

rotten_tomato_movie_review.csv

Variable Definitions

1. movie_id: Unique identifier for each movie (matches the id in rotten_tomatoes_movies.csv)
2. reviewId: Unique identifier for each critic review
3. creationDate: The date the review was published
4. criticName: The name of the critic who wrote the review
5. isTopCritic: A boolean value indicating if the critic is considered a top critic
6. originalScore: The score provided by the critic
7. reviewState: The status of the review (e.g., fresh, rotten)
8. publicationName: The name of the publication where the review was published
9. reviewText: The name of the publication where the review was published
10. scoreSentiment: The sentiment of the critic's score (e.g., positive, negative, neutral)
11. reviewUrl: The URL of the original review on Rotten Tomatoes

Summary Statistics

	movie_id	reviewId	creationDate	criticName	isTopCritic	originalScore	reviewState	publicationName	reviewText	scoreSentiment
count	1444963	1.444963e+06	1444963	1444963	1444963	1009745	1444963	1444963	1375738	1444963
unique	69263	NaN	8510	15510	2	1729	2	2707	1359771	2
top	parasite_2019	NaN	2000-01-01	Emanuel Levy	False	3/5	fresh	New York Times	Parental Content Review	POSITIVE
freq	954	NaN	48083	13274	1008156	116711	963799	19853	236	963799
mean	NaN	9.035203e+06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	2.575716e+07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	1.000000e+00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	1.610366e+06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	2.200337e+06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	2.587024e+06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

rotten_tomato_movies.csv

Variable Definitions

1. Movie_id: Unique identifier for each movie (matches the id in rotten_tomatoes_movies.csv)
2. title: title of the movie
3. audienceScore: The average score given by regular viewers
4. tomatoMeter: The percentage of positive reviews from professional critics
5. rating: The movie's age-based classification (e.g., G, PG, PG-13, R)
6. ratingContent: Content leading to the rating classification
7. releaseDateTheater: The date the movie was released in theaters
8. releaseDateStreaming: The date the movie was released to streaming
9. runtimeMinutes: The duration of the movie in minutes
10. genre: The movie's genre(s)
11. originalLanguage: The original language of the movie
12. director: the movie's director
13. writer: the movie's writer(s)

14. boxOffice: The movie's total box office revenue
15. distributor: The company responsible for distributing the movie
16. soundMix: The audio format(s) used in the movie

	movie_id	title	audienceScore	tomatoMeter	rating	ratingContents	releaseDateTheaters	releaseDateStreaming	runtimeMinutes	genre	originalLanguage	director	writer	boxOffice	distributor	soundMix
count	143258	142891	73248.000000	33877.000000	13991	13991	30773	79420	129431.000000	132175	129400	139041	90116	14743	23001	15917
unique	142052	126403	NaN	NaN	10	8353	12062	4726	NaN	2912	112	62206	67274	4863	3693	551
top	junction_48	The Stranger	NaN	NaN	R	["Language"]	2018-09-14	2017-05-22	NaN	Drama	English	Unknown Director	Jing Wong	\$1.1M	Paramount Pictures	Surround
freq	2	17	NaN	NaN	7734	365	37	1232	NaN	27860	85034	3544	48	118	994	4075
mean	NaN	NaN	55.674967	65.770346	NaN	NaN	NaN	NaN	93.708578	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	24.553648	28.023203	NaN	NaN	NaN	NaN	28.129175	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	0.000000	0.000000	NaN	NaN	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	37.000000	45.000000	NaN	NaN	NaN	NaN	84.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	57.000000	73.000000	NaN	NaN	NaN	NaN	92.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	76.000000	89.000000	NaN	NaN	NaN	NaN	103.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN