Microsoft Movie Studio Project

BY TITUS MUTUKU

OVERVIEW

- Microsoft, is a multinational corporation that was founded in 1975 and serves as the world's leading software developer.

- As an ambitious entity that seeks to offer solutions to global challenges, the company has identified a gap in movie production and it seeks to fill the gap by opening a subsidiary movie studio.

- It is worth highlighting that Microsoft's business domain is broad and encompasses a wide range of products, services, and initiatives in the technology industry hence venturing into movie production is an added avenue for business.

- It is primarily known for its software products, cloud services, and hardware devices hence delving into the film industry will not only promote diversification but also promote economic growth while creating employment opportunities.

- Most importantly, the new venture will generate income for the entity thus improving its performance and while surpassing the set financial goals.

Business Case

❑ Creating a successful movie studio in today's competitive film industry requires a deep understanding of current market trends and audience preferences.

❑ One of the most ideal actionable insight based on the types of films that have been performing well at the box office is ensuring that directors consider the global market when creating films.

❑ Having highlighted that, it is imperious to acknowledge that the global entertainment industry is experiencing robust growth, with the demand for original content reaching unprecedented levels.

❑ Microsoft aims to leverage its brand, resources, and technology to enter this lucrative sector by establishing Microsoft Movie Studios.

Business Problem

- Microsoft Movie Studios needs to determine which movie genres and runtime durations are most likely to attract a large and engaged audience.

- Ensuring the production of high-quality films is essential for maintaining audience satisfaction and reputation.

The General Business Understanding

- The aim of this project is to create a thriving movie production entity that produces high-quality, diverse content for various distribution channels, including theaters, streaming platforms, and television networks.

- This project is pivotal in leveraging data-driven insights to optimize various aspects of movie production, distribution, and audience engagement.

- This is because the global entertainment industry is valued at $2 trillion and continues to expand (Guttmann, 2023). Streaming platforms, international markets, and diverse content preferences offer substantial growth prospects.

- The Microsoft brand, and technological prowess of the company can create a competitive advantage that can set the new movie studio venture apart in the industry.

- Additionally, integration with Microsoft's existing ecosystem can enhance user engagement and content delivery. The following objectives will play a role in shaping the new investment;

  1. To determine the most popular genres among viewers, aiding in strategic genre selection for future movie projects.

  2. To identify directors whose movies consistently receive a high number of votes and positive ratings.

  3. To explore viewer preferences regarding movie runtime and identify the ideal runtime for maximizing engagement.

  4. To uncover which director-genre combinations yield the most favorable results in terms of audience engagement.

  5. To investigate the correlation between movie runtime and genre preferences among viewers.

  6. To develop a quality assessment metric considering both user ratings and votes, facilitating content quality evaluation.

Data Understanding
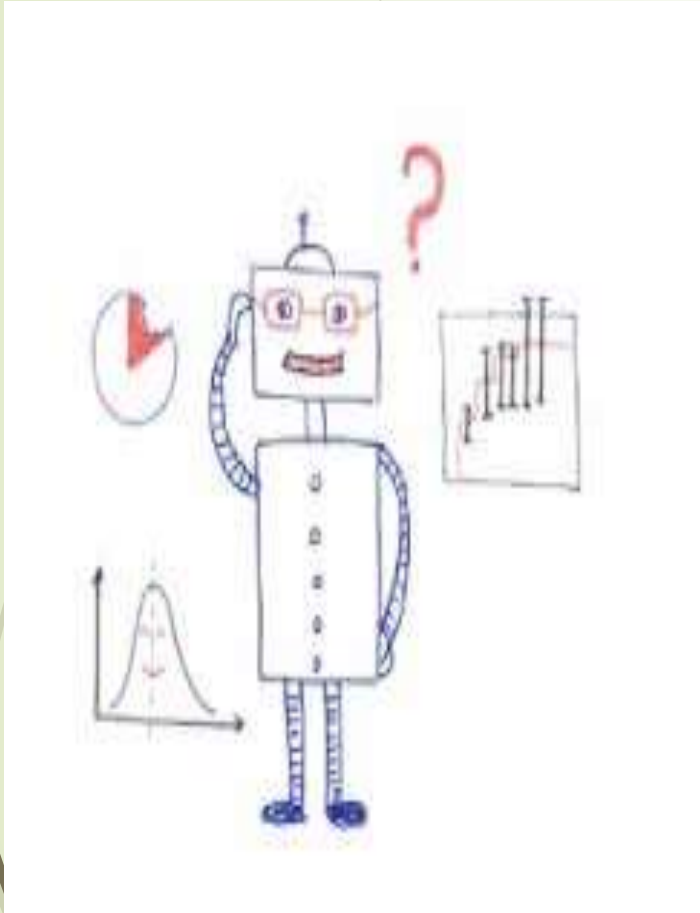
- Because of the wealth of data available that can help understand the film production industry, selecting the most ideal dataset to work with was important.

- The process of data cleaning began with the exploration of the available datasets.

- After an in-depth comparison, I decided to work with the 'im.db' database. I will be analyzing a database that contains several tables.

- The most important tables in the database will be 'Movie basics' and 'Movie ratings'.

- Each table has various fields which anchor critical information necessary for analysis and later conduct visualization that will play a critical role in drawing my conclusions, insights, and recommendations necessary for Microsoft movie studios.

- The fields for the 'Movie basics' table include; Movie_id, Primary title, Original title, Start_year, Runntime_minutes, and most importantly, the 'Genres' field.

- The second table (Movie_ratings) contains three major fields; the Movie _id column, Average rating column, and the Number_of_votes column.

- Here, I was able to gain a deep understanding of the dataset's characteristics. I completed the data cleaning process where I addressed the issue of missing values, dropped unnecessary columns, edited the field titles, and shaped the tables in a more attractive manner for analysis.

Data Analysis

Figure 1: Top 40 Movies by Average Rating and genres
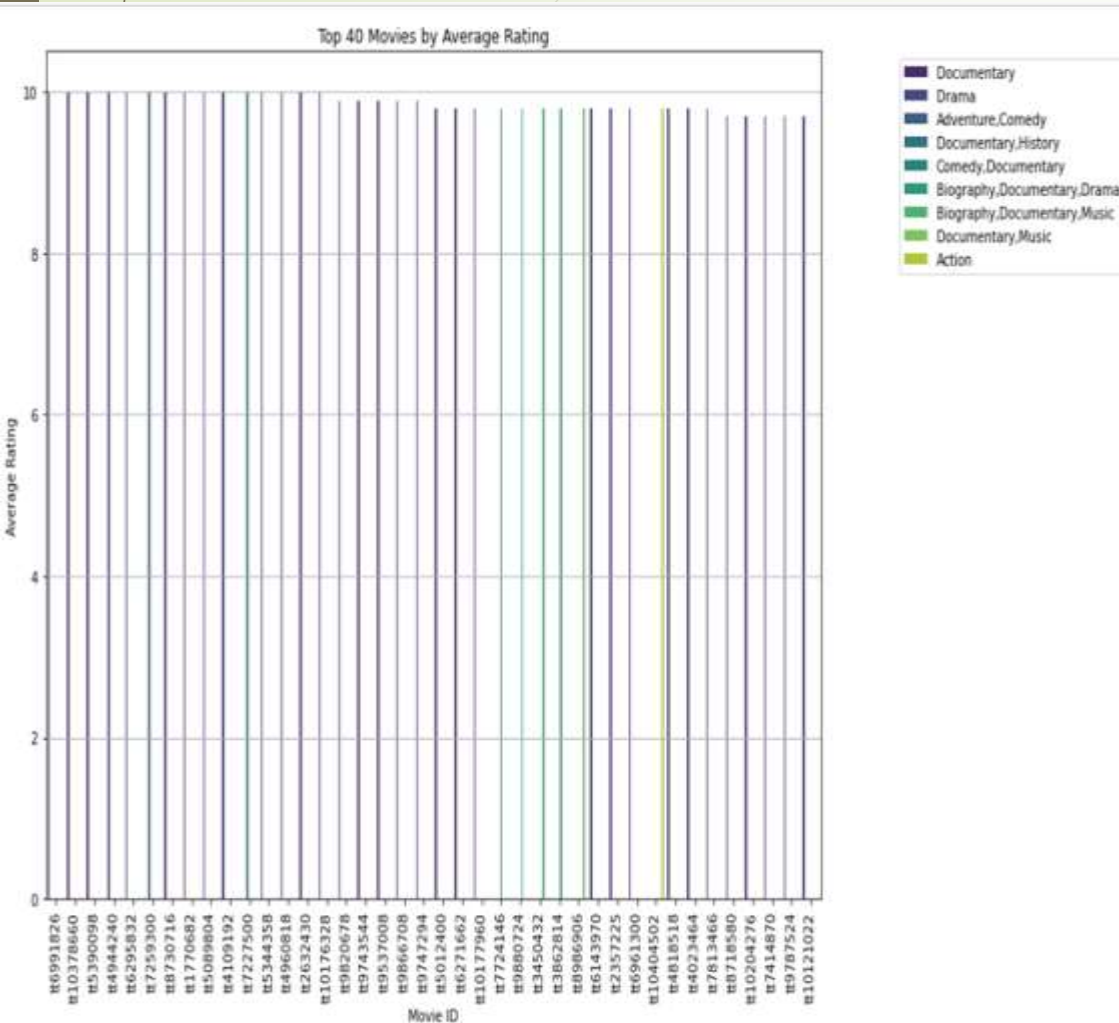
- After settling on the 'im.db', the first step was to access the data in a pandas DataFrame format where I generated the 'Movie_basics' and the 'Movie_ratings' tables.

- This was followed by calculating the value count to check the categorical variable distribution of various genres in the 'Genre' field.

- After cleaning the tables, I joined them and the process of data visualization kicked in. The first graph was generated using the seaborn module where I created a vertical bar graph.

- Because of the length of the data in each field, a sample population of 40 variables was selected based on average ratings and genres.

- The goal was to select the top 40 movies in the merged DataFrame. From the graph, all the top for movies and their corresponding genres have a rating of more than 9 in a scale of 10. The most preferred genres in the order of preference include;

  - Documentary

  - Drama

  - Adventure, Comedy

  - Documentary, History

  - Comedy, Documentary

  - Biography, Documentary Drama

  - Biography, Music

  - Action

Average Ratings

A Histogram Showcasing the Symmetry of Average Ratings

Distribution of Average Ratings
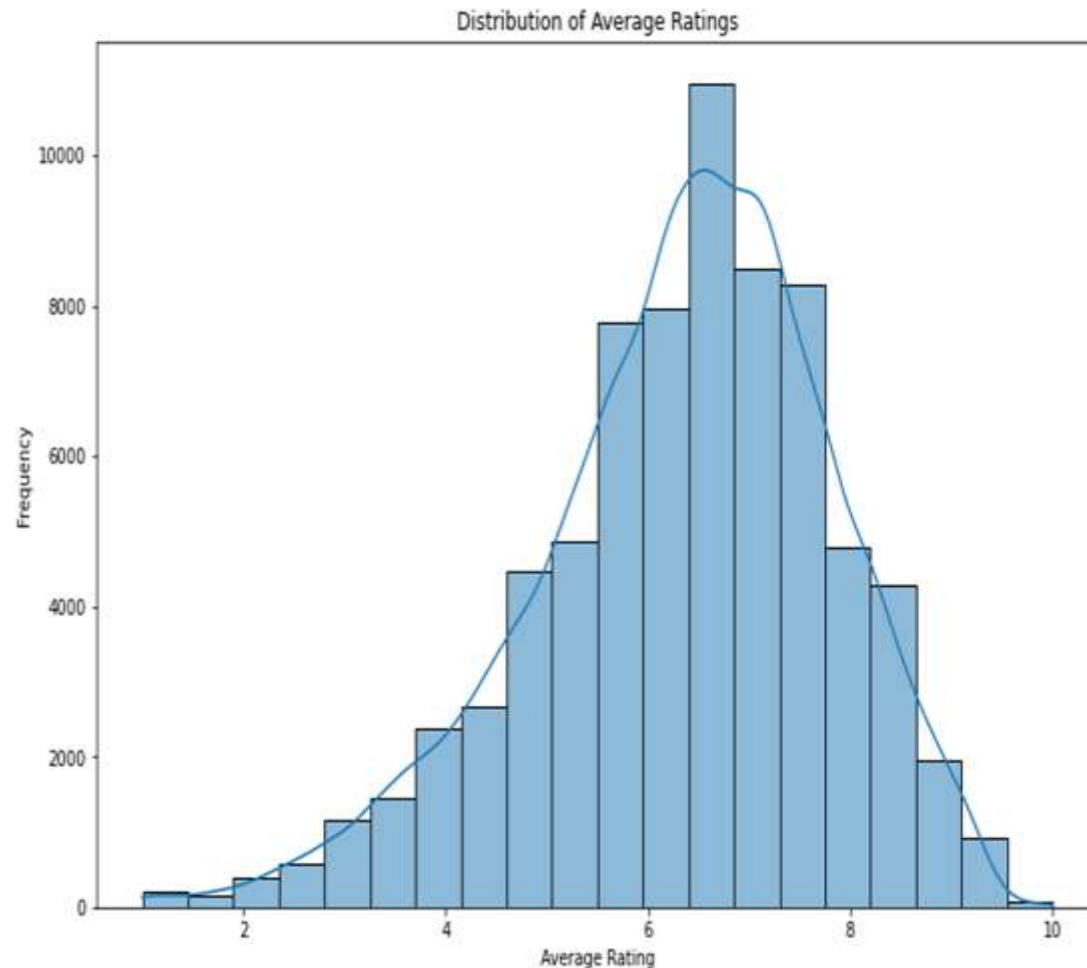
- The second visualization generated a histogram of average ratings whereby the analysis sought to establish how skewed the ratings were for the purpose of confirming how biased the audience could have been in order to make sound decisions as far as the new movie studio is concerned.

- The symmetrical distribution typically suggests that the ratings for the movies are evenly distributed around a central value.

- Therefore, the studio should use this information as a starting point for further analysis and decision-making and for understanding the nature of the target audience.

# Establishing the correlation Between Average ratings and Number of votes

A scatter plot of average ratings vs. number of votes

Average Ratings vs. Number of Votes
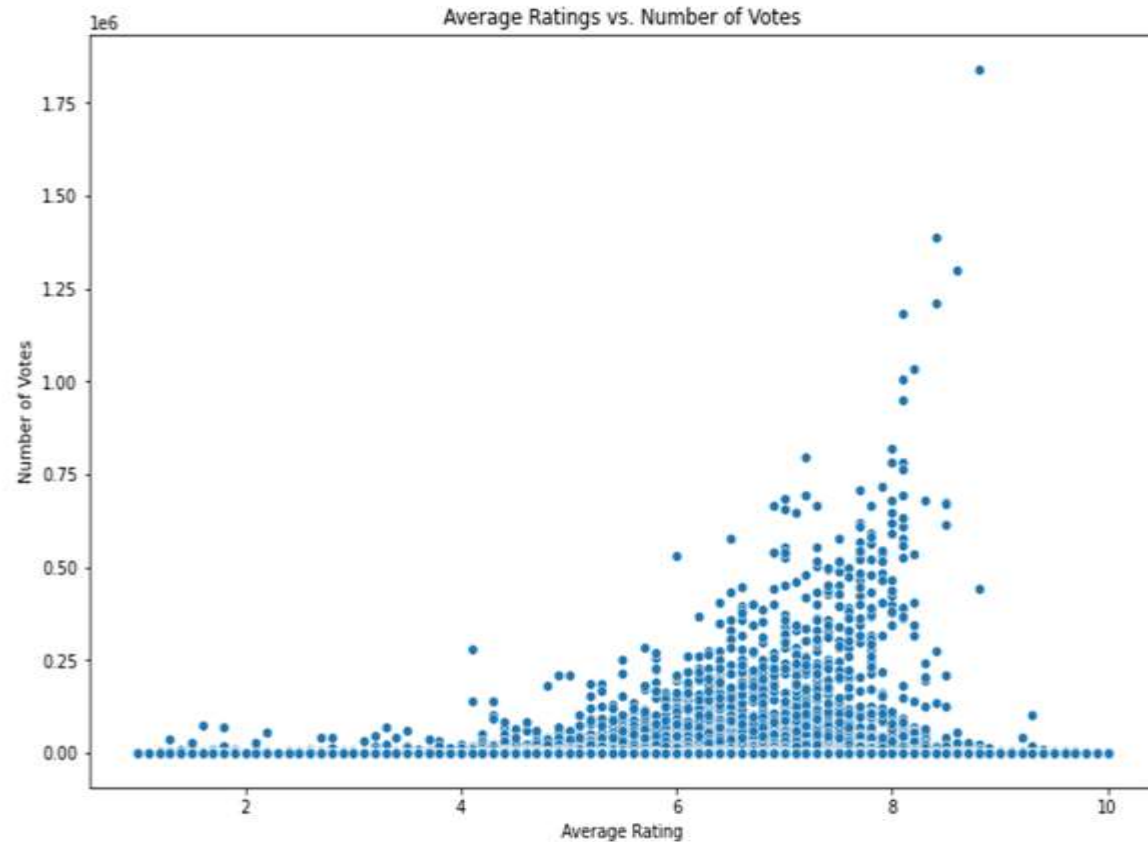
- The third visualization sought to establish whether there is correlation between the average rating and the number of votes for each movie on the basis of their respective genres.

- Analyzing the relationship between average

- e ratings and the number of votes provided valuable insights into audience engagement and the impact of audience size on perceived content quality.

- From the scatter graph, it is evident that the top rated movies and their respective genres tend to receive the most votes.

- It is therefore advisable for the manager of the new Microsoft movie studio to put more focus on the highly regarded genres when planning for production to attract positive responses and establish the studio is the go-to studio for contemporary movie consumers.

A Countplot of Movie Production by Year

A countplot of movie production by year

Yearly Movie Production Trends
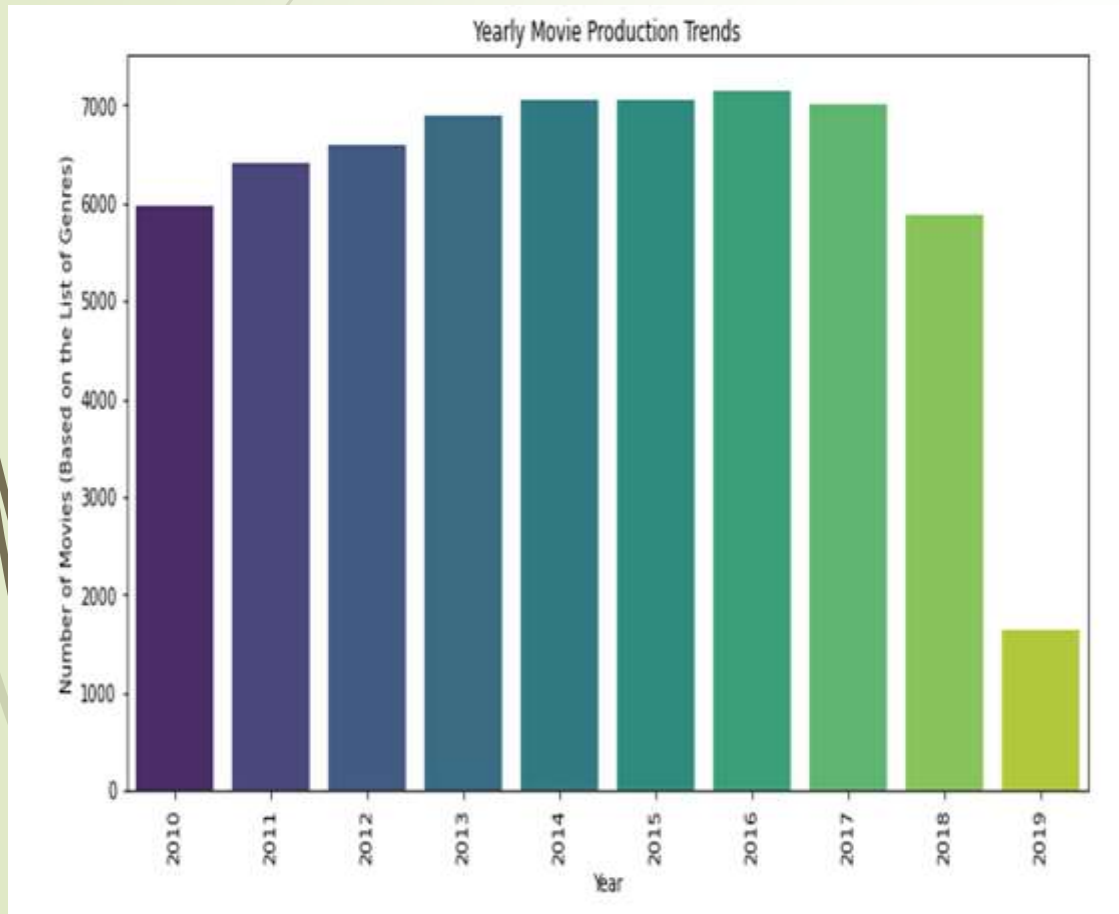
- Finally, the analysis sought to establish the trend of movie production per year.

- To establish this line of thought, a count plot of movie production by year was created. This visualization established the number of movies produced each year.

- It was established that movie production was on the rise up until 2019 whereby the industry witnessed a sudden decline.

- For justifiable conclusion, there needs to be done for the year 2020, 2021, and 2022 to determine whether the decline was subjective.

- The countplot allows the studio to visualize how the number of movies produced has evolved over time.

- This provides insights into trends and patterns in the studio's production history.

- This understanding can inform strategic decisions related to resource allocation, content strategy, and audience engagement, ultimately helping Microsoft Movie Studios succeed in the film industry.

- The decline signifies the need for huge resource allocation to outdo competitors and the probable decline.

CONCLUSION

- The project confirms that there is a strong correlation between genres, user ratings, and number of votes.

- The three can be effectively used to identify genres associated with high-quality content.

- By developing a content quality assessment metric, Microsoft Movie Studio can make data-driven decisions to improve the quality of its productions.

RECOMMENDATIONS & NEXT STEPS

Recommendations

- Put more emphasis on the highly regarded genres when planning for production to attract positive responses and establish the studio is the go-to studio for contemporary movie consumers.

- By observing production trends, the studio can make informed decisions regarding resource allocation for future projects. For example, during periods of increased production, more resources may be needed for casting, marketing, and distribution.

- Invest in technology to engage the target audience continuously

Next Steps

- Consider more analysis to establish the poorly performing genres and establish the reason to their poor performance.

- Establish the role of technology in improving the performance of contemporary movie studios.

- Predict the undesirable trends in movie production for the contemporary audience.

Thank You!

- Email: tituskilonzomutuku@gmail.com

- Github : TitusKilonzoMutuku