

apelicier / Phase-2-Project

<> Code

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

Your repository details have been saved.

Use this template

Movie studio data analysis

GPL-3.0 license

0 stars 9 forks 0 watching Branches Activity

Tags

Public template repository · Forked from [learn-co-curriculum/DS-Phase-2-Project-Template](#)

1 Branch 0 Tags Go to file

t

Go to file

+

Add file

Code

This branch is **10 commits ahead of** learn-co-curriculum/DS-Phase-2-Project-Template:main .

Contribute

Sync fork

apelicier Update README.md

ec86abc · 6 minutes ago

Data	template	2 years ago
Images	third commit	1 hour ago
Notebooks	third commit	1 hour ago
Presentation	second commit	1 hour ago
zippedData	third commit	1 hour ago
.gitignore	template	2 years ago
Final.html	third commit	1 hour ago
Final.ipynb	third commit	1 hour ago
LICENSE	Initial commit	2 years ago
README.md	Update README.md	6 minutes ago
notebook.pdf	third commit	1 hour ago

README

License

Movie Studio Data Analysis

Project Overview

This project focuses on providing data-driven insights to a new movie studio, guiding their content creation strategy to maximize box office success. By analyzing historical movie data, we aim to identify key characteristics of films that perform well financially.

Business Problem

My company is venturing into original video content with a new movie studio but lacks experience in film production. The objective is to explore current box office trends to understand what types of films are most successful and translate these findings into actionable recommendations for the studio head.

Data Sources

The analysis leverages data from two primary sources:

[Box Office Mojo](#): Provides domestic and foreign gross revenue figures for movies.

[IMDB](#): A SQLite database containing movie metadata such as titles, release years, genres, average ratings, and number of votes. (Specifically, movie_basics and movie_ratings tables are used).

Methodology

Data Acquisition: Loaded Box Office Mojo data from a compressed CSV and extracted/connected to the IMDB SQLite database.

Data Preparation:

- Merged IMDB's movie_basics and movie_ratings tables on movie_id.
- Cleaned and converted gross revenue columns to numeric types, handling missing values by filling with zero.
- Cleaned movie titles for merging across datasets by converting to lowercase and removing non-alphanumeric characters.
- Merged the cleaned Box Office Mojo data with the combined IMDB data, primarily on cleaned movie titles.
- Handled missing genres and filtered out movies with zero total gross.

Exploratory Data Analysis (EDA):

- Analyzed average worldwide gross by genre to identify top-performing categories.
- Investigated the relationship between audience ratings, number of votes, and total gross, focusing on movies with a significant number of votes.

- Examined correlations between key financial and audience metrics.

Business Recommendations:

Formulated three concrete, data-backed recommendations for the movie studio based on the analytical findings.

Key Findings & Recommendations

My analysis revealed several critical insights:

Genre Dominance: Adventure, Action, and Sci-Fi genres consistently show the highest average worldwide gross.

Recommendation 1:

Prioritize the development of films within Adventure, Action, and Sci-Fi genres to maximize box office potential.

Audience Engagement is Key: A high number of audience votes (indicating broad engagement) is a strong predictor of higher worldwide gross, even more so than just a high average rating.

Recommendation 2:

Invest in films that can generate widespread interest and encourage broad audience participation and discussion.

Broad Appeal Wins: The highest-grossing films often resonate with a very wide audience, suggesting the importance of universal themes.

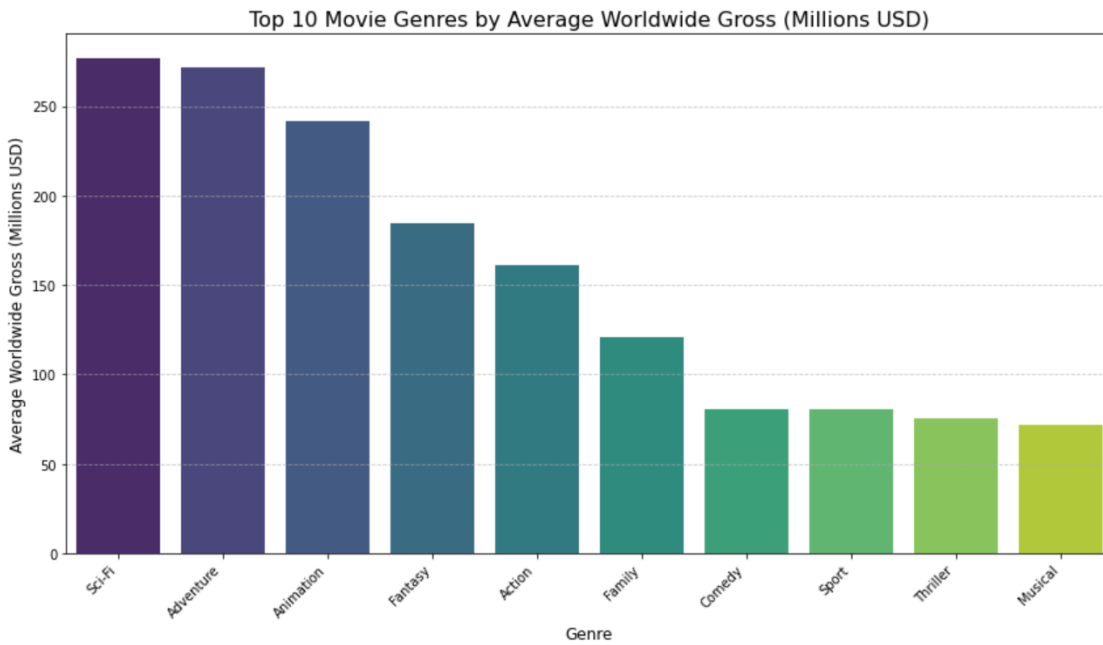
Recommendation 3:

Focus on producing movies with broad audience appeal and themes that can attract a diverse and large demographic.

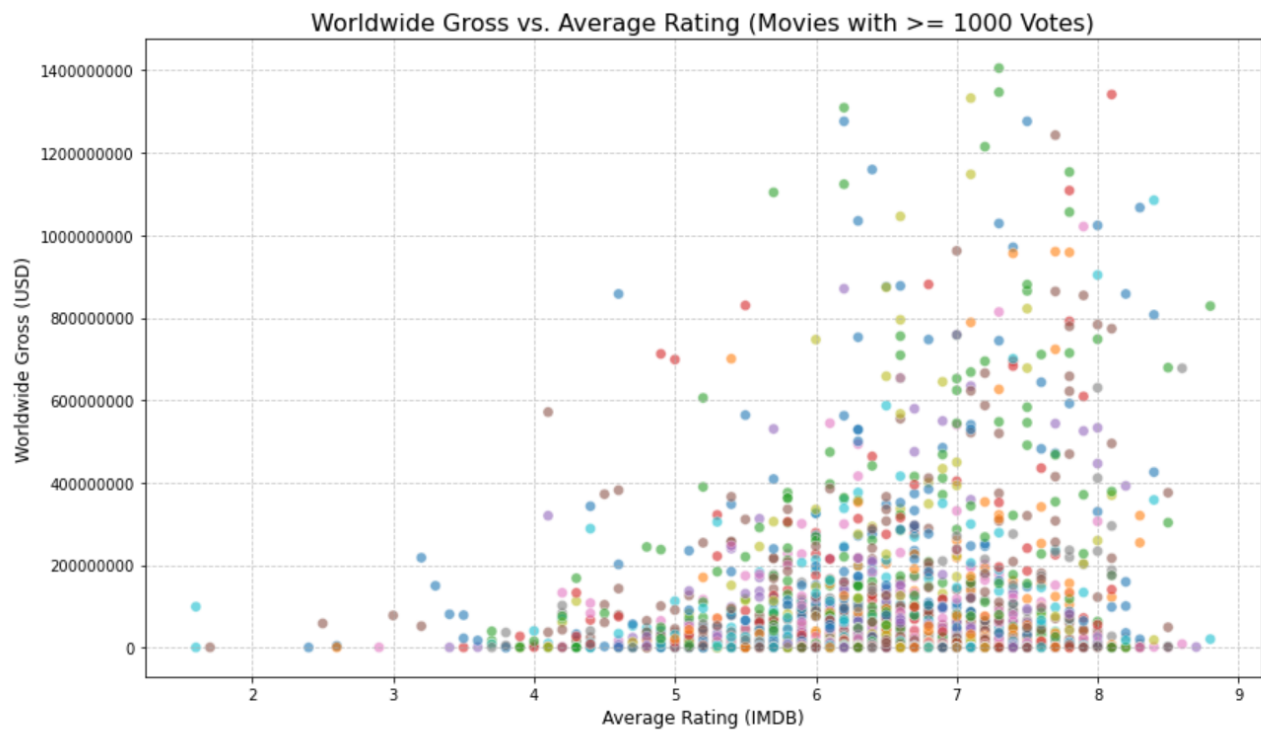
Visualizations

The project includes several visualizations to support the findings:

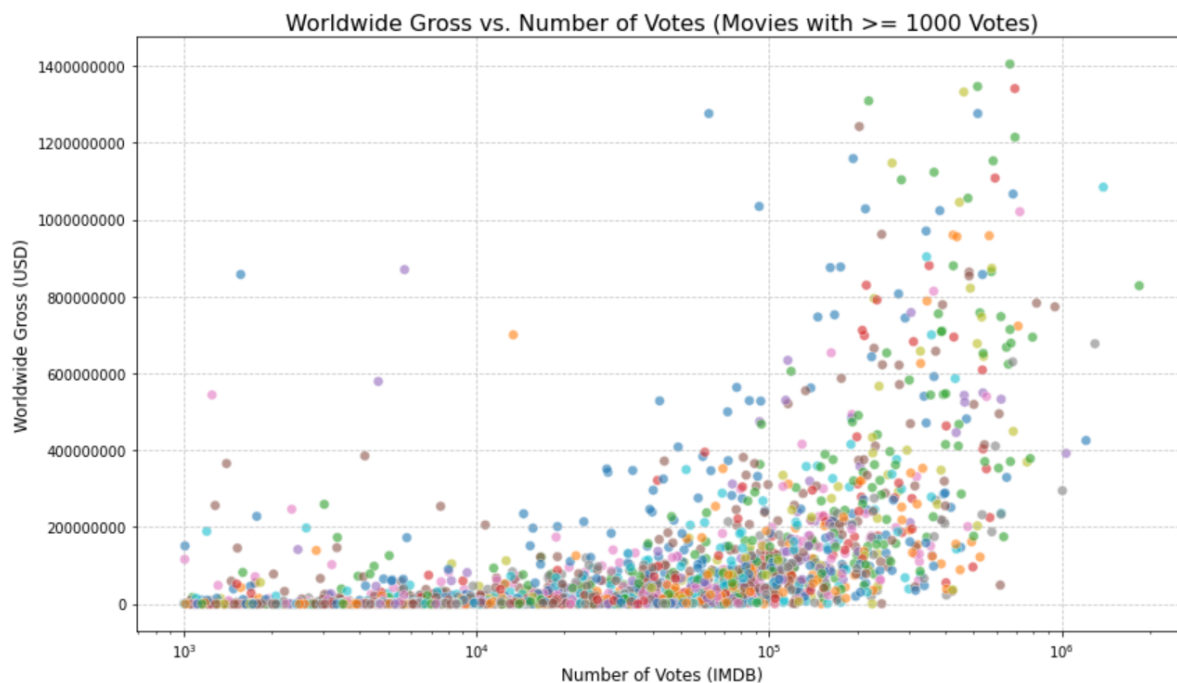
Bar Chart: [Top 10 Movie Genres by Average Worldwide Gross\(.PNG\)](#).



Scatter Plot: [Worldwide Gross vs. Average Rating\(.PNG\)](#).

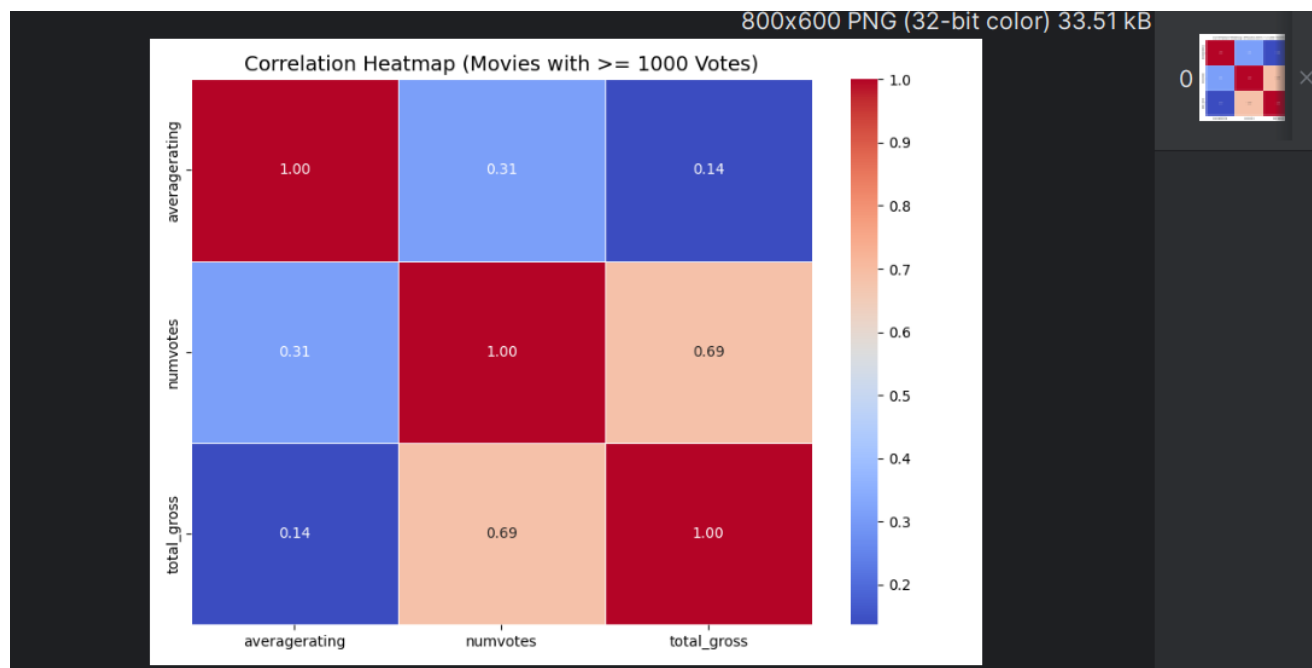


Scatter Plot: [Worldwide Gross vs. Number of Votes\(.PNG\)](#).



Visualization 3: Worldwide Gross vs. Number of Votes generated.

Heatmap: [Correlation between Average Rating, Number of Votes, and Total Gross\(.PNG\)](#).



Deliverables

Jupyter Notebook: [Final \(.ipynb\)](#) (and its PDF export).

Jupyter Notebook_pdf: [notebook \(.pdf\)](#) Pdf notebook

Non-technical Presentation: [presentation \(.pdf\)](#) (in pdf format).

GitHub Repository: [This repository itself](#):

GitHub Repository_pdf: [github \(.pdf\)](#)

Next Steps for the Studio

- Conduct deeper market research within identified top genres.
- Develop concepts that prioritize broad audience appeal and potential for high engagement.
- Strategize talent acquisition and marketing efforts to align with these findings.

Contact

[PELICIER Antonine]

[antoninepelicier10@gmail.com]



Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● HTML 70.2% ● Jupyter Notebook 29.8%

Suggested workflows

Based on your tech stack



SLSA Generic generator

Generate SLSA3 provenance for your existing release workflows

Configure



Jekyll using Docker image

Package a Jekyll site using the jekyll/builder Docker image.

Configure

[More workflows](#)

[Dismiss suggestions](#)