PHASE 1 PROJECT PRESENTATION

By Cleve Mwebi

BUSINESS PROBLEM

Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

PROBLEM ANALYSIS

To determine what types of films are currently performing best at the box office, I will analyze data from sources like IMDB. By looking at metrics like genre, gross revenues, average rating and many others, I can identify trends and patterns. My analysis would provide actionable insights like targeting specific genres that overindex with certain demographics, optimizing production budgets based on genre, and selecting ideal release dates. This data-driven approach will inform what types of films Microsoft should greenlight to maximize their chances of box office success.

DATASETS

3 datasets will be used in tackling the aforementioned business problems namely:

1. bom.movie_gross.csv – contains data on title of the movie, domestic and foreign revenue accrued, and year of movie production.
2. imdb.title.basics.csv – contains data on title of the movie, runtime minutes and genres
3. imdb.title.ratings.csv – contains the movie id, average rating of the movie and the number of votes for every movie

QUESTIONS

These datasets aforementioned will be used to answer the following questions:

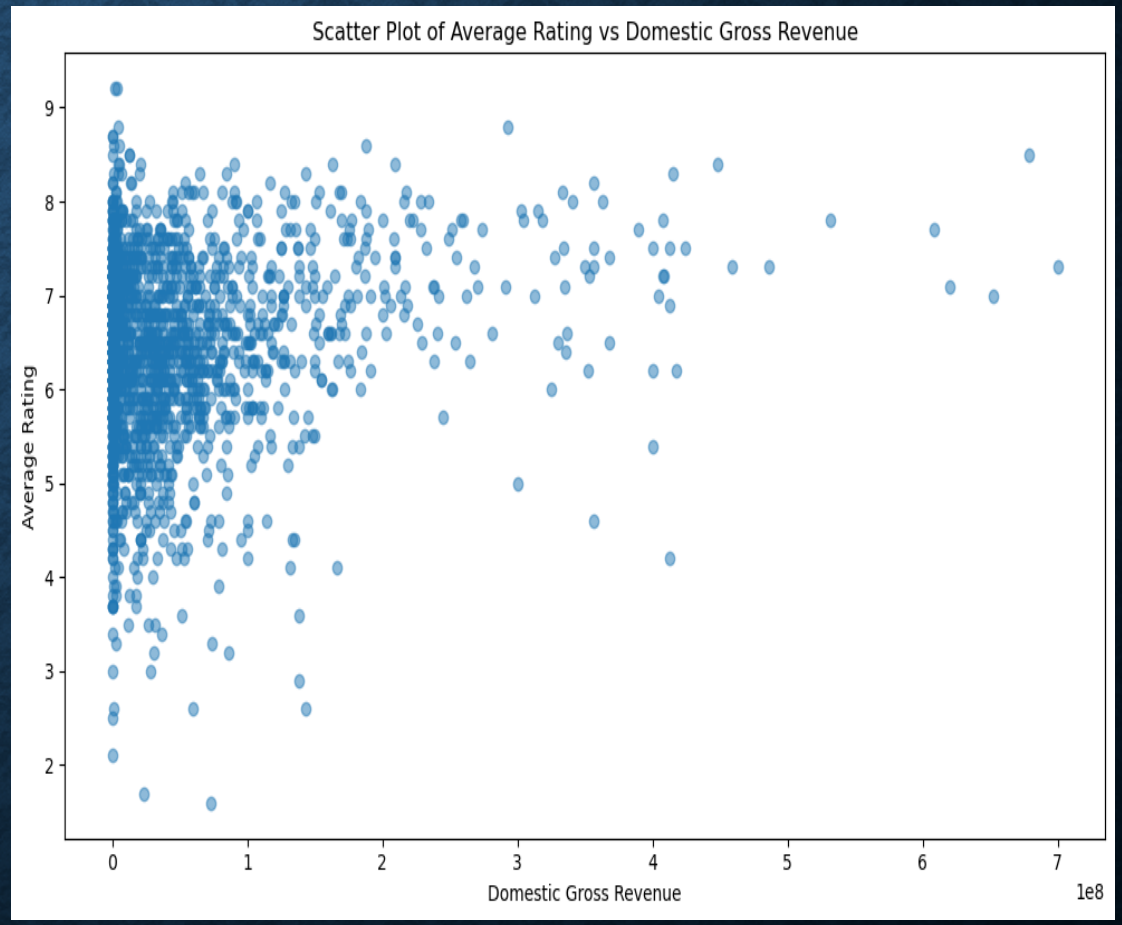
1. What are the most popular movie genres?
2. Which movie genres have the highest average ratings?
3. What is the relationship between movie budget and revenue?
4. Which movie studios have produced the most successful films?

METHODOLOGY

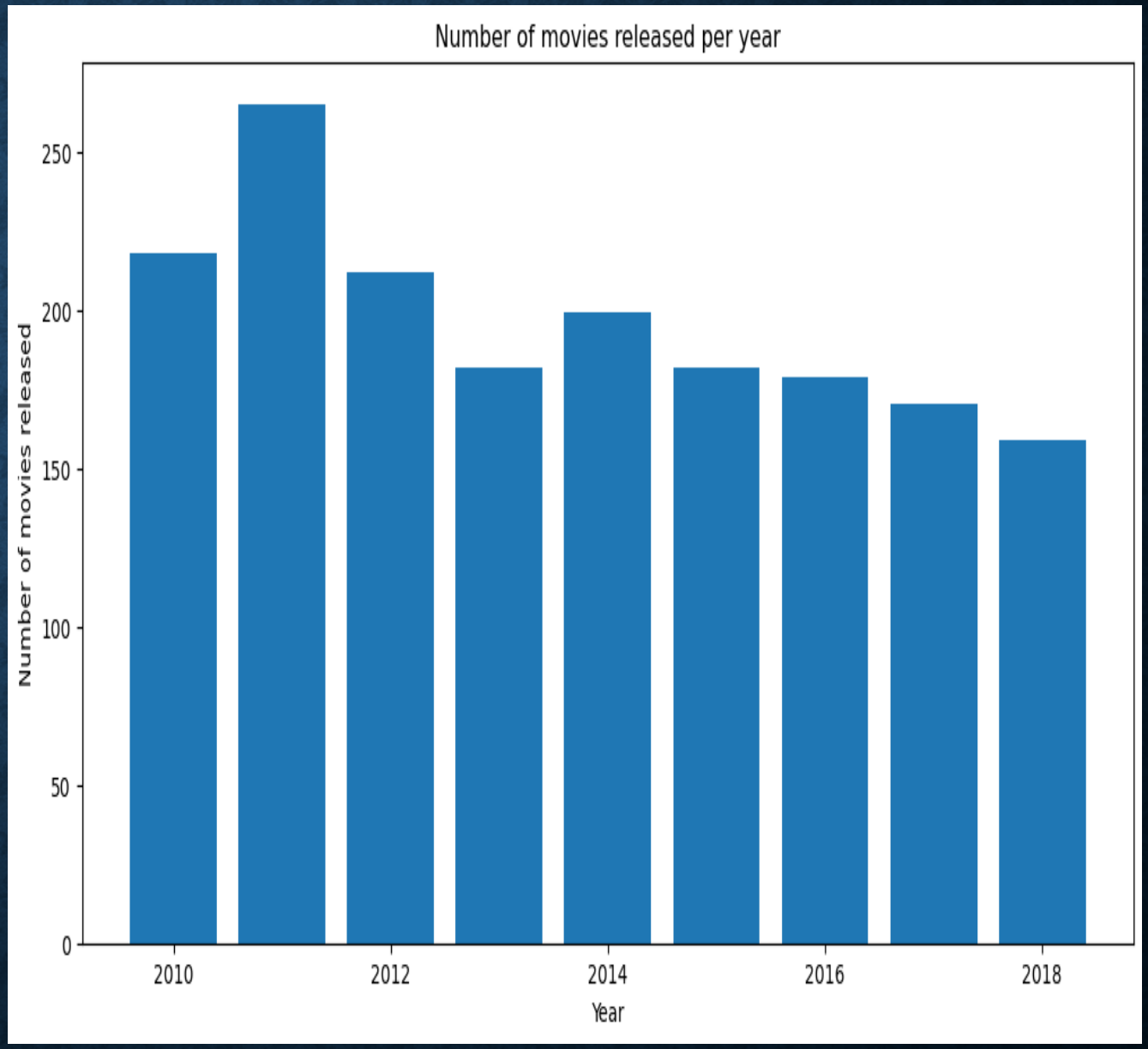
- The project made use of both descriptive and analytic techniques in conducting its EDA.
- This was done with the help of different python packages such as pandas, numpy, matplotlib, seaborn and others.
- Some of the aspects of analysis involved description of trends through visualization and data modelling

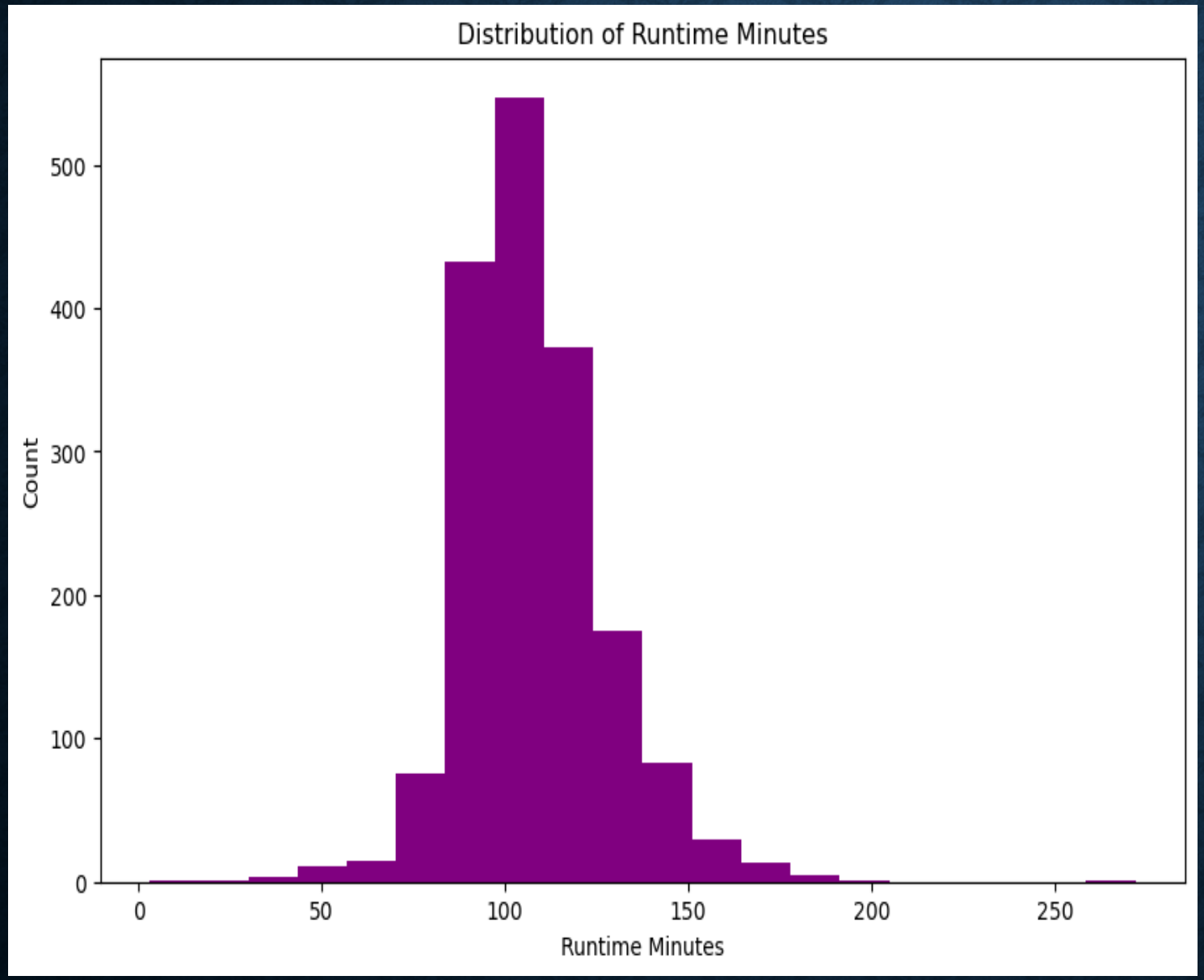
RESULTS

- The scatter plot shows a weak positive correlation between ratings and revenue, meaning higher rated films tend to earn more. However, the distribution of revenue is heavily skewed right with most movies clustered at lower levels and a few high-grossing blockbusters outliers on the right. This indicates ratings have a minor impact on earnings, but box office hits substantially outperform most other releases.



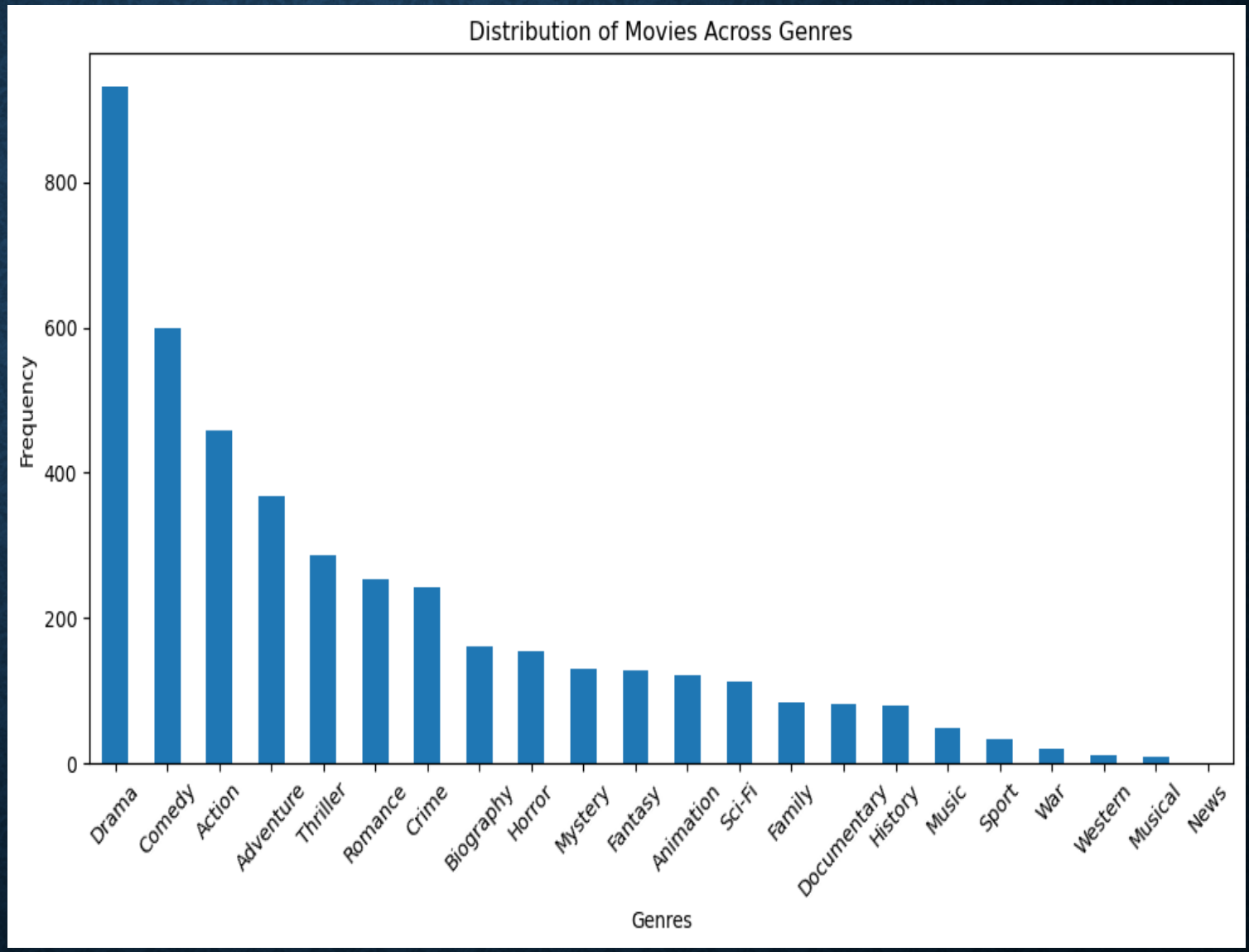
The bar chart illustrates a downward trend in the quantity of movies released annually, with the highest number hitting theaters in 2011 followed by a steady decline over the subsequent years. Analyzing these release patterns and trends can provide helpful competitive intelligence for Microsoft when it comes to understanding the evolving landscape of the movie industry.





- The histogram displaying movie runtimes in minutes demonstrates that most films have a duration between 80 and 120 minutes, with the highest frequency of movies having a runtime of approximately 100 minutes.

The bar chart depicts the distribution of movie genres illustrates that Drama, Comedy, and Action are the most prevalent categories. These films account for the highest percentages of total movies released. Analyzing the popularity of genres can provide helpful marketing insights for media companies aiming to cater their content and messaging to audience preferences.



CONCLUSION

- In summary, this analysis offers valuable business insights about profitable genres, foreign markets, and other industry trends that can empower data-driven decision-making.
- However, limitations exist regarding external factors, time frame, and regional constraints.
- Future enhancements could incorporate broader data sources, wider time periods, and global markets to improve scope. Additionally, machine learning algorithms could be deployed to predict movie success based on genre, runtime, cast, budget, release date, and other attributes.
- Though imperfect, these findings provide a meaningful perspective into key drivers and patterns in the complex movie industry landscape.
- Expanding the analytical approaches will offer more comprehensive intelligence for long-term strategy.

RECOMMENDATIONS

- Examine domestic and international box office earnings distributions. Identify variables like genre, studio, release year, and ratings that drive higher revenues.
- Explore genre popularity over time among audiences. Analyze genre distribution across movies and determine consistently in-demand categories.
- Study runtime patterns and trends. Calculate average or optimal durations and determine correlations between length, ratings, and box office earnings.
- Review historical release volume fluctuations. Pinpoint years with more or fewer releases and factors causing shifts in production levels.
- Investigate connections between average ratings and potential factors including revenue, genre, runtime, and release year. Look for influential relationships.

Overall, analyzing these key areas will provide data-driven insights into audience preferences, performance predictors, and industry trends to inform strategic decisions for Microsoft's movie ventures.