P1: Understanding the background and data

To understand how fandango operates, read the below mentioned article.

[Be cautious of online movie rating, especially Fandango's]

There are two csv files:

- One with fanango star ratings(fandango_scrape.csv)
- One with other site ratings such as IMDb, Rotten Tomatoes, Metacritics etc.(all_sites_scores.csv)

All_sites_scores.csv

Column | Definition

--- | -----------

FILM | The film in question

RottenTomatoes | The Rotten Tomatoes Tomatometer score for the film

RottenTomatoes_User | The Rotten Tomatoes user score for the film

Metacritic | The Metacritic critic score for the film

Metacritic_User | The Metacritic user score for the film

IMDB | The IMDb user score for the film

Metacritic_user_vote_count | The number of user votes the film had on Metacritic

IMDB_user_vote_count | The number of user votes the film had on IMDb

Fandango_scrape.csv

Column | Definiton

FILM | The movie

STARS | Number of stars presented on Fandango.com

RATING | The Fandango ratingValue for the film, as pulled from the HTML of each page. This is the actual average score the movie obtained.

VOTES | number of people who had reviewed the film at the time we pulled it.

P2: Exploring Fandango displayed scores versus true user ratings

Complete Fandango data analysis is done below.

1. Detailed description of the Fandango data

`fandango_scrape.csv` contains every film 538 pulled from Fandango.

Column | Definiton

FILM | The movie

STARS | Number of stars presented on Fandango.com

RATING | The Fandango ratingValue for the film, as pulled from the HTML of each page. This is the actual average score the movie obtained.

VOTES | number of people who had reviewed the film at the time we pulled it.

Data Info :

```
RangeIndex: 504 entries, 0 to 503
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   FILM    504 non-null    object
 1   STARS   504 non-null    float64
 2   RATING  504 non-null    float64
 3   VOTES   504 non-null    int64
dtypes: float64(2), int64(1), object(1)
memory usage: 15.9+ KB
```

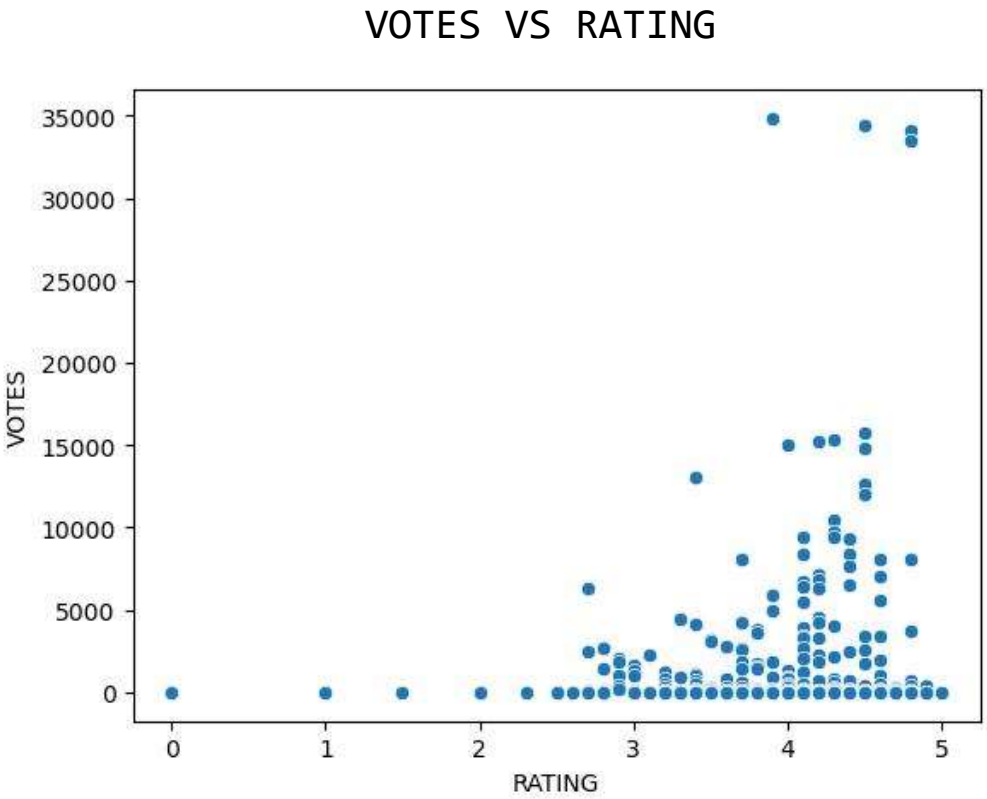
Statitical Description of fandango numerical data

	STARS	RATING	VOTES
count	504.000000	504.000000	504.000000
mean	3.558532	3.375794	1147.863095
std	1.563133	1.491223	3830.583136

min	0.000000	0.000000	0.000000
25%	3.500000	3.100000	3.000000
50%	4.000000	3.800000	18.500000
75%	4.500000	4.300000	189.750000
max	5.000000	5.000000	34846.000000

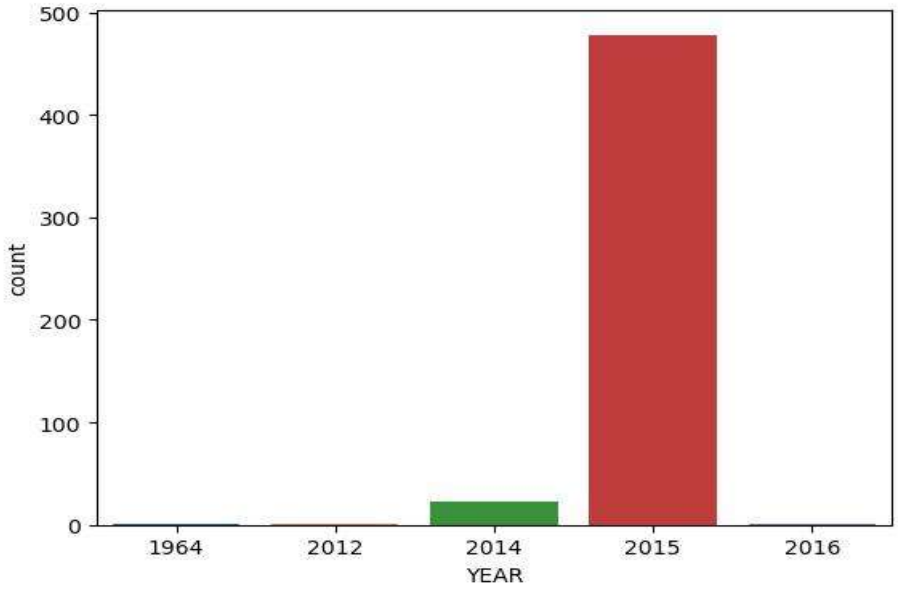
2. Creating a relationship between the popularity of the movie and the rating.

- A scatter plot is used



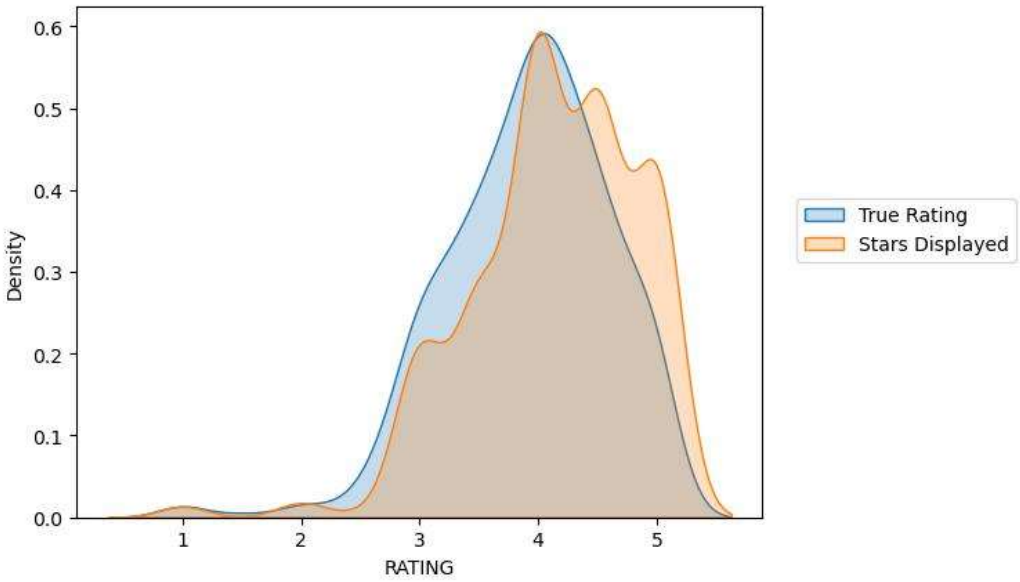
3. Count of Movies Versus Year listed on the website

Number of Movies VS Year



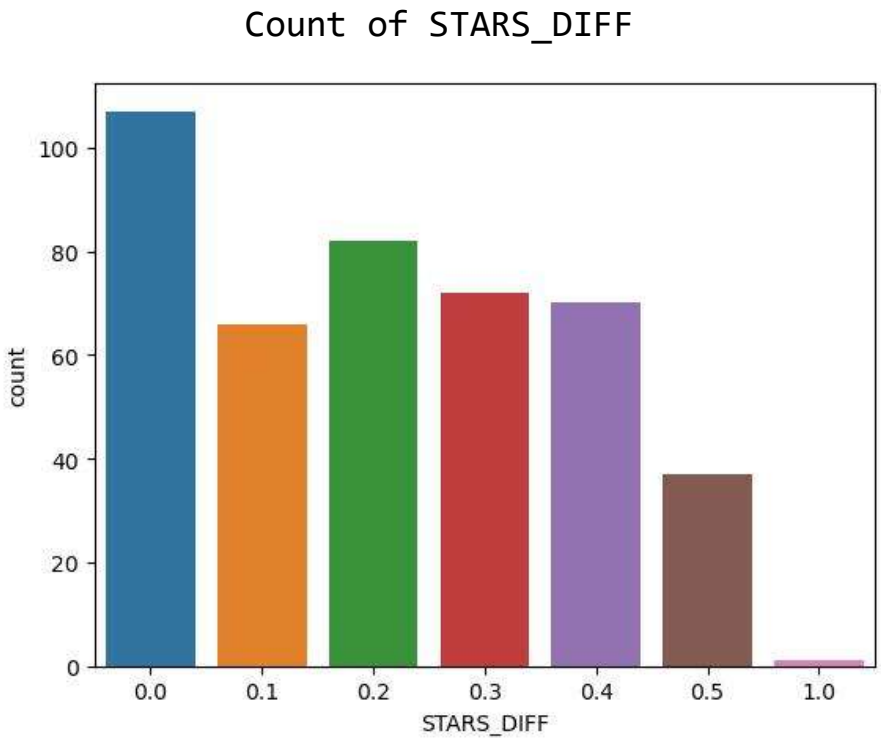
4. Plotting the discrepancy between the true rating and the Star rating

KDE Plots of True Rating and Stars Displayed



Using pa, the number of movies with zero votes were found out to be 69. Hence these movies will not have stars to display. Since these movies do not have star rating, they have been exempted from the below graph.

5. Count of the discrepancy



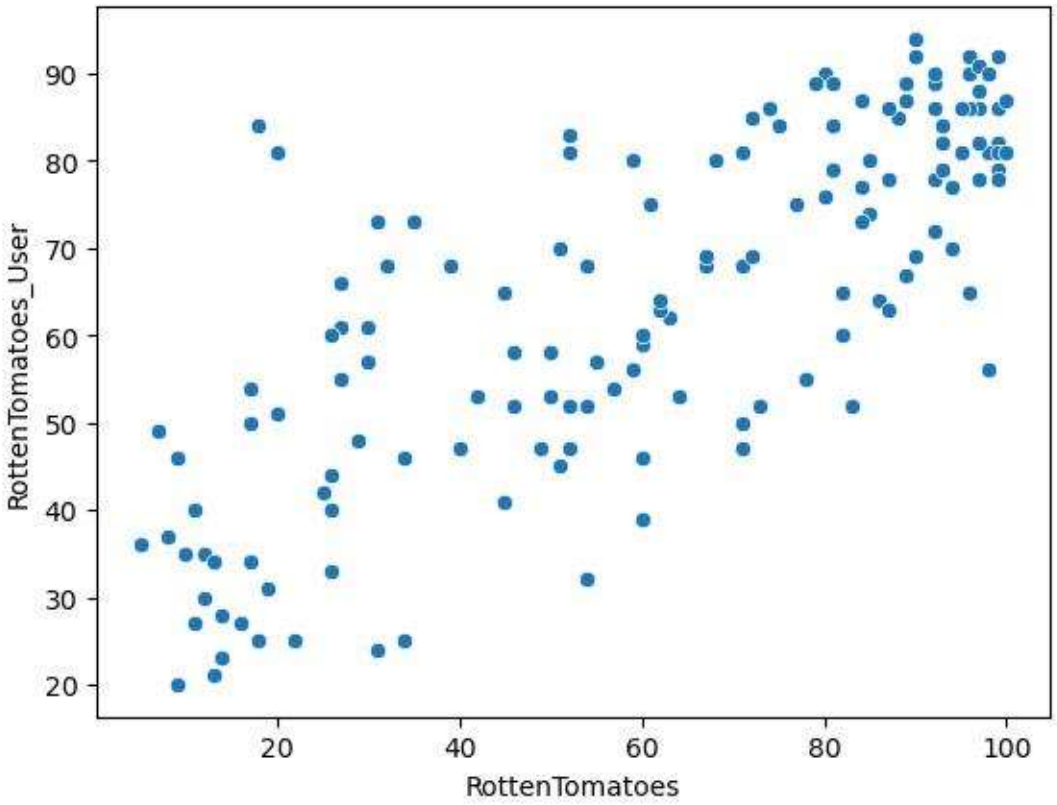
P3: Comparison of Fandango's ratings to other sites

Detailed data description of all_sites_scores.csv

```
Data columns (total 8 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   FILM                        146 non-null    object
 1   RottenTomatoes              146 non-null    int64
 2   RottenTomatoes_User         146 non-null    int64
 3   Metacritic                  146 non-null    int64
 4   Metacritic_User             146 non-null    float64
 5   IMDB                        146 non-null    float64
 6   Metacritic_user_vote_count  146 non-null    int64
 7   IMDB_user_vote_count        146 non-null    int64
```

Reviewing Rotten Tomatoes(RT) site:

Scatterplot exploring the relationship between RT Critic ratings and RT user ratings.



Rotten Tomatoes rating difference:

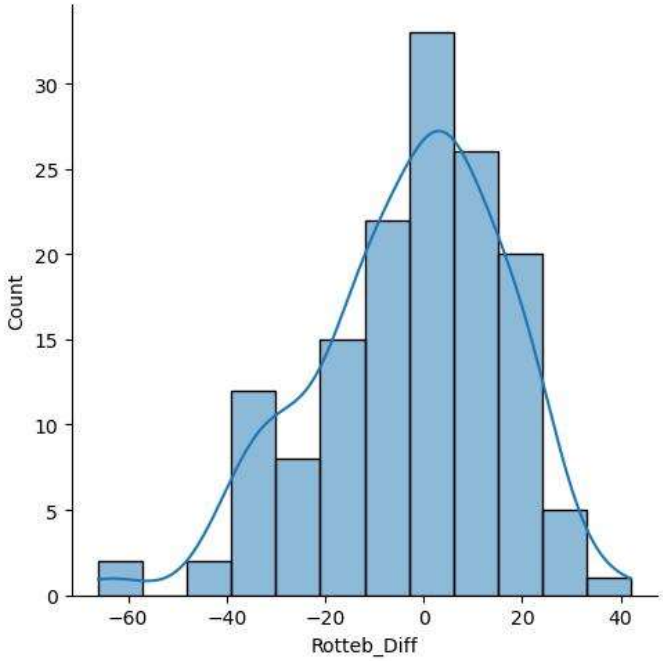
Rotten_Diff = Rotten_critic_score – Rotten_user_score

Rotten_Diff here is Critics - User Score. So values closer to 0 means aggrement between Critics and Users. Larger positive values means critics rated much higher than users. Larger negative values means users rated much higher than critics.

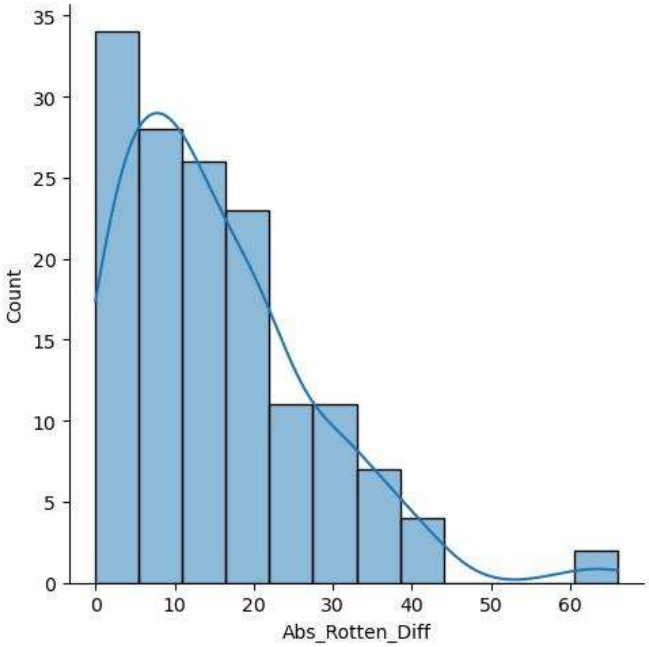
Since we're dealing with differences that could be negative or positive, first take the absolute value of all the differences, then take the mean. This would report back on average to absolute difference between the critics rating versus the user rating.

The mean absolute rotten_diff score is 15.096

Distribution plot of Rotten_Diff_score

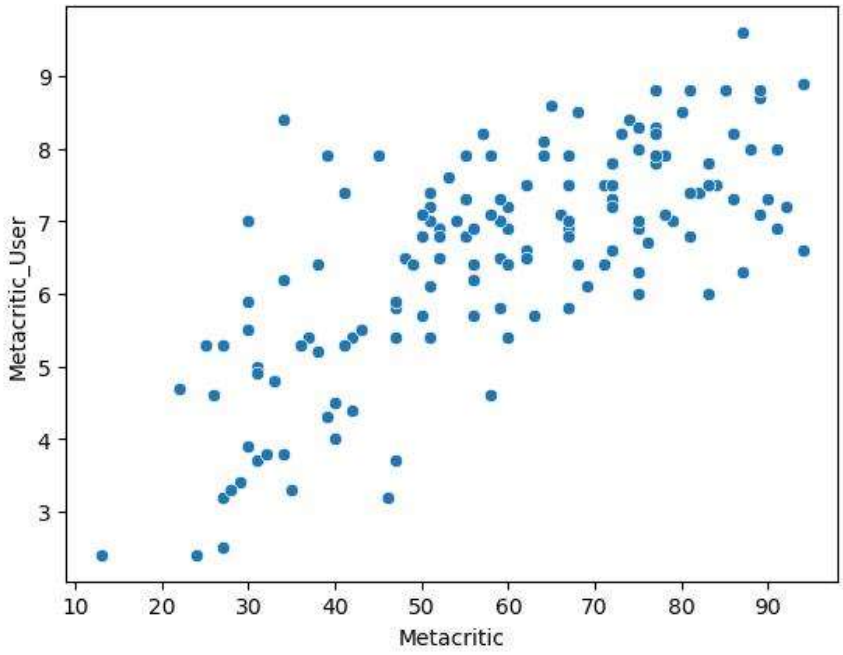


Distribution plot of mean Rotten_Diff_score



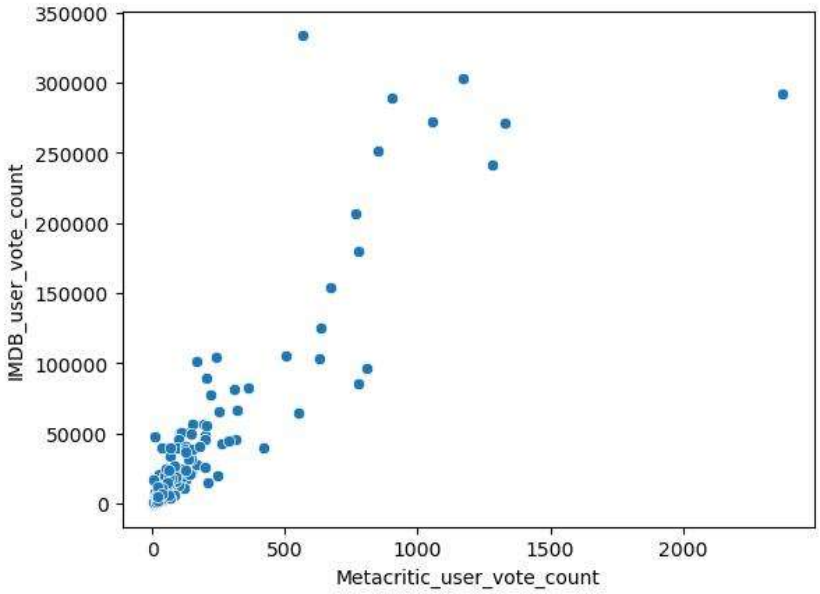
Reviewing MetaCritic site:

Scatterplot exploring the relationship between MetaCritic ratings and MetaCritic user ratings.



Reviewing IMDb site:

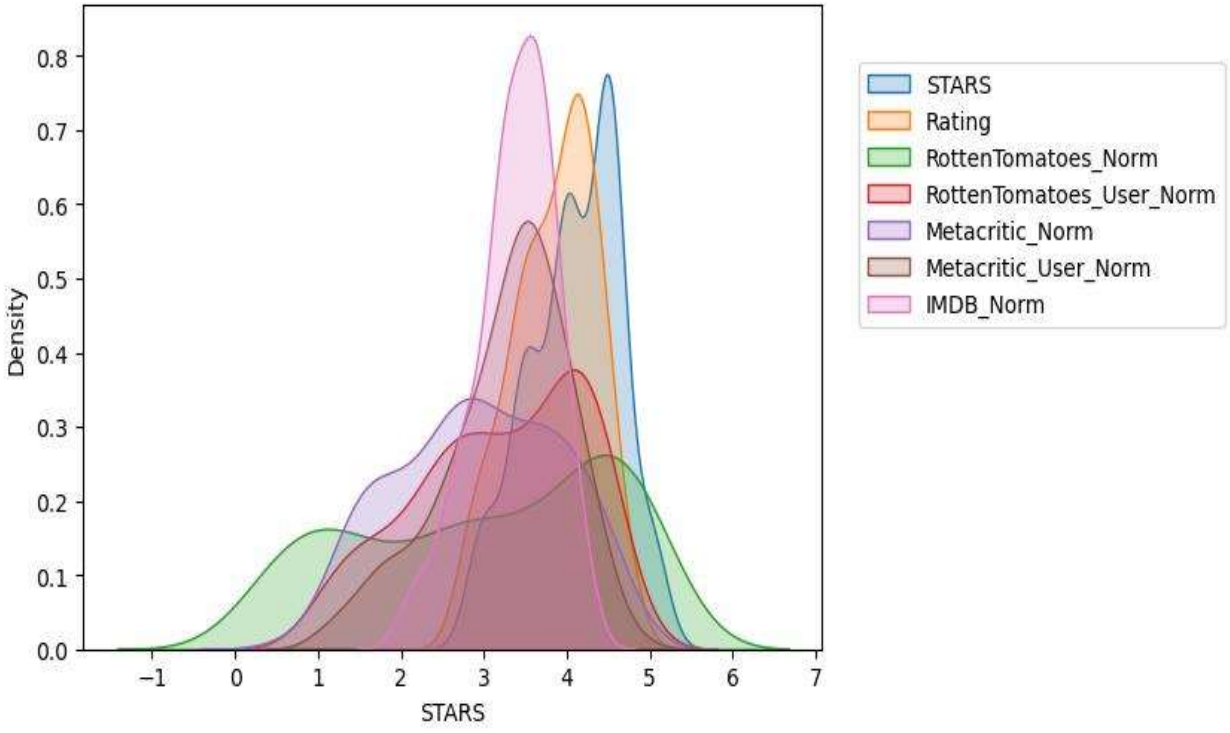
Scatterplot exploring the relationship between MetaCritic user vote count and IMDb user vote count



Fandango Scores vs all sites scores:

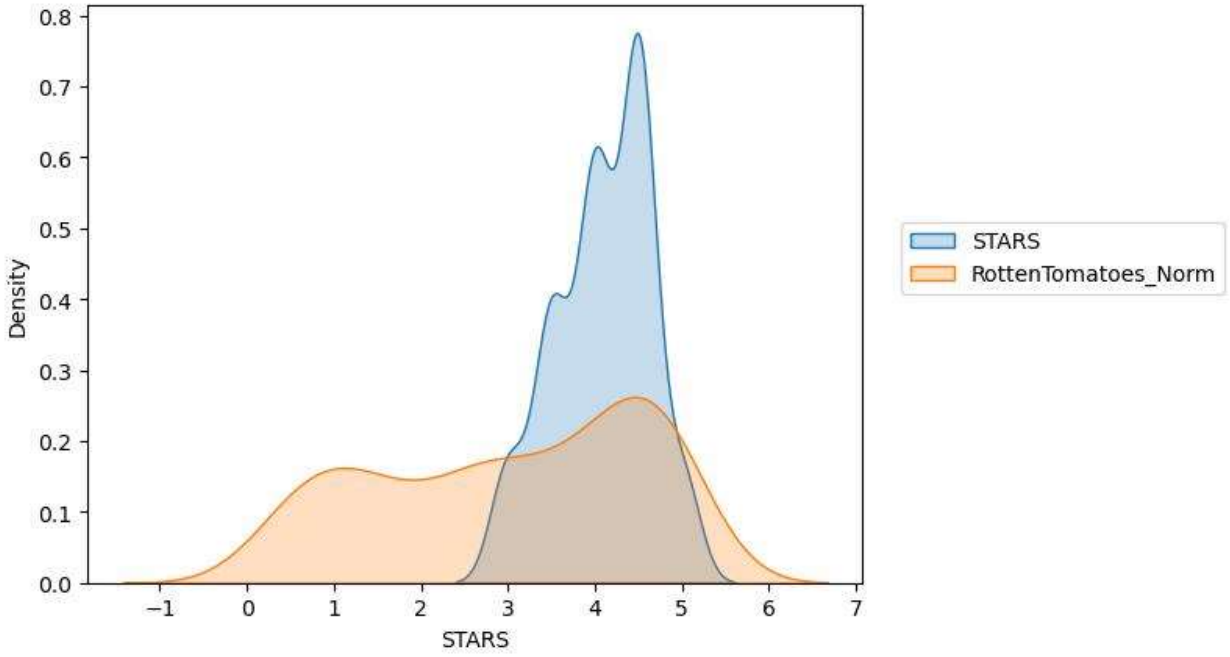
Since RT,Metacritic, and IMDB don't use a score between 0-5 stars like Fandango does. In order to do a fair comparison, we need to normalize these values so they all fall between 0-5 stars and the relationship between reviews stays the same.

Plot comparing the distributions of normalized ratings across all sites.

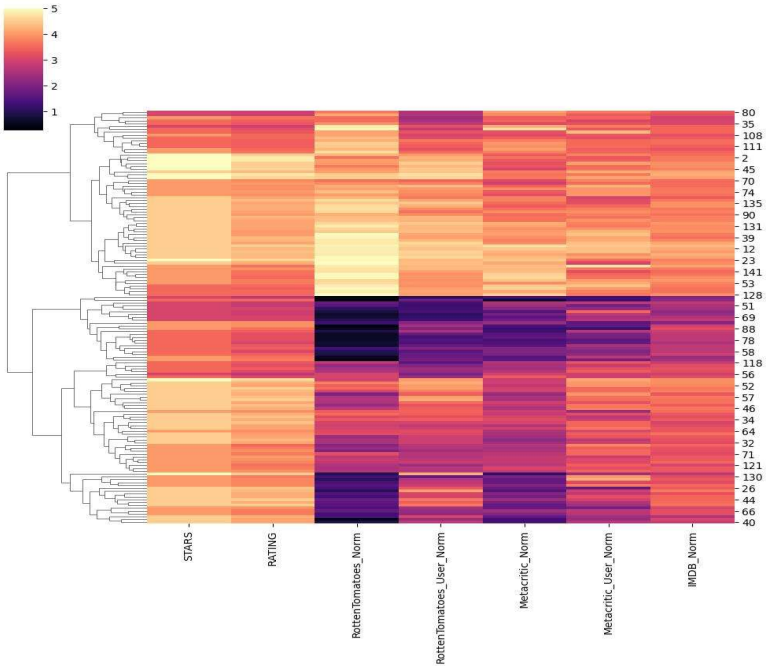


Clearly Fandango has an uneven distribution. We can also see that RT critics have the most uniform distribution. Let's directly compare these two.

KDE plot that compare the distribution of RT critic ratings against the STARS displayed by Fandango.



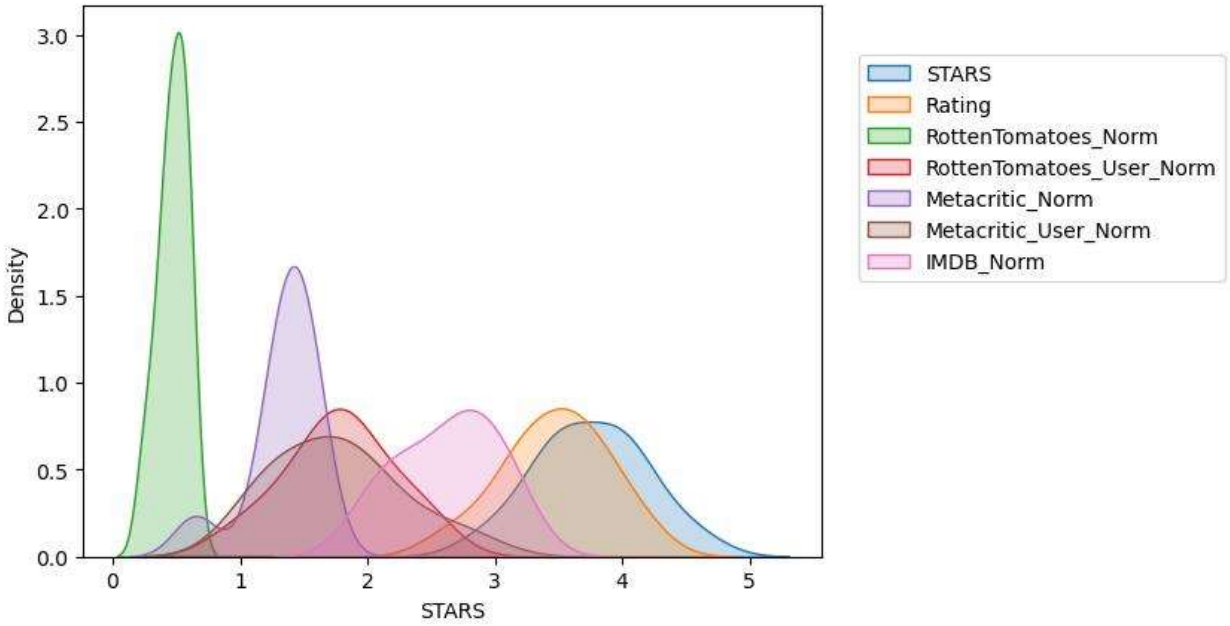
Clustermap visualization of all normalized scores.



The differences in ratings, highly rated movies should be clustered together versus poorly rated movies. Clearly Fandango is rating movies much higher than other sites, especially considering that it is then displaying a rounded up version of the rating.

Since Rotten Tomatoes critic has the most even distribution, we shall try to visualize the top to worst movie ratings as per Rotten Tomatoes critic across all sites.

KDE plot that compare the distribution of top 10 worst movie ratings as per Rotten Tomato across all sites.



Final Thoughts:

Clearly Fandango is rating movies much higher than the critic rating. Their rating round-off algorithm seems to be misplaced, trying to mislead it's customers by displaying higher star rating than the actual critic rating in order to increase the sale of movie tickets on their website.