Microsoft New Movie Studio Exploratory Analysis

Prepared By Mucui Bernard ,March 2023

Executive Summary

Microsoft is launching a new movie studio and has commissioned an exploratory analysis to help them understand the current market trends and make data-driven decisions about the type of movies to produce. The analysis focused on three key factors: the original language of movies, their return on investment (ROI), and their vote average.

After analyzing the data, several trends were identified. The analysis found that English-language movies are the most popular. In terms of ROI, it was found that movies with higher budgets tend to perform better at the box office, but there is also a significant degree of variability in this relationship. Finally, the analysis found a positive correlation between a movie's vote average and its box office earnings, indicating that higher-rated movies tend to perform better financially.

These findings have important implications for Microsoft's new movie studio. By focusing on producing English-language movies and investing in high-budget productions that have the potential for high ROI, the studio can increase its chances of success. Additionally, producing high-quality movies that receive high vote averages is likely to result in financial success.

Overall, the exploratory analysis provides valuable insights into the movie industry and will enable Microsoft to make informed decisions about the type of movies to produce for its new studio.

Outline

- Business Problem Statement
- Data characteristics
- Methodology
- Results
- Key Findings
- Conclusions/recommedations

Business Problem

Microsoft is launching a new movie studio and wants to understand the current market trends to make informed decisions about the type of movies to produce. In this exploratory analysis, we will focus on the original language, return on investment (ROI), and vote average to identify popular movie trends and help Microsoft make data-driven decisions."

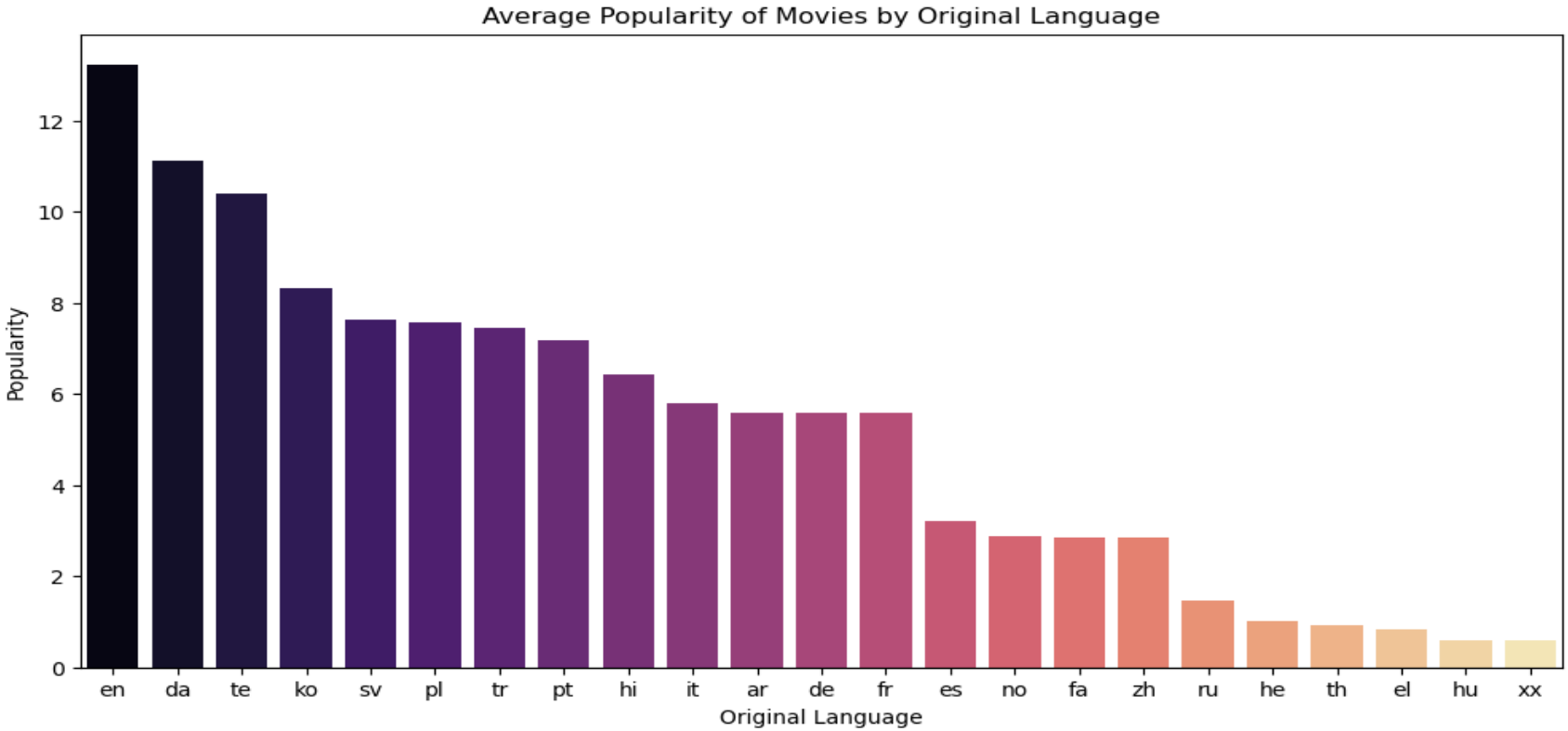
Data Characteristics

- The data was obtained from publicly available sources such as Box Office Mojo, IMDB, Rotten Tomatoes, The Movie DB, and The Numbers.
    - These datasets contain essential information about movies, such as box office earnings, budget, original language, budget, release date, ratings etc.
    - The original language variable includes over 20 different languages, with English being the most common language used in production.
    - The budget and revenue variables have a wide range of values, with some movies having very high budgets and revenues, while others have much lower budgets and revenues.
    - The data set includes some missing values, particularly in the budget and revenue variables, which were imputed using the mean values for the respective variables.
    - The data set contains outliers in some variables, particularly in the budget and revenue variables, which were addressed using logarithmic transformation to improve the normality of the data.

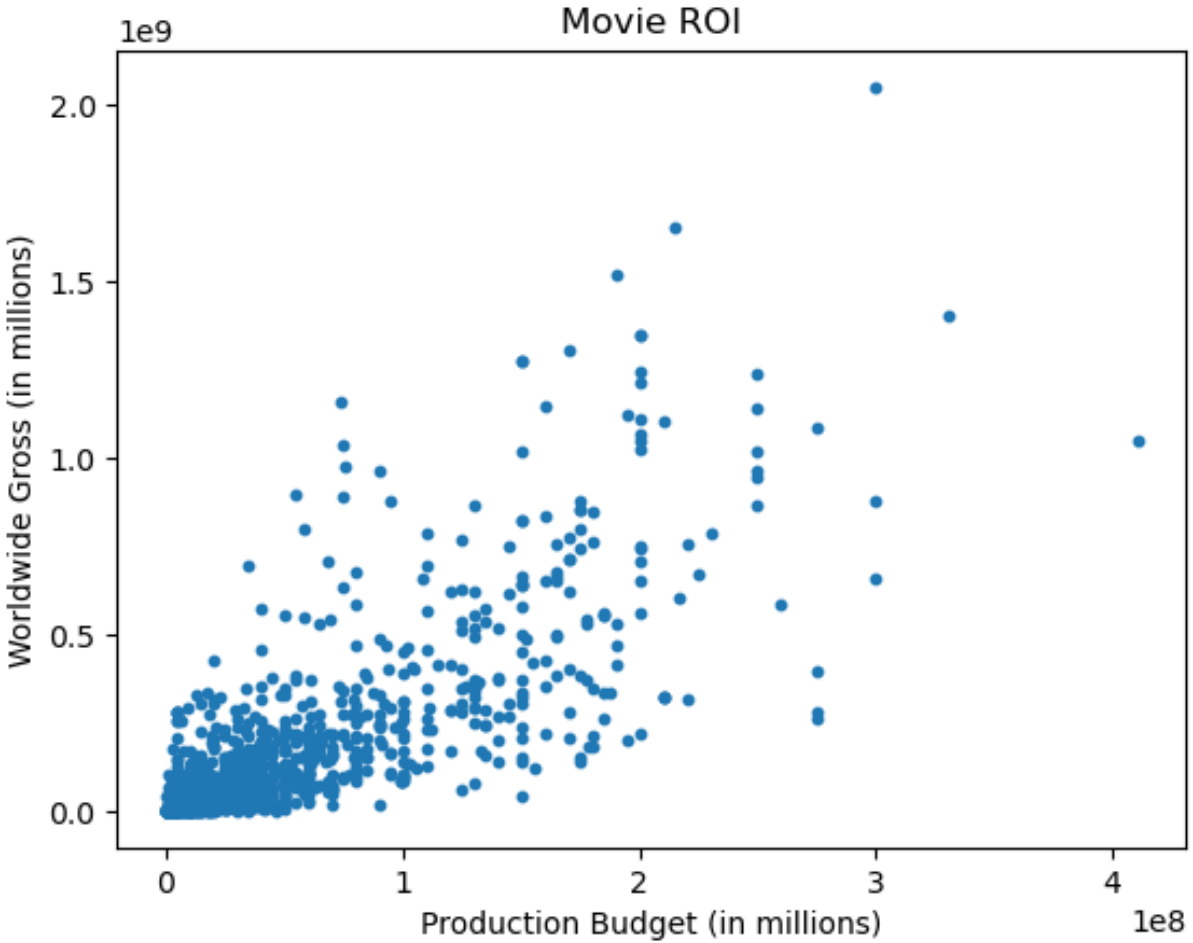
Methodology

- This project involved an exploratory data analysis approach to analyze the movie data obtained from publicly available sources such as The Movie Database (TMDb) and IMDB. The data was cleaned, preprocessed, and merged into a single data frame (df2) for analysis.
- To address the research questions, descriptive statistics and visualizations such as histograms, scatter plots, and box plots were used to explore the relationship between the variables of interest.
- In addition, correlation analysis was performed to quantify the strength and direction of the relationship between variables.
- Python programming language and several data analysis libraries such as pandas, seaborn, and matplotlib were used to perform the data analysis.
- This approach allowed us to gain important insights into the relationship between variables of interest and answer our research questions formulated.

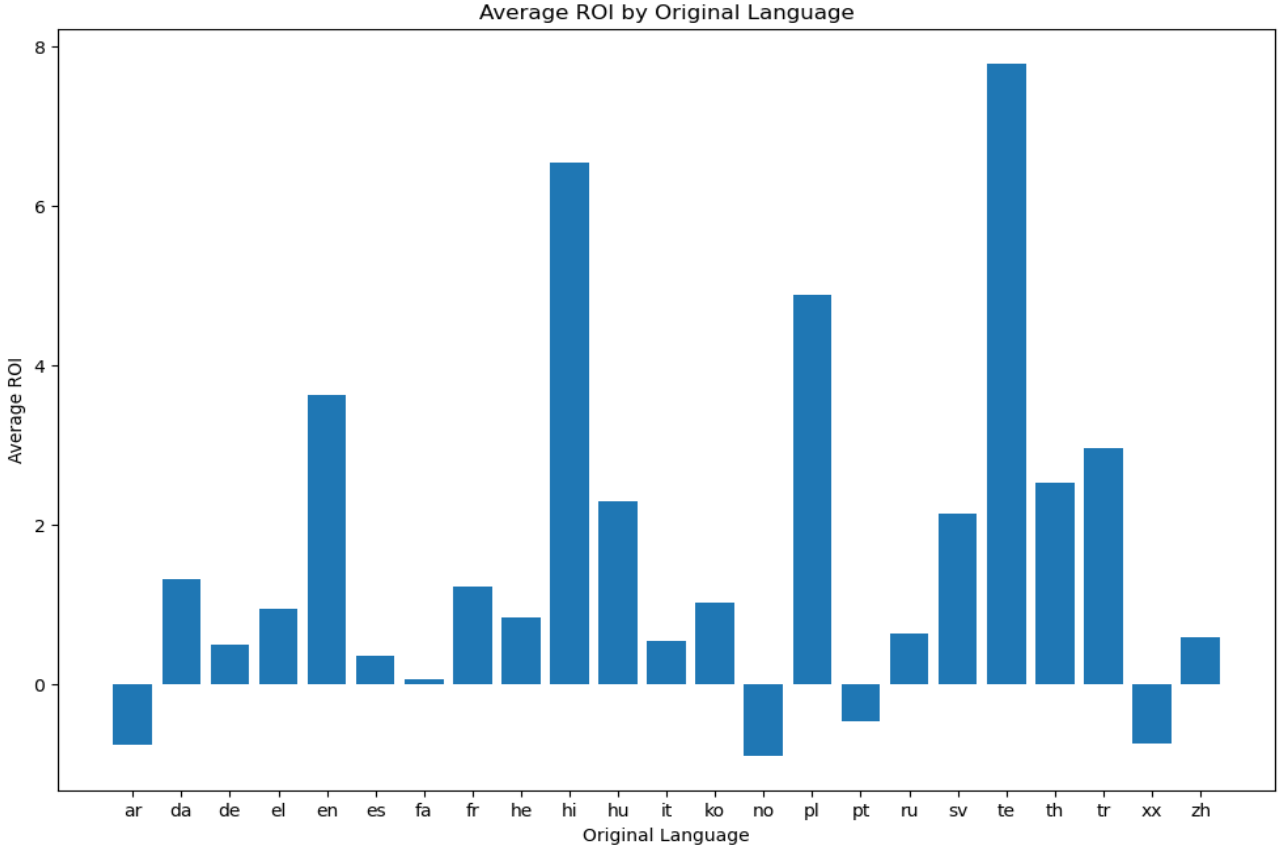
Results/Outcome of the Analysis



From the bar chart , It's clearly evident that English language is by far the most common language for movie production in the dataset



- we can observe a positive correlation ,i.e. movies with higher budgets tend to have higher box office performance.
- However, there are also many data points with low budgets and high revenue, suggesting that a high production budget does not guarantee high box office success.
- We tend to think performance of such movies with low budget success might have been due to other factors such as marketing and the quality of the movie etc.

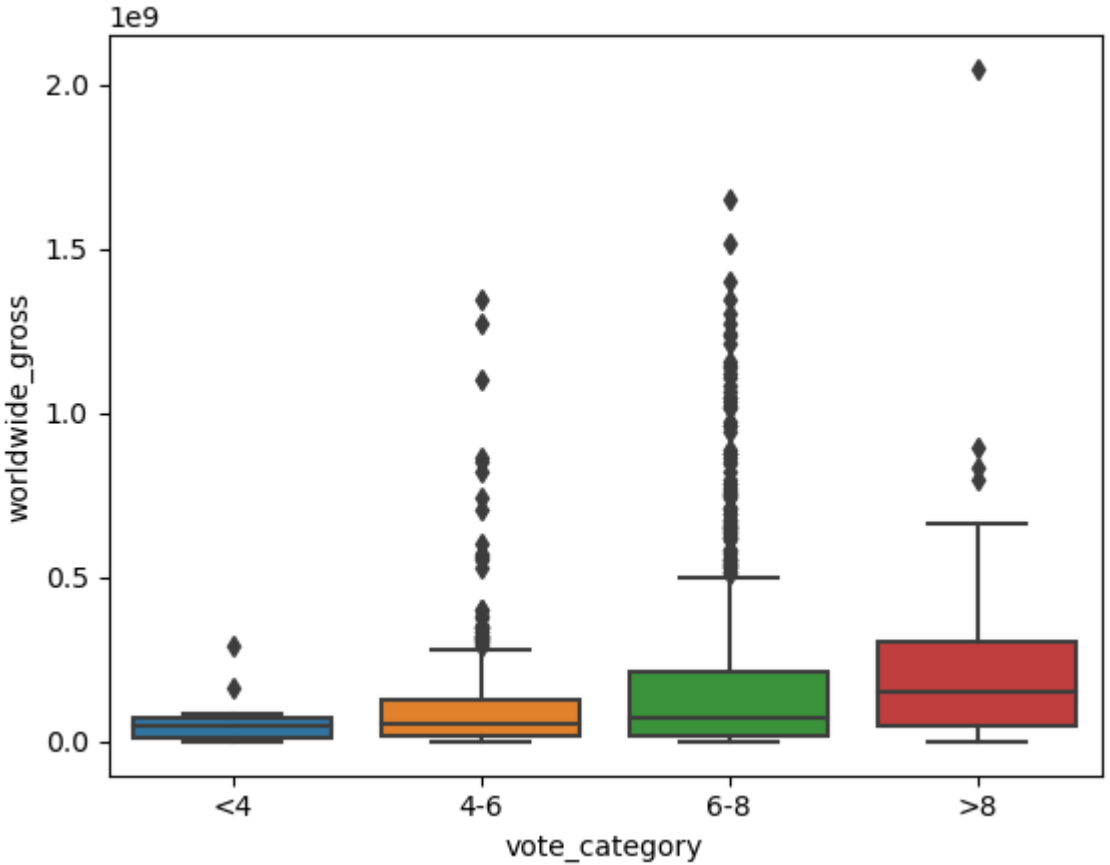


We also note the following:

- English language is the most popular language for the movie however based on ROI ,there are some languages that have a higher ROI.

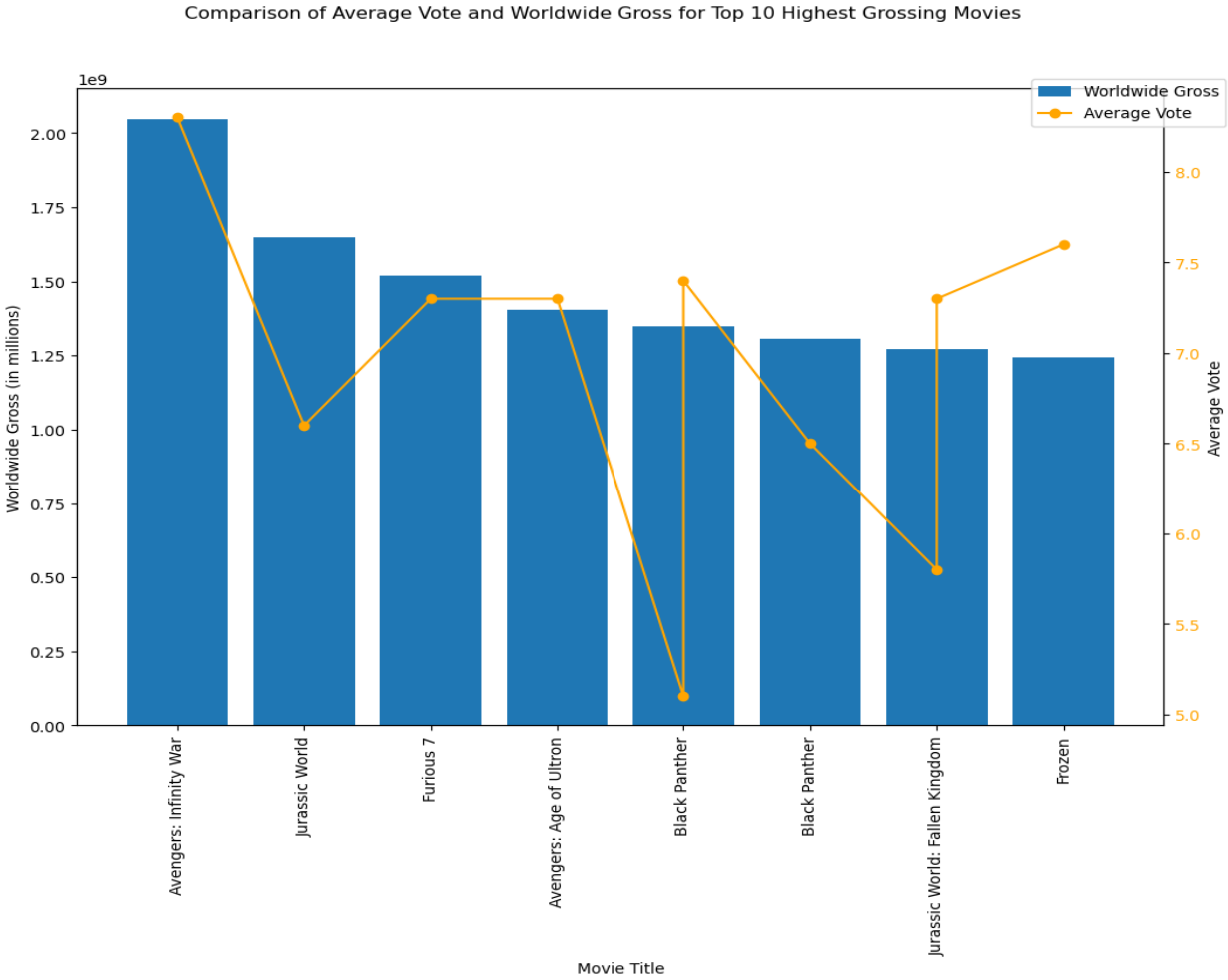
- Telegu is the language with the highest return on investments.

Movie rating vs Worldwide Gross



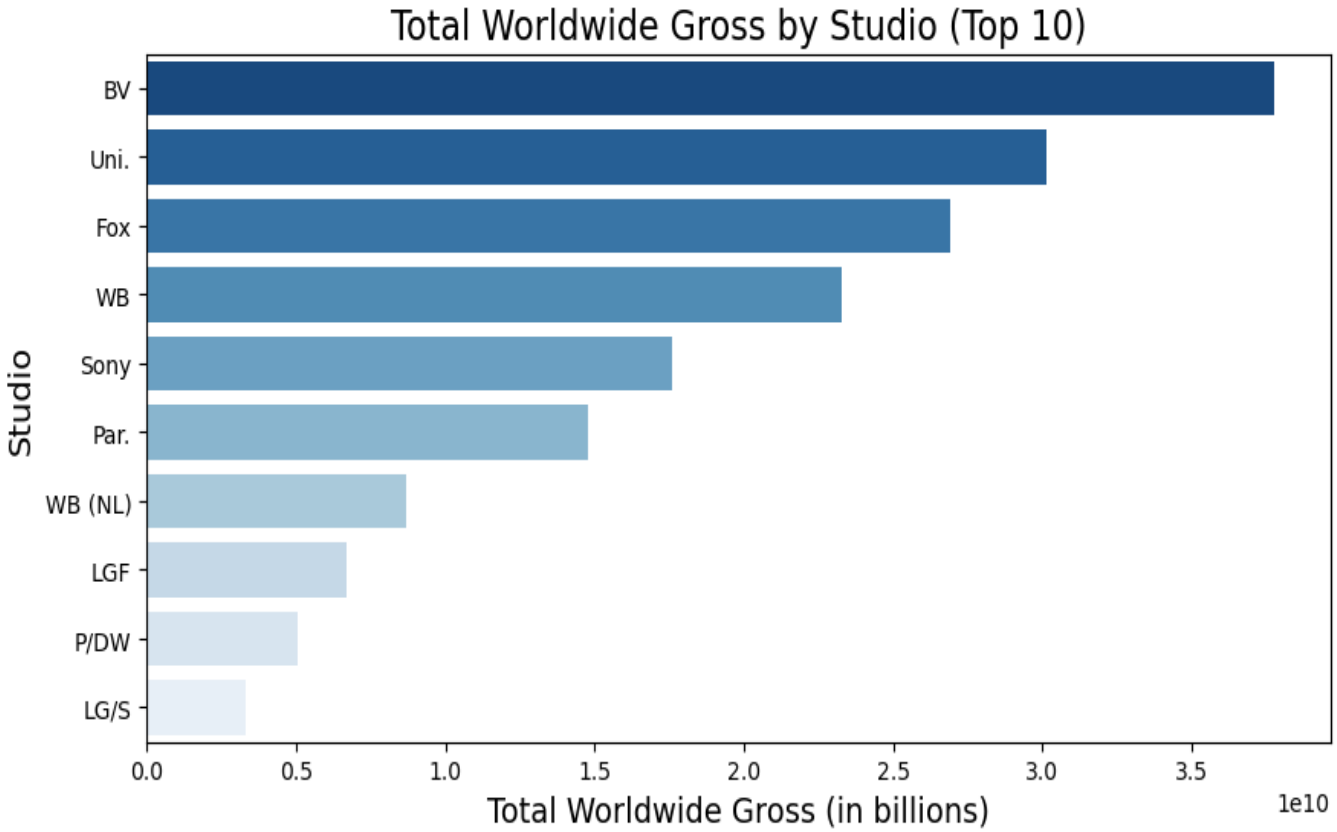
We made the following observations:

- there are a large number of movies in the "6-8" rating category, which suggests that movies with moderate to high ratings are more popular among viewers.
- Conversely, number of movies in the "<4" rating category, it suggests that movies with low ratings are less popular among viewers.
- There are quite a number of outliers in the range 6-8, further analysis need to be done to ascertain why is the case.



From the bar charts, we can derive a few important insights.

- The highest grossing movies are not necessarily the ones with the highest average vote.
- The top 10 highest grossing movies have a relatively high average vote. This suggests that popular movies tend to be well-liked by viewers as well.
- Other movies, such as "Frozen", have relatively lower worldwide gross compared but have a high average vote, indicating that they were popular among viewers despite not being as commercially successful.



The bar chart shows the top 10 studios by worldwide gross revenue

- BV and Uni tops the list

Key findings

Original Language:

- English is the most popular language for movie production.
- Movies produced in English tend to have higher popularity scores compared to other languages.
- However, movies produced in Spanish tend to have higher average ROI compared to movies produced in other languages.

Return on Investment (ROI):

- There is a weak positive correlation between a movie's budget and its worldwide gross.
- Movies with budgets between 20 *million and* 100 million tend to have the highest ROI.
- Movies with higher vote averages tend to have higher ROI.

Vote Average:

- There is a weak positive correlation between a movie's vote average and its worldwide gross.
- Movies with higher vote averages tend to have higher worldwide gross earnings.
- However, the correlation between vote average and worldwide gross is weaker for movies with budgets greater than $100 million.

Key Recommendations

Having analyzed the data carefully, we do hereby make the following recommendations based on the problem statement.

- Microsoft to consider producing movies in English language since it's the most popular language but also to translate to other languages like Telegu, as they tend to have higher average ROI. This could be a good strategy to diversify their portfolio and reach a wider audience.
- When considering a movie’s production budget, it is important to note that there is a weak positive correlation between budget and worldwide gross. Therefore, it is important to balance the budget with other factors such as the movie's potential appeal to audiences and critical reception.
- When deciding which movies to produce, Microsoft should pay attention to a movie's vote average as it is positively correlated with worldwide gross. However, they should also be cautious about investing in movies with extremely high budgets, as the correlation between vote average and worldwide gross is weaker for these movies.

Thank You!

Email: mucuibernard2015 @gmail.com
GitHub: @MucuiBen
LinkedIn: https://www.linkedin.com/in/bernard-mucui-961b1481/