**Purpose/Goal**

Should you decide to buy the distribution rights of the most popular movie? And you want to make use of the popular rating site like rotten tomatoes, Metacritic, IMDB and Fandango to assist in making that decision? You landed in the right place. Here is my *analysis to recommend you the best/better movie rating site* to facilitate your decision the best.

**About data**

This analysis is based on the sample of rating provided for 146 movies in the year 2014 and 2015. The ratings are provided by popular movie rating sites like Fandango, Metacritic, Rotten Tomatoes, and IMDB. This includes the ratings by user and critics. IMDB has the ratings only by the users. The data also includes the number of votes by the users for the particular movie.

*Initial analysis of the data*

* + Different scaling used by the rating sites to rate the movies. E.g. Rotten Tomatoes, Metacritic used the scale of 1 to 100; Metacritic Users and IMDB used the scale of 1 to 10; Fandango used the scale of 0 to 5. Hence the data is normalized to the scale of 0 to 5 to provide an unbiased result.
  + There are no missing or duplicate records in the data.

**Methods and Tools**

Below are the methods and tools used for the analysis;

* + Descriptive Statistics
  + Normal Distribution
  + Correlation
  + Python, JMP, and Tableau

**Analysis**

The analysis is mainly based on the ratings by the critic's as they are the one who views the movie in different aspects like the proper stage setting, directing plot, character development, cinematography, editing, and special effects etc. I believe that the user ratings might be influenced by different factors like demographics, genres, actors, their personal opinion etc. For instance, the person who likes horror and action movies may provide the high rating, the rating of the romantic movie is influenced by genders. They also may rate the movie by only watching the trailer of the movie and posters. Hence ratings by the users may be biased.

Though I have supported the rating by critics, let get started with descriptive statistics of the ratings by both critics and users.

*Descriptive Statistics*

Below are the fundamental descriptive statistics performed on the data. This includes mean, standard deviation, 1st quartile, 2nd quartile/ median, 3rd quartile, minimum, and maximum value.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Fandango**  **\_Stars** | **Fandango**  **\_Rating value** | **Rotten Tomatoes** | **Rotten Tomatoes**  **\_Users** | **Metacritic** | **Metacritic**  **\_User** | **IMDB** |
| **Mean** | 4.09 | 3.85 | 3.04 | 3.19 | 2.94 | 3.26 | 3.37 |
| **Standard Deviation** | 0.54 | 0.50 | 1.51 | 1.00 | 0.98 | 0.76 | 0.48 |
| **Minimum** | 3.00 | 2.70 | 0.25 | 1.00 | 0.65 | 1.20 | 2.00 |
| **First Quartile-25%** | 3.50 | 3.50 | 1.56 | 2.50 | 2.18 | 2.85 | 3.15 |
| **Second Quartile - 50%** | 4.00 | 3.90 | 3.18 | 3.33 | 2.95 | 3.43 | 3.45 |
| **Third Quartile - 75%** | 4.50 | 4.20 | 4.45 | 4.05 | 3.75 | 3.75 | 3.70 |
| **Maximum** | 5.00 | 4.80 | 5.00 | 4.70 | 4.70 | 4.80 | 4.30 |

Let’s get started with findings from this statistic:

**Fandango Stars**

|  |  |  |  |
| --- | --- | --- | --- |
| **Count** | **Mean** | **Median** | **Standard Deviation** |
| 146 | 4.09 | 4.00 | 0.54 |

Breaking down the above fundamental statistics,

***Count:*** Total **146** movies are rated.

***Mean:*** The average rating for 146 movies is **4.09**. The movies which have a rating of 4.09 will be considered to be an average movie. It's not good or bad.

***Median:*** The median is the center point in the dataset. The median 4.00 tells that the movies which have the rating above 4.00 fall into the category of good/better movie and movies with below 4.00 fall into the category of bad/worse movie.

***Standard Deviation:*** This tells how spread the ratings are from the mean. The Std.Dev of 0.54 shows that about 68% of movie ratings fall within one standard deviation (i.e.) in the range of **3.46 to 4.54.**

Other rating site follows similar break down with the difference in the value. Hence going forward, I will highlight the major findings for other rating sites.

**Rotten Tomatoes**

It has the larger standard deviation (1.51) compared to other rating sites. (i.e.) the data is widely spread, and we can assume that the distribution of the ratings doesn't have a bell-shaped curve which I will analyze later.

**Metacritic**

I am assuming that the ratings follow the normal curve as the mean (2.94) and the median (2.95) are almost same.

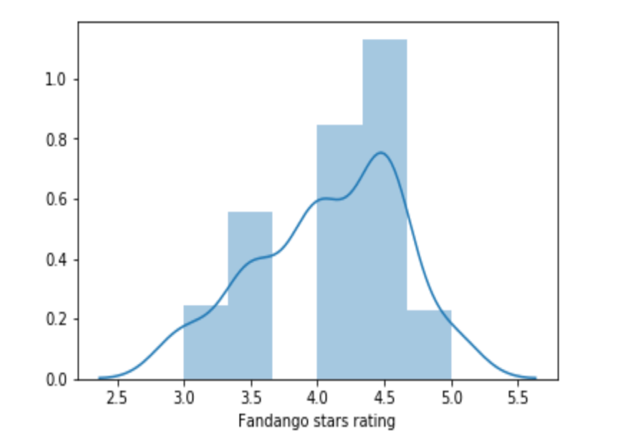
**IMDB**

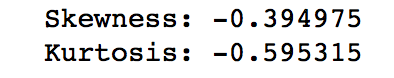
By comparing the quartile values (3.15, 3.45, and 3.70) I could say most of the ratings fall almost near to average and above average. (i.e.) the ratings are left-skewed. Mostly all the movies are rated average or above average.

**Normal Distribution**

Why normal distribution? In real-world movies released in a particular year can be categorized as an average movie, below average and above average. Based on this assumption we can say films follows the normal distribution. Hence if any of the movie rating sites follow the same or near normal bell-shaped curve we can conclude that site is trustworthy for making any decision. For example, in the scale of 0-5, let’s assume 3 is the average rating, then movies fall below 3 are bad and movies falls above 3 are good.

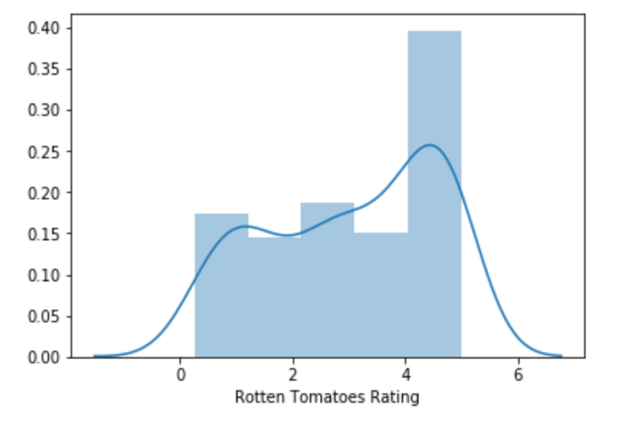
Let’s examine whether these movie rating site follows the bell-shaped curve (i.e.) to check the rating are normally distributed.

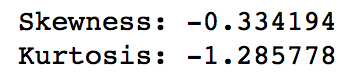
*Fandango Stars*

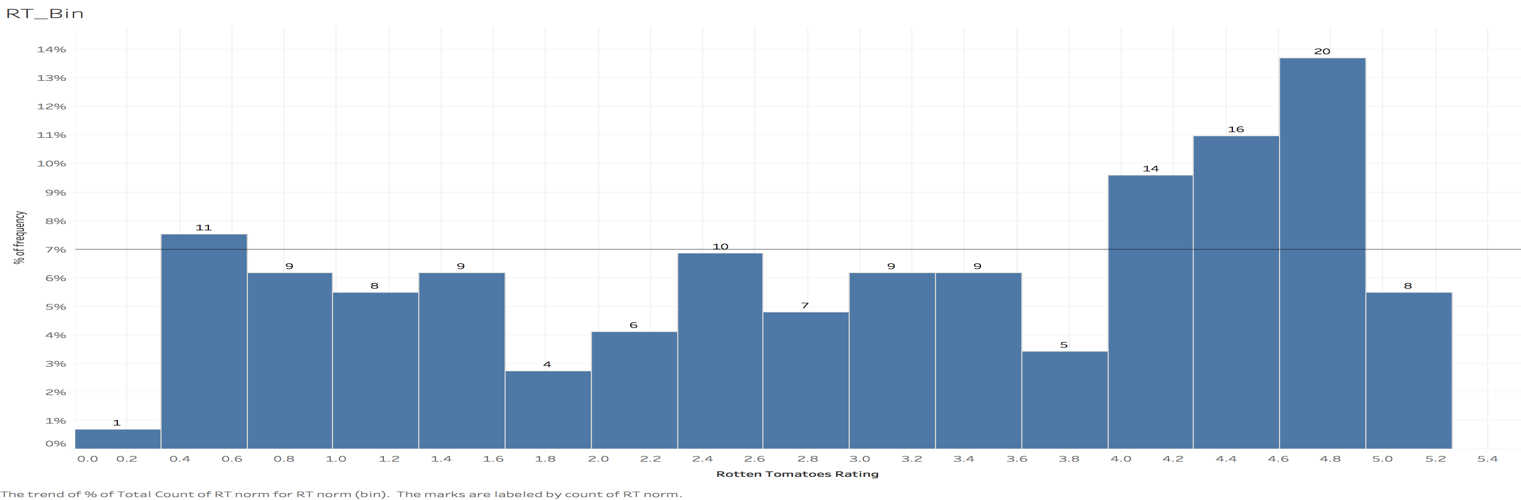
The figure shows the frequency distribution for fandango stars rating. It explains that the rating doesn’t follow the normal distribution. The skewness value (-0.40) substantiate that the data is left-skewed. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. Additionally, if you look at the descriptive statistics the minimum value only starts from 3 which is near to mean (4.09). Hence the rating seems to be biased (i.e.) It is not normally distributed.

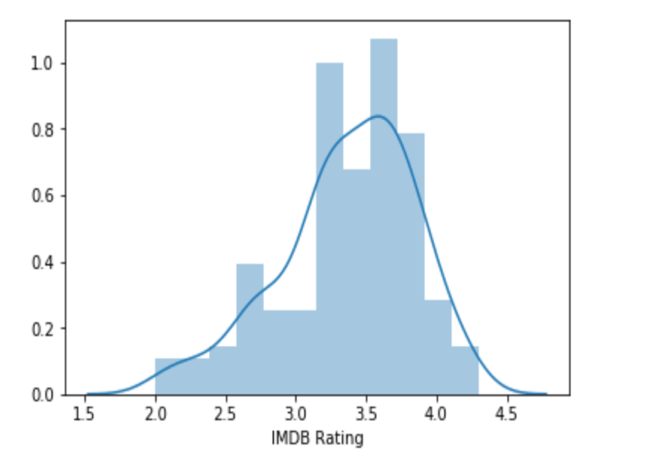
Also, by exploring the data it is evident that all the ratings are rounded to the nearest value. Example, their rating value has only 3, 3.5, 4, 4.5 and 5. They may round the 2.6 rating to 3.0 which is not certain. Their rating approach is not clear. It seems to be they rating all the movies as average or above average movies which are not convincing in the real-time scenario. Hence, **we cannot rely on Fandango Stars.**

*Rotten Tomatoes*

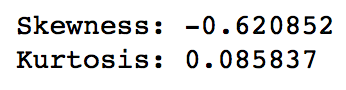
The figure shows the frequency distribution for Rotten Tomatoes rating. It seems that the rating is equally distributed. It is not normally distributed. As the center is flattened, we could say most of the movies are rated either below average or above average. As stated in descriptive statistics, the standard deviation is higher. Hence the data is widely spread which substantiate the rating doesn't follow the normal bell-shaped curve. It is left-skewed as per the skewness value (-0.33). Also attached the histogram from the tableau to show the pattern.



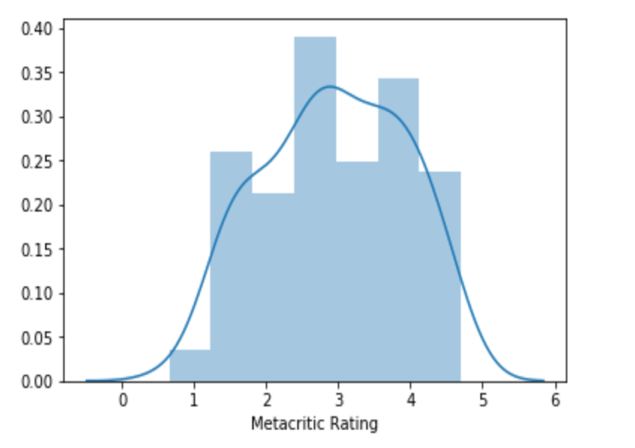
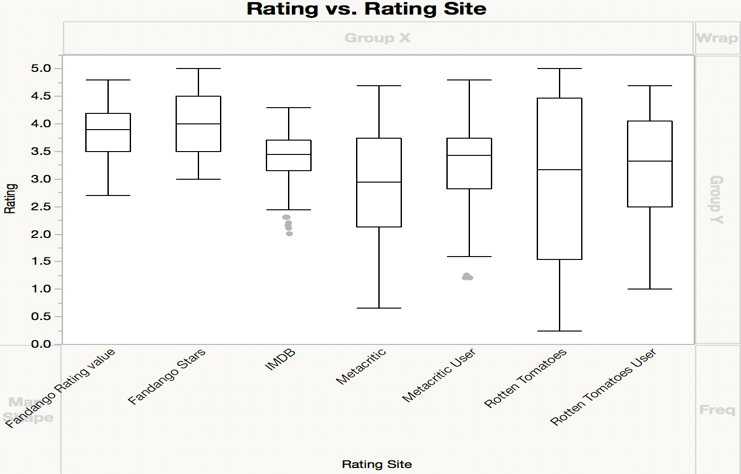


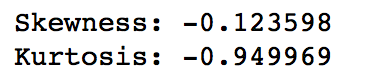
*IMDB*

The figure shows the frequency distribution for IMDB rating. Looking at the graph, it is clear that the ratings are left-skewed. The Factors such as Mean (3.37) is less than the median (3.45), negative skewness value (-0.62) explains that the data is not symmetric. As per the quartile value stated in the descriptive statistics, 50 % of the data falls above average. (i.e.) Most the movies are rated as above average which doesn't make any sense.



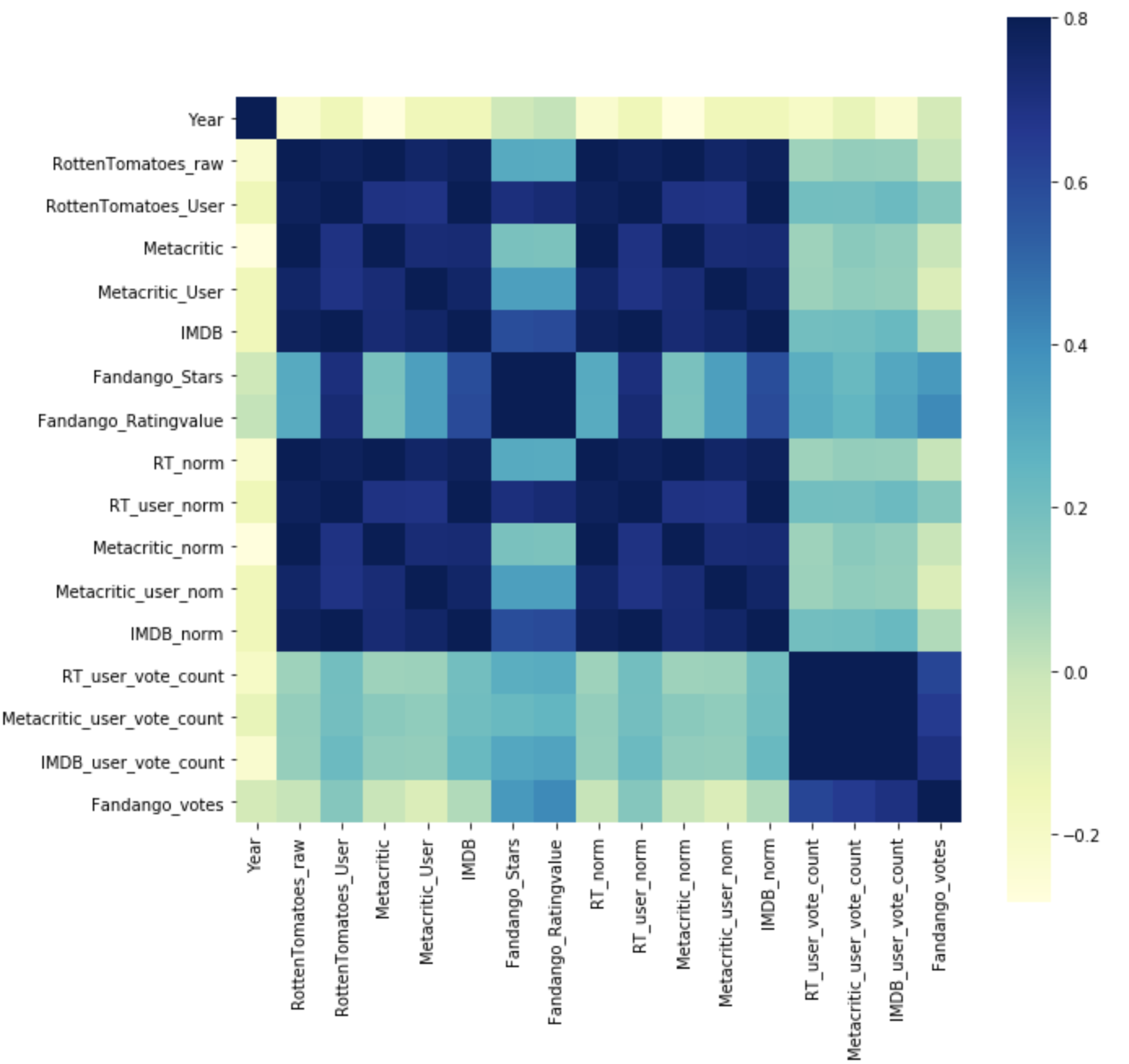
*Metacritic*

The figure shows the frequency distribution for Metacritic rating. Compared to other graph's it seems to be normally distributed. As the mean (2.94) and median (2.95) are almost same, we could say the data is symmetric. The skewness value (-0.1) also supports this claim as it is almost zero. The data is properly distributed from 0.65 to 4.70. Below is the box plot to show the distribution of Metacritic rating. As per the below plot, Metacritic has the proper distribution of data compared to others without any outliers.



**Correlation**

From the above analysis, it is clear that Fandango stars have the least preference compared to others. Hence, as a next step, we need compare which movie rating site has the least correlation compared to fandango stars.

Below is the correlation matrix/Table determined using python.

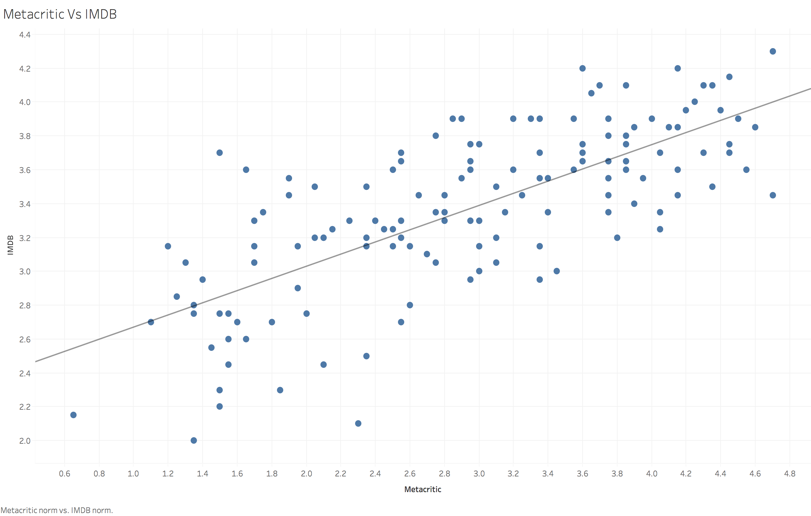
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Fandango Stars** | **Fandango Rating value** | **Rotten Tomatoes** | **Metacritic** | **Metacritic User** | **IMDB** |
| **Fandango Stars** | 1 | 0.959709 | 0.293988 | 0.181124 | 0.337497 | 0.587295 |
| **Fandango Rating value** | 0.959709 | 1 | 0.290047 | 0.177204 | 0.336945 | 0.600354 |
| **Rotten Tomatoes** | 0.293988 | 0.290047 | 1 | 0.95736 | 0.754296 | 0.779671 |
| **Metacritic** | 0.181124 | 0.177204 | 0.95736 | 1 | 0.723459 | 0.727298 |
| **Metacritic User** | 0.337497 | 0.336945 | 0.754296 | 0.723459 | 1 | 0.756265 |
| **IMDB** | 0.587295 | 0.600354 | 0.779671 | 0.727298 | 0.756265 | 1 |

As per the above table, Metacritic has the least correlation value compared with Fandango Stars.

From the above analysis, we can nail down to Metacritic. Hence, to further substantiate the conclusion, let's compare the critics rating with user rating to see the correlation between them.

As per the below correlation table, the value is **0.72** which shows they highly correlated. This proves that the critics rating balances the user rating also.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Fandango Stars** | **Fandango Rating value** | **Rotten Tomatoes** | **Metacritic** | **Metacritic User** | **IMDB** |
| **Fandango Stars** | 1 | 0.959709 | 0.293988 | 0.181124 | 0.337497 | 0.587295 |
| **Fandango Rating value** | 0.959709 | 1 | 0.290047 | 0.177204 | 0.336945 | 0.600354 |
| **Rotten Tomatoes** | 0.293988 | 0.290047 | 1 | 0.95736 | 0.754296 | 0.779671 |
| **Metacritic** | 0.181124 | 0.177204 | 0.95736 | 1 | 0.723459 | 0.727298 |
| **Metacritic User** | 0.337497 | 0.336945 | 0.754296 | 0.723459 | 1 | 0.756265 |
| **IMDB** | 0.587295 | 0.600354 | 0.779671 | 0.727298 | 0.756265 | 1 |

As a final step, let's also compare the Metacritic with IMDB as it rating is purely based on user and also it has a huge volume of user votes.

The correlation value between them is **0.73** which also shows they are having a similar pattern.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Fandango Stars** | **Fandango Rating value** | **Rotten Tomatoes** | **Metacritic** | **Metacritic User** | **IMDB** |
| **Fandango Stars** | 1 | 0.959709 | 0.293988 | 0.181124 | 0.337497 | 0.587295 |
| **Fandango Rating Value** | 0.959709 | 1 | 0.290047 | 0.177204 | 0.336945 | 0.600354 |
| **Rotten Tomatoes** | 0.293988 | 0.290047 | 1 | 0.95736 | 0.754296 | 0.779671 |
| **Metacritic** | 0.181124 | 0.177204 | 0.95736 | 1 | 0.723459 | 0.727298 |
| **Metacritic User** | 0.337497 | 0.336945 | 0.754296 | 0.723459 | 1 | 0.756265 |
| **IMDB** | 0.587295 | 0.600354 | 0.779671 | 0.727298 | 0.756265 | 1 |

**Conclusion**

From the above analysis, my recommendation would be ***Metacritic***, the reason being

1. The critic's ratings are spread across below average to above average with more rating falling under the average category, which reasonable for real case scenario (Proved via normal distribution)
2. Fandango being the least preferred among others, the correlation value of Metacritic with Fandango is very less, which clearly substantiate they are not following a similar pattern.
3. Metacritic critic’s rating and user ratings are highly correlated, which shows that critic rating matched with user rating as well.
4. Since the Metacritic user voting count is less, So I have also checked the correlation with IMDB. The correlation value is high as expected.

It is evident from all the above, that the ***Metacritic is the better movie rating*** site compared to others.