

Ogam-Alvin /
Phase2Project

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings



☆ 1 star

🍴 3 forks

👁 1 watching

🌿 2 Branches

🏷 0 Tags

🔄 Activity

🌐 Public repository

🌿 2 Branches

🏷 0 Tags

🔍 Go to file

t

Go to file

+

Add file ▾

Code

⋮

Ogam-Alvin

presentation upload

a51ea35 · 15 hours ago

🕒

📁 zippedData	commit before merge	2 days ago
📄 .gitattributes	Tracking Large Files	5 days ago
📄 .gitignore	Added .gitignore file	yesterday
📄 MOVIE STUDIO ANALYSIS PD...	presentation upload	15 hours ago
📄 Movies Recommendation Ana...	Tableau pdf	17 hours ago
📄 README.md	Dashboard link	17 hours ago
📄 movie_recomendation.ipynb	Final notebook	2 days ago

📖 README

MOVIE STUDIO ANALYSIS



Project Overview

This project will use exploratory data analysis to generate insights for a business stakeholder. Here is an interactive dashboard for this analysis [Click here](#)

Business Problem

Since the client sees big companies creating original video content and they want to get in on the fun, they want to create a new movie studio. The aim of this project is to make tangible business suggestions on movie production on what types of films are currently doing the best at the box office. This process will involve analysing different movie datasets to gain more information and enable better bussiness suggestions to the client. These findings must then translate into actionable insights that the head of the company's new movie studio can use to help decide what type of films to create.

The main business questions this project sought to answer were:

- What is the total profit made by movies based on the genres and which genres have the most profit
- What are the most popular genres among customers?
- Which season(s) is best to release a movie in order to maximize profits.

Data Understanding and Analysis

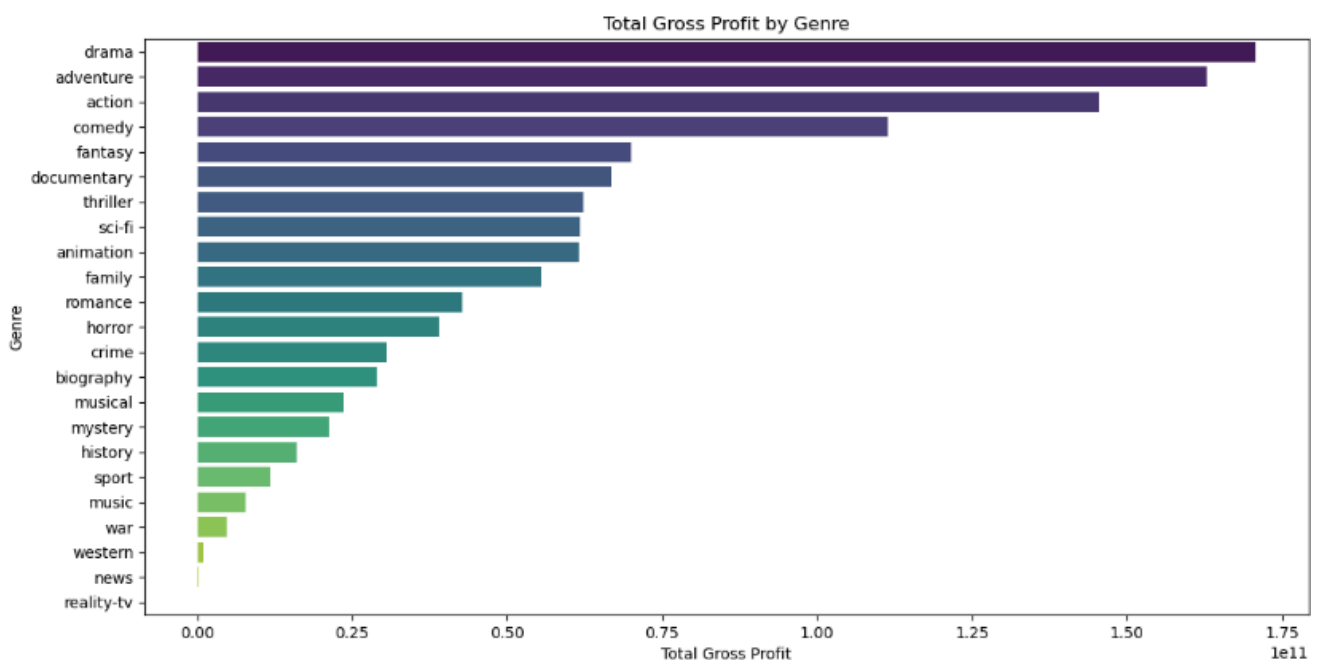
For this project we will be having 6 data sources:

1. bom.movie_gross.csv: This is a file that contains information regarding a movies name, both domestic and foreign gross amount and the year it was released
2. im.db: This is a zipped folder that contains a database called im which contains data regarding a movie. This data includes what category it falls under, the director, actors involved, movie ratings, running time and many more
3. rt.movie_info.tsv: This file has information the movie rating for theatrical releases e.g. 'R' 'PG 13' e.t.c. it also tells more about the movie like genre, director, theatrical release date, dvd release date, runtime, box office gross and studio
4. rt.reviews.tsv: This is a list of reviews from reviewers and the ratings they gave, it also gives info on who published the review
5. tmdb.movies.csv: Here we have genre ids, movie titles, language the movie is in, its popularity, release date, vote count and the average vote
6. tn.movie_budgets.csv: This file contains a movies production budget, domestic gross and worldwide gross

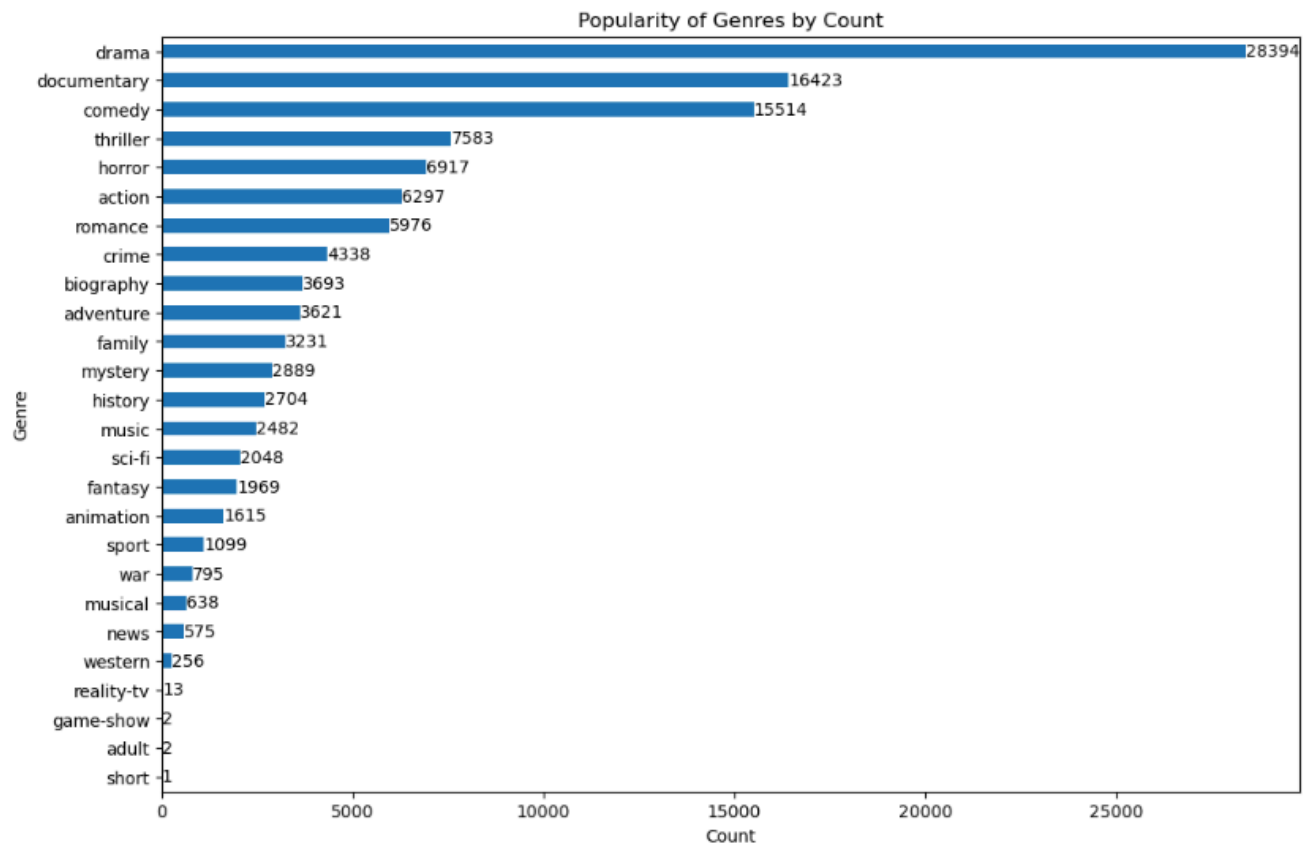
Data Description

In order to take a closer look at the data sources, all the necessary libraries were first imported. All the above datasets were then loaded and analyzed to gain better insight on the data. One of the files was a zipped file which had to be unzipped first in order to access the data within it. After loading the datasets, it was easier to find related columns that could be used to connect two or more tables.

The total gross profit by genre was determined using a series of code and the results displayed in a graph as shown below for better visualization:



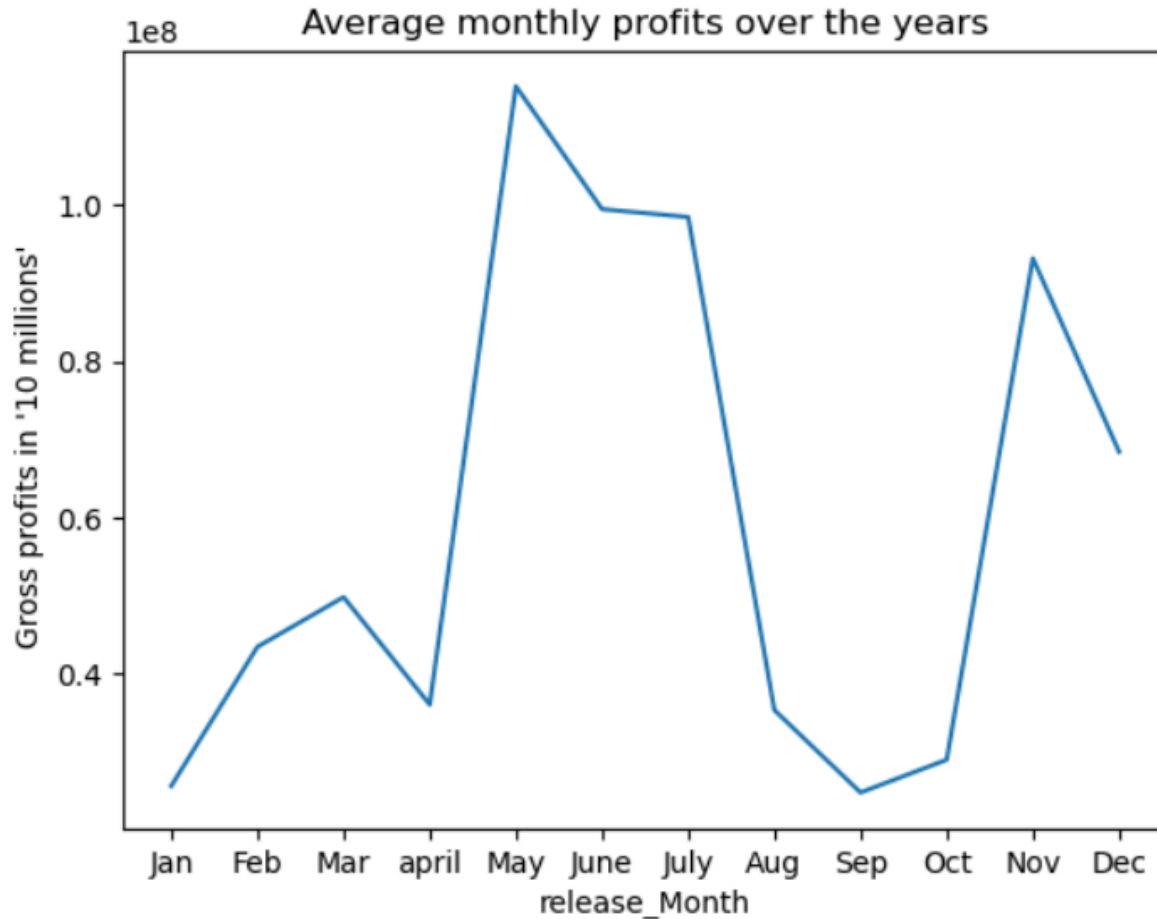
The most popular movie genre among customers was analyzed by looking at the ratings given for different movies. The best way to obtain this information was merging the movie_basics table with the movie_ratings table. To display the results, this bar graph showing popularity of Genres by Count was plotted.



From the graph above it was noted that:

- There is no other genre more popular than drama due to its large following
- The top 5 genres with a reasonable popularity count are : Drama, Documentary, Comedy, Thriller and horror, while only the first three have a popularity count of over 15k

The most appropriate time to release a movie in order to make the most amount of profit was analyzed by looking at the movie budgets table. A line graph showing the average profits of each month over the years was plotted as shown for better visualization.



From the above graph we can deduce that;

- There is high growth in interest in between the month of April and May and the same for October to November
- Movies released in the months May, June, July and November have yielded the most gross profits.
- Movies released in January, April, August, September and October have yielded the least profits.

Conclusion

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 3



Ogam-Alvin



Sharon-Mukami



GraceRKimotho

Languages

● Jupyter Notebook 100.0%