Saadia-ramo / **phase2_project7**

Code Pull requests Actions Projects Wiki Security Insights Settings

0 stars 1 fork 0 watching Branches Activity
Tags

Public repository · Forked from Marylyne-Ingwe/phase2_project7

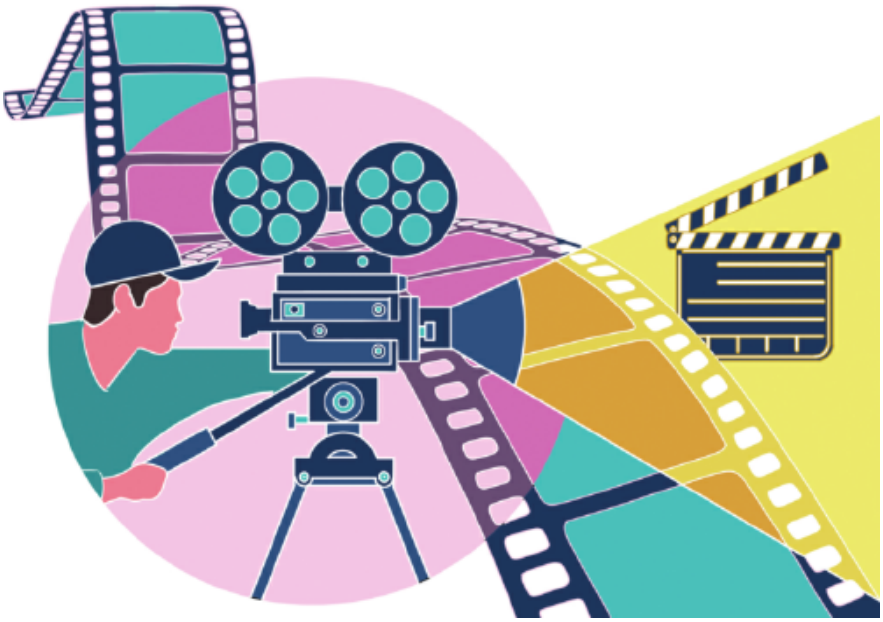
1 Branch 0 Tags Go to file t Go to file Add file Code

This branch is 1 commit ahead of `Marylyne-Ingwe/phase2_project7:main` . Contribute Sync fork

Saadia-ramo change presentation name 6fe77b8 · 2 minutes ago

zippedData	change presentation name	2 minutes ago
.gitattributes	Track im.db with Git LFS	last week
.gitignore	project updates	5 days ago
README.md	Update README.md	2 days ago
final.ipynb	Rename notebook to final.ipynb	2 days ago
presentation.pdf	change presentation name	2 minutes ago

README



Movie Market Analysis: Data-Driven Insights for Strategic Movie Production

Authors :

Marylyne

Cindy

Trevor

Halima

Project Overview

This project focuses on analyzing movie market data to deliver **strategic insights** for launching a new movie studio.
By leveraging **Python** and **SQL**, we explore factors such as **box office performance, audience ratings, budgets, genres, and release timing** to uncover what makes movies successful.

The aim is to translate data-driven insights into **practical recommendations** that stakeholders in the film industry can use to minimize risk and maximize profitability.

In today's entertainment industry, major players like **Netflix** and **Amazon Studios** are reshaping the landscape through original content. Inspired by their success, our company seeks to establish a new studio with a strong **data-informed foundation**.

Business Understanding

Real-World Problem

The film industry is high-stakes: millions of dollars are invested in production, yet only a fraction of films become profitable.
Challenges include:

- Huge **production & marketing costs**
- Rapidly shifting **audience preferences**
- Technological disruption & streaming
- Fierce competition among studios

Our company's challenge is to enter this space with **limited experience** but high ambition. To succeed, we must understand **what drives success** and **how to avoid failures**.

Stakeholders

- 🎬 **Head of Movie Studio** – Strategic production guidance
- 💰 **Investment Team** – Profitable, low-risk financial decisions
- 🏗️ **Operations Team** – Planning & resource allocation
- ⚖️ **Risk Management Team** – Lessons from past failures & successes
- 📊 **Monitoring & Evaluation Team** – Metrics for ongoing improvements
- 📣 **Marketing Team** – Target audience insights for campaigns
- 👥 **Audience** – Better content aligned with preferences

Business Value

Our analysis provides competitive advantages:

- **Market Alignment** – Understanding audience-driven genres & trends
- **Investment Efficiency** – Smarter budget allocation
- **Content Strategy** – Building a focused, profitable film portfolio
- **Competitive Positioning** – Benchmarking against leaders
- **Risk Reduction** – Avoiding historical pitfalls

Project Goals

- Identify **box office success drivers**
- Understand **market trends**

- Develop **actionable insights** for production
- Support **investment decisions**
- Lay a foundation for a **data-driven studio strategy**

Key Business Questions

1. How does a movie's **budget** affect revenue?
2. Which **genres and directors** consistently perform best?
3. What patterns emerge from **audience ratings and runtimes**?
4. Does **release timing** significantly impact success?

Data Understanding and Analysis

Data Sources

- Box Office Mojo – Revenues
- IMDb – Ratings & movie basics
- Rotten Tomatoes – Reviews & critic sentiment
- The Movie DB – Metadata
- The Numbers – Budgets & grosses

Key Datasets

- **im.db** – SQLite database with 146,144 movies
- **bom.movie_gross.csv.gz** – Box office gross (3,387 movies)
- **tn.movie_budgets.csv.gz** – Production budgets & worldwide grosses (5,782 movies)

Data Overview

IMDb Dataset

- **Movie Basics**: 146k records (titles, genres, runtime, release year)
- **Ratings**: 73k records (average rating + votes)
- **Missing Data**: ~22% runtimes, ~4% genres

Box Office Mojo Dataset

- **Records**: 3,387 (2010–2018)
- **Fields**: title, studio, domestic gross, foreign gross, year
- **Missing**: 1,350 foreign gross values (~40%)

Budget Dataset (The Numbers)

- **Records**: 5,782 movies
- **Fields**: budget, domestic gross, worldwide gross
- **Quality**: Complete, no missing values

Tools Used

- **Python** – Core analysis
- **Pandas / NumPy** – Data manipulation
- **Matplotlib / Seaborn** – Visualization
- **SQLite** – Querying IMDb data
- **Jupyter Notebook** – Interactive environment

Analysis Approach

- **Data Cleaning** – Fixing missing values & duplicates
- **Exploratory Analysis** – Genre, budget, and rating patterns
- **Statistical Analysis** – Correlations with revenue
- **Visualization** – Clear insights through graphs
- **Business Insights** – Translating findings into recommendations

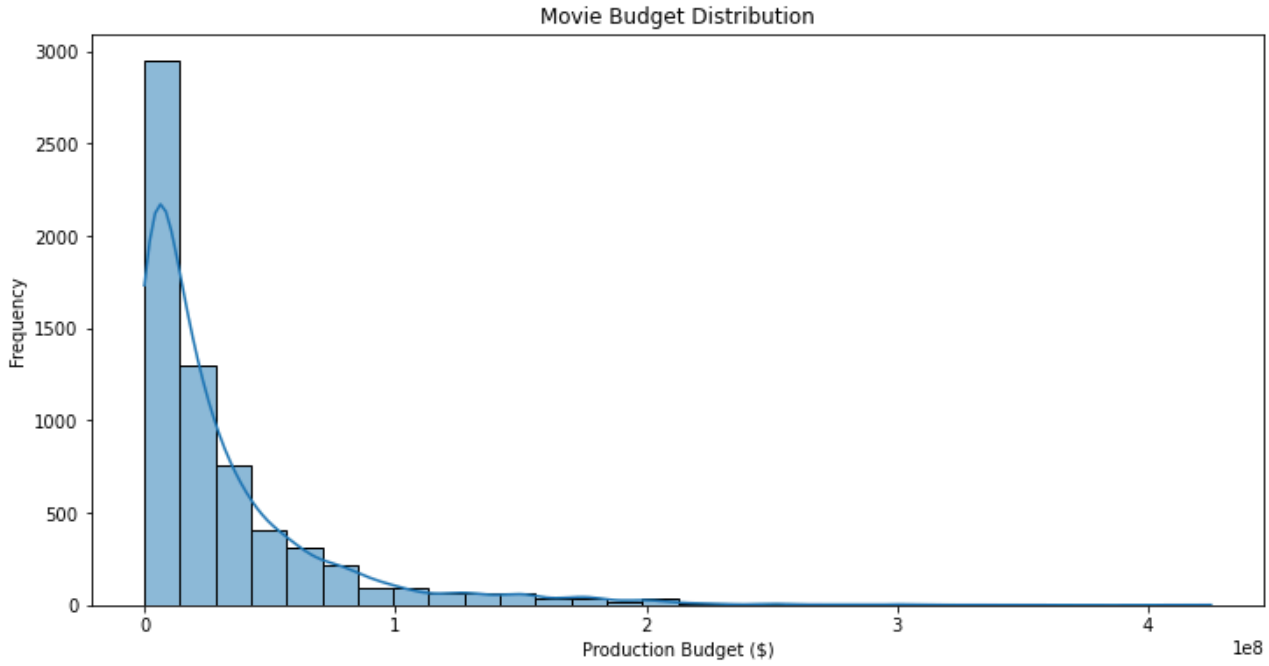
Key Analysis Areas

- 📊 **Genre Performance** – Best & worst financial performers
- 💵 **Budget vs Revenue** – ROI across budget ranges
- 🗓️ **Seasonal Trends** – Release timing strategies
- ⭐ **Rating Impact** – Influence of critic & audience scores
- 🏆 **Market Competition** – Benchmarking major studios
- ⏳ **Runtime Optimization** – Ideal length by genre

Visualizations

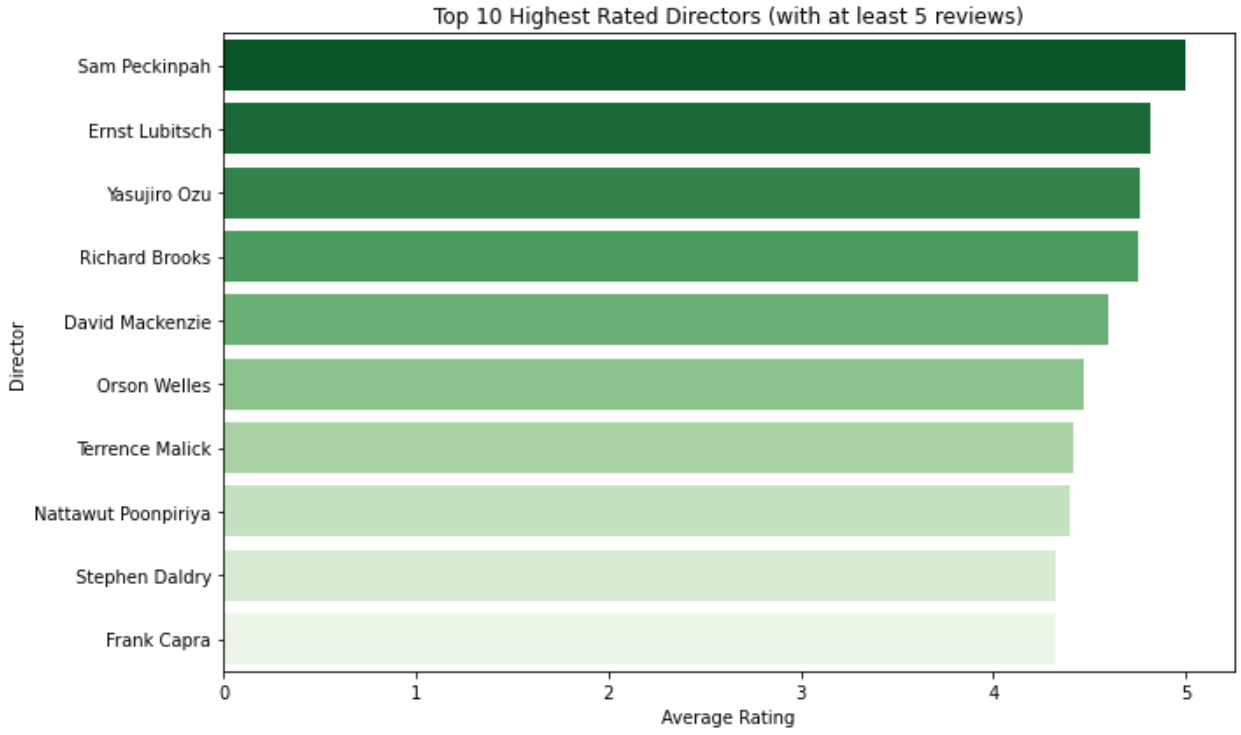
1. Genre vs ROI

Genres like **Drama, Family, Fantasy** show strong ROI.



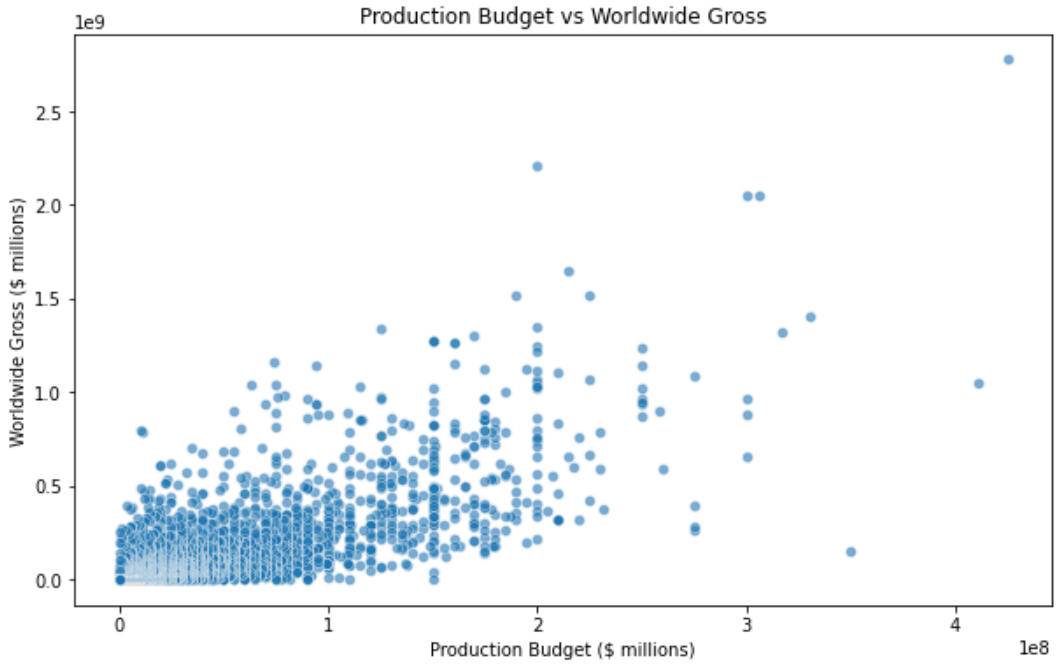
2. Top 10 Directors (by Ratings)

Certain directors consistently deliver strong films.



3. Revenue Analysis

Action and Sci-Fi dominate worldwide revenue.



📈 Statistical Analysis & Tests

To validate our insights, we applied:

1. **Correlation Analysis** → Checked linear relationships (e.g., budget vs. revenue, ratings vs. revenue).
2. **Hypothesis Testing (t-tests, ANOVA)** →
   - Tested if certain **genres** significantly outperform others in revenue.
   - Compared **ratings groups** (IMDb > 7 vs. < 7) for differences in gross revenue.
3. **Regression Models** →
   - **Linear regression**: Predicted revenue based on budget, ratings, and runtime.
   - **Multiple regression**: Combined financial and audience predictors.

🔑 Why regression?
Regression helps **quantify the effect size** of predictors like budget or ratings on box office revenue, allowing data-backed investment decisions rather than assumptions.

🔑 Key Findings

1. **Bigger Budgets Usually Mean Bigger Profits** – Movies that spend more on production often make more money, though smart mid-budget films can also perform well.
2. **Directors Influence Success** – Well-known and experienced directors consistently deliver strong box office results.
3. **Winning Genres** – Action, Sci-Fi, and Adventure movies attract the most audiences and dominate earnings.

4. **Release Timing Matters** – Films released during **holidays and summer** seasons perform better than those released at other times.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%