What Drives Box Office Hits?

Annie Cen, Rohan Saklani, Jasmine Chu, Alexis Adzich, Cassia Ramelb, Kevin Hamakawa

UCLA Department of Statistics and Data Science

Objectives

- Identify key factors contributing to box office success (high gross revenue).
- Examine the impact of actor influence and movie genre on revenue by testing relevant hypotheses.
- Provide actionable insights for studios and stakeholders.

We used a dataset with the following characteristics:

- **Source**: Dataset with top 1000 movies (IMDB)
- **Key Features**: Released_Year, Certificate, Runtime, Genre, IMDB_Rating, Overview, Meta_score, Director, Star1,Star2, Star3, Star4, Decade, No_of_votes, Gross, Sentiment

Preprocessing Steps

We performed several preprocessing steps before analyzing our dataset:

1. Cleaned missing values in revenue and budget.
2. Categorized genres into primary classifications.
3. Standardized diversity scores for analysis.
4. Added Sentiment Column
5. Removed poster link column

These steps allow us to ensure the dataset is clean, structured, and optimized for meaningful analysis, improving the reliability and interpretability of our results.

Descriptive Statistics

Key insights from our data exploration include:

- **Mean Domestic Box Office Gross Revenue:** $60,513,599
- **Max Domestic Box Office Gross Revenue:** $936,662,225 (Star Wars: Episode VII - The Force Awakens)
- **Mean IMDB Rating:** 7.95
- **Max IMDB Rating:** 9.3 (The Shawshank Redemption)
- **Mean Runtime:** 122.89 minutes (~2 hours)
- **Actor in most movies:** Robert De Niro

Testable Hypotheses

Popular Actors and Revenue:

- **H_0:** Revenue is consistent across popular actors (appeared in 10+ films)
- **H_a:** There is a significant difference in revenue across top actors.

Actor-Pairs Influence on Revenue:

- **H_0:** Revenue is consistent across top actor-pairs.
- **H_a:** There is a significant difference in revenue across top actor-pairs.

Exploratory Data Analysis

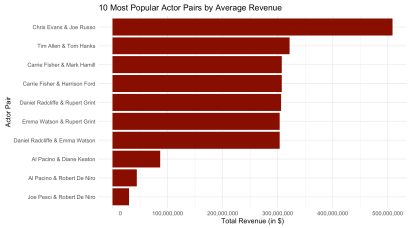

Figure 3: Top 10 Actor Pairs by Avg Revenue

Conclusion

- Our statistical tests gave us the conclusion that the inclusion of **popular actors** in a film make a statistically significant impact on the film's overall **revenue**.
- Furthermore, it is important for movie studios and directors to take **pairs of actors** into consideration, and understanding the pair's historical success in the box office.
- Last, the development of franchises and the continuity of similar clusters of actors can be a contributing factor in a movie's success.

Introduction

Understanding what drives box office is critical for decision-making in the film industry. By analyzing factors like genre, budget, star power, and runtime, we aim to discover patterns that distinguish successful films. This project combines statistical analysis with data visualization to explore these relationships.

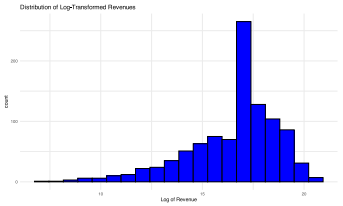

Figure 1: Distribution of Log-Transformed Revenue

Methods

- **Hypothesis Testing:** A grouped ANOVA will be used to assess differences in revenue across multiple groups (e.g., genre, director) to determine if these factors significantly impact revenue.
- **Sentiment Analysis:** Analyze the influence of star power and movie overviews on revenue.
- **Confounding Factors:** Consider and control for confounding variables that may influence the relationship between predictors and revenue.

Results

We conducted a **grouped ANOVA** to investigate whether there are statistically significant differences in revenues among the actors. Our p-value came out to be .00000335, indicating that we can reject our null hypothesis of revenue being consistent across popular actors.

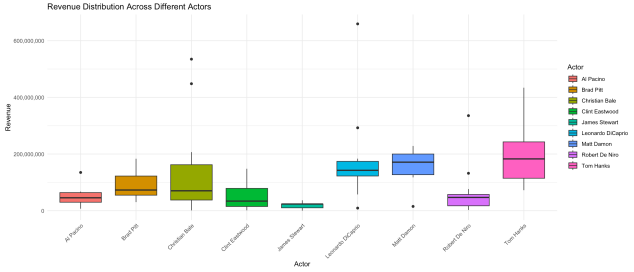

Figure 2: Revenue Distribution Across Actors

This box-plot highlights some of the key differences in distribution between some of the top actors. It indicates which top stars to potentially target when trying to achieve the highest revenue.

References

- Kaggle: IMDB Dataset of Top 1000 Movies and TV Shows
- IMDB Official Website