

INTRODUCTION

In this project, we aim to provide valuable insights to Microsoft's new movie studio by conducting exploratory data analysis on the current trends in the box office. As Microsoft ventures into the world of original video content, it is crucial to understand the factors that contribute to the success of films. By analyzing various attributes such as genre, gross earnings, and ratings, we will uncover patterns and correlations that can guide the decision-making process for the creation of successful movies.

BUSINESS UNDERSTANDING

The problem at hand is that Microsoft's new movie studio lacks knowledge and experience in the film industry. To ensure a successful entry into the market, it is essential to gain an understanding of the factors that influence a film's performance at the box office. By identifying the types of films that are currently thriving and exploring the relationship between various variables, we can provide actionable insights to guide Microsoft in deciding the most promising genres, gross earnings , and ratings for the best performing movies for their upcoming movie productions.

DATA UNDERSTANDING

This involves identifying and collecting data of interest from sources after agreeing on analytical questions and identifying the objective for analysis. The data for analysis for this project have come from the zippedData provided and its contents are

im.db.zip -Zipped SQLite database -movie_basics and movie_ratings tables are most relevant
bom.movie_gross.csv.gz -Compressed CSV file

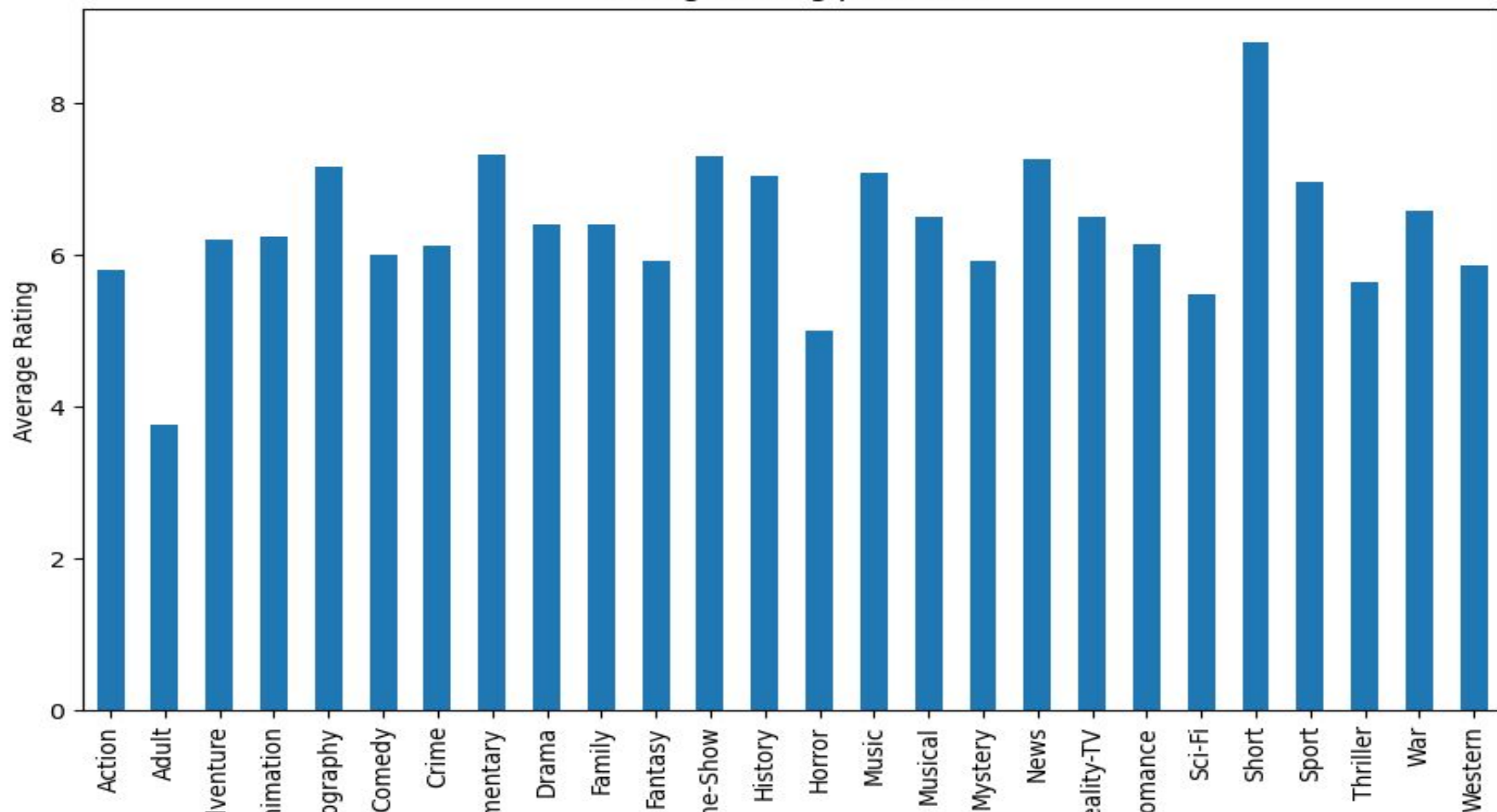
The target variables for this analysis are popularity, gross earnings, genres, and audience ratings. These variables are all important and necessary to address Microsoft's business problem. Audience rating , gross earnings and release year are numeric variables (though gross earnings is continuous) and genres are categorical.

DATA ANALYSIS

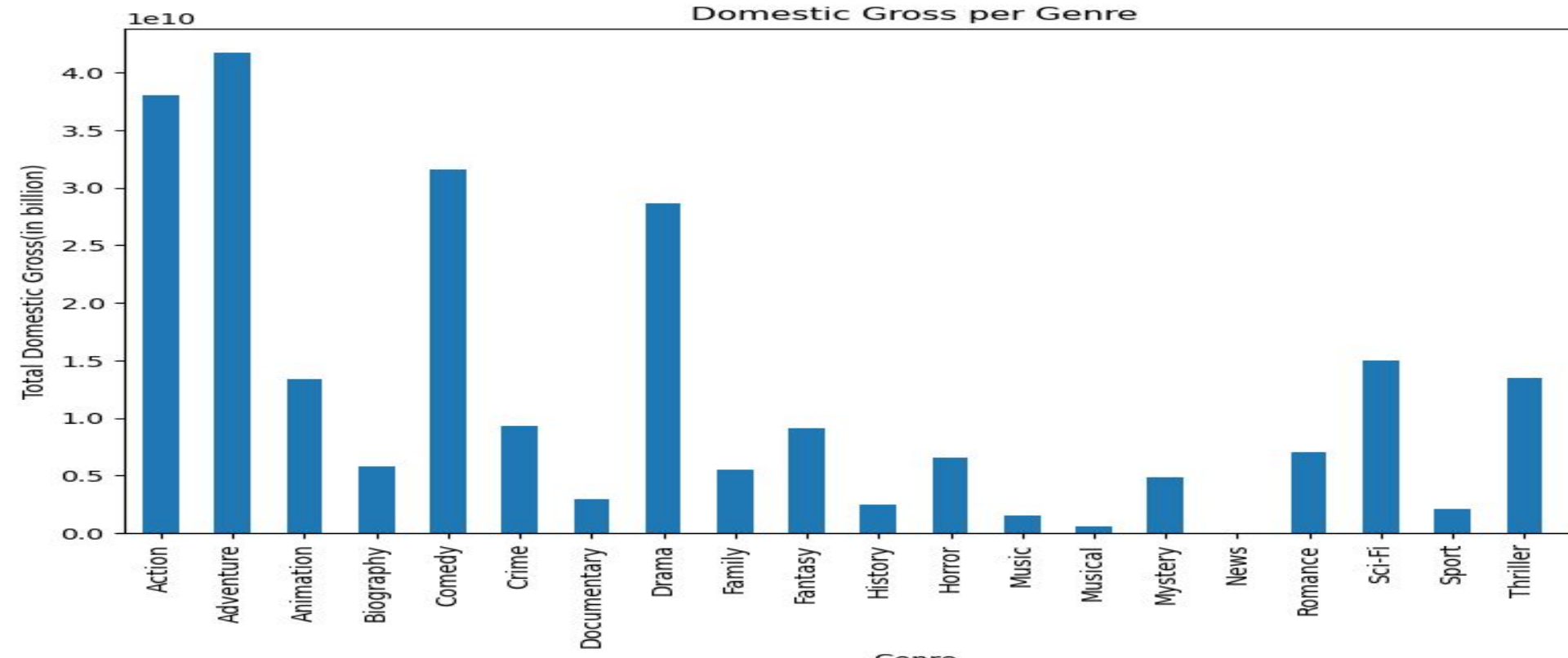
After performing data cleaning and preparation the foundation for data analysis has been set. This involves examining and understanding the structures, patterns and relationships within dataset. It also involves calculating basic descriptive statistics, creating plots like bar graphs visualizations to focus on specific segments of data. Bar graph visualization suited this analysis and thus was employed in visualizing every aspect of data required.

Analysis of average rating per genre

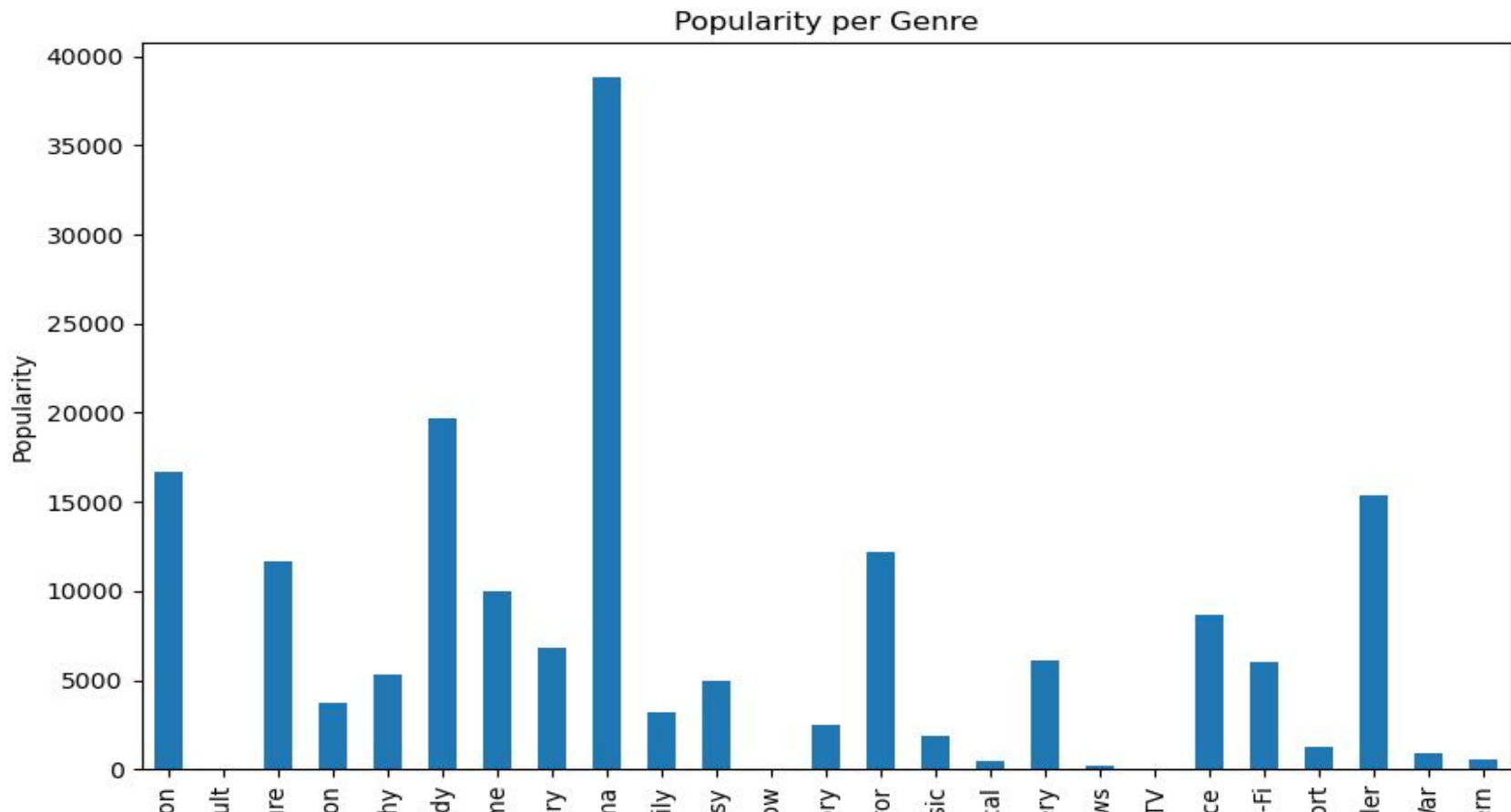
Average Rating per Genre



Domestic gross per genre



Movie popularity per genre



conclusion

- Short movies genre had the highest average rating whereas Adult genre had the least average rating.
- Action movies genre and adventure genre had relatively high domestic gross earnings. News, War and Western genres had the least domestic gross earnings.
- Drama movies genre were the most popular whereas Game_Show, News and Reality-TV were least popular.

Recommendation

For successful sustained domestic gross earnings overtime Microsoft new movie studio should consider production of Action and Adventure movies.

-For creation of popular movies Microsoft new studio should consider production of Drama Drama movies.Movie popularity generates buzz on social media and in turn win awards and also large audiences can contribute to the financial success of the movie.

-For movies with the best rating Microsoft new studio should consider creating movies of Short genre.Best rating attracts more viewers,makes the movie recognized thus boosting its reputation and also good rating attract lucrative opportunities and financial benefits through increased sales.