Analyzing the Relationship Between Release Timing, Box Office Success and Award Season Success

Introduction

This project aims to investigate the relationship between a movie's release timing and its box office success, examining how these factors influence its chances of receiving award season nominations and wins.

Does releasing a movie during awards season increase its chances of winning awards, regardless of box office performance? Do high-grossing movies have a better chance of winning awards, regardless of release timing? What role does the timing of a movie’s release play in its journey to becoming a cultural phenomenon or an award-winning masterpiece? As a movie enthusiast and someone who's interested in data science, this project seeks to unravel these questions by exploring the relationship between a movie’s release timing, its box office performance, and its success during award season.

By analyzing publicly available data, I aim to uncover patterns and trends that can shed light on the strategies studios use to maximize both financial returns and critical acclaim.

Data Collection & Preprocessing

Box Office Mojo, which tracks box office performance across domestic and international markets. Web Scraping is used.

TMDB API is used for release times and the months are categorized into release seasons for further analysis. I only included movies released between 2016-2020 and with at least 10 votes.

Then I normalized the titles and merged on normalized titles and years.

I will use Kaggle for Award Show Data:

The Oscar Award Dataset: https://www.kaggle.com/datasets/unanimad/the-oscar-award

Golden Globe Awards: <https://www.kaggle.com/datasets/unanimad/golden-globe-awards>

BAFTA Awards: <https://www.kaggle.com/datasets/unanimad/bafta-awards>

Standardized columns, counted how many awards each movie won and merged, filled missing values with a 0. Calculated the total number of awards won. Then merged all of my data together and filtered it further to include only movies with total awards won > 0. Because there are a lot of movies that won no award at all.

Methodology

metin, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

There's variability in the average number of awards won among award-winning movies across different months. Months typically associated with Awards Season (like October, November, and December) seem to have higher average award counts among the award-winning films.

metin, çizgi, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

There doesn't appear to be a very strong, tight linear relationship between worldwide gross and the number of awards won, even among award-winning films. While there might be a general trend where movies with higher gross tend to win more awards, there are many award-winning movies with moderate gross and a high number of awards, and some very high-grossing movies with a relatively low number of awards

diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Box Office Performance by Release Season (Left Plot):

Similar to the plot with all movies, the median worldwide gross for award-winning movies doesn't vary drastically by season. However, the Holiday Season and Summer Blockbuster seasons still show some of the highest outliers in terms of worldwide gross, even within this group of award winners. This suggests that while award success can happen in any season, the massive box office hits that also win awards are more likely to be released in these traditional blockbuster periods.

Award Success by Release Season (Right Plot):

This plot is particularly insightful. While all movies in this dataset won at least one award, the Awards Season and Holiday Season clearly show higher medians and upper quartiles for the total number of awards won. Movies released in Awards Season show the highest median and the largest spread, with many outliers winning a significant number of awards. Holiday Season award winners also tend to win more awards on average than those from Spring or Summer. Spring and Summer Blockbuster award winners tend to win fewer awards overall, even if they were recognized at least once.

metin, ekran görüntüsü, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

The Holiday Season has the highest average number of awards won (around 1). The Awards Season, Summer Blockbuster, and Spring seasons have much lower average award counts (closer to 0).

metin, ekran görüntüsü, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

While the trendline shows a general upward slope, movies with higher worldwide gross tend to win more awards, the points are very scattered around the line. Some movies with relatively lower worldwide gross still managed to win a moderate number of awards (points further up on the left side).

Hypothesis Testing

**Null Hypothesis (H₀)**: A movie's release timing (e.g., summer blockbuster season, award season, holiday season, spring) and box office success have no significant impact on its likelihood of receiving award season nominations or wins.

**Alternative Hypothesis (H₁)**: A movie’s release timing (e.g., summer blockbuster season, award season, holiday season) and box office success significantly influence its chances of receiving award season nominations or wins.

Performed Two-Way ANOVA,

Based on the Two-Way ANOVA results, we reject null hypothesis (H₀) and support alternative hypothesis (H1). Both a movie's release timing and its box office success (specifically, whether it is high-grossing or not) have a statistically significant impact on its likelihood of receiving award season nominations or wins.

Performed Chi-Square Test

Chi-Square statistic (2441.940): A large value, suggesting a strong association. p-value (0.00000): This p-value is extremely small (much less than 0.05)

For movies that are not high-grossing ('False' for Is\_High\_Gross), the proportion that won an award is very low (around 0.024) For movies that are high-grossing ('True' for Is\_High\_Gross), the proportion that won an award is significantly higher (around 0.558)

Is there a significant linear relationship between 'Worldwide Gross' and 'Total\_Awards\_Won' for award-winning movies?

R-squared: 0.004

This confirms the very weak linear relationship between Worldwide Gross and the total number of awards won among award-winning movies in this dataset.

Machine Learning Models

K-nearest neighbors:

The model performs very well at identifying movies that are not High Award (class 0), which is expected given the large number of movies in this category.

However, its performance on the minority class ('High Award', class 1) is less strong. It correctly identifies about 52% of the actual High Award movies (Recall), and when it predicts a movie is High Award, it's correct about 67% of the time (Precision). The F1-score of 0.59 provides a balanced view of its performance on the 'High Award' class.

This suggests that while the model has some ability to identify movies likely to win a high number of awards based on release season and worldwide gross, there is room for improvement, especially in correctly identifying all the actual High Award movies.

Decision Tree

The Decision Tree Classifier shows slightly better performance on the minority class ('High Award') compared to the k-Nearest Neighbors model. It has higher precision, recall, and F1-score for class 1. This means it's better at identifying actual High Award movies and is more accurate when it predicts a movie will be High Award.

While the performance on the minority class is improved, there are still a significant number of false negatives (movies that were High Award but predicted as Low Award - 52).

Random Forest

The Random Forest Classifier performs quite similarly to the Decision Tree on the minority class ('High Award'), with slightly different trade-offs between precision and recall. Both the Decision Tree and Random Forest models show better performance on the minority class compared to the k-NN model, as indicated by their higher precision, recall, and F1-scores for class 1.

Conclusion

This study provides evidence that both a film’s release timing and its box office performance significantly influence its success during the award season. Statistical analysis, including Two-Way ANOVA and Chi-Square tests, supports the conclusion that these two variables are not independent of a movie’s likelihood of receiving award nominations or wins. Specifically, films released during the traditional Awards Season (October to December) and those that are high-grossing show a statistically significant advantage in award recognition.

However, the analysis also reveals important nuances. The relationship between worldwide gross and total awards won is weakly linear, as indicated by a very low R-squared value. This suggests that while financial success contributes to award recognition, it is not a strong standalone predictor. Certain films with moderate earnings manage to achieve high critical acclaim, while some commercially successful films receive limited recognition from award bodies.

Machine learning models trained on release timing and box office features further illustrate the complexity of predicting award success. Although the Decision Tree and Random Forest classifiers outperformed the k-Nearest Neighbors model in identifying high-award-winning films, their overall performance—particularly in minimizing false negatives—indicates room for improvement. These findings underscore the multifaceted nature of award season success, which likely involves additional factors such as genre, critical reception, marketing strategy, and cultural impact.

In conclusion, while release timing and box office performance are important factors in award season outcomes, they are not the only variables that determine a film’s critical success. Future research could incorporate more qualitative features such as reviews, social media sentiment, and studio prestige to build more comprehensive predictive models of award success.

References

Box Office Mojo, TMDB API, Kaggle Award Season Data

Python libraries: pandas, matplotlib, seaborn, scikit-learn, numpy