

tedronoh-14 / movie-studio-insights

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

A comprehensive data analysis project that extracts actionable insights from movie industry datasets (IMDB, Box Office Mojo, etc.) to guide strategic decisions for launching a new movie studio. The analysis explores genres, budgets, runtimes, ratings, and box office performance to identify key trends and recommend data-driven production strategies.

1 star

3 forks

1 watching

Branches

Activity

Tags

Public repository

1 Branch

0 Tags

Go to file

t

Go to file

+

Add file

Code

tedronoh-14 Update README.md4348e5f · 5 days ago🕒

README.md	Update README.md	5 days ago
bom.movie_gross.csv	Add files via upload	5 days ago
rt.movie_info.tsv	Add files via upload	5 days ago
rt.reviews.tsv	Add files via upload	5 days ago
studio_insights.ipynb	Add files via upload	5 days ago
tmdb.movies.csv	Add files via upload	5 days ago
tn.movie_budgets.csv	Add files via upload	5 days ago

README

Studio Insights

Project Overview

Our company is keen on investing in the movie industry. We have decided to create a new movie studio but do not know anything about creating movies. We are tasked with exploring the types of films currently doing the best at the box office. We must then translate those findings into actionable insights that the head of our company's new movie studio can use to help decide what type of films to create.

Project Goals

- Examine historical box office performance across various genres, budgets, revenues and release dates.
- Identify key trends that contribute to a movie’s commercial success.
- Recommend data-driven strategies to guide the creation and marketing of new films.

Why This Project Matters

- **Competitive Landscape:** Major streaming platforms and traditional studios are heavily investing in original content.
- **High Stakes:** Movie production requires significant capital investment; data-driven decisions can mitigate financial risks.
- **Opportunity:** By analyzing what works (and what doesn't), we tailored film production choices to meet audience demand.

Business Understanding

Stakeholder

- **Primary Stakeholder:** The Head of the New Movie Studio at our company
 - **Role:** Oversees film production, marketing, and distribution strategies.
 - **Interest:** Seeks actionable insights on which film projects to greenlight and how to position them for maximum box office returns.

Business Questions

1. Which Genres Are Most Profitable?

- Did certain genres consistently yield higher box office revenues or returns on investment (ROI)?
- Which genres showed growth or decline in recent years?

2. What Budget Ranges Correlate with Success?

- How did production budgets relate to box office returns?
- Was there an optimal budget range that maximized ROI?

3. Which Time of Year Was Best for Movie Releases?

- Were certain months or seasons historically more profitable?

Data Understanding

Data Sources

1. Box Office Mojo (CSV)

- **File:** bom.movie_gross.csv.gz
- **Columns (typical):**
 - title : Movie title.
 - studio : Studio that produced or distributed the film.
 - domestic_gross : Revenue from the domestic market (US/Canada).
 - foreign_gross : Revenue from international markets.
 - year : Release year.

2. Movie budgets (CSV)

Data Descriptions

- movie_basics

- `movie_id` : Unique identifier (often `tmdb_id` in IMDB's original data).
 - `primary_title` : The main title of the movie.
 - `original_title` : Original title if different from the main one.
 - `start_year` : Year the movie was first released or produced.
 - `runtime_minutes` : Length of the movie in minutes.
 - `genres` : One or more genres (e.g., "Action, Comedy").
- **Other Tables (e.g., `principals`, `directors`, `writers`)**
 - Provided additional details about cast and crew (optional for a basic analysis).
 - **`bom.movie_gross.csv.gz`**
 - `title` : Movie title.
 - `studio` : Studio that produced or distributed the film.
 - `domestic_gross` : Revenue from the domestic market (US/Canada).
 - `foreign_gross` : Revenue from international markets.
 - `year` : Release year.

Data Transformation

Formatting Worldwide Gross for Readability

To make the revenue figures more interpretable, we converted the `release_date` to a datetime object, extracted the month, grouped the data by month, and transformed the `worldwide_gross` figures from raw numbers into millions. This makes the data easier to understand at a glance (e.g., displaying \$50M instead of 50000000).

Below is the code snippet demonstrating this transformation:

```
# Convert release_date to datetime and extract month
movie_budgets['release_date'] = pd.to_datetime(movie_budgets['release_date'])
movie_budgets['release_month'] = movie_budgets['release_date'].dt.month

# Group by month and calculate mean worldwide gross
monthly_revenue = movie_budgets.groupby('release_month')['worldwide_gross'].mean().reset_index()

# Convert worldwide gross to millions for readability
monthly_revenue['worldwide_gross_millions'] = monthly_revenue['worldwide_gross'] / 1e6

# Display the transformed data
print(monthly_revenue)
```



Data Limitations

- **Data Gaps:**
Some movies did not have complete information (e.g., missing budgets or missing foreign gross).
- **Merge Complexity:**
The IMDB database used unique IDs, while Box Office Mojo used movie titles. Matching them required careful joining on title and year.
- **Outliers:**
Certain blockbuster films (e.g., *Avengers*) or extremely low-budget films might have skewed averages.

- **Time Range:**

We ensured that the years in each dataset overlapped; otherwise, we might have been comparing different time periods.

Data Preparation Strategy

Merging

- **Merge Approach:**

We merged the IMDB data with Box Office Mojo on movie title and release year (using an approximate matching technique when necessary).

In cases where merging proved too challenging or resulted in a limited subset of data, we maintained the datasets separately.

Cleaning

- **Handling Missing Values:**

We removed or imputed missing values where appropriate.

- **Data Type Conversion:**

We converted currency columns to numeric types for accurate analysis.

- **Genre Processing:**

We decided whether to split genres into multiple rows for more granular analysis or keep them as a single string based on the specific requirements of our analysis.

Feature Engineering

- **Profit/ROI Calculation:**

We created a `profit` or `ROI` column when budget data was available.

- **Primary Genre Extraction:**

We extracted the primary genre when multiple genres were listed (e.g., "Action" from "Action,Comedy").

EXPLORATORY DATA ANALYSIS

Genre Profitability

- Action and Sci-Fi genres dominate in profitability, generating the highest average box office returns.
- These genres tend to have strong international appeal, contributing significantly to foreign gross revenue.

Budget vs. Box Office Performance

- High-budget films generally achieve better returns, but only when supported by strategic marketing and distribution.
- While low-budget films can be profitable, their success rate is lower, and they often rely on niche audiences or viral marketing.

Foreign Market Considerations

- Movies with global appeal tend to perform better financially, with foreign box office earnings contributing a major share of total revenue.
- Films with minimal dialogue and strong visual storytelling (e.g., action-packed sequences) often succeed internationally.

Release Timing Impact

- Movies released during peak seasons (holidays, summer) see significantly higher box office earnings.
- Avoiding release dates that coincide with major blockbusters can improve a film’s performance by reducing direct competition.

Recommendations

Focus on High-Performing Genres

- The studio should prioritize Action and Sci-Fi films, as they offer the best return on investment.
- Consider hybrid genres (e.g., Action-Thriller, Sci-Fi-Adventure) to attract wider audiences.

Budget Optimization

- Allocate sufficient funds for both production and marketing, as both are critical for box office success.
- If producing lower-budget films, ensure a strong storyline, unique marketing, or a dedicated fanbase to boost profitability.

Global Market Strategy

- Invest in projects with cross-cultural appeal to maximize foreign revenue.
- Consider international partnerships for distribution to tap into foreign markets.

Strategic Release Scheduling

- Schedule releases during peak seasons (summer, holiday periods) for maximum audience turnout.
- Avoid competing with major franchise films unless the movie has a unique selling proposition.

By following these insights and recommendations, the new movie studio can improve its chances of creating successful, high-grossing films while minimizing financial risks.



Releases

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

Packages

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

Languages

● Jupyter Notebook 100.0%