

# Rotten Tomatoes and the Oscars: A Data Analysis Project

## Background

It's hard to deny the cultural relevance of Rotten Tomatoes, a review website that aggregates critics' ratings and calculates a score that's either "fresh" or "rotten". But how accurate is the self-proclaimed "most trusted measurement of quality for movies" really? Enter another cultural giant: the Oscars. Regarded by many as the most prestigious awards ceremony in the entertainment industry, the ceremony's top prize of "Best Picture" is widely viewed as the highest honor a movie can receive. In this data analysis, we set out to discover how closely these two variables are related. Can Rotten Tomatoes scores be used as a reliable predictor of a film's Oscar success?

## Data

All data used in this project is taken from Kaggle. The analysis begins from three raw datasets:

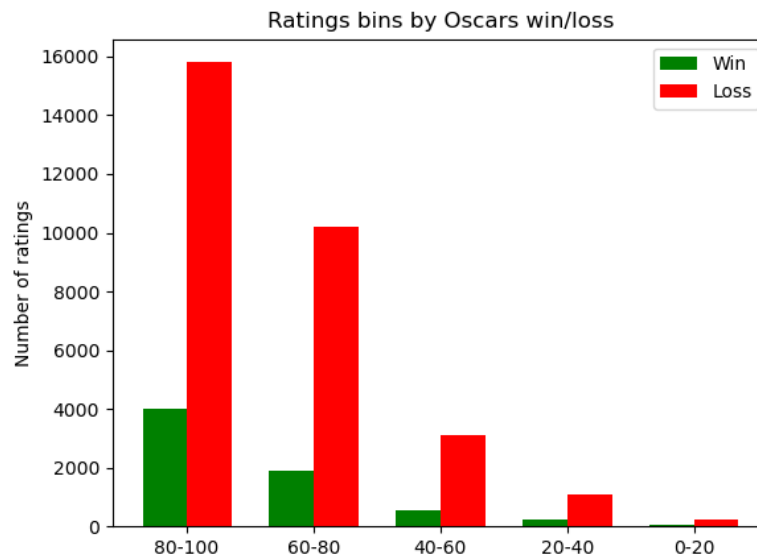
1. [rotten\\_tomatoes\\_movies.csv](#), which provides aggregated Rotten Tomatoes scores
2. [rotten\\_tomatoes\\_critic\\_reviews.csv](#), which provides individual critic reviews
3. [the\\_oscar\\_award.csv](#), which provides all historical Oscar nominations and wins

Our analysis pipeline downloads these three Kaggle datasets, merges them, and cleans the final merged dataset for analysis. Those interested in viewing and/or analyzing the data on their local machine should follow the process described [here](#).

## Results

***Historically, how well have rotten tomatoes critic scores correlated with "Best picture" Oscar wins?***

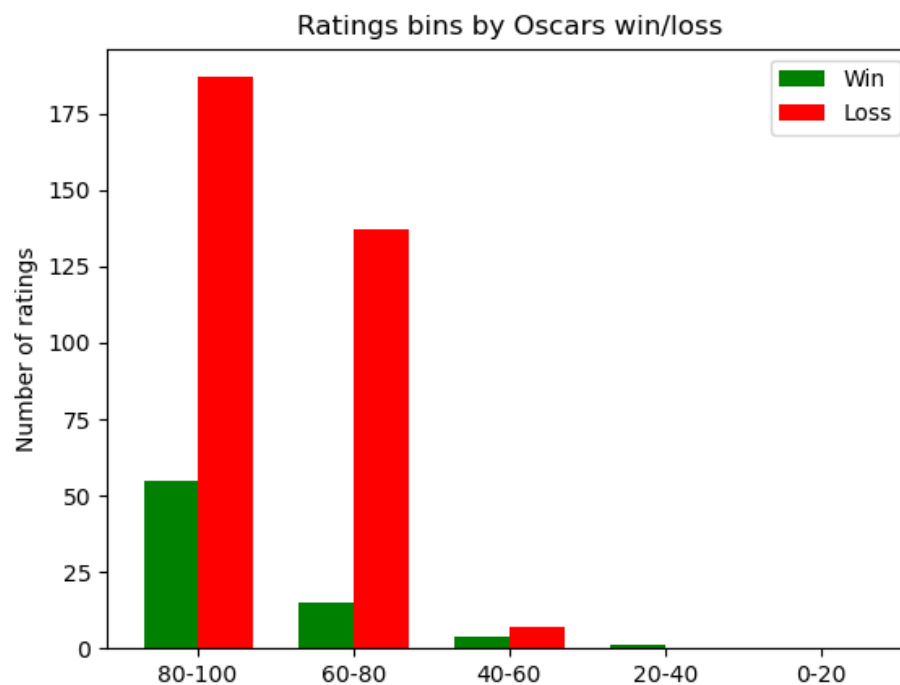
To find the answer to this first research question, we ran a simple logistic regression of critic rotten tomatoes score to best picture wins. When plotting the data, you see that many winning movies have low Oscar scores.



It seemed like there might be an outlier issue, with even very popular movies getting low ratings from a few critics. For example, here are the critic ratings for the film "American Beauty", which won the best picture Oscar in 1999.

	review_score	num_reviews
0	100.0	47.0
4	95.0	9.0
12	94.0	1.0
15	92.0	1.0
3	90.0	9.0
6	88.0	4.0
1	87.5	11.0
11	85.0	1.0
2	80.0	11.0
5	75.0	7.0
14	73.0	1.0
10	70.0	2.0
9	62.5	3.0
8	60.0	3.0
7	50.0	4.0
13	40.0	1.0
16	25.0	1.0

So we decided to take the average of Rotten Tomatoes scores instead, which gave the following breakdown:



In the first runs of the model, we noticed that the model was only ever predicting losses. Therefore, we increased the class weights to ensure that the model was predicting at least one Best Picture win. We ran some hyperparameter optimization to ensure that we were picking the class weights with the highest test accuracy.

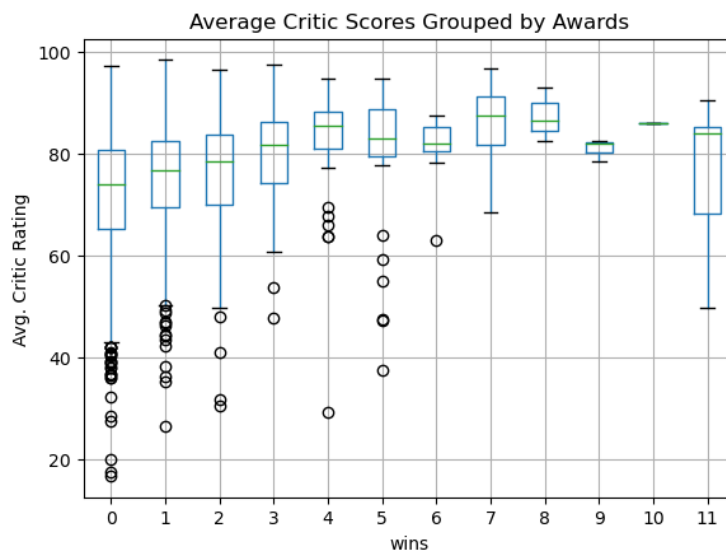
However, even with these improvements, our best any-win model coefficient was 0.023. When the model was allowed to predict no wins, its regression coefficient was 0.028. This tells us that Rotten Tomatoes critic score is not a strong predictor of a Best Picture Oscars win.

The highest any-win model test accuracy we saw was 83.3%, while the no-win model was 84.3%. This tells us that predicting no wins at all leads to higher performance with this model architecture.

To improve these results, you might need to choose a more complicated model than logistic regression, or add in other explanatory variables.

### ***Historically, are Rotten Tomatoes scores a good predictor of Oscars success?***

To answer this second question we first define “Oscars success”. We defined Oscar success in two ways. Firstly, winning an award at the Oscars, this definition is simply true or false depending on if the film won any award at the Oscars. Secondly, we defined success in terms of the number of awards a film won at the Oscars, with more awards being more success.



Looking at the distribution of ratings across levels of Oscar’s success it initially appeared as though Rotten Tomatoes critic reviews may be predictive of Oscar’s success

With these definitions we conducted correlation and regression analysis. Looking at our correlation analysis we found that the average critic score was not highly correlated with either the number of awards won nor general success. Below is a heatmap showing this analysis as well as a few other metrics we investigated.



### Ideas for Further Analysis

The analysis contained in this repository only scratches the surface of what can be discovered with this data. Here are some potential questions for future research:

- Does another model type (besides logistic regression) more accurately fit the relationship between Rotten Tomatoes scores and Oscar wins?
- Which Oscar award most strongly correlates with Rotten Tomatoes scores?
- Is there an interaction effect between genre and Rotten Tomatoes score on Oscar wins?
- How closely do audience scores correlate with critic scores?
- Historically, which critics have most accurately predicted Oscar success?
- How do outside factors (budget, box office sales, etc.) affect Oscar win probability?
- Do scores from top critics correlate more strongly with Oscar wins than non-top critics?