Data Insights for Microsoft's Film-making Studio

Table of contents

01. Overview

Project Overview:

In this project, I aim to assist Microsoft in leveraging data-driven insights to establish its new venture into film production. As the company embarks on establishing its movie studio, my objective is to conduct exploratory data analysis to identify trends and patterns in the film industry, specifically focusing on box office performance. By analyzing datasets from various sources such as Box Office Mojo, IMDb, Rotten Tomatoes, and TheMovieDB, I seek to provide actionable recommendations to guide Microsoft in making informed decisions regarding the types of films to produce.

02. Business Understanding

The film industry is constantly evolving, with new trends emerging and audience preferences shifting. As major companies like Netflix, Disney, and Amazon invest heavily in original content creation, Microsoft recognizes the opportunity to enter the realm of film production. However, with limited experience in the movie-making business, Microsoft faces the challenge of determining what types of films will resonate with audiences and achieve success at the box office.

Stakeholders

Microsoft, in particular the management team in charge of the company's recent film production initiative, is the project's main stakeholder. The insights derived from this analysis will also be beneficial to important decision-makers in Microsoft entertainment division, such as producers, directors, and marketing executives. These stakeholders can choose the right films to make, manage resources wisely, and create efficient marketing campaigns by knowing consumer preferences and industry trends.

Questions the project aims to answer?

The project aims to answer several key questions to address the business problem faced by Microsoft in its new venture into film production. These questions include:

- What are the current trends in the film industry, particularly in terms of box office performance and audience preferences?
- Which genres are currently the most popular among audiences, and how do they vary in terms of box office revenue and critical reception?
- What factors contribute to the success of a film at the box office, and how do they differ across genres?
- How do audience ratings on platforms like IMDb and Rotten Tomatoes correlate with box office performance, and which platforms are most influential in driving audience engagement?
- Are there specific production strategies or marketing approaches that have proven successful in maximizing box office revenue and audience satisfaction?

03. Data Understanding

Introduction to the Dataset:

In my project, I leverage multiple datasets sourced from reputable platforms in the film industry to gain insights into various aspects of movie production, performance, and reception. These datasets provide a rich source of information essential for understanding the dynamics of the film industry and informing strategic decision-making for Microsoft's entry into film production.

The primary datasets utilized in our analysis include:

- The Movie Database (TMDB) Dataset
- Movie Budget Dataset
- Info_data Dataset
- Reviews data Dataset

Data Sets Overview

tmdb_movies Dataset:

- This dataset contains information about movies sourced from The Movie Database (TMDB).
- Features likely include movie titles, genres, release dates, production companies, budgets, and revenue.
- It serves as a valuable source for analyzing movie attributes, such as genre trends, budget allocations, and revenue generation.

movie_budget Dataset:

- ★ The dataset provides information on movie budgets and financial performance.
- ★ Features include movie titles, production budgets, domestic and worldwide box office revenue, and release dates.
- ★ It allows for the analysis of box office success relative to production costs and provides insights into profitability and return on investment (ROI) for movies.

Data Sets Overview

info_data Dataset:

- This dataset contains movie information from Rotten Tomatoes (RT), including movie titles, genres, ratings, and runtime.
- Features may also include production companies, directors, and cast information.
- It complements other datasets by providing critical metadata and ratings from the Rotten Tomatoes platform, facilitating analysis of critical reception and audience engagement.

reviews_data Dataset:

★ The dataset contains reviews of movies from Rotten Tomatoes, including critic and audience reviews.

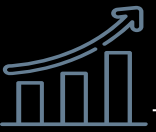
★ Features may include review texts, ratings, and reviewer information.

★ It provides detailed feedback on movie performance and audience sentiment, aiding in understanding audience preferences and critical acclaim.

Data Limitation

Limitations of the Data:

- The data may be subject to biases and inaccuracies inherent in self-reported or user-generated content, such as user ratings on IMDb or Rotten Tomatoes.
- Data availability and completeness may vary across sources, potentially limiting the scope of the analysis or introducing biases in the results.
- The dataset may not capture all relevant factors influencing a movie's success, such as marketing strategies, distribution channels, or cultural influences.



04. Data Analysis

- After loading the datasets, I made sure everything was neat and tidy, like making sure all the data was in the right order and using the same units and terms throughout. This helps us understand the data better.
- Next, I checked if there were any missing values in the datasets and figured out the best way to deal with them. Sometimes we filled in the gaps and sometimes we just left them out.
- Then, I dug into the data to find out what kinds of movies people like and how well they do at the box office. By looking at things like popularity and earnings, we could see which movies were really hitting the mark with audiences.
- I also took a closer look at different movie genres to see which ones were getting the most attention from viewers and making the most money. This helped us understand what kinds of movies people are really into.
- We used some techniques to analyze the data, like looking at averages, making graphs, and seeing how different variables relate to each other. These visual tools helped us spot trends and correlations in the data.

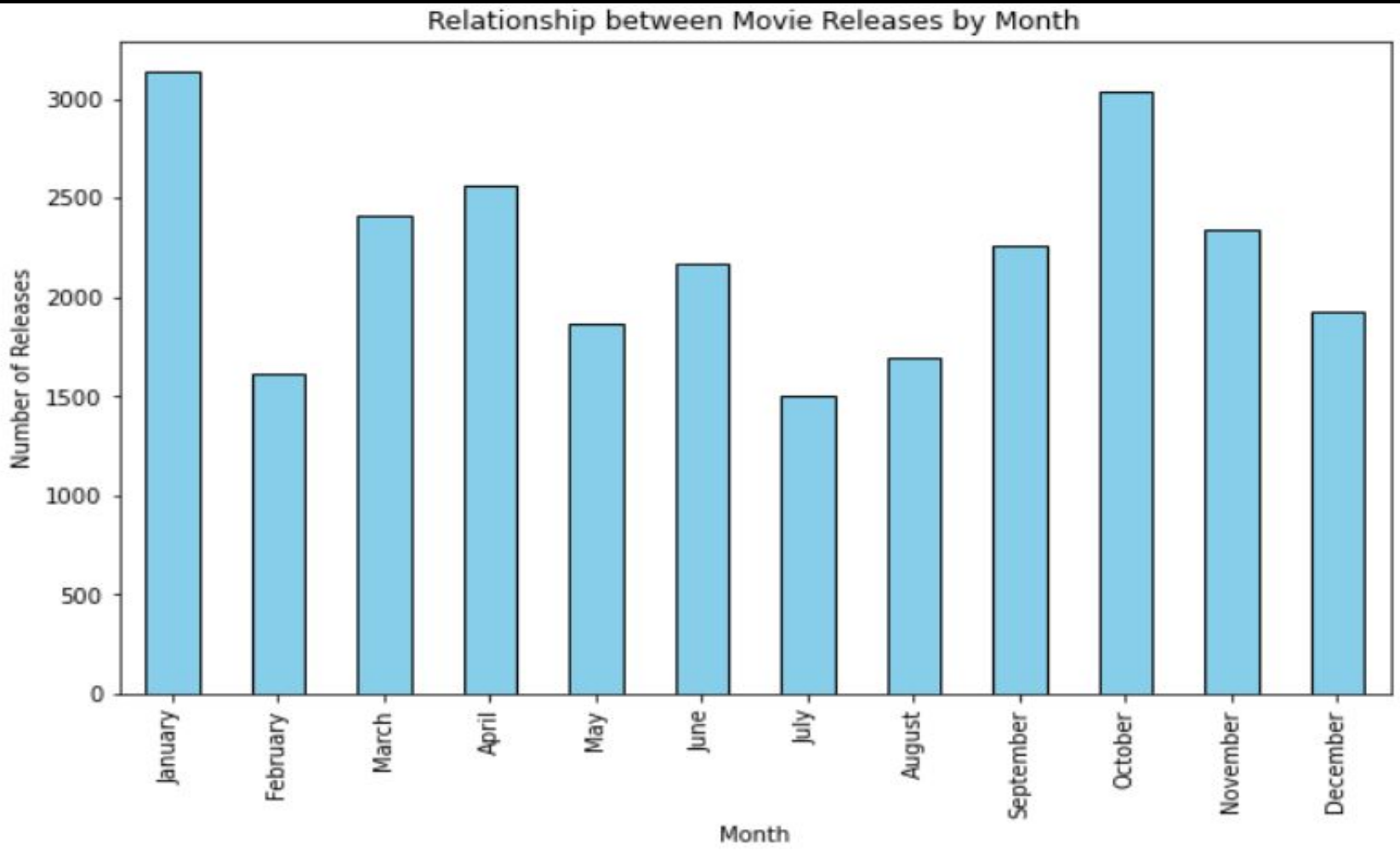
05. (i) Recommendations

Movie Releases Strategy

Microsoft should strategically schedule movie releases by analyzing release patterns and competition dynamics. By targeting peak months with fewer competing movies, Microsoft can maximize audience attention and box office results.

Additionally, by avoiding direct competition and adjusting release schedules in response to competitors' activities, Microsoft can enhance market positioning and increase the chances of success.

Analyzing audience participation and preferences allows Microsoft to tailor content offers accordingly, increasing engagement and viewership. Overall, strategic release scheduling is essential for optimizing box office impact and ensuring the success of Microsoft's movie productions.


Relationship between Movie Releases by Month
Number of Releases
0
500
1000
1500
2000
2500
3000
January
February
March
April
May
June
July
August
September
October
November
December
Month

(ii) Recommendations

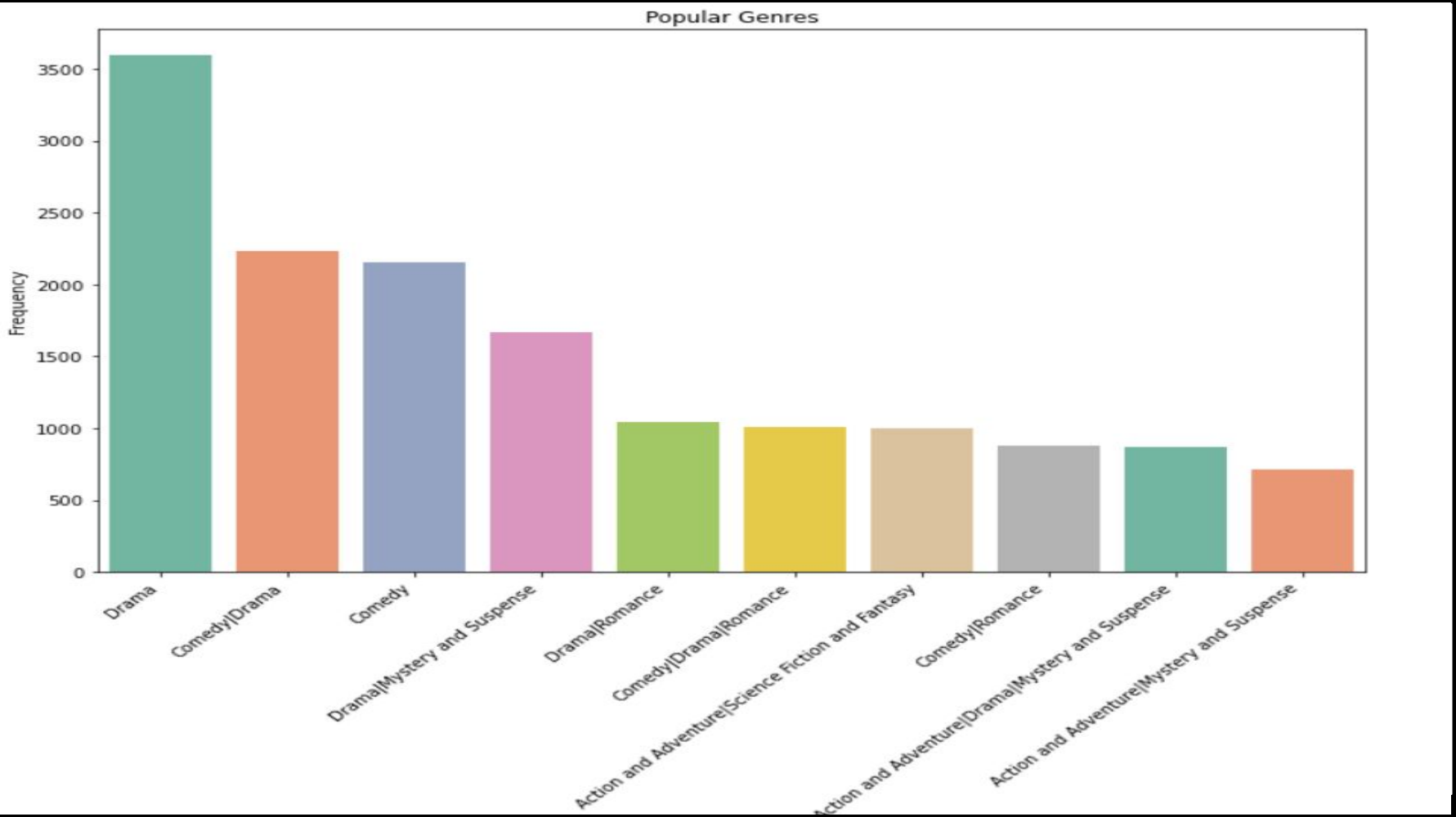
Genre-Centric Film Strategy

Microsoft should consider focusing on the development of films within popular genres such as drama, comedy, and mystery.

By understanding the frequency of these genres in the dataset and recognizing their popularity among viewers, Microsoft can maximize audience engagement and box office success.

This approach involves strategic genre selection, aiming to capture audience attention and capitalize on market demand while maintaining creative flexibility to deliver unique content.

Overall, prioritizing these genres in Microsoft's film production strategy presents an opportunity to establish a strong presence in the industry and compete effectively with established players.


Popular Genres
Frequency
3500
3000
2500
2000
1500
1000
500
0
Drama
Comedy|Drama
Comedy
Drama|Mystery and Suspense
Drama|Romance
Comedy|Drama|Romance
Action and Adventure|Science Fiction and Fantasy
Comedy|Romance
Action and Adventure|Drama|Mystery and Suspense
Action and Adventure|Mystery and Suspense

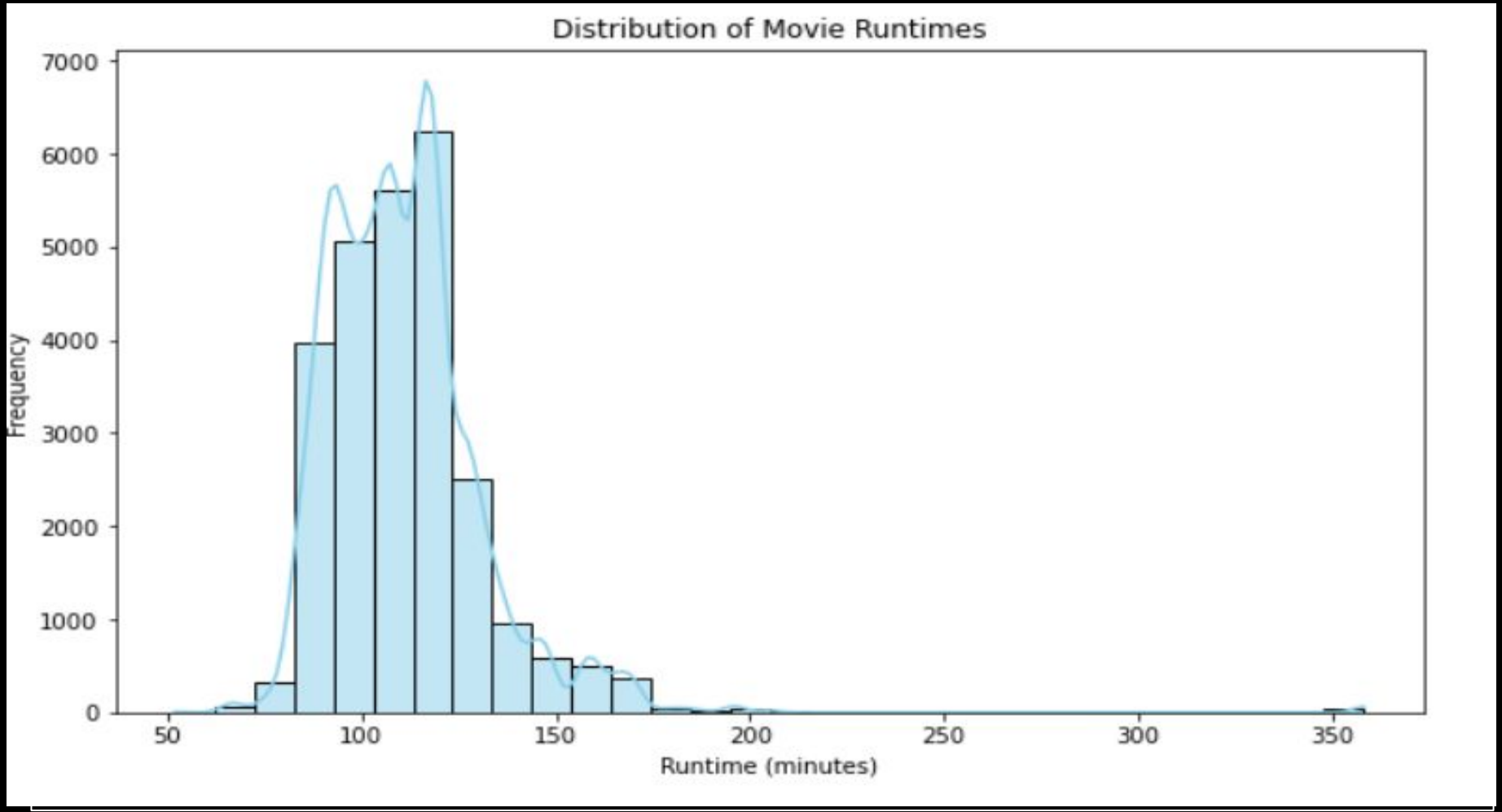
(iii) Recommendations

Optimize Film Runtimes for Audience Engagement

The recommendation advises Microsoft to optimize film runtimes based on insights from the distribution of movie runtimes in the dataset.

By analyzing average runtimes, identifying trends and anomalies, assessing audience preferences, and tailoring film duration accordingly, Microsoft can enhance audience engagement and satisfaction.

This approach ensures that Microsoft's film productions align with audience expectations, leading to increased box office success and long-term viability in the film industry.


Distribution of Movie Runtimes
Frequency
Runtime (minutes)
0
1000
2000
3000
4000
5000
6000
7000
50
100
150
200
250
300
350

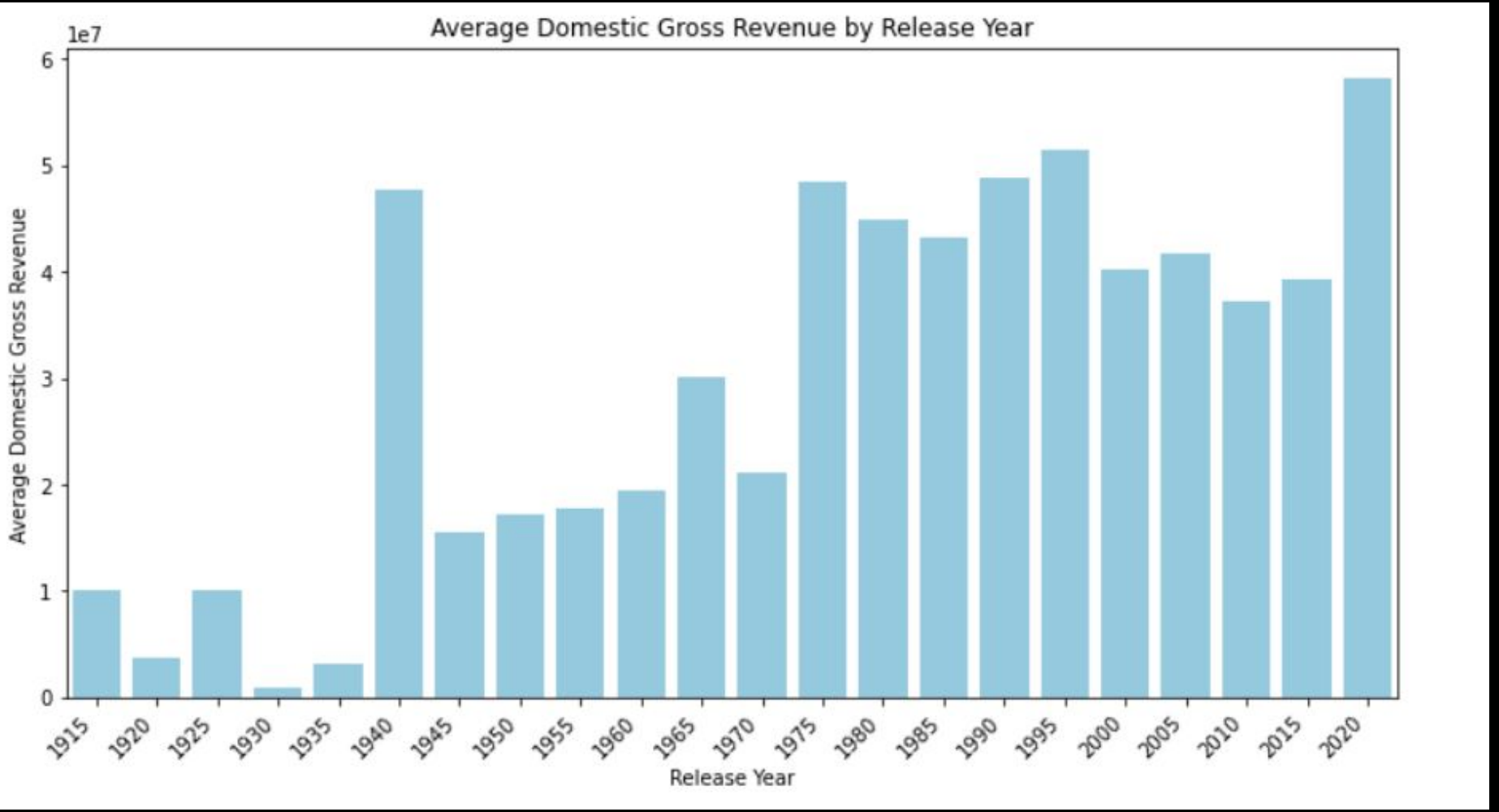
(iv) Recommendations

Strategic Planning Based on Revenue Trends

The recommendation advises Microsoft to leverage insights from the general trend of total gross revenue over various release dates to inform strategic planning in the film production industry.

By conducting long-term revenue analysis, forecasting future revenue, adapting strategies based on revenue trends, strategically positioning film releases, and continuously monitoring industry dynamics, Microsoft can optimize its positioning in the market, maximize revenue generation, and foster long-term success.

This data-driven approach ensures Microsoft remains agile and responsive to market dynamics, enhancing its competitive advantage and profitability in the industry.


Average Domestic Gross Revenue by Release Year
1e7
Average Domestic Gross Revenue
0
1
2
3
4
5
6
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Release Year

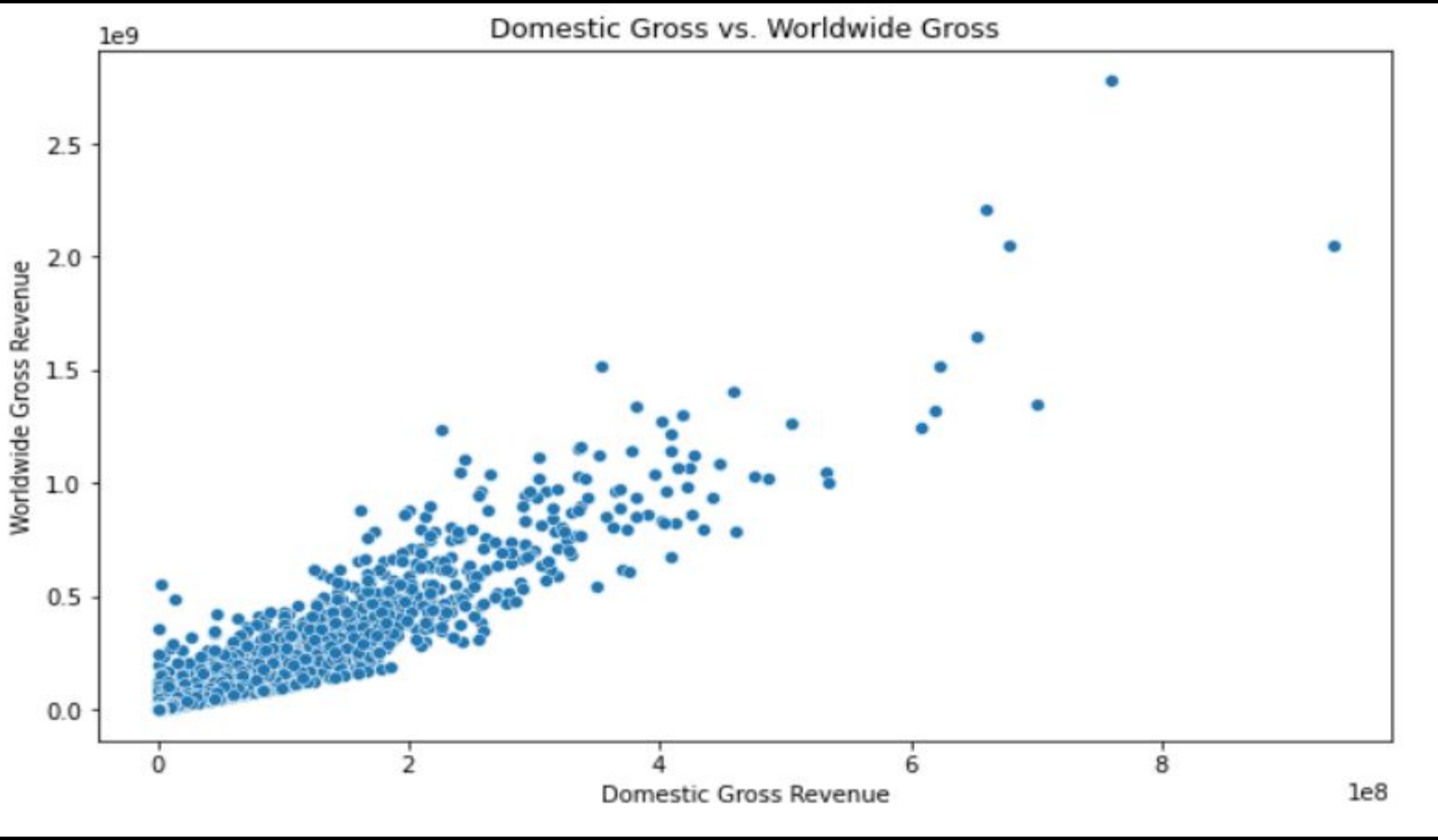
(v) Recommendations

Optimize Revenue Distribution Strategy

The recommendation advises Microsoft to optimize its revenue distribution strategy based on insights derived from the scatter plot depicting the relationship between domestic and global gross revenue for specific films.

By conducting market analysis, targeting marketing strategies, forging strategic partnerships, investing in content localization, and adopting a data-driven approach, Microsoft can maximize revenue generation both domestically and internationally.

This data-driven approach ensures that Microsoft's films resonate with diverse audiences worldwide, leading to increased box office success and long-term profitability in the film industry.


Domestic Gross vs. Worldwide Gross
1e9
2.5
2.0
1.5
1.0
0.5
0.0
Worldwide Gross Revenue
0
2
4
6
8
Domestic Gross Revenue
1e8

Thanks!

Do you have any questions?

HAWKINS MURITHI KOBIA

https://www.linkedin.com/in/hawkins-murithi-635b502aa/