

RomeC-137 / Phase-G-2-Project

<> Code

Issues

Pull requests 1

Actions

Projects

Wiki

Security

Insights

Settings

0 stars

2 forks

0 watching

Branches

Activity

Tags

Public repository

Your main branch isn't protected

Protect this branch from force pushing or deletion, or require status checks before merging. [View documentation.](#)

Protect this branch

Dismiss

4 Branches

0 Tags

Go to file

t

Go to file

+

Add file

Code

RomeC-137 Update .gitignore 4ae71ca · 37 minutes ago

Images	completion of the README file.	12 hours ago
Phase-G-2-Project	my first commit	yesterday
.gitignore	Update .gitignore	37 minutes ago
README.md	completion of the README file.	3 hours ago
bom.movie_gross.csv	This is the analysis that has been don...	3 days ago
business investment strategy.i...	Add files via upload	yesterday
presentation.pdf.pptx	presentation.pdf.pptx	9 hours ago
rt.movie_info.tsv	This is the analysis that has been don...	3 days ago
rt.reviews.tsv	This is the analysis that has been don...	3 days ago
tmdb.movies.csv	This is the analysis that has been don...	3 days ago
tn.movie_budgets.csv	This is the analysis that has been don...	3 days ago
zipped folder.zip	adding zipped file	3 days ago

README

Analysis of the Film Industry for Business Expansion

Recognizing the growing trend of major corporations producing original content, our company is embarking on the creation of a successful movie studio. To overcome our limited experience in film production, this project leverages a comprehensive analysis of various film industry datasets. Our objective is to extract data-driven insights that will illuminate the most profitable, well-received, and promising areas for investment. By integrating these datasets, we aim to identify current industry trends and provide actionable recommendations to guide our studio's film prioritization strategy.

Business Understanding

The evolving landscape of streaming, global cinema, and digital distribution has transformed film production into a complex intersection of art and high-stakes business. To understand the drivers of film success, a thorough analysis of box office performance, critical reviews, audience preferences, and market trends is essential.

This analysis will address key business questions to inform strategic decision-making:

- **What genres are most consistently praised by critics?**
- **Which genres offer the highest potential for return on investment (ROI)?**
- **Are there directors whose films consistently achieve strong international performance?**
- **What languages are most commonly associated with successful films?** By answering these questions, we aim to empower the new studio team with reliable data to go ahead with projects that resonate with audiences and maximize financial returns.

Analysis Data

This project utilizes a powerful combination of industry datasets, each contributing unique insights to our understanding of the global film landscape:

- **The Numbers:** Providing crucial financial details such as production budgets and revenues, enabling accurate ROI calculations.
- **Box Office Mojo:** Offering data on gross earnings and overall box office performance.
- **Rotten Tomatoes:** Supplying critic and audience scores, along with valuable insights into review sentiment.
- **TheMovieDB:** Delivering rich metadata on movies, encompassing genres, languages, production countries, popularity metrics, and more.
- **IMDb Links:** Providing unique identifiers to facilitate cross-referencing of movies across all datasets.

1 Exploration of Data

This phase emphasizes a thorough understanding of the dataset by assessing its structure and identifying key columns essential for our analysis. We investigate potential relationships within the data and evaluate its cleanliness to determine if any areas require attention.

To achieve this, each team member explores different subsets of the data, focusing on understanding the distribution and data types while identifying any obvious issues, such as missing values or inconsistencies. The objective is to uncover initial patterns, pinpoint issues, and identify opportunities for further exploration, ultimately forming hypotheses that will guide the subsequent steps in our analysis.

2 Data Preparation & Cleaning

Following the initial exploration, we shift our focus to preparing the data for more in-depth analysis. This involves accurately merging various datasets to ensure that relationships are correctly explored.

We address any missing values through imputation, manage outliers, and carry out additional cleaning tasks, such as removing duplicates and resolving inconsistencies.

By structuring the data appropriately, we enable a deeper exploration of relationships while ensuring that the datasets are accurate, consistent, and ready for analysis. This stage is crucial for ensuring that the data is in a clean and usable state for modeling and deriving deeper insights.

Data Analysis

Data Visualization

Team members investigate visual trends associated with genres, ratings, ROI, and other variables. Various visualization techniques, such as bar plots, box plots, histograms, and scatter plots, are employed to reveal patterns. The visual insights are cross-validated and shared within the team to enhance hypotheses and interpretations.

Business Recommendations

The results are synthesized into clear, actionable insights for the new movie studio. All team members work together to ensure that recommendations align with the business objectives. The focus is on assisting decision-makers in prioritizing **genres**, **directors**, **languages**, and **film types** based on factors such as profitability, audience reception, and global appeal.

Visualizations

The following are the key visualizations that contributed to our understanding of the film industry:

Understanding the Relationship between Genres and Ratings

 [Top Genres by Average Rating \(Audience vs Critic\)](#)

Understanding the Relationship between Genre and ROI

 [Normalized ROI per Genre](#)

Understanding the Relationship between Language and Popularity

 [Popularity Distribution by Top 20 Languages](#)

Understanding the Relationship between Directors and Worldwide Gross Earnings

 [Top 10 Directors by Worldwide Gross](#)

Summary

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 4



RomeC-137



masunta



KabiraKabira7 KABIRA



S-909

Languages

● Jupyter Notebook 100.0%