

mkiganane20 / Group3_Phase_Two_Project

<> Code Issues Pull requests Actions Projects Wiki Security Insights Setting

0 stars

3 forks

1 watching

Branches

Activity

Tags

Public repository

1 Branch

0 Tags

Go to file

t

Go to file

+

Add file

Code

mkiganane20

Jupyter_notebook_pdf

2284198 · 4 minutes ago

<div></div> dsc-phase-2-project-v3	first discussion	3 days ago
<div></div> Data_Driven_Insights_Movie_...	presentation_pdf_group 3	27 minutes ago
<div></div> README.md	Update README.md	18 minutes ago
<div></div> group3_phase_two_project - ...	Jupyter_notebook_pdf	4 minutes ago
<div></div> group3_phase_two_project.ip...	Updated Analysis	1 hour ago
<div></div> im.zip	Database	3 days ago

README

Phase 2 Project Description

Data-Driven Insights for a New Movie Studio

PHASE TWO PROJECT - GROUP 3

Authors

- Myrajoy Kiganane
- Austine Otieno
- Endalkachew Dessalegne
- Danton Kipngeno

1. Introduction

This project focuses on exploratory data analysis (EDA) to uncover key insights that will inform the strategy of a newly established movie studio. By analyzing film industry data, the goal is to identify trends that contribute to success in the industry. The findings will provide data-driven recommendations to help the studio make informed decisions about the types of films to produce, maximizing the likelihood of financial success.

1.1 Project Overview

1.1.1 Business Understanding

A company is launching a new movie studio but lacks expertise in the film industry. To help guide its production strategy, we will analyze successful films and extract actionable insights that will aid decision-making.

1.1.2 Business Problem

The studio needs to determine:

- Which film genres perform best at the box office?
- What factors impact a movie's ratings and audience engagement?
- How can data-driven insights shape production choices for success?

1.1.3 Key Business Questions

This project seeks to answer the following:

- Which movie genres have the highest average rating?
- Which movie genres have the highest number of votes?
- What are the top five movie genres that are highly voted for according to the start year?
- What is the trend of movie popularity over time?
- How does language influence the average rating of movies?
- How does region influence the number of votes of movies?
- What impact does the runtime (in minutes) have on the average rating?

2. Data Understanding

2.1 Data Preprocessing

2.1.1 The Data

To ensure a comprehensive analysis, we retrieved data from the IMDb website. The IMDb dataset consists of eight tables, but for this project, we focus on the following three key tables:

- **movie_basics** → Contains movie information such as genres and runtime.
- **movie_ratings** → Contains movie average ratings and number of votes.
- **movie_akas** → Contains details on region and language.

The tables share a common column, `movie_id` , which serves as the primary key for merging the datasets.

3. Technologies Used

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- **Jupyter Notebook** for data analysis and visualization

4. Usage

1. Open Jupyter Notebook.
2. Navigate to the notebook file (`.ipynb`) and run the cells step by step.
3. Explore the data analysis, visualizations, and insights.

5. Project Structure

<code>/data/</code>	- Contains datasets used for analysis
<code>/notebooks/</code>	- Jupyter notebooks with data analysis
<code>/scripts/</code>	- Python scripts for preprocessing & modeling
<code>README.md</code>	- Project documentation



6. Key Insights & Recommendations

- Identify high-performing genres based on revenue trends.
- Analyze budget-to-revenue ratios to determine profitable investments.
- Assess market trends to guide production decisions.

7. Conclusion

This exploratory data analysis has provided valuable insights into the factors that contribute to the success of movies. By analyzing IMDb data, we identified key trends in movie ratings, audience engagement, and genre popularity. The findings suggest that certain genres consistently perform well at the box office and receive higher audience ratings. Additionally, factors such as language, region, and runtime influence a movie's success.

These insights will help the new movie studio make data-driven decisions regarding film production, ensuring a higher likelihood of financial success. Moving forward, the studio should focus on high-performing genres, optimize budget allocation based on revenue trends, and stay aligned with market demands to maximize audience engagement and profitability.

8. Contributors

- Austine Otieno

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 4



mkiganane20 Myrajoy Kiganane



Endalkmin



dantonki



Austine-pixel

Languages



Jupyter Notebook 100.0%