

Sentiment Analysis for Theatrical Reviews

By Kai Tamashiro

The Questions

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?





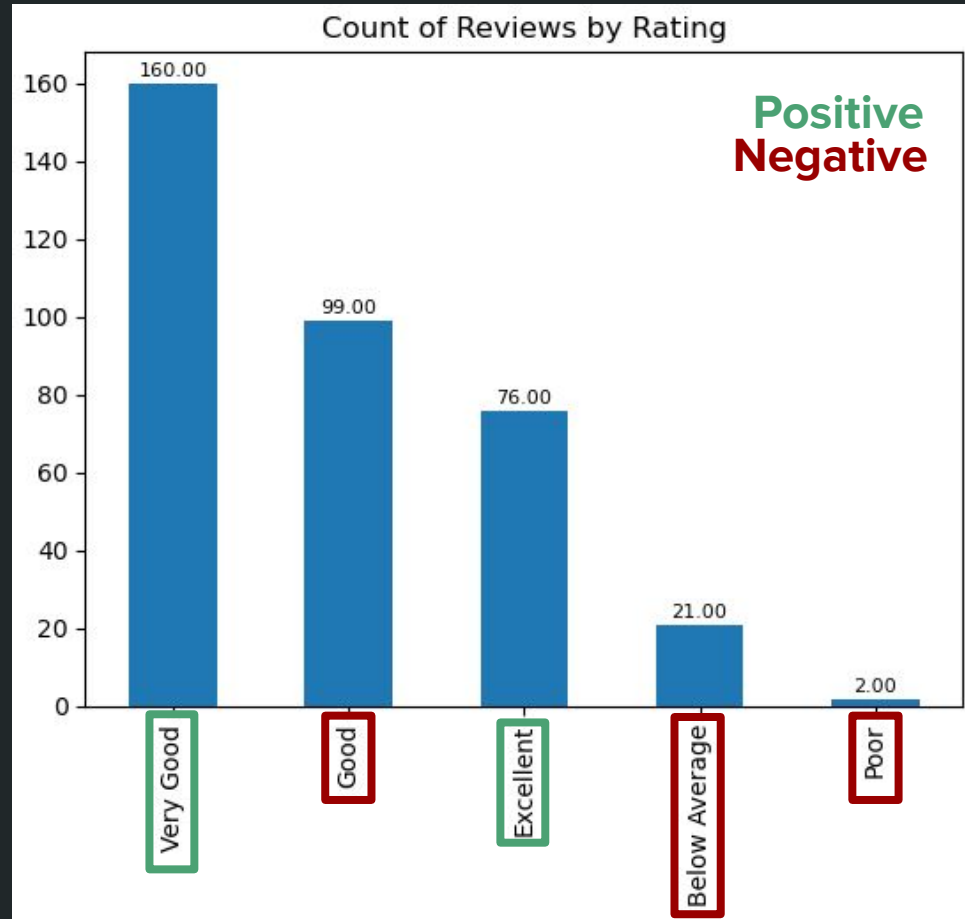
Scrapy

**New York
Theatre
Guide**

358 reviews

PLAYBILL®

~50,000 rows



Genres and Show Examples

Classic - *Cats, Chicago, Fiddler on the Roof*

Comedy - *Mrs. Doubtfire, Mean Girls, The Play That Goes Wrong*

Contemporary - *Kinky Boots, The Visitor, The Headlands*

Dark Comedy - *Into the Woods, Little Shop of Horrors, Kimberly Akimbo*

Drama - *The Kite Runner, Life of Pi, Harry Potter and the Cursed Child*

Family Friendly - *Frozen, Aladdin, The Little Prince*

Historical - *Titanic, Richard III, Rock and Roll Man*

Jukebox - *A Beautiful Noise, Once Upon a One More Time, & Juliet*

Revue - *Bob Fosse's Dancin', Assassins, Forbidden Broadway*

Satire - *Wives, The Thanksgiving Play, Ain't No Mo'*

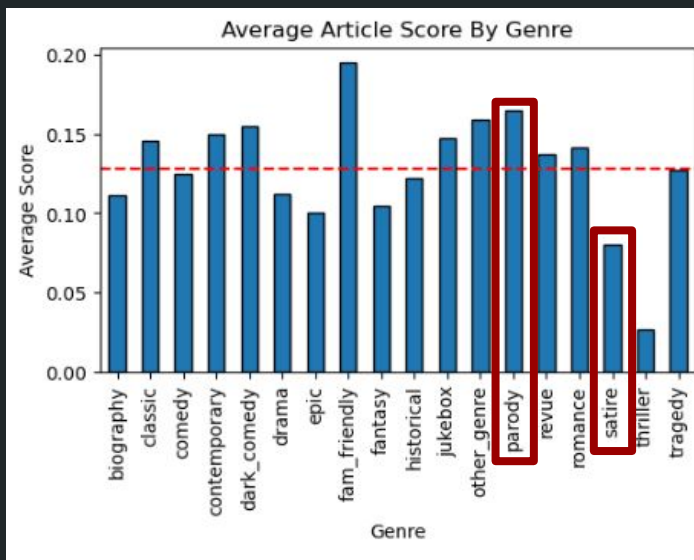
Thriller - *Grey House, Selling Kabul, Our Dear Dead Drug Lord*

Tragedy - *Hamlet, The Best We Could, The Tragedy of Macbeth*

Other genre - *Hadestown, Hamilton, Here Lies Love*

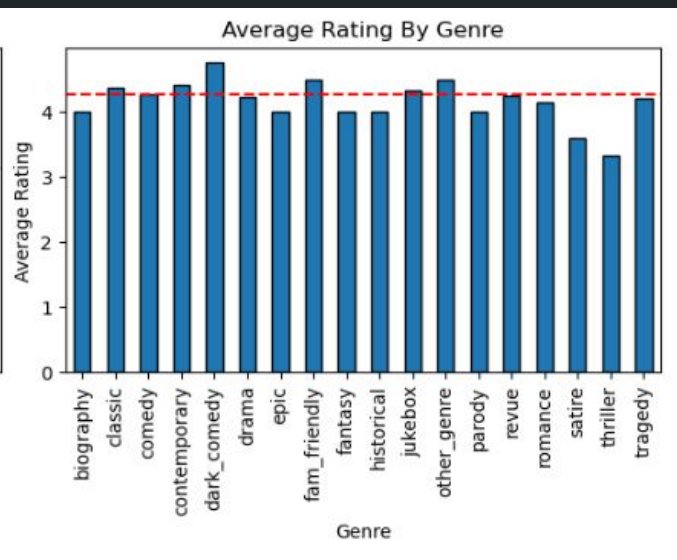
Article Sentiment Score vs. Show Rating

Higher variance



Tells a better story

Lower variance



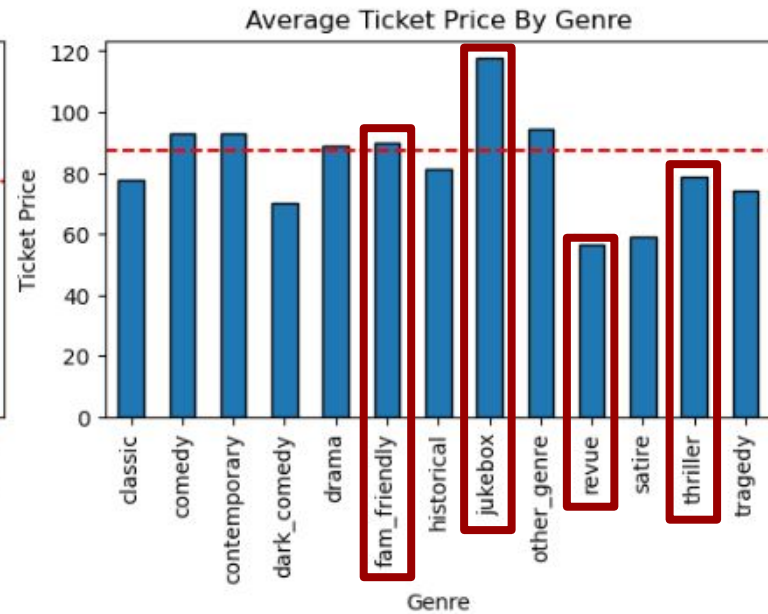
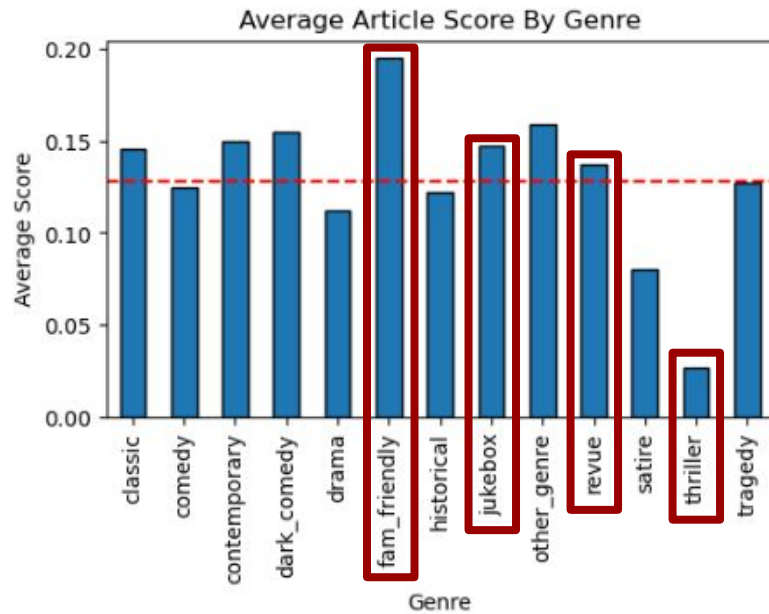
Not a lot to infer



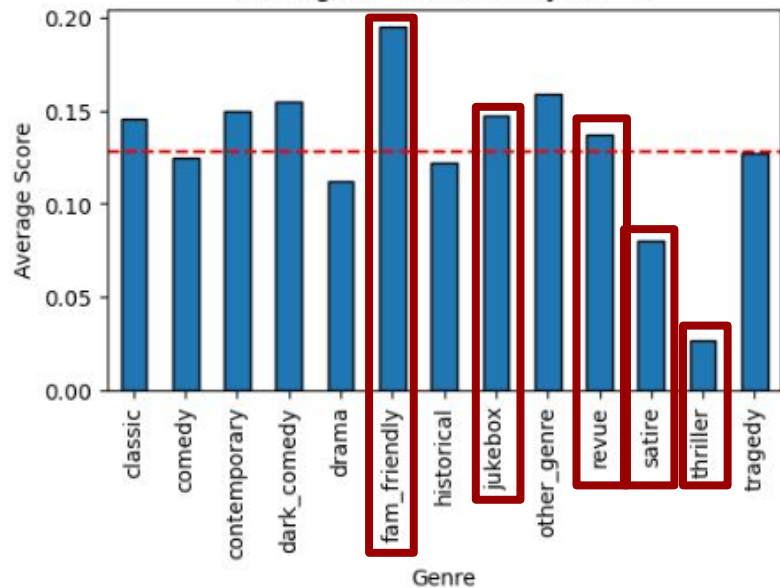
A heatmap visualization showing the correlation coefficients between various movie features. The features are listed on the y-axis, and the correlation coefficients are shown in a grid. The 'article_score' column is highlighted with a green oval, indicating it is the focus of the analysis.

week_gross	-0.0015	0.063	0.35	0.15
avg_ticket	-0.043	0.025	0.26	0.17
top_ticket	-0.12	-0.03	0.24	0.13
seats_sold	-0.031	0.076	0.42	0.14
seats_in_theater	-0.064	0.051	0.34	0.099
num_performances	0.12	0.14	0.36	0.069
capacity_filled	0.035	0.064	0.2	0.15
	rating	fixed_sent	article_score	article_subj

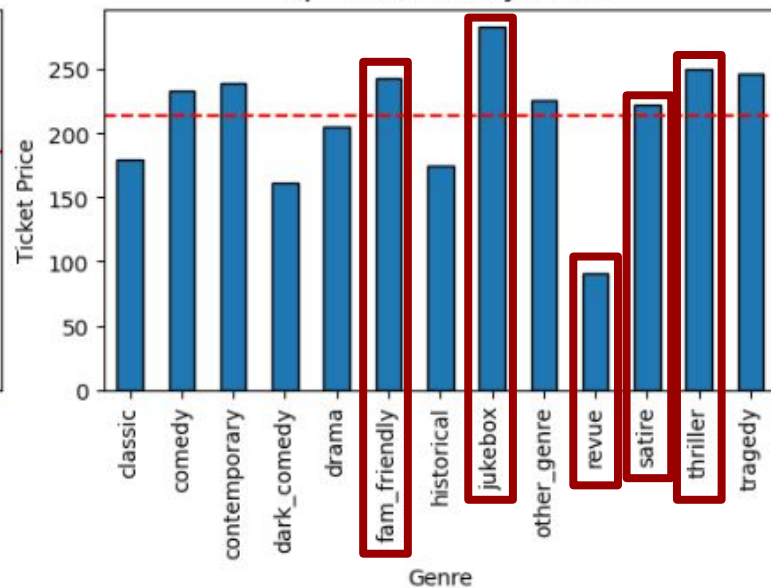
Our focus

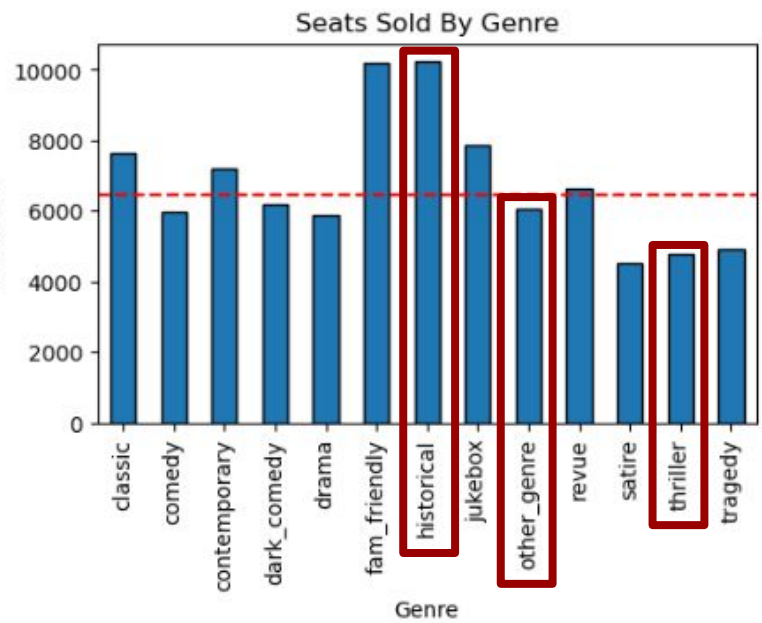
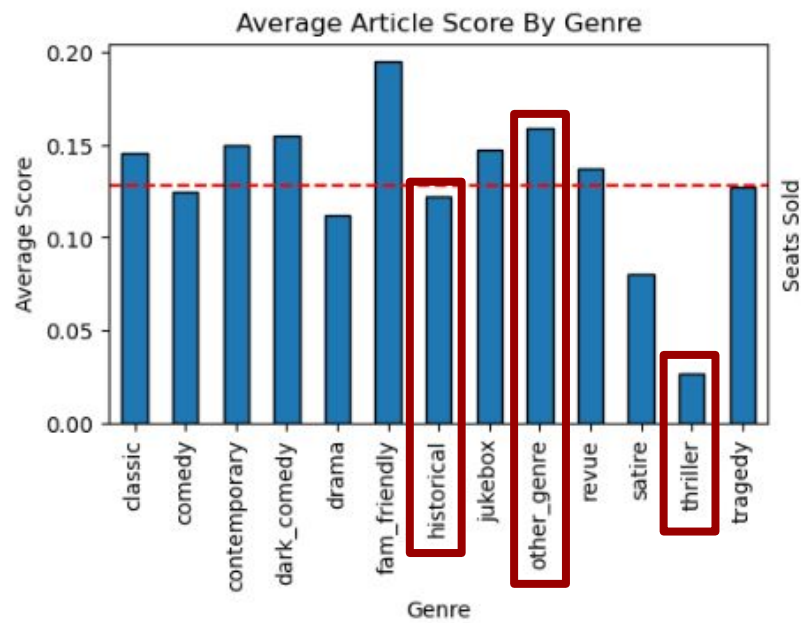


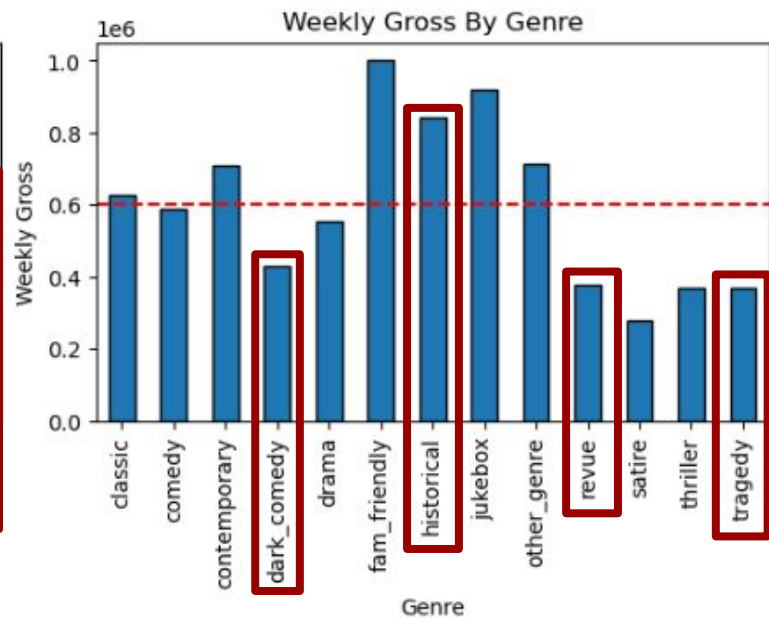
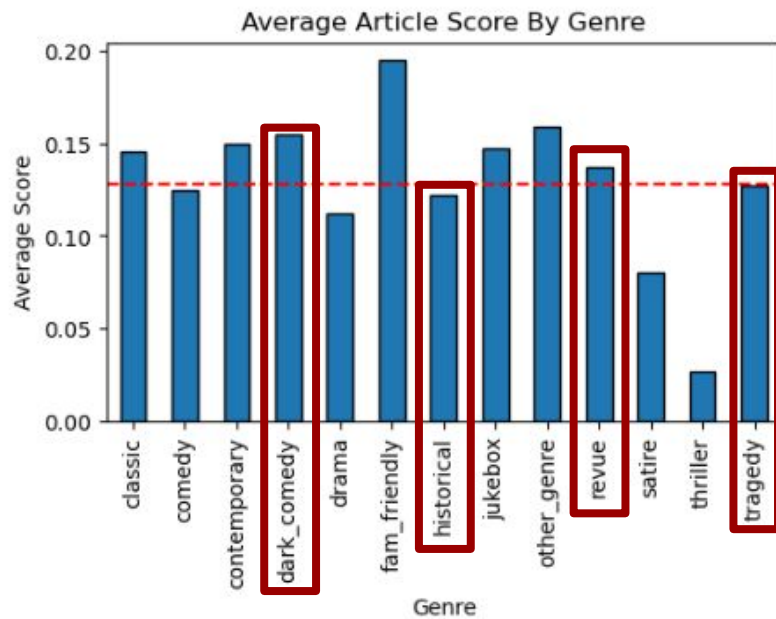
Average Article Score By Genre



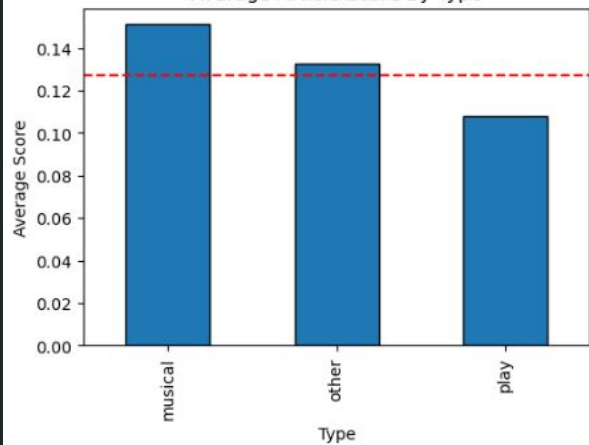
Top Ticket Price By Genre



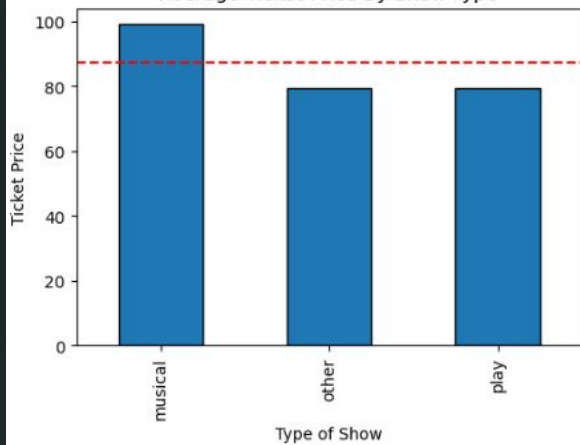




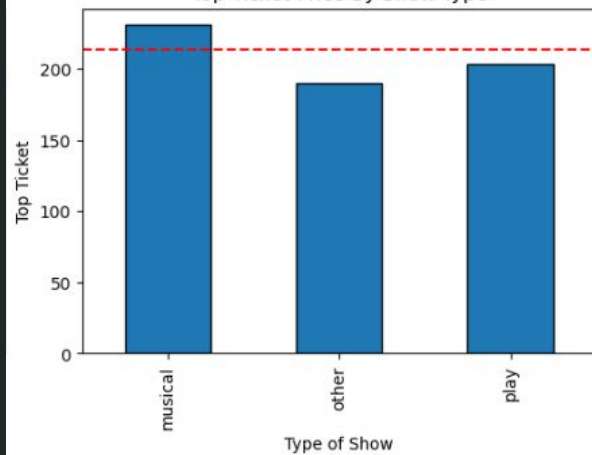
Average Article Score By Type



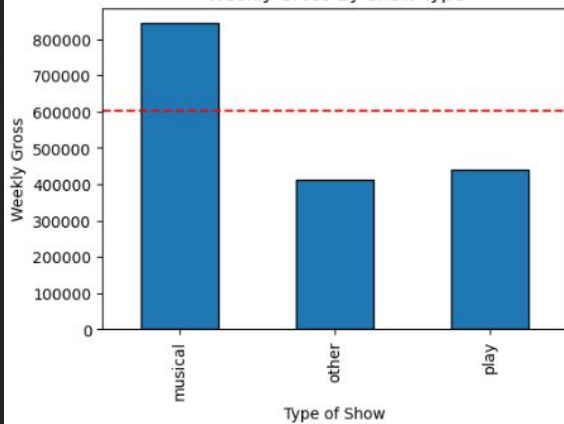
Average Ticket Price By Show Type



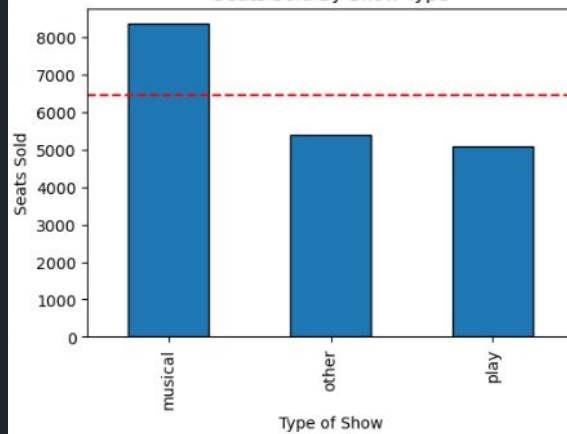
Top Ticket Price By Show Type



Weekly Gross By Show Type



Seats Sold By Show Type



What's next?

- Crawl more websites
- Reviews from critics AND audience
- Use different sentiment analysis model

