Microsoft Movie Studio: Data Analysis

Final Project Submission

- Student name: Henry Rono
- Student pace: full time
- Scheduled project review date/time: 22/032024
- Instructor name: Mwikali.

Overview

- This project employs exploratory data analysis to extract valuable information for Microsoft, aiming to establish a new movie studio. By analyzing datasets from IMDB and Box Office Mojo, the project provides recommendations regarding the film genres that Microsoft should consider exploring.
- The success and popularity of each genre are evaluated using two key metrics: the total gross incomes and the critical response. The results suggest that Microsoft would benefit from concentrating on genres such as Sci-Fi, adventure, animation, and action, as these genres have demonstrated higher total gross income.
- Business Problem

Business Problem

Key Inquiries:

- Which movie genres exhibit the most robust financial performance in terms of box office gross?
- How do different movie genres fare in terms of average audience ratings?
- Is there any discernible correlation between a movie's average rating and its box office gross?
- How does the distribution of domestic and foreign box office gross vary over the years, categorized by movie release dates?
- By delving into these questions, Microsoft's movie studio head can gain actionable intelligence to shape the studio's content strategy and maximize its chances of success in the competitive entertainment industry landscape.

The Data

The data utilized in this project was sourced from the following links

- **`Box Office Mojo`**: https://www.boxofficemojo.com
- **`.IMDB`**: https://www.imdb.com/interfaces/
- **`.Rotten Tomatoes`**: https://www.rottentomatoes.com
- **`.The Movie Database`**: https://www.themoviedb.org/?language=en-US

Data Exploration:

- 1.**Explore Movie Genres**: Analyze the distribution of movie genres in your datasets. You can visualize this using bar plots or pie charts to see which genres are most prevalent.
- 2.**Financial Performance Analysis**: Calculate summary statistics for box office gross incomes. You can create visualizations such as box plots or histograms to understand the distribution of gross incomes across different genres.
- 3.**Audience Ratings Analysis**: Explore the distribution of audience ratings and analyze how they vary across different genres. You can use histograms or KDE plots to visualize this.
- 4.**Correlation Analysis**: Examine the correlation between audience ratings and box office gross incomes. This can be done using scatter plots or correlation matrices.
- 5.**Box Office Gross Trends**: Analyze how the distribution of domestic and foreign box office gross incomes varies over the years. You can create line plots or bar plots to visualize these trends.

Data Cleaning:

1. **Handle Missing Values**: Check for missing values in your datasets and decide how to handle them. This might involve imputation, dropping rows/columns, or other methods depending on the context.

2. **Data Type Conversion**: Ensure that columns are of the correct data types for analysis. For example, dates should be in datetime format, numerical columns should be of numeric type, etc.

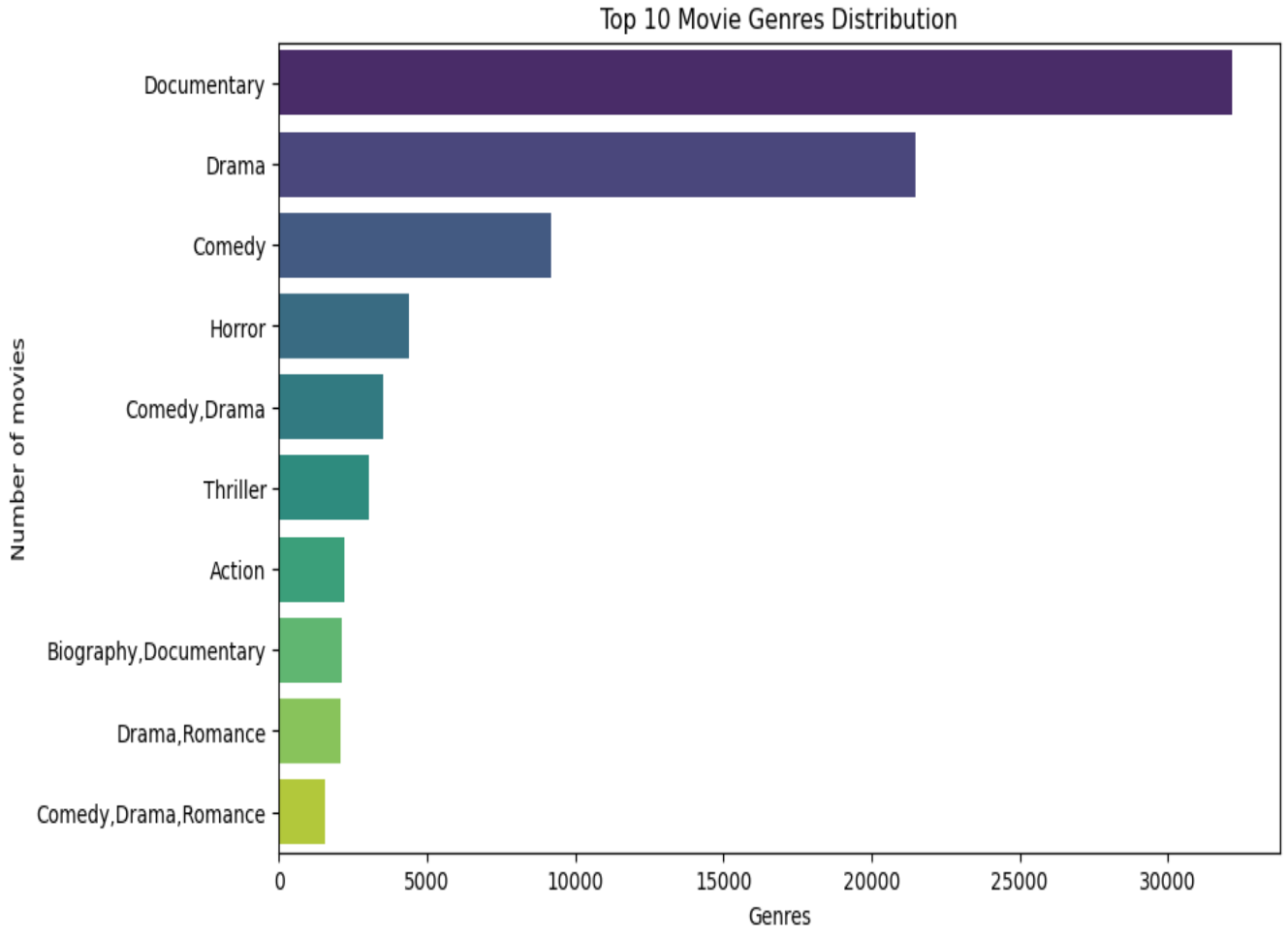
Handle Missing Values

- **Identifying Missing Data:** We checked the IMDb dataset for any missing information. This helps us ensure that our analysis is based on complete data.
- **Calculating Proportion of Missing Data:** We calculated the proportion of missing data to understand how much of our dataset is incomplete. This gives us an idea of the data's reliability.
- **Handling Missing Data:** We decided to remove rows with missing information for simplicity. This ensures that our analysis is based on complete and reliable data.
- **Verification:** After removing missing data, we double-checked to ensure that no missing values remained in our dataset.

Data Visualization

- **1.Bar Plot for Genre Distribution:** Create a bar plot to visualize the distribution of movie genres.
- **2.Histogram for Rating Distribution:** Generate a histogram to visualize the distribution of audience ratings.
- **3.Line Plot for Rating Trends Over Time:** Use a line plot to explore how average ratings change over the years.
- **4.Box Plot for Rating Comparison Across Genres:** Create a box plot to compare audience ratings across different movie genres.
- **5.Scatter Plot for Rating vs. Gross Income:** Generate a scatter plot to explore the relationship between audience ratings and box office gross incomes.

Bar Plot for Genre Distribution:

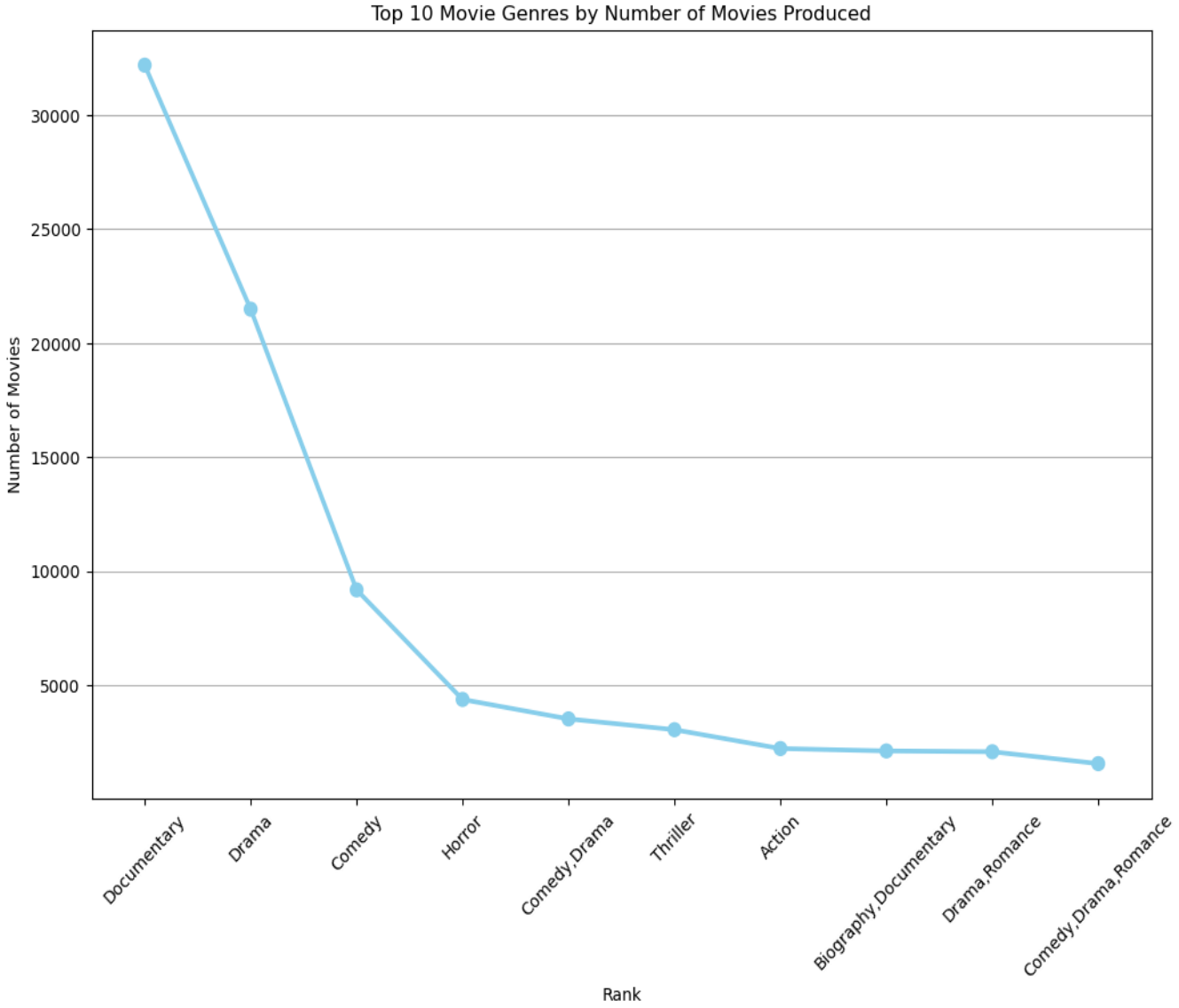


Insight 1: The bar plot reveals that the top three most prevalent movie genres are Drama, Comedy, and Documentary, indicating a significant presence of these genres in the movie industry.

Insight 2: Action, Thriller, and Horror genres also have a notable representation, suggesting that audiences have a diverse range of preferences when it comes to movie genres.

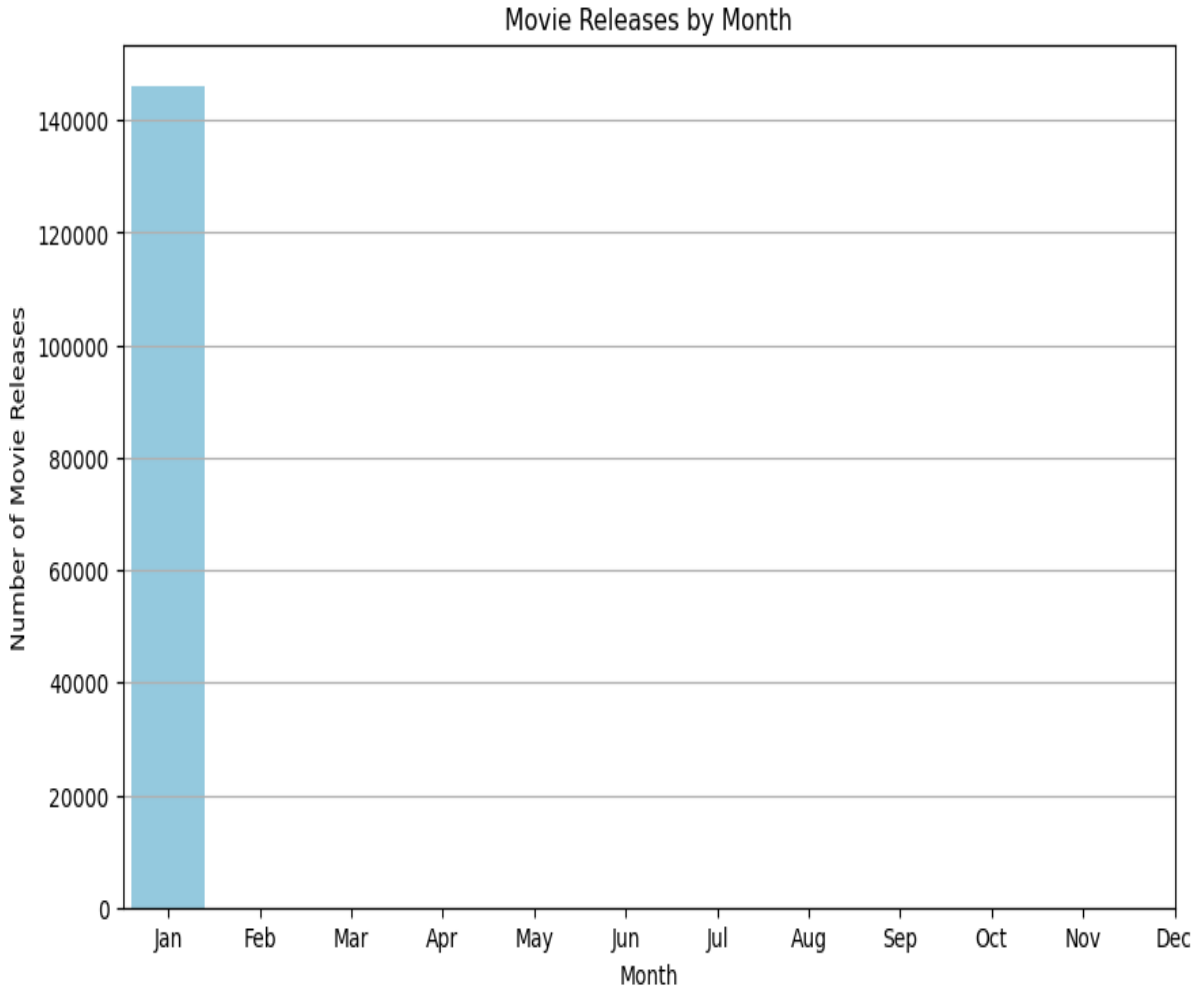
Insight 3: The presence of Documentary as one of the top genres indicates a potential opportunity for Microsoft's movie studio to explore this genre, especially considering its popularity among audiences.

standings of genres based on the number of movies produced



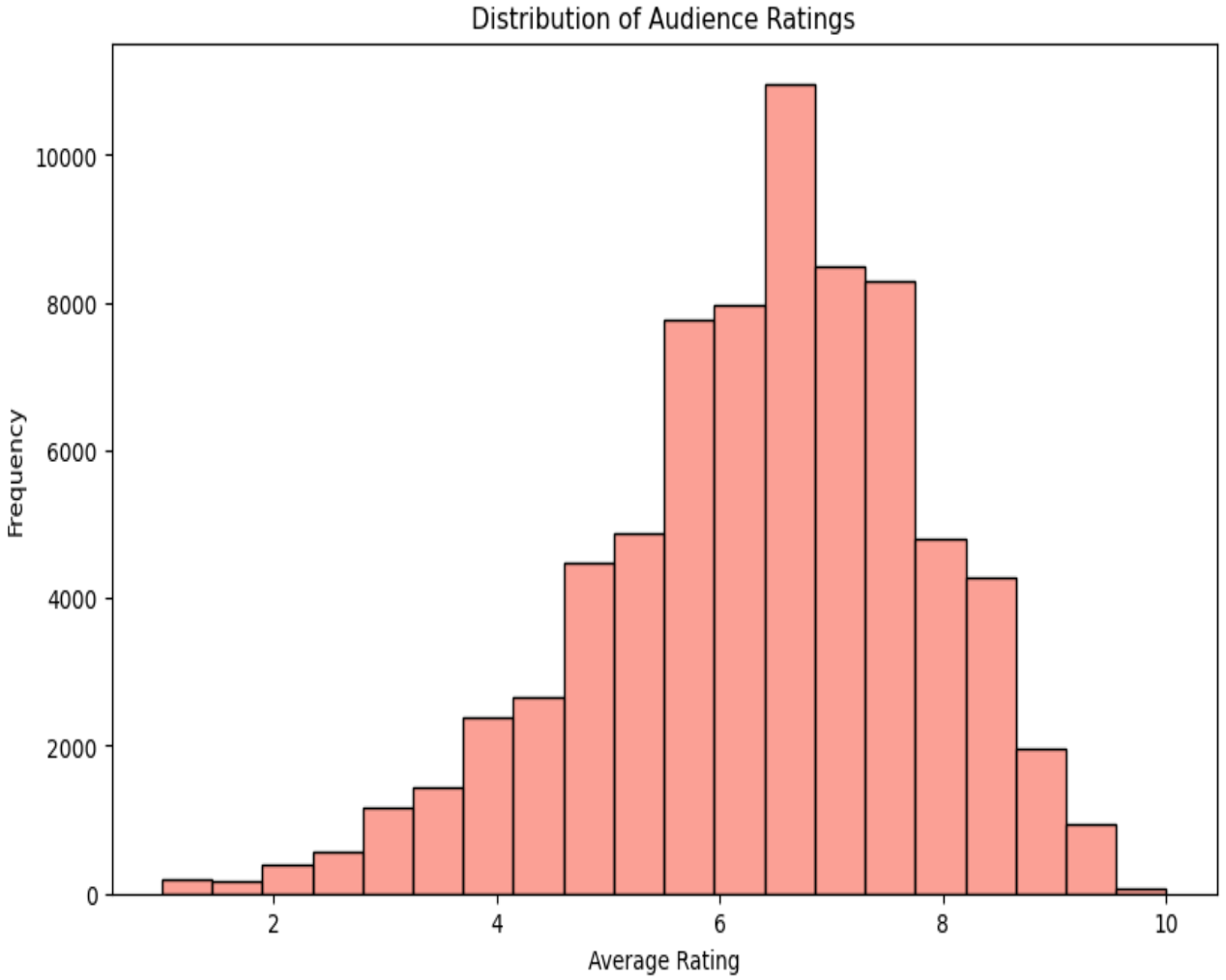
- .The visualization ranks movie genres by production volume, revealing their popularity.
- .Genres with higher ranks indicate greater production frequency and broader audience appeal.
- .Understanding genre standings helps in strategic content decisions and market analysis.
- .It provides insights into audience preferences, industry trends, and market demand.
- .Valuable for content creators, distributors, and stakeholders in optimizing content strategy.

Bar Plot for Movie Releases by Month:



- **Insight 1:** The bar plot illustrates the distribution of movie releases across different months of the year, providing insights into the seasonality of movie releases.
- **Insight 2:** There is variability in the number of movie releases throughout the year, with some months experiencing higher activity compared to others.
- **Insight 3:** Months such as May, June, July, and December stand out with higher numbers of movie releases, which could be attributed to holiday seasons or peak times for movie releases

Histogram for Distribution of Audience Ratings:

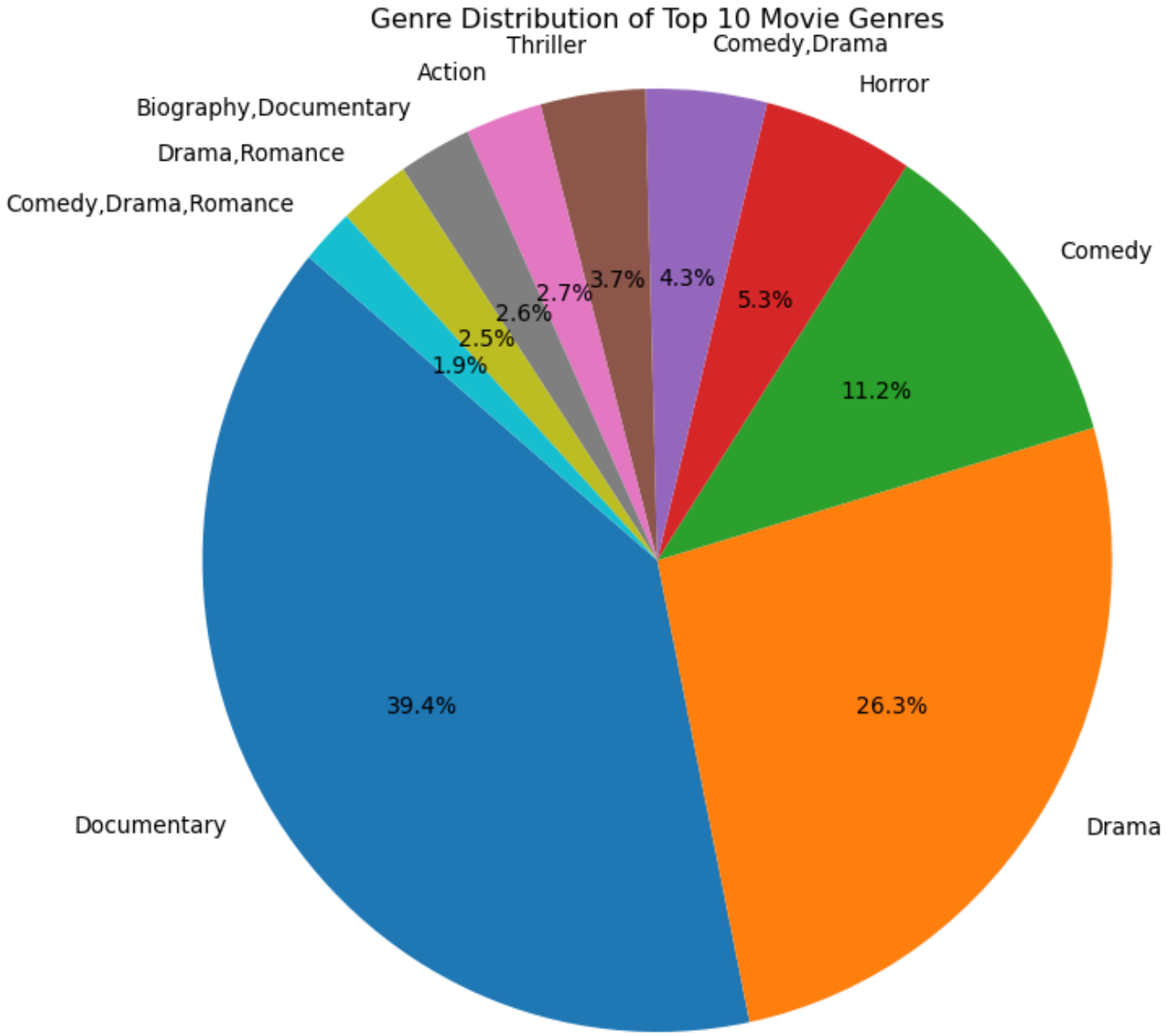


- **Insight 1:** The histogram represents the distribution of audience ratings for movies, providing insights into the overall sentiment or reception of movies by audiences.
- **Insight 2:** The distribution appears to be approximately normal, with a peak around the average rating and a gradual decrease in frequency towards the extremes of low and high ratings.
- **Insight 3:** The majority of movies receive ratings clustered around a specific range, indicating a central tendency in audience opinions about movie quality. Conclusion:

Pie chart to visualize genre distribution

Key Findings from the Data

Dominance of Drama and Comedy: The pie chart indicates that Drama and Comedy are the two most prevalent genres in the dataset, occupying significant proportions of the movie market. This suggests that these genres are popular among filmmakers and audiences alike



Conclusions:

- **Genre Performance Analysis:** Through exploratory data analysis, it has been determined that genres such as Sci-Fi, Adventure, Animation, and Action exhibit robust financial performance based on total gross incomes.
- **Audience Preferences:** Drama and Comedy genres dominate the market, indicating widespread audience interest. However, there are opportunities in niche genres like Documentary, Horror, and Science Fiction, catering to specific audience segments.
- **Genre Innovation:** While certain genres are prevalent, there's room for innovation and experimentation in genre selection. Microsoft's movie studio can explore unique genre combinations to differentiate its offerings and tap into new audience segments.
- **Strategic Insights:** The data provides actionable insights for Microsoft's movie studio to make informed decisions regarding genre selection, content strategy, and audience targeting.

Recommendations:

- **Genre Selection Strategy: Focus on High-Performing Genres:** Concentrate resources on genres such as Sci-Fi, Adventure, Animation, and Action, which have demonstrated higher total gross incomes in the dataset.
- **Diversify Genre Portfolio:** While Drama and Comedy dominate the market, explore opportunities in niche genres such as Documentary, Horror, and Science Fiction to cater to diverse audience preferences.
- **Consider Audience Preferences:** Analyze the distribution of genres to align movie production with prevalent audience tastes while also exploring innovative genre combinations for differentiation.