

Box Office Success Factors Analysis

Joby Varghese

May 13, 2023

Summary

Analysis of certain key datasets from the IMDB website reveals the following observations that Microsoft can take into consideration:

1. 'Average rating of a movie' and 'runtime of a movie' exhibit a weak positive correlation.
2. In general, the genres of Documentary, News, and Biography are rated the highest, while Horror, Sci-Fi, and Thriller are rated the lowest.
3. In terms of profits, BV, Fox, and Uni rank in the top three, while ORF, W/Dim, and STX rank among the bottom three.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

1. The lack of experience that Microsoft has in the movie industry and its eagerness to get into the film production industry.
2. Uncertainty regarding the type of movies that will be successful in the current market.
3. It is necessary to develop content that will appeal to audiences and generate revenue.
4. The risk of investing a significant amount of money in a project that may or may not perform well at the box office.
5. Developing a strong brand identity in the movie industry is essential for competing with established studios.

Data

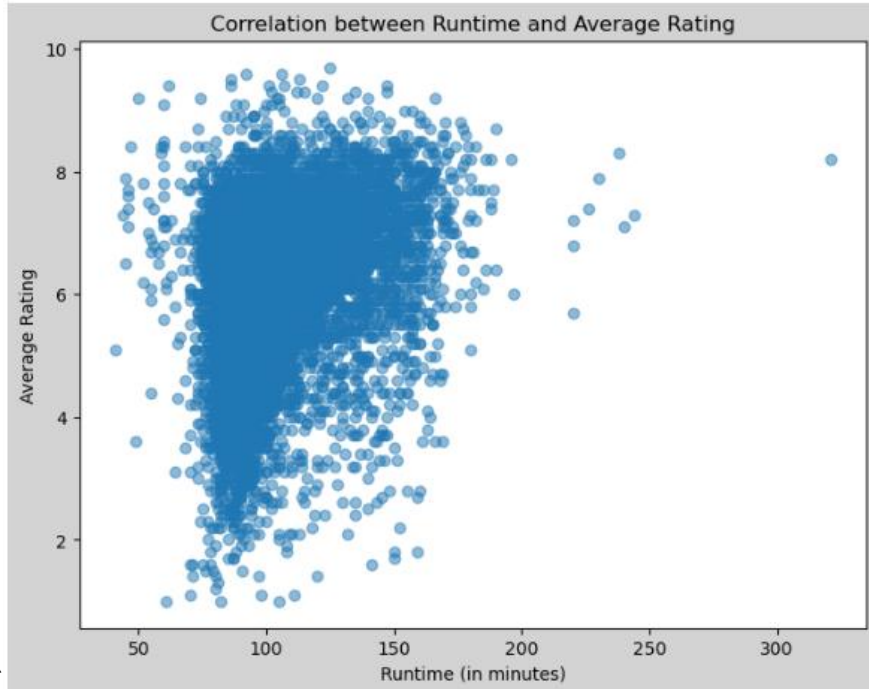
Datasets from the IMDB website were considered for analysis:

1. IMDB.title.basics - Information regarding movie titles, year of release, runtime minutes, and genres. (For the period 2010-2023)
2. IMDB.title.ratings - Information regarding a particular movie's number of votes and average rating.
3. BOM.movie_gross - Information about the studios that produced the films and their revenue. (For the period 2010-2018)
4. TN.movie_budgets - Data pertaining to the budget and revenue of each movie. (For the period 1915 to 2019)

Methods

1. Datasets 'imdb.title.basics', 'imdb.title.ratings', 'bom.movie_gross' & 'tn.movie_budgets' have been cleaned and assigned the names df1, df2, df3 & df4.
2. The data was preprocessed by identifying and replacing missing values with central tendency values, dropping rows with missing values, and converting column data types (e.g., object to datetime, float, or int).
3. A new dataframe 'masterdata1' was created by merging dataframes df1 and df2 using the left join method.
4. The dataframes df1 and df3 were merged using a left join and a new dataframe 'masterdata2' was created for further analysis.
5. Using left join, df3 and df4 were merged and a new dataframe 'masterdata3' was created for further analysis.

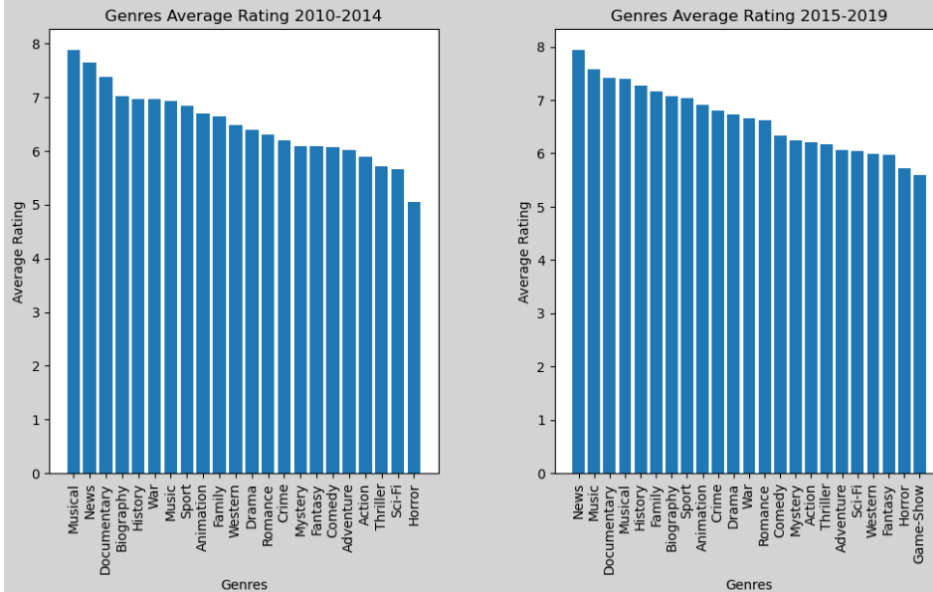
Results



Correlation between Runtime & Average Rating

- Correlation Coefficient of 0.317
- Infer a weak positive relationship.
- Slight increase in the average rating when the movie runtime is increased.
- Relationship is not strong.

Results

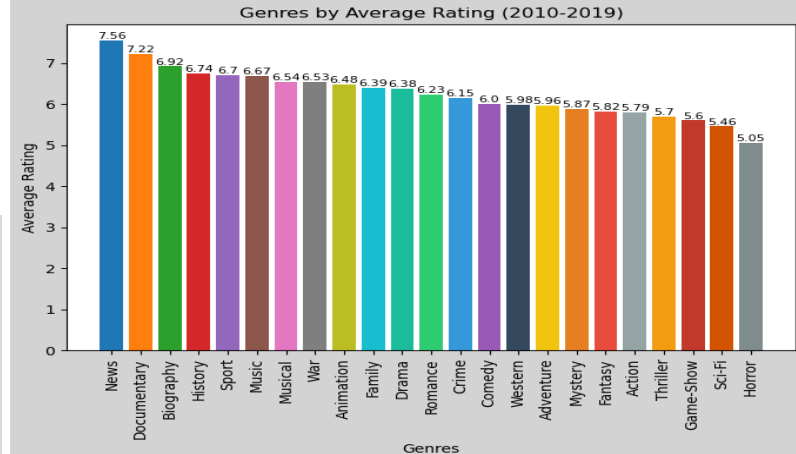


Genres – Average Rating 2010-2014 & 2015-2019

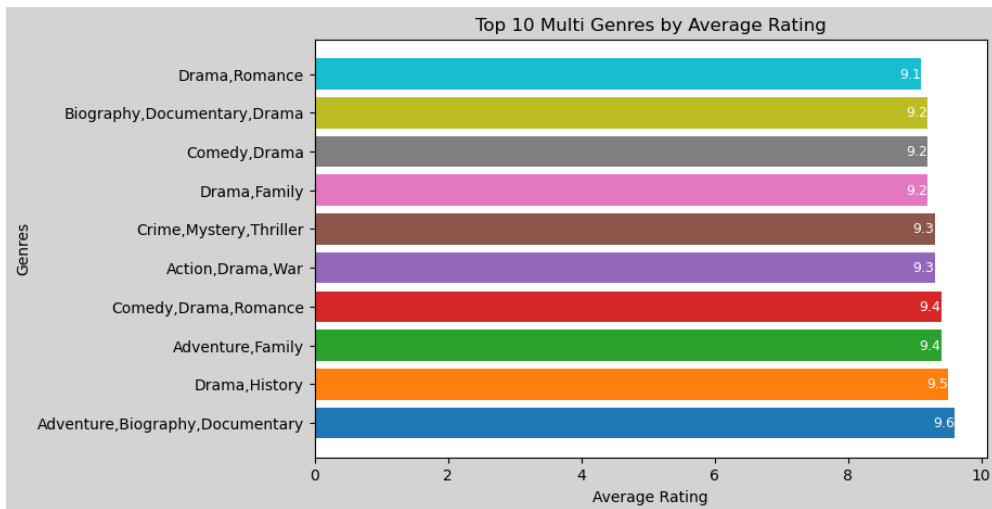
- It appears from both graphs that the genres listed here have maintained their average ratings within a range of 5 to 8.
- The viewership for these genres has not changed significantly for these two groups of years.

Top Rated Individual Genres

News, Documentary, Biography, History, Sport, Music, Musical, War, Animation, Family, Drama, Romance, Crime, Comedy, Western, Adventure, Mystery, Fantasy, Action, Thriller, Gameshow, Sci-Fi, Horror



Results



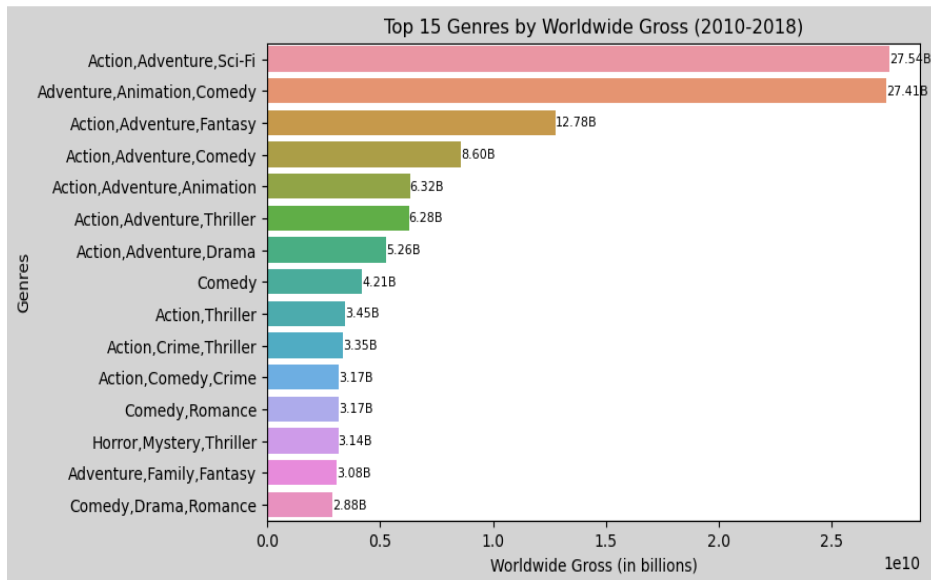
Movies with multiple genres – Average Rating 2014-2019

- Top-rated genres are rated between 9.1 and 9.6, which indicates they are highly regarded.
- These statistics indicate that the viewers' tastes and preferences encompass a wide range of highly acclaimed movies in diverse genres.

Top Rated Movies with Multiple Genres

Adventure Biography Documentary, Drama History, Adventure Family, Comedy Drama Romance, Action Drama War, Crime Mystery Thriller, Drama Family, Comedy Drama, Biography Documentary Drama, Drama Romance

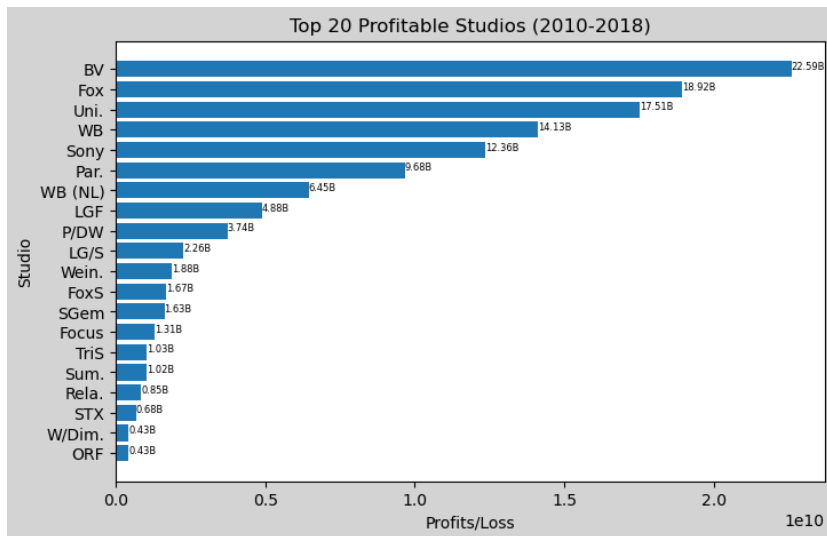
Results



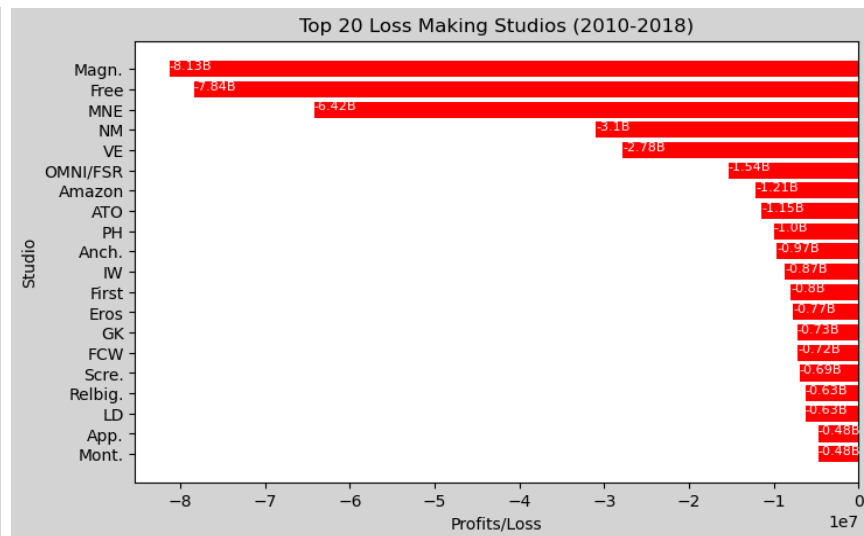
Movies with multiple genres – Average Rating 2014-2019

- Among the genres with the highest grossing films are action, adventure, and comedy.
- Action-Adventure-Sci-Fi is the highest grossing movie genre with a worldwide gross of 27.5 billion dollars.
- Outside of action and adventure, comedy is the only genre to make it to the top ten grossing genres.

Results



- With 22.59 billion dollars in profits, BV (Buena Vista) has the highest profits among all studios.
- ORF (Open Road Films) is at the twentieth spot with 426.03 million USD.
- A significant difference in profitability exists between the top 5 studios and the bottom 5 studios.



- The largest loss was experienced by Magnolia Pictures, which was approximately -81 million dollars.
- Several other studios, including Free, MNE, and NM, also suffered significant losses.
- Amazon and ATO are the only two major studios on this list to report losses, which could suggest that they are struggling in the movie industry in comparison to their other activities.

Conclusions

1. Movies in top-rated genres should be considered by Microsoft.
2. Focus on action-adventure-sci-fi and comedy genres.
3. Invest in marketing and distribution to achieve profitability similar to top studios.
4. Recognize the limitations of the available data and conduct further research.
5. A greater amount of data and the maintenance of its sanctity will improve the accuracy and completeness of the analysis.

Thank You!

Email: grace.joby@gmail.com

GitHub: @VargheseJoby

LinkedIn: [linkedin.com/in/joby-varghese-516](https://www.linkedin.com/in/joby-varghese-516)