MICROSOFT MOVIE STUDIO ANALYSIS: INSIGHTS AND RECOMMENDATIONS

NAME: ANNAH MUKETHE

PRESENTATION OBJECTIVES

PROJECT INTRODUCTION

- Overview
- Business Understanding
- Methodology

DATA UNDERSTANDING

- Data Understanding
- Data preprocessing
- Data Analysis
- Data Visualization

SUMMARY

- Recommendations
- Next Steps
- Q&A

PROJECT INTRODUCTION

For this project, we aim to use exploratory data analysis to generate insights for a business stakeholder.

OVERVIEW:
This project uses exploratory data analysis (EDA) to help Microsoft's new movie studio understand successful movie characteristics. By analyzing a comprehensive dataset, the project aims to provide actionable insights for producing financially successful movies that resonate with audiences.

THE DATA:

The data utilized in this project was sourced from the following link:

- IMDB: https://www.imdb.com/interfaces/

BUSINESS UNDERSTANDING

BUSINESS PROBLEM:

Microsoft is entering the movie industry with a new studio but lacks experience. To succeed, they aim to understand what types of films perform best at the box office. This understanding will guide their film production decisions, ensuring they enter the market competitively.

OBJECTIVES:

- **Analyze Market Trends:** Identify the genres, themes, and characteristics of movies that are currently performing well.

- **Determine Key Success Factors:** Understand the elements that contribute to high performance, such as runtime, release period, and critical reception.

- **Provide Actionable Insights:** Translate the findings into strategic recommendations for the types of films Microsoft should create to maximize their chances of success in the competitive movie industry.

PROJECT METHODOLOGY

1. **Objective Definition:** Clearly outlined the project goals to understand what types of films are performing well at the box office to guide Microsoft's new movie studio.

2. **Data Collection:** Compiled data from IM.db movie databases to ensure comprehensive coverage of movie details.

3. **Data Understanding:** Conducted a preliminary analysis to get an overview of the data, identify key columns, and detect missing values.

4. **Data Preprocessing:** Addressed missing values, removed duplicates, and standardized data formats. Also, created new features like `genre_categories` and extracted useful information such as the release decade.

5. **Exploratory Data Analysis (EDA):**
   - **Descriptive Statistics:** Summarized basic statistics to understand the distribution and central tendencies of the data.
   - **Visualization**: Used various plots (e.g., histograms, bar charts, scatter plots) to visualize data distributions and relationships.
   - **Correlations:** Analyzed correlations between variables to identify significant relationships.

6. **Data Analysis:**
   - **Univariate Analysis:** Investigated the distribution of single variables to understand their characteristics.
   - **Bivariate Analysis:** Explored relationships between two variables to uncover patterns and insights.
   - **Multivariate Analysis:** Conducted complex analyses to understand the interactions between multiple variables and their combined effect on movie performance.

7. **Results Interpretation:** Summarized key insights from the data analysis, highlighting the most influential factors for movie success.

8. **Recommendations:** Provided actionable recommendations based on data-driven insights to help guide Microsoft's new movie studio's film production decisions.

DATA UNDERSTANDING, PREPROCESSING AND VISUALIZATION

DATA INSPECTION AND UNDERSTANDING

Dataset Overview:

Source:

Collected data from movie databases that is the IM.DB database, including tables like movie_basics, directors, known_for, movie_akas, movie_ratings, persons, principals, and writers.

Content:

Data includes details about movies such as titles, release years, genres, ratings, and number of votes.

Data Size:

The dataset has 73,856 rows and 8 columns.

Identified missing values in runtime_minutes (7,620 missing) and genres (804 missing).

DATA PREPROCESSING

Handling Missing Values:

Imputed missing runtime_minutes using the mean value of the column.
Dropped missing values for the genre column.

Feature Engineering:

Created the genre_categories column to group similar genres together for better analysis.

Data Transformation:

Normalized numerical columns like runtime_minutes to ensure consistent scales for analysis.

DATA ANALYSIS

Descriptive Statistics:

- Calculated summary statistics for numerical columns.
- Analyzed the distribution of movies numbers across different release years.

Correlation Analysis:

- Examined correlations between key variables such as average rating, number of votes, and runtime minutes.

Genre Performance:

- Evaluated which genres received the highest average ratings and most votes.

- Assessed the impact of runtime and release decade on movie ratings.

Multivariate Analysis:

- Conducted multivariate analysis to understand how variables like number of votes, runtime minutes, and average rating influence genre categories.

VISUALIZATIONS

Distribution Plots:

- Plotted the distribution of runtime minutes, number of votes and average rating to identify patterns and outliers.
- Plotted a bar plot to show number of movies produced in each year.

Genre Analysis:

- Created bar charts to show the number of movies and average ratings per genre.
- Created scatter plot to show genre categories and number of votes.
- Created a scatter plot to show genre categories and average ratings.
- Used bar plot to show movies with highest votes.

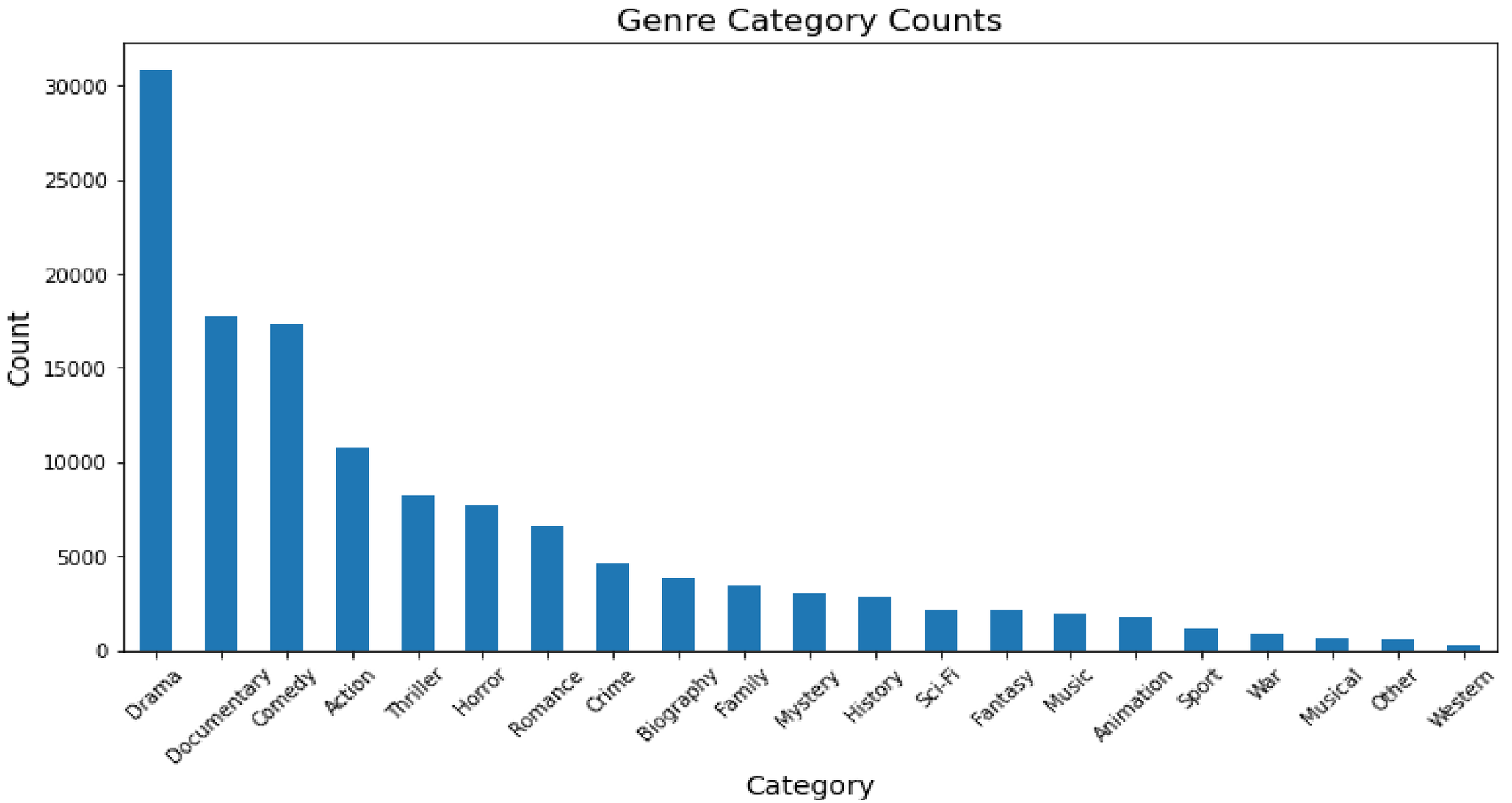
Trend Analysis:

Visualized the trend of average movie ratings over genres categories and number of votes.
Used pair plots to understand the relationship between number of votes, runtime minutes and average rating.

KEY FINDINGS AND INSIGHTS

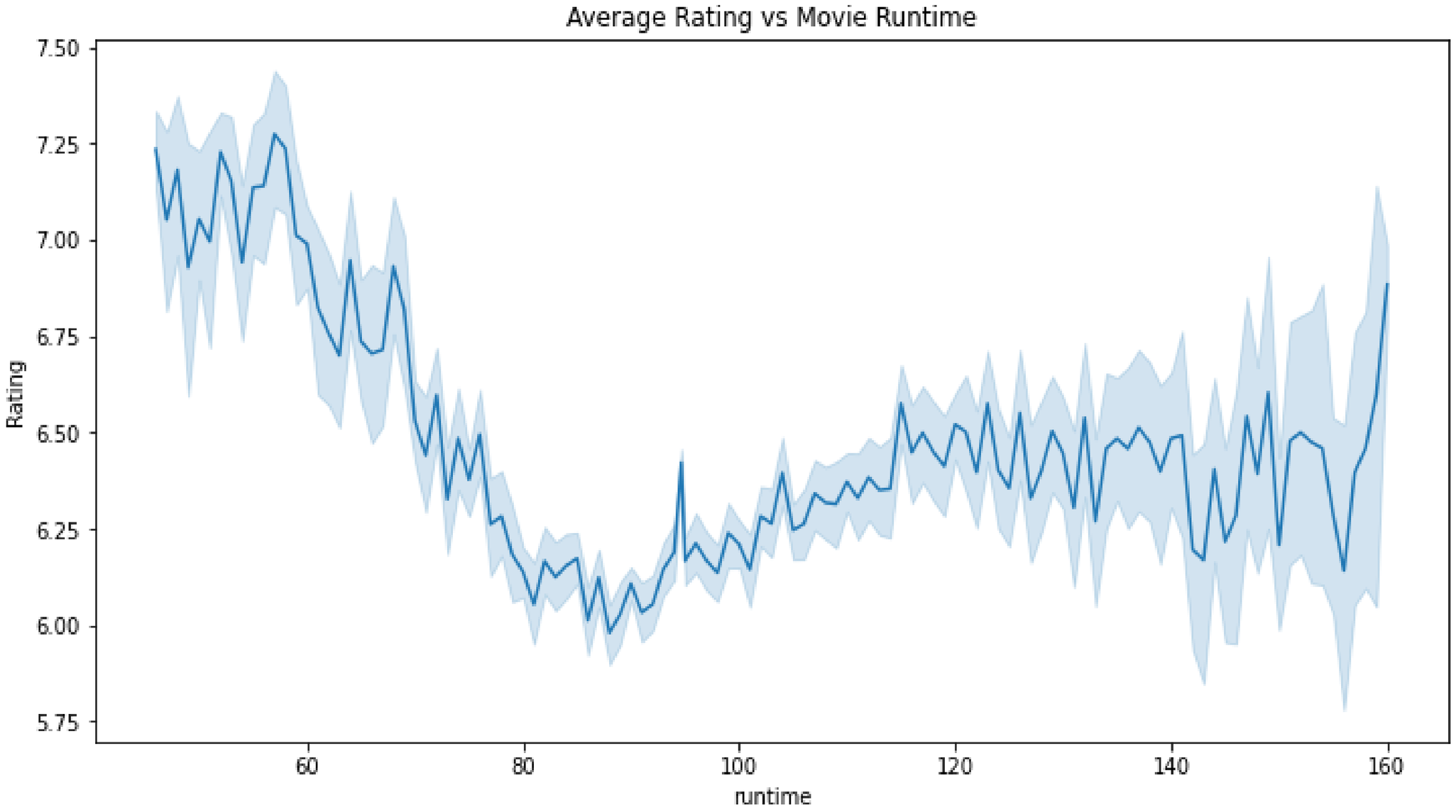
GENRE CATEGORIES

From the analysis, we observed that movies of the drama category were majority than those of western genre.



GENRE CATEGORIES VS NUMBER OF VOTES

From the plot below it is evident that movie runtime contributes to movie ratings. Movies with short runtime minutes have high ratings than those with long runtime.



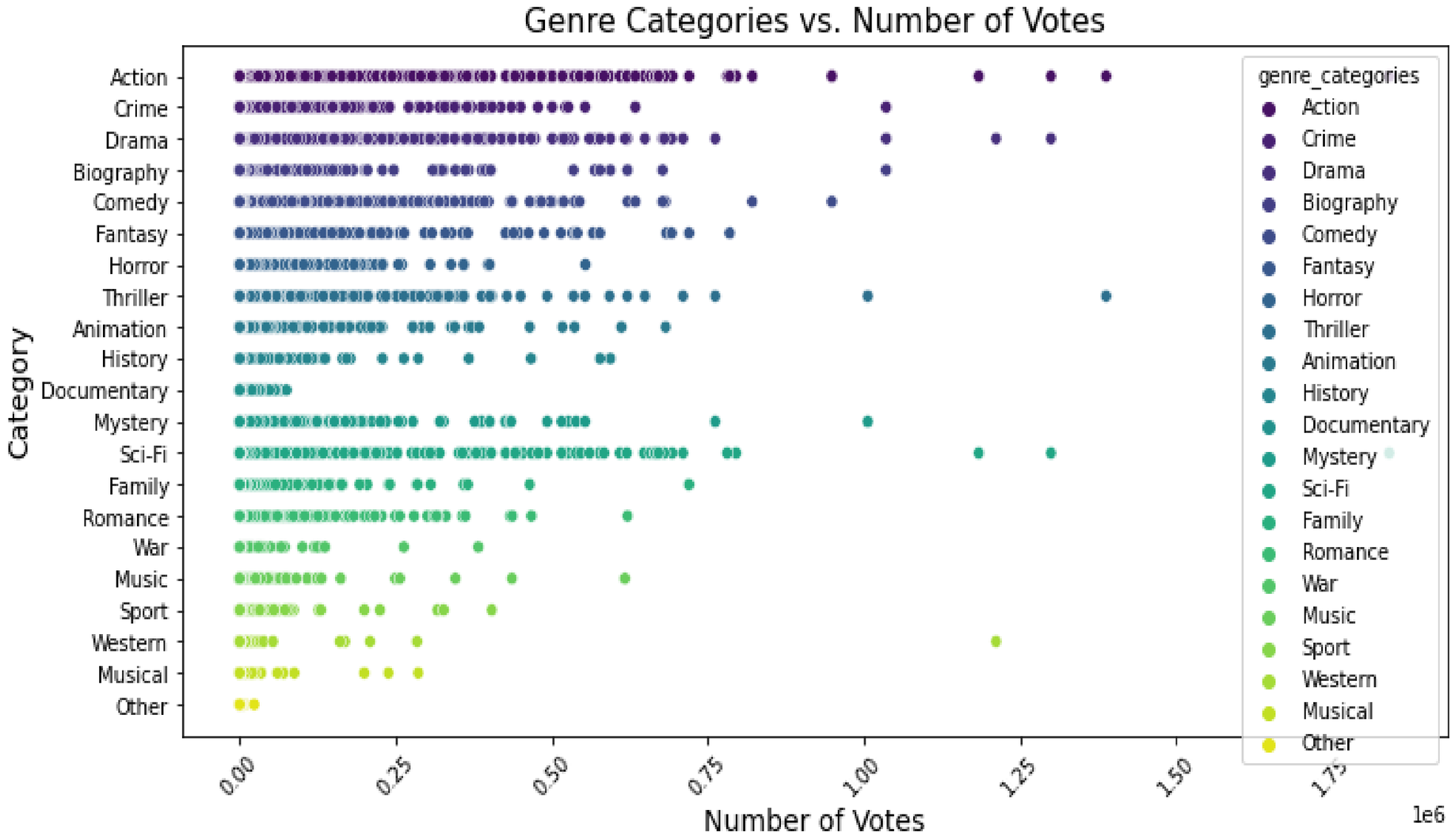
GENRE CATEGORIES VS NUMBER OF VOTES

INSIGHT 1

It is apparent from the plot below, that movies in genres such as action, crime, drama, thriller, sci-fi, and comedy receive a notably high number of votes. These genres appear to consistently attract votes compared to others

INSIGHT 2

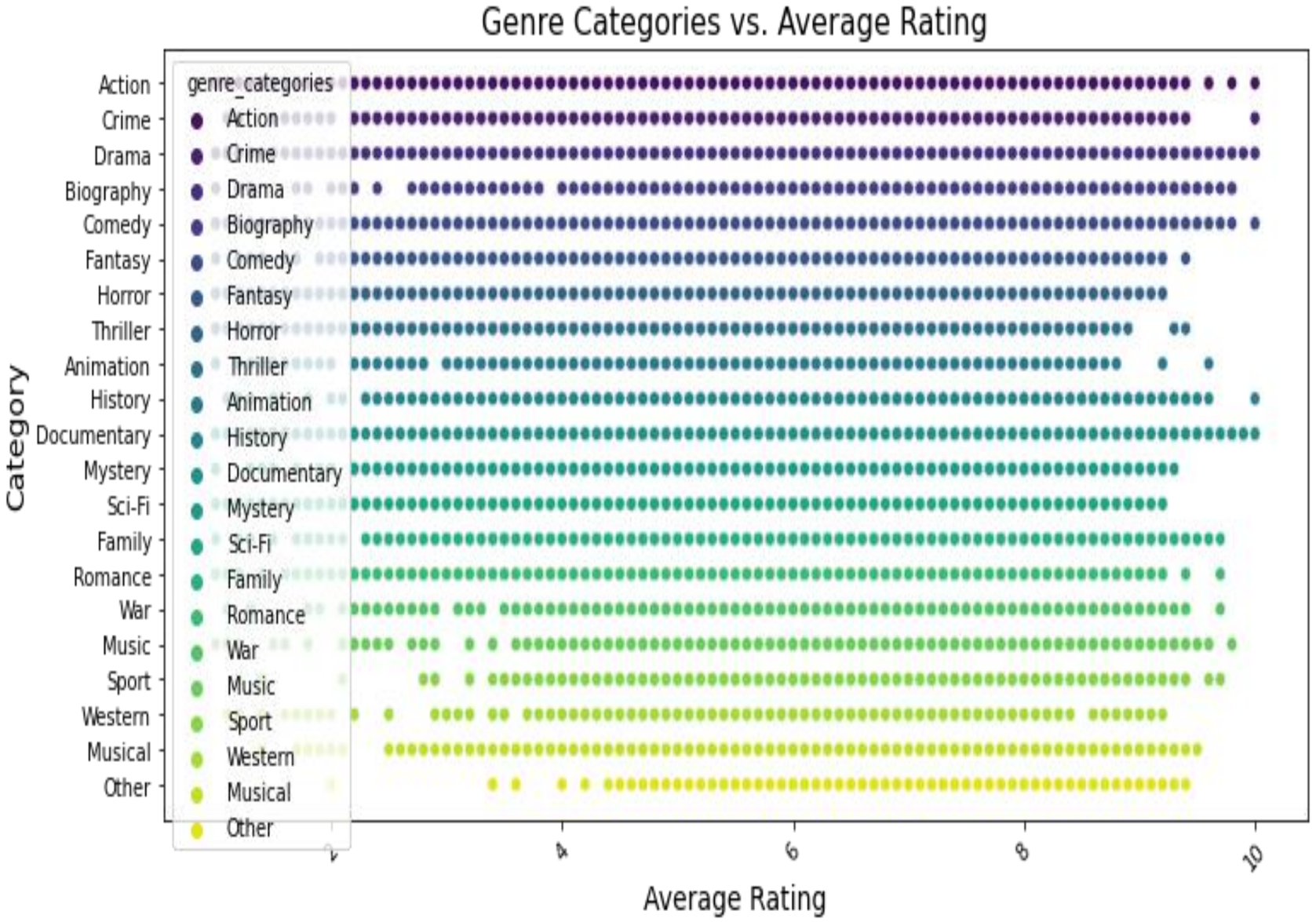
Documentaries, war, musical and sports show relatively low number of votes.



GENRE CATEGORIES VS MOVIE AVERAGE RATINGS

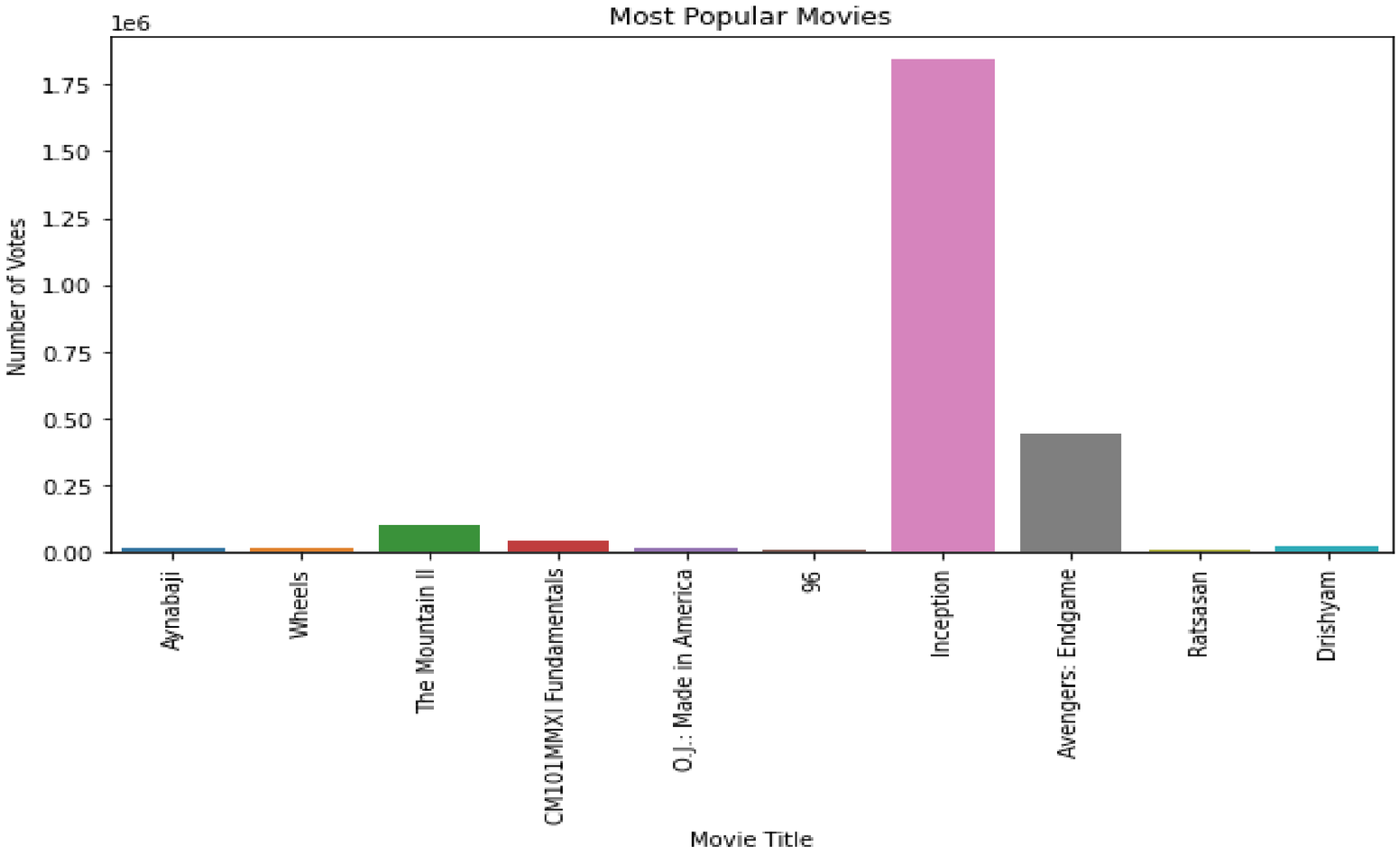
INSIGHT 1

The drama genre, along with documentary, shows a consistency in ratings across a range from low to high. Action, crime, drama, comedy, documentary, and history have also received extremely high ratings of 10. Drama appears to garner a significant amount of attention.



TOP TEN MOVIES AGAINST NUMBER OF VOTES

From the plot below, we see that the movie with the highest number of votes is 'Inception'. This film falls into the action genre category, with elements of sci-fi. Both genres have previously demonstrated high ratings and received numerous votes. The combination of these genres in one movie has resulted in a film with high votes and an average rating of 8.8, which is also relatively high.'



primary_title	Inception
start_year	2010
runtime_minutes	148.0
genres	Action, Adventure, Sci-Fi
averagerating	8.8
numvotes	1841066
genre_categories	[Action, Action, Sci-Fi]

RECOMMEDATIONS AND NEXT STEPS

RECOMMEDATIONS

Based on the analysis and findings, here are some recommendations for Microsoft's new movie studio:

Focus on Action, Drama, and Sci-Fi Genres: The analysis revealed that movies in the action, drama, and sci-fi genres tend to receive higher ratings and more votes, indicating their popularity among audiences. Microsoft should consider producing films in these genres or combining elements from these genres to increase their chances of success.

Maintain a Balanced Runtime: The analysis showed that movies with shorter runtimes tend to receive higher ratings and more votes. However, it's essential to strike a balance and not compromise the storytelling or character development. A runtime between 90 and 120 minutes could be an ideal range to aim for, as it allows for a compelling narrative without becoming too lengthy.

Experiment with Genre Combinations: The highly successful movie "Inception" demonstrated the potential of combining genres like action and sci-fi. Microsoft should explore innovative genre combinations that can captivate audiences and offer unique experiences. This approach can help their films stand out in a crowded market.

Invest in Marketing and Promotion: While the analysis focused primarily on the attributes of successful movies, effective marketing and promotion campaigns are also crucial for driving audience awareness and interest. Microsoft should allocate appropriate resources for marketing and promotion to ensure their films reach their target audiences effectively.

CONCLUSION

This exploratory data analysis provided valuable insights into the movie industry and identified key characteristics of successful films. By analyzing genres, runtime, ratings, and popularity, the analysis revealed important trends and patterns that can guide Microsoft's entry into movie production.

The key findings were:

1) Genres like action, drama, and sci-fi tend to resonate well with audiences, receiving higher ratings and more votes.

2) Movies with shorter runtimes (around 90-120 minutes) and innovative genre combinations (e.g., combining action and sci-fi) have the potential to captivate viewers and achieve box office success.

Based on these insights, several recommendations were outlined for Microsoft's new movie studio, including:

- Focusing on producing films in the action, drama, and sci-fi genres, or experimenting with combinations of these genres.
- Maintaining a balanced runtime between 90 and 120 minutes.
- Exploring innovative genre combinations to offer unique experiences.

NEXT STEPS:

1. Continuously monitor and adapt to changing audience preferences, explore additional data sources (e.g., social media, box office data).

2. Conduct further analysis on movie budgets, production teams, and distribution strategies.

THANK YOU