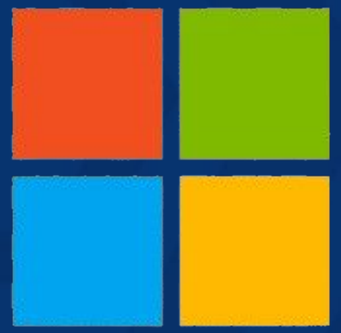
MicroSoft Movie Studio Data Analysis

Presentation

By Julius Kinyua
Moringa School Project



Microsoft

Welcome To Presentation

I'm Julius, and I'll be sharing with you the data analysis process and findings.

Follow me at on X @juliuskczar to learn more.

Table Of Contents

- Business Understanding
- Data Understanding
- Data Cleaning
- Data Analysis
- Conclusion
- Recommendations
- Next Steps

1

Business Understanding

- Microsoft has started a new movie studio
- Need to understand best performing Movie

Main Sources of Data

- IMDb data
- Movies data
- Movie reviews
- Movie budget
- Among others

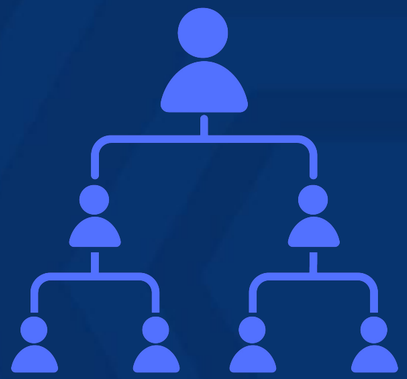
Business Problems to Address

Genres

Which movie genres should the new movie studio focus on?

Market

Which Market should the movie studio target?



Removals

Which movie genres should the movie studiio avoid?

2

Data

Our Data analysis process relies on two main sources of data including:

- IM.db - Containing movie ratings & genre including ratings and number of votes per movie
- The Box Office Mojo - Provide info on the domestic and foreign gross

Data Cleaning

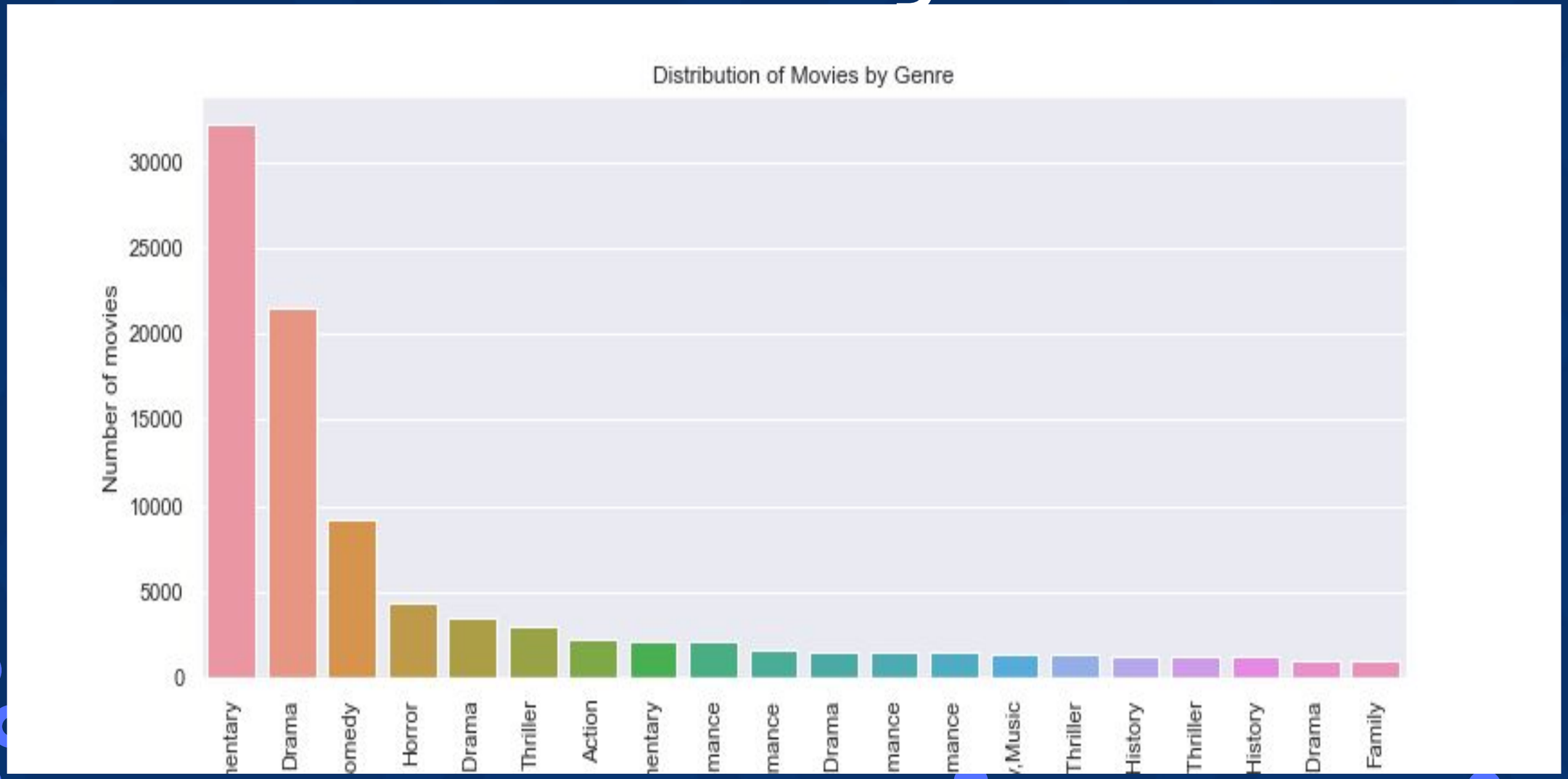
Null Variables

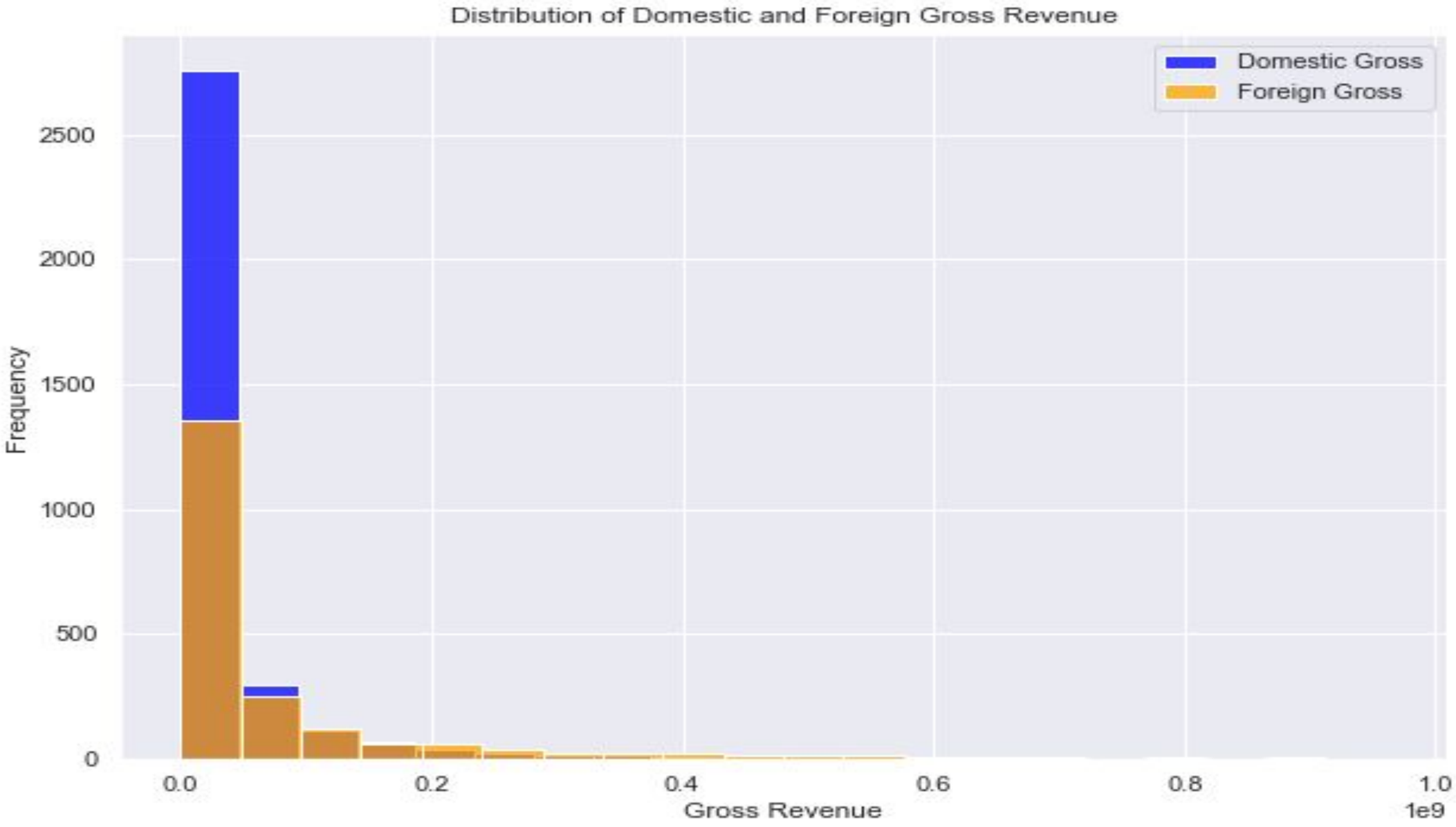
Check for null values in the data

Duplicates

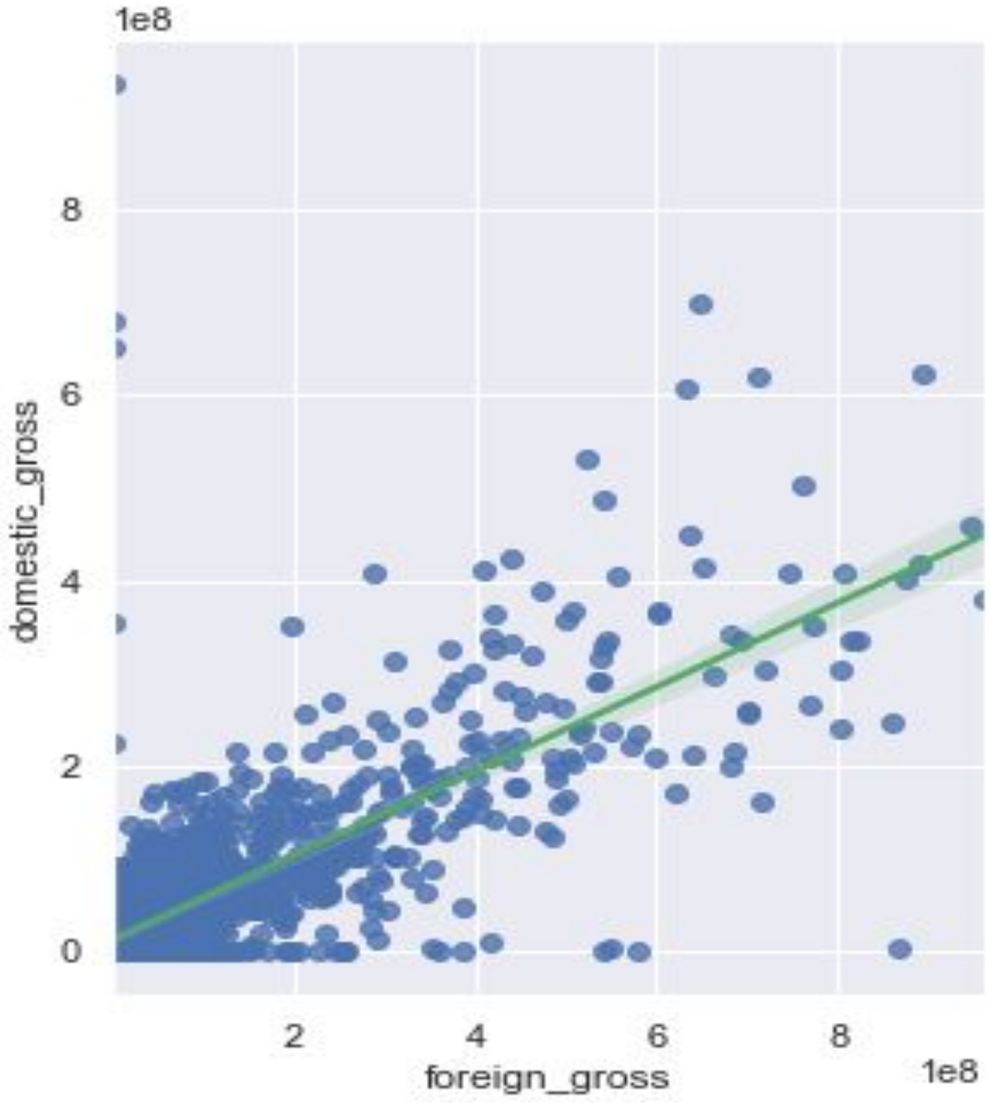
Remove Duplicate values

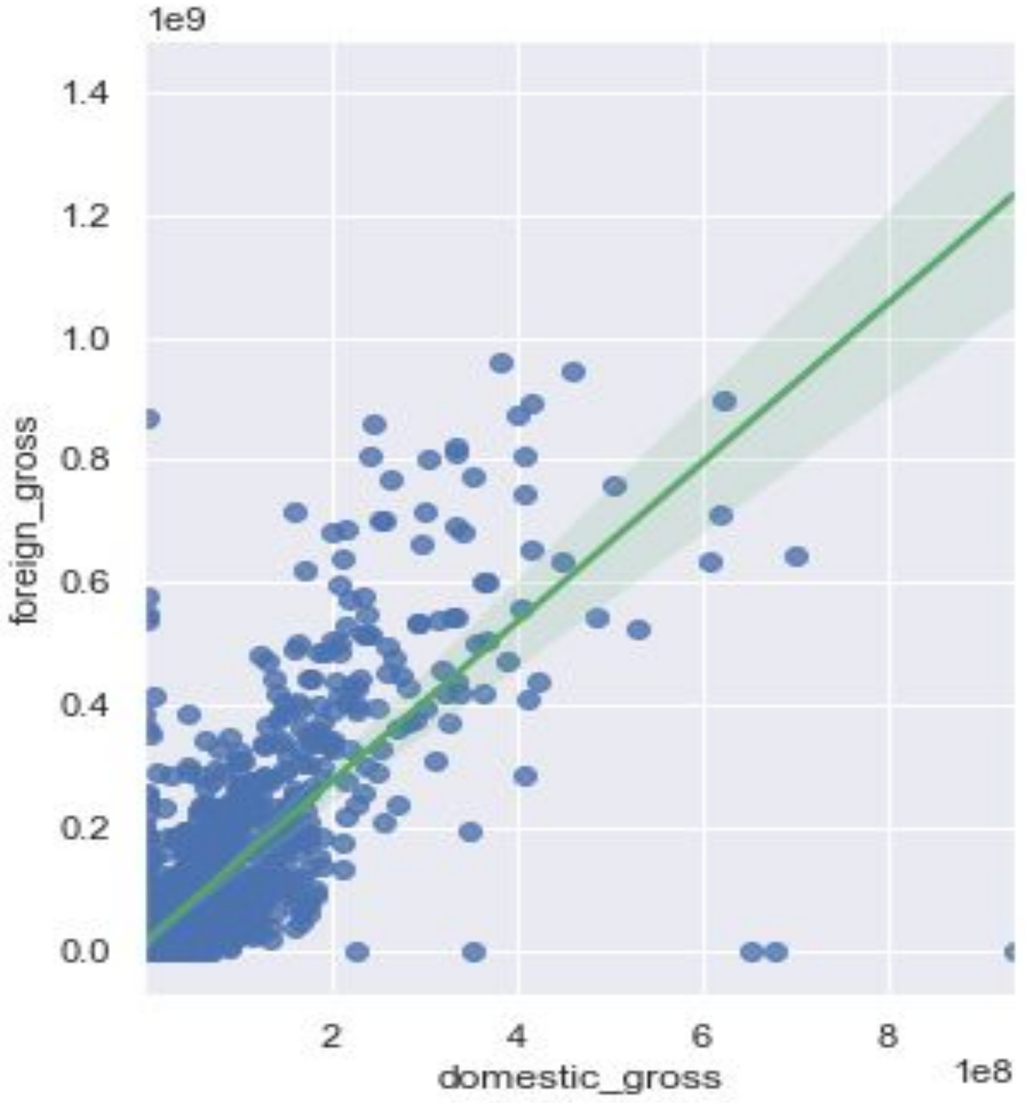
Data Analysis




Distribution of Domestic and Foreign Gross Revenue
Domestic Gross
Foreign Gross
Frequency
2500
2000
1500
1000
500
0
0.0
0.2
0.4
0.6
0.8
1.0
Gross Revenue
1e9

Bivariate Analysis





Conclusion

The 20 most prevalent movie genres include:

- Documentary
- Drama
- Comedy
- Horror
- Thriller
- Action
- Biography
- Romance
- Music
- History
- Family

Cont …

- Typical movie length is around 86 minutes
- Most movies receive domestic gross amounts below the average.
- Films with higher domestic gross tend to have larger foreign gross amounts compared to those with lower domestic gross.
- The majority of movies are rated around 6, and typically receive a low number of votes.
- Genres with fewer results are outlined in part 2 of the data analysis process.

Recommendations

- Choose movie genres from the most popular categories and those that typically receive the highest average votes.
- Aim for movie lengths exceeding the average of 86 minutes.
- Focus primarily on the domestic market, even though foreign markets yield higher income.
- Avoid genres with low average vote counts, as they may indicate a limited audience appeal.

NEXT STEPS

Further analysis could be conducted to identify more precise steps for enhancing the film production process:

- Expanding the dataset
- Procuring revenue data
- Investigating actor-director correlations

Thank You!