

Sentiment Analysis of Theatrical Reviews

Are we able to find correlation between a critic's review of a show and the show's financial success? How much of an effect does genre or show type have, if any? These were the questions guiding this project, one that I am afraid was too broad of a scope to accurately analyze in a short amount of time. Though the results of my analysis are inconclusive, I was able to see this project through from problem identifying and data collecting to the end where I visualized my findings. This project was a good exercise in solidifying the skills I have learned thus far and will inform future decisions on projects I want to take on.

Because there are no datasets readily available that hold the kind of reviews I was looking to analyze, I had to find a way to collect the data myself and transform it into something that I could run through various Python packages and libraries. I used Scrapy to gather reviews from NewYorkTheatreGuide and financial data from Playbill. I scraped 418 articles from NYTG and over 50,000 entries from Playbill. I would have liked to scrape more reviews from other websites, but because of my time constraint and the amount of time I was taking to learn Scrapy and get clean crawls, I had to stop at one website. Because I have this experience now, in the future if I decide to tackle a similar project I feel more confident that I'll be able to have a wider selection of sources that I get my data from.

I was not able to find strong evidence one way or another that there were factors of the reviews that influenced how a show did on Broadway, and it's likely because I simply didn't have enough data. There was some weak correlation with an article's sentiment score and the weekly gross and seats sold, so there is some promise there that reviews do have a positive influence on these factors, but nothing can be concluded from what I have right now.

I also took a look at whether genre and show type had an influence on a show's success and got back mixed results - again, probably a signal that I should have collected more data. There is clearly compelling evidence that any sort of 'Family Friendly' show will do well, especially if it's a musical, but any conclusions further than that cannot be made in confidence.