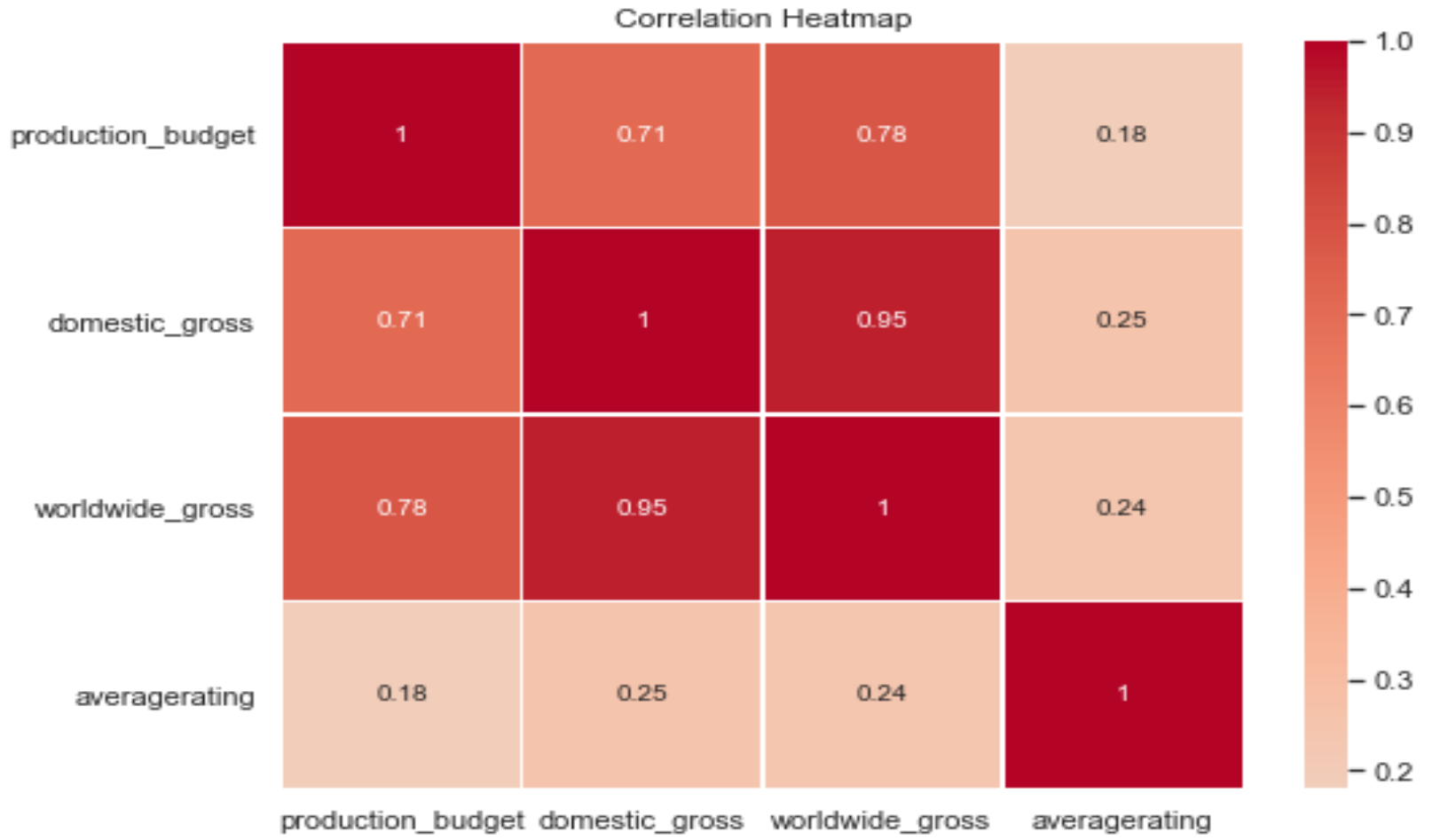
Budget and **Worldwide Gross**: The correlation coefficient is **0.78**. There is a **strong positive correlation** between the movie's budget and its worldwide gross. This indicates that movies with higher budgets are more likely to have significantly higher worldwide grosses.

Domestic Gross and **Worldwide Gross**: The correlation coefficient is **0.95**. There is a **very strong positive correlation** between a movie's domestic gross and its worldwide gross. This means that movies that perform well in the domestic market are highly likely to perform similarly well in the worldwide market.

Domestic Gross and **Average Rating**: The correlation coefficient is **0.25**. There is a **weak positive correlation** between a movie's domestic gross and its average rating. This suggests that movies with higher domestic grosses may slightly tend to have slightly higher average ratings, but the relationship is not very strong.

Worldwide Gross and **Average Rating**: The correlation coefficient is **0.24**. There is a **weak positive correlation** between a movie's worldwide gross and its average rating. This indicates that movies with higher worldwide grosses may slightly tend to have slightly higher average ratings, but the relationship is not very stro

Variable correlation heat-map



Conclusion

1. **Recommendations for the Business:**

Based on the analysis conducted, the following recommendations are made for Microsoft's new movie studio:
a.) **Focus on Popular Genres**: Allocate resources to produce movies in genres that are currently performing well based on box office gross and audience ratings. This can increase the likelihood of commercial success and positive audience reception.
The genres that are likely to perform well are:

- ❑ Action
- ❑ Adventure
- ❑ Animation
- ❑ Comedy
- ❑ Documentary
- ❑ Drama
- ❑ Fantasy
- ❑ Sci-fi
- ❑ Thriller

Conclusion

1. Recommendations for the Business (continued…):

b.) **Optimize Release Timing**: Use the insights from seasonal trends to strategically plan movie releases. Target months with historically higher average worldwide gross and audience ratings to maximize revenue and audience reach.
The best months of release are:

- August
- December
- July
- June
- May
- November

c.) **Budget Allocation**: Analyze the relationship between movie budgets and box office revenue to make informed decisions about budget allocation for different movie projects. This can help optimize resources and manage financial risks.

- The safest budget range is: **$170,001,400 - $255,001,400**, with the highest profitability index (**96.96%**).

Conclusion

1. Recommendations for the Business (continued…):

<u>Other Recommendations:</u>

d.) **Emphasize Audience Preferences**: Analyze user ratings and reviews to understand audience preferences and align movie production with content that resonates well with the target audience.

e.) **Collaborate with Industry Experts**: Engage with industry experts and professionals to complement data-driven insights with creative expertise and market knowledge, ensuring a well-rounded approach to movie production.

Conclusion

2. Limitations of the Analysis:

While the analysis provides valuable insights, certain limitations should be considered:

a.) **Data Quality**: The accuracy and completeness of the analysis depend on the quality of the data used. Incomplete or biased data could impact the reliability of the results.

b.) **External Factors**: The movie industry is influenced by various external factors, such as competitor strategies, economic conditions, and cultural events, which are not directly captured in the analysis.

c.) **Subjective Nature**: Movie success involves subjective elements such as artistic creativity, storytelling, and audience emotional connection, which cannot be fully captured through data analysis alone.

Conclusion

3. Future Improvements:

To enhance the project in the future, consider the following:

a.) **Validation and Testing**: Validate the analysis results using external datasets or real-world performance. Testing the model on a separate hold-out dataset can help assess its generalizability.

b.) **Incorporate Audience Sentiment Analysis**: Analyze audience sentiments and reviews using natural language processing techniques to gain deeper insights into audience preferences and reactions.

c.) **Market Research**: Conduct market research and audience surveys to complement data analysis and gather qualitative feedback on movie preferences and expectations.

d.) **Continuous Monitoring**: Continuously monitor movie performance and audience reactions to adapt strategies and improve decision-making based on real-time data.

e.) **Expand Data Sources**: Consider incorporating data from other reliable sources, such as audience demographics and social media trends, to gain a comprehensive understanding of the target audience.

Thank you!

Email: fredrickisaac142@gmail.com
GitHub: https://github.com/FREDRICKKYEKI
LinkedIn: https://www.linkedin.com/in/fredrick-kyeki-554177205/