Title: Predicting Oscars "Best Picture" Award from Google Trends Signals
Neha Kulshreshtha (nk442), Armita Jamshidi (aaj34), Yuqi Li (yl3698)

Online Data/Resources
For this project, we utilized Google Trends, which allows us to analyze what is trending. We got data on the Oscars awards from the official site: https://www.oscars.org/oscars/ceremonies/2022.

The Question
We aim to use search data (Google Trends) about Oscar-nominated films to predict the winner of the Academy's Best Picture Annual Award. We predicted that the Google search popularity of the movie name and the lead actor would be a signal to predict if a movie would win the Oscar award over the other nominated films. As we will discuss below, we initially created a binary tree classification model to solely use movie and actor Google search score, but we found that this was not effective enough and did some further research to create a model between Google Trends Score and Box Office Revenue for each film and determine what, if any, correlation exists.

We used data from the Oscars for the years 2021-2024, and set our time ranges in the Google Trends query for each year's Oscars according to the day that the nominations were announced up until the day before the winners were disclosed. One thing to note is that the Google Trends scores are scaled from 0-100 according to search volume and we did not need to perform any normalization on these scores (as the website states): "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term."

We first examined the search data for "Oscar" and related terms and discovered that among the search data, "Oscar" is a more well-known name that is searched by the public.
Actually, the Oscars and the Academy Awards are essentially the same thing. "Oscar" is the nickname given to the golden statuettes awarded at the Academy Awards, an annual event hosted by the Academy of Motion Picture Arts and Sciences. The awards ceremony recognizes excellence in cinematic achievements in the film industry.

Methods:
The goal of the prediction model was to determine the winning movie for the 2024 Oscars purely based on Google search trends of each of the nominations. To do this, only search trends from the announcement date of the nominations (January 23rd, 2024) until the award ceremony (March 9th, 2024) were accounted for. Google trend search was used to obtain a standardized outline of how each of the Oscar nominees compared in Google searches. We hypothesized that the nominee with the maximum Google search value between the nomination announcement date

and the award announcement would be the Oscar winner for *Best Picture* that respective year. We also wanted to understand if there was a correlation between the search trends of the lead actor in each respective movie and the Oscar winner for *Best Picture.* To understand how accurate our prediction model was, we identified the top 5 most-searched films out of the nomination list for the years 2021-2024. This was done by taking comparing the movies' search results on Google Trends.

After the top 5 most searched nominee films were identified, their trend search ranking was compared from the day that the nominees were announced to the award winner for each respective year. As well, only Google search trends within the United States were included in the data. Then, both the average standardized comparison of the nominee's search rankings and the maximum search ranking for each film, were calculated.

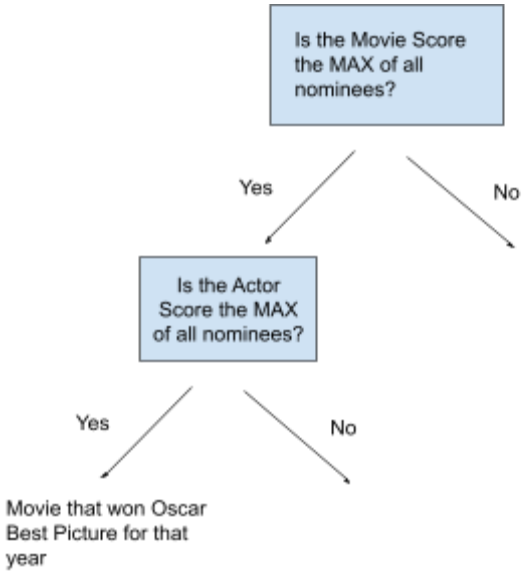
In 2024, there were 56 days of Google search rankings that were included when comparing each of the 5 nominees: Oppenheimer, American Fiction, The Holdovers, Poor Things, and Killers of the Flower Moon. The lead actors included Cillian Murphy, Jeffrey Wright, Paul Giamatti, Margot Robbie, and Leonardo DiCaprio, each starring in their respective films listed above.

In 2023, there were 47 days between the nomination announcement and the award announcement. The top 5 nominees included: Top Gun: Maverick, All Quiet on the Western Front, Everything Everywhere All at Once, and Avatar: The Way of Water. The lead actors, respectively, included Tom Cruise, Felix Kammerer, Michelle Yeoh, and Sam Worthington.

In 2022, there were 46 days between the Oscars and the nomination announcement. The top 5 films included CODA, Dune, Part One, Belfast, Don't Look Up, and Nightmare Alley. The respective actors included Emilia Jones, Timothee Chalamet, Caitriona Balfe, Jennifer Lawrence, and Rooney Mara.

Lastly, there remained a 43-day difference in 2021 between the nomination announcement and the awards ceremony. The top 5 films included Nomandland, Promising Young Women, Judas and the Black Messiah, and The Trial of the Chicago 7. The respective actors included Frances McDonald, Carey Mulligan, Daniel Kaluuya, and Sascha Baron Cohen.

For each of these data ranges, we calculated the average and maximum values of the Google Trends score for both the movie name as the queried keyword and the actor name as the query. We then built a model that classifies a given movie for a given year as the Oscar winner or not.



Graph 1.

The model is seen in Graph 1, above. It uses both the Movie and Actor scores to check if the movie won the Best Picture award for that year and selects the movie that has the highest of both of these variables. Following this model, however, our model only correctly predicts the 2024 data, which is only 25% accuracy. We then realized that the model is too restrictive, because the actor score is not necessarily correlated with the Oscar success of the movie, and rather the Movie score is a better predictor of a movie's success. We attribute this to other factors that affect an actor's popularity, such as their personal life or other films that they are featured in. We will elaborate further in the conclusion section.

Conclusion

Part 1: Google Search Trends of Oscar-Nominated Films and Lead Actors, 2021-24

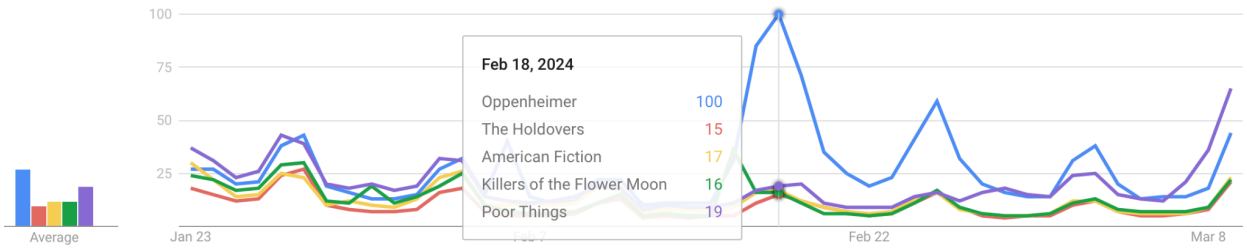


Figure 1: Google Trend Comparison of 2024 Oscar Nominee Film Titles

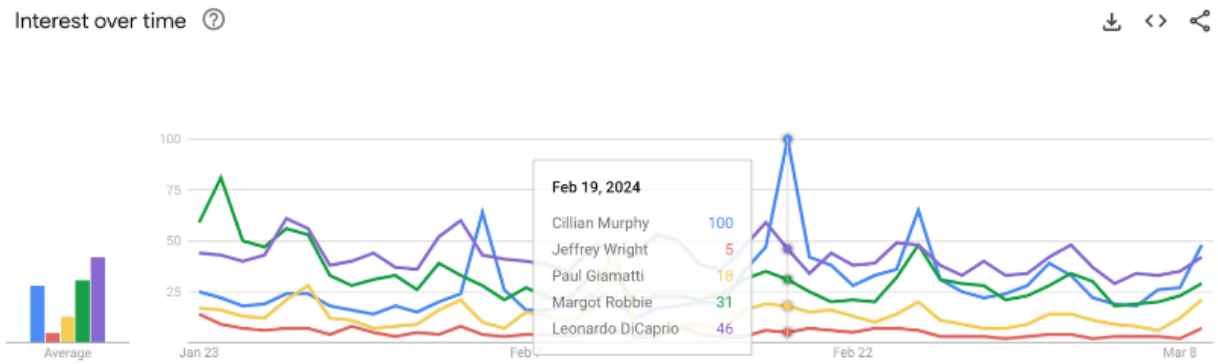


Figure 2: Google Trend Comparison of 2024 Lead Actors of Oscar-Nominated Films

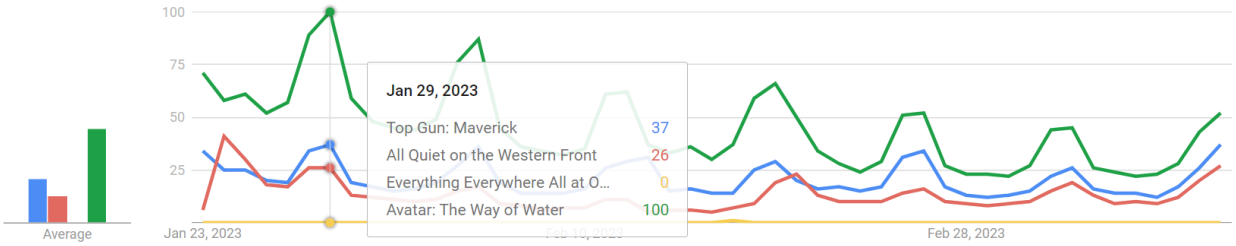


Figure 3: Google Trend Comparison of 2023 Oscar Nominee Film Titles

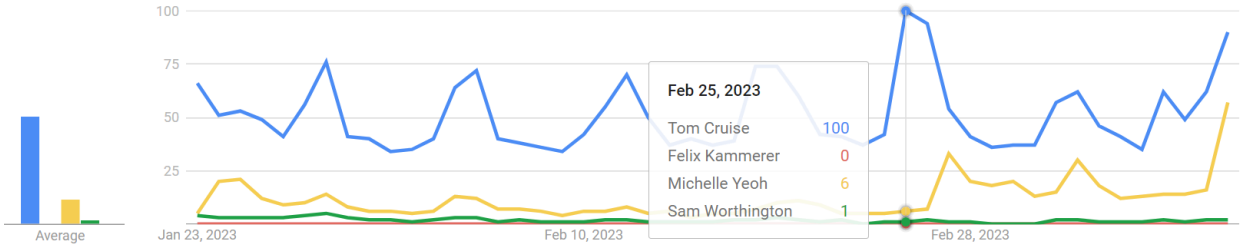


Figure 4: Google Trend Comparison of 2023 Lead Actors of Oscar-Nominated Films

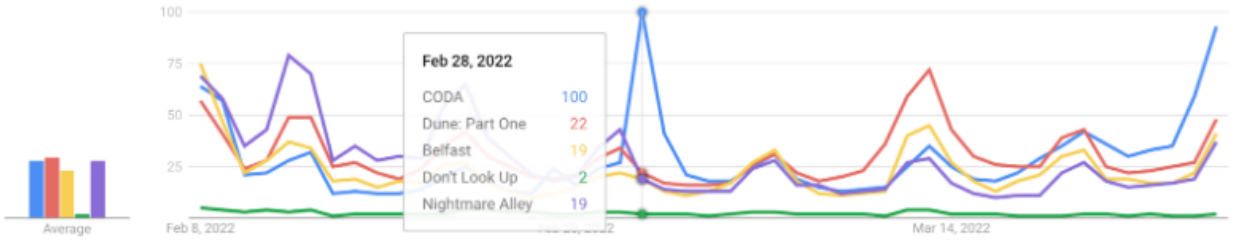


Figure 5: Google Trend Comparison of 2022 Oscar Nominee Film Titles

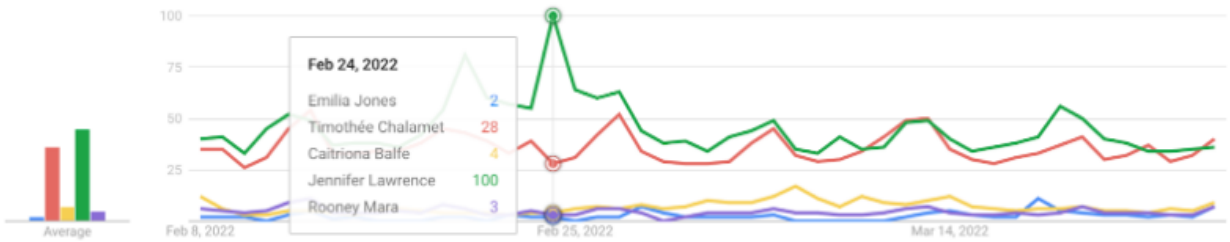


Figure 6: Google Trend Comparison of 2022 Lead Actors of Oscar-Nominated Films

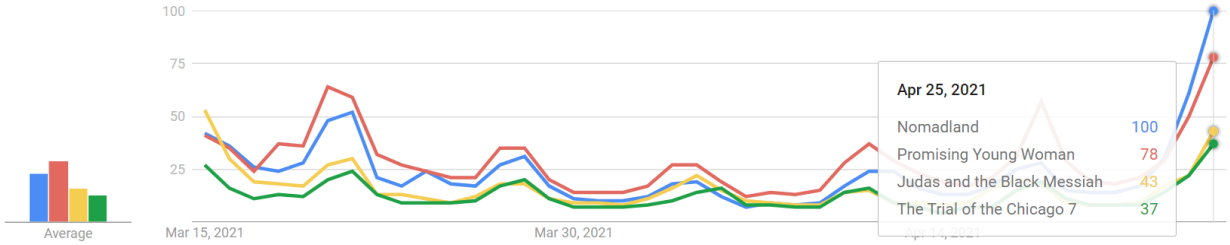


Figure 7: Google Trend Comparison of 2021 Oscar Nominee Film Titles

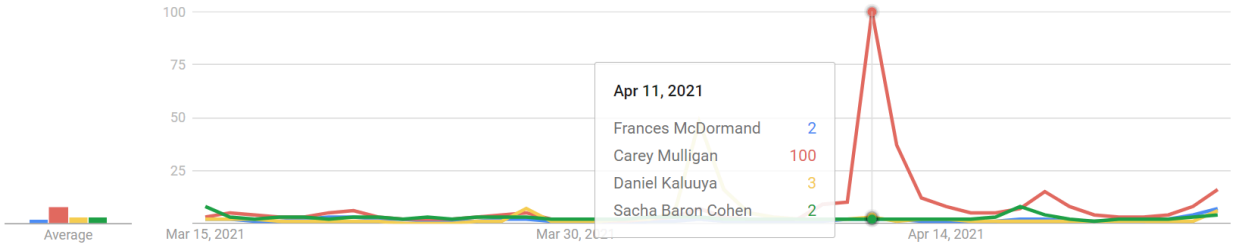


Figure 8: Google Trend Comparison of 2021 Lead Actors of Oscar-Nominated Films

Part 2:

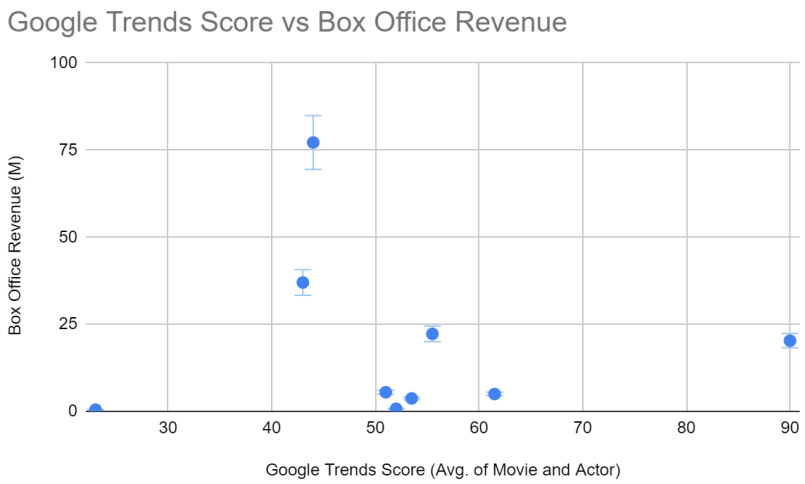


Figure 9: Google Trends Score vs Box Office Revenue Scatterplot

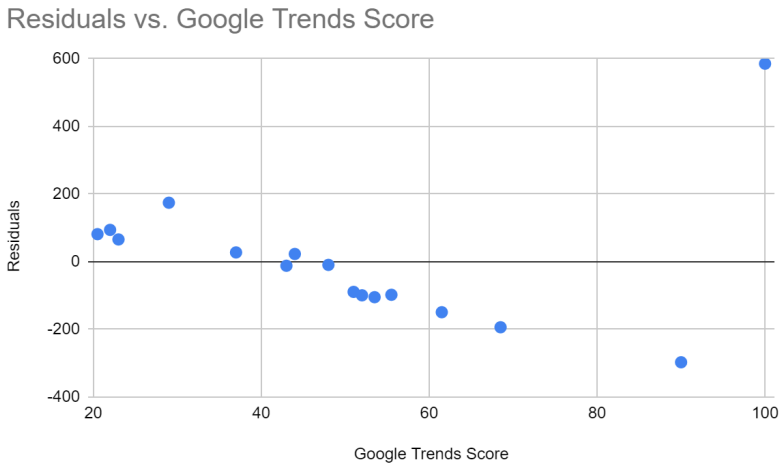


Figure 10: Residuals for Google Trends Score

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.549728							
R Square	0.3022008							
Adjusted R	0.252358							
Standard E	19.595774							
Observatic	16							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regressior	1	2328.188304	2328.19	6.06308	0.273859545			
Residual	14	5375.921071	383.994					
Total	15	7704.109375						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	45.244164	5.252091723	8.6145	5.7E-07	33.97954744	56.50878026	33.97954744	56.50878026
X Variable	0.0528682	0.021470798	2.46233	0.27386	0.006817953	0.098918516	0.006817953	0.098918516

Figure 10: Statistical Analysis of Google Trends vs Residuals Data

Discussion:

Part 1

Figures 1-8 depict the Google search trends for the top 5 searched Oscar-nominated films between 2021-24. In 2024 the winner of *Best Film* was Oppenheimer with lead actor Cillian Murphy. This coincides with our prediction that the title and lead actor with the maximum search trend between the nominee and award date will be the winner of the Oscars.

In 2023 the winner of *Best Film* was *Everything Everywhere All at Once* with lead actor Michelle Yeoh. This does not align with our prediction. This hypothesis was based on the assumption that the winner of the Oscars represents a democratic vote of who the U.S. population believes to have the best film. While Google search does not necessarily cause positive reviews, we predicted there would be a correlation. The lowest-searched film could have been selected due to the biases of the Oscar Awards committee, who have in the past been known for posing favoritism towards certain groups and certain movie topics.

In 2022 the winner of the award was CODA, with lead actor Emily Jones. While the movie title did indeed have the greatest Google search ranking, Emily Jones did not. This could mean that the leading actor's search rankings are not a good indicator of the popularity of the movie. Especially given that many movie actors starts remain influential people (and in this case, Jennifer Lawrence is a multi-movie star and had the most Google trends), the film's star likely is not a good indication of the film's success.

In 2021, Nomadland, with lead actor Frances McDormand, won *Best Film*. Like CODA, Nomandland had the largest Google search ranking, but Frances McDomand had a Google

search ranking that was severely under Carey Mulligan. Once again we see that lead actor may not be the best indicator of film success.

To ideate other ways that we can predict movie success, we decided to determine how movie revenue and film title search trends correlate. This led to the second part of our movie popularity exploration.

Part 2:

Given that the model in the first part was not very accurate, we attempted to determine if the model was better suited to predict success as measured by the metric of box office revenue. The independent variable is the score as calculated by the average of the Google Trends score for the movie name and the Google Trends score for the actor name. The dependent variable that we wished to predict is the Box office revenue (M). We assembled the datapoints which were the nominees and winning movies from all 4 years, and removed any outliers such as Barbie and Avataar, since the revenues for these were in the billions, not millions. We plotted this to determine if there is a linear relationship between the variables (Figure 9). From this, it was evident that there is no linear correlation, but we wanted to determine this analytically using a residual plot.

To create the residuals plot, we determined the slope and coefficient of the data to build a linear formula and then calculated the residuals as found by taking the difference of the actual revenue with predicted revenue. We then created the residuals plot (Figure 10) and determined that this data cannot be fitted with a linear regression as the residuals are not randomly distributed around 0 and instead, there is a general negative linear relationship in the residuals.

As seen in Figure 10, the p-value, 5.7E-5, is above the critical value of 0.05, which means that we fail to reject the null hypothesis, which states that there is not a linear relationship between the variables. This is in line with our findings above, from which we know that the model that uses Google Trend scores to predict revenue cannot be fitted with a linear regression model.

Conclusion

We began our exploration into our research question (Can Google Trends data about Oscar-nominated films and actors predict the winner of the Academy's Best Picture Annual Award). Upon some preliminary analysis in building a binary decision tree classifier, we quickly realized that our model was not accurate as it had an error rate of only 75%. This is likely due to there being other conflating factors that influence a Google Trends score of a movie or actor. For example, in 2023, "Everything, Everywhere All At Once" won the Oscars award but Tom Cruise, the lead in "Top Gun" had a Trends score of 100 while Michelle Yeoh, the lead in the former, only had a score of 57. This may be attributed to factors such as his general popularity

and other factors that may arise from his personal life. Furthermore, as seen in part 2 of the exploration, we utilized the average of these Google Trends scores to determine if this could predict Box Office revenue for a movie, and by way of linear regression, we realized that it could not. We may attribute this non-linear relationship to, again, similar factors that may have affected Revenue, such as effort put into marketing, how long the movie has been out, etc.