



A large, semi-transparent orange graphic is positioned on the left side of the slide. It features a white square pattern that forms the outline of the letter 'Z' and the word 'PROJECT'. The graphic has a soft, rounded shape and is surrounded by a subtle, multi-colored gradient of purple, orange, and red.

By Limo Dennis, Ronoh Peter, Osoro
Delvin, Thuo Boniface and Gitau Joy

Business **PROBLEM**

A company aims to start a new movie studio and wants to compete with major studios in creating original content. They lack insight into which types of films perform best at the box office and need data-driven recommendations to make informed production decisions.





Dataset Used:

- **Budget Data:** Movies' release dates, production budgets, domestic and worldwide gross revenues.
- **Movie Ratings Data:** Genres, average ratings, popularity scores, and vote counts.

Method:

- Conducted Exploratory Data Analysis (EDA), statistical analysis, and visualizations, followed by predictive modeling, to derive actionable insights.

Thesis Statement: This project aims to determine the factors that drive box office success, including budget allocation, genre choice, and pre-release popularity.



Data OVERVIEW

01

Budget Dataset:

- Contains data on 5,782 movies, with columns including:
 - Production Budget: Ranging up to \$425 million.
 - Domestic Gross: Highest being \$936 million.

Worldwide Gross: Maximum reached \$2.77 billion.

02

Movie dataset

Dataset: 26,517 movies

Key Columns:

- genre_ids: Genre identifiers for analysis
- original_language: Language diversity insights
- popularity: Measures audience appeal
- vote_average & vote_count: Indicators of quality and engagement

03

im.db Dataset:

26,517 entries including genre, popularity score, and average ratings for quality assessment.



METHODS FOR DATA EXPLORATION

- **Data Cleaning:** Duplicates removed; missing values handled to maintain data integrity.
- **Data Transformation:** Converted revenue columns to integers for analysis consistency, created Foreign Gross and Profit columns.
- **Data Merging:** Combined data from multiple sources (budget and ratings datasets) to create a unified dataset for comprehensive analysis.

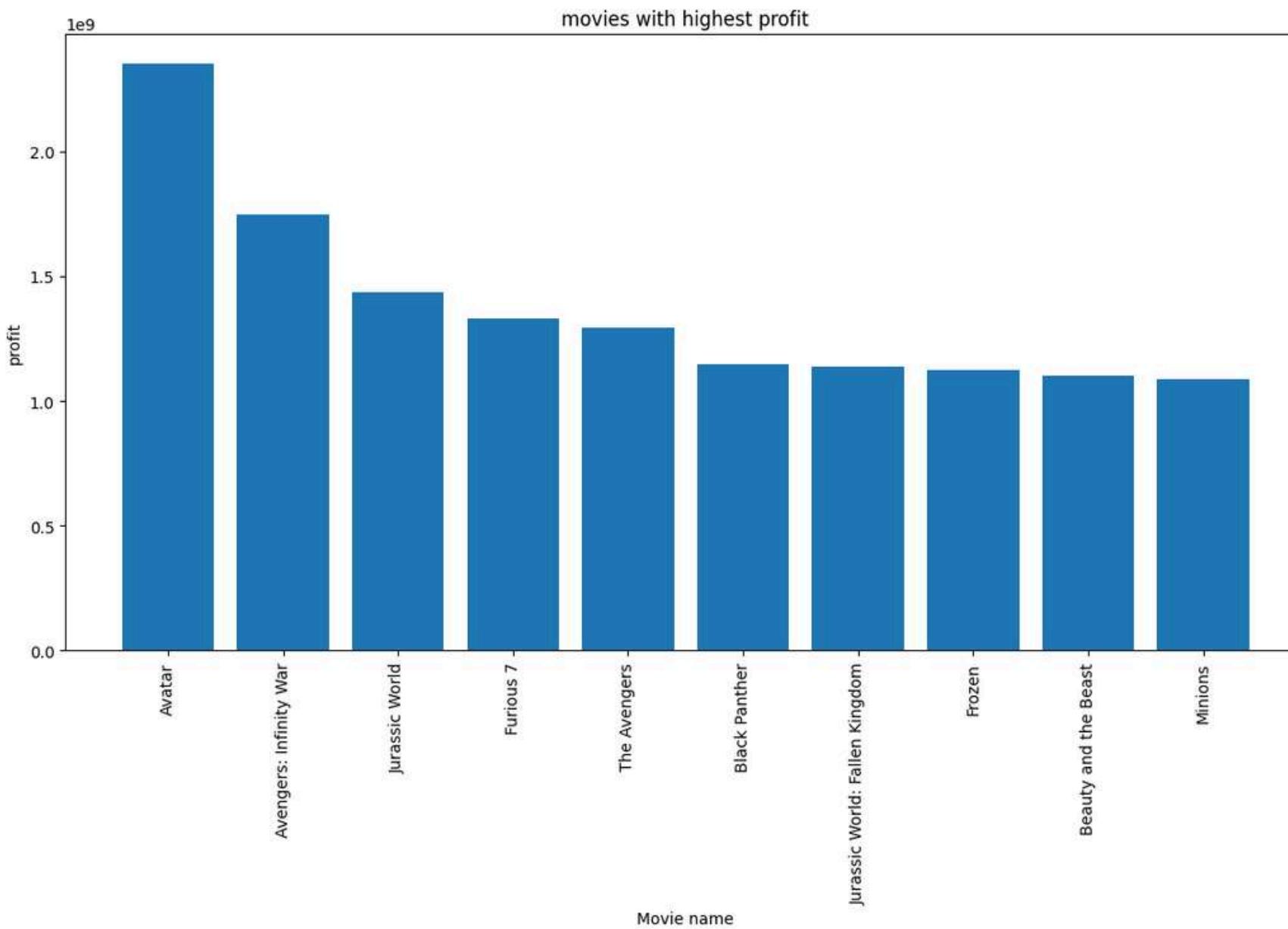


VISUALIZATIONS

- **Top Profitable Movies:**
 - Bar graph showing "Avatar" as the highest-grossing movie worldwide, with a profit of over \$2 billion.
- **Heatmap Analysis:**
 - Showcased strong correlations among Production Budget, Domestic Gross, and Worldwide Gross.
 - Genres like Drama are frequently associated with higher ratings.
- **Genre Popularity:**
 - Drama emerged as the most common genre, followed by Documentary, indicating audience interest.



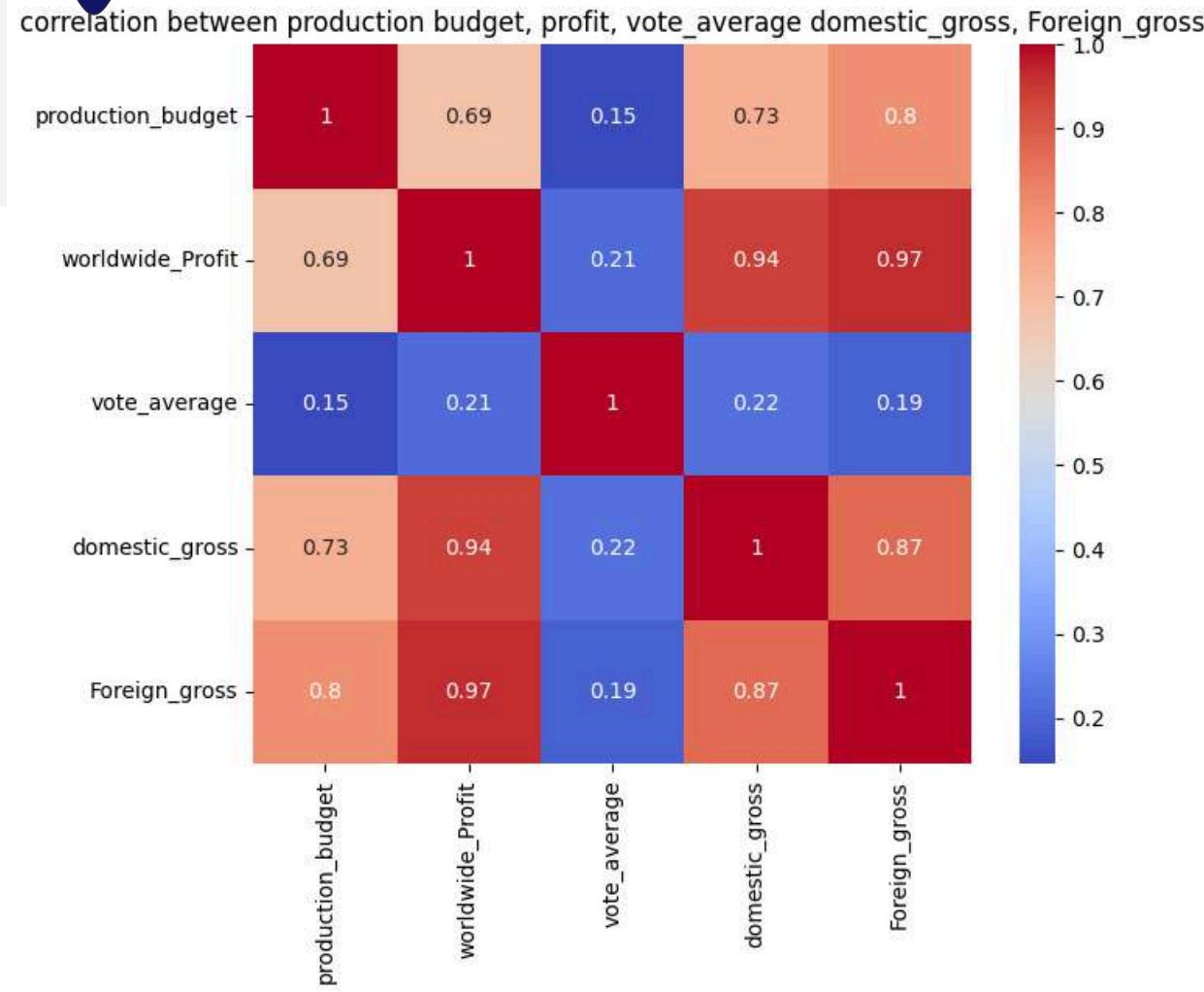
1



TOP MOVIES

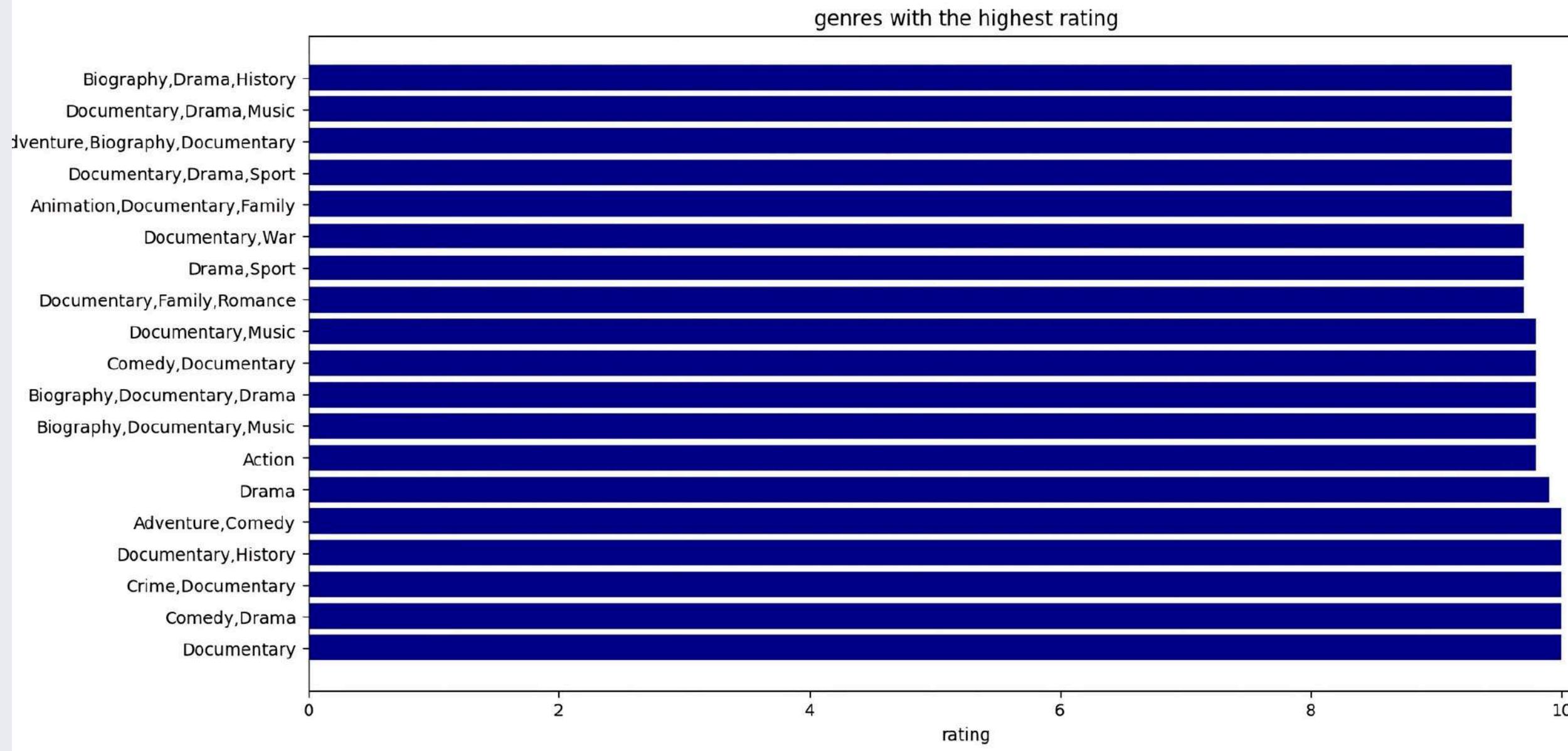
This bar chart displays the movies with the highest profit, showing that Avatar leads with the highest profit, followed by Avengers: Infinity War and Jurassic World, indicating these titles generated significant financial success compared to other high-grossing films.

2



HEATMAP

This heatmap shows the correlation between various film-related metrics, indicating strong positive correlations between production budget, worldwide profit, domestic gross, and foreign gross, while vote average has a weaker correlation with these financial metrics.



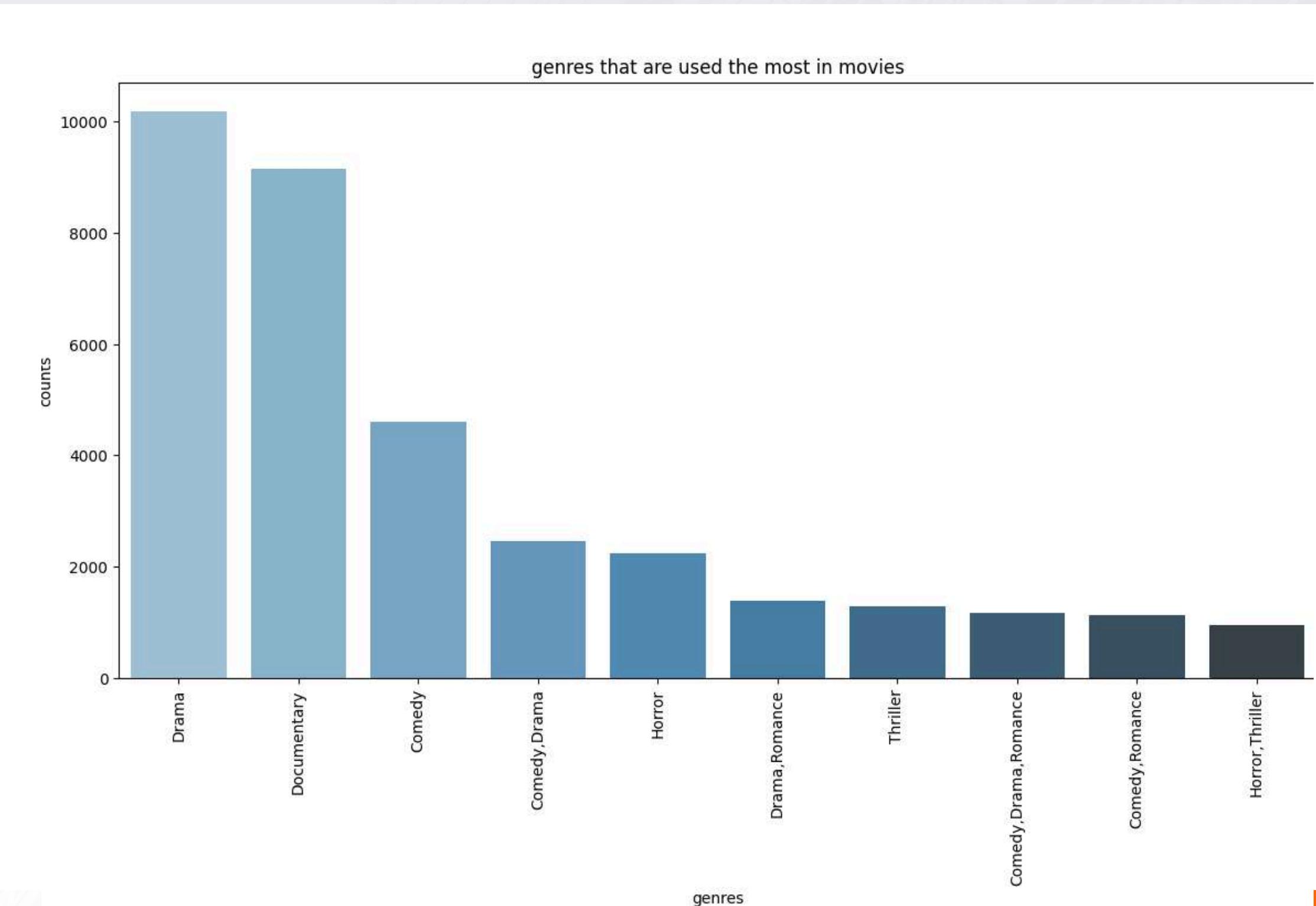
PERFORMANCE OF DIFFERENT GENRES

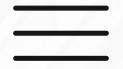
The bar chart displays a range of genres that have achieved the highest rating of 10, indicating that these genres are tied for the top rating rather than showing a range of average ratings. Each genre listed, which includes combinations like "Biography, Drama, History" and "Documentary, Drama, Music," has received a perfect rating.

5

PERFORMANCE OF DIFFERENT GENRES

The bar chart displays the most frequently used movie genres, with Drama and Documentary leading significantly, followed by Comedy, while genres that combine multiple elements, such as Comedy/Drama/Romance and Horror/Thriller, appear less frequently, suggesting a strong audience preference for traditional genres over hybrid ones.



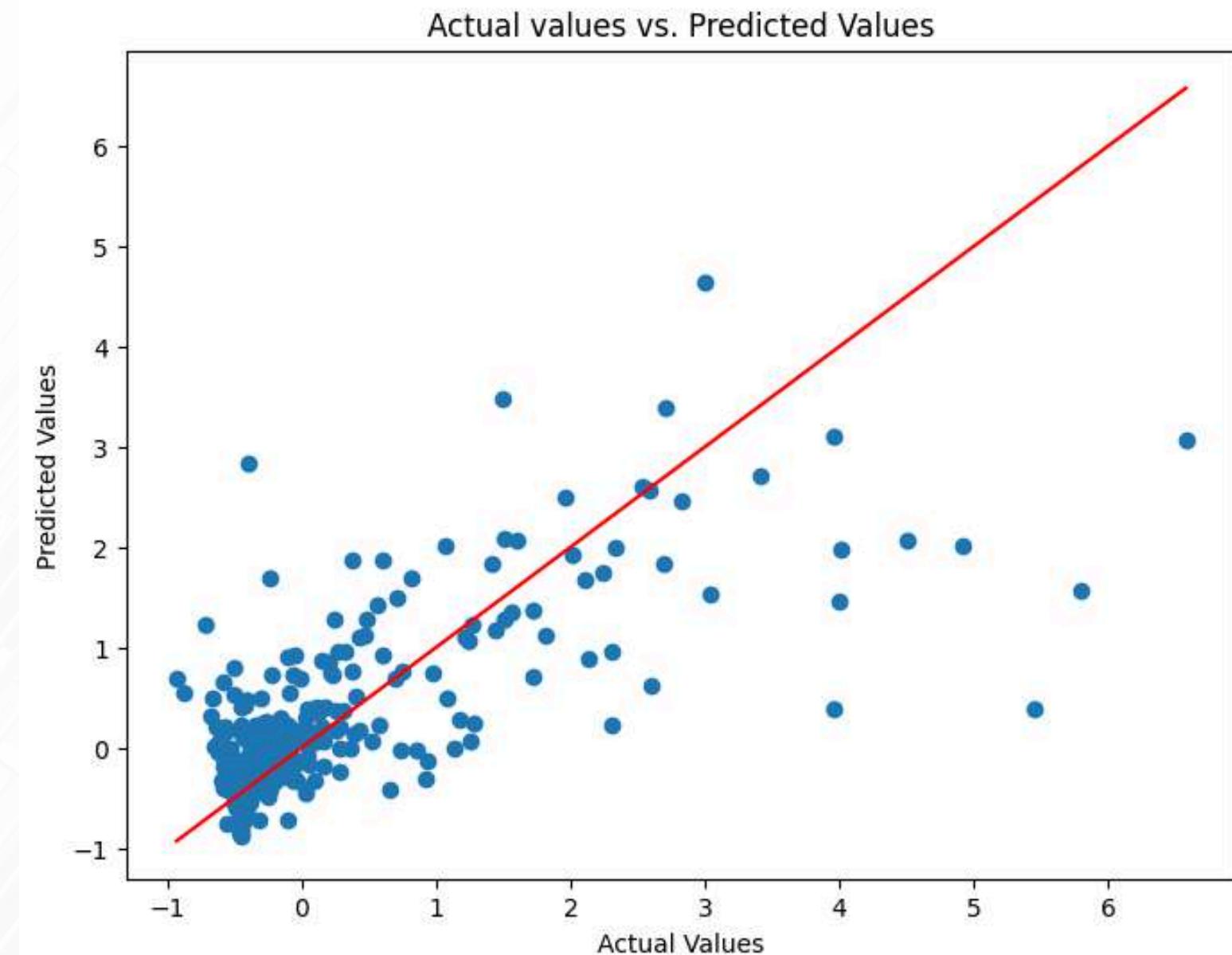


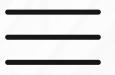
MODELING

- **Regression Analysis:** Examined the impact of Production Budget and Popularity on Worldwide Profit.
 - Positive correlation found between popularity scores and worldwide profits.
 - Independent Variables: Production Budget and Popularity.
 - Dependent Variable: Worldwide Profit.
- **Model Validation:** Assessed relationships through scatter plots and linear regression, confirming the importance of budget and popularity in predicting box office success.

MODEL ANALYSIS

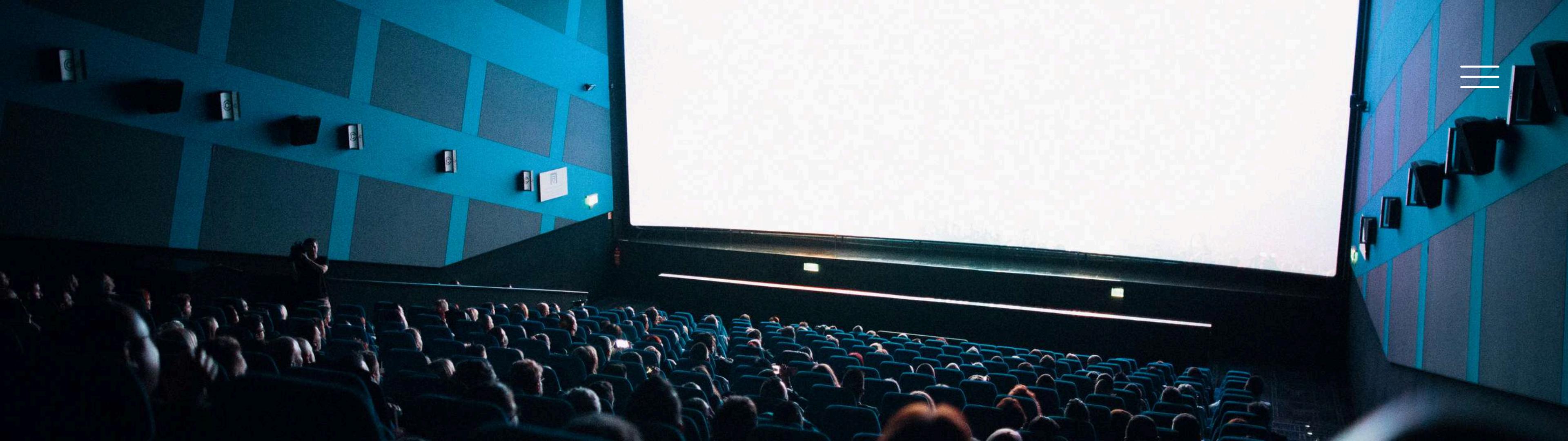
The scatter plot shows the relationship between actual and predicted values, with points clustering along the red diagonal line, indicating that the model performs well when points are closer to this line, although some variability and outliers suggest areas for improvement in prediction accuracy.





FINDINGS

- **Key Insights:**
 - High-budget movies tend to earn significantly in domestic markets and even more in foreign markets.
 - Pre-release popularity boosts worldwide earnings, indicating marketing is crucial.
 - Drama and Documentary genres consistently receive higher ratings, making them safer choices.
- **Profitability Trends:** Investments in production budget directly increase profitability due to strong box office returns, particularly in high-budget films.



CONCLUSION

- The analysis identifies three core drivers of success: budget, popularity, and genre.
 - **Budget and Popularity:** High-budget, high-popularity films tend to achieve the most substantial worldwide profits.
 - **Genre:** Prioritizing high-rated genres like Drama and Documentary aligns with audience preferences.



RECOMMENDATIONS

- **Focus on Popular Genres:** Allocate resources towards producing high-rated genres like Drama and Documentary.
- **Invest in High Budgets:** Prioritize larger budgets to increase domestic and worldwide returns.
- **Boost Marketing for Popularity:** Invest in pre-release marketing to drive up popularity, which correlates strongly with box office performance.

Thank
YOU

