

Overview

Background of the business

The movie industry is a large and diverse sector of the entertainment sector that creates, markets, and presents movies. It has captivated audiences of all ages and backgrounds for more than a century, making it one of the most significant and profitable sectors in the world. Key players in movie industry include Genre, Actors, Producers etc.

Microsoft is a technology firm with a wide range of product lines and industry verticals. Microsoft's commercial domains can be generally divided into numerous categories: Operating Systems, Office Productivity Suite, Cloud Services, Developer Tools, Security: Microsoft provides a variety of cybersecurity solutions,

Business Understanding

Microsoft wants to join the fun as they see so many large firms producing original video content. They've chosen to start a new movie company, but they have no experience making movies. I am expected to research the genres of movies that are currently performing the best at the box office. The next step is to translate your research into practical insights that the head of Microsoft's new film division can use to guide the production of films.

Objectives.

- 1) Are there specific months or seasons when movies tend to perform better?
- 2) Are there any noticeable trends in movie popularity?
- 3) Is higher production budget correlated with higher revenue?

I have analysed the attributes that Microsoft needs to have in order to succeed in the industry. I have based my analysis on.

- 1) The month the movies were released, to get the best time in a year to release a movie.
- 2) The highest rated movies and popularity
- 3) The production budgets.

Data Understanding

I used two datasets for my analysis. The data set I used are tmdb.movies.csv and tn.movies_budgets.csv.

The `tmdb.movies.csv` has 26517 rows and 10 columns. With variables describing genre, popularity, vote average & vote count, and release date, the Movie Database (TMDB) dataset contains over 26,000 films that were published between 1930 and 2020. Data preparation was done by:

- 1) Checking missing values and duplicates, none were present.
- 2) Checking for the constituency of the dataset by checking for outliers
- 3) If the data format is correct. I found outliers in `vote_average` column.

The `tn.movies_budget.csv` has 5782 rows and 6 columns. Includes data on 3,400 films that were released between 2010 and 2018 and includes studio, domestic, and international gross revenue characteristics. `df_movie_budgets` provide data on 5,800 films that were released between 1915 and 2020 and includes fields for production budget, release date, domestic gross revenues, and global gross revenues. Data preparations was done by:

- 1) Checking missing values and duplicates, none were present in the dataset.
- 2) The data format in the columns that had numbers was incorrect and it had to be corrected.
- 3) Columns that were unnecessary were dropped.

Data Analysis

The year with the most movie releases is 2015, and the month with the most releases is January, according to the "`tmdb.movies.csv`" dataset. It implies that 2015 experienced a lot of new movie releases, with January being especially active.

There is a correlation between an increase in a movie's popularity and an increase in its profit margin, according to the "`tn.movies_budget.csv`" dataset. This suggests that a movie tends to make greater margins of profit as it increases in popularity. Medium budgets tend to have lower popularity scores while high and very high budget films have greater popularity. Several variables, such as production value, marketing, genre, and critical acclaim, have an impact on the relationship between film budgets, popularity ratings, and ROI. The observations made emphasizes the variety of the movie industry, where various budget ranges can produce a range of outcomes in terms of appeal and revenue. Because of this intricacy, investing in film production requires serious planning and decision-making.

The correlation between Return on Investment between popularity is linear, meaning that as Return on Investment increases, popularity tends to increase as well. In other words, movies with higher Return on Investment are generally more popular.

Recommendations

The year 2015 had the most released movies and January had the highest number of movie releases. This observation might indicate a particular trend in the movie industry for a specific time period.

It is interesting to note that in the analysis, high budget movies are the most successful in terms of popularity. This finding suggests a positive correlation between the budget size of a film and its popularity, meaning that as the budget increases the likelihood of a movie being popular among audiences also increases because of high production quality, marketing and promotions.

Next Step

The average vote for a movie appears to be evenly distributed across budget sizes, even though popularity tends to rise with budget size, indicating that other factors—like movie quality, for instance—that were not covered in this research affect the voted rating. I want to investigate how, among other things, the average vote is impacted by rating, runtime, and directors.

I'd like to learn more about how genre success has changed over time, specifically whether some genres have had faster growth than others. Also, what genres will be successful in the future.

I'd also like to see more detailed geographic revenue data. Are there some areas of the world that have produced more revenues than others?