Set Up For Success

By: Felix, Swathi, Rachel, Ryan

Background on the Oscars

What are the Oscars:
- Also known as the Academy Awards
- Annual ceremony to recognize the film industry, founded in 1927

Why they matter:
- Considered the highest honor in film industry
- Boosts careers of winners & positively impacts box office sales

Problem Statement

We are a team of data scientists consulting producers and production companies who want to increase their chances of winning an Oscar. We will provide insights into whether a movie is likely to win in at least one of the 23 Oscar categories. Using multiple datasets of past Oscar-nominated and winning films, we will build a binary classification model to predict Oscar successes based on features such as genre, runtime, director, etc. The model will be evaluated using accuracy, and F1 score. The goal accuracy of the model is to outperform our base accuracy score.

Data Overview

- Obtained 2 datasets, containing a variety of features, regarding nominated and winning Oscar films

- Final merged data frame contains 4642 rows and 89 columns

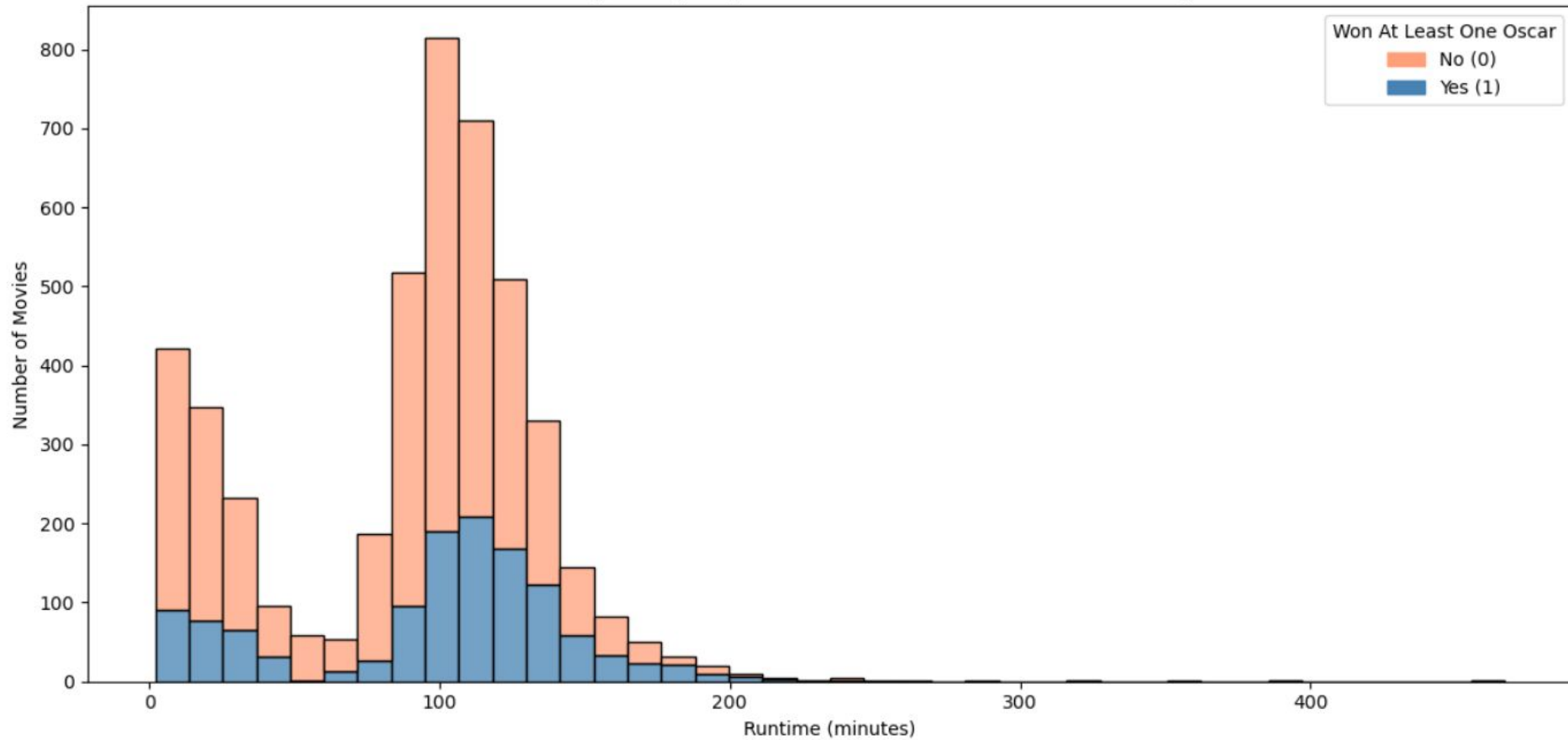- Analyzing data from the years 1928 – 2024

Approach

Data cleaning:
- Removed unnecessary columns and cleaned text
- Mapped award categories
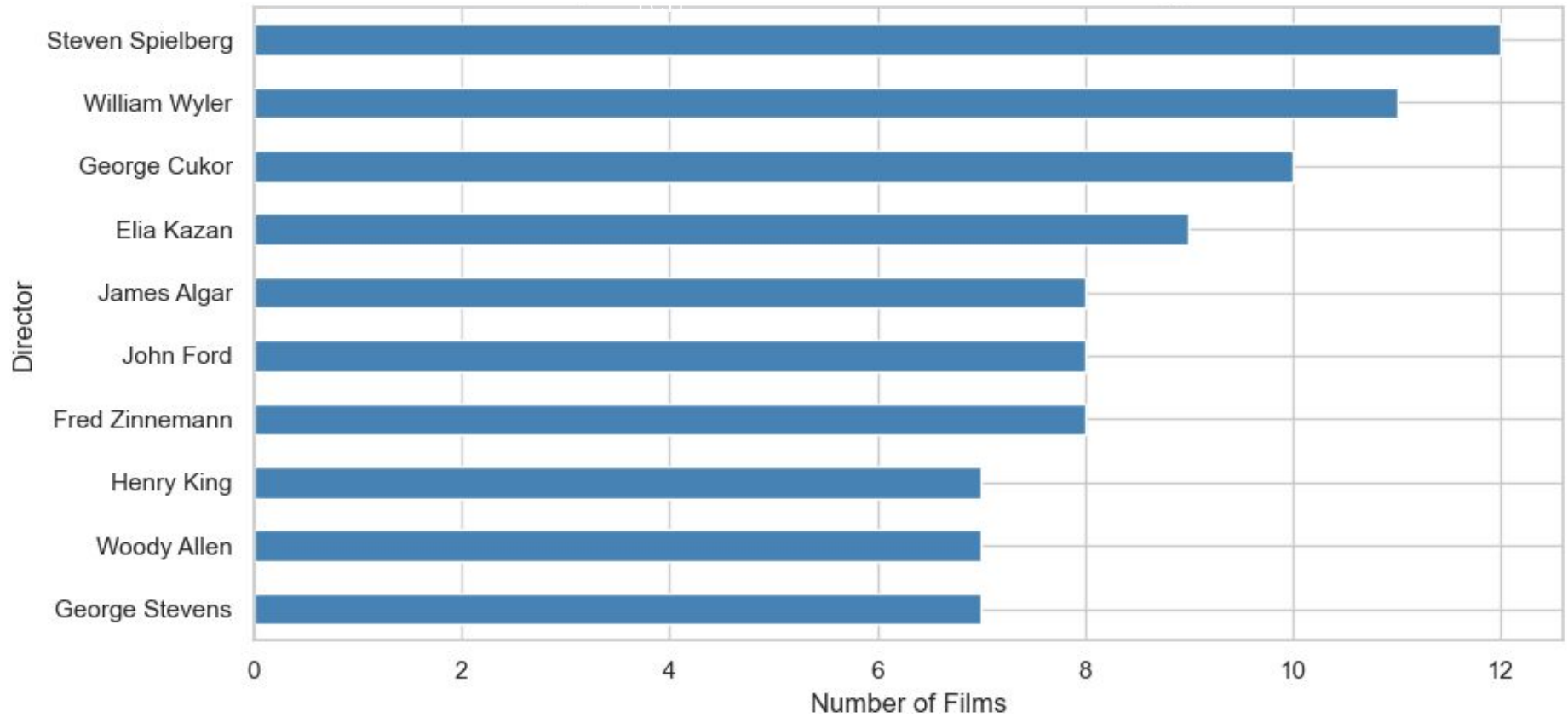- Merged the 2 dataframes on film title

Feature Engineering:
- One hot encoded award category wins & nominations
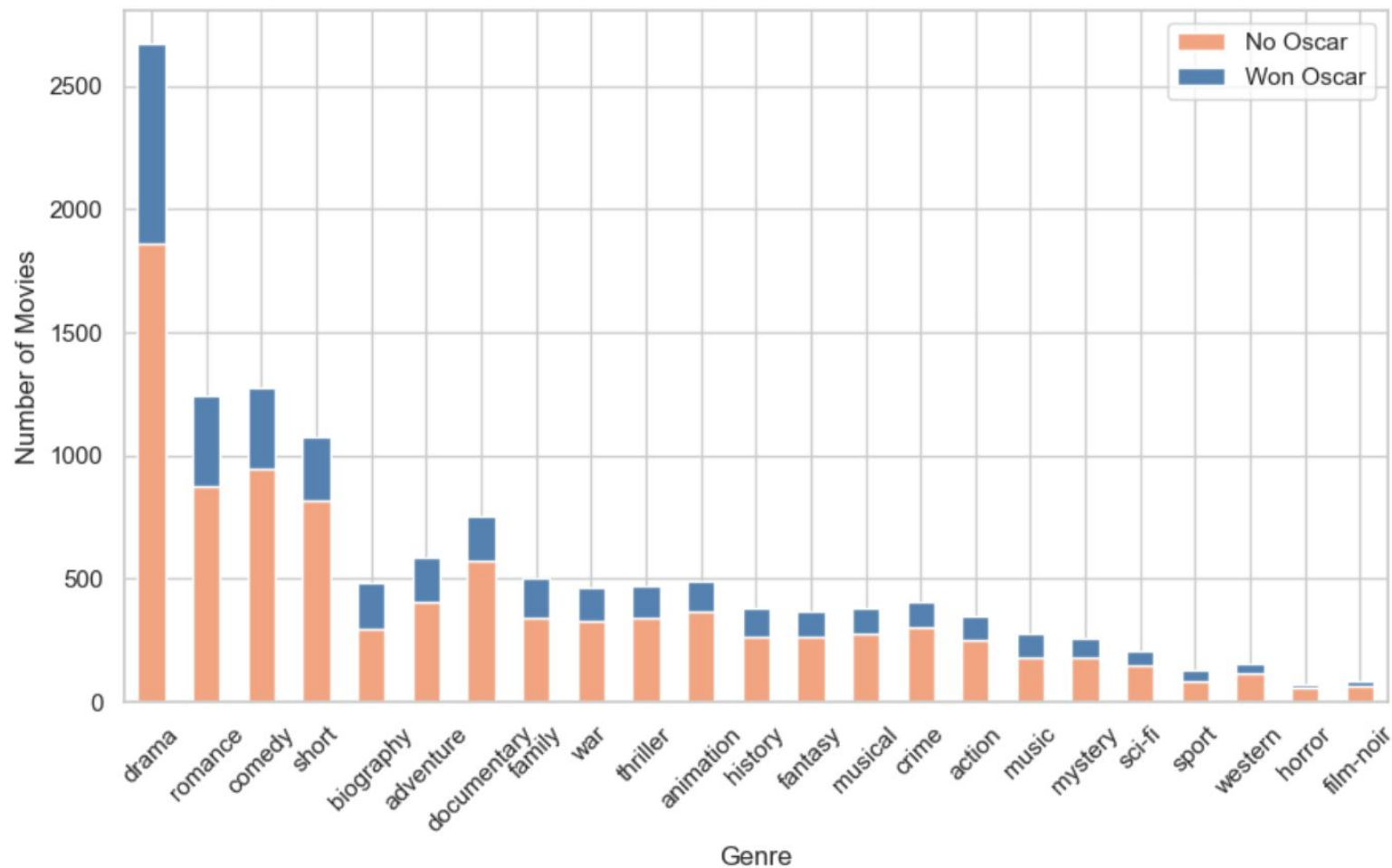- One hot encoded genre categories

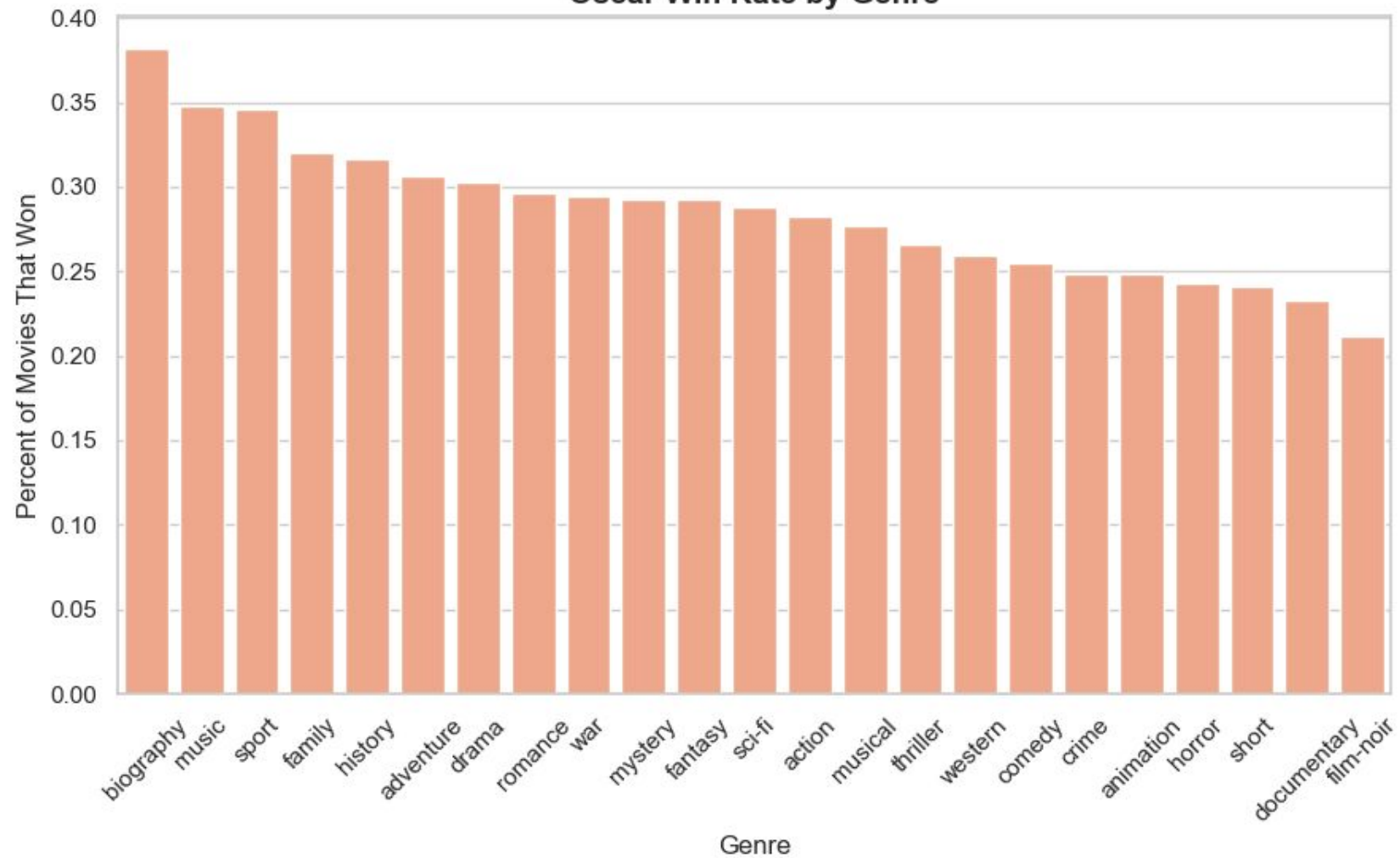Oscar-winning films (blue) have a runtime of 120–150 minute range.
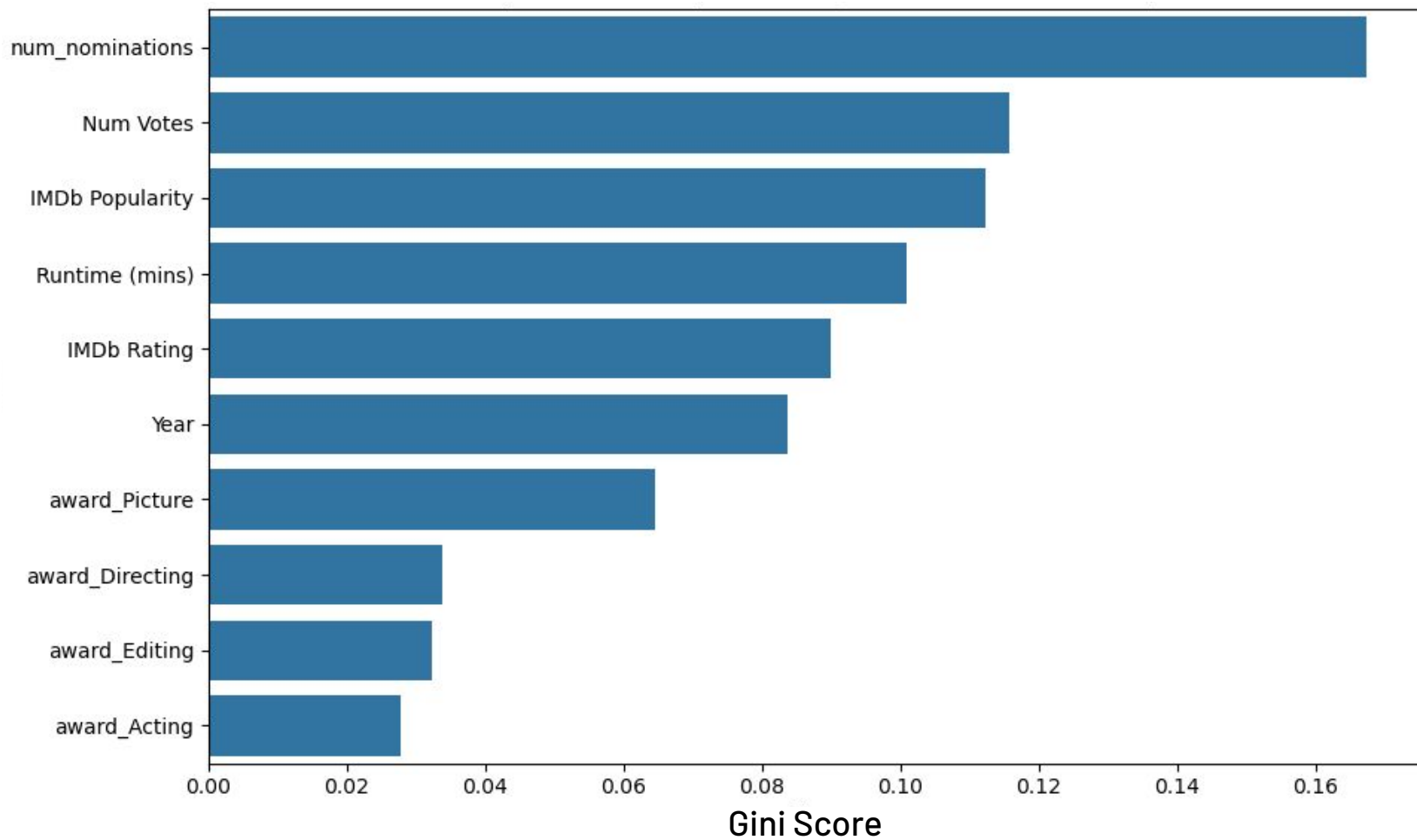
Ten Directors to Hire

A horizontal bar chart titled "Ten Directors to Hire" showing Number of Films per Director:

Director	Number of Films
Steven Spielberg	12
William Wyler	11
George Cukor	10
Elia Kazan	9
James Algar	8
John Ford	8
Fred Zinnemann	8
Henry King	7
Woody Allen	7
George Stevens	7

More Oscar Competition in Drama, Romance, Comedy, and Shorts

Oscar Win Rate by Genre

The Number of Nominations is the Highest Correlated Feature

Model Evaluation

Baseline Model: Score of 27% predicting class 1 (oscar win)

Chosen model: Random Forest Classifier

Challenges:
- Class imbalance
- Number of observations

Model Performance:
- Accuracy: 80 %
- F1 Score: 62 %

Recommendations/Conclusions

Did our model meet our Goals?
- Base accuracy score: 27 %
- Model accuracy score: 80 %

Recommendations for producers and production companies
- Focus on biography, sport, & music genres
- A movie length ranging from 120-150 minutes
- Hire the top 3 directors: Steven Spielberg, William Wyner, George Cukor

Improvements and refinement of model
- More data features: leading actors, predicted box office earnings, etc.

Questions?