MICROSOFT DATA ANALYSIS PROJECT

Microsoft Movie Analysis

Professional Pitching presentation

BY JOSEPH KINUTHIA

PROBLEM

The project aims to analyze the movie industry data to provide insights for Microsoft's movie analysis project. The dataset includes information on movie titles, release years, genres, ratings, votes, studios, and domestic and foreign gross revenues.

PROBLEM ANALYSIS

The business problem we are trying to solve is to help Microsoft, who wants to create a new movie studio, to identify the types of films that are currently performing well at the box office. By understanding which types of films are currently successful, Microsoft can make data-driven decisions about what types of films to produce.

DATASETS

We will be analyzing two datasets: "movie_basics" and "movie_ratings". From these datasets, we will try to answer questions such as:

Research Questions

* What are the most popular movie genres?
* Which movie genres have the highest average ratings?
* What is the relationship between movie budget and revenue?
* Which movie studios have produced the most successful films?

DATASETS CONT..

For this project, we are using two datasets: "movie_basics" and "movie_ratings".
***
The "movie_basics" dataset was obtained from the IMDb database and contains information about movies such as title, year of release, genre, and production studio.
***
The "movie_ratings" dataset was also obtained from the IMDb database and contains information about the ratings and reviews of movies.
***
"""

VISION

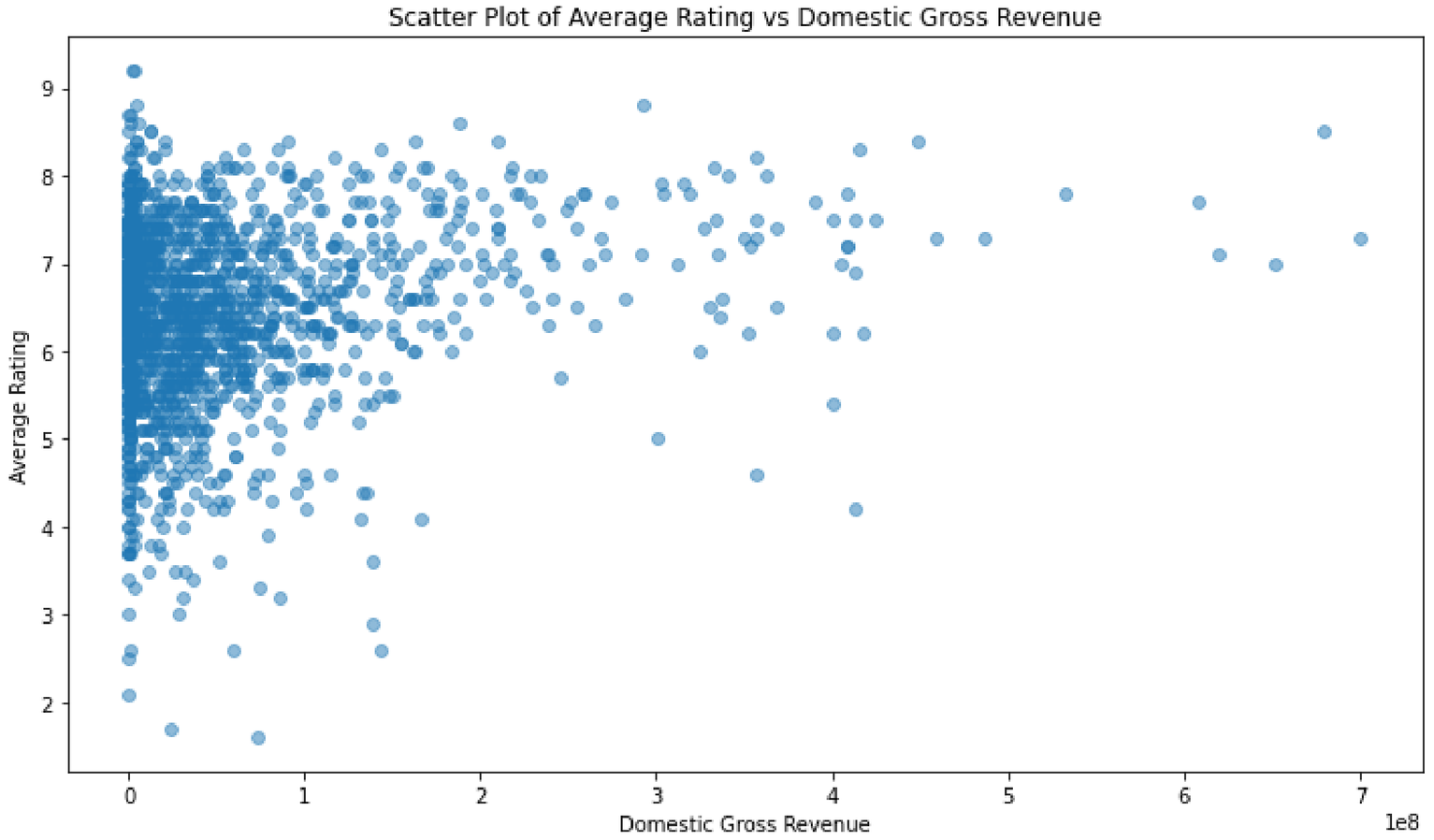
METHODS

 # Methods

This project uses descriptive analysis, including a description of trends using visualization and data modeling.

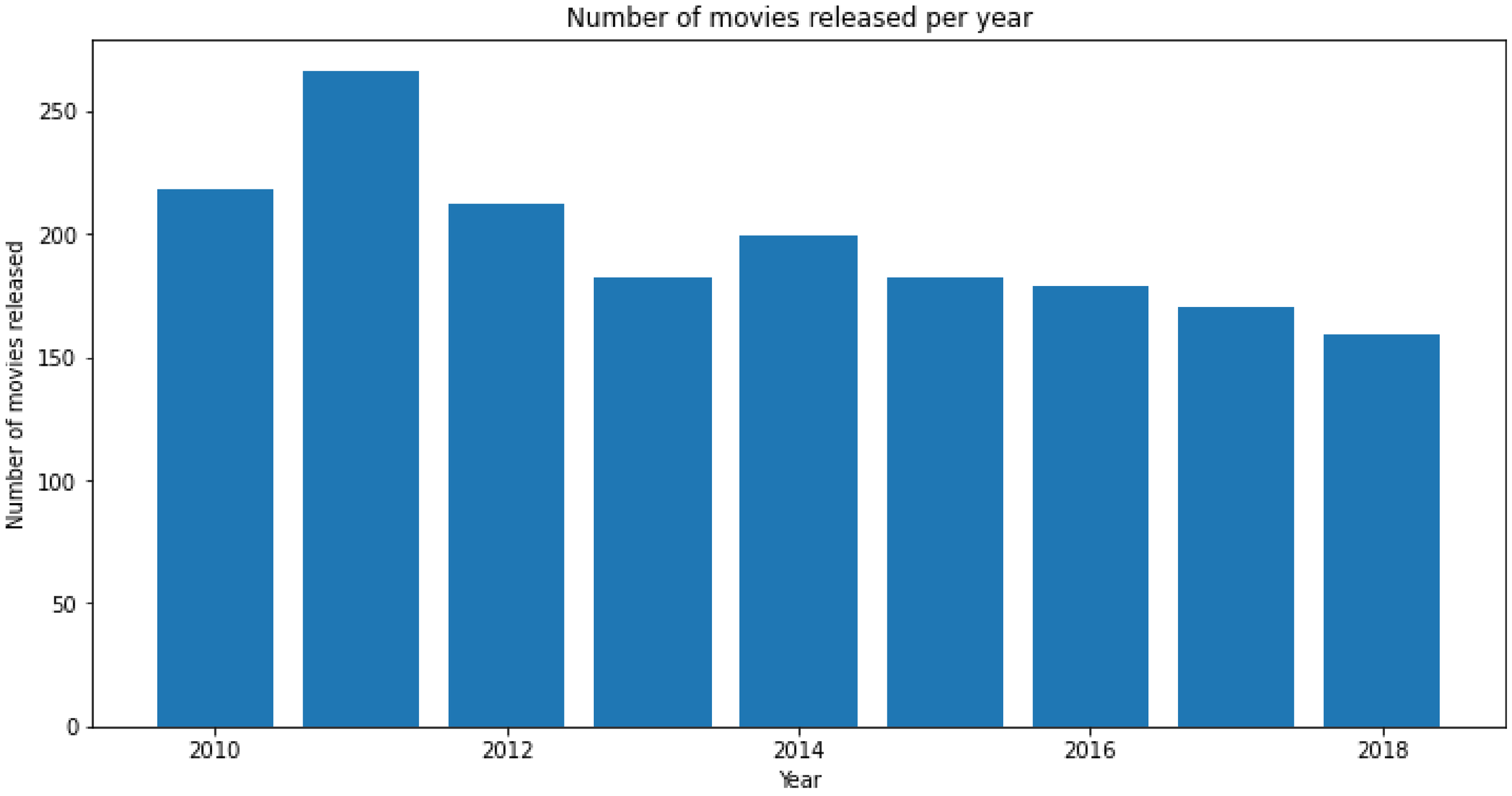
Results

The scatter plot shows a weak positive correlation between the average rating and domestic gross revenue of movies. There is a skew or imbalance in the distribution of domestic gross revenue. This means that most of the movies have lower domestic gross revenue, while only a few movies have higher domestic gross revenue.

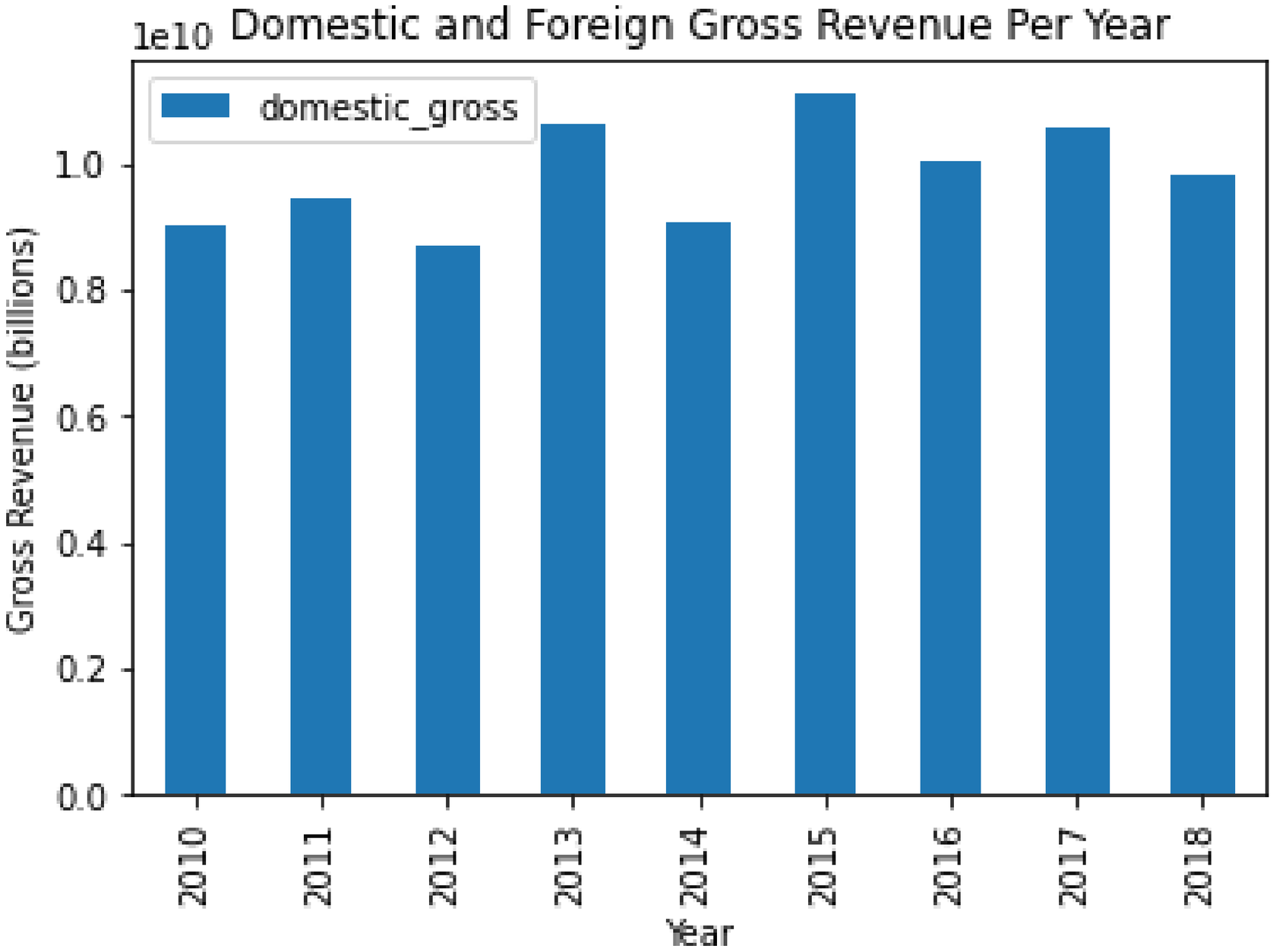


Created by
Team

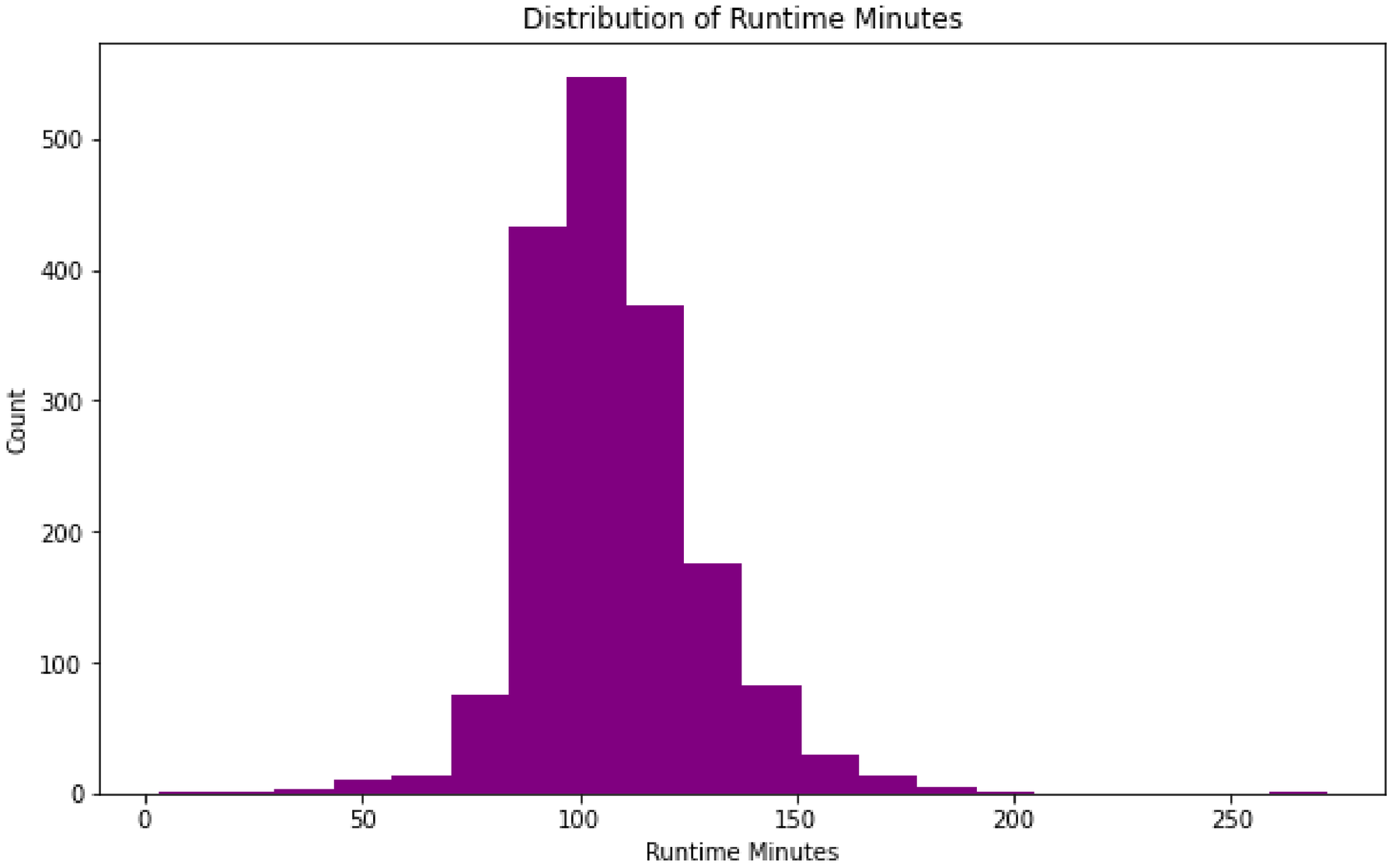
The bar chart shows a decreasing trend in the number of movies released over the years, with a peak in 2011. This trend can be useful for businesses to understand the competition and market trends.



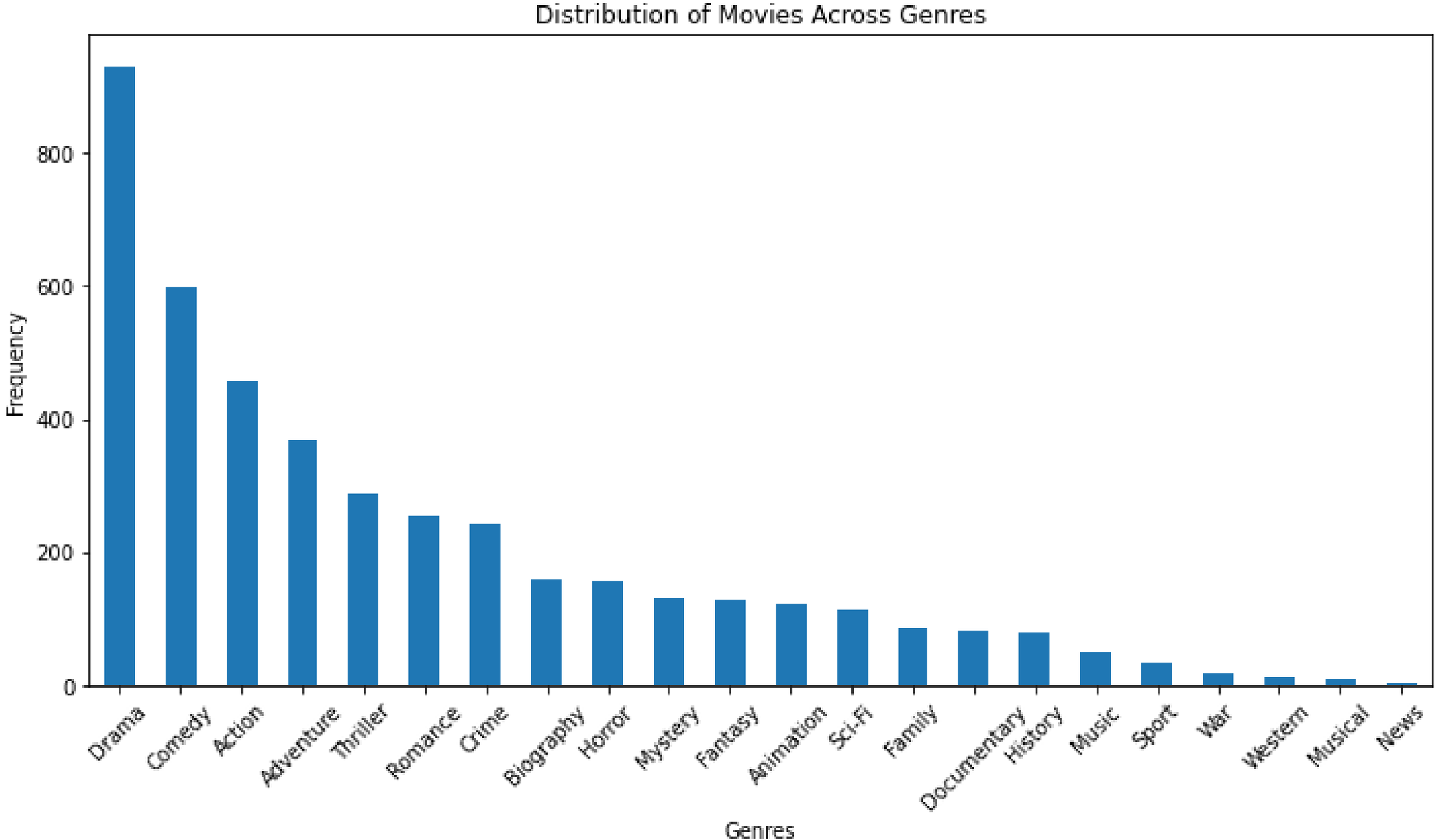
The stacked bar chart shows that domestic revenue is consistently higher than foreign revenue. It also shows that there was a dip in revenue in 2012.



Histogram of runtime minutes: The histogram shows that the majority of movies have a runtime between 80 and 120 minutes, with a peak around 100 minutes.



Bar plot of genre distribution: The bar plot of genre distribution shows that the most popular genres of movies are Drama, Comedy, and Action. This can be useful for businesses to understand the preferences of their target audience.



Conclusions

Overall, our analysis provides valuable insights into the movie industry. The business can use these insights to make data-driven decisions, such as identifying the most profitable genres and focusing on increasing foreign revenue.

However, there are limitations to our analysis. For example, we did not consider the impact of external factors such as competition and economic conditions on the movie industry. Also, the dataset is limited to a specific time frame and region, which may not be representative of the entire industry.

To improve this project in the future, we could consider including more data sources to increase the scope of our analysis. We could also use machine learning algorithms to predict the success of a movie based on its genre, runtime, and other factors.

Based on the evaluation of the visualizations, here are some recommendations for the Microsoft Movie Analysis project:

Revenue Analysis: Explore the distribution of domestic and foreign gross revenue for movies. Analyze the factors that contribute to higher revenue, such as genre, studio, release year, and average rating.
Genre Preferences: Investigate the popularity of different genres among the audience. Analyze the distribution of movies across genres and identify the most popular genres over time.
Runtime Analysis: Examine the distribution of movie runtimes and identify patterns or trends. Determine the average or preferred runtime among viewers and analyze whether there is any correlation between runtime and average ratings or revenue.
Release Year Analysis: Study the trends in movie releases over the years. Identify periods of higher or lower movie production and analyze the factors that contribute to these trends.
Average Rating Analysis: Investigate the relationship between average ratings and other factors like revenue, genre, and runtime.