

















 **0** stars  **0** forks  **1** watching  **1** Branch  **0** Tags  Activity

 Public repository

<div>  George-Chira add readme <div>30c7bf8 · 2 minutes ago</div>  </div>		
 .gitattributes	Update .gitattributes	1 hour ago
 .gitignore	Add .gitignore to exclude imdb dire...	10 hours ago
 Github.pdf	Remove new-repo directory and sta...	16 minutes ago
 Movie Studio Project .pdf	Track .db files with Git LFS	13 hours ago
 Notebook.pdf	pdf uploaded	10 hours ago
 Presentation ppt.pdf	pdf uploaded	10 hours ago
 Project.ipynb	final	14 hours ago
 README.md	add readme	2 minutes ago
 bom.movie_gross.csv	exploration	2 days ago
 tn.movie_budgets.csv	exploration	2 days ago

<https://github.com/George-Chira/phase-one-project>

- [Feature Engineering](#)
- [Visualization](#)
- [Key Findings](#)
- [Recommendations](#)
- [Conclusion](#)
- [Appendices](#)

Introduction

This analysis focuses on identifying key factors that contribute to box office success. Data was sourced from IMDB, Box Office Mojo, and TheMovieDB to provide a robust foundation for insights.

Data Sources and Justification

- **IMDB:** Comprehensive metadata including movie titles, genres, cast, crew, and user ratings.
- **Box Office Mojo:** Detailed box office revenue data, essential for understanding financial performance.
- **TheMovieDB:** Extensive information on movies, including production budgets and popularity metrics.

Data Collection and Preparation

- **Data Collection:** Data was gathered from the aforementioned sources.
- **Data Cleaning:** Missing values were handled, data formats standardized, and datasets merged.
- **Handling Missing Values:** Different strategies were employed, such as filling with median values.
- **Standardizing Data Formats:** Dates and currencies were standardized.
- **Merging Datasets:** Combined on common keys like movie titles and release dates.

Data Exploration

Exploratory data analysis (EDA) was conducted to understand data distributions and relationships using summary statistics and visualizations.

Feature Engineering

New features were created to enhance predictive power, including calculating ROI for each movie and profitability per genre.

Visualization

Visualizations such as bar charts, scatter plots, and heatmaps were used to present findings clearly.

Key Findings

- **Top-Performing Genres:** Animation, Adventure, and Sci-Fi are top genres in terms of earnings.
- **Returns on Investment:** Horror has the highest median ROI but with significant variability.
- **Optimal Budget Range:** \$50M to \$150M budget range tends to yield the best results.
- **Importance of Ratings:** High movie ratings correlate with box office success.

Recommendations

- **Investment Strategy:** Focus on high-performing genres like Animation and Sci-Fi.
- **Risk Management:** Manage risks by acknowledging genre variability.
- **Diversification:** Maintain a mix of high-risk/high-reward and stable genres.
- **ROI Optimization:** Prioritize high ROI genres and optimal budget ranges.
- **Project Selection Criteria:** Use high ratings as key criteria for project selection.

Conclusion

By implementing these recommendations, Microsoft's movie studio can strategically position itself for success, leveraging data-driven insights to navigate the competitive movie industry.

Appendices

Additional details and supporting information can be found in the appendices.



Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

