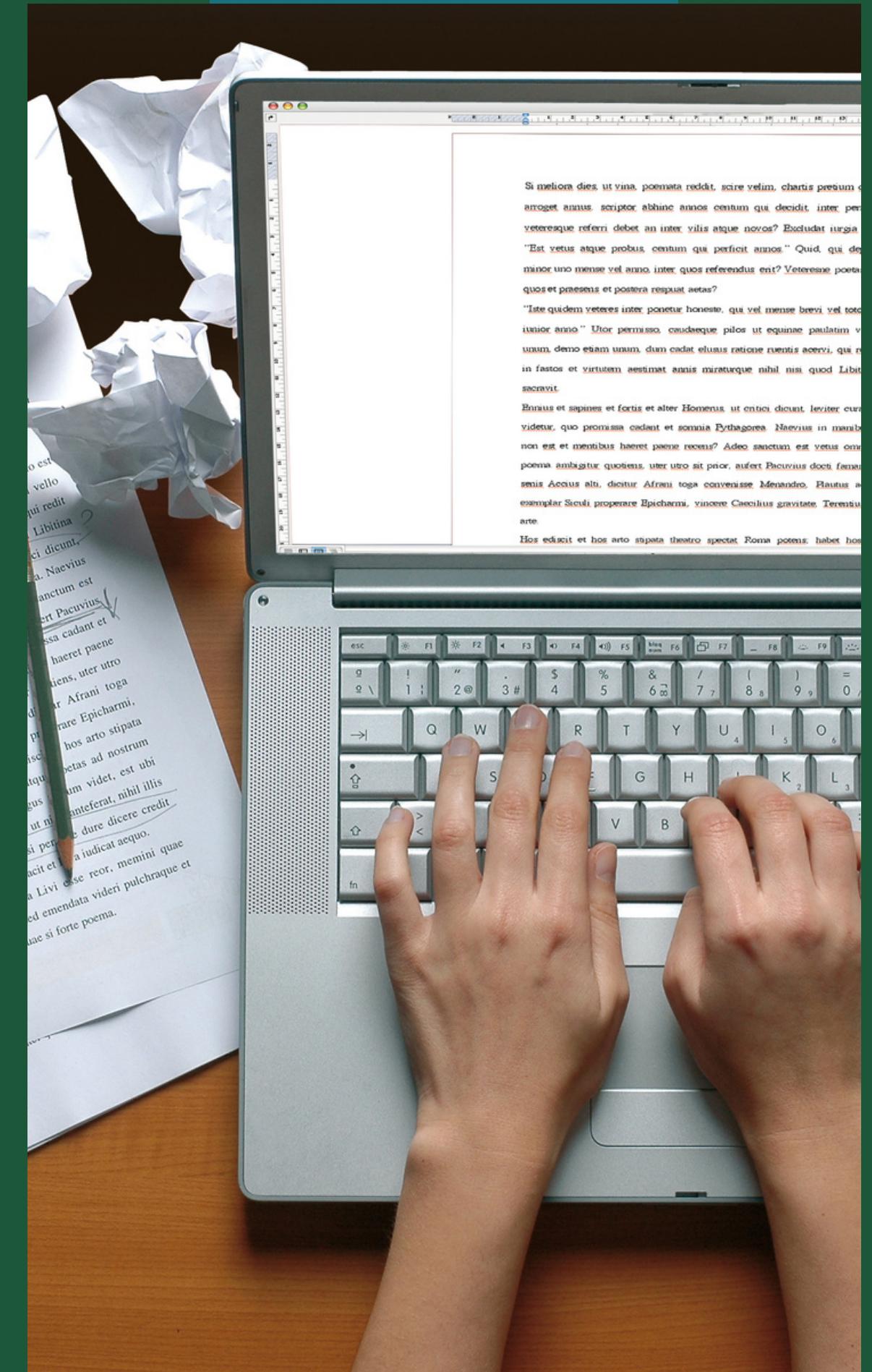




Matt Nelson

Presentation 2023

# LITERARY AWARD ANALYSIS



# Poets&Writers

DONATE | NEWSLETTER

PUBLISH YOUR  
WRITING

PROMOTE YOUR  
WRITING

FIND YOUR  
COMMUNITY

POETS & WRITERS  
MAGAZINE

[Home](#) › [Articles](#) › [May/June 2023](#) › Literary Prizes Under Scrutiny

## Literary Prizes Under Scrutiny

by Jen DeGregorio

NEWS AND TRENDS

From the [May/June 2023](#) issue of

Poets & Writers Magazine



# Problem

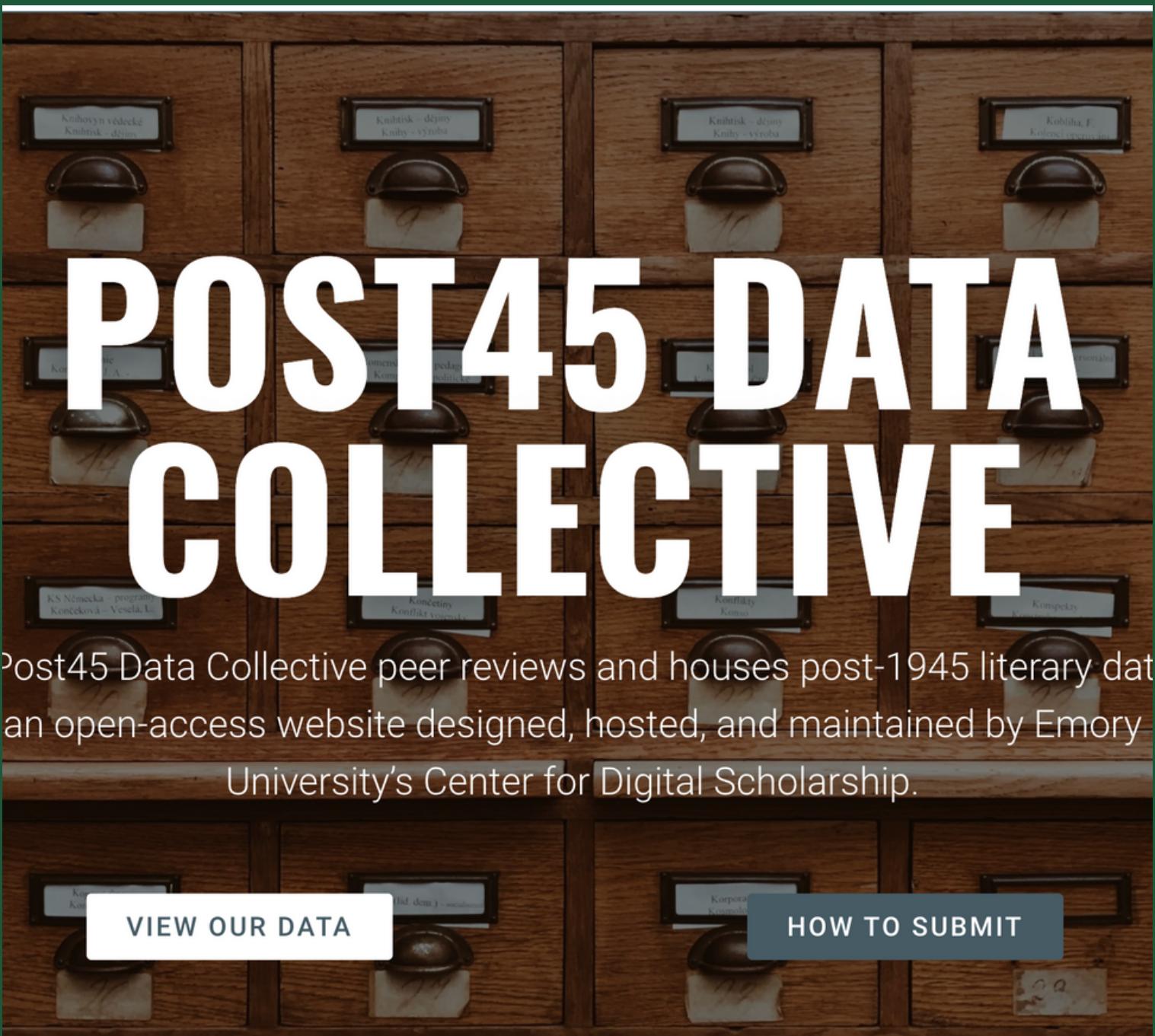
After reading in “Poets & Writers” about the work Juliana Spahr, Stephanie Young, and Claire Grossman did regarding the question “Who gets to be a writer?,” I decided to see what I could discover. Their question revolves around educational privilege as a placeholder for racial inequity.

80%

of recent major literary prizes winners have a graduate degrees.



# Dataset



# POST45 DATA COLLECTIVE

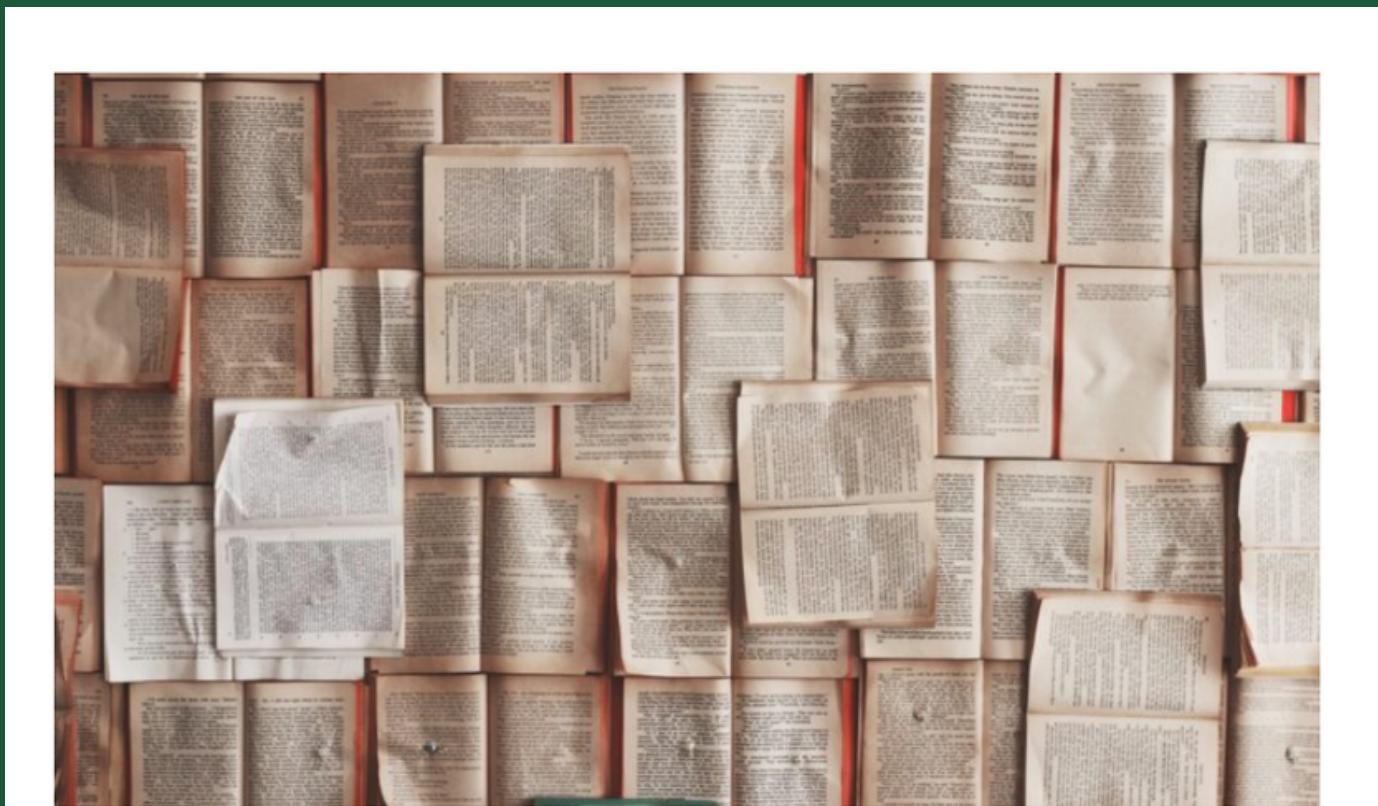
Post45 Data Collective peer reviews and houses post-1945 literary data an open-access website designed, hosted, and maintained by Emory University's Center for Digital Scholarship.

[VIEW OUR DATA](#)

[HOW TO SUBMIT](#)

*The Post45 Data Collective:  
Index of Major Literary Prizes in the US*

*Although this dataset does not have racial identifiers, it does have all major winners and judges of literary prizes over \$10,000 for over 100 years*



## THE INDEX OF MAJOR LITERARY PRIZES IN THE US

Authors: Claire Grossman, Juliana Spahr, and Stephanie Young

The Index of Major Literary Prizes in the US includes datasets to the winners and judges of prizes for prose, poetry, or unspecified genre between 1918 and 2020 with a purse of \$10,000 and over.

# Early Findings

## Judges > Winners

There were nearly double the amount of judges to winners within the dataset, which makes sense since awards often have multiple judges. For my purposes, I split the data et.

## University of Iowa

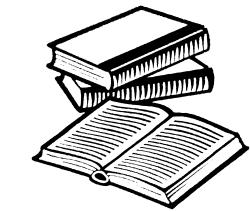
One of the oldest MFA programs in the country, the amount of winners with an MFA from the University of Iowa beat out the rest of the MFA programs.

## No Genre > Poetry > Prose

44% of the awards were given without a specific genre of the writing or the author. 31% of the awards went to poetry works. 25% of the awards went to prose works.

## \$10,000

By far the most amount of prizes given were for \$10,000, and the least amount of awards were given for \$200,000

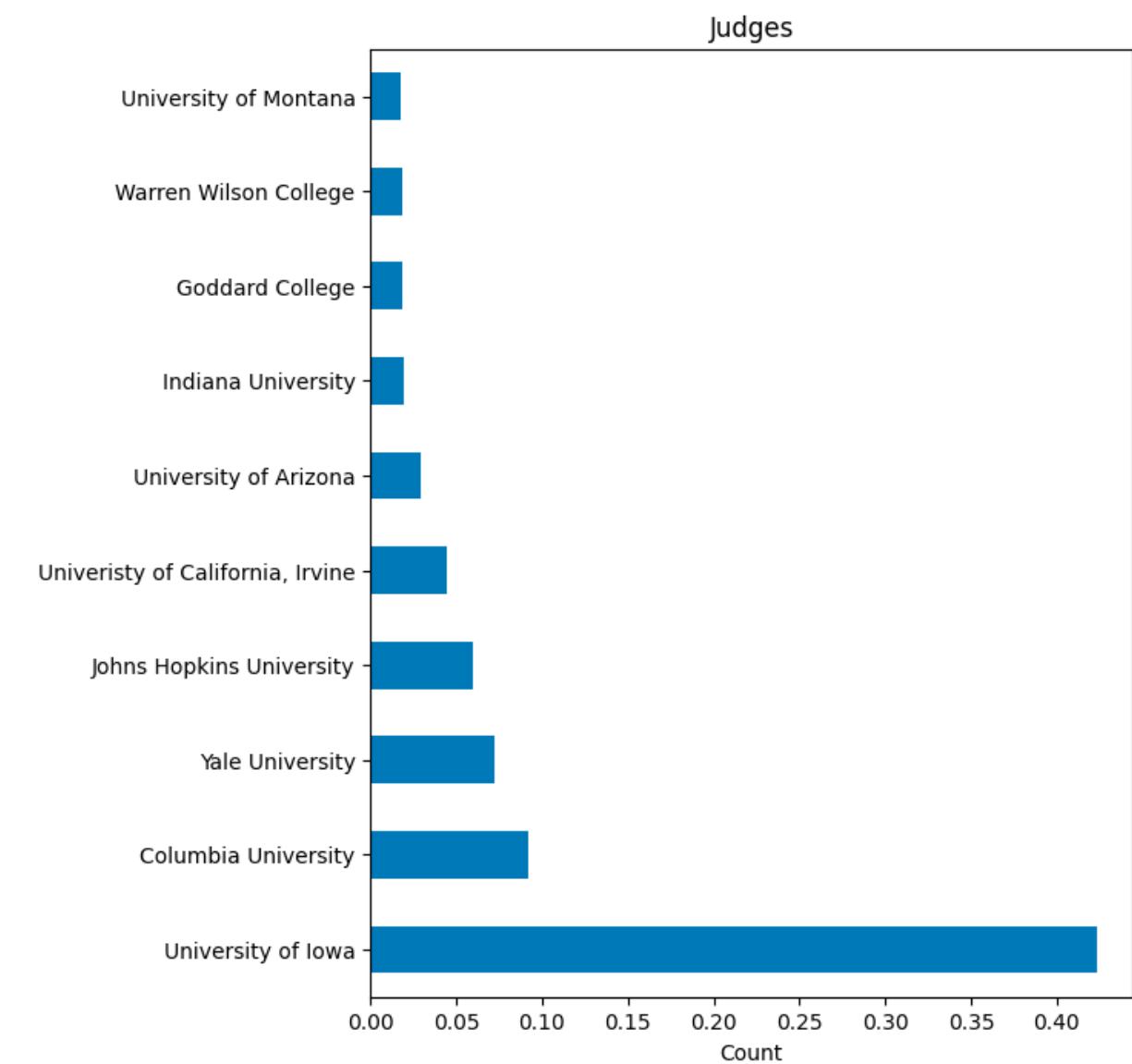
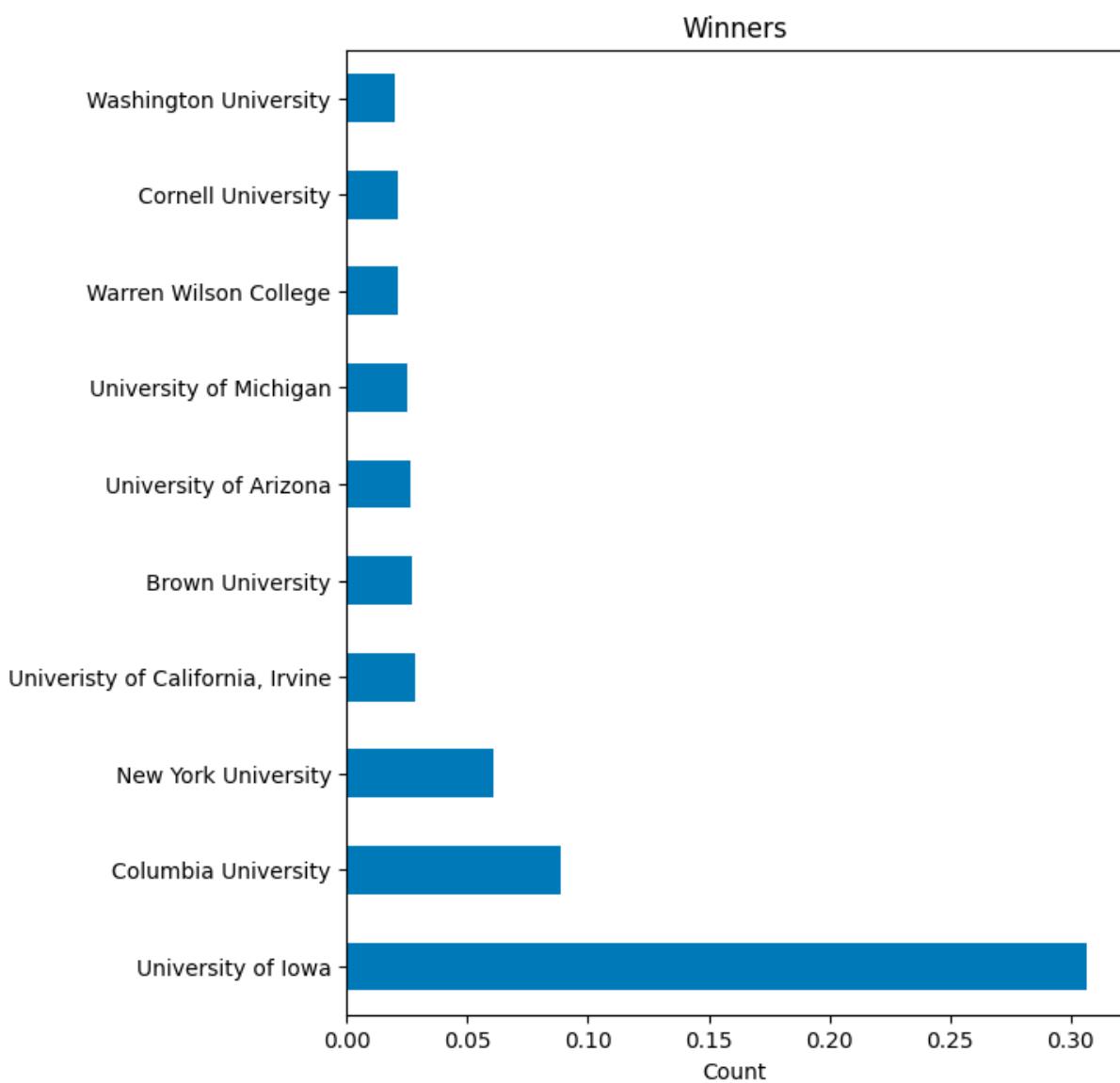


## Literary Awards

# The Supremacy of Iowa

Iowa's MFA program, also known as the Iowa's Writers Workshop, is renowned for being one of the best writing programs in the world. Looking at both the Winners and the Judges, we can see that after normalizing the dataset, over 30% of the winners with MFA degrees come from Iowa and more than 40% of the Judges with MFA degrees have one from Iowa.

Could there be a bias? Does going to Iowa secure you a place in the pantheon of literary award winners?



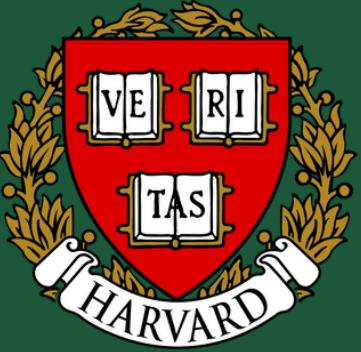
# MONEY BREAKDOWN

Sum totals won by different categories



Iowa

People with an MFA from Iowa have won over \$13,000,000 dollars in literary prizes. The next nearest is Columbia with just over \$4,000,000



Harvard

People with a graduate degree from Harvard have amassed close to \$12,000,000 in literary awards.



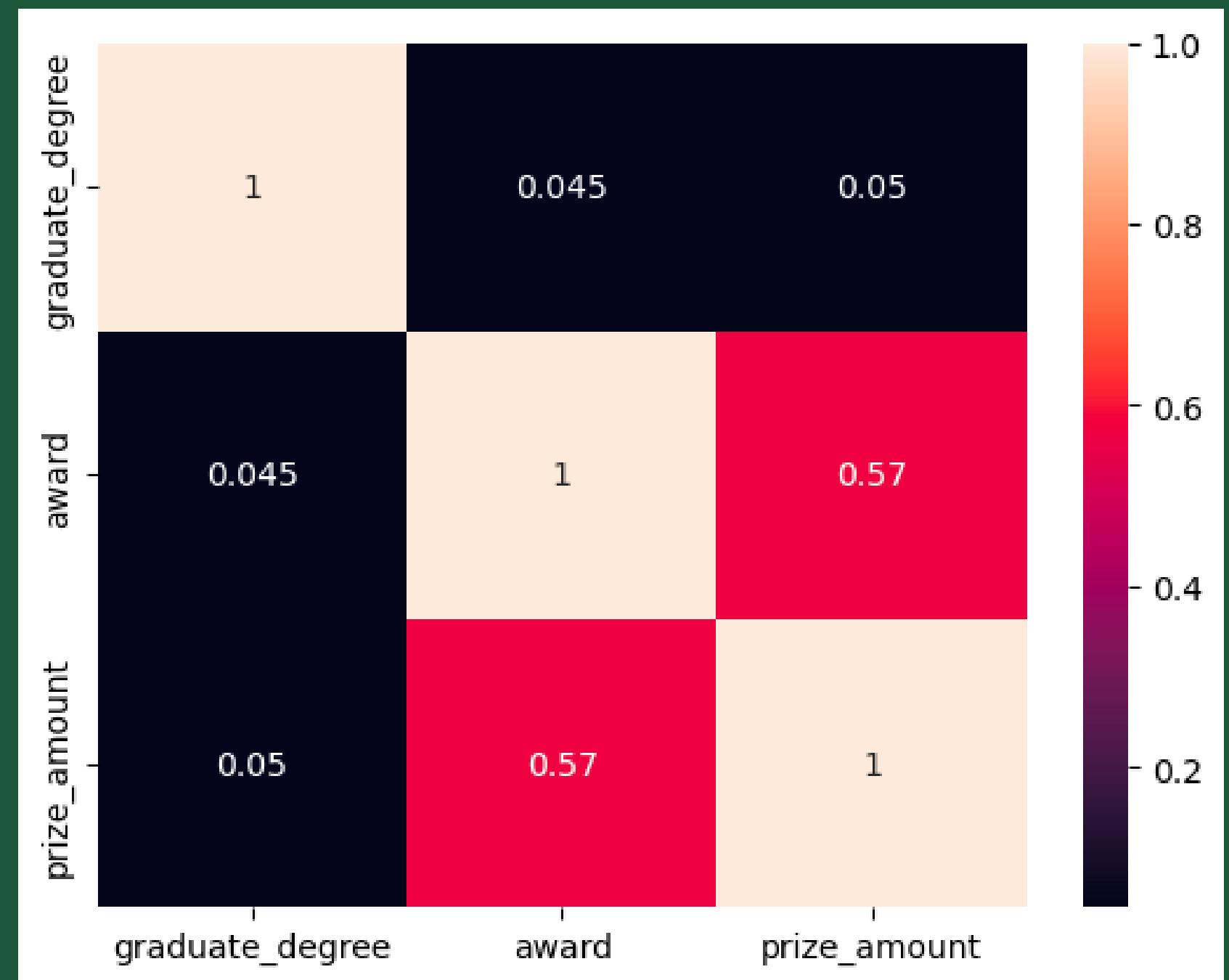
John Keene

This poet has made the most amount of money of anyone in the dataset with \$1,010,000. He has a degree from Harvard as well as an MFA from NYU.

# Correlation Heatmap of Numerical features

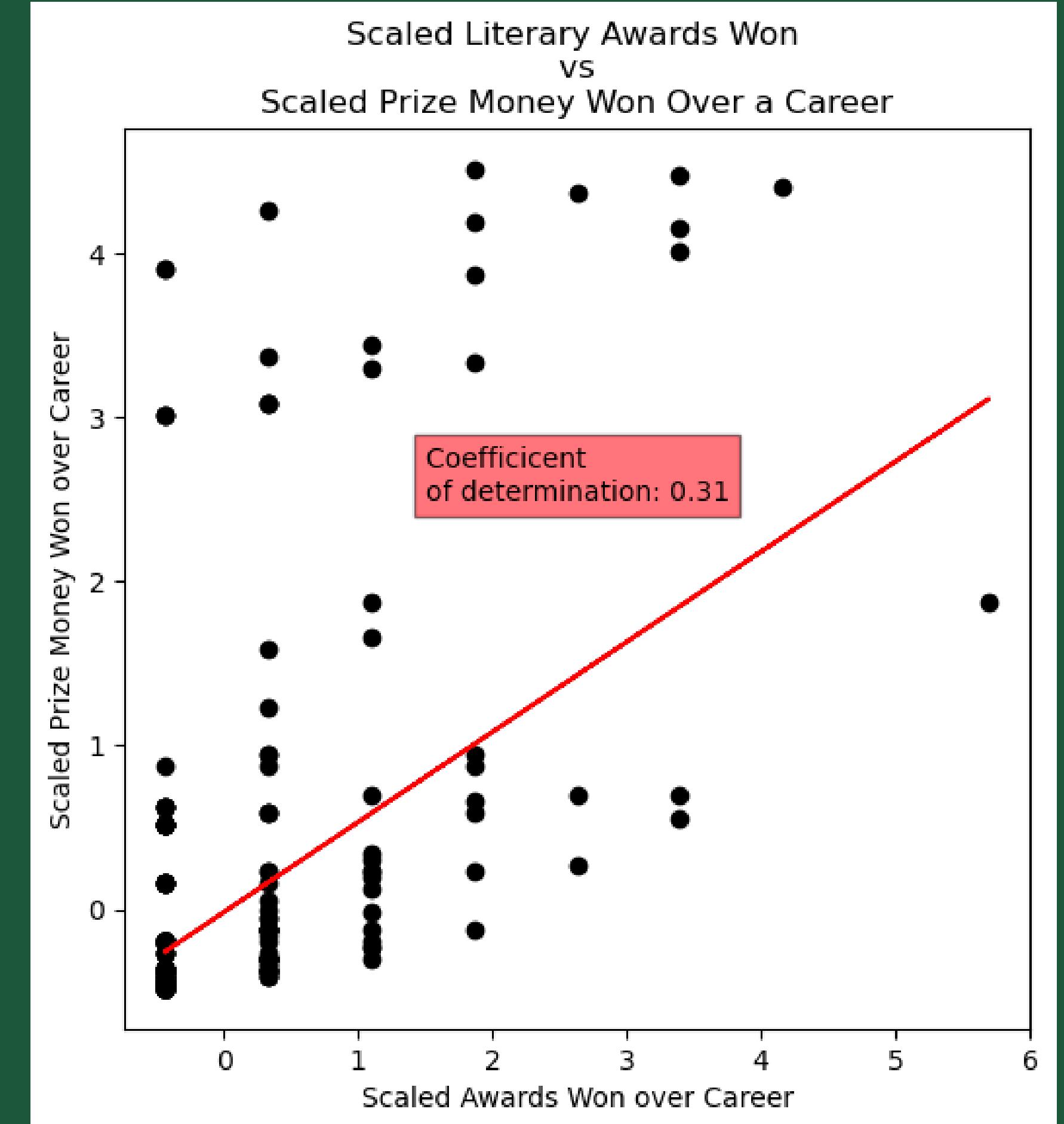
But what really is the driver for making a lot of money? For the numerical variables (after one-hot encoding graduate degrees), there is a natural correlation between the amount of awards won and the prize amounts. But just having a graduate degree didn't seem that important.

What does account for getting a lot of literary awards?



# Correlation Of Awards Won and Amount Won

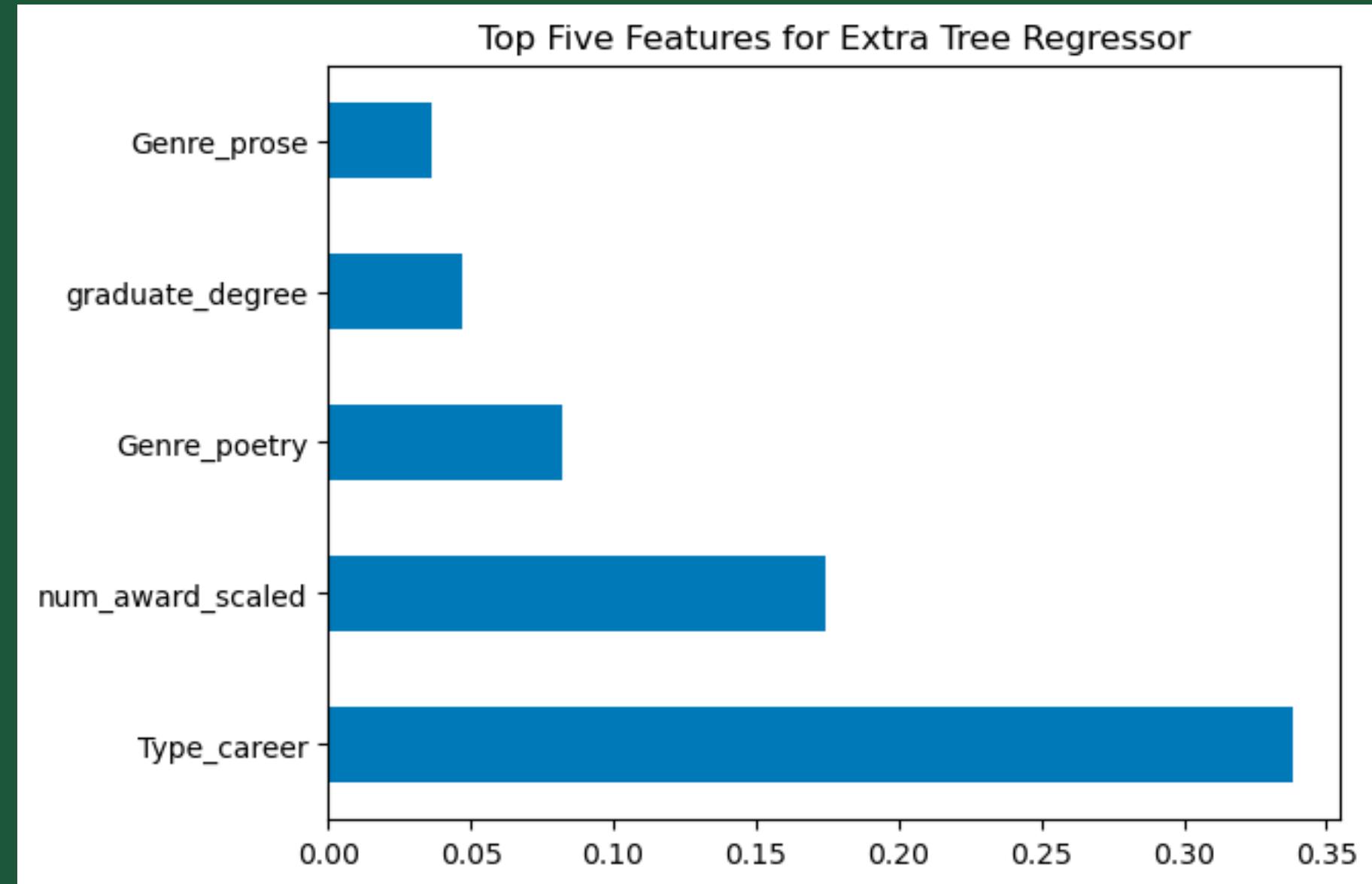
The first model I made was a linear regression tracking the correlation between scaled awards won and scaled amount won over an author's career. The coefficient of determination was 0.31, not as high as I would have imagined.



Model: 1  
RSS: 14.4%

# ExtraTree Regressor

I tried a few different Machine Learning models, but one of the first was the ExtraTree Regressor. The top five features for the target variable of sum of award winnings were the type of award (career), the number of awards won, the genre (poetry), having a graduate degree, and then another genre (prose).



The following graphs show the different models' features as well as their R-squared score (RSS)

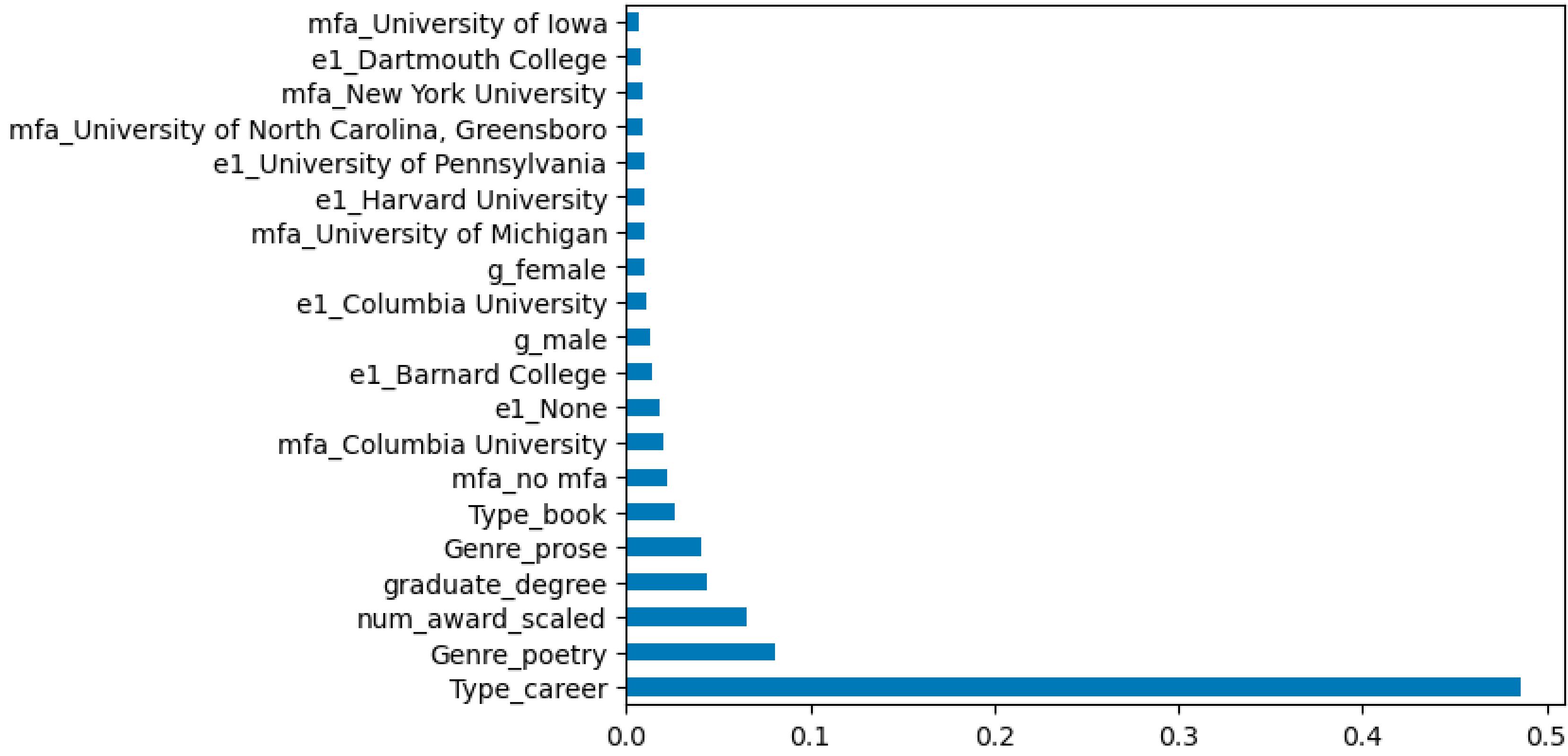
The only problem was the low R-squared score: 14.4%

# Random Forest Features

Model: 2

RSS: 24.2%

Top Twenty Features of Random Forest Regressor

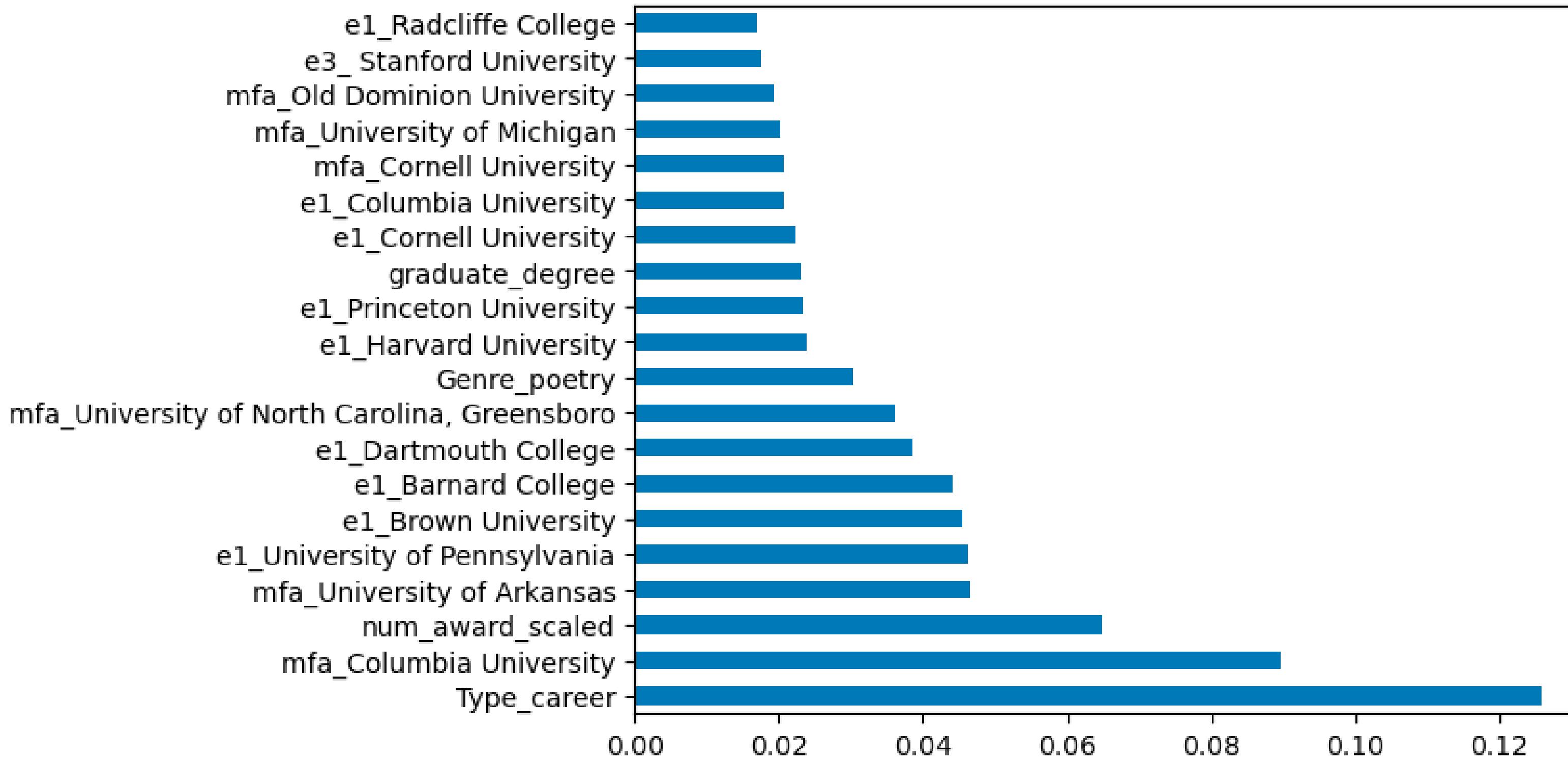


# XGBoost Features

Model: 3

RSS: 14.5%

Top Twenty Features of Importance for XGBoost

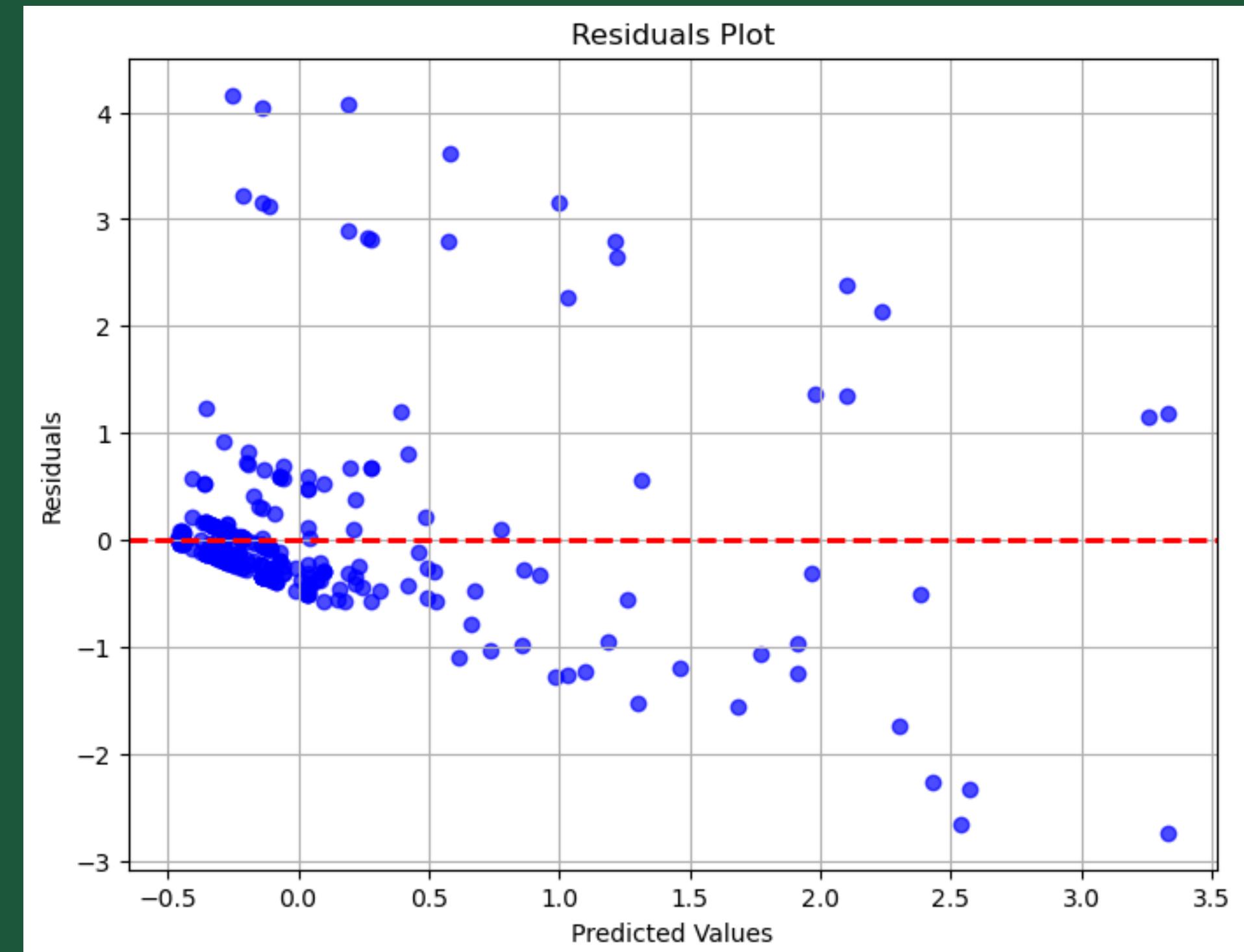


# Random Forest Features

Model: 4  
RSS: 31.7%

Since the Random Forest Regressor was the best model, I decided to tune the hyperparameters and got a better RSS.

Here is the residual plot of the predicted values vs the residuals from the actuals. The closer to the red line, the better!

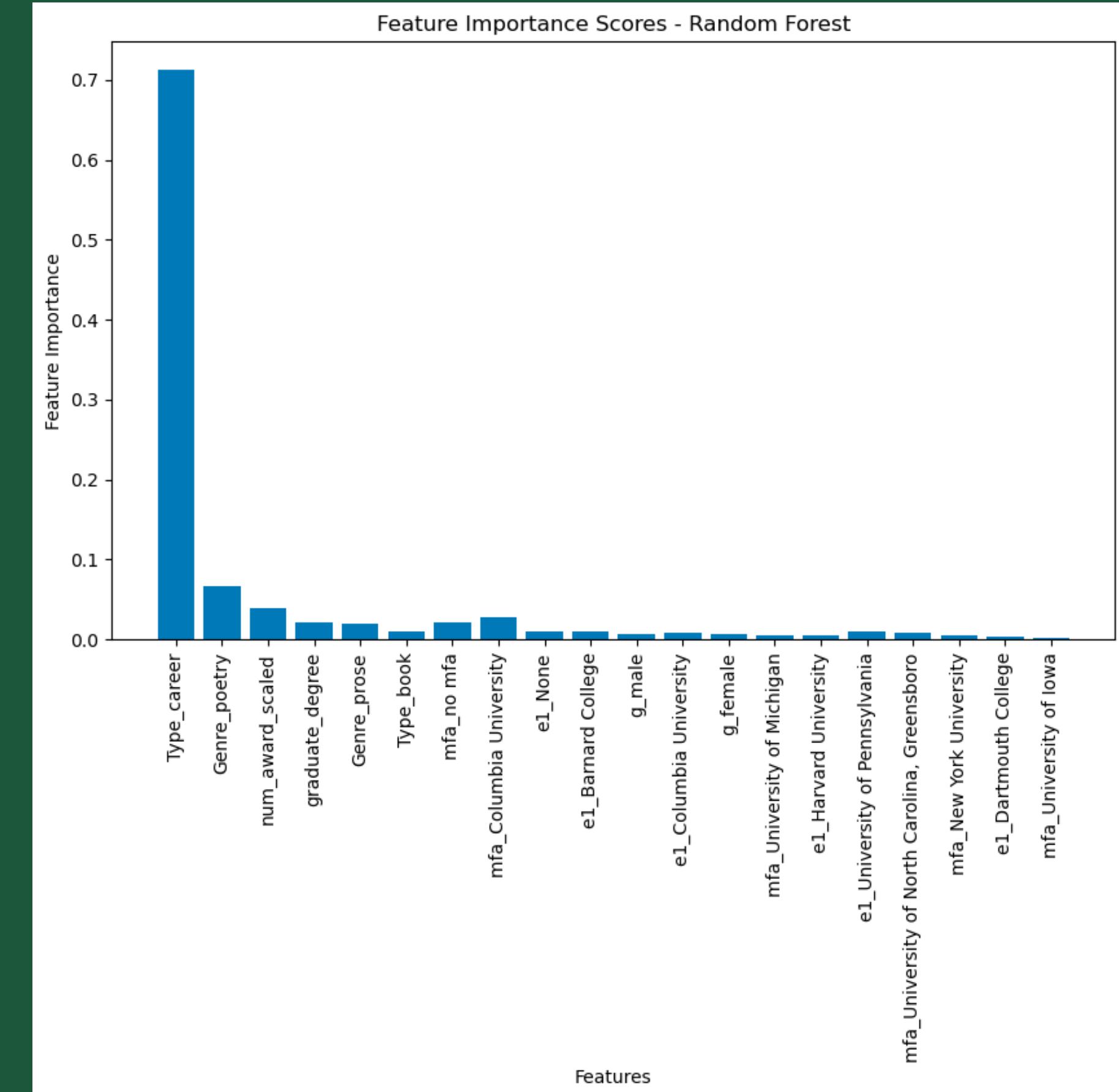


# Random Forest Features

Model: 4  
RSS: 31.7%

The top five features for the Random Forest were the type of award (career), genre (poetry), the number of awards, having a graduate degree, and genre (prose).

Having an MFA from Iowa barely makes the top 20 features.



# Now/ Future

01

It turns out that the biggest determinant within my model for the sum of money won in a career is the amount of career awards an author wins. Potentially, if someone wins a career award, they might also have a lot more book awards. Or perhaps the awards with the most money are for career achievement.

02

Seeing that Iowa was in the bottom of the top 20 features for my Random Forest Regressor, I would like to try an unsupervised classification model to see if it could predict which MFA program someone went to based on other features including amount of money won over a career.

