# **Impact of Film Genre on Oscar Recognition**

Final Project Report

Spring 2025

MA 705 Data Science with Nathan Carter

Tim Nolan, Katherine Avendano, Mike Wang

**Executive Summary**

This project examines how film genre influences the likelihood of winning an Academy Award. Using data publicly available from Kaggle and IMDb. We merged Oscar award data with genre information for films released from 1990 to 2024. Our analysis aimed to support a film production company’s strategic investment decision by identifying genres with higher Oscar success rates.

We performed Chi-square tests to assess the significance of genre-category win rates, and calculated phi coefficients to measure effect size. Our findings reveal that certain genre-category combinations significantly increase the likelihood of Oscar success. These insights are particularly valuable for allocating awards campaign budgets and selecting film projects that optimize prestige and visibility.

**Introduction**

The Academy Awards (Oscars) represent one of the most prestigious recognitions in the film industry. Recognition at the Oscars can significantly elevate a film's visibility, enhance a studios reputation, and bring long-term commercial performance. For a new film production company aiming to gain prestige, understanding what kinds of films are most likely to win Oscars is a vital strategic consideration.

**Research Question**

How does film’s genre influence the likelihood of winning an Oscar, and does this relationship vary across award categories?

**Scenario:**

As a data analyst for a new production company, you are tasked with helping the executive team decide which types of films to invest in. The team’s goal is to increase the studio’s prestige and visibility. Recent industry reports suggest that earning an Academy Award nomination or win can significantly impact a studio’s reputation and audience reach, while also balancing the need for commercial success. The executives have allocated a budget specifically for an Oscars campaign, including marketing efforts, red carpet appearances, and promotional materials, and are seeking data-driven insights on which film genres and award categories would best maximize their chances of success during awards season.

**Data Preparation**

**Data Sources**

**The Oscar Award Dataset (Kaggle)**: Includes detailed records of Oscar nominations and wins from 1927 through 2025.

**IMDb Genre Dataset**: Extracted using Python web scraping techniques.

**Preparation Steps:**

1. Loaded the Oscar Awards dataset.
2. Built a subset of the data only including recent awards (where film release date is from 1990 to 2024). Focused the analysis to maintain relevance to current Academy voting patterns and trends.
3. Extracted film genre information by web scraping IMDb data and loading in the dataset
4. Remove rows from movie genres dataset where the Genre column is missing.
5. Create a dictionary of the missing values.
6. Merge the dictionary of missing values.
7. Merged the Oscar Awards with the IMDb genres data on film titles with an inner join. This final cleaned dataset contained over 4,000 film records, with each record including the film’s primary genre and Oscar win status.
8. Verify the merge with matching validations.
9. Import libraries for analyses.
10. Conduct statistical tests.
11. Print output of Statistical tests.
12. Build Graphical Interpretations of the results.

**Methodology**

**Variables**

* **Film Year**: The year of the films release.
* **Ceremony Year**: The year the award was presented.
* **Ceremony**: The ordinal number of the ceremony (e.g., 95th Academy Awards).
* **Category**: The award category (e.g., Best Picture, Actor).
* **Cannon Category:** The specific award subcategory (e.g., Actor in a leading role).
* **Name**: The individual person nominated.
* **Film**: The title of the film associated with the nomination.
* **Winner**: A Boolean indicating whether the nominee won the award.
* **IMDb ID**:The unique identifier ID associated with a particular film.
* **Genres:** The genres a particular film has (often more than one).

**Independent Variable:** film genres

**Dependent Variable:** Oscar win status (1 = winner, 0 = not winner)

**Statistical Methods**

* **Chi Square Test:** To determine whether win rates differ significantly by genre-category pairs.
  + **Null Hypothesis (H0)**: There is no association between film genre and Oscar win probability in a given award category (i.e., win rates across genres are completely independent of category).
  + **Alternative Hypothesis (H₁)**: There is a significant association between film genre and Oscar win probability in a given award category (i.e., certain genre-category combinations have win rates significantly different from what would be expected by chance).
* **Phi Coefficient**: To measure the effect size of those differences.

Data was grouped by genre and cannon category to assess relationships between these variables and Oscar success. A significance level of α = 0.05 was used for all statistical tests to determine whether to reject or accept the null hypothesis.

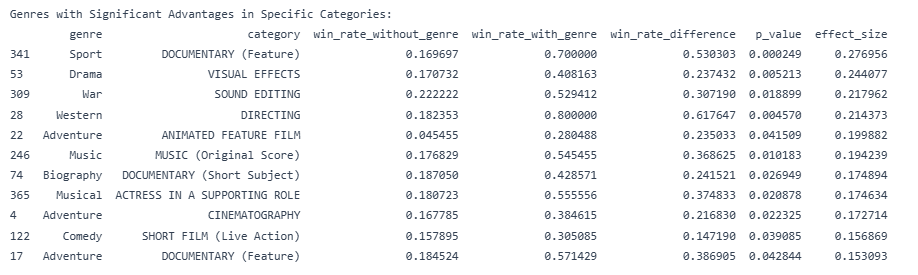
**Results & Analysis**

**Chi Square Test**

We found a statistically significant relationship between genre and win likelihood in specific award categories. The p-value was < 0.05, allowing us to reject the null hypothesis of no relationship between film genre and Oscar win probability. This indicates that the observed difference in win rates for specific genre-category combinations are likely not due to chance.

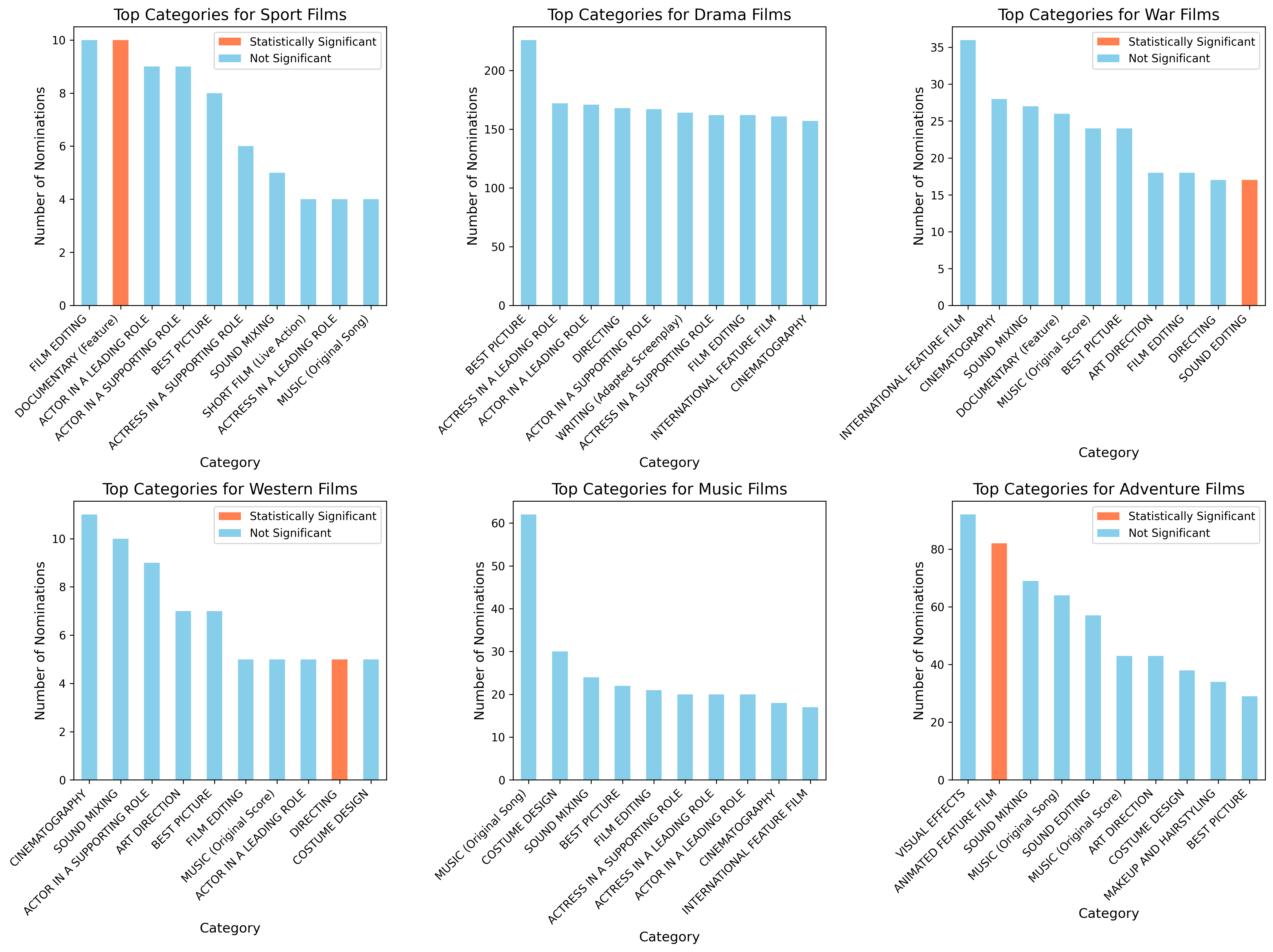
**Visual Interpretations**

**Figure 1:** Statistical Analysis of Genre-Category Relationships

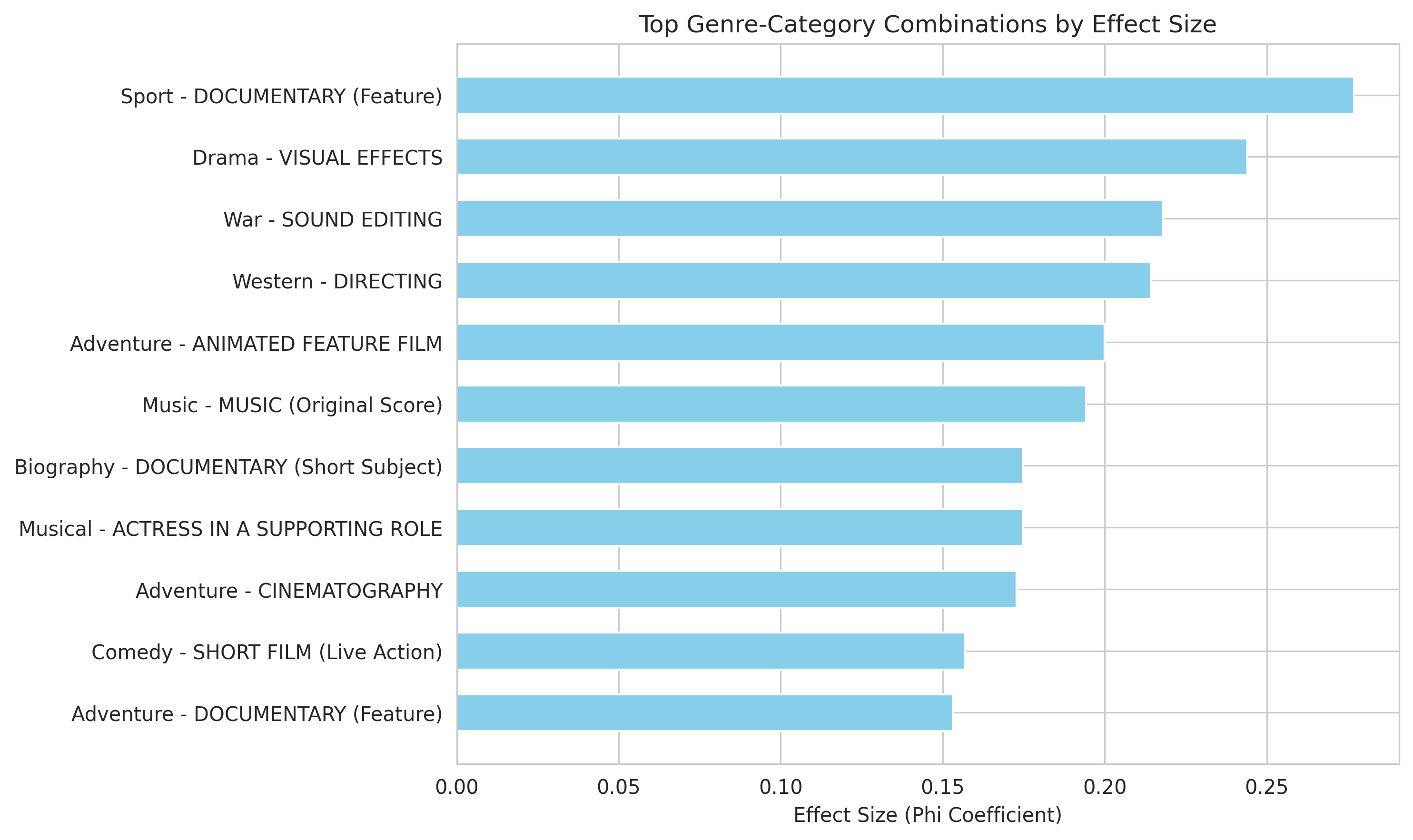


This table presents the complete results of our statistical analysis, showing Chi-square test p-values and phi coefficients for all tested genre-category combinations that were statistically significant at the (p < 0.05) indicating non-random associations.

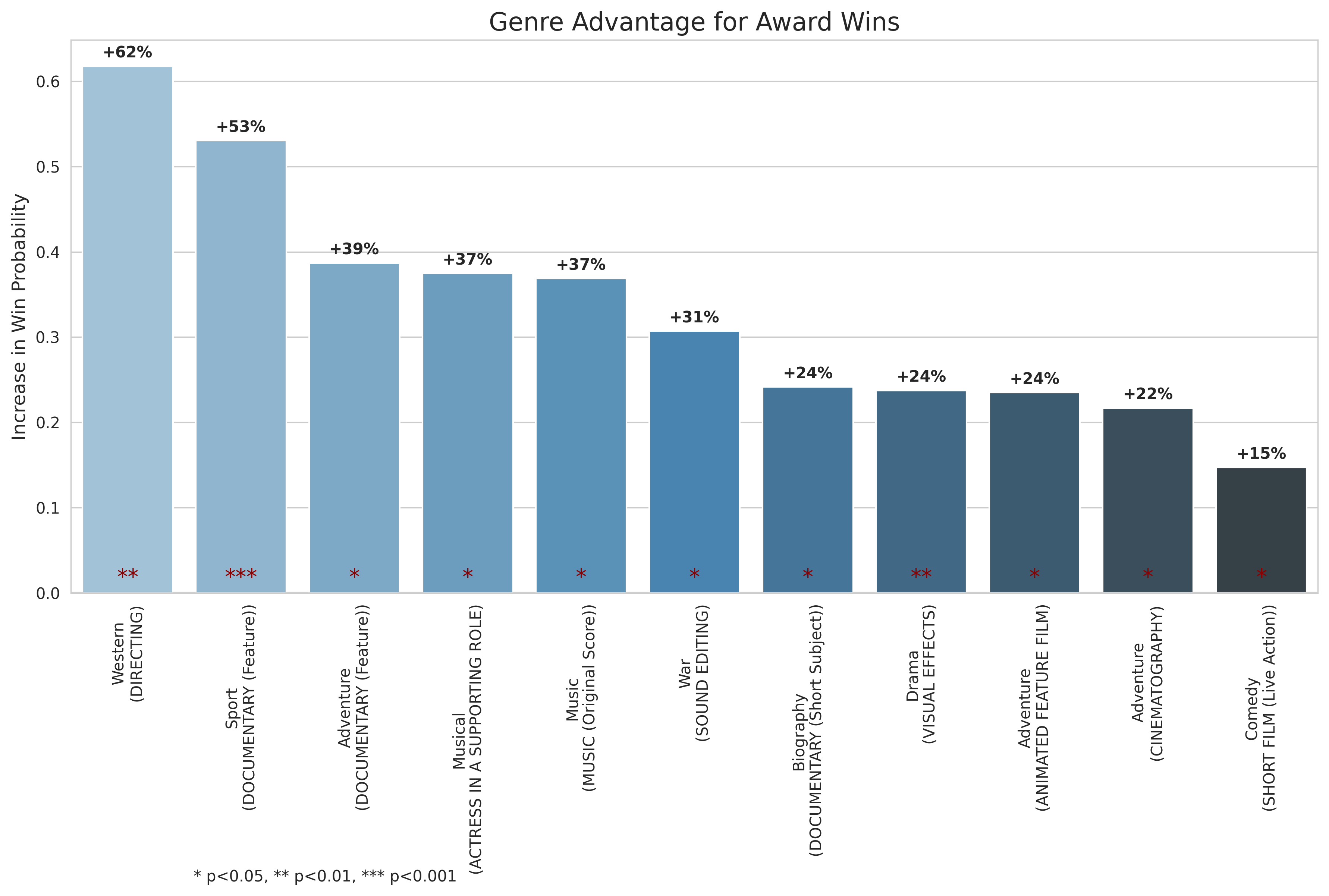
**Figure 2:** Nomination Distribution and Win Significance for Top 6 Genres

  
  
  
  
This multi-panel visualization shows nomination counts across award categories for the six genres with the strongest statistical relationships to Oscar wins. Orange highlighted boxes indicate statistically significant genre-category combinations with elevated win rates. Note that for Drama and Music genres, their most statistically significant category pairs do not appear among their top 10 most nominated categories, revealing that these are unexpected opportunities.

**Figure 3: Top Genre-Category Combinations Ranked by Effect Size**

  
This bar chart ranks the genre-category combinations with the strongest statistical relationships (highest phi coefficients), indicating which specific combinations have the greatest deviation from expected win rates. The higher the effect size, the stronger the predictive relationship between that genre-category pairing and Oscar success.

**Figure 4: Win Probability Advantage by Genre-Category Combinations**



This heat map visualizes the increased probability of winning an Oscar for specific genre-category combinations. The color gradient indicates advantage magnitude (lighter blue = higher advantage), while asterisks denote statistical significance levels (\* p<0.05, \*\* p<0.01, \*\*\* p<0.001). This visualization allows for an easy identification of the most advantageous genre-category strategic choices.

**Key Takeaways & Recommendations**

Based on the results of this analysis, we recommend investing in genres with the highest boost in advantage for winning an Oscar:

1. Westerns for Directing
2. Sports for Documentary Feature
3. Adventure for Documentary Feature

These genre-category combinations demonstrated the greatest increase in win probability compared to random chance. The effect sizes (phi coefficients) for these pairings indicate substantial strategic advantages when campaigning for these specific award categories.

Beyond these top three, our analysis identified 11 statistically significant genre-category combinations that could be strategically viable options. The production company should consider their existing catalog’s strengths and risk tolerance when selecting which combinations to pursue. For a studio such as this which is less established, it would be advised to stick to genre-category pairs with the highest statistical advantage. However, even combinations with moderate effect sizes might represent worthwhile opportunities to take advantage of, particularly if they align with their current creative expertise or a studio’s future planning direction.

**Limitations & Future Direction**

This analysis has several limitations, but suggest avenues for future research:

1. **Time Series Trends:** As Academy preferences may shift, ongoing analysis should be conducted to stay aligned with evolving trends. It could be important to see how these changes have shifted across various decades. This paper exclusively looks at the total count since 1990, and did not evaluate it as a time series object.
2. **Box Office Numbers:** This study did not include any confounding factors that could have stem from box office success. This could be an interesting direction for future work as studios have increasingly been moving towards straight-to-streaming releases or shortening theatrical releases for the public.
3. **Seasonality within Theatrical Releases:** There has been longstanding beliefs that movies released during certain months of the year perform better at the box office. That way the films theatrical release aligns with award-season marketing and builds strengths to maximize visibility and competitiveness before a film is forgotten. This could have been a confounding variable in our study, as we did not account for month, but instead the year in which it was released. Further data on this could be beneficial in seeing if time of year for release impacts award wins and potentially has different impacts on genre-category pairs.

**Conclusion**

Our findings demonstrate that film genre significantly impacts Oscar success across multiple award categories. Certain combinations consistently increase the likelihood of winning an Academy Award, providing important strategic insights for production companies seeking industry recognition.

**Deepnote Project**: [Deepnote Notebook](https://deepnote.com/workspace/ma-705-1-57c0152d-99a8-43e3-a897-a7e621980d86/project/Data-Science-Final-Project-Official-aa4c38a4-93c5-41d1-843f-4c2dc2bc3205/notebook/Homework-4-fec0044f07fd46a6aa6af1b89c4d0cad?utm_source=share-modal&utm_medium=product-shared-content&utm_campaign=notebook&utm_content=aa4c38a4-93c5-41d1-843f-4c2dc2bc3205)

**Kaggle Dataset**: [The Oscar Award Dataset](https://www.kaggle.com/datasets/unanimad/the-oscar-award)

**IMDb Non-Commerical Dataset:** [IMDb Genres Dataset](https://developer.imdb.com/non-commercial-datasets/)