

فرض کنید قصد دارید برای تماشای فیلم بیرون بروید. تا چه حد می توانید به نقد و امتیازهای آنلاین اعتماد کنید؟ به ویژه اگر همان شرکتی که امتیاز را نمایش می دهد، همچنین از طریق فروش بلیط فیلم درآمد کسب کند. آیا آنها تمایلی به امتیاز دادن بالاتر از حد واقعی فیلم ها دارند؟

هدف

هدف شما تکمیل وظایف زیر بر اساس مقاله ۵۳۸ و بررسی حصول نتیجه ای مشابه است. برای تعیین اینکه آیا امتیازهای فاندوango در سال ۲۰۱۵ برای فروش بیشتر بلیط، تمایلی به امتیازدهی بهتر به فیلم ها داشته است یا خیر، باید از مهارت های پانداس و تجسم سازی خود استفاده کنید.

| |
|--|
| |
| |
| وظایف ذکر شده با فونت درشت را تکمیل کنید |
| |
| |

بخش اول: درک پیشینه و داده ها

تکلیف: مطالعه این مقاله: <http://fivethirtyeight.com/features/fandango-movies-ratings>: /Be Suspicious Of Online Movie Ratings, Especially Fandango’s

بخش اول: درک پیشینه و داده ها (ادامه)

تکلیف: پس از مطالعه مقاله، این دو جدول را که نمای کلی از دو فایل csv را که با آنها کار خواهیم کرد، ارائه می دهد، مطالعه کنید:

داده ها

این داده ها پشت پرده مقاله "به ارزیابی های آنلاین فیلم، به ویژه فاندانگو مشکوک باشید": Be Suspicious Of Online Movie Ratings, Especially Fandango’s: <https://github.com/fivethirtyeight/data> در دسترس است: <http://fivethirtyeight.com/features/fandango-movies-ratings> است که به طور رایگان در گیت هاب ۵۳۸ در دسترس است: <https://github.com/fivethirtyeight/data>. دو فایل csv وجود دارد، یکی با امتیاز ستاره ای فاندانگو و امتیاز نمایش داده شده، و دیگری با داده های تجمعی برای رتبه بندی فیلم از سایت های دیگر مانند متاکریتیک، IMDb و راتن تومیتوز.

| |
|----------------------|
| all_sites_scores.csv |
| |
| (|
| |
| all_sites_scores.csv |

فایل `all_sites_scores.csv` شامل تمام فیلم‌هایی است که دارای امتیاز راتن تومیتوز، امتیاز کاربر راتن تومیتوز، امتیاز متاکریتیک، امتیاز کاربر متاکریتیک و امتیاز IMDb هستند و همچنین حداقل 30 نقد و بررسی از طرفداران در فاندانگو دارند. داده های فاندانگو در تاریخ 24 آگوست 2015 جمع آوری شده است.

| 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_scape.csv

فایل fandango_scape.csv شامل تمام فیلم هایی است که ۵۳۸ از فاندانگو استخراج کرده است.

| 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. |

تکلیف: هر کتابخانه ای که فکر می کنید استفاده خواهید کرد را وارد کنید.**

In [173...# IMPORT HERE!

In [174...

بخش دوم: بررسی امتیازهای نمایش داده شده فاندانگو در مقابل امتیازهای واقعی کاربران

ابتدا بیاپید به بررسی امتیازهای فاندانگو بپردازیم تا ببینیم آیا تحلیل ما با نتیجه گیری مقاله مطابقت دارد.

تکلیف: سلول زیر را برای خواندن فایل fandango_scrape.csv اجرا کنید

In [175...fandango = pd.read_csv("fandango_scrape.csv")

تکلیف: ویژگی ها و ابتدای DataFrame را بررسی کنیAR تنظیم کند

In [176...

Out[176...

| | FILM | STARS | RATING | VOTES |
|---|-----------------------------|-------|--------|-------|
| 0 | Fifty Shades of Grey (2015) | 4.0 | 3.9 | 34846 |
| 1 | Jurassic World (2015) | 4.5 | 4.5 | 34390 |
| 2 | American Sniper (2015) | 5.0 | 4.8 | 34085 |
| 3 | Furious 7 (2015) | 5.0 | 4.8 | 33538 |
| 4 | Inside Out (2015) | 4.5 | 4.5 | 15749 |

In [177...

```
<class 'pandas.core.frame.DataFrame'>
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
```

In [178...

Out[178...

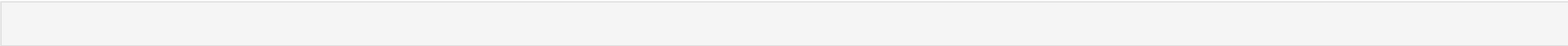
| | STARS | RATING | VOTES |
|-------|------------|------------|--------------|
| count | 504.000000 | 504.000000 | 504.000000 |
| mean | 3.558532 | 3.375794 | 1147.863095 |
| std | 1.563133 | 1.491223 | 3830.583136 |
| min | 0.000000 | 0.000000 | 0.000000 |
| 25% | 3.500000 | 3.100000 | 3.000000 |
| 50% | 4.000000 | 3.800000 | 18.500000 |
| 75% | 4.500000 | 4.300000 | 189.750000 |
| max | 5.000000 | 5.000000 | 34846.000000 |

تکلیف: بیایید رابطه بین محبوبیت یک فیلم و امتیاز آن را بررسی کنیم. یک نمودار پراکنده ایجاد کنید که رابطه بین امتیاز و تعداد آراء را نشان دهد. برای تنظیم سبک بصری بر اساس ترجیح خود راحت باشید.

In [179...

```
# CODE HERE
```

In [180...

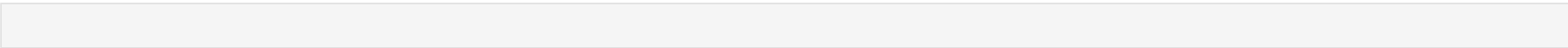


تکلیف: همبستگی بین ستون ها را محاسبه کنید.

In [181...

```
# CODE HERE
```

In [182...



Out[182...

| | STARS | RATING | VOTES |
|--------|----------|----------|----------|
| STARS | 1.000000 | 0.994696 | 0.164218 |
| RATING | 0.994696 | 1.000000 | 0.163764 |
| VOTES | 0.164218 | 0.163764 | 1.000000 |

تکلیف: با فرض اینکه هر ردیف در ستون عنوان FILM دارای فرمت یکسانی است:

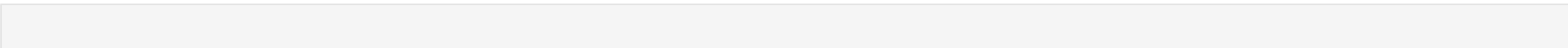
Film Title Name (Year)

یک ستون جدید ایجاد کنید که بتواند سال را از رشته های عنوان حذف کند و این ستون جدید را به عنوان YEAR تنظیم کند

In [183...

```
# CODE HERE
```

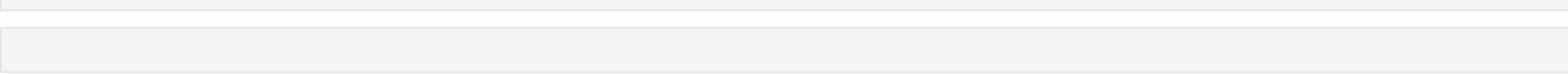
In [184...



In [185...

```
#CODE HERE
```

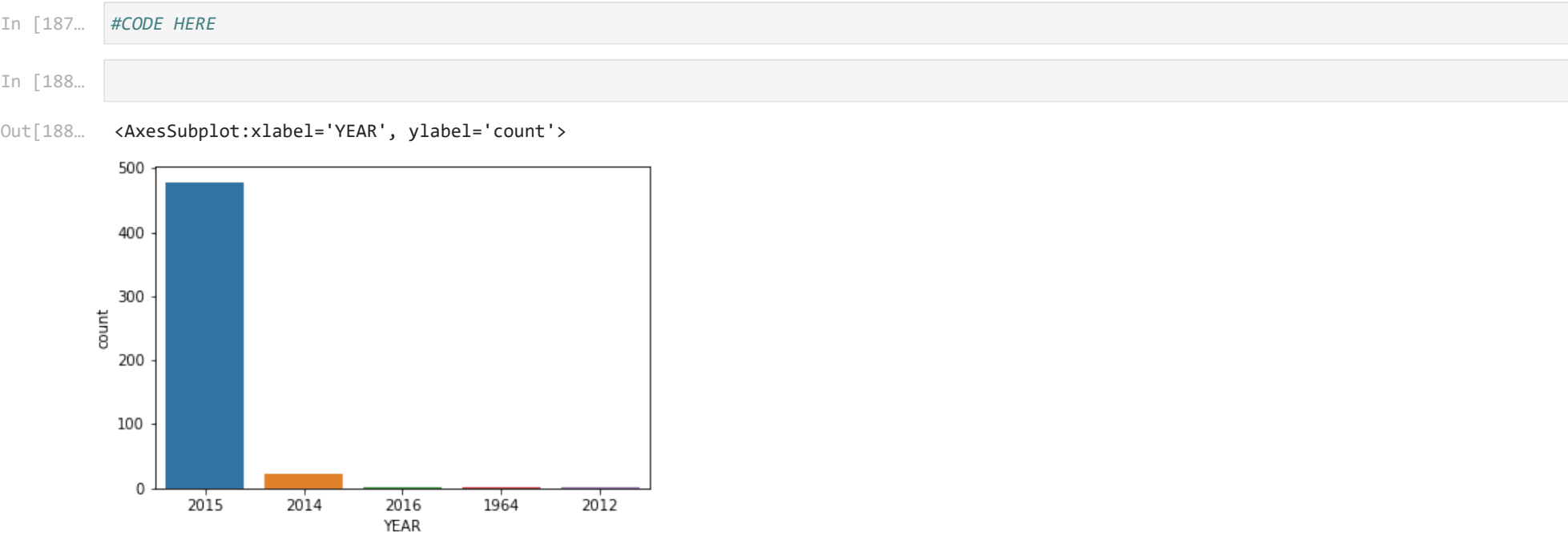
In [186...



Out[186...

| | |
|--------------------------|-----|
| 2015 | 478 |
| 2014 | 23 |
| 2016 | 1 |
| 1964 | 1 |
| 2012 | 1 |
| Name: YEAR, dtype: int64 | |

تکلیف: تعداد فیلم ها را در هر سال با یک نمودار نمایش دهید.



تکلیف: 10 فیلم با بیشترین تعداد آراء کدامند؟

In [189... #CODE HERE

In [190...

Out[190...

| | FILM | STARS | RATING | VOTES | YEAR |
|---|--|-------|--------|-------|------|
| 0 | Fifty Shades of Grey (2015) | 4.0 | 3.9 | 34846 | 2015 |
| 1 | Jurassic World (2015) | 4.5 | 4.5 | 34390 | 2015 |
| 2 | American Sniper (2015) | 5.0 | 4.8 | 34085 | 2015 |
| 3 | Furious 7 (2015) | 5.0 | 4.8 | 33538 | 2015 |
| 4 | Inside Out (2015) | 4.5 | 4.5 | 15749 | 2015 |
| 5 | The Hobbit: The Battle of the Five Armies (2014) | 4.5 | 4.3 | 15337 | 2014 |
| 6 | Kingsman: The Secret Service (2015) | 4.5 | 4.2 | 15205 | 2015 |
| 7 | Minions (2015) | 4.0 | 4.0 | 14998 | 2015 |
| 8 | Avengers: Age of Ultron (2015) | 5.0 | 4.5 | 14846 | 2015 |
| 9 | Into the Woods (2014) | 3.5 | 3.4 | 13055 | 2014 |

تکلیف: چند فیلم دارای رای صفر هستند؟

In [191... #CODE HERE

In [192...

Out[192... 69

تکلیف: با حذف هر فیلمی که رای صفر دارد، یک دیتافریم فقط از فیلم های نقد شده ایجاد کنید.

In [193... #CODE HERE

In [194...

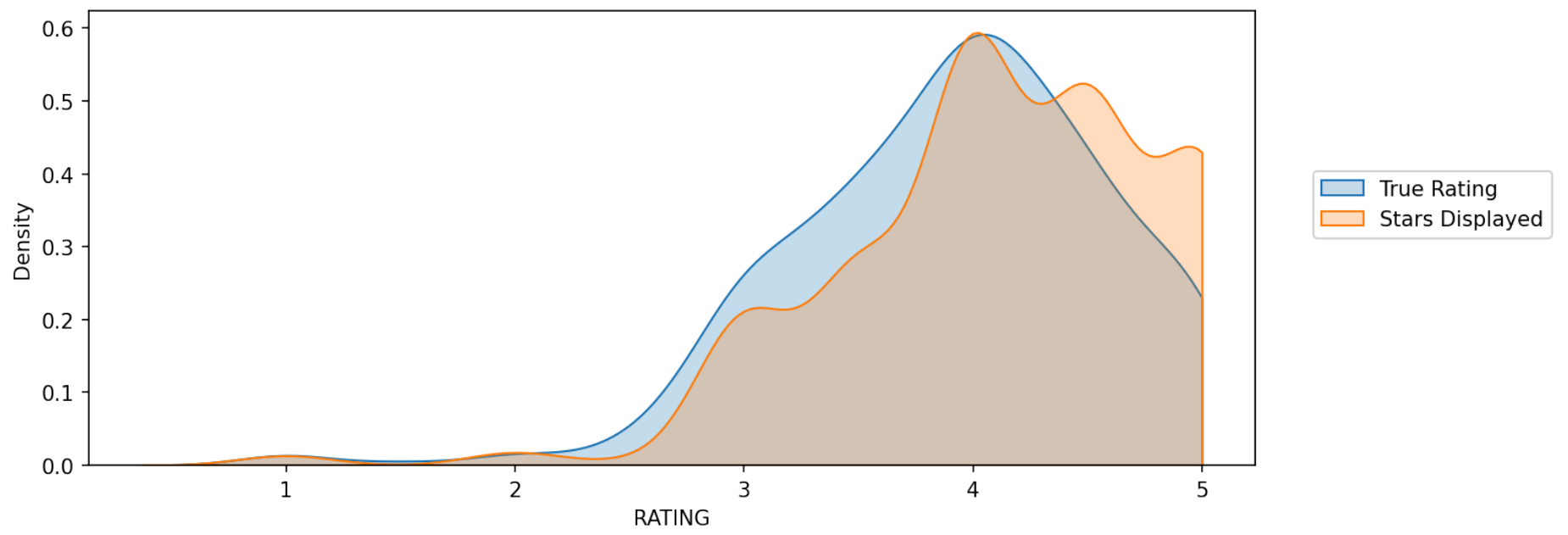
همانطور که در مقاله ذکر شد، به دلیل نمایش HTML و امتیاز ستاره ای، امتیاز واقعی کاربر ممکن است کمی با امتیازی که به کاربر نشان داده می شود متفاوت باشد. بیایید این تفاوت را در توزیع ها تجسم کنیم.

تکلیف: یک نمودار KDE (یا چندین نمودار kdeplot) ایجاد کنید که توزیع امتیازاتی که نمایش داده می شوند (STARS) را در مقابل امتیاز واقعی آراء (RATING) نشان دهد. توزیع ها را بین 0 تا 5 محدود کنید.

In [195... #CODE HERE

In [196...

Out[196... <matplotlib.legend.Legend at 0x1aa0110cdc8>



(۵)

تکلیف: حالا بیایید این اختلاف را به طور واقعی کمیت کنیم. یک ستون جدید از اختلاف بین رتبه بندی نمایش داده شده STARS در مقابل رتبه بندی واقعی RATING ایجاد کنید.

این اختلاف را با STARS-RATING محاسبه کنید و این اختلافات را به نزدیکترین نقطه اعشار گرد کنید زیر بخوانید

In [197... `#CODE HERE`

In [198...

In [199...

Out[199...

| | FILM | STARS | RATING | VOTES | YEAR | STARS_DIFF |
|-----|-----------------------------|-------|--------|-------|------|------------|
| 0 | Fifty Shades of Grey (2015) | 4.0 | 3.9 | 34846 | 2015 | 0.1 |
| 1 | Jurassic World (2015) | 4.5 | 4.5 | 34390 | 2015 | 0.0 |
| 2 | American Sniper (2015) | 5.0 | 4.8 | 34085 | 2015 | 0.2 |
| 3 | Furious 7 (2015) | 5.0 | 4.8 | 33538 | 2015 | 0.2 |
| 4 | Inside Out (2015) | 4.5 | 4.5 | 15749 | 2015 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... |
| 430 | That Sugar Film (2015) | 5.0 | 5.0 | 1 | 2015 | 0.0 |
| 431 | The Intern (2015) | 5.0 | 5.0 | 1 | 2015 | 0.0 |
| 432 | The Park Bench (2015) | 5.0 | 5.0 | 1 | 2015 | 0.0 |
| 433 | The Wanted 18 (2015) | 5.0 | 5.0 | 1 | 2015 | 0.0 |
| 434 | Z For Zachariah (2015) | 5.0 | 5.0 | 1 | 2015 | 0.0 |

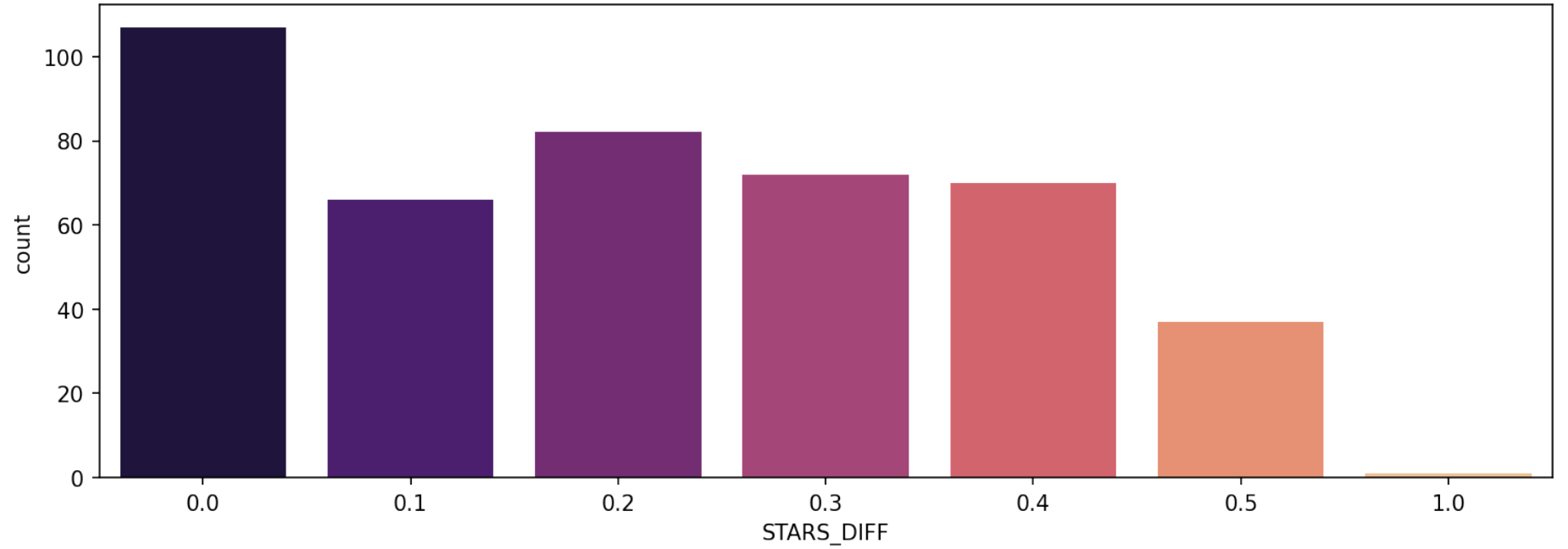
rows × 6 columns 435

تکلیف: یک نمودار تعداد را برای نمایش تعداد دفعاتی که یک اختلاف خاص رخ می دهد ایجاد کنید:

In [200... `#CODE HERE`

In [201...

Out[201... `<AxesSubplot:xlabel='STARS_DIFF', ylabel='count'>`



تکلیف: ما از نمودار می بینیم که یک فیلم امتیاز نمایش داده شده ای داشت که بیش از 1 ستاره با امتیاز واقعی آن اختلاف داشت! کدام فیلم این اختلاف نزدیک به 1 ستاره را داشت؟

```
In [202... #CODE HERE

In [203...

Out[203...
      FILM  STARS  RATING  VOTES  YEAR  STARS_DIFF
381  Turbo Kid (2015)      5.0      4.0      2   2015      1.0
```

بخش سوم: مقایسه امتیازهای فاندانگو با سایر سایت ها

حالا بیااید امتیازهای فاندانگو را با سایر سایت های فیلم مقایسه کنیم و ببینیم چگونه مقایسه می شوند.

تکلیف: فایل "all_sites_scores.csv" را با اجرای سلول زیر بخوانید

```
In [204... all_sites = pd.read_csv("all_sites_scores.csv")

تکلیف: ستون ها، اطلاعات و توضیحات دیتافریم را بررسی کنیلا محاسبه کنید.
```

```
In [ ]:

In [205...

Out[205...
      FILM  RottenTomatoes  RottenTomatoes_User  Metacritic  Metacritic_User  IMDB  Metacritic_user_vote_count  IMDB_user_vote_count
0  Avengers: Age of      74                86        66          7.1      7.8                1330                271107
1  Cinderella (2015)      85                80        67          7.5      7.1                249                65709
2  Ant-Man (2015)        80                90        64          8.1      7.8                627                103660
3  Do You Believe?      18                84        22          4.7      5.4                 31                3136
   (2015)
4  Hot Tub Time          14                28        29          3.4      5.1                 88                19560
   Machine 2 (2015)
```

```
In [206...

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 146 entries, 0 to 145
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
dtypes: float64(2), int64(5), object(1)
memory usage: 9.2+ KB
```

```
In [207...

Out[207...
      RottenTomatoes  RottenTomatoes_User  Metacritic  Metacritic_User  IMDB  Metacritic_user_vote_count  IMDB_user_vote_count
count      146.000000      146.000000    146.000000     146.000000    146.000000      146.000000      146.000000
mean       60.849315      63.876712    58.808219      6.519178     6.736986      185.705479     42846.205479
std        30.168799      20.024430    19.517389      1.510712     0.958736      316.606515     67406.509171
min         5.000000      20.000000    13.000000      2.400000     4.000000       4.000000      243.000000
25%        31.250000      50.000000    43.500000      5.700000     6.300000      33.250000      5627.000000
50%        63.500000      66.500000    59.000000      6.850000     6.900000      72.500000     19103.000000
75%        89.000000      81.000000    75.000000      7.500000     7.400000     168.500000     45185.750000
max       100.000000      94.000000    94.000000      9.600000     8.600000     2375.000000     334164.000000
```

راتن تومیتوز

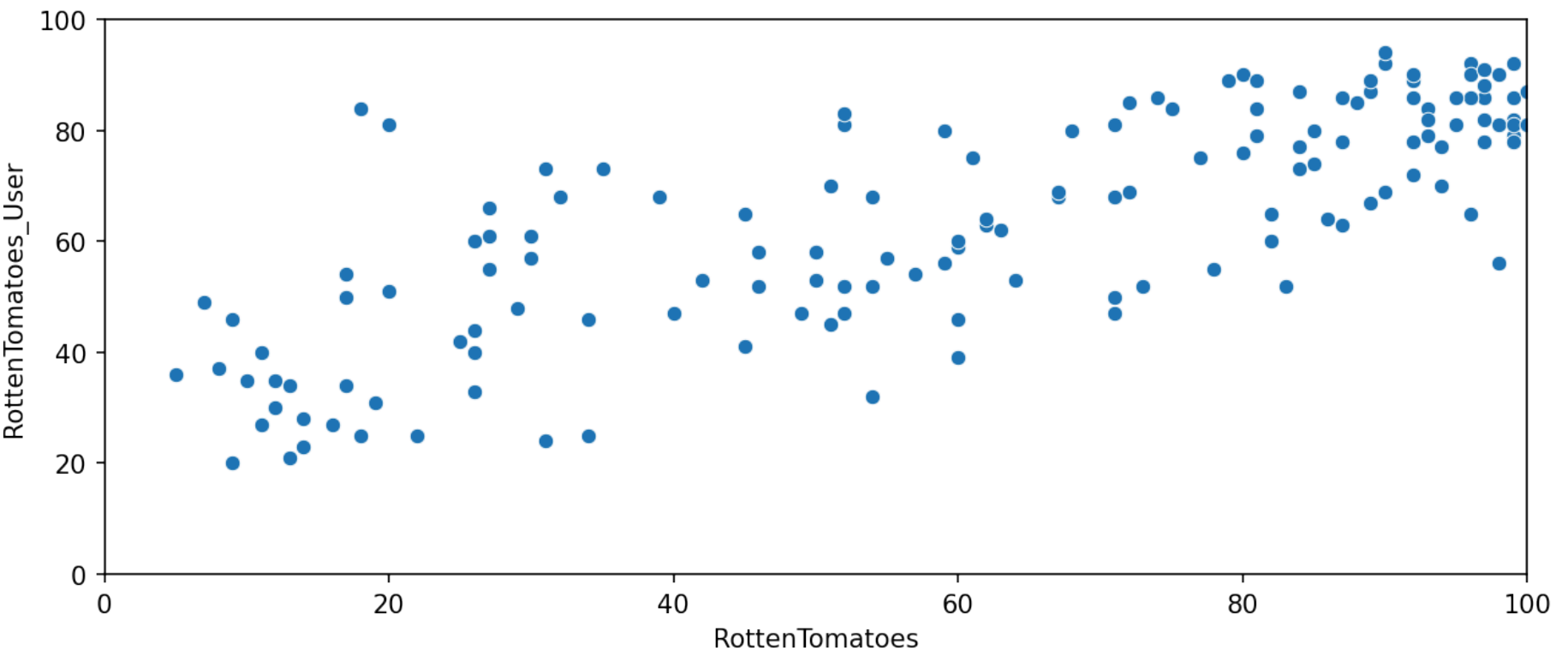
ابتدا نگاهی به راتن تومیتوز بیندازیم. RT دارای دو مجموعه نقد است، نقدهای منتقدان (رتبه بندی هایی که توسط منتقدان رسمی منتشر می شود) و نقدهای کاربران.

تکلیف: یک نمودار پراکنده برای بررسی رابطه بین نقدهای منتقدان RT و نقدهای کاربران RT ایجاد کنید.

```
In [208... # CODE HERE

In [209...

Out[209... (0.0, 100.0)
```



بیاپید این اختلاف را با مقایسه رتبه بندی منتقدان و رتبه بندی کاربران RT کمیت کنیم. این کار را با `RottenTomatoes-RottenTomatoes_User` محاسبه می کنیم. توجه: `Rotten_Diff` در اینجا به معنای منتقدان - امتیاز کاربران است. بنابراین مقادیر نزدیک به 0 به معنای توافق بین منتقدان و کاربران است. مقادیر مثبت بزرگتر به این معنی است که منتقدان بسیار بالاتر از کاربران امتیاز داده اند. مقادیر منفی بزرگتر به این معنی است که کاربران بسیار بالاتر از منتقدان امتیاز داده اند.

تکلیف: یک ستون جدید بر اساس اختلاف بین رتبه بندی منتقدان و رتبه بندی کاربران برای رتن تومیتوز ایجاد کنید. این کار را با `RottenTomatoes-RottenTomatoes_User` محاسبه کنید.

```
In [210... #CODE HERE

In [211...
```

حالا بیاپید میانگین اختلاف کلی را مقایسه کنیم. از آنجایی که ما با اختلافاتی سروکار داریم که می تواند منفی یا مثبت باشد، ابتدا مقدار مطلق همه اختلافات را بگیرید، سپس میانگین را بگیرید. این گزارش به طور متوسط بر اختلاف مطلق بین رتبه بندی منتقدان و رتبه بندی کاربران بازگردد.

تکلیف: میانگین اختلاف مطلق بین امتیازات RT و امتیازات کاربران RT را به شرح بالا محاسبه کنید.

```
In [212... # CODE HERE

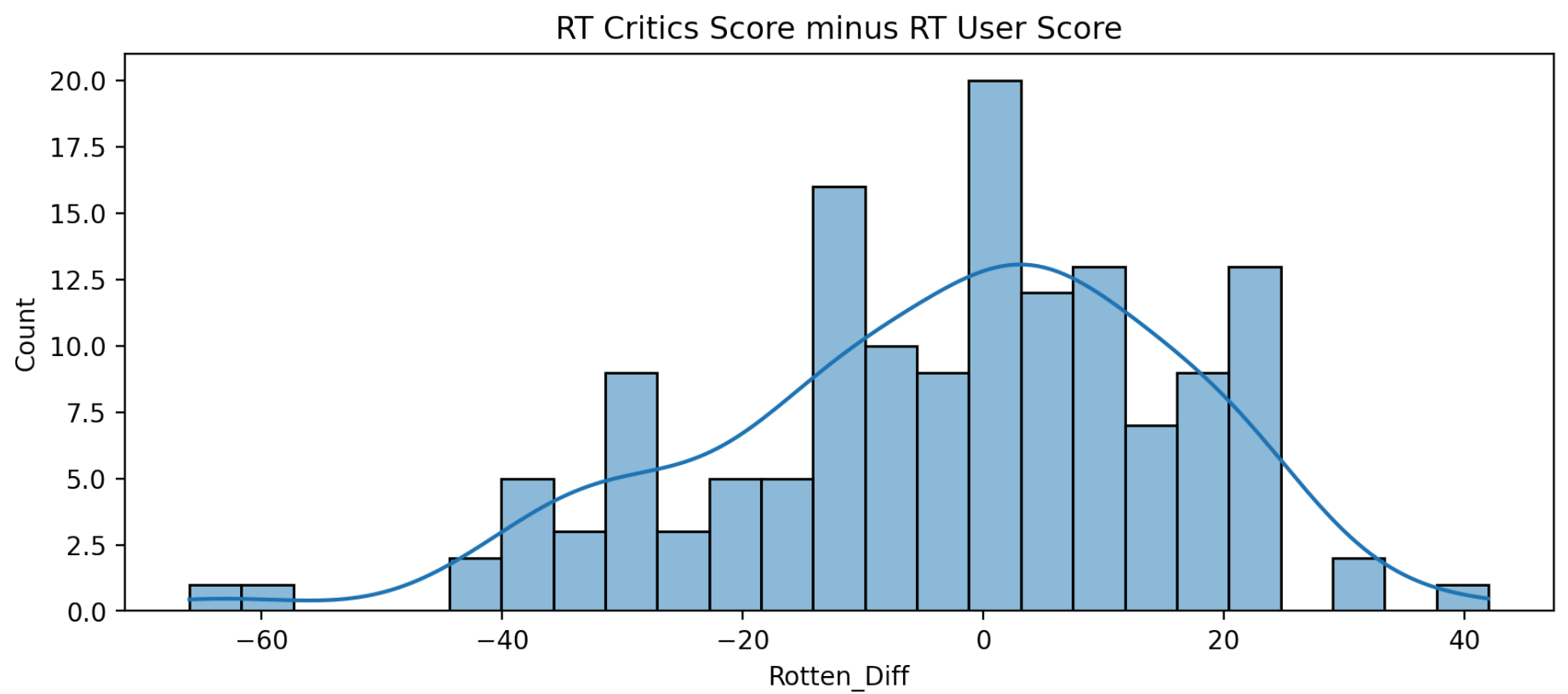
In [213...

Out[213... 15.095890410958905
```

****تکلیف: توزیع اختلافات بین امتیاز منتقدان RT و امتیاز کاربران RT را رسم کنید. در این نمودار توزیع باید مقادیر منفی وجود داشته باشد. برای نمایش این توزیع از نمودار KDE یا هیستوگرام استفاده کنید.**

```
In [214... #CODE HERE

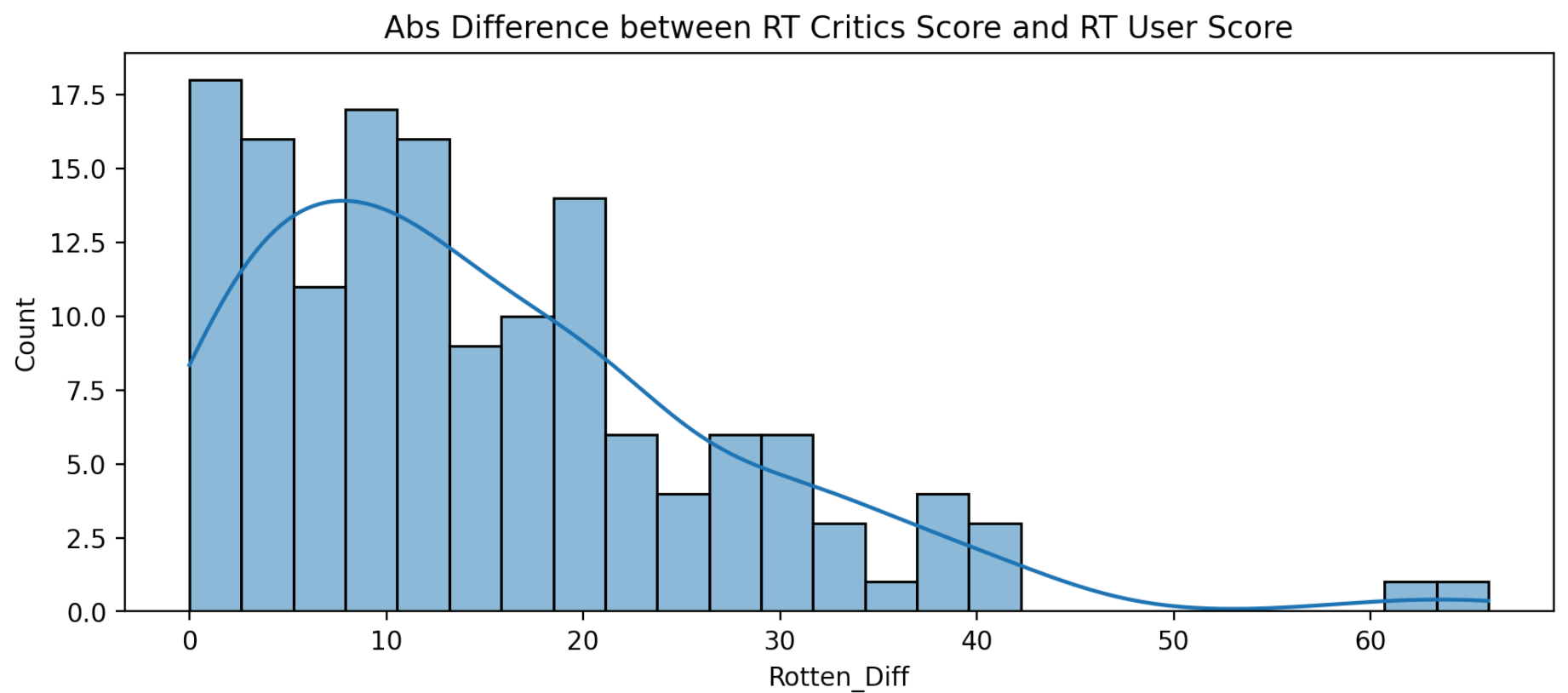
In [215...
```



تکلیف: حالا توزیعی را ایجاد کنید که مقدار مطلق اختلاف بین منتقدان و کاربران در راتن تومیتوز را نشان دهد.کاربران متاکریتیک است؟**

In [216... `#CODE HERE`

In [217...



(

بگذارید فیلم‌هایی را که باعث ایجاد بزرگترین اختلافات می‌شوند را پیدا کنیم. ابتدا، 5 فیلم برتر با بیشترین اختلاف منفی بین کاربران و منتقدان RT را نشان دهید. از آنجایی که ما اختلاف را به صورت امتیاز منتقدان - امتیاز کاربران محاسبه کردیم، مقادیر منفی بزرگ نشان می‌دهد که کاربران به طور متوسط امتیاز بسیار بالاتری به فیلم داده اند تا منتقدان.ند:

تکلیف: 5 فیلم برتر که کاربران به طور متوسط بالاتر از منتقدان امتیاز داده اند کدامند:

In [218... `# CODE HERE`

In [219...

Users Love but Critics Hate

| | FILM | Rotten_Diff |
|-----|---------------------------|-------------|
| 3 | Do You Believe? (2015) | -66 |
| 85 | Little Boy (2015) | -61 |
| 105 | Hitman: Agent 47 (2015) | -42 |
| 134 | The Longest Ride (2015) | -42 |
| 125 | The Wedding Ringer (2015) | -39 |

تکلیف: حالا 5 فیلم برتر را که امتیاز منتقدان به طور متوسط بالاتر از کاربران است، نشان دهید.

```
In [220... # CODE HERE

In [221...

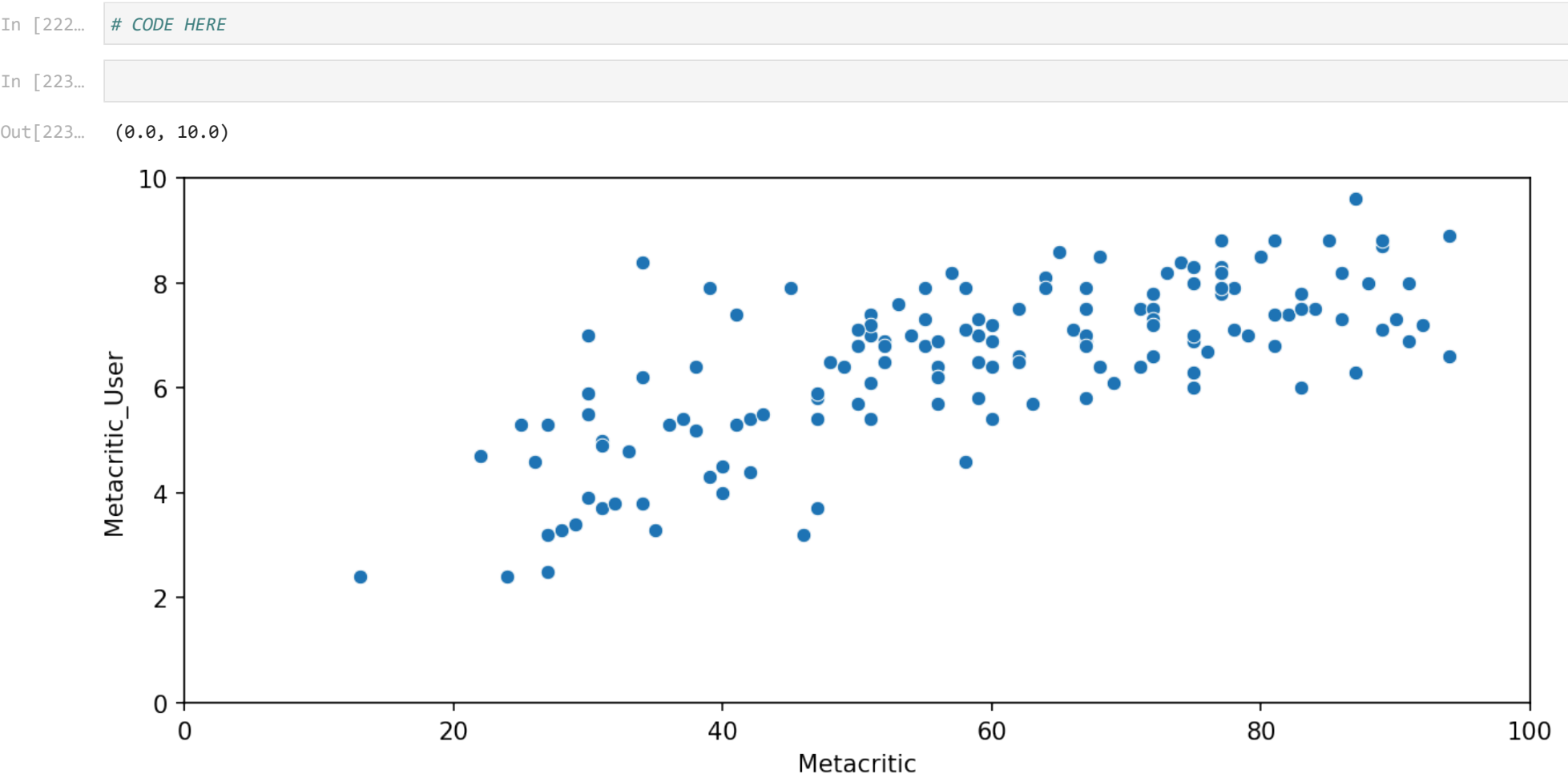
Critics love, but Users Hate

Out[221...
      FILM  Rotten_Diff
69      Mr. Turner (2014)      42
112     It Follows (2015)      31
115    While We're Young (2015)  31
37    Welcome to Me (2015)      24
40    I'll See You In My Dreams (2015)  24
```

متاکریتیک

حالا نگاهی گذرا به رتبه بندی های متاکریتیک بیندازیم. متاکریتیک همچنین امتیاز متوسط کاربران را در مقابل امتیاز رسمی نمایش داده شده آنها نشان می دهد.

تکلیف: یک نمودار پراکنده از امتیاز متاکریتیک در مقابل امتیاز کاربر متاکریتیک نمایش دهید.



IMDB

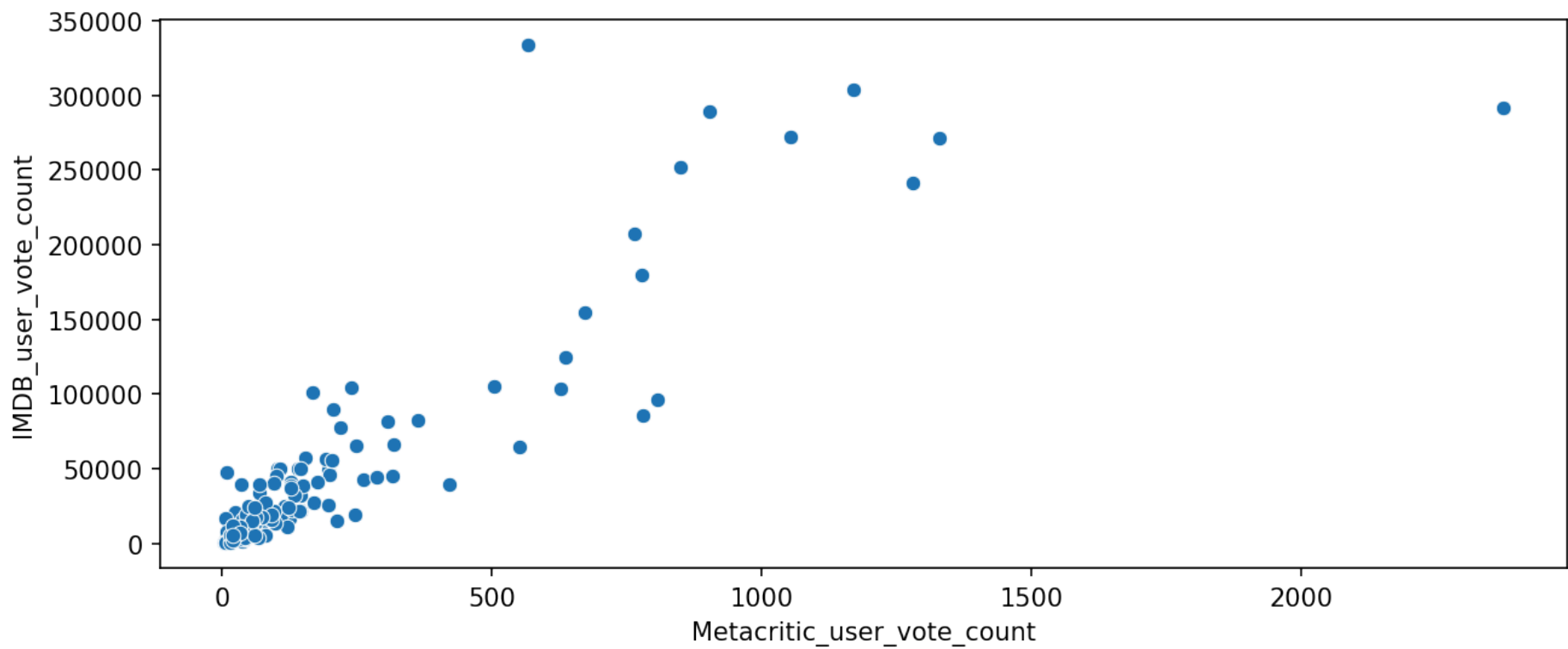
در نهایت بیایید IMDB را بررسی کنیم. توجه داشته باشید که هر دو Metacritic و IMDB تعداد رای ها را گزارش می دهند. بیایید فیلم های محبوب ترین را تحلیل کنیم.

وظیفه: یک نمودار پراکندگی برای رابطه بین تعداد رای ها در MetaCritic در مقابل تعداد رای ها در IMDB ایجاد کنید.

```
In [224... #CODE HERE

In [225...

Out[225... <AxesSubplot:xlabel='Metacritic_user_vote_count', ylabel='IMDB_user_vote_count'>
```



توجه داشته باشید که در اینجا دو مقدار پرت (outlier) وجود دارد. فیلمی با بالاترین تعداد آراء کاربران در IMDB، تنها حدود 500 امتیاز متاکریتیک دارد. این فیلم کدام است؟

وظیفه: کدام فیلم بالاترین تعداد آراء کاربران IMDB را دارد؟

In [226...]

#CODE HERE

In [227...]

Out[227...]

| | FILM | RottenTomatoes | RottenTomatoes_User | Metacritic | Metacritic_User | IMDB | Metacritic_user_vote_count | IMDB_user_vote_count | Rotten_Diff |
|----|---------------------------|----------------|---------------------|------------|-----------------|------|----------------------------|----------------------|-------------|
| 14 | The Imitation Game (2014) | 90 | 92 | 73 | 8.2 | 8.1 | 566 | 334164 | -2 |

وظیفه: کدام فیلم بالاترین تعداد آراء کاربران متاکریتیک را دارد؟

In [228...]

#CODE HERE

In [229...]

Out[229...]

| | FILM | RottenTomatoes | RottenTomatoes_User | Metacritic | Metacritic_User | IMDB | Metacritic_user_vote_count | IMDB_user_vote_count | Rotten_Diff |
|----|---------------------------|----------------|---------------------|------------|-----------------|------|----------------------------|----------------------|-------------|
| 88 | Mad Max: Fury Road (2015) | 97 | 88 | 89 | 8.7 | 8.3 | 2375 | 292023 | 9 |

امتیازدهی فاندانگ در مقابل سایر وبسایت‌ها

در نهایت، بیا ببینیم بررسی کنیم که آیا فاندانگ به طور مصنوعی امتیازهای بالاتری را برای افزایش فروش بلیط نمایش می‌دهد یا خیر.

وظیفه: جدول فاندانگ را با جدول تمام وبسایت‌ها ادغام کنید. از آنجایی که برخی از فیلم‌های فاندانگ نقدهای بسیار کم یا هیچ نقدی ندارند، همه فیلم‌های جدول فاندانگ در

جدول تمام وبسایت‌ها وجود ندارند. ما فقط می‌خواهیم فیلم‌هایی را که در هر دو DataFrame هستند مقایسه کنیم، بنابراین یک الحاق داخلی بر اساس ستون‌های FILM انجام

دهید.

In [230...]

#CODE HERE

In [231...]

In [232...]

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 145 entries, 0 to 144
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   FILM                                145 non-null    object
1   STARS                               145 non-null    float64
2   RATING                              145 non-null    float64
3   VOTES                               145 non-null    int64
4   YEAR                                145 non-null    object
5   RottenTomatoes                      145 non-null    int64
6   RottenTomatoes_User                 145 non-null    int64
7   Metacritic                          145 non-null    int64
8   Metacritic_User                     145 non-null    float64
9   IMDB                                145 non-null    float64
10  Metacritic_user_vote_count          145 non-null    int64
11  IMDB_user_vote_count                145 non-null    int64
12  Rotten_Diff                          145 non-null    int64
dtypes: float64(4), int64(7), object(2)
memory usage: 15.9+ KB
```

In [233...

Out[233...

| | FILM | STARS | RATING | VOTES | YEAR | RottenTomatoes | RottenTomatoes_User | Metacritic | Metacritic_User | IMDB | Metacritic_user_vote_count | IMC |
|---|-----------------------------|-------|--------|-------|------|----------------|---------------------|------------|-----------------|------|----------------------------|-----|
| 0 | Fifty Shades of Grey (2015) | 4.0 | 3.9 | 34846 | 2015 | 25 | 42 | 46 | 3.2 | 4.2 | 778 | |
| 1 | Jurassic World (2015) | 4.5 | 4.5 | 34390 | 2015 | 71 | 81 | 59 | 7.0 | 7.3 | 1281 | |
| 2 | American Sniper (2015) | 5.0 | 4.8 | 34085 | 2015 | 72 | 85 | 72 | 6.6 | 7.4 | 850 | |
| 3 | Furious 7 (2015) | 5.0 | 4.8 | 33538 | 2015 | 81 | 84 | 67 | 6.8 | 7.4 | 764 | |
| 4 | Inside Out (2015) | 4.5 | 4.5 | 15749 | 2015 | 98 | 90 | 94 | 8.9 | 8.6 | 807 | |

نرمال سازی ستون ها به امتیاز و رتبه بندی فاندانگ (۵-۰)

توجه داشته باشید که RT، متاکریتیک و IMDB از امتیازی بین ۵-۰ ستاره مانند فاندانگ استفاده نمی کنند. برای انجام یک مقایسه منصفانه، نیاز به *نرمال سازی* این مقادیر داریم تا همگی بین ۵-۰ ستاره قرار بگیرند و ارتباط بین نقدها ثابت بماند.

وظیفه: ستون های نرمال شده جدیدی برای تمام امتیازات ایجاد کنید تا با محدوده ۵-۰ ستاره ای که در فاندانگ نشان داده شده مطابقت داشته باشد. راه های زیادی برای انجام این کار وجود دارد.

In [234...

```
# CODE HERE
```

In [235...

In [236...

In [237...

In [238...

Out[238...

| | FILM | STARS | RATING | VOTES | YEAR | RottenTomatoes | RottenTomatoes_User | Metacritic | Metacritic_User | IMDB | Metacritic_user_vote_count | IMC |
|---|-----------------------------|-------|--------|-------|------|----------------|---------------------|------------|-----------------|------|----------------------------|-----|
| 0 | Fifty Shades of Grey (2015) | 4.0 | 3.9 | 34846 | 2015 | 25 | 42 | 46 | 3.2 | 4.2 | 778 | |
| 1 | Jurassic World (2015) | 4.5 | 4.5 | 34390 | 2015 | 71 | 81 | 59 | 7.0 | 7.3 | 1281 | |
| 2 | American Sniper (2015) | 5.0 | 4.8 | 34085 | 2015 | 72 | 85 | 72 | 6.6 | 7.4 | 850 | |
| 3 | Furious 7 (2015) | 5.0 | 4.8 | 33538 | 2015 | 81 | 84 | 67 | 6.8 | 7.4 | 764 | |
| 4 | Inside Out (2015) | 4.5 | 4.5 | 15749 | 2015 | 98 | 90 | 94 | 8.9 | 8.6 | 807 | |

وظیفه: اکنون یک DataFrame با نام norm_scores ایجاد کنید که فقط شامل امتیازات نرمال شده باشد. هر دو ستون STARS و RATING از جدول اصلی فاندانگ را نیز در این DataFrame قرار دهید.

In [239...

```
#CODE HERE
```

In [240...

In [241...

Out[241...

| | STARS | RATING | RT_Norm | RTU_Norm | Meta_Norm | Meta_U_Norm | IMDB_Norm |
|---|-------|--------|---------|----------|-----------|-------------|-----------|
| 0 | 4.0 | 3.9 | 1.2 | 2.1 | 2.3 | 1.6 | 2.1 |
| 1 | 4.5 | 4.5 | 3.6 | 4.0 | 3.0 | 3.5 | 3.6 |
| 2 | 5.0 | 4.8 | 3.6 | 4.2 | 3.6 | 3.3 | 3.7 |
| 3 | 5.0 | 4.8 | 4.0 | 4.2 | 3.4 | 3.4 | 3.7 |
| 4 | 4.5 | 4.5 | 4.9 | 4.5 | 4.7 | 4.4 | 4.3 |

مقایسه توزیع امتیازات در سراسر وبسایت‌ها

حالا به سوال اصلی می‌رسیم! آیا فاندانگ امتیازهای غیرعادی بالایی را نشان می‌دهد؟ ما قبلاً می‌دانیم که امتیاز RATING نمایش داده شده را بالاتر از STARS قرار می‌دهد، اما آیا خود امتیازات از میانگین بالاتر هستند؟

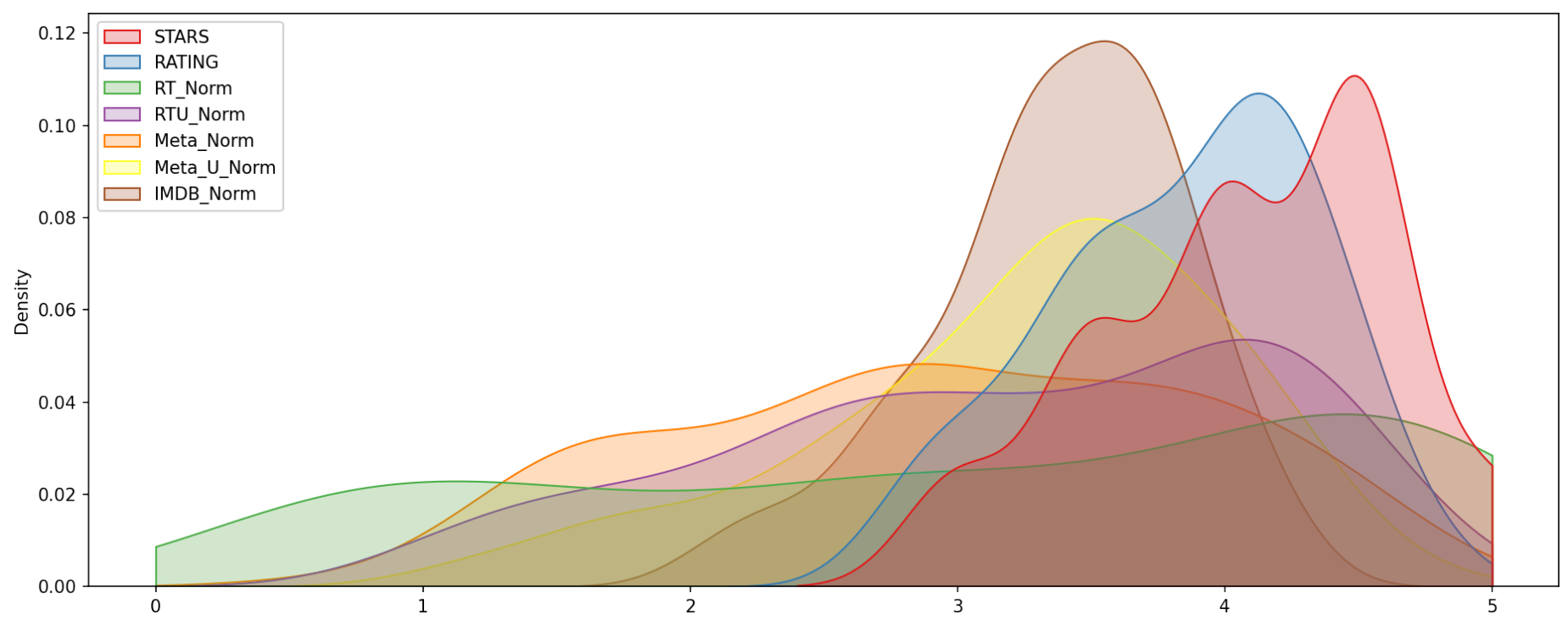
وظیفه: نموداری برای مقایسه توزیع امتیازات نرمال شده در سراسر وبسایت‌ها ایجاد کنید. روش‌های زیادی برای انجام این کار وجود دارد، اما برای پیدا کردن روش‌های ساده و سریع برای نمایش این موضوع، مستندات Seaborn KDEplot را بررسی کنید. نگران نباشید اگر قالب نمودار شما دقیقاً مشابه ما نباشد، مهم این است که تفاوت‌های توزیع به وضوح مشخص باشد.

In [242...

```
#CODE HERE
```

In [243...

In [244...

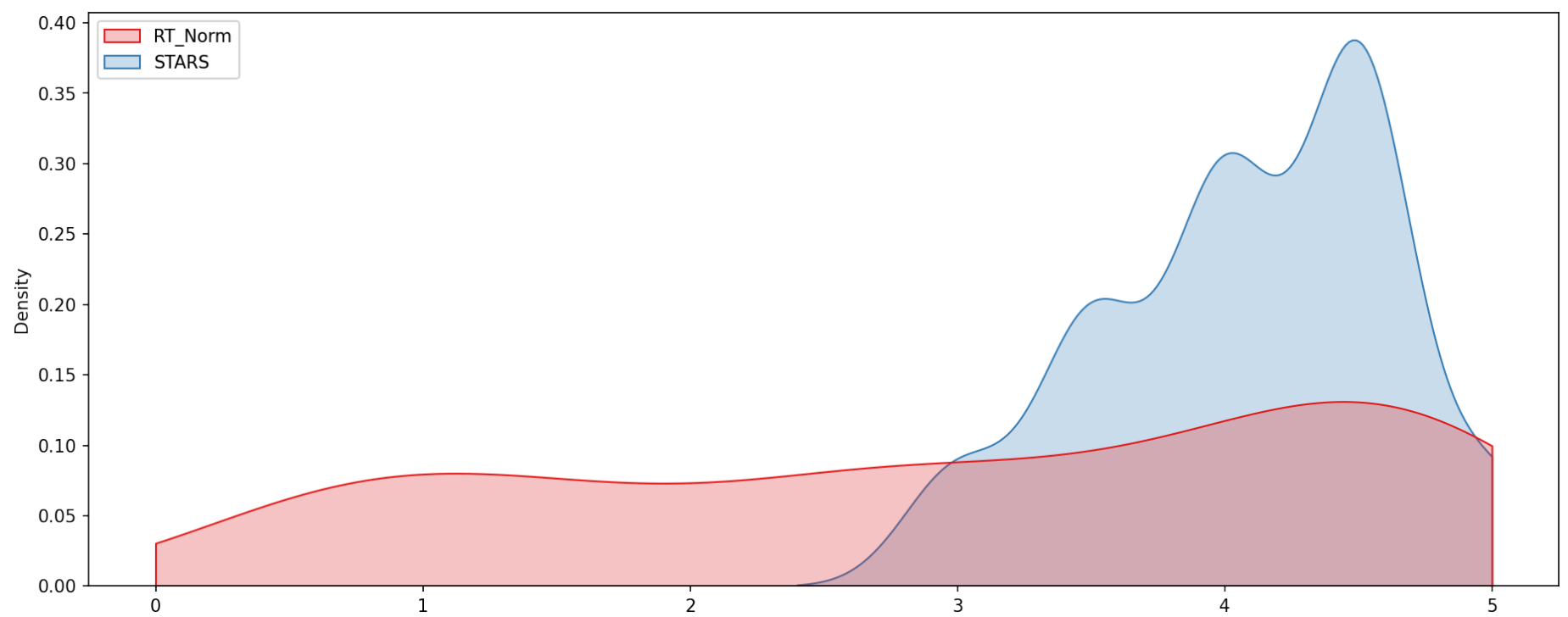


توزیع امتیازات فاندانگ با توزیع امتیازات منتقدان RT به طور واضح، فاندانگ توزیع ناهمگونی دارد. همچنین می توانیم ببینیم که منتقدان RT توزیع یکنواخت تری دارند. بیایید مستقیماً این دو را با هم مقایسه کنیم.

وظیفه: یک نمودار KDE ایجاد کنید که توزیع امتیازات منتقدان RT را با STARS نمایش داده شده توسط فاندانگ مقایسه کند.

In []: `#CODE HERE`

In [167...

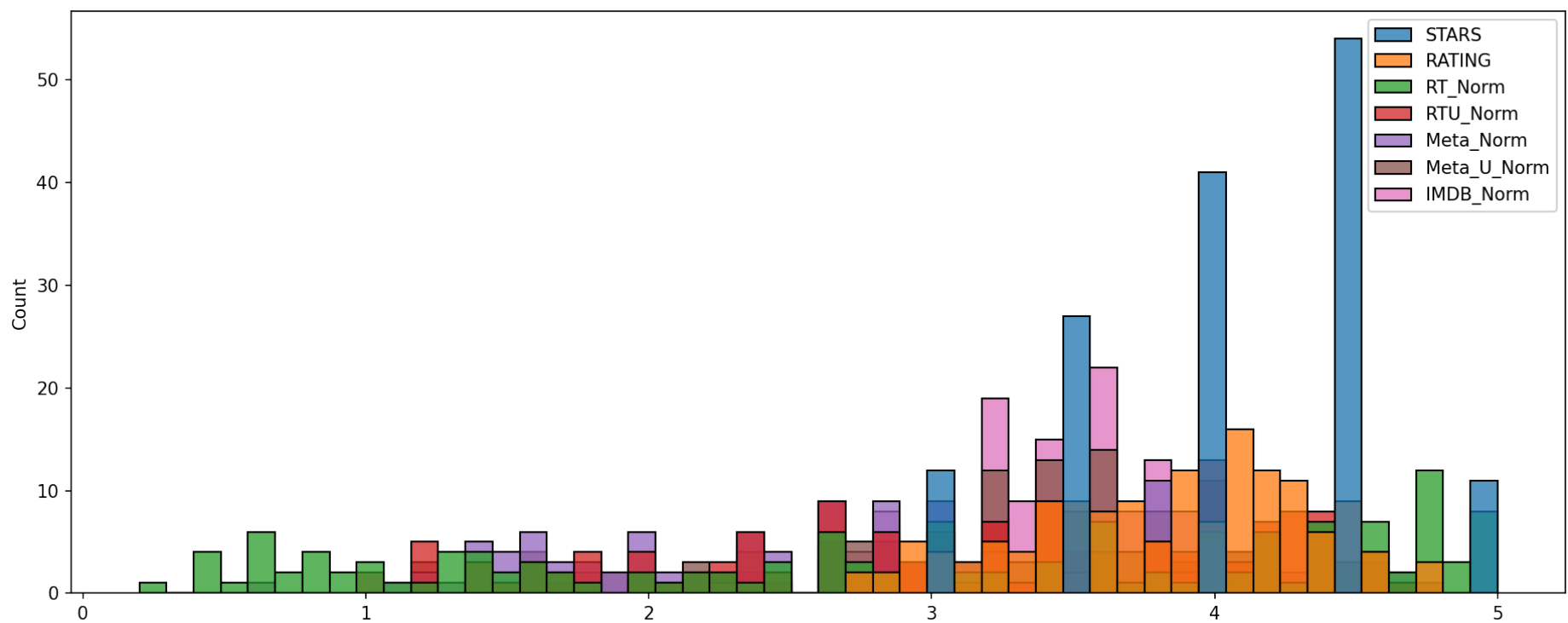


وظیفه اختیاری: یک نمودار هیستوگرام برای مقایسه همه امتیازات نرمال شده ایجاد کنید.

In []: `#CODE HERE`

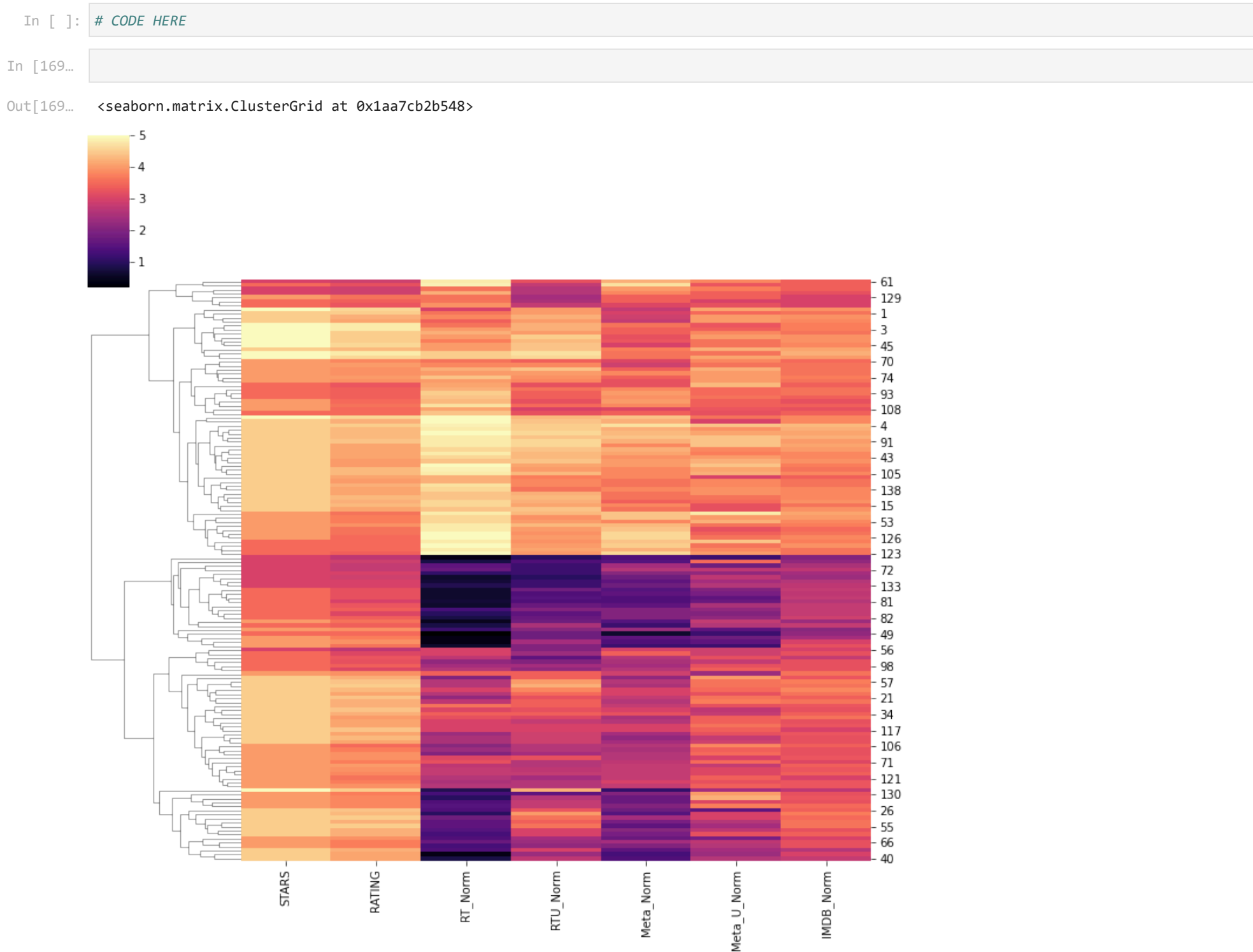
In [168...

Out[168... `<AxesSubplot:ylabel='Count'>`



بدترین فیلم‌ها در تمام پلتفرم‌ها چگونه رتبه‌بندی می‌شوند؟

وظیفه: یک خوشه بندی (clustermap) از تمام امتیازات نرمال شده ایجاد کنید. به تفاوت‌های امتیازدهی توجه داشته باشید، فیلم‌های با امتیاز بالا باید با هم گروه بندی شوند و برعکس فیلم‌های با امتیاز پایین. توجه: این خوشه بندی نیازی به داشتن عنوان فیلم‌ها به عنوان اندیس ندارد، برای خوشه بندی می‌توانید عناوین فیلم‌ها را حذف کنید.



وظیفه: به طور واضح، فاندانگ فیلم‌ها را بسیار بالاتر از سایر وبسایت‌ها امتیازدهی می‌کند، به خصوص با در نظر گرفتن این موضوع که امتیاز گرد شده را نمایش می‌دهد. بیاپید ۱۰ فیلم با بدترین رتبه را بررسی کنیم. بر اساس امتیازات منتقدان Rotten Tomatoes، کدام ۱۰ فیلم پایین‌ترین امتیاز را دارند؟ امتیازات نرمال شده در سراسر پلتفرم‌ها برای این فیلم‌ها چقدر است؟ برای مشاهده نتایج، ممکن است لازم باشد ستون FILM را دوباره به DataFrame امتیازات نرمال شده خود اضافه کنید.

```
In [245...]: # CODE HERE
```

In [246...]

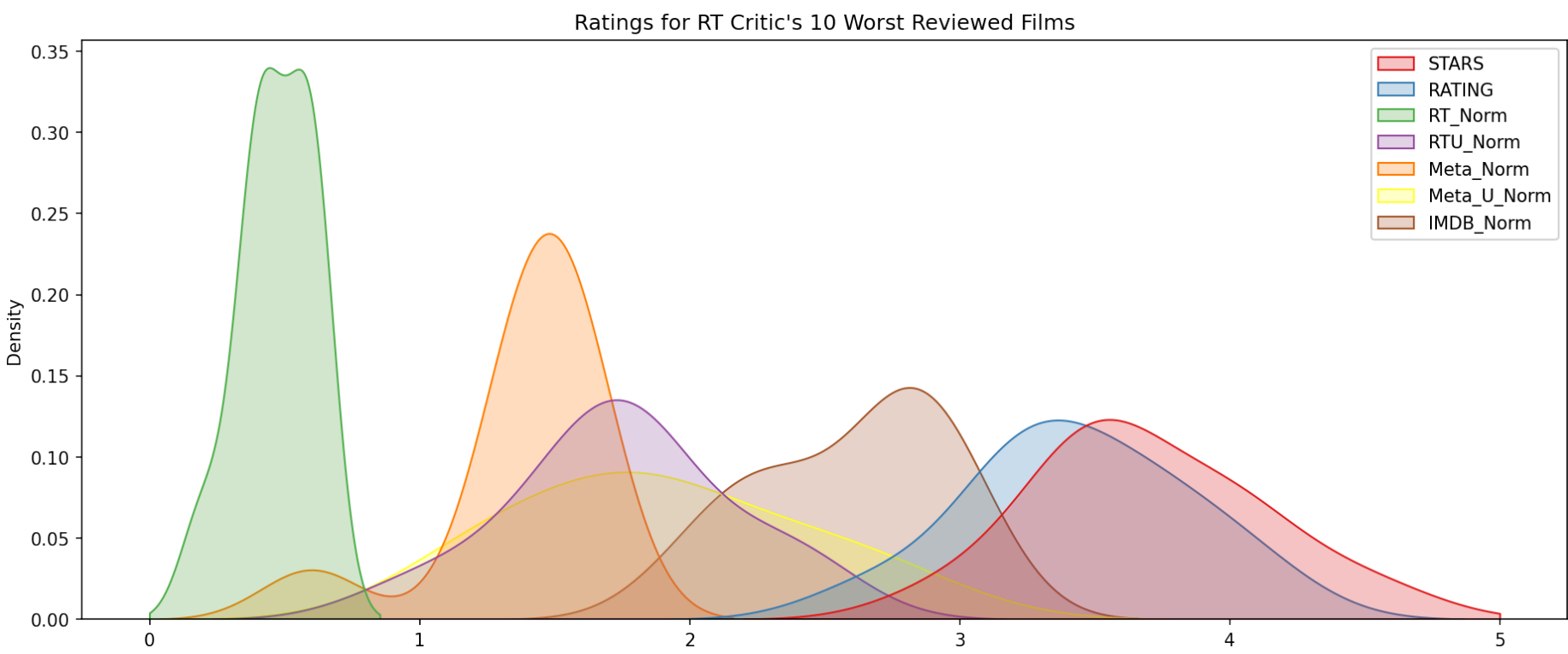
In [248...]

| Out[248...] | STARS | RATING | RT_Norm | RTU_Norm | Meta_Norm | Meta_U_Norm | IMDB_Norm | FILM |
|-------------|-------|--------|---------|----------|-----------|-------------|-----------|-------------------------------|
| 49 | 3.5 | 3.5 | 0.2 | 1.8 | 0.6 | 1.2 | 2.2 | Paul Blart: Mall Cop 2 (2015) |
| 25 | 4.5 | 4.1 | 0.4 | 2.3 | 1.3 | 2.3 | 3.0 | Taken 3 (2015) |
| 28 | 3.0 | 2.7 | 0.4 | 1.0 | 1.4 | 1.2 | 2.0 | Fantastic Four (2015) |
| 54 | 4.0 | 3.7 | 0.4 | 1.8 | 1.6 | 1.8 | 2.4 | Hot Pursuit (2015) |
| 84 | 4.0 | 3.9 | 0.4 | 2.4 | 1.4 | 1.6 | 3.0 | Hitman: Agent 47 (2015) |
| 50 | 4.0 | 3.6 | 0.5 | 1.8 | 1.5 | 2.8 | 2.3 | The Boy Next Door (2015) |
| 77 | 3.5 | 3.2 | 0.6 | 1.8 | 1.5 | 2.0 | 2.8 | Seventh Son (2015) |
| 78 | 3.5 | 3.2 | 0.6 | 1.5 | 1.4 | 1.6 | 2.8 | Mortdecai (2015) |
| 83 | 3.5 | 3.3 | 0.6 | 1.7 | 1.6 | 2.5 | 2.8 | Sinister 2 (2015) |
| 87 | 3.5 | 3.2 | 0.6 | 1.4 | 1.6 | 1.9 | 2.7 | Unfinished Business (2015) |

وظیفه نهایی: توزیع امتیازات را در تمام سایت ها برای 10 فیلم بدترین نمایش دهید.

In []: # CODE HERE

In [251...]



نتیجه گیری نهایی: شگفت انگیز است! فاندانگ برای فیلمهایی که به طور واضح بد هستند، امتیاز ۳-۴ ستاره نشان می‌دهد! به بزرگترین تخلفکار، Taken 3!

دقت کنید. فاندانگ برای فیلمی با میانگین امتیاز ۱.۸۶: <https://www.youtube.com/watch?v=tJrflmRCHJ0>

https://en.wikipedia.org/wiki/Taken_3#Critical_response در سایر پلتفرمها، امتیاز ۴.۵ ستاره را در وبسایت خود نشان می‌دهد!

```
In [253...
Out[253... STARS          4.5
           RATING        4.1
           RT_Norm        0.4
           RTU_Norm       2.3
           Meta_Norm      1.3
           Meta_U_Norm    2.3
           IMDB_Norm      3
           FILM           Taken 3 (2015)
           Name: 25, dtype: object

In [254... 0.4+2.3+1.3+2.3+3
Out[254... 9.3

In [255... 9.3/5
Out[255... 1.86
```
