Film Industry Analysis

By Stephen Waweru

Overview

The purpose of this project was to recommend the best genre or combination of genres for the new film business. We use data science methods to clean, structure and analyse the data to provide in-depth understanding of the different genres and their relevance in the current movie market.

This will assist the business in making the right first step towards success in this competitive buisnes stage.

The data used comes from three acclaimed websites that compile data on the current trends in revenue, ratings, budget and human resource in the film industry. These websites are;

1. im.db from the International Movies Data Base website (link: https://www.imdb.com/)
2. tn.movie_budgets.csv.gz from The Numbers website (link: https://www.the-numbers.com/)
3. bom.movie_gross.csv.gz from Box Office Mojo website (link: https://www.boxofficemojo.com/)

Definition Of Terms

We used the factors below to measure success based on how they affect profitability:

1. **Genre** - This is the type of film in terms of the stylistic categories that organize films based on criteria such as the setting, characters, plot, mood, tone, and theme.
2. **Release Date** - The time of day, the day, the month or year or a combination of, in which the release copy my be published.
3. **Director** - a person who controls a film's artistic and dramatic aspects and visualizes the screenplay (or script) while guiding the film crew and actors in the fulfillment of that vision.
4. **Production Budget** - details all the projected costs of making a film, including talent and crew, equipment, location, wardrobe, construction, transportation and post-production expenses, to name a few
5. **Average Rating** - This the rating of the movie based on. It is the sum of all the scores divided by the number of votes.

Methods Used

1. **Spearman's Rank Correlation** is a statistical test used to measure the strength and direction of the association between two variables.

We'll use this to test the significance of each of the aforementioned factors except Genre for their effect on our profitability in the market represented by the data.

2. **The Kruskal-Wallis H test** is a non-parametric method used to determine whether there are statistically significant differences between the distributions of three or more independent groups.

We'll use this to test which combination of genres is most significant to our profitability represented by the data.

Data Analysis: Genre Combination

Profit by Genre
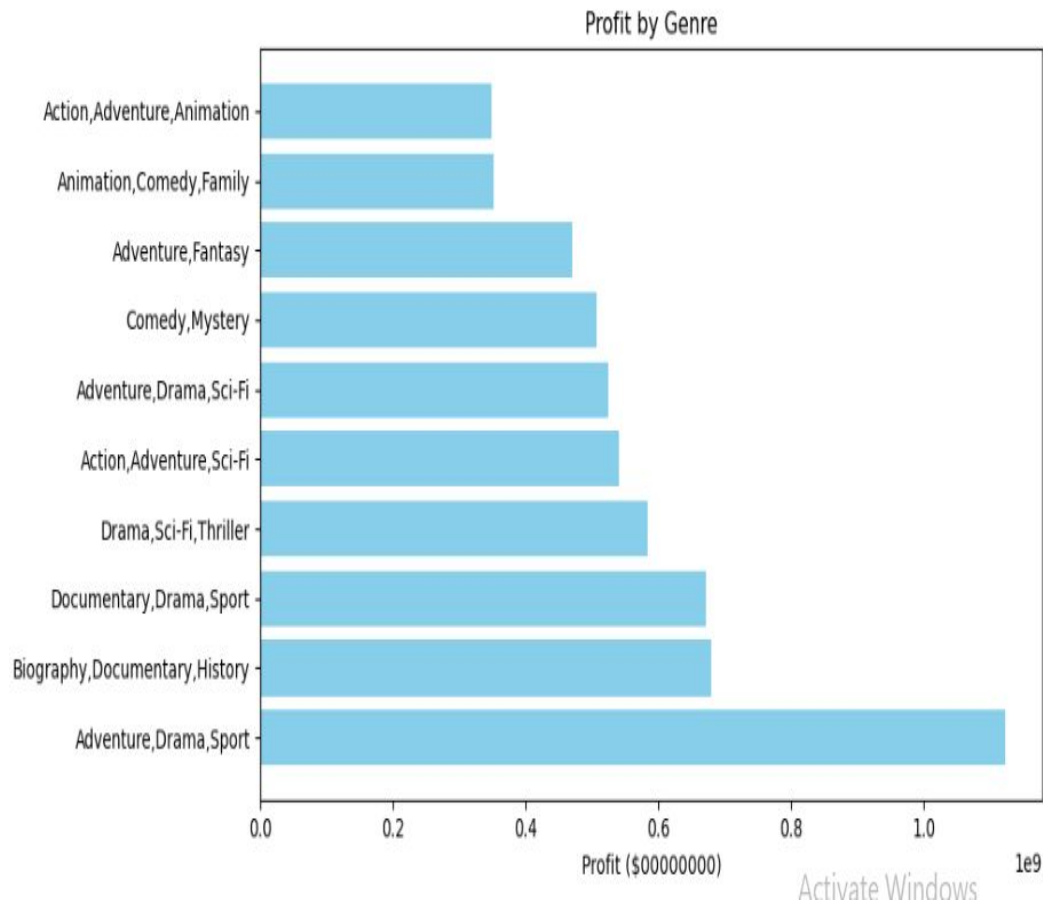
There is a consistent performance of Adventure Action combinations in terms of average gross profit.

The Kruskal-Wallis H test test indicates that we reject the null hypothesis, which poses that the median gross profits are the same across all genre combinations.

This result suggests that there are significant differences in gross profit among the different genre combinations in the dataset. In other words, the **genre combination of a film appears to have a substantial impact on its profitability.**

Data Analysis: Release Date

Average Monthly Profit
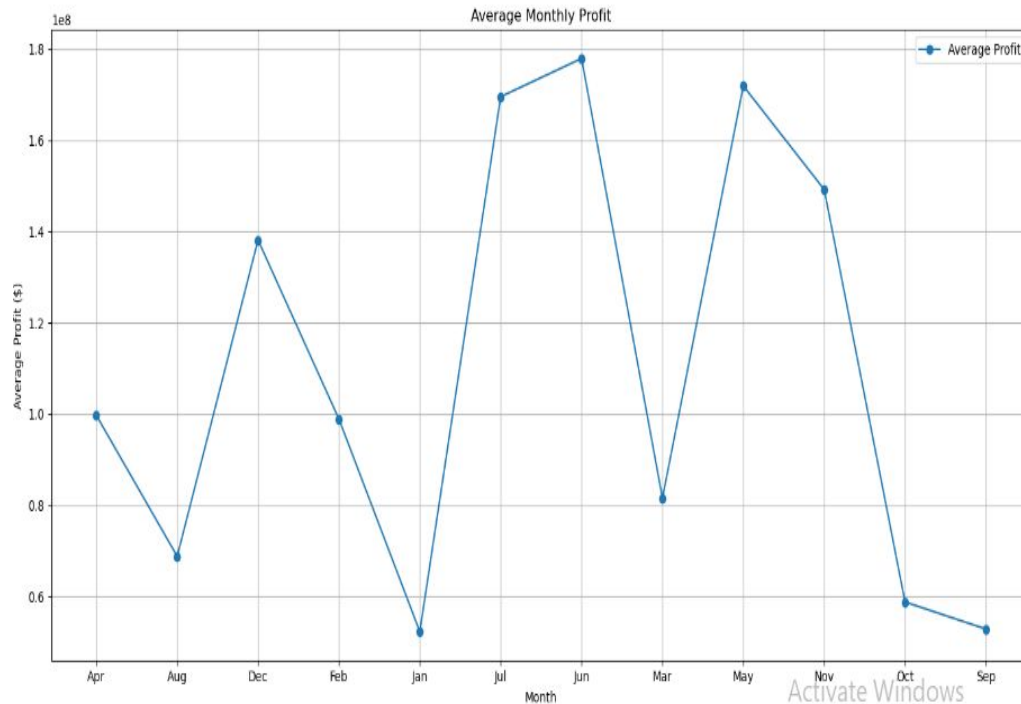
The visualisation suggest that certain moths have high significance in the average profitability, however, based on the Spearman's Rank test **there is no significant relationship between the release month or release day and the profitability of the films in your dataset**.

Data Analysis: Directors

director	count	gross_profit mean	gross_profit median
Steven Spielberg	7	1.829057e+08	129748880.0
Clint Eastwood	6	1.379160e+08	54133150.0
Francis Lawrence	5	4.042383e+08	488986787.0
Jon M. Chu	5	1.531137e+08	208099711.0
Tim Story	5	7.815801e+07	84827316.0
Ridley Scott	5	2.281130e+08	141521247.0
Jaume Collet-Serra	5	7.134972e+07	61985431.0
Zack Snyder	5	2.941840e+08	355945209.0
Steven Soderbergh	5	5.649705e+07	15264271.0
Pierre Coffin	4	8.549363e+08	929472292.5
James Wan	4	6.779971e+08	642447390.5
Antoine Fuqua	4	1.045980e+08	108166030.0
Christopher Nolan	4	5.840451e+08	588452008.5

The number of films done per director seems to have a correlation with the profitability as shown in the table on the left.

Spearman's test confirms this with showing **directors who have directed more films tend to be associated with higher gross profits**, indicating that experience in directing may contribute to a film's profitability.

Data Analysis: Production Budget
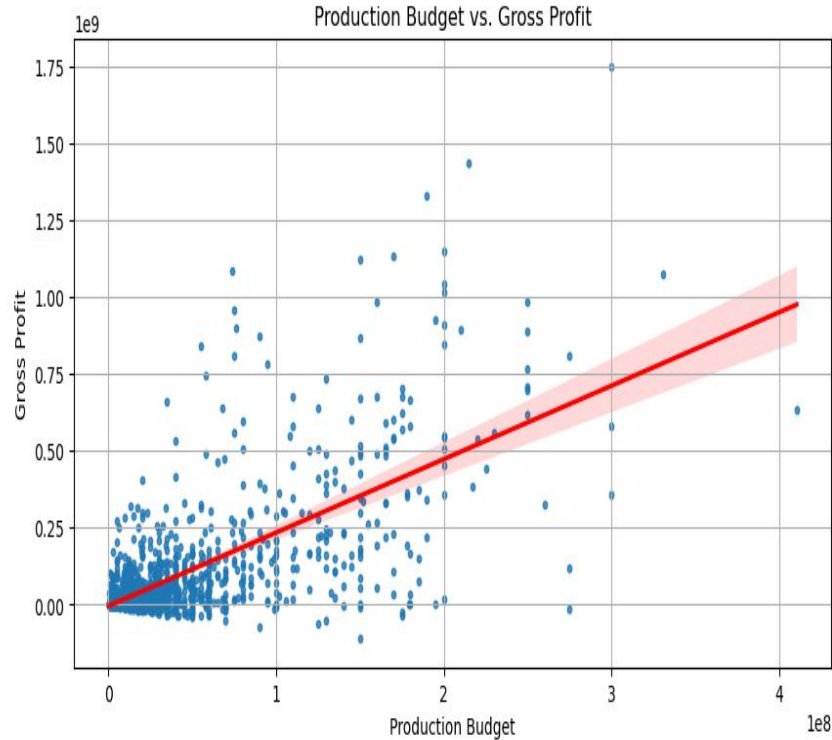
The visualisation suggest a linear relationship between production budget and profitability and Spearman's correlation confirms this with **films with higher production budgets tend to generate higher gross profits**. However, while the correlation is strong, it doesn't necessarily imply causation, as other factors could also contribute to profitability.

Data Analysis: Average Ratings

The table suggest ahigh correlation with average rating showing the highest profit with films rated 7.5 and above.

Spearman's Rank test for correlation coefficient indicates the strength and direction of the relationship between average rating and gross profit. A positive coefficient **suggests that higher average ratings are associated with higher gross profits**, while a negative coefficient indicates an inverse relationship.

averagerating	count	mean	median
6.3	56	1.028522e+08	45055701.5
6.6	54	1.291332e+08	34500439.5
6.8	51	7.142417e+07	27000339.0
6.2	50	1.125849e+08	44618823.5
7.0	48	1.469819e+08	33525986.0
6.5	48	1.033555e+08	46096244.5
6.4	46	5.900833e+07	21366985.0
7.2	44	1.332261e+08	22954369.0
6.7	42	9.288485e+07	40210703.0
7.3	38	1.508347e+08	36924341.5
7.1	38	1.159145e+08	40695658.0
5.7	35	8.260822e+07	31158601.0
6.9	35	9.027258e+07	18073377.0
5.8	31	5.252623e+07	13374700.0
6.1	31	4.864317e+07	14811293.0
5.9	30	3.850675e+07	21828452.5
6.0	29	1.101556e+08	72678948.0
7.4	28	1.413950e+08	25366817.0
7.6	26	1.294557e+08	65415984.5
7.5	24	2.044020e+08	78989930.0

Recommendations

Based on the tests carried out, we would suggest the best way to proceed would be;

1. Focus on films with an Action Adventure combination with an additional theme of Drama, Comedy Sci Fi and Sport genres.
2. Hire experienced directors with more than 1 film in their professional background preferably Steven Spielberg.
3. Use part of the production budget as marketing to influence the rating of the film before release to get a rating of 7.5 and above.

I believe these recommendations will be the best first step in creating a successful film buisness based on the data provided.

Question And Answer Section

Thank you for viewing my presentation. I leave the floor open for any questions

My Linkedin Profile: Stephen.Waweru