

FILM INDUSTRY INSIGHTS: DATA ANALYSIS SUMMARY

PROJECT 1



PROJECT OBJECTIVE

- To analyze and merge datasets from IMDB, Rotten Tomatoes and The Numbers to uncover movie patterns in movie success, both critically and financially.
- The findings need to be translated into actionable insights.

DATA SOURCES

- IMBD- This contains user ratings, genres and release years of films and movies.
- Rotten Tomatoes- Contains critic/audience scores.
- The Numbers- Has box office revenue and studio information.

The Data

In the folder `zippedData` are movie datasets from:

- [Box Office Mojo](#)
- [IMDB](#)
- [Rotten Tomatoes](#)
- [TheMovieDB](#)
- [The Numbers](#)

METHODOLOGY

- Structured workflow using CRISP-DM for reliability.
- Data cleaning: Dealing with missing values, merging conflicts and datasets and also column standardization.
- Visualizations with matplotlib, seaborn and pandas plots.

KEY INSIGHTS

- Independent studios show niche success despite lower budgets.
- The financial status both domestic and foreign differ depending on the genres and also studio.

PRESENTATION STYLE

- Clear graphs for rating comparisons, revenue by genre, and studio performance.
- Documented code and modular functions for transparency.
- Summary findings in code for easier understanding of the visuals used.

BUSINESS VALUE

This project equips film industry stakeholders with data-backed recommendations to:

- Prioritize genres for investment.
- Align marketing strategy with audience trends.
- Benchmark performance against competitors.