

veenaledi / DS-Phase-2-Project

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

☆ 0 stars

🍴 0 forks

👁 0 watching

🔗 Branches

📈 Activity

🏷 Tags

🌐 Public repository

1 Branch

0 Tags







Go to file

Go to file



+

Add file

Code

 Your Name	Final Changes	2247ca4 · 2 hours ago	
	.gitignore	First commit	2 days ago
	Merged.csv	First commit	2 days ago
	Movies.ipynb	Fixed Pending Errors	2 hours ago
	README.md	Final Changes	2 hours ago

README



Film Making Entry Analysis Report

Author

Neema Naledi

Overview

This project analyzes box office data to provide strategic insights for a new movie studio entering the film production market. The analysis examines financial performance, genre trends, and production budgets to identify the most profitable opportunities in the film industry.

- Tableau Dashboard:
[\[https://public.tableau.com/app/profile/naledi.ambuku/viz/FilmMakingEntryAnalysisDashboard/TheDashboard?publish=yes\]](https://public.tableau.com/app/profile/naledi.ambuku/viz/FilmMakingEntryAnalysisDashboard/TheDashboard?publish=yes)

Business Understanding

The Problem: The company is establishing a new movie studio to expand into original content production. An analysis of current box office trends is underway to identify high-performing film genres and translate these findings into strategic creative direction.

Objectives: The primary business goal is to maximize profitability while minimizing financial risk in the highly competitive film industry.

Success Criteria: Determine optimal budget allocation strategies. Achieve 15-20% profit margins on initial film releases.

Data Understanding

Data sets used:[tmdb.movies.csv.gz] - Contains movie data including genres, release dates, and ratings
[tn.movie_budgets.csv.gz] - Contains production budgets and worldwide gross revenue data

Reasons for data selection:

- Valid production budget values
- Complete budget and revenue information
- Release dates within analysis timeframe

Data Description

1. TMDB Dataset:

- Records: ~26,000 movies
- Key Variables: title, genres, release_date, popularity, vote_average
- Time Period: 1960-2020

2. TN Dataset:

- Records: ~5,000 movies
- Key Variables: movie, production_budget, domestic_gross, worldwide_gross
- Time Period: 2010-2019

Outliers:

- Present in budget and revenue data (blockbuster films)
- I decided to keep Outliers since they are a vital part of the analysis.

Merged Data:

- Data was merged using inner join to the id column then duplicates were removed from the title column.
- It is easier to analyze a smaller dataset rather than a large dataset with multiple empty spaces.

Data Preparation

1. Data Cleaning

- Decompression of .csv.gz files
- Removal of duplicates and missing values
- Currency formatting applied to relevant budget and revenue fields
- Outliers were kept for better and accurate analysis
- Conversion of dates and runtime fields to standardized formats
- Removal of unnecessary columns to make merging easier

2. Added Calculated Columns

- $\text{profit} = \text{worldwide_gross} - \text{production_budget}$
- $\text{roi} = (\text{profit} / \text{production_budget}) \times 100$

Modeling

Visual Data Analysis:

- Production Budget vs Worldwide Gross Revenue scatter plots
- Average Movie Gross Revenue by Release Year
- Top 15 Movies by Return on Investment
- Average Domestic vs International Revenue
- Top 20 Highest Grossing Movies Worldwide
- Seasonal Release Patterns
- Budget Ranges and Success Rates
- Top Performances: High Budget vs Low Budget Success Charts
- Vote Average vs Box Office Performance

Evaluation

1. Holiday (Nov-Dec): Strong performance for family-friendly content.
2. Summer (May-Aug): 35% higher average gross than other seasons
3. Analysis based on past performance may not predict future trends
4. Streaming services impact not fully captured

Prerequisites

Getting started

1. Fork Create a fork.
 2. Clone
- Type: git clone (you can clone using either *SSH key* or the *HTTPS*) [(<https://github.com/veenaledi/DS-Phase-2-Project.git>)]

Key Findings

1. High-budget films (>\$100M) can achieve massive returns but carry significant financial risk
2. Low-budget productions (<\$10M) can yield extraordinary return of investment percentages
3. Summer releases (May-August) show higher average gross revenue
4. Original content requires careful genre selection and budget management

Reccomendations

1. Plan major releases for summer and holiday windows
2. Use smaller films to test audience response before major franchise investments
3. Develop franchise-ready content with sequel and merchandising potential
4. Reserve high budgets (>\$100M) only for franchise properties with strong IP foundation
5. Coordinate marketing spend with budget tier - allocate 50-60% of production budget for marketing on high-budget films
6. Consider international appeal in script development and casting decisions

Testing

To run the cells press ctrl+shift You'll need to download the datasets required You can get the datasets from: [\[https://github.com/learn-co-curriculum/dsc-phase-2-project-v3/tree/main/zippedData\]](https://github.com/learn-co-curriculum/dsc-phase-2-project-v3/tree/main/zippedData)

Technologies Used

- Python: Primary programming language
- Pandas: Data manipulation and analysis
- Matplotlib: Data visualization
- Jupyter Notebook: Development environment
- Git: Commit and push to remote repository

Contributions

Contributions to the Film Making Entry Analysis Report are welcome! If you have any suggestions, bug fixes, or additional features you'd like to add to the dashboard, please feel free to submit a pull request or open an issue.



Releases

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

Packages

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

Languages

● Jupyter Notebook 100.0%