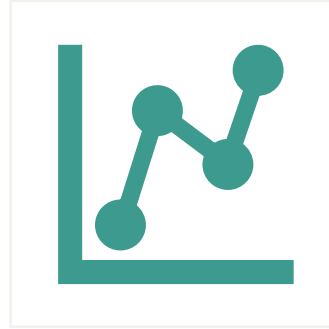# EXPLORATORY DATA ANALYSIS FOR MICROSOFT'S NEW MOVIE STUDIO

Insights for Creating Successful Movies

Keziah Gicheha

3/06/2024

Project Overview

This project uses exploratory data analysis to generate insights for Microsoft.

Tasked with exploring the performance of other films in the box office market, actionable insights have been drawn that aim to be useful to the head of Microsoft's new movie studio to help them decide what type of films to create.

OBJECTIVES

To analyze current box office trends and generate actionable insights for Microsoft's new movie studio

Business Problem

Initiative: Microsoft is seeking to venture into the film industry and wants to create a new movie studio and original video content.

Challenge: Lack of experience and knowledge in movie creation.

Goal: Identify successful film types to guide Microsoft's new movie studio.

Data Description & Understanding

Two datasets were used for this EDA (Exploratory Data Analysis)

1. IMDB database

2. Box Office Mojo – CSV

Key columns and their importance

IMDB tables utilized – Movie _basics table(columns : movie_id, ordering,primary_title, start-year, runtime_minutes genres.)

And

movie_ratings table (movie_id, averagerating, numvotes)

Box office movies columns used – title, domestic gross, foreign gross, genres, average rating

Sources of data

https://www.imdb.com/

https://www.boxofficemojo.com/

Methodology

1. Data cleaning and preprocessing

It was established that both datasets had missing values

Movies basics

runtime_minutes column – 21% of that column had missing values(31,739/773,855 rows)

Original title column – 0.014% (21 /773,855)

Genres column – 3.7% (5408/773,855)

Box office movies

Studio 5/3386 (0.15%)

 domestic_gross   28 /3386 (0.8%)

foreign_gross 1350/3386 (40%)

Methodology(continued)

2. Handling missing values

- Both datasets were cleaned, and null/missing values were removed

- As my rule of thumb, the columns with missing values that are less than 5%, and the rows with the missing values will be dropped, as this will not significantly impact the analysis.

- In both datasets, columns with missing data of more than 20% were imputed using the median to fill in the missing values.

- The median was used, in this case, there were outliers in both datasets - to prevent skewness

Exploratory Data Analysis (EDA)

1. Which movies have the highest average ratings?

id	primary_title	genres	averagerating
922	Love on a Leash	Documentary,Family,Romance	9.7
921	Foosballers	Comedy,Documentary,Sport	9.7
920	Lost Conquest	Comedy,Documentary,Fantasy	9.4
919	Sunakali	Documentary,Family,Musical	9.3
918	The Long Count Teaser	History,Sport	9.2

Exploratory Data Analysis (EDA)

Which movies have the lowest ratings?

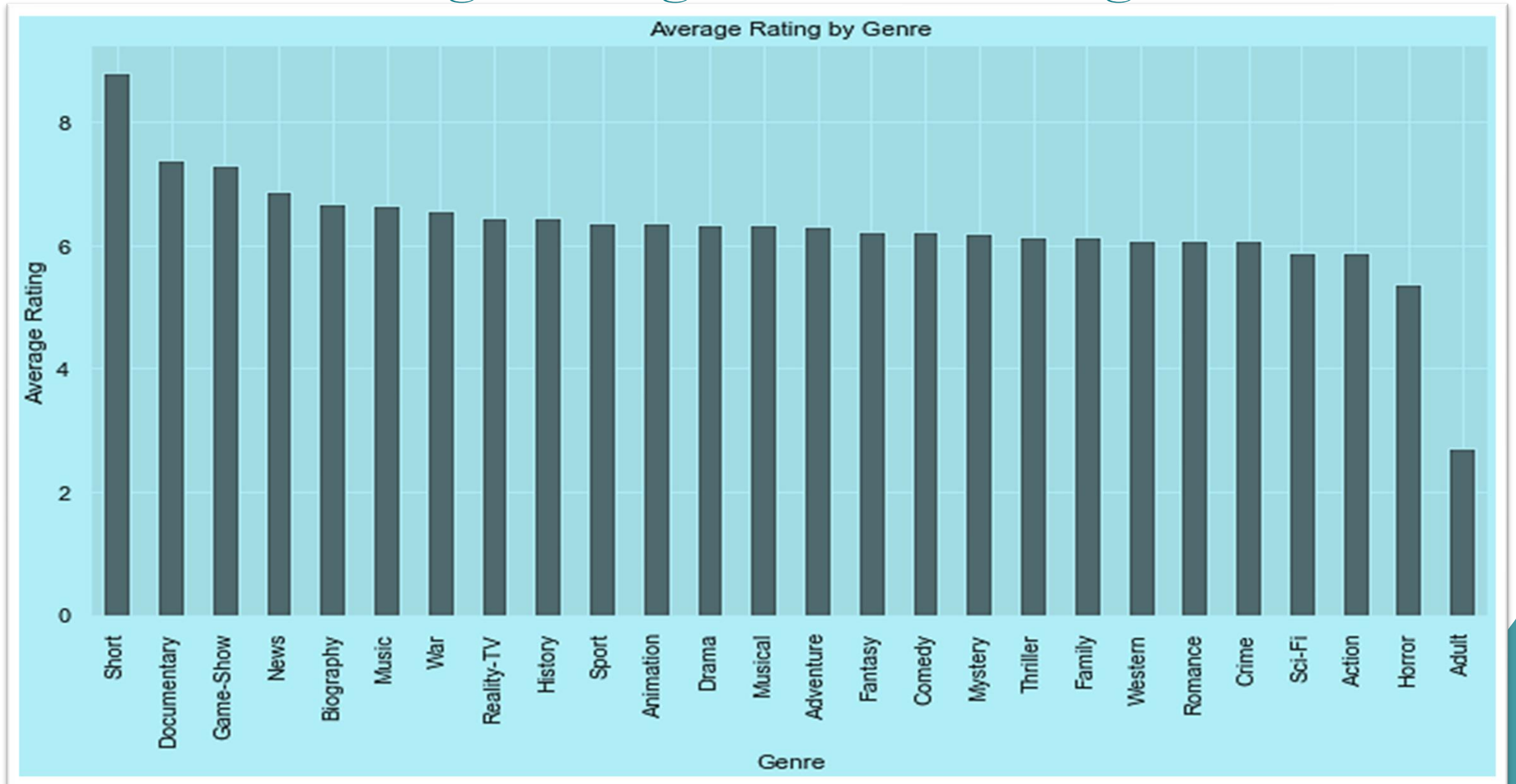|	primary_title	genres	averagerating
0	Maldito Amor	Comedy,History,Horror	1.1
1	Futures Past	Action,Biography,History	1.3
2	Eyal Hirrifah	Comedy,Musical,Sport	1.4
3	The Devil's Flower	Mystery,Romance	1.7
4	Foodfight!	Action,Animation,Comedy	1.9

2. Distribution of Movie Ratings
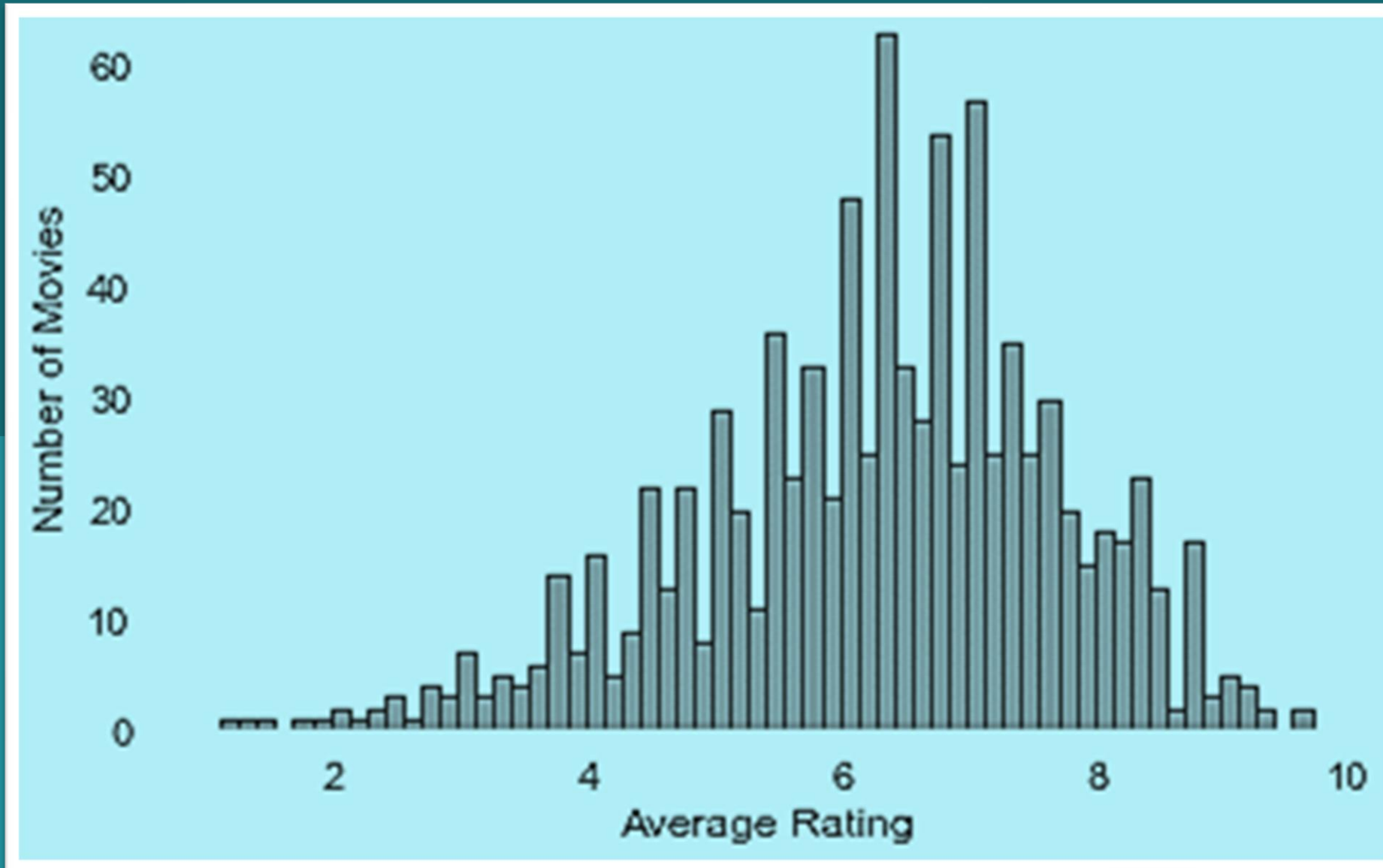
3. Most Common Movie Genres
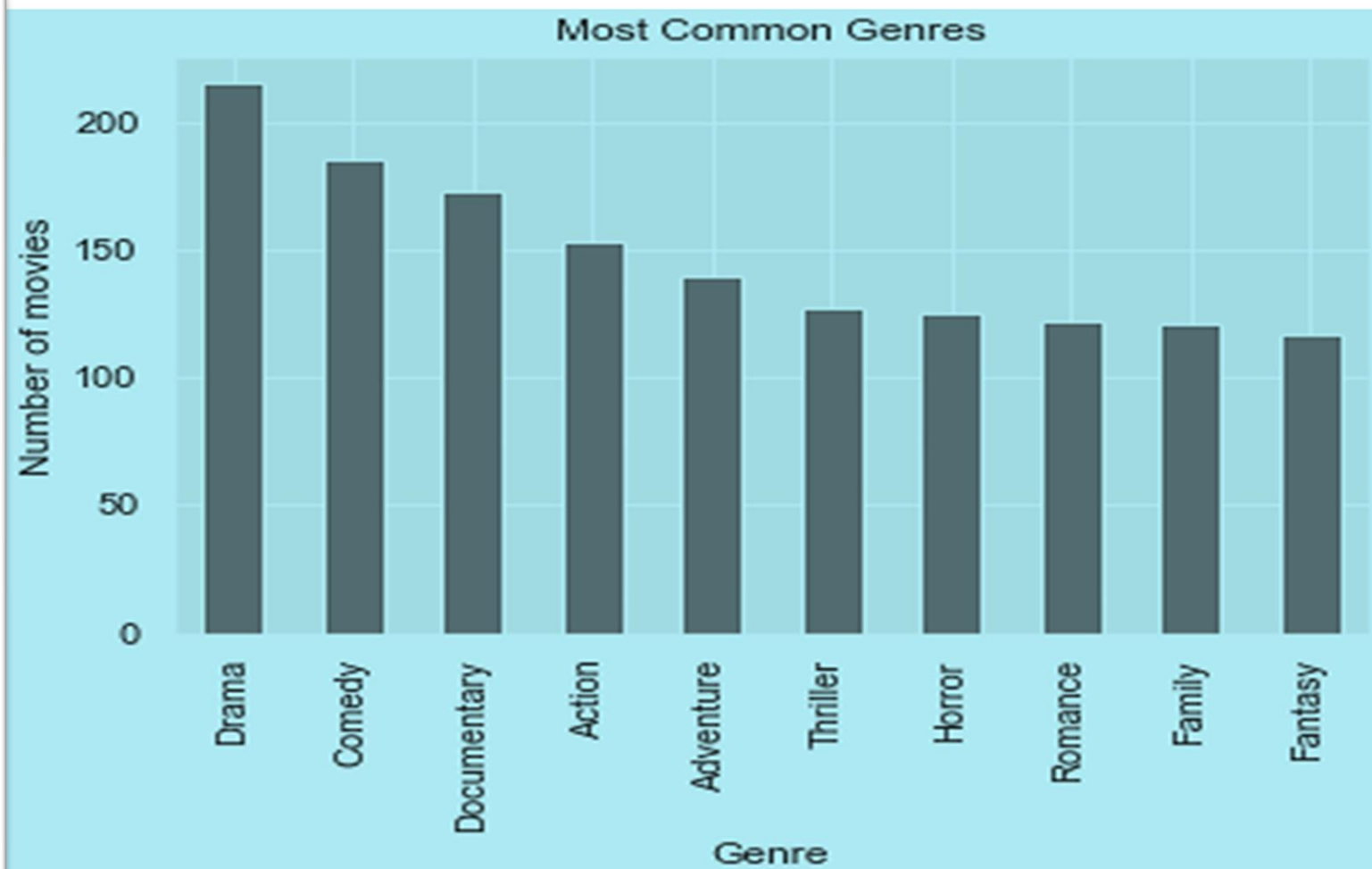
4. Average rating based on each genre?
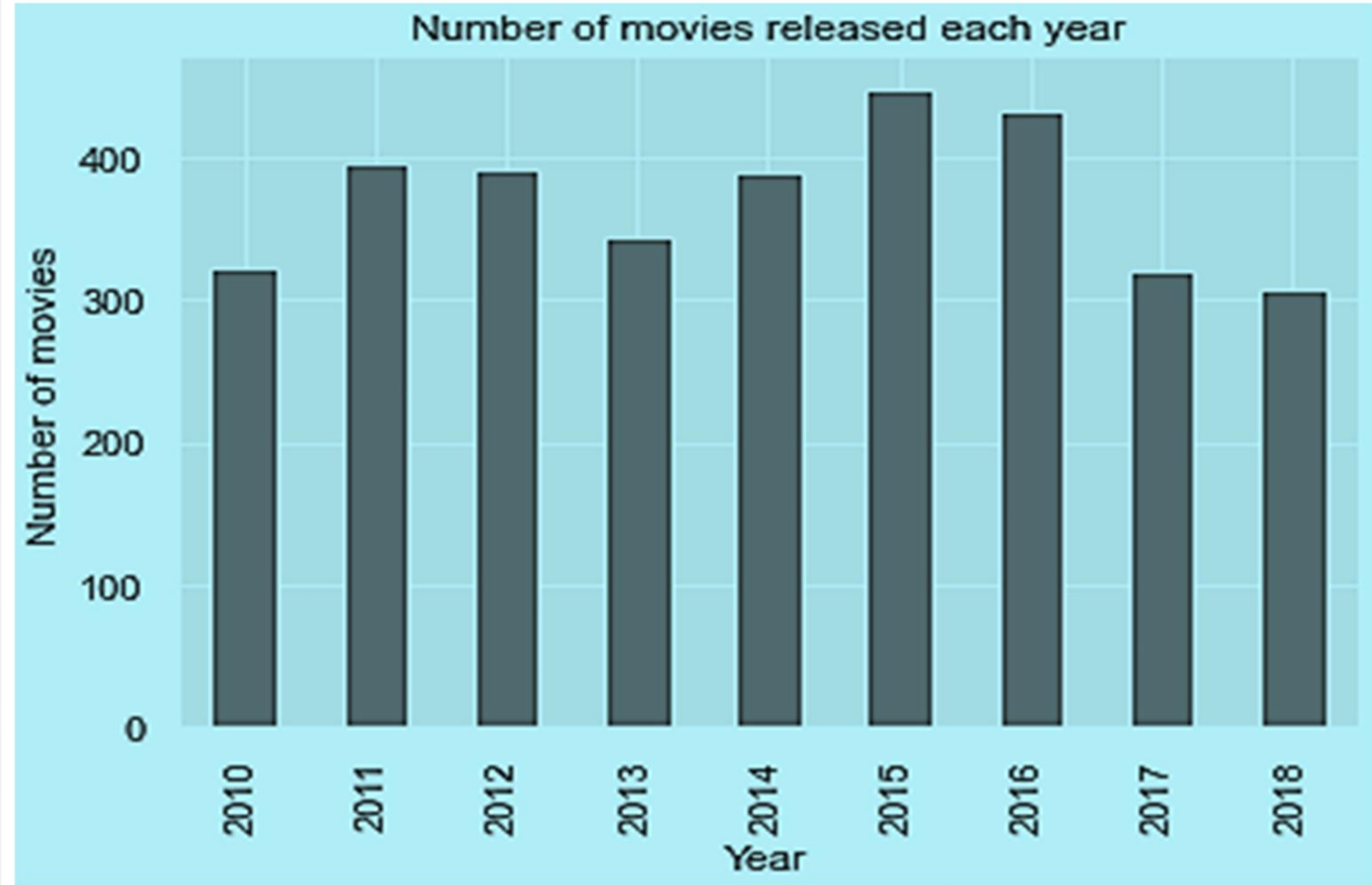

Average Rating by Genre

5. Distribution of Foreign Gross
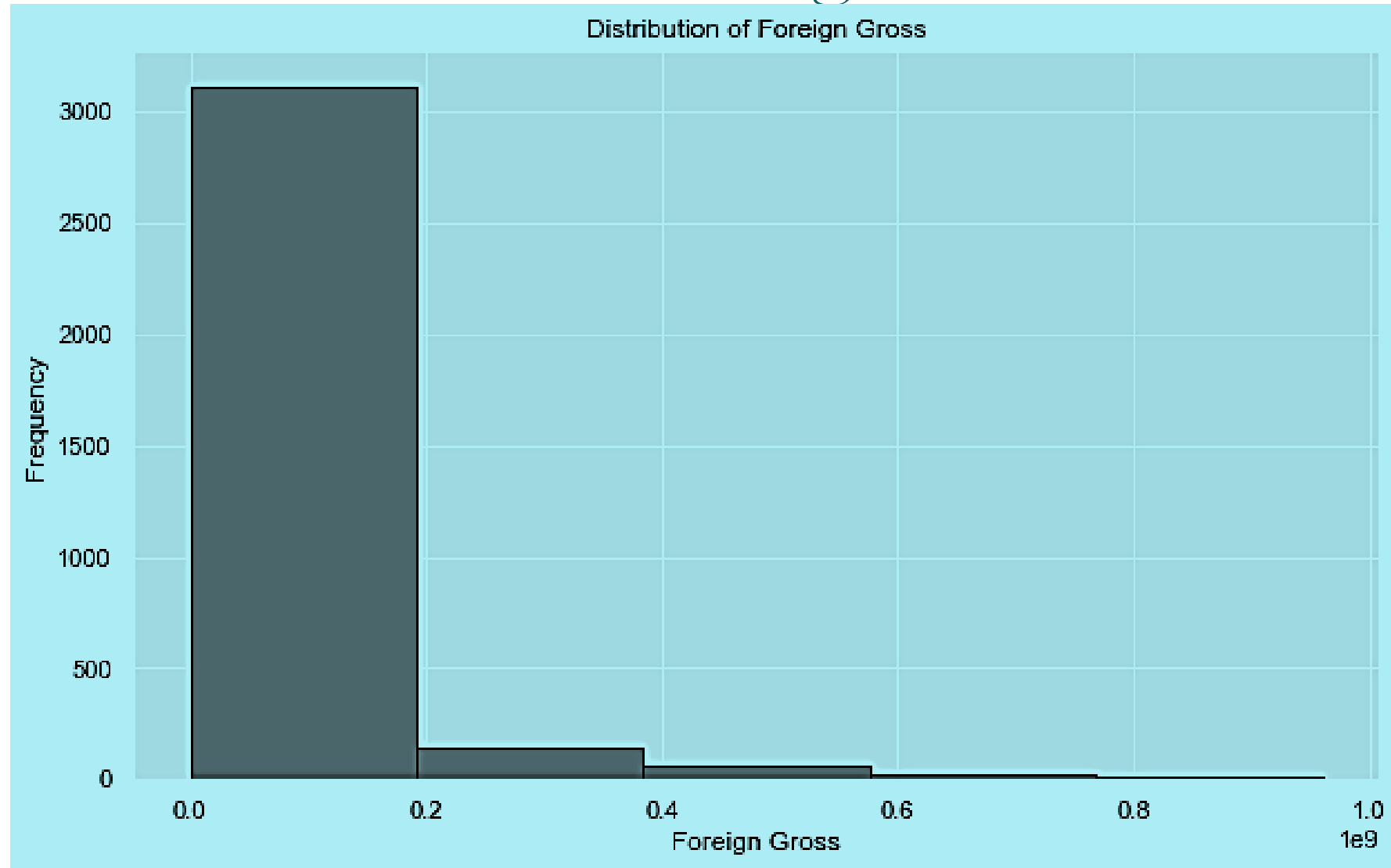
6. Distribution of movies per year

7. Comparison of Gross revenues

Gross Domestic Revenue by Studio

Average Domestic Gross by Studio (Top 30 Studios)
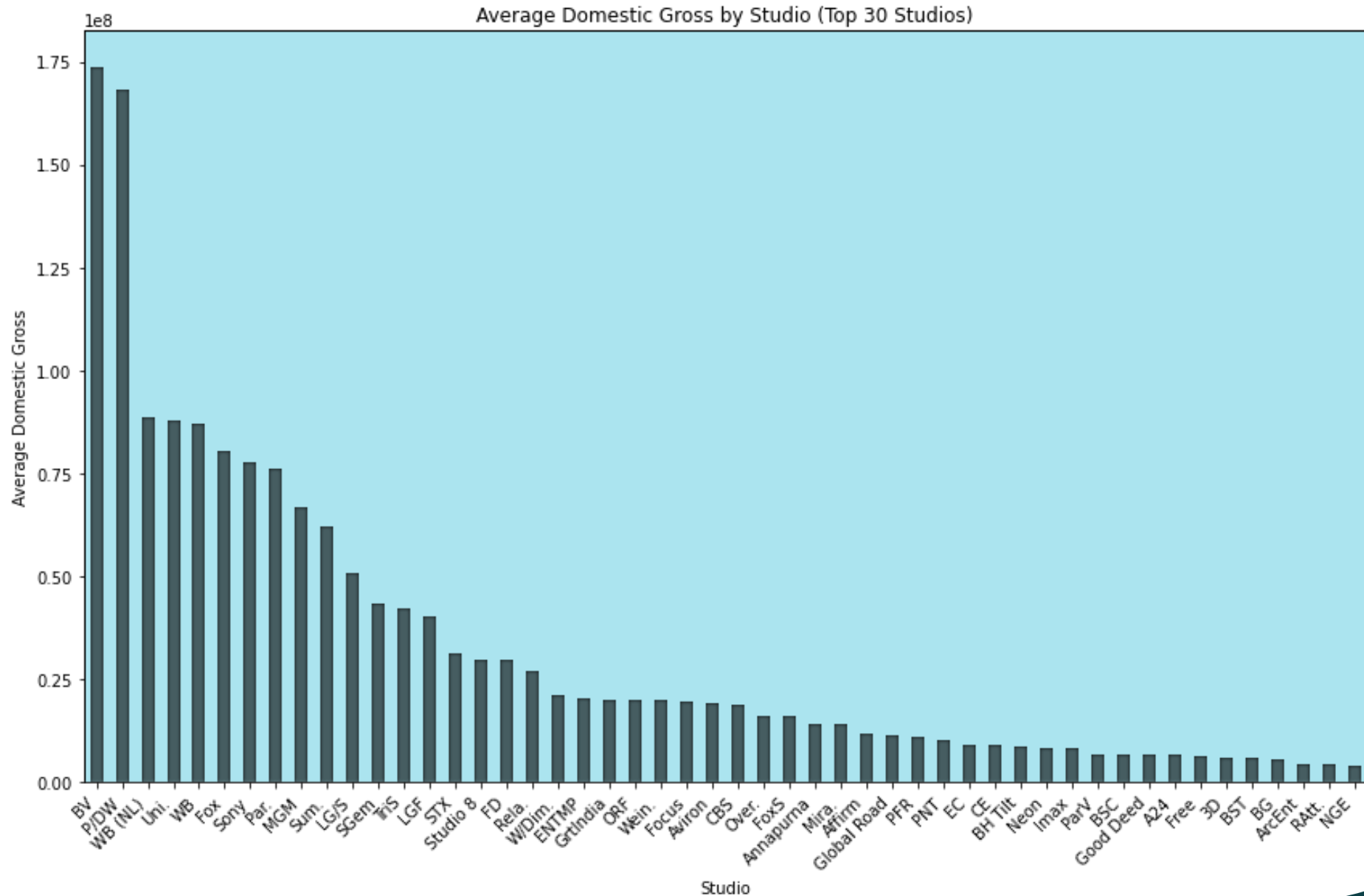
Gross Foreign Revenue by Studio

Average Foreign Gross by Studio (Top 30 studios)
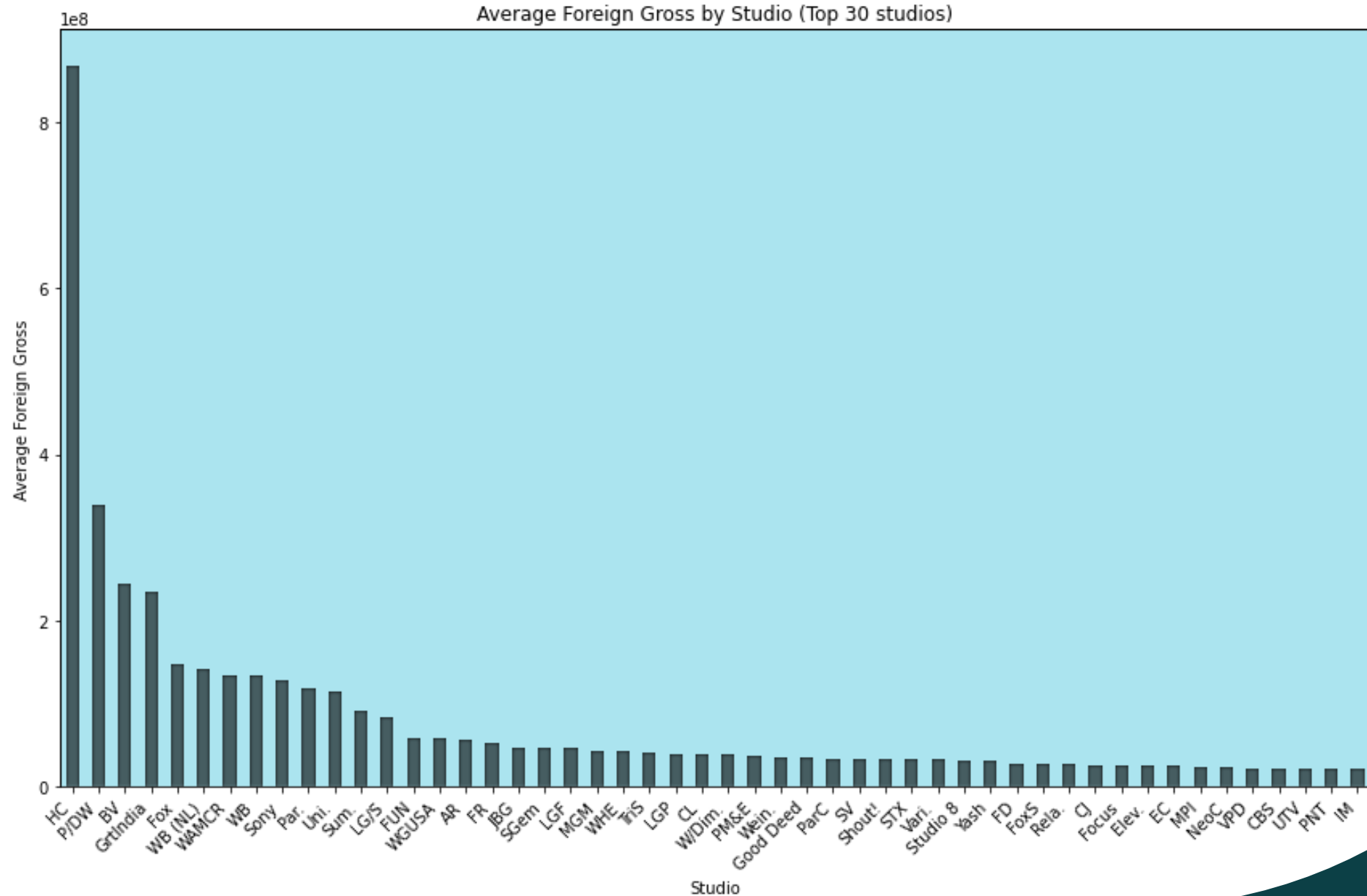
Top-rated movies by Gross revenues

id	TITLE	TOTAL_GROSS
727	Marvel's The Avengers	1.518900e+09
1875	Avengers: Age of Ultron	1.405400e+09
3080	Black Panther	1.347000e+09
328	Harry Potter and the Deathly Hallows Part 2	1.341500e+09
2758	Star Wars: The Last Jedi	1.332600e+09
3081	Jurassic World: Fallen Kingdom	1.309500e+09
1127	Frozen	1.276400e+09

Recommendations

- Focus on popular movie genres with a high rating – Invest in genres that are consistently performing well at the box office(Documentaries, Action, Drama, Adventure, Comedy,)

- Benchmark on studios that have a high domestic gross and foreign gross(BV, P\DW, WB, Uni,)- possibly collaborate in new movies with them.

- Adapt to Trends: Monitor and adapt to changing genre trends to stay relevant.

Next steps

- Further analysis of specific sub-genres or niche markets

- Research on production costs vs. revenue for different genres

- Pilot projects or collaborations to gain industry experience

Conclusion

- In conclusion, this venture stands out to be a wise business venture, as the revenue to be drawn from this is promising in both the domestic and gross numbers.
- Focus on the recommended movie genres, documentaries, action and comedy
- Collaborate with current leading studios for pilots.

Presenter : Keziah Gicheha
www.linkedin.com/in/gichehawambui

Thank you!
Any Questions?