Stella-Kiarie / **Group-5-Project**

Code Issues Pull requests Actions Projects Wiki Security Insights Settin

Movie Success Analysis Project.

View license

Contributing

0 stars 3 forks 0 watching Branches Activity
Tags

Public repository

5 Branches 0 Tags Go to file t Go to file Add file Code

Stella-Kiarie Merge pull request #21 from Stella-Kiarie/Stella 8797aad · 4 minutes ago

images	Updates	yesterday
zippedData	Added Files and ReadMe	2 days ago
.gitattributes	Added im.db using Git LFS	2 days ago
.gitignore	ignore anaconda_projects folder	7 hours ago
CONTRIBUTING.md	added files and readme	2 days ago
CRISP DM - Group 5.pdf	Updates	5 hours ago
LICENSE.md	add files and readme	2 days ago
README.md	Updated README.md	5 minutes ago
cleaned_movie_data.csv	updated notebook	2 days ago
index.ipynb	updated dataset in notebook	2 days ago
movie_data_erd.jpeg	add files and readme	2 days ago
presentation.pdf	updated changes	6 hours ago
student.ipynb	updated notebook	9 hours ago

README Contributing License

MOVIE SUCCESS ANALYSIS PROJECT

OVERVIEW

This project explores the entertainment industry as the industry has been evolving rapidly as major plays aim to produce original contents for streaming Platforms. In the project, we leveraged movie revenue data from Box office Mojo and IMDB database to analyze the movie ratings and characteristics. The objective of this project is to conduct a comprehensive exploratory data analysis to identify patterns, relationships and factors contributing towards a movie success within the entertainment industry. The insights and findings gathered in this project can be translated to actionable recommendations for companies within the industry.

Business Understanding

A company plans to launch a new movie studio but has no information on the performance of various movies at the box office. For the company to remain competitive, the company needs an understanding on best movie performances to make informed investment decisions. This project will explore historical data such as movie ratings, movie revenue and movie characteristics to guide the company on which movie studio to invest in and the type of films to produce.

Project Goal

Goal: To identify the key factors that drive movie box office success and to provide data driven insights to guide in the new movie studio and production venture.

Business Questions

1. Which are the highest performing movie genres?
2. What is the impact of audience ratings on box office performance?
3. What are the key factors that determine movie success? factors such as runtime, genre and release timing
4. How have the box office performance changed over the years?

Data Understanding and Analysis

To achieve the goal of providing data driven insights on a new movie studio venture, two datasets were used in the project.

1. bom.movie_gross.csv.gz — Box Office Mojo data with revenue information.i.e domestic and foreign gross
2. im.db — IMDB SQLite database containing movie information such as titles, genres, and ratings.

The two datasets are relevant in the project as :

1. BOM has revenue information for which revenue is one of our target variable
2. IMDB contains movie attributes which are more important for our objectives

domestic_gross , foreign gross --------- help determine movie earnings

average ratings, numvotes ---------- Determine the impact audience response on box office movie success runtime_minutes, genre, studion ----------insights on the production features

First, the IMBD database contains multiple tables, however, only the movie_basics and movie_ratings contained relevant data for our analysis. The data on movie_ratings was good as did not contain duplicates or missing values but the movie_basics contained missing values which reqiuired cleaning by dropping rows or imputing . On the other hand, the Bom revenue data was messy and required cleaning such as dealing with missing values and duplicates.

Eventually, we created a merge between the bom revenue dataset and the imbd datasets to generated a cleaned dataset for our analysis. The merge generated a data with 3109 rows and 13 columns.

Hypothesis Testing

To determine whether there is a significant relationship between between movie genres and box office success, we conducted a Chi-Square Test of independence.

Null Hypothesis (H_0): There is no relationship between movie genre and box office success.

Alternative Hypothesis (H_1): There is a relationship between movie genre and box office succes

The results of the Chi-Square are shown below: Chi-Square Test Results

Chi2 Statistic: 736.6494756816546

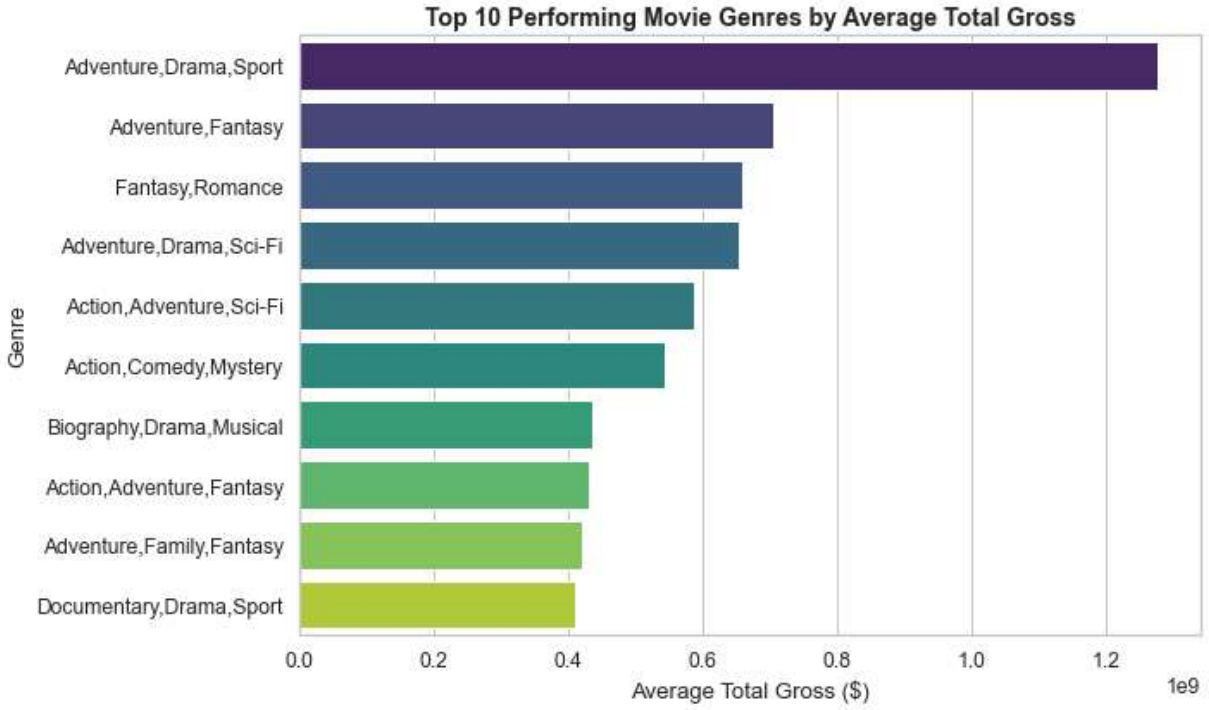
p-value: 1.0623260172850567e-34

Degrees of Freedom: 321

From the results, the Chi-Square-Test has a statistic of 736.6 and a p- value of 1 0623260172850567e-34 which is below the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is a significant relationship between movie genre and box office success.

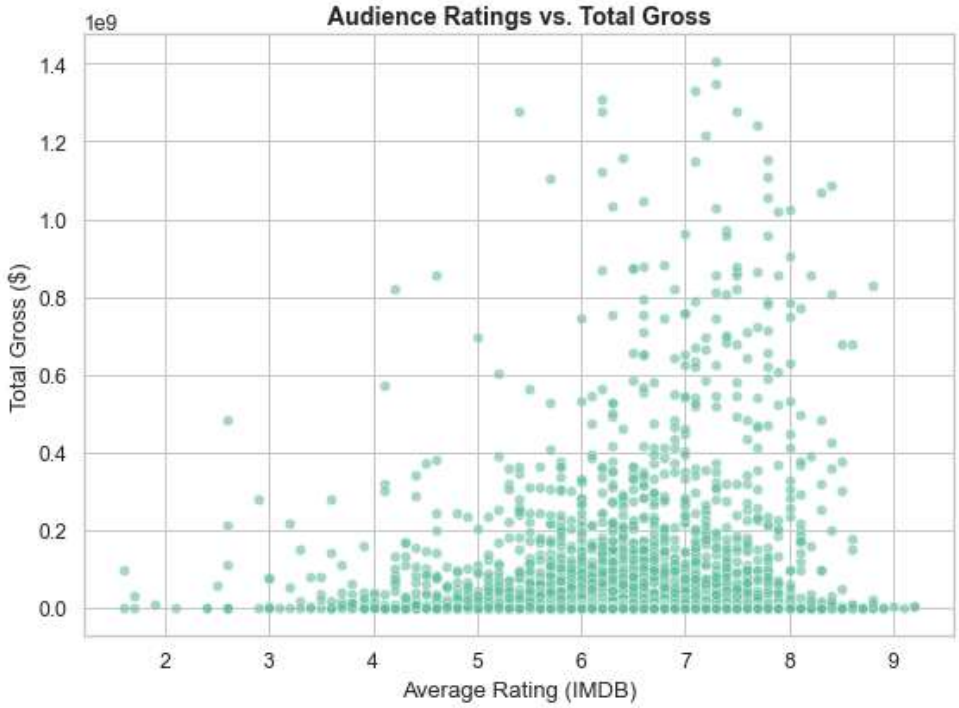
Visualizations

1. Top Performing Genres



From the visualization above, Adventure and Action-based genres dominate box office performance, especially when combined with elements of Drama, Fantasy, or Sci-Fi.

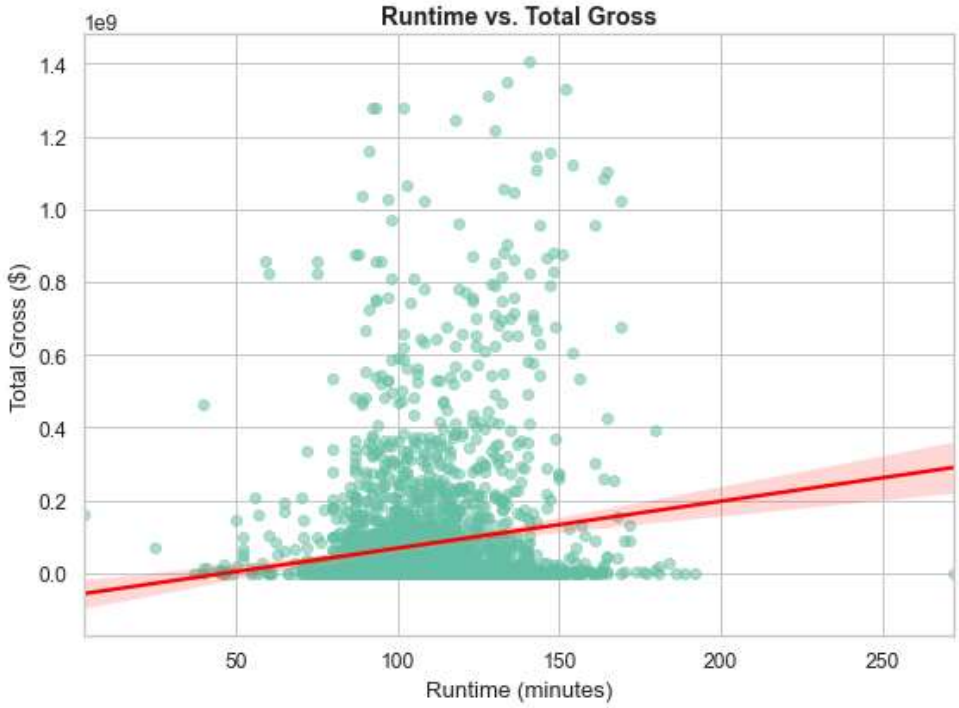
2. Ratings as a factor in Box Office success



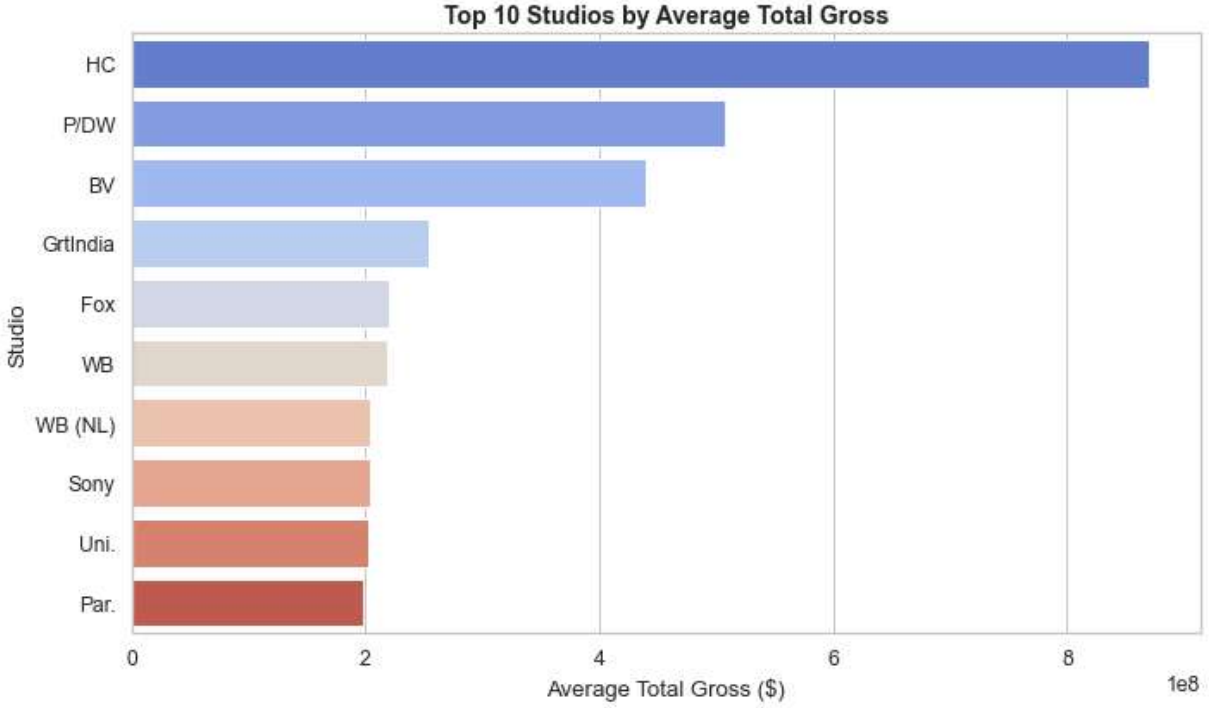
The scatter plot shows a positive relationship between audience ratings and total box office gross. Movies with ratings between 6 and 9 on IMDB tend to earn significantly higher revenues, while poorly rated films (below 5.0) rarely achieve major financial success. This indicates that audience satisfaction and perceived quality strongly influence a movie's commercial performance films that resonate well with viewers are more likely to generate higher box office returns.

3. Runtime as a factor in Box office success

The visualization below reveals a slight positive relationship between runtime and total box office gross. Movies with runtimes between 90 and 130 minutes generally achieve higher earnings, suggesting that audiences prefer films that are long enough to develop a story but not excessively lengthy. Extremely short or overly long films tend to earn less, indicating that finding an optimal runtime balance is key to maximizing box office performance.



4. Top performing Studios



The visualizations shows that a few major studios dominate the box office market. Studios such as HC, P/DW, and BV achieve the highest average total gross, far outperforming others. This suggests that established studios benefit from larger production budgets, strong marketing strategies, and brand loyalty. Meanwhile, smaller studios like Sony, Universal, and Paramount still maintain consistent performance but on a smaller scale. Overall, studio reputation and financial capacity play a significant role in driving box office success

Here is a link to the tableau visualization:
https://public.tableau.com/app/profile/kumati.dapash5591/viz/MoviesSuccessAnalysis/MovieSuccessAnalysis?publish=yes

Conclusion

To conclude, exploratory data analysis reveals that movie success is largely driven by genre choice, audience ratings, and production strategy. High-grossing films typically belong to popular hybrid genres, are well-rated by audiences, and have optimal runtimes. These findings suggest that creativity, quality, and data-driven decision-making should guide the company's new movie production strategy.

Recommendations

Following the critical analysis on Box office Mojo and IMDB, the following recommendatio would be best strategies to plan when venturing into the movie studio business:

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 4

- Stella-Kiarie Stella_kiarie
- Morvine-otieno
- Dapash7397-star
- mutiedoris

Languages

- **Jupyter Notebook** 100.0%